Convex Optimization in Infinite Dimensional Spaces*

Sanjoy K. Mitter

Department of Electrical Engineering and Computer Science, The Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, USA
mitter@mit.edu

Summary. The duality approach to solving convex optimization problems is studied in detail using tools in convex analysis and the theory of conjugate functions. Conditions for the duality formalism to hold are developed which require that the optimal value of the original problem vary continuously with respect to perturbations in the constraints only along feasible directions; this is sufficient to imply existence for the dual problem and no duality gap. These conditions are also posed as certain local compactness requirements on the dual feasibility set, based on a characterization of locally compact convex sets in locally convex spaces in terms of nonempty relative interiors of the corresponding polar sets. The duality theory and related convex analysis developed here have applications in the study of Bellman–Hamilton Jacobi equations and Optimal Transportation problems. See Fleming–Soner [8] and Villani [9].

Keywords: Convexity, Optimization, Convex Conjugate Functions? Quantum Detection.

Introduction

The duality approach to solving convex optimization problems is studied in detail using tools in convex analysis and the theory of conjugate functions. Conditions for the duality formalism to hold are developed which require that the optimal value of the original problem vary continuously with respect to perturbations in the constraints only along feasible directions; this is sufficient to imply existence for the dual problem and no duality gap. These conditions are also posed as certain local compactness requirements on the dual feasibility set, based on a characterization of locally compact convex sets in locally convex spaces in terms of nonempty relative interiors of the corresponding polar sets. The duality theory and related convex analysis developed here have applications in the study of Bellman–Hamilton Jacobi equations and Optimal Transportation problems. See Fleming–Soner [8] and Villani [9].

* Support for this research was provided by the Department of Defense MURI Grant: Complex Adaptive Networks for Cooperative Control Subaward #03-132, and the National Science Foundation Grant CCR-0325774.

V.D. Blondel et al. (Eds.) Recent Advances in Learning and Control, LNCIS 371, pp. 161–179, 2008.
springerlink.com

1 Notation and Basic Definitions

This section assumes a knowledge of topological vector spaces and only serves to recall some concepts in functional analysis which are relevant for optimization theory. The extended real line $[-\infty, +\infty]$ is denoted by $\overline{R}$. Operations in $\overline{R}$ have the usual meaning with the additional convention that

$$(+\infty) + (-\infty) = (-\infty) + (+\infty) = +\infty$$

Let X be a set, $f\colon X \to \overline{R}$ a map from X into $[-\infty, +\infty]$. The *epigraph* of f is

$$\text{epi}f \triangleq \{(x,r) \in X \times R\colon r \geq f(x)\} \ .$$

The *effective domain* of f is the set

$$\text{dom}f \triangleq \{x \in X\colon f(x) < +\infty\} \ .$$

The function f is *proper* iff $f \not\equiv +\infty$ and $f(x) > -\infty$ for every $x \in X$. The *indicator* function of a set $A \subset X$ is the map $\delta_A\colon X \to \overline{R}$ defined by

$$\delta_A(x) = \begin{cases} +\infty & \text{if} \ \ x \notin A \\ 0 & \text{if} \ \ x \in A \end{cases} \ .$$

Let X be a vector space. A map $f\colon X \to \overline{R}$ is *convex* iff epi$f$ is a convex subset of $X \times R$, or equivalently iff

$$f(\epsilon x_1 + (1-\epsilon)x_2) \leq \epsilon f(x+1) + (1-\epsilon) f(x_2)$$

for every $x_1, x_2 \in X$ and $\epsilon \in [0,1]$. The *convex hull* of f is the largest convex function which is everywhere less than or equal to f; it is given by

$$\begin{aligned} \text{co}f(x) &= \sup\{f'(x)\colon f' \text{ is convex } X \to \overline{R}\ ,\ f' \leq f\} \\ &= \sup\{f'(x)\colon f' \text{ is linear } X \to \overline{R}\ ,\ f' \leq f\} \end{aligned}$$

Equivalently, the epigraph of co$f$ is given by

$$\text{epi}(\text{co}f) = \{(x,r) \in X \times R\colon (x,s) \in \text{coepi}f \ \ \text{for every } s > r\} \ ,$$

where coepi$f$ denotes the convex hull of epi$f$.

Let X be a topological space. A map $f\colon X \to \overline{R}$ is lower semicontinuous} (ℓsc) iff epi$f$ is a closed subset of $X \times R$, or equivalently iff $\{x \in X\colon f(x) \leq r\}$ is a closed subset of X for every $r \in R$. The map $f\colon X \to \overline{R}$ is ℓsc at x_0 iff given any $r \in (-\infty, f(x_0))$ there is a neighborhood N of x_0 such that $r < f(x)$ for every $x \in N$. the *lower semicontinuous hull* of f is the largest lower semicontinuous functional on X which everywhere minorizes f, i.e.

$$\ell\text{sc}f(x) = \sup\{f'(x)\colon f' \text{ is } \ell\text{sc}X \to \overline{R}\ ,\ f' \leq f\} = \liminf_{x' \to x} f'(x)$$

Equivalently, epi(ℓsc$f$) = cl(epi$f$) in $X \times R$.

A *duality* $\langle X, X^* \rangle$ is a pair of vector spaces X, X^* with a bilinear form $\langle \cdot, \cdot \rangle$ on $X \times X^*$ that is separating, i.e. $\langle x, y \rangle = 0 \forall y \in X^* \Rightarrow x = 0$ and $\langle x, y \rangle = 0 \forall x \in X \Rightarrow x = 0$. Every duality is equivalent to a Hausdorff locally convex space X paired with its topological dual space X^* under the natural bilinear form $\langle x, y \rangle \triangleq y(x)$ for $x \in X, y \in X^*$. We shall also write $xy \equiv \langle x, y \rangle \equiv y(x)$ when no confusion arises.

Let X be a (real) Hausdorff locally convex space (HLCS), which we shall always assume to be real. X^* denotes the topological dual space of X. The *polar* of a set $A \subset X$ and the *(pre-)polar* of a set $B \subset X^*$ are defined by[1]

$$A^0 \triangleq \{y \in X^* \colon \sup_{x \in A} \langle x, y \rangle \leq 1\}$$

$$^0B \triangleq \{x \in X^* \colon \sup_{y \in B} \langle x, y \rangle \leq 1\} \quad .$$

The *conjugate* of a functional $f \colon X \to \overline{R}$ and the *(pre-)conjugate* of a functional $g \colon X^* \to \overline{R}$ are defined by

$$f^* \colon X^* \to \overline{R} \colon y \mapsto X^* \sup_{x \in X} [\langle x, y \rangle - f(x)]$$

$$g^* \colon X \to \overline{R} \colon x \mapsto X \sup_{y \in Y} [\langle x, y \rangle - g(y)] \quad .$$

If X is a HLCS there are several topologies on X which are important. By τ we denote the original topology on X; by the definition of equicontinuity τ is precisely that topology which has a basis of 0-neighborhoods consisting of polars of equicontinuous subsets of X^*. The *weak topology* $w(X, X^*)$ is the weakest topology compatible with the duality $\langle X, X^* \rangle$, i.e. it is the weakest topology on X for which the linear functionals $x \mapsto \langle x, y \rangle, y \in X^*$ are continuous. Equivalently, $w(X, X^*)$ is the locally convex topology on X generated by the seminorms $x \mapsto |\langle x, y \rangle|$ for $y \in X^*$; it has a basis of 0-neighborhoods given by polars of finite subsets of X^*. The *Mackey topology* $m(X, X^*)$ on X is the strongest topology on X compatible with the duality $\langle X, X^* \rangle$[2]; it has a 0-neighborhood basis consisting of polars of all $w(X^*, X)$-compact convex[3] subsets of X^*. The *strong topology* $s(X, X^*)$ is the strongest locally convex topology on X that still has a basis consisting of $w(X, X^*)$-closed sets; it has as 0-neighborhood basis all $w(X, X^*)$-closed convex absorbing subsets of X, or equivalently all polars of $w(X^*, X)$-bounded subsets of X^*. We shall often write w, m, s for $w(X, X^*), m(X, X^*), s(X, X^*)$, and also w^* for $w(X^*, X)$. The strong topology need not be compatible with the duality $\langle X, X^* \rangle$. In general we have

[1] We use the convention $\sup \emptyset = -\infty, \inf \emptyset = +\infty$. Hence $\emptyset^0 = X^*$.

