16

Conjugate Duality and Optimization

R. TYRRELL ROCKAFELLAR
University of Washington, Seattle

CBMS-NSF
REGIONAL CONFERENCE SERIES
IN APPLIED MATHEMATICS

SPONSORED BY
CONFERENCE BOARD OF
THE MATHEMATICAL SCIENCES

SUPPORTED BY
NATIONAL SCIENCE
FOUNDATION

Conjugate Duality and Optimization

CBMS-NSF REGIONAL CONFERENCE SERIES IN APPLIED MATHEMATICS

A series of lectures on topics of current research interest in applied mathematics under the direction of the Conference Board of the Mathematical Sciences, supported by the National Science Foundation and published by SIAM.

GARRETT BIRKHOFF, *The Numerical Solution of Elliptic Equations*
D. V. LINDLEY, *Bayesian Statistics, A Review*
R. S. VARGA, *Functional Analysis and Approximation Theory in Numerical Analysis*
R. R. BAHADUR, *Some Limit Theorems in Statistics*
PATRICK BILLINGSLEY, *Weak Convergence of Measures: Applications in Probability*
J. L. LIONS, *Some Aspects of the Optimal Control of Distributed Parameter Systems*
ROGER PENROSE, *Techniques of Differential Topology in Relativity*
HERMAN CHERNOFF, *Sequential Analysis and Optimal Design*
J. DURBIN, *Distribution Theory for Tests Based on the Sample Distribution Function*
SOL I. RUBINOW, *Mathematical Problems in the Biological Sciences*
P. D. LAX, *Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves*
I. J. SCHOENBERG, *Cardinal Spline Interpolation*
IVAN SINGER, *The Theory of Best Approximation and Functional Analysis*
WERNER C. RHEINBOLDT, *Methods of Solving Systems of Nonlinear Equations*
HANS F. WEINBERGER, *Variational Methods for Eigenvalue Approximation*
R. TYRRELL ROCKAFELLAR, *Conjugate Duality and Optimization*
SIR JAMES LIGHTHILL, *Mathematical Biofluiddynamics*
GERARD SALTON, *Theory of Indexing*
CATHLEEN S. MORAWETZ, *Notes on Time Decay and Scattering for Some Hyperbolic Problems*
F. HOPPENSTEADT, *Mathematical Theories of Populations: Demographics, Genetics and Epidemics*
RICHARD ASKEY, *Orthogonal Polynomials and Special Functions*
L. E. PAYNE, *Improperly Posed Problems in Partial Differential Equations*
S. ROSEN, *Lectures on the Measurement and Evaluation of the Performance of Computing Systems*
HERBERT B. KELLER, *Numerical Solution of Two Point Boundary Value Problems*
J. P. LASALLE, *The Stability of Dynamical Systems* - Z. ARTSTEIN, *Appendix A: Limiting Equations and Stability of Nonautonomous Ordinary Differential Equations*
D. GOTTLIEB AND S. A. ORSZAG, *Numerical Analysis of Spectral Methods: Theory and Applications*
PETER J. HUBER, *Robust Statistical Procedures*
HERBERT SOLOMON, *Geometric Probability*
FRED S. ROBERTS, *Graph Theory and Its Applications to Problems of Society*
JURIS HARTMANIS, *Feasible Computations and Provable Complexity Properties*
ZOHAR MANNA, *Lectures on the Logic of Computer Programming*
ELLIS L. JOHNSON, *Integer Programming: Facets, Subadditivity, and Duality for Group and Semi-Group Problems*
SHMUEL WINOGRAD, *Arithmetic Complexity of Computations*
J. F. C. KINGMAN, *Mathematics of Genetic Diversity*
MORTON E. GURTIN, *Topics in Finite Elasticity*
THOMAS G. KURTZ, *Approximation of Population Processes*

JERROLD E. MARSDEN, *Lectures on Geometric Methods in Mathematical Physics*
BRADLEY EFRON, *The Jackknife, the Bootstrap, and Other Resampling Plans*
M. WOODROOFE, *Nonlinear Renewal Theory in Sequential Analysis*
D. H. SATTINGER, *Branching in the Presence of Symmetry*
R. T MAM, *Navier-Stokes Equations and Nonlinear Functional Analysis*
MIKLÓS CSÖRGO, *Quantile Processes with Statistical Applications*
J. D. BUCKMASTER AND G. S. S. LUDFORD, *Lectures on Mathematical Combustion*
R. E. TARJAN, *Data Structures and Network Algorithms*
PAUL WALTMAN, *Competition Models in Population Biology*
S. R. S. VARADHAN, *Large Deviations and Applications*
KIYOSI ITÔ, *Foundations of Stochastic Differential Equations in Infinite Dimensional Spaces*
ALAN C. NEWELL, *Solitons in Mathematics and Physics*
PRANAB KUMAR SEN, *Theory and Applications of Sequential Nonparametrics*
LÁSZLÓ LOVÁSZ, *An Algorithmic Theory of Numbers, Graphs and Convexity*
E. W. CHENEY, *Multivariate Approximation Theory: Selected Topics*
JOEL SPENCER, *Ten Lectures on the Probabilistic Method*
PAUL C. FIFE, *Dynamics of Internal Layers and Diffusive Interfaces*
CHARLES K. CHUI, *Multivariate Splines*
HERBERT S. WILF, *Combinatorial Algorithms: An Update*
HENRY C. TUCKWELL, *Stochastic Processes in the Neurosciences*
FRANK H. CLARKE, *Methods of Dynamic and Nonsmooth Optimization*
ROBERT B. GARDNER, *The Method of Equivalence and Its Applications*
GRACE WAHBA, *Spline Models for Observational Data*
RICHARD S. VARGA, *Scientific Computation on Mathematical Problems and Conjectures*
INGRID DAUBECHIES, *Ten Lectures on Wavelets*
STEPHEN F. MCCORMICK, *Multilevel Projection Methods for Partial Differential Equations*
HARALD NIEDERREITER, *Random Number Generation and Quasi-Monte Carlo Methods*
JOEL SPENCER, *Ten Lectures on the Probabilistic Method, Second Edition*
CHARLES A. MICCHELLI, *Mathematical Aspects of Geometric Modeling*
ROGER T MAM, *Navier–Stokes Equations and Nonlinear Functional Analysis, Second Edition*
GLENN SHAFER, *Probabilistic Expert Systems*
PETER J. HUBER, *Robust Statistical Procedures, Second Edition*
J. MICHAEL STEELE, *Probability Theory and Combinatorial Optimization*
WERNER C. RHEINBOLDT, *Methods for Solving Systems of Nonlinear Equations, Second Edition*
J. M. CUSHING, *An Introduction to Structured Population Dynamics*
TAI-PING LIU, *Hyperbolic and Viscous Conservation Laws*
MICHAEL RENARDY, *Mathematical Analysis of Viscoelastic Flows*
GÉRARD CORNUÉJOLS, *Combinatorial Optimization: Packing and Covering*

This page intentionally left blank

R. Tyrrell Rockafellar
University of Washington, Seattle

Conjugate Duality and Optimization

siam.
SOCIETY FOR INDUSTRIAL AND APPLIED MATHEMATICS
PHILADELPHIA

10 9 8 7 6 5 4

ISBN 0-89871-013-8

Contents

Preface ix
1. The role of convexity and duality 1
2. Examples of convex optimization problems 6
3. Conjugate convex functions in paired spaces 13
4. Dual problems and Lagrangians 18
5. Examples of duality schemes 23
6. Continuity and derivatives of convex functions 30
7. Solutions to optimization problems 38
8. Some applications 45
9. Calculating conjugates and subgradients; integral functionals 52
10. More applications 64
References 73

This page intentionally left blank

Preface

Several books have recently been published describing applications of the theory of conjugate convex functions to duality in problems of optimization. The finite-dimensional case has been treated by Stoer and Witzgall [25] and Rockafellar [13] and the infinite-dimensional case by Ekeland and Temam [3] and Laurent [9]. However, these are long works concerned also with many other issues.

The purpose of these lecture notes is to provide a relatively brief introduction to conjugate duality in both finite- and infinite-dimensional problems. The material is essentially to be regarded as a supplement to the book *Convex Analysis* [13]; the same approach and results are presented, but in a broader formulation and more selectively. However, the order of presentation differs from [13]. I have emphasized more from the very beginning the fundamental importance of the concepts of Lagrangian function, saddle-point and saddle-value. Instead of first outlining everything relevant about conjugate convex functions and then deriving its consequences for optimization, I have tried to introduce areas of basic theory only as they became needed and their significance for the study of dual problems more apparent. In particular, general results on the calculation of conjugate functions have been postponed nearly to the end, making it possible to deduce more complete versions of the formulas by means of the duality theory itself. I have also attempted to show just where it is that convexity is needed, and what remains true if certain convexity or lower-semicontinuity assumptions are dropped.

The notation and terminology of [13] have been changed somewhat to make an easier introduction to the subject. Thus the concepts of "bifunction" and "convex process" have been omitted, even though they are needed in the larger picture to see how the results on optimization problems fit in with other aspects of duality that have long been a mainstay in mathematical analysis. The duality theorem for linear programming problems, for instance, turns out to be an analogue of an algebraic identity relating a linear transformation and its adjoint. For more on this point of view and its possible fertility for applications such as to mathematical economics, see [24].

A number of general examples drawn from nonlinear (including nonconvex) programming, approximation, stochastic programming, the calculus of variations and optimal control, are discussed in tandem with the theory. These examples are obviously not meant to cover in a representative manner all the possible areas of application of conjugate duality, but rather to illustrate and motivate the main lines of thought. They underline especially the fact that, as soon as one admits stochastic or dynamic elements into a model, one is likely to arrive at a problem in infinite dimensions. Moreover, such problems typically involve so-called integral functionals (even in "discrete" models, since an infinite series can be viewed as an integral over a discrete space of indices). For this reason, I have devoted some of the theoretical discussion specifically to convex integral functionals and their special properties.

Thanks are due to the Johns Hopkins University, the National Science Foundation, and the Conference Board of Mathematical Sciences for sponsoring the Regional Conference in June, 1973, which precipitated these notes. The participants in the conference deserve no small credit for the final form of the material, as it reflects many of their questions and observations. I am especially grateful to Professor J. Elzinga of Johns Hopkins who, in addition to all his many duties as organizer of the conference, managed in equally fine fashion to take on the job of supervising the typing and personally proofreading and filling in the symbols in the original version of these notes. I also want to thank the Air Force Office of Scientific Research very much for its support of research which went into these notes under Grant AFOSR-72-2269 and earlier.

Seattle

August 1973 R. TYRRELL ROCKAFELLAR

Conjugate Duality and Optimization

R. Tyrrell Rockafellar

1. The role of convexity and duality. In most situations involving optimization there is a great deal of mathematical structure to work with. However, in order to get to the fundamentals, it is convenient for us to begin by considering only the following *abstract optimization problem*: minimize $f(x)$ for $x \in C$, where C is a subset of a real linear space X, and $f: C \to [-\infty, +\infty]$.

If we define (or redefine) f so that $f(x) = +\infty$ for $x \notin C$, then minimizing f over C is equivalent to minimizing the new f over all of X. Thus no generality is lost in our abstract model if we restrict attention to the case where $C = X$. This conceptual and notational simplification is one of the main reasons for admitting extended-real-valued functions to the theory of optimization. However, once the step is taken, many other technical advantages and insights are also found to accrue. Functions having the value $-\infty$ as well as $+\infty$ arise in natural ways too, as will be seen later.

The function $f: X \to [-\infty, +\infty]$ is *convex* if its *epigraph*,

$$\operatorname{epi} f = \{(x, \alpha) \mid x \in X, \alpha \in R, \alpha \geqq f(x)\},$$

is convex as a subset of the linear space $X \times R$. Then the *effective domain* of f,

$$\operatorname{dom} f = \{x \in X \mid f(x) < +\infty\},$$

is also a convex set, and minimizing f over X is equivalent to minimizing f over $\operatorname{dom} f$. In particular, if f is a real-valued function on a set $C \subset X$ and we define $f(x) = +\infty$ for $x \notin C$, then f is convex as an extended-real-valued function on X if and only if (i) C is convex, and (ii) f is convex relative to C in the classical sense that the inequality

$$f((1-\lambda)x_1 + \lambda x_2) \leqq (1-\lambda)f(x_1) + \lambda f(x_2) \quad \text{if} \quad 0 < \lambda < 1 \tag{1.1}$$

always holds. A convex function is *proper* if it is the extension in this way of a real-valued function on a nonempty set C, i.e., if $\operatorname{dom} f \neq \emptyset$ and $f(x) > -\infty$ for all $x \in X$.

Convex optimization problems, that is, problems which can be expressed as above with f convex, have many pleasant properties:

(a) Local versus global analysis. One does not have to contend computationally or theoretically with isolated local minima or "stationary points" which do not solve the overall problem.

(b) Optimality conditions. It is relatively easy to determine criteria for an optimal solution which are not only necessary but sufficient. Even in the absence of differentiability in the ordinary sense, convexity makes possible a thorough treatment of such conditions in terms of one-sided directional derivatives and "subgradients."

(c) Existence and uniqueness of solutions. The compactness of various sets involved in existence theorems can often be assured more simply using the special continuity and growth properties of convex functions. In some cases involving integrals, convexity is even necessary for compactness. "Strict" convexity is often an easily verified criterion for uniqueness.

(d) Convergence of algorithms. Many special facts are available in this context, and convergence from an arbitrary starting point can often be guaranteed. Algorithms can also be based on features, like duality, which are hardly present in the nonconvex case.

(e) Duality. This subject at its fullest, depending in an essential way on convexity, leads to an enrichment of all aspects of the analysis of optimization problems.

A concept which is extremely fruitful is that of an *abstract optimization problem depending on parameters*: minimize $F(x, u)$ over all $x \in X$. The parameter vector u ranges over a space U. For example, u might be time, or it might be a random vector expressing uncertainty in the data defining the problem. Or u might simply represent certain variables whose effect on the problem is of interest. Note that the "essential constraints" in the abstract problem also depend in general on u, in the sense that the effective domain of the function $x \to F(x, u)$ depends on u.

Of particular importance in this situation is the study of the corresponding *optimal value function*

$$\varphi(u) = \inf_{x \in X} F(x, u), \qquad u \in U. \tag{1.2}$$

Typically we are interested in the continuity or directional derivative properties of φ at some point $\bar{u}$. (In most cases we can take U to be another real linear space and $\bar{u} = 0$.) The continuity may be crucial to the stability of some algorithm. The directional derivatives may be needed in a "sensitivity analysis" of marginal values from an economic viewpoint. Here also convexity enters in strongly, due to the following fact.

THEOREM 1. *Let X and U be real linear spaces, and let $F : X \times U \to [-\infty, +\infty]$ be convex. Then the function φ in* (1.2) *is also convex* (*but not necessarily proper, even if F is proper*).

Proof. Let E be the image of epi F under the projection $(x, u, \alpha) \to (u, \alpha)$. Then

$$\operatorname{epi} \varphi = \{(u, \alpha) \in U \times R \mid (u, \beta) \in E \text{ for all } \beta > \alpha\}.$$

Since epi F is a convex set by definition, and convexity is preserved under projections, we know E is convex. The formula for epi φ then yields the fact that epi φ is a convex set. Therefore, φ is a convex function.

Thus, in the case of an abstract *convex* optimization problem depending *convexly* on parameters (i.e., with $F(x, u)$ jointly convex in x and u), all the special results on the continuity and differentiability of convex functions can be used in the study of φ.

The theory of dual optimization problems, the main object of our attention below, also involves the notion of a minimization problem depending on parameters, but in a different fashion. Basically, this theory is concerned with representing a given minimization problem as "half" of a minimax problem whose saddle-value exists. We proceed to describe this fundamental idea in some detail.

Let us start with a function of the form

$$K: X \times Y \to [-\infty, +\infty],$$

where X and Y are arbitrary sets, and define

$$f(x) = \sup_{y \in Y} K(x, y), \tag{1.3}$$

$$g(y) = \inf_{x \in X} K(x, y). \tag{1.4}$$

Consider the two optimization problems:

$$\text{minimize} \quad f(x) \quad \text{over all} \quad x \in X, \tag{1.5}$$

$$\text{maximize} \quad g(y) \quad \text{over all} \quad y \in Y. \tag{1.6}$$

It is elementary that

$$f(x) \geqq K(x, y) \geqq g(y) \quad \text{for all} \quad x \in X, y \in Y, \tag{1.7}$$

and consequently,

$$\inf_{x \in X} \sup_{y \in Y} K(x, y) = \inf_{x \in X} f(x) \geqq \sup_{y \in Y} g(y) = \sup_{y \in Y} \inf_{x \in X} K(x, y). \tag{1.8}$$

If equality holds in (1.8), the common value is called the *saddle-value* of K. The saddle-value exists in particular if there is a *saddle-point* of K, i.e., a pair $(\bar{x}, \bar{y})$ such that

$$K(x, \bar{y}) \geqq K(\bar{x}, \bar{y}) \geqq K(\bar{x}, y) \quad \text{for all} \quad x \in X, y \in Y, \tag{1.9}$$

the saddle-value then being $K(\bar{x}, \bar{y})$. This is part of the following result.

THEOREM 2. *A pair* $(\bar{x}, \bar{y})$ *satisfies the saddle-point condition* (1.9) *if and only if* $\bar{x}$ *solves problem* (1.5), $\bar{y}$ *solves* (1.6), *and the saddle-value of* K *exists, i.e., one has*

$$\inf_{x \in X} f(x) = \sup_{y \in Y} g(y). \tag{1.10}$$

Proof. We can rewrite (1.9) as the condition

$$f(\bar{x}) = K(\bar{x}, \bar{y}) = g(\bar{y}).$$

The equivalence is then immediate from (1.8).

Because of Theorem 2, problems (1.5) and (1.6) are said to be *dual* to each other. Each represents "half" of the problem of finding a saddle-point for K. The latter problem corresponds, of course, to a certain *game*: player I chooses an element of X, player II chooses an element of Y, the choices are revealed simultaneously, and then there is a payoff of $K(x, y)$ from I to II. Problems (1.5) and (1.6) are the optimal strategy problems for players I and II, respectively.

Although classes of optimization problems dual to each other can be generated in the above manner by considering various classes of functions K, and some important examples of duality can be obtained this way, it is really the opposite pattern of construction that is of greatest interest and potentiality in optimization. Thus the fundamental question is this, starting from a problem of the form (1.5) where f is some extended-real-valued function on a space X, how can we introduce a space Y and a "meaningful" function K on $X \times Y$ so that (1.3) holds? Having determined such a K, we have at once a dual problem given by (1.6) and (1.4), along with a game-theoretic interpretation of the duality and a saddle-point criterion for optimality (Theorem 2). For K to be "meaningful," there should be some natural interpretation of $K(x, y)$ in terms of the initial problem (1.5), and K should belong to some class of functions for which the existence of saddle-values is at least not beyond hope. For most purposes, one would also want K to be fairly "concrete" and open to direct manipulation.

Further motivation for minimax representations of minimization problems is found in computation. Suppose, starting with (1.5), we have constructed a function K satisfying (1.3); then we have an expression of f as the "envelope" of a collection of functions $x \to K(x, y)$, $y \in Y$. Let us consider the problem

$$\text{(1.11)} \qquad \text{minimize} \quad K(x, y) \quad \text{over all} \quad x \in X$$

as an optimization problem parametrized by $y \in Y$. The corresponding optimal value function, expressing the infimum as dependent on y, is the function g in (1.4). Since (1.7) holds, we can think of (1.11) as a sort of "lower representative" of problem (1.5), with

$$\text{(1.12)} \qquad \inf_{x \in X} f(x) \geqq \inf_{x \in X} K(x, y).$$

Consider now the case where, for some $\bar{y} \in Y$, one actually has

$$\text{(1.13)} \qquad \inf_{x \in X} f(x) = \inf_{x \in X} K(x, \bar{y}).$$

Then, in view of the inequality $f(x) \geqq K(x, \bar{y})$, every $\bar{x}$ minimizing f also minimizes $K(\cdot, \bar{y})$. In other words, the solutions to (1.5) can be found among the solutions to (1.11) for $y = \bar{y}$. This fact could be significant for computation, for example, if $K(x, \bar{y})$ is easier to minimize than $f(x)$ and attains its minimum at a unique point. But how can a $\bar{y}$ satisfying (1.13), if any, be determined? This is easy to answer from (1.8), observing that the right side of (1.13) is just $g(\bar{y})$: one has (1.13) if and only if $\bar{y}$ solves the dual problem (1.6) and the saddle-value of K exists.

Thus if f and K are related by (1.3) and the saddle-value of K exists, we have a *dual approach* to minimizing f: first we maximize $g(y)$ to obtain y, and then we minimize $K(x, \bar{y})$ to obtain $\bar{x}$. Of course, in practice this dual approach would more likely be implemented in the following manner: given $y^k \in Y$, determine x^k by minimizing $K(x, y^k)$ to within some tolerance, and then construct y^{k+1}

(using local information about K at (x^k, y^k), say, and perhaps data from previous iterations). The aim would be to execute this so that the sequence y^k tends to a point $\bar{y}$ maximizing g, while x^k tends to a point $\bar{x}$ minimizing $K(x, \bar{y})$. Note that for this purpose an expression for g more "concrete" than (1.4) might not be needed.

Naturally, many details must be investigated in an individual case to see whether an effective algorithm can be built on the dual approach. However, the possible rewards could be great, since it might be much easier to solve the problems (1.11) than the original problem (1.6), due to some "decomposition" property of K or the fact that various constraints implicit in (1.6) (through allowing $+\infty$ as a value of f) might not be present in (1.11).

At all events, it is clear that, for both theoretical and computational purposes, we may want to explore for a given f many different "meaningful" functions K such that (1.3) holds. Thus we do not want to think of a particular problem (1.5) as having a fixed dual problem associated with it, however interesting or traditional the association may be (e.g., linear programming). In fact, there is potentially a great multiplicity of duals. The question is how to construct them so as to obtain desired properties.

The theory of conjugate convex functions furnishes a fairly complete answer to this question, at least in general terms. It establishes an essentially one-to-one correspondence between representations (1.3), where $K(x, y)$ is convex in x and concave in y, and representations

$$f(x) = F(x, 0), \tag{1.14}$$

where $F: X \times U \to [-\infty, +\infty]$ is convex, that is, embeddings of the given problem (1.5) in a convexly parametrized family of convex problems:

$$\text{minimize} \quad F(x, u) \quad \text{over all} \quad x \in X. \tag{1.15}$$

Furthermore, the theory brings to light a remarkable new aspect of duality by demonstrating a close relationship (in fact a "dualism") between the properties of the function g in the dual problem (1.6) which are important to maximization and the properties of the optimal value function

$$\varphi(u) = \inf_{x \in X} F(x, u)$$

which are involved in "sensitivity" analysis at $u = 0$. Through the study of this relationship, one finds that the saddle-value of the function K in question "usually" exists.

Since the principal theorems known about the existence of the saddle-value of a function K require that $K(x, y)$ be convex in x and concave in y, or almost so, it is evident that this scheme is capable of generating most of the functions K that are "meaningful" for the basic theory of dual problems. Representations (1.14) with F convex are quite easy to obtain, provided of course that f itself is convex. Diverse examples are given in the next section.

Although the main results of conjugate function theory apply only to optimization problems of convex type, there are many applications also to nonconvex problems. Some of these concern the derivation of necessary conditions for optimality. Others arise because, in the course of some algorithm, a nonconvex problem is at each step approximated locally by a convex problem, perhaps as a direction-finding mechanism for a method of descent. One nonconvex application will be discussed in some detail below (Examples 3, 3′, 3″).

2. Examples of convex optimization problems. We begin by reviewing some facts and notations convenient in formulating problems and verifying convexity.

Let X denote a real linear space. The *indicator* ψ_C of a set $C \subset X$ is defined by

$$\psi_C(x) = \begin{cases} 0 & \text{if } x \in C, \\ +\infty & \text{if } x \notin C. \end{cases} \tag{2.1}$$

This is a convex function if and only if C is a convex set.

If $f: X \to [-\infty, +\infty]$ is convex, then for each $\alpha \in [-\infty, +\infty]$ the level sets $\{x | f(x) \leqq \alpha\}$ and $\{x | f(x) < \alpha\}$ are convex.

If f is convex, then so are the functions $g(x) = f(x + a)$ where $a \in X$, $h(x) = f(x) + \alpha$, where $\alpha \in R$, and $k(x) = \lambda f(x)$, where $\lambda > 0$. Also, the function $r(z) = f(Az)$ is convex on a linear space Z if $A: Z \to X$ is a linear transformation.

If f_1 and f_2 are convex on X, then $f = f_1 + f_2$ is convex, provided the convention $(+\infty) + (-\infty) = +\infty$ is used in the addition. (The other conventions for handling $+\infty$ and $-\infty$ are all obvious.) However, this special convention is risky in some contexts, particularly where one might get it mixed up with the opposite convention $(+\infty) + (-\infty) = (-\infty)$ needed for concave functions. (A function g is *concave* if $f = -g$ is convex.) Therefore it is often wise to steer away from situations where $+\infty$ and $-\infty$ would have to be added. This can be done for example by arranging that only *proper* convex functions be added together, as suffices for most purposes anyway.

Some optimization problems require the consideration of infinite or "continuous" sums of convex functions. These can be represented by an integral over a measure space (S, Σ, σ) with σ nonnegative:

$$I(x) = \int_S h(x, s)\sigma(ds). \tag{2.2}$$

More generally, one is led to the study of *integral functionals*

$$I: \mathscr{X} \to [-\infty, +\infty] \tag{2.3}$$

of the form

$$I(x) = \int_S h(x(s), s)\sigma(ds), \tag{2.4}$$

where $\mathscr{X}$ is some linear space of functions $x: S \to X$. Note that the functional (2.2) can be regarded as the special case of (2.4) where $\mathscr{X}$ is the space of *constant* functions.

There are a number of technical questions associated with integral functionals, not the least of which is the question of when such a functional is well-defined. We shall adopt the point of view that I in (2.4) is well-defined on $\mathscr{X}$ if for each function $x \in \mathscr{X}$ the function $s \to h(x(s), s)$ is measurable. Then if there exists a real-valued function $\alpha : S \to R$ which is summable with respect to σ (in the usual sense) and satisfies $\alpha(s) \geqq h(x(s), s)$ almost everywhere, the integral (2.4) has an unambiguous, classical value (finite or $-\infty$), and this is what we assign to $I(x)$; if such a function α does not exist, we set $I(x) = +\infty$.

To get a better grip on this concept, we can make use of the fact that in applications X typically has a topological structure and consequently a measurability structure: the Borel sets. The integrand $h : X \times S \to [-\infty, +\infty]$ is said to be *measurable on $X \times S$ (relative to the Borel structure on X)* if h is measurable with respect to the σ-algebra on $X \times S$ generated by the sets $B \times T$, where B is a Borel set in X, and $T \in \Sigma$. Then $h(x(s), s)$ is indeed measurable in s whenever $x(s)$ is measurable in s, since the latter measurability implies that of the transformation $s \to (x(s), s)$.

Integral functionals will be discussed further in § 9. For present purposes, we merely state an immediate consequence of these definitions regarding convexity.

THEOREM 3. *The integral functional I given by* (2.4) *is well-defined on $\mathscr{X}$ in the above sense if, relative to the Borel structure on X generated by some topology, the integrand $h : X \times S \to [-\infty, +\infty]$ is measurable and the functions $x : S \to X$ are all measurable. If in addition h is convex in the X argument, then I is a convex function on $\mathscr{X}$.*

Another useful construction: if $f_1, \cdots, f_m$ are convex functions on X, then so is

$$g(x) = \max \{f_1(x), \cdots, f_m(x)\}. \tag{2.5}$$

Similarly, if $f : X \times W \to [-\infty, +\infty]$ is convex in the X argument, then the function

$$g(x) = \sup_{s \in S} h(x, s) \tag{2.6}$$

is convex.

Example 1. (Nonlinear programming.) Let $f_0, f_1, \cdots, f_m$ be real-valued convex functions on a nonempty convex set C in the linear space X. The problem is to minimize $f_0(x)$ over all $x \in C$ satisfying $f_i(x) \leqq 0$, $i = 1, \cdots, m$. Abstract representation: minimize f over X, where

$$f(x) = \begin{cases} f_0(x) & \text{if } x \in C \text{ and } f_i(x) \leqq 0, \quad i = 1, \cdots, m, \\ +\infty & \text{for all other } x. \end{cases} \tag{2.7}$$

There are many ways of introducing parameters "convexly." The "ordinary" way is to define $F(x, u)$ for $u = (u_1, \cdots, u_m) \in R^m$ by

$$F(x, u) = \begin{cases} f_0(x) & \text{if } x \in C \text{ and } f_i(x) \leqq u_i, \quad i = 1, \cdots, m, \\ +\infty & \text{for all other } x. \end{cases} \tag{2.8}$$

Then $F : X \times R^m \to (-\infty, +\infty]$ is convex and $F(x, 0) = f(x)$.

