JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 109, No. 1, pp. 99–121, APRIL 2001

Nonlinear Lagrangian Theory for Nonconvex Optimization[1]

C. J. GOH[2] AND X. Q. YANG[3]

Communicated by F. A. Potra

Abstract. The Lagrangian function in the conventional theory for solving constrained optimization problems is a linear combination of the cost and constraint functions. Typically, the optimality conditions based on linear Lagrangian theory are either necessary or sufficient, but not both unless the underlying cost and constraint functions are also convex.

We propose a somewhat different approach for solving a nonconvex inequality constrained optimization problem based on a nonlinear Lagrangian function. This leads to optimality conditions which are both sufficient and necessary, without any convexity assumption. Subsequently, under appropriate assumptions, the optimality conditions derived from the new nonlinear Lagrangian approach are used to obtain an equivalent root-finding problem. By appropriately defining a dual optimization problem and an alternative dual problem, we show that zero duality gap will hold always regardless of convexity, contrary to the case of linear Lagrangian duality.

Key Words. Inequality constraints, nonlinear Lagrangian, nonconvex optimization, sufficient and necessary conditions, zero duality gap.

1. Introduction

Consider the following inequality constrained optimization problem:

$$(\mathrm{P}_0)\quad \inf_{x\in X} \; f_0(x),$$

$$\text{s.t.}\quad f_i(x)\leq \theta_i, \qquad i=1,2,\ldots,m,$$

[1]This work was supported by the Australian Research Council and a Research Grant from the Research Committee of The Hong Kong Polytechnic University. We are grateful to the referees for detailed comments and suggestions on an earlier version of this paper.

[2]Research Fellow, Department of Mathematics and Statistics, University of Western Australia, Nedlands, Western Australia, Australia.

[3]Assistant Professor, Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, PRC.

where the functions f_i: $\mathbb{R}^n \to \mathbb{R}$, $i = 0, 1, 2, \ldots, m$, are continuously differentiable but not necessarily convex, and where X is a subset of $\mathbb{R}^n$. The constraint $x \in X$ is intended to include simple constraints, such as simple bounds or linear constraints. Without loss of generality, we assume that the parameter vector

$$\theta = [\theta_1, \ldots, \theta_m] \in \operatorname{int} \mathbb{R}^m_+.$$

Define the set of all feasible solutions to be

$$X_0 = \{x \in X \mid f_i(x) \leq \theta_i, i = 1, 2, \ldots, m\}.$$

Let

$$g(x) = [f_1(x), f_2(x), \ldots, f_m(x)].$$

The family of perturbed problems associated with problem P_0 is defined by

$$(P_y) \quad \inf_{x \in X} \quad f_0(x)$$

$$\text{s.t.} \quad g(x) \leqq y,$$

where the vector $y = [y_1, y_2, \ldots, y_m]$ is a perturbation to the parameter vector θ of the original problem P_0. When $y = \theta$, the perturbed problem reduces to the original problem P_0. Let the perturbation function w: $\mathbb{R}^m \to \mathbb{R}$ be defined by

$$w(y) = \inf\{f_0(x) \mid g(x) \leqq y, x \in X\}.$$

Using the conventional notion that $\inf \varnothing = +\infty$, w has an effective domain

$$\operatorname{dom}(w) = \{y \mid \exists x \in X, \text{ s.t. } g(x) \leqq y\}.$$

Clearly, the perturbation function w is a monotone nonincreasing function of y. Define the epigraph of $w(y)$ as the set

$$\operatorname{epi}(w) = \{[y, y_0] \mid y \in \operatorname{dom}(w), y_0 \geq w(y)\} \subset \mathbb{R}^{m+1}.$$

The conventional way of tackling this problem theoretically is via the Lagrangian approach; see for example Ref. 1. To distinguish the conventional Lagrangian approach from our proposed approach, we shall call the former as the linear Lagrangian theory, consistently with the fact that the Lagrangian function is a linear combination of the cost and constraint functions,

$$L(x, \lambda) = f_0(x) + \sum_{i=1}^{m} \lambda_i [f_i(x) - \theta_i]. \tag{1}$$

Essentially, this reduces a constrained problem to an unconstrained problem.

The linear Lagrangian theory leads to well-known optimality conditions and duality results, which can be found in any classical book on the theory of optimization such as Ref. 2. Most notably:

(i) the saddle-point theorem is a sufficient condition for optimality;
(ii) the Kuhn–Tucker condition is a necessary condition for optimality under a constraint qualification;
(iii) if we let the dual function be defined by

$$\phi_L(\lambda) = \inf_{x \in X} L(x, \lambda)$$

and the dual optimization problem be defined by

$$\sup_{\lambda \geq 0} \phi_L(\lambda),$$

then the standard weak duality result asserts that

$$\phi_L(\lambda) \leq f_0(x), \qquad \text{for all } \lambda \geq 0 \text{ and for all feasible } x.$$

When all of the f_i, $i = 0, 1, \ldots, m$, are convex, then it is well-known that:

(a) the saddle-point sufficient condition is also necessary under the condition that $X = \mathbb{R}^n$ and a Slater constraint qualification holds;
(b) the Kuhn–Tucker necessary condition is also sufficient;
(c) zero duality gap holds under a suitable constraint qualification (Ref. 3), i.e.,

$$\sup_{\lambda \geq 0} \phi_L(\lambda) = \inf_{x \in X_0} f_0(x).$$

With the convexity assumption, the minimum solution can be obtained easily using either one of the three above results. The problem is that, if any one of f_i, $i = 0, 1, \ldots, m$, is not convex, then the optimal solution may not satisfy the saddle-point sufficient condition; a solution obtained by the Kuhn–Tucker necessary condition may not be optimal; and the duality gap may be nonzero, i.e.,

$$\inf_{x \in X_0} f_0(x) > \sup_{\lambda \geq 0} \phi_L(\lambda);$$

consequently the primal–dual method may fail (see Ref. 4).

A unified study in terms of conjugate duality for various convex programming problems is presented in Ref. 5. Extensions on the zero duality gap without the convexity assumptions can be found in the literature. The conjugate framework of Ref. 5 was extended to cater for nonconvex optimization problems in Ref. 6. Moreover, it was shown in Ref. 7 that a zero duality gap exists for an optimization problem with an infinite number of inequality constraints if one assumes the regularity and convexlike property. In Ref. 8, a pth power transformation was introduced to guarantee a zero duality gap for an optimization problem which is not necessarily convex. In Ref. 9, zero duality gaps were also established for nonconvex optimization problems in terms of a quasi-conjugate function. More recently a unified scheme of augmented Lagrangian functions was proposed in Ref. 10 to guarantee the zero duality gap.

