unconventional instead of accepted terms. Without giving a lengthy list of examples, the author defines the waiting time of a customer as including the service time of that customer, contrary to accepted practice. The erlang unit is mentioned without a definition and an elementary abelian argument is labeled tauberian.

The shortcomings of this book are not merely stylistic. The proofs of several theorems are inadequate. In discussing the classical result that the output process of an $M/M/1$ queue is Poisson, the author shows only that the times between three successive departures are independent and negative exponentially distributed. This is not only insufficient, but the author's subsequent statement that the theorem is not valid for more general systems, is incorrect. P. J. Burke's theorem was indeed proved for the $M/M/s$ queue.

When there is a choice of several classical arguments the author has a propensity for selecting the least informative approach as in his presentation of the $M/G/1$ model. A number of formulas are poorly aligned and a lengthy proof ends in mid-sentence on p. 135. Apart from all other considerations, the book would have benefited from greater editorial care.

In summary, except as an accessible reference to the author's own research, this book cannot be recommended as reading material on the classical queueing models. This is unfortunate. There is a definite need for clear and unified expositions of the theory of queues, which provide a broad synthesis of an interesting but overly ramified field.

MARCEL F. NEUTS

BULLETIN OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 83, Number 3, May 1977

Optimization theory, the finite dimensional case, by Magnus R. Hestenes, John Wiley and Sons, New York, London, Sydney, Toronto, 1975, xiii + 447 pp., \$24.95.

The central problem of *nonlinear programming*, which is one of the four or five main areas within *mathematical programming*, can be stated as follows:

$$(\mathrm{P})\qquad \begin{aligned} &\text{infimize} && f_0(x)\\ &\text{subject to} && f_i(x) \leqslant 0, \quad i = 1, \ldots, r,\\ & && g_i(x) = 0, \quad i = 1, \ldots, s,\\ & && x \in S. \end{aligned}$$

In (P), the functions f_i, g_i map $S \subseteq R^n$ to R, $x = (x_1, \ldots, x_n) \in R^n$, and typically $S = R^n$ or S is a compact, or convex, or an open subset of R^n.

When $r = 0$, i.e., when all constraints in (P) are equalities, and suitable differentiability conditions are imposed, (P) becomes the calculus optimization problem that is a standard topic of most two-semester calculus sequences. Indeed, much of the work in nonlinear programming, which is not aimed at obtaining specific algorithms, is a continuation of classical investigations.

1. A concept of abstract duality. A new "twist" on (P) is provided by the fairly recent generalized dual problems, which can be abstractly formulated as follows, with $S \neq \varnothing$ an arbitrary set.

A class of functions $\Sigma \neq \varnothing$ is specified, and the domain of each function $\sigma \in \Sigma$ is to contain the image set

$$(1)\quad \mathrm{IM} = \left\{ (v_1, \ldots, v_r, u_1, \ldots, u_s) \in R^{r+s} \;\middle|\; \begin{array}{l} \text{for some } x \in S, \\ f_i(x) \leq v_i, i = 1, \ldots, r; \\ g_i(x) = u_i, i = 1, \ldots, s \end{array} \right\}.$$

Each $\sigma \in \Sigma$ maps into R, is required to have $\sigma(0, \ldots, 0) \geq 0$ and to be monotone nonincreasing (m.n.i.) in each v_i, i.e.,

$$(2)\quad \begin{array}{c} v_1' \geq v_1 \text{ and } \cdots \text{ and } v_r' \geq v_r \text{ implies} \\ \sigma(v_1', \ldots, v_r', u_1, \ldots, u_s) \leq \sigma(v_1, \ldots, v_r, u_1, \ldots, u_s). \end{array}$$

Then the dual problem is taken to be that of determining

$$(\mathrm{D})\quad v(D) = \max_{\sigma \in \Sigma} \inf_{x \in S} L(x, \sigma)$$

where the "generalized Langrangian" is defined by

$$(3)\quad L(x, \sigma) = f_0(x) - \sigma(f_1(x), \ldots, f_r(x), g_1(x), \ldots, g_s(x)).$$

For some dual problems, "max" is replaced by "sup" in (D).

What is sought for in dual problems (D) is equality of value with (P), i.e., if (P) is consistent and has finite value $v(\mathrm{P})$, one wants the duality equality to hold:

$$(\mathrm{DE})\quad v(\mathrm{P}) = v(\mathrm{D}).$$

One always is assured of one direction, namely

$$(4)\quad v(\mathrm{P}) \geq v(\mathrm{D})$$

by simply noting that, for any $\sigma \in \Sigma$, assuming (P) consistent we have

$$(5)\quad \begin{aligned} \inf_{x \in S} L(x, \sigma) &\leq \inf \left\{ L(x, \sigma) \;\middle|\; \begin{array}{l} x \in S \text{ and } f_i(x) \leq 0, i = 1, \ldots, r; \\ g_i(x) = 0, i = 1, \ldots, s \end{array} \right\} \\ &\leq \inf \left\{ f_0(x) \;\middle|\; \begin{array}{l} x \in S \text{ and } f_i(x) \leq 0, i = 1, \ldots, r; \\ g_i(x) = 0, i = 1, \ldots, s \end{array} \right\} \\ &= v(\mathrm{P}). \end{aligned}$$

The kind of dual problem (D), that we consider here, differs from its primal (P) in that its infimization is free of any functional constraints. Note, however, that the variables in (D) overlap with those of (P), and that (D) is a mixed ("max inf") problem rather than a pure ("max") problem. Nevertheless, for (P) a linear program, this dual (D) is the usual one (see §3 below).

