Conditions for optimality in multi-stage stochastic programming problems

EINDHOVEN UNIVERSITY OF TECHNOLOGY

Department of Mathematics

PROBABILITY THEORY, STATISTICS, OPERATIONS RESEARCH, AND SYSTEMS THEORY GROUP

Memorandum COSOR 79-05

Conditions for optimality in multi-stage stochastic programming problems

by

Luuk Groenewegen*) and Jaap Wessels

*) Rijkswaterstaat, Data Processing Division, Rijswijk (Z.H.).

Eindhoven, March 1979

The Netherlands

Conditions for optimality in multi-stage stochastic programming problems

by

Luuk Groenewegen*) and Jaap Wessels**)

Summary. In this paper it is demonstrated how necessary and sufficient conditions for optimality of a strategy in multi-stage stochastic programs may be obtained without topological assumptions. The conditions are essentially based on a dynamic programming approach. These conditions - called conserving and equalizing - show the essential difference between finite-stage and ∞-stage stochastic programs.
Moreover, it is demonstrated how a recursive structure of the problem can give a reformulation of the conditions. These reformulated conditions may be used for the construction of numerical solution techniques.

1. Introduction

In this paper it will be shown how it is possible for a very general class of multi-stage stochastic decision problems to give necessary and sufficient conditions for the optimality of a strategy. Since we will not introduce topological assumptions, it is not possible to give duality assertions. So, the theory will be based on primal properties of the decision problems. In some sense the theory is a generalization of dynamic programming. The theory will also show why the step from a finite-stage problem to an infinite-stage problem is a difficult one. It is also demonstrated for which structures the optimality conditions may be formulated locally in time. Such a formulation facilitates computation considerably. Since many stochastic programming problems do not have such a structure, they present essential computational difficulties. However, in some cases it is feasible to reformulate the problem in order to give it this special structure.

Actually, the theory which will be presented here, can be generalized to noncooperative dynamic games in continuous time. This more general theory has been worked out by Groenewegen in his doctoral dissertation and will be published by him as a monograph [5]. See also [6]. In continuous time the set-up must be less constructive, since no continuous-time version of the

*) Rijkswaterstaat, Data Processing Division, Rijswijk (Z.H.), the Netherlands.
**) Eindhoven University of Technology, Dept. of Mathematics, Eindhoven, the Netherlands.

Ionescu Tulcea construction for making a probability space from transition probabilities is available.

The search for necessary and sufficient conditions for the optimality of a strategy in a rather generally formulated multi-stage decision problem has not been triggered by the idea that new or better conditions for specific problems can be found. The main drive has been, that is is worthwhile to make clear what the well-known conditions have in common and what the essential circumstances are for these conditions to work.
As stated, the conditions for optimality that will be presented in this paper may be seen as an outgrowth of the dynamic programming approach and therefore its traces go back to Bellman's optimality principle. The intrinsic difficulties for the characterization of optimality in infinite-stage decision problems have been discovered and solved for gambling houses by Dubins and Savage [1] and by Sudderth [12]. They show that for their type of problems an extra condition is required to guarantee optimality. The standard condition (called *conservingness*) says that the strategy should maintain its potential reward over the stages. The extra condition (called *equalizingness*) says that the strategy should cash its potential reward in the long run. This has been generalized to Markov decision processes by Hordijk [8]. For rather general multi-stage stochastic decision processes the characterization has been given independently by Kertz and Nachman [9], however, they need a topological structure and obtain the result in a rather indirect and unnecessary difficult way.

The set-up of the multi-stage stochastic programming problem, as it will be formulated in section 2, bears the traces of its dynamic programming background. However, it should be clear that e.g. the rather general type of stochastic programming problems from Rockafellar and Wets [11] fit into this structure. In fact, the dynamic programming set-up not only facilitates the formulation of optimality conditions, it also facilitates the formulation of essential structural properties of the problem like non-anticipativity of the strategies.
Section 3 contains the characterization of optimal strategies for multi-stage stochastic programs. In section 4 this characterization is reformulated in terms of local quantities for the situation that the problem has a recursive structure. Section 5 is devoted to some additional remarks.

2. The multi-stage stochastic programming problem

In this section we will formulate the basic model for the theory which will be developed in subsequent sections. As stated in the introduction, the model has a dynamic programming flavour, but it is essentially more general than the usual model for Markov decision processes. It is also more general than the rather general Markovian models of e.g. Hinderer [7] and of Furukawa and Iwamoto [3]. We will come back on this aspect at the end of this section.

