Discussiones Mathematicae
Probability and Statistics 23 (2003) 5–44

STOCHASTIC DYNAMIC PROGRAMMING WITH RANDOM DISTURBANCES

Regina Hildenbrandt

Technical University Ilmenau, Institute of Mathematics
PSF 10 05 65, D–98684 Ilmenau, Germany

Abstract

Several peculiarities of stochastic dynamic programming problems where random vectors are observed before the decision is made at each stage are discussed in the first part of this paper. Surrogate problems are given for such problems with distance properties (for instance, transportation problems) in the second part.

Keywords: Stochastic dynamic programming, Markov decisions process, dominant policy, distance properties, (partial) certainty equivalence principle.

2000 Mathematics Subject Classification: 90C39, 90C40.

1. Introduction

It is assumed for a lot of concepts in the theory of stochastic dynamic programming that random disturbances are observed after the decision is made at each stage (for instance, compare Bertsekas [2], Schneeweiss [16], Dinkelbach [4]). Problems for which this is assumed we denote as DB models. (DB stands for "decision before".) Problems where the random disturbances are observed before the decision is made at each stage we call DA models. (DA stands for "decision after".) We began to study DA models by a stochastic dynamic transportation problem (compare [8], [9]). Not very much can be found about problems of DA models. Something is included in the book by Sebastian and Sieber, [17]. There the situations of incomplete information are described by means of operators as a starting point for further investigations (compare [17], 2.7 with $n = 1$). Dreyfus and Law give an example in relation to certainty equivalence and an example of a stochastic equipment inspection and replacement model, where some components of the random vector are observed after the decision is made (in a usual way) but some components are observed before (compare [5], p. 189 resp. 137). On the one hand, DA models belong to the extensive group of stochastic dynamic programming problems, but on the other hand DA models show peculiarities. The complexity of such problems (compare the inspection/replacement problem by Dreyfus and Law) is one motivation for further considerations of DA models. We work out several special qualities of DA models in the first part of the present paper. (It is not complicated to combine DA with DB models.)

A "certainty equivalence principle" is formulated and proved in some cases of DA models with linear dynamics and quadratic criteria in Section 2.

In Section 3, Markov decision processes, which result from DA models under appropriate assumptions are investigated. There, the corresponding decisions are characterized by a "simple" structure. The transition probability matrices differ only by two elements from the corresponding "neighbouring" decisions.

In the second part of the paper we consider Markov decision processes with "distance properties" (in a natural way such properties are found in flow problems), which result from DA models. We give surrogate problems for the calculation of approximate solutions. These surrogate problems are based on the distance properties and the DA models.

At the end of the paper (Section 5), the theoretical investigations of the preceding sections are applied to the above-mentioned stochastic dynamic transportation problem.

1.1. The DA model

In the following we use

$N \in \mathbb{N} \cup \{\infty\}$	the horizon
$t \in \{1, 2, ..., N\}$	numbers of stages
S	state space
$s \in S$	states
B	disturbance space
$w \in B$	random disturbances
A	decision space
$x \in A$	decisions (resp. controls)

(The questions of measurability are mostly omitted. At the beginning let S and A be Borel spaces and let the values of w be elements of a Borel space. Later on we often assume $S \subseteq \mathbb{Z}^n$ (or $\mathbb{R}^n$) and so on. We use the same notations for the random vectors and their realizations.)

The above data are written with inferior indices t in order to attach the date to the stages t.

Further on

$K_t : S_t \times B_t \times A_t \to \mathbb{R}_+$	stage - cost (respectively - return) functions
$G_t : S_t \times B_t \times A_t \to S_{t+1}$	transition function

denote (measurable) functions.

Decision spaces A_t can depend on previous states and disturbances.

Now, we introduce the basic problem of a DA model:

Let DA models be closed-loop optimization problems (i. e. feedback control, cf. [2], I, p.4 or [12], 2.4): More precisely that means that we postpone making the decision x_t until the last possible moment (time t) when the current state s_t and (in the case of a DA model) the realization of the random vector w_t will be known. We assume that an initial state $s_1 \in S_1$ and an initial realization w_1 of the random disturbances are given.

A policy

$$F = \{x_1(s_1, w_1), x_2(s_2, w_2), \ldots, x_N(s_N, w_N)\}$$

is to be found so that

$$\underset{w_2,\ldots,w_N}{E} \left(\sum_{t=1}^{N} K_t(s_t, w_t, x_t) | s_1, w_1 \right)$$

$$= K_1(s_1, w_1, x_1) + \underset{w_2,\ldots,w_N}{E} \left(\sum_{t=2}^{N} K_t(s_t, w_t, x_t) | w_1, s_2 \right) \rightarrow min$$

subject to the constraints

$$s_t \in S_t, t = 2, \cdots, N,$$

$$x_t \in A_t(s_t, w_t), t = 1, \cdots, N,$$

(dependences $A_t(\overline{s_t}, w_t)$ with $\overline{s_t} = \{s_1, \ldots, s_t\}$ are thinkable, too)

$$s_{t+1} = G_t(s_t, w_t, x_t), t = 1, \ldots, N-1 \text{ (dynamic constraints).}$$

(The objective function always exists as $K_t \geq 0$, but it may have the value ∞ without any additional assumptions.) We assume that the distribution functions and the densities of the sequence of disturbances $\{w_t : t = 1, \ldots, N\}$ are known and that all (following) conditional expected values exist.

Remarks. The dependence of A_t on w_t is a peculiarity of DA models. In DA models more information is known before the decisions are made at each stage as in the usual DB models, namely $x_t \in A_t(s_t, \mathbf{w_t})$.

a) b)

Feedback control
DA models

Feedback control
DB models
(with analogous symbols)

Figure 1

Of course, DA models are stochastic dynamic programming problems, too. When a decision x_t is made, then the realizations $w_{t+1}, w_{t+2}, \cdots$ of the disturbances at the next stages are not known. And the cost of the next stages depends on $s_{t+1} = G_t(s_t, w_t, x_t)$, too.

The optimal value function for the remaining periods and the functional

Further on, we use $F_t = \{x_t(s_t, w_t), x_2(s_2, w_2), \ldots, x_N(s_N, w_N)\}, t = 1, ..., N$ for any admissible policy F and the symbol $\overline{w_t} := (s_1, w_1, \ldots, w_t)$ (an admissible policy $F = \{x_1(s_1, w_1), x_2(s_2, w_2), \ldots, x_N(s_N, w_N)\}$ means $x_{t'} \in A_{t'}(s_{t'}, w_{t'}) \forall s_{t'} \in S_{t'}, \forall t' = 1, ..., N$).

The optimal value function for the remaining periods $t, \ldots, N$ is

$$
\begin{aligned}
f_t(s_t\ , \overline{w_t}) &= \min_{F_t} \quad \underset{w_{t+1},\ldots,w_N}{E} \quad \left(\sum_{t'=t}^{N} K_{t'}(s_{t'}, w_{t'}, x_{t'}) | \overline{w_t} \right) \\
(1) \qquad &= \min_{F_t} \left(K_t(s_t, w_t, x_t) + \underset{w_{t+1},\ldots,w_N}{E} \left(\sum_{t'=t+1}^{N} K_{t'}(s_{t'}, w_{t'}, x_{t'}) | \overline{w_t} \right) \right)
\end{aligned}
$$

for $t = 1, ..., N-1$

and

$$f_N(s_N, \overline{w_N}) = \min_{F_N} K_N(s_N, w_N, x_N)$$

for DA models.

We define

$$f_{N+1} \equiv 0. \tag{2}$$

The functional equation

$$f_t(s_t, \overline{w_t}) = \min_{x_t \in A_t(s_t, w_t)} \Big(K_t(s_t, w_t, x_t) + \underset{w_{t+1}}{E} \left(f_{t+1}(s_{t+1}, \overline{w_{t+1}}) | \overline{w_t} \right) \Big) \tag{3}$$

$t = N, \ldots, 1$

follows.

If an optimal policy exists, the functional equation can be proved directly by means of mathematical induction (compare Sebastian and Sieber [17], the general formula (2.188) and the upper remarks on page 147.):

Proof.

Compare $f_N(s_N, \overline{w_N}) := \min_{F_N} K_N(s_N, w_n, x_N)$ for $t = N$.

Step 1.

(begin with the mathematical induction $t = N - 1$)

$$f_{N-1}(s_{N-1}, \overline{w_{N-1}})$$

$$:= \min_{F_{N-1}} \Big(K_{N-1}\left(s_{N-1}, w_{N-1}, x_{N-1}\right) + \underset{w_N}{E} \left(K_N(s_N, w_N, x_N) \mid \overline{w_{N-1}} \right) \Big)$$

(Compare (1) for $t = N - 1$.)

$$= \min_{\substack{x_{N-1} \in A_{N-1}(s_{N-1}, w_{N-1}) \\ x_N \in A_N(s_N, w_N)}} \Big(K_{N-1}(s_{N-1}, w_{N-1}, x_{N-1}) +$$

$$\underset{w_N}{E} \left(K_N(s_N, w_N, x_N) \mid \overline{w_{N-1}} \right) \Big)$$

$$= \min_{x_{N-1}\in A_{N-1}(s_{N-1},w_{N-1})} \Big\{ K_{N-1}(s_{N-1},w_{N-1},x_{N-1}) +$$

$$\min_{x_N\in A_N(s_N,w_N)} \Big(\underset{w_N}{E} \left(K_N(s_N,w_N,x_N) \mid \overline{w_{N-1}} \right) \Big) \Big\}.$$

(There $\min_{x_N\in A_N(s_N,w_N)} \ldots$ means in detail $\min_{\substack{x_N(w_N)\in A_N(s_N,w_N)\\ \forall\ w_N \in B_N}} \ldots$.)

