ALGORITHMIC IMPLICATIONS OF DUALITY IN STOCHASTIC PROGRAMS

Julia L. Higle

*Department of Systems and Industrial Engineering,
The University of Arizona,
Tucson, AZ 85721, USA
julie@sie.arizona.edu*

Suvrajeet Sen

As above

Abstract In this paper, we study primal and dual formulations of multistage stochastic programs (SP). Using a dual formulation, we discuss a decomposition/cutting plane algorithm that can be used to solve such problems. The algorithm, which is based on a scenario decomposition derived from the dual statement of the problem, is best viewed as a conceptual algorithm. Nevertheless, it lends itself to the use of sampled data, and enhancements necessary to produce a computationally viable method are discussed.

1. INTRODUCTION

Stochastic programming (SP) is a powerful modeling tool 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 (see, e.g., Cariño et al [1994], Sen, Doverspike, and Cosares [1994]).

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]). In this paper, we present an easily accessible development of a dual problem for a multi-

The original version of this chapter was revised: The copyright line was incorrect. This has been corrected. The Erratum to this chapter is available at DOI: [10.1007/978-0-387-35514-6_15](https://doi.org/10.1007/978-0-387-35514-6_15)

stage SP. This accessibility is due, in large part, to the use of stochastic analogs of deterministic constructs that are already well understood in the mathematical programming literature. Based on this setting, we present a decomposition/cutting plane algorithm that may be used to solve a multistage stochastic program. The problem is decomposed by “scenario”, as in Rockafellar and Wets [1991], rather than by time stage, as in Birge [1985] and Gassmann [1990]. 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.

This paper is organized as follows. In §2, we present a generic formulation of a multistage stochastic program, and suggest alternate representations of the so-called “nonanticipativity constraints”. These constraints model the time-staged evolution of information, and are characteristic components of a stochastic program. In §3, we present a dual representation of a stochastic program. In §4, we propose a decomposition/cutting plane algorithm for the solution of stochastic programs. It may be best to think of this as a “conceptual algorithm”, from which a computationally viable method can be obtained. As such, some of the enhancements necessary to obtain a viable algorithm are also discussed in §4. Concluding remarks may be found in §5. While this paper is largely expository in nature, with most of the mathematical details omitted, the reader is referred to Higle and Sen [1997] for a fully detailed presentation of the duality results in §3, and to Sen, Higle and Rayco [1999] for a full presentation of a viable algorithm, along with preliminary computational results.

2. PRIMAL PROBLEM

In what follows, we consider a problem in which “decisions”, which we denote as x , and random variables are interwoven over time. An initial decision is made, after which an outcome of a random variable is observed. In response to the observation, a subsequent decision is made, after which another outcome is observed, 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) information, denoted ω_1 . Information beyond the first stage is uncertain, and is modeled 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 $x = \{x_t\}$ and $\omega = \{\omega_t\}$ are associated with decisions and outcomes, respectively. In this sense, x_t indicates a decision made in stage t and ω_t indicates an observation made

in stage t . The sequence of random variables $\tilde{\omega} = \{\tilde{\omega}_t\}_{t=1}^T$ is defined on a probability space $(\Omega, \mathcal{A}, \mathcal{P})$, and in the stochastic programming literature, any realization ω is termed as a *scenario*. 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 upon $\tilde{\omega}$, a feasible choice of x is nonetheless dependent upon $\tilde{\omega}$. Thus, for each possible scenario $\omega \in \Omega$, there is a set of feasible solutions, $X(\omega)$, and an objective function $g(x, \omega)$ which influences the choice of x . When a particular variable, such as x is explicitly defined via the consideration of all possible scenarios in Ω , we will denote it as $\{x(\omega)\}_{\omega \in \Omega}$. Otherwise, we will simply denote it as x . Finally, throughout our development, we will assume that all vectors are appropriately dimensioned. Consequently, for each possible scenario $\omega \in \Omega$, we have the following “scenario problem”, also known as the “wait and see problem”:

$$\begin{array}{ll} \text{Min} & g(x, \omega) \\ \text{s.t.} & x \in X(\omega). \end{array} \quad (P_\omega)$$

Let $x(\omega)$ denote the solution to the scenario problem (P_ω) . Note that the “scenario solutions”, $\{x(\omega)\}_{\omega \in \Omega}$, represent the result of posterior optimization. That is, $x(\omega)$ would be an optimal response if one knew for certain that scenario ω would occur. The difficulty, of course, is that one is rarely (if ever) graced with such knowledge, and while $x(\omega^1)$ is an optimal response to ω^1 , it may be a disastrous response to ω^2 . Thus, a model that leads to a more balanced response, one whose cost under all scenarios has been explicitly considered, is necessary.

