STOCHASTIC DYNAMIC OPTIMIZATION APPROACHES AND COMPUTATION [1]

Pravin Varaiya and Roger J-B Wets
University of California, Berkeley-Davis

In: *Mathematical Programming, Recent Developments and Applications,* M. Iri & K. Tanabe (eds), Kluwer Academic Publisher, 1989. pp. 309-332.

The description of stochastic dynamical optimization models that follows is intended to exhibit some of the connections between various formulations that have appeared in the literature, and indicate some of the difficulties that must be overcome when trying to adapt solution methods that have been successfully applied to one class of problems to an apparently related but different class of problems. The emphasis will be on *solvable* models.

We begin with the least dynamical versions of stochastic optimization models, one- and two-stage models then consider discrete time models, and conclude with continuous time models.

1 ONE-STAGE MODELS

We consider the following simple one-stage stochastic optimization problem:

$$\begin{array}{lll} \text{minimize} & E\{h_0(z,\xi)\} & \\ \text{subject to} & h_i(z) \leq 0, & i = 1,\ldots,s, \\ & h_i(z) = 0, & i = s+1,\ldots,m, \\ & z \in Z \subset R^n & \end{array}$$

where ξ is a random vector with support $\Xi \subset R^N$ and distribution P. We are looking for a vector z^* that is feasible, i.e., belongs to

$$S = \{z \in Z | h_i(z) \leq 0,\ i = 1,\ldots,s; h_i(z) = 0,\ i = s+1,\ldots,m\},$$

and minimizes $E\{h_0(\cdot,\xi)\}$ on S. Of course, this is just a special instance of a nonlinear programming problem. Indeed, after integration, the objective can be rewritten as

$$\text{minimize } Eh_0(z),$$

[1]Supported in part by grants of the National Science Foundation and the Air Force Office of Scientific Research

where for each z,

$$Eh_0(z) \ := \int_{\Xi} h_0(z,\xi)\, dP(\xi).$$

Such a function is called an *expectation functional*; the study of its properties is a major theme of the theory of stochastic programming. However, even this "simple" stochastic optimization problem cannot be solved by standard nonlinear optimization algorithms. The problem is with the evaluation of Eh_0 or its (sub)gradient. There are a few cases that can be managed:

1. when the function $h_0(z,\cdot)$ is separable so that

$$\int h_0(z,\xi)\, dP = \sum_{j=1}^{N} \int_{\Xi_j} h_{0j}(z,\xi_j)\, dP_j(\xi_j)$$

   (with P_j the marginal distribution function),

2. when $\Xi = \{\xi^1, \ldots, \xi^L\}$ is finite and L is not too large, then

$$\int h_0(z,\xi)\, dP = \sum_{\ell=1}^{L} p_\ell h_0(z,\xi^\ell)$$

   (where $p_\ell = P[\xi = \xi^\ell]$),

3. if h_0 is convex, sufficiently smooth, easy enough to evaluate and P is a multidimensional normal, Gamma or Dirichlet distribution function.

The first case simply reflects the fact that univariate calculus, as well as one-dimensional numerical integration routines, are well developed. That is definitely not the case for multivariate calculus and multidimensional numerical integration. In the second case, the evaluation of Eh_0, or its gradient, at a point z is reduced to evaluating $h_0(z,\xi^l)$, or its gradient $\nabla_z f_0(z,\xi^\ell)$, for each ξ^ℓ in Ξ. And, in the third case, there are specific subroutines (developed by Hungarian computer scientists for stochastic programming problems) that combine Monte-Carlo techniques with some of the specific properties of those distributions. Because sampling is involved, the evaluation of $h_0(z,\xi)$ at any point ξ in Ξ should be "cheap" enough; unfortunately that is seldom the case in the most important applications.

Because of this state of affairs, the research in stochastic programming has been concerned with either identifying classes of models that fit in those "solvable" categories, designing reliable and efficient solutions procedures for such problems, or developing theories and procedures that would allows us to solve any problem by solving approximating problems that belongs to the "solvable" categories.

One version of the one-stage model that has received limited attention in the literature is the case when the probability distribution of ξ depends on z. In terms of the essential objective, the problem would take on the following form:

$$\text{find } z^* \text{ that minimizes } \int h(z,\xi)\, dP(\xi;z).$$

Again, this is *just* a nonlinear optimization problem and an evaluation of the objective at any point z is not more complicated than it was before. What has changed are the properties of the function:

$$z \mapsto Eh(z) = \int h(z,\xi)\, dP(\xi;z).$$

For example, when P does not depend on z, the convexity of Eh follows immediately from the convexity of $h(\cdot,\xi)$ for all ξ. That is no longer the case when P depends on z. Similarly, the (sub)gradients of Eh can no longer be obtained by the (relatively) simple formula:

$$\partial Eh = \int \partial h(\cdot,\xi)\, dP(\xi).$$

The stochastic approximation-like techniques, e.g., stochastic quasi-gradient methods, can no longer be used to find (almost surely) a solution, at least not in the form in which these techniques have been used up to now. In fact, in this situation, the properties of Eh may very well have nothing in common with those of $h(\cdot,\xi)$.

The challenge would not be so much to design general solution procedures for this (richer, but ungainly) class of problems, but to identify those that possess properties that would still allow us to use "classical" solution procedures. Clearly, it all has to do with the type of dependence of P on z. For example, if P is defined on R^N, and $P(\xi;z) = Q(\xi + Hz)$, where Q is a probability distribution function and H is a (given) matrix of the appropriate size, the problem takes on the following form (after a simple change of variables):

$$\text{find } z^* \text{ that minimizes } \int h(z,\zeta - Hz)\, dQ(\zeta).$$

The properties of Eh will thus depend on the properties of $h(z,\xi)$ viewed as a function of (z,ξ) jointly.

