Thesis for the Degree of Licentiate of Engineering

Optimization of opportunistic replacement activities in deterministic and stochastic multi-component systems

Niclas Andréasson

CHALMERS | GÖTEBORG UNIVERSITY

Department of Mathematics
Chalmers University of Technology and Göteborg University
Göteborg, Sweden 2004

Optimization of opportunistic replacement activities in deterministic and stochastic multi-component systems
NICLAS ANDRÉASSON

ISSN 0347–2809/No 2004:24
Department of Mathematics
Chalmers University of Technology and Göteborg University
SE-412 96 Göteborg
Sweden
Telephone +46 (0)31–772 1000

Chalmers University of Technology
Göteborg, Sweden 2004

Abstract

The thesis deals with the development and analysis of optimization models for computing optimal replacement schedules for systems comprising deterministic and stochastic parts. The main application is in the maintenance activity of aircraft engines. There is a large fixed cost associated with taking the engine to the workshop, so if a part fails and must be replaced it is important to take the opportunity into consideration to also replace non-failed parts.

Optimization models for systems consisting of only deterministic parts are developed. The real structure of the aircraft engine is taken into account by representing each module of the engine by a directed graph; these graphs are utilized in the formulation of linear integer programming models for complete module based aircraft engines.

The linear integer programming models are then used in order to formulate two-stage stochastic models for systems consisting of both deterministic and stochastic parts.

The convex hull of the set of feasible solutions to one of the linear integer programming models is studied in detail. It is shown that the replacement polytope is full-dimensional under general assumptions. Also, several of the inequalities in the original formulation of the model are facet-defining. They are, however, not sufficient to completely describe the convex hull. A new class of facets is developed through an example.

Contents

Acknowledgements

First of all, I am grateful for the always-present support from my advisors Michael Patriksson and Ann-Brith Strömberg.

At Volvo Aero Corporation I would like to thank Torgny Almgren, Peter Bernsteen, Torbjörn Hillerström, and Fredrik Plyhm. Also, special thanks go to all the people that patiently answered my questions regarding the maintenance activity.

Further, I would like to thank Fredrik Altenstedt for good discussions on dynamic programming.

I would also like to thank all my friends and colleagues at Mathematical Sciences for contributing to an inspiring working atmosphere.

Finally, I want to thank my parents for their encouragement.

This research was supported by NFFP (Nationella Flygtekniska Forskningsprogrammet), grant NFFP485, which is hereby greatly acknowledged.

Göteborg, April 2004

Niclas Andréasson

Chapter 1

Introduction

1.1 The value of optimization in the maintenance activity

A current trend in service workshops in the aircraft industry is to offer the complete undertaking of the maintenance of all engines belonging to the customer. According to such a corresponding contract, by paying a fixed price per year the customer is ensured of a working fleet of engines throughout the contract period. When the maintenance contract has been signed the profit of the workshop from this contract depends, of course, on the least cost at which the necessary maintenance can be performed.

An aircraft engine consists of thousands of parts. Some of the parts are safety-critical, which means that if they fail there will be an engine breakdown, possibly with catastrophic consequences. Therefore, the safety-critical parts have fixed estimated lifetimes, and must be replaced at the very latest when these are reached. Hence we consider the safety-critical parts as deterministic. All the other parts of the engine are considered as stochastic.

When a deterministic lifetime is reached or when a stochastic part fails the engine must be taken out of service and sent to the workshop in order to perform the required replacements. Due to economies of scale (for example, large fixed costs at each maintenance occasion independent of what is replaced), this unpleasant event is at the same time considered as an opportunity for preventive replacements of non-failed stochastic parts and deterministic parts that have not reached their respective lifetimes. It is, however, not at all clear which parts to replace in order to minimize the total maintenance cost over the contract period.

In this thesis we develop maintenance optimization models for the minimization of the total expected cost for having a functioning aircraft engine (consisting of deterministic and stochastic parts) during a finite time period (the contract period). The output from these models are replacement schedules for each maintenance occasion, that is, when the aircraft engine is taken to the service workshop, for whatever reason, the optimization models are used in order to decide which parts should be replaced in order to minimize the total expected maintenance cost during the remaining contract period.

1.2 Outline of the thesis

Chapter 2 describes the maintenance activity at Volvo Aero Corporation site in Trollhättan, Sweden. The discussion is mainly focused on the military maintenance, since the optimization models are mainly intended to support the maintenance of the military RM12 engine. In order to illustrate how an optimization model for the maintenance activity can be used, we describe the model of a hypothetical maintenance situation.

The first application of the optimization models developed will be to the maintenance of the RM12 engine. The structure of the RM12 engine is described in Chapter 3. We show how to represent each module of the engine by a directed graph, and investigate how to compute desired input to the optimization models by computing minimum Steiner trees in these graphs.

In Chapter 4 we develop maintenance optimization models for systems consisting of deterministic parts only. First, a dynamic programming model is presented, and then several linear integer programming models are developed.

In Chapter 5 the maintenance optimization models presented in Chapter 4 are generalized to systems consisting of both deterministic and stochastic parts.

In order to be able to solve real-sized replacement problems we need to develop specialized solution methods. One such solution method is to completely describe the convex hull of the set of feasible solutions with linear inequalities and then use standard linear programming software for solving the problem. In Chapter 6 we study the convex hull of the set of feasible solutions to one of the deterministic optimization models presented in Chapter 4. The model studied is basic in the sense that if we can find a good solution method for solving it, then it can potentially be used in order to construct good solution methods for all the other linear models presented in the Chapters 4 and 5.

In Chapter 7 we present some illustrative examples with the maintenance optimization models from the Chapters 4 and 5. First, we illustrate how the fixed costs to take the engine to the workshop affect the replacement schedule, and then we compare the optimal solutions from two of the stochastic optimization models.

The literature in the area of maintenance of multi-component systems that can be useful when modelling the maintenance of an aircraft engine is surveyed in Chapter 8. The main part of the literature considers systems comprising stochastic parts exclusively, which means that it is extremely hard to compute optimal replacement schedules for systems containing many parts; the policies developed are used to find heuristic solutions only. However, for an aircraft engine the main part (about 75%) of the components considered are deterministic. Hence, our aircraft engine application is in fact a more structured problem than those considered in the literature, which we utilize in the development of the optimization models in this thesis.

1.3 Contributions of the thesis

The thesis contains contributions to both operations research and mathematical programming.

Operations research: The main part of the thesis deals with the development of optimization models for the maintenance of multi-component systems consisting of deterministic and stochastic parts. In the main part of the related literature one assumes that

- the systems consist of stochastic parts only;
- the time horizon is infinite; and
- a policy is used to find a replacement scheme.

Also, from the literature it turns out that it is extremely hard to find an optimal replacement schedule when the number of parts is large, and hence different replacement policies are developed. Such policies reduce the complexity of the problems, but the solutions found are most often not optimal. Further, the literature points out that the case of a finite time horizon is even harder than the infinite time horizon case.