Results regarding the dual program (D) all depend on some analysis of the *value function*, also called the perturbation function:

$$(6)\quad \mathrm{val}(v_1, \ldots, v_r, u_1, \ldots, u_s) = \inf \left\{ f_0(x) \;\middle|\; \begin{array}{l} x \in S \text{ and } f_i(x) \leq v_i, \\ i = 1, \ldots, r; \\ g_i(x) = u_i, i = 1, \ldots, s \end{array} \right\}.$$

If $(v_1, \ldots, v_r, u_1, \ldots, u_s) \notin \mathrm{IM}$, the value function is defined to be $+\infty$ for this argument. If $\mathrm{val}(v_1^0, \ldots, v_r^0, u_1^0, \ldots, u_s^0) = -\infty$ for any $(v_1^0, \ldots, v_r^0, u_1^0, \ldots, u_s^0) \in \mathrm{IM}$, clearly (DE) is impossible for $v(\mathrm{P})$ finite, regardless of how the family Σ is chosen, since then for any $\sigma \in \Sigma$ we have

$$
\begin{aligned}
\inf_{x \in S} L(x, \sigma) &\leqslant \inf\left\{ L(x, \sigma) \,\middle|\, \begin{array}{l} x \in S \text{ and } f_i(x) \leqslant v_i^0, i = 1, \ldots, r; \\ g_i(x) = u_i^0, i = 1, \ldots, s \end{array} \right\} \\
&\leqslant -\sigma(v_1^0, \ldots, v_r^0, u_1^0, \ldots, u_s^0) \\
&\quad + \inf\left\{ f_0(x) \,\middle|\, \begin{array}{l} x \in S \text{ and } f_i(x) \leqslant v_i^0, i = 1, \ldots, r; \\ g_i(x) = u_i^0, i = 1, \ldots, s \end{array} \right\} \\
&= -\infty.
\end{aligned} \tag{7}
$$

In all further discussion in §1, we use the *blanket assumption* that $\mathrm{val}(-, -)$ has no values of $-\infty$. This is a substantive restriction, since it fails for the convex program with finite value

$$\inf(-y) \text{ subject to } (x^2 + y^2)^{1/2} - x \leqslant 0. \tag{8}$$

Using this blanket assumption, in the case that $v(\mathrm{P})$ is finite, it can be shown that (DE) holds if and only if Σ contains at least one support $\sigma^* \in \Sigma$. By a *support* we mean a m.n.i. function with $\sigma^*(0, \ldots, 0) \geqslant 0$ satisfying identically

$$
\begin{aligned}
\mathrm{val}(v_1, \ldots, v_r, u_1, \ldots, u_s) - \sigma^*(v_1, \ldots, v_r, u_1, \ldots, u_s) &\geqslant v(\mathrm{P}), \\
\text{for} (v_1, \ldots, v_r, u_1, \ldots, u_s) &\in \mathrm{IM}.
\end{aligned} \tag{9}
$$

In fact, the optima to (D), when (DE) holds, are precisely such supports σ^*. Moreover, if x^* is an optimum of (P) and σ^* is an optimum in (D), and (DE) holds, then the following saddle-point condition holds:

$$L(x, \sigma^*) \geqslant L(x^*, \sigma^*) \geqslant L(x^*, \sigma) \quad \text{for all } x \in S, \sigma \in \Sigma. \tag{SP}$$

A converse to (SP) is also valid, in the following sense. If Σ contains the function Z which is identically zero, and if the unboundedness condition

$$\sup_{\sigma \in \Sigma} L(x, \sigma) = +\infty \tag{UBD}$$

holds whenever x is not feasible in (P), then (SP) implies that x^* is optimal in (P), and σ^* is optimal in (D), and (DE) holds.

(All claims in the last several paragraphs are exercises in elementary algebra!)

Our approach above follows Gould [8] closely. The main point here is this, that in order to obtain (DE) and (SP), etc., one does not need to know the graph $v_0 = \mathrm{val}(-, -)$ exactly: one needs instead to simply know one function $v_0 = v(\mathrm{P}) + \sigma(-, -)$ which lies entirely below this graph and touches it at $(v(\mathrm{P}), 0, 0)$.

Some of the ideas occurring in connection with generalized Lagrangians appear in a simpler form in Everett's seminal paper [6], where the ordinary linear Lagrangian (see §2 below) is discussed from the perspective of linear affine supports for $\mathrm{val}(-, -)$.