2 TWO-STAGE MODELS

In addition to a (first stage) decision z_1, this model allows for a second stage or *recourse* decision z_2 that is taken *after* full or partial information is obtained

about the values of the random components of the problem. The problem can be formulated as follows:

$$\begin{array}{lll} \text{minimize} & f_{10}(z_1) + E\{f_{20}(z_2(\xi), \xi)\} & \\ \text{subject to} & f_{1i}(z_1) \leq 0, & \text{for } i = 1, \ldots, m_1, \\ & f_{2i}(z_1, z_2(\xi), \xi) \leq 0 \ a.s., & \text{for } i = 1, \ldots, m_2, \end{array}$$

where the function z_2 can depend (measurably) on ξ in a way that is consistent with the information that will be available in the second stage, i.e., when taking the recourse decision. A much more detailed discussion of the the modeling of the information process will follow; for the time being let us assume that full information is available before choosing z_2.

If we define

$$f_0(z, \xi) := f_{10}(z) + \inf_{z_2}[f_{20}(z_2, \xi) | f_{2i}(z, z_2, \xi) \leq 0, i = 1, \ldots, m_2],$$

and

$$f_i(z) := f_{1i}(z) \text{ for } i = 1, \ldots, m_1,$$

we see that, at least from a theoretical viewpoint, the two-stage model can be analyzed in the framework provided by the one-stage model as long as we allow for a sufficiently general class of functions f_0, viz., infinite-valued (to account for the cases when for given z_1 and some ξ there is no z_2 that satisfies the second-stage constraints) and nondifferentiable (the infimal value of a mathematical program is seldom a differentiable function). Because, it covers a large number of applications, and because it is in some sense the first hurdle that must be mastered when considering dynamical optimization models, much of the algorithmic research in stochastic programming has been oriented at solving two-stage (recourse) models.

At first sight, the two-stage model may appear very restricted in its dynamical aspects. However, it is important to keep in mind that "stages" do not necessarily refer to time units. They correspond to stages in the decision process. The variable z_1 refers to all the decisions that must be taken *before* there will be any information about the values to be assigned to the random elements of the problem. The variables z_2 model *all* the decisions that will be made *after* the available information about these values will be collected. For example, z_1 could represent a sequence of decisions (control actions) to be made over a given time horizon, say $z_{11}, ..., z_{1t}, ..., z_{1T}$, and $z_2 = (z_{21}, ..., z_{2t}, ..., z_{2T})$, representing a similar sequence of decisions used to correct the basic trend set by the z_1-variables. Each one of the z_{2t} refers to a decision to be made at time t in response to the situation that would result from choosing z_1 and obtaining information about the random events that can be observed up to time t. Such models could be called *dynamical* two-stage models. As a special case, we could have $z_1 = z_{11}, ..., z_{1t}$, and $z_2 = z_{2,t+1}, ..., z_{2T}$,

which would correspond to a mid-course correction. And, of course, there is no need to restrict oneself to discrete time.

Let us now turn to the case when the recourse decision must be made under less than full information. Before we start, let us stress the fact that although one may not observe ξ, there are many cases when the *observations* made allow us to recover enough information about the values of ξ that one can still refer to it as *full* information. This has sometimes been the source of some confusion between the "stochastic programming" formulation and the "stochastic optimal control" formulation. A typical, and simple, example could go as follows: instead of ξ, we observe the "state" x_1 of the system, with the state defined by a relation of the form:

$$x_1 = \varphi(z_1) + \xi.$$

In such cases, instead of viewing the recourse decision as a function of ξ, we could equally well think of it as a function of the "state" of the system.

If only partial information will be available, let $\mathcal{G}$ be the (sub)field of events that could be observed before taking the recourse decision; let $\mathcal{A}$ be the field of all events generated by ξ. In these terms, partial information would mean that $\mathcal{G}$ is a proper subcollection of $\mathcal{A}$. Since the recourse decision z_2 can only depend on the information that will become available, it must be $\mathcal{G}$-adapted or, equivalently $\mathcal{G}$-measurable. Moreover, in evaluating the performance of a particular decision, only those events that lie in $\mathcal{G}$ can be taken into account, thus rather than using $f_{20}(z_2, \xi)$ as the objective function of the recourse problem, we would replace it by

$$E\{f_{20}(z_2, .) \,|\, \mathcal{G}\}(\xi).$$

Also, feasibility of a recourse function z_2 can only be checked up to events that lie in $\mathcal{G}$. Thus, a feasible first stage decision is one that satisfies the first stage constraints $f_{1i}(z_1) \leq 0, i = 1, .., m_1$ and to which one can associate a $\mathcal{G}$-measurable function z_2 such that almost surely satisfies:

$$f_{2i}(z_1, z_2(\xi), \xi) \leq 0, i = 1, ..., m_2.$$

This latter condition, may or may not impose restrictions on the choice of z_1 beyond those already imposed by the first stage constraints. If it does, one refers to them as *induced constraints*. Otherwise, the problem is said to have *relatively complete recourse*. This can also be expressed in terms of a certain property ($\mathcal{G}$-nonanticipativity) of the multifunction determined by the constraints; we shall return to this in the context of the multistage models.

Although the observations may very well depend on z_1, so far, we have only dealt with the case when the information available about the values taken on by the random quantities of the problem do *not* depend on the first stage decision. The

solution of the two-stage model, defined at the beginning of this section, can be found by first finding z_1^* the optimal solution of the (finite dimensional) nonlinear program:

$$\text{minimize } \ f_0(z_1) \ \text{ subject to } \ f_i(z_1) \leq 0, \ i = 1, \ldots, m_1,$$

with the functions $f_i, i = 0, \ldots, m_1$, as defined above, and then solving for each ξ (in the support of the probability measure), the deterministic nonlinear program:

$$\text{minimize } \ E\{f_{20}(z_2, \cdot) \,|\, \mathcal{G}\}(\xi) \ \text{ subject to } \ f_{2i}(z_1^*, z_2, \xi) \leq 0, i = 1, \ldots, m_2.$$

As long as as there is a consistent rule for choosing the optimal solution when there are multiple (optimal) solutions, this will define an optimal $\mathcal{G}$-measurable function z_2^*. In most applications, only the here-and-now decision, i.e., the first stage decision, is of interest, and then there is no need to explicitly calculate the optimal z_2 function.

