R. Hildenbrandt

DA Stochastic Dynamic Programming, Stochastic Dynamic Distance Optimal Partitioning Problems and Partitions-Requirements-Matrices

Cuvillier Verlag Göttingen
Internationaler wissenschaftlicher Fachverlag

DA Stochastic Dynamic Programming,

Stochastic Dynamic Distance Optimal Partitioning Problems and

Partitions-Requirements-Matrices

R. Hildenbrandt

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

1. Aufl. - Göttingen: Cuvillier, 2010

978-3-86955-608-6

Nonnenstieg 8, 37075 Göttingen
Telefon: 0551-54724-0
Telefax: 0551-54724-21
www.cuvillier.de

1. Auflage, 2010
Gedruckt auf säurefreiem Papier

978-3-86955-608-6

DA Stochastic Dynamic Programming, Stochastic Dynamic Distance Optimal Partitioning Problems and Partitions-Requirements-Matrices

---

R. Hildenbrandt

Technical University Ilmenau, Institute of Mathematics

For my family

I want to thank Mary Donahue for her patent and careful language corrections of this book.

Contents

1 Introduction **7**

1.1 Chapter 2 Contents 9

1.2 Chapter 3 Contents 11

1.3 Chapter 4 Contents 13

2 DA Stochastic Dynamic Programming with Random Disturbances **15**

2.1 The DA Model 16

2.2 The Certainty Equivalence Principle 24

2.3 DA Models as Markov Decision Processes 32

2.3.1 Markov Decision Processes: Models and Properties . . 33

2.3.2 DA Models as Markov Decision Processes under Appropriate Assumptions 36

2.3.2.1 The Structure of Decisions within DA Markov Decision Processes 39

2.3.2.2 An Optimality Criterion for DA MDP and an "Almost-Partial Order" of the States . . . 42

2.3.3 DA Models with Special Properties 51

2.3.3.1 Notes on DA Models with Distance Properties 51

2.3.3.2 The Dominant Policy 64

2.3.4 Cost-Parametric Analysis of DA Markov Decision Processes 71

2.3.4.1 Some Properties of Parametric Markov Decision Processes 72

2.3.4.2 Cost-Parametric DA Markov Decision Processes 77

2.3.4.3 Remarks on the Parametrization of DA Markov Decision Processes 92
2.3.5 Remarks on the Solutions of DA Markov Decision Processes 93

3 The Problem of Stochastic Dynamic Distance Optimal Partitioning (SDDP) 97
3.1 Preliminary Notes on Non-Balanced Transportation Problems 98
3.2 Model Formulation 105
3.3 Properties of SDDP Problems and their Characteristic Quantities and Conversion Numbers 114
3.4 Characterization of Special Cases of SDDP Problems 131
3.4.1 Average Reward Functions of SDDP Problems with Identical Basic Costs 135
3.4.2 Reduction of SDDP Problems with Identical Basic Costs and Independent and Identically Distributed Requirements 140
3.5 Notes on the Solution Methods of SDDP Problems 145

4 Partitions-Requirements-Matrices 153
4.1 Arrangement of the Partitions 155
4.2 Definitions of Partitions-Requirements-Matrices and Initial Properties and Results 163
4.2.1 General Partitions-Requirements-Matrices 164
4.2.2 Partitions-Requirements-Matrices 166
4.2.3 Partitions-Requirements-Matrices for Equivalent Lattices of Partitions 173
4.3 The Computation of PRMs by means of Permutations of Perturbed Partitions 176
4.4 Limits of Partitions-Requirements-Matrices 200
4.4.1 Classification of the Lattices of the Restricted Partitions 201
4.4.2 Limits of Partitions-Requirements-Matrices with regard to Sets of Sparse Partitions 213
4.4.3 Limits of Partitions-Requirements-Matrices with regard to Sets of Non-Truncated Heavy Partitions . . 218

