SIAM J. CONTROL AND OPTIMIZATION
Vol. 29, No. 4, pp. 968–998, July 1991

014

AN EXACT PENALIZATION VIEWPOINT OF CONSTRAINED OPTIMIZATION*

JAMES V. BURKE†

Abstract. In their seminal papers Eremin [*Soviet Mathematics Doklady*, 8 (1966), pp. 459–462] and Zangwill [*Management Science*, 13 (1967), pp. 344–358] introduce a notion of exact penalization for use in the development of algorithms for constrained optimization. Since that time, exact penalty functions have continued to play a key role in the theory of mathematical programming. In the present paper, this theory is unified by showing how the Eremin–Zangwill exact penalty functions can be used to develop the foundations of the theory of constrained optimization for finite dimensions in an elementary and straightforward way. Regularity conditions, multiplier rules, second-order optimality conditions, and convex programming are all given interpretations relative to the Eremin–Zangwill exact penalty functions. In conclusion, a historical review of those results associated with the existence of an exact penalty parameter is provided.

Key words. exact penalty functions, calmness, constraint qualification, optimality conditions, convex programming

AMS(MOS) subject classifications. 49A42, 49D30, 49D37, 90D30

1. Introduction. In their seminal papers Eremin [23] and Zangwill [75] introduced a notion of exact penalization for use in the development of algorithms for nonlinear constrained optimization. This notion of exact penalization is the natural extension of the so-called big-M method of linear programming (see Charnes, Cooper, and Henderson [14, §4] for the earliest reference known to us) to nonlinear programming. Since that time, exact penalty functions have continued to play a key role in the theory of mathematical programming. Within the algorithmic sphere, the history of these functions is quite rich, even though their use has been, and still is, a topic of controversy. The root of this controversy is the nondifferentiable nature of these functions. From an algorithmic viewpoint, this nondifferentiability can induce the so-called Maratos effect (a phenomenon that prevents rapid local convergence). A great deal of effort has been devoted to overcoming this difficulty, leading to the development of the so-called watchdog technique [12] and second-order correction techniques [19], [28], [26], [29], and others. Other authors, in an effort to avoid the problems associated with nondifferentiability, have introduced entirely different classes of exact penalty functions that are differentiable [5], [30], [34], [60], and [69]. The research in this area continues at a rapid pace and the controversies over the use of nondifferentiable exact penalty functions in algorithms are far from nearing resolution. This paper can, in many ways, be viewed as a contribution to this discussion. However, our approach is from a rather different perspective. We do not discuss algorithms at all, rather we demonstrate how the Eremin–Zangwill exact penalty functions can be used to develop the foundations of the theory of constrained optimization in an elementary and straightforward way. In doing so, we show how all of the fundamental notions and results in constrained optimization can be derived from the Eremin–Zangwill exact penalty functions, from regularity conditions such as calmness [15], [66], to the

*Received by the editors September 16, 1986; accepted for publication (in revised form) June 1, 1990.

†Mathematics Department, GN-50, University of Washington, Seattle, Washington 98195. The work of the author was supported in part by National Science Foundation grant DMS–8803206 and by Air Force Office of Scientific Research grant AFOSR–860080.

existence of Lagrange multipliers, to second-order necessary and sufficient conditions for optimality. The derivation of these results by means of the Eremin–Zangwill exact penalty functions is by no means strained or artificial, quite the contrary, the proofs are often simplified at the expense of obtaining a more powerful result. Thus, our goal in this endeavor is not to demonstrate the viability of these penalty functions for use in algorithmic development, but rather to demonstrate their role vis-á-vis the foundations of the theory and to provide an interpretation for many of the familiar objects in this theory in terms of the corresponding objects associated with these penalty functions. Hopefully, one consequence of these investigations is that the practical significance of these penalty functions can be more accurately assessed.

We begin §2 by reviewing some of the fundamental results and concepts associated with constrained optimization. We discuss calmness, regularity, constraint qualifications, and their relationships vis-á-vis exact penalization. This section contains all of the first-order results related to the existence of Kuhn–Tucker [43] multipliers. In §3 we show how exact penalization techniques can be used to derive a multiplier theorem in the absence of a constraint qualification. This multiplier rule is reminiscent of the one given by John [42]. Second-order results are obtained in §4. The case of convex programming is studied in §5, and in §6 we provide a historical review of the literature on the existence of a finite exact penalty parameter. The approach to the theory of constrained optimization from the viewpoint of exact penalization is also the theme of Fletcher [29, §14.3], Garcia-Palomares [31], and Rockafellar [64]. A very nice survey of exact penalization techniques in general is given by Fletcher [27]. The present paper is based on Burke [9], wherein several further results and generalizations are obtained.

The notation that we employ is for the most part standard; however, a partial list is provided for the reader's convenience. Let X be a real normal linear space and let X^* be its topological dual. The spaces X and X^* are paired in duality by the continuous bilinear form

$$\langle x^*, x\rangle := x^*(x)$$

defined on $X^* \times X$. Given $x_1, x_2 \in X$ the line segment joining them is denoted by

$$[x_1, x_2] := \{\lambda x_1 + (1-\lambda)x_2 : \lambda \in [0,1]\}.$$

Let C be a subset of X. Then $\mathrm{cl}(C)$ is the closure of C, $\mathrm{int}(C)$ is the interior of C, and $\mathrm{ri}(C)$ is the interior of C relative to its affine hull, i.e., the smallest closed affine set containing C. The core of C, denoted $\mathrm{core}(C)$, is the set of all point $z \in C$ such that every line through z contains a line segment $[z_1, z_2]$ with $z \in [z_1, z_2] \subset C$ and $z_1 \neq z \neq z_2$. In finite dimensions, we have $\mathrm{core}(C) = \mathrm{int}(C)$. The *polar* of C is given by

$$C^0 := \{x^* \in X^* : \langle x^*, x\rangle \leq 1 \text{ for all } x \in C\}$$

and the *positive conjugate* of C is $C^* := -C^0$. The *recession cone* of C is

$$\mathrm{rec}(C) := \{y \in X : C + y \subset cl(C)\}$$

and the cone generated by C is

$$\mathrm{cone}(C) := \cup_{\lambda \geq 0} \lambda C,$$

where for any two subsets S_1 and S_2 of X and any two scalars $\alpha, \beta \in \mathbb{R}$ we have

$$\alpha S_1 + \beta S_2 := \{\alpha s_1 + \beta s_2 : s_1 \in S_1, s_2 \in S_2\}.$$

The *support* and *indicator* functions for C are given, respectively, as

$$\psi^*(x^*|C) := \sup\{\langle x^*, y\rangle : y \in C\}$$

and

$$\psi(x|C) := \begin{cases} 0, & \text{if } x \in C \\ +\infty, & \text{otherwise.} \end{cases}$$

The *barrier cone* of C is

$$\text{bar}(C) := \{x^* \in X^* : \psi^*(x^*|C) < \infty\}$$

and the relation

$$\text{rec}(C) = [\text{bar}(C)]^0$$

holds if X is reflexive.

A *multifunction* T mapping X into Y where Y is another real normal linear space, written $T : X \rightrightarrows Y$, is a mapping of X whose values are subsets of Y. The *domain* of T is the set $\text{dom}(T) := \{x \in X : T(x) \neq \emptyset\}$ and the *graph* of T is $\text{graph}\,(T) := \{(x, y) | y \in T(x)\}$. T is said to be *upper semicontinuous* if graph (T) is closed in $X \times Y$ under the product topology. The space $\mathcal{L}(X, Y)$ is the space of continuous linear maps from X to Y. Given $T \in \mathcal{L}(X, Y)$ we write

$$\text{ran}(T) := \{y \in Y : \exists x \in X \text{ with } y = Tx\}$$

and

$$\ker(T) := \{x \in X : Tx = 0\}.$$

If X and Y are finite-dimensional, then, with respect to fixed bases for X and Y, one can identify $\mathcal{L}(X, Y)$ with $\mathbb{R}^{m \times n}$ the set of $m \times n$ matrices, where $\dim(X) = n$ and $\dim(Y) = m$. The *adjoint* of $A \in \mathcal{L}(X, Y)$ is the uniquely defined mapping $A^* \in \mathcal{L}(Y^*, X^*)$ for which

$$\langle A^* y^*, x\rangle = \langle y^*, Ax\rangle$$

for all $(y^*, x) \in Y^* \times X$. In finite dimensions we have $A^* = A^T$.

Let $f : X \to \overline{\mathbb{R}}$ where $\overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$, we write

$$\begin{aligned} \text{dom}(f) &:= \{x \in X : f(x) < +\infty\}, \\ \text{lev}_f(x) &:= \{y \in X : f(y) \leq f(x)\}, \text{ and} \\ \text{epi}(f) &:= \{(\mu, x) : f(x) \leq \mu\}. \end{aligned}$$

We say that f is *lower semicontinuous* if epi(f) is a closed set. If f is Lipschitz near a point $x \in X$, then the *Clarke generalized directional derivative*,

$$f^0(x; d) := \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{f(y + tv) - f(y)}{t},$$

exists at x for every $d \in X$.

The norm on X is denoted $\|\cdot\|$ and its unit ball is $\mathbb{B} := \{x : \|x\| \leq 1\}$. The dual norm is given by $\|x^*\|_0 := \psi^*(x^*|\mathbb{B})$ and its unit ball is $\mathbb{B}^0$. The distance function for a set $C \subset X$ is given by

$$\text{dist}(y|C) := \inf\{\|y - x\| : x \in C\}.$$

For $C \subset X^*$, the dual distance function is denoted

$$\text{dist}_0(y|C) := \inf\{\|y - x\|_0 : x \in C\}.$$

In finite dimensions the Euclidean norm plays a special role and is denoted by $\|\cdot\|_2$ with corresponding distance function $\text{dist}_2(\cdot|C)$. The distance function for a set $C \subset X$ is Lipschitz with Lipschitz constant 1, and so its Clarke generalized directional derivative exists at every point in all directions. Based on this observation we define the *tangent cone* to a point $x \in C$ by

$$T(x|C) := \{d \in X : \text{dist}(\cdot|C)^0(x; d) = 0\}$$

with the *normal cone* defined via polarity

$$N(x|C) := T(x|C)^0.$$

For convex sets, these objects reduce to the usual notions of tangent and normal cone. In finite dimensions one can also define the *limiting proximal normal cone* at a point $x \in C$ by $\widehat{N}(x|C) := \{\lambda \lim v_i/\|v_i\| : \lambda \geq 0, v_i \perp C \text{ at } x_i \to x,\ v_i \to 0\}$, where one writes $v \perp C$ at y to mean that $y \in \text{cl}(C)$ and $v = y^1 - y$ with $\|y^1 - y\|_2 = \text{dist}_2(y|C)$. One has that $N(x|C)$ is the closed convex hull of $\widehat{N}(x|C)$.

Given $f : X \to \overline{\mathbb{R}}$ the *generalized subdifferential* of f at $x \in \text{dom}(f)$ is given by

$$\partial f(x) := \{x^* \in X^* : (-1, x^*) \in N((f(x), x)|\text{epi}(f))\},$$

the *asymptotic subdifferential* is

$$\partial^\infty f(x) := \{x^* \in X^* : (0, x^*) \in N((f(x), x)|\text{epi}(f))\},$$

the *limiting proximal subdifferential* is

$$\widehat{\partial} f(x) := \{x^* \in X^* : (-1, x^*) \in \widehat{N}((f(x), x)|\text{epi}(f))\},$$

and the *asymptotic limiting proximal subdifferential* is

$$\widehat{\partial}^\infty f(x) := \{x^* \in X : (0, x^*) \in \widehat{N}((f(x), x)|\text{epi}(f))\}.$$

Clearly, $\partial^\infty f(x) = \text{rec}(\partial f(x))$ whenever $\partial f(x) \neq \emptyset$. The *generalized directional derivative* of f is then defined to be

$$f^0(x; v) := \psi^*(v|\partial f(x))$$

with $f^0(x; v) := -\infty$ if $\partial f(x) = \emptyset$. This notation is consistent with that of the Clarke subdifferential for locally Lipschitz functions.

The function f is said to be *subdifferentially regular* at a point $x \in \text{dom}(f)$ if

$$\liminf_{\substack{u \to v \\ t \downarrow 0}} \frac{f(x + tu) - f(x)}{t} = f^0(x; v)$$

for all $v \in X$, in which case

$$f^0(x; v) = f'(x; v) := \lim_{t \downarrow 0} \frac{f(x + tv) - f(x)}{t}.$$

A function $F : X \to Y$ has *Frechet derivative* $F'(x) \in \mathcal{L}(X, Y)$ at $x \in X$ if

$$F(y) = F(x) + F'(x)(y - x) + o(\|y - x\|),$$

where $\lim_{y \to x} o(\|y - x\|)/\|y - x\| = 0$. The mapping F is *strictly differentiable* at $x \in X$ if there exists $F'_s(x) \subset \mathcal{L}(X, Y)$ such that

$$\lim_{\substack{x' \to x \\ t \downarrow 0}} \frac{F(x' + tv) - F(x')}{t} = F'_s(x)v$$

for all v in X. If $f : X \to \overline{R}$ is strictly differentiable at a point $x \in \text{dom}(f)$, then $\partial f(x) = \{f'_s(x)\}$.

If both X and Y are finite-dimensional and $F : X \to Y$ is locally Lipschitz, then F is almost everywhere differentiable in the sense of Lebesgue measure. The *generalized Jacobian* of F at a point $x \in X$, denoted $\partial F(x)$, is the convex hull of all operators in $\mathcal{L}(X, Y)$ obtained as the limit of sequences of the form $\{F'(x_i)\}$ where $x_i \to x$ and $F'(x_i)$ exists at each x_i. Again, if F is strictly differentiable at x, then $\partial F(x) := \{F'_s(x)\}$.

Let $f : X \to \overline{\mathbb{R}}$ and $C \subset X$. We write

$$\arg\min\{f(x) : x \in C\} := \{x \in C : f(x) = \min\{f(x) : x \in C\}\}$$

and define $\arg\max\{f : x \in C\}$ similarly. A local minimum of *radius* ε for the problem $\min\{f(x) : x \in C\}$ is any point $x \in C$ such that $f(x) \le f(y)$ for all $y \in C \cap (x + \varepsilon\mathbb{B})$.

For more information about the objects defined above see [15]–[17], [54], and [65]–[68].

2. The fundamentals: calmness, regularity, and exact penalization. Let X and Y be normal linear spaces and consider the problem

$$(\mathcal{P}) \qquad \begin{array}{l} \text{minimize } f(x) \\ \text{subject to } g(x) \in C, \end{array}$$

where $f : X \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$, $g : X \to Y$, and C is a closed subset of Y. We begin with a discussion of regularity conditions that allow the development of general multiplier rules for $\mathcal{P}$. One of the weakest such conditions was proposed by Rockafellar and is known as calmness.

DEFINITION 2.1. Let f, g, X, Y, and C be as in the statement of $\mathcal{P}$ and consider the perturbed problems

$$(\mathcal{P}_u) \qquad \begin{array}{l} \text{minimize } f(x) \\ \text{subject to } g(x) \in C + u. \end{array}$$

Let $\overline{x} \in X$ and $\overline{u} \in Y$ be such that $g(\overline{x}) \in C + \overline{u}$ and $\overline{x} \in \text{dom}(f) := \{x \in X : f(x) < +\infty\}$. The problem $\mathcal{P}_{\overline{u}}$ is said to be *calm* at $\overline{x}$ if there are constants $\overline{\alpha} \ge 0$ and $\varepsilon > 0$ such that for every pair $(x, u) \in X \times Y$ with $\|x - \overline{x}\| \le \varepsilon$ and $g(x) \in C + u$ we have

$$f(x) + \overline{\alpha}\|u - \overline{u}\| \ge f(\overline{x}). \tag{2.1}$$

The constants $\overline{\alpha}$ and ε are called the *modulus* and *radius* of calmness for $\mathcal{P}_{\overline{u}}$ at $\overline{x}$, respectively.

