Ann Oper Res (2006) 142: 129–146
DOI 10.1007/s10479-006-6165-z

Multistage stochastic convex programs: Duality and its implications

Julia L. Higle · Suvrajeet Sen

Abstract In this paper, we study alternative primal and dual formulations of multistage stochastic convex programs (SP). The alternative dual problems which can be traced to the alternative primal representations, lead to stochastic analogs of standard deterministic constructs such as conjugate functions and Lagrangians. One of the by-products of this approach is that the development does not depend on dynamic programming (DP) type recursive arguments, and is therefore applicable to problems in which the objective function is non-separable (in the DP sense). Moreover, the treatment allows us to handle both continuous and discrete random variables with equal ease. We also investigate properties of the expected value of perfect information (EVPI) within the context of SP, and the connection between EVPI and nonanticipativity of optimal multipliers. Our study reveals that there exist optimal multipliers that are nonanticipative if, and only if, the EVPI is zero. Finally, we provide interpretations of the retroactive nature of the dual multipliers.

Keywords Stochastic Programming · Duality · EVPI

1. Introduction

Stochastic programming (SP) is a powerful modeling paradigm that allows decision making models to incorporate uncertain parameters. One of the main strengths of the SP methodology is its ability to consider the impact of a variety of scenarios when evaluating a proposed solution, in contrast to the more restrictive approach of deterministic optimization models, in which only a single scenario is considered. Also, despite the large scale nature of stochastic optimization models, several successful applications of SP models have been reported in the literature (e.g., Cariño et al, 1994; Sen, Doverspike and Cosares, 1994). Notwithstanding these successes, there remain some conceptual and computational barriers which restrict our current understanding of SP models and algorithms. In an effort to overcome some of these

This work was supported by NSF grant DMII-9414680

J. L. Higle (✉) · Suvrajeet Sen
SIE Dept. University of Arizona, Tucson, AZ 85721
e-mail: julie@sie.arizona.edu

barriers, this paper is devoted to characterizations of dual problems for multistage stochastic convex programs.

In order to preview our results in an economic context, consider an SP model that attempts to study national farm output by minimizing total expected cost of production subject to demand constraints. It is not difficult to envision a multistage stochastic program in which the states of nature ("wet" or "dry") are incorporated using random variables that evolve over time. Note that farmers devise plans for planting prior to observing the state of nature. Crop yields are a consequence of the eventual state of nature and the planting decisions adopted earlier in the season. Since planting decisions are made prior to observing the state of nature, they are said to be nonanticipative. The dual problem we study focuses on relaxing primal constraints that impose nonanticipativity of planting decisions. The dual variables provide a "tax system" in which taxes (collected) and subsidies (paid out) are required to balance each other out across the various scenarios for future weather patterns. Consequently the SP dual requires that from any point in time, the conditional expected value of taxes minus subsidies in future years must be zero. When interpreted in this setting, it is not surprising that the taxes and subsidies depend on the state of nature. For instance, if a certain year is classified as a "dry year", then farmers may be entitled to subsidies on certain crops, whereas, in "wet years", taxes may be levied. Since the precise rates for any given year are applied only after the season (wet or dry) is observed, the rates (taxes/subsidies) are anticipative. It follows that the dual variables studied in this paper are anticipative. We note that this conclusion, which we illustrate with a simple computational example, is at odds with previously published suggestions that at optimality, such variables are nonanticipative (Dempster, 1981, 1988).

As in other areas of optimization, duality has implications for both SP modeling as well as the development of SP algorithms (see e.g. Rockafellar and Wets, 1991; Higle and Sen, 1996a). Our focus in this paper is essentially conceptual; we examine equivalent forms of primal and dual multistage stochastic programs in which information regarding uncertain parameters unfolds over time. Within our framework, we make no distinctions regarding the nature of the random variables involved; discrete and continuous random variables are considered under a common umbrella. Although algorithms typically work with discretizations of continuous distributions (e.g., Birge, 1985a; Rockafellar and Wets, 1991, 1992; Mulvey and Ruszczynski, 1995), this discretization is a potential source of error when the continuous nature of the random variables is essential to model validity. From a computational viewpoint, such error analysis is also useful for approximations of SP (e.g. Birge 1982, 1985b; and Zipkin, 1980) as well as successive refinement algorithms such as those presented in Frauendorfer (1992). More recently, Frauendorfer (1996) has applied two-stage duality in a recursive manner to show convergence of a multistage successive refinement algorithm.

Wright (1994) develops symmetric dual problems for multistage stochastic linear programs which permit both discrete and continuous random variables. Our approach is more direct, and in line with the papers of Rockafellar and Wets, 1976a, b, 1992). The earlier paper (Rockafellar and Wets, 1976a) develops the dual problem using recursive arguments, as in dynamic programming. The more recent paper (Rockafellar and Wets, 1992) is algorithmically motivated, and deals only with the case of discrete random variables. While our setup also focuses on the nonanticipativity requirements of the primal, our proof is based directly on stochastic analogs of deterministic mathematical programming. Hence, no DP recursion is invoked in our proofs. An important by-product of this approach is that we are able to handle instances in which the DP recursion does not apply (e.g., when the stagewise returns are non-separable). We also observe that our treatment of duality does not distinguish between discrete and continuous random variables. All of this is made possible by studying the multistage stochastic convex programs in infinite dimensional spaces. Thanks to the work of Rockafellar, Clarke, Hiriart-Urruty and others (see Clarke, 1983) subdifferential calculus in this setting is well understood, and leads to a much more comprehensive treatment than available from previous studies. Furthermore, we provide a clarification of the connection between EVPI (the expected value of perfect information) and the nonanticipativity multipliers. In particular, we provide a counter-example which establishes that contrary to previous assertions (e.g., Dempster, 1981, 1988), the multipliers associated with the nonanticipativity restrictions are anticipative except for the extremely special case in which perfect information has no value. Furthermore, this example also counters Dempster's claim regarding a supermartingale structure associated with the nonanticipativity multipliers (Dempster, 1981, 1988).

This paper is organized as follows. In §2, we present a generic formulation for a multistage stochastic program. Following a discussion of the nature of the nonanticipativity requirement, we offer two alternate representations of these constraints: the state vector formulation and the mean vector formulation. Assuming convexity of the objective function and the feasible set, in §3 we present a stochastic version of a multistage conjugate dual, as well as a stochastic Lagrangian dual. As may be expected, the two dual problems are equivalent, and more importantly, strong duality holds between these problems and the alternative primal problems in §2. In §4, we illustrate the anticipative nature of the dual variables, using an example for which *all* optimal dual solutions are anticipative. From this example, we present a relationship between the dual variables and the expected value of perfect information (EVPI). In addition, we use this example to note that the dual solutions do not, in general, have an established martingale form. This section highlights the points of divergence between our results and those in Dempster (1981). Finally, in §5, we present various interpretations of the dual multipliers, and our conclusions.

2. Primal formulations

In what follows, we consider a problem in which "decisions", which we denote as x, and random data, which we denote as $\tilde{\omega}$, are interwoven over time. An initial decision is made, after which relevant data are observed. In response to the observation, a subsequent decision is made, after which another observation is made, etc. As a result of the multistage nature of the problems that we consider, our model is one in which both randomness and decisions evolve over time. In stage 1, we have the current (certain) data, denoted ω_1. Data beyond the first stage is uncertain and is modelled through a sequence of random variables $\tilde{\omega}_2, \dots, \tilde{\omega}_T$. We use the index t to denote a stage in the decision problem, $t = 1, \dots, T$, whereas x and $\tilde{\omega}$ are associated with decisions and data, respectively. In this sense, x_t indicates a decision made in stage t and ω_t indicates a realization of the data obtained in stage t. In general, the random data in stage t is denoted as $\tilde{\omega}_t$. The stochastic data process, $\tilde{\omega} = \{\tilde{\omega}_t\}_{t=1}^T$, is defined on a probability space $\{\Omega, \mathcal{A}, \mathcal{P}\}$. Although we consider "randomness" as exogenous to the problem, so that a particular choice of $x = \{x_t\}_{t=1}^T$ does not have a distributional impact on $\tilde{\omega}$, a feasible choice of x is nonetheless dependent upon $\tilde{\omega}$. Thus, for each possible data realization $\omega \in \Omega$, there is a set of feasible solutions, $X(\omega)$, and an objective function $g(x, \omega)$ which influences the choice of x. Finally, throughout our development we will assume that all vectors are appropriately dimensioned and that with probability one, $g(\cdot, \tilde{\omega})$ is a convex function and $X(\tilde{\omega})$ is a convex set.