In general, all of this is no longer possible if the probability distribution of the random quantities depends on the first stage decision, or if the information (derived from the observations) depends on z_1.

To indicate that the (sub)field of events depends on z_1, let us denote it by $\mathcal{G}(z_1)$. The two-stage problem is then to find a pair (z_1, z_2) in

$$Z_{\mathcal{G}} := \{z_1 \in R^{n_1}, \ z_2 \ \ \mathcal{G}(z_1)-\text{measurable}\}$$

that satisfies the constraints and minimize the objective function as defined above. The space $Z_{\mathcal{G}}$ is no longer a linear space (as was the case when the field of information did not depend on z_1), in general it is neither convex (not even connected), nor closed. The nonlinearities introduced by the dependence of the information field on z_1 have changed the essence of the problem, and usually, it is a much more difficult problem to solve. The solution cannot be found, as before, by solving (in sequence) finite dimensional optimization problems. The optimal first stage decision cannot be found without finding an explicit description of the associate (optimal) second-stage decision function. There are examples in the literature (not exactly formulated in these terms), beginning with one due to Witsenhausen, that illustrate all of these difficulties. The fact that the problem becomes so complicated may suggest that there is a need to consider more carefully its formulation.

We shall return to this in the context of stochastic control models.

For purposes of illustration, let us consider a simple example: let

$$f_{20}(z_2, \xi) = q \cdot z_2,$$

and for $i = 1, \ldots, m_2$,

$$f_{2i}(z_1, z_2, \xi) = T_i z_1 + W_i z_2 - h_i(\xi),$$

where T_i, W_i are (fixed) vectors, and h_i is a random variable. Assuming that we have observed h, to find the optimal recourse decision, the problem that needs to be solved is a linear program. And, from parametric programming, we know that there is a piecewise linear function of $h - Tz_1$ that yields the optimal recourse decision. If we do not observe h, or equivalently the "state" $h - Tz_1$, but instead information is some (nonlinear) function of $h - Tz_1$, then, in general, we loose the piecewise linearity of the optimal recourse decision with respect to the state. In order to be able to deal with such problems, we may very well want to restrict the class of acceptable second-stage decision functions to those that that depend on a finite number of parameters.

There is also the question of the dependence of the probability measure on the first-stage decision. We already discussed this in the framework of the one-stage model. The situation is not any different here. There are no new conceptual or theoretical difficulties, beyond those that we mentioned in Section 1, except that we may have to deal with complications generated by the dependencies of P on z_1 and by the restriction of z_2 to the class of functions that are $\mathcal{G}(z_1)$ measurable.

3 MULTISTAGE MODELS

Conceptually, multistage models are straightforward extensions of two-stage models. There are a few technical details that need to be taken care off, but most assertions one can make about such models follow from those that have been established for two-stage models. However, it does pay to analyze in more detail the dynamical aspects of the problem. The real challenge comes from having to deal with what has been called "the curse of dimensionality" in the design of solution procedures. We shall begin with a rather general formulation whose main virtue is that it is simple from a notational and conceptual viewpoint. As in the previous section, we start with the case when the information (inferred from the observations) and the distribution of the random quantities do not depend on past decisions. Once more, let us stress the fact that we do not exclude the possibility of having the observation values depend on earlier decisions (controls).

Although stages of a multistage stochastic optimization problem do not necessarily correspond to time periods, let us use $t = \{1, \ldots, T\}$ to denote the stage-index and refer to it, by abuse of language, as "time". Let ξ_t denote the random quantities that are observed at stage t *before* we have to make our decision, i.e., the t-th stage decision function z_t can depend on all past observations $\xi^t := \{\xi_s, s = 1, \ldots, t\}$.

With $T = 2$ and ξ_1 a degenerate random vector (i.e., whose distribution is concentrated at one point), we recover the two-stage model; the variables denoted

ξ then, are now called ξ_2. We are now allowing for the possibility that the problem considered in Section 2, was actually one of a possible collection of problems obtained after observing ξ_1. This slight generalization of the model comes from a shift in the type of questions that we like to see answered. In the two-stage model, the emphasis was on calculating an optimal first-stage decision, and this is still the case for many multistage problems, but for another wide range of models the accent will be on finding an optimal decision (control) rule that could be applied at *all* stages.

The random quantities of the problem will again be denoted by ξ with $\xi = (\xi_1, \ldots, \xi_T)$. The dependence of the (recourse) decision on past observations can be expressed in the following terms: let $(\Xi, \mathcal{A}, P)$ be the underlying probability space and let $\mathcal{B}_t$ be the (σ-)field of events generated by the observations up to time t; this corresponds to the σ-field generated by the random vector ξ^t. The dependence of z_t on the past observation can thus be expressed in terms of the measurability of z_t with respect to $\mathcal{B}_t$, in other words, z_t must be $\mathcal{B}_t$-adaptable.

The constraints that are explicitly included in the formulation of the problem, will be represented by a multifunction:

$$\Gamma(t, \xi) := \{z^t = (z_1, \ldots, z_t) \text{ that satisfy the } t-\text{th stage constraints}\}.$$

(We use, somewhat indiscriminately, z_t to designate a function from Ξ into the decision space, say R^{n_t}, and a point in its range.)