[2] A topology τ_0 on the vector space $X$is *compatible* with the duality $\langle X, X^* \rangle$ iff $(X, \tau_0)^* = X^*$, i.e. the space of all continuous linear functionals on X with the τ_0-topology may be identified with X^*.

[3] The word "convex" here may not be omitted unless X is a barrelled space. In general there may be $w(X^*, X)$-compact subsets of X^* whose closed convex hulls are not compact for the $w(X^*, X)$ topology.

$w(X,X^*) \subset \tau \subset m(X,X^*) \subset s(X,X^*)$. For a *convex* set A, however, it follows from the Hahn–Banach separation theorem that A is closed iff A is $w(X,X^*)$-closed iff A is $m(X,X^*)$-closed. More generally

$$w - \mathrm{cl}A = \mathrm{cl}A = m - \mathrm{cl}A \supset s - \mathrm{cl}A$$

when A is convex. Similarly, if a convex function $f\colon X \to \overline{R}$ is $m(X,X^*)$-ℓsc then it is ℓsc and even $w(X,X^*)$-lsc. It is also true that the bounded sets are the same for every compatible topology on X.

Let X be a HLCS and $f\colon X \to \overline{R}$. The *conjugate function* $f^*\colon X^* \to \overline{R}$ is convex and $w(X^*,X)$-ℓsc since it is the supremum of the $w(X^*,X)$-continuous affine functions $y \mapsto \langle x,y\rangle - f(x)$ over all $x \in \mathrm{dom}f$. Similarly, for ${}^*g\colon X \to R$ is convex and lsc. The conjugate functions f^*, ${}^*g$ never take on $-\infty$ values, unless they are identically $-\infty$ or equivalently $f \equiv +\infty$ or $g \equiv +\infty$. Finally, from the Hahn–Banach separation theorem it follows that

$$^*(f^*) = \ell\mathrm{scco}f \tag{1}$$

whenever f has an affine minorant, or equivalently whenever $f^* \equiv +\infty$; otherwise $\ell\mathrm{scco}f$ takes on $-\infty$ values and $f^* \equiv +\infty, {}^*(f^*) \equiv -\infty$.

The following lemma is very useful.

Lemma 1. *Let X be a HLCS, and let $f\colon X \to \overline{R}$. Then* $\mathrm{co}(\mathrm{dom}f) = \mathrm{dom}(\mathrm{co}f)$. *If $f^* \not\equiv +\infty$, then* $\mathrm{clcodom}f = \mathrm{cldom}^*(f^*)$.

A *barrelled space* is a HLCS X for which every closed convex absorbing set is a 0-neighborhood; equivalently, the $w(X^*,X)$-bounded sets in X^* are conditionally $w(X^*,X)$-compact. It is then clear that the $m(X,X^*)$ topology is the original topology, and the equicontinuous sets in X^* are the conditionally w^*-compact sets. Every Banach space or Frechet space is barrelled, by the Banach–Steinhaus theorem.

We use the following notation. If $A \supset X$ where X is HLCS, then $\mathrm{int}A$, $\mathrm{cor}A$, $\mathrm{ri}A$, $\mathrm{rcor}A$, $\mathrm{cl}A$, $\mathrm{span}A$, $\mathrm{aff}\,A$, $\mathrm{co}A$ denote the *interior* of A, the *algebraic interior* of A or *core*, the *relative interior* of A, the *relative core* or *algebraic interior* of A, the *closure* of A, the *span* of A, the *affine hull* of A, and the *convex hull* of A. By *relative interior* of A we mean the interior of A in the relative topology of X on $\mathrm{aff}\,A$; that is $x \in \mathrm{ri}A$ iff there is a 0-neighborhood N such that $(x+N)\cap \mathrm{aff}\,A \subset A$. Similarly, $x \in \mathrm{rcor}A$ iff $x \in A$ and $A - x$ absorbs $\mathrm{aff}\,A - x$, or equivalently iff $x + [0,\infty] \cdot A \supset A$ and $x \in A$. By *affine hull* of A we mean the smallest (not necessarily closed) affine subspace containing A; $\mathrm{aff}\,A = A + \mathrm{span}(A-A) = x_0 + \mathrm{span}(A - x_0)$ where x_0 is any element of A.

Let A be a subset of the HLCS X and B a subset of X^*. We have already defined $A^0, {}^0B$. In addition, we make the following useful definitions:

$$\begin{aligned} A^+ &\triangleq \{y \in X^*\colon \langle x,y\rangle \geq 0 \forall x \in A\} \\ A^- &\triangleq -A^+ = \{y \in X^*\colon \langle x,y\rangle \leq 0 \forall x \in A\} \\ A^\perp &\triangleq A^+ \cap A^- = \{y \in X^*\colon \langle x,y\rangle = 0 \forall x \in A\}\ . \end{aligned}$$

Similarly, for $B \subset X^*$ the sets ${}^+B, {}^-B, {}^\perp B$, are defined in X in the same way. Using the Hahn–Banach separation theorem it can be shown that for $A \subset X, {}^0(A^0)$ is the smallest closed convex set containing $A \cup \{0\}$; ${}^+(A^+) = {}^-(A^-)$ is the smallest closed convex cone containing A; and ${}^\perp(A^\perp)$ is the smallest closed subspace containing A. Thus, if A is nonempty[4] then

$$\begin{aligned} {}^0(A^0) &= \text{clco}(A \cup \{0\}) \\ {}^+(A^+) &= \text{cl}[0,\infty) \cdot \text{co}A \\ {}^\perp(A^\perp) &= \text{clspan}A \\ A + {}^\perp((A-A)^\perp) &= \text{cl aff}\, A \ . \end{aligned}$$

2 Some Results from Convex Analysis

A detailed study of convex functions, their relative continuity properties, their sub-gradients and the relation between relative interiors of convex sets and local equicontinuity of polar sets is presented in the doctoral dissertation of S.K. Young [1977], written under the direction of the present author. In this section, we cite the relevant theorems needed in the sequel. The proofs may be found in the above-mentioned reference.

Theorem 1. *Let X be a HLCS, $f\colon X \to \overline{\mathbb{R}}$ convex and M an affine subset of X with the induced topology, $M \supset \text{dom} f$.*

Let $f(\cdot)$ be bounded above on a subset C of X where $\text{ri}C \neq \emptyset$ and $\text{aff}\, C$ is closed with finite co-dimension in M. Then, $\text{rcorcodom} f \neq \emptyset$, $\text{co}f$ restricted to $\text{rcorcodom} f$ is continuous and $\text{aff dom} f$ is closed with finite co-dimension in M. Moreover, $f^* \equiv +\infty$ of $\exists x_0 \in X$, $r_0 > -f(x_0)$, such that $\{y \in X^* | f^*(y) - \langle x_0, y\rangle \le r_0\}$ is $w(X^*, X)/M^\perp$ locally bounded.

Proposition 1. *Let $f\colon X \to \overline{R}$ convex be a function on the HLCS X. The following are equivalent:*

(1) $y \in \partial f(x_0)$
(2) $f(x) \ge f(x_0) + \langle x - x_0, y\rangle \ \forall x \in X$
(3) x_0 solves $\inf_x [f(x) - xy]$, i.e. $f(x_0) - \langle x_0, y\rangle = \inf_x [f(x) - \langle x, y\rangle]$
(4) $f^*(y) = \langle x_0, y\rangle - f(x_0)$
(5) $x_0 \in \partial f^*(y)$ and $f(x_0) = {}^*(f^*)(x_0)$.