To develop such a model, it is important to first note that scenarios that share a common history up to some point in time must implement the same decision at that time. That is, if we want to be able to implement a decision at a particular time, it cannot depend on information that will only become available at a later time. This requirement leads to constraints known as the nonanticipativity constraints. If we let $\mathcal{N}$ denote the set of nonanticipative solutions, we may formulate a stochastic programming model as follows:

$$\begin{array}{ll} \text{Min} & E[g(x(\tilde{\omega}), \tilde{\omega})] \\ \text{s.t.} & x(\tilde{\omega}) \in X(\tilde{\omega}) \quad a.s. \\ & x(\tilde{\omega}) \in \mathcal{N}. \end{array}$$

We will assume throughout that the constraints, $x(\tilde{\omega}) \in X(\tilde{\omega})$ have a property known as relatively complete recourse. 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 cannot be observed until some later time. With this assumption, the scenarios are only coupled through the nonanticipativity constraints. (This constraint qualification is common in the stochastic programming literature.) When formulating a stochastic program, relatively complete recourse can be incorporated through the judicious use of penalties for infeasibilities.

To model the nonanticipativity requirements, let $\mathcal{H}_t$ denote the operator that projects any multistage entity onto the space corresponding to the first t stages. Then

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

reflects the evolution of scenario ω through the first t periods. Let Ω_t denote the set of possible “truncations” in period t , or

$$\Omega_t = \{\underline{\omega}_t \mid \mathcal{H}_t\omega = \underline{\omega}_t \text{ for some } \omega \in \Omega\}.$$

Note that $\Omega_T = \Omega$. Next, define the point-to-set map $\mathcal{H}_t^{-1}$, which identifies sets of scenarios with common truncations in period t . That is,

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

Note that for each t , $\{\mathcal{H}_t^{-1}\underline{\omega}_t \mid \underline{\omega}_t \in \Omega_t\}$ forms a partition of Ω . 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, ω^1 and ω^2 are indistinguishable in period t , and nonanticipativity requires that $\mathcal{H}_tx(\omega^1) = \mathcal{H}_tx(\omega^2)$. We may use this notion to express the nonanticipativity constraints in terms of state variables $\{z_t\}_{t=1}^T$, where $z_t : \Omega_t \rightarrow \mathbb{R}^n$. Then the state variable representation of the nonanticipativity constraints may be represented as:

$$x_t(\tilde{\omega}) - z_t(\mathcal{H}_t\tilde{\omega}) = 0 \quad \text{wp1}, \quad \forall t. \quad (1a)$$

Note that these constraints will ensure that all scenarios that share a common history at period t yield a common solution in period t as well. Alternatively, we may express the nonanticipativity constraints using conditional expectations. That is, we note that if $\tilde{\omega}'$ is defined on the same probability space as $\tilde{\omega}$, then

$$x_t(\tilde{\omega}) - E[x_t(\tilde{\omega}') \mid \tilde{\omega}' \in \mathcal{H}_t^{-1}(\mathcal{H}_t\tilde{\omega})] = 0 \quad \text{wp1}, \quad \forall t \quad (1b)$$

will ensure that $x(\tilde{\omega})$ is nonanticipative. Either (1a) or (1b) may be used to represent the nonanticipativity requirements. There are additional representations as well, and the choice of representation will typically be guided by the solution method adopted. For the purposes of our algorithmic presentation, we will adopt the state variable representation,

(1a). Finally, for ease of exposition, we define the following extended valued function

$$\varphi(x, \omega) = \begin{cases} g(x, \omega) & \text{if } x \in X(\omega) \\ \infty & \text{otherwise.} \end{cases} \quad (2)$$

With this function, we may now specify a multistage stochastic program as follows:

$$\begin{aligned} \text{Min}_x \quad & E[\varphi(x(\tilde{\omega}), \tilde{\omega})] \\ \text{s.t.} \quad & x(\tilde{\omega}) - z_t(\mathcal{H}_t \tilde{\omega}) = 0 \quad \text{wp1, } t = 1, \dots, T. \end{aligned} \quad (P)$$

3. DUAL PROBLEM

As with the primal problem (P), the dual problem that we study is valid for both continuous and discrete random variables. In this section, we propose a stochastic analog of conjugate 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]).