4.5 Further Results from Elements of Partitions-Requirements-Matrices 224
4.5.1 Elements of PRMs in Dependence on Variables n and k_0 225
4.5.2 The Elements of the Last Row and the Last Column of PRMs in the Case that $n \geq su$ 249
4.6 Poisson Equations for PRMs and the Monotonicity of their Solutions 254
4.6.1 Poisson Equations for PRMs 256
4.6.2 Partial Results 257
4.6.2.1 PRMs for Sets of Restricted Partitions which are Equivalent with regard to the Poisson Equation 257
4.6.2.2 The Dominance Condition and PRMs with regard to
m-Totally Ordered Sets of Partitions 259
4.6.2.3 PRMs with regard to Sets of Restricted Partitions with at most 4 Partitions 262
4.6.3 The Poisson Equations with regard to Sets of Sparse Partitions with Sufficiently Large n 264
4.6.4 The Poisson Equations with regard to Sets of
Non-Truncated Heavy Partitions with Sufficiently Large n or k_0 272
4.7 Conclusion for SDDP Problems with Identical Basic Cost and Independent and Identically Distributed Requirements 280

Final observations on SDDP problems **282**
References **284**

Glossary of Symbols and Abbreviations **287**

Index **296**

Chapter 1

Introduction

The **S**tochastic **D**ynamic **D**istance Optimal **P**artitioning (SDDP) problem - an Operations Research problem - was the motivation for the investigations presented in this book.

As evident from the name of the problem, investigations in two different mathematical fields were necessary for its treatment, i.e. in stochastic dynamic programming and in combinatorics (”Partitioning”).

This book therefore, apart from the introduction, covers the following three chapters

2 DA Stochastic Dynamic Programming with Random Disturbances,

3 The Problem of Stochastic Dynamic Distance Optimal Partitioning (SDDP),

4 Partitions-Requirements-Matrices (PRMs).

DA (”decision after”) stochastic dynamic programming with random disturbances is characterized by the fact that these random disturbances are observed before the decision is made at each stage.

In the past only very moderate attention was given to problems with this characteristic (see also Section 1.1).

Examples of DA models are SDDP problems and certain inspection-replacement problems. (Also refer to connections with k-server problems and metric task systems at the end of Section 1.2.)

In Chapter 2 specific properties of DA stochastic dynamic programming problems are worked out for theoretical characterization and for more efficient solution strategies of such problems.

In order to understand this chapter, and the book altogether, previous knowledge about stochastic dynamic programming and Markov decision processes (MDPs) is useful, however not absolutely necessary since the concerned models are developed from scratch. (Basic knowledge can be found in [7], [12], [28], [16] or [31].)

In Chapter 3 we formulate and discuss in detail the problem of **S**tochastic **D**ynamic **D**istance Optimal **P**artitioning (**SDDP**).

SDDP problems are extremely complex.

Superordinately regarded, SDDP problems are DA stochastic dynamic programming problems (**S**tochastic **D**ynamic DP).

It requires a certain initial effort, however, in order to compute the real input data for the DA stochastic dynamic programming problem (SD **D**istance optimal P).

Furthermore, the problem shows combinatorial aspects (SDD **P**artitioning).

The understanding for the formulation of the problem and the basic methods of its solution requires knowledge from Section 2.1 (at least from the beginning of this section) and absolutely from Section 2.3.

However, an important statement concerning certain SDDP problems is proven at the end of Chapter 4, only after several combinatorial considerations.

Partitions-requirements-matrices (PRMs) (Chapter 4) are matrices of transition probabilities of SDDP problems which are formulated as Markov decision processes (MDPs).

PRMs "in the strict meaning" include optimal decisions of certain SDDP problems, as is shown toward the end of Chapter 4.

PRMs (in the strict meaning) themselves represent interesting (almost self-evident) combinatorial structures, which are not otherwise found in literature.

We therefore ensure that the treatise of Chapter 4 can essentially be understood independent of Chapters 2 and 3.

Relationships to Chapter 3 specifically marked and they can be omitted if one is only interested in PRMs.

Retrospectively, in relation to the topic of ”optimal dominant policies” of MDPs, PRMs in the strict meaning include policies of certain SDDP problems for which the ”condition of dominance” is typically infringed on, however only to a slight extent such that a generalization of the concept of ”dominant policies” seems possible.

We now discuss the contents of the chapters in more detail.

1.1 Chapter 2 Contents

In Section 2.1 we introduce the DA model of stochastic dynamic programming with random disturbances and give the corresponding functional equation.

In Section 2.2 a ”certainty equivalence principle” is formulated and also proven in cases of DA models with linear dynamics and quadratic criteria.

Markov decision processes which result from DA models under appropriate assumptions (DA MDPs) are investigated in Section 2.3.

In literature the state space, which is used for DA MDPs, is the cross product set of the origin state space and the disturbance space.