At the optimum solution x^*, the conventional Lagrangian (linear in f_0 and f_i) can be interpreted geometrically as a supporting hyperplane to the set epi(w). When the functions f_i, $i=0, 1, \ldots, m$, are convex, then the epigraph of w is convex, and a supporting hyperplane exists at every point of the perturbation function $w(y)$ or the lower boundary of epi(w). When any one of f_i, $i=0, 1, \ldots, m$, is not convex, then epi(w) is no longer convex, and a supporting hyperplane may not exist at some point of the perturbation function. This is the reason why the saddle-point theorem is only sufficient but not necessary. Furthermore, one can also interpret the Lagrangian dual function as the vertical intercept of the supporting hyperplane to the perturbation function with the normal λ when $y=\theta$ (see Fig. 8.3 of Ref. 11). Thus, the duality gap is the difference between this vertical intercept and $f_0(x^*)$. If the perturbation function is nonconvex, then it is easy to see that the duality gap cannot be zero if the supporting hyperplane is not supporting the perturbation function at $y=\theta$.

The nonlinear Lagrangian approach proposed in this paper is motivated by an intuitively clear geometrical observation. Since the perturbation function is monotone nonincreasing, every point of the perturbation function can be supported by a shifted cone obtained by shifting the negative orthant in $\mathbb{R}^{m+1}$ (see Fig. 1). This simple observation begs the question; is there a dual formulation such that the duality gap is always zero regardless of convexity? Other than the known results in Refs. 10 and 12, this paper offers alternative affirmative answers to these questions by using a weighted Chebyshev norm. Like the conventional linear Lagrangian theory, the constrained problem is reduced essentially to an unconstrained problem using a nonlinear Lagrangian function (effectively, a scalarization of the cost and constraint functions). Indeed, in this case, the Lagrangian function is now a nonlinear function of the cost and constraint functions. It is worth noting that our approach shares some similarity with the one in Ref. 13 where a nonlinear programming problem was reformulated equivalently as a minimax optimization problem. Yet, our approach provides new models for a

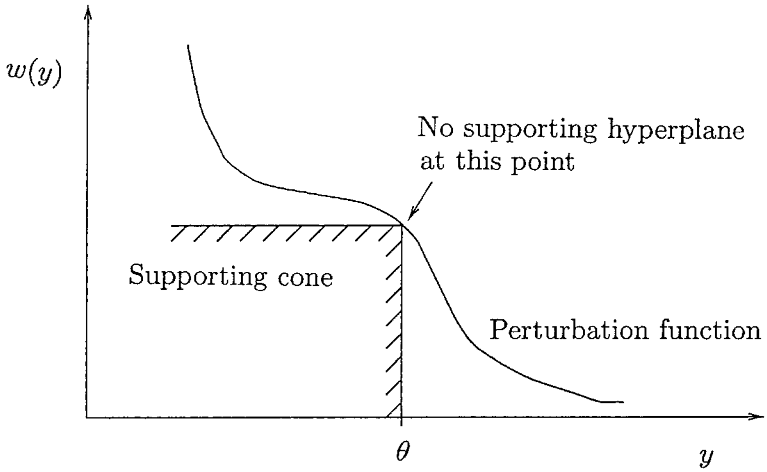

Fig. 1. Perturbation function of nonconvex problem.

large class of optimization problems (namely, convex composite optimization), which have been studied extensively in Refs. 12 and 13–16.

The outline of the paper is as follows. In Section 2, we define a nonlinear Lagrangian function based on the weighted Chebyshev norm, derive an equivalent optimality condition, and show how it can be used to construct a method for solving nonconvex optimization problems. In Section 3, we present a duality theory based on the nonlinear Lagrangian function defined in Section 2. In Section 4, an alternative nonlinear Lagrangian function is introduced, and another alternative version of the duality result is derived, requiring substantially less assumptions than the previous version. A simple illustrative example is given in Section 5; we summarize the strength and limitations of the theory in the conclusions of Section 6.

2. Nonlinear Lagrangian and Equivalent Optimality Condition

The key analytical tool used here is the weighted Chebyshev norm (strictly speaking, this does not qualify as a norm), which is a scalar-valued function mapping $\mathbb{R}^{m+1}$ to $\mathbb{R}_+$. Given $z \in \mathbb{R}^{m+1}$ and a weight vector $e \in \operatorname{int} \mathbb{R}_+^{m+1}$, the weighted Chebyshev norm of z is defined as

$$\xi_e(z) = \max_{0 \le i \le m} \{z_i/e_i\}. \tag{2}$$

It is convenient to normalize the weight vector such that the first component $e_0 = 1$. Define the set

$$E = \{e \in \operatorname{int} \mathbb{R}_+^{m+1} \mid e_0 = 1, e_i > 0, i = 1, 2, \ldots, m\}. \tag{3}$$

Note that there is no reason why we cannot weigh the vector z by multiplying each component z_i by some weight $\lambda_i > 0$ as in the case of a linear Lagrangian. However, division of z_i by e_i affords a clearer geometrical interpretation. Note that the Chebyshev norm is a limiting case ($p = \infty$) of the Minkowski metric,

$$\rho(z; e) = \left\{ \sum_{i=0}^{m} [z_i/e_i]^p \right\}^{1/p}. \tag{4}$$

There is more than one way of defining the nonlinear Lagrangian function. We begin with the simpler version defined below; an alternative version is discussed in Section 4.

Definition 2.1. Given $x \in X, e \in E$, the nonlinear Lagrangian function for the primal constrained optimization problem P_0 is defined by

$$\mathscr{L}(x, e) = \xi_e(f(x)) = \max_{0 \le i \le m} \{f_i(x)/e_i\}, \tag{5}$$

where

$$f(x) = [f_0(x), f_1(x), f_2(x), \ldots, f_m(x)] \in \mathbb{R}^{m+1}. \tag{6}$$

Remark 2.1. A much more general Lagrangian function than the nonlinear Lagrangian function given in Definition 2.1 is defined in Ref. 17. Consider the following primal optimization problem:

(AP) $\inf h(G)$,

where the constraint set G is a subset of a given set F, and where $h: F \to \bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ is an extended-value function. The dual optimization problem of (AP) is defined as follows:

(AD) $\inf \lambda(w)$,

in which w is assumed to be the dual constraint set, and the function $\lambda(w)$ is defined by

$$\lambda(w) = \inf_{y \in F} L^{G,h}(y, w), \qquad w \in W,$$

where $L^{G,h}: F \times W \to \bar{\mathbb{R}}$ is called an abstract Lagrangian function for (AP). However, in Ref. 17, there is no study about the zero duality gap property between the primal and dual problems. It is clear that the nonlinear Lagrangian function $\mathscr{L}(x, e)$ is a special case of the abstract Lagrangian function.