Throughout our development, we will introduce a number of “variables” which correspond to measurable functions of random variables. For example, $\{x_t(\omega)\}_{\omega \in \Omega}$, or equivalently, $x(\tilde{\omega})$, is one such measurable function mapping Ω to $\mathbb{R}^n$. We assume that the functions $x \in \mathcal{L}^\infty$ are essentially bounded and $\mathcal{P}$ -measurable. Let $\mathcal{L}^1$ denote the space of functions that are integrable with respect to $\mathcal{P}$. Given $\xi \in \mathcal{L}^1$, it is convenient to define the following linear operation.

$$\xi \circ x = \int_{\Omega} \xi(\omega)^\top x(\omega) \mathcal{P}(d\omega). \quad (3)$$

Maintaining an eye toward the stochastic properties of the mathematical program under consideration, it is interesting to note that this inner product is equivalently stated as

$$\xi \circ x = E[\xi(\tilde{\omega})^\top x(\tilde{\omega})], \quad (4)$$

which we recognize as the expected value of the traditional counterpart from deterministic mathematical programming. Throughout our development, we use the representations in (3) and (4) interchangeably.

Consider the function φ defined in (2), and its conjugate function

$$\varphi^*(\xi, \omega) = \sup_{x \in \mathbb{R}^n} \{\xi^\top x - \varphi(x, \omega)\}. \quad (5)$$

Relying on the state variable representation of the nonanticipativity constraints, the following problem

$$\begin{aligned} & \sup_{\xi \in \mathcal{L}^1} -E[\varphi^*(\xi(\tilde{\omega}), \tilde{\omega})] \\ \text{s.t. } & E[\xi_t(\tilde{\omega}) \mid \tilde{\omega} \in \mathcal{H}_t^{-1}(\mathcal{H}_t \tilde{\omega}')] = 0 \quad \text{wp1, } t = 1, \dots, T \end{aligned} \quad (D)$$

is dual to (P), as indicated in the following result from Higle and Sen [1997].

Theorem 1 *Let $\varphi(\cdot, \tilde{\omega})$, as defined in (2), be a convex normal integrand, and assume that (P) has relatively complete recourse. Let v_p and v_d denote the optimal values of (P) and (D), respectively. Then*

(a) $v_p \geq v_d$.

(b) *If (P) possesses an optimal solution denoted $(\hat{x}, \hat{z})$, and $\partial\varphi(\hat{x}(\tilde{\omega}), \tilde{\omega})$ is nonempty (wp1), then there exists $\hat{\xi}(\tilde{\omega}) \in \partial\varphi(\hat{x}(\tilde{\omega}), \tilde{\omega})$, wp1, such that*

$$E[\hat{\xi}_t(\tilde{\omega}) \mid \tilde{\omega} \in \mathcal{H}_t^{-1}(\mathcal{H}_t \tilde{\omega}')] = 0 \quad (\text{wp1}),$$

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

Proof (See Higle and Sen [1997]).

It is interesting to note that (D) may be construed as a *stochastic conjugate dual*. Dual feasibility requires only that the conditional means of the dual vectors be zero at each decision point. For notational convenience, we let Ξ denote the set of dual feasible solutions. In the following section, we use this dual statement of the multistage stochastic programming problem to develop a cutting plane approximation of the problem.

4. A CUTTING PLANE ALGORITHM

Our algorithmic development depends on the dual representation, (D). Assuming the existence of an optimal solution, we restate the problem as

$$\begin{aligned} & \text{Max}_{\xi \in \mathcal{L}^1} E[\text{Min}_{x(\tilde{\omega})} \{\varphi(x(\tilde{\omega}), \tilde{\omega}) - \xi^\top(\tilde{\omega})x(\tilde{\omega})\}] \\ \text{s.t. } & E[\xi_t(\tilde{\omega}) \mid \tilde{\omega} \in \mathcal{H}_t^{-1}(\mathcal{H}_t \tilde{\omega}')] = 0 \quad \text{wp1, } t = 1, \dots, T \end{aligned}$$

Notice that in this representation, the scenarios are linked via the outer maximization, while the inner minimization is separable by scenario. As such, the inner minimization can be decomposed by scenario, with the outer maximization providing the coordination necessary to identify an optimal solution.