Now, we use the relation $\min_x E\{\phi(x)\} = E\left\{\min_x \phi(x)\right\}$.

$$= \min_{x_{N-1}\in A_{N-1}(s_{N-1},w_{N-1})} \Big\{ K_{N-1}(s_{N-1},w_{N-1},x_{N-1}) +$$

$$\underset{w_N}{E} \Big(\min_{x_N\in A_N(s_N,w_N)} K_N(s_N,w_N,x_N) \mid \overline{w_{N-1}} \Big) \Big\}$$

$$= \min_{x_{N-1}\in A_{N-1}(s_{N-1},w_{N-1})} \Big(K_{N-1}(s_{N-1},w_{N-1},x_{N-1}) + \underset{w_N}{E} \left(f_N(s_N,\overline{w_N}) \mid \overline{w_{N-1}} \right) \Big).$$

Step $N - t^*$:

Now, let us assume

$$(*) \qquad f_t(s_t,\overline{w_t}) = \min_{x_t\in A(s_t,w_t)} \Big(K_t(s_t,w_t,x_t) + \underset{w_{t+1}}{E} \left(f_{t+1}(s_{t+1},\overline{w_{t+1}}) \mid \overline{w_t} \right) \Big)$$

for $t = N, N-1, \ldots, t^*+1$ $(t^*+1 > 1)$.

We will prove the functional equation for $t = t^*$:

$$f_{t^*}(s_{t^*},\overline{w_{t^*}}) := \min_{F_{t^*}} \Big(K_{t^*}(s_{t^*},w_{t^*},x_{t^*}) +$$

$$\underset{w_{t^*+1},\ldots,w_N}{E} \Big(\sum_{t'=t^*+1}^{N} K_{t'}(s_{t'},w_{t'},x_{t'}) \mid \overline{w_{t^*}} \Big) \Big)$$

(compare (1))

$$= \min_{\substack{x_{t^*} \in A_{t^*}(s_{t^*}, w_{t^*}) \\ \vdots \\ x_N \in A_N(s_N, w_N)}} \left(K_{t^*}(s_{t^*}, w_{t^*}, x_{t^*}) + \sum_{t'=t^*+1}^{N} \underset{w_{t'}, \ldots, w_N}{E} (K_{t'}(s_{t'}, w_{t'}, x_{t'}) \mid \overline{w_{t^*}}) \right)$$

$$= \min_{\substack{x_{t^*} \in A_{t^*}(s_{t^*}, w_{t^*}) \\ \vdots \\ x_N \in A_N(s_N, w_N)}} \Big\{ K_{t^*}(s_{t^*}, w_{t^*}, x_{t^*}) + \underset{w_{t^*+1}, \ldots, w_N}{E} (K_{t^*+1}(s_{t^*+1}, w_{t^*+1}, x_{t^*+1})$$

$$+ \underset{w_{t^*+2}, \ldots, w_N}{E} (K_{t^*+2}(s_{t^*+2}, w_{t^*+2}, x_{t^*+2})$$

$$+ \cdots + \underset{w_N}{E} (K_N(s_N, w_N, x_N) \mid \overline{w_{N-1}}) \mid \ \ldots \ \mid \overline{w_{t^*+1}}) \mid \overline{w_{t^*}}) \Big\}$$

$$= \min_{x_{t^*} \in A_{t^*}(s_{t^*}, w_{t^*})} \Big\{ K_{t^*}(s_{t^*}, w_{t^*}, x_{t^*}) +$$

$$\underset{w_{t^*+1}, \ldots, w_N}{E} \Big(\min_{x_{t^*+1} \in A_{t^*+1}(s_{t^*+1}, w_{t^*+1})} \Big(K_{t^*+1}(s_{t^*+1}, w_{t^*+1}, x_{t^*+1})$$

$$+ \cdots + \underset{w_N}{E} \Big(\min_{x_N \in A_N(s_N, w_N)} K_N(s_N, w_N) \mid \overline{w_{N-1}} \Big) \mid \ \ldots \Big) \mid \overline{w_{t^*}} \Big) \Big\} .$$

Now, we use (*) for $t = N, N-1, \ldots, t^* + 1$.

$$= \min_{x_{t^*} \in A_{t^*}(s_{t^*}, w_{t^*})} \Big(K_{t^*}(s_{t^*}, w_{t^*}, x_{t^*}) + \underset{w_{t^*+1}}{E} (f_{t^*+1}(s_{t^*+1}, \overline{w_{t^*+1}}) \mid \overline{w_{t^*}}) \Big).$$

The ,,DA decision functions" and some other definitions (which are based on DA models)

In DA models the state s_{t+1} is (for a given s_t, w_t) completely determined by the decision (as opposed to DB models). Thus we can introduce:
The DA decision sets

$$\hat{A}_t(s_t, w_t) := \{ s' \mid s' = G_t(s_t, w_t, x_t) \ \textit{with} \ x_t \in A_t(s_t, w_t) \} \tag{4}$$

for a given $s_t \in S_t$, $w_t \in B_t$, internal cost

$$\begin{aligned}(5)\qquad &\hat{c}(s_t, w_t, s') := min\left\{K_t(s_t, w_t, x_t)|x_t : s' = G(s_t, w_t, x_t)\right\}\\ &\text{with } s' \in \hat{A}_t(s_t, w_t)\end{aligned}$$

and DA decision functions

$$\begin{aligned}(6)\qquad &\hat{d}_t :\ S_t \times B_t \rightarrow S_{t+1}\\ &\text{with}\quad \hat{d}_t(s_t, w_t) = s' \in \hat{A}_t(s_t\ , w_t).\end{aligned}$$

Finally, we use

Definition 1. The set of DA decision functions is the set

$\hat{D}_t := \{\hat{d}_t|\ \hat{d}_t :\ S_t \times B_t \rightarrow S_{t+1} \text{ with } \hat{d}_t(s_t, w_t) \in \hat{A}_t(s_t\ , w_t)\}$

for a given S_t, B_t, S_{t+1} and DA decision sets $\hat{A}_t$.

Separate maps

$$(7)\qquad \left.\begin{aligned}&(s_t, w_t) \rightarrow s'\ (by\ \hat{d})\\ &\text{thatmeans } \hat{d}_t(s_t, w_t) = s'\end{aligned}\right\}$$

for $s_t\ \in S_t\ , w_t \in B_t$ are called separate decision.

If S_t and B_t are finite sets, then $\hat{d}_t$ will include $|S_t|\ \cdot\ |B_t|$ separate decisions (where $|S_t|$ resp.$|B_t|$ denote the numbers of elements of the sets S_t resp. B_t).

In this way Figure 1 a) can be replaced with

2. The certainty equivalence principle

For a lot of DB models with quadratic cost functionals and linear dynamics (so called quadratic linear problems) it is possible to replace the random disturbances with their expected values and to solve the yielded deterministic problems. The solutions are the same (certainty equivalence principle). We can use a similar method for DA models.

At first, let us consider the following example.

Example 1. We consider the stochastic dynamic programming problems

$$E\left(\sum_{t=1}^{N=3}\left((x_t)^2+(s_t)^2\right)\right)\to min,$$

where s_1/ resp. s_1 and w_1 are given

and $s_{t+1}=s_t+w_t+x_t,$

$x_t\in\mathbb{R}.$

(There $\{w_t\}_{t=1,2,3}$ is a sequence of independent random disturbances.)

Since the decision spaces $(A_t(s_t,w_t)=)\mathbb{R}$ (at each stage) are independent of w_t, we can classifiy such stochastic dynamic programming problems as DA models or as DB models (with the same data, but $x_t(s_t,\mathbf{w_t})$ for DA models and $x_t(s_t)$ for DB models).

The optimal solution to the problem of the DB model is

$x_N=x_3=0$

$x_{N-1}=x_2=\frac{-s_2-E(w_2)}{2}$

$x_{N-2}=x_1=\frac{-3s_1-E(w_2)-3E(w_1)}{5}$

(We can calculate this by means of the Bellman-principle or the certainty equivalence principle.)

The optimal solution of the DA model is

$x_N = x_3 = 0,$

$x_{N-1} = x_2 = \frac{-s_2 - w_2}{2},$

$x_{N-2} = x_1 = \frac{-3s_1 - E(w_2) - 3w_1}{5}$

(at the beginning we have calculated this by means of the Bellmann-principle, compare (3)).

Obviously, the minimal expected costs for the DA model are not greater than these costs for the DB model since every policy of the DB model is possible for the DA model, too ($A_t(s_t, w_t)$ are independent of w_t).

Example 1 shows us the relationship between the solutions of the DB model and the DA model, respectively.

Now, we want to generalize the results of the example.

Quadratic-linear-problems

Let us assume

$$S_t = \mathbb{R}^N, \ t = 1, \ldots, N,$$

$$A_t = \mathbb{R}^q, \ t = 1, \ldots, N.$$

The dynamic constraints are

$$s_{t+1} = \Phi_t s_t + \Gamma_t x_t + \Pi_t w_t \ \ for \ \ t = 1, \ldots, N^1 \tag{8}$$

with given matrices Φ_t, Γ_t and Π_t and a given s_1 and moreover with given w_1 for DA model (the symbols are taken from the pattern of Schneeweiss [16], 11.3).

The types of these matrices are determined by the types of the states, disturbances and decisions.

If $z_t = \binom{w_t}{1}, v_t = \binom{s_t}{z_t}, y_t = \binom{x_t}{v_t}$ and $T_t = (\Gamma_t, \Phi_t, \Pi_t, 0)$

are used, then (8) has got the form

$s_{t+1} = T_t y_t.$

Finally, the cost functional is

$$E\left\{\sum_{t=1}^{N} y_t^T W_{t,yy} y_t\right\} \to min,$$

where the matrices $W_{t,yy}$ have the following structure

$$W_{t,yy} = \begin{pmatrix} W_{t,xx} & W_{t,xv} \\ W_{t,vx} & W_{t,vv} \end{pmatrix} = \begin{pmatrix} W_{t,xx} & W_{t,xs} & W_{t,xz} \\ W_{t,sx} & W_{t,ss} & W_{t,sz} \\ W_{t,zx} & W_{t,zs} & W_{t,zz} \end{pmatrix} =$$

$$= \begin{pmatrix} W_{t,xx} & W_{t,xs} & W_{t,xw} & W_{t,x1} \\ W_{t,sx} & W_{t,ss} & W_{t,sw} & W_{t,s1} \\ W_{t,wx} & W_{t,ws} & W_{t,ww} & W_{t,w1} \\ W_{t,1x} & W_{t,1s} & W_{t,1w} & W_{t,11} \end{pmatrix}$$

with regard to v_t, s_t and y_t.

