DUALITY AND OPTIMALITY IN MULTISTAGE STOCHASTIC PROGRAMMING

R. T. Rockafellar *

University of Washington, Seattle

Abstract. A model of multistage stochastic programming over a scenario tree is developed in which the evolution of information states, as represented by the nodes of a scenario tree, is supplemented by a dynamical system of state vectors controlled by recourse decisions. A dual problem is obtained in which multipliers associated with the primal dynamics are price vectors that are propagated backward in time through a dual dynamical system involving conditional expectation. A format of Fenchel duality is employed in order to have immediate specialization not only to linear programming but extended linear-quadratic programming. The resulting optimality conditions support schemes of decomposition in which a separate optimization problem is solved at each node of the scenario tree.

* This work was supported in part by the National Science Foundation under grant DMS–9500957.

1. Introduction

In all branches of optimization, duality draws on convexity. In stochastic programming, another important ingredient to duality is dynamical structure. Such structure describes the evolution of information about the random elements in a problem's environment, and to that extent is essential to the very concept of optimization under uncertainty, but it can also be developed with respect to other ways that the past and future affect the present. Concepts of "state" and "control" are useful in this. Controls correspond to decisions to be made, whereas states summarize the current primal or dual status of the system at the time of making those decisions.

A distinguishing feature of the multistage model of stochastic programming to be laid out here is that the dynamics of uncertainty, discretized as a scenario tree in which nodes represent probabilistic states of information, is supplemented by a linear dynamical system of vectors representing auxiliary aspects of state. The relations in this system can be treated as constraints to which multiplier vectors are assigned, and those vectors become dual states in a dual dynamical system.

A forerunner to this model was developed in Rockafellar and Wets [1], but with the evolution of probabilistic information described by a sequence of fields of measurable sets in a discrete probability space. The notion of a scenario tree, adopted here instead, adds concreteness without serious loss of generality. In associating the nodes of the tree with decision stages, the crucial property of nonanticipativity of decisions is made automatic. Moreover, the role of conditional expectations in the dynamics of prices is clarified. Independently, Korf [2] has found an equivalent expression for that role in the measurability context of [1], but the scenario tree approach brings it out quite naturally.

Another difference between the model in [1] and the one here is that the framework of linear or extended linear-quadratic programming has been generalized to the extended Fenchel duality format, which is more flexible and less cluttered. Simultaneously, some features of the cost structure in [1], such as ways of writing initial and terminal costs, have been simplified. The problem should therefore be easier to understand and work with.

A powerful property of the optimality conditions in [1] emerges again here. It is seen in how the dynamical systems of primal and dual state vectors lead to a decomposition in which a small-scale optimization problem, depending on those vectors, is solved at each individual node of the scenario tree in order to obtain the optimal controls. This differs both from the classical Dantzig-Wolfe type of decomposition (which, in extension to convex programming, is tied to separability of the Lagrangian in the primal argument for fixed values of the dual argument) and from Benders decomposition (which concerns cost-to-go functions and their subgradients). It relates instead to a characterization of primal and dual optimal solution pairs as saddle points of a generalized Lagrangian function that is the sum of two "sub-Lagrangians." One of these breaks into separate expressions for each node of the scenario tree, while the other is a bi-affine representation of the dynamics of the primal and dual state vectors.

Decomposition of this third kind dates back, in deterministic rather than stochastic settings, to Rockafellar [3], [4]. In stochastic programming computations it has recently been utilized by Salinger [5]. A corresponding numerical application to the deterministic case of the model in [1] has been carried out by Eckstein and Ferris [6]. The backward-forward splitting methods applied by Chen [7] to deterministic problems would be suited to this kind of decomposition also; for related results on the general convergence of such methods, see [8].

The Lagrangian saddle point scheme in [1] relies on a "reduced" formulation of the underlying problem, in which only the controls appear as independent variables. Here, in further contrast to [1], we look at a "full" formulation in tandem with the reduced formulation. In that way additional insights are obtained about the interpretation of the primal and dual state vectors.

2. Scenario Tree and Dynamics

We begin with a *scenario tree* based on a finite set I of nodes i. One of the nodes, $i = 0$, stands for the here-and-now. Every other node $i \neq 0$ has a unique predecessor node, denoted by i_-, and a transition probability $\tau_i > 0$, which is the probability that i will be the successor to i_-. The successors to any node i, denoted generically by i_+, form the set $I_+(i) \subset I$; the transition probabilities of those nodes add to 1. Thus, i_+ can be viewed as a discrete random variable over $I_+(i)$ with distribution given by the probabilities τ_{i_+}. Nodes i with $I_+(i) = \emptyset$ are called *terminal*; they constitute the set $T \subset I$.

In tracing back from any node i through its predecessors to 0, we trace in reverse a sequence of realizations of the discrete random variables associated with the transition probabilities. It's convenient to think of i as standing for this history of realizations, and to define π_i to be the probability of the particular sequence occurring. Such probabilities are generated recursively by

$$\pi_0 = 1, \qquad \pi_i = \tau_i \pi_{i_-} \text{ for } i \neq 0. \tag{2.1}$$

In the case of a node $i \in T$, the history of realizations corresponds to a path from the root 0 of the tree all the way to one of its "leaves" and is called a *scenario*. The probabilities π_i for $i \in T$ obviously add to 1 and provide a distribution for the set of all scenarios.

In many, or most, situations it may be helpful to view the node set I as partitioned into subsets I_k designating *stages* $k = 0, 1, \ldots, N$, where $I_0 = \{0\}$, $I_N = T$, and the successor to a node in I_k belongs to I_{k+1}. Mathematically, however, there's no actual need for that, so stage notation will be left out in the interests of simplicity. Further details about the scenario tree can always be brought in when it matters.

Every information state $i \in I$ is viewed as providing the opportunity for a decision to be made. We model this as the choice of a vector $u_i \in I\!R^{n_i}$; the vector u_0 gives rise to the "here-and-now" decision, whereas the vectors u_i for $i \neq 0$ give "recourse" decisions. Optimization revolves around these elements, which will be termed the *controls* of the system, but it's important that the decision environment in state i be able to be molded by the decisions taken in the states leading up to i. The mechanism for this is provided by the introduction for each $i \in I$ of a *state* vector $x_i \in I\!R^{d_i}$ and letting the states and controls be governed by a dynamical system

$$x_0 = a \text{ (given)}, \qquad x_i = F_i(x_{i_-}, u_{i_-}) \text{ for } i \neq 0. \tag{2.2}$$

The optimization environment in information state i is modeled then by a cost expression $f_i(x_i, u_i)$ (oriented toward minimization), in which the vector x_i acts as a parameter element supplying the influence from the past. This cost is allowed to be ∞ as a means of incorporating constraints without having to appeal at once to some particular constraint structure and its burdensome notation; a vector u_i is only considered feasible relative to x_i if $f_i(x_i, u_i) < \infty$. The forms of F_i and f_i will be specialized in due course, but it's useful for now to approach the situation more generally.

In attaching F_i, f_i and x_i by subscript to the information state i, we take the position that these elements are known to the decision maker upon reaching i; thus too, the function $f_i(x_i, \cdot)$ giving the costs (and implicit constraints) in choosing u_i is known. (Any random variables that enter the description of f_i are regarded as having been built into the specification of the transition probabilities encountered on the way from 0 to i.) Observe however that although F_i is supposed to be known upon reaching i, it might not have been known at the predecessor node i_-, when u_{i_-} had to be chosen. The explanation again is that in passing from i_- to i the dynamics in (2.2) may involve the realization of some additional aspect of uncertainty. Alternatively these dynamics can be written as

$$x_{i_+} = F_{i_+}(x_i, u_i) \text{ for } i \notin T, \quad \text{with } x_0 = a, \tag{2.3}$$

in which the role of i_+ as a random variable ranging over $I_+(i)$ is more apparent.