Suppose that actions have to be selected at subsequent stages or time instants numbered by $t = 0,1,2,\ldots$. At each stage some variable is observed. According to our dynamic programming set-up we call the actual value of this variable the state of the system. This state is supposed to be an element of a given set X, which might be a different one for different stages; however, for simplicity of notations we will take the same state space X for all stages. In stochastic programming terminology one would say that the state at time t is the random observation of stage t. X is supposed to be endowed with a σ-field $\mathcal{X}$. After the observation of the state of the systen at a certain stage, one has to select an action from an action space A. Without extra difficulty this action space might depend on the stage number, however, we will not incorporate that feature. The action space A is supposed to be endowed with a σ-field $\mathcal{A}$.
Especially for recourse problems one needs the following aspect of the model. It is not necessarily true that at all stages the same actions are allowed, in recourse problems for instance the set of allowed actions may depend on all preceding observations and actions. Therefore, we suppose that for each stage t a subset L_t of $\times_{\tau=0}^{t} (X \times A)$ has been given (L_t is supposed to be an element of the product-σ-field). The interpretation of L_t is such that $(x_0,a_0,\ldots,x_t,a_t) \in L_t$ means that the action $a_t \in A$ is admissible if $x_\tau \in X$ were the observations at the corresponding stages for $\tau = 0,\ldots,t$ and $a_\tau \in A$ were the selected actions at the stages $\tau = 0,\ldots,t-1$. L_t should be such that for any sequence $x_0,a_0,\ldots,x_t$ there is at least one admissible action.

Now we are able to introduce the concept of strategy. This concept should be defined in such a way that the selected action at some stage can depend on the previous observations and actions. Moreover, we will define it in such a way that mixed actions are allowed.

A strategy $s = (s_0, s_1, \ldots)$ is a sequence of transition probabilities (in stochastic programming terminology: recourse probabilities) such that s_t is a transition probability from $\times_{\tau=0}^{t-1} (X \times A) \times X$ (with the appropriate product-σ-field) to A. This means that $s_t(x_0, a_0, \ldots, a_{t-1}, x_t; .)$ is a probability measure on A. Naturally, we require that this measure is concentrated on the set of admissible actions for $x_0, a_0, \ldots, x_t$. It also means that $s_t(.; A')$ is measurable for any $A' \in \mathcal{A}$.
Note that the non-anticipativity requirement has been built in quite naturally.

In a sensible multi-stage decision model a strategy and a starting state determine the probabilistic properties of the process. Therefore we have to introduce now the propulsion or transition mechanism of the system for given strategy and starting state or starting distribution. An appropriate way of doing this is by assuming a transition probability p_t for every stage t, such that p_t is a transition probability from $\times_{\tau=0}^{t} (X \times A)$ (but essentially L_t) to X. Now p_t gives for the sequence $(x_0, a_0, \ldots, x_t, a_t)$ of observations and actions a probability measure for the observation or state at stage $t+1$. Using the alternate transition mechanisms of the strategy (s_t) and the propulsion mechanism or observational device (p_t), we can easily construct a probability measure on $H = \times_{\tau=0}^{\infty} (X \times A)$ which describes the process of observations and actions properly:

This measure $\mathbb{P}_{x,s}$ (where x is a given starting state and s a given strategy) is uniquely determined by its values for the finite cylinder-sets $H' = X_0 \times A_0 \times \ldots \times A_{t-1} \times X_t \times A \times X \times \ldots$

$$\mathbb{P}_{x,s}(H') := \int_{A_0} s_0(x; da_0) \int_{X_1} p_0(x, a_0; dx_1) \ldots \int_{X_t} p_{t-1}(x, a_0, \ldots, a_{t-1}; dx_t).$$

That this probability measure $\mathbb{P}_{x,s}$ is the only appropriate one for our purposes is a consequence of a theorem of Ionescu Tulcea (see Neveu [10] th. V.1.1 and its corollaries).

So, for any starting distribution ν on X (we suppose ν to be fixed from now on) we have a probability measure $\mathbb{P}_s$ for every strategy s on H which describes our process properly

$$\mathbb{P}_s(H') := \int_X \mathbb{P}_{x,s}(H') \nu(dx) \quad ,$$

where H' is any subset of H, measurable with respect to the product-σ-field. Expectations with respect to this probability measure will be denoted by $\mathbb{E}_s$.