However, such a state space is markedly larger than the original state space.

Moreover, corresponding matrices of transition probabilities would have many zeros, in general. An analogous situation is found in linear programming: the classical transportation problem which can be solved by the Simplex algorithm. Special solution methods for this transportation problem have been developed (for example the ”MODI-method”, refer to [30], Section 2.8.9).

In Section 2.3 we keep the origin state space when modelling DA models as MDPs. In this way special structures of decisions follow.

Here, the corresponding decisions are characterized by a "simple" structure. The transition probability matrices differ by only two elements for corresponding "neighbouring" decisions.

An effect of this structure of decisions is that optimal decisions imply an "almost-partial order" of the states, if the underlying average one-step reward functions do not depend on the decisions.

Thus, the solution of a DA MDP by solving a corresponding parameterized DA MDP in terms of a continuation of the solutions of the parameterized problem arises as one variant for solving DA MDPs, for which the Howard algorithm (policy iteration) is adapted (Section 2.3.4). For this, the underlying internal costs and hence the average one-step reward functions are considered in dependence on one parameter such that these costs do not depend on the decisions for the initial parameter. Then, the adapted Howard algorithm yields a purposeful computation for the solution. Furthermore, under certain additional conditions, this solution method is a greedy algorithm.

Section 2.3.3 includes special considerations of DA MDPs with "distance properties" and "dominant policies".

"Distance properties" can also be found in flow problems, metric task system or k-server problems. In particular, we use the statements of this section for SDDP problems.

The "dominance of Markov chains" can be seen in Daley 68 (see [10]).
We can apply this concept to Markov chains which correspond to policies of MDPs. However, if we want to transfer this concept to the MDPs themselves then convenient properties are also required for the average one-step reward functions (and for the corresponding policies).
If dominant policies should also be optimal, further strong conditions (which contain comparisons of any feasible policies with the dominant policy) are required.
The question which follows is: can we find (useful) MDPs which fulfil all of these conditions?

A certain kind of equipment replacement models with dominant policies can be found in Puterman [31]. However, in these models only two different decisions are possible.

The chance of finding MDPs with more than two decisions which fulfil these conditions is better for MDPs which are based on DA models, due to their decision structures.

Some SDDP problems have optimal dominant policies (Section 4.6.2.2).
For other SDDP problems we will consider the above-mentioned interesting effect in which the conditions of dominance are infringed on, however only to a slight extent.

The state spaces of SDDP problems are inherently finite. Therefore, we will also concentrate our efforts on finite-state models in Chapter 2. Notes on countable-state models can be found in Puterman [31]; more information can be found here at the beginning of Section 2.3.

1.2 Chapter 3 Contents

In Chapter 3 the "Problem of Stochastic Dynamic Distance Optimal Partitioning (SDDP)" is described in detail. Possibilities and methods of its exact or approximate solution are discussed.

A problem in industry, which contains an optimal conversion of moulds, supplied the origin of investigations.
Essentially, SDDP problems include the following practical facts:

- A fixed number of machines is given. (∗)
  (Moulds are also conceivable.)
- Different types of parts can be produced by these machines. For this purpose the machines have to be converted to states, which in accordance with the types of the parts. Costs are incurred. (∗∗)
- The production takes place in successive stages (periods).
- In a single stage, one part (at most) can be produced by one machine.
- At each stage a requirement of parts (of several types) is to be met.

  Initially, probability functions of the requirements are given.

The realizations of the requirements are known at the beginning of the stages (before decisions about conversions of machines have to be made).

· The objective is to minimize the expected cost of the conversions over all stages (or the average expected cost per stage). (To accomplish this we must decide which machine should be converted into which state in each stage.)

Thus, SDDP problems are DA stochastic dynamic programming problems.

More specifically, from a mathematical view point, we could designate this practical problem as a stochastic dynamic transportation problem, since throughout the stages feasible solutions of transportation problems must be determined (see $(**)$). (We have also used this designation in previous papers.)

Here, however designating this problem as a stochastic dynamic distance optimal partitioning problem (SDDP) seems more appropriate. Partitioning means partitions of the number of machines into numbers of machines which are in the same state. The number of machines is therefore constant (see $(*)$).

We will thus use this designation in the future.