Remark 2.2. The Lagrangian function of Definition 2.1 is also a special case of the augmented Lagrangian given in Ref. 9. That is, $\mathscr{L}(x, e)$ satisfies the following property:

$$\mathscr{L}(x, e) = f_0(x),$$

as it is assumed that

$$f_0(x) > 0, \qquad \forall x \in X.$$

Moreover, a direct minimax reformulation of problem (P_0) is given in Ref. 13 as follows: problem (P_0) is equivalent to

$$\min \max\{f_0(x) - f(x^*), f_1(x), \ldots, f_m(x)\},$$

if an optimal solution of (P_0), namely x^*, is known.

We now present an equivalent optimality condition, in terms of an unconstrained problem, which is both sufficient and necessary. This optimality condition leads directly to a conceptually simple method for solving convex or nonconvex constrained optimization problems.

Theorem 2.1. Sufficient and Necessary Condition for Optimality without Convexity. Let x^* be the (global) optimal solution of problem P_0, and let $\theta_0 = f_0(x^*) > 0$. Then, x^0 solves problem P_0 if and only if x^0 solves the following unconstrained problem:

$$(P_1) \quad \inf_{x \in X} \mathscr{L}(x, \hat{\theta}) = \inf_{x \in X} \max_{0 \le i \le m} \{f_i(x)/\theta_i\},$$

where $\hat{\theta} = [\theta_0, \theta_1, \ldots, \theta_m]$.

Proof. Let us first prove that

$$\inf_{x \in X} \max_{0 \le i \le m} \{f_i(x)/\theta_i\} = 1.$$

If $x \in X$ is feasible, then

$$f_0(x) \ge f_0(x^*) \quad \text{and} \quad f_j(x) \le \theta_j, j = 1, \ldots, m.$$

Thus,

$$\max_{0 \le i \le m} \{f_i(x)/\theta_i\} \ge 1.$$

If $x \in X$ is infeasible, then $\exists j \in \{1, 2, \ldots, m\}$ such that $f_j(x) > \theta_j$, implying that

$$\max_{0 \le i \le m} \{f_i(x)/\theta_i\} > 1.$$

Moreover,

$$\max_{0\le i\le m} \{f_i(x^*)/\theta_i\} = 1.$$

Thus, the claim holds.

Necessary Condition. If x^0 solves problem P_0, then x^0 must be feasible, and $f_0(x^0) = f_0(x^*)$. Hence,

$$\max_{0\le i\le m} \{f_i(x^0)/\theta_i\} = 1;$$

i.e., x^0 solves problem P_1.

Sufficient Condition. If x^0 does not solve problem P_0, then either x^0 is infeasible or x^0 is feasible and $f_0(x^0) > f_0(x^*)$. If x^0 is infeasible, then $\exists j \in \{1, 2, \ldots, m\}$ such that $f_j(x^0) > \theta_j$, implying that

$$\max_{0\le i\le m} \{f_i(x^0)/\theta_i\} > 1. \tag{7}$$

If x^0 is feasible and $f_0(x^0) > f_0(x^*)$, then inequality (7) still holds. Thus, in both cases, x^0 does not solve problem P_1. □

Remark 2.3. Problem P_1 in Theorem 2.1 is actually called a convex composite problem where the cost function $\mathscr{L}(x, \hat{\theta})$ is a convex composite function of the form $g(F(x))$, with a convex function $g(y) = \max_{0\le i\le m} y_i$; see Refs. 15 and 16.

The above theorem leads to a root-finding method for solving the nonconvex inequality constrained optimization problem P_0. Consider the following scalar function of a scalar parameter $\theta_0 \in \mathbb{R}_+ \backslash \{0\}$, recalling that θ_i, $i = 1, 2, \ldots, m$, are fixed parameters of the problem P_0:

$$\psi(\theta_0) = \inf_{x\in X} \max_{0\le i\le m} \{f_i(x)/\theta_i\}. \tag{8}$$

Under appropriate assumptions, the function ψ has some nice properties which are summarized as follows.

Theorem 2.2. Assume that X_0 is compact and that, if X is unbounded, then

$$\lim_{\|x\|\to+\infty} f_0(x) = +\infty.$$

Let $\theta_0 > 0$. Assume that problem P_0 admits a solution x^* with $f_0(x^*) > 0$ and that, for any $\theta_0 > f_0(x^*)$, there exists $x \in X$ such that $f_i(x) < \theta_i$, $f_0(x^*) < f_0(x) < \theta_0$. Then, ψ has the following properties:

(i) $0 < \theta_0 < f_0(x^*) \Rightarrow \psi(\theta_0) > 1;$
(ii) $\theta_0 > f_0(x^*) \Rightarrow \psi(\theta_0) < 1;$
(iii) ψ is a monotone nonincreasing function of θ_0.

Thus, solving (P_0) is equivalent to the problem of finding the root of the equation $\psi(\theta_0) = 1$.

Proof.

(i) Assume that $0 < \theta_0 < f_0(x^*)$. It is clear that

$$\psi(\theta_0) = \inf_{x \in X} \max_{0 \le i \le m} \{f_i(x)/\theta_i\}$$

$$= \min\left\{ \inf_{x \in X_0} \max_{0 \le i \le m} \{f_i(x)/\theta_i\},\ \inf_{x \in X\ X_0} \max_{0 \le i \le m} \{f_i(x)/\theta_i\} \right\}.$$

Thus,

$$\inf_{x \in X_0} \max_{0 \le i \le m} \{f_i(x)/\theta_i\}$$

$$\ge \inf_{x \in X_0} \max_{0 \le i \le m} \{f_0(x^*)/\theta_0, f_1(x)/\theta_1, \ldots, f_m(x)/\theta_m\} > 1. \tag{9}$$

For any $\delta > 0$, the set $X \backslash X_0$ can be decomposed into

$$X \backslash X_0 = \bar{X}_0^{> \delta} \cup \bar{X}_0^{< \delta},$$

where

$$\bar{X}_0^{> \delta} = \{x \in X \backslash X_0 : \exists i,\ f_i(x)/\theta_i > 1 + \delta\},$$

$$\bar{X}_0^{< \delta} = \{x \in X \backslash X_0 : \forall i,\ f_i(x)/\theta_i \le 1 + \delta\}.$$

Then,

$$\inf_{x \in \bar{X}_0^{> \delta}} \max_{0 \le i \le m} \{f_i(x)/\theta_i\} \ge 1 + \delta > 1. \tag{10}$$

Moreover, as

$$\lim_{\|x\| \to +\infty} f_0(x) = +\infty,$$

there is a $\Delta > 0$ such that

$$\inf_{x \in \bar{X}_0^{< \delta}} \max_{0 \le i \le m} \{f_i(x)/\theta_i\} = \inf_{\substack{x \in \bar{X}_0^{< \delta} \\ \|x\| \le \Delta}} \max_{0 \le i \le m} \{f_i(x)/\theta_i\}.$$