1.1. *Sensitivity analysis*. Suppose that $\sigma \in \Sigma$ and that

$$\inf_{x \in S} L(x, \sigma) = L(x^0, \sigma) \tag{10}$$

with $x^0 \in S$. Then it is not hard to show that x^0 is optimal for the following problem, which is like (P) except for the change in the right-hand-side:

$$\text{(P)}' \qquad \begin{array}{ll} \text{infimize} & f_0(x) \\ \text{subject to} & f_i(x) \leqslant v_i^0, \quad i = 1, \ldots, r, \\ & g_i(x) = u_i^0, \quad i = 1, \ldots, s. \end{array}$$

In (P)′, we have set $v_i^0 = f_i(x^0)$ and $u_i^0 = g_i(x^0)$ for all i.

In the course of solving (P), often several problems of type (10) are solved, and by the previous remark one thereby obtains some information about changes in the optimum to (P) as the right-hand-side is varied. Since usually only a few constraints of (P) are exact, while other constraints can often be "fudged" somewhat in the applications, this sensitivity information can be significant and useful.

Gould provides sensitivity results of this type **[8]** in a similarly general setting. Everett appears to be the first to have emphasized the importance of this sensitivity information for the usual linear Lagrangian (see §2 below).

2. The linear Langragian and convex programs. Most of what is known about dual problems (D) is for the convex programming problem, in which each function f_i in (P) is convex and each function g_i is linear affine on R^n. Here the function class Σ is taken to be the class of m.n.i. linear functions, i.e.

$$\sigma(v_1, \ldots, v_r, u_1, \ldots, u_s) = \sum_{i=1}^{r} \lambda_i v_i + \sum_{i=1}^{s} \theta_i u_i, \qquad \lambda_i \leqslant 0, i = 1, \ldots, r. \tag{11}$$

Note that (UBD) holds for the class (11).

For this case we have the following sufficient condition for (DE) in convex programming with the linear class (11), when (P) is consistent and has finite value $v(\mathrm{P})$: (1) S is convex; (2) There is a "Slater point" x^0 in the relative interior of S, which satisfies all the contraints of (P), for which the inequality

$$f_i(x^0) < 0 \tag{12}$$

holds for all f_i that are not linear affine, $i \neq 0$. As a consequence, when all functions f_i, $i \neq 0$, and g_i are linear affine, the (DE) always holds if $v(\mathrm{P})$ exists and is finite.

A simple example from the literature shows why some kind of "constraint qualification" like (12) is, in general, needed for (DE) even in the convex case. Take as (P) the convex program

$$\text{(8)}' \qquad \inf e^{-y} \text{ subject to } (x^2 + y^2)^{1/2} - x \leqslant 0.$$

Clearly, this program (8)′ has $v(\mathrm{P}) = 1$, since $y = 0$ is forced by the constraint. However, for any scalar $\lambda = \lambda_1 \leqslant 0$,

$$\inf_{(x,y) \in R^2} \left(e^{-y} - \lambda\left((x^2 + y^2)^{1/2} - x\right)\right) \leqslant 0, \tag{13}$$

hence $v(\mathrm{D}) = 0$ is attained at $\lambda = 0$.

The sufficient condition (12) given above, for (DE) to hold in convex programs, yields an alternate version of the Kuhn-Tucker Saddle Point Theorem. This theorem states that, under certain hypotheses, to every optimum x^* of (P) there is an optimum σ^* of (D) such that (SP) holds; and conversely, (SP) implies that x^* and σ^* are optima to their respective problems. The theorem has provided much of the impetus for the work in duality, and is closely related to the Fenchel Duality Theorem which is stated in terms of convex conjugate functions (see Magnanti's note **[14]**).

The textbooks of Rockafellar **[21]**, and Stoer and Witzgall **[24]** provide well-known improvements of Fenchel's result that are due to the authors, in the conjugate function framework, as well as extensive information about the linear Lagrangian (11) in the convex case.

3. The quadratic Lagrangian. The requirement (DE) for a "correct" dual problem (D) is weaker than requiring the optimal "activity levels" x^* emerging from (P) (when the optimum in (P) is attained) to be those also optimal in

$$(\mathrm{D})_{\sigma^*} \qquad \inf_{x \in S} L(x, \sigma^*)$$

whenever σ^* is optimal in the dual, i.e.,

$$v(\mathrm{D}) = \inf_{x \in S} L(x, \sigma^*). \tag{14}$$

To be specific, an optimum x^* to (P) does indeed provide an optimum to $(\mathrm{D})_{\sigma^*}$ when (DE) holds (recall (SP)), but the converse may fail: some optima to $(\mathrm{D})_{\sigma^*}$ may not even be feasible in (P).