Thus the *multistage recourse* problem, is to find

$$\begin{array}{ll} \text{for } t = 1, \ldots, T, & z_t \ \mathcal{B}_t - \text{measurable}, \\ \text{for } t = 1, \ldots, T, & z^t \in \Gamma(t, \xi), , \\ \text{that minimizes} & E\{h_0((z_1(\xi), \ldots, z_T(\xi)), \xi)\}. \end{array}$$

Most of the theory developed for one- and two-stage models can be applied to the multistage problem to obtain the basic properties of the deterministic equivalent problem, a number of useful characterizations of the optimal solutions (linearity, piecewise linearity, etc.), as well as necessary and sufficient optimality conditions. However, as already mentioned earlier, one is also interested in the dynamical properties of the solution, in particular in the role played by the dynamical restrictions on the z_t that comes from the $\mathcal{B}_t$-measurability condition.

Let Z be the space of all ($\mathcal{A}$-measurable) functions $z := (z_1, \ldots, z_T)$ defined on Ξ such that for all t, z_t is $\mathcal{B}_t$-measurable; such functions will be called *nonanticipative*. It is a linear subspace of the space of $\mathcal{A}$-measurable functions. From this simple observation follows an important optimality criterion: assuming that the problem at hand satisfies a "standard" constraint qualification, and the constraint-multifunction is nonanticipative, a necessary condition for optimality of z^*, that

is also sufficient in the convex case, is that there exist multipliers $p = (p_1, \ldots, p_T)$ defined on Ξ, orthogonal to Z, i.e., such that

$$E\{p_t(\cdot)|\mathcal{B}_t\} = 0 \ a.s., \ \text{for } t = 1, \ldots, T,$$

and for almost all ξ:

$$z^*(\xi) \in \operatorname{argmin}\{h_0(z, \xi) - p(\xi) \cdot z | z^t \in \Gamma(t, \cdot) \ a.s., \ \text{for } t = 1. \ldots, T\}.$$

Knowledge of these multipliers would reduce the problem to one of pointwise minimization. One can interpret these multipliers as a price system associated with the *nonanticipativity* restrictions; a beautiful economic interpretation of these multipliers in terms of insurance prices has been sketched out by I. Evstigneev from C.E.M.I.(Moscow).

To state the optimality condition, we mentioned the concept of *nonanticipativity* of the constraint multifunction. By this one means the following: at any time time t there are no constraints induced on z^t beyond those already imposed by $\Gamma_s, s = 1, \ldots, t$; i.e., there are no constraints induced by potential future infeasibilities. This means: if z^t satisfies all the constraints up to time t, there exist functions $z_{t+1}, \ldots, z_T$, such that the resulting z is feasible for the multistage recourse problem. We referred to this, in Section 2, as *relatively complete recourse*. By deriving the induced constraints and including them explicitly in the formulation of the problem, any multistage recourse problem can be reduced to one with relatively complete recourse. However, deriving the induced constraint is not necessarily an easy task, and thus the general optimality theory must (and does) make provisions for the case when Γ is not necessarily nonanticipative, and the solution procedures must (and do) cope with the presence of these induced constraints (by introducing feasibility cuts).

In the choice of a solution technique, we have at our disposal all the experience gained from the study of one- and two-stage models, but all the difficulties that we have encountered so far are compounded by the fact that the number of possible realizations is exponentially increasing with the number of stages, the so-called "curse of dimensionality". The only possible remedy is *decomposition*. Decomposition not only with respect to possible realizations, but also, whenever possible, with respect to time (i.e., stages).

We have seen that introducing the multipliers associated with the nonanticipativity constraints, suggests a potential decomposition with respect to the sample (realization) space. This and the notion of an *average problem* have lead to the *aggregation principle* which allows us to solve any multistage recourse problem, by solving (repeatedly) deterministic versions of the original problem for particular realizations of ξ, sometimes called *scenarios*. The basic idea is captured in the *hedging* algorithm.

3.1 Partial Information, etc.

If instead of observing, or being able to infer, the values assumed by ξ^t, the information to which we have access determines a field $\mathcal{G}_t$, a strict subset of σ-field $\mathcal{B}_t$ of possible events, the (recourse) decision must now be $\mathcal{G}_t$-measurable. Let $Z_{\mathcal{G}}$ be the subspace of Z consisting of all $\mathcal{A}$-measurable functions z so that for all t, z_t is $\mathcal{G}_t$-measurable. This is a linear subspace of Z. The same arguments, and the same conditions as before, except for $\mathcal{B}_t$-nonanticipativity of the constraint-multifunction replaced by $\mathcal{G}_t$-nonanticipativity, will yield the following optimality criterion: if z^* solves the multistage recourse problem, there exist multipliers $q = (q_1, \ldots, q_T)$ defined on Ξ such that

$$E\{q_t(\cdot) \,|\, \mathcal{G}_t\} = 0 \; a.s., \text{ for } t = 1, \ldots, T,$$

and for almost all ξ:

$$z^*(\xi) \in \operatorname{argmin}\{h_0(z, \xi) - q(\xi) \cdot z | z^t \in \Gamma(t, \xi), \text{for } t = 1. \ldots, T\}.$$

These conditions are of the same nature as those we already know for the full information case, the only differences are the stronger constraint qualification (nonanticipativity of $\Gamma(t, \cdot)$) with respect to $\mathcal{G}_t$, and the fact that now conditional expectation of q_t is taken with respect to a coarser σ-field. Again there is a rich economic interpretation that can be attached to these multipliers. If p corresponds to the multipliers associated with full information, then $q - p$ yields a price system that could be used to determine if it would be desirable or not, to seek full information; one could think of these multipliers as an information price-system.

As for the two-stage model, it is not always possible to express the information collected (from observations) independently of past decisions. We need to consider also the case when the information fields $\mathcal{G}_t$ depend on $z^{t-1} = (z_1, \ldots, z_T)$; we then write $\mathcal{G}_t(z^{t-1})$. And all the difficulties mentioned in connections with the two-stage model are still all present, except more so. The mathematical complexity generated by asking even the simplest of questions about such models is mind-boggling.