For $[f_0(x^*) - \theta_0]/2 > 0$, from the compactness of X_0, choose $\delta > 0$ so small that, for any $x \in \bar{X}_0^{< \delta}$ and $\|x\| \le \Delta$, there exists $y \in X_0$ such that

$$|f_0(x) - f_0(y)| \leq [f_0(x^*) - \theta_0]/2.$$

Then,

$$\begin{aligned} f_0(x) &\geq f_0(y) - [f_0(x^*) - \theta_0]/2 \\ &\geq f_0(x^*) - [f_0(x^*) - \theta_0]/2 \\ &= [f_0(x^*) + \theta_0]/2. \end{aligned}$$

Therefore,

$$f_0(x)/\theta_0 \geq [f_0(x^*) + \theta_0]/2\theta_0, \qquad \forall x \in \bar{X}_0^{<\delta}, \|x\| \leq \Delta;$$

i.e.,

$$\begin{aligned} \inf_{\substack{x \in \bar{X}_0^{<\delta} \\ \|x\| \leq \Delta}} \max_{0 \leq i \leq m} f_i(x)/\theta_i &\geq \inf_{x \in \bar{X}_0^{<\delta}} f_0(x)/\theta_0 \\ &\geq [f_0(x^*) + \theta_0]/2\theta_0 > 1. \end{aligned} \tag{11}$$

Combining (9)–(11), we see that $\psi(\theta_0) > 1$.

(ii) By the assumption, if $\theta_0 > f_0(x^*)$, then there exists $x \in X$ such that

$$f_i(x) < \theta_i, i = 1, 2, \ldots, m, \quad \text{and} \quad \theta_0 > f_0(x) > f_0(x^*).$$

Consequently,

$$\max_{0 \leq i \leq m} \{f_i(x)/\theta_i\} < 1, \tag{12}$$

and hence,

$$\psi(\theta_0) = \inf_{x \in X} \max_{0 \leq i \leq m} \{f_i(x)/\theta_i\} < 1. \tag{13}$$

(iii) Let $\theta_0^1 \geq \theta_0^2$. Then,

$$f_0(x)/\theta_0^1 \leq f_0(x)/\theta_0^2, \qquad \forall x \in X. \tag{14}$$

This implies that

$$\begin{aligned} &\max\{f_0(x)/\theta_0^1, f_i(x)/\theta_i, i = 1, 2, \ldots, m\} \\ &\leq \max\{f_0(x)/\theta_0^2, f_i(x)/\theta_i, i = 1, 2, \ldots, m\}, \end{aligned}$$

and hence,

$$\psi(\theta_0^1) \leq \psi(\theta_0^2). \tag{15}$$

(iv) If $\theta_0 = f_0(x^*)$, it follows from Theorem 2.1 that $\psi(\theta_0) = 1$. Now assume that $\psi(\theta_0) = 1$. Then, it follows from (i) and (ii) that $\theta_0 = f_0(x^*)$. □

Remark 2.4. The continuity of ψ was studied in Ref. 18 and the references cited therein. As the continuity of θ_0 is not important for our approach, we do not go into details.

Remark 2.5. Note that, if $\theta_0 < f_0(x^*)$, then the corresponding minimizing x in (8) is unfeasible. Similarly, if $\theta_0 \geq f_0(x^*)$, then the corresponding minimizing x in (8) is feasible.

Remark 2.6. The assumption in Theorem 2.2 that, for any $\theta_0 > f_0(x^*)$, there exists $x \in X$ such that $f_i(x) < \theta_i$ and $\theta_0 > f_0(x) > f_0(x^*)$ looks stronger as it depends on a parameter θ_0. However, the following example shows that the above assumption is necessary to assure Theorem 2.2(ii). Let

$$f_0(x) = x^2 + 1,$$

$$f_1(x) = (x+1)^3 + 1, \qquad \text{if } x < -1,$$

$$f_1(x) = (x-1)^3 + 1, \qquad \text{if } x > 0,$$

$$f_1(x) = 1, \qquad \text{otherwise.}$$

This assumption does not hold, and neither does Theorem 2.2(ii), since $\varphi(\theta_0) = \text{const} = 1$ for all $\theta_0 \in [1, 2]$. Without this assumption, a weaker result is obtained in Theorem 2.3.

Because of these special properties of the function ψ, under the assumptions given in Theorem 2.2, solving the constrained optimization problem P_0 is now equivalent to a rather simple problem: find the unique root of a monotone decreasing (scalar) function of a scalar variable:

(P_2) Find $\theta_0^* = f_0(x^*)$ such that $\psi(\theta_0^*) = 1$.

At first look, this appears to be a trivial problem, since it is almost effortless to solve numerically for the root of a monotone [strictly monotone at the root by virtue of Theorem 2.2(iii)] scalar function of a scalar variable. However, in practice, this is made nontrivial by the fact that each function call of ψ requires the solution of an unconstrained minimax problem, which in itself is not a trivial problem. Nevertheless, there exist several effective methods and software packages that deal with minimax optimization effectively [for example, see Ref. 19 or the minimax function in the optimization toolbox of MATLAB (Ref. 20)], and these will be fully exploited.

A weaker version of Theorem 2.2 is as follows.

Theorem 2.3. Let f_0 be uniformly continuous on X and $\theta_0 > 0$. Assume that problem P_0 admits a solution x^* with $f_0(x^*) > 0$. Then:

(i) $0 < \theta_0 < f_0(x^*) \Rightarrow \psi(\theta_0) > 1$;
(ii) $\theta_0 \geq f_0(x^*) \Rightarrow \psi(\theta_0) \leq 1$;
(iii) ψ is a monotone nonincreasing function of θ_0;
(iv) $\theta_0 = f_0(x^*)$ if and only if θ_0 is the least root of $\psi(\theta_0) = 1$.

Proof. The assertions (i) and (iii) follow according to the proof of Theorem 2.2.

(ii) Let $x = x^*$. Then,

$$\max_{0 \leq i \leq m} \{f_i(x^*)/\theta_i\} \leq 1,$$

and hence,

$$\psi(\theta_0) = \inf_{x \in X} \max_{0 \leq i \leq m} \{f_i(x)/\theta_i\} \leq 1.$$

(iii) Assume $\theta_0 = f_0(x^*)$. From Theorem 2.1, $\psi(\theta_0) = 1$. It follows from (i) that θ_0 is the least root of $\psi(\theta_0) = 1$. If θ_0 is the least root of $\psi(\theta_0) = 1$, then from (i), $\theta_0 \geq f_0(x^*)$. It is clear that $\psi(f_0(x^*)) = 1$. Thus, $\theta_0 = f_0(x^*)$. □

So, without the assumptions given in Theorem 2.2, solving the constrained optimization problem P_0 is now equivalent to a slightly harder problem, that of finding the least root of a monotone decreasing (scalar) function of a scalar variable:

(P_3) Find the least root θ_0^* such that $\psi(\theta_0^*) = 1$.