Consider, for example, the linear class (11) in the linear programming case

$$\begin{array}{lll} \text{infimize} & cx & \\ \text{subject to} & a^i x - b_i \leqslant 0, & i = 1, \ldots, r, \\ & g^i x - h_i = 0, & i = 1, \ldots, s \end{array} \tag{15}$$

with c, a^i, $g^i \in R^n$ and b_i, $h_i \in R$. Here we have $S = R^n$ and

$$L(x, \sigma) = \left(c - \sum_{i=1}^{r} \lambda_i a^i - \sum_{i=1}^{s} \theta_i g^i\right)x + \left(\sum_{i=1}^{r} \lambda_i b_i + \sum_{i=1}^{s} \theta_i h_i\right), \tag{16}$$

and therefore (DE) implies simply

$$\begin{gathered} c - \sum_{i=1}^{r} \lambda_i a^i - \sum_{i=1}^{s} \theta_i g^i = 0, \qquad \lambda_i \leqslant 0 \text{ for } i = 1, \ldots, r, \\ \sum_{i=1}^{r} \lambda_i b_i + \sum_{i=1}^{s} \theta_i h_i = v(\mathrm{P}) \end{gathered} \tag{15}$$

which places no constraints on $x \in R^n$ whatever. Here (15)′ is simply the ordinary dual to (15) in the linear programming theory (note by (5) that the last equation in (15)′ is equivalent to maximizing $\Sigma_i \lambda_i b_i + \Sigma_i \theta_i h_i$).

The development of generalized dual problems (D) and generalized Langrangians (3) has been prompted by both the desire to treat nonconvex instances of (P) à lá Kuhn and Tucker, and to improve upon the properties of the ordinary linear Lagrangian (3), (11) for the convex programming problem. Intuitively speaking, what one cannot accomplish with linear forms (11) might well be possible with quadratic forms; hence we consider the class Σ of all quadratic functions

$$(17)' \quad \sigma'(v_1, \ldots, v_r, u_1, \ldots, u_s) = \sum_{i=1}^{r} (\lambda_i v_i - \rho v_i^2) + \sum_{i=1}^{s} (\theta_i u_i - \rho u_i^2)$$

where $\rho \geqslant 0$. Now (17)′ is not quite correct, since it is not m.n.i.; to make σ' into an m.n.i. function σ we set

$$(18) \quad \sigma(v_1, \ldots, v_r, u_1, \ldots, u_s) = \sup\{\sigma'(v_1', \ldots, v_r', u_1, \ldots, u_s) \mid v_i' \geqslant v_r, i = 1, \ldots, r\}$$

and an easy calculation gives, for $\rho > 0$,

$$(17) \quad \sigma(v_1, \ldots, v_r, u_1, \ldots, u_s) = \sum_{i=1}^{r} \left\{ \lambda_i \max\left\{ v_i, \frac{\lambda_i}{2\rho} \right\} - \rho \max^2\left\{ v_i, \frac{\lambda_i}{2\rho} \right\} \right\} + \sum_{i=1}^{s} (\theta_i u_i - \rho u_i^2).$$

If $\rho = 0$ then $\sigma = \sigma'$ provided all $\lambda_i \leqslant 0$; otherwise σ is infinite (hence invalid).

The earliest results regarding (17) were obtained by Arrow and Solow [**2**], Hestenes [**12**], [**13**], Powell [**19**], and Haarhoff and Buys [**11**]. Rockafellar [**22**], [**23**], has done an extensive investigation of this quadratic Lagrangian from the perspective of the dual program (D), and has established that it can obtain the duality equality (DE) in many nonconvex settings where (DE) is false for the linear Lagrangian (11). When (DE) holds for a convex program Rockafellar has shown that the optima in $(\mathrm{D})_{\sigma^*}$ are the optima of (P), when the optimum σ^* of the dual has $\rho > 0$ (as can be arranged whenever (DE) holds).

By choosing all $\theta_i = 0$ in the case $r = 0$, we obtain one "penalty method" for solving (P) that has been attributed to Courant.

For a study of a large family of generalized Lagrangians, including the quadratic Lagrangian (17), see [**1**] and [**15**]. In particular, Mangasarian provides several classes Σ of differentiable m.n.i. functions σ, which are useful in problems of type (10).

3.1. *Quadratic Lagrangians and derivatives of* val(−, −). The reason for considering quadratic forms (17)′ with purely diagonal entries

$$-\rho\left(\sum v_i^2 + \sum u_i^2\right)$$

as the quadratic part, is that a support need only get under the graph of $v_0 = \mathrm{val}(-, -)$; hence, whatever can be accomplished with a general quadratic form can be done by (17)′.

By the same reasoning, if $\lambda_1^*, \ldots, \lambda_r^*, \theta_1^*, \ldots, \theta_s^*, \rho^*$ are the parameters of a quadratic Lagrangian (17) that is optimal in (D), so are $\lambda_1^*, \ldots, \lambda_r^*, \theta_1^*, \ldots, \theta_s^*, \rho$ for $\rho > \rho^*$. In particular, if (P) is a convex program for which (DE) holds for the linear Lagrangian, and $\lambda_1^*, \ldots, \lambda_r^*, \theta_1^*, \ldots, \theta_s^*$ are the parameters of an optimum to (D) with the Lagrangian (11), then $\lambda_1^*, \ldots, \lambda_r^*, \theta_1^*, \ldots, \theta_s^*, \rho$ are the parameters of an optimum in (D) for the quadratic Lagrangian for any $\rho > 0$.