Let $W_{t,yy}$ be symmetric matrices (without loss of generality) and let $W_{t,xx}$ be positive definite. Further on , let all matrices V_{xx} be positive definite which are calculated by means of the backward dynamic programming procedure.

Quadratic-linear-problems can be classified as DA models or as DB models with the same data, but $x_t(s_t)$ for DB models and $x_t(s_t, \mathbf{w_t})$ for DA models (compare Example 1).

Theorem 1 (Certainty equivalence principle). *Let a quadratic-linear DB model and a quadratic-linear DA model with the same data be given.*

Further on let

$$x_N = 0$$

$$x_t = \varphi(E(w_t), E(w_{t+1}), \cdots, E(w_{N-1}) \quad), t = N-1, \cdots, 1$$

be a representation of an optimal solution of the quadratic-linear DB model.

Then

$$x_N = 0$$

$$x_t = \varphi(w_t, E(w_{t+1}), \cdots, E(w_{N-1})), t = N-1, \cdots, 1$$

is an optimal solution of the quadratic-linear DA model.

Proof. The above symbols and the following representations are taken from the pattern of Schneeweiss [19] (compare 11.3) and they are applied to the DA models.

The functional equation for this DA problem is

$$(*1) \qquad \begin{aligned} f_t(s_t, \overline{w_t}) = \min_{x_t} \Big\{ y_t^T W_{t,yy} y_t + \underset{w_{t+1}}{E} \Big\{ f_{t+1}(s_t, \overline{w_{t+1}}) | \overline{w_t} \Big\} \Big\} \\ t = N, \cdots, 1 \\ f_{N+1} \equiv 0 \end{aligned}$$

(compare(4)).

Step 1.

(begin with the mathematical induction $t = N$)

$$\begin{aligned} f_N(s_N, w_N) &= \min_{x_N} \left(y_N^T W_{N,yy} y_N \right) \\ (*2) \qquad &= \min_{x_N} \left(x_N^T W_{N,xx} x_N + 2 x_N^T W_{N,xv} v_N + v_N^T W_{N,vv} v_N \right). \end{aligned}$$

$$(*3) \qquad x_N^* = -(W_{N,xx})^{-1} W_{N,xv} v_N$$

$$(*3a) \qquad = -\left(W_{N,xx}\right)^{-1} \left(W_{N,xs} s_N + W_{N,xw} w_N + W_{N,x1}\right)$$

is the optimal x_N for (*2), since $W_{N,xx}$ is positive definite.

If we use (*3) in (*2), it follows

$$
\begin{aligned}
& f_N(s_N, \overline{w_N}) \\
& = -v_N^T W_{N,xv}^T (W_{N,xx})^{-1} W_{N,xv} v_N + v_N^T W_{N,vv}, v_N \\
& = s_N^T Q_N s_N + 2 s_N^T \bar{\beta}_N + GA_N(w_N)
\end{aligned}
$$

with

$$Q_N = W_{N,ss} - W_{N,xs}^T (W_{N,xx})^{-1} W_{N,xs},$$

$$\tilde{\beta}_N = (W_{N,sz} - W_{N,xs}^T (W_{N,xx})^{-1} W_{N,xz}) z_N,$$

$$GA_N(w_N) = \tilde{\gamma}_N = z_N^T W_{N,zz} z_N - z_N^T W_{N,xz}^T (W_{N,xx})^{-1} W_{N,xz} z_N.$$

Further on, let β_N and γ_N denote as follows:

$$
\begin{aligned}
\beta_N &= E\{\tilde{\beta}_N | \overline{w_{N-1}}\} \\
&= (W_{N,sz} - W_{N,xs}^T (W_{N,xx})^{-1} W_{N,xz}) E\{z_N | \overline{w_{N-1}}\}, \\
\gamma_N &= E\{z_N^T W_{N,zz} z_N | \overline{w_{N-1}}\} - \hat{z}_N^T W_{N,xz}^T (W_{N,xx})^{-1} W_{N,xz} \hat{z}_N, \\
\hat{z}_N &= E\{z_N | \overline{w_{N-1}}\}.
\end{aligned}
$$

Step $N - t + 2$:
Now, let us assume

$$(*4) \qquad f_t(s_t, \overline{w_t}) = s^T Q_t s_{t-1} + 2 s_t^T \tilde{\beta}_t + GA_t(w).$$

On the one hand, we will prove

$$(*5) \qquad f_{t-1}(s_{t-1}, \overline{w_{t-1}}) = s_{t-1}^T Q_{t-1} s_{t-1} + 2 s_{t-1}^T \tilde{\beta}_{t-1} + GA_{t-1}(w)$$

for the optimal expected value function at stage $t-1$, where

$$Q_t := V_{t,ss} V_{t,xs}^T (V_{t,xx})^{-1} V_{t,xs},$$

$$\tilde{\beta}_t := (V_{t,sz} - V_{t,xs}^T (V_{t,xx})^{-1} V_{t,xs}) z_t.$$

Further on, let $\beta_t, \tilde{\gamma}_t$ and γ_t denote as follows:

$$\beta_t = E\{\tilde{\beta}_t | \overline{w_{t-1}}\},$$

$$\tilde{\gamma}_t := z_t^T V_{t,zz} z_t - z_t^T V_{t,xz}^T (V_{t,xx})^{-1} V_{t,xz} z_t,$$

$$\gamma_t := E\{z_t^T V_{t,zz} z_t | \overline{w_{t-1}}\} - \hat{z}_t^T V_{t,xz}^T (V_{t,xx})^{-1} V_{t,xz} \hat{z}_t,\ \hat{z}_t = E\{z_t | \overline{w_{t-1}}\}.$$

There the sub-matrices $V_{t,ij} (i,j = x, v, z, w, s, 1)$ are calculated from

$$y_t^T V_{t,yy} y_t = y_t^T W_{t,yy} y_t + y_t^T T_t^T Q_{t+1} T_t y_t + 2 y_t^T \beta_{t+1} + \gamma_{t+1},$$

where the initials are

$$Q_{N+1} = 0, \tilde{\beta}_{N+1} = 0, \gamma_{N+1} = 0, \tilde{\gamma}_{N+1} = 0.$$

On the other hand, we will show that

$$\begin{aligned} x_{t-1}^* &= - (V_{t-1,xx})^{-1} V_{t-1,xv} v_{t-1} \\ &= - (V_{t-1,xx})^{-1} (V_{t-1,xs} s_{t-1} + V_{t-1,xz} z_{t-1}) \end{aligned}$$

is an optimal decision at stage $t-1$.

Proof.
$(*1)$ for $t-1$ and (*4) yield

$$
\begin{aligned}
&f_{t-1}(s_{t-1}, w_{t-1})\\
&= \min_{x_{t-1}} \{y_{t-1}^T\, W_{t-1,yy} y_{t-1} + E\{f_t(T_{t-1}y_{t-1}, \overline{w_t}) | \overline{w_{t-1}}\}\}\\
&= \min_{x_{t-1}} \{y_{t-1}^T\, W_{t-1,yy}\, y_{t-1} +\\
&\qquad E\{y_{t-1}^T\, T_{t-1}^T\, Q_t\, T_{t-1}\, y_{t-1} \;+\; 2y_{t-1}^T\, T_{t-1}^T\, \tilde{\beta}_t + GA_t(w) \mid \overline{w_{t-1}}\}\}\\
&= \min_{x_{t-1}} \{y_{t-1}^T\, W_{t-1,yy}\, y_{t-1} \;+ y_{t-1}^T T_{t-1}^T\, Q_t\, T_{t-1}\, y_{t-1} \;+\; 2y_{t-1}^T\, T_{t-1}\, \beta_t\\
&\qquad + E\{GA_t(w) \mid \overline{w_{t-1}}\}\}\\
&\quad = \min_{x_{t-1}} \{y_{t-1}^T\, V_{t-1,yy}\, y_{t-1} \;-\; \gamma_t + E\{GA_t\,(w) \mid \overline{w_{t-1}}\}\}
\end{aligned}
$$

$$
\begin{aligned}
(*6)\qquad &= \min_{x_{t-1}} \{x_{t-1}^T\, V_{t-1,xx}\, x_{t-1} + 2x_{t-1}^T\, V_{t-1,xv}\, v_{t-1} + v_{t-1}\, V_{t-1,vv}\, v_{t-1}\\
&\qquad -\gamma_t + E\{GA_t(w) \mid \overline{w_{t-1}}\}\}.
\end{aligned}
$$

$$
\begin{aligned}
&x_{t-1}^* = -(V_{t-1,xx})^{-1} V_{t-1,xv}\, v_{t-1}\\
(*7)\qquad &\quad = -(V_{t-1,xx})^{-1}(V_{t-1,xs} z_{t-1} + V_{t-1,xs}\, s_{t-1})
\end{aligned}
$$

follows for positive definite $V_{t-1,xx}$.

(*7) placed in (*6) implies

$$
\begin{aligned}
&f_{t-1}(s_{t-1}, w_{t-1})\\
&= \;-\, v_{t-1}\, V_{t-1,xv}^T\, (V_{t-1,xx})^{-1}\, V_{t-1,xv}\, v_{t-1} \;+\; v_{t-1}^T\, V_{t-1,vv}\, v_{t-1}\\
&\qquad -\, \gamma_t \;+\; E\{GA_t(w) \mid \overline{w_{t-1}}\}
\end{aligned}
$$

$$= s_{t-1}^T\, Q_{t-1}\, s_{t-1} + 2 s_{t-1}^T\, \tilde{\beta}_{t-1} + \gamma_{t-1} - \gamma_t + E\{GA_t(w) \mid \overline{w_{t-1}}\}$$

$$= s_{t-1}^T\, Q_{t-1}\, s_{t-1} + 2 s_{t-1}^T\, \tilde{\beta}_{t-1} + GA_{t-1}(w).$$

Now, we compare the optimal decisions (*7) of the quadratic-linear DA models with the optimal decisions of the quadratic-linear DB models, see Schneeweiss [19], 11.3.