Example 2. (Nonlinear programming.) The problem is the same as in Example 1 (same f too), but a more complete system of parameters is introduced. Let each f_i be expressed in the form

$$f_i(x) = h_i(A_i x + a_i) + l_i(x), \tag{2.9}$$

where l_i is an *affine* (linear-plus-constant) function on X, A_i is a linear transformation from X to another linear space Z_i, h_i is a convex function on Z_i, and $a_i \in Z_i$. For

$$u = (u_1, \cdots, u_m, z_0, \cdots, z_m) \in R \times \cdots \times R \times Z_0 \times \cdots \times Z_m \triangleq U, \tag{2.10}$$

define

$$F(x,u) = \begin{cases} h_0(A_0 x + a_0 - z_0) + l_0(x) & \text{if} \quad x \in C \quad \text{and} \\ \qquad h_i(A_i x + a_i - z_i) + l_i(x) \leqq 0, & i = 1, \cdots, m, \\ +\infty \quad \text{for all other } x. \end{cases} \tag{2.11}$$

Again F is convex with $F(x,0) = f(x)$. This scheme turns out to be particularly useful in special cases like exponential programming (also called geometric programming, although the latter term is now often used in a much wider sense), or quadratically-constrained quadratic programming, where a representation (2.9) is part of the basic structure of the problem. In exponential programming one has

$$h_i(z_i) = \log\left(\sum_{k=1}^{n_i} \exp z_{ik}\right) \quad \text{for} \quad z_i \in R^{n_i} \triangleq Z_i \tag{2.12}$$

(the logarithm might be omitted), while in quadratic programming

$$h_i(z_i) = \frac{1}{2}\sum_{k=1}^{n_i} z_{ik}^2 \quad \text{for} \quad z_i \in R^{n_i} \triangleq Z_i. \tag{2.13}$$

(Analogues involving infinite-dimensional spaces Z_i can also be formulated easily.)

Example 3. (Nonlinear programming.) Everything is the same as in Example 1, but for a fixed $r > 0$ we define

$$F(x,u) = \begin{cases} f_0(x) + r|u|^2 & \text{if} \quad x \in C \quad \text{and} \quad f_i(x) \leqq u_i, \quad i = 1, \cdots, m, \\ +\infty \quad \text{for all other } x, \end{cases} \tag{2.14}$$

where $|u|$ denotes the Euclidean norm. This alteration may seem pointless, but we shall show that it leads to a dual problem with strikingly different properties. Moreover, it opens up computational applications of duality in *nonconvex* programming.

Example 4. (Nonlinear programming with infinitely many constraints.) The problem is to minimize $f_0(x)$ over the set

$$\{x \in C | h(x,s) \leqq 0 \text{ for all } s \in S\}, \tag{2.15}$$

where f_0 is a real-valued convex function on the nonempty convex set $C \subset X$, S is an arbitrary set, and $h: X \times S \to [-\infty, +\infty]$ is convex in the X argument. The corresponding abstract convex optimization problem has

$$(2.16) \qquad f(x) = \begin{cases} f_0(x) & \text{if } x \text{ belongs to (2.15)}, \\ +\infty & \text{for all other } x. \end{cases}$$

The parametrization analogous to Example 1 is to let U be a linear space of functions $u: S \to R$ and define

$$(2.17) \qquad F(x, u) = \begin{cases} f_0(x) & \text{if } \quad x \in C \quad \text{and} \quad h(x, s) \leqq u(s) \qquad \text{for all } s \in S, \\ +\infty & \text{for all other } x. \end{cases}$$

Parameters could also be introduced more complicatedly, as in Example 2, say.

Example 5. (Chebyshev approximation.) Let $h_i: [0, 1] \to R$ be continuous for $i = 0, 1, \cdots, m$. The problem is to minimize

$$f(x) = \left\| h_0 - \sum_{i=1}^{m} x_i h_i \right\| = \max_{0 \leqq t \leqq 1} \left| h_0(t) - \sum_{i=1}^{m} x_i h_i(t) \right|$$

over all possible coefficient vectors $x = (x_1, \cdots, x_m) \in R^m \triangleq X$. Note that f is a finite convex function which is *not differentiable*: this serves to emphasize the fact that, even for the purpose of treating classical optimization problems, a theory able to handle nondifferentiable functions is desirable. A "convex" parameterization may be introduced here by defining

$$(2.18) \qquad F(x, u) = \left\| h_0 - \sum_{i=1}^{m} x_i h_i + u \right\|$$

for $u \in \mathscr{C}[0, 1] \triangleq U$.

Example 6. (Stochastic programming.) Let (S, Σ, σ) be a probability space, let $h: X \times S \to [-\infty, +\infty]$ be convex in the X argument, where X is some linear topological space, and let C be a closed convex subset of X. We are interested in the abstract optimization problem:

$$\text{minimize} \quad h(x, s) \quad \text{over all} \quad x \in C,$$

which we imagine as depending on s as a random element whose statistical distribution is known. The difficulty is that x must be chosen before s has been observed. The problem really facing us is therefore that of minimizing the expectation

$$(2.19) \qquad f(x) = \int_S h(x, s)\sigma(ds)$$

over all $x \in X$. We assume h is measurable on $X \times S$ (relative to the Borel structure on X), so that f is well-defined and convex by Theorem 3. If we set

$$(2.20) \qquad F(x, u) = \int_S h(x - u(s), s)\sigma(ds) + \psi_C(x) \quad \text{for} \quad u \in U,$$

where U is some linear space of measurable function $u: S \to X$, then F is likewise well-defined and convex, and $F(x, 0) = f(x)$.

Example 7. (Stochastic programming.) The problem is the same as in Example 6, but we introduce the structure

$$h(x, s) = \inf_{w \in W} h_0(x, w, s), \tag{2.21}$$

where W is another linear topological space, and h_0 is of the form

$$h_0(x, w, s) = \begin{cases} f_0(x, w, s) & \text{if } \ w \in D \ \text{ and } \ f_i(x, w, s) \leqq 0, \ i = 1, \cdots, m, \\ +\infty & \text{otherwise.} \end{cases} \tag{2.22}$$

Here D is a closed convex subset of W, and f_i is a function defined on $X \times W \times S$ for $i = 0, 1, \cdots, m$. The interpretation is that w is a *recourse* that can be selected after the random element s has been observed.

If the functions f_i are convex in (x, w), then the same is true of h_0, and therefore $h(x, s)$ is convex in x by Theorem 1. It is not obvious, however, what assumptions on the functions f_i imply that h is measurable on $X \times S$ as needed for applying Theorem 3. Certainly if every f_i is measurable on $X \times W \times S$ (relative to the Borel structure on $X \times W$), then the same is true of h_0. But operation (2.21) is questionable, as far as preserving measurability is concerned.

This is typical of the difficulties encountered in applications involving integral functionals. A lot of work frequently goes into establishing the measurability of integrands defined in complicated ways. Fortunately there is extensive machinery available for this purpose, much of it taking special advantage of convexity. But trying to describe the details here would carry us too far from our main topic: see [19] and the references given there.

For the case at hand, let it suffice to say that h will be measurable on $X \times S$ if $f_i(x, w, s)$ is a finite convex function of (x, w) for each $s \in S$ and a measurable function of s for each $(x, w) \in X \times W$, and if the spaces X and W are finite-dimensional. (This can be demonstrated from results in [17].)

Now let U be a linear space of measurable functions

$$u: s \to (u_1(s), \cdots, u_m(s)) \in R^m$$

and set

$$F(x, u) = \int_S H(x, u(s), s)\sigma(ds) + \psi_C(x), \tag{2.23}$$

where $H(x, u(s), s)$ is the infimum of $f_0(x, w, s)$ over all $w \in D$ satisfying

$$f_i(x, w, s) \leqq u_i(s) \quad \text{for} \quad i = 1, \cdots, m. \tag{2.24}$$

Our assumptions on the functions f_i imply actually that $H(x, u, s)$ is not only convex as a function of $(x, u) \in X \times R^m$, but also measurable in (x, u, s) relative to the Borel structure on $X \times R^m$. (Namely, let $\bar{f}_i(x, u, w, s) = f_i(x, w, s) - u_i$ for $i = 1, \cdots, m$, and let $\bar{f}_0(x, u, w, s) = f_0(x, w, s)$. These functions are finite, convex in (x, u, w) and measurable in s, and when the construction (2.22), (2.21) is applied to them one gets H in place of h. Therefore H is measurable for the same reason that h is under these assumptions.) It follows from Theorem 3 that the functional

$$F : X \times U \to [-\infty, +\infty]$$

is well-defined and convex.

We thus have a different "convex parameterization" of the problem in Example 6. Of course, both types of parameterization could be introduced simultaneously.

Example 8. (Calculus of variations.) The problem is to minimize

$$f(x) = \int_0^1 L(x(t), \dot{x}(t), t)\, dt + l(x(0), x(1)) \tag{2.25}$$

over the linear space $\mathscr{A}_n[0, 1]$ consisting of the absolutely continuous functions $x : [0, 1] \to R^n$. (Here $\dot{x}$ denotes the derivative of x, which can be represented by a measurable function whose values are uniquely determined except on a set of measure zero.) The extended-real-valued functions $L(\cdot, \cdot, t)$ and l on $R^n \times R^n$ are assumed convex, and L is assumed measurable on $R^n \times R^n \times [0, 1]$ relative to the Borel structure on $R^n \times R^n$ and the Lebesgue structure on $[0, 1]$. Then f is well-defined and convex (Theorem 3). It deserves emphasis that this problem may involve constraints, represented namely by allowing L and l to have the value $+\infty$. Thus, for instance, if l is the indicator of a point pair (c_0, c_1), that is,

$$l(x(0), x(1)) = \begin{cases} 0 & \text{if } x(0) = c_0 \text{ and } x(1) = c_1, \\ +\infty & \text{if } x(0) \neq c_0 \text{ or } x(1) \neq c_1, \end{cases} \tag{2.26}$$

the problem is equivalent to minimizing

$$\int_0^1 L(x(t), \dot{x}(t), t)\, dt$$

over the class of all arcs $x \in \mathscr{A}_n[0, 1]$ satisfying the endpoint conditions $x(0) = c_0$ and $x(1) = c_1$.

A significant way to parameterize the problem "convexly," as it turns out, is to set

$$F(x : z, a) = \int_0^1 L(x(t) - z(t), \dot{x}(t), t)\, dt + l(x(0) - a, x(1)), \tag{2.27}$$

where a ranges over R^n and z ranges over the space $\mathscr{L}_n^\infty[0, 1]$ consisting of all the essentially bounded, measurable functions from $[0, 1]$ to R^n.

This example could be generalized by replacing R^n by some infinite-dimensional space.

Example 9. (Optimal control.) The problem we consider is to minimize

$$\int_0^1 f_0(x(t), w(t), t)\, dt + l_0(x(0), x(1)) \tag{2.28}$$

over all the functions $x \in \mathscr{A}_n[0, 1]$ and $w \in \mathscr{L}_n^1[0, 1]$ satisfying

$$\dot{x}(t) = A(t)x(t) + w(t) \quad \text{a.e.,} \tag{2.29}$$

$$f_i(x(t), w(t), t) \leqq 0 \quad \text{a.e.,} \qquad i = 1, \cdots, m, \tag{2.30}$$

$$l_j(x(0), x(1)) \leqq 0, \qquad j = 1, \cdots, r. \tag{2.31}$$

Here $A(t)$ is a matrix, and the functions $f_i(\cdot, \cdot, t)$ and l_j are finite and convex on $R^n \times R^n$. We assume Lebesgue measurability with respect to t; it can be shown that this yields all the measurability properties needed [21]. (No extra generality would be gained in replacing the term $w(t)$ in (2.29) by $B(t)w(t)$.) Conditions (2.30) can in part represent constraints purely on the states $x(t)$ or purely on the controls $w(t)$.

The problem can be expressed in the form of Example 8 simply by setting

$$L(x, \dot{x}, t) = \begin{cases} f_0(x, \dot{x} - Ax, t) & \text{if } f_i(x, \dot{x} - Ax, t) \leqq 0 \\ & \qquad \text{for } i = 1, \cdots, m, \\ +\infty & \text{otherwise,} \end{cases} \tag{2.32}$$

$$l(x(0), x(1)) = \begin{cases} l_0(x(0), x(1)) & \text{if } l_j(x(0), x(1)) \leqq 0 \text{ for } j = 1, \cdots, r, \\ +\infty & \text{otherwise.} \end{cases} \tag{2.33}$$

The same parameters as in Example 8 can then be introduced. However, for purposes of illustration we indicate instead a different "convex" parameterization:

$$F(x, u) = \int_0^1 f_0(x(t), \dot{x}(t) - A(t)x(t), t)\, dt + l_0(x(0), x(1)) \tag{2.34}$$

if the conditions

$$\begin{aligned} f_i(x(t), \dot{x}(t) - A(t)x(t), t) &\leqq u_i(t) \quad \text{a.e.,} \qquad i = 1, \cdots, m, \\ l_j(x(0), x(1)) &\leqq u_j', \qquad j = 1, \cdots, r, \end{aligned} \tag{2.35}$$

are satisfied, but $F(x, u) = +\infty$ otherwise. The components of u are thus real-valued functions u_i of certain types on $[0, 1]$ and real numbers u_j'.

Example 10. (Partial differential equations.) Let Ω be a region in R^m with boundary Γ. As a generalization of Example 8, we have the problem of minimizing

$$f(x) = \int_\Omega L(x(t), \nabla x(t), t)\, dt + \int_\Gamma l(x(s), s)\, ds \tag{2.36}$$

over (say) the space of all continuously differentiable functions $x : \Omega \to R$, where $L(\cdot, \cdot, t)$ is a convex function on $R \times R^n$ for each $t \in \Omega$, and $l(\cdot, s)$ is a convex function on R^n for each $s \in \Gamma$. For instance, if

$$L(x(t), \nabla x(t), t) = \tfrac{1}{2}|\nabla x(t)|^2, \tag{2.37}$$

$$l(x(s), s) = \begin{cases} 0 & \text{if } x(s) = \xi(s), \\ +\infty & \text{otherwise,} \end{cases} \tag{2.38}$$

we have the classical Dirichlet problem in which the integral $\frac{1}{2}\int_\Omega |\nabla x(t)|^2\, dt$ is minimized over the functions x which agree with the given function ξ on the boundary of Ω. The solution to this problem solves an important partial differential equation. Many other instances can also be given where minimizing (2.36) corresponds to solving a classical partial differential equation. This is due to the fact that such equations in physical situations are often derived from variational principles. It is no surprise, therefore, that the study of (2.36) in the general case turns out to correspond to the analysis of a broad class of generalized partial differential equations and the dual variational principles associated with them.

"Convex" parameterization can be effected in a manner parallel to Example 8:

$$F(x; z, a) = \int_\Omega L(x(t) - z(t), \nabla x(t), t)\, dt + \int_\Gamma l(x(s) - a(s), s)\, ds, \tag{2.39}$$

where $z:\Omega \to R^n$ and $a:\Gamma \to R^n$ are summable functions. A model resembling Example 9 can also be investigated. The functions x, instead of being real-valued, can be allowed to have values in a suitable linear space X.

3. Conjugate convex functions in paired spaces. A notion fundamental to the study of duality is that the linear functions on a given linear space can themselves be regarded as elements of a linear space. In the finite-dimensional case, the linear functions on R^n can be identified with elements of R^n in terms of the usual inner product: every linear function is of the form $x \to x \cdot v$ for some $v \in R^n$. The infinite-dimensional case is, of course, not so simple, because suitable "inner products" are not always present. Furthermore, the class of linear functions one is interested in may vary from problem to problem, depending on topological considerations.

A convenient way of handling the general situation is the idea of paired spaces. A pairing of two (real) linear spaces X and V is a (real-valued) bilinear form $\langle \cdot, \cdot \rangle$ on $X \times V$; the expression $\langle x, v \rangle$ thus behaves much like an inner product, except that the x argument is restricted to X and v to V. The pairing associates with each $v \in V$ a linear function

$$\langle \cdot, v \rangle : x \to \langle x, v \rangle \tag{3.1}$$

on X and with each $x \in X$ a linear function

$$\langle x, \cdot \rangle : v \to \langle x, v \rangle \tag{3.2}$$

on V. A topology on X is *compatible* with the pairing if it is a locally convex topology such that the linear functions (3.1) are all continuous and *every* continuous linear function on X can be represented in this form for some $v \in V$. Compatible topologies on V are likewise the locally convex topologies such that the continuous linear functions on V are the ones of form (3.2). When we say that X and V are *paired spaces*, we mean that a particular pairing has been singled out, *and X and V have been equipped with compatible topologies with respect to the pairing.*

For those concerned mainly with finite-dimensional aspects, the only example which needs to be kept in mind is

$$X = R^n = V, \quad \langle x, v \rangle = x \cdot v. \tag{3.3}$$

A common infinite-dimensional example occurs when X and V are linear spaces of R^n-valued functions on a measure space (S, Σ, σ) and

$$\langle x, v \rangle = \int_S x(s) \cdot v(s)\, \sigma(ds). \tag{3.4}$$

Of course, X and V must be such that the function $x \to x(s) \cdot v(s)$ is (measurable and) summable for every $x \in X$ and $v \in V$. Sometimes the topologies to choose are not altogether obvious. For instance, suppose that $X = \mathcal{L}^1(S, \Sigma, \sigma)$ and $V = \mathcal{L}^\infty(S, \Sigma, \sigma)$. The norm topology on X is compatible with the pairing (3.4), as is the weak topology X possesses as a Banach space. The norm topology on V is *not* compatible, since there generally exist continuous linear functions on $\mathcal{L}^\infty$ which are norm-continuous but not representable through (3.4) by elements of $\mathcal{L}^1$. However, the weak topology induced on $\mathcal{L}^\infty$ by $\mathcal{L}^1$ is a compatible topology on V.

Various topologies compatible with a given pairing always exist and can be generated systematically. For the details we must defer to texts on functional analysis [8].

Henceforth let X and V be paired spaces. The closed half-spaces in X are then the various sets of the form $\{x | \langle x, v \rangle \leqq \beta\}$, where $\beta \in R$, $v \in V$, $v \neq 0$. It is a basic theorem that a subset C of X is a closed convex set if and only if C can be expressed as the intersection of a collection of such closed half-spaces. More generally, if C is any subset of X, then the intersection of all the closed half-spaces containing C is cl co C, where co C is the convex hull of C and cl denotes topological closure. This fact, exploited in the context of epigraph sets, leads as we shall see to the idea of the function f^* *conjugate* to a given function f on X, which is essentially a description of the "nonvertical" closed half-spaces in $X \times R$ containing epi f.

It is noteworthy that, since the class of closed convex sets can be described directly in terms of the pairing, it depends only on the pairing and not on the particular compatible topology on X which has been selected.

Let $f: X \to [-\infty, +\infty]$. The *convex hull* co f is the greatest convex function $\leqq f$. Geometrically, the epigraph of co f is obtained from the convex hull of the epigraph of f:

$$\operatorname{epi} \operatorname{co} f = \{(x, \alpha) | (x, \beta) \in \operatorname{co} \operatorname{epi} f \text{ for all } \beta > \alpha\}. \tag{3.5}$$

The function f is *lower-semicontinuous* (l.s.c.) if the set $\{x | f(x) \leqq \alpha\}$ is closed for all $\alpha \in R$. This is equivalent to the condition that epi f be closed as a subset of $X \times R$. The *l.s.c. hull* lcs f is the greatest l.s.c. function $\leqq f$. Thus

$$\operatorname{epi} \operatorname{lsc} f = \operatorname{cl} \operatorname{epi} f, \tag{3.6}$$

or equivalently

$$\operatorname{lsc} f(x) = \liminf_{x' \to x} f(x'). \tag{3.7}$$

Note that f is l.s.c. if and only if $f = \operatorname{lsc} f$.

A slightly modified operation turns out to be even more useful in convex analysis: we define the *closure* $\operatorname{cl} f$ of f by

$$\operatorname{cl} f(x) = \begin{cases} \operatorname{lsc} f(x) & \text{for all } x \text{ if } \operatorname{lsc} f(x) > -\infty \text{ for all } x, \\ -\infty & \text{for all } x \text{ if } \operatorname{lsc} f(x) = -\infty \text{ for some } x. \end{cases} \tag{3.8}$$

We say f is *closed* if $f = \operatorname{cl} f$, i.e., if f is an l.s.c. function nowhere having the value $-\infty$, or if f is the constant function $-\infty$.

THEOREM 4. *Let $f: X \to [-\infty, +\infty]$ be convex. Then* $\operatorname{lsc} f$ *and* $\operatorname{cl} f$ *are convex. If* $\operatorname{lsc} f$ *has a finite value at some point, then* $\operatorname{lsc} f$ *and* f *are proper, and* $\operatorname{cl} f = \operatorname{lsc} f$. *Otherwise,* $\operatorname{lsc} f$ *is of the form*

$$\operatorname{lsc} f = \begin{cases} -\infty & \text{if } x \in \operatorname{cl} \operatorname{dom} f, \\ +\infty & \text{if } x \notin \operatorname{cl} \operatorname{dom} f. \end{cases} \tag{3.9}$$

In the latter event we still have $\operatorname{cl} f(x) = \operatorname{lsc} f(x)$, *unless* $x \notin \operatorname{cl} \operatorname{dom} f \neq \emptyset$ *(then* $\operatorname{cl} f(x) = -\infty$ *but* $\operatorname{lsc} f(x) = +\infty$).

Proof. The convexity is obvious from (3.6) and (3.8). To obtain everything else, we need only verify that if $\operatorname{lsc} f(\bar{x}) = -\infty$ and

$$x \in \operatorname{dom} \operatorname{lsc} f = \operatorname{cl} \operatorname{dom} f,$$

then $\operatorname{lsc} f(x) = -\infty$. Let α and β be arbitrary real numbers with $\alpha \geqq \operatorname{lsc} f(x)$. The pairs (x, α) and $(\bar{x}, \beta)$ belong to the epigraph of $\operatorname{lsc} f$, which is convex, and hence so does $(1 - \lambda)(x, \alpha) + \lambda(\bar{x}, \beta)$, $0 \leqq \lambda \leqq 1$. In other words,

$$\operatorname{lsc} f((1 - \lambda)x + \lambda\bar{x}) \leqq (1 - \lambda)\alpha + \lambda\beta \quad \text{if} \quad 0 < \lambda \leqq 1.$$

Since β is arbitrary, we must have

$$\operatorname{lsc} f((1 - \lambda)x + \lambda\bar{x}) = -\infty \quad \text{if} \quad 0 < \lambda \leqq 1.$$

Letting $\lambda \downarrow 0$, we see from the lower-semicontinuity of $\operatorname{lsc} f$ that $\operatorname{lsc} f(x) = -\infty$.

The *conjugate* of a function $f: X \to [-\infty, +\infty]$ is the function $f^*: V \to [-\infty, +\infty]$ defined by

$$f^*(v) = \sup\{\langle x, v\rangle - f(x) \mid x \in X\}. \tag{3.10}$$

In the same pattern, the conjugate of a function $h: V \to [-\infty, +\infty]$ is defined by

$$h^*(x) = \sup\{\langle x, v\rangle - h(v) \mid v \in V\}. \tag{3.11}$$

The *biconjugate* of $f: X \to [-\infty, +\infty]$ is the conjugate f^{**} of $f^*: V \to [-\infty, +\infty]$:

$$f^{**}(x) = \sup\{\langle x, v\rangle - f^*(v) \mid v \in V\}. \tag{3.12}$$

The operation $f \to f^*$ is called the *Fenchel transform*.

THEOREM 5. *Let $f: X \to [-\infty, +\infty]$ be arbitrary. Then the conjugate f^* is a closed convex function on V, and one has $f^{**} = \text{cl co} f$. Similarly starting with a function on V. In particular, the Fenchel transform induces a one-to-one correspondence $f \to h$ (where $h = f^*$ and $f = h^*$) between the closed convex functions on X and the closed convex functions on V.*

Proof. By definition f^* is the pointwise superemum of the collection of continuous affine functions $V \to \langle x, v\rangle - \alpha$, $(x, \alpha) \in \text{epi} f$. From this it follows that f^* is convex and l.s.c., actually closed. We have $(v, \beta) \in \text{epi} f^*$ if and only if the continuous affine function $x \to \langle x, v\rangle - \beta$ satisfies

$$f(x) \geqq \langle x, v\rangle - \beta \quad \text{for all} \quad x \in X,$$

i.e., the epigraph of this affine function (a certain closed half-space in $X \times R$ of "nonvertical" type) contains epi f. Thus epi f^{**} is the intersection of all the "nonvertical" closed half-spaces in $X \times R$ containing epi f. An elementary argument, based on the characterization of closed convex sets mentioned above, shows that this means $f^{**} = \text{cl co} f$.

Some basic examples:

(a) For the indicator ψ_C of a set $C \subset X$ as in (2.1), ψ_C^* is the *support function* of C:

$$\psi_C^*(v) = \sup\{\langle x, v\rangle \mid x \in C\}. \tag{3.13}$$

If C is a cone, then ψ_C^* is in fact the indicator ψ_{C^*} where C^* is the *polar* cone:

$$C^* = \{v \mid \text{for all } x \in X, \langle x, v\rangle \leqq 0\}. \tag{3.14}$$

Of course, if C is a subspace we have

$$C^* = C^\perp = \{v \mid \text{for all } x \in X, \langle x, v\rangle = 0\}. \tag{3.15}$$

In this sense the polarity correspondence for cones and the "orthogonality" correspondence for subspaces may be regarded as special cases of the conjugacy correspondence.

(b) In the finite-dimensional case of (3.3), let f be a finite, differentiable convex function such that for each $v \in R^n$ the equation $\nabla f(x) = v$ has a unique solution $x \in R^n$. Denoting the solution by $s(v)$, we have

$$f^*(v) = s(v) \cdot v - f(s(v)). \tag{3.16}$$

The latter formula is said to define the *Legendre transform*, and f^* is thus the *Legendre conjugate of* f. It can be shown that f^* likewise is a finite, differentiable convex function such that for each $x \in R^n$ the equation $\nabla f^*(v) = x$ has a unique solution $s^*(x)$. (In fact $\nabla f^*(v) = s(v)$, so that $s^*(x) = \nabla f(x)$.) Therefore, f is in turn the Legendre conjugate of f^*. The Legendre transform can, to some extent, be generalized to infinite-dimensional spaces X and to differentiable functions given *only on a subset* of X [1, § 5], [13, § 26].

To illustrate (a) in the finite-dimensional case (3.3), we determine the conjugate of $f(x) = |x|$ (Euclidean norm). Since

$$|x| = \sup\{\langle x, v\rangle \mid |v| \leqq 1\},$$

we have $f = \psi_C^*$, where C is the unit ball; therefore $f^* = \psi_C^{**} = \psi_C$. Another example: if f is the indicator of the nonnegative orthant R^n_+, then f^* is the indicator of the nonpositive orthant R^n_- (the polar of R^n_+). An illustration of (b) is the case where

(3.17) $$f(x) = \tfrac{1}{2}x \cdot Qx + a \cdot x + \alpha, \qquad x \in R^n,$$

with $a \in R^n$, $\alpha \in R$, and the matrix Q symmetric and positive definite. Then

(3.18) $$f^*(v) = \tfrac{1}{2}(v - a) \cdot Q^{-1}(v - a) - \alpha.$$

An example not quite subsumed by (b), but obtained by the same kind of direct calculation, is

(3.19) $$f(x) = \log(e^{x_1} + \cdots + e^{x_n}), \qquad x \in R^n.$$

One has (using the convention $0 \log 0 = 0$)

(3.20) $$f^*(v) = \begin{cases} v_1 \log v_1 + \cdots + v_n \log v_n & \text{if } v_i \geqq 0, \quad v_1 + \cdots + v_n = 1, \\ +\infty & \text{otherwise,} \end{cases}$$

and $f^{**} = f$. (This will have applications to Example 2; cf. (2.13).)

It is fair to say that most of the "primary" examples of conjugate convex functions follow from (a) or (b). But many other important examples can be generated from these by means of various operations such as at the beginning of § 2. The relationship between these operations and conjugacy will be discussed in § 9.

It is clear from the remarks about minimax problems and duality in § 1 that *concave* functions will sometimes need to be treated on an equal footing with convex functions. For concave $g: X \to [-\infty, +\infty]$, we define (not *too* ambiguously)

(3.21) $$\operatorname{dom} g = \{x \in X \mid g(x) > -\infty\},$$

(3.22) $$\operatorname{epi} g = \{(x, \alpha) \mid x \in X, \alpha \in R, \alpha \leqq g(x)\}.$$

The *upper-semicontinuous hull* usc g has

(3.23) $$\operatorname{usc} g(x) = \limsup_{x' \to x} g(x'),$$

while the *closure* cl g is given by

(3.24) $$\operatorname{cl} g(x) = \begin{cases} \operatorname{usc} g(x) & \text{for all } x \text{ if } \operatorname{usc} g(x) < +\infty \text{ for all } x, \\ +\infty & \text{for all } x \text{ if } \operatorname{usc} g(x) = +\infty \text{ for all } x. \end{cases}$$

To distinguish this operation from (3.8) when confusion might arise, we may refer to it as *upper* closure, as opposed to *lower* closure. Note that the *closed concave functions* are the constant functions $+\infty$ and $-\infty$ and the upper-semicontinuous concave functions g which are *proper* ($g(x) < +\infty$ for all x, and $g(x) > -\infty$ for at least one x).