In our aircraft application the time horizon is finite and the number of parts is large (about 50), so if all of the parts had been stochastic it would have been necessary to use replacement policies. However, about 75% of the components considered in an aircraft engine are deterministic, so our problem is more structured than the completely stochastic systems considered in the literature.

Our idea in this thesis is to model the deterministic system with a linear integer programming model, based on the replacement model presented in [50], and then to use this model in order to formulate a two-stage stochastic model for the system with both deterministic and stochastic parts. This approach has not been found in the literature.

Further, we consider the real structure of the aircraft engine by representing each module by a directed graph, and then utilize these graphs in the formulation of the linear integer programming models. This means that we take into account that in order to replace a certain part often other parts must be removed before we can reach it, and hence these parts can be opportunistically replaced at no effective work-cost. This kind of models has not been found in the literature.

In future research it would be interesting to develop appropriate replacement policies for systems consisting of both deterministic and stochastic parts and compare them with the two-stage model developed in this thesis.

Mathematical programming: We make a polyhedral study of the convex hull of the set of feasible solutions to the linear integer programming model for the replacement problem presented in [50]. We refer to this convex hull as the replacement polytope. We show that the replacement polytope is full-dimensional under general assumptions. Also, we show that if the variables associated with the fixed costs in the model are fixed to integers, then the polyhedron arising from the continuous relaxation is integral. The inequalities in the original formulation are studied and we show that several of them are facet-defining. Further, we show that the inequalities in the original formulation are not sufficient in order to completely describe the replacement polytope. By using Chvátal–Gomory rounding we construct a new class of valid inequalities and show that these inequalities in some cases are facet-defining.

Chapter 2

The maintenance activity at Volvo Aero Corporation

2.1 Introduction

This chapter describes the maintenance activity at Volvo Aero Corporation (VAC) site in Trollhättan, Sweden. The discussion is mainly focused on the military maintenance, since the optimization model is mainly intended to support the maintenance of the military RM12 engine. We discuss different forms of maintenance contracts and appropriate objectives and constraints in an optimization model. In order to illustrate how an optimization model for the maintenance activity can be used, we describe the model of a hypothetical maintenance situation.

It should be noted that the description below is based on a large number of interviews with people on different positions at VAC. Since everyone had his/her own personal view of the company it was hard to get a clear picture of the maintenance activity. Hence the presentation might be slightly less concise than ideal. However, the purpose is not to give a completely accurate description of the maintenance activity, but to give some background to the replacement problem that arises when the maintenance is performed, which in turn motivates the optimization models to be set up in the subsequent chapters.

2.2 Military maintenance

VAC manufacture and maintain the RM12 engine, which is the engine of the military aircraft JAS 39 Gripen. JAS is mainly used by the Swedish Air Force (SAF), whose fleet encompasses about 200 RM12 engines. The discussion below is mainly restricted to the relationship between VAC and SAF. The RM12 engine is quite new, so the maintenance activity has not yet begun in earnest. We will, however, describe the maintenance activity as it is planned to work in the future, since this is what the optimization model is intended to support.

A current trend in service workshops in the aircraft industry is to offer the complete undertaking of the maintenance of all engines belonging to the customer. According to such a corresponding contract, by paying a fixed price per year the customer is ensured of a working fleet of engines throughout the contract period. This type of contract demands accurate estimates of all the costs involved in order to be advantageous for both the workshop and the customer. The maintenance agreement between VAC and SAF will be of this kind. When the maintenance contract has been signed the profit of VAC from this contract depends, of course, on the least cost at which the necessary maintenance can be performed. This kind of complete undertaking contracts is the main motive for developing maintenance optimization models.

2.2.1 Properties of the RM12 engine

The RM12 engine consists of several modules, each comprising several components. (The structure of the RM12 engine is discussed in detail in Chapter 3.) The modules can be removed, replaced, and shipped to and from the workshop separately. If this fact is used in a proper manner savings can be achieved in the maintenance activity. (The modular concept is briefly discussed in [53], and it is pointed out that the transportation costs as well as the required number of spare engines can be reduced if the fact that the engine is composed of modules is used properly.) When a component is to be replaced the module that comprises it must be sent to the service workshop. SAF has a stock of replacement modules according to their policy of preparedness.

Some of the parts (about 50) of the RM12 engine are safety-critical. This means that if they fail there will be an engine breakdown, possibly with catastrophic consequences. The lifetimes of the safety-critical parts are estimated by the Department of Solid Mechanics at VAC. The probability that a safety-critical part fails before its estimated lifetime is over is lower than one per mille. (In fact, the lifetime is computed so that this is the case.)

The lifetime is measured in terms of numbers of *cycles*. The accumulated number of cycles of the engine at a certain point in time depends on the load profile during the use of the engine. (When the engine is driven hard the number of cycles accumulates faster than if the engine is driven carefully.) The real lifetime (in hours) of an engine part depends, therefore, on the load profile during use. The real lifetime can be estimated by using the average usage. This type of information is provided by SAF, and is specified before a contract is signed.

Each safety-critical part must be replaced before it reaches its estimated lifetime, and the probability that it fails before this happens is extremely low. Hence the safety-critical parts can be considered as *deterministic*.

Definition 2.1 (deterministic part) *A* deterministic part *is defined as a safety-critical one, and its lifetime is fixed.* ■

Of course, all of the parts of an aircraft engine are stochastic in the sense that we do not know the actual lifetimes of them. However, as we have defined the safety-critical parts as *deterministic* we will only refer to the non-safety-critical parts as *stochastic*.

Definition 2.2 (stochastic part) *The parts of the engine that are not safety-critical are called* stochastic. ■

The lifetimes of the stochastic parts are represented by failure distributions. How to compute such failure distributions is investigated in [124]. With our definition there exist several thousands of stochastic parts. However, only a few of them will be of interest in the optimization models, namely the ones that are expensive and/or require a large amount of work to replace. About 25% of the parts considered in the RM12 engine are stochastic.

2.2.2 The maintenance activity

In each engine there are sensors at different locations that continuously measure, for example, pressure, temperature, number of ignitions, and number of cycles accumulated for each part. For each engine SAF keeps a record of this data. For optimization modelling purposes the most important part of this data is the age of the different parts of the engine. Since the lifetimes of the deterministic parts are given in cycles, which also determines their age, it will be evident from the record when a deterministic part has (nearly) reached its lifetime and must be replaced. Further, a failure of a stochastic part is discovered through inspection or by the monitoring instruments in the aircraft.

A need for maintenance (or replacement) appears when a stochastic part fails or a deterministic part reaches its lifetime. When this happens SAF will place a maintenance order at VAC. The module containing the component that is to be replaced then must be removed from the engine and sent to the service workshop.

When a module arrives at the service workshop at VAC an inspection is performed. At this inspection advanced techniques, such as fibre optics, are used to check the status of the components inside the module. The module is then disassembled to the level required (which depends on the arrival inspection and the record of the module), and the parts removed are cleaned and further inspected.