In order to compare strategies one needs a criterion. Therefore, we introduce a measurable utility function r on H. As a criterion we might use the expected utility

$$v(s) := \mathbb{E}_s r$$

and hence we assume r to be quasi-integrable with respect to all measures $\mathbb{P}_s$. We also need the conditionally expected utilities given actions and observations until some stage. We therefore assume that $\mathbb{P}_{h_t,s}$ is a fixed version of the probability measure $\mathbb{P}_s$ conditioned with respect to $\mathcal{H}_t$ and denoted by $\mathbb{P}_s^{\mathcal{H}_t}$, where $h_t = (x_0,a_0,\ldots,a_{t-1},x_t)$ and $\mathcal{H}_t$ is the product-σ-field in $\times_{\tau=0}^{t-1} (X \times A) \times X$. So, now we can also speak about the value of a strategy s given the history h_t:

$$v_t(h_t,s) := \begin{cases} \mathbb{E}_{h_t,s} r & \text{if } r \text{ is quasi-integrable} \\ -\infty & \text{otherwise .} \end{cases}$$

Note that r is quasi-integrable with respect to $\mathbb{P}_{h_t,s}$ for $\mathbb{P}_s$-almost all h_t. So, the proviso in the definition of $v_t(h_t,s)$ has no practical meaning.

As optimality criterion we would like to choose:

the strategy s^* is optimal if

$$v_0(h_0,s^*) = \sup_s v_0(h_0,s) \qquad \text{for } \nu\text{-almost all } h_0 \in X .$$

A strategy which is optimal in this sense also maximizes the function v. Let us denote $\sup_s v_t(h_t,s)$ by $w_t(h_t)$, then the definition of optimality becomes

$$v_0(h_0,s^*) = w_0(h_0) \qquad \text{for } \nu\text{-almost all } h_0 \in X .$$

One might think that a strategy which is optimal in this sense also maximizes $v_t(h_t,s)$. So, the question is: does an optimal strategy s satisfy for all $t = 1,2,\ldots$

$$v_t(h_t,s) = w_t(h_t) \qquad \text{for } \mathbb{P}_s\text{-almost all } h_t?$$

In order to prove this, one is tempted to suppose the contrary for some t and to use this for the construction of a strategy which is better than s. However, for this kind of construction one needs a selection type argument. This type of argument requires some topological structure. This structure can be made in several ways, each allowing application of a different selection theorem. Since that type of structure would not be used any further in this paper, we prefer it to extend the definition of optimality in such a way that this point is circumvented:

Definition. The strategy s is optimal, if it satisfies for all $t = 0,1,2,\ldots$

$$v_t(h_t,s) = w_t(h_t) \quad \text{for } \mathbb{P}_s\text{-almost all } h_t .$$

Note: another way to circumvent this difficulty is by formulating the conservingness condition in terms of expectations instead of almost everywhere (the criterion then also needs a slight revision). However, then one requires a selection type argument to prove that an optimal strategy is also pointwise optimal for almost all starting states.

Now, we can return to our remark about the generality of the model at the beginning of this section. Our model is definitely more general because of the complete lack of topological requirements. Formally, it is also more general because of the non-Markovian structure of the transition and action mechanism. Moreover, it is formally more general because of the nonrecursiveness of the reward structure (compare section 4). However, these last three aspects can also be brought in the models of e.g. Hinderer [7] and Furukawa and Iwamoto [3] by incorporating the history of the process into the state and by splitting the rewards in additive parts. So, in this respect our model is only slightly more general. However, since we don't need such tricks, it is more direct and more natural.

3. The characterization of optimal strategies

For any strategy s we have for $\mathbb{P}_s$-almost all h_t

(3.1) $$v_t(h_t,s) = \mathbb{E}_{h_t,s}r(h) = \mathbb{E}_{h_t,s}\mathbb{E}_{h_{t+1},s}r(h) = \mathbb{E}_{h_t,s}v_{t+1}(h_{t+1},s) .$$

If s is optimal, we have moreover for any τ and $\mathbb{P}_s$-almost all h

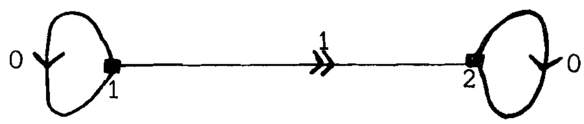
(3.2) $$w_\tau(h_\tau) = v_\tau(h_\tau,s) .$$