So far, these definitions correspond to what happens if we pass to $-g$, apply the definition in the convex case, then reconvert to concavity, multiplying by -1. However, we do not define the conjugate in this way as $-f^*$ for $f = -g$. Rather we set

$$g^*(v) = \inf\{\langle x, v\rangle - g(x) | x \in X\}, \tag{3.25}$$

$$g^{**}(v) = \inf\{\langle x, v\rangle - g^*(v) | v \in V\}, \tag{3.26}$$

and this yields instead

$$g^*(v) = -f^*(-v) \quad \text{for} \quad f = -g. \tag{3.27}$$

We call this operation "taking the conjugate in the concave sense," as opposed to the "convex sense." Usually it is apparent from the context which sense is intended. The obvious analogues of Theorems 4 and 5 hold, with $g^{**} = \operatorname{cl} g$ for any concave function g.

4. Dual problems and Lagrangians. Following the pattern explained earlier, we take as the "primal" problem an abstract optimization problem of the form

$$\text{minimize} \quad f \text{ over } X, \tag{P}$$

where X is a linear space, and we suppose that a representation

$$f(x) = F(x, 0), \quad F: X \times U \to [-\infty, +\infty] \tag{4.1}$$

has been specified, where U is another linear space. Everything below depends on the particular choice of U and F. To express duality, we take X to be paired with a space V and U to be paired with a space Y. In applications, the choice of these pairings may also have a big effect. Of main interest are the cases where $F(x, u)$ is a closed convex function of $u \in U$ for each $x \in X$, or more restrictively, a closed convex function of x and u jointly. (The second property implies the first.)

We define the *Lagrangian* function K on $X \times Y$ by

$$K(x, y) = \inf\{F(x, u) + \langle u, y\rangle | u \in U\}. \tag{4.2}$$

Thus the function $y \to K(x, y)$ is the conjugate in the concave sense of the function $u \to -F(x, u)$. If the latter function is closed concave, the conjugacy is reciprocal (Theorem 5 in the concave case), i.e., the function $u \to -F(x, u)$ is the conjugate in the concave sense of the function $y \to K(x, y)$. Thus

$$F(x, u) = \sup\{K(x, y) - \langle u, y\rangle | y \in Y\} \quad \text{if } F(x, \cdot) \text{ is closed convex.} \tag{4.3}$$

The definition of the Lagrangian K is motivated by Example 1 in § 2, which yields (with $F(x, u)$ closed convex in u):

$$K(x, y) = \begin{cases} f_0(x) + y_1 f_1(x) + \cdots + y_m f_m(x) & \text{if} \quad x \in C, y \in R^m_+, \\ -\infty \quad \text{if} \quad x \in C, y \notin R^m_+, \\ +\infty \quad \text{if} \quad x \notin C. \end{cases} \tag{4.4}$$

THEOREM 6. *The Lagrangian $K(x, y)$ is closed concave in $y \in Y$ for each $x \in X$, and if $F(x, u)$ is closed convex in u one has*

$$f(x) = \sup_{y \in Y} K(x, y). \tag{4.5}$$

Conversely, if K is any extended-real-valued function on $X \times Y$ such that (4.5) *holds, and if $K(x, y)$ is closed concave in y, then K is the Lagrangian associated with a uniquely determined representation* (4.1) *having $F(x, u)$ closed convex in u, namely F given by the formula in* (4.3).

Furthermore, assuming $F(x, u)$ is closed convex in u, one has $K(x, y)$ convex in x if and only if $F(x, u)$ is convex jointly in x and u.

Proof. Except for the last assertion, everything follows via Theorem 5, formula (4.5) being the case of (4.3) with $u = 0$. If $K(x, y)$ is convex in x, the function

$$(x, u) \to K(x, y) - \langle u, y \rangle$$

is convex in x for each $y \in Y$, and hence, as the pointwise supremum of a collection of such functions by (4.3), $F(x, u)$ is convex in (x, u). On the other hand, if F is convex the function

$$(x, u) \to F(x, u) + \langle u, y \rangle$$

is convex; since the function $x \to K(x, y)$ is the infimum of this in u, it too is convex (Theorem 1).

In view of (4.5), we define the *dual problem* to (P) (relative to the representation (4.1)) as

$$\text{(D)} \qquad \text{maximize} \quad g \quad \text{over } Y,$$

where

$$g(y) = \inf_{x \in X} K(x, y). \tag{4.6}$$

The next result is perhaps the central theorem about dual problems. It relates g to the optimal value function

$$\varphi(u) = \inf_{x \in X} F(x, u), \tag{4.7}$$

which, as we recall from Theorem 1, is convex in u if F is convex in (x, u).

THEOREM 7. *The function g in* (D) *is closed concave. In fact, taking conjugates in the concave sense we have $g = (-\varphi)^*$, and hence $-g^* = \operatorname{cl} \operatorname{co} \varphi$. Thus*

$$\sup_{y \in Y} g(y) = (\operatorname{cl} \operatorname{co} \varphi)(0), \tag{4.8}$$

whereas

$$\inf_{x \in X} f(x) = \varphi(0). \tag{4.9}$$

In particular, if $F(x, u)$ is convex in (x, u) we have $-g^* = \operatorname{cl} \varphi$, and indeed

$$\sup_{y \in Y} g(y) = \liminf_{u \to 0} \varphi(u) \tag{4.10}$$

except in the case where $0 \notin \mathrm{cl}\,\mathrm{dom}\,\varphi \neq \varnothing$ *and the function* $\mathrm{lsc}\,\varphi$ *is nowhere finite-valued.* (*In the exceptional case, the limit in* (4.10) *is* $+\infty$, *as are the quantities* (4.9), *but* $g(y) \equiv -\infty$, *so that the supremum in* (4.10) *is* $-\infty$.)

Proof. We have by definition

$$\begin{aligned} g(y) &= \inf_{x\in X} \inf_{u\in U} \{F(x,u) + \langle u, y\rangle\} \\ &= \inf_{u\in U} \inf_{x\in X} \{F(x,u) + \langle u, y\rangle\} \\ &= \inf_{u\in U} \{\langle u, y\rangle + \varphi(u)\}, \end{aligned}$$

and the rest is immediate from Theorem 4 and Theorem 5.

The importance of Theorem 7 is that it transforms the question of whether

$$\inf_{x\in X} f(x) = \sup_{y\in Y} g(x), \tag{4.11}$$

and the related question of whether the saddle-value of the Lagrangian K exists, into the question of whether the optimal value function φ satisfies

$$\varphi(0) = \mathrm{cl}\,\mathrm{co}\,\varphi(0). \tag{4.12}$$

In the convex case, i.e., $F(x, u)$ convex in (x, u), the latter question reduces essentially to whether

$$\varphi(0) = \liminf_{u\to 0} \varphi(u). \tag{4.13}$$

At all events, (4.12) is conceptually more open to direct analysis than (4.11), because everything can be viewed in terms of the geometry of epigraphs, their convex hulls and closures. The heart of "conjugate" duality is this geometric outlook, combined with notions and terminology from analysis.

It is relatively easy in the convex case to formulate conditions guaranteeing (4.13) and consequently (4.12). (This is not to say that the conditions are always easy to verify in applications.) The convex case will be treated in detail below. Nevertheless, it should be realized that the convexity of $F(x, u)$ in (x, u) is not an absolute prerequisite for (4.12) to hold. While it is highly unlikely without this convexity for φ to be a convex function (only a couple of special instances are known), one may still be able to arrange for (4.12) in some nonconvex problems (P) through a careful choice of the representation (4.12). In this way, "nonconvex" applications of conjugate duality may be obtained; see Example 3′ in § 5 and Example 3″ in § 8.

We have shown how to pass from a problem (P) to a "meaningful" dual problem (D) by means of a representation (4.1). But for the duality to be truly symmetric the process should be reversible. Thus there should be a representation

$$g(y) = G(y, 0), \quad G: Y \times V \to [-\infty, +\infty], \tag{4.14}$$

which similarly generates, as the dual for (D), the original problem (P).

The natural representation to investigate, in view of (4.3), (4.5), and (4.6), is

$$G(y, v) = \inf\{K(x, y) - \langle x, v\rangle | x \in X\}; \tag{4.15}$$

this does at least satisfy (4.14). Since (D) is a maximization problem, the *optimal value function* we associate with G is

$$\gamma(v) = \sup_{y \in Y} G(y, v). \tag{4.16}$$

Substituting (4.2) into (4.15), we see that

$$\begin{aligned} G(y, v) &= \inf_{x,u} \{F(x, u) + \langle u, y\rangle - \langle x, v\rangle\} \\ &= -F^*(v, -y), \end{aligned} \tag{4.17}$$

where F^* is the conjugate of F (in the convex sense) with respect to the natural pairing induced between $X \times U$ and $V \times Y$:

$$\langle (x, u), (v, y)\rangle = \langle x, v\rangle + \langle u, y\rangle .$$

THEOREM 7'. *The function G is concave and closed, while γ is concave. If F is closed convex (jointly in x and u), we have in parallel with* (4.17):

$$F(x, u) = \begin{cases} \sup_{y,v} \{G(y, v) + \langle x, v\rangle - \langle u, y\rangle\} \\ -G^*(-x, u). \end{cases} \tag{4.18}$$

Then, taking conjugates in the convex sense yields $f = (-\gamma)^*$, and hence $-f^* = \mathrm{cl}\,\gamma$. Thus, while on the one hand

$$\sup_{y \in Y} g(y) = \gamma(0), \tag{4.19}$$

we also have, assuming F is closed convex,

$$\inf_{x \in X} f(x) = \limsup_{v \to 0} \gamma(v), \tag{4.20}$$

the only exception to the latter being when $0 \notin \mathrm{cl\,dom}\,\gamma \neq \emptyset$ and $\mathrm{usc}\,\gamma$ is nowhere finite-valued. (In the exceptional case, the limit in (4.20) *is $-\infty$, as are the quantities* (4.19), *but $f(x) \equiv +\infty$, so that the infimum in* (4.20) *is $+\infty$.)*

Proof. Theorem 5 yields from (4.17) the fact that G is concave and closed, and

$$\mathrm{cl\,co}\,F(x, u) = F^{**}(x, u) = -G^*(-x, u). \tag{4.21}$$

This implies (4.18) when F is closed convex. The concavity of γ follows from the concavity of G by Theorem 1. Setting $u = 0$ in (4.18) gives us

$$f(x) = \sup_{v} \{\gamma(v) + \langle x, v\rangle\} = (-\gamma)^*(x),$$

and hence $f^* = (-\gamma)^{**} = \mathrm{cl}\,(-\gamma)$, since $-\gamma$ is convex. Thus $-f^* = \mathrm{cl}\,\gamma$, implying

$$\mathrm{cl}\,\gamma(0) = -\sup_{x \in X} \{\langle x, 0\rangle - f(x)\} = \inf_{x \in X} f(x).$$

Everything else follows from Theorem 4.

COROLLARY 7'A. *If F is convex and closed, the relations* (4.10) *and* (4.20) *both hold, except in the degenerate case where all of the following properties are present:*

the limit in (4.10) *is* $-\infty$, $0 \notin \operatorname{cl} \operatorname{dom} \varphi$ (*implying the infimum in* (4.10) *is* $+\infty$), $\operatorname{lsc} \varphi(u) = -\infty$ *for all* $u \in \operatorname{cl} \operatorname{dom} \varphi$, *the limit in* (4.20) *is* $+\infty$, $0 \notin \operatorname{cl} \operatorname{dom} \gamma$ (*implying the supremum in* (4.20) *is* $-\infty$), *and* $\operatorname{usc} \gamma(v) = +\infty$ *for all* $v \in \operatorname{cl} \operatorname{dom} \gamma$.

Proof. This results from combining Theorems 7 and 7′.

The degenerate case described in Corollary 7′A is indeed possible. An example may be obtained from Example 1 in the case of an inconsistent linear programming problem whose dual is also inconsistent.

We conclude this section with some remarks on the nature of the Lagrangian functions K furnished by Theorem 6 and their role in general minimax theory.

Starting from the representation (4.14) of the maximization problem (D), it is natural to define in analogy with (4.2) a corresponding *dual* Lagrangian $\tilde{K}$ by

$$\tilde{K}(x, y) = \sup \{G(y, v) + \langle x, v \rangle | v \in V\}. \tag{4.22}$$

Inasmuch as $G(y, v)$ is closed concave in (y, v) (Theorem 7′), we have $\tilde{K}(x, y)$ closed convex in x and concave in y (parallel version of Theorem 6) and

$$G(y, v) = \inf \{\tilde{K}(x, y) - \langle x, v \rangle | x \in X\} \tag{4.23}$$

(Theorem 5 applied to (4.22)). In particular (setting $v = 0$ in (4.23)):

$$g(y) = \inf_{x \in X} \tilde{K}(x, y). \tag{4.24}$$

On the other hand, suppose F is closed convex. We can combine (4.22) with (4.18) to obtain

$$F(x, u) = \sup \{\tilde{K}(x, y) - \langle u, y \rangle | y \in Y\}, \tag{4.25}$$

and therefore

$$f(x) = \sup_{y \in Y} \tilde{K}(x, y). \tag{4.26}$$

Thus $\tilde{K}$ gives rise to the same problems (P) and (D) that K does, and even the same representations (4.1) and (4.14).

What then is the relationship between $\tilde{K}$ and the function K, which according to Theorem 6 is convex in x and closed concave in y? This is answered in a simple way using Theorem 5: formulas (4.22) and (4.15) tell us

$$\tilde{K}(x, y) = \operatorname{cl}_x K(x, y) \quad \text{(lower closure)}, \tag{4.27}$$

while (4.25) and (4.2) tell us

$$K(x, y) = \operatorname{cl}_y \tilde{K}(x, y) \quad \text{(upper closure)}. \tag{4.28}$$

However, we cannot hope to have $\tilde{K} = K$, except in unusual cases.

For example, assume in Example 1 that the set C is closed and the functions f_i are all l.s.c. on C. Then $F(x, u)$ is closed convex in (x, u). As we have noted, the corresponding Lagrangian K is given by (4.4). Instead of $\tilde{K} = K$, we have from (4.27):

$$\tilde{K}(x, y) = \begin{cases} f_0(x) + y_1 f_1(x) + \cdots + y_m f_m(x) & \text{if} \quad x \in C, y \in R^m_+, \\ +\infty & \text{if} \quad x \notin C, y \in R^m_+, \\ -\infty & \text{if} \quad y \notin R^m_+. \end{cases} \tag{4.29}$$

In this case the discrepancy between $\tilde{K}$ and K is "trivial," but other examples are known where, even in finite-dimensional spaces, the two functions can differ at special points where both are finite-valued, or where one is finite and the other infinite. In infinite-dimensional spaces, the relationship can be quite complicated.

The details of the relationship can be found in [13], [16]. The point we wish to make here is that the continuity properties of Lagrangian functions are not so elementary. One must be careful in theoretical developments not to make some arbitrary restriction for mathematical convenience (such as assuming $K(x, y)$ is l.s.c. in x and u.s.c. in y), because the effect on the class of representations (4.1) and (4.14) may be disagreeable, or hard to trace.

Another point is that different convex-concave functions can correspond to essentially the same minimax problem and dual problems (P) and (D). They therefore must be lumped together in a nontrivial way into equivalence classes. This idea is especially important in infinite-dimensional applications, since the "equivalence" may get rid of what otherwise seem to be troublesome technical ambiguities in how to express the Lagrangian.

A final remark is that virtually every reasonable minimax problem of "convex-concave" type corresponds to some Lagrangian K and hence to a dual pair of problems (P) and (D) in the relationships above. To illustrate without getting into too many technicalities, suppose for instance that K_0 is a real-valued function on $C \times D \subset X \times Y$, where C and D are nonempty closed convex sets, $K_0(x, y)$ is l.s.c. in $x \in C$ and u.s.c. in $y \in D$. Define

$$K(x, y) = \begin{cases} K_0(x, y) & \text{if } x \in C, y \in D, \\ -\infty & \text{if } x \in C, y \notin D, \\ +\infty & \text{if } x \notin C. \end{cases} \tag{4.30}$$

Then the minimax problem for K on $X \times Y$ is equivalent to the one for K_0 on $C \times D$. Moreover K is closed concave in y and convex in x, and hence by Theorem 6 it is the Lagrangian corresponding to the F defined by (4.3) (and f defined by (4.1) or (4.5)). The lower-semicontinuity of $K_0(x, y)$ in x actually implies via (4.3) that F is *closed* convex. Theorems 7 and 7′ are therefore both applicable to the minimax problem.

Thus "convex-concave" minimax theory is essentially *equivalent* to the theory of dual optimization problems, as presented here. For more on this see Example 13 in § 8.

5. Examples of duality schemes. The duality scheme in § 4 will now be discussed in terms of the examples in § 2 and other general models. The treatment of Examples 6–10 is postponed however until § 10, since it requires the results in § 9 concerning integral functionals. Some of the examples below will be analyzed further in § 8.

Example 1′. (Nonlinear programming.) Since the Lagrangian K for Example 1 is given by (4.4), the dual objective function is

$$g(y) = \begin{cases} \inf_{x \in C} \{f_0(x) + y_1 f_1(x) + \cdots + y_m f_m(x)\} & \text{if } y \geqq 0, \\ -\infty & \text{if } y \ngeqq 0. \end{cases} \tag{5.1}$$

Thus the condition $y \geqq 0$ is an implicit constraint in (D). But there may be other implicit constraints, since the infimum could be $-\infty$ for some values of $y_i \geqq 0$. There are many special cases of interest where g and these implicit constraints can be determined more explicitly, but to illustrate the ideas it will be enough to see what happens in the linear programming case

$$f_i(x) = \langle x, b_i \rangle + \beta_i, \qquad i = 0, 1, \cdots, m, \tag{5.2}$$

with $C = X$. Then we may calculate the whole function G from (4.15): if $y \ngeqq 0$ we have $G(v, y) = -\infty$, while if $y \geqq 0$, then

$$\begin{aligned} G(v, y) &= \inf_{x \in X} \left\{ \langle x, b_0 \rangle + \beta_0 + \sum_{i=1}^{m} y_i[\langle x, b_i \rangle + \beta_i] - \langle x, v \rangle \right\} \\ &= \inf_{x \in X} \left\{ \beta_0 + \sum_{i=1}^{m} y_i \beta_i + \left\langle x, b_0 + \sum_{i=1}^{m} y_i b_i - v \right\rangle \right\} \\ &= \begin{cases} \beta_0 + \sum_{i=1}^{m} y_i \beta_i & \text{if} \quad b_0 + \sum_{i=1}^{m} y_i b_i = v, \\ -\infty & \text{if} \quad b_0 + \sum_{i=1}^{m} y_i b_i \neq v. \end{cases} \end{aligned} \tag{5.3}$$

The dual problem is therefore the linear programming problem that one would expect:

$$\begin{gathered} \text{maximize} \quad \beta_0 + \sum_{i=1}^{m} y_i \beta_i \quad \text{subject to} \\ y_i \geqq 0, \quad b_0 + \sum_{i=1}^{m} y_i b_i = 0. \end{gathered} \tag{5.4}$$

Moreover the dual optimal value function γ gives the maximum as a function of a vector $v \in V$ replacing the 0 in the equation constraint.

It is not hard to see in this setting why the classical duality theorem for linear programming problems is valid. Assuming for simplicity that $X = R^n = V$, one notes that the epigraph of the function F can be described as the intersection of a finite collection of half-spaces (the epigraph of f_0 and the "vertical half-spaces" corresponding to the conditions $f_i(x) - u_i \leqq 0$), i.e., it is a *polyhedral* convex set. The epigraph of φ is obtained from the epigraph of F through the projection $(x, u, \alpha) \to (u, \alpha)$. But the projected image of a polyhedral convex set is polyhedral, hence closed [13, § 19]. Therefore φ is l.s.c., and we have $\varphi(0) = \text{cl co}\, \varphi(0)$ unless $0 \notin \text{dom}\, \varphi$ and φ is identically $-\infty$ on $\text{dom}\, \varphi$. Similarly, γ is a "polyhedral" concave function, so that $\gamma(0) = \text{cl}\, \gamma(0)$ unless $0 \notin \text{dom}\, \gamma$ and γ is identically $+\infty$ on dom g. In particular, the optimal values in (P) and (D) are equal, if the problems are not both "inconsistent."

Example 2′. (Nonlinear programming.) We consider the parameteric representation function F of Example 2 of § 2 in the more explicit form where $l_i(x) = \langle x, b_i \rangle + \beta_i$. Let Z_i be paired with W_i. Direct calculation from (4.2) yields for

$y = (y_1, \cdots, y_m, w_0, \cdots, w_m)$ the Lagrangian expression

$$(5.5)\qquad K(x,y) = \begin{cases} \langle A_0x + a_0, w_0\rangle + \langle b_0, x\rangle + \beta_0 \\ \quad + \sum_{i=1}^m [\langle A_ix + a_i, w_i\rangle + y_i(\langle b_i, x\rangle + \beta_i)] \\ \quad - h_0^*(w_0) - (y_1h_1)^*(w_1) - \cdots - (y_mh_m)^*(w_m) \\ \qquad\qquad \text{if } (y_1, \cdots, y_m) \geqq 0, \\ -\infty \quad \text{if} \quad (y_1, \cdots, y_m) \ngeqq 0. \end{cases}$$

This is linear in x. Taking the supremum in x and using the identity $\langle A_ix, w_i\rangle = \langle x, A_i^*w_i\rangle$, where $A_j^*: W_i \to X$ is the adjoint of A_i, we see that the dual problem amounts to

$$(5.6)\qquad \begin{aligned} &\text{maximize} \quad \beta_0 + \sum_{i=1}^m y_i\beta_i + \sum_{i=1}^m \langle a_i, w_i\rangle \\ &\qquad\qquad - h_0^*(w_0) - (y_1h_1)^*(w_1) - \cdots - (y_mh_m)^*(w_m) \\ &\text{subject to} \quad y_i \geqq 0, \quad b_0 + \sum_{i=1}^m y_ib_i + \sum_{i=0}^m A_i^*w_i = 0. \end{aligned}$$

What is especially interesting about this dual is that the constraints are linear, except for possible implicit constraints describing the effective domains of the conjugate functions h_0^* and $(y_ih_i)^*$. In a number of important cases, for instance in quadratic or exponential programming (formulas (2.12) and (2.13) respectively, see also (3.17)–(3.20)), we are able to calculate the conjugates explicitly and determine that the effective domains are elementary polyhedral convex sets, or "essentially" so. In these cases (D) is essentially a *linearly constrained* problem, even though (P) is nonlinearly constrained. Special properties of the functions h_i then lead to the result that, if (P) has a feasible solution, the dual optimal value function λ is *closed* concave, and hence in particular the optimal values in (P) and (D) coincide [20]. For more examples in this vein, see [27].

Example 3′. (Nonconvex programming.) We consider the problem of Examples 1 and 3 in § 2, but without the assumption that the functions f_i are convex. The representation functions F in the two examples are still closed convex in u; we denote the one in (2.8) by F and the one in (2.14) by F_r. The corresponding optimal value functions satisfy

$$(5.7)\qquad \varphi_r(u) = \varphi(u) + r|u|^2, \qquad r \geqq 0.$$

One sees geometrically by way of epigraphs that, while φ may be far from satisfying (4.12), φ_r may satisfy it for r sufficiently large, thereby yielding a "nonconvex" duality theorem. We shall return to this possibility in more specific terms in Example 3″ in § 8.

The Lagrangian K_r associated with F_r for $r > 0$ is easily calculated right from (4.2), and it turns out to be

$$(5.8)\qquad K_r(x,y) = \begin{cases} f_0(x) + \sum_{i=1}^m \lambda_r(f_i(x), y_i) & \text{if} \quad x \in C, \\ +\infty \quad \text{if} \quad x \notin C, \end{cases}$$

where

$$\lambda_r(f_i(x), y_i) = \begin{cases} y_i f_i(x) + r f_i(x)^2 & \text{if } f_i(x) \geqq -y_i/2r, \\ -y_i^2/4r & \text{if } f_i(x) \leqq -y_i/2r. \end{cases} \tag{5.9}$$

Notice the absence of any nonnegativity condition on y_i like the one in (4.4).

Example 4′. (Nonlinear programming with generalized constraints.) To modify the model in Example 4 of § 2 slightly, let U denote a linear space in which a certain nonempty convex cone Q has been singled out as the "nonpositive orthant" for a partial ordering, and consider the problem

$$\text{(P)} \qquad \text{minimize} \quad f_0(x) \quad \text{subject to} \quad x \in C, \quad \Phi(x) \in Q,$$

where $C \subset X$ is convex, $f_0 : C \to R$ is convex, and where $\Phi : C \to U$ is convex in the sense that

$$\Phi((1-\lambda)x + \lambda x') - (1-\lambda)\Phi(x) - \lambda\Phi(x') \in Q, \qquad 0 < \lambda < 1. \tag{5.10}$$

We suppose U is paired with a space Y in such a way that Q is closed. Selecting the parametric representation

$$F(x, u) = \begin{cases} f_0(x) & \text{if} \quad x \in C \quad \text{and} \quad \Phi(x) - u \in Q, \\ +\infty & \text{otherwise}, \end{cases} \tag{5.11}$$

which is convex in (x, u), closed in u, we obtain from (4.2) the Lagrangian

$$K(x, y) = \begin{cases} f_0(x) + \langle \Phi(x), y \rangle & \text{if} \quad x \in C, \quad y \in Q^*, \\ -\infty & \text{if} \quad x \in C, y \notin Q^*, \\ +\infty & \text{if} \quad x \notin C, \end{cases} \tag{5.12}$$

where Q^* is the polar of Q. Thus, similar to Example 1′ above, the dual consists of maximizing

$$g(y) = \inf_{x \in C} \{ f_0(x) + \langle \Phi(x), y \rangle \}, \qquad y \in Q^*. \tag{5.13}$$

If for instance the set C is compact and the functions f_0 and Φ are continuous, it is simple to prove that the (convex) optimal value function φ is lower-semicontinuous and proper, implying the optimal values in (P) and (D) coincide. Deeper results of this sort, involving weaker conditions of compactness, will be established in § 7 (see parts (d) and (e) of Theorem 18′ in conjunction with Theorem 17′). Other results use generalizations of the Slater condition; see Example 4″ in § 8.

Infinite linear programming is covered by this example. In other specific cases, one might approach the model along the lines of Examples 2, 2′. The idea in Examples 3, 3′, can also lead to something.

Example 11. (Fenchel duality.) This is a useful general model for many applications involving linear operators. The primal problem is of the form

$$\text{(P)} \qquad \text{minimize} \quad h(x) - k(Ax), \quad x \in X,$$

where h is a proper convex function on X, k is a proper concave function on U, and $A: X \to U$ is a linear operator. We take

$$F(x, u) = h(x) - k(Ax + u). \tag{5.14}$$

Then F is convex, and the corresponding convex-concave Lagrangian is

$$K(x, y) = \begin{cases} h(x) + k^*(y) - \langle Ax, y\rangle & \text{if} \quad h(x) < +\infty, \\ +\infty \quad \text{if} \quad h(x) = +\infty. \end{cases} \tag{5.15}$$

In terms of the adjoint linear transformation $A^*: Y \to V$, we obtain

$$G(y, v) = k^*(y) - h^*(A^*y + v) \tag{5.16}$$

and the dual problem

(D) $\qquad$ maximize $\quad k^*(y) - h^*(A^*y), \quad y \in Y.$

Elementary choices of the functions h and k turn these problems into familiar ones in linear programming, quadratic programming, etc.