If $f(\cdot)$ is convex and $f(x_0) \in R$, then each of the above is equivalent to

(6) $f'(x_0\,; x) \ge \langle x, y\rangle \ \forall x \in X$.

Theorem 2. *Let $f\colon X \to \overline{R}$ convex be a function on the HLCS X, with $f(x_0)$ finite. Then the following are equivalent:*

(1) $\partial f(x_0) \neq \emptyset$
(2) $f'(x_0\,; \cdot)$ is bounded below on a 0-neighborhood in X, i.e. there is a 0-neighborhood N such that $\inf_{x \in N} f'(x_0\,; x) > -\infty$

[4] If $A = \emptyset$, then ${}^0(A^0) = {}^+(A^+) = {}^\perp(A^\perp) = \{0\}$.

(3) $\exists 0 - \text{nbhd } N,\ \delta > 0 \text{ st } \inf_{\substack{x \in N \\ 0 < t < \delta}} \frac{f(x_0 + tx) - f(x_0)}{t} > -\infty$

(4) $\liminf_{x \to 0} f'(x_0\,;\,x) > -\infty$

(5) $\liminf_{\substack{x \to 0+ \\ t \to 0}} \frac{f(x_0 + tx) - f(x_0)}{t} > -\infty$

(6) $\exists\ y \in X^* \text{ st } f(x_0 + x) - f(x_0) \geq \langle xy \rangle\ \forall x \in X.$

If X is a normed space, then each of the above is equivalent to:

(7) $\exists M > 0 \text{ st } f(x_0 + x) - f(x_0) \geq -M|x|\ \forall x \in X$

(8) $\exists M > 0\ , \epsilon > 0 \text{ st whenever } |x| \leq \epsilon\ ,\ f(x_0 + x) - f(x_0) \geq -M|x|$

(9) $\liminf_{|x| \to 0} \frac{f(x_0 + x) - f(x_0)}{|x|} > -\infty$

Definition 1. *The* recession function f_∞ *of a function* $f\colon X \to \overline{\mathbb{R}}$ *is defined to be*

$$f_\infty = \sup_{y \in \mathrm{dom} f^*} \langle x, y \rangle\ .$$

Proposition 2. *Let $f\colon X \to \overline{\mathbb{R}}$ be a convex* ℓsc *proper function on the HLCS X. Then $f_\infty(x)$ is given by each of the following:*

(1) $\min\{r \in R\colon (x, r) \in (\mathrm{epi} f)_\infty\}$

(2) $\sup_{a \in \mathrm{dom} f} \sup_{t > 0} [f(a + tx) - f(a)]/t$

(3) $\sup_{t>0}[f(a + tx) - f(a)]/t$ *for any fixed* $a \in \mathrm{dom} f$

(4) $\sup_{a \in \mathrm{dom} f} [f(a + x) - f(a)]$

(5) $\sup_{y \in \mathrm{dom} f^*} \langle x, y \rangle.$

In (1), the minimum is always attained (whenever it is not $+\infty$), since $(\mathrm{epi} f)_\infty$ *is a closed set.*

Theorem 3. *Let X be HLCS, $f\colon X \to \overline{R}$ convex. Assume* $\mathrm{riepi} f \neq \emptyset$. *Then $f(\,\cdot\,)$ is continuous relative to* $\mathrm{aff\,dom} f$ *on* $\mathrm{rcordom} f$, *and the following are equivalent for a point $x_0 \in X$:*

(1) $f(\,\cdot\,)$ *is relatively continuous at* $x_0 \in \mathrm{dom} f$

(2) $x_0 \in \mathrm{rcordom} f$

(3) $\mathrm{dom} f - x_0$ *absorbs* $x_0 - \mathrm{dom} f$

(4) $\forall x \in \mathrm{dom} f,\ \exists \epsilon > 0 \text{ st } (1 + \epsilon)x_0 - \epsilon x \in \mathrm{dom} f$

(5) $[\mathrm{dom} f - x_0]^- \subset [\mathrm{dom} f - x_0]^\perp \equiv \{y \in X^*\colon y \equiv$ *constant on* $\mathrm{dom} f$

(6) $[\mathrm{dom} f - x_0]^-$ *is a subspace*

(7) $\{y \in X^*\colon (f^*)_\infty(y) - x_0 y \leq 0\}$ *is a subspace*

(8) $x_0 \in \mathrm{dom} f$ *and* $\{y \in X^*\colon f^*(y) - x_0 y \leq r\}_\infty$ *for some* $r \geq -f(x_0)$

(9) $\partial f(x_0) \neq \emptyset$ *and* $(\partial f(x_0))_\infty$ *is a subspace*

(10) $\partial f(x_0)$ *is nonempty and* $w(X^*, \mathrm{aff\,dom} f - x_0)$*-compact.*

3 Duality Approach to Optimization

3.1 Introduction

The idea of duality theory for solving optimization problems is to transform the original problem into a "dual" problem which is easier to solve and which has the same value as the original problem. Constructing the dual solution corresponds to solving a "maximum principle" for the problem. This dual approach is especially useful for solving problems with difficult implicit constraints and costs (e.g. state constraints in optimal control problems), for which the constraints on the dual problem are much simpler (only explicit "control" constraints). Moreover the dual solutions have a valuable sensitivity interpretation: the dual solution set is precisely the subgradient of the change in minimum cost as a function of perturbations in the "implicit" constraints and costs.

Previous results for establishing the validity of the duality formalism, at least in the infinite-dimensional case, generally require the existence of a feasible interior point ("Kuhn–Tucker" point) for the implicit constraint set. This requirement is restrictive and difficult to verify. Rockafellar [5, Theorem 11] has relaxed this to require only continuity of the optimal value function. In this chapter we investigate the duality approach in detail and develop weaker conditions which require that the optimal value of the minimization problem varies continuously with respect to perturbations in the implicit constraints only along feasible directions (that is, we require relative continuity of the optimal value function); this is sufficient to imply existence for the dual problem and no duality gap. Moreover we pose the conditions in terms of certain local compactness requirements on the dual feasibility set, based on results characterizing the duality between relative continuity points and local compactness.

To indicate the scope of our results let us consider the Lagrangian formulation of nonlinear programming problems with generalized constraints. Let U, X be normed spaces and consider the problem

$$P_0 = \inf\{f(u) \colon u \in C, g(u) \le 0\}$$

where C is a convex subset of U, $f \colon C \to R$ is convex, and $g \colon C \to X$ is convex in the sense that

$$g(tu_1 + (1-t)u_2) \le tg(u_1) + (1-t)g(u_2) \ , \ u_1, u_2 \in C \ , \ t \in [0,1] \ .$$

We are assuming that X has been given the partial ordering induced by a nonempty closed convex cone Q of "positive vectors"; we write $x_1 > x_2$ to mean $x_1 - x_2 \in Q$. The dual problem corresponding to P_0 is well-known to be

$$D_0 = \sup_{y \in Q^+} \inf_{u \in C} [f(u) + \langle g(u), y \rangle] \ ;$$

this follows from (6) below by taking $L \equiv 0$, $x_0 = 0$, and

$$F(u,x) = \begin{cases} f(u) \text{ if } u \in C, g(u) \le x \\ +\infty \text{ otherwise.} \end{cases} \tag{2}$$

We also remark that it is possible to write

$$P_0 = \inf_u \sup_y (u, y) \qquad D_0 = \sup_y \inf_u (u, y)$$

where we have defined the Lagrangian function by

$$\ell(u,y) = \begin{cases} +\infty & \text{if } u \in C \\ f(u) - \langle g(u), y\rangle & \text{if } u \in C, y \in Q^- \\ -\infty & \text{if } u \in C, y \in Q^- \ . \end{cases}$$

In analyzing the problem P_0, we imbed it in the family of perturbed problems

$$P(x) = \inf\{f(u) \colon u \in C, g(u) \leq x\} \ .$$

It then follows that the dual problem is precisely the second conjugate of P_0 evaluated at 0: $D_0 =^* (P^*)(0)$. Moreover if there is no duality gap ($P_0 = D_0$) then the dual solution set is the subgradient $\partial P(0)$ of $P(\cdot)$ at 0. The following theorem summarizes the duality results for this problem.