Within the stochastic programming literature, a realization of $\tilde{\omega}$ is commonly referred to as a scenario. For each scenario $\omega \in \Omega$, we may define a problem, which we refer to as the "scenario problem", as follows

$$\text{Min}\{g(x, \omega) \mid x \in X(\omega) \subseteq \Re^n\}. \qquad (P_\omega)$$

Note that (P_ω) is stated as a typical mathematical program with $x \in \Re^n$. By considering all possible data scenarios, one could define the following, which is often referred to as the "wait and see" problem

$$E[\text{Min}\{g(x(\tilde{\omega}), \tilde{\omega}) \mid x(\tilde{\omega}) \in X(\tilde{\omega}) \subseteq \Re^n \text{ a.s.}\}]. \qquad (1)$$

Note the explicit representation of the dependence of the "decision", x, upon the data $\tilde{\omega}$. Formally, we have $x \in \mathcal{L}^\infty(\Omega, \mathcal{A}, \mathcal{P}, \Re^n)$. Note that (1) offers a model of "posterior" optimization, in which optimization occurs *after* the data sequence, $\tilde{\omega}$, has been revealed (hence the term "wait and see"). A solution to (1) may be described as a vector that is a function of $\tilde{\omega}$. The problem is separable in ω, so that it decomposes into the collection of problems $\{P_\omega\}_{\omega \in \Omega}$, each of which may be solved independently of the others. As such, solutions to (1) allow the sequence of decisions made to vary with the scenario. Of course, these solutions will be somewhat optimistic in that they are derived with full knowledge of the manner in which the future will unfold. Perhaps more importantly, such decisions cannot be implemented because one must know the complete evolution of the data sequence over all T stages before *any* decision can be implemented. In order for these plans to be implementable, we must ensure that scenarios that share a common history up to some point in time implement the same decision at that time. Hence, it is necessary to add constraints which ensure that a decision made in stage t depends only upon the information regarding the data process which is available at that time. These constraints are known as the nonanticipativity constraints (aka, "implementability constraints"), and form a characteristic component of a stochastic programming model. If we let $\mathcal{N}$ denote the linear subspace of $\mathcal{L}^\infty(\Omega, \mathcal{A}, \mathcal{P}, \Re^n)$ consisting of all nonanticipative elements, we may amend (1) to formulate a stochastic programming model as follows:

$$\begin{aligned} \text{Min } & E[g(\ x(\tilde{\omega}), \tilde{\omega})] \\ & x(\tilde{\omega}) \in X(\tilde{\omega}) \qquad \text{a.s.} \\ & x(\tilde{\omega}) \in \mathcal{N} \end{aligned} \qquad (SP)$$

There are a variety of ways in which the nonanticipativity requirements may be modeled. Formally, $x_t(\cdot)$ should be $\mathcal{A}_t$-measurable, where $\mathcal{A}_t$ is the sub-σ-algebra generated by $(\tilde{\omega}_1, \tilde{\omega}_2, \ldots, \tilde{\omega}_t)$. In essence, the constraint $x(\tilde{\omega}) \in \mathcal{N}$ ensures that while $x_t(\omega) \in \Re^{n_t}$ (where $n = \sum_{t=1}^T n_t$), the decision made in period t under the scenario ω, may vary with $\omega_1, \ldots, \omega_t$, it must be conditionally independent of $\omega_{t+1}, \ldots, \omega_T$. Notationally, let $\mathcal{H}_t$ denote an operator that truncates a sequence at the t^{th} stage. Then

$$\mathcal{H}_t\omega = (\omega_1, \omega_2, \ldots, \omega_t)$$

reflects the evolution of the scenario ω through the first t periods. Alternatively, we see that $\mathcal{H}_t\omega$ yields the 'history' associated with scenario ω available in the tth stage. Moreover, if $x = (x_1, x_2, \ldots, x_T)$, denotes a sequence of decisions, then $\mathcal{H}_t x = (x_1, x_2, \ldots, x_t)$ denotes the subsequence of decisions implemented through the tth stage. In this sense, $\mathcal{A}_t$ is the $\mathcal{P}$-completed σ-field defined by the history $\mathcal{H}_t\tilde{\omega}$ of the data process. Next, define the point-to-set map

$$\mathcal{H}_t^{-1}\underline{\omega}_t = \{\omega \in \Omega \mid \mathcal{H}_t\omega = \underline{\omega}_t\},$$

so that $\mathcal{H}_t^{-1}\underline{\omega}_t$ is the set of all possible realizations of $\tilde{\omega}$ whose history at t is $\underline{\omega}_t$. Suppose that $\omega^i \in \Omega$, $i = 1, 2$. Clearly, if the first t components of ω^1 and ω^2 are identical, then $\omega^2 \in \mathcal{H}_t^{-1}(\mathcal{H}_t\omega^1)$. In this case, nonanticipativity requires that $\mathcal{H}_t x(\omega^1) = \mathcal{H}_t x(\omega^2)$. Equivalently, we require

$$x_s(\omega^1) = x_s(\omega^2) \quad s = 1, \ldots, t. \tag{2}$$

Note that

$$\mathcal{H}_t\omega^1 = \mathcal{H}_t\omega^2 \iff \mathcal{H}_s\omega^1 = \mathcal{H}_s\omega^2 \quad \forall s \le t,$$

so that the requirement in (2) can equivalently be stated as

$$x_s(\omega^1) = x_s(\omega^2) \quad \text{whenever} \quad \mathcal{H}_s\omega^1 = \mathcal{H}_s\omega^2.$$

There are a variety of ways in which this requirement may be modeled. In this paper, we consider the implications associated with two alternate representations of the nonanticipativity constraints. The first, which we refer to as the "state vector" representation involves the introduction of "state variables" for each possible value of $\mathcal{H}_t\tilde{\omega}$, $t = 1, \ldots, T$. To begin, let $z \in \mathcal{L}^\infty(\Omega, \mathcal{A}, \mathcal{P}, \Re^n)$, where $z = (z_t)_{t=1}^T$ and $z_t : \mathcal{H}_t\Omega \to \Re^{n_t}$. We refer to $z_t(\mathcal{H}_t\omega)$ as the period t state variable when the history of the data process at that time is $\mathcal{H}_t\omega$. That is, there is one such variable for each possible "state" of $\tilde{\omega}$ at each stage. Nonanticipativity is ensured by constraining the appropriate subsets of the decision variables, $\{\{x_t(\omega)\}_{\omega\in\Omega}\}_{t=1}^T$, to be equal to these state variables. Note that in case of two stage problems with finitely many outcomes, this state variable formulation reduces to the "split variable" formulation (Dempster, 1988). The state vector representation of the nonanticipativity constraints may be written as $x(\tilde{\omega}) - z(\tilde{\omega}) = 0$ a.s., or equivalently

$$x_t(\tilde{\omega}) - z_t(\mathcal{H}_t\tilde{\omega}) = 0 \quad \text{a.s.,} \quad t = 1, \ldots, T. \tag{3}$$

Note that (3) explicitly requires that $x_t(\omega)$ be almost surely constant for all scenarios that share a common history through period t.

We assume that the constraints, $x(\tilde{\omega}) \in X(\tilde{\omega})$, have *relatively complete recourse*, so that if $X(\tilde{\omega}) = X_1(\tilde{\omega}) \times X_2(\tilde{\omega}) \times \ldots \times X_T(\tilde{\omega})$, then $X_t(\tilde{\omega})$ is $\mathcal{A}_t$-measurable. That is, we assume that if $x_t(\tilde{\omega})$ appears to be feasible on the basis of all decisions and observations made through time t, it cannot be rendered infeasible as a result of some event that can only be observed at a later time. This constraint qualification is common in the stochastic programming literature, and appears, for example, in Rockafellar and Wets (1976a,b, 1982), Dempster (1988), and Wets (1989).