The definition of the adjoint A^* can involve some subtleties in infinite dimensions, particularly if, as often happens when dealing with differential operators, one wants to encompass transformations A that are only densely defined. If A is everywhere-defined and continuous, the relation

$$\langle Ax, y\rangle = \langle x, A^*y\rangle \quad \text{for all } x, y, \tag{5.17}$$

furnishes the definition of A^*: for each $y \in Y$, the continuous linear functional $x \to \langle Ax, y\rangle$ corresponds to an element $v \in V$, and this v is what is denoted by A^*y. To ensure that v is uniquely determined, it is necessary to assume the pairing has the property that

$$\langle x, v\rangle = 0 \quad \text{for all} \quad x \in X \quad \text{implies} \quad v = 0. \tag{5.18}$$

This property can always be arranged, if needed, by identifying as equivalent any elements of V which induce the same linear functional on X. The companion property,

$$\langle u, y\rangle = 0 \quad \text{for all} \quad y \in Y \quad \text{implies} \quad u = 0, \tag{5.19}$$

is required in showing that the adjoint A^{**} is in turn well-defined and coincides with A. *Hence we always assume both properties* (5.18) *and* (5.19) *tacitly in any discussion of adjoints.*

If the linear transformation A is not everywhere-defined or continuous, a suitable A^* still exists, provided that the domain of definition dom A is a *dense* subspace of X and the set

$$\text{graph } A = \{(x, u) | x \in \text{dom } A, u = Ax\} \tag{5.20}$$

is closed in $X \times U$. Namely, for each $y \in Y$ the linear functional $x \to \langle Ax, y\rangle$ on dom A may or may not be continuous. If it is, there exists (by the density of dom A) a unique $v \in V$ such that $\langle Ax, y\rangle = \langle x, v\rangle$ for all $x \in \text{dom } A$; we denote this v by A^*y and say that $y \in \text{dom } A^*$. It can be shown that then dom A^* is a dense subspace of Y, the transformation $A^*: \mathrm{dom}\, A^* \to V$ is linear, and graph A^* is closed in $Y \times V$. Moreover, the adjoint A^{**} in the sense of the same definition satisfies $A^{**} = A$. (This all follows easily through consideration of the annihilator spaces (graph $A)^{\perp}$ and (graph $A)^{\perp\perp}$.)

Of course in the case just described, where A or A^* may not be everywhere defined, it is necessary to modify formulas (5.14) to

$$F(x,u) = \begin{cases} h(x) - k(Ax+u) & \text{if} \quad x \in \mathrm{dom}\, A, \\ +\infty & \text{if} \quad x \notin \mathrm{dom}\, A. \end{cases} \tag{5.21}$$

Then (5.15) becomes in turn

$$K(x,y) = \begin{cases} h(x) + k^*(y) - \langle Ax, y\rangle & \text{if} \quad x \in \mathrm{dom}\, h \cap \mathrm{dom}\, A, \\ +\infty & \text{otherwise}, \end{cases} \tag{5.22}$$

and we have

$$G(y,v) = \begin{cases} k^*(y) - h^*(A^*y+v) & \text{if} \quad y \in \mathrm{dom}\, A^*, \\ -\infty & \text{if} \quad y \notin \mathrm{dom}\, A^*. \end{cases} \tag{5.23}$$

It is interesting to observe that the dual Lagrangian in this situation is

$$\tilde{K}(x,y) = \begin{cases} h(x) + k^*(y) - \langle x, A^*y\rangle & \text{if} \quad y \in \mathrm{dom}\, k^* \cap \mathrm{dom}\, A^*, \\ -\infty & \text{otherwise}. \end{cases} \tag{5.24}$$

As a matter of fact, the two functions

$$K_0(x,y) = \begin{cases} \langle Ax, y\rangle & \text{if} \quad x \in \mathrm{dom}\, A, \\ +\infty & \text{if} \quad x \notin \mathrm{dom}\, A, \end{cases} \tag{5.25}$$

$$\tilde{K}_0(x,y) = \begin{cases} \langle x, A^*y\rangle & \text{if} \quad y \in \mathrm{dom}\, A^*, \\ -\infty & \text{if} \quad y \notin \mathrm{dom}\, A^*, \end{cases} \tag{5.26}$$

obey the closure relations (4.27) and (4.28) and thus form an example of convex-concave functions equivalent to each other in the sense alluded to towards the end of § 4. Without awareness of this, the possible differences between (5.25) and (5.26) could be a source of perplexity.

The Fenchel duality model can easily be generalized to problems of the form

$$\text{minimize} \quad q(x, Ax) \quad \text{over all} \quad x \in \mathrm{dom}\, A, \tag{P$'$}$$

where q is a convex function on $X \times U$, by setting

$$F(x,u) = \begin{cases} q(x, Ax+u) & \text{if} \quad x \in \mathrm{dom}\, A, \\ +\infty & \text{if} \quad x \notin \mathrm{dom}\, A. \end{cases} \tag{5.27}$$

One then has

$$K(x,y) = \begin{cases} Q(x,y) - \langle Ax, y\rangle & \text{if} \quad x \in \mathrm{dom}\, A, \\ -\infty & \text{if} \quad x \notin \mathrm{dom}\, A, \end{cases}$$

where

$$Q(x, y) = \inf_u \{q(x, u) + \langle u, y \rangle\} \tag{5.28}$$

and

$$G(y, v) = \begin{cases} -q^*(-A^*y + v, y) & \text{if } y \in \operatorname{dom} A^*, \\ -\infty & \text{if } y \notin \operatorname{dom} A^*. \end{cases} \tag{5.29}$$

The dual problem is thus:

(D′) maximize $-q^*(-A^*y, y)$ over all $y \in \operatorname{dom} A^*$.

We remark that problem (P′) can be viewed as one of minimizing a convex function q over a certain subspace of the linear space $X \times U$, namely over the graph of A. In the same light, (D′) consists of maximizing $-q^*$ over the "orthogonal" subspace $(\text{graph } A)^\perp$. The next example explores this idea in more direct terms.

Example 12. (Complementary duality.) As an abstract model, let us consider a problem of the form

(P_0) minimize $F(z)$ subject to $z \in X$ (subspace) $\subset Z$,

where $F: Z \to [-\infty, +\infty]$ and the subspace X is closed. (Z is a linear space paired with a linear space W.) Suppose (as is certainly true if Z is finite-dimensional or a Hilbert space) that there is a closed subspace U of Z *complementary* to X, i.e., such that each $z \in Z$ can be represented uniquely as a sum $x + u$, where $x \in X$ and $u \in U$ depend continuously on z. Then Z can be identified with $X \times U$, and the problem (P_0) can be rewritten as

$$\text{minimize} \quad F(x, u) \quad \text{subject to} \quad x \in X, \quad u = 0, \tag{5.30}$$

or in other words,

(P) minimize $f(x) = F(x, 0)$ over all $x \in X$.

But this is precisely our general model with parameters.

The "annihilator" subspaces $Y = X^\perp$ and $V = U^\perp$ in W are likewise complementary to each other, so that W can be identified with $U \times Y$: for $z = (x, u)$ and $w = (v, y)$ we have

$$\langle z, w \rangle = \langle x, v \rangle + \langle u, y \rangle.$$

Regarding X as paired with V, and U as paired with Y, the dual associated with (P) is

(D) maximize $g(y) = G(y, 0)$ over all $y \in Y$,

where

$$G(y, v) = -F^*(-v, y). \tag{5.31}$$

But this can be written also as

(D_0) maximize $-F^*(w)$ subject to $w \in X^\perp \subset W$.

Thus the dual of minimizing a function F over a subspace X is to minimize F^* over $X^\perp$. Moreover, *this is a description valid for all our dual pairs of problems* (P) *and* (D), since they can always be reformulated as (P_0) and (D_0) by setting $Z = X \times U$ and $W = V \times Y$ (identifying X with $X \times \{0\}$, etc.). However, it is not as effective as a basic scheme for the theory of duality for several reasons. It does not lead unambiguously to a Lagrangian function and associated minimax problem (with corresponding optimality conditions and game-theoretic interpretation of duality), because these depend on the particular choice of U. Nor does it provide as natural a setting for the study of optimal value functions like φ and γ, which turn out to be so important. The complementary subspace formulation can also be a conceptual stumbling block in applications where there is no obvious subspace at hand (such as Example 1). However, it is often useful in theoretical analysis for its simplicity.

Incidentally, one can always pass from (P_0) to (D_0) in terms of Fenchel duality, without relying on the existence of a subspace complementary to X. Taking the linear transformation A in Example 11 to be the identity (identifying the spaces in question), we get dual pairs of problems of the general type

$$(P_1) \qquad \text{minimize} \quad h(z) - k(z), \quad z \in Z,$$

$$(D_1) \qquad \text{maximize} \quad k^*(w) - h^*(w), \quad w \in W,$$

where Z and W are paired spaces. (These are the original problems introduced by Fenchel in the finite-dimensional case—the starting point of conjugate duality.) If we now set $h = F$ and $k = -\psi_X$ (indicator), where X is a subspace of Z, we get the (concave) conjugate $k^* = -\psi_{X^\perp}$, and thus (P_0) and (D_0). An immediate generalization is to take X to be a cone: then $X^\perp$ is replaced by the polar cone X^*.

This discussion demonstrates that all the schemes, Fenchel duality, complementary duality, the original Fenchel duality, and the "parametric" conjugate duality presented in these notes, are capable of generating the same pairs of dual problems. The differences lie in flexibility in application, richness of results and interpretations, and potential relevance in the analysis of nonconvex problems.

6. Continuity and derivatives of convex functions. We have seen that the study of the minimum and maximum in a dual pair of problems and of the existence of the saddle-value in a minimax problem can be reduced very broadly to questions about the continuity properties of certain convex and concave functions at the origin. It happens also that the study of solutions to such problems is intimately connected with the differentiability properties of the same functions. In this section we review the basic results of convex analysis concerning continuity and differentiability, in preparation for their application.

Throughout, we assume that U and Y are paired spaces, and that $\varphi : U \to [-\infty, +\infty]$ is an arbitrary function. We denote the topological interior of a set $C \subset U$ by int C and define

$$(6.1) \qquad \operatorname{core} C = \{u \in C \mid \forall u' \in U, \exists \varepsilon > 0, \forall \lambda \in [-\varepsilon, +\varepsilon], u + \lambda u' \in C\}.$$

For *convex* sets C, we have $\text{int}\, C = \text{core}\, C$ under any *one* of the following conditions: (a) $\text{int}\, C \neq \emptyset$, (b) $U = R^n$, (c) C is closed and U is a Banach space (or "barrelled" space) in the compatible topology in question [8].

THEOREM 8. *The two conditions*

(i) φ *is (finitely) bounded on a neighborhood of some point of* U,

(ii) $\text{int epi}\, \varphi \neq \emptyset$,

are equivalent and imply $\text{int dom}\, \varphi \neq \emptyset$. *If* φ *is convex, they imply* φ *is continuous on* $\text{core dom}\, \varphi = \text{int dom}\, \varphi$.

Proof. See [11], for example.

COROLLARY 8A. *If* $U = R^n$ *and* φ *is convex, then* φ *is continuous on* $\text{core dom}\, \varphi = \text{int dom}\, \varphi$. *In particular, every finite convex function on* R^n *is closed.*

Proof. If $\bar{u} \in \text{core dom}\, \varphi$, we can find points $a_1, \cdots, a_m$ in $\text{dom}\, \varphi$ such that the set $N = \text{int co}\, \{a_1, \cdots, a_m\}$ contains $\bar{u}$. Choose a real number α such that $\alpha \geqq \varphi(a_i)$ for $i = 1, \cdots, m$. The pairs (a_i, α) all belong to $\text{epi}\, \varphi$, which is convex, and hence so do all the pairs (u, α) for $u \in N$. Thus φ is bounded above by α on N.

COROLLARY 8B. *If* U *is a Banach space (or barrelled space) and* φ *is convex and l.s.c. (or closed), then* φ *is continuous on* $\text{core dom}\, \varphi = \text{int dom}\, \varphi$.

Proof. Let $\bar{u} \in \text{core dom}\, \varphi$ and $\varphi(\bar{u}) < \alpha < +\infty$. Let $C = \{u | \varphi(u) \leqq \alpha\}$. Then C is closed convex and $\bar{u} \in C$. An elementary argument, invoking the properties of convex functions along the line segments, shows in fact that $\bar{u} \in \text{core}\, C$ and hence $\bar{u} \in \text{int}\, C$. Thus φ is bounded above on a neighborhood of $\bar{u}$.

A subset D of Y is said to be *bounded* if every continuous linear functional on Y has a finite upper bound on D, i.e., if the *support function*

$$\psi_D^*(u) = \sup_{y \in D} \langle u, y \rangle \tag{6.2}$$

satisfies $\psi_D^*(u) < +\infty$ for all $u \in U$. In the finite-dimensional case ($U = R^n = Y$, $\langle u, y \rangle = u \cdot y$) closed bounded sets are compact, but the infinite-dimensional situation is subtler. It is not always true even that closed bounded subsets of Y are compact relative to the *weak topology* induced on Y by the pairing with U (that is, the coarsest topology under which the linear functions $y \to \langle u, y \rangle$ are continuous, a topology known always to be compatible with the pairing).

Another concept becomes useful: D is said to be *equicontinuous* if the support function ψ_D^* is not just $< +\infty$ everywhere, but actually bounded above on some neighborhood of the origin in U. (In view of Theorem 8, this condition is equivalent to ψ_D^* being continuous on U.) The "equicontinuity" refers to the family of linear functionals $\langle \cdot, y \rangle$ on U corresponding to elements $y \in D$. One might prefer, instead of this standard terminology, to speak of D being *fully bounded (in relation to the topology assigned to* U).

An important theorem in functional analysis asserts that if D is closed, equicontinuous and *convex*, then D is indeed compact relative to the weak topology induced on Y by U.

Incidentally, boundedness of D alone is sufficient to imply that ψ_D^* is bounded above on all equicontinuous subsets of U (the latter being defined analogously).

Of course, in the finite-dimensional case "boundedness" and "equicontinuity" are equivalent, as are "compactness" and "weak compactness."

THEOREM 9. *Let h be any closed convex function on Y other than the constant function* $-\infty$ (*for instance* $h = \varphi^*$, *unless* $\varphi \equiv +\infty$). *Let D be a nonempty bounded subset of Y. Then*

$$\inf\{h(y) \mid y \in D\} > -\infty. \tag{6.3}$$

If in addition D is closed, convex and equicontinuous, the infimum is attained at some point of D.

Proof. There exists u such that

$$+\infty > h^*(u) = \sup\{\langle u, y\rangle - h(y) \mid y \in Y\}$$

(Theorem 5). Let α be a real number, $\alpha \geqq h^*(u)$. Then

$$h(y) \geqq \langle u, y\rangle - \alpha \quad \text{for all} \quad y \in Y,$$

and hence the infimum (6.3) is bounded below by

$$\inf\{\langle u, y\rangle - \alpha \mid y \in D\} = -\alpha - \sup\{\langle -u, y\rangle \mid y \in D\} = -\alpha - \psi_D^*(-u) > -\infty.$$

To prove the second assertion, we denote the infimum (6.3) by $\bar{\beta}$, assuming for nontriviality that $\bar{\beta} < +\infty$, and we observe that the sets

$$\{y \in D \mid h(y) \leqq \beta\}, \qquad \bar{\beta} < \beta < +\infty, \tag{6.4}$$

are nonempty, closed, convex and equicontinuous. Hence they are compact in the weak topology induced on Y by U. Since a nest of nonempty compact sets has a nonempty intersection, it follows that the set $\{y \in D \mid h(y) \leqq \bar{\beta}\}$ is nonempty.

A word of caution: it is not true in the infinite-dimensional case that even a continuous (finite) convex function is necessarily bounded *above* on bounded sets. Counterexamples exist already for Hilbert spaces.

The next result expresses a duality between the preceding concepts of continuity and boundedness. An obvious application is to the situation in Theorem 7, where for the convex optimal value function φ, one has $\varphi^*(y) = -g(-y)$, and g is the concave function being maximized in (D). Similarly the situation in Theorem 7′, where for the concave optimal value function γ associated with (D) we have $(-\gamma)^* = f$.

THEOREM 10. (a) *If* $0 \in \operatorname{core} \operatorname{dom} \varphi$, *then the level sets*

$$\{y \in Y \mid \varphi^*(y) \leqq \beta\}, \quad \beta \in R, \tag{6.5}$$

are bounded. Conversely, if one of these level sets for $\beta > \inf \varphi^*$ *is bounded, then* $0 \in \operatorname{core} \operatorname{dom} \varphi^{**}$.

(b) *If* φ *is* (*finitely*) *bounded above on a neighborhood of* 0, *then the level sets* (6.5) *are all equicontinuous* (*also closed, convex, and hence compact in the weak topology induced on Y by U*). *Conversely, if one of these level sets for* $\beta > \inf \varphi^*$ *is equicontinuous, then* $0 \in \operatorname{core} \operatorname{dom} \varphi^{**}$ *and* φ^{**} *is continuous at* 0.

(c) *In the finite-dimensional case with* φ *convex, one has* $0 \in \operatorname{core} \operatorname{dom} \varphi$ *if and only if: the level sets* (6.5) *are all compact, and either one of them is nonempty or* $\varphi(0) < +\infty$. *Moreover, the level sets are all compact if any one of them is nonempty and does not include a half-line.*

Proof. See, for example, [11], [2], [1, Theorem 2].

For closed convex functions φ, "if and only if" versions of (a) and (b) in Theorem 11 can be stated, since $\varphi^{**} = \varphi$.

We turn now to differentiability properties. Suppose φ is convex and u is a point at which φ is finite. Then for arbitrary $u' \in U$ the difference quotient

$$[\varphi(u + \lambda u') - \varphi(u)]/\lambda, \qquad \lambda > 0, \tag{6.6}$$

makes sense. This expression is convex as a function of u' for fixed λ, and one can show easily from convexity that it is also nondecreasing as a function of $\lambda > 0$ for fixed u'. Therefore the *directional derivative*

$$\varphi'(u; u') = \lim_{\lambda \downarrow 0} [\varphi(u + \lambda u') - \varphi(u)]/\lambda \tag{6.7}$$

is well-defined (possibly $-\infty$ or $+\infty$), the limit being the same as the infimum over all $\lambda > 0$. Moreover $\varphi'(u; u')$ is *convex* in u' and satisfies

$$\varphi'(u; \alpha u') = \alpha\varphi'(u; u') \quad \text{for} \quad \alpha > 0. \tag{6.8}$$

If the limit (6.7) depends continuously on u' and exists in the two-sided sense (for $\lambda \to 0$), i.e.,

$$\varphi'(u; -u') = -\varphi'(u; u') \quad \text{for all} \quad u', \tag{6.9}$$

then $\varphi'(u; u')$ is a continuous linear function of u' and hence corresponds to some $y \in Y$. This y is called the *gradient* of φ at u and denoted by $\nabla\varphi(u)$; one then has

$$\varphi'(u; u') = \langle u', \nabla\varphi(u)\rangle \quad \text{for all} \quad u' \in U. \tag{6.10}$$

The existence of $\nabla\varphi(u)$ entails $u \in \operatorname{core} \operatorname{dom} \varphi$.

While gradients do not always exist, the directional derivatives above can nevertheless be characterized in terms of a more general notion. If φ is a convex function finite at u, an element $y \in Y$ is called a *subgradient* of φ at u if

$$\varphi'(u; u') \geqq \langle u', y\rangle \quad \text{for all} \quad u' \in U, \tag{6.11}$$

or equivalently (since the limit in (6.7) is the same as the infimum over $\lambda > 0$) if

$$\varphi(u + \lambda u') \geqq \varphi(u) + \lambda\langle u', y\rangle \quad \text{for all} \quad u' \in U, \qquad \lambda > 0.$$

The latter condition can better be written as

$$\varphi(u') \geqq \varphi(u) + \langle u' - u, y\rangle \quad \text{for all } u' \in U, \tag{6.12}$$

and this makes sense when $\varphi(u)$ is not finite, even when φ is not convex. Therefore (6.12) is adopted as the definition of y being a *subgradient* at u in the general case; the set of all such elements y is denoted by $\partial\varphi(u)$. Geometrically, when $\varphi(u)$ is finite, (6.12) says that the epigraph of the affine function

$$u' \to \varphi(u) + \langle u' - u, y\rangle$$

is a supporting half-space to the epigraph of φ at the point $(u, \varphi(u))$.

It must be emphasized that the definitions of "gradient" and "subgradient" depend on the choice of the pairing of U with a space Y. This is a potential source of confusion in some infinite-dimensional applications, where different pairings may be equally suitable or needed for different purposes.

The subgradient set $\partial\varphi(u)$ may be nonempty, but it is always closed and convex in Y. Indeed, (6.12) defines $\partial\varphi(u)$ as the set of points y satisfying a certain system of (continuous) linear inequalities, i.e., as the intersection of a collection of closed half-spaces.

THEOREM 11. *Let φ be convex, and let u be a point where φ is finite.*

(a) *If φ is continuous at u, then $\partial\varphi(u)$ is nonempty and bounded (in fact equicontinuous, hence weakly compact), and*

$$\varphi'(u;u') = \max\{\langle u',y\rangle | y \in \partial\varphi(u)\} \quad \text{for all } u'. \tag{6.13}$$

(b) *More generally, the conjugate of the convex function $\theta(u') = \varphi'(u;u')$ is the indicator of the set $\partial\varphi(u)$, and hence the support function of $\partial\varphi(u)$ (which is the conjugate of this indicator) is* cl θ:

$$\text{cl}\,\theta(u') = \sup\{\langle u',y\rangle | y \in \partial\varphi(u)\} \quad \text{for all } u'. \tag{6.14}$$

(c) *In particular, if φ satisfies one of the equivalent conditions* (i) *or* (ii) *in Theorem* 8 *and u' is such that*

$$u + \lambda u' \in \text{core dom}\,\varphi \text{ for some } \lambda \geqq 0, \tag{6.15}$$

then θ is continuous at u', and $\varphi(u;u')$ can be written in place of cl $\theta(u')$ *in* (6.14).

Proof. The first assertion in (b) is easily verified by direct computation using (6.8) and the defining condition (6.11) for $y \in \partial\varphi(u)$. One then invokes Theorem 5. Next, (c) is obtained from Theorem 8 and the fact that

$$h(u') \leqq [\varphi(u + \lambda u') - \varphi(u)]/\lambda, \qquad \lambda > 0.$$

Then (a) follows, at least with "sup," because (6.15) is fulfilled for every u'; moreover, $\varphi'(u;u')$ is everywhere continuous in u'. Thus $\partial\varphi(u)$ is a set whose support function is continuous everywhere, i.e., $\partial\varphi(u)$ is equicontinuous. Since $\partial\varphi(u)$ is also closed and convex, it is weakly compact. If $\partial\varphi(u)$ were empty, its support function would be identically $-\infty$, contrary to $\varphi'(u;0) = 0$. Thus the supremum in (6.13) is indeed attained.

How are $\partial\varphi^*$ and $\partial\varphi^{**}$ related to $\partial\varphi$? Definition (6.12) of "subgradient" gives us

$$\partial\varphi(u) = \{y | \langle u,y\rangle - \varphi^*(y) \geqq \varphi(u)\}, \tag{6.16}$$

where

$$\langle u,y\rangle - \varphi^*(y) \leqq \varphi(u) \quad \text{for all } u, y, \tag{6.17}$$

by the definition of φ^*. Dually,

$$\partial\varphi^*(y) = \{u | \langle u,y\rangle - \varphi^{**}(u) \geqq \varphi^*(y)\}, \tag{6.18}$$

where

$$\langle u, y\rangle - \varphi^{**}(u) \leqq \varphi^*(y) \quad \text{for all } u, y, \tag{6.19}$$

by the definition of φ^{**}. Since $(\varphi^{**})^* = (\varphi^*)^{**} = \varphi^*$ (Theorem 5, φ^* being closed), we also have

$$\partial\varphi^{**}(u) = \{u|\langle u, y\rangle - \varphi^*(y) \geqq \varphi^{**}(u)\}. \tag{6.20}$$

These relations and the inequality $\varphi^{**}(u) \leqq \varphi(u)$ yield the following theorem.

THEOREM 12. *If for a particular $u \in U$ one has $\partial\varphi(u) \neq \emptyset$, then $\varphi(u) = \varphi^{**}(u)$. Moreover, if $\varphi(u) = \varphi^{**}(u)$, then $\partial\varphi(u) = \partial\varphi^{**}(u)$* and

$$\partial\varphi(u) = \{y|u \in \partial\varphi^*(y)\}. \tag{6.21}$$

COROLLARY 12A. *If $\varphi = \varphi^{**}$, then the multifunction $\partial\varphi : u \to \partial\varphi(u)$ is the inverse of the multifunction $\partial\varphi^* : y \to \partial\varphi^*(y)$, in the sense that*

$$y \in \partial\varphi(u) \quad \textit{if and only if} \quad u \in \partial\varphi^*(y). \tag{6.22}$$

Theorem 5 will be of greatest interest to us in the case where $u = 0$. Then

$$\varphi^{**}(0) = \sup_{y \in Y}\{\langle 0, y\rangle - \varphi^*(y)\} = -\inf_{y \in Y}\varphi^*(y), \tag{6.23}$$

while by definition (6.12) applied to φ^*,

$$0 \in \partial\varphi^*(y) \Leftrightarrow \varphi^*(y') \geqq \varphi^*(y) \quad \text{for all} \quad y' \in Y. \tag{5.24}$$

Thus, we have the following corollary.

COROLLARY 12B. *If $\varphi(0) = \varphi^{**}(0)$, as is true in particular if $\partial\varphi(0) \neq \emptyset$, then*

$$\partial\varphi(0) = \{y|\varphi^* \text{ attains its (global) min at } y\}. \tag{6.25}$$

Next we describe refinements of (6.25) showing that the actual differentiability of φ at 0 is closely connected with the convergence of *minimizing sequences* for φ^*, that is, sequences y_k in Y, $k = 1, 2, \cdots$, with

$$\lim_{k \to \infty} \varphi^*(y_k) = \inf \varphi^*. \tag{6.26}$$

Example. Let us determine the subgradients of the indicator ψ_C of a nonempty convex set $C \subset U$. By definition, the relation $y \in \partial\psi_C(u)$ means that

$$\psi_C(u') \geqq \psi_C(u) + \langle u' - u, y\rangle \quad \text{for all} \quad u' \in C, \tag{6.27}$$

or in other words that

$$u \in C \quad \text{and} \quad \langle u' - u, y\rangle \leqq 0 \quad \text{for all } u' \in C. \tag{6.28}$$

The elements y satisfying this are said to be *normal* (in the sense of convex analysis) to C at u, and they form a closed convex cone. Thus $\partial\psi_C(u)$ is the *normal cone* to C at u. (This cone is taken to be the empty set if $u \notin C$, whereas it degenerates to $\{0\}$ if $u \in \text{int } C$.)

We can use this fact and Corollary 12A to characterize also the subgradients of the support function

$$\psi_D^*(u) = \sup_{y \in D} \langle u, y \rangle, \qquad u \in U, \tag{6.29}$$

when D is a nonempty closed convex subset of Y. Since $\psi_D^{**} = \psi_D$, we have

$$y \in \partial\psi_D^*(u) \quad \text{if and only if} \quad u \in \partial\psi_D(y). \tag{6.30}$$

Therefore $\partial\psi_D^*(u)$ consists of the points $y \in D$ (if any) such that u is normal to D at y. An important special case is ψ_D^* a norm on U, D being the unit ball for the dual norm.

Recall that in the finite-dimensional case the function φ is said to be *differentiable* at u if $\varphi(u)$ is finite and not only does the gradient $\nabla\varphi(u)$ exist, i.e., a vector y such that

$$\langle u', y \rangle = \lim_{\lambda \downarrow 0} \frac{\varphi(u + \lambda u') - \varphi(u)}{\lambda} = \varphi'(u; u') \quad \text{for all } u', \tag{6.31}$$

but also: the difference quotients in (6.31), as functions of u', converge *uniformly* on every *bounded* set.

In the infinite-dimensional case, this property is called *strong* (or *Fréchet*) *differentiability*. However, there are other concepts of differentiability which are sometimes more relevant. These correspond to substituting other classes of sets for the bounded sets in the uniformity requirement. In particular, we shall simply say that φ is *differentiable* (*in relation to the designated topology on* Y!) if the convergence of the difference quotients is uniform on all *equicontinuous* sets of U. This reduces then to the standard concept in finite dimensions.

THEOREM 13. *Let φ be convex and, on a neighborhood of* 0, *lower-semicontinuous.*

(a) *φ is differentiable at* 0 *with $\nabla\varphi(0) = \bar{y}$, if and only if $\bar{y}$ has the property that every minimizing sequence for φ^* converges to $\bar{y}$.*

(b) *The gradient $\nabla\varphi(0)$ exists and equals $\bar{y}$, if and only if $\bar{y}$ has the property that every minimizing sequence for φ^* converges to $\bar{y}$ in the weak topology induced on Y by U.*

(c) *The properties in* (a) (*and* (b)) *imply that $\bar{y}$ is the unique element of $\partial\varphi(0)$, and in the finite-dimensional case of $U = R^n = Y$ they are actually equivalent to the latter. This is true even without the hypothesis of lower-semicontinuity.*

Proof. See [1], [13].

A useful criterion for differentiability of φ at a point u is the following: φ is convex, and there exists a function φ_0 (not necessarily convex) such that $\varphi_0(u') \geq \varphi(u')$ for all u', $\varphi_0(u) = \varphi(u)$, and φ_0 is differentiable at u; then $\nabla\varphi(u) = \nabla\varphi_0(u)$. (This can be argued from the definition of differentiability and the fact that $\varphi_0'(u; u') \geq \varphi'(u; u')$ for all u', which by convexity implies equality for all u' and hence $\nabla\varphi_0(u) \in \partial\varphi(u)$.)