The family of perturbed problems $\mathcal{P}_u$ is said to be calm at $\overline{u}$ if

$$\liminf_{u \to \overline{u}} \frac{V(u) - V(\overline{u})}{\|u - \overline{u}\|} > -\infty, \tag{2.2}$$

where

$$V(u) := \begin{cases} +\infty, & \text{if } \{x : g(x) \in C + u\} = \emptyset \\ \min\{f(x) : g(x) \in C + u\}, & \text{otherwise} \end{cases}$$

is the value function for the family $\mathcal{P}_u$.

Remarks. (1) This definition for $\mathcal{P}_{\overline{u}}$ to be calm at $\overline{x}$ varies from the definition that is usually given (eg., see Clarke [15, Def. 6.4.1]); however, in Burke [8, §2], it is shown that they are equivalent when g is continuous at $\overline{x}$.

(2) Observe that if $\mathcal{P}_{\overline{u}}$ is calm at $\overline{x}$, then $\overline{x}$ is necessarily a local solution to $\mathcal{P}_{\overline{u}}$, and if $\mathcal{P}_u$ is calm at $\overline{u}$, then for any solution $\overline{x}$ to $\mathcal{P}_{\overline{u}}, \mathcal{P}_{\overline{u}}$ is calm at $\overline{x}$.

(3) The notion of calmness is closely related to the notion of a Φ_1-subdifferential introduced in Dolecki and Rolewicz [20].

The calmness hypothesis is quite weak and in many situations is easily verified. In finite dimensions, calmness holds on a dense subset of the perturbations.

PROPOSITION 2.1. (1) (Clarke [15, Prop. 6.4.5]) *Suppose that* $Y := \mathbb{R}^m, C := \mathbb{R}^m_-$, *and* $f := f_0 + \psi(\cdot|S)$ *with* $S \subset X$ *nonempty and closed, and* $f_0 : X \to \mathbb{R}$ *and* $g : X \to \mathbb{R}^m$ *locally Lipschitzian. If* $V(u)$ *is finite for all* u *near* 0, *then for almost all* u *in a neighborhood of the origin the problem* $\mathcal{P}_u$ *is calm.*

(2) (Burke [8, Prop. 3.1]) *Suppose that* Y *is finite-dimensional,* f *is lower semicontinuous, and* g *is continuous. If* $\overline{u} \in Y$ *and* $\gamma > 0$ *are such that* V *is bounded on* $\overline{u} + \gamma\mathbb{B}$, *then* $\mathcal{P}_u$ *is calm on a dense subset of* $\overline{u} + \gamma\mathbb{B}$.

From (2.2) it is clear that calmness is a weak variational property of the value function V. A condition of this type is always required for establishing the existence of multipliers. It is remarkable that the notion of calmness at a solution to $\mathcal{P}_{\overline{u}}$ is equivalent to the existence of a finite exact penalty parameter.

THEOREM 2.1 (Burke [8, Thm. 1.1]). *Let* $\overline{x} \in X$ *and* $\overline{u} \in Y$ *be such that*

$$g(\overline{x}) \in C + \overline{u} \quad \text{and} \quad \overline{x} \in \operatorname{dom}(f).$$

Then $\mathcal{P}_{\overline{u}}$ *is calm at* $\overline{x}$ *with modulus* $\overline{\alpha} \geq 0$ *and radius* $\varepsilon > 0$ *if and only if* $\overline{x}$ *is a local minimum of radius* ε *for*

$$P_{\overline{u},\alpha}(x) := f(x) + \alpha \operatorname{dist}(g(x)|C + \overline{u})$$

for all $\alpha \geq \overline{\alpha}$, *that is,*

$$P_{\overline{u},\alpha}(\overline{x}) \leq P_{\overline{u},\alpha}(x)$$

for all $x \in \overline{x} + \varepsilon\mathbb{B}$ *and* $\alpha \geq \overline{\alpha}$.

Remark. The fact that calmness implies the existence of an exact penalty parameter is also established in Clarke [15] and Dolecki and Rolewicz [20]. However, the reverse implication and the precision of this correspondence is first established in [8].

Thus, at this early juncture we see that the Eremin–Zangwill exact penalty functions play a fundamental role in the theory. Under the calmness hypothesis we can obtain multiplier rules for $\mathcal{P}$ by first invoking Theorem 2.1 and then applying the pertinent calculus rules of an appropriate subdifferential (e.g., the Clarke subdifferential [15]–[17], the Michel–Penot subdifferential [53], the limiting proximal subdifferential [65]–[66], etc.)

We present two sample results based on the subdifferential calculus developed in Clarke [15] and Rockafellar [66], [68].

THEOREM 2.2. (1) *Suppose $\mathcal{P}$ is calm at $\overline{x} \in X$, g is strictly differentiable at $\overline{x}$ with strict derivative $g'_s(\overline{x})$, and $\partial f(\overline{x}) \neq \emptyset$. Then there is a $y \in N(g(\overline{x})|C)$ such that*

$$0 \in \partial f(\overline{x}) + g'_s(\overline{x})^* y.$$

(2) *If X and Y are finite-dimensional, $\mathcal{P}$ is calm at $\overline{x} \in X$, f is lower semicontinuous near $\overline{x}$, and g is Lipschitzian near $\overline{x}$, then there exists $y \in N(g(\overline{x})|C)$ such that*

$$0 \in \partial f(\overline{x}) + \partial g(\overline{x})^* y.$$

Proof. (1) By Theorem 2.1, $\overline{x}$ is a local minimum for $P_\alpha(x) := f(x) + \alpha \operatorname{dist}(g(x)|C)$ for all α sufficiently large. Hence $0 \in \partial P_\alpha(\overline{x})$ for all $\alpha \geq \overline{\alpha}$ for some $\overline{\alpha} \geq 0$. By [68, Cor. 2],

$$\partial P_\alpha(\overline{x}) \subset \partial f(\overline{x}) + \alpha \partial \,[\operatorname{dist}(g(\cdot)|C)](\overline{x}).$$

From [15, Prop. 2.4.2],

$$N(g(\overline{x})|C) = cl[\cup_{\lambda \geq 0} \lambda \partial \,[\operatorname{dist}(\cdot|C)](g(\overline{x}))]. \tag{2.3}$$

Consequently, by the chain rule [15, Thm. 2.3.10],

$$0 \in \partial f(\overline{x}) + g'_s(\overline{x})^* N(g(\overline{x})|C),$$

from which the result follows.

(2) This is an immediate consequence of Rockafellar [66, Cor. 5.2.3] and inclusion (2.3). ☐

Remarks. (1) To incorporate an abstract constraint of the form $x \in S \subset X$ we simply replace f by $f + \psi(\cdot|S)$.

(2) We do not claim that the results in Theorem 2.2 are original. Results similar to these can be found elsewhere in the literature, e.g., [1], [3], [15]–[17], [39]–[41], [45], [54], [63]–[66]. However, the proofs that we provide are different from those that are usually provided, due to the explicit dependence on Theorem 2.1.

Various conditions can be found in the literature that ensure that the calmness hypothesis is satisfied. All of these conditions are related to the regularity of the constraint systems of the form

$$g(x) \in C \quad \text{and} \quad x \in S \subset X. \tag{2.4}$$

DEFINITION 2.2. System (2.4) is said to be *regular* at a solution x_0 if there exist constants $\kappa > 0$ and $\varepsilon > 0$ such that

$$\operatorname{dist}(x|\Omega(u)) \leq \kappa \operatorname{dist}(g(x)|C + u)$$

for all $x \in (x_0 + \varepsilon \mathbb{B}) \cap S$ and $u \in \varepsilon \mathbb{B}$ where

$$\Omega(u) := \{x \in X : g(x) \in C + u, x \in S\}.$$

The constant κ is called the *modulus of regularity* for (2.4) at x_0.

Remark. This and more general notions of regularity for (2.4) are studied by several authors, e.g., [1], [4], [7], [15], [20], [48], [49], [51], [52], [61], [62], [66]–[68], [73].

Calmness and regularity are related via Clarke's elementary exact penalization theorem.

THEOREM 2.3 (Clarke [15, Prop. 2.4.3]). *Let $f : X \to \overline{\mathbb{R}}$ be Lipschitz of rank κ on a set $T \subset X$. Let $\overline{x}$ belong to a set $\Omega \subset T$ and suppose that f attains a minimum over Ω at $\overline{x}$. Then for any $\widehat{\kappa} \geq \kappa$, the function $\pi_{\widehat{\kappa}}(x) := f(x) + \widehat{\kappa}\ \mathrm{dist}(x|\Omega)$ attains a minimum over T at $\overline{x}$. If $\widehat{\kappa} > \kappa$ and Ω is closed, then any other point minimizing $\pi_{\widehat{\kappa}}$ over T must also lie in Ω.*

We have the following elementary corollaries to Theorem 2.3.

COROLLARY 2.3.1. *Consider the problem $\mathcal{P}$ with g continuous and $f := f_0 + \psi(\cdot|S)$ for some $f_0 : X \to \mathbb{R}$ and some $S \subset X$ closed and nonempty. Suppose that $\overline{x} \in S$ is a local solution to $\mathcal{P}$ at which the system (2.4) is regular with modulus κ_1 and near which f_0 is Lipschitz of rank κ_2, then $\overline{x}$ is a local minimum of $P_\alpha(x) := f(x) + \alpha\ \mathrm{dist}(g(x)|C)$ for all $\alpha \geq \kappa_1\kappa_2$. If $\alpha > \kappa_1\kappa_2$, then there is a neighborhood of $\overline{x}$ such that any other local minimum $\widehat{x}$ of $P_\alpha(x)$ within this neighborhood is such that $f(\overline{x}) = f(\widehat{x})$ and $g(\widehat{x}) \in C$.*

Proof. Let $\varepsilon > 0$ be such that f_0 is Lipschitz of rank κ_2 on $\overline{x} + \varepsilon\mathbb{B}$, the defining inequality for regularity holds for all $x \in (\overline{x} + \varepsilon\mathbb{B}) \cap S$ and $u \in \varepsilon\mathbb{B}$, and $f(\overline{x}) \leq f(x)$ for all $x \in \{z : g(z) \in C\} \cap (\overline{x} + \varepsilon\mathbb{B})$. Set $\Omega := \Omega(0) \cap (\overline{x} + \varepsilon\mathbb{B})$ and note that since g is continuous, the set Ω is closed. By Theorem 2.3, $\pi_{\widehat{\kappa}}(x)$ attains a minimum over $\overline{x} + \varepsilon\mathbb{B}$ at $\overline{x}$ for $\widehat{\kappa} \geq \kappa_2$, and if $\widehat{\kappa} > \kappa_2$, then any other minimum of $\pi_{\widehat{\kappa}}$ over $\overline{x} + \varepsilon\mathbb{B}$ must also lie in Ω. Then, for every $\delta \in (0, \frac{1}{3}\varepsilon)$ and $y \in \overline{x} + \frac{1}{3}\varepsilon\mathbb{B}$, there is a $z_\delta \in \Omega(0)$ such that $\|y - z_\delta\| \leq \mathrm{dist}(y|\Omega(0)) + \delta \leq \frac{2}{3}\varepsilon$. Hence $\|z_\delta - \overline{x}\| \leq \|y - z_\delta\| + \|y - \overline{x}\| \leq \varepsilon$ so that $\mathrm{dist}(y|\Omega) \leq \mathrm{dist}(y|\Omega(0)) + \delta$. Letting $\delta \downarrow 0$ we find that $\mathrm{dist}(y|\Omega) = \mathrm{dist}(y|\Omega(0))$ for all $y \in \overline{x} + \frac{1}{3}\varepsilon\mathbb{B}$. The result now follows from the definition of regularity with the neighborhood of $\overline{x}$ being $\overline{x} + \frac{1}{3}\varepsilon\mathbb{B}$. □

COROLLARY 2.3.2. *Consider the problem $\mathcal{P}$ and let f, g, and $\overline{x}$ be as in Corollary 2.3.1. Then $\mathcal{P}$ is calm at $\overline{x}$.*

Proof. This is an immediate consequence of Corollary 2.3.1 and Theorem 2.1. □

Remark. Dolecki and Rolewicz [20] obtain a result similar to Corollary 2.3.1 in a more general setting by using somewhat different techniques. Their result is based upon the notion of an upper Hausdorff semicontinuous multifunction.

Conditions yielding the regularity of the constraint system (2.4) have been studied by many authors [1], [4], [7], [15], [20], [48], [49], [51], [52], [61], [62], [66]–[68], [73]. The first and most famous of these results is the Lyusternik theorem [48]. An excellent discussion of a variety of these regularity results is given in Borwein [7]. In the mathematical programming literature such conditions are often called constraint qualifications, e.g., the Mangasarian–Fromovitz constraint qualification [51], [52]. In his thesis, Maguregui [49, Chap. 2], introduced the constraint qualification

$$0 \in \mathrm{core}(g(x_0) + g'(x_0)(S - x_0) - C). \tag{2.5}$$

THEOREM 2.4 (Maguregui [49, Chap. 2]). *Suppose that X and Y are Banach spaces, the sets $S \subset X$ and $C \subset Y$ are nonempty, closed, and convex, and $g : X \to Y$ is strictly differentiable at $\overline{x} \in S$. If $g(\overline{x}) \in C$ and (2.5) is satisfied, then system (2.4) is regular at $\overline{x}$.*

Remarks. (1) Using the constraint qualification

$$g'_s(x_0)T(x_0|S) - T(g(x_0)|C) = Y, \tag{2.6}$$

Borwein [7, Thm. 4.3] show that the convexity assumption on the sets C and S can be removed if we instead assume that the sets S and C are epi-Lipschitzian (in the sense of Rockafellar [68]) at x_0 and $g(x_0)$, respectively.

(2) If X and Y are finite-dimensional or if C and S are convex, then the conditions (2.5), (2.6), and

$$\ker([g_s'(x_0)^T, I]) \cap [N(g(x_0)|C) \times N(x_0|S)] = \{0\} \tag{2.7}$$

are all equivalent. Moreover, if $S = \mathbb{R}^n$, and $C := \mathbb{R}^s \times \{0\}_{R^{m-s}}$, all of the conditions (2.5)–(2.7) are equivalent to the Mangasarian–Fromovitz constraint qualification.

(3) In the finite-dimensional case, Borwein [7, Thm. 3.2] has shown that we can generalize (2.7) to

$$\ker([\partial g(x_0)^T, I]) \cap [N(g(x_0)|C) \times N(x_0|S)] = \{0\}, \tag{2.8}$$

where

$$\ker([\partial g(x_0)^T, I]) := \{(y,z) \in \mathbb{R}^m \times \mathbb{R}^n : 0 \in \partial g(x_0)^T y + z\},$$

and still guarantee the regularity of system (2.4).