Because the search for an optimal solution will necessarily require, at each iteration a total description of $\xi \mapsto z_t(\xi)$ for all t, the challenge created by *this* formulation of the multistage recourse model may be, for ever, beyond our computational capabilities, unless one replaces the decision space and the sample space by a discrete set. In this discrete case, finding the optimal solution becomes a question of enumerating all possibilities, and this can be organized via dynamic programming techniques. And even that is only possible if the number of decisions in each time step (stage) is rather limited. One other approach is to replace the search for an optimal z^* with the search for the best z in a given class. We return to this in the context of the the stochastic optimal control model.

Finally, we could also have to deal with the dependency of the probability distribution on past decisions z^{t-1}. The added complexity is a function of the form of the relationship between P and z and the properties of h_0 and Γ, when viewed as functions of (z, ξ), just like for the one-stage model.

3.2 Stochastic Optimal Control Models

As we shall see, the formulation of the discrete-time stochastic optimal control model is very similar in nature to that of the multistage recourse models. However, the relationship between these models has not always been very well understood. The basic reason is motivation: the concept of solution is somewhat different in both models. The multistage recourse model is, in many cases, only concerned with z_1, the other decisions are of little interest. The stages 2 to T are only included in the problem to help evaluate the costs that may result from a particular choice of z_1. To the contrary, most of the motivation for the research on stochastic control problems comes from a class of applications where it is the decision *rule* (to be used in all time periods) that is of interest, i.e., the rule that will allow us to pass from observations to decisions. Hence, the insistence of finding a rule that depends on the observed (or estimated) state and not on the information we may infer about the underlying stochastic phenomena. This is only possible if there is a certain similarity between the stages. From a theoretical viewpoint, neither the multistage recourse model nor the stochastic optimal control model is a special case of the other, but there are fundamental differences when it comes to what practitioners will identify as "solvable" problems. Algorithmic research on multistage recourse models is oriented towards mathematical programming techniques, whereas the solution technique favored in the stochastic control literature is dynamic programming. This places natural limitations on the type of problems that can be approached in either way.

We consider the following formulation of a *discrete time, finite horizon, stochastic optimal control* problem:

$$x_t = f_t(x_{t-1}, u_t, \zeta_t^1), \quad t = 1, \ldots, T$$

with initial state x_0 about which we may only have probabilistic information. The variables x_t denotes the *state* of the system, u_t is the *control*, and ζ_t^1 models the system's disturbances (with given probability distribution). The *observations* $y^t = (y_1, \ldots y_t)$ that are available to the controller at time t are related to the state of the system by:

$$y_t = k_t(x_{t-1}, \zeta_t^2), \quad t = 1, \ldots, T$$

where ζ_t^2 are disturbances that affect the observations (again with known probability distribution). The choice of a control law is subject to *system* constraints

(*state-space* constraints and *control* constraints):

$$x_t \in X_t, \qquad u_t \in U_t, \quad t = 1, \ldots, T,$$

and *information* constraints:

$$\text{for } t = 1, \ldots T, \ u_t \text{ is } \mathcal{Y}_t - \text{measurable},$$

where $\mathcal{Y}_t$ is the σ-field generated by the observations, i.e., $\mathcal{Y}_t = \sigma\{y_s \,|\, s \le t\}$.

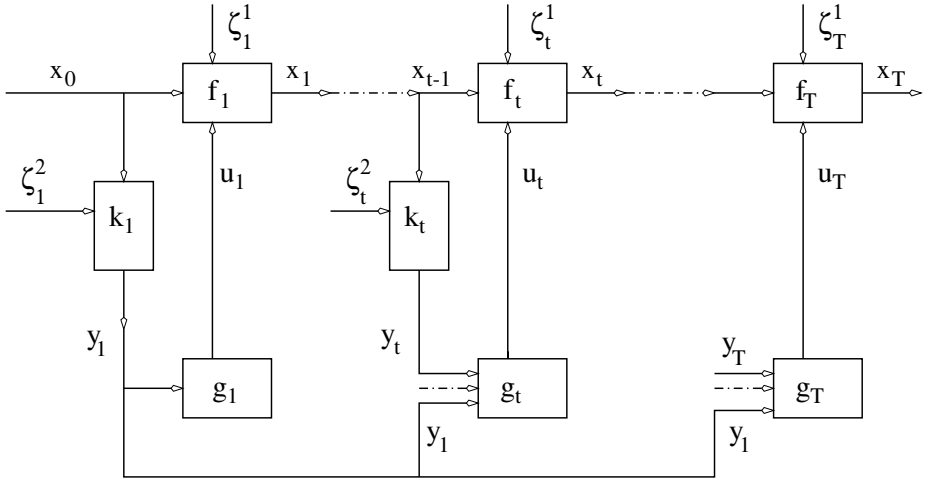

Figure 1: A controlled stochastic system

The choice of the control u_t must be a (measurable) function of the observations, let us denote it g_t,

$$u_t := g_t(y^t) = g_t(y_1, \ldots, y_t) \in U_t.$$

The vector-valued function

$$g = \{g_1, \ldots, g_T\}$$

is called the *feedback law*. Given g, we can define stochastic processes $\{x_t^g\}$, $\{y_t^g\}$, $\{u_t^g\}$ with

$$\begin{aligned} x_t^g &= f_t(x_{t-1}^g, u_t^g, \zeta^1), \\ y_t^g &= k_t(x_{t-1}^g, \zeta_t^2), \\ u_t^g &= g_t(y_1^g, \ldots, y_t^g). \end{aligned}$$

In the ensuing development, we usually drop the reference to g when referring to u, x or y but it is implicitly always there. Figure 1 gives a block diagram representation of the dynamics of the system.