To begin, let

$$v(\xi, \omega) = \min_x \{\varphi(x, \omega) - \xi^\top x\} \leq \varphi(x, \omega) - \xi^\top x \quad \forall x, \omega$$

with equality holding if $x \in \operatorname{argmin}_x \{\varphi(x, \omega) - \xi^\top x\}$. Thus, given $\xi(\tilde{\omega})$ and $x(\tilde{\omega})$, we have

$$E[v(\xi(\tilde{\omega}), \tilde{\omega})] \leq E[\varphi(x(\tilde{\omega}), \tilde{\omega})] - E[\xi(\tilde{\omega})^\top x(\tilde{\omega})] = E[\varphi(x(\tilde{\omega}), \tilde{\omega})] - \xi \circ x$$

with equality holding if $x(\tilde{\omega}) \in \operatorname{argmin}_{x(\tilde{\omega})} \{\varphi(x(\tilde{\omega}), \tilde{\omega}) - \xi(\tilde{\omega})^\top x(\tilde{\omega})\}$, with probability one. Note that $E[\varphi(x(\tilde{\omega}), \tilde{\omega})]$ is constant with respect to ξ . Thus, this affine bound provides the basis for a cutting plane approximation to the expected value of the inner minimization.

The Basic Scenario Decomposition method may be stated as follows:

<p><u>Step 0. Initialize</u> $\varepsilon > 0$, an error tolerance, is given.</p> <p>$k \leftarrow 0$, $\xi_k(\omega) \leftarrow 0$ for all $\omega \in \Omega$, $v^k \equiv +\infty$.</p>
<p><u>Step 1. Solve the Subproblem</u> $k \leftarrow k + 1$.</p> <p>Let $x^k(\omega) \in \operatorname{argmin}_x \{\varphi(x, \omega) - \xi^{k-1}(\omega)^\top x\}$ wp1.</p>
<p><u>Step 2. Update the Master Problem</u> Let $\alpha^k = E[\varphi(x^k(\tilde{\omega}), \tilde{\omega})]$, and add the constraint $v \leq \alpha^k - \xi \circ x^k$ to the master problem.</p>
<p><u>Step 3. Solve the Master Problem</u> Let (v^k, ξ^k) denote a solution to:</p> $\begin{array}{ll} \text{Max} & v \\ \text{s.t.} & v \leq \alpha^j - \xi \circ x^j \quad j = 1, \dots, k \\ & \xi \in \Xi \end{array}$ <p>If $\alpha^k - \xi^k \circ x^k - v^k < \varepsilon$, then stop. Otherwise, continue from Step 1.</p>

One recognizes that the Basic Scenario Decomposition algorithm is Kelley's cutting plane method applied to the dual statement of the multi-stage stochastic program. As such, it follows that the iterates $\{\xi^k(\tilde{\omega})\}_{k=1}^\infty$ converge to an optimal solution. We note, however, that there are several impediments to the practical use of this method.

By way of example, note that in Step 1 of the method, we require that $x^k(\tilde{\omega})$ be an optimal solution with probability one. If $\tilde{\omega}$ is a discrete random variable, this requires the solution of the indicated problem for

all possible scenarios. Otherwise, it is necessary to obtain an optimal solution for all scenarios outside a set of probability measure zero. Clearly, as the number of possible scenarios increases, each execution of Step 1 becomes a substantial task. Related to this is the fact in Step 3 of the method, the columns may be described as: $\{\xi(\omega)\}_{\omega \in \Omega}$. Again, as the number of scenarios increases, the column dimension grows quickly out of hand. Additionally, as with all cutting plane methods, the proliferation of cuts in the master program will become problematic. Because of these observations, as the number of scenarios increases, the ability to actually execute Step 3 diminishes.

Both of these problems are related to the specification of Ω . Of course, when dealing with large sample spaces, there is a natural tendency to look toward sample-based methods. Indeed, these have been successfully used in the solution of large scale two-stage stochastic programs. In particular, Stochastic Decomposition (SD) and its variants, which use a successive sampling scheme (as opposed to a fixed sample) is the only such method which is capable of identifying an optimal solution to the problem with probability one (see, for example Higle and Sen [1991, 1996a]). SD begins with a small sample size, and increases it as iterations progress. In this manner, Steps 1 and 3 may be executed rather quickly in the early iterations, when one expects the iterates, (ξ^k) to be far from an optimal solution. As iterations progress, the information obtained via the sample becomes increasingly accurate, as do the resulting cutting planes and solutions. We note that safeguards are required to prevent cutting planes derived in the early iterations (i.e., when the sample-based information is most likely to be inaccurate) from interfering with the search for an improved solution. Within the context of the Scenario Decomposition method, note that the use of sampled observations of Ω will impact the definition of dual feasibility. Moreover, as iterations progress and larger sample sizes are used, cut proliferation and the corresponding growth in the master problem column dimension becomes problematic.