Theorem 4. *Assume P_0 is finite. The following are equivalent:*

(1) $P_0 = D_0$ *and* D_0 *has solutions*
(2) $\partial P(0) \neq \emptyset$
(3) $\exists \hat{y} \in Q^+$ st $P_0 = \inf_{u \in C}[f(u) + \langle g(u), \hat{y}\rangle]$
(4) $\exists \epsilon > 0, M > 0$ st $f(u) \geq P_0 - M|x|$ *whenever* $u \in C, |x| \leq \epsilon, g(u) \leq x$.

If (1) is true then $\hat{u}$ is a solution for P_0 iff $\hat{u} \in C, g(u) \leq 0$, and there is a $\hat{y} \in Q^+$ satisfying

$$f(u) + \langle g(u), \hat{y}\rangle \geq f(\hat{u}) \forall u \in C \ ,$$

in which case complementary slackness holds, i.e. $\langle g(\hat{u}), \hat{y}\rangle = 0$, and $\hat{y}$ solves D_0.

Proof. This follows directly from Theorem 6 with F defined by (2). ∎

We remark here that criterion (4) is necessary and sufficient for the duality result (1) to hold, and it is crucial in determining how strong a norm to use on the perturbation space X (equivalently, how large a dual space X^* is required in formulating a well-posed dual problem).

The most familiar assumption which is made to insure that the duality results of Theorem 4 hold is the existence of a Kuhn–Tucker point:

$$\forall \overline{u} \in C \text{ st } -g(\overline{u}) \in\in Q$$

This is a very strong requirement, and again is often critical in determining what topology to use on the perturbation space X. More generally, we need only require that $P(\cdot)$ is continuous as 0. Rockafellar has presented the following result [5]: if U is the normed dual of a Banach space V, if X is a Banach space, if g is lower semicontinuous in the sense that

$$\text{epi} g \triangleq \{(u,x) \colon g(u) \leq x\}$$

is closed in $U \times X$ (e.g. if g is continuous), then the duality results of Theorem 4 hold whenever

$$0 \in \text{core}[g(C) + Q] \ .$$

In fact, it then follows that $P(\cdot)$ is continuous at 0. The following theorem relaxes this result to relative continuity and also provides a dual characterization in terms of local compactness requirements which are generally easier to verify.

Theorem 5. *Assume P_0 is finite. The following are equivalent:*

(1) $\text{aff}[g(C) + Q]$ *is closed; and* $0 \in \text{rcor}[g(C) + Q]$, *or equivalently* $\forall u \in C, \forall x \leq g(u) \exists \epsilon > 0$ *and* $u_1 \in C$ st $g(u_1) + \epsilon x \leq 0$.

(2) $Q^+ \cap g(C)^+$ *is a subspace M; and there is an $\epsilon > 0$, and $x_1 \in X$, and $r_1 \in R$ such that* $\{y \in Q^+ \colon \inf_{|v| \leq \epsilon} \sup_{u \in C} [f(u) + g(u)y - uv] > r_1\}$ *is nonempty and $w(X^*, X)/M$-locally bounded.*

If either of the above holds, then $P(\cdot)$ is relatively continuous at 0 and hence Theorem 4 holds. Moreover the dual solutions have the sensitivity interpretation

$$P'(0; x) = \max\{\langle x, y \rangle : y \text{ solves } D_0\}$$

where the maximum is attained and $P'(0; \cdot)$ denotes the directional derivative of the optimal value function $P(\cdot)$ evaluated at 0.

Proof. This follows directly from Theorem 9. ■

3.2 Problem Formulation

In this section we summarize the duality formulation of optimization problems. Let U be a HLCS of controls; X a HLCS of states; $u \mapsto Lu + x_0$ an affine map representing the system equations, where $x_0 \in X$, and $L \colon U \to X$ is linear and continuous; $F \colon U \times X \to \overline{R}$ a cost function. We consider the *minimization problem*

$$P_0 = \inf_{u \in U} F(u, Lu + x_0) \ , \tag{3}$$

for which feasibility constraints are represented by the requirement that $(u, Lu + x_0) \in \text{dom} F$. Of course, there are many ways of formulating a given optimization problem in the form (3) by choosing different spaces U, X and maps L, F; in general the idea is to put explicitly, easily characterized costs and constraints into the "control" costs on U and to put difficult implicit constraints and costs into the "state" part of the cost where a Lagrange multiplier representation can be very useful in transforming implicit constraints to explicit constraints. The dual variables, or multipliers will be in X^*, and the dual problem is an optimization in X^*.

In order to formulate the dual problem we consider a family of *perturbed problems*

$$P(x) = \inf_{u \in U} F(u, Lu + x) \tag{4}$$

where $x \in X$. Note that if $F: U \times X \to \overline{R}$ is convex then $P: X \to \overline{R}$ is convex; however F ℓsc does not imply that P is lsc. Of course $P_0 = P(x_0)$. We calculate the conjugate function of P

$$P^*(y) = \sup_x [\langle x, y\rangle - P(x)] = \sup_{ux,} [\langle x, y\rangle - F(u, Lu + x)] = F^*(-L^*y, y) \quad . \quad (5)$$

The dual problem of $P_0 = P(x_0)$ is given by the second conjugate of P evaluated at x_0, i.e.

$$D_0 =^* (P^*)(x_0) = \sup_{y \in X^*} [\langle x_0, y\rangle - F^*(-L^*y, y)] \quad (6)$$

The feasibility set for the dual problem is just $\mathrm{dom}P^* = \{y \in X^*: (-L^*y, y) \in \mathrm{dom}F^*\}$. We immediately have

$$P_0 \equiv P(x_0) \geq D_0 \equiv^* (P^*)(x_0) \quad . \quad (7)$$

Moreover, since the primal problem P_0 is an infimum, and the dual problem D_0 is a supremum, and $P_0 \geq D_0$, we see that if $\hat{u} \in U, \hat{y} \in X^*$ satisfy

$$F(\hat{u}, L\hat{u} + x_0) = \langle x_0, \hat{y}\rangle - F^*(-L^*\hat{y}, \hat{y}) \quad (8)$$

then $P_0 = D_0 = F(\hat{u}, L\hat{u} + x_0)$ and (assuming $P_0 \in R$) $\hat{u}$ is optimal for P, $\hat{y}$ is optimal for D. Thus, the existence of a $\hat{y} \in X^*$ satisfying (8) is a sufficient condition for optimality of a control $\hat{u} \in U$; we shall be interested in condition under which (8) is also necessary. It is also clear that any "dual control" $y \in X^*$ provides a lower bound for the original problem: $P_0 \geq \langle x_0, y\rangle - F^*(-L^*y, y)$ for every $y \in X^*$.

The duality approach to optimization problems P_0 is essentially to vary the constraints slightly as in the perturbed problem $P(x)$ and see how the minimum cost varies accordingly. In the case that F is convex, $P_0 = D_0$ or no "duality gap" means that the perturbed minimum costs function $P(\cdot)$ is ℓsc at x_0. The stronger requirement that the change in minimum cost does not drop off too sharply with respect to perturbations in the constraints, i.e. that the directional derivative $P'(x_0\,;\ \cdot)$ is bounded below on a neighborhood of x_0, corresponds to the situation that $P_0 = D_0$ and the dual problem D_0 has solutions, so that (8) becomes a necessary and sufficient condition for optimality of a control $\hat{u}$. It turns out that the solution of D_0 when $P_0 = D_0$ are precisely the element of $\partial P(x_0)$, so that the dual solutions have a sensitivity interpretation as the subgradients of the change in minimum cost with respect to the change in constraints.