The stochastic programming problem we consider has two formulations, fundamentally equivalent yet different, and it will be important to distinguish between them. In the *full* formulation, the problem is

$$(\mathcal{P}_0^+)\qquad \begin{array}{l} \text{minimize } \sum_{i\in I} \pi_i f_i(x_i,u_i) \text{ over all } x_i \text{ and } u_i \\ \text{subject to the relations (2.2) as constraints.} \end{array}$$

Implicit in this, as already mentioned, are the conditions $f_i(x_i,u_i) < \infty$, without which the expression being minimized would have the value ∞. In the *reduced* formulation, the x_i's are regarded as dependent rather than independent variables:

$$(\mathcal{P}_0^-)\qquad \begin{array}{l} \text{minimize } \sum_{i\in I} \pi_i f_i(x_i,u_i) \text{ over all } u_i, \text{ where} \\ x_i \text{ stands for an expression in prior control vectors.} \end{array}$$

The expressions in question are generated recursively from (2.2).

The two problems are obviously equivalent in the sense that vectors $\bar{u}_i$ and $\bar{x}_i$ furnish an optimal solution to $(\mathcal{P}_0^+)$ if and only if the controls $\bar{u}_i$ furnish an optimal solution to $(\mathcal{P}_0^-)$ and the states $\bar{x}_i$ furnish the corresponding trajectory obtained from them by "integrating" the dynamics (2.2). Both of these formulations are useful. Problem $(\mathcal{P}_0^-)$ has the advantage of being "smaller" and conceptually leaner, but $(\mathcal{P}_0^+)$ promotes the exploration of dual state vectors y_i, which come out as multipliers for the relations in (2.2) as constraints.

Theorem 1 (basic convexity)**.** *As long as the functions f_i are convex and the mappings F_i are affine, both $(\mathcal{P}_0^+)$ and $(\mathcal{P}_0^-)$ are problems of convex optimization—where a convex function is minimized subject to constraints describing a convex feasible set.*

The proof of this fact is elementary; we record the theorem for the perspective it offers, since convexity will be needed for duality. Note that the convexity depends heavily on the transition probabilities being unaffected by the decisions that are to be made over time; we are dealing with constants τ_i instead of variable expressions $\tau_i(x_{i_-},u_{i_-})$. Problems $(\mathcal{P}_0^+)$ and $(\mathcal{P}_0^-)$ would still make sense if such variable transition probabilities were allowed, with the π_i's then turning into expressions in prior states and controls as generated through (2.1) and (2.2). Convexity would be lost, however, and with it the prospect of being able to use duality-based decomposition methods of solution to compensate for the extremely large scale of problems in stochastic programming.

3. Duality Background

From Theorem 1 it's clear that, for our purposes, the mappings F_i should be affine, but what structure should be introduced in the cost functions f_i to bring out duality most conveniently? Because the f_i's are extended-real-valued, constraint structure is at stake as well. We want multistage models of linear programming type to be covered nicely, and also quadratic programming analogs, for instance. Even in ordinary quadratic programming, however, there is trouble over duality. Unlike the situation in linear programming, one can't dualize a quadratic programming problem and expect to get another quadratic programming problem.

The kind of Lagrangian duality that is available from conventional formulations of convex programming with equality and inequality constraints is too narrow for the task now facing us and suffers from the further drawback that such formulations tend to emphasize "hard constraints," whereas the needs of stochastic programming may often be better served by penalty expressions. The Fenchel scheme of duality comes to the rescue here. It's much more flexible, yet just as explicit in key cases. In particular, it gets around the quadratic programming difficulty by way of "extended linear-quadratic programming," which handles penalties and even box constraints with ease. The ideas behind Fenchel duality will be reviewed now as background to presenting, in the next section, more structured versions of problems $(\mathcal{P}_0^+)$ and $(\mathcal{P}_0^-)$. A fresh treatment of such duality in more detail is available now in [9, Chap. 11].

Recall that an extended-real-valued convex function ϕ on $I\!\!R^n$ is *proper* if it nowhere takes on $-\infty$ and is not the constant function ∞. The function ϕ^* *conjugate* to a proper convex function ϕ is defined by

$$\phi^*(w) = \sup_{u \in \boldsymbol{R}^n} \big\{ u{\cdot}w - \phi(u) \big\}.$$

It's always proper convex and lsc (lower semicontinuous). As long as ϕ itself is lsc, the function ϕ^{**} conjugate to ϕ^* is in turn ϕ:

$$\phi(w) = \sup_{w \in \boldsymbol{R}^n} \big\{ u{\cdot}w - \phi^*(w) \big\}.$$

Although it may be hard in some cases to come up with a more explicit formula for ϕ^* than the definition, there are cases where it's easy, and they go a long way toward making conjugate functions a practical tool in duality schemes.

The *extended* Fenchel duality scheme that will serve as our basic guide concerns a proper lsc convex function ϕ on $I\!\!R^n$, another such function ψ on $I\!\!R^m$, a matrix $D \in I\!\!R^{n \times m}$, and vectors $p \in I\!R^n$ and $q \in I\!R^m$. The primal problem has the form

$$(\mathcal{P}) \qquad \text{minimize } \Phi(u) := p{\cdot}u + \phi(u) + \psi(q - Du) \text{ over } u \in I\!R^n,$$

while the dual problem is

$$(\mathcal{D}) \qquad \text{maximize } \Psi(v) := q{\cdot}v - \psi^*(v) - \phi^*(D^*v - p) \text{ over } v \in I\!R^m,$$

where D^* is the transpose of D. Implicit constraints come out of the effective domains $\operatorname{dom}\phi := \{u \in I\!R^n \,|\, \phi(u) < \infty\}$ and $\operatorname{dom}\psi := \{z \in I\!R^m \,|\, \psi(z) < \infty\}$. The implicit feasible set in $(\mathcal{P})$ is the convex set consisting of the vectors u that satisfy

$$u \in \operatorname{dom}\phi, \quad q - Du \in \operatorname{dom}\psi. \tag{3.1}$$

Similarly, the implicit feasible set in $(\mathcal{D})$ is described by

$$v \in \operatorname{dom}\psi^*, \quad D^*v - p \in \operatorname{dom}\phi^*. \tag{3.2}$$

Examples will be considered after the main results about this pairing of problems are stated.

For problems in this format, the constraint qualifications needed to obtain duality in general are expressed in terms of the notion of the relative interior "ri" of a convex set. Such constraint qualifications turn out not to be needed for functions that are *piecewise linear-quadratic*. A proper convex function ϕ is said to fall into that category if $\operatorname{dom}\phi$ is a polyhedral (convex) set on which ϕ is given by a linear-quadratic formula, or a union of finitely many such sets on which ϕ is given by such formulas. By a linear-quadratic formula we mean a polynomial function of degree at most 2; linear functions and constant functions are a special case. For instance if ϕ is the indicator δ_U of a polyhedral set U (i.e., has the value 0 on U and ∞ elsewhere), then in particular, ϕ is piecewise linear-quadratic, although the full generality of the definition isn't utilized.