Prior to introducing the second representation of the nonanticipativity constraints, we note that throughout this paper we will depend heavily upon arguments derived using conditional expectations. Suppose that $f : \Omega \to \Re^m$ is a $\mathcal{P}$-measurable function. Then

$$E[f(\tilde{\omega})] = \int_\Omega f(\omega)\mathcal{P}(d\omega).$$

We may write the expectation by conditioning on $\mathcal{H}_t\omega$, so that

$$\begin{aligned} E[f(\tilde{\omega})] &= \int_{\Omega} E[f(\tilde{\omega}) \mid \tilde{\omega} \in \mathcal{H}_t^{-1}(\mathcal{H}_t\omega)]\mathcal{P}(d\omega) \\ &= E\{E[f(\tilde{\omega}) \mid \tilde{\omega} \in \mathcal{H}_t^{-1}(\mathcal{H}_t\tilde{\omega}')]\}, \\ &= E\{E[f(\tilde{\omega}') \mid \tilde{\omega}' \in \mathcal{H}_t^{-1}(\mathcal{H}_t\tilde{\omega})]\}, \end{aligned}$$

where $\tilde{\omega}$ and $\tilde{\omega}'$ are defined on the same probability space.

The second representation of the nonanticipativity constraints, which we refer to as the "mean vector" representation replaces the state vector $z_t(\mathcal{H}_t\tilde{\omega})$ in (3) with the conditional expectation of the vectors with which it is associated. That is,

$$x_t(\tilde{\omega}) - E\big[x_t(\tilde{\omega}') \mid \tilde{\omega}' \in \mathcal{H}_t^{-1}(\mathcal{H}_t\tilde{\omega})\big] = 0 \quad \text{a.s.,} \quad t = 1, \ldots, T, \tag{4}$$

where $\tilde{\omega}$ and $\tilde{\omega}'$ are defined on the same probability space. One can easily note that both (3) and (4) ensure that $x_t(\omega)$ is constant for almost every $\omega \in \mathcal{H}_t^{-1}(\mathcal{H}_t\omega')$, for almost every $\omega' \in \Omega$ and thus specify the same set of nonanticipative solutions to (SP). It follows that equivalent problems result when either of them are used in the formulation of (SP).

In order to focus our study on the implications of the specific form of the nonanticipativity constraints used, we introduce the following extended real-valued function

$$\phi(x(\omega), \omega) = \begin{cases} g(x(\omega), \omega) & \text{if } x(\omega) \in X(\omega) \\ \infty & \text{otherwise.} \end{cases} \tag{5}$$

With this function, we may now specify the state vector and mean vector formulations of a multistage stochastic program as follows.

State vector formulation

$$\min_{(x,z)\in\mathcal{L}^\infty(\Omega,\mathcal{A},\mathcal{P},\Re^{2n})} E[\phi(x(\tilde{\omega}), \tilde{\omega})] \tag{P–SV}$$
$$\text{s.t.} \quad x_t(\tilde{\omega}) - z_t(\mathcal{H}_t\tilde{\omega}) = 0 \quad \text{a.s.,} \quad t = 1, \ldots, T.$$

Mean vector formulation

$$\min_{x\in\mathcal{L}^\infty(\Omega,\mathcal{A},\mathcal{P},\Re^{n})} E[\phi(x(\tilde{\omega}), \tilde{\omega})] \tag{P–MV}$$
$$\text{s.t.} \quad x_t(\tilde{\omega}) - E\big[x_t(\tilde{\omega}') \mid \tilde{\omega}' \in \mathcal{H}_t^{-1}(\mathcal{H}_t\tilde{\omega})\big] = 0 \quad \text{a.s.,} \quad t = 1, \ldots, T.$$

The mean vector formulation is the more common problem addressed in the SP literature, and appears for example in Dempster (1988). However in the development that follows, the state vector formulation provides a more convenient avenue into duality.

3. Dual problems

As with the primal problems P-SV and P-MV, the dual problems we study are valid for both continuous and discrete random variables. In this section, we propose stochastic analogs of conjugate, as well as Lagrangian, dual problems. One of the key features that distinguish these duals from their deterministic counterparts is the role played by multipliers associated with the nonanticipativity constraints (Wets, 1975). In the development that follows, it will be convenient to study the duality between the stochastic conjugate dual and the primal problem stated as P-SV under the assumption of convexity. Consequently, we will refer to this dual as D-SV. On the other hand, the connections between the stochastic Lagrangian dual will be more readily apparent via its relationship to P-MV. Consequently, we shall refer to the stochastic Lagrangian dual as D-MV.

Throughout our development, we will introduce a number of "variables" which are actually measurable functions of random variables. For example, $\{x(\omega)\}_{\omega\in\Omega}$, or equivalently, $x(\tilde{\omega})$, is one such measurable function mapping Ω to $\Re^n$. For notational convenience, let $\mathcal{L}_n^\infty$ and $\mathcal{L}_n^1$ denote $\mathcal{L}^\infty(\Omega, \mathcal{A}, \mathcal{P}, \Re^n)$ and $\mathcal{L}^1(\Omega, \mathcal{A}, \mathcal{P}, \Re^n)$, respectively. For $\xi \in \mathcal{L}_n^1$ and $x \in \mathcal{L}_n^\infty$ we define the following linear operation:

$$\xi \circ x = \int_\Omega \xi(\omega)^\top x(\omega)\mathcal{P}(d\omega) = E[\xi(\tilde{\omega})^\top x(\tilde{\omega})], \tag{6}$$

which we recognize as the expected value of the traditional counterpart from deterministic mathematical programming.

A Stochastic Conjugate Dual

Consider the function ϕ defined in (5) and its conjugate function

$$\phi^*(\sigma, \omega) = \sup_{x\in\Re^n} \{\sigma^\top x - \phi(x, \omega)\}, \tag{7}$$

where $\sigma = (\sigma_1, \ldots, \sigma_T) \in \Re^n$. In what follows, we assume that $\phi(\cdot, \omega)$ is a convex function on its effective domain and verify that the following problem is dual to P-SV.

$$\begin{aligned} &\sup_{\sigma\in\mathcal{L}_n^1} && -E[\phi^*(\sigma(\tilde{\omega}), \tilde{\omega})] && \text{(D} - \text{SV)} \\ &\text{s.t.} && E\big[\sigma_t(\tilde{\omega}) \mid \tilde{\omega} \in \mathcal{H}_t^{-1}(\mathcal{H}_t\tilde{\omega}')\big] = 0 \quad \text{a.s.} \quad t = 1, \ldots, T, \end{aligned}$$

where $\tilde{\omega}$ and $\tilde{\omega}'$ are defined on the same probability space. The rationale for calling it a dual problem is found in Theorem 3. We begin the development with the following lemma which provides a characterization of the primal and dual feasible solutions for the state vector formulation.