3. Nonlinear Lagrangian Duality Theory

We now present a duality theory with zero duality gap regardless of convexity. Let the function $\gamma: E \to \mathbb{R}$ be defined as follows:

$$\gamma(e) = \inf_{x \in X} \mathscr{L}(x, e) = \inf_{x \in X} \max_{0 \leq i \leq m} \{f_i(x)/e_i\}. \tag{16}$$

Recall that E is defined in (3).

Definition 3.1. Nonlinear Lagrangian Dual Function for Problem P_0, First Version. Assuming that the solution x^* for problem P_0 exists and $f_0(x^*) > 0$, then the nonlinear Lagrangian dual function is defined as

$$\phi(e) = \begin{cases} \gamma(e), & \text{if } e_i \geq \theta_i/f_0(x^*), \ \forall i = 1, 2, \ldots, m, \\ -\infty, & \text{otherwise.} \end{cases} \tag{17}$$

Note that there is a peculiar feature of this dual function; namely, given some $e \in E$, $\phi(e)$ cannot be computed, since $f_0(x^*)$ is not known beforehand. Fortunately, this does not restrict the usefulness of this duality theory, as we will demonstrate in Section 5. In Section 4, we shall present another version of the duality function which is compatible. For the rest of this section, the noncomputable dual function ϕ as defined in (17) will be used.

Definition 3.2. Dual Optimization Problem to Problem P_0, First Version. Here,

$$(D_0) \quad \sup_{e \in E} \phi(e).$$

The following proposition shows that the maximization of ϕ with respect to e occurs in a relative interior point of E. As a result, max may be used instead of sup in problem D_0.

Proposition 3.1. The dual function ϕ is maximized at $e = e^*$.

Proof. Clearly, the dual function cannot be maximized if it takes on the value $-\infty$. We need only to establish that $\gamma(e)$ is a monotone nonincreasing function for all $e \in E$ such that

$$e_i \geq \theta_i / f_0(x^*), \forall i = 1, 2, \ldots, m.$$

Given $e^1, e^2 \in E$, $e^1 \leq e^2$, it is clear that, for a given $x \in X$,

$$f_i(x)/e_i^1 \geq f_i(x)/e_i^2, \qquad \forall i = 0, 1, \ldots, m;$$

hence,

$$\max_{0 \leq i \leq m} \{f_i(x)/e_i^1\} \geq \max_{0 \leq i \leq m} \{f_i(x)/e_i^2\},$$

and consequently,

$$\gamma(e^1) = \min_{x \in X} \max_{0 \leq i \leq m} \{f_i(x)/e_i^1\} \geq \gamma(e^2) = \min_{x \in X} \max_{0 \leq i \leq m} \{f_i(x)/e_i^2\}.$$

Thus, the dual function must be maximized at the smallest possible e without taking the value $-\infty$, which is at $e = e^*$. □

Theorem 3.1. Weak Duality. The following inequality holds:

$$\phi(e) \leq f_0(x), \qquad \forall e \in E, \forall x \in X_0. \tag{18}$$

Proof. If there exists k such that

$$e_k < \theta_k / f_0(x^*),$$

then

$$\phi(e) = -\infty < f_0(x), \qquad \forall x,$$

and (18) holds trivially. Hence, we may assume that

$$e_i \geq \theta_i / f_0(x^*), \qquad \forall i = 1, 2, \ldots, m. \tag{19}$$

Then,

$$\begin{aligned}\phi(e) &= \gamma(e) \\ &= \min_{x\in X} \max_{0\leq i\leq m} \{f_i(x)/e_i\} \\ &\leq \max\{f_0(x^*), f_1(x^*)/e_i, \ldots, f_m(x^*)/e_i\} \\ &\leq \max\{f_0(x^*), [f_1(x^*)/\theta_i] f_0(x^*), \ldots, [f_m(x^*)/\theta_i] f_0(x^*)\} \\ &\leq f_0(x^*) \\ &\leq f_0(x).\end{aligned} \tag{20}$$

Theorem 3.2. Zero Duality Gap. $f_0(x^*) = \phi(e^*)$.

Proof. By definition,

$$\phi(e^*) = \min_{x\in X} \mathscr{L}(x, e^*). \tag{21}$$

Consider the two cases (i) and (ii) below.

(i) If x is feasible, then

$$f_i(x)/\theta_i \leq 1, \forall i = 1, 2, \ldots, m, \quad \text{and} \quad f_0(x) \geq f_0(x^*);$$

this implies that

$$\begin{aligned}\mathscr{L}(x, e^*) &= \max\{f_0(x), [f_1(x)/\theta_1] f_0(x^*), \ldots, [f_m(x)/\theta_m] f_0(x^*)\} \\ &= f_0(x^*).\end{aligned}$$

(ii) If x is infeasible, then $\exists k$ such athat $f_k(x)/\theta_k > 1$; hence,

$$\mathscr{L}(x, e^*) \geq f_0(x^*).$$

Thus, in both case (i) and case (ii), we have

$$\mathscr{L}(x, e^*) \geq f_0(x^*), \qquad \forall x \in X.$$

From (21), we conclude that

$$\min_{x\in x} \mathscr{L}(x, e^*) = \phi(e^*) \geq f_0(x^*). \tag{22}$$

Finally, (22) and the weak duality together imply that $\phi(e^*) = f_0(x^*)$. □

4. Alternative Version of Nonlinear Lagrangian Duality Theory

We now present an alternative version of nonlinear Lagrangian duality based on a computable dual function. It has the advantage over the previous version in that it does not require the assumption of existence of an optimal solution to problem P_0.

Definition 4.1. Let $\theta_0 = 0$. Given $d \in E$ and $x \in X$, the alternative nonlinear Lagrangian function for problem P_0 is defined by

$$\mathscr{L}'(x, d) = \max_{0\leq i\leq m} \{[f_i(x) - \theta_i]/d_i\}.$$

Now, we present a computable dual function based on the alternative nonlinear Lagrangian function. This leads to similar weak and strong dual results as in the previous section.

Definition 4.2. Alternative Dual Function for Problem P_0. Let $\theta_0 = 0$. The alternative dual function for problem P_0 is defined as

$$\phi'(d) = \inf_{x\in X} \max_{0\leq i\leq m} \{[f_i(x) - \theta_i]/d_i\},$$

where $d \in E$.