An important fact is this: if a proper convex function is piecewise linear-quadratic, its conjugate function is piecewise linear-quadratic as well. Thus, if ϕ and ψ are piecewise linear-quadratic in $(\mathcal{P})$, the same is true of ϕ^* and ψ^* in $(\mathcal{D})$. We refer to this as the *piecewise linear-quadratic* case in Fenchel duality.

Theorem 2 (extended Fenchel duality)**.**

(a) *The relation* $\inf(\mathcal{P}) = \sup(\mathcal{D}) < \infty$ *is guaranteed under the primal constraint qualification that*

$$\exists\, u \textit{ satisfying } \quad u \in \operatorname{ri}\operatorname{dom}\phi, \quad q - Du \in \operatorname{ri}\operatorname{dom}\psi. \tag{3.3}$$

Then too, unless the common optimal value is $-\infty$ *(so* $(\mathcal{D})$ *has no feasible solution),* $(\mathcal{D})$ *is sure to have an optimal solution.*

(b) *The relation* $\inf(\mathcal{P}) = \sup(\mathcal{D}) > -\infty$ *is guaranteed under the dual constraint qualification that*

$$\exists\, v \textit{ satisfying } \quad v \in \operatorname{ri}\operatorname{dom}\psi^*, \quad D^*v - p \in \operatorname{ri}\operatorname{dom}\phi^*. \tag{3.4}$$

Then too, unless the common optimal value is ∞ *(so* $(\mathcal{P})$ *has no feasible solution),* $(\mathcal{P})$ *is sure to have an optimal solution.*

(c) *In the piecewise linear-quadratic case, the primal and dual constraint qualifications are superfluous and can be replaced simply by the feasibility conditions in (3.1) and (3.2), respectively. In that case, therefore,*

$$\inf(\mathcal{P}) = \sup(\mathcal{D}) \quad \textit{unless} \quad \inf(\mathcal{P}) = \infty \textit{ and } \sup(\mathcal{D}) = -\infty,$$

(i.e., unless neither $(\mathcal{P})$ *nor* $(\mathcal{D})$ *has a feasible solution). Moreover, when the common optimal value is finite, both problems have an optimal solution.*

Proof. The basic facts in (a) and (b) go back all the way to Rockafellar [10]. The piecewise linear-quadratic case, while partially covered earlier, was recently established in its full scope in Rockafellar and Wets [9; cf. 11.42]. □

Among the special cases to note here, *linear programming* duality corresponds to taking ϕ to be the indicator of $\mathbb{R}^n_+$ and ψ to be the indicator of $\mathbb{R}^m_-$, so that $(\mathcal{P})$ comes out as minimizing $p{\cdot}u$ subject to $u \geq 0$ and $Du \geq q$. Then ϕ^* and ψ^* are the indicators of $\mathbb{R}^n_-$ and $\mathbb{R}^m_+$, so that $(\mathcal{D})$ consists of maximizing $q{\cdot}v$ subject to $v \geq 0$ and $D^*v \leq p$. This is covered by part (c) of Theorem 2.

The orthants here could be replaced by other convex cones. (The function conjugate to the indicator of a cone is the indicator of the polar cone.) More interesting for stochastic programming, however, is the case where $\mathbb{R}^n$ is replaced by some *box* (a product of closed intervals, bounded or unbounded). When $\phi = \delta_U$, ϕ^* is the support function σ_U of U, and for a box U that is bounded this means ϕ^* is piecewise linear (and easy to write down explicitly). Even more to the point is the case where ψ is such a support function σ_V of a box V, so that $\psi^* = \delta_V$. The term $\psi(q - Du)$ in $(\mathcal{P})$ corresponds then to a linear penalty expression in place of, say, a constraint like $q - Du \leq 0$. There are rich possibilities.

A handy tool in this respect is that of the function $\theta_{V,Q}$ on $\mathbb{R}^m$ defined in terms of a nonempty polyhedral set $V \subset \mathbb{R}^m$ and a symmetric positive *semi*-definite matrix $Q \in \mathbb{R}^{m\times m}$ ($Q = 0$ allowed) by

$$\theta_{V,Q}(z) = \sup_{v\in V} \left\{ z{\cdot}v - \tfrac{1}{2} v{\cdot}Qv \right\}. \tag{3.5}$$

This means that $\theta_{V,Q}$ is the convex function conjugate to $\delta_V + j_Q$, where $j_Q(v) = \frac{1}{2}v{\cdot}Qv$. Since $\delta_V + j_Q$ falls in the category of piecewise linear-quadratic functions, the same is true of $\theta_{V,Q}$. For various useful choices of V and Q it's possible to make this linear-quadratic structure of $\theta_{V,Q}$ quite explicit. Analogously, functions $\theta_{U,P}$ can be introduced on $I\!R^n$ as the piecewise linear-quadratic conjugates of functions $\delta_U + j_P$ for a polyhedral set $U \subset I\!R^n$ and symmetric positive *semi*-definite matrix $P \in I\!R^{n\times n}$.

In taking $\phi = \delta_U + j_P$ and $\psi = \theta_{V,Q}$, so that $\phi^* = \theta_{U,P}$ and $\psi^* = \delta_V + j_Q$, we obtain the following pair of problems from $(\mathcal{P})$ and $(\mathcal{D})$:

$$(\mathcal{P}') \qquad \text{minimize } p{\cdot}u + \tfrac{1}{2}u{\cdot}Pu + \theta_{V,Q}(q - Du) \text{ over } u \in U,$$

$$(\mathcal{D}') \qquad \text{maximize } q{\cdot}v - \tfrac{1}{2}v{\cdot}Qv - \theta_{U,P}(D^*v - p) \text{ over } v \in V.$$

This is the duality scheme of *extended linear-quadratic programming*. It too is governed by part (c) of Theorem 2. Linear programming comes out when U and V are cones while P and Q are zero matrices. As another example, conventional quadratic programming would consist of minimizing $p{\cdot}u + \frac{1}{2}u{\cdot}Pu$ subject to $u \in U$ and $Du \geq q$, where U is $I\!R^n_+$ or perhaps some other box. A problem of such type can't be dualized within that format, but in the framework of extended linear-quadratic programming the dual problem consists of maximizing $q{\cdot}v - \theta_{U,P}(D^*v - p)$ over $v \geq 0$. (The implicit constraint $D^*v - p \in \operatorname{dom}\theta_{U,P}$ combines with $v \geq 0$ to produce the implicit feasible set in this dual.)

Because stochastic programming is our subject here, it's worth mentioning that piecewise linear-quadratic functions of type $\theta_{V,Q}$ were first introduced in a stochastic programming context, in Rockafellar and Wets [11]. This was motivated by the convenience of such functions in furnishing penalty expressions in a readily dualizable form. Penalty substitutes for constraints are especially welcome when dealing with uncertainty. The format of extended linear-quadratic programming in $(\mathcal{P}')$ and $(\mathcal{D}')$ comes from [11] as well. Examples of the many special problem statements covered by it were subsequently presented in [3; §§2,3]; for stochastic programming, see also [12], where the separable case of functions $\theta_{V,Q}$ is well described.