Lemma 1. *If $x \in \mathcal{L}_n^\infty$ and $\sigma \in \mathcal{L}_n^1$ satisfy the constraints of P-SV and D-SV, respectively, then $E[\sigma(\tilde{\omega})^\top x(\tilde{\omega})] = 0$ (equivalently, $\sigma \circ x = 0$).*

Proof: Note that $\sigma(\omega)^\top x(\omega) = \sum_{t=1}^T \sigma_t(\omega)^\top x_t(\omega)$, so that

$$\begin{aligned} E[\sigma(\tilde{\omega})^\top x(\tilde{\omega})] &= \sum_{t=1}^T E\big[\sigma_t(\tilde{\omega})^\top x_t(\tilde{\omega})\big] \\ &= \sum_{t=1}^T E\left\{E\big[\sigma_t(\tilde{\omega})^\top x_t(\tilde{\omega}) \,\big|\, \tilde{\omega} \in \mathcal{H}_t^{-1}(\mathcal{H}_t\tilde{\omega}')\big]\right\}, \end{aligned}$$

where $\tilde{\omega}$ and $\tilde{\omega}'$ are defined on the same probability space. By hypothesis, $x(\tilde{\omega})$ is feasible to P-SV, so that $x_t(\omega)$ is almost surely constant on subsets of Ω for which the scenarios share a common history through period t. That is, for some $z_t(\cdot)$, it follows that $x_t(\omega) = z_t(\mathcal{H}_t\omega')$

for almost every $\omega \in \mathcal{H}_t^{-1}(\mathcal{H}_t\omega')$, for almost every $\omega' \in \Omega$. Thus, we have

$$\begin{aligned} E\big[\sigma_t(\tilde{\omega})^\top x_t(\tilde{\omega}) \,\big|\, \tilde{\omega} \in \mathcal{H}_t^{-1}(\mathcal{H}_t\omega)\big] &= E\big[\sigma_t(\tilde{\omega})^\top z_t(\mathcal{H}_t\omega) \,\big|\, \tilde{\omega} \in \mathcal{H}_t^{-1}(\mathcal{H}_t\omega)\big] \\ &= E\big[\sigma_t(\tilde{\omega}) \,\big|\, \tilde{\omega} \in \mathcal{H}_t^{-1}(\mathcal{H}_t\omega)\big]^\top z_t(\mathcal{H}_t\omega) \end{aligned}$$

for almost every $\omega \in \Omega$, $t = 1, \ldots, T$. In addition, $\sigma(\tilde{\omega})$ is feasible to D-SV, which ensures that $E[\sigma_t(\tilde{\omega}) \mid \tilde{\omega} \in \mathcal{H}_t^{-1}(\mathcal{H}_t\omega)] = 0$, for almost every $\omega \in \Omega$. Thus, for $t = 1, \ldots, T$,

$$E\big[\sigma_t(\tilde{\omega})^\top x_t(\tilde{\omega}) \,\big|\, \tilde{\omega} \in \mathcal{H}_t^{-1}(\mathcal{H}_t\omega)\big] = 0$$

for almost every $\omega \in \Omega$. It follows that

$$E[\sigma_t(\tilde{\omega})^\top x_t(\tilde{\omega})] = 0, \quad t = 1, \ldots, T$$

and thus

$$E[\sigma(\tilde{\omega})^\top x(\tilde{\omega})] = 0.$$ □

Next we characterize the normal cone associated with the feasible solutions to the constraints in (3). Of course, since (3) involves only linear equality constraints, this cone is identical for all feasible solutions.

Lemma 2. *Let $x = (x_1, \ldots, x_T)$ and $z = (z_1, \ldots, z_T)$, such that $x_t \in \mathcal{L}^\infty_{n_t}$ and $z_t \in \mathcal{L}^\infty_{n_t}$ ($\sum_{t=1}^T n_t = n$), and*

$$S = \{(x, z) \mid x_t(\tilde{\omega}) - z_t(\mathcal{H}_t\tilde{\omega}) = 0 \quad \text{a.s. } t = 1, \ldots, T\}.$$

Let $\eta^x = (\eta^x_1, \ldots, \eta^x_T)$, $\eta^z = (\eta^z_1, \ldots, \eta^z_T)$, with $\eta^x_t \in \mathcal{L}^1_{n_t}$ and $\eta^z_t \in \mathcal{L}^1_{n_t}$. Let $\eta = (\eta^x, \eta^z)$ and define

$$N_S = \big\{(\eta^x, \eta^z) \in \mathcal{L}^1_{2n} \,\big|\, E\big[\eta^x_t(\tilde{\omega}) + \eta^z_t(\tilde{\omega}) \,\big|\, \tilde{\omega} \in \mathcal{H}_t^{-1}(\mathcal{H}_t\tilde{\omega}')\big] = 0, \; a.s. \quad t = 1, \ldots, T\big\}$$

where $\tilde{\omega}$ and $\tilde{\omega}'$ are defined on the same probability space. Then N_S is the normal cone to S at any point $(x, z) \in S$.

Proof: Suppose that $(x, z) \in S$, and that $\tilde{\omega}$ and $\tilde{\omega}'$ are defined on the same probability space, $(\Omega, \mathcal{A}, \mathcal{P})$. Using (6) we have

$$\begin{aligned} (\eta^x, \eta^z) \circ (x, z) &= \eta^x \circ x + \eta^z \circ z \\ &= \sum_{t=1}^T \big(\eta^x_t \circ x_t + \eta^z_t \circ z_t\big) \\ &= \sum_{t=1}^T E\big[\eta^x_t(\tilde{\omega})^\top x_t(\tilde{\omega}) + \eta^z_t(\tilde{\omega})^\top z_t(\mathcal{H}_t\tilde{\omega})\big] \\ &= \sum_{t=1}^T E\big\{E\big[\eta^x_t(\tilde{\omega})^\top x_t(\tilde{\omega}) + \eta^z_t(\tilde{\omega})^\top z_t(\mathcal{H}_t\tilde{\omega}) \mid \tilde{\omega} \in \mathcal{H}_t^{-1}(\mathcal{H}_t\tilde{\omega}')\big]\big\}. \end{aligned}$$

Since $(x, z) \in S$, $x_t(\tilde{\omega}) = z_t(\mathcal{H}_t\tilde{\omega})$ a.s., and thus $x_t(\omega) = z_t(\mathcal{H}_t\omega) = z_t(\mathcal{H}_t\omega')$ for almost every $\omega \in \mathcal{H}_t^{-1}(\mathcal{H}_t\omega')$, for almost every $\omega' \in \Omega$, $t = 1, \ldots, T$. Thus,

$$\begin{aligned} &E\left[\eta_t^x(\tilde{\omega})^\top x_t(\tilde{\omega}) + \eta_t^z(\tilde{\omega})^\top z_t(\mathcal{H}_t\tilde{\omega}) \,\middle|\, \tilde{\omega} \in \mathcal{H}_t^{-1}(\mathcal{H}_t\omega')\right] \\ &\quad = E\left[\left(\eta_t^x(\tilde{\omega}) + \eta_t^z(\tilde{\omega})\right) \,\middle|\, \tilde{\omega} \in \mathcal{H}_t^{-1}(\mathcal{H}_t\omega')\right]^\top z_t(\mathcal{H}_t\omega') \end{aligned}$$

Thus,

$$(\eta^x, \eta^z) \circ (x, z) = 0 \quad \forall (x, z) \in S$$

if, and only if

$$E\left[\eta_t^x(\tilde{\omega}) + \eta_t^z(\tilde{\omega}) \,\middle|\, \tilde{\omega} \in \mathcal{H}_t^{-1}(\mathcal{H}_t\tilde{\omega}')\right] = 0 \quad \text{a.s.} \quad t = 1, \ldots, T$$

and the result follows. □

With Lemmas 1 and 2, we may now establish the primal-dual relationship between P-SV and D-SV. The duality result presented below draws upon the extended calculus presented in Clarke (1983) (see Section 2.9). We note that in this development the subdifferential is a subset of $\mathcal{L}_n^1$.

Theorem 1. *Let $\phi(\cdot, \tilde{\omega})$, as defined in (5), be a convex normal integrand, and assume that P-SV has relatively complete recourse. Let v_p and v_d denote the optimal values of P-SV and D-SV, respectively. Then*

a) $v_p \geq v_d$.

b) Let P-SV possess an optimal solution denoted $(\hat{x}, \hat{z})$, and assume that $\partial\phi(\hat{x}(\tilde{\omega}), \tilde{\omega})$ is non-empty (a.s). Then there exists $\hat{\sigma}(\tilde{\omega}) \in \partial\phi(\hat{x}(\tilde{\omega}), \tilde{\omega})$ a.s., such that

$$E\left[\hat{\sigma}_t(\tilde{\omega}) \,\middle|\, \tilde{\omega} \in \mathcal{H}_t^{-1}(\mathcal{H}_t\tilde{\omega}')\right] = 0 \quad \text{(a.s.)},$$

where $\tilde{\omega}$ and $\tilde{\omega}'$ are defined on the same probability space. Furthermore, $-E[\phi^*(\hat{\sigma}(\tilde{\omega}), \tilde{\omega})] = v_d = v_p$.