When the module is disassembled in order to replace a failed part, often other parts must be removed; the structure of the modules of the RM12 engine is discussed in detail in Chapter 3. Hence we get an opportunity to replace parts at no effective work-cost. Further, the fixed costs associated with taking the module to the workshop can also motivate the replacement of non-failed parts.

A report is written, which contains a discussion on the status of the engine as well as a maintenance proposal. This report is sent to SAF and further discussions are conducted until SAF and VAC agree on which components to replace.

The components are replaced by other components, some of which are stored in a warehouse at VAC. The stock consists of both new and used components. A used component is one that has been in service but has been replaced before it reached its lifetime, or a failed component that has been repaired, which means that it has some "life" left. Some of the components are not produced by VAC, but by General Electric, Pratt & Whitney, or Rolls Royce. The delivery times for these components are long and depend, for example, on the supply (in the world) of several exclusive materials.

2.3 Civil maintenance

The most important civil engines for VAC are PW100 and JT8 (the engine of the aircraft types Boeng 727, 737, and DC9). These engines are not manufactured by VAC but by General Electric, Pratt & Whitney, and Rolls Royce. The civil maintenance is controlled by the governmental authorities of the countries where the aircraft operates. These authorities set up rules and regulations for how often the engine has to be taken to the service workshop and what service actions must be taken. The requirements are based on hours of use and the load that the parts have been subject to during use.

The largest airlines manage their own maintenance, so the civil customers of VAC are small and medium sized airline companies. These companies sometimes choose different service workshops from time to time, and at each maintenance occasion they clearly specify what the workshop shall do. In this situation an optimization model such as the ones developed in this thesis is not very useful.

However, as the optimization models that are developed and analyzed in the forthcoming chapters are intended to support a complete undertaking of the maintenance activity during a specified period of time, they can potentially be used in order to construct advantageous offers that hopefully make civil customers to sign this kind of contracts.

2.4 An optimization model for aircraft maintenance

The foundations of the aircraft engine maintenance industry are

- safety;
- availability; and
- cost minimization.

Obviously safety requirements give rise to constraints in the maintenance activity, for example that regular inspections must be performed and that the safety-critical parts must be replaced before they fail. (A moral evaluation of safety and aircraft maintenance is found in [65].) For the RM12 engine safety is regarded by requiring the deterministic parts to be replaced a reasonable amount of time before their computed lifetimes are reached.

The availability of an aircraft can have many meanings, but its most simple definition is the number of flight hours divided by the number of hours at the workshop. The availability can also be conditioned to different time periods. For example, if a customer only uses a certain aircraft during certain weeks, it is required to have an engine available only during these weeks.

An optimization model for the maintenance of an aircraft engine may focus on maximizing availability, minimizing the total cost, or perform a tradeoff between availability and cost. For an aircraft engine it seems more appropriate to minimize cost rather than to maximize availability: if focus is on availability there is a risk that we advocate extending intervals of operation (of the aircraft engine) with weak statistical evidence, which might cause a failure of the engine; on the other hand, if focus is on cost, we can indicate the importance of avoiding failures by proposing that costs must include, for example, insurance premiums, expected costs of compensating those injured and the relatives to those killed, and even the expected loss of failure trade and goodwill.

A general maintenance optimization model can be summarized as follows:

Objectives:
Minimize cost, maximize availability, or optimize a weighted sum of the cost and availability objectives.

Constraints:
Safety requirements, availability requirements, a maximum cost level, etc., to ensure that the solution is a maintenance operation that makes sense.

2.5 The maintenance activity: An example

We give a description of a hypothetical maintenance situation, where we follow the chain of actions and decision-making from the drawing up of the contract to its realization. The purpose is to explain how an optimization model can be useful, and in each situation we describe how an optimization model might support the decision-maker. Observe that the discussion is focused on the engine maintenance, which is assumed to be independent of the maintenance of the other parts of the aircraft. Of course this is a simplification, but it is partly motivated by the fact that often an engine can easily be removed and replaced by another engine. Availability may then be expressed in terms of the number of working replacement engines at each point in time.

2.5.1 Contracts and maintenance costs

Before the maintenance activity can begin a contract must be signed between VAC and the customer. The forms of the contracts differ between customers;

the variations are often associated with how much a customer is willing to pay to get a certain level of availability.

When the availability requirement is established, the minimization of the total cost for maintaining this availability level during the specified time period is of primary interest. The customer presents his availability requirements to several service workshops and the one that offers the lowest price will get the contract. (In order to be allowed to perform aircraft maintenance the workshop must have a certificate from the government authorities; partly for this reason, the quality of the work does not differ very much between different workshops.) In order to be both competitive and ensure a sufficient profit, it is of obvious main interest for the workshop to make a good estimate of the cost for maintaining a fleet of engines at a certain availability level during the contract period.

The cost for the maintenance depends on the type of engine. Each engine consists of several thousands of parts which fail at different times with different probabilities. The parts of the engine are partitioned into deterministic and stochastic ones according to the Definitions 2.1 and 2.2. The need for maintenance (and its cost) is triggered by events that depend, for example, on the

- running profile (civil airplanes require less maintenance than military airplanes, since they are more carefully flown);
- location (a desert climate, or low flight over sea, exposes an engine to stronger wear than, for example, the Swedish inland climate);
- safety requirements (which are due to laws in countries where the airline operates); and
- unplanned occurrences and accidents (for example, birds going through the engine and causing heavy damages).

To determine the maintenance cost is therefore a complex problem. The Utopia is to have an optimization model whose input consists of

- the number of engines in the fleet;
- the costs of all the activities involved when the maintenance is performed, including all the logistics (stock, transportation, etc.);
- the availability requirements;
- all the parameters given above (running profile, location, safety requirements, etc.);
- the lifetimes of the deterministic parts; and
- probability distributions for the lifetimes of the stochastic parts.

The output from such a model would be a replacement scheme that minimizes the expected cost for having a working fleet of engines during the specified time period. A replacement scheme consists of the points in time when the engine is taken to the workshop, and what parts to replace at each such occasion. Unfortunately, to take all of these aspects into consideration is, practically speaking, impossible, not only from a modelling point of view, but also since some of the data that would be required does not exist (for example, the probability distributions for time to failure for some stochastic parts). However, there is data available describing the lifetimes of the deterministic parts, the failure distributions for some of the stochastic parts, the material cost for new parts, and the amount of time required to replace the parts. An optimization model that utilizes this data gives an approximate cost for maintaining an engine during a specific time period.

2.5.2 Performing maintenance

We assume that a contract has been signed which requires the complete undertaking of the maintenance of a fleet of engines at a fixed cost per month. For simplicity we also assume that the service workshop is free to decide when to take a certain engine to the workshop and what to replace when having it there. Of course, this can be done in several ways: The most simple way is to run every engine until a part fails, then take the broken engine to the workshop and replace the failed part. This policy is, however, not allowed according to the safety requirements; each deterministic part has to be replaced, at the latest, when its computed lifetime is reached even if it has not failed. Note, however, that in current practice, an engine will not be taken to the workshop unless at least one stochastic part is failed or a deterministic lifetime has been reached.