Although Fenchel duality isn't based on Lagrange multipliers, at least of the traditional variety, a Lagrangian function plays a crucial role nonetheless. This Lagrangian in the general case of $(\mathcal{P})$ and $(\mathcal{D})$ is

$$\begin{aligned} L(u,v) = p{\cdot}u + \phi(u) + q{\cdot}v - \psi^*(v) - v{\cdot}Du \\ \text{on } U \times V, \text{ where } U := \operatorname{dom}\phi,\ V := \operatorname{dom}\psi^*. \end{aligned} \tag{3.6}$$

In the extended linear-quadratic programming format of problems $(\mathcal{P}')$ and $(\mathcal{D}')$, the generalized Lagrangian that takes on the right role is

$$L(u,v) = p{\cdot}u + \tfrac{1}{2}p{\cdot}Pu + q{\cdot}v - \tfrac{1}{2}v{\cdot}Qv - v{\cdot}Du \quad \text{ on } U \times V. \tag{3.7}$$

The Lagrangians associated with extended linear-quadratic programming are thus the functions obtained by restricting some convex-concave linear-quadratic function to a product of nonempty polyhedral sets.

Note that the Lagrangian in each case isn't a function with unspecified domain, but a triple (L, U, V). This entire triple enters the picture through the way that the objective functions in the two problems can be recovered from L, U and V by

$$\Phi(u) = \begin{cases} \sup_{v\in V} L(u,v) & \text{when } u \in U \\ \infty & \text{when } u \notin U, \end{cases} \tag{3.8}$$

$$\Psi(v) = \begin{cases} \inf_{u\in U} L(u,v) & \text{when } v \in V \\ -\infty & \text{when } v \notin V. \end{cases} \tag{3.9}$$

It also enters in saddle point characterizations of optimality, as in the next theorem. Recall that $(\bar u, \bar v)$ is a *saddle point of* L *on* $U \times V$ when

$$\begin{cases} \bar u \in U,\ \bar v \in V, \text{ and} \\ L(u,\bar v) \geq L(\bar u, \bar v) \geq L(\bar u, v) \text{ for all } u \in U,\ v \in V. \end{cases} \tag{3.10}$$

Theorem 3 (Lagrangians in Fenchel duality)**.** *In the circumstances of Theorem 2 in which* $\inf(\mathcal{P}) = \sup(\mathcal{D})$, *a pair* $(\bar u, \bar v)$ *is a saddle point of the Lagrangian* L *over* $U \times V$ *in (3.6) if and only if* $\bar u$ *is an optimal solution to* $(\mathcal{P})$ *and* $\bar v$ *is an optimal solution to* $(\mathcal{D})$*. This saddle point property of* $(\bar u, \bar v)$ *is equivalent to the subgradient conditions*

$$D^*\bar v - p \in \partial\phi(\bar u), \qquad q - D\bar u \in \partial\psi^*(\bar v). \tag{3.11}$$

Proof. The original saddle point characterization of optimal solutions in Fenchel duality was developed in [13], where the corresponding subgradient conditions were first given as well. More recently see also [9; Chap. 11]. □

The saddle point characterization of optimality assists in interpreting $\bar v$ as a "generalized multiplier vector" associated with the term $\psi(q - Du)$ in $(\mathcal{P})$. This perspective is opened further in [14].

The formulas in (3.8) and (3.9) for the objectives in $(\mathcal{P})$ and $(\mathcal{D})$ in terms of L, U and V lead to a general principle that can help us, in more complicated situations in other notation, to ascertain whether a given optimization problem fits the Fenchel format, and if so, what the corresponding dual problem is. All we need to know is whether the function of u (say) that is being minimized in the given problem can be expressed by *the right side of* (3.8) through a function $L(u,v)$ of the type in (3.6)—and that too can be viewed very broadly: $L(u,v)$ need only be *the difference between two lsc proper convex functions of u and v separately (as restricted to their effective domains U and V), plus some expression that's bi-affine in u and v* (i.e., affine in u for fixed v as well as affine in v for fixed u). Once this has been confirmed, we can identify the dual with the problem of maximizing the function of v given by *the right side of* (3.9), and the theorems above can be applied.

The point here is that we can bypass having to write down what the vectors p and q and the matrix D are in a given case in order to invoke Fenchel duality. The objective in the dual problem can be deduced straight from the right side of (3.9) without that distraction. For stochastic programming in particular, that will be the most expedient approach to dualization.

4. Stochastic Programming Duality

The stage is set now for the specialization of the general stochastic programming problems $(\mathcal{P}_0^+)$ and $(\mathcal{P}_0^-)$ to models that are able to take advantage of extended Fenchel duality. We choose the mappings F_i to be linear in the notation

$$F_i(x_{i_-}, u_{i_-}) = A_i x_{i_-} + B_i u_{i_-} \tag{4.1}$$

for matrices A_i and B_i, so that the vectors x_i have the dynamics

$$\begin{cases} x_0 = a, \\ x_i = A_i x_{i_-} + B_i u_{i_-} \quad \text{for } i \neq 0. \end{cases} \tag{4.2}$$

We take the cost functions f_i to have the form

$$f_i(x_i, u_i) = p_i \cdot u_i + \phi_i(u_i) + \psi_i(q_i - C_i x_i - D_i u_i) \tag{4.3}$$

for matrices C_i and D_i, vectors p_i and q_i, and lsc proper convex functions ϕ_i and ψ_i on $\mathbb{R}^{n_i}$ and $\mathbb{R}^{m_i}$, respectively. As the subscripting by i indicates, all these elements are regarded as known to the decision maker once the information state i has been reached.

(Affine mappings $F_i(x_{i_-}, u_{i_-}) = A_i x_{i_-} + B_i u_{i_-} + b_i$ could be handled in place of the linear ones in (4.2), but the additional notation gets cumbersome. Anyway, no real generality would be gained, because the vector b_i could equally well be incorporated as an extra column of B_i for which the corresponding component of u_{i_-} has to be 1, as secured implicitly through the specification of the effective domain of the corresponding ϕ_i.)

Let $u = \{u_i\}_{i\in I}$ be the "supervector" of controls u_i, and similarly let $x = \{x_i\}_{i\in I}$. We have $u \in I\!R^n := \Pi_{i\in I} I\!R^{n_i}$ and $x \in I\!R^d := \Pi_{i\in I} I\!R^{d_i}$. Let $X : I\!R^n \to I\!R^d$ be the affine mapping defined by

$$X(u) = \text{ the primal state trajectory generated from } u \text{ by the dynamics (4.2).}$$

Our stochastic programming problem in its *full* formulation is then

$$(\mathcal{P}^+) \qquad \begin{array}{l} \text{minimize } \Phi^+(x,u) := \displaystyle\sum_{i\in I} \pi_i \big[p_i{\cdot}u_i + \phi_i(u_i) + \psi_i(q_i - C_i x_i - D_i u_i)\big] \\ \text{over } x \text{ and } u, \text{ subject to } x - X(u) = 0, \end{array}$$

whereas in its *reduced* formulation it is

$$(\mathcal{P}^-) \qquad \begin{array}{l} \text{minimize } \Phi^-(u) := \displaystyle\sum_{i\in I} \pi_i \big[p_i{\cdot}u_i + \phi_i(u_i) + \psi_i(q_i - C_i x_i - D_i u_i)\big] \\ \text{over } u, \text{ where } x = X(u). \end{array}$$