The above matrices $V_{t,yy}$ are the same as the corresponding matrices of Schneeweiss.

Then x_t^* corresponds to u_k^* of Schneeweiss except for w_t in v_t. In Schneeweiss we find there $E\{w_t \mid \overline{w_{t-1}}\}(\hat{=}\hat{r}^k)$ (compare [19], pages 162/163).

With that, the Theorem 1 is proved.

Remarks.

- Above we use the denotation "certainty equivalence principle" (for DA models) because of the relationships between the solutions of DB model and DA model, respectively.
- Dreyfus and Law have another opinion about such a conception (compare [5], pages 275, 276). But the remark on page 276 is very short. The calculations and considerations are not sufficiently given in detail.
- Sebastian and Sieber, [17], deal with quadratic-linear-problems, too (see 2.8.3.3). But the disturbances do not take place in the cost functional and an interpretation of the calculations is not given.

3. DA models as Markov decision processes under appropriate assumptions

In what follows, we assume an infinite horizon, that means $N = \infty$. The average expected cost per stage will be minimized. Further on, we demand stationary properties. The functions and sets K_t, G_t, A_t, B_t, S_t are the same at each stage. We write K, G and so on. Let B, S and A be finite sets. $q(\omega)(q : B \to (0,1))$ denote the probabilities of the random disturbances. These $q(\cdot)$ are the same at every stage, too. Finally, we assume that the components w_i of w are realized independent of each other.

Then the problem

$$\overline{\lim_{n\to\infty}}\frac{1}{n}E\left(\sum_{t=1}^{n}K(s_t,w_t,x_t)\right)\to inf,$$

where s_1 and w_1 are given

$$\text{and}\quad \left.\begin{array}{l} s_{t+1}=G(s_t,w_t,x_t) \\ s_{t+1}\in S \end{array}\right\} t=1,2\cdots$$

$$x_t\in A(s_t,w_t), t=1,2,\cdots$$

remains to be solved.

We want to represent this problem as a Markov decision process. We can do that in two ways. First, we could use section 1.1 in link with Girlich, Köchel and Kuenle [6], page 36, respectively Neumann and Morlock [15], page 618. A state space $S\times B$ would follow. But here we will directly convert the above problem into a Markov decision process. On the one hand, then the corresponding state space is at once S and on the other hand, peculiarities of DA models can be better characterized (compare 3.1). (The disturbance space B and the probability q merely serve to calculate the transition probabilities of the Markov decision process.)

Now, the Markov decision processes are introduced by means of (stationary) Markov control models $MCM:(N,S,A^M,p,\gamma)$ with $N=\infty$, the state spaces S,A^M the sets of action spaces $A^M(S)$, the transition probabilities p and the average (one-step) reward functions γ (compare [7], 1.2). $A^M(s),p$ and γ are constructed by means of $\hat{d}$ and $\hat{c}$ (compare (4), (5) and (6) and Definition 1) in the following way:

(9) $A^M(s)=\{d(s):=\{\hat{d}(s,w^1)\ ,\ \hat{d}(s,w^2),\cdots,\hat{d}(s,w^{IBI})\}|\hat{d}\in\hat{D}\},s\in S$

(where $|B|$ denotes the number of elements of the set B)

(10) $$p(s'|s,d)=\sum_{w:s'=\hat{d}(s,w)}q(w)$$

$$\gamma(s,d) = \sum_{s'} \sum_{w:s'=\hat{d}(s,w)} \hat{c}(s,w,s')q(w)$$

$$= \sum_{s'} \left(\sum_{w:s'=\hat{d}(s,w)} \hat{c}(s,w,s') \frac{q(w)}{p(s'|s,d)} \right) p(s'|s,d). \tag{11}$$

We define the cost

$$c^d(s,s') = \sum_{w:s'=\hat{d}(s,w)} \hat{c}(s,w,s') \frac{q(w)}{p(s'|s,d)}. \tag{12}$$

The relation

$$\gamma(s,d) = \sum_{s'} c^d(s,s')p(s'|s,d) \tag{13}$$

follows.

Now, let us note $S = \{s^1, \cdots, s^m\}$ and $p^d_{fl} = p(s^l|s^f, d)$.

Under the assumption that the stationary distributions

$(p^{d,\infty}_f)_{f=1,\dots,m}$ with

$$\lim_{t\to\infty} \left(\ (p^d_{fl})_{\binom{f=1,\dots,m}{l=1,\dots,m}} \ \right)^t = \begin{pmatrix} p^{d,\infty}_1 & \dots & p^{d,\infty}_m \\ \vdots & & \vdots \\ p^{d,\infty}_1 & \dots & p^{d,\infty}_m \end{pmatrix}$$

exist, an optimal policy d is to be found so that

$$dk = \gamma(s^1,d)p^{d,\infty}_1 + \cdots + \gamma(s^m,d)p^{d,\infty}_m \to min. \tag{14}$$

Now, we represent two special cases for cost ((a) is valid for a stochastic dynamics transportation problem, compare [9])

(a) $\hat{c}(s,\omega,s')$ do not depend on ω.
That means

$$\hat{c}(s,\omega,s') = \hat{c}(s,s') \text{ for each w with } s' \in \hat{A}(s,w). \tag{15}$$

Then (20) yields

$$c^d(s,s') = \hat{c}(s,s') \left(\sum_{w:s'=\hat{d}(s,\omega)} \frac{q(\omega)}{p(s'|s,d))} \right) = \hat{c}(s,s').$$

(b) $\hat{c}(s,\omega,s')$ do not depend on s'.

That means

$$\hat{c}(s,\omega,s') = \hat{c}(s,\omega) \text{ for each } s' \in \hat{A}(s,w). \tag{16}$$

Then

$$\gamma(s,d) = \sum_{\omega} \hat{c}(s,w)q(\omega) \text{ do not depend on } d. \tag{17}$$

3.1. The structure of decisions within DA models

Definition 2.
$d^1 \in A^M, d^2 \in A^M$ will be called neighbouring, if a unique $s^0 \in S$ and a unique $w^0 \in B$ exist with

$d^1(s) \equiv d^2(s)$ for each $s \in S$ and $s \neq s^0$,

$\hat{d}^1(s^0,w) = \hat{d}^2(s^0,w)$ for each $w \in B$ and $w \neq w^0$,

$\hat{d}^1(s^0,w^0) \neq \hat{d}^2(s^0,w^0)$.

(That means d^1 and d^2 are only different in one separate decision (compare (7))).

Theorem 2.
I *Let $d \in A^M$ and $\tilde{d} \in A^M$. Then a sequence $d = d^0, d^1, d^2, \cdots, d^v = \tilde{d}$ of neighbouring decisions d^i, d^{i+1} $(0 \leq i \leq v-1)$ exists with $d^i \in A^M$.*

II *Now, let $d \in A^M, \bar{d} \in A^M$ be neighbouring with the different separate decisions:*

$$\hat{d}\left(s^f, w\right) = s^l, \qquad \hat{\bar{d}}\left(s^f, w\right) = s^{\bar{l}} \quad \left(l \neq \bar{l}\right). \tag{18}$$

Then the following relations hold - in regard to the transition probabilities,

the average reward functions resp. the cost:

a)

$$\begin{aligned} p_{fl}^{\bar{d}} &= p_{fl}^{d} - q(w) \\ p_{f\bar{l}}^{\bar{d}} &= p_{f\bar{l}}^{d} + q(w) \\ p_{rv}^{\bar{d}} &= p_{rv}^{d} \text{ for } (f,l) \neq (r,v) \neq \left(f,\bar{l}\right) \quad \text{(cf. (10))}. \end{aligned} \tag{19}$$

That means the corresponding matrices of transition probabilities are only different in two elements (of a row)!

b)

$$\begin{aligned} \gamma\left(s^f, \bar{d}\right) &= \gamma\left(s^f, d\right) + q(w)\left(\hat{c}\left(s^f, w, s^{\bar{l}}\right) - \hat{c}\left(s^f, w, s^l\right)\right) \\ \gamma\left(s^l, \bar{d}\right) &= \gamma\left(s^l, d\right) \quad \text{for } \ l \neq f \end{aligned} \tag{20}$$

resp.

c)

$$\begin{aligned} c^{\bar{d}}\left(s^f, s^l\right) &= \left(c^d\left(s^f, s^l\right) - \frac{\hat{c}\left(s^f, w, s^l\right)}{p_{fl}^d} q(w)\right) \frac{p_{fl}^d}{p_{fl}^{\bar{d}}} \\ c^{\bar{d}}\left(s^f, s^{\bar{l}}\right) &= \left(c^d\left(s^f, s^{\bar{l}}\right) + \frac{\hat{c}\left(s^f, w, s^{\bar{l}}\right)}{p_{f\bar{l}}^d} q(w)\right) \frac{p_{f\bar{l}}^d}{p_{f\bar{l}}^{\bar{d}}} \\ c^{\bar{d}}\left(s^r, s^v\right) &= c^d\left(s^r, s^v\right) \text{ for } (f,l) \neq (r,v) \neq \left(f,\bar{l}\right). \end{aligned} \tag{21}$$

Of course, the computation of the stationary distribution is more complicated. In general, it is $p_r^{d,\infty} \neq p_r^{\bar{d},\infty}$ for all r. But the differences of $p_r^{d,\infty}$ and $p_r^{\bar{d},\infty}$ "are greater" for $r = l$ and $r = \bar{l}$ than for other $r \in \{1, \cdots, m\}$ and moreover we have:

Theorem 3. *Let P^d and $P^{\bar{d}}$ be two stochastic matrices with positive elements. Let P^d differ from $P^{\bar{d}}$ only for two elements in the following manner*

$$p_{fl}^d > p_{fl}^{\bar{d}} \quad \textit{and} \quad p_{f\bar{l}}^d < p_{f\bar{l}}^{\bar{d}}.$$

Then corresponding relations are true for the stationary distributions $P^{d,\infty}$ and $P^{\bar{d},\infty}$ belonging to P^d and $P^{\bar{d}}$:

$$p_l^{d,\infty} > p_l^{\bar{d},\infty} \quad \textit{and} \quad p_{\bar{l}}^{d,\infty} < p_{\bar{l}}^{\bar{d},\infty}.$$

(Cf. proof and remarks in [10] or [9], Section 3.2.3).