COROLLARY 2.4.1. *Let the hypotheses of Theorem* 2.4 *hold and consider the problem* $\mathcal{P}$ *with* g *continuous and* $f := f_0 + \psi(\cdot|S)$, *where* $f_0 : X \to \mathbb{R}$ *Lipschitz near* $\overline{x}$. *Then* $\mathcal{P}$ *is calm at* $\overline{x}$, *or equivalently,* $\overline{x}$ *is a local minimum for* P_α *for all* α *sufficiently large. Moreover, there is a threshold value of* α, *say* $\overline{\alpha}$, *and a neighborhood* U *of* $\overline{x}$ *such that if* $\alpha > \overline{\alpha}$, *then any other local minimum of* P_α, $\widehat{x} \in U$, *must satisfy* $f(\overline{x}) = f(\widehat{x})$ *and* $g(\widehat{x}) \in C$.

Proof. This is an immediate consequence of Theorem 2.1, Corollary 2.3.1, and Theorem 2.4. □

Remarks. (1) Corollary 2.4.1 extends Han and Mangasarian [33, Thm. 4.4] where it is assumed that $X = \mathbb{R}^n, Y = \mathbb{R}^m$, $S = \mathbb{R}^n$, and $Y := \mathbb{R}^s_- \times \{0\}_{\mathbb{R}^{m-s}}$, f_0 and g are continuously differentiable and $\overline{x}$ is a strict local solution to $\mathcal{P}$.

(2) Dolecki and Rolewicz [20, Thm. 2.1] obtain a result similar to Corollary 2.4.1 in a somewhat more general setting. Their result is based upon the notion of locally controllable image nearly inner approximations (inia).

In finite dimensions it is possible to strengthen the result in Corollary 2.4.1 by dropping the requirement that S be convex. Clarke establishes this in [15, Cor. 5, p. 244]. It can also be established by methods that place exact penalty techniques within a broader context of convex composite optimization. In convex composite optimization one studies the problem

$$(\mathcal{Q}) \qquad \text{minimize} \quad q(x)$$

with $q := f + h \circ g$ where $f : X \to \overline{\mathbb{R}}$ and $g : X \to Y$ are as in the statement of $\mathcal{P}$, and $h : Y \to \overline{\mathbb{R}}$ is convex. If C is convex, then $P_\alpha(x) := f(x) + \alpha \operatorname{dist}(g(x)|C)$ is an example of a convex composite function. The following result concerning $\mathcal{Q}$ is a modest extension of a result originally due to Burke and Poliquin [11, Thm. 3.1].

THEOREM 2.5. *Consider the problem* $\mathcal{Q}$ *where* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *is lower semicontinuous,* $g : \mathbb{R}^n \to \mathbb{R}^m$ *is locally Lipschitz, and* $h : \mathbb{R}^m \to \overline{\mathbb{R}}$ *is lower semicontinuous and convex.*

Let $\overline{x} \in \operatorname{dom}(q)$ *and suppose that*

$$\begin{bmatrix} 0 \in \partial^\infty f(\overline{x}) + \partial g(\overline{x})^* y \\ y \in N(g(\overline{x})|\operatorname{dom}(h)) \end{bmatrix} \Longleftrightarrow y = 0. \tag{2.9}$$

Define

$$q_\alpha(x) := f(x) + h_\alpha(g(x)) \tag{2.10}$$

with

$$h_\alpha(y) := \inf\{h(z) + \alpha\|y - z\| : z \in \mathbb{R}^m\}. \tag{2.11}$$

If $\overline{x} \in \mathrm{dom}(q)$ *is a local solution to* $\mathcal{Q}$, *then there is an* $\overline{\alpha} > 0$ *such that* $\overline{x}$ *is a local minimizer for* $q_\alpha(x)$ *for all* $\alpha \geq \overline{\alpha}$.

Remarks. (1) The proof of Theorem 2.5 is rather technical, and so is relegated to Appendix A.

(2) The operation employed in (2.11) is known as the infimal convolution of h and $\alpha\|\cdot\|$, and is written $h \,\square\, \alpha\|\cdot\|$. In general, we have

$$\mathrm{epi}[h_1 \,\square\, h_2] = \mathrm{epi}(h_1) + \mathrm{epi}(h_2)$$

for any two convex functions h_1 and h_2. Consequently, $h_1 \,\square\, h_2$ is always convex.

(3) Note that $\mathrm{dom}(h_\alpha) = \mathbb{R}^m$ even if $\mathrm{dom}(h) \neq \mathbb{R}^m$. Hence $\mathrm{dom}(q_\alpha) = \mathrm{dom}(f)$.

If the set C in problem $\mathcal{P}$ is convex, then $\mathcal{P}$ can be seen as an instance of $\mathcal{Q}$ by taking $h := \psi(\cdot|C)$. In this case we have

$$\begin{aligned} h_\alpha(y) :&= \inf\{\psi(z|C) + \alpha\|y - z\| : z \in \mathbb{R}^m\} \\ &= \alpha\ \mathrm{dist}(y|C), \end{aligned}$$

and so

$$q_\alpha(x) = P_\alpha(x) := f(x) + \alpha\ \mathrm{dist}(g(x)|C).$$

Thus Theorem 2.5 can be used to provide conditions under which a finite exact penalty parameter α exists. Condition (2.9) is just another constraint qualification. In particular, if $f := f_0 + \psi(\cdot|S)$ with f_0 locally Lipschitz and S closed, then (2.8) and (2.9) are equivalent and we recover Clarke's result [15, Cor. 5, p. 244] as a special case. Constraint qualifications of the type (2.9) were originally formulated by Rockafellar in [66] and [68]. These comments yield the following corollary to Theorem 2.5.

COROLLARY 2.5.1. *Let* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, *and* $\overline{x} \in \mathbb{R}^n$ *be as in the statement of Theorem* 2.5 *and consider problem* $\mathcal{P}$. *If* (2.9) *holds with* $h := \psi(\cdot|C)$ *where is* C *nonempty closed and convex, then* $\mathcal{P}$ *is calm at* $\overline{x}$, *or equivalently,* $\overline{x}$ *is a local minimum for* P_α *for all* α *sufficiently large. Moreover, there is a threshold value of* α, *say* $\overline{\alpha}$, *and a neighborhood* U *of* $\overline{x}$ *such that if* $\alpha > \overline{\alpha}$, *then any minimum* $\widehat{x}$ *of* P_α *on* U *must satisfy* $f(\overline{x}) = f(\widehat{x})$ *and* $g(\widehat{x}) \in C$.

Proof. In light of the comments preceding the statement of the result, we need only prove the last part of the result. To this end let $\overline{\alpha}$ be any value of α for which $\overline{x}$ is a local minimum of P_α, and let U be any neighborhood of $\overline{x}$ such that $P_{\overline{\alpha}}(\overline{x}) \leq P_{\overline{\alpha}}(x)$ for all $x \in U$. If $\widehat{\alpha} > \overline{\alpha}$, then $P_{\overline{\alpha}}(\overline{x}) \leq P_{\widehat{\alpha}}(x)$ for all $x \in U$. If $\widehat{x}$ is any other minimum of $P_{\widehat{\alpha}}$ on U, then

$$P_{\widehat{\alpha}}(\widehat{x}) = f(\overline{x}) \leq f(\widehat{x}) + \frac{\overline{\alpha} + \widehat{\alpha}}{2}\ \mathrm{dist}(g(\widehat{x})|C).$$

Consequently, $\mathrm{dist}(g(\widehat{x})|C) = 0$ and $f(\overline{x}) = f(\widehat{x})$. □

Remark. For the case in which $C := \mathbb{R}^s_- \times \{0\}_{\mathbb{R}^{m-s}}$ and both f and g are continuously differentiable, this result was first obtained by Han and Mangasarian in [33, Thm. 4.4]. Rosenberg [71, Prop. 1] later generalized Han and Mangasarian's result to the case in which $C := \mathbb{R}^s_- \times \{0\}_{R^{m-s}}$, $f := f_0 + \psi(\cdot|S)$, where f_0 and g are locally Lipschitz and S is nonempty and closed.

In this section we have obtained multiplier rules for $\mathcal{P}$ via the exact penalty function P_α and the calmness hypothesis. We call these multipliers Kuhn–Tucker multipliers. Given $x \in \Omega := \{x \in \operatorname{dom}(f) : g(x) \in C\}$, we denote these multipliers by

$$\text{K-T}(x) := \{y \in N(g(x)|C) : 0 \in \partial f(x) + \partial g(x)^* y\}, \tag{2.12}$$

where $\partial g(x)$ is always taken to be $g'_s(x)$ in the infinite-dimensional setting. This set is always closed and may be empty. It should be noted that this is an extension of the usual theory of Kuhn–Tucker multipliers; that is, if f and g are continuously differential and $C := \mathbb{R}^s_- \times \{0\}_{\mathbb{R}^{m-s}}$, then K-T (x) consists precisely of the usual Kuhn–Tucker multipliers for $\mathcal{P}$ at x [43], [51], [70].

PROPOSITION 2.6. *Suppose X and Y are normed linear spaces, C is a closed subset of Y, $f : X \to \overline{\mathbb{R}}$, and $g : X \to Y$. Let $\overline{x} \in \operatorname{dom}(P_\alpha)$ be such that $g(\overline{x}) \in C$.*

(1) *If g is strictly differentiable at $\overline{x}$ and $0 \in \partial P_\alpha(\overline{x})$ for some $\alpha \geq 0$, then K-T $(\overline{x}) \neq \emptyset$.*

(2) *If X and Y are finite-dimensional, g is Lipschitz near $\overline{x}$, and $0 \in \partial P_\alpha(x)$ for some $\alpha \geq 0$, then K-T $(\overline{x}) \neq \emptyset$.*

(3) *If f is subdifferentially regular at $\overline{x}$, g is strictly differentiable at $\overline{x}$, C is convex, and K-T $(\overline{x}) \neq \emptyset$, then $0 \in \partial P_\alpha(\overline{x})$ for all $\alpha >$ $\operatorname{dist}_0(0|K\text{-}T(\overline{x}))$ (or $\alpha \geq$ $\operatorname{dist}_0(0|K\text{-}T(\overline{x}))$ if Y^* is separable).*

Proof. (1) This follows directly from [68, Thms. 2 and 3] and inclusion (2.3).

(2) This follows from [66, Cor. 5.2.3].

(3) The proof is by [68, Thms. 2 and 3],

$$\alpha \partial P_\alpha(\overline{x}) = \partial f(\overline{x}) + \alpha g'_s(\overline{x})^* \partial\, [\operatorname{dist}(\cdot|C)](g(\overline{x})).$$

Let $y \in K - T(\overline{x})$ be such that $\|y\|_0 < \alpha$. If Y^* is separable we can choose $\|y\|_0 = \alpha$ [21], [74]. Then $0 \in \partial P_\alpha(\overline{x})$ since, by (3.5),

$$\partial\, [\operatorname{dist}(\cdot|C)](g(\overline{x})) = \mathbb{B}^0 \cap N(g(\overline{x})|C).$$

Remark. Proposition 2.6 extends similar results found in Garcia-Palomares [31, §4], Han and Mangasarian [33, §4], Lasserre [44], Polak, Mayne, and Wardi [59, §3], and Rosenberg [71]. All of these results apply to the finite-dimensional case with $C := \mathbb{R}^s_- \times \{0\}_{\mathbb{R}^{m-s}}$. They obtain results for other norms by appealing to the equivalence of norms in finite-dimensions.

It is well known that K-T $(\overline{x})$ may be empty even if $\overline{x}$ is a local solution to $\mathcal{P}$. Nevertheless, more general multiplier rules can be established in this case. The such result is attributed to John [42]. In the next section, we generalize this result to $\mathcal{P}$.

3. A John type multiplier rule for $\mathcal{P}$.

In this section we consider the problem $\mathcal{P}$ with $f := f_0 + \psi(\cdot|S)$, $f_0 : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ locally Lipschitz, $S \subset \mathbb{R}^n$ nonempty and closed, and $C \subset \mathbb{R}^m$ nonempty, closed, and convex, and derive a multiplier rule that does not depend on calmness. For this purpose let $\overline{x}$ be a local solution of radius ε to $\mathcal{P}$, and for each $\delta \geq 0$, consider the function $\theta_\delta : \mathbb{R} \times \mathbb{R}^n \to \overline{\mathbb{R}}$ given by

$$\theta_\delta(x) := \operatorname{dist}[(f_0(x), g(x))|C_{\overline{x},\delta}] + \psi(\cdot|S \cap (\overline{x} + \varepsilon\mathbb{B})), \tag{3.1}$$

where

$$C_{\overline{x},\delta} := (f_0(\overline{x}) - \delta + \mathbb{R}_-) \times C \subset \mathbb{R} \times \mathbb{R}^m. \tag{3.2}$$

It is assumed that the norm chosen for $\mathbb{R} \times \mathbb{R}^m$ is such that $\|(\xi, 0)\| = |\xi|$. Observe that for each $\delta \geq 0$ we have

$$\theta_\delta(\overline{x}) \leq \delta + \inf\theta_\delta(x) \tag{3.3}$$

and if $\delta > 0$, then

$$0 < \inf\theta_\delta(x) < +\infty. \tag{3.4}$$

Thus, in particular, $\overline{x}$ is a global minimum for θ_0. The function θ_0 is a kind of exact penalty function for $\mathcal{P}$. It is similar to the Eremin–Zangwill penalty functions except that no a priori assumptions are required for $\overline{x}$ to be a global minimum for θ_0. Exact penalty functions of this type were considered by Morrison [55] in the case where $C = \{0\}$, $S = \mathbb{R}^n$, and where $\mathbb{R}^m$ is given the Euclidean norm. In this setting, Morrison showed how we can apply the methods of nonlinear least squares to solve $\mathcal{P}$. Further discussion of these penalty functions is given in Fletcher [27].

By applying the appropriate rules of the subdifferential calculus to θ_0, we can obtain a multiplier rule for $\mathcal{P}$. Unfortunately, such a direct application yields a rather uninteresting multiplier rule because of the nature of the subdifferential of the distance function $\operatorname{dist}[\cdot|C_{\overline{x},\varepsilon}]$.