In the full version the equations in (4.2) are taken as a system of linear constraints on the vector variables x_i and u_i. In the reduced version, though, x_i merely stands for an affine expression in the control vectors associated with the information states leading up to i. Those expressions can be generated out of (4.2) by a chain of substitutions, but it won't actually be necessary to do that in order to make effective use of $(\mathcal{P}^-)$. The implicit conditions for feasibility in both problems can anyway be written as

$$u_i \in \operatorname{dom}\phi_i, \quad q_i - C_i x_i - D_i u_i \in \operatorname{dom}\psi_i, \quad x = X(u), \tag{4.4}$$

whichever point of view is being adopted. Obviously $(\mathcal{P}^+)$ and $(\mathcal{P}^-)$ are equivalent in the sense that $\inf(\mathcal{P}^+) = \inf(\mathcal{P}^-)$ and

$$(\bar{x},\bar{u}) \text{ solves } (\mathcal{P}^+) \iff \bar{u} \text{ solves } (\mathcal{P}^-) \text{ and } \bar{x} = X(\bar{u}). \tag{4.5}$$

Dualization will proceed first with the full primal problem $(\mathcal{P}^+)$. The full dual problem $(\mathcal{D}^+)$ obtained in this way will have a reduced form $(\mathcal{D}^-)$, which will be shown later to be dual to the reduced problem $(\mathcal{P}^-)$ with respect to a reduced Fenchel scheme. Let

$$\begin{array}{l} L_i(u_i,v_i) := p_i{\cdot}u_i + \phi_i(u) + q_i{\cdot}v_i - \psi_i^*(v_i) - v_i{\cdot}D_i u_i \\ \qquad\qquad \text{on } U_i \times V_i := [\operatorname{dom}\phi_i] \times [\operatorname{dom}\psi_i^*] \subset I\!R^{n_i} \times I\!R^{m_i}. \end{array} \tag{4.6}$$

Let $v = \{v_i\}_{i\in I}$ in $I\!R^m := \Pi_{i\in I} I\!R^{m_i}$ and define

$$U = \Pi_{i\in I} U_i \subset I\!R^n, \qquad V = \Pi_{i\in I} V_i \subset I\!R^m. \tag{4.7}$$

The sets U and V are convex.

In terms of $y_i \in \mathbb{R}^{d_i}$ and $y = \{y_i\}_{i\in I} \in \mathbb{R}^d$, we take as the *full Lagrangian* L^+, associated with $(\mathcal{P}^+)$, the expression

$$L^+(x,u;y,v) := \sum_{i\in I} \pi_i \big[L_i(u_i,v_i) - v_i{\cdot}C_i x_i\big] + \sum_{i\neq 0} \pi_i y_i{\cdot}\big[x_i - A_i x_{i_-} - B_i u_{i_-}\big] + y_0{\cdot}[x_0 - a]$$
$$\text{on}\ \ [\mathbb{R}^d \times U] \times [\mathbb{R}^d \times V]. \tag{4.8}$$

The vectors $y_i \in \mathbb{R}^{d_i}$ are multipliers in the traditional sense for the equations in (4.2) as constraints in $(\mathcal{P}^+)$ (so that, in overview, y is a multiplier for the constraint $x - X(u) = 0$). The vectors v_i, on the other hand, will act as generalized multipliers, in the sense of Fenchel duality, for the ψ_i terms in the objective of $(\mathcal{P}^+)$.

The principle set down at the end of §3 will guide our effort at dualization. We apply this principle by thinking of $L^+(x,u;y,v)$ as $L^+(u',v')$ for $u' = (x,u) \in U' = \mathbb{R}^d \times U$ and $v' = (y,v) \in V' = \mathbb{R}^d \times V$. Calculating $\sup_{v'\in V'} L^+(u',v')$ as on the right side of (3.8), we get the function Φ^+ in $(\mathcal{P}^+)$ as restricted by the dynamics in (4.2): namely, whereas $\Phi^+(x,u) = \infty$ when $u \notin U$, we have for $u \in U$ that

$$\sup_{(y,v)\in \boldsymbol{R}^d\times V} L^+(x,u;v,y) = \begin{cases} \Phi^+(x,u) & \text{when } x = X(u), \\ \infty & \text{when } x \neq X(u). \end{cases} \tag{4.9}$$

Next we observe that L^+ has the form required for Fenchel duality: it's the difference between the lsc proper convex functions

$$\phi^+(u') = \sum_{i\in I} \phi_i(u_i), \qquad (\psi^+)^*(v') = \sum_{i\in I} \psi_i^*(v_i) \tag{4.10}$$

(not really depending on x and y), as restricted to their effective domains U' and V', plus an expression that's affine in u' for fixed v' and affine in v' for fixed u'. It follows by our principle that, in the Fenchel duality framework, $(\mathcal{P}^+)$ is the primal problem associated with L^+ on $U' \times V'$, and moreover that the corresponding dual problem can be obtained by developing the expression

$$\inf_{u'\in U'} L^+(u',v') = \inf_{(x,u)\in \boldsymbol{R}^d\times U} L^+(x,u;y,v) \tag{4.11}$$

on the right side of (3.9).

This calculation is facilitated by a notation for *conditional expectation* in a state i with respect to its successor states i_+. We'll set

$$E_i\{w_{i_+}\} := \sum_{i_+\in I_+(i)} \tau_{i_+} w_{i_+} \tag{4.12}$$

when a vector w_{i_+} depends on i_+.

The trick now is to rewrite the linear and bilinear terms in $L^+(x,u;y,v)$ from the (i_-,i) mode to the (i,i_+) mode, in which

$$x_{i_+} = A_{i_+}x_i + B_{i_+}u_i \text{ for } i \notin T. \tag{4.13}$$

Denoting the transposes of A_i, B_i and C_i by A_i^*, B_i^* and C_i^*, it's easy to see in this way, through (2.1), that

$$\begin{aligned}
\sum_{i\neq 0} \pi_i y_i{\cdot}A_i x_{i_-} &= \sum_{i\notin T}\left[\sum_{i_+\in I_+(i)} \pi_{i_+} y_{i_+}{\cdot}A_{i_+}x_i\right]\\
&= \sum_{i\notin T}\left[\sum_{i_+\in I_+(i)} \pi_i \tau_{i_+} x_i{\cdot}A_{i_+}^* y_{i_+}\right] = \sum_{i\notin T}\pi_i x_i{\cdot}E_i\{A_{i_+}^* y_{i_+}\},\\
\sum_{i\neq 0} \pi_i y_i{\cdot}B_i u_{i_-} &= \sum_{i\notin T}\left[\sum_{i_+\in I_+(i)} \pi_{i_+} y_{i_+}{\cdot}B_{i_+}u_i\right]\\
&= \sum_{i\notin T}\left[\sum_{i_+\in I_+(i)} \pi_i \tau_{i_+} u_i{\cdot}B_{i_+}^* y_{i_+}\right] = \sum_{i\notin T}\pi_i u_i{\cdot}E_i\{B_{i_+}^* y_{i_+}\},
\end{aligned}$$

and consequently, for potential use in the context of (4.8), that

$$\begin{aligned}
&-\sum_{i\in I}\pi_i v_i{\cdot}C_i x_i + \sum_{i\neq 0}\pi_i y_i{\cdot}\big[x_i - A_i x_{i_-} - B_i u_{i_-}\big] + y_0{\cdot}[x_0 - a]\\
&\quad= \sum_{i\notin T}\pi_i x_i{\cdot}\big[y_i - E_i\{A_{i_+}^* y_{i_+}\} - C_i^* v_i\big] + \sum_{i\in T}\pi_i x_i{\cdot}\big[y_i - C_i^* v_i\big]\\
&\qquad - \sum_{i\notin T}\pi_i u_i{\cdot}E_i\{B_{i_+}^* y_{i_+}\} - y_0{\cdot}a.
\end{aligned} \tag{4.14}$$