So, we obtain by combining (3.1) with (3.2) for $\tau = t, t+1$ for any optimal strategy s:

(3.3) $$w_t(h_t) = \mathbb{E}_{h_t,s}w_{t+1}(h_{t+1}) \quad \text{for } \mathbb{P}_s\text{-almost all } h_t .$$

(3.3) formulates a martingale property for the sequence $\{w_t(h_t)\}_{t=0}^{\infty}$. Because of the conservational character of the formula (3.3), we will call a strategy which satisfies (3.3) for any t a *conserving* strategy. So, we have proved that any optimal strategy is conserving and the question arises whether the reverse is true or not.
A simple example shows that the reverse is not true in the ∞-stage case.

Counterexample.



In this deterministic example there are 2 states, with 2 actions allowed in state 1 (resulting in a return to 1 and a transition to state 2 respectively) and only one action in state 2. Each action provides the reward as given with the appropriate arc. Now, the strategy "stay in state 1" is conserving. It never looses its prospective gain, but it also never cashes this gain.

So, what should be added to the conservingness property in order to guarantee optimality of a strategy, is some condition enforcing the cashing of prospective rewards. It should be noted here, that in finite-stage dynamic programming problems and also in many ∞-stage dynamic programming problems (e.g. discounted problems) the solution techniques are essentially based on (3.3), which shows that those problems don't need an extra condition.

A simple formulation of such a cashing condition is the following:

(3.4) $$\lim_{t\to\infty} \mathbb{E}_s[w_t(h_t) - v_t(h_t,s)] = 0 \ .$$

If a strategy s satisfies (3.4), we say that s is *equalizing* (implicitly its definition presupposes the existence of the relevant integrals). In finite-stage stochastic programs (i.e. after a fixed number of stages the system is in some absorbing state where nothing happens anymore) any strategy is equalizing.
For an optimal strategy s we have

$$\mathbb{E}_s[w_t(h_t) - v_t(h_t,s)] = \mathbb{E}_s[w_t(h_t) - w_t(h_t)] = 0 \ ,$$

hence any optimal strategy is equalizing.

Theorem. A strategy s is optimal if and only if it is conserving and equalizing.

Proof. It only remains to be proved that a strategy which is conserving and equalizing is also optimal.
Suppose s is conserving and equalizing.
The conservingness (martingale property) implies for $\tau > t$

$$w_t(h_t) = \mathbb{E}_{h_t,s} w_\tau(h_\tau) \quad \text{for } \mathbb{P}_s\text{-almost all } h_t .$$

Hence $\mathbb{E}_s w_t(h_t) = \mathbb{E}_s w_\tau(h_\tau)$ for all t,τ.
So (3.4) implies

$$\mathbb{E}_s w_t(h_t) = \lim_{\tau\to\infty} \mathbb{E}_s v_\tau(h_\tau,s) = \lim_{\tau\to\infty} \mathbb{E}_s r(h) = \mathbb{E}_s r(h) = \mathbb{E}_s v_t(h_t,s) .$$

Since

$$w_t(h_t) \geq v_t(h_t,s) ,$$

it follows that

$$w_t(h_t) = v_t(h_t,s) \quad \text{for } \mathbb{P}_s\text{-almost all } h_t .$$ □

Many dynamic programming problems are solved by using the conservingness requirement for optimal strategies. This is possible since in many problems (e.g. finite-stage or discounted problems) all strategies are equalizing. This is not so natural in the typical ∞-stage stochastic programming set-up. Therefore, there is an essential difference between finite-and infinite-stage stochastic programs. But even finite-stage stochastic programs are numerically difficult. This is caused by another difference between dynamic programs and stochastic programs. In dynamic programs we find some sort of recursive structure which makes it possible to reformulate (3.3) in one-period quantities. This is not always the case for stochastic programs as will be demonstrated in the subsequent sections. However - as will be pointed out in section 5 - problems may be reformulated as recursive problems. The prize for this consists mainly of a more extensive state space.