When the engine is taken to the workshop in order to replace a failed stochastic part or a deterministic part that has reached its lifetime, it is possible to also replace stochastic parts that have not yet failed and deterministic parts that have not yet reached their computed lifetimes. This is called *opportunistic maintenance* [45] and is mainly motivated by the fact that there is a large fixed cost associated with taking the engine to the workshop which is independent of which parts that are replaced (such a fixed cost is often referred to as "economies of scale").

When the engine is at the workshop the stochastic parts are inspected and their respective conditions are estimated. Based on this estimation and historical data the failure distributions for the stochastic parts can be computed. The failure distributions and the remaining lifetimes of the deterministic parts are main inputs to the optimization model; other inputs are material costs of new parts, work-cost to replace parts, computed lifetimes of new deterministic parts, and failure distributions for new stochastic parts. The optimization model then computes what to replace at the specific maintenance occasion in order to minimize the total expected maintenance cost during the remaining contract period.

Every time a stochastic part fails or a deterministic part reaches its lifetime we perform a reoptimization to compute what to replace at this and later maintenance occasions depending on future scenarios. The longer the engine has been in service the more we know about the stochastic parts (that is, we know if they have failed or not) and thus the failure distributions will change and must be recomputed.

2.5.3 The end of the contract period

When the time period for the contract runs out it is to the advantage of the workshop if the remaining lifetimes of the parts of the engine are as short as possible (at least if a sequel contract has not been signed). It is, for example, unfavourable to replace parts in the engine with new ones just before the contract runs out, since the workshop cannot get any benefits from such a service if the customer chooses another workshop for the next contract period. Of course, the earlier a customer signs a new contract, the better the maintenance activity can be planned. A reasonable policy for the workshop is to give the customer a discount if a new contract is signed with them well before the current contract runs out.

An optimization model for the aircraft maintenance aims to minimize the total expected cost during a given time period, so if it is set to the contract time the model will tend to make use of the parts as well as is possible (that is, so that the total expected cost during the contract period is minimized). This, however, does not necessarily mean that the lifetimes of the parts are as small as possible when the time period runs out.

So far we have not mentioned that the real value of the engine at the end of the contract period actually depends on the remaining lifetimes of the parts. The required status of the engine at the end of the contract period can be given in the contract, and then the status must be considered as a constraint in the optimization model (in fact, this is a type of availability constraint). It is also possible to give the engine a value (for the workshop) at the end of the contract period that depends on the remaining lifetimes of the parts. How this value shall be computed must be given in the contract, so that the workshop can take this into consideration when the maintenance activity is planned.

2.6 Conclusions

We have presented the maintenance activity at Volvo Aero Corporation. The presentation is based on studies and interviews that were performed at VAC during late 2001 and early 2002. By now, two years later, things might have been changed. For example, when the interviews were performed the current maintenance contract with the Swedish Air Force did not have the form of a complete undertaking of the maintenance of their aircraft engines. However, such a contract will be signed during 2005 and hence new information will become available regarding the form of contracts.

Therefore, part of the description of the maintenance activity in this chapter will soon be obsolete, and future research should begin with a new interview study at VAC. Such a study should focus on contract forms, since they are crucial for the development of appropriate optimization models for the maintenance.

Chapter 3

The structure of an aircraft engine

3.1 Introduction

Consider the system consisting of four parts which is illustrated in Figure 3.1.

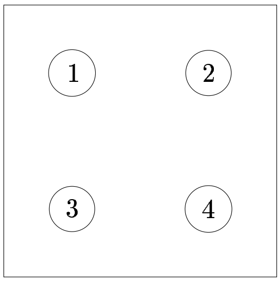

Figure 3.1: A system consisting of four parts.

Assume that we want to remove a specific part of this system. The simplest case is that every part can be removed independently of the others. A part of the system can, however, be *dependent* on other parts in the sense that other parts must be removed before we can reach it. Figure 3.1 tells nothing about possible dependencies between the parts. We can, however, indicate dependences by adding arcs in the figure. For example, to indicate that in order to remove part 2 part 1 must be removed, we can add an arc from part 1 to part 2, as done in Figure 3.2.

In the same manner we add more arcs to indicate further dependencies between the parts. Consider, for example, the system in Figure 3.3. Here, we can interpret the dependencies in two fundamentally different ways. The first interpretation is that in order to remove part 2 *at least one* of the parts 1, 3,

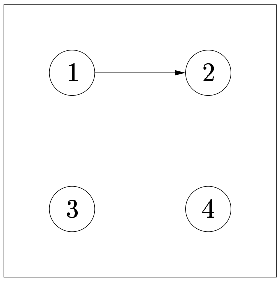

Figure 3.2: A system consisting of four parts, where part 1 must be removed in order to remove part 2.

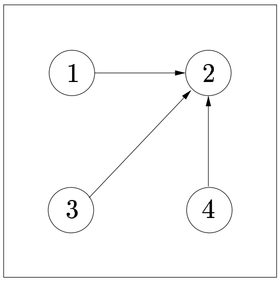

Figure 3.3: A system consisting of four parts dependent of each other.

and 4 must be removed, and the second interpretation is that in order to remove part 2 *all of* the parts 1, 3, and 4 must be removed.

This ambiguity can be eliminated by introducing variables and constraints. To illustrate how to do this we introduce a binary variable for each part, namely,

$$x_i = \begin{cases} 1, & \text{if part } i \text{ is removed,} \\ 0, & \text{otherwise.} \end{cases}$$

Now, if the correct interpretation of Figure 3.3 is that in order to remove part 2 all of the parts 1, 3, and 4 must be removed, then we introduce the constraints

$$x_2 \leq x_1, \tag{3.1a}$$
$$x_2 \leq x_3, \tag{3.1b}$$
$$x_2 \leq x_4. \tag{3.1c}$$

On the other hand, if the correct interpretation is that in order to remove part 2 at least one of the parts 1, 3, and 4 must be removed, then we introduce the constraint

$$x_2 \leq x_1 + x_3 + x_4. \tag{3.2}$$

We have shown how to model and illustrate systems that consist of parts that are dependent of each other. Many real world problems have this structure. We give some examples next.

Example 3.1 (pit mine production [36]) In order to mine ore from a pit at a certain level, in certain cases first one must mine ore according to a 45 degree slope rule. This means that the ore above the ore that are to be mined first must be removed so that there is a free 45 degree slope in each direction (see Figure 3.4).

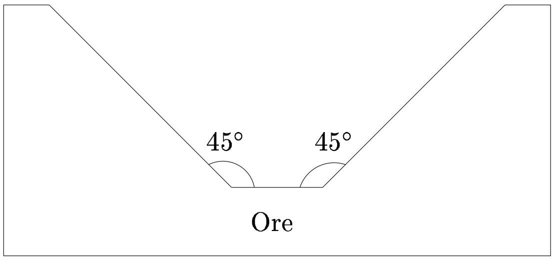

Figure 3.4: Illustration of the 45 degree slope rule.