PROPOSITION 3.1. *Let Γ be a nonempty, closed, convex subset of a normed linear space X. Then* $\operatorname{dist}(y|\Gamma)$ *is a convex function whose subdifferential is*

$$\partial \operatorname{dist}(y|\Gamma) := \begin{cases} \mathbb{B}^0 \cap N(y|\Gamma), & \text{if } y \in \Gamma \\ (\operatorname{bdry} \mathbb{B}^0) \cap N(y|\Gamma + \operatorname{dist}(y|\Gamma)\mathbb{B}), & \text{otherwise.} \end{cases} \tag{3.5}$$

If Γ is not assumed to be convex, then

$$N(y|\Gamma) = \operatorname{cl}[\cup_{\lambda \geq 0} \lambda\partial \operatorname{dist}(y|\Gamma)]. \tag{3.6}$$

Proof. In the convex case with $y \in \Gamma$, the formula $\partial \operatorname{dist}(y|\Gamma) = \mathbb{B}^0 \cap N(y|\Gamma)$ is elementary and well known. When Γ is convex and $y \notin \Gamma$, the formula is derived in Burke [9, §2]. The final formula (3.6) is due to Clarke [15, Prop. 2.4.2]. □

Thus a direct application of the chain rule [15, Thm. 2.3.10] to θ_0 would yield, according to Proposition 3.1, the trivial inclusion

$$0 \in \mathbb{R}_+\partial f_0(\overline{x}) + \partial g(\overline{x})^* N(g(\overline{x})|C) + N(\overline{x}|S).$$

This is the reason for including the perturbation δ in definition (3.1). Due to inequality (3.3) we can apply Ekeland's variational principle [22] to obtain, for each $\delta > 0$, the existence of an $x_\delta \in (\overline{x} + \varepsilon\mathbb{B}) \cap S$ satisfying

$$\|\overline{x} - x_\delta\| \leq \sqrt{\delta},$$

and

$$\theta_\delta(x) + \sqrt{\delta}\|x - x_\delta\| > \theta_\delta(x_\delta)$$

for all $x \neq x_\delta$. Hence $\overline{x}$ is a strict global minimum of the function

$$\widehat{\theta}_\delta(x) := \theta_\delta(x) + \sqrt{\delta}\|x - x_\delta\|.$$

Now, by (3.4), we have $\theta_\delta(x_\delta) > 0$ and so $(f_0(x_\delta), g(x_\delta)) \notin C_{\overline{x},\delta}$. Thus, when we apply the appropriate rules of the subdifferential calculus (Rockafellar [66, Cor. 5.2.3]) to the inclusion $0 \in \partial\widehat{\theta}_\delta(x_\delta)$, we obtain the existence of an $x_\delta \geq 0$ and

$$y_\delta \in N(g(x_\delta)|C + \text{ dist}(g(x_\delta)|C)\mathbb{B})$$

with $\|(\lambda_\delta, y_\delta)\|_0 = 1$ such that

$$0 \in \lambda_\delta \partial f_0(x_\delta) + \partial g(x_\delta)^* y_\delta + N(x_\delta|S) + \sqrt{\delta}\mathbb{B}^0$$

for all δ with $\sqrt{\delta} < \varepsilon$. Consequently, any cluster point $(\overline{\lambda}, \overline{y})$ of $\{(\lambda_\delta, y_\delta)\}$ as $\delta \downarrow 0$ must satisfy

$$\|(\overline{\lambda}, \overline{y})\|_0 = 1, \tag{3.7}$$

$$\overline{\lambda} \geq 0,\ \overline{y} \in N(g(\overline{x})|C), \tag{3.8}$$

and

$$0 \in \overline{\lambda}\partial f_0(\overline{x}) + \partial g(\overline{x})^* \overline{y} + N(\overline{x}|S). \tag{3.9}$$

We have just proved the following theorem.

THEOREM 3.1. *Let f, g, s, C, and $\overline{x}$ be as given at the beginning of this section. Then there exist multipliers $\overline{\lambda} \geq 0$ and $\overline{y} \in N(g(\overline{x})|C)$ such that* (3.7)–(3.9) *hold.*

With a bit of work this result can be obtained from Clarke [15, Thm. 6.1.1]. Moreover, the proof that we provide has a certain similarity to Clarke's proof. We included this proof since it is simpler and more direct. Furthermore, it illustrates the intimate relationship between the multipliers and the subgradient of the distance function at $(f_0(\overline{x}), g(\overline{x}))$.

Note that if the multiplier $\overline{\lambda}$ in (3.7) is nonzero, then $\overline{\lambda}^{-1}\overline{y} \in \text{K-T}(\overline{x})$, i.e.,

$$\text{K-T}(\overline{x}) := \{\overline{\lambda}^{-1}\overline{y} : (\overline{\lambda}, \overline{y}) \text{ satisfy } (3.7)\text{–}(3.9) \text{ with } \overline{\lambda} \neq 0\}. \tag{3.10}$$

Moreover, if f and g satisfy the conditions of part (2) of Proposition 2.6, then

$$0 \in \partial P_{\overline{\lambda}^{-1}}(\overline{x}).$$

Thus the magnitude of $\overline{\lambda}$ is inversely related to the magnitude of an exact penalty parameter for $\mathcal{P}$. The multipliers $(\overline{\lambda}, \overline{y})$, for which $\overline{\lambda} = 0$, are of great significance in the analysis of $\mathcal{P}$. We call these multipliers *Fritz John multipliers* and denote them by

$$\begin{aligned} FJ(x) : &= \{\mu\overline{y} : \mu \geq 0, (0, \overline{y}) \text{ satisfies } (3.7)\text{—}(3.9)\ \} \\ &= \ker([\partial g(x)^T, I]) \cap (N(g(x)|C) \times N(x|S)), \end{aligned}$$

where

$$\ker([\partial g(x)^T, I]) := \{(y, z) \in \mathbb{R}^m \times \mathbb{R}^n : 0 \in \partial g(x)^T y + z\}.$$

Observe that $FJ(x)$ is a nonempty, closed, and convex cone for every $x \in S$ with $g(x) \in C$. Moreover, if K-T $(x) \neq \emptyset$, and g is strictly differentiable at x, then $FJ(x) = \text{rec}(\text{K-T}(x))$. Clarke [15] refers to the Kuhn–Tucker and Fritz John multipliers as the normal and abnormal multipliers, respectively.

According to Theorem 3.1, one is guaranteed of the existence of Kuhn–Tucker multipliers at a local solution $\overline{x}$ to $\mathcal{P}$ if $FJ(\overline{x}) = \{0\}$, or equivalently, if

$$\ker[\partial g(\overline{x})^T, I] \cap (N(g(x)|C) \times N(x(S)) = \{0\}.$$

This condition is precisely the constraint qualification (2.8) and (2.9) of the previous section. Thus we see that condition (2.8) is truly a fundamental property for constrained optimization. It is a natural condition under which we obtain both constraint regularity and the the existence of Kuhn–Tucker multipliers. For this reason, we will refer to (2.8) as the *basic constraint qualification* throughout the remainder of the paper.

PROPOSITION 3.2. *Let f, g, and C be as given in the beginning of this section and let $x \in S$ be such that $g(x) \in C$ and K-T $(x) \neq \emptyset$.*

(1) *If the basic constraint qualification* (2.8) *is satisfied at x, then K-T (x) is compact.*
(2) *If g is strictly differentiable at x, then* K-T (x) *is convex and* $\text{rec}(\text{K-T}(x)) = FJ(x)$, *in which case* K-T (x) *is compact if and only if the basic constraint qualification* (2.8) *is satisfied at x.*
(3) *If $\overline{x}$ is a local solution to $\mathcal{P}$ at which the basic constraint qualification* (2.8) *is satisfied, then* K-T $(\overline{x})$ *is nonempty.*

Proof. (1) If K-T (x) is not compact, then it contains an unbounded sequence $\{y_i\} \subset N(g(x)|C)$. For each $i = 1, 2, \cdots$, there exists vectors $v_i \in \partial f_0(x)$ and $z_i \in N(x|S)$ and a matrix $J_i \in \partial g(x)$ such that

$$0 = v_i + J_i^T y_i + z_i.$$

With no loss in generality, we can assume that $(y_i, z_i)(\|y_i\| + \|z_i\|)^{-1} \to (\overline{y}, \overline{z})$ and $J_i \to \overline{J}$ with $\|\overline{y}\| + \|\overline{z}\| = 1$, $\overline{y} \in N(g(x)|C), \overline{z} \in N(x|S)$, and $\overline{J} \in \partial g(x)$. But then $0 = \overline{J}^T \overline{y} + \overline{z}$ so that $FJ(x) \neq \{0\}$, a contradiction. Hence K-T (x) is compact.

(2) The convexity of K-T (x) and the equivalence $\text{rec}(\text{K-T}(x)) = FJ(x)$ follow directly from the definitions. Thus the equivalence of (2.8) with the compactness of K-T (x) follows immediately from [70, Thm. 8.4].

(3) This follows from the preceding discussion. □

Remark. Proposition 3.2 extends a well-known result of Gauvin [32]. Another generalization of Gauvin's result is obtained in Nguyen, Strodiot, and Mifflin [56], where it is assumed that $C := \mathbb{R}^s_- \times \{0\}_{\mathbb{R}^{m-s}}$ and that the s components of g are Lipschitz.

4. Second-order optimality conditions for $\mathcal{P}$. The second-order results of this section are based on the second-order theory for convex composite optimization developed in Burke [10] and Burke and Poliquin [11]. If $f := f_0 + \psi(\cdot|S)$ with $f_0 : \mathbb{R}^n \to \mathbb{R}$ and the sets $S \subset \mathbb{R}^n$ and $C \subset \mathbb{R}^m$ are taken to be nonempty, closed and convex, then the exact penalty functions P_α and θ_0 defined in §§2 and 3, respectively, are convex composite functions. Thus we can apply the results of [10], [11] directly to these functions. The theorems obtained in this way are very much in the spirit of those established in Levitin, Miljutin, and Osinolovski [45], Ioffe [39]–[41], Ben–Isreal, Ben-Tal, and Zlobec [4], and Rockafellar [63]–[64]. These results are distinguished by their use of the entire set of multipliers rather than a single vector of multipliers as is the case in the classical theory of second-order optimality conditions (e.g., see Hestenes [35]–[36], Pennisi [57], and Fiacco and McCormick [25]). Let us begin by reviewing the pertinent results in [66] and [11].

THEOREM 4.1 (Rockafellar [66, Cor. 5.2.3]). *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is lower semicontinuous, $g : \mathbb{R}^n \to \mathbb{R}^n$ is locally Lipschitz, and $\overline{x} \in \operatorname{dom}(q)$, where $q(x) := f(x) + h(g(x))$, is such that (2.9) holds. If $\overline{x}$ is a local minimum of q, then the set of multipliers*

$$M_Q(\overline{x}) := \{y \in \partial h(\cdot)(g(\overline{x})) : 0 \in \partial f(\overline{x}) + \partial g(\overline{x})^* y\} \tag{4.1}$$

is nonempty.

THEOREM 4.2 (Burke and Poliquin [11, Thm. 4.2]). *Let $\overline{x} \in S \subset \mathbb{R}^n$ be such that $f_0 : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are twice continuously differentiable near $\overline{x}$. Moreover, let $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ be lower semicontinuous and convex with $g(\overline{x}) \in \operatorname{dom}(h)$, and suppose that S is closed and convex. Set $q := f_0 + \psi(\cdot|S) + h \circ g$.*

(1) *If $\overline{x}$ is a local minimum for q at which the basic constraint qualification (2.7) is satisfied, then $M_Q(\overline{x}) \neq \emptyset$ and*

$$\max\{d^T(\nabla^2 f_0(\overline{x}) + \nabla^2_{xx}(\langle y, g(\overline{x})\rangle))d : y \in M_Q(\overline{x})\} \geq 0 \tag{4.2}$$

for all $d \in \operatorname{cl}(K_Q(\overline{x})) \cap T(\overline{x}|S)$

$$K_Q(x) := \{d \in \mathbb{R}^n : \exists \bar{t} > 0 \quad \text{such that } h(g(x) + tg'(x)d) \leq h(g(x)) \forall t \in (0, \bar{t})\}. \tag{4.3}$$

(2) *If $M_Q(\overline{x}) \neq \emptyset$ and*

$$\sup\{d^T(\nabla^2 f_0(\overline{x}) + \nabla^2_{xx}(\langle y, g(\overline{x})\rangle))d : y \in M_Q(\overline{x})\} > 0 \tag{4.4}$$

for all $d \in D_Q(\overline{x}) \backslash \{0\}$ where

$$D_Q(x) := \{d \in \mathbb{R}^n : q'(x; d) \leq 0\}, \tag{4.5}$$

then there is a $\gamma > 0$ such that

$$q(x) \geq q(\overline{x}) + \gamma \|x - \overline{x}\|^2$$

for all x near $\overline{x}$.

PROPOSITION 4.3. *Let $\overline{x} \in S \subset \mathbb{R}^n$ be such that $f_0 : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable near $\overline{x}$. Moreover, let $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ be lower semicontinuous and convex with $g(\overline{x}) \in \operatorname{dom}(h)$, suppose that S is closed, and set $q := f_0 + \varphi(\cdot|S) + h \circ g$. If*

$$\operatorname{ran}\begin{bmatrix} g'(\overline{x}) \\ I \end{bmatrix} \bigcap \operatorname{ri}[T(g(\overline{x})|\operatorname{lev}_h(g(x))) \times T(\overline{x}|S)] \neq \emptyset \tag{4.6}$$

and

$$\operatorname{cone}(\partial h(\cdot)(g(\overline{x}))) = N(g(\overline{x})|\operatorname{lev}_h(g(\overline{x}))), \tag{4.7}$$

then $D_{\mathcal{Q}}(\overline{x}) = \overline{K_{\mathcal{Q}}(\overline{x})} \cap T(\overline{x}|S)$. *Moreover, if* $0 \notin \partial h(\cdot)(g(\overline{x}))$, *then* (4.7) *is satisfied, and if the basic constraint qualification* (2.8) *holds, then* (4.6) *is satisfied.*

Proof. All but the very last statement is established in [11, Prop. 5.1]. For the last statement, we take polars in (2.7) to obtain

$$\operatorname{ran}\begin{bmatrix} g'(\overline{x}) \\ I \end{bmatrix} + (T(g(\overline{x})|\operatorname{lev}_h(g(\overline{x}))) \times T(\overline{x}|S)) = \mathbb{R}^m.$$

Now, for any subspace W and closed convex cone K the condition $W + K = \mathbb{R}^m$ implies that $W \cap \operatorname{ri}(K) \neq \emptyset$ by a simple separation argument. This establishes the result. □

We now apply these results to $\mathcal{P}$. The result is a sufficiency theorem which does not require a constraint qualification. The result is obtained by applying Theorem 4.2 to the function θ_0.

THEOREM 4.4. *Let* S *and* C *be nonempty closed convex subsets of* $\mathbb{R}^n$ *and* $\mathbb{R}^m$, *respectively, and let* $\overline{x} \in S$ *be such that* $f_0 : \mathbb{R}^n \to \mathbb{R}$ *and* $g : \mathbb{R}^n \to \mathbb{R}^m$ *are twice continuously differentiable near* $\overline{x}$ *and* $g(\overline{x}) \in C$. *Set* $f := f_0 + \psi(\cdot|S)$ *and consider the problem* $\mathcal{P}$. *If the set of multipliers*

$$M_{\mathcal{P}}(\overline{x}) := \{(\lambda, y) \in \mathbb{R} \times \mathbb{R}^m : (3.7)\text{–}(3.9) \text{ are satisfied}\} \tag{4.8}$$

is nonempty and

$$\max\{d^T(\lambda\nabla^2 f_0(\overline{x}) + \nabla^2_{xx}(\langle y, g(\overline{x})\rangle))d : (\lambda, y) \in M_{\mathcal{P}}(\overline{x})\} > 0 \tag{4.9}$$

for every $d \in D_{\mathcal{P}}(\overline{x})$ *where*

$$D_{\mathcal{P}}(\overline{x}) := \{d \in T(\overline{x}|S) : \nabla f_0(\overline{x})^T d \le 0, g'(\overline{x})d \in T(g(\overline{x})|C)\}, \tag{4.10}$$

then there is a $\gamma > 0$ *such that*

$$f_0(x) \ge f_0(\overline{x}) + \gamma\|x - \overline{x}\|^2 \tag{4.11}$$

for every $x \in \Omega := \{x \in S : g(x) \in C\}$ *near* $\overline{x}$.