4. Stochastic programs with recursive structure

In most multi-stage stochastic programs the utility function r is the sum of rewards (or costs) for the individual stages. Usually, these single-stage rewards only depend on local quantities like the actions at that stage. So, the influence of one stage on total rewards is completely determined by local quantities, which gives the utility some sort of Markovian property. Unfortunately, it is also typical for stochastic programs that the allowed action set at stage n depends on the proceedings at the foregoing stages. This de-

pendency is not usual in dynamic programming problems which clarifies the difference in numerical solution possibilities.
In this section we will introduce a general recursive structure for multi-stage stochastic programming problems and show how such a structure simplifies the concepts conserving and equalizing. The idea of recursiveness and its basic meaning for the use of dynamic programming ideas in more general multi-stage decision problems stems essentially from Furukawa and Iwamoto [3].

<u>Definition 4.1.</u> The multi-stage stochastic programming problem is called t-recursive for some t if

a) the transition probabilities p_τ and the sets of admissible actions at stage τ do not depend on the state-action history $x_0, a_0, x_1, \dots, x_{t-1}, a_{t-1}$ before stage t for all $\tau \geq t$.
b) r(h) can be separated as follows

$$r(h) = \theta(h_t) + X(h_t)\rho(\zeta^t(h)) \ ,$$

where θ is integrable, X is nonnegative and integrable, ρ is quasi-integrable with respect to $\mathbb{P}_s$ for every s and ζ is the shift operator for histories, so $\zeta(x_0, a_0, x_1, a_1, \dots) = (x_1, a_1, \dots)$.

This concept of t-recursiveness makes it possible to formulate the tail of a multi-stage decision problems as a multi-stage decision problem only depending on the new starting state.

<u>Lemma 4.1.</u> If the multi-stage stochastic programming problem is t-recursive for some t, then

$$v_t(h_t, s) = \theta(h_t) + X(h_t) v_{[t]}(x_t, s(h_t)) \tag{4.1}$$

$$w_t(h_t) = \theta(h_t) + X(h_t) w_{[t]}(x_t) \tag{4.2}$$

for all strategies s and all h_t.
Here $s(h_t)$ is the strategy for the decision problem from stage t on, which applies s as if h_t preceded; $v_{[t]}$ is the value for the problem from time t on with ρ as utility function and it depends on the starting state at stage t and on the strategy for the tail problem; similarly $w_{[t]}$ is the optimal expected utility for the tail problem as a function of the starting state at stage t.

This lemma is trivial and the formal proof only requires a somewhat more formal introduction of the tail problem with utility function ρ.
The interesting aspect of the lemma is, that (4.2) suggests the optimality principle from dynamic programming. Namely, if one tries to find a strategy s with

$$v_t(h_t,s) = w_t(h_t) \quad \text{almost surely ,}$$

then one has to find s such that

$$v_{[t]}(x_t,s(h_t)) = w_{[t]}(x_t) \quad \text{almost surely.}$$

These properties may be used systematically, if the problem is t-recursive for all $t \in T$, where it is desirable that the functions ρ are strongly related. For this purpose we introduce the concept of recursiveness.

Definition 4.2. The multi-stage stochastic programming problem is called *recursive* if

a) it is t-recursive for all $t \in T$; now the separating functions are called $\theta_t^{[0]}$, $\chi_t^{[0]}$, $\rho^{[t]}$.
b) $\rho^{[t]}$ satisfies:

$$\rho^{[0]} = r$$

$$\rho^{[t]}({}_t h) = \theta_\tau^{[t]}({}_t h_\tau) + \chi_\tau^{[t]}({}_t h_\tau)\rho^{[\tau]}(\zeta^{\tau-t}{}_t h) \quad \text{for } \tau \geq t ,$$

where ${}_t h = (x_t,a_t,x_{t+1},\ldots)$, ${}_t h_\tau = (x_t,a_t,x_{t+1},\ldots,x_\tau)$, $\theta_\tau^{[t]}$ is integrable, $\chi_\tau^{[t]}$ is nonnegative and integrable with respect to $\mathbb{P}_s$ for every s.

Except for some trivialities this decomposition is unique. It is also apparent that this decomposition implies a relation between the functions $\theta_\tau^{[t]}$, $\chi_\tau^{[t]}$. These relations show somewhat more explicitly that $\theta_{t+1}^{[t]}$ can be interpreted as a single-stage reward function and $\chi_{t+1}^{[t]}$ as some sort of discount factor for the appropriate stage.