(In this way we also emphasize the conceptual distinguishment of the designation of our problem from the typical stochastic dynamic transportation problems, see Arnold [4].) [1]

In this mathematical model, partitions of integers are the "states" of the DA stochastic dynamic programming problems (ordered partitions in general and unordered partitions in the case of certain reduced SDDP problems).

Partitioning the integers as "states" involves the combinatorial aspects of SDDP problems, which can also be observed in "matrices of transition probabilities" and "average one-step reward functions" of SDDP problems, modelled as DA MDPs.

It can therefore, only in Chapter 4 by means of combinatorial consideration, be shown that decisions for feasible states with least square sums of

[1] Further comments in connection with transportation problems and corresponding references can be found in the preface of [22].

their parts are in every case optimal for special SDDP problems.

Partitions of integers as states of DA MDPs require an enormous amount of storage space for the corresponding computer programs.

Furthermore, many transportation problems have to be solved (see (∗∗)) in order to compute "average one-step reward functions" for the SDDP problems, modelled as DA MDPs.

Thus, investigations of inherent characteristic structures of SDDP problems are also important as a basis for heuristics.

Finally, we refer to connections of SDDP problems with other problems in operations research and informatics such as stochastic dynamic facility location problems (refer to [27]) or metric task systems and more specific k-server problems, see [8], Chapter 10 and [5], for instance.

Since the current request, which is to be fulfilled, is known (and without knowing the future requests) k-server problems can also be initially labeled as a certain kind of DA model. Furthermore, distance properties are also assumed for k-server problems. However, on-line algorithms are often the center of attraction for consideration of k-server problems.

In contrast, we assume probability functions for requirements of SDDP problems and consider SDDP problems as stochastic dynamic programming problems with the aim to minimize the expected cost or the average expected cost per stage. Typical characteristics of SDDP problems as stochastic dynamic programming problems, in particular Markov decision process, are worked out.

Furthermore, let us note that we consider a number of machines which are in the same state (in the terms of k-server problems, on the same point), in general, and many machines must convert at the beginning of each equidistant stage.

1.3 Chapter 4 Contents

Partitions-Requirements-Matrices (PRMs) are the main topic of Chapter 4.

If SDDP problems are modelled as DA MDPs, then the matrices of transition probabilities are called "general PRMs".

The strict meaning of PRMs assumes that the costs of converting the machines into different types are identical and the requirements are identically distributed. Then in every case decisions for feasible states with least square sums of their components lead to PRMs (in the strict meaning).

The definition of PRMs (in the strict meaning) includes that PRMs can be initially computed by means of simple enumeration, however a laborious method. In addition, there is a main difficulty to deal with: No formulas are known for most of the elements in PRMs. Due to this lack of formulas, PRMs themselves represent interesting (almost self-evident) combinatorial structures.

Properties which are associated with SDDP problems (modelled as DA MDPs), besides the search for effective methods to compute the elements of PRMs, are in the realm of investigation of PRMs (in the strict meaning) in this chapter.

Thus in Section 4.6 so-called "Poisson equations" are considered. That their solutions are "monotone" is shown in many cases. This means that, in every case, decisions for feasible states with least square sums of their components are optimal for the corresponding SDDP problems.

The above-mentioned SDDP problems, for which the "condition of dominance" is infringed on, however only to a slight extent, are also in this set of SDDP problems.

A more detailed specification of the content of Chapter 4 can be found at the beginning of this chapter.

Chapter 2

DA Stochastic Dynamic Programming with Random Disturbances

It is assumed for many concepts in the theory of stochastic dynamic programming that random disturbances are observed after the decision is made at each stage. (For instance, refer to Bertsekas [7], Schneeweiss [33], Dinkelbach [11].)

We denote problems for which this is assumed as "Decision Before" models (DB models).

Conversely, we call problems where random disturbances are observed before the decision is made at each stage "Decision After" models (DA models).

We began to take notice of DA models with our investigation of **S**tochastic **D**ynamic **D**istance Optimal **P**artitioning (SDDP) problems [1] (see [19], [20], [22]).

In general, not much information exists dealing only with DA modelled problems.
We can find some, however, included in a book by Sebastian and Sieber [34]. Here, situations in which incomplete information is given are described by

[1] In previous papers, SDDP problems were termed stochastic dynamic transportation problems, see also Section 1.2.

means of operators as starting points for further investigations (see [34], 2.7 with $n = 1$).