Before stating the above remarks in a precise way, we define the Hamiltonian and Lagrangian functions associated with the problem P_0. We denote by $F_u(\cdot)$ the functional $F(u, \cdot): x \to F(u, x): X \to \overline{R}$, for $u \in U$. The *Hamiltonian* function $H: U \times X^* \to \overline{R}$ is defined by

$$H(u, y) = \sup_{x \in X} [\langle x, y\rangle - F(u, x)] = F_u^*(y) \quad . \quad (9)$$

Proposition 3. *The Hamiltonian H satisfies:*

(1) $({}^*H_u)(x) = {}^*(F_u^*)(x)$
(2) $({}^*H_u)^*(y) = H_u(y) = F_u^*(y)$
(3) $F^*(v,y) = \sup_u[\langle u,v\rangle + H(u,y)] = (-H(\cdot\,,y))^*(v)$.

Moreover $H(u,\cdot)$ *is convex and* $w^* - \ell\mathrm{sc}\ X^* \to \overline{R}$; $H(\cdot\,,y)$ *is concave* $U \to \overline{R}$ *if* F *is convex; if* $F(u,\cdot)$ *is convex, proper, and* $\ell\mathrm{sc}$ *then* $H(\cdot\,,y)$ *is concave for every* y *iff* F *is convex.*

Proof. The equalities are straightforward calculations. $H(u,\cdot)$ is convex and ℓsc since $({}^*H_u)^* = H_u$. It is straightforward to show that $-H(\cdot\,,y)$ is convex if $F(\cdot)$ is convex. On the other hand if $({}^*F_u)^* = F_u$ and $H(\cdot\,,y)$ is concave for every $y \in X^*$, then

$$F(u,x) = {}^*(F_u^*)(x) = {}^*H_u(x) = \sup_y[xy - H(u,y)]$$

is the supremum of the convex functionals $(u,x) \mapsto \langle x,y\rangle - H(u,y)$ and hence F is convex. ■

The *Lagrangian* function $\ell\colon U \times X^*$ to $\overline{R}$ is defined by

$$\begin{aligned}\ell(u,y) = \inf_x[F(u,Lu+x_0+x) - \langle x,y\rangle] &= \langle Lu+x_0,y\rangle - F_u^*(y)\\ &= \langle Lu+x_0,y\rangle - H(u,y)\ . \qquad (10)\end{aligned}$$

Proposition 4. *The Lagrangian ℓ satisfies*

(1) $\inf_u \ell(u,y) = \langle x_0,y\rangle - F^*(-L^*y,y)$
(2) $D_0 \equiv {}^*(P^*)(x_0) = \sup_y \inf_u \ell(u,y)$
(3) ${}^*(-\ell_u)(x) = {}^*(F_u^*)(Lu+x_0+x)$
(4) $P_0 \equiv P(x_0) = \inf_u \sup_y \ell(u,y)$ *if* $F_u = {}^*(F_u^*)$ *for every* $u \in U$.

Moreover $\ell(u,\cdot)$ *is convex and* w^**-*ℓsc $X^* \to \overline{R}$ *for every* $u \in U$; $\ell(\cdot)$ *is convex* $U \times X^* \to \overline{R}$ *if* F *is convex; if* $F_u = {}^*(F_u^*)$ *for every* $u \in U$ *then* ℓ *is convex iff* F *is convex.*

Proof. The first equality (1) is direct calculation; (2) then follows from (1) and (4). Equality (3) is immediate from (10); (4) then follows from (3) assuming that ${}^*(F_u^*) = F_u$. The final remarks follow from Proposition 3 and the fact that $\ell(u,y) = \langle Lu+x_0,y\rangle - H(u,y)$. ■

Thus from Proposition 4 we see that the duality theory based on conjugate functions includes the Lagrangian formulation of duality for inf-sup problems. For, given a Lagrangian function $\ell\colon U \times X^* \to \overline{R}$, we can define $F\colon U \times X \to \overline{R}$ by $F(u,x) = {}^*(-\ell_u)(x) = \sup_y[\langle x,y\rangle + \ell(u,x)]$, so that

$$\begin{aligned}P_0 &= \inf_u \sup_y \ell(u,y) = \inf_u F(u,0)\\ D_0 &= \sup_y \inf_u \ell(u,y) = \sup_y -F^*(0,y)\ ,\end{aligned}$$

which fits into the conjugate duality framework.

For the following we assume as before that U, X are HLCS's; $L: U \to X$ is linear and continuous; $x_0 \in X$; $F: U \times X \to \overline{R}$. We define the family of optimization problems $P(x) = \inf_u F(u, Lu+x)$, $P_0 = P(x_0)$, $D_0 = \sup_y[\langle x, y\rangle - F^*(-L^*y, y)] =^* (P^*)(x_0)$. We shall be especially interested in the case that $F(\cdot)$ is convex, and hence $P(\cdot)$ is convex.

Proposition 5. (no duality gap). *It is always true that*

$$P_0 \equiv P(x_0) \geq \inf_u \sup_y \ell(u,y) \geq D_0 \equiv \inf_u \sup_y \ell(u,y) \equiv^* (P^*)(x_0) \; . \qquad (11)$$

If $P(\cdot)$ is convex and D_0 is feasible, then the following are equivalent:

(1) $P_0 = D_0$

(2) $P(\cdot)$ *is* ℓsc *at* x_0, *i.e.* $\lim_{x \to x_0} \inf P(x) \geq P(x_0)$

(3) $\sup_{F \text{ finite} \subset X^*} \inf_{\substack{u \in U \\ x \in Lu + x_0 + {}^0F}} F(u,x) \geq P_0$

These imply, and are equivalent to, if $F_u =^* (F_u^*)$ for every $u \in U$,

(4) ℓ *has a saddle value, i.e.* $\inf_u \sup_y \ell(u,y) = \sup_y \inf_u \ell(u,y)$.

Proof. The proof is immediate since $P_0 = P(x_0)$ and $D_0 =^* (P^*)(x_0)$. Statement (4) follows from Proposition 4 and Eq. (11). ∎

Theorem 6. (no duality gap and dual solutions). *Assume P_0 is finite. The following are equivalent:*

(1) $P_0 = D_0$ *and* D_0 *has solutions*
(2) $\partial P(x_0) \neq \emptyset$
(3) $\exists \hat{y} \in Y \mathrm{st} P_0 = \langle x_0, \hat{y}\rangle - F^*(-L^*\hat{y}, \hat{y})$
(4) $\exists \hat{y} \in Y \mathrm{st} P_0 = \inf_u \ell(u, \hat{y})$.

If $P(\cdot)$ is convex, then each of the above is equivalent to

(5) $\exists 0 -$ neighborhood $N \mathrm{st} \inf_{x \in N} P'(x_0; x) > -\infty$
(6) $\liminf_{x \to 0} P'(x_0; x) > -\infty$
(7) $\liminf_{x \to 0 + t \to 0} \dfrac{P(x_0 + tx) - P_0}{t}$

$$\equiv \sup_{N = 0-\mathrm{nbhd}} \inf_{t > 0} \inf_{x \in N} \inf_{u \in U} \frac{F(u, Lu + x_0 + tx) - P_0}{t} > -\infty \; .$$

If $P(\cdot)$ is convex and X is a normed space, then the above are equivalent to:

(8) $\exists \epsilon > 0, M > 0$ st $F(u, Lu + x_0 + x) - P_0 \geq -M|x| \forall u \in U, |x| \leq \epsilon$.
(9) $\exists \epsilon > 0, M > 0$ st $\forall u \in U, |x| \leq \epsilon, \delta > 0 \exists u' \in U$ st $F(u, Lu + x_0 + x) - F(u', Lu' + x_0) \geq -M|x| - \epsilon$.

Moreover, if (1) is true then $\hat{y}$ solves D_0 iff $\hat{y} \in \partial P(x_0)$, and $\hat{u}$ is a solution for P_0 iff there is a $\hat{y}$ satisfying any of the conditions (1′)–(3′) *below. The following statements are equivalent:*

$(1')\hat{u}$ *solves* P_0, $\hat{y}$ *solves* D_0, *and* $P_0 = D_0$
$(2')F(\hat{u}, L\hat{u} + x_0) = \langle x_0, \hat{y}\rangle - F^*(-L^*\hat{y}, \hat{y})$
$(3')(-L^*\hat{y}, \hat{y}) \in \partial F(\hat{u}, L\hat{u} + x_0)$.