Proof:

a) If D-SV is infeasible, $v_d = -\infty$ and the result follows. Similarly, if P-SV is infeasible, $v_p = +\infty$ and the result follows. Thus, suppose that x and σ are feasible in P-SV and D-SV, respectively. It follows from the definition of ϕ^* in (7) that

$$\begin{aligned} &\phi^*(\sigma(\omega), \omega) \geq \sigma(\omega)^\top x(\omega) - \phi(x(\omega), \omega) \quad \forall\, \omega \in \Omega \\ &\quad \Rightarrow E[\phi^*(\sigma(\tilde{\omega}), \tilde{\omega})] \geq E[\sigma(\tilde{\omega})^\top x(\tilde{\omega})] - E[\phi(x(\tilde{\omega}), \tilde{\omega})]. \end{aligned}$$

As a result of Lemma 1, feasibility of σ and x ensures that $E[\sigma(\tilde{\omega})^\top x(\tilde{\omega})] = 0$, so that

$$\begin{aligned} &E[\phi^*(\sigma(\tilde{\omega}), \tilde{\omega})] \geq -E[\phi(x(\tilde{\omega}), \tilde{\omega})] \\ &\Rightarrow E[\phi(x(\tilde{\omega}), \tilde{\omega})] \geq -E[\phi^*(\sigma(\tilde{\omega}), \tilde{\omega})] \end{aligned}$$

for all feasible x and σ, and thus

$$v_p \geq v_d. \tag{8}$$

b) For notational convenience, let $\Phi(x) = E[\phi(x(\tilde{\omega}), \tilde{\omega})]$ and note that $\partial\Phi(x) \subset \mathcal{L}^1_n$. For $(x, z) \in \mathcal{L}^\infty_{2n}$, let

$$\psi(x, z) = \begin{cases} 0 & \text{if } x_t(\tilde{\omega}) - z_t(\mathcal{H}_t\tilde{\omega}) = 0, \quad a.s., \quad t = 1, \dots T \\ \infty & \text{otherwise.} \end{cases}$$

Note that (x, z) is feasible to P-SV if, and only if, $\psi(x, z) = 0$. Furthermore, $(\hat{x}, \hat{z})$ is an optimal solution to P-SV, if, and only if, it is an optimal solution to

$$\operatorname*{Min}_{(x,z)\in\mathcal{L}^\infty_{2n}} \ \Phi(x) + \psi(x, z).$$

Let $\partial_x\phi$ and $\partial_z\phi$ denote the projection of $\partial\phi$ on the x and z coordinates, respectively. Then following Clarke (1983), we have

$$0 \in (\partial\Phi(\hat{x}) + \partial_x\psi(\hat{x}, \hat{z}), \partial_z\psi(\hat{x}, \hat{z})).$$

Here the "0" denotes an element in $\mathcal{L}^1_{2n}$ that is equal to zero almost surely. From convex analysis, it is well known that $\partial\psi(\hat{x}, \hat{z}) = N_S$, the normal cone associated with the set S which provides the state-variable formulation of non-anticipativity (see Lemma 2). Thus, there exists $(\eta^x, \eta^z) \in N_S$ and $\hat{\sigma} \in \mathcal{L}^1_n$ such that $\hat{\sigma} \in \partial\Phi(\hat{x})$ almost surely, and

$$(\hat{\sigma}(\tilde{\omega}) + \eta^x(\tilde{\omega}) \ , \ \eta^z(\tilde{\omega})) = 0 \quad \text{a.s.}$$

Thus,

$$\hat{\sigma}_t(\tilde{\omega}) + \eta^x_t(\tilde{\omega}) = 0 \quad \text{a.s.,} \quad t = 1, \dots, T$$

and

$$\eta^z_t(\tilde{\omega}) = 0 \quad \text{a.s.,} \quad t = 1, \dots, T.$$

Appealing to Lemma 2, we see that

$$\begin{aligned} \hat{\sigma}_t(\tilde{\omega}) &= -\eta^x_t(\tilde{\omega}) \quad \text{a.s.,} \quad t = 1, \dots, T \\ \Rightarrow E\big[\hat{\sigma}_t(\tilde{\omega}) \,\big|\, \tilde{\omega} \in \mathcal{H}_t^{-1}(\mathcal{H}_t\tilde{\omega}')\big] &= -E\big[\eta^x_t(\tilde{\omega}) \,\big|\, \tilde{\omega} \in \mathcal{H}_t^{-1}(\mathcal{H}_t\tilde{\omega}')\big] \\ &= E\big[\eta^z_t(\tilde{\omega}') \,\big|\, \tilde{\omega} \in \mathcal{H}_t^{-1}(\mathcal{H}_t\tilde{\omega}')\big] \qquad \text{a.s.} \quad t = 1, \dots, T \\ &= 0 \qquad \text{a.s. } t = 1, \dots, T, \end{aligned}$$

so that $\hat{\sigma}$ is feasible to D-SV. Finally, given our assumption of relatively complete recourse and the finiteness of $\Phi(\hat{x})$, Rockafellar and Wets (1982) ensures that

$$\partial\Phi(\hat{x}) = \int_\Omega \partial\phi(\hat{x}(\omega), \omega)\mathcal{P}(d\omega),$$

so that $\hat{\sigma}(\tilde{\omega}) \in \partial\phi(\hat{x}(\tilde{\omega}), \tilde{\omega})$ a.s. It follows that

$$\hat{x}(\tilde{\omega}) \in \underset{\mathcal{L}_n^\infty}{\operatorname{argmax}}\{\hat{\sigma}(\tilde{\omega})^\top x(\tilde{\omega}) - \phi(x(\tilde{\omega}), \tilde{\omega})\}, \qquad \text{a.s.}$$

so that

$$\phi^*(\hat{\sigma}(\tilde{\omega}), \tilde{\omega}) = \hat{\sigma}(\tilde{\omega})^\top \hat{x}(\tilde{\omega}) - \phi(\hat{x}(\tilde{\omega}), \tilde{\omega}) \qquad \text{a.s.}$$

Thus,

$$-v_d \leq E[\phi^*(\hat{\sigma}(\tilde{\omega}), \tilde{\omega})] = E[\hat{\sigma}(\tilde{\omega})^\top \hat{x}(\tilde{\omega})] - E[\phi(\hat{x}(\tilde{\omega}), \tilde{\omega})] \leq E[\hat{\sigma}(\tilde{\omega})^\top \hat{x}(\tilde{\omega})] - v_p.$$

From Lemma 1, $E[\hat{\sigma}(\tilde{\omega})^\top \hat{x}(\tilde{\omega})] = 0$, so that $v_d \geq v_p$. In combination with (8), it follows that

$$v_d = -E[\phi^*(\hat{\sigma}(\tilde{\omega}), \tilde{\omega})] = E[\phi(\hat{x}(\tilde{\omega}), \tilde{\omega})] = v_p. \qquad \square$$

Note that the stochastic programming constraint qualification of relatively complete recourse implies that no induced constraints are necessary to ensure feasibility, so that the operations of expectation and subdifferentiation may be interchanged. For an example that violates these conditions, we refer the reader to Wets (1989), where multipliers associated with induced constraints become necessary.

A Stochastic Lagrangian Dual

Recall that it is a trivial matter to establish the equivalence between the two primal statements of (SP), P-SV and P-MV. By the same token, there is an equivalent dual problem that can be motivated by a certain Lagrangian dual associated with P-MV, which we denote as D-MV.

For $\mu \in \mathcal{L}_n^1$, we define $\bar{\mu}$ as follows:

$$\bar{\mu}(\omega) = \{\bar{\mu}_t(\mathcal{H}_t\omega)\}_{t=1}^T, \text{ where } \bar{\mu}_t(\omega) = E\left[\mu_t(\tilde{\omega}) \,\middle|\, \tilde{\omega} \in \mathcal{H}_t^{-1}(\mathcal{H}_t\omega)\right]. \tag{9}$$

That is, $\bar{\mu}(\omega)$ yields the stagewise conditional expectations associated with $\mu(\tilde{\omega})$, given the scenario ω. Note that with this definition, the constraints (4) in P-MV may equivalently be stated as

$$x(\tilde{\omega}) - \bar{x}(\tilde{\omega}) = 0 \quad \text{a.s.}$$

The following Lemma will prove useful in establishing a Lagrangian dual for P-MV.