Dreyfus and Law give an example in relation to certainty equivalence and also an example of a stochastic equipment inspection and replacement model, where some components of the random vector are observed after the decision is made (as usual) but some components are observed before (see [12], pages 189 and 137).

(The k-server problems mentioned at the end of Section 1.3 also show the "DA" property.)

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 (refer here also to the inspection/replacement problem by Dreyfus and Law) is one aspect of the motivation for the further consideration of DA models.

An introduction to the extended content of Chapter 2 has already been given in Section 1.1.

2.1 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 (or controls)

(Questions of measurability are skipped for the most part. In the beginning, let S and A be Borel spaces and let the values of w be elements of a Borel space. Afterward we often assume $S \subseteq \mathbb{Z}^n$ (or $\mathbb{R}^n$) and so on. We will use the same notations for the random vectors and their realizations.)

The above data are written with the subscript t in order to attach the time to the stages t.

Furthermore,

$K_t : S_t \times B_t \times A_t \to \mathbb{R}_+$ stage - cost (or - return) functions

$G_t : S_t \times B_t \times A_t \to S_{t+1}$ transition functions

denote (measurable) functions.

Decision spaces A_t can depend on previous states and disturbances.

We now introduce the basic problem of the DA model:

(DAP):

Let DA models be closed-loop optimization problems (i.e. feedback control, refer to [7], I, page 4 or [27], Section 2.4): More precisely, this 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

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

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

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 also conceivable),

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

(The objective function always exists when $K_t \geq 0$, but it may have the value ∞ without some 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 2.1.1. *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 than in the usual DB models, namely* $x_t \in A_t(s_t, \mathbf{w_t})$.

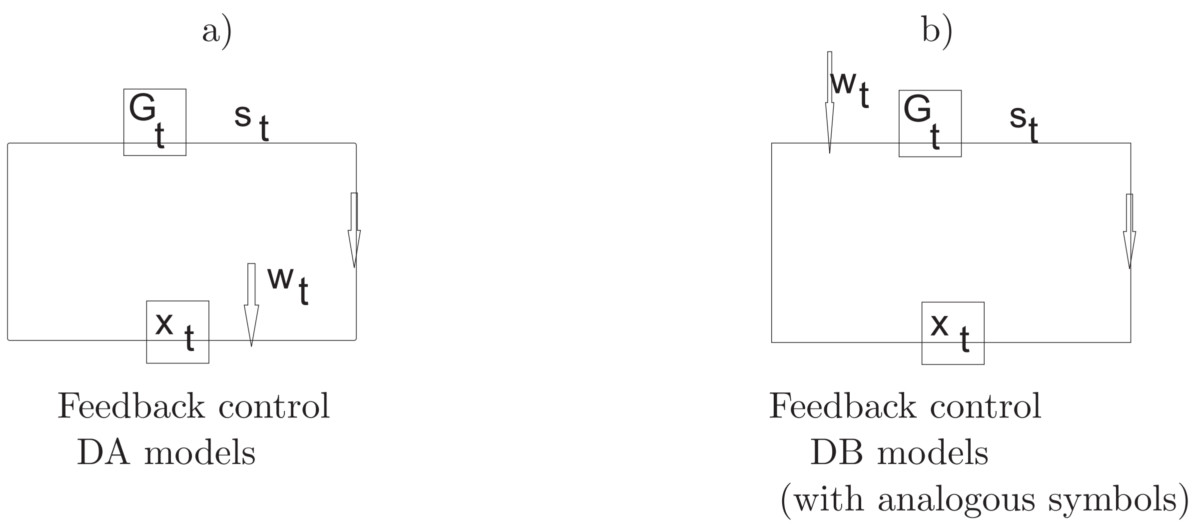

Figure 2.1.1.

Of course DA models are also stochastic dynamic programming problems. 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. The cost of the next stages also depends on $s_{t+1} = G_t(s_t, w_t, x_t)$.

The Optimal Value Function for the Remaining Periods and the Functional Equation

We use $F_t = \{x_t(s_t, w_t), x_{t+1}(s_{t+1}, w_{t+1}), \ldots, x_N(s_N, w_N)\}$, 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 $x_{t'} \in A_{t'}(s_{t'}, w_{t'})\ \forall\ s_{t'} \in S_{t'},\ \forall\ t' \in \{1, ..., N\}$.)