The objective is to choose a feedback control law g^* that minimizes costs (or maximize performance):

$$J(g) := E^g\{\sum_{t=1}^{T} c_t(x_{t-1}, u_t, \zeta_1^t) + \Phi(x_T)\} := E\{\sum_{t=1}^{T} c_t(x_{t-1}^g, u_t^g, \zeta_t^1) + \Phi(x_T)\}.$$

The function Φ plays the role of a terminal condition.

The relation between this model and the multistage recourse model is immediate. Indeed, simply set $z_t := (x_t, u_t)$, $\xi_t := (\zeta_t^1, \zeta_t^2)$, $\xi_0 := x_0$, $\xi := (\xi_0, \ldots, \xi_T)$, $Z_t = X_t \times U_t$,

$$\Gamma(t, \xi) := \{z^t \in Z_t \,|\, x_t = f_t(x_{t-1}, u_t, \zeta_t^1)\},$$

and

$$h_0(z, \xi) := \sum_{t=1}^{T} c_t(x_{t-1}, u_t, \zeta_t^1) + \Phi(x_T).$$

The information constraint, which in the case of the stochastic optimal control model is explicitly included in the model in terms of a feedback law, would in the case of the multistage recourse model take the form: z_t must be $\mathcal{B}_t(z^t)$-measurable, where $\mathcal{B}_t(z^t) := \mathcal{Y}_t$.

There are thus no significant differences between these two models, at least as far as formulation goes. Certainly, any general theoretical result known about any one of these models, has a counterpart for the other one. To cite just a couple of examples, the optimality conditions mentioned earlier can easily be reformulated so that they apply to the stochastic control model. Similarly, qualitative results obtained about the value function of stochastic control problems could be applied to the corresponding class of multistage recourse problems. There are a few results that admit easy translation, whereas others are not so readily adaptable. There are two major features of stochastic control models that are not explicitly included in the recourse models. However, the differences are more a matter of perception (and formulation) than factual. First, the stochastic control model includes an explicit expression for the observation process, and second we are to use a feedback law based directly on the actual observations (rather than on the information gathered about "nature": ξ).

As for multistage recourse problems, the major classifications for stochastic control models is based on the type of feedback that will be called for, or/and the level of information that will be available to the controller.

OPEN LOOP : No information is collected that would enable us to adjust earlier decisions. This corresponds to having $y_t \equiv h_t \equiv 0$ for $t = 1, \ldots, T$. The

selection of u_t, can as well be made from the very outset. We could extend this model to include those cases that allow for "local" adjustments, i.e., adjustment that are made at time t that do not affect the selected trajectory but try to remedy local deviations from a desired state. This latter case is then of the same nature as the dynamical two-stage model mentioned in Section 2. Such models are sometimes used with a *rolling horizon*, however the use of such an approach cannot always be recommended, since it arbitrarily ignores feedback (or recourse) possibilities that are inherent to all stochastic optimization problems. One further restriction would be to insist on *myopic* controls.

COMPLETE INFORMATION : Full information is available about the state, i.e.

$$y_t = k_t(x_{t-1}, \zeta_t^2) = x_{t-1};$$

we refer to this case as *full state-information*. This should not be confused with what we have called full information in the framework of the multistage recourse model. In fact, full state-information, may or may not correspond to the full information case. A *nice* case when one can identify full state-information with full information, is when (ζ^1 and ζ^2 are strongly correlated):

$$\begin{aligned} y_t &= k_t(x_{t-1}, \zeta_t^2) = \zeta_t^1, \\ x_t &= f_t(x_{t-1}, u_t) + \zeta_t^1. \end{aligned}$$

If in addition, the random variables ζ_t are time-independent, then dynamic programming techniques can be used as a solution technique. This is the first time that we encounter in our discussion, this independence condition. This is *not* a modeling choice, but one dictated by the solution technique; more about this later.

PARTIAL INFORMATION : This is the general case. Let us stress once more that this does not correspond to what we have been calling partial information in the context of the multistage recourse model; to make sure that this distinction is not lost, we shall refer to this case as *partial state-information*. Here again is it is possible to appeal to dynamic programming techniques for finding the optimal feedback law. Instead of using the state of the system we rely on on an extended notion of state, viz., conditional distributions (on the state-space) will play the role of the state. These conditional distributions are sometimes called *hyperstates* or *information states*.

FEEDFORWARD : In this case the information available at time t, is either ζ_t^1 or a function of ζ_t^1, in other words the information is a random variable

strongly correlated with ζ_t^1. If we take ζ_t^2 to be such a variable, then in terms of the stochastic optimal control problem, we could think of it as the case when

$$y_t = \zeta_t^2.$$

We receive direct information about the underlying stochastic phenomena. Without any need to adjust the information collected, we are in the framework of the multistage recourse model with full or partial information.

RESTRICTED FEEDBACK : Rather than allowing for g to be just any measurable function of the observations, we may want to restrict the class of admissible feedback laws to a particular (parametrized) class of functions. We already discussed this option in the context of the multistage recourse model. From a computational view point, this looks very attractive. But, before we really can use this approach, there are many unresolved theoretical questions that deserve serious investigation. More precisely, we need to characterize, as well as possible, the properties of optimal feedback and obtain error bounds when restrictions are placed on the class of admissible controls. Note that there are some models for which the optimal law is known and can be characterized in terms of a finite number of parameters, e.g., (s, S)-policies, impulse controls, certain bang-bang situations, etc..

The stochastic optimal control model may also include a *filtering* equation, i.e. a process used to analyze the observations in order to obtain an estimate of the state of the system. Instead of using the data that comes from the observations, we are to use the filtered data. If the filter is known *a priori*, then our formulation already allows for such a possibility, we simply define k_t appropriately and take y_t to be the filtered data. If, we are allowed to choose both an optimal control and an optimal filter, the problem is not so simple. In a few cases, one can appeal to the *Separation Lemma* which allows us to first calculate an optimal filter, and use it (redefining k_t) to calculate the optimal feedback law. In general, the situation is unfortunately much more complex. Although this is an important issue, we shall not be concerned with it here; we implicitly assume that we are using raw data (observations) or if it is filtered data (state estimates) the function k_t has been defined so as to include the filtering process.