Lemma 3. *Suppose that $\mu \in \mathcal{L}_n^1$ and $x \in \mathcal{L}_n^\infty$ and that $\bar{\mu}$ and $\bar{x}$ are defined from μ and x, respectively, as in* (9)*. Then using* (6)

$$\mu \circ \bar{x} = \bar{\mu} \circ \bar{x} = \bar{\mu} \circ x.$$

Proof: Let $\tilde{\omega}$ and $\tilde{\omega}'$ be defined on the same probability space. Then

$$\begin{aligned}
\mu \circ \bar{x} &= E[\mu^\top(\tilde{\omega})\bar{x}(\tilde{\omega})] \\
&= \sum_{t=1}^{T} E\left[\mu_t^\top(\tilde{\omega})\bar{x}_t(\tilde{\omega})\right] \\
&= \sum_{t=1}^{T} E\left\{E\left[\mu_t^\top(\tilde{\omega})\bar{x}_t(\tilde{\omega}) \,\middle|\, \tilde{\omega} \in \mathcal{H}_t^{-1}(\mathcal{H}_t\tilde{\omega}')\right]\right\} \\
&= \sum_{t=1}^{T} E\left\{E\left[\mu_t^\top(\tilde{\omega}) \,\middle|\, \tilde{\omega} \in \mathcal{H}_t^{-1}(\mathcal{H}_t\tilde{\omega}')\right]\bar{x}_t(\mathcal{H}_t\tilde{\omega}')\right\} \\
&= \sum_{t=1}^{T} E\left\{\bar{\mu}_t^\top(\tilde{\omega}')\bar{x}_t(\tilde{\omega}')\right\} \\
&= \bar{\mu} \circ \bar{x}.
\end{aligned}$$

A symmetric argument yields $\bar{\mu} \circ x = \bar{\mu} \circ \bar{x}$, and the result follows. □

As a result of Lemma 4, $\mu \circ (x - \bar{x}) = (\mu - \bar{\mu}) \circ x$. Thus, from P-MV we define the following Lagrangian function.

$$L(\mu, \omega) = \sup_{x \in \Re^n} \{(\mu(\omega) - \bar{\mu}(\omega))^\top x - \phi(x, \omega)\}. \tag{10}$$

We offer the following equivalent dual

$$\sup_{\mu \in \mathcal{L}_n^1} -E[L(\mu(\tilde{\omega}), \tilde{\omega})]. \tag{D−MV}$$

The equivalence between D-SV and D-MV is easily established. Note that for any σ that is feasible to D-SV, $\bar{\sigma}(\tilde{\omega}) = 0$ almost surely, where $\bar{\sigma}$ is defined from σ as in (9). Moreover, for a given μ, $\mu - \bar{\mu}$ is feasible to D-SV. Thus, solutions to the unconstrained D-MV are easily converted to feasible solutions to D-SV. This observation, coupled with the fact that (7) and (10) yield

$$L(\mu, \omega) = \phi^*(\mu(\omega) - \bar{\mu}(\omega), \omega)$$

establishes the equivalence of D-SV and D-MV. Given the equivalence of the primal problems P-SV and P-MV as well as the equivalence of D-SV and D-MV, Theorem 3 ensures that D-MV is a dual for the multistage stochastic convex program. For the sake of completeness, we state this result as a corollary to Theorem 3.

Corollary 1. *Let $\phi(\cdot, \tilde{\omega})$, as defined in* (5), *be a convex normal integrand, and assume that P-MV has relatively complete recourse. Let v_p and v_d denote the optimal values of P-MV and D-MV, respectively.*

a) If $\hat{\sigma}$ solves D-SV, then $\hat{\sigma}$ solves D-MV. Similarly, if $\hat{\mu}$ solves D-MV, then $\hat{\mu} - \bar{\hat{\mu}}$ solves D-SV.

b) Let P-MV possess an optimal solution denoted $\hat{x}$, and assume that $\partial\phi(\hat{x}, \tilde{\omega})$ is non-empty (a.s.). Then there exists $\hat{\mu} \in \mathcal{L}_n^1$ such that $\hat{\mu}(\tilde{\omega}) - \bar{\hat{\mu}}(\tilde{\omega}) \in \partial\phi(x(\tilde{\omega}), \tilde{\omega})$ a.s., and $-E[L(\hat{\mu}, \tilde{\omega})] = v_d = v_p$.

4. Nonanticipativity and the expected value of perfect information

The dual variables σ and μ may be seen as multipliers for the nonanticipativity constraints for the primal problems P-SV and P-MV respectively. In interpreting these multipliers, it is of interest to study whether these quantities are nonanticipative. Dempster (1981) was among the first to address this question, and concluded the existence of multipliers that are nonanticipative. In the following, we illustrate that in general, the multipliers σ and μ are anticipative, and moreover, nonanticipativity of these multipliers arises only under extremely restrictive circumstances.

Example. Consider a three stage LP, which serves as the basis for our example:

$$\text{Min} \sum_{t=1}^{3} c_t x_t$$

$$\text{s.t. } x_t - x_{t+1} \geq 0 \quad t = 1, 2$$

$$-1 \leq x_t \leq 1 \quad t = 1, 2, 3$$

Within this example, the objective coefficients are random variables, with $c_1 = 1$, $c_2 \in \{-1, 1\}$, and $c_3 \in \{-1, 1\}$, so that there are four possible outcomes for the vector of cost coefficients,

$$(c_1, c_2, c_3) \in \Omega = \{(1, -1, -1),\ (1, -1, 1),\ (1, 1, -1)\ (1, 1, 1)\}.$$

Notationally, we will denote these scenarios as corresponding to $\{\omega^i\}_{i=1}^4$. That is, $c_2(\omega^1) = -1$, while $c_3(\omega^2) = 1$, etc. With regard to the evolution of the random variables, we note that $\mathcal{H}_1\omega^i = 1$, $i = 1, 2, 3, 4$, $\mathcal{H}_2\omega^i = (1, -1)$, $i = 1, 2$, $\mathcal{H}_2\omega^i = (1, 1)$, $i = 3, 4$, and $\mathcal{H}_3\omega^i = \omega^i$, $i = 1, 2, 3, 4$. We note also that

$$\mathcal{H}_1^{-1}(1) = \{(1, -1, -1),\ (1, -1, 1),\ (1, 1, -1),\ (1, 1, 1)\} = \{\omega^i\}_{i=1}^4 \tag{11a}$$

$$\mathcal{H}_2^{-1}(1, -1) = \{(1, -1, -1),\ (1, -1, 1)\} = \{\omega^1, \omega^2\} \tag{11b}$$

$$\mathcal{H}_2^{-1}(1, 1) = \{(1, 1, -1),\ (1, 1, 1)\} = \{\omega^3, \omega^4\}. \tag{11b}$$

In the stochastic programming formulation, we associate decision variables with each possible outcome $\{(x_1(\omega^i), x_2(\omega^i), x_3(\omega^i))\}_{i=1}^4$. Thus, from the groupings of the possible outcomes associated with the cost coefficients in (11), nonanticipativity restrictions require that

$$x_1(\omega^1) = x_1(\omega^2) = x_1(\omega^3) = x_1(\omega^4)$$

$$x_2(\omega^1) = x_2(\omega^2) \quad \text{and} \quad x_2(\omega^3) = x_2(\omega^4).$$

These restrictions arise from the commonality of the data sequence (i.e., $c_1(\omega^i) = 1$ for all i, while $c_2(\omega^1) = c_2(\omega^2)$, and $c_2(\omega^3) = c_2(\omega^4)$). Note that the nonanticipativity requirements depend upon the structure of the possible evolution of the realizations of the data, and are not dependent upon the probability distribution associated with these realizations. In our example, we assume that all four outcomes are equally likely.