Figure 2. If d and $\bar{d}$ are different in the kind of (18), then the average expected costs per stage are different, first of all, in the marked terms

$$\sum_{v=1}^{m} \underbrace{\gamma(s^v, d)}_{=\text{ for } w \neq f} p_v^{d,\infty} = \cdots + \gamma(s^f, d) \cdot p_f^{d,\infty} + \cdots + \underbrace{\gamma(s^l, d)}_{=} \cdot p_l^{d,\infty} + \cdots + \underbrace{\gamma(s^{\bar{l}}, d)}_{=} \cdot p_{\bar{l}}^{d,\infty} + \cdots$$

$$\sum_{v=1}^{m} \overbrace{\gamma(s^v, \bar{d})} p_v^{\bar{d},\infty} = \cdots + \gamma(s^f, \bar{d}) \cdot p_f^{\bar{d},\infty} + \cdots + \overbrace{\gamma(s^l, \bar{d})} \cdot p_l^{\bar{d},\infty} + \cdots + \overbrace{\gamma(s^{\bar{l}}, \bar{d})} \cdot p_{\bar{l}}^{\bar{d},\infty} + \cdots$$

An Algorithm for approximate solutions of the Markov decision processes from this section which is based on the above structures can be found in [10].

3.2. Special considerations

3.2.1. The dominant policy

In this subsection, we want to show that it is easy to inspect Markov decision processes which are based on DA-models in the case of monotonicity.

Theorem 4. *Let a stationary Markov decision process with states $s^1, \cdots, s^m$ be given and let the transition probability matrices $P^{\bar{d}}$ and P^d belong to $\bar{d} \in A^M$ and $d \in A^M$ respectively. Furthermore, the following conditions are assumed to be fulfilled, where*

$I_v = \{h_{v-1}+1, h_{v-1}+2, \cdots, h_v\}, v = 1, 2, \cdots, r$

are sets of indices with $h_0 = 0, h_{v-1} < h_v, h_r = m$:

(C1) *(monotonicity or domination)*

$$\sum_{v=1}^{\bar{r}} \sum_{l\in I_v} p^{\bar{d}}_{h_1 l} \geq \sum_{v=1}^{\bar{r}} \sum_{l\in I_v} p^{\bar{d}}_{h_2 l} \geq \cdots \geq \sum_{v=1}^{\bar{r}} \sum_{l\in I_v} p^{\bar{d}}_{h_r l}, \ \forall \bar{r} = 1, \cdots, r$$

(C2) *(comparison)*

$$\sum_{v=1}^{\bar{r}} \sum_{l\in I_v} p^{\bar{d}}_{\bar{y} l} \leq \sum_{v=1}^{\bar{r}} \sum_{l\in I_v} p^{d}_{\bar{y} l} \quad for \quad \bar{y} \in I_y, \ \forall \bar{r} = 1, \cdots, r \ \forall y = 1, \cdots, r$$

(C3) *(reduction)*

$$\sum_{\bar{y}\in I_y} p^{\bar{d}}_{h_{q-1}+1\bar{y}} = \sum_{\bar{y}\in I_y} p^{\bar{d}}_{h_{q-1}+2\bar{y}} = \cdots = \sum_{\bar{y}\in I_y} p^{\bar{d}}_{h_q \bar{y}}, \forall q = 1, \cdots, r \forall y = 1, \cdots, r$$

(a) *Then* (C1), (C3) *are valid for each power* $(P^{\bar{d}})^t (t = 1, 2, \cdots)$ *of the matrix* $P^{\bar{d}}$ *and* (C2) *is valid for each power* $(P^{\bar{d}})^t, (P^d)^t (t = 1, 2, \cdots)$, *too.*

(b) *Let additionally the stationary distributions* $P^{\bar{d},\infty}$ *and* $P^{d,\infty}$ *exist and let the reward functions belong to* $\bar{d}$ *and* d*, respectively, fulfil the conditions*

$(Cr1, a)$ $\gamma(s^{h_1}, \bar{d}) \geq \gamma(s^{h_2}, \bar{d}) \geq \cdots \geq \gamma(s^{h_r}, \bar{d})$

$(Cr1, b)$ $\gamma(s^{h_{v-1}+1}, \bar{d}) = \gamma(s^{h_{v-2}+2}, \bar{d}) = \cdots = \gamma(s^{h_v}, \bar{d}) \ \forall v = 1, \cdots, r$

$(Cr2)$ $\gamma(s^l, \bar{d}) \leq \gamma(s^l, d) \ \forall l = 1, \cdots, m.$

Then the inequality

$$\gamma(s^1,\bar{d})p_1^{\bar{d},\infty}+\cdots+\gamma(s^m,\bar{d})p_m^{\bar{d},\infty}\le\gamma(s^1,d)p_1^{d,\infty}+\cdots+\gamma(s^m,d)_m^{d,\infty}$$

is valid.

Remarks.

- Of course, $I_q=\{q\}, q=1,\cdots,m$ is possible in Theorem 4. The definition of I_q, (C3), (Cr1,b) serves the reduction of the Markov decision process if the conditions are fulfiled.
- Conditions for Markov-chains in the way of (C1) and (C2) we find in a paper of Daley, [3], occurring for the first time.
- The proof and further comments are to be found in [9], p. 97 and in [11], p. 15 or in [9], section 3.3.3.

Definition 3. $\bar{d}\in A^M$ is called a dominant policy of a stationary Markov decision process if the conditions of Theorem 2 are valid with regard to $\bar{d}$ and to any $d\in A^M$.

(Of course, a dominant policy is an optimal policy.)

Now, let a Markov decision process be given, which results from the DA model. On the whole we investigate the existence of (and construct) a dominant policy in the following way.

1. We compute d^0 that

(22)
$$\begin{gathered}d^0(s^f,w)=s^l\\ \text{with}\\ \hat{c}(s^f,w,s^l)=min\{\hat{c}(s^f,w,s^{l'})|s^{l'}\in\hat{A}(s^f,w)\}\\ \text{for } f=1,\cdots,m,\ w\in B\end{gathered}$$

and $\gamma(s^f,d^0)$ by means of (11) (first equation).

Then the condition (Cr2) of Theorem 4 is fulfiled (cf. the first equation of (11) and (20)).

2. We number the states in a new way that

(23)
$$\gamma(s^{\lambda_1},d^0)\ge\gamma(s^{\lambda_2},d^0)\ge\cdots\ge\gamma(s^{\lambda_m},d^0)$$

$\{\lambda_1, \cdots, \lambda_m\} = \{1, \cdots, m\}$ (condition (Cr1,a)!).

3. Further on, we check the conditions (C1) and (C2) of Theorem 4.

   Either not all conditions are valid or d^0 is an optimal dominant policy ($\bar{d} = d^0$).

(If d_0 (from 1.) is not unique or any equations can be found in (23) then additional conditions are to be laid down.)

Remark. First of all Theorem 4 and dominant policies play a role in stationary Markov decision processes which result from DA models:

Property II a) (from Theorem 2) can be used to construct a strategy so that (C2) (of a dominant policy) is satisfied (if the numbering of the states is adequate) and property II b) (from Theorem 2) can be used to construct a strategy so that (Cr2) is satisfied (compare 2., too).

If both strategies are conform and (C1) is fulfilled, then a dominant policy exists. Markov decision processes which are not based on DA models with dominant policies are hard to be found.

3.2.2. Problem with special stage-cost and a "partial certainty equivalent principle"

Now, let K, G be defined that

$$K : S \times \mathbb{R}^m \times A \to \mathbb{R}_+,$$

$$G : S \times \mathbb{R}^m \times A \to S.$$

Except for this fact we take as a basis the problem as at the beginning of Section 3 with finite or infinite horizon (where the stationary distribution should exist). Further on, we assume that the state-cost are independent of the decisions, that means

$$K : S \times B \to \mathbb{R}_+ \text{ (and } K : S \times \mathbb{R}^m \to \mathbb{R}_+, \text{too).}$$

This is the special case (b) from Section 3, compare (17a).

And

$$\gamma(s,d) = \sum_w K(s,w)q(w) = E(K(s,w)) = \sum \hat{c}(s,w)q(w) =: \gamma(s)$$

follows from the representation as stationary Markov decision process with $c(s,w) = \hat{c}(s,w,s')$ are independent of d (compare (17a), too).

In this section, we will interpret an answer to the following question: Under which conditions are the optimal solutions of the "surrogate problems"

$$min\{K(s', E(w)) | s' \in \hat{A}(s,w)\} (s \in S, w \in B)$$

optimal decisions $d(s,w)$ for the above DA model, respectively for the above Markov decision problem?

At first we will notice that the following condition is sufficient:

(Co) Condition of orders (sufficient condition) A numbering of states $\{s^1, s^2, \cdots, s^m\}$ exists so that

$$K(s^1, E(w)) \geq K(s^2, E(w)) \geq \cdots \geq K(s^m, E(w)) \tag{24}$$

and

$$\min_{d_1,\cdots,d_n} \left\{ \left(\left(\sum_{t=0}^{n} \prod_{t'=0}^{t} P^{d_{t'}} \right) \gamma \right)_1 \right\} \geq \min_{d_1,\cdots,d_n} \left\{ \left(\left(\sum_{t=0}^{n} \prod_{t'=0}^{t} P^{d_{t'}} \right) \gamma \right)_2 \right\}$$

$$\geq \cdots \geq \min_{d_1,\cdots,d_n} \left\{ \left(\left(\sum_{t=0}^{n} \prod_{t'=0}^{t} P^{d_{t'}} \right) \gamma \right)_m \right\} \tag{25}$$

for $n = 0, 1, \cdots$

(There $\gamma = \begin{pmatrix} \gamma(s^1) \\ \vdots \\ \gamma(s^m) \end{pmatrix}$ and $P^{d^0} := I$, cf. [11]).