There is a substantial literature devoted to the characterization of optimality centered around the Hamilton-Jacobi-Bellman equation (discrete or continuous time versions). The suggested solution methods for stochastic control problems are mostly based on solving that equation. They range from discretization (of state-space, controls and possible realizations) to Monte-Carlo simulations passing through finite element approximations of the Hamilton-Jacobi-Bellman equation.

We shall only discuss the "discrete" case, and this in the setting of full or partial state-information; for simplicity's sake, we also assume that there no state-space constraints, i.e., no constraints of the type $x_t \in X_t$.

This approach relies on a crucial assumption that has not been needed up to now:

Assumption: The random variables $x_0, \zeta_t^1, \zeta_t^2, \ldots, \zeta_T^1, \zeta_T^2$ are mutually *independent*.

This has the following implication: for all g,

$$\begin{aligned} P^g\{x_t \in D \;\; | \, x_{t-1}, \ldots, x_0, u_t, \ldots, u_1\}, & \\ = P\{x_t \in D \,|\, x_{t-1}, u_t\} \quad & \text{independent of } g, \\ = P\{\zeta_t^1 \in Q(x_{t-1}, u_t)\} & \end{aligned}$$

where

$$Q(x_{t-1}, u_t) := \{\zeta \,|\, f_t(x_{t-1}, u_t, \zeta) \in D\}.$$

We can reformulate the problem in terms of the following equivalent *Markov Decision Problem*: given the "controlled transition probabilities"

$$P(d\, x_{t-1} \,|\, x_{t-1}, u_t)$$

and the observation channel transition probabilities,

$$P(d\, y_t \,|\, x_{t-1}),$$

find $g = (g_1, \ldots, g_T)$, that minimizes

$$E^g \sum_{t=1}^{T} \hat{c}_t(x_{t-1}, u_t),$$

where

$$\hat{c}_t(x, u) := \int c_t(x, u, \zeta_t^1) P(d\, \zeta_t^1).$$

3.2.1 Full state-information

Now, if for all $t = 1, \ldots, T$, full state-information is available, i.e., $y_t \equiv x_t$, we define recursively the real-valued functions:

$$V_T(x_T), \ldots, V_0(x_0),$$

by

$$\begin{aligned} V_T(x_T) &:= \Phi(x_T), \\ &\vdots \\ V_t(x_t) &:= \min_{u \in U_t}\{\hat{c}_t(x_t, u) + \int V_{t+1}(x_{t+1}) P(d\, x_{t+1} \,|\, x_t, u)\} \end{aligned}$$

with $\hat{c}_0 \equiv 0$. Then

$$V_t(x) = \min_g \{E^g \sum_{s=t}^{T} \hat{c}_s(x_s.u_s) \,|\, x_t = x\}.$$

If

$$g_t^*(x) \in \operatorname*{argmin}_{u \in U_t}\{\hat{c}_t(x, u) + \int V_{t+1}(x_{t+1}) P(d\, x_{t+1} \,|\, x, u)\},$$

then

$$u_t = g_t^*(x_{t-t}), \text{ for } t = 1, \ldots, T.$$

is the optimal feedback law. In particular, note that u_t is Markovian, in that it only depends on $x_{t-1} = y_t$ and not on earlier observations $y_{t-1}, \ldots, y_1$.

3.2.2 Partial state-information

When only partial information is available, i.e., $y_t \neq x_{t-1}$, let

$$v^t := (y^t, u^{t-1}),$$

denote the information available when choosing u_t. Fix the feedback law g, and define

$$\pi_t^g(d\, x \,|\, v^t) := P\{x_t^g \in d\, x \,|\, v^t\}.$$

A fact which is of crucial importance to the development that follows is that π_t^g does not depend on g. It can be shown that there exists an operator S_t, sometimes called a *'filter'*, such that for $t = 1, \ldots, T$,

$$\pi_{t+1}(\cdot \,|\, v^{t+1}) = S_t[\pi_t(\cdot \,|\, v^t), y_{t+1}, u_t]$$

and

$$\pi_1(d\, x \,|\, v^1) = P\{x_0 \in d\, x \,|\, y_1\}.$$

Let Π be the space of all probability distributions on the state-space. For example, if $x_t \in \{1, \ldots, I\}$, then

$$\Pi = \{\pi_1, \ldots, \pi_I \,|\, \sum_1^I \pi_i = 1, \pi_i \geq 0\}.$$

In a manner similar to that used in the full state-information case, we define real-valued functions, but on Π, the hyperstate-space:.

$$\begin{aligned} V_T(\pi) &:= E\{\Phi(x) \,|\, \pi_T(\cdot \,|\, v^T) = \pi\}, \\ &\vdots \\ V_t(\pi) &= \min_{u \in U_t} E\{\hat{c}_t(x,u) + V_{t+1}(S_t[\pi, y_{t+1}, u]) \,|\, \pi_t(\cdot \,|\, v^t) = \pi\}. \end{aligned}$$

Then, for all g,

$$V_t(\pi_t(\cdot \,|\, v^t)) \leq E^g\{\sum_{s=t}^{T} c_s(x_{s-1}, u_s) \,|\, (y^t, u^{t-1})\},$$

and

$$u_t = g_t^*(\pi_t(\cdot \,|\, v^t))$$

is the optimal feedback law, where g_t^* is the argument that yields the minimum in the expression that defines V_t.