The pit can be represented by quadratic blocks, as in Figure 3.5.

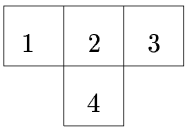

Figure 3.5: Illustration of a pit partitioned into quadratic blocks.

For example, in order to mine ore from block 4 in Figure 3.5 one must first mine the ore in the blocks 1, 2, and 3 according to the 45 degree slope rule. The dependencies are of the type described by the constraints (3.1). ■

Example 3.2 (house building) In order to build the walls of a house, the foundation of the house must first be built, and in order to install windows or to build the roof, the walls must be constructed, etc. This is a classic case of a project planning problem [133]. ■

Example 3.3 (facility location [139]) Given a set of potential depots and a set of clients, if we want to deliver to a client from a certain depot, then that depot first must be opened. ■

Example 3.4 (road maintenance) In order to repair a water leakage, perhaps asphalt and gravel must be removed. ■

Example 3.5 (engine maintenance) In order to remove a specific part of an engine, often other parts have to be removed. This is often the case for aircraft engines. The dependencies are here described by the constraint type (3.2). ■

We get a slightly generalized system if we assume that the system is composed of modules and that each module consists of several parts. This is illustrated in Figure 3.6.

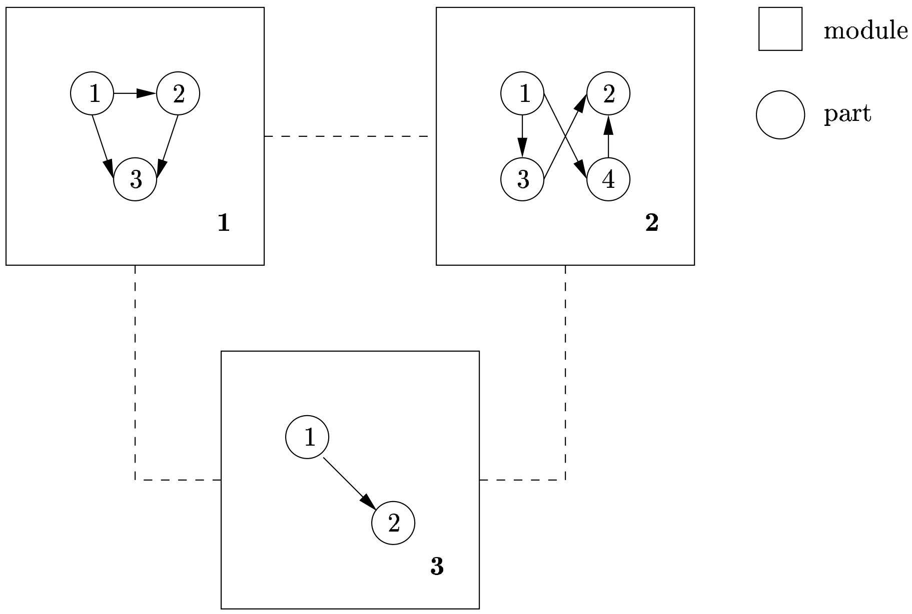

Figure 3.6: Illustration of a system that is composed of modules each consisting of several parts.

In order to remove a part, first the module that contains it must be removed. The dependencies between the modules are different from them between the parts. The system in Figure 3.6 can be divided between the modules according to the dashed lines, that is, between the modules 1 and 2, 2 and 3, and 1 and 3. In order to remove a module one must divide the system so that the module is released. For example, in order to remove module 1 in Figure 3.6 the engine must be divided between the modules 1 and 2, and 1 and 3. Observe that this is generally not the case for the parts. For example, part 2 in module 2 in Figure 3.6 can not be removed by just removing the parts 3 and 4; part 1 must also be removed even if there is no immediate connection between the parts 1 and 2.

Modern aircraft engines, such as the RM12 engine, can be represented by this kind of modular structure.

3.2 The structure of the RM12 engine

In the subsequent chapters we develop optimization models that are intended to support the maintenance of aircraft engines. The first application of these models will be the RM12 engine, and hence we describe the structure of this engine now. In the same manner as in the previous section the RM12 engine can be represented by modules containing parts, and the dependencies between the parts and between the modules can be indicated as was described above.

We note that each module is in fact possible to represent by a directed graph; we will refer to the structure of each module in terms of such a graph. The graph shows how the different parts are related to each other. As was mentioned above, however, it is not always clear how we should interpret the graph. The interpretation for the RM12 engine should be that in order to remove a specific part at least one of the parts with arcs pointing towards it must be removed. For example, in the case of Figure 3.3 at least one of the parts 1, 3, and 4 must be removed in order to remove part 2.

Section 3.2.1 presents the modules of the RM12 engine and the *deterministic* parts in each module. Then a graphical representation of the structure of the deterministic parts of the RM12 engine is given in Section 3.2.2. The data that is included in the graphical model stems from discussions with people at VAC that work with the maintenance of the RM12 engine and from M.Sc. studies by Theander [128] at VAC. The *stochastic* parts are omitted, since for the moment we do not have reliable data for any stochastic part. In the future, however, failure distributions for some of the stochastic parts will be available. These parts should then be included in the graph.

The dynamic programming model and one of the integer linear programming models (Model III) to be set up in Chapter 4 need as input data the cost for man-hours required to replace each specific group of parts inside each module. For the RM12 engine we know the cost for man-hours required to remove a specific part given that we have reached the part, so we can compute the (minimal) man-hour cost required to remove a specific group of parts by solving a Steiner tree problem. How to do this is discussed in detail in Section 3.3.

3.2.1 The components of the RM12 engine

The RM12 engine consists of seven modules: gear box, compressor, fan, after burner, low pressure turbine, high pressure turbine, and burner. The modules are represented by numbered boxes in Figure 3.7 below. The fact that the engine is composed of modules brings availability advantages. Namely, if a specific part fails and must be replaced, we only have to send the module containing it, and not the whole engine, to the workshop. This module can then be replaced by another module (of the same sort), and the engine can be directly put into the aircraft again. For this reason SAF keeps a stock of replacement modules.

We denote each module by a number according to Table 3.1. To remove a module requires a certain amount of time, measured in man-hours. In the last column of Table 3.1 the removal time for each module is given. These removal

Table 3.1: Numbering and removal times of the modules.

No.	Module	Removal time (h)
1	Gear box	4
2	Compressor	1
3	Fan	6
4	After burner	1
5	Low pressure turbine (LPT)	1
6	High pressure turbine (HPT)	1
7	Burner	1

times assume that the modules are independent of each other, so, for example, to remove module 1 (the gear box) takes 4 man-hours and to remove the modules 1, 2, and 3 takes 11 man-hours (4+1+6). This seems to be a simplification of the real structure of the RM12 engine (see Section 3.4), but corresponds to the characteristics of the data received from VAC. The removal times are based on estimations given by the maintenance workers at VAC.