A converse to the result of the last paragraph also holds for a convex program (P). In this case, it is easy to show that val(−, −) is an extended convex function and, hence, that its epigraph

$$(19)\ \text{epi(val)} = \{(v_0, v_1, \ldots, v_r, u_1, \ldots, u_s) | v_0 \geq \text{val}(v_1, \ldots, v_r, u_1, \ldots, u_s)\}$$

is a convex set. Suppose (DE) holds and σ^* with parameters $\lambda_1^*, \ldots, \lambda_r^*$, $\theta_1^*, \ldots, \theta_s^*$, ρ^* is optimal in (D). Then since $\text{val}(-,-) - \sigma^*(-,-) \geq v(\text{P})$ is an identity, it follows that epi(val) has no inner points in common with the convex hypograph given by

$$(20)\quad \text{hypo}(\sigma') = \left\{ (v_0, v_1, \ldots, v_r, u_1, \ldots, u_s) \middle| \begin{array}{l} v_0 \leq \sum_{i=1}^{r} (\lambda_i^* v_i - \rho^* v_i^2) \\ \quad + \sum_{i=1}^{s} (\theta_i^* u_i - \rho^* u_i^2) + v(\text{P}) \end{array} \right\}.$$

By standard results, there is a hyperplane separating epi(val) from hypo(σ') (not strictly) which touches hypo(σ') at (v(P), 0, 0). From the quadratic nature of σ' in (17)$'$, this hyperplane can only be the linear part of (17)$'$; therefore the linear Lagrangian (11) also yields (DE) with parameters $\lambda_1^*, \ldots, \lambda_r^*, \theta_1^*, \ldots, \theta_s^*$. See **[21]**.

Whether or not the constraints of (P) are convex, the "near disjointness" of epi(val) from hypo(σ') when (DE) holds, together with the nature of the unique hyperplane touching hypo(σ') at (v(P), 0, 0), yields uniform lower bounds on the directional derivatives of val$(-,-)$ at (0, 0), etc.

For instance, it is easy to see that, when val$(-,-)$ has a derivative at (0, 0) and (DE) holds, this derivative must be $\Sigma\lambda_i^* v_i + \Sigma\theta_i^* u_i$. Now quite frequently, $f_i(x) = f_i^0(x) - b_i^0$, where b_i^0 is the quantity available of the ith factor of "production", and $f_i^0(x)$ is the quantity of the ith factor needed to maintain "activity level x." Given this economic interpretation of (P), if $\pi_i > 0$ is the (externally given) current market price of the ith factor, we have the following diagonostic obtained from the directional rates of change of val$(-,-)$: if $\pi_i < (-\lambda_i^*)$, then it is advantageous to purchase at least some small additional amount of factor i; if $\pi_i > (-\lambda_i^*)$, it is advantageous to sell a small amount of factor i.

From a global perspective, this diagnostic may seem odd, since it refers to a problem (P) with the right-hand-sides constrained by fixed factor availabilities b_i^0. One might view it as advantageous to have the factor availabilities introduced as variables, and to have their levels, along with those of the activities, optimally set by a larger programming problem. However, this freedom to choose levels may not be present in models of short-run problems; and the model (P) itself may fail to be globally accurate, due to many factors, such as local linearization of a (globally little known) demand curve. It is therefore of value to have a diagnostic that provides directions for local improvement. Moreover, the diagnostic is particularly useful since in most applications very many λ_i^* are zero. This diagnostic has been fruitfully employed by managers in connection with the linear program (15) for up to two decades in some industries.

4. Duality and stationarity conditions. When (DE) holds, for every support σ^*, we have the "complementary slackness" condition

$$\text{(CS)}\qquad \sigma^*(f_1(x^*), \ldots, f_r(x^*), 0, \ldots, 0) = 0$$

which is part of what are called the "Kuhn-Tucker necessary conditions" on derivatives, the remaining conditions being

$$\text{(KT)} \quad \begin{aligned} &\operatorname{grad} f_0(x^*) - \sum_{i=1}^{r} \frac{\partial \sigma^*}{\partial v_i}(v_1^*, \ldots, v_r^*, 0, \ldots, 0) \operatorname{grad} f_i(x^*) \\ &\quad - \sum_{i=1}^{s} \frac{\partial \sigma^*}{\partial u_i}(v_1^*, \ldots, v_r^*, 0, \ldots, 0) \operatorname{grad} g_i(x^*) = 0 \end{aligned}$$

when the chain-rule applies. In (CS) and (KT), x^* is any optimum in (P) that is in the interior of S, and $v_i^* = f_i(x^*)$, $i = 1, \ldots, r$.

(CS) is an algebraic exercise. To obtain (KT), note that $x = x^*$ is a minimum to $L(x, \sigma^*)$ by (SP). By differentiating $L(x, \sigma^*)$ at $x = x^*$, we obtain (KT); when the indicated derivatives do not exist, useful information of the type (KT) can often still be obtained, from noting that all directional derivatives of $L(x, \sigma^*)$ at $x = x^*$ are $\geqslant 0$.