Proof. Consider part (2) of Theorem 4.2 as it applies to the function θ_0 defined in (3.1). We begin by defining the functions f_0, g, and h and the set S that appear in Theorem 4.2. For the sake of clarity, we denote these functions and set as $f_{0_{(4.2)}}$, $g_{(4.2)}$, $h_{(4.2)}$, and $S_{(4.2)}$, respectively. For the remainder of the proof the functions f_0 and g, and the set S, will refer to those that are given in the statement of Theorem 4.4. With this notation we define $f_{0_{(4.2)}} \equiv 0$, $g_{(4.2)} := (f_0, g)$, $h_{(4.2)} :=$ $\operatorname{dist}[\cdot|C_{\overline{x},0}]$, and $S_{(4.2)} := S \cap (\overline{x} + \varepsilon\mathbb{B})$. The set $M_{\mathcal{Q}}(\overline{x})$ is given by

$$\{(\lambda, y) : \|(\lambda, y)\|_0 \le 1 \text{ and } (3.8)\text{–}(3.9) \text{ hold for } f_{0_{(4.2)}} \text{ and } g_{(4.2)}\} \supset M_{\mathcal{P}}(\overline{x}),$$

and the set $D_{\mathcal{Q}}(\overline{x})$ is given by

$$\begin{aligned} &\{d : \psi^*(d|N(\overline{x}|S)) + \psi^*(d|[0,1]\nabla f_0(x) + g'(x)^T(\mathbb{B}^0 \cap N(g(x)|C))) \le 0\} \\ =&\{d \in T(\overline{x}|S) : \lambda\nabla f_0(\overline{x})^T d + y^T g'(\overline{x})d \le 0 \ \forall \lambda \in [0,1], y \in \mathbb{B}^0 \cap N(g(\overline{x})|C)\} \\ =&\{d \in T(\overline{x}|S) : \nabla f_0(\overline{x})^T d \le 0, g'(\overline{x})d \in T(g(\overline{x})|C)\} \\ =&D_{\mathcal{P}}(\overline{x}), \end{aligned}$$

where the line follows by choosing the norm on $\mathbb{R} \times \mathbb{R}^m$ to be $|\xi| + \|y\|_0$ for every $(\xi, y) \in \mathbb{R} \times \mathbb{R}^m$. Since inequality (4.9) implies inequality (4.4), we have the existence of $\gamma > 0$ such that

$$\theta_0(x) \geq \theta_0(\overline{x}) + \gamma\|x - \overline{x}\|^2 \tag{4.12}$$

for all x near $\overline{x}$, where by the theorem is proved. □

Remarks. (1) The theorem actually establishes inequality (4.12), which is stronger than inequality (4.11).

(2) We could just as well have used the multiplier set $M_{\mathcal{Q}}(\overline{x})$ in (4.9), but, since the maximum is positive, both of these multiplier sets yield the same value in (4.9).

Unfortunately, without a constraint qualification, the same trick cannot be applied to obtain a second-order necessary condition for $\mathcal{P}$. The problem is that $M_{\mathcal{P}}(\overline{x}) \subsetneq M_{\mathcal{Q}}(\overline{x})$ with $(0, 0) \in M_{\mathcal{Q}}(\overline{x})$. Consequently (4.2) is valid for all $d \in \mathbb{R}^n$ and it does not imply (4.9) with the weak inequality. On the other hand, if the sets C and S are polyhedral convex, then such a result can be established (e.g., see [35] and [36] or [25]).

If one is willing to assume the basic constraint qualification (2.7), then, by applying Theorem 4.2 to P_α both second-order necessary and sufficient conditions for $\mathcal{P}$ can be obtained. To establish this result, we require the following lemma.

LEMMA 4.5. *Let X and Y be normed linear spaces and let C be a nonempty closed convex subset of Y. Moreover, let $\overline{x} \in X$, $f : X \to \overline{\mathbb{R}}$, and $g : X \to Y$ be such that $\partial f(\overline{x}) \neq \emptyset$, f is subdifferentially regular at $\overline{x}$, $g(\overline{x}) \in C$, and g is strictly differentiable at $\overline{x}$. If the set* K-T $(\overline{x})$ *is nonempty, then*

$$\begin{aligned}\{d \in X : P_\alpha^0(\overline{x}; d) \leq 0\} &= \{d \in X : f^0(\overline{x}; d) = 0, g_s'(\overline{x})d \in T(g(\overline{x})|C)\} \\ &=: D_{\mathcal{P}}(\overline{x})\end{aligned}$$

for all $\alpha >$ $\mathrm{dist}_0(0|K\text{-}T(\overline{x}))$.

Proof. The hypotheses and Rockafellar [68, Thms. 2 and 3] imply that

$$\partial P_\alpha(\overline{x}) = \partial f(\overline{x}) + \alpha g_s'(\overline{x})^*(\mathbb{B}^0 \cap N(g(\overline{x})|C)).$$

Thus, if $\alpha >$ $\mathrm{dist}_0(0|\text{K-T}(\overline{x}))$, then clearly

$$D_{\mathcal{P}}(\overline{x}) \subset \{d \in X : P_\alpha^0(\overline{x}; d) \leq 0\}.$$

On the other hand, let $d \in \{d : P_\alpha^0(\overline{x}; d) \leq 0\}$. Then for each $y \in$ K-T $(\overline{x})$ with $\|y\|_0 < \alpha$ there is a $z \in \partial f(\overline{x})$ such that $0 = z + g_s'(\overline{x})^* y$. Hence,

$$\begin{aligned}0 &\geq P_\alpha^0(\overline{x}; d) \\ &\geq \langle z + \alpha g_s'(\overline{x})^* \frac{y}{\|y\|_0}, d\rangle \\ &= (1 - \frac{\alpha}{\|y\|_0})\langle z, d\rangle,\end{aligned}$$

and so $f^0(\overline{x}; d) \geq \langle z, d\rangle \geq 0$. But $0 \in N(g(\overline{x})|C) \cap \mathbb{B}^0$ so that $f^0(\overline{x}; d) \leq 0$. Consequently, $f^0(\overline{x}; d) = 0$ and $g_s'(\overline{x})d \in T(g(\overline{x})|\mathbb{B})$. □

Remark. The set $D_{\mathcal{P}}(\overline{x})$ given above is the obvious generalization of the set defined in (4.10) to which it reduces under the hypotheses of Theorem 4.4.

THEOREM 4.6. *Let S and C be nonempty closed convex subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, and let $\overline{x} \in S$ be such that $f_0 : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are twice continuously differentiable near $\overline{x}$, $g(\overline{x}) \in C$, and the basic constraint qualification (2.7) holds. Set $f := f_0 + \psi(\cdot|S)$ and consider the problem $\mathcal{P}$.*

(1) *If $\overline{x}$ is a local solution to $\mathcal{P}$, then*

$$\max\{d^T(\nabla^2 f_0(\overline{x}) + \nabla^2_{xx}(\langle y, g(\overline{x})\rangle))d : y \in K\text{-}T(\overline{x})\} \geq 0 \tag{4.13}$$

for all $d \in D_{\mathcal{P}}(\overline{x})$.

(2) *If $K\text{-}T\ (\overline{x}) \neq \emptyset$ and*

$$\max\{d^T(\nabla^2 f_0(x) + \nabla^2_{\overline{x}x}(\langle y, g(\overline{x})\rangle)))d : y \in K\text{-}T(\overline{x})\} > 0 \tag{4.14}$$

for all $d \in D_{\mathcal{P}}(\overline{x})\backslash\{0\}$, then for each

$$\alpha > \overline{\alpha} := \max\{\|y\|_0 : y \in K\text{-}T(\overline{x})\} \tag{4.15}$$

there are scalars $\varepsilon > 0$ and $\gamma > 0$ such that

$$P_\alpha(x) \geq P_\alpha(\overline{x}) + \gamma\|x - \overline{x}\|^2$$

for all $x \in \overline{x} + \varepsilon\mathbb{B}$, and

$$f_0(x) \geq f_0(\overline{x}) + \gamma\|x - \overline{x}\|^2$$

for all $x \in \Omega \cap (\overline{x} + \varepsilon\mathbb{R})$ where

$$\Omega := \{x \in S : g(x) \in C\}.$$

Proof. In $\mathcal{Q}$ take $h := \alpha\ \text{dist}(\cdot|C)$. Then, by Proposition 4.3 and Lemma 4.5,

$$D_{\mathcal{Q}}(\overline{x}) = \text{cl}(K_{\mathcal{Q}}(\overline{x})) \cap T(\overline{x}|S) = D_{\mathcal{P}}(\overline{x}) \tag{4.16}$$

as long as $\alpha > \|y\|_0$ for some $y \in \text{K-T}(\overline{x})$. Moreover, by part (1) of Proposition 3.2, the set K-T $(\overline{x})$ is compact. Hence if K-T $(\overline{x}) \neq \emptyset$, then $\overline{\alpha}$ is finite and for any $\alpha \geq \overline{\alpha}$ one has

$$M_{\mathcal{Q}}(\overline{x}) = \text{K-T}(\overline{x}). \tag{4.17}$$

(1) By Corollary 2.5.1, there is an $\alpha > 0$ such that $\overline{x}$ is a local minimum for P_α. Taking $h := \alpha\ \text{dist}(\cdot|C)$ in $\mathcal{Q}$, we get $M_{\mathcal{Q}}(\overline{x}) \subset \text{K-T}(\overline{x})$, where α is chosen so that $\alpha >\ \text{dist}_0(0|\text{K-T}(\overline{x}))$. The result then follows from (4.16) and part (1) of Theorem 4.2.

(2) By taking $\alpha > 0$ to satisfy (4.15) and by observing (4.16) and (4.17), the result is an immediate consequence of part (2) of Theorem 4.2 with $h := \alpha\ \text{dist}(\cdot|C)$. □

Remark. For the case in which C and S are polyhedral convex, Theorem 4.6 is also obtained by Ioffe [39]–[41], Ben–Israel, Ben–Tal, and Zlobec [4], and Rockafellar [63]–[64].

In Theorem 4.6 we obtain second-order necessary and sufficient conditions for $\mathcal{P}$ from the corresponding second-order conditions for P_α. This approach is the reverse of that which is usually taken in the literature. In particular, Charalambous [13, Thm. 2], Han and Mangasarian [33, Thm. 4.6], and Lasserre [44, Thm. 2] essentially show that if K-T $(\overline{x}) \neq \emptyset$ and the second-order sufficiency condition of Pennisi [57, Thm. 3.3] holds for some $y \in$ K-T$(\overline{x})$, then $\overline{x}$ is a strict local minimum for P_α for all $\alpha > \|y\|_0$. These results do not require the imposition of the basic constraint qualification (2.7). On the other hand, they do require the application of a stronger second-order sufficiency condition. In the next result, we obtain a result, paralleling those of Charalambous, Han and Mangasarian, and Lasserre.

THEOREM 4.7. *Let S, C, $\overline{x}$, f_0, and g be as in the statement of Theorem* 4.6, *except that the basic constraint qualification may fail to hold at $\overline{x}$. If there exists $\overline{y} \in K\text{-}T(\overline{x})$ such that*

$$d^T(\nabla^2 f_0(\overline{x}) + \nabla^2_{xx}(\langle \overline{y}, g(\overline{x})\rangle))d > 0$$

for every $d \in D_{\mathcal{P}}(\overline{x})\backslash\{0\}$, then for each $\alpha > \|y\|_0$ there are scalars $\varepsilon > 0$ and $\gamma > 0$ such that

$$P_\alpha(x) \geq P_\alpha(\overline{x}) + \gamma\|x - \overline{x}\|^2$$

for all $x \in \overline{x} + \varepsilon\mathbb{B}$ and

$$f_0(x) \geq f_0(\overline{x}) + \gamma\|x - \overline{x}\|^2$$

for all $x \in \Omega \cap (\overline{x} + \varepsilon\mathbb{B})$.

Proof. For this choice of α (4.16) holds and by part (3) of Proposition 2.6, $0 \in \partial P_\alpha(\overline{x})$ with $y \in M_{\mathcal{Q}}(\overline{x})$ where $h := \alpha\ \mathrm{dist}(\cdot|C)$. Hence, the result again follows directly from part (2) of Theorem 4.2. □

Before leaving this section we obtain yet another sufficiency result for $\mathcal{P}$. It is a first-order sufficiency result and is a direct consequence of Lemma 4.5. The result is similar to results by Howe [37], Rosenberg [71, Thm. 3], and Bazaraa and Goode [3, Thms. 2.1, 2.2, 3.1, and 4.1].

THEOREM 4.8. *Let X, Y, $\overline{x}$, f, and g be as in the statement of Lemma* 4.5 *where it is further assumed that X is finite-dimensional. If the set $K\text{-}T\ (\overline{x})$ is nonempty and $D_{\mathcal{P}}(\overline{x}) = \{0\}$, then there are scalars $\varepsilon > 0$ and $\gamma > 0$ such that*

$$P_\alpha(x) \geq P_\alpha(\overline{x}) + \gamma\|x - \overline{x}\|$$

for all $x \in (\overline{x} + \varepsilon\mathbb{B})$ and $\alpha >$ $\mathrm{dist}_0(0|K\text{-}T(\overline{x}))$, and

$$f(x) \geq f(\overline{x}) + \gamma\|x - \overline{x}\|$$

for all $x \in \{x : g(x) \in C\} \cap (\overline{x} + \varepsilon\mathbb{B})$.

Proof. From Lemma 4.5, $P^0_\alpha(\overline{x}; d) > 0$ for all $d \neq 0$. By Rockafellar [68, Thms. 2 and 3], $P^0_\alpha(\overline{x}; d) = P'_\alpha(\overline{x}; d)$. The result now easily follows with $\gamma = \inf\{P'_\alpha(\overline{x}; d) : \|d\| = 1\} > 0$. □

5. Convex programming. Eremin and Zangwill originated the study of exact penalization in the context of convex programming. In this section, we extend this theory to the problem $\mathcal{P}$. The step in this process is to establish an equivalence between the problem $\mathcal{P}$ and a problem $\widetilde{\mathcal{P}}$ to which the classical theory of convex programming applies [35], [36], [38], [43], [46], [48], [51], [70], [72]. To this end, let X be a real normed linear space, Y a real reflexive Banach space, $C \subset X$ and $S \subset X$ be nonempty, closed, and convex, and set

$$\widetilde{C} := \mathrm{cl}\{(\lambda, \lambda y) : \lambda \geq 0, y \in C\},$$

where the closure is taken with respect to the product topology on $\mathbb{R} \times Y$.

Given $g : X \to Y$ we define $G : X \to \mathbb{R} \times Y$ by

$$G(x) := (-1, -g(x))$$

for all $x \in X$. Consider the constrained optimization problem

$$(\widetilde{\mathcal{P}}) \qquad \begin{array}{l} \text{minimize } f(x) \\ \text{subject to } G(x) \leq 0, \end{array}$$

where $f := f_0 + \psi(\cdot|S)$ with $f_0 : X \to \mathbb{R}$ a convex function and where "$\leq$" denotes the partial order induced on $\mathbb{R} \times Y$ by $\widetilde{C}$, i.e., $\underset{\sim}{y_1} \leq y_2$ if and only if $y_2 - y_1 \in \widetilde{C}$. Observe that $x \in X$ solves $\mathcal{P}$ if and only if x solves $\widetilde{\mathcal{P}}$. We now develop a purely convex theory for $\mathcal{P}$ based upon that which already exists for $\widetilde{\mathcal{P}}$.