Each module may contain several deterministic parts, as shown in Table 3.2. Here, each part is given a specific number. (Number 4 in the compressor module is not a part but an operation required to take apart the roller. Note that the roller must be taken apart in order to reach the two spools, the disk, and the shaft.) In Figure 3.7 each part is represented by a circle and the dependencies between them are indicated by arcs which should be interpreted according to (3.2).

Estimations of the amount of time required to remove the different parts are given in the right-most column of Table 3.2. These times should be interpreted as the time required to remove a specific part given that we have already reached it according to the graphical representation; see Example 3.6 for a discussion on how to reach a part. It should be noted that this time is independent of the path used to reach the part; again, this seems to be a simplification of the real structure of the RM12 engine (see Section 3.4), but it corresponds to the characteristics of the data received from VAC.

3.2.2 A graphical representation of the RM12 engine

In order to remove a specific part from a module it is most often necessary to remove other parts as well. Based on the real structure of the RM12 engine, Figure 3.7 gives a graphical illustration of the structure of the deterministic parts.

Each arc is associated with a working cost, measured in man-hours. (This cost arises from the removal times according to Table 3.2.) Observe that a characteristic of the data is that every arc that points towards a certain node has the same cost.

Table 3.2: Deterministic parts and their removal times.

No.	Module	No.	Deterministic part	Time (h)
1	Gear box	1	Aggregate	4
2	Compressor	1	Case, fan	4
		2	Case, combustion	4
		3	Roller	6
		4	Roller-split	4
		5	Spool	4
		6	Disk	0
		7	Spool	0
		8	Shaft, front	0
3	Fan	1	Bearing, roller	0
		2	Inlet	0.5
		3	Roller	5.5
		4	Blade, stage 1	0.5
		5	Blade, stage 2	0.5
		6	Blade, stage 3	0.5
		7	Disk, stage 1	0.5
		8	Disk, stage 2	1
		9	Disk, stage 3	0.5
		10	Shaft, after	1
4	After burner	1	Holder	4
		2	Liner	1
5	Low pressure turbine	1	Stator	0.5
		2	Exhaust frame	0
		3	Roller	0.5
		4	Seal segment, HPT	1
		5	Case	1.5
		6	Nozzle segment	1
		7	Shaft, conical	2
		8	Seal, air	1
		9	Disk	0
		10	Blade	0.5
6	High pressure turbine	1	Roller	1.5
		2	Shaft, LPT	1.5
		3	Shaft, HPT	1.5
		4	Plate, cooling, rear	2
		5	Bearing, roller 4	2
		6	Seal, rotating, air	0.5
		7	Blade	0.5
		8	Shaft	1
		9	Plate, cooling, front	2
		10	Disk	0
7	Burner	1	Liner combustion	1
		2	Nozzle segment HPT	1

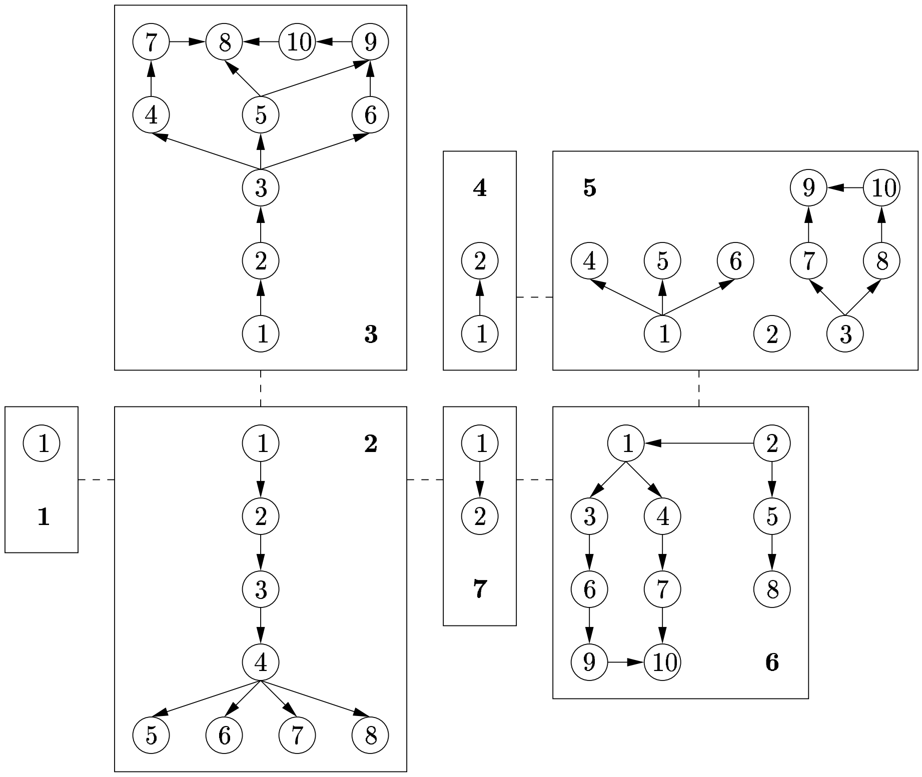

Figure 3.7: The graphical representation of the deterministic parts of the RM12 engine.

Example 3.6 (interpretation of the graph representation) Often there are several ways to reach a specific part. Figure 3.7 shows that to reach part 9 in module 3 one has to first remove part 5 *or* part 6. Hence there are two possible ways to remove part 9 in module 3. One is to remove the parts 1, 2, 3, and 5 (in this order), and finally part 9; the other is to remove the parts 1, 2, 3, and 6 (in this order), and finally part 9. From Table 3.2 we see that the costs associated with the two paths are the same, namely 7 man-hours. ■

3.3 On the computation of the desired input data by using Steiner trees

Some of the mathematical optimization models to be set up in Chapter 4 require man-hour data for every possible combination of parts in each module as input. To compute these data we treat the modules separately. As can be seen in Figure 3.7 each module can be represented as a directed graph. Each arc is associated with a cost (representing man-hours), and we denote the cost of an arc (u, v) from node u to node v by c_{uv}.

Let $D = (V, A)$ be a directed graph where V is the set of nodes (the circles in Figure 3.7) and A is the set of directed arcs (the arcs in Figure 3.7). For each combination of parts inside a module the minimal amount of man-hours required to replace these parts needs to be computed.

Example 3.7 (man-hours required to replace a group of parts) Consider a module for which the structure of its parts is represented by the directed graph in Figure 3.8.

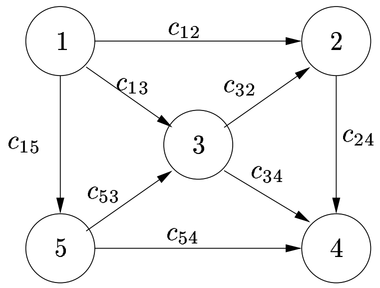

Figure 3.8: The directed graph used in Example 3.7.