For the linear (11) class as Σ, (CS) is equivalent to the logical conditions

$$\text{(21)} \qquad f_i(x^*) < 0 \text{ implies } \lambda_i = 0, \qquad i = 1, \ldots, r.$$

I.e., only those inequality constraints that are "tight" at the optimum ($f_i(x^*) = 0$) can occur with nonzero $\partial\sigma^*/\partial u_i$ in (KT).

A frequently-used approach for the numerical solution of (P) is to solve the necessary conditions (CS), (KT) for the linear Lagrangian (11) (often solving together with the functional constraints of (P)), thus by passing the unconstrained optimizations of type (10). For $r = 0$, this is the procedure that we advocate to our undergraduate students; except for homework exercises deliberately chosen for algebraic simplifications that render solutions trivial, this technique is simply a partial reduction of (P) to solutions of nonlinear equations. The latter need not be that easy to obtain! As in our calculus classes, second-order conditions are used to partially determine when the necessary conditions are also sufficient for (local) optima.

4.1. *Stationarity conditions without duality.* A primary reason for interest in these necessary conditions (CS) and (KT), is that their applicability is potentially broader than that of the whole development for the dual problem (D), since these conditions remain necessary for a much wider class of optimization problems than those for which dual programs have been established. For instance, for $s = 0$, by a result of John, both (CS) and (KT) are valid for the class Σ of linear functions (11), under appropriate constraint qualifications, if only the functions f_i are continuously differentiable–no convexity is assumed here.

Some limitations of the approach via stationarity conditions alone, are the lack of sensitivity analysis, the fact that programs (P) arising in practice may involve nondifferentiable functions (see **[25]** and the volume in which it appears), and the difficulties in determining which solutions to (KT) provide optima for (P).

Conceptually, the abstract dual problems (D) go part of the way in explaining why (KT) is valid for the linear class Σ in so many settings. In fact, the partial derivatives $\partial\sigma^*/\partial v_i$ and $\partial\sigma^*/\partial u_i$ in the "general" formula (KT) are simply scalars λ_i and θ_i; since σ^* is m.n.i, we will also have all $\partial\sigma^*/\partial u_i \leqslant 0$. Moreover, for this derivation one needs x^* to be simply the optimum to a "local" dual problem (D), i.e., a dual where S is an open neighborhood with $x^* \in S$.

Chapter 5 of Hestenes' textbook is devoted to interrelations between (CS) and (KT) for the linear class Σ, and local quadratic dual programs (D) with Σ the quadratic class. A strengthened version of the second-order sufficient conditions, for solutions to linear (KT) to be local optima to (P), is shown to insure that a local quadratic dual problem is valid at an interior point x^*. Similar results have been obtained by Mangasarian **[15]** and by Rockafellar **[22]**.

Rockafellar provides other valuable information on local quadratic dual problems when $s = 0$, under very weak assumptions (e.g., f_0 continuous). He shows that, for (DE) to hold in a local dual problem, it is both necessary and sufficient that there exists a function $\sigma'(v_1, \ldots, v_r)$ such that the identity

$$\text{val}(v_1, \ldots, v_r) - \sigma'(v_1, \ldots, v_r) \geqslant v(\mathrm{P}) \tag{22}$$

holds in some neighborhood of the origin, where σ' is of class $C^{(2)}$ and $\sigma'(0) = 0$. Furthermore, if one changes "max" to "sup" in (D), the requirement that σ' is of class $C^{(2)}$ can be replaced by simply requiring σ' to be continuous.

For a different perspective on the linear Kuhn-Tucker conditions, see **[9]**.

5. Other comments on the textbook. The author's text provides a treatment of the dual problem (D) for the convex programming problem with the linear class (11), and states a widely-used version of the Kuhn-Tucker Saddle Point Theorem. While building to this result, the author establishes a number of basic results on polarity, linear inequality systems, and convex cones, which are of broad applicability in all areas of mathematical programming.

A highlight of the author's results on the quadratic Lagrangian is his statement of an algorithm for (P) when S is compact, $r = 0$, and certain additional assumptions hold. This algorithm is his well-known "method of multipliers", in which a sequence of problems of the type

$$\inf_{x \in S} L(x, \rho_0 + \Delta_q), \tag{23}$$

$q = 1, 2, \ldots$, are successively solved, with $\Delta_q \geqslant \Delta_0 > 0$ and $\theta_1^{(1)}, \ldots, \theta_s^{(q)}$ updated by a rule of the form

$$\theta_i^{(q+1)} = \theta_i^{(q)} - 2\Delta_q g_i(x^{(q+1)}). \tag{24}$$

In (24), $x^{(q+1)}$ is an optimum to (23). If $\rho^{(0)}$ is properly chosen, the sequence $x^{(q)}$ converges to an optimum to (P). Clearly, if the work on generalized duality and generalized Lagrangians is to be of value in practice, we shall need more information and positive results on methods like the author's (24). For an extension of Hestenes' algorithm to the case $r > 0$, see Mangasarian **[15]**.