Convergence conditions of the type in Theorem 13 are known as *rotundity* properties of φ^* at the point $\bar{y}$. They are closely related to strict convexity, which likewise plays a role dual to differentiability but in a more global sense. In the finite-dimensional case, for example, if φ is a finite convex function such that φ^* is also finite everywhere, the global differentiability of φ is equivalent to the strict convexity of φ^*. There are various generalizations of this to infinite-dimensional spaces or functions that are not finite everywhere [1, § 5], [13, § 26]. These are involved in the study of the *Legendre* transformation mentioned in § 3.

Incidentally, the continuity properties of the gradient function $\nabla\varphi : u \to \nabla\varphi(u)$ are greatly simplified when φ is convex [1, § 4], [13, § 25]. In fact, $\nabla\varphi$ is always continuous as a function from the set $\{u | \nabla\varphi(u) \text{ exists}\}$ to the space Y endowed with the weak topology induced by U (which is just the ordinary topology on R^n in the finite-dimensional case). If $\nabla\varphi$ is defined on a neighborhood of u, it is continuous at u (with respect to the designated topology on Y) if and only if φ is differentiable at u (in relation to the same topology).

The subgradient multifunction

$$\partial\varphi : u \to \partial\varphi(u) \subset Y \tag{6.32}$$

has many interesting properties. The following facts will not be needed later in these notes, but we mention them anyway.

A multifunction $T: U \to Y$ is called a *monotone* operator if

$$y_i \in T(u_i) \quad \text{for} \quad i = 1, 2 \quad \text{implies} \quad \langle u_1 - u_2, y_1 - y_2 \rangle \geqq 0. \tag{6.33}$$

It is a *cyclically monotone* operator if

$$\begin{gathered} y_i \in T(u_i) \quad \text{for} \quad i = 1, \cdots, m \quad \text{implies} \\ \langle u_2 - u_1, y_1 \rangle + \langle u_3 - u_2, y_2 \rangle + \cdots + \langle u_1 - u_m, y_m \rangle \leqq 0. \end{gathered} \tag{6.34}$$

It is a *maximal* monotone operator if it is monotone and its graph

$$G(T) = \{(u, y) | y \in T(u)\} \tag{6.35}$$

is not properly included in the graph of any other monotone operator; similarly *maximal* cyclically monotone operator.

THEOREM 14. *Suppose U is a Banach space (in the designated "compatible" topology). In order that a multifunction $T: U \to Y$ be of the form $T = \partial\varphi$ for some closed proper convex function φ, it is necessary and sufficient that T be a maximal cyclically monotone operator. Then φ is unique up to an additive constant, and T is also a maximal monotone operator.*

If in addition $U = Y = H$, where H is a Hilbert space and $\langle u, y \rangle$ is the inner product in H, the mapping $(u, y) \to u + y$ is a homeomorphism of $G(\partial\varphi)$ onto H.

Proof. See [15], [13, § 24], [10].

The last assertion says that the graph of $\partial\varphi$ for φ closed, proper, convex is geometrically very much like the graph of a continuous mapping; for example, in the case of $U = Y = R^n$ it is an n-dimensional set without "gaps" or "edges."

The theorem also characterizes a class of maximal monotone operators T such that solving the "equation" $0 \in T(u)$ is equivalent to minimizing a convex function φ (since φ attains its minimum at u if and only if $0 \in \partial\varphi(u)$). The study of such "equations" has assumed some importance in recent years in connection with partial differential equations, and it is easy to see why. In Example 10, for instance, if the precise formulation is effected in such a way that the convex function f is closed and proper on a Banach space, then ∂f is a maximal monotone operator. The elements x minimizing f can be interpreted as the solutions to some generalized partial differential equation, as already remarked. For more on monotone operators and partial differential equations, see [2].

Theorem 14 also has some significance in axiomatics, for example, in mathematical economics, where one wants to deduce that certain correspondences can be interpreted as resulting from an "optimality principle."

7. Solutions to optimization problems. Returning to the notation of § 4 (with $f, F, \varphi, K, g, G, \gamma$), we employ the theory of continuity and subgradients of convex functions to derive the existence of solutions to problems (P) and (D) and to characterize them in terms of directional derivatives of the optimal value functions φ and γ.

The optimal values in (P) and (D) are denoted by inf (P) (or inf f) and sup (D) (or sup g). We say that $\bar{x}$ *solves* (P) if $\bar{x}$ minimizes f (globally) on X. Note that by the definition of ∂f:

$$\bar{x} \textit{ solves } (\mathrm{P}) \textit{ if and only if } \quad 0 \in \partial f(\bar{x}). \tag{7.1}$$

If f is not identically $+\infty$, this entails $\bar{x} \in \operatorname{dom} f$. The elements of dom f are called the *feasible solutions* to (P), since in applications they satisfy the explicit or implicit constraints that are present (cf. Example 1 and others). The situation where f is identically $+\infty$ is interpreted as meaning that the constraints in (P) are inconsistent. In terms of optimal values, this is signalled by the equation $\inf(\mathrm{P}) = +\infty$. Certainly there is no interest in minimizing the constant function $+\infty$, so it may seem odd that in this case the terminology just adopted has us say that every point of X solves (P), when such points are not even feasible solutions! However, this is a matter of technical convenience in the statement of theorems. In cases of expository hardship, we can say that an element $\bar{x}$ solves (P) *properly* when we want to indicate that $\bar{x}$ is also feasible, i.e., that $\inf(\mathrm{P}) < +\infty$. It is not desirable to eliminate the case $\inf(\mathrm{P}) = +\infty$ in some a priori fashion, since the more general results, in placing conditions on inf (P), provide us in fact with important information on whether the constraints in (P) are consistent or inconsistent.

We say similarly that $\bar{y}$ *solves* (D) if $\bar{y}$ maximizes the concave function g over Y. This is equivalent to $0 \in \partial(-g)(\bar{y})$. However, for notational elegance and symmetry we prefer to introduce for concave functions the concepts parallel to those for convex functions. Thus for instance we define $u \in \partial g(y)$ to mean that

$$g(y') \leqq g(y) + \langle u, y' - y \rangle \quad \text{for all } y'.$$

It will be clear from the context when this definition is intended, rather than the previous one (for example, the concave functions g, G, γ of $K(x, \cdot)$ are involved). One can speak of "subgradients in the concave sense," if necessary. The results in § 6 all carry over with only the obvious changes of sign and directions of inequalities. At any rate, we are enabled to say that

(7.2) $\bar{y}$ *solves* (D) *if and only if* $0 \in \partial g(\bar{y})$.

Subgradient notation is also useful in connection with the Lagrangian function K. We define $(v, u) \in \partial K(x, y)$ to mean that $v \in \partial_x K(x, y)$ and $u \in \partial_y K(x, y)$, that is,

$$K(x', y) \geqq K(x, y) + \langle x' - x, v \rangle \quad \text{for all} \quad x' \in X,$$

$$K(x, y') \leqq K(x, y) + \langle u, y' - y \rangle \quad \text{for all} \quad y' \in Y.$$

Thus

(7.3) $(0, 0) \in \partial K(\bar{x}, \bar{y})$ *if and only if* $(\bar{x}, \bar{y})$ *is a saddle-point of* K.

Of course, the subgradient conditions (7.1), (7.2) and (7.3) have little content at this stage and appear as trivial consequences of the definitions. What will give them the content are the results to be discussed in § 9 on how to calculate the subgradients of particular functions using various structural properties. After all, even the classical result of the calculus that the directive of a function must vanish, as a necessary condition for a minimum or maximum, would be largely "tautological," without all the rules such as for calculating the derivative of a sum of two functions.

In anticipation of the later developments, we dub the relation $(0, 0) \in \partial K(\bar{x}, \bar{y})$ the *(abstract) Kuhn–Tucker condition* for (P) (associated with the particular choice of F in the representation (4.1)). While the saddle-point criterion for optimality has a game-theoretic quality, the Kuhn–Tucker condition involves the "vanishing" of a certain multifunction which in specific cases can hopefully be described in considerable detail. We shall see that in Example 1 the abstract Kuhn–Tucker condition does yield the classical conditions of Kuhn and Tucker. In Example 8 it yields variants of the Euler–Lagrange condition and transversality condition of the calculus of variations, and so forth.

The initial theorem of this section summarizes in the new terminology the information already gained in Theorems 2, 7 and 7′ about the possible equality of the optimal values in (P) and (D).

THEOREM 15. *The implications*

$$\text{(a)} \Leftrightarrow \text{(b)} \Rightarrow \text{(c)} \Rightarrow \text{(d)}$$

hold among the following conditions:

(a) $\inf(\text{P}) = \sup(\text{D})$;

(b) $\varphi(0) = \operatorname{cl} \operatorname{co} \varphi(0)$;

(c) *the saddle-value of the Lagrangian* K *exists*;

(d) $\gamma(0) = \operatorname{cl} \gamma(0)$.

Assuming $F(x, u)$ *is closed convex in* u, *one has the equivalence of* (a), (b) *and* (c). *Assuming* $F(x, u)$ *is closed convex in* (x, u), *one has the equivalence of* (a), (b), (c) *and* (d).

Furthermore, the implication (e) $\Rightarrow$ (f) *holds for the following conditions, with actual equivalence if* $F(x, u)$ *is closed convex in* u:

(e) $\bar{x}$ *solves* (P), $\bar{y}$ *solves* (D), *and* $\inf(\mathrm{P}) = \sup(\mathrm{D})$:

(f) *the pair* $(\bar{x}, \bar{y})$ *satisfies the Kuhn–Tucker condition.*

Proof. Let $\bar{F}$ be the function on $X \times U$ obtained by taking the closure of the convex hull of the function $u \to F(u, x)$ for each x. Consider the problems: $(\bar{\mathrm{P}})$ minimize $\bar{f}(x) = \bar{F}(x, 0)$ over all $x \in X$, $(\bar{\bar{\mathrm{P}}})$ minimize $\bar{\bar{f}}(x) = F^{**}(x, 0)$ over all $x \in X$, in terms of the duality corresponding to the indicated parametric representations. It is evident that $(\bar{\mathrm{P}})$ and $(\bar{\bar{\mathrm{P}}})$ yield the same (D), g, G and γ as (P), and furthermore $(\bar{\mathrm{P}})$ yields the same Lagrangian K as (P). Using the inequality $F \geqq \bar{F} \geqq \bar{\bar{F}}$, we obtain from Theorems 7 and 7′ (applied also to $\bar{F}$ and $\bar{\bar{F}}$ in place of F) the relations

$$\begin{aligned} \operatorname{cl} \operatorname{co} \varphi(0) &= \sup(\mathrm{D}) = \sup g(y) = \gamma(0) \\ &\leqq \operatorname{cl} \gamma(0) = \inf(\bar{\bar{\mathrm{P}}}) \leqq \inf(\bar{\mathrm{P}}) = \inf_x \bar{f}(x) \\ &\leqq \inf_x f(x) = \inf(\mathrm{P}) = \varphi(0). \end{aligned}$$

On the other hand, we have

$$f(x) \geqq \bar{f}(x) = \sup K(x, y) \quad \text{and} \quad g(y) = \inf_x K(x, y)$$

from Theorem 6 (applied to $\bar{F}$) and the definition of g. The desired implications are obvious from this and the definition of "saddle-value" and "saddle-point."

COROLLARY 15A. *Suppose it is known that* $\inf(\mathrm{P}) = \sup(\mathrm{D})$, *where the supremum is attained. Then* $\bar{x}$ *solves* (P) *if and only if there exists* $\bar{y}$ *such that* $(\bar{x}, \bar{y})$ *satisfies the Kuhn–Tucker condition.*

The next theorem provides the basis for applying the theory of directional derivatives of convex functions to the study of solutions to (D). Condition (c) in the theorem, as will be recalled from § 1, is fundamental to dual methods of solving (P) using K.

THEOREM 16. *The following conditions on an element* $\bar{y}$ *of* Y *are equivalent*:

(a) $\bar{y}$ *solves* (D), *and* $\sup(\mathrm{D}) = \inf(\mathrm{P})$;

(b) $-\bar{y} \in \partial\varphi(0)$;

(c) $\inf_{x \in X} K(x, \bar{y}) = \inf_{x \in X} f(x)$.

Proof. Recalling from Theorem 7 that $-g(y) = \varphi^*(-y)$ and $\sup(\mathrm{D}) = -\varphi^{**}(0)$, we can rewrite (a) as

$$0 \in \partial\varphi^*(-y) \quad \text{and} \quad \varphi^{**}(0) = \varphi(0).$$

The equivalence of (a) and (b) is then asserted by Corollary 12B. The equivalence of (a) and (c) is clear from rewriting (c) as $g(\bar{y}) = \inf(\mathrm{P})$ (cf. (4.6)), since $g(y) \leqq \inf(\mathrm{P})$ for all y.

COROLLARY 16A. *Assuming the quantity* $\varphi(0) = \inf(\mathrm{P})$ *is finite and* F *is convex (so that* φ *is convex), we have the equivalence of*:

(a) $\inf(\mathrm{P}) = \sup(\mathrm{D})$, *and there exists at least one* $\bar{y}$ *solving* (D);

(b) $\liminf_{u' \to u} \varphi'(0; u)$ *is finite for at least one* $u \in U$.

Proof. This is immediate from Theorem 11 (b) (and Theorem 4, applied to the convex function $\theta(u) = \varphi'(0; u)$).

By virtue of the equivalence of (a) and (b) in Theorem 16, we obtain in terms of the directional derivatives $\varphi'(0; u)$ an even fuller description of the vectors $\bar{y}$ solving (D). We shall not try to list all the details, but limit ourselves to the choicest case, where φ is actually convex and continuous at 0.

THEOREM 17. *Suppose F is convex, and φ (also convex under this assumption) is bounded above on a neighborhood of* 0. *Then*

(a) $\inf(\mathrm{P}) = \sup(\mathrm{D})$, *and there exists at least one $\bar{y}$ solving* (D).

(b) *In fact, for every real β the set $\{y | g(y) \geqq \beta\}$ is closed, bounded and convex, actually equicontinuous and hence weakly compact (i.e., in the weak topology induced on Y by U). Thus every maximizing sequence for* (D) *has weak cluster points, and every such cluster point $\bar{y}$ solves* (D).

(c) *Assuming of course that the optimal values in* (a) *are not* $-\infty$ (*so that* $\varphi(0)$ *is finite*), *we have*

$$\varphi'(0; u) = \max\{\langle u, -\bar{y}\rangle | \bar{y} \text{ solves (D)}\} \text{ for all } u \in U. \tag{7.4}$$

(d) *A vector $\bar{y}$ solves* (D) *uniquely if and only if* $\bar{y} = -\nabla\varphi(0')$, *i.e.*,

$$\varphi'(0; u) = -\langle u, \bar{y}\rangle \quad \text{for all} \quad u \in U, \tag{7.5}$$

and in this event every maximizing sequence for (D) *converges weakly to* $\bar{y}$.

(e) *The maximizing sequences for* (D) *all actually converge in the designated topology on Y, if and only if φ is differentiable at* 0 (*in relation to that topology*).

Proof. We know from Theorem 8 that φ is continuous at 0, with $\varphi(0) < +\infty$. If $\varphi(0) = -\infty$, then $\partial\varphi(0) = Y$, $\varphi^* \equiv +\infty$, and since $g(y) = -\varphi^*(-y)$ (Theorem 7) everything is trivial (Theorem 16). Therefore assume $\varphi(0)$ is finite. Then (a) and (c) are asserted by Theorem 11 (a), in conjunction with Theorem 16. Theorem 10 (b) gives us (b), while Theorem 13 gives (d) and (e).

An interesting generalization of Theorem 17 (c) to the nonconvex case may be found in [4, § 7].

To complete the chain of ideas, we now furnish some convenient criteria for the hypothesis of Theorem 17 to be satisfied. Note that these criteria, used with Corollary 15A (cf. (a) of Theorem 17), are capable of yielding a vast array of "Kuhn–Tucker theorems" which characterize the solutions to various convex optimization problems. Criteria (c) and (e) have not appeared in any general form previously in the literature.

THEOREM 18. *Assuming F is convex, each of the following conditions is sufficient for the convex function φ to be bounded above on a neighborhood of* 0 *and hence continuous at* 0.

(a) *There is an $x \in X$ such that the function $u \to F(u, x)$ is bounded above on a neighborhood of* 0. (*Or more generally, the function $u \to F(u, \theta(u))$ is bounded above on a neighborhood of* 0 *for some mapping* θ.)

(b) $U = R^n = Y$, *and* $0 \in \operatorname{core} \operatorname{dom} \varphi$.

(c) *U and V are both Banach spaces* (*in the designated "compatible" topologies!*), *F is closed and* $0 \in \operatorname{core} \operatorname{dom} \varphi$.

(d) $U = R^n = Y$, *and at least one of the level sets* $\{y|g(y) \geqq \beta\}$ *is nonempty and bounded.*

(e) F *is closed, and there exist a neighborhood* N *of* 0 *in* V *and a number* β *such that the set*

$$\{y \in Y | v \in N, G(y, v) > \beta\} \tag{7.6}$$

is nonempty and equicontinuous.

Proof. The sufficiency of (a) is obvious from the definition (4.7) of φ, while that of (b) and (d) follows from Corollary 8A and Theorem 10 (c) (using $-g(y) = \varphi^*(-y)$). In verifying the sufficiency of (c) and (e), we investigate for an arbitrary convex neighborhood N of 0 in V the nature of the function

$$g_N(y) = \sup\{G(y, v) | v \in N\}. \tag{7.7}$$

Of course g_N is concave, since G is concave: apply Theorem 1 to the convex function

$$h(y, v) = -G(y, v) + \psi_N(v),$$

using the fact that

$$-g_N(y) = \inf\{h(y, v) | v \in V\}.$$

Furthermore,

$$\begin{aligned} g_N^*(u) &= \inf\{\langle u, y\rangle - g_N(y) | y \in Y\} \\ &= -\sup\{G(y, v) - \langle u, y\rangle | y \in Y, v \in N\}. \end{aligned} \tag{7.8}$$

We claim next that

$$-g_N^*(u) \geqq \varphi(u) \quad \text{if} \quad \sup(\text{D}) > -\infty \quad \text{or if} \quad u \in \operatorname{dom}\varphi. \tag{7.9}$$

To prove this, fix any $u_0 \in U$ and consider the problem (P^0) of minimizing f^0 over X, where

$$f^0(x) = F^0(x, 0), \quad F^0(x, u) = F(x, u_0 + u). \tag{7.10}$$

Here F^0 is again convex and closed, and the duality theory of § 4 may be invoked. The corresponding dual problem (D_0) has (by direct calculation from the definitions)

$$g^0(y) = G^0(y, 0), \quad G^0(y, v) = G(y, v) - \langle u_0, y\rangle, \tag{7.11}$$

and consequently

$$\gamma^0(v) = \sup\{G(y, v) - \langle u_0, y\rangle | y \in Y\}. \tag{7.12}$$

Inasmuch as $\varphi^0(u) = \varphi(u_0 + u)$, we know from Corollary 7'A that the two relations

$$\gamma^0(0) = \liminf_{u \to u_0} \varphi(u), \tag{7.13}$$

$$\varphi(u_0) = \limsup_{v \to 0} \gamma^0(v) \leqq \sup_{v \in N} \gamma^0(v) = -g_N^*(u_0) \tag{7.14}$$

hold, if $\varphi(u_0) \neq +\infty$ (as implied by $u_0 \in \operatorname{dom} \varphi$) or if $\gamma^0(0) > -\infty$ (as implied by $\sup(\mathrm{D}) > -\infty$, i.e., $G(y, 0) > -\infty$ for at least one y). Thus (7.9) is true as claimed. *Notice from* (7.14) *that if u_0 is a point with $\varphi(u_0)$ finite, then the supremum $-g_N^*(u_0)$* must also be finite, at least if N is chosen sufficiently small.

We are now in a position to establish (e). Under this condition, it is true in particular that the level set $\{y | g(y) \geqq \bar{\beta}\}$ is nonempty and bounded, where we choose $\bar{\beta}$ to be any number satisfying $\beta < \bar{\beta} < \sup(\mathrm{D})$. Since g is a closed concave function we may conclude (Theorems 9 and 7) that

$$+\infty > \sup(\mathrm{D}) = \operatorname{cl} \varphi(0) > -\infty.$$

Hence there must exist at least one u_0 with $\varphi(u_0)$ finite, so that, as just observed, we can assume $-g_N^*$ is finite somewhere. Then g_N^* is a closed *proper* concave function, implying $g_N^{**} = \operatorname{usc} g_N$. The level set $\{y | g_N(y) \geqq \bar{\beta}\}$ is equicontinuous by hypothesis, with

$$\beta < \sup_{y \in Y} g_N(y) = \sup_{y \in Y} \operatorname{usc} g_N(y), \tag{7.15}$$

and the latter inequality gives us

$$\{y | \operatorname{usc} g_N(y) \geqq \bar{\beta}\} = \operatorname{cl} \{y | g_N(y) \geqq \bar{\beta}\}. \tag{7.16}$$

Therefore g^{**} has a nonempty, equicontinuous level set. It follows from Theorem 10(c) that the convex function $-(g_N^{**})^* = -g_N^*$ is bounded above in some neighborhood of 0. Then φ is bounded above in the same neighborhood due to (7.9), and this is what we set out to prove.

Finally, we attack (c). We can assume $N = \{v | \, \|v\| \leqq \varepsilon\}$ for some $\varepsilon > 0$. Then from (7.8) and the inequality

$$G(y, v) - \langle u, y \rangle \leqq F(x, u) - \langle x, v \rangle, \tag{7.17}$$

which is a consequence of (4.17), we have

$$\begin{aligned} -g_N^*(u) &\leqq \sup_{\|v\| \leqq \varepsilon} \{F(x, u) - \langle x, v \rangle\} \\ &= F(x, u) + \|x\| \quad \text{for all} \quad x \in X. \end{aligned}$$

Thus

$$-g_N^*(u) \leqq \inf_{x \in X} \{F(x, u) + \|x\|\} < +\infty \quad \text{for all} \quad u \in \operatorname{dom} \varphi. \tag{7.18}$$

But $-g_N^*$ is a closed convex function on a Banach space U, so $-g_N^*$ is continuous on core dom$(-g_N^*)$ by virtue of Corollary 8B. Hence $-g_N^*$ is continuous on core dom φ by (7.18). Since (7.9) holds, we conclude that φ is bounded above on some neighborhood of 0 if $0 \in$ core dom φ.

An important advantage of the symmetry inherent in conjugate duality is that dual versions of the preceding results can be added to the repertory with no extra effort.

THEOREM 16′. *If F is convex and closed, the following conditions on an element $\bar{x}$ of X are equivalent*:

(a) $\bar{x}$ *solves* (P), *and* $\inf(\mathrm{P}) = \sup(\mathrm{D})$;

(b) $-\bar{x} \in \partial\gamma(0)$;

(c) $\sup_{y \in Y} K(\bar{x}, y) = \sup_{y \in Y} g(y)$.

COROLLARY 16′A. *Assuming the quantity $\gamma(0) = \sup(\mathrm{D})$ is finite and F is convex and closed, we have the equivalence of*:

(a) $\inf(\mathrm{P}) = \sup(\mathrm{D})$, *and there exists at least one $\bar{x}$ solving* (P);

(b) $\limsup_{v' \to v} \gamma'(0; v')$ *is finite for at least one $v \in V$.*

THEOREM 17′. *Suppose F is convex and closed and γ is bounded below on some neighborhood of* 0. *Then*

(a) $\inf(\mathrm{P}) = \sup(\mathrm{D})$, *and there exists at least one $\bar{x}$ solving* (P).

(b) *In fact, for every real α the set $\{x \mid f(x) \leqq \alpha\}$ is closed, bounded and convex, actually equicontinuous and hence weakly compact (i.e., in the weak topology induced on X by V). Thus every minimizing sequence for* (P) *has weak cluster points, and every such cluster point $\bar{x}$ solves* (P).

(c) *Assuming of course that the optimal values in* (a) *are not $+\infty$ (so that $\gamma(0)$ is finite), we have*

$$\gamma'(0; v) = \min\{\langle -\bar{x}, v\rangle \mid \bar{x} \text{ solves (P)}\} \quad \textit{for all} \quad v \in V. \tag{7.19}$$

(d) *A vector $\bar{x}$ solves* (P) *uniquely if and only if $\bar{x} = -\nabla\gamma(0)$, i.e.,*

$$\gamma'(0; v) = -\langle \bar{x}, v\rangle \quad \textit{for all} \quad v \in V, \tag{7.20}$$

and in this event every minimizing sequence for (P) *converges weakly to $\bar{x}$.*

(e) *The minimizing sequences in* (d) *all converge in the designated topology on X, if and only if γ is differentiable at* 0 *in relation to that topology.*

THEOREM 18′. *Assuming F is convex and closed, each of the following conditions is sufficient for the concave function γ to be bounded below on a neighborhood of* 0 *and hence continuous at* 0.

(a) *There is a $y \in Y$ such that the function $v \to G(y, v)$ is bounded below on a neighborhood of* 0.

(b) $X = R^m = V$, *and* $0 \in \operatorname{core} \operatorname{dom} \gamma$.

(c) *U and V are both Banach spaces (in the designated "compatible" topologies), and* $0 \in \operatorname{core} \operatorname{dom} \gamma$.

(d) $X = R^m = V$, *and at least one of the level sets $\{x \mid f(x) \leqq \alpha\}$ is nonempty and bounded.*

(e) *There exist a neighborhood N of* 0 *in U and a number α such that the set*

$$\{x \in X \mid \exists u \in N, F(x, u) < \alpha\} \tag{7.21}$$

is nonempty and equicontinuous.

Since the choice of the compatible topology on V is at our discretion in most applications, we usually want to choose it as strong as possible, so as to increase the chances of γ being continuous at 0. It happens that there is always a unique strongest topology meeting our compatibility requirement. This is the *Mackey* *topology* on V induced by the pairing with X, which can be defined as the coarsest topology on V such that all the support functions ψ_C^* of weakly compact convex sets C in X are continuous [8]. It has by definition the property that every convex set whose closure is weakly compact is equicontinuous, and this fact can be used in Theorem 18′.

COROLLARY 18′A. *Suppose that F is convex and closed, and that condition* (e) *of Theorem* 18′ *holds, but with equicontinuity replaced by the assumption that the closure of the set in question is weakly compact. Then γ is bounded around* 0 *relative to the Mackey topology on V, and hence the conclusions of Theorem* 17′ *are valid if interpreted in that topology.*

Proof. The only question in need of attention is whether the set (7.21) might also have to be assumed convex, inasmuch as the assertion preceding the corollary referred just to convex sets. However, the assumption can be omitted, because if the property holds for a given N it also holds for an arbitrary smaller N which can be taken to be convex. When N it convex, the set (7.21) is convex, because it is the image of the convex set

$$\{(x, u) \mid u \in N, F(x, u) \leqq \alpha\}$$

under the projection $(x, u) \to x$.

The Mackey topology induced on U by Y can be used similarly in Theorems 17 and 18.

8. Some applications.

Example 1″. (Convex programming.) We continue with Examples 1 of § 2 and 1′ of § 5 under the assumption that the convex set C is closed ($\neq \emptyset$) and the convex functions f_i are l.s.c. Then F is closed, convex and proper. Since

$$\varphi(u) = \inf\{f_0(x) \mid x \in C, f_i(x) \leqq u_i \text{ for } i = 1, \cdots, m\}, \tag{8.1}$$

it is clear that

$$\operatorname{dom} \varphi = \{u \in R^m \mid \exists x \in C, f_i(x) \leqq u_i \text{ for } i = 1, \cdots, m\}. \tag{8.2}$$

The sufficient condition (a) of Theorem 18 is equivalent to the so-called *Slater condition*:

$$\text{there exists} \quad x \in C \quad \text{with} \quad f_i(x) < 0, \qquad i = 1, \cdots, m. \tag{8.3}$$

Under the Slater condition, therefore, all the properties of Theorem 17 are present.

Then, for example, in the case of a unique dual solution $\bar{y}$ we have an interpretation of the Lagrange multipliers corresponding to a saddle-point of K as derivatives:

$$\bar{y}_i = -\left.\frac{\partial \varphi}{\partial u_i}\right|_{u=0}, \qquad i = 1, \cdots, m. \tag{8.4}$$

The subdifferential form of the Kuhn–Tucker condition will be discussed in § 10.

It is easy to see that the Slater condition is *equivalent* to having $0 \in \operatorname{core} \operatorname{dom} \varphi$.

For the properties in Theorem 17′ all to hold, we have from Corollary 18′A the sufficient condition that for some $\varepsilon > 0$ and $\alpha > \inf(\mathrm{P})$, the set

$$\{x \in C \mid f_0(x) \leqq \alpha, f_1(x) \leqq \varepsilon, \cdots, f_m(x) \leqq \varepsilon\}, \tag{8.5}$$

is equicontinuous in X. If $X = R^n$, this can according to Theorem 18′ (d) be weakened to the condition that for some real number α the set

$$\{x \in C \mid f_0(x) \leqq \alpha, f_1(x) \leqq 0, \cdots, f_m(x) \leqq 0\} \tag{8.6}$$

is nonempty and bounded.