These imply, and are equivalent to, if $F(u, \cdot)$ *is proper convex* ℓsc $X \to \overline{R}$ *for every* $u \in U$, *the following equivalent statements:*

$(4')0 \in \partial\ell(\cdot, \hat{y})(\hat{u})$ *and* $0 \in \partial(-\ell(\hat{u}, \cdot))(\hat{y})$, *i.e.* $(\hat{u}, \hat{y})$ *is a saddlepoint of* ℓ, *that is* $\ell(\hat{u}, y) \leq \ell(\hat{u}, \hat{y}) \leq \ell(u, \hat{y})$ *for every* $u \in U, y \in X^*$).

$(5')L\hat{u} + x_0 \in \partial H(\hat{u}, \cdot)(\hat{y})$ *and* $L^*\hat{y} \in \partial(-H(\cdot, \hat{y}))(\hat{u})$, *i.e.* $\hat{y}$ *solves* $\inf_y[H(\hat{u}, y) - \langle L\hat{u} + x_0, y\rangle]$ *and* $\hat{u}$ *solves* $\inf_u[H(u, \hat{y}) + \langle u, L^*\hat{y}\rangle]$.

Proof. (1) => (2). Let $\hat{y}$ be a solution of $D_0 =^* (P^*)(x_0)$. Then $P_0 = \langle x_0, \hat{y}\rangle - P^*(\hat{y})$. Hence $P^*(\hat{y}) = \langle x_0, \hat{y}\rangle - P(x_0)$ and from Proposition 1, (4) => (1) we have $y \in \partial P(x_0)$.

(2) => (3). Immediate by definition of D_0.

(3) => (4) => (1). Immediate from (11).

If $P(\cdot)$ is convex and $P(x_0) \in R$, then (1) and (4)–(9) are all equivalent by Theorem 2. The equivalence of (1′)–(5′) follows from the definitions and Proposition 5. ∎

Remark. *In the case that* X *is a normed space, condition (8) of Theorem 6 provides a necessary and sufficient characterization for when dual solutions exists (with no duality gap) that shows explicitly how their existence depends on what topology is used for the space of perturbations. In general the idea is to take a norm as weak as possible while still satisfying condition (8), so that the dual problem is formulated in as nice a space as possible. For example, in optimal control problems it is well known that when there are no state constraints, perturbations can be taken in e.g. an* L_2 *norm to get dual solutions* y *(and costate* $-L^*y$*) in* L_2, *whereas the presence of state constraints requires perturbations in a uniform norm, with dual solutions only existing in a space of measures.*

It is often useful to consider perturbations on the dual problem; the duality results for optimization can then be applied to the dual family of perturbed problems. Now the dual problem D_0 *is*

$$-D_0 = \inf_{y \in X^*}[F^*(-L^*y, y) - \langle x_0, y\rangle \quad .$$

In analogy with (4) we define perturbations on the dual problem by

$$D(v) = \inf_{y \in X^*}[F^*(v - L^*y, y) - \langle x_0, y\rangle] \ , \quad v \in U^* \tag{12}$$

Thus $D(\cdot)$ *is a convex map* $U^* \to \overline{R}$, *and* $-D_0 = D(0)$. *It is straightforward to calculate*

$$({}^*D)(u) = \sup_v[\langle u, v\rangle - D(v)] =^* (F^*)(u, Lu + x_0) \quad .$$

Thus the "dual of the dual" is

$$-{}^*(D^*)(0) = \overset{*}{\inf_{u \in U}}(F^*)(u, Lu + x_0) \quad . \tag{13}$$

In particular, if $F =^* (F^*)$ *then the "dual of the dual" is again the primal, i.e.* dom^*D *is the feasibility set for* P_0 *and* $-^*(D^*)(0) = P_0$*. More generally, we have*

$$P_0 \equiv P(x_0) \geq -^*(D^*)(0) \geq D_0 \equiv -D(0) \equiv^* (P^*)(0) \ . \tag{14}$$

3.3 Duality Theorems for Optimization Problems

Throughout this section it is assumed that U, X are HLCS's; $L\colon U \to X$ is linear and continuous; $x_0 \in X$ and $F\colon U \times X \to \overline{R}$. Again, $P(x) = \inf_u F(u, Lu+x_0+x)$, $P_0 = P(x_0)$,

$$D_0 =^* (P^*)(x_0) = \sup_{y \in X^*} [\langle x_0, y\rangle - F^*(-L^*y, y)] \ .$$

We shall be interested in conditions under which $\partial P(x_0) \neq \emptyset$; for then there is no duality gap and there are solutions for D_0. These conditions will be conditions which insure the $P(\cdot)$ is relatively continuous at x_0 with respect to $\mathrm{aff\,dom}P$, that is $P \upharpoonright \mathrm{aff\,dom}P$ is continuous at x_0 for the induced topology on $\mathrm{aff\,dom}P$. We then have

$$\begin{aligned}
&\partial P(x_0) \neq \emptyset \\
&P_0 = D_0 \\
&\text{the solution set for } D_0 \text{ is precisely } \partial P(x_0) \\
&P'(x_0; x) = \max_{y \in \partial P(x_0)} \langle x, y\rangle \ .
\end{aligned} \tag{15}$$

This last result provides a very important sensitivity interpretation for the dual solutions, in terms of the rate of change in minimum cost with respect to perturbations in the "state" constraints and costs. Moreover if (15) holds, then Theorem 6, $(1')$–$(5')$, gives necessary and sufficient conditions for $\hat{u} \in U$ to solve P_0.

Theorem 7. *Assume* $P(\cdot)$ *is convex (e.g. F is convex). If* $P(\cdot)$ *is bounded above on a subset* C *of* X*, where* $x_0 \in \mathrm{ri}C$ *and* $\mathrm{aff}\,C$ *is closed with finite codimension in an affine subspace* M *containing* $\mathrm{aff\,dom}P$*, then (15) holds.*

Proof. From Theorem 1, (1b) => (2b), we know that $P(\cdot)$ is relatively continuous at x_0. ■

Corollary 1. (Kuhn–Tucker point). *Assume* $P(\cdot)$ *is convex (e.g. F is convex). If there exists a* $\overline{u} \in U$ *such that* $F(\overline{u}, \cdot)$ *is bounded above on a subset* C *of* X*, where* $L\overline{u} + x_0 \in \mathrm{ri}\ C$ *and* $\mathrm{aff}\,C$ *is closed with finite codimensions in an affine subspace* M *containing* $\mathrm{aff\,dom}\ P$*, then (15) holds. In particular, if there is a* $\overline{u} \in U$ *such that* $F(\overline{u}, \cdot)$ *is bounded above on a neighborhood of* $L\overline{u} + x_0$*, then (15) holds.*

Proof. Clearly

$$P(x) = \inf_u F(u, Lu + x) \leq F(\overline{u}, L\overline{u} + x) \ ,$$

so Theorem 1 applies. ■

The Kuhn–Tucker condition of Corollary 1 is the most widely used assumption for duality [4]. The difficulty in applying the more general Theorem 7 is that, in cases where $P(\,\cdot\,)$ is not actually continuous but only relatively continuous, it is usually difficult to determine aff dom P. Of course,

$$\text{dom}P = \bigcup_{u \in U} [\text{dom}F(u,\,\cdot\,) - Lu] \ ,$$

but this may not be easy to calculate. We shall use Theorem 1 to provide dual compactness conditions which insure that $P(\,\cdot\,)$ is relatively continuous at x_0.