In addition to the main developments of the text discussed above, the author provides considerable information on quadratic forms, a subject which is helpful throughout nonlinear programming. Also, the author uses the quadratic Lagrangian (17) as a point of departure to discuss a few results on penalty methods. Finsler's Lemma is established from the perspective of penalty functions.

The textbook has been designed to appeal to users of optimization theory in mathematics and related fields, as well as to specialists. The elementary versions of many results are often presented before full generality is obtained, and proofs are entirely rigorous. In terms of developing motivation for results, by means of a series of examples which illustrate many facets of the problems discussed, the text is unexcelled in its area.

6. A few general references. For an overview of nonlinear programming which discusses important topics we have omitted here, the reader may wish to consult Mangasarian's survey **[16]**; also Oettli **[17]** has an extensive bibliography of the area.

Due to space limitations, we have not discussed the considerable literature that is concerned with techniques for unconstrained function minimization, even though these techniques are essential for the solution of problems like (10) when $S = R^n$. This nonlinear programming topic is not directed toward extending classical results. Instead, the focus is on computational issues, such as avoiding the n^2 function evaluations of a Hessian matrix at each iteration, which occurs in the "obvious" iterative method for solving minimizations in twice continuously differentiable functions (i.e., moving toward the minimum of the approximating quadratic form). In designing algorithms which are computationally tractable, and have excellent asymptotic convergence behavior, one encounters considerable mathematical difficulty toward establishing their convergence and convergence rate. For a survey of this topic, see **[20]**.

Regarding the stationarity approach to nonlinear programming, the text **[18]** is a general reference for many solution techniques for nonlinear equations and for nonlinear minimizations. Typically, the techniques discussed in **[18]** and **[20]** yield local convergence, so that the user must supply a "good" starting point for these iterative methods.

Research in *complementarity theory* and fixed point computation, another area of mathematical programming, has lead to a new algorithms for solving certain nonlinear equations, which are globally convergent to at least one solution to these equations. Eaves and Scarf **[5]** have achieved a unification of many results and techniques, which is extremely concise and lucid, and can also serve as an introduction to that area. Earlier unifications (e.g., **[10]**) stressed the use of combinatorial principles, such as an "algorithmic version" of a highly abstracted Sperner's lemma, as the main device toward substantially limiting the search which is (implicitly) involved in the complementarity techniques. In the treatment of Eaves and Scarf, these discrete combinatorial principles, as well as the results previously derived from them, are, surprisingly enough, all obtained as consequences of entirely elementary results on solutions of piecewise-linear equations on the continuum. For some complementarity algorithms for convex programs, see **[10]** and the references cited there.

As we mentioned above, the convex programming problem is one special case of (P) for which a substantial body of information exists, which of course utilizes arguments that are quite specialized to the case studied. Another well-understood special case of (P) is *geometric programming* **[4]**, which was motivated by a study of the kind of nonlinearities occurring in chemical processes. The characteristic format of a geometric program is the minimization of a sum of positive monomials (with possibly negative or fractional exponents) in positive variables, subject to several functions of the same type not exceeding unity. Clearly, this format is broad enough to allow a representation of nonlinear functions by regression fittings on monomials, provided that the positivity of monomial coefficients is realized. Geometric programming has been used in applications as distant from chemistry as marketing. The geometric inequality allows a partial reduction of geometric programs to linear programs, sufficient to permit a specialized dual construction and to provide the foundation for several computational approaches; see [3].

The textbook [7] provides information on several penalty methods. However, the literature on these methods is very extensive, and much has been done since the publication of [7]. Advances in this area are usually closely related to work on augmented or generalized Lagrangians, so an examination of the references in our references below will lead to much of the penalty-method literature.

ACKNOWLEDGEMENT. We thank C. E. Blair for several suggestions on an earlier draft of this review.

REFERENCES

1. K. J. Arrow, F. J. Gould and S. M. Howe, *A general saddle point result for constrained optimization*, Math. Programming **5** (1973), 225–234. MR **48** #7983.

2. K. J. Arrow and R. M. Solow, *Gradient methods for constrained maxima with weakened assumptions*, Studies in Linear and Nonlinear Programming (K. J. Arrow, L. Hurwicz and H. Uzawa, editors), vol. II, Stanford Univ. Press, Stanford, Calif., 1958, pp. 166–176. MR **21** #7115.

3. R. J. Duffin, *Linearizing geometric programs*, SIAM Rev. **12** (1970), 211–227. MR **44** #3702.

4. R. J. Duffin, E. L. Peterson and C. Zener, *Geometric programming: Theory and application*, Wiley, New York and London, 1967. MR **35** #5225.

5. B. C. Eaves and H. Scarf, *The solution of systems of piecewise linear equations*, Mathematics of Operations Research **1** (1976), 1–27.

6. H. Everett III, *Generalized Lagrange multiplier method for solving problems of optimum allocation of resources*, Operations Res. **11** (1963), 399–417. MR **27** #2340.

7. A. V. Fiacco and G. P. McCormick, *Nonlinear programming: Sequential unconstrained minimization techniques*, Wiley, New York and London, 1968. MR **39** #5152.