LEMMA 5.1. *Let $G : X \to \mathbb{R} \times Y$ and $\widetilde{C} \subset \mathbb{R} \times Y$ be as given above. Then the following conditions are equivalent.*

(1) *G is convex with respect to $\widetilde{C}$; i.e., $G(\lambda x + (1-\lambda)y) \leq \lambda G(x) + (1-\lambda)G(y)$ for every $x, y \in X$ and $\lambda \in [0,1]$.*

(2) *g is concave with respect to* $\text{rec}(C)$; *i.e.,* $g(\lambda x + (1-\lambda)y) - [\lambda g(x) + (1-\lambda)g(y)] \in \text{rec}(C)$ *for every $x, y \in X$ and $\lambda \in [0,1]$.*

(3) *For each $y \in \text{bar}(C)$ the mapping $g_y : X \to \mathbb{R}$, given by $g_y(\cdot) := \langle y, g(\cdot)\rangle$, is convex.*

Moreover, each of the above conditions imply that the distance function $\text{dist}(g(\cdot)|C)$ *is convex.*

Proof. (1) $\iff$ (2): Let $x_1, x_2 \in X$ and choose $\lambda \in [0,1]$. Then G is convex with respect to $\widetilde{C}$ if and only if

$$\lambda(-1, -g(x_1)) + (1-\lambda)(-1, -g(x_2)) - (-1, -g(\lambda x_1 + (1-\lambda)x_2)) \in \widetilde{C},$$

or equivalently

$$g(\lambda x_1 + (1-\lambda)x_2) - [\lambda g(x_1) + (1-\lambda)g(x_2)] \in \text{rec}(C)$$

since $\text{rec}(C) = \{y : (0, y) \in \widetilde{C}\}$. This is equivalent to saying that g is concave with respect to $\text{rec}(C)$.

(2) $\iff$ (3): The mapping g is concave with respect to $\text{rec}(C)$ if and only if for every $x_1, x_2 \in X$ and $\lambda \in [0,1]$

$$g(\lambda x_1 + (1-\lambda)x_2) - [\lambda g(x_1) + (1-\lambda)g(x_2)] \in (\text{bar}(C))^0$$

since $[\text{bar}(C)]^0 = \text{rec}(C)$. This is equivalent to saying that

$$\langle y, g(\lambda x_1 + (1-\lambda)x_2)\rangle \leq \langle y, \lambda g(x_1) + (1-\lambda)g(x_2)\rangle$$

for every $x_1, x_2 \in X$, $\lambda \in [0,1]$, and $y \in \text{bar}(C)$; i.e., $\langle y, g(\cdot)\rangle$ is convex for every $y \in \text{bar}(C)$.

Finally, if any one of (1)–(3) hold, then clearly (3) is valid. Hence, for every $y \in \text{bar}(C) = \text{dom}(\psi^*(\cdot|C))$ the function

$$\langle y, g(\cdot)\rangle - \psi^*(y|C)$$

is convex. Therefore,

$$\operatorname{dist}(g(\cdot)|C) := \sup\{\langle y, g(\cdot)\rangle - \psi^*(y|C) : y \in \mathbb{B}^0\}$$

is convex since it is the supremum of a collection of convex functions. □

Remark. If C is bounded, then g is concave with respect to the $\operatorname{rec}(C)$ if and only if g is affine, and if $C := \mathbb{R}^s_- \times \{0\}_{\mathbb{R}^{m-s}}$, then g is concave with respect to $\operatorname{rec}(C)$ if and only if g_i is convex for $i = 1, \cdots, s$ and g_i is affine for $i = s+1, \cdots, m$.

LEMMA 5.2. *Let $\widetilde{L} : S \times \widetilde{C}^* \to \mathbb{R}$ be the standard Lagrangian for $\widetilde{\mathcal{P}}$ where $\widetilde{C}^* := -\widetilde{C}^0$, i.e.,*

$$\widetilde{L}(x, z) := f(x) + \langle z, G(x)\rangle,$$

and define $L : S \times Y^* \to \mathbb{R}$ by

$$L(x, y) := f(x) + \langle y, g(x)\rangle - \psi^*(y|C).$$

Suppose that $f := f_0 + \psi(\cdot|S)$ with $f_0 : X \to \mathbb{R}$ convex and $g : X \to Y$ is concave with respect to $\operatorname{rec}(C)$ so that both $\widetilde{L}$ and L are convex–concave saddle functions by the previous lemma. Then $(x_0, (\xi_0, -y_0)) \in S \times \widetilde{C}^*$ is a saddle point for $\widetilde{L}$ if and only if (x_0, y_0) is a saddle point for L in which case $\xi_0 = \psi^*(y_0|C)$, $y_0 \in N(g(x_0)|C)$, and $g(x_0) \in C$.

Proof. By direct computation we verify that

$$\widetilde{C}^* := \{(\xi, -y) | (\xi, y) \in \operatorname{epi}(\psi^*(\cdot|C))\}.$$

If $(x_0, (\xi_0, -y_0))$ is a saddle point for $\widetilde{L}$, then, in particular, $x_0 \in S$ and

$$\langle(\xi, -y), (-1, -g(x_0))\rangle \le \langle(\xi_0, -y_0), (-1, -g(x_0))\rangle$$

for every $(\xi, y) \in \operatorname{epi}(\psi^*(\cdot|C))$, or equivalently,

$$\langle y, g(x_0)\rangle - \xi \le \langle y_0, g(x_0)\rangle - \xi_0 \tag{5.1}$$

for every $(\xi, y) \in \operatorname{epi}(\psi^*(\cdot|C))$. But this can occur if and only if

$$\xi_0 = \psi^*(y_0|C), g(x_0) \in C, \langle y_0, g(x_0)\rangle = \psi^*(y_0|C),$$

and

$$L(x_0, y) \le L(x_0, y_0)$$

for every $y \in \operatorname{bar}(C)$. To see this, set $y = y_0$ in (5.1) to get $\xi_0 = \psi^*(y_0|C)$, next set $y = y_0 + z$ in (5.1) to get $\langle z, g(x_0)\rangle \le \psi^*(z|C)$ for all $z \in \operatorname{bar}(C)$, and so $g(x_0) \in C$. Finally, having $g(x_0) \in C$ we obtain from (5.1) that

$$0 \le \langle y_0, g(x_0)\rangle - \psi^*(y_0|C) \le 0.$$

The reverse implication is obvious.

By employing the fact that $\xi_0 = \psi^*(y_0|C)$, $g(x_0) \in C$, and $\psi^*(y_0|C) = \langle y_0, g(x_0)\rangle$, we obtain from the other half of the saddle point inequalities for $\widetilde{L}$ that

$$L(x_0, y_0) = \widetilde{L}(x_0, (\xi_0, -y_0)) \le f(x) + \langle y_0, g(x)\rangle - \psi^*(y_0|C)$$

for every $x \in S$, or equivalently,

$$L(x_0, y_0) \leq L(x, y_0)$$

for all $x \in S$ whereby the lemma is established. $\square$

Having obtained the equivalence of the saddle point conditions for $\widetilde{L}$ and L, we can now simply translate the saddle point results for $\widetilde{\mathcal{P}}$ into similar results for $\mathcal{P}$. In this way, we obtain the following two results from [46, Cor. 1, p. 219] and [46, Thm. 2, p. 221], respectively.

THEOREM 5.3. *Let X be a real normed linear space and Y a real reflexive Banach space, let $S \subset X$ and $C \subset Y$ be nonempty, closed, and convex, and suppose that $f := f_0 + \psi(\cdot|S)$ with $f_0 : X \to \mathbb{R}$ convex, and $g : X \to Y$ is concave with respect to* $\text{rec}(C)$, *and there is an $x \in S$ such that $g(x) \in \text{int}(C)$. If $\overline{x}$ solves $\mathcal{P}$, then there is a $\overline{y} \in N(g(\overline{x})|C)$ such that $(\overline{x}, \overline{y})$ is a saddle point for $L(x, y)$.*

Remark. If X and Y are finite-dimensional, then we need only assume that there is an $x \in S$ such that $g(x) \in \text{ri}(C)$.

THEOREM 5.4. *Let X, Y, S, C, g, and f be as in the statement of Theorem* 5.3. *If there exists an $\overline{x} \in S$ and $\overline{y} \in \text{bar}(C)$ such that $(\overline{x}, \overline{y})$ is a saddle point for $L(x, y)$, then $\overline{x}$ solves $\mathcal{P}$.*

Further results of this type can also be obtained. Theorems 5.3 and 5.4 are presented only to give the flavor of what can be said in the convex case. In this setting, the most natural notion of a Kuhn–Tucker multiplier is derived from that of a saddle point of L. Thus, for the convex case, we extend the definition of K-T (x) as follows;

$$\text{K-T}(x) := \{y \in \text{bar}(C) : (x, y) \text{ is a saddle point for } L\}.$$

Our primary result on exact penalization in the convex case now follows.

THEOREM 5.5. *Let X, Y, S, C, f, and g be as in the statement of Theorem* 5.3, *let $\overline{x} \in S$, and consider the following two conditions:*

(A) *f is continuous near $\overline{x}$ and g is strictly differentiable at $\overline{x}$.*

(B) *X and Y are finite-dimensional and g is Lipschitz near $\overline{x}$.*

The following statements are equivalent:

(1) *$\mathcal{P}$ is calm at $\overline{x}$.*

(2) *$\overline{x}$ is a global minimum of P_α for all α sufficiently large.*

Moreover, if either (A) *or* (B) *holds, then* (1) *and* (2) *are equivalent to*

(3) *K-T $(\overline{x}) \neq \emptyset$.*

Furthermore, given $\overline{y} \in$ K-T$(\overline{x})$, then $\overline{x}$ is a global minimum for P_α for all $\alpha \geq \|\overline{y}\|_0$ and if $\alpha >$ $\text{dist}_0(0|\text{K-T}(\overline{x}))$, then

$$\arg\min\{P_\alpha(x) : x \in X\} \subset \arg\min\{f(x) : g(x) \in C\}.$$

Proof. By Lemma 5.1, P_α is a convex function for all $\alpha \geq 0$; consequently, any local minimum of P_α is a global minimum of P_α. Therefore, the equivalence of (1) and (2) is a consequence of Theorem 2.1.

The proof that (3) is equivalent to (1) and (2) is essentially identical under the two hypotheses (A) and (B), except that we use [66, Cor. 5.2.3] in the finite-dimensional case and [68, Thms. 2 and 3] in the infinite-dimensional case. Hence, we provide the proof only when (A) is assumed.

We begin by assuming (2) and showing that (3) holds. From (2) there is a $\overline{y} \in \partial \operatorname{dist}(\cdot|C)(g(\overline{x}))$ with $0 \in \partial f(\overline{x}) + \alpha g_s'(\overline{x})^* \overline{y}$, or equivalently, there is a $\overline{y} \in N(g(\overline{x})|C)$ such that

$$0 \in \partial_x L(\overline{x}, \overline{y}),$$

since $\partial \operatorname{dist}(\cdot|C)(y(\overline{x})) = \mathbb{B}^0 \cap N(g(\overline{x})|C)$ by Proposition 3.1. Consequently,

$$L(\overline{x}, \overline{y}) \leq L(x, \overline{y})$$

for all $x \in X$, since $L(x, \overline{y})$ is convex in x by Lemma 5.1. Finally, since

$$\begin{aligned} \psi(g(\overline{x})|C) &= \langle \overline{y}, g(\overline{x}) \rangle - \psi^*(\overline{y}|C) \\ &= \sup\{\langle y, g(\overline{x}) \rangle - \psi^*(y|C) : y \in Y\}, \end{aligned}$$

we have that

$$L(\overline{x}, y) \leq L(\overline{x}, \overline{y})$$

for all $y \in Y$.

Next we assume that (3) holds and establish (2). Since $L(\overline{x}, y) \leq L(\overline{x}, \overline{y})$ for all $y \in Y$ we know that

$$0 = \psi(g(\overline{x})|C) = \langle \overline{y}, g(\overline{x}) \rangle - \psi^*(\overline{y}|C).$$

Next, let $x \in S$ and choose $\alpha \geq \|\overline{y}\|_0$, then

$$\begin{aligned} P_\alpha(\overline{x}) &= L(\overline{x}, \overline{y}) \\ &\leq L(x, \overline{y}) \\ &\leq \sup\{L(x, y) : y \in \alpha\mathbb{B}\} \\ &= f(x) + \alpha \sup\{\langle y, g(x) \rangle - \psi^*(y|C) | y \in \mathbb{B}^0\} \\ &= P_\alpha(x). \end{aligned}$$

Hence $\overline{x}$ is a global minimum for $P_\alpha(x)$ for all $\alpha \geq \|\overline{y}\|_0$.

To prove the last statement of the theorem choose $\overline{y} \in \text{K-T}(\overline{x})$ such that $\alpha > \|\overline{y}\|_0$. Setting $\overline{\alpha} = \|\overline{y}\|_0$, we know that $0 \in \partial P_{\overline{\alpha}}(\overline{x})$ and $0 \in \partial P_\alpha(\overline{x})$ so that $\overline{x}$ is a global minimum for both $P_{\overline{\alpha}}$ and P_α. Thus, in particular, $\arg\min\{P_\alpha(x) : x \in X\} \neq \emptyset$. Let $\widetilde{x} \in \arg\min\{P_\alpha(x) : x \in X\}$, we need to show that $\widetilde{x} \in \arg\min\{f(x) : g(x) \in C\}$. For this, it is sufficient to show that $f(\widetilde{x}) \leq f(\overline{x})$ and $g(\widetilde{x}) \in C$. Due to the nature of $\overline{x}$ and $\widetilde{x}$ we have

$$f(\widetilde{x}) + \alpha \operatorname{dist}(g(\widetilde{x})|C) \leq f(\overline{x}) + \alpha \operatorname{dist}(g(\overline{x})|C)$$

and

$$f(\overline{x}) + \overline{\alpha} \operatorname{dist}(g(\overline{x})|C) \leq f(\widetilde{x}) + \overline{\alpha} \operatorname{dist}(g(\widetilde{x})|C).$$

By adding these inequalities we find that

$$(\alpha - \overline{\alpha}) \operatorname{dist}(g(\widetilde{x})|C) \leq (\alpha - \overline{\alpha}) \operatorname{dist}(g(\overline{x})|C).$$

Hence $g(\widetilde{x}) \in C$ and $f(\widetilde{x}) \leq f(\overline{x})$. □

Remark. The form of Theorem 5.5 is based on Rosenberg [71, Thm. 2]. This result extends similar results appearing in Eremin [23], Zangwill [75], Pietrzykowski

[58], Luenberger [47], Charalambous [13], Han and Mangasarian [33], Lasserre [44], Garcia-Palomares [31], Rosenberg [71], and Bertsekas [6].

6. Historical review. In this section we attempt to provide a chronology of those results that establish the existence of an exact penalty parameter. We apologize for any omission or oversight.

It seems apparent that the big-M method for linear programming is the precursor of exact penalization techniques for nonlinear programming, especially since the initial results were obtained for the convex programming case. However, we are uncertain that this was indeed the motivation. Our earliest reference for the big-M method is Charnes, Cooper, and Anderson [14, §4]. The precise origins of the method are unknown to us. Our earliest reference for exact penalization in nonlinear programming is Eremin [23]. In this paper, Eremin considers the case of convex programming with $C = \mathbb{R}^s_- \times \{0\}_{\mathbb{R}^{m-s}}$ and $S = \mathbb{R}^n$. In [23, Thm. 2], he shows that if $\overline{y} \in$ K-T$(\overline{x})$, then $\overline{x}$ is a global minimum for P_α whenever $\alpha > \|y\|_0$ when $\mathbb{R}^m$ is endowed with the ℓ_1 norm. At essentially the same time, Zangwill [75] published his well-known paper. Zangwill considered the case of convex programming with $C = \mathbb{R}^m_-$ and $S = \mathbb{R}^n$ and showed that if $\overline{x}$ solved $\mathcal{P}$ and $g(x_0) \in \text{int}(C)$, then $\overline{x}$ minimized P_α for all $\alpha > (f(x_0) - f(\overline{x}) + 1)(\max[g_i(x_0) : i = 1, \cdots, m])^{-1}$. This result can be used to show that K-T $(\overline{x}) \neq \emptyset$, and so is somewhat deeper than Eremin's result.