Assume that we want to calculate the minimum amount of man-hours required to replace the parts 2 and 4. From the figure we see that in order to remove any part, part 1 must be removed. We call such a part/node an *entrance node*. A feasible solution to this problem is a set of arcs that contains directed paths from the entrance node (i.e., node 1) to each of the nodes 2 and 4. (A feasible solution then indicates paths to removing the parts 2 and 4.) Three examples of feasible solutions are: $\{(1, 2), (1, 3), (3, 4)\}$, $\{(1, 2), (2, 4)\}$, and $\{(1, 5), (5, 3), (3, 2), (5, 4)\}$.

As in this example, often several feasible solutions exist and an optimal feasible solution is one that minimizes the sum of the corresponding arc-costs. The sum of arc-costs for an optimal feasible solution then equals the minimum amount of man-hours required to replace the parts 2 and 4. ■

Remark 3.8 (entrance nodes for aircraft engines) In Example 3.7 we mentioned the term *entrance node*. For the graph representing an aircraft engine there may exist several entrance nodes, but we can always add a fictitious node and arcs from this node to all of the possible entrance nodes. Hence we can assume that there exists only one entrance node. In the mathematical description of the Steiner tree problem we will call the entrance node the *root*. ■

The problem described in Example 3.7 is an instance of a *Steiner tree problem*, which is an $\mathcal{NP}$-hard problem [73]. Different solution methods for the Steiner tree problem have been developed, for example, dynamic programming ([52] and [81]), Lagrangian relaxation ([13] and [14]), dual-ascent methods ([140]), and branch-and-cut methods ([33] and [76]). Surveys on the Steiner tree problem are found in [138], [87], [66], and [67]. Since we only consider small instances such that $|V| \leq 10$, we consider them to be practically solvable by standard linear integer programming software. The linear integer programming model we employ originates from Koch and Martin [76].

In mathematical notation, the Steiner tree problem is given by the following: Let $D = (V, A)$ be a directed graph with arc costs $c \in \mathbb{R}_+^{|A|}$. Consider a node set $T \subseteq V$, a root $r \in T$ (according to Remark 3.8 we can always assume that there exists only one root) and let the set $V(S)$ consist of all the nodes of V that are tail or head of any arc in S. The Steiner tree problem is to find a set $S \subseteq A$ of arcs such that the subgraph $(V(S), S)$ contains a directed path from r to t for all $t \in T \setminus \{r\}$ (such a subgraph $(V(S), S)$ is called a Steiner tree) and such that the sum of the arc-costs for the arcs in the set S is minimized.

To obtain an integer programming formulation we introduce, for each arc $a \in A$, a variable x_a indicating whether arc a is in the solution set ($x_a = 1$) or not ($x_a = 0$). The vector $x \in \{0,1\}^{|A|}$ collects these values into a characteristic vector of the solution. Further, given a set $C \subseteq A$ of arcs and a characteristic vector $x \in \{0,1\}^{|A|}$, $x(C)$ is defined as

$$x(C) = \sum_{a \in C} x_a.$$

Now, for a given node set $T \subseteq V$ and a root $r \in V$ consider the linear integer program

$$\begin{array}{lll} \text{minimize} & c^{\mathrm{T}} x & \qquad\qquad \mathrm{STP}(T, r) \\ \text{subject to} & x(\delta(W)) \geq 1, \quad W \in F(T, r), & \\ & \qquad x \in \{0,1\}^{|A|}, & \end{array}$$

where $F(T, r) = \{\, W \subseteq V \mid W \ni r,\, W \cap T \neq T \,\}$ is the sets of nodes which contain r but not all nodes in T, and $\delta(W) = \{\, (u, v) \in A \mid u \in W,\, v \in V \setminus W \,\}$ is the set of arcs with tail in W and head in the complement of W.

Example 3.9 (linear integer programming formulation of the STP) Consider an instance of $\mathrm{STP}(T, r)$ with (see Figure 3.9)

$$\begin{aligned} V &= \{1, 2, 3, 4\}, \\ A &= \{(1,2), (1,3), (2,3), (2,4), (3,4)\}, \\ T &= \{1, 4\}, \\ r &= 1. \end{aligned}$$

The integer linear programming problem $\mathrm{STP}(T, r)$ contains one constraint for each $W \in F(T, r)$. In this example, $F(T, r) = \{\{1\},\ \{1,2\},\ \{1,3\},\ \{1,2,3\}\}$,

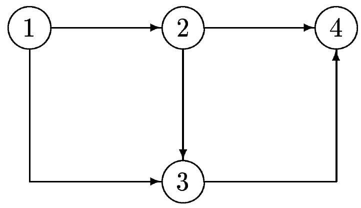

Figure 3.9: Graph for Example 3.9.

so STP(T, r) can be formulated as the problem to

$$
\begin{array}{llllllll}
\text{minimize} & c_{12}x_{12} & +c_{13}x_{13} & +c_{23}x_{23} & +c_{24}x_{24} & +c_{34}x_{34} & & \\
\text{subject to} & x_{12} & +x_{13} & & & & \geq 1, \\
 & & x_{13} & +x_{23} & +x_{24} & & \geq 1, \\
 & x_{12} & & & & +x_{34} & \geq 1, \\
 & & & & x_{24} & +x_{34} & \geq 1, \\
 & \multicolumn{5}{l}{x_{ij} \in \{0,1\}, \quad (i,j) \in A.}
\end{array}
$$

■

Next, we show that a solution to STP(T, r) solves the Steiner tree problem and vice versa.

Proposition 3.10 *The characteristic vector x of a Steiner tree is a feasible solution to* STP(T, r). *Conversely, a feasible solution of* STP(T, r) *describes a Steiner tree.*

Proof. Assume that x is a characteristic vector of a Steiner tree and consider a set $W \subseteq V$ such that $r \in W$ and $W \cap T \neq T$. Then, $x(\delta(W)) \geq 1$ must hold since otherwise there is no path from W to any of the nodes in $(V \setminus W) \cap T$, which contradicts that x is a characteristic vector of a Steiner tree.

On the other hand, suppose that x is a feasible solution to STP(T, r) and $t \in T \setminus \{r\}$. Let $W = V \setminus \{t\}$. We have that $x(\delta(W)) \geq 1$, so there is at least one arc from W to t, say (u, t). Similarly we have that there exists at least one arc from $V \setminus \{t, u\}$ to some of the nodes t and u, which means that there is a path from $V \setminus \{t, u\}$ to t. This argument can then be repeated until a path from r to t is found. Hence x is a characteristic vector of a Steiner tree. ■

We can find the minimal amount of man-hours required to replace the specific groups of parts by solving the integer program STP(T, r) for each subset of nodes $T \subseteq V$. It should be noted, however, that if the total number of nodes is n, then the number of combinations consisting of $m \leq n$ nodes is

$$\binom{n}{m} = \frac{n!}{m!(n-m)!}.$$

Therefore, the total number of combinations becomes

$$\sum_{j=1}^{n} \binom{n}{j}.$$

Clearly, the number of combinations grows rapidly with the number of nodes. This "combinatorial explosion" is illustrated in Table 3.3.