We remark that (25) for $n = 0$ means

$$\begin{aligned} \gamma(s^1) &= E(K(s^1, w)) \geq \gamma(s^2) \\ &= E(K(s^2, w)) \geq \cdots \geq \gamma(s^m) = E(K(s^m, w)) \end{aligned} \tag{26}$$

and that (25) is valid for dominant policies with suitable numberings of states.

We speak of the certainty equivalent principle of DA models with a special stage- cost subject to the restriction $x_{t+1} \in \hat{A}(s_t, w_{t+1})$ if the "future" w_{t+1} in the functional equation (3) are replaced by $E(w)$ in the corresponding functional equation, that means

$$
\begin{aligned}
f_t^c(s_t, w_t) &= \min_{x_t \in A(s_t, w_t)} \big(K(s_t, w_t) + f_{t+1}^c(s_{t+1}, E(w))\big) \\
&= K(s_t, w_t) + \min_{x_t \in A(s_t, w_t)} f_{t+1}^c(s_{t+1}, E(w)) \\
&= K(s_t, w_t) + \min_{s_{t+1} \in \hat{A}(s_t, w_t)} f_{t+1}^c(s_{t+1}, E(w)) \\
f_{N+1}^c &\equiv 0
\end{aligned}
\tag{27}
$$

and (3) (in the special case) and (27) have same optimal solutions.

Lemma 5. *Under the following condition* (Cc) *the functional equation* (27) *yields the same optimal decisions, that means the same states* s_{t+1}*, as the surrogate problems* $\min\limits_{s_{t+1} \in \hat{A}(s_t, w_t)} K(s_{t+1}, E(w))$ *for* $t \leq N-1$.

(Cc) Compensation condition of the stage-cost

If $\tilde{s}^* \in \operatorname{argmin} \Big\{K(\tilde{s}, E(w)) | \tilde{s} \in \hat{A}(s, w)\Big\}$, then

$\min \Big\{K(\tilde{\tilde{s}}, E(w)| \tilde{\tilde{s}} \in \hat{A}(\hat{A}(s, w), E(w))\Big\}$

$= \min \Big\{K(\tilde{\tilde{s}}, E(w))| \tilde{\tilde{s}} \in \hat{A}(\tilde{s}^*, E(w))\Big\}$

for any $s \in S, w \in B$.

Proof.

(27) yields

$$f_N^c(s_N, w_N) \;=\; K(s_N, w_N) + 0 \;=\; K(s_N, w_N)$$

thus

$$f_{N-1}^c(s_{N-1}, w_{N-1}) \;=\; K(s_{N-1}, w_{N-1}) + \min_{s_N \in \hat{A}(s_N, w_N)} K(s_N, E(w_N)).$$

Obviously, this functional equation includes the same optimal decisions (the same optimal states s_N) as the surrogate problem $\min\limits_{s_N \in \hat{A}(s_N, w_N)} K(s_N, E(w))$.

Now we consider the functional equation (27) for any $t \in \{1, \dots, N-2\}$:

$$\begin{aligned} f_t^c(s_t, w_t) \;&=\; K(s_t, w_t) + \min_{s_{t+1} \in \hat{A}(s_t, w_t)} f_{t+1}^c(s_{t+1}, E(w)) \\ &=\; K(s_t, w_t) + \\ &\min_{s_{t+1} \in \hat{A}(s_t, w_t)} \Big[K(s_{t+1}, E(w)) + \min_{s_{t+2} \in \hat{A}(s_{t+1}, E(w))} f_{t+2}^c(s_{t+2}, E(w)) \Big]. \end{aligned}$$

If we use the compensation condition (Cc) with

$\tilde{s} = s_{t+1}$,

$s = s_t, \; w = w_t$

and

$\approx\!\!\!\!s = s_{t+2}$

then

$$f_t^c(s_t, w_t) \;=\; K(s_t, w_t) + K(\tilde{s}^*, E(w)) + \min_{s_{t+2} \in \hat{A}(\tilde{s}^*, E(w))} f_{t+2}^c(s_{t+2}, E(w))$$

with $\tilde{s}^* \in \arg\min\{K(s_{t+1}, E(w)) \mid s_{t+1} \in \hat{A}(s_t, w_t)\}$ follows.

That means the functional equation yields the same optimal decisions (the same optimal states $s_{t+1} = \tilde{s}^*$ as the surrogate problem

$$\min_{s_{t+1} \in \hat{A}(s_{t+1}, w_{t+1})} K(s_{t+1}, E(w)).$$

Definition 4. Let condition (Cc) be valid. If the optimal solutions of the surrogate problems are optimal decisions $d(s, w)$ for the DA models and Markov decision problems of this section, then we call this the partial certainty equivalent principle.

(We call the principle "partial" since condition (Cc) simplifies (27) in the way of Lemma 5).

4. DA models with distance properties

In this section, we consider Markov decision processes as in Section 3 and additionally we assume "distance properties". In a natural way such properties are found, for instance, in flow problems (compare [1] and among other things see Theorem 3.4). We want to give surrogate problems (these surrogate problems are a kind of two-stage-problems), which can be used to solve the Markov decision processes approximately. These surrogate problems are to be used above all, if the state spaces of the Markov decision processes are very immense.

The distance properties include:

$$\text{1.} \qquad \hat{c}(s, w, s') = 0 \text{ if and only if } s \in \hat{A}(s, w) \text{ and } s' = s. \tag{28}$$

$$\text{2.} \qquad \left.\begin{array}{r} \text{Let be } s^v \in \hat{A}(s^l, w^2), s^v \in \hat{A}(s^f, w^1), \\ s^l \in \hat{A}(s^f, w^1) \text{ then} \\ \hat{c}(s^l, w^2, s^v) + \hat{c}(s^f, w^1, s^l) \geq \hat{c}(s^f, w^1, s^v) \\ \text{has to follow.} \\ (triangle - inequality) \end{array}\right\} \tag{29}$$

4.1. Surrogate problems

The surrogate problems are formulated by means of

Definition 5. Let $\bar{s} \in \hat{A}(s,w)$. The set

$B\ OPT(\bar{s}|s,w) =$

$\{\bar{w}|\bar{s}\in\hat{A}(\bar{s},\bar{w}) \text{ and } \nexists \bar{s}'\in\hat{A}(s,w) : \hat{c}(s,w,\bar{s}') < \hat{c}(s,w,\bar{s}) \text{ and } \bar{s}'\in\hat{A}(\bar{s}',\bar{w})(*)\}$

is called the optimum domain to $\bar{s}$.

We remark that $\bar{s} \in \hat{A}(\bar{s},\bar{w})$ includes $\hat{c}(\bar{s},\bar{w},\bar{s}) = 0$ and likewise $\bar{s}' \in \hat{A}(\bar{s}',\bar{w})$ includes $\hat{c}(\bar{s}',\bar{w},\bar{s}') = 0$. The definition is based on the first distance property.

Now, we explain the definition.

If we make the internal decision $\bar{s} = \hat{d}(s,w)$ at a present stage then the optimum domain to $\bar{s}$ includes disturbances $\bar{w}$ (at the next stage) so that 1. the cost at the next stage could be zero, namely $\hat{c}(\bar{s},w,\bar{s})$ 2. the cost $\hat{c}(s,w,\bar{s})$ at the present stage fulfils the "minimum condition" (*).

In the surrogate problems the "probabilities of optimum domains" are maximized over $\hat{A}(s,w)$:

$$(30) \qquad P(w \in BOPT(\bar{s}|s,w)) = \sum_{\bar{w}\in BOPT(\bar{s}|s,w)} q(\bar{w}) \to \max, \quad \bar{s} \in \hat{A}(s,w)$$

(where $q(\bar{w}) = \prod_{i=1}^{n} q^i(\bar{w}_i)$ if w_i are realized independent of each other).

Remark. Let d^o denote a policy which is yielded by the surrogate problems. The inclusion of the "minimum conditions" (*) in the surrogate problems affects the greatness of $\gamma(s,d^o)$ (compare (20)). $P(w \in BOPT(\bar{s}|s,w))$-self is related to the transition probabilities $p(\bar{s}|\bar{s},d^o)$, hence to $p^{d^o,\infty}(\bar{s})$, too. On the whole, the surrogate problems yield great $p^{d^o,\infty}(s)$ for small $\gamma(s,d^o)$ and vice versa. By this we can consider problems (30) as surrogate problems for Markow decision processes of this section (further comments follow in Section 5).

4.2. A certain subset of the problems with distance properties

Problems with (28) and (29) for which both the additional properties

$$\hat{c}(s, w, s')(= \hat{c}(s, s')) \text{ do not depend on } w \tag{31}$$

(compare (15) and (16))

$$\hat{A}(s, w)(= \hat{A}(w)) \text{ do not depend on s} \tag{32}$$

are fulfilled represent a wide subset of problems with distance properties. In the sensible way we assume $\hat{A}(w) \neq \emptyset$ for any $w \in B$.

Without loss of generality we can make the internal decision sets $\hat{A}(w)$ smaller for these problems.

4.2.1. The decision set of feasible states

At first we specify the triangle inequality under assumption (32). (In Section 4.2.1 we do not need (31).)

$$\left.\begin{array}{r} \text{Let } s^v \in \hat{A}\left(w^1\right), s^v \in \hat{A}\left(w^2\right), \text{ and } s^l \in \hat{A}\left(w^1\right) \\ \text{then} \\ c\left(s^l, w^2, s^v\right) + c\left(s^f, w^1, s^l\right) \geq c\left(s^f, w^1, s^v\right) \ \forall s^f \in S \\ \text{has to follow.} \end{array}\right\} \tag{33}$$

We will show that states s^v are not essential for optimal internal decisions $d^*\left(s^f, w^1\right)$, if we have an equality in (33).

For this we define the DA decision sets of feasible states:

Definition 6. Let $s^f \in S, w^1 \in B$ be given.

$$\hat{\hat{A}}\left(s^f, w^1\right) = \Big\{s^v \in \hat{A}\left(w^1\right) | \exists w^2 \in B \text{ with } s^v \in \hat{A}\left(w^2\right):$$

$$\hat{c}\left(s^l, w^2, s^v\right) + \hat{c}\left(s^f, w^1, s^l\right) > \hat{c}\left(s^f, w^1, s^v\right) \ \forall\, s^l \in \hat{A}\left(w^1\right) \text{ with } s^l \neq s^v\Big\}$$

is called DA decision set of feasible states for a given state s^f and a realized disturbance w^1.