Pietrzykowski [58] provides the result for the nonconvex case. He considers the instance of $\mathcal{P}$ where $C = \mathbb{R}^s \times \{0\}_{\mathbb{R}^{m-s}}$ and $S = \mathbb{R}^n$. The analysis that Pietrzykowski gives a reminiscent of Zangwill's. He shows that if $\overline{x}$ is a strict local minimum for $\mathcal{P}$, near which f and g are differentiable and at which $g'(\overline{x})$ is surjective, then $\overline{x}$ is a strict local minimum for P_α for all α sufficiently large. Pietrzykowski's result can be used to show that K-T $(\overline{x}) \neq \emptyset$ under these hypotheses.

Luenberger [47] considers exact penalization in the setting of optimal control. We interpret his result as it applies to $\mathcal{P}$. In this context, Luenberger has $C = \mathbb{R}^m_-$ and $S = \mathbb{R}^n$ and assumes that $\overline{x}$ is a local minimum for $\mathcal{P}$ at which there exists a $\overline{y} \in N(g(\overline{x})|C)$ such that $\overline{x}$ is a local minimum for $L(x, \overline{y})$. Under these circumstances Luenberger shows that $\overline{x}$ is a local minimum for P_α for all $\alpha \geq \|\overline{y}\|_0$ where $\mathbb{R}^m$ is endowed with the ℓ_1 norm. Luenberger's proof is the same as that provided by Eremin. Clearly, Luenberger's result applies in the convex case subject to the appropriate constraint qualification, but it can also be applied to cases in which the second-order sufficiency condition of Pennisi [57] holds. Luenberger himself only states that this result applies "under standard regularity conditions."

Evans, Gould, and Tolle [24] consider the case where $C = \mathbb{R}^m_-$, $S = \mathbb{R}^n$, and f and g are continuously differentiable. In this context, their nondifferentiable exact penalty functions are quite different from the Eremin–Zangwill exact penalty functions. For these new functions they provide some exactness results that are similar in spirit to those of Eremin, Zangwill, and Pietrzykowski.

Howe [37] considers the case in which $C = \mathbb{R}^s_- \times \{0\}_{\mathbb{R}^{m-s}}$, $S = \mathbb{R}^n$, and f and g are continuously differentiable. His result is the appearance of the type of sufficiency result given in Theorem 4.8. He shows that if $D_{\mathcal{P}}(\overline{x}) = \{0\}$, then $\overline{x}$ is a local minimum for P_α for all α sufficiently large.

Bandler and Charalambous [2] consider the same case as Evans, Gould, and Tolle [24] and derive yet another type of nondifferentiable exact penalty function. For this exact penalty function they provide an exactness result that is similar in spirit to those of Eremin, Zangwill, and Pietrzykowski.

Bertsekas [6] investigates the case of convex programming with $C = \mathbb{R}^m_-$ and

$X = \mathbb{R}^n$, and establishes necessary and sufficient conditions for a function of the form

$$\pi(x) = f(x) + \sum_{i=1}^{m} p_i(g_i(x))$$

to be exact for $\mathcal{P}$. If $(\overline{x}, \overline{y})$ is a saddle point for $L(x, y)$ he shows that

$$\lim_{t \to 0^+} \frac{p_i(t)}{t} \geq \overline{y}^{(i)} \quad i = 1, \cdots, m,$$

with

$$\arg\min\{\pi(x)\} = \arg\min\{f(x) : g(x) \in C\},$$

if

$$\lim_{t \to 0^+} \frac{p_i(t)}{t} > \overline{y}^{(i)}.$$

We obtain Eremin's result as a special case. Bertsekas also applies his result to the exact penalty functions of Evans, Gould, and Tolle.

Charalambous [13] is the first to consider more general norms in the construction of P_α. Specifically, Charalambous considers the case $C = \mathbb{R}^m_-, S = \mathbb{R}^n$ where f and g are continuously differentiable. He then utilizes the ℓ_p-norms to form P_α. Charalambous establishes two key results. In the result, he considers the convex programming case and shows that if $(\overline{x}, \overline{y})$ is a saddle point for $L(x, y)$, then $\overline{x}$ is a global minimum for P_α for all $\alpha > \|\overline{y}\|_\circ$. The proof is similar to those of Eremin and Luenberger. Charalambous' second result is the instance of an exact penalization theorem employing Pennisi's [57] second-order sufficiency conditions. He shows that if the second-order sufficiency condition of Theorem 4.7 is satisfied, then $\overline{x}$ is a local minimum for P_α for all $\alpha > \|\overline{y}\|_\circ$.

Dolecki and Rolewicz [20] present the deepest first-order results for exact penalization currently available in the literature. They consider a model problem that is somewhat more general than the problem $\mathcal{P}$ and obtain exact penalty results based on a more general notion of subdifferential. In this context, they obtain one of the implications in Theorem 2.1 and a version of Corollary 2.3.1. The Dolecki–Rolewicz paper represents the attempt to extend exact penalization techniques to the nondifferentiable case in infinite-dimensions.

Perhaps the most widely referenced paper on exact penalization is by Han and Mangasarian [33]. Their paper is the most comprehensive and comprehensible study of the subject available in the literature. Han and Mangasarian consider the case in which $C = \mathbb{R}^s_- \times \{0\}_{\mathbb{R}^{m-s}}$, $S = \mathbb{R}^n$, and f and g are continuously differentiable. One of the most significant contributions of their paper is the relaxation of the first-order conditions under which an exact penalty parameter for $\mathcal{P}$ exists. Specifically, they show that if the Mangasarian–Fromovitz constraint qualification is satisfied at a strict local solution $\overline{x}$ to $\mathcal{P}$, then there exists an $\overline{\alpha} \geq 0$ such that $\overline{x}$ is a local solution to P_α for all $\alpha \geq \overline{\alpha}$. They establish this result for an arbitrary norm by appealing to the equivalence of norms in finite dimensions. This result is an instance of Corollary 2.4.1 (however, Corollary 2.4.1 does not require that $\overline{x}$ be a strict local solution). They also provide a second-order result that is similar to that of Charalambous. Moreover, they establish the equivalence of stationarity conditions for $\mathcal{P}$ and the minimization of P_α, as is done in Proposition 6.2. They conclude by again establishing Eremin's result for the case of convex programming. The penalty functions they consider are a generalization of the Eremin–Zangwill penalty functions and are based on the work of Bertsekas.

The work of Lasserre [44] appears soon after that of Han and Mangasarian. He considers the case in which $C := \mathbb{R}^s_- \times \{0\}_{\mathbb{R}^{m-s}}$, $S = \mathbb{R}^n$, f and g are continuously differentiable, and $\mathbb{R}^m$ is endowed with a weighted ℓ_1 norm. In this case, he establishes a second-order result similar to that of Charalambous. Moreover, he shows that if the active constraint gradients are linearly independent at a local solution to $\mathcal{P}$, then an exact penalty parameter exists for $\mathcal{P}$. This result is different from the corresponding result of Han and Mangasarian, and Pietrzykowski since Lasserre does not assume that the solution is a strict local minimum. Nonetheless, it appears that both of these results are subsumed in the work of Dolecki and Rolewicz. Lasserre also recaptures and extends the result of Luenberger by recognizing the relationship between saddle points of the Lagrangian and local minimum of the exact penalty function.

Fletcher [29, §14.3] considers the same situation as Lasserre. Under the hypothesis that the active constraint gradients are linearly independent, Fletcher [29] is the to recognize the actual equivalence of the first- and second-order optimality conditions for $\mathcal{P}$ and the exact penalty function P_α. Consequently, Fletcher's work is a direct precursor of the results presented in this paper.

Bazaraa and Goode [3] consider the case where $C = \mathbb{R}^m_-$, S is closed, and f_0 and g are continuously differentiable. They establish some extensions to Howe's result using some of the modern techniques of nonsmooth analysis. Moreover, by assuming that S is compact, they obtain global versions of Howe's theorem and give estimates for the value of an exact penalty parameter that are reminiscent of those established by Zangwill.

In [15] Clarke establishes his elementary exact penalization result for the case in which the inclusion constraint $g(x) \in C$ is absent. This result is one of the corner stones of §2 and appears as Theorem 2.3. Clarke's proof should be reviewed by every student of this subject. It is very elementary, requiring only seven short sentences. Clarke also shows that calmness implies the existence of an exact penalty parameter for $\mathcal{P}$ when $C := \mathbb{R}^s_- \times \{0\}_{\mathbb{R}^{m-s}}$.

In [59], Polak, Mayne, and Wardi consider the case where $C = \mathbb{R}^s_- \times \{0\}_{\mathbb{R}^{m-s}}$, $S = \mathbb{R}^n$, f and g_i, $i = 1, \cdots, s$ are locally Lipschitz, and g_i, $i = s+1, \cdots, m$ are continuously differentiable. In this setting, they establish the equivalence of the stationarity conditions for $\mathcal{P}$ and the minimization of P_α for all α sufficiently large. This result is generalized in Proposition 2.6.

Rosenberg [71] considers the case in which $C = \mathbb{R}^s_- \times \{0\}_{\mathbb{R}^{m-s}}$, $S = \mathbb{R}^n$, and f and g are locally Lipschitz functions. He begins by providing local and strict local versions of Clarke's result that calmness implies the existence of an exact penalty parameter. He then reviews the convex programming case and establishes the version of Theorem 4.8, upon which our treatment is based. Rosenberg concludes his study by extending Howe's result to the Lipschitzian case where he provides results that are substantially more general than those of Bazaraa and Goode. For problems of this type he also provides a sharp lower bound for the value of an exact penalty parameter.

Garcia-Palomares [31] examines the case in which $C = \mathbb{R}^s_- \times \{0\}_{\mathbb{R}^{m-s}}$ $S = \mathbb{R}^n$, f and g are continuously differentiable, and $\mathbb{R}^m$ is endowed with the ℓ_∞ norm. The perspective in this paper is quite similar to the one we have taken. His goal is to establish the equivalence between the first- and second-order optimality conditions for $\mathcal{P}$ and P_α. In this regard, he provides versions of some of the results presented in the latter half of §2 and §3. His approach allows a great deal of further insight in the case of the ℓ_∞-norm.

In [50] Mangasarian considers the convex programming case with $C = \mathbb{R}^s_-$, $S = \mathbb{R}^n$, and f and g continuously differentiable. He extends the analysis of Zangwill to provide lower bounds for the value of an exact penalty function under weaker hypotheses.

Recently Conn and Gould [18, 1987] have generalized the ℓ_1 exact penalty function to obtain an exact penalty function for a class of semi-infinite programming problems. They consider both the convex and nonconvex cases, and their results are not covered by those presented in this paper. These new exact penalty functions for semi-infinite programming are quite interesting and deserve much further study.

In [64] Rockafellar studies the case in which $C \subset \mathbb{R}^m$ is the product of intervals, $X \subset \mathbb{R}^n$ is polyhedral convex, and $f_0(x) := \max\{f_{0_j}(x) : j = 1, \cdots, s\}$ where $f_{0_j}, j = 1, \cdots, s$ and g are all continuously differentiable. As in our study, Rockafellar derives the equivalence of first- and second-order optimality conditions for $\mathcal{P}$ and P_α via similar results for convex composite optimization. However, Rockafellar's results rely on the piecewise linear–quadratic case, the theory that he develops in [63].

We conclude by offering our apologies to the many authors we have not mentioned, especially to those who have made significant contributions in the domain of algorithmic development.

A. Appendix. We proceed to establish Theorem 2.5. For this purpose we will need the following lemmas from Burke and Poliquin [11].

LEMMA A1. *Let* $q : \mathbb{R}^n \to \overline{\mathbb{R}}$ *be as given in Theorem* 2.5. *If* $\overline{x} \in \text{dom}(q)$ *is such that* (2.9) *holds, then there is a neighborhood* U *of* $\overline{x}$ *such that* (2.9) *is satisfied at every point of* $\text{dom}(q) \cap U$.

Proof. This is a direct consequence of the upper semicontinuity of $\partial f, \partial f^\infty$, ∂g, and $N(\cdot|\text{dom}(h))$. □

LEMMA A2. *Let* $h : \mathbb{R}^m \to \overline{\mathbb{R}}$ *be as in Theorem* 2.5 *and let* $\{(y_i, z_i)\} \subset$ *graph* (∂h) *be such that* $y_i \to y \in \text{dom}(h)$ *and* $\|z_i\| \uparrow \infty$. *Then every cluster point of the sequence* $\{z_i/\|z_i\|\}$ *is an element of the normal cone to* $\text{dom}(h)$ *at* y.

LEMMA A3. *Let* $h : \mathbb{R}^m \to \overline{\mathbb{R}}$ *and* $h_\alpha : \mathbb{R}^m \to \mathbb{R}$ *be as in Theorem* 2.5. *If* $h_\alpha(\overline{y}) = h(\overline{z}) + \alpha\|\overline{y} - \overline{z}\|$, *where* $\overline{z} \in \text{dom}(h)$, *then* $u \in \partial h_\alpha(\overline{y})$ *if and only if* $u \in \partial h(\overline{z}) \cap (\alpha\mathbb{B}^0)$ *and* $(\overline{y} - \overline{z}) \in N(u|\alpha\mathbb{B}^0)$.

The proof of Theorem 2.5 now follows.

Proof. Let ε, $\delta > 0$ be such that $f(x) \geq f(\overline{x})$ for all $x \in \overline{x} + \varepsilon\mathbb{B}$ and (2.9) is satisfied on $\text{dom}(q) \cap (\overline{x} + \delta\mathbb{B})$. Set

$$\xi := 1 + \max\{\|g(x) - g(\overline{x})\| : x \in \overline{x} + \varepsilon\mathbb{B}\},$$

and define

$$\widetilde{h}_\alpha(y) := \inf\{h(z) + \psi(z|g(\overline{x}) + \xi\mathbb{B}) + \alpha\|y - z\| : z \in \mathbb{R}^m\}.$$

Consider the function

$$\widehat{q}_\alpha(x) := \widetilde{q}_\alpha(x) + \varphi(x) + \psi(x|\overline{x} + \varepsilon\mathbb{B}),$$

where $\widetilde{q}_\alpha := f + \widetilde{h}_\alpha \circ g$ and $\varphi(x) := \text{dist}_2^2(x|\overline{x} + \delta\mathbb{B})$. Observe that arg $\min\widehat{q}_\alpha$ is nonempty as $\widehat{q}_\alpha$ is lower semicontinuous and $\overline{x} + \varepsilon\mathbb{B}$ is compact. Hence, there is a sequence $\alpha_i \uparrow \infty$ for which there is a corresponding sequence $\{x_i\} \subset \overline{x} + \varepsilon\mathbb{B}$ converging to some element $\widehat{x}$ of $\overline{x} + \varepsilon\mathbb{B}$ such that

$$x_i \in \text{arg}\ \min\widehat{q}_{\alpha_i}$$

for each $i = 1, 2, \cdots$. Also, from the lower semicontinuity of h and the compactness of $g(\overline{x}) + \xi\mathbb{B}$, there exists for each $i = 1, 2, \cdots$ a y_i in $\text{dom}(h) \cap (g(\overline{x}) + \xi\mathbb{B})$ such that

$$\widetilde{h}_{\alpha_i}(x_i) = h(y_i) + \alpha_i \|y_i - g(x_i)\|.$$

Clearly,

$$q(\overline{x}) \geq \widehat{q}_{\alpha_i}(x_i) \geq \widetilde{q}_{\alpha_i}(x_i). \tag{A.1}$$

Therefore, as $\alpha_i \uparrow \infty$ we have $\|y_i - g(x_i)\| \to 0$ so that $y_i \to g(\widehat{x})$, and thus eventually $y_i \in \text{int}(g(\overline{x}) + \xi\mathbb{B})$, which implies that $\widetilde{q}_{\alpha_i}(x_i) = q_{\alpha_i}(x_i)$. From (A.1) we also obtain that $g(\widehat{x}) \in \text{dom}(h) \cap (g(\overline{x}) + \xi B)$, $\widehat{x} \in \overline{x} + \varepsilon\mathbb{B}$, and

$$q(\overline{x}) \geq q(\widehat{x}) + \varphi(\widehat{x}) \geq q(\widehat{x}).$$

But since $\widehat{x} \in \overline{x} + \varepsilon\mathbb{B}$, the hypotheses imply that $q(\overline{x}) = q(\widehat{x})$ and $\widehat{x} \in \overline{x} + \delta\mathbb{B}$.