The vectors x_i can be perceived now as multipliers for the constraints associated with the dynamical system

$$\begin{cases} y_i = C_i^* v_i & \text{for } i \in T,\\ y_i = E_i\{A_{i_+}^* y_{i_+}\} + C_i^* v_i & \text{for } i \notin T, \end{cases} \tag{4.15}$$

in which the vectors y_i can be interpreted as *dual* states, propagating backward in time in response to the vectors v_i as *dual* controls. Let $Y : \mathbb{R}^m \to \mathbb{R}^d$ be the affine mapping defined by

$$Y(v) = \text{ the dual state trajectory generated from } v \text{ by the dynamics (4.15).}$$

In expressing $L^+(x,u;y,v)$ in terms of the right side of (4.14) in place of the left, and performing the minimization in (4.11), we now get, as the *full* dual problem,

$$(\mathcal{D}^+) \qquad \begin{array}{l} \text{maximize} \\ \Psi^+(y,v) := \sum_{i\in I} \pi_i \big[q_i{\cdot}v_i - \psi_i^*(v_i) - \phi^*\big(E_i\{B_{i_+}^* y_{i_+}\} + D_i^* v_i - p_i\big)\big] - a{\cdot}y_0 \\ \text{over } y \text{ and } v, \text{ subject to } y - Y(v) = 0. \end{array}$$

Here the equations in (4.15) are taken as a system of linear constraints on the vector variables y_i and v_i. In analogy with the foregoing we can immediately also write down a corresponding *reduced* dual problem, namely

$$(\mathcal{D}^-) \qquad \begin{array}{l} \text{maximize} \\ \Psi^-(v) := \sum_{i\in I} \pi_i \big[q_i{\cdot}v_i - \psi_i^*(v_i) - \phi^*\big(E_i\{B_{i_+}^* y_{i_+}\} + D_i^* v_i - p_i\big)\big] - a{\cdot}y_0 \\ \text{over } v, \text{ where } y = Y(v). \end{array}$$

The feasibility conditions can be written for both problems as

$$v_i \in \operatorname{dom}\psi_i^*, \quad E_i\{B_{i_+}^* y_{i_+}\} + D_i^* v_i - p_i \in \operatorname{dom}\phi_i^*, \quad y = Y(v). \tag{4.16}$$

It's clear that $(\mathcal{D}^+)$ and $(\mathcal{D}^-)$ are equivalent in the sense that $\inf(\mathcal{D}^+) = \inf(\mathcal{D}^-)$ and

$$(\bar{y},\bar{v}) \text{ solves } (\mathcal{D}^+) \iff \bar{v} \text{ solves } (\mathcal{D}^-) \text{ and } \bar{y} = Y(\bar{v}). \tag{4.17}$$

Theorem 4 (Fenchel scheme in the full model). *The full problems $(\mathcal{P}^+)$ and $(\mathcal{D}^+)$ are dual to each other in the extended Fenchel sense with Lagrangian L^+ on $[\mathbb{R}^d \times U] \times [\mathbb{R}^d \times V]$. In this, the piecewise linear-quadratic case is the one in which all the convex functions ϕ_i and ψ_i (or equivalently ϕ_i^* and ψ_i^*) are piecewise linear-quadratic. The primal constraint qualification in (3.3) comes out as the strict feasibility condition*

$$\exists\, u_i \in \operatorname{ri}\operatorname{dom}\phi_i \;\; \textit{with} \;\; q_i - C_i x_i - D_i u_i \in \operatorname{ri}\operatorname{dom}\psi_i \;\; \textit{for } x = X(u),\; i \in I, \tag{4.18}$$

whereas the dual constraint qualification in (3.4) corresponds to the strict feasibility condition

$$\exists\, v_i \in \operatorname{ri}\operatorname{dom}\psi_i^* \;\; \textit{with} \;\; E_i\{B_{i_+}^* y_{i_+}\} + D_i^* v_i - p_i \in \operatorname{ri}\operatorname{dom}\phi_i^* \;\; \textit{for } y = Y(v),\; i \in I. \tag{4.19}$$

Proof. The preceding derivation has shown that these problems fit the framework of Fenchel duality in which ϕ^+ and $(\psi^+)^*$ are the functions in (4.10). In terms of the vector variable $v' = (y,v)$ being dual to $z' = (w,z)$, ψ^+ itself is given by

$$\psi^+(w,z) = \begin{cases} \sum_{i\in I} \psi_i(z_i) & \text{when } w = 0, \\ \infty & \text{when } w \neq 0. \end{cases}$$

Feasibility in this problem, which in the Fenchel scheme takes the form

$$(x,u) \in \operatorname{dom} \phi^+, \qquad M(x,u) \in \operatorname{dom} \psi^+, \tag{4.20}$$

for a certain affine transformation M, has to correspond to (4.4); it's apparent that M must be the transformation that takes (x,u) to $(M_1(x,u), M_2(x,u))$ with $M_1(x,u)$ the element $w = x - X(u)$ and $M_2(x,u)$ the element $z = \{z_i\}_{i\in I}$ with $z_i = q_i - C_i x_i - D_i u_i$. The relative interior of a product of convex sets is the product of their relative interiors, so in replacing $\operatorname{dom} \phi^+$ and $\operatorname{dom} \psi^+$ in (4.20) by $\operatorname{ri} \operatorname{dom} \phi^+$ and $\operatorname{ri} \operatorname{dom} \psi^+$ we simply replace the sets $\operatorname{dom} \phi_i$ and $\operatorname{dom} \psi_i$ in (4.4) by $\operatorname{ri} \operatorname{dom} \phi_i$ and $\operatorname{ri} \operatorname{dom} \psi_i$. This confirms that the strict feasibility conditions in (4.18) do correspond to the constraint qualification obtained for $(\mathcal{P}^+)$ through the theory in §3.

In like manner, the strict feasibility conditions in (4.19) can be seen to correspond to the dual constraint qualification (3.4) as applied to $(\mathcal{D}^+)$. □

The direct connection between the reduced problems $(\mathcal{P}^-)$ and $(\mathcal{D}^-)$ can now be brought to light. For this purpose we define

$$l(u,v) = \text{ common value of both sides of (4.14) when } x = X(u) \text{ and } y = Y(v),$$

so that

$$l(u,v) = -\sum_{i\in I} \pi_i v_i \cdot C_i x_i \quad \text{for } x = X(u) \tag{4.21}$$

but at the same time

$$l(u,v) = -\sum_{i\notin T} \pi_i u_i \cdot E_i\{B^*_{i_+} y_{i_+}\} - y_0 \cdot a \quad \text{for } y = Y(v). \tag{4.22}$$

The value $l(u,v)$ is affine in its dependence on u for fixed v, as well as affine in its dependence on v for fixed u. Next we define the *reduced* Lagrangian L^- by

$$L^-(u,v) := \sum_{i\in I} \pi_i L_i(u_i, v_i) + l(u,v) \quad \text{on } \; U \times V, \tag{4.23}$$

where U and V are given still by (4.7).