Definition 4.3. Alternative Dual Optimization Problem to Problem P_0. Here,

$$(D_1) \quad \sup_{d\in E} \phi'(d).$$

Theorem 4.1. Weak Duality. Let $x \in X_0$ and $d \in E$. Then,

$$\phi'(d) \leq f_0(x). \tag{23}$$

Proof. It follows from the feasibility of x that

$$f_0(x) \geq \max_{0\leq i\leq m} \{[f_i(x) - \theta_i]/d_i\}.$$

By Definition 4.2, we have

$$\max_{0\le i\le m}\{[f_i(x)-\theta_i]/d_i\}\ge\phi'(d).$$

Then, the conclusion holds. □

Theorem 4.2. Zero Duality Gap. Assume that

$$f_0(x)\ge 0,\qquad \forall x\in X_0,$$

that X_0 is compact, and that, if X is unbounded, then

$$\lim_{\|x\|\to+\infty} f_0(x)=+\infty.$$

Then,

$$\inf_{x\in X_0} f_0(x)=\sup_{d\in E}\phi'(d). \tag{24}$$

Proof. If (24) does not hold, then by the weak duality (23), there exists $\epsilon>0$ such that

$$\inf_{x\in X_0} f_0(x)-\epsilon\ge\phi'(d),\qquad \forall d\in E. \tag{25}$$

For $\epsilon/5>0$, from the definition of infimum, there exists $\hat{x}\in X_0$ such that

$$f_0(\hat{x})\ge\inf_{x\in X_0} f_0(x)\ge f_0(\hat{x})-\epsilon/5.$$

It is clear that, from (25),

$$f_0(\hat{x})-\epsilon\ge\inf_{x\in X_0} f_0(x)-\epsilon\ge\phi'(d),\qquad \forall d\in E. \tag{26}$$

Then, by the assumptions of the theorem,

$$\begin{aligned}\phi'(d)&=\min\left\{\inf_{x\in X_0}\max_{0\le i\le m}\{[f_i(x)-\theta_i]/d_i\},\ \inf_{x\in X\,X_0}\max_{0\le i\le m}\{[f_i(x)-\theta_i]/d_i\}\right\}\\ &=\min\left\{\inf_{x\in X_0} f_0(x),\ \inf_{x\in X\,X_0}\max_{0\le i\le m}\{[f_i(x)-\theta_i]/d_i\}\right\}\\ &\ge\min\left\{f_0(\hat{x})-\epsilon/5,\ \inf_{x\in X\,X_0}\max_{0\le i\le m}\{[f_i(x)-\theta_i]/d_i\}\right\}.\end{aligned} \tag{27}$$

From (26)–(27), we have

$$\phi'(d)\ge\inf_{x\in X\,X_0}\max_{0\le i\le m}\{[f_i(x)-\theta_i]/d_i\},\qquad \forall d\in E. \tag{28}$$

If follows from the compactness of X_0 that there exists $\delta > 0$ such that, if $x \in X \setminus X_0$ and $d(x, X_0) \leq \delta$, then

$$-\epsilon/2 < f_0(x) - f_0(x_2) < \epsilon/2, \qquad \text{for some } x_2 \in X_0.$$

Note that

$$f_0(x_2) \geq \inf_{x \in X_0} f_0(x) \geq f_0(\hat{x}) - \epsilon/5.$$

Then,

$$f_0(\hat{x}) - 4\epsilon/5 < f_0(x). \tag{29}$$

From (28)–(29), we have

$$\phi'(d) \geq \inf_{\substack{x \in X\ X_0,\\ d(x, X_0) > \delta}} \max_{0 \leq i \leq m} \{[f_i(x) - \theta_i]/di\}, \qquad \forall d \in E. \tag{30}$$

As

$$\lim_{\|x\| \to +\infty} f_0(x) = +\infty,$$

there exists $\Delta > 0$ such that

$$\phi'(d) \geq \inf_{\substack{x \in X\ X_0,\\ d(x, X_0) > \delta \\ \|x\| \leq \Delta}} \max_{0 \leq i \leq m} \{[f_i(x) - \theta_i]/d_i\}, \qquad \forall d \in E.$$

Thus, there exists $\beta > 0$ such that, for any $x \in X \setminus X_0$, $\|x\| \leq \Delta$, and $d(x, X_0) > \delta$,

$$f_i(x) - \theta_i > \beta, \qquad \text{for some } i.$$

Define

$$B_i = \{x \in X \setminus X_0 : \|x\| \leq \Delta, d(x, X_0) > \delta, f_i(x) - \theta_i > \beta\}.$$

Let

$$\hat{d}_i = \begin{cases} \min_{x \in B_i} \{[f_i(x) - \theta_i]/[f_0(\hat{x}) - \epsilon/2]\}, & \text{if } B_i \neq \emptyset, \\ 1, & \text{if } B_i = \emptyset. \end{cases}$$

It is clear that

$$\hat{d} = (1, \hat{d}_1, \ldots, \hat{d}_m) \in E.$$

Moreover,

$$\begin{aligned} \phi'(\hat{d}) &\geq \inf_{\substack{x \in X\ X_0,\\ d(x, X_0) > \delta \\ \|x\| \leq \Delta}} \max_{0 \leq i \leq m} \{[f_i(x) - \theta_i]/\hat{d}_i\} \\ &\geq \max_{0 \leq i \leq m} [f_i(w) - \theta_i]/\hat{d}_i - \epsilon/2, \end{aligned}$$

where, for some $w \in X \setminus X_0$,

$$d(w, X_0) > \delta, \|w\| \le \Delta.$$

Now, there exists k_w such that $w \in B_{k_w}$ and

$$\max_{0 \le i \le m} [f_i(w) - \theta_i]/\hat{d}_i \ge [f_{k_w}(w) - \theta_{k_w}]/\hat{d}_{k_w}.$$

So,

$$\phi'(\hat{d}) \ge [f_{k_w}(w) - \theta_{k_w}]/\hat{d}_{k_w} - \epsilon/2,$$

and hence

$$\phi'(\hat{d}) \ge [f_{k_w}(w) - \theta_{k_w}]\left\{\min_{v \in B_{kw}} [f_{k_w}(v) - \theta_{k_w}]/[f_0(\hat{x}) - \epsilon/2]\right\}^{-1} - \epsilon/2.$$

Since $w \in B_{k_w}$, we have

$$\phi'(\hat{d}) \ge f_0(\hat{x}) - \epsilon.$$

This contradicts (26) and the proof is complete. □

Corollary 4.1. If there exists $x^0 \in X_0$ and $d^* \in E$ such that

$$f_0(x^0) = \phi'(d^*),$$

then x^0 is an optimal solution for problem P_0 and d^* is an optimal solution for problem D_1.

Proof. This follows from Theorem 4.2. □

Remark 4.1. The Lagrangian function $\mathscr{L}'(x, e)$ defined in Definition 4.1 is again a special case of the augmented Lagrangian given in Ref. 10. That is, $\mathscr{L}'(x, e)$ satisfies the property

$$\mathscr{L}'(x, e) = f_0(x),$$

as it is assumed that

$$f_0(x) > 0, \qquad \forall x \in X.$$

However, our proof for the zero duality gap property is different from that of Ref. 10.