Let K be a convex balanced $w(U,U^*)$-compact subset of U; equivalently, we could take $K = {}^0N$ where N is a convex balanced $m(U^*,U)$-0-neighborhood in U^*. Define the function $g\colon X^* \to \overline{R}$ by

$$g(y) = \inf_{v \in K^0} F^*(v - L^*y, y) \ . \tag{16}$$

Note that g is a kind of "smoothing" of $P^*(y) = F^*(-L^*y, y)$ which is everywhere majorized by P^*. The reason why we need such a g is that $P(\,\cdot\,)$ is not necessarily ℓsc, which property is important for applying compactness conditions on the levels sets of P^*; however ${}^*g$ is automatically ℓsc and ${}^*g$ dominates P, while at the same time ${}^*g$ approximates P.

Lemma 2. *Define $g(\,\cdot\,)$ as in (16). Then*

$$({}^*g)(x) \le \inf_u [F(u, Lu + x] + \sup_v \in K^0 \langle u, v\rangle] \ .$$

If $F = {}^*(F^*)$, then $P(x) \le ({}^*g)(x)$ for every $x \in \text{dom}P$. Moreover

$$\text{dom}{}^*g \supset \bigcup_{u \in \text{span}K} [\text{dom}F(u,\,\cdot\,) - Lu] \ .$$

Proof. By definition of ${}^*g$, we have

$$({}^*g)(x) = \sup_y \sup_{v \in K^0} [\langle x, y\rangle - F^*(v - L^*y, y)] \ .$$

Now for every $u \in U$ and $y \in Y$, $F^*(v - L^*y, y) \ge \langle u, v - L^*y\rangle + \langle Lu + x, y\rangle - F(u, Lu + x) = \langle u, v\rangle + \langle x, y\rangle - F(u, Lu + x)$ by definition of F^*. Hence for every $u \in U$,

$$\begin{aligned}({}^*g)(x) \le \sup_{v \in K^0} [F(u, Lu + x) - \langle u, v\rangle] &= F(u, Lu + x) + \sup_{v \in -K^0} \langle u, v\rangle \\ &= F(u, Lu + x) + \sup_{v \in K^0} \langle u, v\rangle\end{aligned}$$

where the last equality follows since K^0 is balanced. Thus we have proved the first inequality of the lemma.

Now suppose $F =^* (F^*)$ and $x \in \mathrm{dom}P$. Since K^0 is a $m(U^*,U)$-0-neighborhood, we have

$$\begin{aligned}(^*g)(x) &= \sup_{v\in K^0}\sup_y[\langle x,y\rangle - F^*(v-L^*y,y)]\\ &\geq \limsup_{v\to 0}\sup_y[\langle x,y\rangle - F^*(v-L^*y,y)]\\ &= \liminf_{v\to 0}\inf_y[F^*(v-L^*y,y)-\langle x,y\rangle]\ ,\end{aligned}$$

where the $\liminf_{v\to 0}$ is taken in the $m(U^*,U)$-topology. Define

$$h(v) = \inf_y[F^*(v-L^*y,y)-\langle x,y\rangle]\ ,$$

so that

$$(^*g)(x) \geq -\liminf_{v\to 0} h(v)\ .$$

Now

$$\begin{aligned}(^*h)(u) &= \sup_v\sup_y[\langle u,v\rangle - F^*(v-L^*y,y)+\langle x,y\rangle]\\ &= {}^*(F^*)(u,Lu+x) = F(u,Lu+x)\ .\end{aligned}$$

Hence $P(x) < +\infty$ means that $\inf_u F(u,Lu+x) < +\infty$, i.e. $^*h \not\equiv +\infty$, so that we can replace the lim inf by the second conjugate:

$$(^*g)(x) \geq -\liminf_{v\to 0} h(v) = -(^*h)^*(0) = \inf_u F(u,Lu+x) = P(x)\ .$$

The last statement in the lemma follows from the first inequality in the lemma. For

$$\begin{aligned}x \in \bigcup_{u\in \mathrm{span}K}[\mathrm{dom}F(u,\cdot)-Lu] &\text{ iff } \exists u\in[0,\infty)\cdot K \text{ st } F(u,Lu+x)<+\infty\ ,\\ &\text{iff } \exists u \text{ st } \sup_{v\in K^0}\langle u,v\rangle < +\infty\end{aligned}$$

and

$$\begin{aligned}&F(u,Lu+x) < +\infty \quad (\text{since } K =^0 (K^0))\\ &\text{iff } \exists u \text{ st } F(u,Lu+x) + \sup_{v\in K^0}\langle u,v\rangle < +\infty\end{aligned}$$

and this implies that $x \in \mathrm{dom}^*g$. Hence $\mathrm{dom}^*g \subset \bigcup_{u\in\mathrm{span}K}[\mathrm{dom}F(u,\cdot)-Lu]$.

Note that $\mathrm{dom}P$ is given by $\bigcup_{u\in U}[\mathrm{dom}F(u,\cdot)-Lu]$. ■

Theorem 8. *Assume* $F =^* (F^*)$, $P_0 < +\infty$, *and there is a* $w(U,U^*)$*-compact convex subset* K *of* U *such that* $\mathrm{span}K \subset \bigcup_{x\in X}\mathrm{dom}F(\cdot,x)$. *Suppose*

(1) $\{y\in X^* : (F^*)_\infty(-L^*y,y) - \langle x_0,y\rangle \leq 0\}$ *is a subspace* M;
(2) $\exists m(U^*,U)$*-0-neighborhood* N *in* U^*, *an* $x_1 \in X$, *an* $r_1 \in R$ *such that* $\{y\in X^* : \inf_{v\in N} F^*(v-L^*y,y) - \langle x,y\rangle < r_i\}$ *is nonempty and locally* $^\perp M$*-equicontinuous for the* $w(X^*,X)$*-topology.*

Then aff dom$P$ *is closed,* $P(\cdot) \uparrow$ aff dom$P$ *is continuous at* x_0 *for the induced topology on* aff dom$P$*, and (15) holds.*

Proof. We may assume that K is balanced and contains N^0 by replacing K with co bal$(K \cup N^0) =^0 (K^0 \cap -K^0 \cap N \cap -N)$. Define $g(\cdot)$ as in (16). We first show that dom$P$ = dom$^*g$. Now

$$\text{dom}P = \bigcap_{u \in U} [\text{dom}F(u, \cdot) - Lu] = \bigcup_{u \in \text{span}K} [\text{dom}F(u, \cdot) - Lu] \subset \text{dom}^*g$$

by Lemma 2. But also by Lemma 2 we have $P(x) \leq (^*g)(x)$ for every $x \in X$ (since dom$P \subset$ dom$^*g$), so dom$P \supset$ dom$^*g$ and hence dom$P$ = dom$^*g$.

This also implies that cldom$^*(P^*)$ = cldom$^*g$, since cldom$^*(P^*)$ = cldom$P$ by Lemma 1 (note $P^* \not\equiv +\infty$ since P^* has a nonempty level set by hypothesis 2). Hence by Definition 1 of recession functions we have $(P^*)_\infty = g_\infty - ((^*g)^*)_\infty$. A straightforward calculation using Proposition 2, and the fact that $P^*(y) = F^*(-L^*y, y)$ yields

$$g_\infty(y) = (P^*)_\infty y = (F^*)_\infty(-L^*y, y) \ .$$

Now $M = \{y \in X^* : g_\infty(y) - \langle x_0, y\rangle \leq\} = [\text{dom}g - x_0]^-$ is a subspace, hence $M = [\text{dom}g - x_0]^\perp$ and $x_0 + ^\perp M$ is a closed affine set containing dom$g$. But hypothesis (2) then implies that riepi$^*g \neq \emptyset$ and aff dom$g$ is closed with finite codimension in $x_0 + ^\perp M$, by Theorem 1. Moreover, by Theorem 3, $^*g(\cdot)$ is actually relatively continuous at x_0. Now $^\perp M =^\perp ([\text{dom}g - xc]^\perp) = \text{cl aff dom}^*g - x_0$; since aff dom$^*g$ is a closed subset of $x_0 +^\perp$ = cl aff dom$^*g$, we must have aff dom$^*g$ = cl aff dom$^*g$. Finally, since dom$P$ = dom$^*g$ and $P \leq^* g$, $P(\cdot)$ is bounded above on a relative neighborhood of x_0 and hence is relatively continuous at x_0. ∎

We shall be interested in two very useful special cases. One is when U is the dual of a normed space V, and we put the $w^* = w(U, V)$ topology as the original topology on U; for then $U^* \cong V$ and the entire space U is the span of a $w(U, V)$-compact convex set (namely the unit ball in U). Hence, if $U = V^*$ where V is a normed space, and if $F(\cdot)$ is convex and $w(U \times X, V \times X^*) - \ell$sc, then conditions (1) and (2) of Theorem 6 are automatically sufficient for (1) to hold.