Sampled Scenario Decomposition (SSD) is an algorithmic methodology based on the Basic Scenario Decomposition scheme described in this paper. SSD is designed specifically to address the issues identified above. As the name indicates, it is based on the use of sampled observations of $\tilde{\omega}$. In addition, as iterations progress it incorporates a row and column aggregation scheme designed to control the size of the master program. Additional features are also included. We refer the interested reader to Sen, Higle and Rayco [1999] for a full description of the method, its properties, and preliminary computational results.

5. CONCLUSIONS

In this paper, we have studied a dual approach to multi-stage stochastic programming problems. This development is closely tied to deterministic conjugate duality, with the only additional requirement in the stochastic case being the inclusion of constraints requiring the conditional expectation of dual multipliers to be zero. By interpreting dual multipliers as subgradients of the cost-to-go function, dual feasibility requires the natural optimality condition. This approach is not only elegant, but leads to algorithms in which we can avoid stage-by-stage recursions. The advantage of doing so is that we can address a larger class of problems, especially those in which the objective function is not separable by stage. This approach is also amenable to the use of sample-based algorithms, although the size of the master problems can become arbitrarily large. By using appropriate aggregation schemes, we are able to maintain approximations of reasonable size. Such computations are reported in a forthcoming paper.

References

- [1] J.R. Birge (1985), Decomposition and partitioning methods for multistage stochastic linear programs, *Operations Research*, 22, pp. 989–1007.
- [2] D. Cariño, T. Kent, D. Myers, C. Stacy, M. Sylrams, A. Turner, K. Watanabe and W. Ziemba (1994), The Russell–Yasuda Kasai Model: An asset/liability model of a Japanese insurance company using multi-stage stochastic programming, *Interfaces*, 24, pp. 29–49.
- [3] M.A.H. Dempster (1981), The expected value of perfect information in the optimal evolution of stochastic systems, in M. Arato, E. Vermes, and A.V. Balakrishnan (eds.), *Stochastic Differential Systems*, Lecture Notes in Information and Control, 36, Springer, Berlin, pp. 25–40.
- [4] M.A.H. Dempster (1988), On stochastic programming II: Dynamic problems under risk, *Stochastics*, 25, pp. 15–42.
- [5] H.G. Gassmann (1990), MSLiP: A computer code for the multi-stage stochastic linear programming problem, *Mathematical Programming*, 47, pp. 407–423.
- [6] J.L. Higle and S. Sen (1991), Stochastic Decomposition: An Algorithm for Two Stage Linear Programs with Recourse, *Mathematics of Operations Research*, 16, no. 3, pp. 650–669.

- [7] J.L. Higle and S. Sen (1996a), *Stochastic Decomposition: A Statistical Method for Large Scale Stochastic Linear Programming*, Kluwer Academic Publishers, Dordrecht.
- [8] J.L. Higle and S. Sen (1996b), Duality and statistical tests of optimality for two stage stochastic programs, *Mathematical Programming*, (Series B), v75, pp. 257–275.
- [9] J.L. Higle and S. Sen (1997), Multi-stage Stochastic Programs: Duality and its Implications, submitted to *Mathematical Programming*.
- [10] R.T. Rockafellar and R.J-B. Wets (1991), Scenarios and policy aggregation in optimization under uncertainty, *Mathematics of Operations Research*, 16, pp. 119–147.
- [11] S. Sen, R.D. Doverspike and S. Cosares (1994), Network planning with random demand, *Telecommunications Systems*, 3, pp. 11–30.
- [12] S. Sen, J.L. Higle and B. Rayco (1999), A Scenario Stochastic Decomposition Algorithm for Multistage Stochastic Programming, SIE Department, University of Arizona, Tucson, AZ 85721.
- [13] R.J-B. Wets (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, pp. 350–361.
- [14] R.J-B. Wets (1989), Stochastic Programming, in: G.L. Nemhauser, A.H.G. Rinnooy Kan, and M.J. Todd (eds.) *Handbook of Operations Research: Optimization*, North-Holland, Amsterdam, pp. 573–629.