5. Illustrative Example

The following example is a modified version of a problem taken from Ref. 8:

$$\min_{x \geqq 0} \quad f_0(x) = 1 - x_1 x_2,$$
$$\text{s.t.} \quad f_1(x) = x_1 + 4x_2 \leq 1.$$

It is easy to see geometrically that, at optimality, the inequality constraint is active; therefore, the optimal solution occurs at

$$x^* = [1/2, 1/8] \quad \text{and} \quad f_0(x^*) = 15/16,$$

so that

$$e^* = [1, 16/15].$$

The perturbation function can be computed easily to be

$$w(y) = 1 - y^2/16,$$

which is clearly nonconvex. Thus, if a linear Lagrangian function

$$L(x, \lambda) = f_0(x) + \lambda f_1(x)$$

is used, then the dual function

$$\phi_L(\lambda) = \min_{x \geqq 0} L(x, \lambda) = \min_{x \geqq 0} 1 - x_1 x_2 + \lambda(x_1 + 4x_2 - 1)$$

will always have a value of $\phi_L(\lambda) = -\infty$, since L is minimized at $x = [\infty, \infty]$ for all $\lambda \geq 0$. As a result, the duality gap $f_0(x^*) - \phi_L(\lambda^*) = \infty$.

An application of the proposed nonlinear Lagrangian approach yields the following dual function and alternative dual function:

$$\phi(e) = \begin{cases} \min_{x \in X} \max\{1 - x_1 x_2, (x_1 + 4x_2)/e\}, & \text{if } e \geq e^* = 16/15, \\ -\infty, & \text{otherwise,} \end{cases}$$

$$\phi'(d) = \min_{x \in X} \max\{1 - x_1 x_2, (x_1 + 4x_2 - 1)/d\}.$$

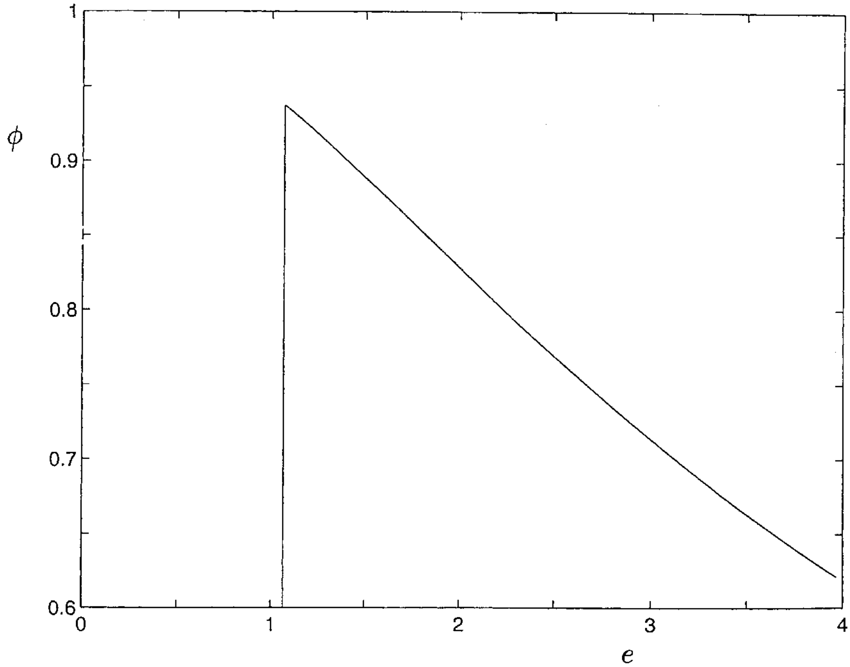

Fig. 2. Dual function, Version 1.

These dual functions are plotted in Figs. 2–3 respectively. Note that, for both dual functions,

$$\max_{e>0} \phi(e) = \sup_{d>0} \phi'(d) = f_0(x^*) = 15/16;$$

therefore, the duality gap is zero, even though the perturbation function is nonconvex.

The function ψ as discussed in Section 2 is plotted for a range of values of θ_0 in Fig. 4. It crosses the line $\psi(\theta_0) = 1$ at $\theta_0^* = 15/16 = f_0(x^*)$, almost linearly. An application of the secant method starting from arbitrary values of θ_0 at significant distance away from the minimum (e.g., 0.1 or 3), using the minimax function of the optimization toolbox of MATLAB (Ref. 20), typically converges to the optimum solution after about 4 iterations.

6. Concluding Remarks

In conclusion, we shall point out an additional feature of the proposed approach, as well as discussing some of the practical computational limitations and directions for further research.

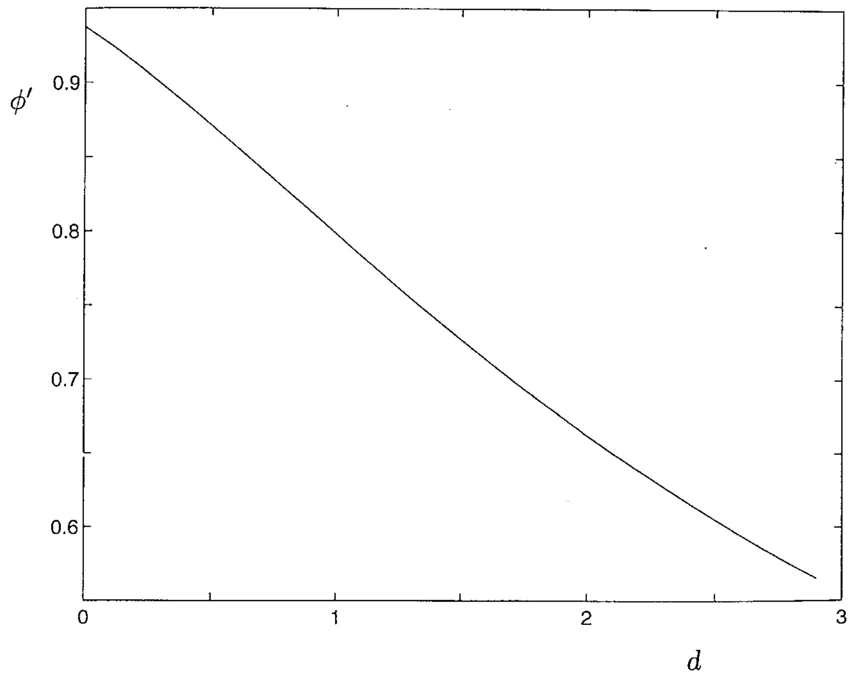

Fig. 3. Dual function, Version 2.

First, it should be noted that many of the results obtained in this paper may still hold even if the set X is discrete (finite or infinite). In this case, the optimality conditions obtained from the traditional Lagrangian theory are never sufficient and necessary, and the duality gap may be nonzero, even if all the underlying functions are convex. However, the nonlinear Lagrangian approach may expect some difficulty in solving discrete minimax problems. This may open up new directions in solving discrete constrained optimization problems.