(We notice that $\hat{\hat{A}}(s^f, w^1) = \{s^v\}$ if $\hat{A}(w^1) = \{s^v\}$. For this we choose $w^2 = w^1$ in Definition 6. Further on, it is $s^v \in \hat{\hat{A}}(s^f, w^1)$ if $\hat{c}(s^f, w^1, s^v) = \min_{s^l \in \hat{A}(w^1)} \hat{c}(s^f, w^1, s^l)(*)$, that means $\hat{\hat{A}}(z^f, w^1) \neq \emptyset$. Take into consideration that $\hat{c}(s^l, w^2, s^v) \neq 0$ for $s^l \neq s^v$.)

Theorem 6. *The minimum of the average expected cost per stage of a Markov decision process which results from DA models with distance properties* (28) *and* (29) *and property* (32) *will not increase, when* $\hat{\hat{A}}(s, w)$ *is used instead of* $\hat{A}(w)$ *for* A^M *(compare* (32), (9)*).*

Proof. Let $s_1, (= \tilde{s}_1)$ be some initial state. Further on let be given any sequences $w_t, t = 1, 2, \ldots$ with $w_t \in B$ and $\tilde{s}_t, t = 2, 3 \ldots$ with $\tilde{s}_{t+1} \in \hat{A}(w_t)$ for $t = 1, 2, \ldots$.

We will construct a sequence $s_t, t = 2, 3, \ldots$ with $s_{t+1} \in \hat{\hat{A}}(s_t, w_t)$ that

$$\sum_{t'=1}^{t} \hat{c}(\tilde{s}_{t'}, w_{t'}, \tilde{s}_{t'+1}) \geq \sum_{t'=1}^{t} \hat{c}(s_{t'}, w_{t'}, s_{t'+1}) \tag{34}$$

for $t = 1, 2, \ldots$.

It is possible that the orginated policy is not stationary. But it is well known that then an optimal stationary policy with the same set of action spaces exists, too. Now, we construct successively $s_{t+1} \in \hat{\hat{A}}(s_t, w_t)$ by

$$s_{t+1} = \begin{cases} \tilde{s}_{t+1}, \text{ if } \tilde{s}_{t+1} \in \hat{\hat{A}}(s_t, w_t) \\ s_{t+1}, \text{ if } \tilde{s}_{t+1} \notin \hat{\hat{A}}(s_t, w_t) \text{ and } \hat{c}(s_{t+1}, w_{t+1}, \tilde{s}_{t+1}) \\ +\hat{c}(s_t, w_t, s_{t+1}) = \hat{c}(s_t, w_t, \tilde{s}_{t+1}). \end{cases} \tag{35}$$

Such $s_{t+1} \in \hat{\hat{A}}(s_t, w_t)$ exists in the second case since S is finite, compare at the beginning of Section 3.) We show that $s_t, t = 1, 2, \dots$ fulfil (34) by means of mathematical induction:

Obviously, the inequality

$$\hat{c}\,(s_1, w_1, \tilde{s}_2) \geq \hat{c}\,(s_1, w_1, s_2)$$

is valid, compare (35).

Now, we assume that the inequality (34) is right for $1, 2, \dots, t$. In the case $\tilde{s}_{t+1} = s_{t+1}$ the inequality (34) for $t+1$ follows from

$$\hat{c}\,(\tilde{s}_{t+1}\,(= s_{t+1})\,, w_{t+1}, \tilde{s}_{t+2}) \geq c\,(s_{t+1}, w_{t+1}, s_{t+2})\,,$$

compare (35).

Finally, we consider cases with

$$\tilde{s}_{t_1} = s_{t_1},\ t_1 < t+1$$

and

$$\tilde{s}_{t''} \neq s_{t''},\ t_1 < t'' \leq t+1.$$

We compute in this way that we alternately use (35) and the triangle inequality:

$$\begin{aligned}
&\underline{\hat{c}(s_{t_1}, w_{t_1}, \tilde{s}_{t_1+1})} + \hat{c}(\tilde{s}_{t_1+1}, w_{t_1+1}, \tilde{s}_{t_1+2}) + \hat{c}(\tilde{s}_{t_1+2}, w_{t_1+2}, \tilde{s}_{t_1+3}) + \cdots + \\
&\qquad + \hat{c}(\tilde{s}_{t+1}, w_{t+1}, \tilde{s}_{t+2}) \\
&= \underline{\hat{c}(s_{t_1+1}, w_{t_1+1}, \tilde{s}_{t_1+1})} + \hat{c}(s_{t_1}, w_{t_1}, s_{t_1+1}) + + \underline{\hat{c}(\tilde{s}_{t_1+1}, w_{t_1+1}, \tilde{s}_{t_1+2})} + \\
&\qquad + \hat{c}(\tilde{s}_{t_1+2}, w_{t_1+2}, \tilde{s}_{t_1+3}) + \cdots + \hat{c}(\tilde{s}_{t+1}, w_{t+1}, \tilde{s}_{t+2}) \\
&\geq \underline{\hat{c}(s_{t_1+1}, w_{t_1+1}, \tilde{s}_{t_1+2})} + \hat{c}(s_{t_1}, w_{t_1}\ s_{t_1+1}) + \hat{c}(s_{t_1+2}, w_{t_1+2}, s_{t_1+3}) + \cdots + \\
&\qquad \hat{c}(\tilde{s}_{t+1}, w_{t+1}, \tilde{s}_{t+2}) \\
&\geq \underline{\hat{c}(s_{t_1+2}, w_{t_1+2}, \tilde{s}_{t_1+2})} + \hat{c}(s_{t_1+1}, w_{t_1+1}, s_{t_1+2}) + \hat{c}(s_{t_1}, w_{t_1}, s_{t_1+1}) + \\
&\qquad + \underline{\hat{c}(\tilde{s}_{t_1+2}, w_{t_1+2}, \tilde{s}_{t_1+3})} + \cdots + \hat{c}(\tilde{s}_{t+1}, w_{t+1}, \tilde{s}_{t+2})
\end{aligned}$$

$$\geq \underline{\hat{c}(s_{t_1+2}, w_{t_1+2}, \tilde{s}_{t_1+3})} + \hat{c}(s_{t_1+1}, w_{t_1+1}, s_{t_1+2}) + \hat{c}(s_{t_1}, w_{t_1}, s_{t_1+1}) + \cdots + \hat{c}(\tilde{s}_{t+1}, w_{t+1}, \tilde{s}_{t+2})$$

$$\vdots$$

$$\geq \underline{\hat{c}(s_{t+1}, w_{t+1}, \tilde{s}_{t+1})} + \hat{c}(s_t, w_t, s_{t+1}) + \cdots + \hat{c}(s_{t_1+1}, w_{t_1+1}, s_{t_1+2}) + \hat{c}(s_{t_1}, w_{t_1}, s_{t_1+1}) + \underline{\hat{c}(\tilde{s}_{t+1}, w_{t+1}, \tilde{s}_{t+2})}$$

$$\geq \underline{\hat{c}(s_{t+1}, w_{t+1}, \tilde{s}_{t+2})} + \hat{c}(s_t, w_t, s_{t+1}) + \cdots + \hat{c}(s_{t_1}, w_{t_1}, s_{t_1+1})$$

$$\geq \hat{c}(s_{t+2}, w_{t+2}, \tilde{s}_{t+2}) + \hat{c}(s_{t+1}, w_{t+1}, s_{t+2}) + \hat{c}(s_t, w_t, s_{t+1}) + \cdots + \hat{c}(s_{t_1}, w_{t_1}, s_{t_1+1})$$

$$\geq \hat{c}(s_{t+1}, w_{t+1}, s_{t+2}) + \hat{c}(s_t, w_t, s_{t+1}) + \cdots + \hat{c}(s_{t_1}, w_{t_1}, s_{t_1+1}).$$

4.2.2. A note on transition probabilities

Lemma 7. *The properties*

i) $s \in \hat{A}(w) \Rightarrow \hat{\hat{A}}(s, w) = \{s\}$

ii) $p(s|s, d) = \sum_{w: s \in \hat{A}(w)} q(w)$ *for* d *with* $\hat{d}(s', w') \in \hat{\hat{A}}(s', w') \forall s' \in S, w' \in B$

(that means $p(s|s, d) =: p(s|s)$ *is independent of* d*)*

are valid for problems of Section 4.2 *with* (28), (29), *for which especially the conditions* (31) *and* (32) *are fulfilled.*

Proof. At first, we remark that the triangle-inequality has the representation:

$$\hat{c}\left(s^l, s^v\right) + \hat{c}\left(s^f, s^l\right) \geq \hat{c}\left(s^f, s^v\right) \quad \textit{for any} \; ; s^f \in S \tag{36}$$

and $s^l \in \hat{A}(w), s^v \in \hat{A}(w)$ for any $w \in \hat{A}(w)$.

i) Let $s \in \hat{A}(w)$ and $s' \in \hat{A}(w)$ with $s' \neq s$. $\hat{c}(s,s') + \hat{c}(s,s) = c(s,s')$ follows from $\hat{c}(s,s) = 0$ (compare (28)). Hence $s' \notin \hat{\hat{A}}(s,w)$ and $\{s\} = \hat{\hat{A}}(s,w)$.

ii) Equation (10) and property i) yield

$$p(s|s,d) = \sum_{w:s=\hat{d}(s,w)} q(w) = \sum_{w:z\in\hat{A}(w)} q(b) =: p(s|s).$$

4.2.3. Problems with special internal cost

We consider problems of Section 5.2 with (28), (29), (31), (32) in the special case that

$$\text{(37)} \qquad \hat{c}\left(s, \bar{s}^1\right) = \cdots = \hat{c}\left(s, \bar{s}^v\right) \text{ for any } s, w \text{ and } \{\bar{s}^1, \cdots, \bar{s}^v\} = \hat{\hat{A}}(s,w)\,.$$

We can use such a special problem to investigate the quality of the surrogate problems for classes of problems with distance properties. (Compare the example in Section 5.)