Corollary 16A leads to a very complete characterization of the existence of an optimal multiplier vector $\bar{y}$ in the case where $\inf(\mathrm{P})$ is finite. One uses the fact that a convex function on R^m whose l.s.c. hull has the value $-\infty$ somewhere must itself have the value $-\infty$ somewhere [13, § 7]. The existence thus fails if and only if there is a vector u yielding the directional derivative $\varphi'(0; u) = -\infty$.

Example 3″. (Nonconvex programming.) With the functions f_i not necessarily convex, as in Example 3′, let us assume that $\varphi(u)$ (given by (8.1)) is actually twice-continuously-differentiable on a neighborhood of $u = 0$ in R^m, as well as globally bounded below on R^m. While these conditions would be virtually impossible to arrange in advance, they turn out to be satisfied very commonly in applications: namely, where $X = R^n$, (P) has a unique solution at a point $\bar{x} \in \operatorname{int} C$, the functions f_i are twice-continuously-differentiable, and certain second-order *sufficient* conditions for a minimum (involving a Lagrange multiplier vector $\bar{y}$) are satisfied. (To get the boundedness, C could be replaced by a bounded subset, if necessary.) At any rate, under these assumptions, if r is sufficiently large the function

$$\varphi_r(u) = \varphi(u) + r|u|^2 \tag{8.7}$$

will be strictly *convex* on a neighborhood of $u = 0$, since the matrix of second partial derivatives will be positive-definite. If r is chosen still larger, it can be seen (and here is where the boundedness assumption on φ comes in) that the nonconvex portions of φ will eventually be "pushed out of the way" to the extent that

$$\operatorname{cl} \operatorname{co} \varphi_r(u) = \varphi_r(u) \quad \text{for all } u \text{ near } 0. \tag{8.8}$$

Thus for r sufficiently large we shall have

$$\inf(\mathrm{P}) = \sup(\mathrm{D}_r),$$

(D_r) being the corresponding dual problem expressible through the general formula (4.6) in terms of the Lagrangian K_r in (5.8). Moreover, (D_r) has a unique optimal solution:

$$\bar{y} = -\nabla\varphi_r(0) = -\nabla\varphi(0). \tag{8.9}$$

The hypothesis of Theorem 16 is satisfied in particular; thus an element $\bar{x}$ solves (P) if and only if $(\bar{x}, \bar{y})$ satisfies the abstract Kuhn–Tucker condition, i.e., is a *global* saddle-point of K_r. The latter implies the usual Kuhn–Tucker conditions (see § 10), assuming the functions f_i are suitably differentiable. Thus $\bar{y}$ turns out to be the optimal Lagrange multiplier vector in the ordinary sense.

The potential existence of a global saddle-point $(\bar{x}, \bar{y})$ of K_r can be exploited computationally by algorithms based on the dual approach. Such algorithms resemble penalty methods, except that the "penalty parameter" r need not tend to $+\infty$. More refined criteria for the existence of a saddle-point are also known. However, we must forego the details here; see [23].

Example 4″. (Convex programming with generalized constraints.) The duality results may be applied to the model of Example 4′ of § 4, where the optimal value function is

$$\varphi(u) = \inf\{f_0(x) | x \in C, \Phi(x) - u \in Q\}. \tag{8.10}$$

Much as in the convex programming case above, the effective domain of φ consists of the values of the parameter vector u such that the perturbed constraint system is consistent:

$$\operatorname{dom} \varphi = \{u \in U | x \in C, \Phi(x) - u \in Q\}.$$

Assuming as before the convexity of C, f_0 and the "epigraph"

$$\operatorname{epi} \Phi = \{(x, u) | \Phi(x) - u \in Q\}, \tag{8.11}$$

we have the convexity of the representation function F and consequently the convexity of φ.

It is obvious that condition (a) of Theorem 18, sufficient for the strongest duality results, is satisfied if:

$$\text{there exists} \quad x \in C \quad \text{with} \quad \Phi(x) \in \operatorname{int} Q. \tag{8.12}$$

This is the natural generalization of the Slater condition. In particular (8.12) implies that $\inf(\mathrm{P}) = \sup(\mathrm{D})$, with the supremum attained. In other words, under assumption (8.12) there is a strong form of the "Kuhn–Tucker theorem": a point $\bar{x}$ solves (P) if and only if for some $\bar{y} \in Q^*$ the pair $(\bar{x}, \bar{y})$ is a saddle-point of the Lagrangian function K (given explicitly in Example 4′).

Unfortunately, for many infinite-dimensional applications condition (8.12) is of no use at all, since the cone Q of interest has no points in its core, much less points in its interior. This is the state of affairs, for instance, if $U = \mathscr{L}^2[0, 1]$ and Q consists of the functions u such that $u(t) \leqq 0$ almost everywhere; then core $Q = \emptyset$.

Of course, special properties in a given problem might still enable one to establish by some different argument that φ is bounded above on a neighborhood of 0. Then Theorem 17 could be invoked directly and the same Kuhn–Tucker theorem obtained. However, for general purposes another possibility remains: a substitute for the Slater condition based on criterion (c) in Theorem 18. This requires that U be a Banach space, with Y identifiable algebraically with the dual of U through the pairing, and perhaps more restrictively, that X be paired with a Banach space V in such a way as to be identifiable as the dual of V. To ensure that the representation function F be *closed* as well as convex, the sets C and epi Φ can be assumed to be closed and f_0 lower-semicontinuous. (If C is not closed, the lower-semicontinuity of the $+\infty$ extension of f_0 from C to all of X would suffice. The closedness of epi Φ implies that of the cone Q and expresses a sort of lower-semicontinuity of Φ.) The generalized Slater condition in these circumstances, $0 \in \text{core dom}\,\varphi$, amounts to the following:

$$\text{(8.13)} \qquad \text{for each} \quad u \in U, \quad \text{there exist} \quad \varepsilon > 0 \quad \text{and} \quad x \in C \quad \text{such that} \quad \Phi(x) - \varepsilon u \in Q.$$

Under this condition and the preceding Banach space and closedness assumptions, therefore, the desired Kuhn–Tucker theorem is again valid. This is a new result which has not been published elsewhere.

Of course, one can get the existence of the saddle-value of the Lagrangian, although not a Kuhn–Tucker theorem, from a boundedness condition corresponding to (e) of Theorem 18′.

The technical difficulties connected with generalizing the Slater condition point up one of the deficiencies of the "generalized constraint" model with a cone Q and mapping Φ, namely that Φ must have values in the space U over which the perturbations range. Looking back at the original Example 4, where $\Phi(x) = h(x, \cdot)$ (a function on the index space S), it is not so clear that this is desirable. For example, when S is a compact topological space it may be useful in some cases to restrict the parameter functions u to lie in the Banach space of continuous functions, even though $h(x, s)$ is not everywhere continuous in s. The latter approach leads to results in certain problems in stochastic programming and optimal control (involving piecewise continuous recourse functions or controls, for instance) where the cone model seems to fail, or at least be unduly awkward.

Example 5′. (Chebyshev approximation.) The situation in Example 5 is well understood, but let us see how it can be placed in the framework of the duality theory. The parametric representation

$$\text{(8.14)} \qquad F(x, u) = \|h_0 - x_1 h_1 - \cdots - x_m h_m + u\|, \qquad u \in \mathscr{C}[0, 1],$$

requires us to pair the space $U = \mathscr{C}[0, 1]$ with a suitable space Y. The natural choice is to let Y be the space of Borel measures y on $[0, 1]$, the pairing being

$$\text{(8.15)} \qquad \langle u, y \rangle = \int_0^1 u(t)y(dt).$$

For compatible topologies, one can take the norm topology on U and the weak topology on Y (the latter corresponding to the so-called weak* topology on the dual of U). The Lagrangian may then be calculated using the fact that the conjugate of the uniform norm $\|\cdot\|$ on $[0, 1]$ is the indicator of the unit ball B for the dual norm

$$\text{(8.16)} \qquad \|y\| = |y|([0, 1]).$$

One obtains

$$(8.17)\quad K(x, y) = \begin{cases} \int_0^1 [h_0(t) - x_1 h_1(t) - \cdots - x_m h_m(t)] y(dt) & \text{if } \|y\| \leq 1, \\ -\infty & \text{if } \|y\| \nleq 1. \end{cases}$$

The optimal value function φ gives the distance of $h_0 + u$ from the subspace of $\mathscr{C}[0, 1]$ generated by the functions $h_1, \cdots, h_m$. Trivially φ is finite and continuous (this also follows from condition (a) of Theorem 18), and therefore the properties of Theorem 17 are in force. On the other hand condition (d) of Theorem 18′ is satisfied, so that the properties of Theorem 17′ hold as well.

The dual problem can be calculated explicitly:

$$(\mathrm{D})\quad \begin{aligned} &\text{maximize } \int_0^1 h_0(t) y(dt) \quad \text{over all Borel measures } y \text{ on } [0, 1] \\ &\text{satisfying } \|y\| \leq 1 \quad \text{and} \quad \int_0^1 h_i(t) y(dt) = 0, \qquad i = 1, \cdots, m. \end{aligned}$$

The dual optimal value function $\gamma(v_1, \cdots, v_m)$ gives the maximum under the perturbed constraints

$$(8.18)\qquad \int_0^1 h_i(t) y(dt) = -v_i, \qquad i = 1, \cdots, m.$$

Thus the uniqueness of the coefficients $\bar{x}_i$ in the best approximation corresponds to the differentiability of this function γ at $v = 0$.

The dual problem here resembles an important moment problem in statistics (testing of hypotheses):

$$(\mathrm{D}_0)\quad \begin{aligned} &\text{maximize } \int_0^1 h_0(t) y(dt) \quad \text{over all probability measures } y \text{ on } [0, 1] \\ &\text{satisfying } \int_0^1 h_i(t) y(dt) = 0, \qquad i = 1, \cdots, m. \end{aligned}$$

Our symmetric duality scheme allows us to explore this further. Introducing the parameters v_i as in (8.18) and the corresponding function G_0, we can use the formula (4.18) to calculate a primal representation function F_0 and hence a primal problem (P_0) paired with (D_0). This yields

$$(8.19)\qquad F_0(x, u) = \max_{0 \leq t \leq 1} [h_0(t) - x_1 h_1(t) - \cdots - x_m h_m(t) - u(t)].$$

Thus (P_0) (corresponding to $u = 0$) consists of finding the coefficients $x_1, \cdots, x_m$ such that the maximum of the function $h_0 - x_1 h_1 - \cdots - x_m h_m$ is as low as possible.

Obviously the interval [0, 1] can be replaced by other compact spaces. Different kinds of moment problems can likewise be treated in this format.

For an analysis of the Kuhn–Tucker conditions for the Chebyshev problem and its generalizations, and their relationship with well-known optimality conditions in approximation, see [5], [7], [9].

Example 11′. (Fenchel duality.) The scheme in Example 11 in § 5 has

(8.20) $$\varphi(u) = \inf\{h(x) - k(Ax + u) \mid x \in \operatorname{dom} A\},$$

(8.21) $$\operatorname{dom}\varphi = \operatorname{dom} k - A \operatorname{dom} h,$$

while dually

(8.22) $$\gamma(v) = \sup\{k^*(y) - h^*(A^*y + v) \mid y \in \operatorname{dom} A^*\},$$

(8.23) $$\operatorname{dom}\gamma = \operatorname{dom} h^* - A^* \operatorname{dom} k^*.$$

Here A is a densely defined linear operator whose graph is closed, and A^* is the adjoint operator (see Example 11). Furthermore, h and k are proper convex functions, so the convexity needed in Theorem 18 is present. The Kuhn–Tucker condition, calculated from formula (5.15) for K, is

(8.24) $$A\bar{x} \in \partial k^*(\bar{y}) \quad \text{and} \quad A^*\bar{y} \in \partial h(\bar{x}).$$

The simplest criterion for strong duality which can be obtained is the one corresponding to (a) of Theorem 18:

(8.25) there exists $x \in (\operatorname{dom} h) \cap (\operatorname{dom} A)$ such that k is bounded above in a neighborhood of Ax.

Alternatively, if U and V are Banach spaces in their "compatible" topologies, and if h and k are closed, the condition

(8.26) $$0 \in \operatorname{core}(\operatorname{dom} k - A \operatorname{dom} h)$$

suffices for the conclusions of Theorem 17.

Dual forms of conditions (8.25) and (8.26), in terms of h^*, k^* and A^*, yield under the closedness assumptions on h and k the properties in Theorem 17′. The boundedness conditions ensuring $\inf(\mathrm{P}) = \max(\mathrm{D})$ or $\min(\mathrm{P}) = \sup(\mathrm{D})$ can also be made quite explicit for the model.

Especially useful is the case where $X = U$, $V = Y$, A is the identity, and $k = -\psi_C$, the convex set C being a cone. Working out the results in this case is a good exercise.

Example 13. (Minimax theory.) Let K be an extended real-valued function on $X \times Y$. We are interested in conditions on K which imply the existence of a saddle-point, or at least the saddle-value. Although other approaches are possible (slightly more general in some respects but more restrictive in others), we limit attention to the case where X and Y are linear spaces paired as above with linear spaces U and V respectively, $K(x, y)$ is convex in x and concave in y, and the semicontinuity property

(8.27) $$\operatorname{cl}_y \operatorname{cl}_x K = K$$

is satisfied. In (8.27), cl_x denotes the operation of lower closure applied to $K(x, y)$

as a function of x for each y, while cl_y denotes the operation of upper closure applied to $K(x, y)$ as a function of y for each x. The class of minimax problems covered by these assumptions is considerably broader than might be thought, as explained at the end of § 4.

Property (8.27) is satisfied if K is the Lagrangian corresponding to a function F which is closed convex; see (4.27) and (4.28). Conversely, if K is a convex-concave function satisfying (8.27) then in particular $K(x, y)$ is closed concave in y, and hence K is the Lagrangian corresponding to a convex function F, namely, F given by the formula in (4.3) (Theorem 6). In fact, this function F is also *closed*. (The function $\tilde{K} = \mathrm{cl}_x K$ satisfies (4.22) as a consequence of (4.15), but it also satisfies (4.25) by (4.3), because $\mathrm{cl}_y \tilde{K} = K$; combining these two formulas with (4.17), one sees that $F = F^{**}$.)

Thus our assumptions exactly describe the minimax problems associated with dual problems of the general form (P) and (D) for which $F(x, u)$ is closed convex in x and u jointly. Corresponding minimax theorems can be derived from Theorems 15, 17, 17′ and the conditions in Theorems 18 and 18′. For this purpose one can make direct use of the formulas

(8.28) $$F(x, u) = \sup_{y \in Y} \{K(x, y) - \langle u, y\rangle\},$$

(8.29) $$G(y, v) = \inf_{x \in X} \{K(x, y) - \langle x, v\rangle\},$$

(8.30) $$\operatorname{dom} \varphi = \bigcup_{x \in X} \{u | K(x, \cdot) - \langle u, \cdot\rangle \text{ is bounded above on } Y\}.$$

(8.31) $$\operatorname{dom} \gamma = \bigcup_{y \in Y} \{v | K(\cdot, y) - \langle \cdot, v\rangle \text{ is bounded below on } X\}.$$

To illustrate, the condition:

(8.32) *there exist $x \in X$ and $\alpha \in R$ such that the set $\{y | K(x, y) > \alpha\}$ is nonempty and equicontinuous,*

implies via (8.28) and Theorem 10 (b) that (for the same x) the function $u \to F(x, u)$ is bounded above on a neighborhood of 0. Invoking Theorems 18 (a) and 17 (a), we see therefore that (8.32) implies

(8.33) $$\inf_x \sup_y K(x, y) = \max_y \inf_x K(x, y).$$

Similarly, it can be shown that the condition:

(8.34) *there exist $y \in Y$ and $\beta \in R$ such that the set $\{x | K(x, y) < \beta\}$ is nonempty and equicontinuous*

implies

(8.35) $$\min_x \sup_y K(x, y) = \sup_y \inf_x K(x, y).$$

If both (8.32) and (8.34) hold, a saddle-point $(\bar{x}, \bar{y})$ exists.

Obviously (8.32) holds in particular if the set

(8.36) $$D = \{y \mid \exists x \text{ with } K(x, y) > -\infty\}$$

is nonempty and equicontinuous, while (8.34) holds in particular if the set

(8.37) $$C = \{x \mid \exists y \text{ with } K(x, y) < +\infty\}$$

is nonempty and equicontinuous.

Note that if K is of the form (4.30) discussed at the end of § 4, then the sets C and D in (4.30) coincide with the C and D in (8.36) and (8.37). Thus, for example, under the assumptions preceding (4.30) one has

(8.38) $$\inf_{x \in C} \sup_{y \in D} K_0(x, y) = \max_{y \in D} \inf_{x \in C} K_0(x, y)$$

if D is equicontinuous, while

(8.39) $$\min_{x \in C} \sup_{y \in D} K_0(x, y) = \sup_{y \in D} \inf_{x \in C} K_0(x, y)$$

if C is equicontinuous.

Since the choice of the compatible topologies on X and Y is arbitrary, we can just as well choose them to be the Mackey topologies (see the end of § 7). Thus "equicontinuous" can be replaced by "weakly compact" in all the minimax theorems just stated.

For more discussion of minimax theorems and references, see [16].

9. Calculating conjugates and subgradients; integral functionals. In many situations it is important to have expressions for the conjugate and subgradients of a convex function h which has been constructed from other convex functions whose properties are better known. For example, h might be a sum (finite or infinite) of differentiable convex functions and indicators of certain elementary convex sets. Various conditions sufficient for this or that formula to be valid have been established in the literature (e.g., [6], [13]). Here, however, we take a more general approach in showing how the formulas can be regarded as instances of the duality relations and Kuhn–Tucker conditions studied in the preceding sections. This illuminates the nature of the formulas more clearly and yields, by way of the duality theory, a more comprehensive list of sufficient conditions.

The essential idea is this. Given a convex function h on X and an element $\bar{v} \in V$ (V paired with X), we seek to characterize in some useful way the value $h^*(\bar{v})$ and the elements $\bar{x}$ such that $\bar{v} \in \partial h(\bar{x})$. We therefore investigate the problem

($P_{\bar{v}}$) $$\text{minimize} \quad f(x) = h(x) - \langle x, \bar{v} \rangle \quad \text{over all} \quad x \in X.$$

The optimal value in this problem is $-h^*(\bar{v})$, and its optimal solutions are precisely the elements $\bar{x}$ satisfying $\bar{v} \in \partial h(\bar{x})$. Introducing a parameter vector u in some fashion, we obtain a Lagrangian function $K_{\bar{v}}$ and a dual problem ($D_{\bar{v}}$). Then the relation

(9.1) $$\inf (P_{\bar{v}}) = \sup (D_{\bar{v}}),$$

if valid, provides a formula

$$h^*(\bar{v}) = -\sup(\mathrm{D}_{\bar{v}}). \tag{9.2}$$

Furthermore, in circumstances where the supremum in $(\mathrm{D}_{\bar{v}})$ is sure to be attained, we obtain from Corollary 15A the characterization that $\bar{v} \in \partial h(\bar{x})$ if and only if there exists $\bar{y}$ such that

$$(0, 0) \in \partial K_{\bar{v}}(\bar{x}, \bar{y}). \tag{9.3}$$

A very fundamental case, appropriate for a beginning, is that of

$$h(x) = k(Ax), \quad A: X \to U, \tag{9.4}$$

where k is a proper convex function on a certain locally convex space U (paired with a space Y), and the transformation A is linear. The question is how to express h^* and ∂h in terms of k^*, ∂k and the adjoint transformation $A^*: Y \to V$ (where as always V is paired with X). We assume only that A and A^* are *densely defined* and have *closed graphs* (cf. Example 11 in § 5); if $x \notin \operatorname{dom} A$, we interpret $h(x)$ as $+\infty$ in (9.4).

Fixing $\bar{v} \in V$, we define the (convex) function $F_{\bar{v}}$ on $X \times U$ by

$$F_{\bar{v}}(x, u) = \begin{cases} k(Ax - u) - \langle x, \bar{v} \rangle & \text{if} \quad x \in \operatorname{dom} A, \\ +\infty & \text{if} \quad x \notin \operatorname{dom} A. \end{cases} \tag{9.5}$$

Then

$$F_{\bar{v}}(x, 0) = h(x) - \langle x, \bar{v} \rangle, \tag{9.6}$$

that is, we have a "convex parameterization" of problem $(\mathrm{P}_{\bar{v}})$ as desired. The Lagrangian is

$$K_{\bar{v}}(x, y) = \begin{cases} \langle Ax, y \rangle - \langle x, \bar{v} \rangle - k^*(y) & \text{if} \quad x \in \operatorname{dom} A, \\ +\infty & \text{if} \quad x \notin \operatorname{dom} A. \end{cases} \tag{9.7}$$

The Kuhn–Tucker condition (9.3) therefore reduces to

$$\bar{x} \in \operatorname{dom} A, \quad A\bar{x} \in \partial k^*(\bar{y}), \quad \bar{y} \in \operatorname{dom} A^*, \quad A^*\bar{y} = \bar{v}. \tag{9.8}$$

Furthermore,

$$\begin{aligned} G_{\bar{v}}(y, v) &= \inf_{x \in \operatorname{dom} A} \{\langle Ax, y \rangle - \langle x, \bar{v} \rangle - k^*(y) - \langle x, v \rangle\} \\ &= \begin{cases} -k^*(y) & \text{if} \quad y \in \operatorname{dom} A^* \quad \text{and} \quad A^*y = \bar{v} + v, \\ -\infty & \text{otherwise.} \end{cases} \end{aligned} \tag{9.9}$$

The dual problem $(\mathrm{D}_{\bar{v}})$ is thus:

$$\text{maximize} \quad -k^*(y) \quad \text{over all} \quad y \in \operatorname{dom} A^* \quad \text{with} \quad A^*y = \bar{v}. \tag{9.10}$$

We note finally that the optimal value function is

$$\varphi_{\bar{v}}(u) = -\sup_{x \in \operatorname{dom} A} \{\langle x, \bar{v} \rangle - k(Ax - u)\}, \tag{9.11}$$

and consequently,

$$\text{dom}\,\varphi_{\bar v} = \text{range}\,A - \text{dom}\,k. \tag{9.12}$$

If $\varphi_{\bar v}$ is bounded above on a neighborhood of 0, formulas of the kind we want for h^* and ∂h can be obtained from Theorem 17 (a) and Corollary 15A. The sufficient conditions in Theorem 18 provide the following.

THEOREM 19. *Let $h(x) = k(Ax)$, where k is a proper convex function on U, and $A : X \to U$ is a densely defined linear transformation with closed graph ($h(x)$ interpreted as $+\infty$ for $x \notin \text{dom}\,A$). The two formulas*

(i) $\partial h(x) = A^*\partial k(Ax)$ *for all* x ($\emptyset$ *if* $x \notin \text{dom}\,A$),

(ii) $h^*(v) = \min\{k^*(y) \mid y \in \text{dom}\,A^*,\ A^*y = v\}$ *for all* $v \in V$,

are valid, provided that any one of the following conditions is fulfilled:

(a) *There exists an $x \in \text{dom}\,A$ such that the function k is bounded above on a neighborhood of Ax. (Or more generally, there is a continuous $\theta : U \to$ range A such that the function $u \to k(\theta(u) - u)$ is bounded above on a neighborhood of* 0.)

(b) $U = R^n = Y$, *and* $0 \in \text{core}\,(\text{range}\,A - \text{dom}\,k)$.

(c) *U and V are both Banach spaces (in the designated "compatible" topologies), k is closed and* $0 \in \text{core}\,(\text{range}\,A - \text{dom}\,k)$.

(d) *$U = R^n = Y$, and for some $\bar v \in V$ and real number α the set $\{y \mid A^*y = \bar v, k^*(y) \leqq \alpha\}$ is nonempty and bounded.*

(e) *k is closed, and for some open set M in V and some real number α, the set*

$$\{y \mid y \in \text{dom}\,A^*, A^*y \in M, k^*(y) < \alpha\} \tag{9.13}$$

is nonempty and equicontinuous.

Remark. The condition $0 \in \text{core}\,(\text{range}\,A - \text{dom}\,k)$ is satisfied in particular if the range of A meets core (dom k).

Proof. Fixing an arbitrary $\bar v \in V$, we find that conditions (a), (b) and (c) of Theorem 18 follow, for the problems above, from (a), (b) and (c) of the present theorem. But the latter are independent of $\bar v$. Therefore, if one of them holds, we have $\varphi_{\bar v}$ bounded above on a neighborhood of 0 for all $\bar v$. Hence from Theorem 17 and Corollary 15A we also have

$$\inf(\mathrm{P}_{\bar v}) = \max(\mathrm{D}_{\bar v}) \quad \text{for all } \bar v \tag{9.14}$$

and the characterization

$$\bar v \in \partial h(\bar x) \quad \text{if and only if (9.8) holds for some } \bar y. \tag{9.15}$$

Before discussing (d) and (e), we demonstrate that the properties just mentioned furnish the desired conclusion. Of course (9.14) is identical to formula (ii). On the other hand, since $A\bar x \in \partial k^*(\bar y)$ if and only if $\bar y \in \partial k^{**}(A\bar y)$ (Corollary 12A applied to $\varphi = k^*$), we can write (9.15) as

$$\partial h(\bar x) = A^*\partial k^{**}(A\bar x) \quad \text{for all } \bar x \quad (\emptyset \text{ if } \bar x \notin \text{dom}\,A). \tag{9.16}$$

(The notation A^*S, where S is a subset of Y, denotes the set of all A^*y such that $y \in S \cap \operatorname{dom} A^*$.) To get (ii), we must show ∂k^{**} can be replaced by ∂k in (9.16). Recall from Theorem 12 that $\partial k^{**}(A\bar{x}) = \partial k(A\bar{x})$ if $\partial k(A\bar{x}) \neq \varnothing$ or if $k(A\bar{x}) = k^{**}(A\bar{x})$. It will be enough therefore to show that if $\partial h(\bar{x}) \neq \varnothing$, then $h(\bar{x}) = h_0(\bar{x})$, where

$$(9.17) \qquad h_0(x) = \begin{cases} k^{**}(Ax) & \text{if } \quad x \in \operatorname{dom} A, \\ +\infty & \text{if } \quad x \notin \operatorname{dom} A. \end{cases}$$

Observe that the property of $\varphi_{\bar{v}}$ being bounded above on a neighborhood of 0 is not lost if k is replaced by k^{**}, inasmuch as $k^{**} \leqq k$. Thus, in the situation being considered, formula (9.14) holds also with k replaced by k^{**}. In other words, we have

$$(9.18) \qquad h_0^*(v) = \min \{k^{**}(y) \mid y \in \operatorname{dom} A, A_y^* = v\} \quad \text{for all } v.$$

But $k^{***} = k^*$. This shows, therefore, that $h_0^* = h^*$. Hence, $h \geqq h_0 \geqq h_0^{**}$. If $\partial h(\bar{x}) \neq \varnothing$, then $h(\bar{x}) = h^{**}(\bar{x})$ by Theorem 12. Thus if $\partial h(\bar{x}) \neq \varnothing$, we do have $h(\bar{x}) = h_0(\bar{x})$ as required.

We turn now to condition (d). The boundedness property here is equivalent to condition (d) of Theorem 18 for the problem $(P_{\bar{v}})$. Thus if this property holds for some $\bar{v}$, we may conclude that $\varphi_{\bar{v}}$ is bounded above on a neighborhood of 0. But this implies condition (b) of the present theorem, thus reverting everything to the case already analyzed.

We shall show finally that condition (e) implies for any $\bar{v}$ the existence of a neighborhood N of 0 in V such that the set

$$(9.19) \qquad \{y \mid y \in \operatorname{dom} A^*, A^*y - \bar{v} \in N, k^*(y) < \alpha\}$$

is nonempty and equicontinuous. This property says that condition (e) of Theorem 18 is satisfied for $(P_{\bar{v}})$, so this argument will complete the proof.