Lemma 4.2. Let r be recursive, then (note that ${}_\bullet h_\tau = h_\tau$ and write $\theta_\tau^{[0]}$ as θ_τ, $\chi_\tau^{[0]}$ as χ_τ)

a) $$\theta_\tau(h_\tau) = \sum_{k=1}^{\tau} \{ \prod_{\ell=1}^{k-1} \chi_\ell^{[\ell-1]}({}_{\ell-1}h_\ell)\}\theta_k^{[k-1]}({}_{k-1}h_k) .$$

b) $$\chi_\tau(h_\tau) = \prod_{k=1}^{\tau} \chi_k^{[k-1]}({}_{k-1}h_k) .$$

c) $$\theta_{\tau+1}(h_{\tau+1}) = \theta_\tau(h_\tau) + X_\tau(h_\tau)\theta^{[\tau]}_{\tau+1}({}_\tau h_{\tau+1}) \ .$$

d) $$X_{\tau+1}(h_{\tau+1}) = X_\tau(h_\tau)X^{[\tau]}_{\tau+1}({}_\tau h_{\tau+1}) \ .$$

Now we can try to work out the conservingness and equalizingness conditions for the case of a recursive problem. Since conservingness is the simplest one, we will start with that condition.

Theorem 4.1. If the multi-stage stochastic programming problem is recursive, then a strategy s is conserving if and only if

(4.3) $$w_{[t]}(x_t) = \mathbb{E}_{h_t,s}[\theta^{[t]}_{t+1}({}_t h_{t+1}) + X^{[t]}_{t+1}({}_t h_{t+1})w_{[t+1]}(x_{t+1})] \text{ for all } t$$

and for $\mathbb{P}_s$-almost all h_t.
With this formulation of the conservingness condition we are back to the optimality principle.

Proof. Suppose s is conserving, then the second part of lemma 4.1 implies

(4.4) $$\theta_t(h_t) + X_t(h_t)w_{[t]}(x_t) = \mathbb{E}_{h_t,s}[\theta_{t+1}(h_{t+1}) + X_{t+1}(h_{t+1})w_{[t+1]}(x_{t+1})]$$

for $\mathbb{P}_s$-almost all h_t.
Lemma 4.2 then allows the following reformulation of the right hand side of this equation

$$\mathbb{E}_{h_t,s}\theta_t(h_t) + \mathbb{E}_{h_t,s}X_t(h_t)\{\theta^{[t]}_{t+1}({}_t h_{t+1}) + X^{[t]}_{t+1}({}_t h_{t+1})w_{[t+1]}(x_{t+1})\}.$$

Using this, we obtain (4.3) from (4.4). The reverse assertion is obtained by reversing all the arguments.

Actually, this theorem shows why some finite-stage problems can be solved numerically in an efficient way even if the number of stages is not very small. For problems which are not recursive one can expect numerical difficulties. In section 5 we will return to this aspect.
For the equalizingness condition we have a reformulation which only works if the problem is recursive and moreover tail vanishing.

Definition 4.3. The recursive multi-stage stochastic programming problem is called *tail vanishing* if for all strategies s

(4.5) $$\lim_{t\to\infty} \mathbb{E}_s X_t(h_t)v_{[t]}(x_t,s(h_t)) = 0 \ .$$

Theorem 4.2. If the multi-stage stochastic programming problem is recursive and tail vanishing, then a strategy s is equalizing if and only if

(4.6) $$\lim_{t\to\infty} \mathbb{E}_s X_t(h_t) w_{[t]}(x_t) = 0 .$$

Proof. Let s be equalizing, then (4.1) and (4.2) imply that (3.4) can be rewritten as

$$\lim_{t\to\infty} \mathbb{E}_s[\theta_t(h_t) + X_t(h_t) w_{[t]}(x_t) - \theta_t(h_t) - X_t(h_t) v_{[t]}(x_t, s(h_t))] = 0 .$$

Now (4.5) implies (4.6). The reverse is obtained by reversing the arguments.

5. Final remarks

The characterization of optimal strategies in section 3 shows that there is an essential difference between finite-stage and ∞-stage stochastic programming problems in this sense that for ∞-stage problems an extra condition is added in order to ensure optimality. However, even in ∞-stage problems this equalizing-condition may be redundant. This is for instance true, if there is some sort of strong fading in the decision process (e.g. discounting with bounded single-stage rewards). Also the opposite may be true, since in Markov decision processes the conserving-condition can very well be redundant (e.g. if single-stage rewards are averaged and all states remain attainable from any other state, namely, in that case all strategies are conserving).
Rockafellar and Wets use this conserving-condition in [11] to derive a dynamic programming formulation for a multi-stage stochastic program.
Grinold [4] presents an infinite-stage stochastic linear programming problem in which not all strategies are equalizing.