A limitation of the proposed theory is that it is not clear at present how equality constraints can be treated in the same framework as inequality constraints. Furthermore, in terms of practical computation, we cannot really pretend that we have reduced a difficult problem to a trivial one. This is because solving a minimax problem (which is required by many of the results herein) is by no means easy, even if it is unconstrained. It is our goal henceforth to devise more efficient algorithms based on the proposed nonlinear Lagrangian theory, and test them over a large range of nontrivial examples.

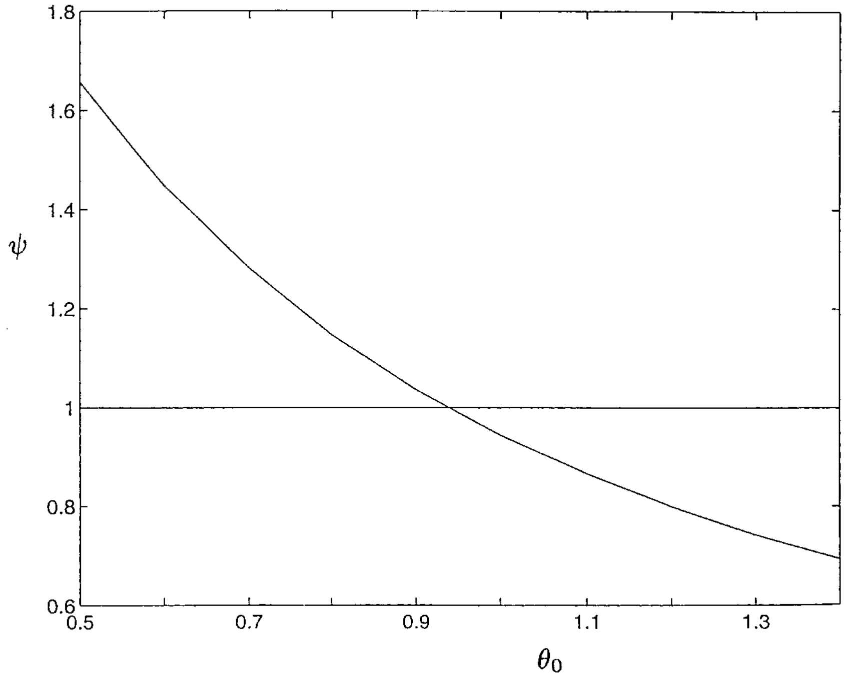

Fig. 4. Solution of a nonlinear equation.

Whether the strong duality (i.e., the Lagrangian multipliers exist and the values of objective functions for the original problem and the dual problem are equal) holds between the original problem and the (alternative) dual problem is still a question. Before answering the question, it is important to have a clear understanding of the role of the parameters e_i or d_i as nonlinear Lagrangian multipliers.

References

1. LUENBERGER, D. G., *Linear and Nonlinear Programming*, 2nd Edition, Addision Wesley, Reading, Massachusetts, 1984.
2. MANGASARIAN, O. L., *Nonlinear Programming*, McGraw-Hill, New York, NY, 1969.
3. CRAVEN, B. D., *Mathematical Programming and Control Theory*, Chapman and Hall, London, England, 1978.
4. KUHN, H. W., and TUCKER, A. W., *Nonlinear Programming*, Proceedings of the 2nd Berkeley Symposium on Mathematical Statistics and Probability, Edited by J. Neyman, University of California Press, Berkeley, California, 1951.

5. ROCKAFELLAR, R. T., *Conjugate Duality and Optimization*, CBMS Lecture Notes Series, SIAM Publications, Philadelphia, Pennsylvania, Vol. 162, 1974.
6. BALDER, E. J., *An Extension of Duality-Stability Relations to Nonconvex Optimization Problems*, SIAM Journal on Control and Optimization, Vol. 15, pp. 329–343, 1977.
7. JEYAKUMAR, V., and WOLKOWICZ, H., *Zero Duality Gaps in Infinite-Dimensional Programming*, Journal of Optimization Theory and Applications, Vol. 67, pp. 87–108, 1990.
8. LI, D., *Zero Duality Gap for a Class of Nonconvex Optimization Problems*, Journal of Optimization Theory and Applications, Vol. 85, pp. 309–324, 1995.
9. THACH, R. T., *Global Optimality Criterion and a Duality with Zero Gap in Nonconvex Optimization*, SIAM Journal on Mathematical Analysis, Vol. 24, pp. 1537–1556, 1993.
10. ROCKAFELLAR, R. T., and WETS, R. J. B., *Variational Analysis*, Springer, Berlin, Germany, 1998.
11. LASDON, L. S., *Optimization Theory for Large Systems*, Series in Operations Research, Macmillan Company, New York, NY, 1970.
12. IOFFE, A., *Necessary and Sufficient Conditions for a Local Minimum, Part 3: Second-Order Conditions and Augmented Duality*, SIAM Journal on Control and Optimization, Vol. 17, pp. 266–288, 1979.
13. HIRIART-URRUTY, J. B., and LEMARECHAL, C., *Convex Analysis and Minimization Algorithms, I*, Series of Comprehensive Studies in Mathematics, Springer Verlag, New York, NY, Vol. 305, 1996.
14. IOFFE, A., *Necessary and Sufficient Conditions for a Local Minimum, Part 1: A Reduction Theorem and First-Order Conditions*, SIAM Journal on Control and Optimization, Vol. 17, pp. 245–250, 1979.
15. BURKE, J. V., *Second-Order Necessary and Sufficient Conditions for Convex Composite NDO*, Mathematical Programming, Vol. 38, pp. 287–302, 1987.
16. YANG, X. Q., *Second-Order Global Optimality Conditions for Convex Composite Optimization*, Mathematical Programming, Vol. 81, pp. 327–347, 1998.
17. SINGER, I., *A General Theory of Dual Optimization Problems*, Journal of Mathematical Analysis and Applications, Vol. 116, pp. 77–130, 1986.
18. BANK, B., GUDDAT, J., KLATTE, D., KUMMER, B., and TAMMER, K., *Nonlinear Parametric Optimization*, Birkhäuser Verlag, Basel, Switzerland, 1983.
19. ZHOU, J. L., and TITS, A. L., *User's Guide for FFSQP Version 3.5: A Fortran Code for Solving Constrained Nonlinear Minimax Optimization Problems, Generating Iterates Satisfying All Inequalities and Linear Constraints,* Technical Report TR-92-107r4, Systems Research Center, University of Maryland, College Park, Maryland, 1992.
20. GRACE, A., *Optimization Toolbox for Use with MATLAB User Guide*, Maths Works, South Natick, Massachusetts, 1990.