3.2.3 Computational implications

We have given a rather detailed description of the theoretical underpinnings of the methods used in practice to solve discrete-time stochastic optimal control problems. The reason is that we want to stress the differences between this approach and that favored for multistage recourse models. In both cases, full or partial state-information, the strategy has been to reduce the control problem to a Markov decision problem. To achieve this and to be able to solve the problem, we had to impose two unwelcome restrictions:

1. time-independence of the random variables plus independence between the disturbances that affect state and observations (althought this latter restriction is inessential),

2. finite state-space, which in turn implies finitely distributed random variables and discrete control space.

These limitations are not always easy to justify in applications. At our present stage of development, that seems to be the price that needs to be paid to build a feedback control law based on information obtained about the state of the system rather than information about the underlying stochastic process.

Unless the state-space is actually discrete and the underlying stochastic process $\{\zeta_t\}_{t=1}^T$ consists of independent random variables, the solution obtained by solving the Markov decision model is, at best, an approximation of the problem at hand.

4 CONTINUOUS-TIME MODELS

We shall be very brief: there is not much to report from a (practical) computational viewpoint. Although the discrete time model did allow for a wide variety of stochastic disturbances, the only case that has really been studied in continuous-time is when the disturbances can be modeled by white noise (although, now, there are also martingale techniques). Defining the variables as the obvious continuous-time analogues of those of the discrete-time models, the *continuous-time* recourse model takes the form:

$$\begin{array}{rcl} \text{minimize} & E\{\int h_0(z_t(\xi), \xi_t)dt\} & \\ \text{subject to} & z_t \in \Gamma(t, \xi) & \text{for all } t, \\ & z_t \ \mathcal{B}_t - \text{measurable} & \text{for all } t, \end{array}$$

where $\xi = (\xi_t)$ is a (continuous-time) stochastic process, $\mathcal{B}_t$ is a σ-field generated by earlier observations that may or may not depend on past decisions. Again the question of the nonanticipativity of the constraint-multifunction needs to be broached, and it plays a role in the type of conditions that can be used to characterize optimal solutions, etc..

The continuous-time version of the stochastic control problem that has received most of the attention in the literature is:

$$\begin{array}{rcl} \text{minimize} & E\{\int c_t(x_t, u_t, \zeta_t^1)dt\} & \\ \text{such that} & d\,x_t = f_t(x_t, u_t)dt + \sigma_1(x_t)d\,\zeta_t^1, & \text{for } t \in [0, T], \\ & d\,y_t = k_t(x_t)dt + \sigma_2(x_t)d\,\zeta_t^2, & \end{array}$$

where ζ_t^1 and ζ_t^2 are Wiener processes (or more generally semi-martingales) that model disturbances that affect system and observations. The variable u_t is the control that is subject to the information constraint:

$$u_t \text{ is } \mathcal{Y}_t - \text{measurable},$$

with $\mathcal{Y}_t$, as before, the σ-field generated by the observations $\{y_s \,|\, s \leq t\}$. There are some technical difficulties with giving a precise meaning to this constraint. To do so, one usually relies on a measure transformation (Girsanov's Lemma).

The continuous-time versions of the multistage recourse model as well as the stochastic optimal control model are (mathematical) analyst's delight. As soon as one goes beyond the quadratic regulator problem (a linear-quadratic model), there are essentially no closed-form solutions and most of the theory has been oriented at finding qualitative characterizations of optimal solutions. One could consult the work of Back and Pliska for the continuous-time recourse model, and that of

Krasovskii, Fleming, Rishel, Kushner, Varaiya, Bensoussan, Evans, Lions (père & fils), Davis, Krylov and many others, for the continuous-time stochastic control model. The most computationally oriented work is probably that of Haussmann (Univ. British Columbia), beginning with his work on the stochastic maximum principle. However, very little success can be reported about the passage from theory to computationally implementable techniques; we exclude here, for obvious reasons, methods based on Monte-Carlo simulations and stochastic approximation techniques (that have a limited range of applicability).

Certain continuous time models have equivalent discrete-time (or discrete state-space) formulation, and sometime this can be exploited to solve (by successive approximations) more complicated problems. Let us give two examples. If the dynamics of the system are described by a continuous-time Markov chain (finite state-space), i.e.,

$$P\{x_{t+dt} = j \,|\, x_t = i, u_t = u\} = p_{ij}(u)dt,$$

it is usually possible to convert to problem to one in discrete time by a technique know as *uniformization*. The second example is a little bit more involved. It is a class of problems studied first by Vermes (Hungary), and at present, under further investigation by Davis (Imperial College). The state at any time t is the sum of a jump process (Markov jumps that occur at random times) and a dynamical system described by an ordinary differential equation that can be controlled. Certain maintenance problems and capacity expansion problems are easy to cast in this mold. Problems of this type can be converted to multistage recourse problems (possibly with an infinite number of stages), where each stage corresponds to the evolution that takes places between jumps and the (recourse) costs are random variables whose values depend on the length of time between jumps.

REFERENCES

[1] Marc H.A. Davis, "Martingale methods in stochastic control," in *Stochastic Control Theory and Stochastic Differential Systems*, M. Kohlmann & W. Vogel, eds., Springer Lecture Notes in Control and Information Sciences 16, Berlin, 1979, 85–117.

[2] Marc H.A. Davis, "Piecewise-deterministic Markov Processes: a general class of non-diffusion stochastic models," *J. Royal Statistical Society, Series B* 46(3) (1984).

[3] Wendel H. Fleming & Raymond W. Rishel, , *Deterministic and Stochastic Optimal Control*, Springer Verlag, Berlin, 1975.

[4] P. R. Kumar & Pravin Varaiya, *Stochastic Systems. Estimation, Identification and Adaptive Control*, Prentice-Hall, Englewood Cliffs, 1986.

[5] Harold J. Kushner,, *Probability Methods for Approximations in Stochastic Control and for Elliptic Equations*, Academic Press, New York, 1977.

[6] Hans S. Witsenhausen, "A counterexample in stochastic control," *SIAM. J. on Control* 6(1) (1968.).