Since the set (9.13) is nonempty, we can select y_0 and v_0 such that

$$(9.20) \qquad y_0 \in \operatorname{dom} A^*, \quad A^*y_0 = v_0 \in M, \quad k^*(y_0) < \alpha.$$

Since M is open, there exists an open convex neighborhood M_0 of 0 in V such that $v_0 + M_0 \subset M$. Choose $\lambda \in (0, 1]$ sufficiently small that $\bar{v} - v_0 \in (1/\lambda)M_0$, and set

$$(9.21) \qquad N = (1/\lambda)M_0 + v_0 - \bar{v}.$$

Then N is a neighborhood of 0, and the set (9.19) is nonempty (it contains y_0). Denote (9.19) by C and (9.13) by B. Our goal is to prove now that C is equicontinuous. For each $y \in C$, we have $A^*y \in (1/\lambda)M_0 + v_0$ by (9.21), so that

$$\lambda(A^*y - v_0) \in M_0.$$

The vector $\lambda y + (1 - \lambda)y_0$, which again belongs to the subspace dom A^*, therefor satisfies

$$(9.22) \qquad \begin{aligned} A^*[\lambda y + (1 - \lambda)y_0] &= \lambda A^*y + (1 - \lambda)v_0 \\ &= v_0 + \lambda(A^*y - v_0) \in v_0 + M_0 \subset M. \end{aligned}$$

It also satisfies

(9.23) $$k^*(\lambda y + (1-\lambda)y_0) \leqq \lambda k^*(y) + (1-\lambda)k^*(y_0) < \lambda\alpha + (1-\lambda)\alpha = \alpha.$$

Thus for each $y \in C$ we have $\lambda y + (1-\lambda)y_0$ in B. In other words,

$$C \subset (1/\lambda)[B - (1-\lambda)y_0].$$

Since B is equicontinuous, and equicontinuity of sets is preserved under translation and scalar multiplication, this relation implies C is equicontinuous.

Remark 1. Part (ii) of Theorem 19 implies that, under any one of the conditions (a)–(e) the function A^*k^* on V defined by

(9.24) $$(A^*k^*)(v) = \inf\{k^*(y) \mid y \in \operatorname{dom} A^*, A^*y = v\}$$

is a *closed* convex function, moreover with the infimum always attained. This result can also be stated in a more direct way, reversing the dual roles of the elements. Thus let q be a closed proper convex function on X and consider the function

(9.25) $$(Aq)(u) = \inf\{q(x) \mid x \in \operatorname{dom} A, Ax = u\}$$

on U. A sufficient condition for Aq to be a closed proper convex function, moreover with the infimum always attained, is according to (a) of Theorem 19 that there exists $y \in \operatorname{dom} A^*$ such that q^* is bounded above on a neighborhood of A^*y. Other sufficient conditions follow similarly from (b), (c), (d) and (e).

Remark 2. Theorem 18′ similarly yields sufficient conditions for the equation $\min(\mathrm{P}_{\bar v}) = \sup(\mathrm{D}_{\bar v})$ to be valid, and hence the formula

(ii′) $$h^*(\bar v) = \inf\{k^*(y) \mid y \in \operatorname{dom} A^*, A^*y = \bar v\}.$$

However, the conditions depend on the particular $\bar v$, except in unusual circumstances, and they do not provide much useful information about ∂h.

Next we treat the important case of the conjugate and subgradient of a sum of convex functions. Our approach is to represent this as a special case of Theorem 19.

THEOREM 20. *Let $h(x) = f_1(x) + \cdots + f_m(x)$, where f_i is a proper convex function on X, $i = 1, \cdots, m$. The two formulas*

(i) $$h(x) = f_1(x) + \cdots + f_m(x) \quad \text{for all} \quad x \in X,$$

(ii) $$h^*(v) = \min\{f_1^*(v_1) + \cdots + f_m^*(v_m) \mid v_1 + \cdots + v_m = v\} \quad \text{for all} \quad v \in V,$$

are valid, provided that any one of the following conditions is fulfilled:

(a) *There exists an $\bar x \in \operatorname{dom} f_1$ in a neighborhood of which the functions $f_2, \cdots, f_m$ are all bounded above. (The role of f_1 could be played by any one of the functions.)*

(b) *$X = R^n = V$ and $0 \in \operatorname{core} W$, where*

(9.26) $$W = \left\{(x_1, \cdots, x_m) \mid x_i \in X, \bigcap_{i=1}^{m} [(\operatorname{dom} f_i) + x_i] \neq \varnothing\right\}.$$

(c) *X is a reflexive Banach space (in the designated "compatible" topology), the functions f_i are closed, and $0 \in \operatorname{core} W$, where W is as in* (b).

(d) $X = R^n = V$, *and for some* $\bar{v}$ *and real* α *the set*

(9.27) $\{(v_1, \cdots, v_m) | v_i \in \operatorname{dom} f_i^*, v_1 + \cdots + v_m = \bar{v}, f_1^*(v_1) + \cdots + f_m^*(v_m) \leqq \alpha\}$

is nonempty and bounded.

(e) *The functions* f_i *are closed, and for some open set* M *in* V *the set*

(9.28) $\{(v_1, \cdots, v_m) | v_i \in \operatorname{dom} f_i^*, v_1 + \cdots + v_m \in M, f_1^*(v_1) + \cdots + f_m^*(v_m) < \alpha\}$

is nonempty and equicontinuous.

Proof. Let $U = X \times \cdots \times X$ (m times) in the natural pairing with $Y = V \times \cdots \times V$. Define $A : X \to U$ by

$$Ax = (x, \cdots, x),$$

and for $u = (x_1, \cdots, x_m)$ set

$$k(u) = k(x_1, \cdots, x_m) = f_1(x_1) + \cdots + f_m(x_m).$$

Then A is a continuous linear transformation with

$$A^*y = v_1 + \cdots + v_m \quad \text{for} \quad y = (v_1, \cdots, v_m),$$

while k is a proper convex function with

$$k^*(y) = k^*(v_1, \cdots, v_m) = f_1^*(v_1) + \cdots + f_m^*(v_m).$$

It remains only to apply Theorem 19. In particular, to get (a) here, let $\theta(u) = (\bar{x} + x_1, \cdots, \bar{x} + x_1)$ for $u = (x_1, \cdots, x_m)$ in (a) of Theorem 19, so that $\theta(u) - u = (\bar{x}, \bar{x} + x_1 - x_2, \cdots, \bar{x} + x_1 - x_m)$.

Remark. The operation described by formula (ii) of Theorem 20 (forming a new function by taking the indicated infimum for each v) is called *inf-convolution* of the functions f_i^*. Notationally: $h^* = f_i^* \,\square \cdots \square\, f_m^*$. In more direct terms, the inf-convolute $f_1 \,\square \cdots \square\, f_m$ is defined by

(9.29) $(f_1 \,\square \cdots \square\, f_m)(x) = \inf\{f_1(x_1) + \cdots + f_m(x_m) | x_1 + \cdots + x_m = x\}.$

This is convex if every f_i is convex. It is easy to verify that even without the hypothesis of convexity, one has

(9.30) $$(f_1 \,\square \cdots \square\, f_m)^* = f_1^* + \cdots + f_m^*.$$

The operations $+$ and $\square$ are thus dual to each other with respect to taking conjugates. Sufficient conditions for $f_1 \,\square \cdots \square\, f_m$ to be a closed convex function, with the infimum in its definition always attained, can be derived from the conditions in Theorem 20 by duality.

Examples. 1. Let k be a proper convex function on X, and consider the problem of minimizing k over a nonempty convex subset C. Let $h(x) = k(x) + \psi_C(x)$, where ψ_C is the indicator of C. The problem is the same as minimizing h over X, and its solutions are thus the points $\bar{x}$ such that $0 \in \partial h(\bar{x})$. Theorem 20 gives sufficient conditions for the latter relation to be expressible as

$$0 \in \partial k(\bar{x}) + \partial \psi_C(\bar{x}).$$

This condition means there is an element $\bar{y}$ normal to C at $\bar{x}$ such that $-\bar{y}$ is a subgradient of k at $\bar{x}$. (See the example in § 6 following Corollary 12B.)

2. To see that condition (c) of Theorem 20 definitely yields something not covered by condition (a), let $X = \mathscr{L}^2[0, 1] = V$ and let f_1 and f_2 be arbitrary closed proper convex functions such that $\operatorname{dom} f_1 = \operatorname{dom} f_2 = \mathscr{L}^2_+$. (Here $\mathscr{L}^2_+$ is the "nonnegative orthant," consisting of the functions which are nonnegative almost everywhere.) Then

$$[(\operatorname{dom} f_1) + x_1] \cap [(\operatorname{dom} f_2) + x_2] \neq \emptyset \quad \text{for all } x_1, x_2,$$

so W is the whole of $\mathscr{L}^2 \times \mathscr{L}^2$. The core condition is thus satisfied, and formulas (i) and (ii) are therefore valid. But there is no question of f_1 or f_2 being bounded above anywhere, because core $\mathscr{L}^2_+ = \emptyset$.

We return now to the integral functionals

$$I(x) = \int_S h(x(s), s)\sigma(ds),: \qquad x \in \mathscr{X}, \tag{9.31}$$

considered in Theorem 3 in § 2. Results about their conjugates and subgradients will in particular enable us to generalize Theorem 20 to certain infinite or "continuous" sums of convex functions (Theorem 23). As earlier, $\mathscr{X}$ denotes a linear space of measurable functions $x : S \to X$. We shall assume that X *is a separable Banach space*, the case of $X = R^n$ being of particular interest. (A Banach space is said to be *separable* if it has a countable dense subset.)

Let $\mathscr{V}$ denote a linear space of measurable functions $v : S \to V$, where V is paired with X and thus identifiable algebraically with the dual of X as a Banach space. There is a natural pairing of $\mathscr{X}$ with $\mathscr{V}$, namely

$$\langle x, v \rangle = \int_S \langle x(s), v(s) \rangle \sigma(ds), \qquad x \in \mathscr{X}, v \in \mathscr{V}, \tag{9.32}$$

provided that the integrand in (9.32) (which is certainly measurable) is in fact summable for every $x \in \mathscr{X}$ and $v \in \mathscr{V}$. This summability will be assumed.

We are interested in calculating the conjugate of the integral functional I with respect to the pairing (9.32). We have by definition,

$$\begin{aligned} I^*(v) &= \sup_{x \in \mathscr{X}} \{\langle x, v \rangle - I(x)\} \\ &= \sup_{x \in \mathscr{X}} \int_S [\langle x(s), v(s) \rangle - h(x(s), s)]\sigma(ds). \end{aligned} \tag{9.33}$$

In approaching the supremum, we would like, for each s, to choose $x(s)$ so that the quantity $\langle x(s), v(s) \rangle - h(x(s), s)$ is as high as possible. If the functions in $\mathscr{X}$ are such that the values $x(s)$ can be specified with enough "independence," we can hope to continue the calculation by equality with

$$\int_S \sup_{x(s) \in X} [\langle x(s), v(s) \rangle - h(x(s), s)]\sigma(ds) = \int_S h^*(v(s), s)\sigma(ds), \tag{9.34}$$

where $h^*(\cdot, s)$ is the conjugate of $h(\cdot, s)$ for each $s \in S$. But this also hinges on the latter integral being well-defined. The measurability of h does not obviously imply that of h^* without some restrictions.

The space $\mathscr{X}$ is said to be *decomposable* if, whenever T is a subset of S of finite measure and $x_0 : T \to X$ is a measurable function whose range is bounded, then for every $x \in \mathscr{X}$ the function

$$x(s) = \begin{cases} x_0(s) & \text{if} \quad s \in T, \\ x(s) & \text{if} \quad s \notin T, \end{cases}$$

also belongs to $\mathscr{X}$. For example, $\mathscr{X}$ is decomposable if $\mathscr{X} = \mathscr{L}^p_X(S, \Sigma, \sigma)$, the space of all measurable functions $x : S \to X$ satisfying

$$\text{(9.35)} \qquad \int_S \|x(s)\|^p \sigma(ds) < +\infty .$$

On the other hand, if S has a topological structure and the functions in $\mathscr{X}$ are all continuous, then $\mathscr{X}$ is typically not decomposable.

Recall that the measure space (S, Σ, σ) is said to be *complete* if every subset of a measurable set of measure zero is itself measurable. It is *totally sigma-finite* if S is the union of countably many sets of finite measure.

THEOREM 21. *Let the integrand h be measurable on $X \times S$ (relative to the Borel structure on X), where the Banach space X is separable and the measure space (S, Σ, σ) is complete and totally sigma-finite. Assume that h is lower-semicontinuous in the X argument, $I(x) < +\infty$ for at least one $x \in \mathscr{X}$ in* (9.31), *where $\mathscr{X}$ is decomposable. Then:*

(a) *The conjugate integrand h^* is measurable on $V \times S$ (relative to the Borel structure on V), and the conjugate of I with respect to the pairing* (9.32) *is*

$$\text{(9.36)} \qquad I^*(v) = \int_S h^*(v(s), s)\sigma(ds), \qquad v \in V.$$

(b) *In particular, the latter integral is well-defined and convex on V, and it is lower-semicontinuous with respect to the weak topology on $\mathscr{V}$ induced by $\mathscr{X}$.*

(c) *An element $v \in \mathscr{V}$ belongs to $\partial I(x)$ if and only if $v(s) \in \partial h(x(s), s)$ for almost every s, where $\partial h(x(s), s)$ is the set of subgradients of $h(\cdot, s)$ at the point $x(s) \in X$.*

(d) *If in addition X is reflexive, $\mathscr{V}$ is decomposable, and $I^*(v) < +\infty$ for at least one $v \in \mathscr{V}$ in* (9.36), *one has*

$$\text{(9.37)} \qquad I^{**}(x) = \int_S \operatorname{cl} \operatorname{co} h(x(s), s)\sigma(hs), \qquad x \in \mathscr{X}.$$

Proof. A proof of (a) can be found in [19]. Assertions (b) and (d) follow from (a) via Theorem 5. As for (c), this is immediate from (9.36) and the following relations (see (6.16) and (6.17)): $I^*(v) \geqq \langle x, v \rangle - I(v)$, with equality if and only if $v \in \partial I(x)$, and similarly

$$h^*(v(s), s) \geqq \langle x(s), v(s) \rangle - h(x(s), s),$$

with equality if and only if $v(s) \in \partial h(x(s), s)$.

Although Theorem 21 applies only to decomposable function spaces $\mathscr{X}$, some results are also known in other cases. For instance if S is a compact space and $\mathscr{X}$ is the space of all continuous R^n-valued functions on S, there is a natural pairing of $\mathscr{X}$ with the space $\mathscr{V}$ of all R^n-valued regular Borel measures on S:

$$\langle x, v\rangle = \int_S x(s)\cdot v(ds). \tag{9.38}$$

The conjugate of an integral functional like I with respect to this pairing can be calculated under certain additional assumptions on h.

As one might expect, if the measure v is absolutely continuous with respect to the underlying (regular Borel) measure σ and therefore expressible by a density function $dv/d\sigma \in \mathscr{L}^1_{R^n}(S, \Sigma, \sigma)$, we get

$$I^*(v) = \int_S h^*((dv/d\sigma)(s), s)\sigma(ds). \tag{9.39}$$

It is not obvious, however, what $I^*(v)$ ought to be when v is a singular measure (i.e., concentrated in a set of σ-measure 0). The answer turns out to be that then

$$I^*(v) = \int_S j((dv/d\theta)(s), s)\theta(ds), \tag{9.40}$$

where θ is an arbitrary nonnegative (regular Borel) measure with respect to which the R^n-valued measure v is absolutely continuous, and $j(\cdot, s)$ is for each $s \in S$ the so-called *recession function* [13] associated with the convex function $h^*(\cdot, s)$. In the general case, where v is the sum of an absolutely continuous component and a singular component, $I^*(v)$ is the sum of the integrals in (9.39) and (9.40) applied to the two components separately. For the details, see [18].

This fact is interesting for many optimization problems, because it indicates the natural way of extending an integral functional on $\mathscr{L}^1_{R^n} = \mathscr{L}^1_{R^n}(S, \Sigma, \sigma)$ to a functional on the larger Banach space consisting of all R^n-valued measures on S, the elements of $\mathscr{L}^1_{R^n}$ being identified with the measures for which they are the density functions. The general measures can be regarded as ideal limits of sequences in $\mathscr{L}^1_{R^n}$ which otherwise would have no limit.

In minimization problems somehow involving an integral functional on $\mathscr{L}^1_{R^n}$, and these are very common, it may be possible by passing to the extended functional to arrange the existence of a generalized solution where otherwise no solution would exist. Similarly, it may be possible to obtain a generalized solution to a dual problem and thereby a generalized Kuhn–Tucker condition which is not only sufficient but necessary for optimality.

A powerful device for calculating the conjugate of I in the case of nondecomposable spaces $\mathscr{X}$ is to identify $\mathscr{X}$ with a subspace of some $\mathscr{L}^p_{X'}(S, \Sigma, \sigma)$ and I with the restriction to $\mathscr{X}$ of an integral functional I'. (The space X' may differ from X.) One then has $I(x) = I'(Ax)$, where A is the embedding mapping, and Theorem 19 map be applied. This depends, of course, on being able to produce the continuity or boundedness properties specified in the hypothesis of Theorem 19.

The following theorem provides some criteria which are useful in this respect. However, some of the more general results that are known entail that $\mathscr{X}$ be identified with a subspace of some $\mathscr{L}^{\infty}_{X}(S, \Sigma, \sigma)$ which is in turn paired, not with $\mathscr{L}^1_{V}(S, \Sigma, \sigma)$, but a large space—its Banach dual. We cannot discuss the technical complications here; see [18], [19].

THEOREM 22. *Let the integrand h be measurable on $X \times S$ (relative to the Borel structure on X), where the Banach space X is separable and reflexive, and the measure space (S, Σ, σ) is complete and totally sigma-finite. Assume that h is lower-semicontinuous and convex in the X argument and that $h(s, x(s))$ is summable in s for every $x \in \mathscr{L}^p_X(S, \Sigma, \sigma)$, where $1 \leqq p \leqq \infty$. Then:*

(a) *The integral functional $I(x) = \int_S h(x(s), s)\sigma(ds)$ is well-defined, finite, convex and everywhere continuous on $\mathscr{L}^p_X(S, \Sigma, \sigma)$. The continuity is not only with respect to the norm, but also in the Mackey topology induced on $\mathscr{L}^p_X(S, \Sigma, \sigma)$ by $\mathscr{L}^q_V(S, \Sigma, \sigma)$ under the pairing (9.32), where $(1/p) + (1/q) = 1$. (The two topologies coincide, except for $p = \infty$, $q = 1$.)*

(b) *The conjugate functional on $\mathscr{L}^q_V(S, \Sigma, \sigma)$ is*

$$I^*(v) = \int_S h^*(v(s), s)\sigma(ds).$$

This convex integral functional has in fact the property that for every $\bar{x} \in \mathscr{L}^p_X(S, \Sigma, \sigma)$ and every $\alpha \in R$ the set

$$\{v \in \mathscr{L}^q_V(S, \Sigma, \sigma) \mid I^*(v) - \langle \bar{x}, v\rangle \leqq \alpha\} \tag{9.41}$$

is equicontinuous and hence weakly compact.

(c) *For $x \in \mathscr{L}^p_X(S, \Sigma, \sigma)$, the condition that $v \in \partial h(x(s), s)$ almost everywhere implies $v \in \mathscr{L}^q_V(S, \Sigma, \sigma)$, and it is equivalent to $v \in \partial I(x)$.*

Remark. The summability property assumed in the theorem is implied for $1 \leqq p < \infty$ by the growth condition

$$h(x, s) \leqq (a/p)\|x\|^p + b(s), \tag{9.42}$$

where a is a positive number and b a summable function. Taking conjugates on both sides of (9.42), one can express the growth condition dually as

$$h^*(v, s) \geqq (1/aq)\|v\|^q - b(s). \tag{9.43}$$

If $p = \infty$ and $X = R^n = V$, it can be shown that the summability property, and indeed all the properties assumed for h, are implied by the following: $h(x, s)$ is a finite convex function of $x \in R^n$ for each $s \in S$ and a summable (measurable) function of $s \in S$ for each $x \in R^n$ [18].

Proof. Certainly I is well-defined, finite and convex (Theorem 3). The formula for I^* is valid by Theorem 21. Part (c) follows from (a) and Theorem 21(c).

We now show that I is lower-semicontinuous in the norm topology. Suppose not. Then there is an element x, a sequence $(x_k)_{k=1}^{\infty}$ converging to x and a number α, such that

$$\int_S h(x_k(s), s)\sigma(ds) \leqq \alpha < \int_S h(x(s), s)\sigma(ds) \quad \text{for all } k. \tag{9.44}$$

Since $\|x_k - x\| \to 0$, we can assume (passing to a subsequence if necessary) that

$$\|x_k(s) - x(s)\| \to 0 \quad \text{a.e.} \tag{9.45}$$

(This is trivial for $p = \infty$, while for $p < \infty$ it follows from the fact that the functions $s \to \|x_k(s) - x(s)\|^p$ converge to 0 as elements of $\mathscr{L}^1_R$. As is well-known, every convergent sequence in $\mathscr{L}^1_R$ has a subsequence converging pointwise almost everywhere.) But (9.45) implies by Fatou's lemma and the lower-semicontinuity of $h(\cdot, s)$ that

$$\begin{aligned}\liminf_{k\to\infty} \int_S h(x_k(s), s)\sigma(ds) &\geqq \int_S \liminf_{k\to\infty} h(x_k(s), s)\sigma(ds)\\ &\geqq \int_S h(x(s), s)\sigma(ds).\end{aligned}$$

This contradicts (9.44) and establishes the lower-semicontinuity of I.

Since I is finite, convex and lower-semicontinuous, and $\mathscr{L}^p_X(S, \Sigma, \sigma)$ is a Banach space, we may conclude that I is continuous (Corollary 8B).

For $p < \infty$, standard theorems of functional analysis assert that the norm topology on $\mathscr{L}^p_X$ is compatible with the pairing (9.32) with $\mathscr{L}^q_V$, and that the norm topology coincides with the induced Mackey topology. Applying Theorem 10 to $\varphi(x) = I(\bar{x} + x)$, we see that the conjugate function $\varphi^*(v) = I^*(v) - \langle \bar{x}, v\rangle$ has equicontinuous, weakly compact level sets for every $\bar{x}$, as claimed in (b).

For $p = \infty$, the rest of the proof is too complicated to be furnished here. It is given in [18, Theorem 2] for $X = R^n = V$. The general case is a straightforward extension of the same argument, making use of the decomposition theorem of [6, Appendix I].

As an application of Theorem 22, we state a version of Theorem 20 for infinite sums of convex functions.

THEOREM 23. *Let*

$$J(x) = \int_S h(x, s)\sigma(ds), \qquad x \in X, \tag{9.46}$$

where X is a separable reflexive Banach space, the measure space (S, Σ, σ) is complete and totally sigma-finite, and h is measurable on $X \times S$ (relative to the Borel structure on X) and convex in the X argument. Assume that $h(x(s), s)$ is summable in s whenever $x(s)$ is a bounded measurable function of s. Then:

(a) *J is a finite, continuous, convex function with $J(x) < +\infty$ everywhere.*

(b) *For each $x \in X$ one has*

$$\partial J(x) = \int_S \partial h(x, s)\sigma(ds), \tag{9.47}$$

or in other words, $v \in \partial J(x)$ if and only if there exists a measurable function w with $w(s) \in \partial h(x, s)$ for almost every s (implying $w \in \mathscr{L}^1_V(S,\Sigma,\delta)$), such that

$$\int_S w(s)\sigma(ds) = v \quad (\textit{strong integral}). \tag{9.48}$$

(c) *The conjugate of* J *on* V *is given by*

$$J^*(v) = \min\left\{\int_S h^*(w(s), s)\sigma(ds) \,|\, w \in \mathscr{L}_V^1 \text{ satisfies } (9.48)\right\}. \tag{9.49}$$

Remarks. If $X = R^n = V$, the summability condition is implied by the assumption that $h(x, s)$ is summable in s for each fixed $x \in X$. Note that the operation (9.49) is a sort of infinite infimal convolution of the convex functions $h^*(\cdot, s)$.

Proof. We follow the same pattern as in Theorem 20. Let $A: X \to \mathscr{L}_X^\infty$ be the linear transformation which assigns to each $x \in X$ the constant function $s \to x$. Then we have $J = I(Ax)$, where I is given by (9.31). The adjoint $A^*: \mathscr{L}_V^1 \to V$ assigns to each function w the element (9.48). To obtain the desired conclusion from Theorem 19, it is enough to know that I is continuous everywhere in the Mackey topology induced on $\mathscr{L}_X^\infty$ by $\mathscr{L}_V^1$. But this is asserted by Theorem 22.

Some versions of Theorem 23 are known which do not require $h(x, s)$ to be finite everywhere in x, but they are more complicated. For these and many other general results on the calculation of conjugates, we refer to [6], [13].

Only one further fact is needed for the applications we want to discuss. In this result, our notation as usual has the space U paired with a space Y, as well as X with V.

THEOREM 24. (a) *Suppose that*

$$\varphi(u) = \inf_{x \in X} F(x, u) \quad \textit{for all} \quad u \in U, \tag{9.50}$$

where $F(x, u)$ *is convex in* (x, u) *(extended-real-valued). Let* $\bar{u} \in U$ *be such that the infimum for* $\varphi(\bar{u})$ *is attained, and let* $\bar{x}$ *denote any one of the elements in* X *at which it is attained. Then the subgradients of the convex function* φ *at* $\bar{u}$ *are given by*

$$\partial\varphi(\bar{u}) = \{y \in Y \,|\, (0, y) \in \partial F(\bar{x}, \bar{u})\}. \tag{9.51}$$

(b) *Suppose that*

$$f(x) = \sup_{y \in Y} K(x, y) \quad \textit{for all} \quad x \in X, \tag{9.52}$$

where $K(x, y)$ *is convex in* x, *concave in* y *(extended-real-valued), and* $\mathrm{cl}_y \mathrm{cl}_x K = K$. *Suppose there exist* $\bar{x} \in X$ *and* $\alpha \in R$ *such that the set* $\{y | K(\bar{x}, y) > \alpha\}$ *is nonempty and equicontinuous. Then the subgradients of the convex function* f *are given by*

$$\partial f(x) = \{v \in V \,|\, \exists y \in M(x) \text{ with } v \in \partial_x K(x, y)\} \quad \textit{for all} \quad x \in X, \tag{9.53}$$

where $M(x)$ *is the set of elements (if any) for which the supremum in* (9.52) *is attained.*

Proof. (a) We have $\varphi(\bar{u}) = F(\bar{x}, \bar{u})$, and hence the subgradient inequality

$$\varphi(u) \geqq \varphi(\bar{u}) + \langle u - \bar{u}, y\rangle \quad \text{for all } u$$

is equivalent to

$$F(x, u) \geqq F(\bar{x}, \bar{u}) + \langle \bar{x} - x, 0\rangle + \langle u - \bar{u}, y\rangle \quad \text{for all } x, u.$$

But the latter says $(0, y) \in \partial F(\bar{x}, \bar{u})$.

(b) As explained in Example 13 in § 8, our assumptions imply that K is the Lagrangian for a dual pair of problems (P) and (D) with F closed convex, and moreover with

$$\inf_x \sup_y K(x, y) = \max_y \inf_x K(x, y).$$

This equation is the same as inf (P) = max (D). In particular then, we have by Corollary 15A that $0 \in \partial f(\bar{x})$ if and only if there exists $\bar{y}$ with $(0, 0) \in \partial K(\bar{x}, \bar{y})$, or in other words, $0 \in \partial_x K(\bar{x}, \bar{y})$ and $\bar{y} \in M(\bar{x})$. More generally, applying the same argument to the function

$$f_{\bar{v}}(x) = \inf_{y \in Y} K_{\bar{v}}(x, y) = f(x) - \langle x, \bar{v} \rangle,$$

where $K_{\bar{v}}(x, y) = K(x, y) - \langle x, \bar{v} \rangle$, we see that $\bar{v} \in \partial f(\bar{x})$ if and only if there exists $\bar{y} \in M(\bar{x})$ with $\bar{v} \in \partial_x K(\bar{x}, \bar{y})$.

Remark. Theorem 24 (b) can be applied to the more general case of

$$f(x) = \max_{s \in S} h(x, s), \qquad x \in X, \tag{9.54}$$

where S is a compact Hausdorff space, $h(x, s)$ is proper convex in x, and $\mathrm{cl}_s\, \mathrm{cl}_x\, h = h$. Let Y denote the space of all (real-valued) regular Borel measures on S and define

$$K(x, y) = \begin{cases} \int_S h(x, s) y(ds) & \text{if} \quad y \in P, \\ -\infty & \text{if} \quad y \notin P, \end{cases} \tag{9.55}$$

where $P = \{y \in Y \mid y \geqq 0, y(S) = 1\}$. It can be shown that the hypotheses of Theorem 24 (b) are fulfilled (with $U = C(S)$). One obtains in this way the result that $v \in \partial f(x)$ if and only if there exists a probability measure $y \in P$ such that

$$v \in \partial \int_S h(x, s) y(ds), \tag{9.56}$$

and such that the support of y is contained in the subset of S where the maximum in (9.54) is attained. To analyze (9.56) further, one can use Theorem 23 or one of its generalizations (see [6]).

10. More applications. The results in § 9 enable us to analyze further the Kuhn–Tucker conditions and dual problems in our earlier examples, particularly those involving integral functionals.