The other case is when X is a barrelled space, so that interior conditions reduce to core conditions for closed sets (equivalently, compactness conditions reduce to boundedness conditions in X^*). For simplicity, we consider only Frechet spaces for which it is immediate that all closed subspaces are barrelled.

Theorem 9. *Assume* $F =^* (F^*)$; $P_0 < +\infty$; X *is a Frechet space or Banach space; and there is a* $w(U, U^*$*-compact convex set* K *in* U *such that* span$K \supset \bigcup_{x \in X} \text{dom}F(\cdot, x)$. *Then the following are equivalent:*

(1) aff dom$P$ *is closed; and* $x_0 \in$ rcordom$P$, *or equivalently* $F(u_0, Lu_0 + x_0 + x) < +\infty \Rightarrow \exists \epsilon > 0$ *and* $u_1 \in U$ st $F(u_1, Lu_1 + x_0 - \epsilon x) < +\infty$.

(2) $\{y \in X^*\colon (F^*)_\infty(-L^*y, y) - \langle x_0, y\rangle \le 0\}$ *is a subspace* M*; and there exists a* $m(U^*, U)$*-0-neighborhood* N *in* U^**, an* $x_1 \in X$*, an* $r_1 \in R$ *such that* $\{y \in X^*\colon \inf_{v\in N} F^*(v - L^*y, y) - \langle x_0, y\rangle < r_1\}$ *is nonempty and* $w(X^*, X)/M$*-locally bounded.*

If either of the above holds, then $P(\cdot) \upharpoonright \operatorname{aff}\operatorname{dom}P$ *is continuous at* x_0 *for the induced metric topology on* $\operatorname{aff}\operatorname{dom}P$ *and (1) holds.*

Proof. We first note that since $\operatorname{span}K \supset \bigcup_{x\in X} \operatorname{dom}F(\cdot, x)$ we have as in Theorem 8 that $\operatorname{dom}P = \operatorname{dom}{}^*g$ and $g_\infty(y) = (P^*)_\infty(y) = (F^*)_\infty(-L^*y, y)$.

(1) => (2). We show that $g(\cdot)$ is relatively continuous at x_0, and then (2) will follow. Now $\operatorname{dom}P = \operatorname{dom}{}^*g$, so $x_0 \in \operatorname{rcordom}P$. Let $W = \operatorname{aff}\operatorname{dom}P - x_0$ be the closed subspace parallel to $\operatorname{dom}P$, and define $h\colon W \to \overline{R}\colon w \to^* g(x_0+w)$. Since ${}^*g$ is ℓsc on X, h is ℓsc on the barrelled space W. But $0 \in \operatorname{coredom}h$ (in W), hence h is actually continuous at 0 (since W is barrelled), or equivalently ${}^*g$ is relatively continuous at x_0. Applying Theorem 3 we now see that M is the subspace $W^\perp$; the remainder of (2) then follows from Theorem 1, since $g(y) = \inf_{v\in N} F^*(v - L^*y, y) \ge ({}^*g)^*(y)$.

(2) => (1). Note that ${}^\perp M$ is a Frechet space in the induced topology, so $w(X^*, X)/M$-local boundedness is equivalent to $w(X^*, X)/M$-local compactness. But now we may simply apply Theorem 8 to get $P(\cdot)$ relatively continuous at x_0 and $\operatorname{aff}\operatorname{dom}P$ closed; of course, (1) follows. ■

Corollary 2. *Assume* $P_0 < +\infty$*;* $U = V^*$ *where* V *is a normed space;* X *is a Frechet space or Banach space;* $F(\cdot)$ *is convex and* $w(U \times X, V \times X^*) - \ell$sc. *Then the following are equivalent:*

(1) $x_0 \in \operatorname{cordom}P \equiv \operatorname{cor}\bigcup_{u\in U}[\operatorname{dom}F(u, \cdot) - Lu]$

(2) $\{y \in X^*\colon (F^*)_\infty(-L^*y, y) - \langle x_0, y\rangle \le 0\} = \{0\}$*; and there is an* $\epsilon < 0$*, an* $x_1 \in X$*, and* $r_1 \in R$ *such that*

$$\{y \in X^*\colon \inf_{|v|\le\epsilon} F^*(v - L^*y, y) - \langle x_0, y\rangle < r_1\}$$

is nonempty and $w(X^*, X)$*-locally bounded.*

(3) There is an $\epsilon > 0$*, an* $r_0 \in R$ *such that*

$$\{y \in X^*\colon \inf_{|v|\le\epsilon} F^*(v - L^*y, y) - \langle x_0, y\rangle < r_0\}$$

is nonempty and $w(X^*, X)$*-bounded.*

If any of the above holds, then $P(\cdot)$ *is continuous at* x_0 *and (1) holds.*

Proof. Immediate. ■

We can also apply these theorems to perturbations on the dual problem to get existence of solutions to the original problem P_0 and no duality gap $P_0 = D_0$.

Corollary 3. *Assume $P_0 > -\infty$; $U = V^*$ where V is a Frechet space or Banach space; X is a normed space; $F(\cdot)$ is convex and $w(U \times X, V \times X^*) - \ell$sc. Suppose $\{u \in U\colon F_\infty(u, Lu + x_0) \leq 0\}$ is a subspace M, and there is an $\epsilon > 0$, an $x_1 \in X$, an $r_1 \in R$ such that*

$$\{u \in U\colon \inf_{|x| \leq \epsilon} F(u, Lu + x_0 + x) < r_1\}$$

is nonempty and $w(U, U^*)/M$-locally compact. Then $P_0 = D_0$ and P_0 has solutions.

Proof. Apply Theorem 9 to the dual problem (12). ■

References

1. Schäfer, H.H.: Topological Vector Space. Springer, Berlin-New York (1971)
2. Schatten, R.: Norm Ideals of Completely Continuous Operators. Springer, Berlin-New York (1960)
3. Treves, F.: Topological Vector Spaces, Distributions and Kernels. Academic Press, New York (1967)
4. Ekeland, I., Teman, R.: Convex Analysis and Variational Problems. North Holland, Amsterdam (1976)
5. Rockafellar, R.T.: Conjugate Duality and Optimization. In: Conference Board of Math. Sci. Series, vol. 16, SIAM Publications (1974), Much of the material presented in this paper represents joint work with S.K. Young as reported in his unpublished doctoral thesis: Convexity and Duality in Optimization Theory, Ph.D. dissertation, Mathematics Department, MIT (1977)
6. Mitter, S.K., Young, S.K.: Integration with respect to Operator Valued Measures with Applications to Quantum Estimation Theory. Annali di Matematica Pura ed Applicata, vol. CXXXVII, pp. 1–39 (1984); and the references cited there. The section on Quantum Detection has not been published before. There has been considerable recent activity in Infinite-Dimensional Mathematical Programming
7. Anderson, E.J., Nash, P.: Linear Programming in Infinite Dimensional Spaces. John Wiley & Sons, New York (1987)
8. Fleming, W.H., Soner, H.M.: Controlled Markov Processes and Viscosity Solutions. Springer, Berlin-New York (1993)
9. Villani, C.: Topics i Optimal Transportation of Graduate Studies in Mathematics, vol. 58. American Mathematical Society, Providence, RI (2003)