8. F. J. Gould, *Extensions of Lagrange multipliers in nonlinear programming*, SIAM J. Appl. Math. **17** (1969), 1280–1297. MR **41** #8031.

9. F. J. Gould and J. W. Tolle, *A necessary and sufficient qualification for constrained optimization*, SIAM J. Appl. Math. **20** (1971), 164–172. MR **46** #3090.

10. ———, *A unified approach to complementarity in optimization*, Report 7312, Graduate School of Business, Univ. of Chicago, February 1973.

11. P. C. Haarhoff and J. D. Buys, *A method for the optimization of a nonlinear function subject to nonlinear constraints*, Comput. J. **13** (1970), 178–184. MR **43** #2832.

12. M. R. Hestenes, *Multiplier and gradient methods*, Computing Methods in Optimization Problems, Vol. 2 (Proc. Conf., San Remo, 1968), Academic Press, New York, 1969, pp. 143–163. MR **41** #7508.

13. ———, *Multiplier and gradient methods*, J. Optimization Theory Appl. **4** (1969), 303–320. MR **42** #6690.

14. T. L. Magnanti, *Fenchel and Lagrange duality are equivalent*, Math. Programming **7** (1974), 253–258. MR **51** #5012.

15. O. L. Mangasarian, *Unconstrained Lagrangians in nonlinear programming*, SIAM J. Control **13** (1975), 772–791.

16. ———, *Nonlinear programming theory and computation*, Handbook of Operations Research, Chapter 6 (Elmaghraby and Moder, editors), van Nostrand Reinhold, New York, 1973.

17. W. Oettli, *Bibliography on nonlinear programming*, Mathematische Optimierung-Grundlagen und Verfahren (Blum and Oettli, editors), Springer-Verlag, Berlin and New York, 1975.

18. J. M. Ortega and W. C. Rheinboldt, *Iterative solution of nonlinear equations in several variables*, Academic Press, New York, 1970. MR **42** #8686.

19. M. J. D. Powell, *A method for nonlinear constraints in minimization problems*, Optimization

(Sympos., Univ. Keele, 1968) (Fletcher, editor), Academic Press, London, 1969, pp. 283–298. MR **42** #7284.

20. ——— , *A view of unconstrained optimization*, C.S.S. 14, Computer Science and Systems Division, Atomic Energy Research Establishment, Harwell, Oxfordshire, England, January 1975.

21. R. T. Rockafellar, *Convex analysis*, Princeton Univ. Press, Princeton, N. J., 1970. MR **43** #445.

22. ——— , *Augmented Lagrange multiplier functions and duality in nonconvex programming*, SIAM J. Control **12** (1974), 268–285.

23. ——— , *A dual approach to solving nonlinear programming problems by unconstrained optimization*, Math. Programming **5** (1973), 354–373. MR **51** #7634.

24. J. Stoer and C. Witzgall, *Convexity and optimization in finite dimensions*. I, Springer-Verlag, Berlin and New York, 1970. MR **44** #3707.

25. P. Wolfe, *A method of conjugate subgradients for minimizing nondifferentiable functions*, Mathematical Programming Study, no. 3, Nondifferentiable Optimization, North-Holland, Amsterdam, 1975, pp. 145–173.

ROBERT G. JEROSLOW

BULLETIN OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 83, Number 3, May 1977

Approximation of functions of several variables and imbedding theorems, by S. M. Nikol'skiĭ, Die Grundlehren der mathematischen Wissenshaften, Band 205, Springer-Verlag, New York, Heidelberg, Berlin, 1975, viii + 420 pp., $46.50.

This book is interesting and important. Although it advertises itself in the introduction as a compact exposition of a number of fundamental questions that have been brought to completion, it is no such thing. In particular, no treatise of 400 plus pages (480 pp. in the original) can be called compact. It is, however, a fascinating description of a mathematical adventure that failed in its main goal but has produced, and continues to produce, much excellent mathematics.

Unfortunately, the editors of this series have given us a seriously flawed English version of the Russian edition that was published in 1969. It has been so badly done that I will take up the question of the translation and its editing in some detail, after I first discuss the text proper.

The book's topic is classes of functions which satisfy various smoothness or differentiability conditions and the determination of which spaces are contained, continuously, in which other spaces or can be mapped, continuously, into such other spaces by appropriate restriction or extension maps. The method used is the method of best approximation by trigonometric polynomials and in the nonperiodic case, by their analogues, the restriction to real space of entire functions of exponential type. We are told in this book how far Nikol'skiĭ and his colleagues got with this topic by the mid 1960's.

The classes of spaces considered are denoted W, H, B, and L spaces. These are decorated in various ways to indicate the amount of smoothness, the L_p-norm in which the smoothness is measured, the space on which the functions are defined, and the "directions" in which the smoothness is measured. Refinements of the notation permit consideration of "anisotropic" spaces where one has different L_p-norms and degrees of smoothness for each direction. Our discussion will mostly be restricted to the isotropic case.

Most of the material in the text is concerned with the case where the functions are defined on all of a Euclidean space $\mathbf{R}^n$. There is some discussion