In section 4, it has been demonstrated how the conserving- and equalizing-condition may be simplified if there is some sort of extra structure in the problem. Regrettably, this recursive structure is usually not available in stochastic programming problems where the subsequent stages have a recourse function. However, Grinold's infinite-stage problem [4] is recursive and many other types of multi-stage stochastic and deterministic decision problems are recursive. The reason that the standard types of recourse problems are not recursive is that the set of allowed actions often depends explicitly on what happened before the current state was attained. For example, the multi-stage stochastic program as formulated by Rockafellar and Wets [11] is not recursive and hence the dynamic programming formulation is not of much numerical use. However, in many types of multi-stage stochastic programs the dependence of the allowed action set on the foregoing stages is of a very simple kind and then it is possible to reformulate the problem in such a way that it becomes recursive. This will be demonstrated in an example.

Example. In stage $n = 1,\dots,N$ a vector $y_n \in \mathbb{R}^m$, $y_n \geq 0$ has to be selected such that

$$A_1 y_1 + \dots + A_n y_n = \xi_n ,$$

where A_k are $m \times \ell$-matrices and $\xi_n \in \mathbb{R}^\ell$ is a random vector which is observed at stage n before the selection of y_n. The vectors ξ_n are independent and have given distributions. The object is to minimize

$$\mathbb{E}(c_1^T y_1 + \dots + c_N^T y_N) .$$

If one chooses in this example $\xi_n - A_1 y_1 - \dots - A_{n-1} y_{n-1} = \xi_n - \xi_{n-1}$ together with ξ_n as state of the system, then the problem becomes recursive with state space $\mathbb{R}^{2\ell}$.

The value of such a trick for numerical purposes should not be underestimated. It causes that the dimension of the dynamic programming problem resulting from the recursive form of the conserving condition does not increase with the number of stages. For this type of dynamic programming problems the numerical techniques have been improved considerably in recent years. However, the size of the problem quickly grows out of hand.

That not all multi-stage decision problems fit into the set-up of this paper may be demonstrated by referring to Evers' monograph [2]. Evers' ∞-stage linear programs, which are deterministic, fit into the set-up with the exception of his optimality criterion. As far as we see, his criterion cannot be written as an expected utility. However, his criterion is nearly equivalent with other criteria which make the problem recursive.

References

[1] L.E. Dubins, L.J. Savage, How to gamble if you must: inequalities for stochastic processes. New York, MacGraw-Hill, 1965.

[2] J.J.M. Evers, Linear programming over an infinite horizon. University of Tilburg Press, Tilburg 1973.

[3] N. Furukawa, S. Iwamoto, Markovian decision processes with recursive reward function. Bull. Math. Statist. 15 (1973), 79-91.

[4] R. Grinold, A new approach to multi-stage stochastic linear programs, pp. 19-29 in R.J.-B. Wets (ed.) Stochastic Systems II, Mathematical Programming Study no. 6, North-Holland Publ. Cy., Amsterdam 1976.

[5] L.P.J. Groenewegen, Characterization of optimal strategies in dynamic games. Mathematical Centre Tract no. 90, Mathematical Centre Amsterdam 1979 (forthcoming).

[6] L.P.J. Groenewegen, J. Wessels, On equilibrium strategies in noncooperative dynamic games. To appear in O. Moeschlin, D. Pallaschke (eds.), Game Theory and Related Topics, North-Holland, Amsterdam 1979.

[7] K. Hinderer, Foundations of non-stationary dynamic programming with discrete time parameter. Berlin, Springer 1970 (Lecture Notes in Oper. Res. & Math. Econ. no. 33).

[8] A. Hordijk, Dynamic programming and Markov potential theory. Mathematical Centre Tract no. 51, Mathematical Centre Amsterdam 1974.

[9] R.P. Kertz, D.C. Nachman, Persistently optimal plans for non-stationary dynamic programming: the topology of weak convergence case. Annals of Probability (to appear).

[10] J. Neveu, Mathematical foundations of the calculus of probability. San Francisco, Holden-Day 1965.

[11] R. Rockafellar, R.J.-B. Wets, Nonanticipativity and $\mathcal{L}'$-martingales in stochastic optimization problems. pp. 170-187 in the same volume as [4].

[12] W.D. Sudderth, On the Dubins and Savage characterization of optimal strategies. Ann. Math. Statist. 43 (1972), 498-507.