Theorem 5 (Fenchel scheme in the reduced model). *The reduced problems $(\mathcal{P}^-)$ and $(\mathcal{D}^-)$ are dual to each other in the extended Fenchel sense with Lagrangian L^- on $U \times V$. In this, the piecewise linear-quadratic case is the one in which all the convex functions ϕ_i and ψ_i (or equivalently ϕ_i^* and ψ_i^*) are piecewise linear-quadratic. Again, the primal* *constraint qualification (3.3) comes out as (4.18), whereas the dual constraint qualification (3.4) comes out as (4.19).*

Proof. Once more we appeal to the principle at the end of §3. The reduced Lagrangian L^- is the difference of the lsc, proper, convex functions

$$\phi^-(u) = \sum_{i\in I} \phi_i(u_i), \qquad [\psi^-]^*(u) = \sum_{i\in I} \psi_i^*(u_i),$$

as restricted to the product of their effective domains, namely $U\times V$, plus terms aggregating to an expression that is affine separately in u and v. Here $[\psi^-]^*$ is conjugate to $\psi^-(u) = \sum_{i\in I} \psi_i(u_i)$. On the basis of the two ways of looking at $l(u,v)$ in (4.21) and (4.22), we calculate that the objective functions specified in $(\mathcal{P}^-)$ and $(\mathcal{D}^-)$ have the form

$$\Phi^-(u) = \sup_{v\in V} L^-(u,v), \qquad \Psi^-(v) = \inf_{u\in U} L^-(u,v),$$

so these problems are indeed the ones that correspond in the Fenchel duality format to the triple (L^-, U, V).

The justification of the constraint qualifications follows the same argument as given in the proof of Theorem 4. □

Theorems 4 and 5 combine immediately with Theorem 2 to give us the following results for stochastic programming problems in either formulation.

Theorem 6 (stochastic programming duality)**.**

(a) *The relation* $\inf(\mathcal{P}^+) = \sup(\mathcal{D}^+) < \infty$ *is guaranteed by (4.18). Then, unless the common optimal value is* $-\infty$ *(so* $(\mathcal{D}^+)$ *has no feasible solution), problem* $(\mathcal{D}^+)$ *is sure to have an optimal solution.*

(b) *The relation* $\inf(\mathcal{P}^+) = \sup(\mathcal{D}^+) > -\infty$ *is guaranteed by (4.19). Then, unless the common optimal value is* ∞ *(so* $(\mathcal{P}^+)$ *has no feasible solution), problem* $(\mathcal{P}^+)$ *is sure to have an optimal solution.*

(c) *In the piecewise linear-quadratic case, the primal and dual constraint qualifications in (4.18) and (4.19) are superfluous and can be replaced simply by the feasibility conditions in (4.4) and (4.16), respectively. In that case,*

$$\inf(\mathcal{P}^+) = \sup(\mathcal{D}^+) \quad \textit{unless} \quad \inf(\mathcal{P}^+) = \infty \textit{ and } \sup(\mathcal{D}^+) = -\infty,$$

(i.e., unless neither $(\mathcal{P}^+)$ *nor* $(\mathcal{D}^+)$ *has a feasible solution). Moreover, when the common optimal value is finite, both problems have an optimal solution.*

(d) *All these results hold equally with the full problems* $(\mathcal{P}^+)$ *and* $(\mathcal{D}^+)$ *replaced by the reduced problems* $(\mathcal{P}^-)$ *and* $(\mathcal{D}^-)$.

Examples of multistage stochastic programming problems that are covered by the results in Theorem 6 are easily generated from the examples in §3. Stochastic *linear* programming is obtained by choosing

$$\phi_i(u_i) = \begin{cases} 0 & \text{if } u_i \in I\!R^{n_i}_+, \\ \infty & \text{if } u_i \notin I\!R^{n_i}_+, \end{cases} \qquad \psi_i(z_i) = \begin{cases} 0 & \text{if } z_i \in I\!R^{m_i}_-, \\ \infty & \text{if } z_i \notin I\!R^{m_i}_-, \end{cases}$$

so that

$$\phi_i^*(w_i) = \begin{cases} 0 & \text{if } w_i \in I\!R^{n_i}_-, \\ \infty & \text{if } w_i \notin I\!R^{n_i}_-, \end{cases} \qquad \psi_i^*(v_i) = \begin{cases} 0 & \text{if } v_i \in I\!R^{m_i}_+, \\ \infty & \text{if } v_i \notin I\!R^{m_i}_+. \end{cases}$$

This model belongs to the piecewise linear-quadratic case, where the theorems are at their best. A much broader model in that category is obtained by choosing

$$\phi_i(u_i) = \begin{cases} \frac{1}{2} u_i{\cdot}P_i u_i & \text{if } u_i \in U_i, \\ \infty & \text{if } u_i \notin U_i, \end{cases} \qquad \psi_i(z_i) = \theta_{V_i,Q_i}(z_i),$$

for nonempty polyhedral sets U_i and V_i and symmetric, positive *semi*definite matrices P_i and Q_i, with θ_{V_i,Q_i} defined as in (3.5). In dualizing one then has

$$\psi_i^*(v_i) = \begin{cases} \frac{1}{2} v_i{\cdot}Q_i v_i & \text{if } v_i \in V_i, \\ \infty & \text{if } v_i \notin V_i, \end{cases} \qquad \phi_i^*(w_i) = \theta_{U_i,P_i}(w_i).$$

This is stochastic *piecewise linear-quadratic* programming.

5. Optimality Conditions and Decomposition

The duality in Theorem 6 has interesting implications for optimality conditions in multistage stochastic programming and how such conditions might be employed in computation.

Theorem 7 (saddle points in stochastic programming).

(a) *In the circumstances of Theorem 6 where* $\inf(\mathcal{P}^+) = \sup(\mathcal{D}^+)$, *one has that*

$$\left.\begin{array}{r} (\bar{x},\bar{u};\bar{y},\bar{v}) \text{ is a saddle point} \\ \text{of } L^+ \text{ on } [I\!R^d \times U] \times [I\!R^d \times V] \end{array}\right\} \iff \left\{\begin{array}{l} (\bar{x},\bar{u}) \text{ solves } (\mathcal{P}^+), \\ (\bar{y},\bar{v}) \text{ solves } (\mathcal{D}^+). \end{array}\right.$$

(b) *In the same circumstances, where equally* $\inf(\mathcal{P}^-) = \sup(\mathcal{D}^-)$, *one has that*

$$\left.\begin{array}{r} (\bar{u},\bar{v}) \text{ is a saddle point} \\ \text{of } L^- \text{ on } U \times V \end{array}\right\} \iff \left\{\begin{array}{l} \bar{u} \text{ solves } (\mathcal{P}^-), \\ \bar{v} \text{ solves } (\mathcal{D}^-). \end{array}\right.$$

(c) *The two saddle point conditions are both equivalent to the following subgradient properties being satisfied:*

$$\left.\begin{aligned} E_i\{B^*_{i_+}\bar{y}_{i_+}\} + D^*_i\bar{v}_i - p_i \;&\in\; \partial\phi_i(\bar{u}_i) \\ q_i - C_i\bar{x}_i - D_i\bar{u}_i \;&\in\; \partial\psi^*_i(\bar{v}_i) \end{aligned}\right\} \text{ for } i \in I, \text{ with } \bar{x} = X(\bar{u}),\ \bar{y} = Y(\bar{v}). \tag{5.1}$$