Using the mean vector formulation, P-MV, we obtain

$$\text{Min} \quad \sum_{i=1}^{4} \frac{1}{4} \left\{ \sum_{t=1}^{3} c_t(\omega^i) x_t(\omega^i) \right\} \tag{12}$$

$$\begin{aligned} \text{s.t.} \qquad & x_t(\omega^i) - x_{t+1}(\omega^i) \geq 0 \quad t = 1, 2\ i = 1, ..., 4 \\ & -1 \leq x_t(\omega^i) \leq 1 \quad t = 1, 2, 3,\ i = 1, ..., 4 \end{aligned}$$

$$x_1(\omega^i) - \tfrac{1}{4} \sum_{i=1}^{4} x_1(\omega^i) = 0 \quad i = 1, ..., 4 \tag{13}$$

$$x_2(\omega^i) - \tfrac{1}{2}(x_2(\omega^1) + x_2(\omega^2)) = 0 \quad i = 1, 2 \tag{14a}$$

$$x_2(\omega^i) - \tfrac{1}{2}(x_2(\omega^3) + x_2(\omega^4)) = 0 \quad i = 3, 4 \tag{14b}$$

Of course, for $i \in \{1, 2, 3, 4\}$, we could simply define

$$X(\omega^i) = \{x \in \Re^3 \mid 1 \geq x_1 \geq x_2 \geq x_3 \geq -1\}$$

and

$$\phi(x, \omega^i) = \begin{cases} \sum_{t=1}^{3} c_t(\omega^i) x_t & \text{if } x \in X(\omega^i) \\ \infty & \text{otherwise.} \end{cases}$$

In this case, we would simply state the problem as $\text{Min}\{\frac{1}{4} \sum_{i=1}^{4} \phi(x(\omega^i), \omega^i)\}$, subject to the constraints (13) and (14). We note also that the nonanticipativity constraints, (13) and (14), contain redundant constraints.

The solution for which $x_t(\omega^i) = -1$ for $t = 1, 2, 3,\ i = 1, ..., 4$ is an optimal solution with an objective value of -1. Of greater interest is the value of the dual multipliers on the constraints (13) and (14). Notationally, let $\mu_t(\omega^i)$ denote the multiplier associated with the nonanticipativity constraint in which $x_t(\omega^i)$ appears outside of the expectation calculations. A dual solution associated with the indicated primal solution is

$$\mu_1(\omega^1) = 0 \quad \mu_1(\omega^2) = -0.25 \quad \mu_1(\omega^3) = 0 \quad \mu_1(\omega^4) = 0$$

$$\mu_2(\omega^1) = 0 \quad \mu_2(\omega^2) = 0.75 \quad \text{and} \quad \mu_2(\omega^3) = -0.25 \quad \mu_2(\omega^4) = 0$$

Note that this dual solution does not yield common values corresponding to any of the groups of constraints (13), (14a) or (14b). That is, even at points at which the cost coefficients are identical, the dual variables disagree. This particular solution is anticipative. In order to determine whether there exists an optimal solution that is nonanticipative, one may append constraints to the dual of (12) to explicitly enforce

$$\mu_1(\omega^1) = \mu_1(\omega^2) = \mu_1(\omega^3) = \mu_1(\omega^4)$$

$$\mu_2(\omega^1) = \mu_2(\omega^2) \quad \text{and} \quad \mu_2(\omega^3) = \mu_2(\omega^4).$$

Note that the above constraints can be equivalently characterized with conditional expectations. In any event, the addition of such constraints renders the objective value (-1) infeasible. It follows that there is no optimal dual solution that is nonanticipative.

Conclusion. In general, the optimal values of the nonanticipativity multipliers are anticipative. Furthermore, using the conditional expectation operator (on these vectors) to enforce nonanticipativity may result in the loss of optimality of dual vectors.

In some cases, a nonanticipative dual solution may result, as the following result indicates.

Theorem 2. *Suppose that the primal-dual pair, P-MV and D-MV, are feasible. Under the conditions of Corollary* 5, *there exists a dual solution, μ^*, that is nonanticipative if, and only if, the expected value of perfect information is zero.*

Proof: First, the expected value of perfect information, EVPI, is defined as the difference between the value of the stochastic program, (SP) and the "wait and see" problem, (1). From the definition of the conjugate function, the value of the "wait and see" problem may be represented as $-E[\phi^*(0, \tilde{\omega})]$. Hence, given a primal optimal solution x^*, we have

$$\text{EVPI} = E[\phi(x^*(\tilde{\omega}), \tilde{\omega})] + E[\phi^*(0, \tilde{\omega})].$$

If μ^* is nonanticipative, then for $t = 1, \ldots, T$,

$$\mu_t^*(\omega) = E\big[\mu_t^*(\tilde{\omega}) \,\big|\, \tilde{\omega} \in \mathcal{H}_t^{-1}(\mathcal{H}_t \omega)\big]$$

for almost every $\omega \in \Omega$. That is, using the definition in (9), nonanticipativity of μ^* is represented by $\mu^*(\tilde{\omega}) = \bar{\mu}^*(\tilde{\omega})$, almost surely. Thus, if μ^* is nonanticipative,

$$\begin{aligned} L(\mu^*, \omega) &= \phi^*(\mu^*(\omega) - \bar{\mu}^*(\omega), \omega) \\ &= \phi^*(0, \omega) \\ \Rightarrow E[L(\mu^*, \tilde{\omega})] &= E[\phi^*(0, \tilde{\omega})]. \end{aligned}$$

It follows that if μ^* is a nonanticipative optimal solution to D-MV, then 0 is an optimal solution to D-SV (note that 0 is the only possible nonanticipative solution to D-SV), and as a result of Theorem 3,

$$\begin{aligned} &E[\phi(x^*(\tilde{\omega}), \tilde{\omega})] = -E[\phi^*(0, \tilde{\omega})] \\ &\qquad \Rightarrow \text{EVPI} = 0. \end{aligned}$$

Of course, the converse is obvious, and hence the result. □

One may verify that the expected value of perfect information in our numerical example is 0.5, which explains the absence of a nonanticipative optimal dual solution. It is also of interest to note that due to the equality of primal and dual optimal values (under the assumptions of Theorem 3), the EVPI can be calculated entirely in terms of the dual problem D-SV. Thus, using the state variable representation,

$$\text{EVPI} = -E[\phi^*(\hat{\sigma}(\tilde{\omega}), \tilde{\omega})] + E[\phi^*(0, \tilde{\omega})],$$

where $\hat{\sigma}(\tilde{\omega})$ denotes an optimal dual solution. One of the advantages of this point of view is that one can obtain bounds on the value of EVPI. That is, for any dual feasible solution, σ,

$$\text{EVPI} \geq -E[\phi^*(\sigma(\tilde{\omega}), \tilde{\omega})] + E[\phi^*(0, \tilde{\omega})]. \tag{15}$$

Table 1 Alternate dual optima

	$\mu_t(\omega^1)$	$\mu_t(\omega^2)$	$\mu_t(\omega^3)$	$\mu_t(\omega^4)$
Solution 1:				
$t = 1$	−0.125	−0.125	0	0
$t = 2$	−0.125	−0.125	−0.125	0.125
Solution 2:				
$t = 1$	−0.125	−0.375	−0.125	−0.125
$t = 2$	−0.375	−0.125	−0.25	0
Solution 3:				
$t = 1$	0	−0.25	0	0
$t = 2$	−0.5	0	−0.25	0
Solution 4:				
$t = 1$	0	−0.25	0	0
$t = 2$	0	−0.5	−0.25	0

We note that the lower bound on EVPI provided by the right hand side of (15) is a function of the dual multipliers. In large scale problems, this bound may be useful if we have good heuristics to generate dual feasible multipliers. Another interesting observation arises from studying the subgradients of the bound. Note that if $x(\omega) \in \operatorname{argmax}\{\sigma(\omega)^\top x - \phi(x, \omega)\}$, then,

$$-E[x(\tilde{\omega})] \tag{16}$$

provides a subgradient of the lower bound on EVPI at σ.