Example 1‴. (Convex programming.) Assume for simplicity, that $C = X$. Thus the functions f_i are everywhere finite and convex, and

$$K(x, y) = \begin{cases} f_0(x) + y_1 f_1(x) + \cdots + y_m f_m(x) & \text{if} \quad y \geqq 0, \\ -\infty & \text{if} \quad y \ngeqq 0, \end{cases} \tag{10.1}$$

where $y = (y_1, \cdots, y_m) \in R^m$. We can use Theorem 20 in § 9 to translate the abstract Kuhn–Tucker condition $(0, 0) \in \partial K(\bar{x}, \bar{y})$ into a more explicit and familiar form. Writing

$$-K(\bar{x}, y) = k(y) + \psi_P(y),$$

where $-k(y) = f_0(\bar{x}) + y_1 f_1(\bar{x}) + \cdots + y_m f_m(\bar{x})$ and P is the nonnegative orthant in R^m, we see from the example following Theorem 20 that the condition $0 \in \partial_y K(\bar{x}, \bar{y})$ holds if and only if the vector

$$-\nabla k(\bar{y}) = (f_1(\bar{x}), \cdots, f_m(\bar{x}))$$

is normal to P at $\bar{y}$. This means that

(10.2) $$\bar{y}_i \geqq 0, \quad f_i(\bar{x}) \leqq 0, \quad \bar{y}_i f_i(\bar{x}) = 0 \quad \text{for} \quad i = 1, \cdots, m.$$

On the other hand, for $\bar{y} \in P$ we have

(10.3) $$\partial_x K(x, \bar{y}) = \partial f_0(\bar{x}) + \partial(\bar{y}_1 f_1)(\bar{x}) + \cdots + \partial(\bar{y}_m f_m)(\bar{x})$$

by Theorem 20, assuming for example that f_i is a continuous function for $i = 1, \cdots, m$ (as follows from Corollary 8A if $X = R^n$). Moreover, for $\bar{y}_i \geqq 0$ and f_i finite one has trivially the relation

(10.4) $$\partial(\bar{y}_i f_i)(x) = \{\bar{y}_i v \mid v \in \partial f_i(x)\} \triangleq \bar{y}_i \partial f_i(x).$$

Formula (10.3) thus allows us to write the condition $0 \in \partial_x K(\bar{x}, \bar{y})$ as

(10.5) $$0 \in \partial f_0(\bar{x}) + \bar{y}_1 \partial f_1(\bar{x}) + \cdots + \bar{y}_m \partial f_m(\bar{x}).$$

The abstract Kuhn–Tucker condition is therefore equivalent to conditions (10.2) and (10.5), assuming f_i is continuous for $i = 1, \cdots, m$. If f_i is differentiable, (10.5) reduces of course to

(10.6) $$\nabla f_0(\bar{x}) + \bar{y}_1 \nabla f_1(\bar{x}) + \cdots + \bar{y}_m \nabla f_m(\bar{x}) = 0.$$

Example 6. (Stochastic programming.) The given problem is

(P) $$\text{minimize} \quad J(x) = \int_S h(x, s)\sigma(ds) \quad \text{over all} \quad x \in C,$$

and the chosen representation involves

(10.7) $$F(x, u) = \int_S h(x - u(s), s)\sigma(ds) + \psi_C(x), \qquad u \in U,$$

where h is measurable on $X \times S$ (relative to the Borel structure on X) and convex in the X argument, and C is a nonempty closed subset of X. In order to apply the theory of integral functionals in § 9 at its fullest, we shall assume X is a separable, reflexive Banach space (whose dual is identifiable with V under the pairing $\langle x, v \rangle$), and

$$U = \mathscr{L}_X^\infty(S, \Sigma, \sigma) \quad \text{and} \quad Y = \mathscr{L}_V^1(S, \Sigma, \sigma),$$

where the measure space is complete and has, of course, $\sigma(S) = 1$. (In the pairing between U and Y, we take the norm topology on Y, but the Mackey topology on U; see the end of § 7.) We shall assume further that $h(x, s)$ is finite and lower-semicontinuous in x for all s, and $h(u(s), s)$ is summable in s for every $u \in U$. (If $X = R^n$, the latter assumption can be weakened to the summability of $h(x, s)$ in s for each $x \in X$.)

These assumptions imply via Theorem 22 that F is a closed convex function on $X \times U$, and indeed $F(x, u)$ is finite and continuous in u in the Mackey topology for each x.

The Lagrangian function can be calculated as

$$K(x, y) = \inf_{u \in U} \left\{ \psi_C(x) + \int_S h(x - u(s), s)\sigma(ds) + \int_S \langle u(s), y(s)\rangle \sigma(ds) \right\}$$

$$= \inf_{w \in U} \left\{ \psi_C(x) + \int_S h(w(s), s)\sigma(ds) + \int_S \langle x - w(s), y(s)\rangle \sigma(ds) \right\},$$

which is to say

$$K(x, y) = \begin{cases} \left\langle x, \int_S y(s)\sigma(ds) \right\rangle - \int_S h^*(y(s), s)\sigma(ds) & \text{if } x \in C, \\ -\infty \quad \text{if } x \notin C. \end{cases} \tag{10.8}$$

This yields in turn

$$\begin{aligned} G(y, v) &= \inf_{x \in X} \{K(x, y) - \langle x, v\rangle\} \\ &= \psi_C^*\left(v - \int_S y(s)\sigma(ds)\right) - \int_S h^*(y(s), s)\sigma(ds). \end{aligned} \tag{10.9}$$

The dual problem is thus:

$$\begin{aligned} &\text{maximize} \quad -\psi_C^*\left(-\int_S y(s)\sigma(ds)\right) - \int_S h^*(y(s), s)\sigma(ds) \\ &\qquad \text{over all} \quad y \in \mathscr{L}_Y^1(S, \Sigma, \sigma). \end{aligned} \tag{D}$$

The continuity property of F is sufficient according to Theorem 18 (a) for all the conclusions of Theorem 17 to hold. In particular, we have

$$\inf(\mathrm{P}) = \max(\mathrm{D}), \tag{10.10}$$

the maximum being attained at $\bar{y}$ if and only if $-\bar{y} \in \partial\varphi(0)$ (Theorem 16).

To gain an interpretation of such vectors $\bar{y}$, let us investigate the significance of the relations

$$-\bar{y} \in \partial\varphi(0) \quad \text{and} \quad F(\bar{x}, 0) = \varphi(0), \tag{10.11}$$

which are equivalent to the Kuhn–Tucker saddle-point condition (Theorem 15). We can write (10.11) as

$$F(x, u) \geqq F(\bar{x}, 0) + \langle u - 0, -\bar{y}\rangle \quad \text{for all } x, u, \tag{10.12}$$

or more specifically in this case: $\bar{x} \in C$ and in addition

$$\int_S [h(x - u(s), s) + \langle u(s), \bar{y}(s)\rangle]\sigma(ds) \geqq \int_S h(\bar{x}, s)\sigma(ds) \tag{10.13}$$

for all $x \in X$ and $u \in \mathscr{L}_X^x$ such that $x - u(s) \in C$ for almost every s.

Consider the following modified problem, where $\bar{y}(s)$ is a "price vector" depending on the random element s. We must still choose x before s is observed, but after the observation we have the recourse of altering x to any more advantageous value $w(s) \in C$ based on our knowledge of s. In doing this we incur the cost $h(w(s), s)$, but also pay the amount $\langle x - w(s), \bar{y}(s)\rangle$, which represents the cost of alteration. The condition above requires $\bar{x}$ and the price structure $s \to \bar{y}(s)$ to be such that under these circumstances, no matter what value of s is observed, it will turn out that no a posteriori alteration of $\bar{x}$ will be advantageous. In other words, the function $\bar{y}$ is an "equilibrium price structure" which exactly reflects in the economic sense the disadvantage of having to choose x without complete information about s.

The Kuhn–Tucker condition in subgradient form is comprised of the two relations $0 \in \partial_x K(\bar{x}, \bar{y})$ and $0 \in \partial_y K(\bar{x}, \bar{y})$, the first of which reduces to:

$$\text{the vector} \quad -\int_S \bar{y}(s)\sigma(ds) \quad \text{is normal to } C \text{ at } \bar{x} \tag{10.14}$$

(see the example in §6 after Corollary 12B). Note that the integral gives the "expected price vector" in the interpretation above. Using the identity

$$\left\langle x, \int_S y(s)\sigma(ds)\right\rangle = \int_S \langle x, y(s)\rangle\sigma(ds), \tag{10.15}$$

we can write the second relation as $\bar{x} \in \partial H(\bar{y})$, where $\bar{x}$ is regarded as a constant function, an element of $\mathscr{L}_X^\infty$, and

$$H(y) = \int_S h^*(y(s), s)\sigma(ds). \tag{10.16}$$

According to Corollary 12A and Theorem 22, this is equivalent to having $\bar{x} \in \partial h^*(\bar{y}(s), s)$ for almost every s, as well as to having

$$\bar{y}(s) \in \partial h(\bar{x}, s) \quad \text{for almost every } s. \tag{10.17}$$

The Kuhn–Tucker condition for our problems can therefore be expressed as (10.14) and (10.17).

In computing a solution to (P), it may be helpful to know the subgradients of the function J. According to Theorem 23, the subgradients of J at x are the "expected price vectors"

$$\int_S y(s)\sigma(ds)$$

corresponding to "price systems" y which satisfy $y(s) \in \partial h(x, s)$ almost everywhere.

In particular, if x is such that the gradient $\nabla h(x, s)$ exists for almost every s, then $\nabla J(x)$ also exists and

$$\nabla J(x) = \int_S \nabla h(x, s)\sigma(ds). \tag{10.18}$$

The expectation (10.18) can in some applications be calculated approximately by sampling the random variable s.

For more about such ideas in stochastic programming, see [26] and the references given there.

Example 7′. (Stochastic programming.) For simplicity, we shall make the assumptions listed in Example 7 in § 2, namely that $X = R^n = V$, $W = R^d$, $f_i(x, w, s)$ is finite, convex in (x, w) and measurable in s, $i = 0, 1, \cdots, m$. Then, as already remarked, the integrand $H(x, u, s)$ in (2.23) is measurable in (x, u, s) and convex in (x, u). For the sake of the theorems in § 9 on integral functionals, we also want the probability space (S, Σ, σ) complete and $H(x, u, s)$ lower-semicontinuous in x. A convenient assumption ensuring lower-semicontinuity is that the set D be compact (the functions f_i being continuous in view of Corollary 8A and the preceding assumptions).

We shall also suppose in fact that $f_0(x, w, s)$ is summable in s for each (x, w), while $f_i(x, w, s)$ for $i = 1, \cdots, m$ is bounded in s for each (x, w). Let $U = \mathscr{L}^\infty_{R^m}(S, \Sigma, \sigma)$ and $Y = \mathscr{L}^1_{R^m}(S, \Sigma, \sigma)$ in the natural pairing (with the norm topology on Y and the weak or Mackey topology, say, on U). Then the convex functional

$$F(x, u) = \int_S H(x, u(s), s)\sigma(ds) + \psi_C(x) \tag{10.19}$$

has the property that for every $x \in C$ there exists $u \in U$ with $F(x, u) < +\infty$. Using this, we can apply Theorem 21 to calculate the Lagrangian function:

$$\begin{aligned} K(x, y) &= \inf_{u\in U} \{F(x, u) + \langle u, y\rangle\} \\ &= \inf_{u\in U} \left\{\psi_C(x) + \langle u, y\rangle - \int_S q_x(u(s), s)\sigma(ds)\right\} \\ &= \begin{cases} \int_S q_x^*(y(s), s)\sigma(ds) & \text{if} \quad x \in C, \\ +\infty \quad \text{if} \quad x \notin C, \end{cases} \end{aligned} \tag{10.20}$$

where $q_x(u, s) = -H(x, u, s)$ and q_x^* is the conjugate in the concave sense. Moreover,

$$\begin{aligned} q_x^*(y, s) &= \inf_{u\in R^m} \{H(x, u, s) + \langle u, y\rangle\} \\ &= \inf_{u\in R^m} \inf_{w\in D} \left\{f_0(x, w, s) + \sum_{i=1}^m u_i y_i \,|\, f_i(x, w, s) \leqq u_i, i = 1, \cdots, m\right\} \\ &= \inf_{w\in D} \inf_{u\in R^m} \left\{f_0(x, w, s) + \sum_{i=1}^m u_i y_i \,|\, f_i(x, w, s) \leqq u_i, i = 1, \cdots, m\right\}, \end{aligned} \tag{10.21}$$

where the inner infimum is

$$(10.22) \qquad k_0(x, w, y, s) = \begin{cases} f_0(x, w, s) + \sum_{i=1}^{m} y_i f_i(x, w, s) & \text{if} \quad y \geqq 0, \\ -\infty \quad \text{if} \quad y \ngeq 0. \end{cases}$$

Therefore, setting

$$(10.23) \qquad k(x, y, s) = \min_{w \in D} k_0(x, w, y, s)$$

we can state that

$$(10.24) \qquad K(x, y) = \begin{cases} \int_S k(x, y(s), s)\sigma(ds) & \text{if} \quad x \in C, \\ +\infty \quad \text{if} \quad x \notin C. \end{cases}$$

This formula leads to the conclusion that $F(x, u)$ is a *closed proper* convex function of $(x, u) \in X \times U$, so that the duality theory in § 4 and § 7 can be invoked symmetrically. To see this, observe first that the integral in (10.24) is $-\infty$ unless $y \geqq 0$ (i.e., $y(s) \geqq 0$ almost everywhere). On the other hand, if $y \geqq 0$, then $k(x, y(s), s)$ is summable in s; this can be demonstrated from the compactness of D and our assumptions on the functions f_i, but we omit the details. It follows that the integral in (10.24) is continuous as a function of $x \in R^n$ if $y \geqq 0$ (Theorem 23 and the remark after it). From (10.20) and part (c) of Theorem 21, we have

$$(10.25) \qquad \begin{aligned} F(x, u) &= \sup_{y \in Y} \{K(x, y) - \langle u, y \rangle\} \\ &= \sup_{y \geqq 0} \left\{\psi_C(x) + \int_S k(x, y(s), s)\sigma(ds) - \langle u, y \rangle\right\}. \end{aligned}$$

The bracketed expression is for each $y \geqq 0$ in $\mathscr{L}^1_{R^m}$ a closed proper convex function of (x, u). We see then that F, as the supremum of a collection of such functions, is closed convex and nowhere $-\infty$. Since we already know $F \not\equiv +\infty$, we conclude F is actually proper.

An explicit form of the subgradient relations in the Kuhn–Tucker condition can be obtained by a fancy calculation employing both parts of Theorem 24 as well as many of the other results in § 9. We only state the outcome: *$\bar{x} \in R^n$ and $\bar{y} \in \mathscr{L}^1_{R^m}$ satisfy the Kuhn–Tucker condition if and only if there exist functions $\bar{p} \in \mathscr{L}^1_{R^n}$, $\bar{w} \in \mathscr{L}^\infty_{R^d}$ and $\bar{r} \in \mathscr{L}^\infty_{R^d}$ such that the following relations all hold*:

$$(10.26) \qquad \bar{y}_i(s) \geqq 0, \quad f_i(\bar{x}, \bar{w}(s), s) \leqq 0, \quad \bar{y}_i(s) f_i(\bar{x}, \bar{w}(s), s) = 0$$

for $i = 1, \cdots, m$ *almost everywhere*;

$$(10.27) \qquad (\bar{p}(s), \bar{r}(s)) \in \left[\partial f_0(\bar{x}, \bar{w}(s), s) + \sum_{i=1}^{m} \bar{y}_i(s)\, \partial f_i(\bar{x}, \bar{w}(s), s)\right]$$

almost everywhere;

(10.28) $-\bar{r}(s)$ *is normal to D at the point* $\bar{w}(s) \in D$

almost everywhere;

(10.29) *the vector* $-\int_S \bar{p}(s)\sigma(ds)$ *is normal to C at the point* $\bar{x} \in C$.

These relations are thus *sufficient* for $\bar{x}$ to solve the primal problem and $\bar{y}$ to solve the dual.

But are the relations necessary? Can every $\bar{x}$ solving (P) be characterized this way? A positive answer to this hinges on being able to show that inf (P) = max (D). If C is bounded, say, we do have min (P) = sup (D) by Theorem 17′ and criterion (d) of Theorem 18′. However, it seems unlikely that inf (P) = max (D) can be established by any of the criteria in Theorem 18. The difficulty is that the optimal value function φ on $U = \mathscr{L}^\infty_{R^m}$ can hardly be expected to be bounded above around 0 in the Mackey topology (or any other topology compatible with the pairing with $Y = \mathscr{L}^1_{R^m}$), due to the compactness of D and the nature of integral functionals.

The fact that the sufficient conditions (10.26)–(10.29) may not be necessary for optimality can also be seen heuristically. Much in the pattern of Example 6′, one can interpret $\bar{y}_i(s)$ as a "price" associated with the constraint $f_i(x, w, s) \leqq 0$. The existence of a $\bar{y}$ of the type described means that an "equilibrium price structure" is possible which distributes the economic effects of the constraints over a subset of S of positive measure. However, there may be only certain crucial values of the random variable s for which the induced constraint on $\bar{x}$:

$$\text{(10.30)} \qquad \text{there exists } w \in D \text{ with } f_i(\bar{x}, w, s) \leqq 0, \qquad i = 1, \cdots, m,$$

is "tight," and these may form a set of measure 0 with respect to σ. The effects of such a constraint cannot be reflected by a function $\bar{y}$, i.e., by a measure on S of the form

$$\text{(10.31)} \qquad \mu(T) = \int_T \bar{y}(s)\sigma(ds).$$

The way out of this difficulty is apparently to generalize the dual problem by admitting measures μ which are not just of the form (10.31), but may have singularities with respect to σ. This returns us to the ideas raised in § 9 after Theorem 21: one can try to work instead with U as the space of all continuous R^m-valued functions on S (the latter assumed to have a compact topological structure) and Y as the space of R^m-valued measures on S.

Example 8′. (Calculus of variations.) The problem is

$$\text{(P)} \qquad \text{minimize} \quad \int_0^1 L(x(t), \dot{x}(t), t)\, dt + l(x(0), x(1)),$$

and we have represented it parametrically by

$$\text{(10.32)} \qquad F(x, u) = \int_0^1 L(x(t) - z(t), \dot{x}(t), t)\, dt + l(x(0) - a, x(1)),$$

where $x \in X = \mathscr{A}_n[0, 1]$ and $u = (z, a) \in U = \mathscr{L}^\infty_n[0, 1] \times R^n$. For the sake of the theory of integral functionals, it is assumed that L is measurable on $R^n \times R^n \times [0, 1]$

relative to the Borel structure on $R^n \times R^n$ and the Lebesgue structure on $[0, 1]$, and that the functions l and $L(\cdot, \cdot, t)$ are l.s.c. convex on $R^n \times R^n$ and nowhere $-\infty$. Let X be paired with $V = \mathscr{L}_n^\infty[0, 1] \times R^n$ by

$$\langle x, (w, b)\rangle = \int_0^1 \dot{x}(t) \cdot w(t)\, dt - x(1) \cdot b, \tag{10.33}$$

and let U be paired with $Y = \mathscr{A}_n[0, 1]$ similarly,

$$\langle (z, a), y\rangle = \int_0^1 z(t) \cdot \dot{y}(t)\, dt + a \cdot y(0). \tag{10.34}$$

The norm topology is taken on V and U and the induced weak topology on X and Y.

Consider the integral functional

$$I(r_1, r_2) = \int_0^1 L(t, r_1(t), r_2(t))\, dt, \qquad (r_1, r_2) \in \mathscr{L}_n^\infty \times \mathscr{L}_n^1. \tag{10.35}$$

Assume $I \not\equiv +\infty$. Pairing the decomposable space $\mathscr{L}_n^\infty \times \mathscr{L}_n^1$ with $\mathscr{L}_n^1 \times \mathscr{L}_n^\infty$ by

$$\langle (r_1, r_2), (q_1, q_2)\rangle = \int_0^1 r_1(t) \cdot q_1(t)\, dt + \int_0^1 r_2(t) \cdot q_2(t)\, dt,$$

we can employ Theorem 21 to calculate the conjugate of I, obtaining

$$I^*(q_1, q_2) = \int L^*(t, q_1(t), q_2(t))\, dt, \quad (q_1, q_2) \in \mathscr{L}_n^1 \times \mathscr{L}_n^\infty. \tag{10.36}$$

This fact enables us to determine an expression for the dual problem (D) in terms of the conjugates L^* and l^*. Specifically, we have

$$\begin{aligned} G(y, (w, c)) &= \inf_{x, (z, a)} \{F(x, (z, a)) - \langle x, (w, b)\rangle + \langle (z, a), y\rangle\} \\ &= -\int_0^1 L^*(\dot{y}(t), y(t) + w(t), t)\, dt - l^*(y(0), -y(1) - b). \end{aligned} \tag{10.37}$$

The dual problem is thus

$$\text{minimize} \quad -\int L^*(\dot{y}(t), y(t))\, dt - l^*(y(0), -y(1)). \tag{D}$$

Remarkably, this has the same form as (P) and even the same sort of parametric representation.

The duality between these two problems (P) and (D) has been studied extensively (see [21], [22] and the references given there). It turns out that the Kuhn–Tucker condition can be expressed as the two conditions

$$(\dot{y}(t), y(t)) \in \partial L(x(t), \dot{x}(t), t) \quad \text{a.e.}, \tag{10.38}$$

$$(y(0), -y(1)) \in \partial l(x(0), x(1)). \tag{10.39}$$

The first of these can also be written in so-called Hamiltonian form:

$$-\dot{x}(t) \in \partial_y H(x(t), y(t), t) \quad \text{and} \quad \dot{y}(t) \in \partial_x H(x(t), y(t), t) \quad \text{a.e.,} \tag{10.40}$$

where H (the *Hamiltonian* function) is defined by

$$H(x, y, t) = \sup_{r \in R^n} \{y \cdot r - L(x, r, t)\}. \tag{10.41}$$

Note that $H(x, y, t)$ is concave in x and convex in y, since $L(x, r, t)$ is convex in (x, r) (Theorem 6).

Although these conditions can be shown to be necessary as well as sufficient for optimality in many important cases, one runs into difficulties similar to those in Example 7 when state constraints on $x(t)$ are implicitly present. This may be gotten around by generalizing the dual problem (and Kuhn–Tucker condition) so as to allow for functions y which are not necessarily absolutely continuous, but only of bounded variation [22].

Example 9′. (Optimal control.) Taking $U = \mathscr{L}_m^\infty \times R^r$ paired with $Y = \mathscr{L}_m^1 \times R^r$, we obtain under broad measurability and summability assumptions resembling those in Example 7′ that

$$\begin{aligned} K(x, y) = {} & \int_0^1 \left[f_0 + \sum_{i=1}^m y_i(s) f_i\right](x(t), \dot{x}(t) - A(t)x(t), t)\, dt \\ & + \left[l_0 + \sum_{j=1}^r y_j' l_j\right](x(0), x(1)) \quad \text{(finite)} \end{aligned} \tag{10.42}$$

if $y_i(s) \geqq 0$ a.e. and $y_j' \geqq 0$, while otherwise $K(x, y) = -\infty$. For fixed y, the problem of minimizing $K(x, y)$ in x thus has the character of Example 8′ and can be analyzed in those terms. Moreover, since $K(x, y)$ is finite in X for $y \geqq 0$ the latter problem is free of implicit constraints.

The idea can be used in a dual approach to solving the control problem, as explained in a more general context in § 1. The basic pattern is that, given $y^k \in Y$ with $y^k \geqq 0$ we minimize the expression (10.42) with respect to x to obtain x^k; then y^k is somehow altered to a new element y^{k+1} as suggested by the information which has been computed. One aims to have x^k and y^k converge respectively to an $\bar{x}$ solving (P) and a $\bar{y}$ with the properties in Theorem 16.

If essential "state constraints" are present among (or induced by) the constraints $f_i(x, w, t) \leqq 0$, the set-up is not adequate for duality theorems of the sharpest sort. Again it is necessary to generalize the dual problem and Lagrangian function to admit "singular" elements: cf. Examples 7′ and 8′.

Example 10′. (Partial differential equations.) This problem is too complicated technically for us to treat here. But the following heuristic comment may be illuminating. If the pattern of Example 8′ is pursued, one obtains as part of the Kuhn–Tucker condition a generalization of the Hamiltonian equations (10.40) in the form

$$(-\operatorname{div} y, \operatorname{grad} x) \in \partial H(x, y, t), \tag{10.43}$$

where again H is related to L by (10.41). This can be regarded as a generalized partial differential "equation." Associated with it is a generalized boundary condition expressed by the part of the Kuhn–Tucker condition corresponding to the function l.

For instance, if L is given by (2.37), we have $H(x, y, t) = \frac{1}{2}|y|^2$, so that (10.43) reduces to the classical equation

$$\operatorname{div} \operatorname{grad} x = 0 \tag{10.44}$$

with $y = \operatorname{grad} x$.

The notions we have discussed above, about extending problems to include "ideal" or "singular" elements as suggested specifically by the theory of integral functionals, are also relevant to such problems involving partial differential equations.

For a rigorous treatment of this class of problems in terms of convexity and duality, see [3].

REFERENCES

[1] E. Asplund and R. T. Rockafellar, *Gradients of convex functions*, Trans. Amer. Math. Soc., 139 (1969), pp. 443–467.

[2] H. Brézis, *Monotonicity methods in Hilbert spaces and some applications to nonlinear partial differential equations*, Contributions to Nonlinear Functional Analysis, E. H. Zarantonello, ed., Academic Press, New York, 1971, pp. 101–156.

[3] I. Ekeland and R. Temam, *Analyse Convexe et Problèmes Variationelles*, Hermann, Paris, 1973.

[4] E. G. Golshtein, *Convex Programming: Elements of the Theory*, Amer. Math. Soc. Transl., 1972.

[5] R. B. Holmes, *A Course in Optimization and Best Approximation*, Springer-Verlag Lecture Note Series, 1972.

[6] A. D. Ioffe and V. L. Levin, *Subdifferentials of convex functions*, Trans. Moscow Math. Soc., 26 (1972), pp. 3–73.

[7] A. D. Ioffe and V. M. Tikhomirov, *Duality of convex functions and extremum problems*, Russian Math. Surveys, 23 (1968), pp. 53–124.

[8] J. L. Kelley and I. Namioka, *Linear Topological Spaces*, Van Nostrand, New York, 1963.

[9] P.-J. Laurent, *Approximation et Optimisation*, Hermann, Paris, 1972.

[10] G. Minty, *Monotone (nonlinear) operators in Hilbert space*, Duke Math. J., 29 (1962), pp. 341–346.

[11] J. J. Moreau, *Functionelles Convexes*, Lecture Notes, Séminaire sur les équations aux derivées partielles, Collège de France, Paris, 1967.

[12] ———, *Weak and strong solutions to dual problems*, Contributions to Nonlinear Functional Analysis, E. H. Zarantonello, ed., Academic Press, New York, 1971, pp. 181–214.

[13] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, N.J., 1969.

[14] ———, *Level sets and continuity of conjugate convex functions*, Trans. Amer. Math. Soc., 123 (1966), pp. 46–53.

[15] ———, *On the maximal monotonicity of subdifferential mappings*, Pacific J. Math., 33 (1970), pp. 209–216.

[16] ———, *Saddle points and convex analysis*, Differential Games and Related Topics, H. W. Kuhn and G. P. Szegö, eds., North-Holland, Amsterdam, 1971, pp. 109–128.

[17] ———, *Measurable dependence of convex sets and functions on parameters*, J. Math. Anal. Appl., 28 (1969), pp. 4–25.

[18] ———, *Integrals which are convex functions, II*, Pacific J. Math., 39 (1971), pp. 439–469.

[19] ———, *Convex integral functionals and duality*, Contributions to Nonlinear Functional Analysis, E. H. Zarantonello, ed., Academic Press, New York, 1971, pp. 215–236.

[20] ———, *Some convex programs whose duals are linearly constrained*, Nonlinear Programming, J. B. Rosen, et al., eds., Academic Press, New York, 1970, pp. 293–322.

[21] ———, *Conjugate convex functions in optimal control and the calculus of variations*, J. Math. Anal. Appl., 32 (1970), pp. 174–222.

[22] ———, *State constraints in convex problems of Bolza*, SIAM J. Control, 10 (1972), pp. 691–715.

[23] ———, *Augmented Lagrange multiplier functions and duality in nonconvex programming*, Ibid., 12 (1974), to appear.

[24] ———, *Convex algebra and duality in dynamic models of production*, Mathematical Models in Economics, J. Łoś and M. W. Łoś, eds., North-Holland, Amsterdam, 1973, pp. 351–378.

[25] J. Stoer and C. Witzgall, *Convexity and Optimization in Finite Dimensions I*, Springer-Verlag, Berlin, 1970.

[26] R. J. B. Wets, *Stochastic programs with fixed recourse: the equivalent deterministic program*, SIAM Rev., to appear.

[27] E. L. Peterson, *Geometric programming and some of its extensions*, Optimization and Design, Avriel, Rijckaert and Wilde, eds., Prentice-Hall, Englewood Cliffs, N.J., 1973.