Proof. This comes from Theorem 3 as applied by way of Theorems 4 and 5. The conditions in (5.1) fall directly out of the saddle point condition for L^+, namely

$$\begin{aligned} &(\bar{x},\bar{u}) \in I\!R^d \times U,\ \ (\bar{y},\bar{v}) \in I\!R^d \times V, \text{ and} \\ &\qquad L^+(x,u;\bar{y},\bar{v}) \ge L^+(\bar{x},\bar{u};\bar{y},\bar{v}) \ge L^+(\bar{x},\bar{u};y,v) \\ &\qquad\qquad \text{for all } \ (x,u) \in I\!R^d \times U,\ \ (y,v) \in I\!R^d \times V. \end{aligned}$$

The maximization of $L^+(\bar{x},\bar{u};y,v)$ in (y,v) can be seen from the formula for L^+ in (4.8) to come down to separate maximizations in the components y_i and v_i. This yields the second set of subgradient relations along with $\bar{x} = X(\bar{u})$. Likewise, by substituting the alternative expression in (4.14) into the formula (4.8), one sees that the minimization of $L^+(x,u;\bar{y},\bar{v})$ in (x,u) corresponds to separate minimizations in x_i and u_i, which furnish the second set of subgradient conditions along with $\bar{y} = Y(\bar{v})$. The fact that (5.1) also describes saddle points $(\bar{u},\bar{u})$ of L^- is evident from (4.6) and (4.17). □

To put a good face on the subgradient conditions in (5.1), we introduce now—in terms of any $\bar{x}$ and $\bar{y}$ in $I\!R^d$, acting as parameter elements—the vectors

$$\bar{p}_i := p_i - E_i\{B^*_{i_+}\bar{y}_{i_+}\}, \quad \bar{q}_i := q_i - C_i\bar{x}_i, \tag{5.2}$$

and an associated family of subproblems, one for each information state $i \in I$:

$(\bar{\mathcal{P}}_i)$ $$\text{minimize } \ \bar{p}_i{\cdot}u_i + \phi_i(u_i) + \psi_i(\bar{q}_i - D_iu_i) \ \text{ in } \ u_i.$$

The dual problem paired with $(\bar{\mathcal{P}}_i)$ in the extended Fenchel format is

$(\bar{\mathcal{D}}_i)$ $$\text{maximize } \ \bar{q}_i{\cdot}v_i - \psi^*_i(v_i) - \phi^*_i(D^*_iv_i - \bar{q}_i) \ \text{ in } \ v_i,$$

and the corresponding Lagrangian is

$$\bar{L}_i(u_i,v_i) := \bar{p}_i{\cdot}u_i + \phi_i(u_i) + \bar{q}_i{\cdot}v_i - \psi_i(v_i) - v_i{\cdot}D_iu_i \ \text{ on } \ U_i \times V_i. \tag{5.3}$$

The facts in Theorems 2 and 3 are available for these problems.

Theorem 8 (problem decomposition by information states). *The optimality conditions in Theorem 7 have the following interpretation:*

$$\begin{cases} \bar{x} = X(\bar{u}), \ \ \bar{y} = Y(\bar{v}), \ \ \textit{and for each } i \in I \\ (\bar{u}_i, \bar{v}_i) \ \textit{ is a saddle point of } \ \bar{L}_i \ \textit{ on } U_i \times V_i. \end{cases} \tag{5.4}$$

Thus, in the piecewise linear-quadratic case in particular, $\bar{u} = \{\bar{u}_i\}$ solves the reduced problem $(\mathcal{P}^-)$ if and only if there exists $\bar{v} = \{\bar{v}_i\}_{i \in I}$ such that, when $\bar{x} = \{\bar{x}_i\}_{i \in I}$ and $\bar{y} = \{\bar{y}_i\}_{i \in I}$ are taken as the trajectories of primal and dual states generated from these controls by the dynamics in (4.2) and (4.15), it turns out that, for every $i \in I$,

$$\begin{cases} \bar{u}_i \ \textit{ is an optimal solution to } (\bar{\mathcal{P}}_i), \\ \bar{v}_i \ \textit{ is an optimal solution to } (\bar{\mathcal{D}}_i). \end{cases}$$

Proof. All we need to observe is that the subgradient conditions in (5.1) are the ones furnished by Theorem 3, as specialized to the problems $(\bar{\mathcal{P}}_i)$ and $(\bar{\mathcal{D}}_i)$. In the piecewise linear-quadratic case, saddle points correspond always to pairs of optimal solutions, as we know from Theorem 2. □

In practical terms, this decomposition result is especially suited to algorithms that in some way alternate between, on the one hand, integrating linear dynamical systems to get states from controls and, on the other hand, solving collections of problems $(\bar{\mathcal{P}}_i)$, perhaps in parallel for various information states $i \in I$. Backward-forward splitting algorithms have just this character; cf. [8]. Other splitting methods can likewise exploit the special Lagrangian structure in (4.23) without relying necessarily on repeated integration of the dynamics; cf. [5] and [6]. It usually wouldn't be required, of course, to solve the corresponding dual problems $(\bar{\mathcal{D}}_i)$ directly, since almost any method for solving $(\bar{\mathcal{P}}_i)$ to get $\bar{u}_i$ would automatically produce $\bar{v}_i$ as an associated multiplier vector.

References

1. R. T. Rockafellar and R. J-B Wets, "Generalized linear-quadratic problems of deterministic and stochastic optimal control in discrete time," SIAM J. Control Opt. 28 (1990), 810–822.
2. L. Korf, "Approximation and solution schemes for dynamic stochastic optimization problems," Ph.D. dissertation, Dept. of Mathematics, University of California at Davis, 1998.

3. R. T. Rockafellar, “Linear-quadratic programming and optimal control,” SIAM J. Control and Opt. 25 (1987), 781–814.

4. R. T. Rockafellar, “Multistage convex programming and discrete-time optimal control,” Control and Cybernetics 17 (1988), 225–246.

5. D. H. Salinger, *A Splitting Algorithm for Multistage Stochastic Programming with Application to Hydropower Scheduling*, Ph.D. dissertation, Dept. of Applied Math., University of Washington, 1997.

6. J. Eckstein and M. C. Ferris, “Operator splitting methods for monotone affine variational inequalities with a parallel application to optimal control,” INFORMS J. on Computing 10 (1998), No. 2.

7. G. H.-G. Chen, *Forward-Backward Splitting Techniques: Theory and Applications*, PhD dissertation, Dept. of Applied Math., University of Washington, 1994.

8. G. H.-G. Chen and R. T. Rockafellar, “Convergence rates in forward-backward splitting,” SIAM J. Optim. 7 (1997), 421–444.

9. R. T. Rockafellar and R. J-B Wets, *Variational Analysis*, Grundlehren der Mathematischen Wissenschaften 317, Springer-Verlag, 1997.

10. R. T. Rockafellar, “Duality theorems for convex functions,” Bull. Amer. Math. Soc. 70 (1964), 189-192.

11. R. T. Rockafellar and R. J-B Wets, “A Lagrangian finite generation technique for solving linear-quadratic problems in stochastic programming,” Math. Programming Studies 28 (1986), 63–93.

12. R. T. Rockafellar and R. J-B Wets, “Linear-quadratic problems with stochastic penalties: the finite generation algorithm,” in *Stochastic Optimization*, V. I. Arkin, A. Shiraev and R. J-B Wets (eds.), Springer-Verlag Lecture Notes in Control and Information Sciences No. 81, 1987, 545-560.

13. R. T. Rockafellar, “Duality and stability for extremum problems involving convex functions,” Pacific J. Math. 21 (1967), 167-187.

14. R. T. Rockafellar, “Lagrange multipliers and optimality,” SIAM Review 35 (1993), 183–238.