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

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

$$= \min_{F_t} \left(K_t(s_t, w_t, x_t) + \underset{w_{t+1},\ldots,w_N}{E} \left(\sum_{t'=t+1}^{N} K_{t'}(s_{t'}, w_{t'}, x_{t'}) | \overline{w_t} \right) \right) \tag{2.1.1}$$

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

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

for DA models.

We define

$$f_{N+1} \equiv 0. \tag{2.1.2}$$

The functional equation

$$f_t(s_t, \overline{w_t}) = \min_{x_t \in A_t(s_t, w_t)} \left(K_t(s_t, w_t, x_t) + \underset{w_{t+1}}{E} \left(f_{t+1}(s_{t+1}, \overline{w_{t+1}}) | \overline{w_t} \right) \right), \tag{2.1.3}$$

$t = N, \ldots, 1$

follows.

In the case that an optimal policy exists the functional equation can be proved directly by means of mathematical induction (refer also to Sebastian and Sieber [34], general formula (2.188) and the upper remarks on page 147):

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

Step 1.

(beginning of mathematical induction $t = N - 1$)

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

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

(see (2.1.1) for $t = N - 1$)

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

$$= \min_{x_{N-1} \in A_{N-1}(s_{N-1}, w_{N-1})} \Big\{ K_{N-1}(s_{N-1}, w_{N-1}, x_{N-1}) + \min_{x_N \in A_N(s_N, w_N)} \Big(\mathop{E}_{w_N}(K_N(s_N, w_N, x_N) \mid \overline{w_{N-1}}) \Big) \Big\}.$$

(Here $\min\limits_{x_N \in A_N(s_N, w_N)} \ldots$ means, in detail, $\min\limits_{x_N(w_N) \in A_N(s_N, w_N)} \ldots$
$\forall \ w_N \in B_N$.)

We now use the relation $\min\limits_x E\{\phi(x)\} = E\left\{\min\limits_x \phi(x)\right\}$.

$$= \min_{x_{N-1} \in A_{N-1}(s_{N-1}, w_{N-1})} \Big\{ K_{N-1}(s_{N-1}, w_{N-1}, x_{N-1}) + \mathop{E}_{w_N} \Big(\min_{x_N \in A_N(s_N, w_N)} K_N(s_N, w_N, x_N) \mid \overline{w_{N-1}} \Big) \Big\}$$

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

Step $N-t^*$:

Let us now assume

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

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

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

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

$$\mathop{E}_{w_{t^*+1},\ldots,w_N}\big(\sum_{t'=t^*+1}^{N}K_{t'}(s_{t'},w_{t'},x_{t'})\mid\overline{w_{t^*}})\Big) \qquad \text{(see (2.1.1))}$$

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

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

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

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

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

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

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

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

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

For subsequent sections we introduce here:

The "DA Decision Functions" and Additional Definitions
(which are based on DA models)

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

$$\hat{A}_t(s_t,w_t) := \{s' \mid s' = G_t(s_t,w_t,x_t) \text{ with } x_t\in A_t(s_t,w_t)\} \tag{2.1.4}$$

for given $s_t \in S_t$, $w_t\in B_t$,
where $s'\in\hat{A}_t(s_t,w_t)$ are called feasible states,

internal costs

$$\begin{aligned}&\hat{c}_t(s_t,w_t,s') := min\left\{K_t(s_t,w_t,x_t)|x_t : s' = G(s_t,w_t,x_t)\right\}\\ &\text{with } s'\in\hat{A}_t(s_t,w_t)\end{aligned} \tag{2.1.5}$$

and DA decision functions

$$\begin{aligned}\hat{d}_t :\ S_t\times B_t &\to S_{t+1}\\ \text{with}\quad \hat{d}_t(s_t,w_t) &= s'\in\hat{A}_t(s_t\ ,w_t).\end{aligned} \tag{2.1.6}$$

Finally, we use

Definition 2.1.1. *The set of DA decision functions is the set*

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

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

In addition, single maps

$$\left.\begin{array}{l} (s_t, w_t) \to s' \ (by\ \hat{d}) \\ \\ \text{this means } \hat{d}_t(s_t, w_t) = s' \end{array}\right\} \qquad (2.1.7)$$

for $s_t \in S_t$, $w_t \in B_t$ *are called single decisions.*

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

With this in mind Figure 2.1.1 a) can be replaced by

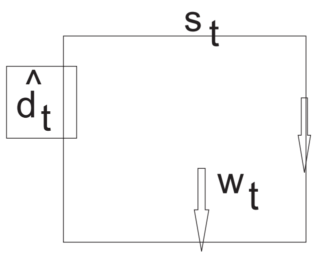

Figure 2.1.2.