We now show that eventually $g(x_i) \in \text{dom}(h)$. Since $x_i \in \arg\min \widehat{q}_{\alpha_i}$ and $x_i \to \widehat{x} \in \overline{x} + \delta\mathbb{B}$, we know that eventually

$$0 \in \partial\widetilde{q}_{\alpha_i}(x_i) + \nabla\varphi(x_i).$$

Hence, by Rockafellar [66, Cor. 5.2.3] and Lemma A.3, eventually there exist $v_i \in \partial f(x_i)$ and $w_i \in \partial\widetilde{h}_{\alpha_i}(g(x_i))$ with

$$w_i \in \partial h(y_i) \text{ and } (g(x_i) - y_i) \in N(w_i | \alpha_i \mathbb{B}^0)$$

(since $N(y_i | g(\overline{x}) + \xi\mathbb{B}) = \{0\}$ as eventually $y_i \in \text{int}(F(\overline{x}) + \xi\mathbb{B})$) such that

$$0 \in v_i + \partial g(x_i)^T w_i + \nabla\varphi(x_i). \tag{A.2}$$

If the sequence $\{(v_i, w_i)\}$ possesses a divergent subsequence $\{(v_i, w_i)\}_J$, then, by Lemma A.2, the sequence $\{(v_i, w_i)/\|(v_i, w_i)\|\}_J$ possesses a cluster point $(\overline{v}, \overline{w})$ with $\overline{v} \in \partial^\infty f(\widehat{x})$, $\overline{w} \in N(g(\widehat{x})|\text{dom}(h))$, and $\|(\overline{v}, \overline{w})\| = 1$. But for such a cluster point $(\overline{v}, \overline{w})$ we obtain from (3.5) that $0 \in \partial^\infty f(\widehat{x}) + \partial g(\widehat{x})^T w$ which contradicts the choice of δ. Thus the sequence $\{(v_i, w_i)\}$ is bounded. Hence for $\overline{\alpha}$ sufficiently large $\{w_i\} \subset \overline{\alpha}\mathbb{B}^0$ so that $N(w_i|\alpha_i\mathbb{B}^0) = \{0\}$ for all i such that $\alpha_i > \overline{\alpha}$. But then $y_i = g(x_i)$ so that $g(x_i) \in \text{dom}(h)$ whenever $\alpha_i > \overline{\alpha}$. Therefore, for all $\alpha_i > \overline{\alpha}$,

$$q(\overline{x}) \geq \widehat{q}_{\alpha_i}(x_i) \geq \widetilde{q}_{\alpha_i}(x_i) = q_{\alpha_i}(x_i) = q(x_i) \geq q(\overline{x}),$$

so that $\overline{x} \in \arg\min \widehat{q}_{\alpha_i}$. Consequently, $\overline{x}$ is also a local minimizer of q_{α_i} for all $\alpha_i > \overline{\alpha}$. □

Remark. The method of proof also shows that if $\overline{x}$ is a strict local minimizer of q, then it is also a strict local minimizer of q_α.

REFERENCES

[1] J. P. AUBIN AND I. EKELAND, *Applied Nonlinear Analysis*, John Wiley, New York, 1984.
[2] J. W. BANDLER AND C. CHARALAMBOUS, *Nonlinear programming using minimax techniques*, J. Optim. Theory Appl., 13 (1974), pp. 607–619.
[3] J. S. BAZARAA AND J. J. GOODE, *Sufficient conditions for a globally exact penalty function without convexity*, Math. Programming Stud., 19 (1982), pp. 1–15.

[4] A. Ben-Israel, A. Ben-Tal, and S. Zlobec, *Optimality in Non-linear Programming, A Feasible Directions Approach*, Wiley–Intersci. Publ., John Wiley, New York, 1981.
[5] D. P. Bertsekas, *Variable metric methods for constrained optimization using differentiable exact penalty functions*, Proc. 18th Annual Allerton Conference on Communication, Control, and Computing, 1980, pp. 584–593.
[6] ———, *Necessary and sufficient conditions for a penalty function to be exact*, Math. Programming, 9 (1975), pp. 87–89.
[7] J. M. Borwein, *Stability and regular points of inequality systems*, J. Optim. Theory Appl., 48 (1986), pp. 9–52.
[8] J. V. Burke, *Calmness and exact penalization*, SIAM J. Control Optim., 1991.
[9] ———, *An exact penalization viewpoint of constrained optimization*, Technical Report ANL/MCS–TM–95, Mathematics and Computer Science Division, Argonne National Laboratories, Argonne, IL, 1987.
[10] ———, *Second order necessary and sufficient conditions for convex composite N.D.O.*, Math. Programming, 38 (1987), pp. 287–302.
[11] J. V. Burke and R. A. Poliquin, *Optimality conditions for nonfinite valued convex composite functions*, Math. Programming, 1991, to appear.
[12] R. M. Chamberlain, C. Lemarechal, H. C. Pederson, and M. J. D. Powel, *The watchdog technique for forcing convergence in algorithms for constrained optimization*, Math. Programming Stud., 16 (1982), pp. 1–17.
[13] C. Charalambous, *A lower bound for the controlling parameter of the exact penalty function*, Math. Programming, 15 (1978), pp. 278–290.
[14] A. Charnes, W. W. Cooper, and A. Henderson, *An Introduction to Linear Programming*, John Wiley, New York, 1953.
[15] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Canad. Math. Soc. Ser. Monographs Adv. Texts, John Wiley, New York, 1983.
[16] ———, *A new approach to lagrange multipliers*, Math. Oper. Res., 1 (1976), pp. 165–174.
[17] ———, *Generalization gradients and applications*, Trans. Amer. Math. Soc., 205 (1975), pp. 247–262.
[18] A. R. Conn and N. I. M. Gould, *An exact penalty function for semi-infinite programming*, Math. Programming, 37 (1987), pp. 19–40.
[19] A. R. Conn and T. Pietrzykowski, *A penalty function method converging directly to a constrained optimum*, SIAM J. Numer. Anal., 14 (1977), pp. 348– 374.
[20] S. Dolecki and S. Rolewicz, *Exact penalties for local minima*, SIAM J. Control Optim. 17 (1979), pp. 596–606.
[21] M. Edelstein, *A note on nearest points*, Quart. J. Math. Oxford, 21 (1970), pp. 403–405.
[22] I. Ekeland, *On the variational principle*, J. Math. Anal. Appl., 47 (1974), pp. 324–353.
[23] I. I. Eremin, *The penalty method in convex programming*, Soviet Math. Dokl., 8 (1966), pp. 459–462.
[24] J. P. Evans, F. J. Gould, and J. W. Tolle, *Exact penalty functions in nonlinear programming*, Math. Programming, 4 (1973), pp. 72–97.
[25] A. V. Fiacco and G. P. McCormick, *Asymptotic conditions for constrained minimization*, RAC–TP–340, Research Analysis Corporation, McLean, VA, 1968.
[26] R. Fletcher, *Practical Methods of Optimization*, second edition, John Wiley, New York, 1987.
[27] ———, *Penalty functions,* in Mathematical Programming: The State of the Art, Bonn 1982, A. Bachem, M. Grötschel, and B. Korte, eds., Springer–Verlag, Berlin, 1983, pp. 87–114.
[28] ———, *A model algorithm for composite nondifferentiable optimization problems*, Math. Programming Study, 17 (1982), pp. 67–76.
[29] ———, *Practical Methods of Optimization, Volume* 2; *Constrained Optimization*, John Wiley, New York, 1981.
[30] ———, *An exact penalty function for nonlinear programming with inequalities*, Math. Programming, 5 (1973), pp. 129–150.
[31] U. M. Garciá–Palomares, *Connections Among Nonlinear Programming, Minimax, and Exact Penalty Functions*, Computer Science Division, Argonne National Laboratories, Technical Memorandum No. 20, 1983.
[32] J. Gauvin, *A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming*, Math. Programming, 12 (1977), pp. 136–138.
[33] S.–P. Han and O. L. Mangasarian, *Exact penalty functions in nonlinear programming*, Math. Programming, 17 (1979), pp. 251–269.
[34] ———, *A dual differentiable exact penalty function*, Math. Programming, 14 (1978), pp. 73–86.

[35] M. R. HESTENES, *Optimization Theory: The Finite Dimensional Case*, Robert E. Krieger, New York, reprint, 1981.

[36] ———, *Calculus of Variations and Optimal Control Theory*, John Wiley, New York, 1966.

[37] S. HOWE, *New conditions for exactness of simple penalty functions*, SIAM J. Control Optim., 11 (1973), pp. 378–381.

[38] L. HURWICZ, *Programming in linear spaces*, in Studies in Linear and Non-Linear Programming, K. J. Arrow, L. Hurwicz, and H. Uzawa, eds., Stanford University Press, Stanford, CA, 1958.

[39] A.D. IOFFE, *Necessary and sufficient conditions for a local minimum. 1: A reduction theorem and order conditions*, SIAM J. Control Optim. 17 (1979), pp. 245–250.

[40] ———, *Necessary and sufficient conditions for a local minimum. 2: Conditions of Levitin-Miljutin-Osmolovskii Type*, SIAM J. Control Optim. 17 (1979), pp. 251–265.

[41] ———, *Necessary and sufficient conditions for a local minimum. 3: Second order conditions and augmented duality*, SIAM J. Control Optim. 17 (1979), pp. 266–288.

[42] F. JOHN, *Extremum Problems with Inequalities as Subsidiary Conditions*, Studies and Essays Presented to R. Courant on his 60th birthday, Wiley-Interscience, New York, 1948, pp. 187–204.

[43] H. W. KUHN AND A. W. TUCKER, *Nonlinear Programming*, Proc. 2nd Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, 1961, pp. 481–492.

[44] J. B. LASSERRE, *Exact penalty functions and Lagrange multipliers*, (R.A.I.R.). Automat./Syst. Anal. Control, 14 (1980), pp. 117–125.

[45] E. S. LEVITIN, A. A. MILJUTIN, AND N. P. OSMOLOVSKII, *On conditions for a local minimum in a problem with constraints*, in Mathematical Economics and Functional Analysis, B. S. Mitjagin, ed., Nauka, Moscow, 1974, pp. 139–202 (In Russian).

[46] D. G. LUENBERGER, *Optimization by Vector Space Methods*, John Wiley, New York, 1969.

[47] ———, *Control problems with kinks*, IEEE Trans. Automat. Control, 15 (1970), pp. 570–575.

[48] L. A. LYUSTERNIK, *On constrained extrema of functionals*, Matem. Sbornik, 41 (1934), pp. 390–401. (In Russian).

[49] J. MAGUREGUI, *Regular Multivalued Functions and Algorithmic Applications*, Ph.D. thesis, University of Wisconsin at Madison, Madison, WI, 1977.

[50] O. L. MANGASARIAN, *Sufficiency of exact penalty minimization*, SIAM J. Control Optim., 23 (1985), pp. 30–37.

[51] ———, *Nonlinear Programming*, Robert E. Kreiger, New York, reprint, 1979.

[52] O. L. MANGASARIAN AND S. FROMOVITZ, *The Fritz John necessary optimality conditions in the presence of equality and inequality constraints*, J. Math. Anal. Appl.,17 (1967), pp. 37–47.

[53] P. MICHEL AND J.-P. PENOT, *Calcul sous-differential pour des fonctions lipschitziennes et non lipschitziennes*, C. R. Acad. Sci. Paris, 298 (1984), pp. 269–272.

[54] B. S. MORDUKHOVICH, *Approximation Methods in Optimization and Control Problems*, Main Physical and Mathematical Editions, 1988, Moscow, Nauka. (In Russian).

[55] D. D. MORRISON, *Optimization by least squares*, SIAM J. Numer. Anal., 5 (1968), pp. 83–88.

[56] V. H. NGUYEN, J.-J. STRODIOT, AND R. MIFFLIN, *On Conditions to Have Bounded Multipliers in Locally Lipschitz Programming*, Math. Programming, 18 (1980), pp. 100–106.

[57] L. PENNISI, *An indirect proof for the problem of Lagrange with differential inequalities as added side conditions*, Trans. Amer. Math. Soc., 74 (1953), pp. 177–198.

[58] T. PIETRZYKOWSKI, *An exact potential method for constrained maxima*, SIAM J. Numer. Anal., 6 (1969), pp. 299–304.

[59] E. POLAK, D. Q. MAYNE, AND Y. WARDI, *On the extension of constrained optimization algorithms from differentiable to non-differentiable problems*, SIAM J. Control Optim., 21 (1983), pp. 179–203.

[60] G. DI PILLO AND L. GRIPPO, *A continuously differentiable exact penalty function for nonlinear programming problems with inequality constraints*, SIAM J. Control Optim., 23 (1985), pp. 72–84.

[61] S. M. ROBINSON, *Stability theory for systems of inequalities. Part* II. *Differentiable nonlinear systems*, SIAM J. Numer. Anal., 13 (1976), pp. 497–513.

[62] ———, *Regularity and stability for convex multivalued functions*, Math. Oper. Res., 1 (1976), pp. 130–143.

[63] R. T. ROCKAFELLAR, *Second-order optimality conditions in nonlinear programming obtained by way of epi-derivatives*, Math. Oper. Res., to appear.

[64] ———, *-and second-order epi-differentiability in nonlinear programming*, Trans. Amer. Math. Soc., 307 (1988), pp. 75–108.

[65] ———, *Proximal subgradients, marginal values, and augmented Lagrangians in nonconvex optimization*, Math. Oper. Res., 6 (1981), pp. 424–436.

[66] ———, *Extensions of subgradient calculus with applications to optimization*, Nonlinear Anal. Theory Meth. Appl., 9 (1985), pp. 665–698.

[67] ———, *Lipschitzian properties of multifunctions*, Nonlinear Anal. TMA, 9 (1985), pp. 867–885.

[68] ———, *Directionally Lipschitzian functions and subdifferential calculus*, Proc. London Math. Soc., 39 (1979), pp. 331–355.

[69] ———, *Augmented Lagrange multiplier functions and duality in nonconvex programming*, SIAM J. Control Optim., 12 (1974), pp. 268–285.

[70] ———, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[71] E. ROSENBERG, *Exact penalty functions and stability in locally Lipschitz programming*, Math. Programming, 30 (1984), pp. 340–356.

[72] M. SLATER, *Lagrange multipliers revisited: a contribution to non-linear programming*, Cowles Commission Discussion Paper, Math., 403, 1950.

[73] C. URSESCU, *Multifunctions with closed convex graph*, Czechoslavak Math. J., 25 (1975), pp. 438–441.

[74] L. P. VLASOV, *Approximative properties of sets in normed linear spaces*, Russian Math. Surveys, 28 (1975), pp. 1–61.

[75] W. I. ZANGWILL, *Nonlinear programming via penalty functions*, Management Sci., 13 (1967), pp. 344–358.