Lemma 8. *The properties*

i) $\hat{c}(s,\bar{s}) < \hat{c}(s,\tilde{s})$ *if* $\{\bar{s},\tilde{s}\} \subseteq \hat{A}(w)$ *for any* w *and* $\bar{s} \in \hat{\hat{A}}(s,w), \tilde{s} \notin \hat{\hat{A}}(s,w)$

ii) $\gamma(s,d)$ *for* d *with* $\hat{d}(s',w') \in \hat{\hat{A}}(s',w') \forall s' \in S, w' \in B$ *do not depend on* d

are valid for problems of Section 4.2.3, *for which especially the condition* (37) *is fulfilled.*

Proof. A proof by contradiction yields i):

i) If $\hat{c}(s,\tilde{s}) \leq \hat{c}(s,\bar{s})$ for $\tilde{s} \notin \hat{\hat{A}}(s,w)$ and $\bar{s} \in \hat{\hat{A}}(s,w)$

then $\approx\!\!\!\!\! s$ with $\hat{c}(s,\tilde{\tilde{s}}) = \min_{s'\in\hat{A}(w)} \hat{c}(s,s'), \tilde{\tilde{s}} \notin \hat{\hat{A}}(s,w)$ ($\tilde{\tilde{s}} = \tilde{s}$ is possible) and $\hat{c}(s,\tilde{\tilde{s}})(\leq \hat{c}(s,\tilde{s})) \leq \hat{c}(s,\bar{s})$ exists.

This is contradictory to the remark *) on Definition 6.

ii) So we can define

$$
\begin{aligned}
c(w,s) :&= min\left\{\hat{c}(s,w,s')|s' \in \hat{\hat{A}}(s,w)\right\} \\
&= \hat{c}(s,w,s'),\ \text{where}\, s' \in \hat{\hat{A}}(s,w)
\end{aligned} \tag{38}
$$

follows from (37). This is analogous to (17a).

And (38) yields

$$
\gamma(s) := \gamma(s,d) = \sum_{w\in B} \hat{c}(s,w)q(w)
$$

$$
(\text{with } \ c(s,w) = \hat{c}(s,w,s') \ \text{ and } \ s' \in \hat{\hat{A}}(s,w)).
$$

5. EXAMPLE: A STOCHASTIC DYNAMIC TRANSPORTATION PROBLEM

(The detailed representation of this problem is to be found in [8] resp. [9].)

Let $n \in \mathbb{N}, n \geq 3, k_o = (k_{o_1}, k_{o_2}, \ldots, k_{o_n})$ with $k_{o_i} \in N$ and $su \in N$ with $su < \sum_{i=1}^{n} k_{o_i}$ be given.

Further on, let

$B_{k_o} = \left\{w \in \mathbb{Z}_+^n | 0 \leq w \leq k_o\right\}$ be the disturbance space,

let $S_{k_o;su} = \left\{s \in \mathbb{Z}_+^n | 0 \leq s \leq k_o, \sum_{i=1}^{n} s_i = su\right\}$ be the state space

and let

$$
A_{k_o;su}(s,w) = \left\{ X \in \mathbb{Z}_+^n \times \mathbb{Z}_+^n \,\middle|\, \begin{array}{l} \sum_{j=1}^{n} x_{ij} \leq z_i, \sum_{i=1}^{n} x_{ij} \ = w_j \forall_j \text{ if } \sum_{j=1}^{n} w_j \leq su \\ \sum_{i=1}^{n} x_{ij} \leq w_j; \sum_{j=1}^{n}, x_{ij} = z_i \forall_i \text{ if } \sum_{j=1}^{n} w_j \ \geq su \end{array} \right\}
$$

be the DA decision sets.

Then we consider the DA model:

$$E\left\{\sum_{t=1}^{N}\sum_{i,j=1}^{n} k_{ij}x_{t,ij}\right\} \to min$$

subject to the constraints

$$x_t \in A_{k_o;su}(s_t, w_t)$$

$$s_{t+1,i} = s_{t,i} - \sum_{j=1}^{n} x_{t,ij} + w_{t,i} \quad \text{for } i = 1, \cdots, n \text{ if } \sum w_{t,j} \le su$$

$$s_{t+1,j} = \sum_{i=1}^{n} x_{t,ij} \qquad \text{for } j = 1, \cdots, n \text{ if } \sum w_{t,j} \ge su$$

where $k_{ij} \in \mathbb{R}^n_+ \times \mathbb{R}^n_+$ with $k_{ii} = 0\ \forall_i, k_{ij} + k_{jl} > k_{il}\ \forall i \neq j \neq l, k_{ij} > 0\ \forall i \neq j$ and the components w_i of disturbances are realized independent of each other.

(The "certainty equivalence principle" is not valid for this problem.)

We can convert this problem for $N = \infty$ in a stationary Markov decision process as in Section 3.

Especially, we have there

$$\hat{c}(s, s') =$$

$$= min\left\{\sum_{i,j=1}^{n} k_{ij}x_{ij}\Big|\sum_{i=1}^{n} x_{ij} = s'_j, \sum_{j=1}^{n} x_{ij} = s_i, x_{ij} \in \mathbb{R}_+\right\} - \text{independent on } d$$

and at first we simplify the decision sets in the way

$$\hat{A}_{k_o;su}(w) = \left\{ s \in S_{k_o;su} \;\middle|\; \begin{array}{l} s = w \text{ if } w \in S_{k_o;su} \\ s \geq w \text{ if } \sum w_j < su \\ s \leq w \text{ if } \sum w_j > su \end{array} \right\}.$$

We can show that the distance properties (28) and (29) are fulfilled for this Markov decision process. The properties (31) and (32) are valid, too. The DA decision set of feasible state has the representation

$$\hat{\hat{A}}_{k_o;su}(s,w) = \left\{ \bar{s} \in S_{k_o;su} \;\middle|\; \begin{array}{l} w_i \leq \bar{s}_i \leq max\{s_i, w_i\} \text{ if } \sum w_i \leq su \\ min\{s_i, w_i\} \leq \bar{s}_i \leq w_i \text{ if} \sum w_i \geq su \end{array} \right\}.$$

The number of states of the state space can grow rather for certain su and k_o. (State spaces with more than fifty thousand millions exist for only $n = 10, k_{o_1} = \cdots = k_{o_{10}} = 19$.) In such cases it seems absolutely necessary to use approximate methods.

The surrogate problems of Section 4.1 are suitable for stochastic dynamic transportation problems with log-concave distributed disturbances (compare [9]).

Under the assumptions $k_{ij}(= const) = 1 \ \forall_{i,j}, k_{o_1} = k_{o_2} = \cdots = k_{o_n}$ we have special problems as in Section 4.2.3 with $\hat{c}(s,\bar{s}) = \frac{1}{2}\sum\limits_{i=1}^{n} |\bar{s}_i - s_i|$ for $\bar{s} \in \hat{\hat{A}}_{k_o;su}$.

In this case we assert that the partial certainty equivalent principle holds (and that the surrogate problems and the Markov decision process have the same optimal decisions for identical discrete log-concave distributed disturbances). Meantime we have shown that for problems with fixed numbers of states but a finite number of such problems.

Dominant policies exist for certain structures of state spaces.

References

[1] R.K. Ahuja, T.L. Magnanti and J.B. Orlin, *Networks Flows.* In: Handbooks in Operations Research and Management Science, Vol. I. Ed. by Neuhauser, G.L.; Rinnooy Kan, A.H.G.; Todd, M.J. Elsevier Science Pub. B.V.; Amsterdam 1989.

[2] D.P. Bertsekas, *Dynamic Programming and Optimal Control, I and II, Athena Scientific*, Belmont, Massachusetts. New York-San Francisco-London 1995.

[3] D.J. Daley, *Stochastic monotone Markov processes*, Z. Wahrsch. Verw. Gebiete, **10** (1968), 305–317.

[4] W. Dinkelbach, *Entscheidungsmodelle*, Springer-Verlag, Berlin-New York 1982.

[5] S.E. Dreyfus and A.M. Law, *The Art and Theory of Dynamic Programming*, Academic Press New York-San Francisco-London 1977.

[6] H.J. Girlich, P.M. Köchel and H.U. Küenle, *Steuerung dynamischer Systeme*, Birkhäuser, Basel 1990.

[7] O. Hernández-Lerma, *Adaptive Markov control Processe*, Springer-Verlag, New York-Berlin 1989

[8] R. Hildenbrandt, *A special stochastic decision problem*, Optimization **28** (1993), 95–110.

[9] R. Hildenbrandt, *Methoden aus ganzzahliger Optimierung und Verbandstheorie zur Behandlung eines stochastischen dynamischen Transportproblems*, Habilitationsschrift, TU Ilmenau 1995, (Libri BoD, 2000).

[10] R. Hildenbrandt, *Zu stochastischen dynamischen Optimierungsproblems*, Preprint No. M17/96, TU Ilmenau 1996.

[11] R. Hildenbrandt, *Notes of stochastic dynamic programming, where the random disturbances are observed before the decision is made at each stage and corresponding problems with distance properties*, Preprint No. M08/00, TU Ilmenau 2000.

[12] P. Kumar and R.P. Varaiya, *Stochastic Systems: Estimation, Identification, and Adaptive Control*, Prentice Hall, New Jersey 1986.

[13] A. Müller, *Integralinduzierte Ordnungen und Metriken auf Mengen von Wahrscheinlichkeitsmaŝen mit Anwendungen bei Markovschen Entscheidungsprozessen*, Dissertation, Universität Karlsruhe 1995.

[14] K. Neumann, *Operations-Research-Verfahren*, Band II Carl Hanser, München 1977.

[15] K. Neumann and M. Morlock, *Operations Research*, Carl Hanser Verlag München, Wien 1993.

[16] C. Schneeweiss, *Dynamisches Programmieren*, Physica-Verlag Würzburg-Wien 1974.

[17] H.J. Sebastian and N. Sieber, *Diskrete dynamische Optimierung*, Geest & K.G. Porting; Leipzig 1981.

Received 15 January 2002
Revised 6 May 2003