We can see that x_t and G_t are combined into $\hat{d}_t$.

(DAP) can then be represented in the following way:

(DAPa):

A policy

$$\{\hat{d}_1(s_1, w_1),\ \hat{d}_2(s_2, w_2),\ \ldots, \hat{d}_N(s_N, w_N)\}$$

is to be found so that

$$\underset{w_2,\ldots,w_N}{E} \left(\sum_{t=1}^{N} \hat{c}_t(s_t, w_t, s_{t+1}) | s_1, w_1 \right) \ \to \ \min$$

subject to the constraints

$$s_t \in S_t,\ t = 2, \cdots, N,$$
$$\hat{d}_t(s_t, w_t) \in \hat{A}_t(s_t, w_t),\ t = 1, \cdots, N,$$
$$s_{t+1} = \hat{d}_t(s_t, w_t),\ t = 1, \ldots, N-1.$$

If (DAPa) exists under the following assumptions, we use the symbol: (**DA$\bar{\text{P}}$a**).

This indicates (DAPa) with

- stationary properties: the sets and functions $B_t, S_t, \hat{A}_t, \hat{d}_t, \hat{c}_t$ are the same at each stage and will be written as $B,\ S$ and so on,
- B and S are finite sets,
- $q(w)(q:\ B \to (0,1))$ denote the probabilities of random disturbances and these $q(\cdot)$ are also the same at every stage.

2.2 The Certainty Equivalence Principle

For many 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 then solve the yielded deterministic problems. The solutions are the same (certainty equivalence principle). We have found a similar statement for DA models.

Let us begin by considering the following example.

Example 2.2.1. *We contemplate the stochastic dynamic programming problems*

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

where $s_1 \in \mathbb{R}$ *or* $s_1 \in \mathbb{R}$ *and* $w_1 \in \mathbb{R}$ *are given*

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

$x_t \in \mathbb{R}.$

Here, $\{w_t\}_{t=1,2,3}$ *is a sequence of independent random disturbances with realizations* $w_t \in \mathbb{R}$.

Since the decision spaces $(A_t(s_t, w_t) =)\mathbb{R}$ *(at each stage) are independent of* w_t*, we can classify 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 of the DB modeled problem is

$$x_N = x_3 = 0$$

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

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

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

The optimal solution of the DA modeled problem is

$$x_N = x_3 = 0$$

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

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

(At the beginning we have calculated this by means of the Bellman-principle, see (2.1.3)*.)*

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

Example 2.2.1 demonstrates the strong relationship between the solutions of the DB and DA models.

We will now generalize the results of the example.

Quadratic-Linear-Problems

Let us assume for (DAP) that

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

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

The dynamic constraints are

$$s_{t+1} = \Phi_t s_t + \Gamma_t x_t + \Pi_t w_t \ \ for \ \ t = 1, \ldots, N \tag{2.2.1}$$

with given matrices Φ_t, Γ_t and Π_t and a given s_1 or given s_1 and w_1. (These symbols are taken from the model in Schneeweiss [33], Section 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 (2.2.1) has the form

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

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. Furthermore, let all matrices V_{xx} which are calculated by means of the backward dynamic programming procedure be positive definite.

Quadratic-linear-problems can be classified as DA models or as DB models with the same data, however $x_t(s_t)$ is used for DB models and $x_t(s_t, \mathbf{w_t})$ for DA models (compare Example 2.2.1).

Theorem 2.2.1. *(Certainty equivalence principle)*
Let a quadratic-linear DB model and a quadratic-linear DA model with the same data be given.

In addition, let

$$x_N = 0,$$

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

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 model in Schneeweiss [33] (see Section 11.3) and they are applied to the DA models here.

The functional equation for this DA problem is

$$f_t(s_t, \overline{w_t}) = \min_{x_t} \Big\{ y_t^T W_{t,yy} y_t + \underset{w_{t+1}}{E} \Big\{ f_{t+1}(s_t, \overline{w_{t+1}}) | \overline{w_t} \Big\} \Big\}$$

$$t = N, \cdots, 1, \qquad (*1)$$

$$f_{N+1} \equiv 0$$

(see (2.1.3)).