Before concluding this section, we return to our example with an eye toward the nature of the dynamic process that governs the values of the dual variables. Of particular interest is the evolution of the dual variables over time along scenarios that share a common history. Dempster (1981, 1988) claims that it is a supermartingale process, although his proof of this claim depends upon the nonanticipativity of the process. We have already seen that the process is anticipative, except in the most trivial of cases. The question remains as to whether or not there is a readily discernible form of dependence among the dual variables. For example, given that ω^1 and ω^2 share a common history through period 2, the relationships between $\mu_1(\omega^i)$ and $\frac{1}{2}(\mu_2(\omega^1) + \mu_2(\omega^2))$ $i = 1, 2$ is of interest (as are the relationships between $\mu_1(\omega^i)$ and $\frac{1}{2}(\mu_2(\omega^3) + \mu_2(\omega^4))$ $i = 3, 4$. If

$$\mu_1(\omega^i) = \begin{cases} \frac{1}{2}(\mu_2(\omega^1) + \mu_2(\omega^2)) & i = 1, 2 \\ \frac{1}{2}(\mu_2(\omega^3) + \mu_2(\omega^4)) & i = 3, 4 \end{cases} \tag{17}$$

the process is a martingale. If the equality in (17) becomes an inequality, the process is a submartingale or supermartingale, depending on the direction of the inequality (see Ross 1983).

As one might expect, there are multiple dual optima associated with this example. In Table 1, we list a few of the alternate dual solutions to the example.

One may easily verify that solution 1 is a martingale solution. Similarly, solution 2 is a submartingale while solution 3 is a supermartingale. Finally, solution 4 does not satisfy a martingale definition. In general, the dual variable process seems to be somewhat arbitrary, insofar as martingales are concerned.

5. Interpretations and conclusions

In this paper, we have studied alternate representations of primal multistage stochastic convex programming problems as well as their duals. We note that the stochastic conjugate function (7) and the stochastic Lagrangian function (10) are convenient analogs of their deterministic counterparts, and hope that these stochastic versions become as popular. We have provided a unified framework for problems involving discrete as well as continuous random variables, and furthermore, the dual problems involve measurable functions of the random variables in the SP model. Consequently, dual approximations can be constructed without appealing to specific primal approximations, thus making it possible to allow optimality tests with "out-of-sample" scenarios as in Higle and Sen (1996b).

In addition to studying dual problems, this paper also clarifies the role of the multipliers associated with the nonanticipativity constraints. Despite previous claims to the contrary, our investigation reveals that these multipliers are, in general, anticipative. Only for the special case in which perfect information has no value are the multipliers nonanticipative. This may come as a surprise to some practitioners of optimization, since anticipative decisions are thought to be unimplementable. In order to put the reader at ease with the possibility of anticipative multipliers, we recall the analogy between them and a tax/subsidy rate in a "tax system". Just as the farmer is given federal subsidies after the outcome (e.g., a flood) is revealed, the nonanticipativity multipliers are obtained after the outcome has been revealed. Thus the marginal values provided by nonanticipativity multipliers are to be implemented on a retroactive basis. Another way to interpret the nonanticipativity multipliers is by likening them to a refund rate (as in a tax system). Clearly, this is feasible.

Finally we comment on the possiblity of interpreting these multipliers as the "marginal EVPI process", as in Dempster (1981). This moniker is something of a misnomer. Note that the dual multipliers help equilibrate plans that may be associated with scenarios, and consequently, each $\sigma(\omega)$ provides a subgradient of an outcome $\phi(x(\omega), \omega)$ at an optimal plan. As in the previous section, it is therefore clear that these multipliers refer to marginal values with respect to changes in plans (x), rather than marginal values with respect to changes in information ($\tilde{\omega}$). Similarly, as shown in Section 4, the subgradients in (16) provide marginal values with respect to σ, rather than $\tilde{\omega}$. Nevertheless, the EVPI as well as its lower bound are separable in σ, and may be used to estimate EVPI.

Acknowledgements We appreciate the feedback we received from Roger Wets and Michael Dempster on an earlier draft of this paper.

References

Birge, J.R. (1982). "The Value of the Stochastic Solution in Stochastic Linear Programs with Fixed Recourse." *Mathematical Programming* 24, 314–325.

Birge, J.R. (1985a). "Decomposition and Partitioning Methods for Multistage Stochastic Linear Programs." *Operations Research* 33, 989–1007.

Birge, J.R. (1985b). "Aggregation in Stochastic Linear Programs." *Mathematical Programming* 31, 25–41.

Cariño, D.R., T. Kent, D.H. Myers, C. Stacy, M. Sylvanus, A. Turner, K. Watanabe, and W.T. Ziemba. (1994). "The Russel-Yasuda Kasai Model: an Asset/liability Model for a Japanese Insurance Company Using Multistage Stochastic Programming." *Interfaces* 24, 29–49.

Clarke, F.H. (1983). *Optimization and Nonsmooth Analysis*, Wiley, New York.
Dempster, M.A.H. (1981). "The Expected Value of Perfect Information in the Optimal Evolution of Stochastic Systems." *Stochastic Differential Systems*, M. Arato, E. Vermes, and A.V. Balakrishnan (eds.), *Lecture Notes in Information and Control* 36, Springer, Berlin, 25–40.
Dempster, M.A.H. (1988). "On Stochastic Programming II: Dynamic Problems Under Risk." *Stochastics* 25, 15–42.
Frauendorfer, K. (1992). *Stochastic Two-Stage Programming*, Springer-Verlag, Berlin.
Frauendorfer, K. (1996). "Barycentric Scenario Trees in Convex Multistage Stochastic Programming." *Mathematical Programming* (Series B), 75, 277–294.
Higle, J.L. and S. Sen. (1996a). *Stochastic Decomposition: A Statistical Method for Large Scale Stochastic Linear Programming*, Kluwer Academic Publishers, Dordrecht.
Higle, J.L. and S. Sen. (1996b). "Duality and Statistical Tests of Optimality for Two Stage Stochastic Prgorams." *Mathematical Programming* (Series B) 75, 257–275.
Mulvey, J.M. and A. Ruszczyński. (1995). "A New Scenario Decomposition Method for Large Scale Stochastic Optimization." *Operations Research* 43, 477–490.
Rockafellar, R.T. and R. J.-B. Wets. (1976a). "Nonanticipativity and $\mathcal{L}^1$ Martingales in Stochastic Optimization Problems." *Mathematical Programming Study* 6, 170–187.
Rockafellar, R.T. and R. J.-B. Wets. (1976b). "Stochastic Convex Programming: Basic Duality." *Pacific J. of Mathematics* 62, 173–195.
Rockafellar, R.T. and R. J.-B. Wets. (1982). "On the Interchange of Subdifferentiation and Conditional Expectation for Convex Functionals. *Stochastics* 7, 173–182.
Rockafellar, R.T. and R. J.-B. Wets. (1991). "Scenarios and Policy Aggregation in Optimization Under Uncertainty." *Math. of Oper. Res.* 16 119–147.
Rockafellar, R.T. and R. J.-B. Wets. (1992). "A Dual Strategy for the Implementation of the Aggregation Principle in Decision Making Under Uncertainty." *Applied Stochastic Models and Data Analysis* 8, 245–255.
Ross, S.M. (1996). *Stochastic Processes, 2^{nd} ed.*, John Wiley and Sons.
Sen, S., R.D. Doverspike, and S.C. Cosares. (1994). "Network Planning with Random Demand." *Telecommunications Systems* 3, 11–30.
Wets, R. J.-B. (1975). On the Relation Between Stochastic and Deterministic Optimization, in: A. Bensoussan and J.L. Lions (eds.), *Control Theory, Numerical Methods and Computer Systems Modeling*, Lecture Notes in Economics and Mathematical Systems, 107, 350–361.
Wets, R. J.-B. (1989). Stochastic Programming, in: G.L. Nemhauser, A.H.G. Rinooy Kan and M.J. Todd (eds.) *Handbooks in Operations Research: Optimization*, 573–629, North-Holland, Amsterdam.
Wright, S.E. (1994). "Primal-dual Aggregation and Disaggregation for Stochastic Linear Programs." *Math. of Oper. Res.* 19, 893–908.
Zipkin, P. (1980). "Bounds for Row Aggregation in Linear Programming." *Operations Research* 28, 903–918.