Table 3.3: The number of nodes versus the number of combinations.

# nodes	**# combinations**
3	7
5	31
10	1,023
15	32,767
20	1,048,575

For us this combinatorial explosion is not a serious problem, since we only have to compute the costs of the Steiner trees *once*. These costs are then introduced as constants in the optimization models.

3.4 Conclusions

In this chapter we have discussed the structure of an aircraft engine in general and the RM12 engine in particular. The presentation is based on interviews and studies that were made during 2002. The main resource was unpublished material from an internal study of the structure of the RM12 engine that was performed at VAC during 1997. However, it was not at all clear how this material should be interpreted. Even though several interviews were performed, many questions regarding the structure of the RM12 engine and the amount of man-hours required to remove parts and modules are still open. For example:

- Is it reasonable to assume that the amount of man-hours required to remove a certain part is independent of the path used to reach it?

- Is it reasonable to assume that the amount of man-hours required to remove a collection of modules equals the sum of the amount of man-hours required for removing each module of the collection separately?

- The graphical representation of the structure of the engine shows that there is always at most one arc (u, v) from one node u to another node v indicating that it is possible to go from node u to node v, but is it never possible to go in the reverse direction, that is, to go from node v to node u?

By now, seven years after the internal study was performed, much may have changed. For example, it is reasonable to think that the construction of the engine has been improved, leading to new structures within the modules. Therefore, before the data presented in this chapter can be used in practice a new study should be performed that updates or validates the graphical representation of the structure of the RM12 engine and the amount of man-hours required to remove parts and modules. Also, the structure presented in this chapter only considers the deterministic parts of the engine, but should also incorporate the stochastic parts in order to enable the use of the stochastic optimization models developed in this thesis.

Chapter 4

Deterministic optimization models

4.1 Introduction

In this chapter we develop maintenance optimization models for systems of the type described in Chapter 3, where all parts have deterministic lifetimes. Hence a part must be replaced before or at the time at which its lifetime is reached. By *replacement problem* we refer to the problem of finding an optimal feasible replacement schedule for the system in question. (Accordingly, in Chapter 6 we use the term *replacement polytope* to denote the convex hull of the set of feasible solutions to the replacement problem).

In all the optimization models we present it is assumed that the time horizon is finite, that is, the system we consider will be discarded when it has been in service for a certain amount of time. In optimization models for the maintenance of aircraft engines this is natural, since the contract period is finite. Of course, most often the aircraft engines will not be discarded when the contract period runs out, but for the maintenance workshop the engines can be regarded as such. However, as was discussed at the end of Section 2.5.3, the conditions of the engine-parts at the end of the contract period can be specified in the maintenance contract. This situation can be represented by constraints in the optimization models and is discussed in Section 4.8.

Assume that the finite time horizon is given by τ. Then it is possible to discretize the time into T equal intervals just by dividing τ by T and then consider the points in time:

$$0, \frac{\tau}{T}, \frac{2\tau}{T}, \ldots, \frac{T\tau}{T}.$$

"Time" is then given by $t = 0, 1, \ldots, T$; the corresponding time axis is given in Figure 4.1. The models in the forthcoming sections will all make use of a discretization of time. From now on, when we talk about time we will actually

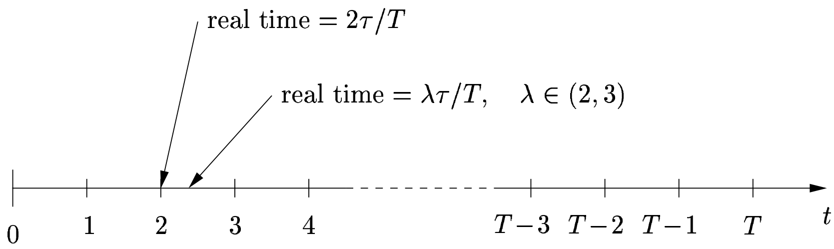

Figure 4.1: Illustration of the discretization of the time.

mean the number of time steps, that is, if we say that a part has lifetime T_1 we will actually mean that it has lifetime $T_1\tau/T$.

Now, consider a system consisting of two deterministic parts, where part 1 has the lifetime 2 and part 2 has the lifetime 3. Further, assume that the time horizon is 5. We illustrate some feasible solutions to the replacement problem in Figure 4.2.

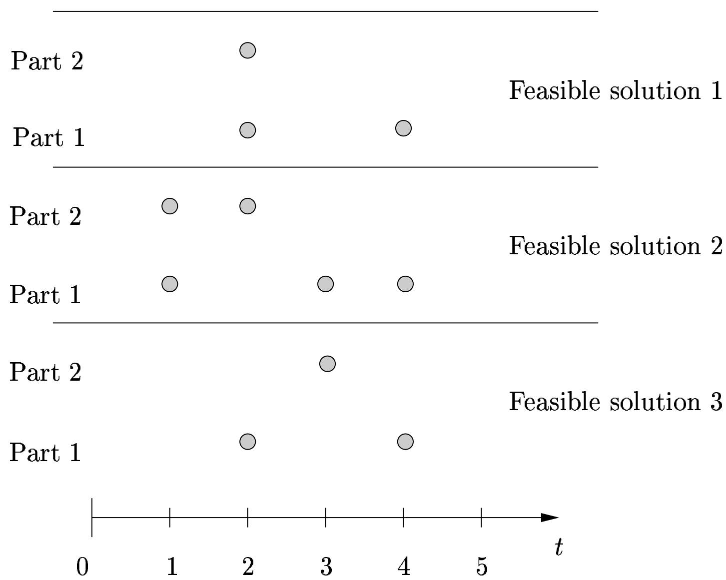

Figure 4.2: Illustration of three feasible solutions for a system with two parts, where the lifetimes are 2 and 3, respectively. (The replacements are illustrated by circles.)

We see that the first and third feasible solutions require three replacements in total, so they are probably to prefer to the second feasible solution which requires five replacements. Further, observe that the first feasible solution only requires two replacement occasions (at the points in time 2 and 4) while the third feasible solution requires three replacement occasions (at the points in time 2, 3, and 4). Typically, there is a large fixed cost associated with each replacement occasion independent of what is actually replaced, and hence the first feasible solution is a good candidate for an optimal solution to the replacement problem.

The replacement problem is a typical example of a *sequential decision process* and hence a natural approach to solve it is to use *dynamic programming*, which is discussed in detail in Section 4.2. The drawback of dynamic programming is that the size of the dynamic programming formulation grows exponentially with the number of parts. Therefore, in order to be able to handle more than about six parts it is necessary to develop other types of models. The dynamic programming formulation, however, has major advantages in modelling of the *stochastic* replacement problem (that is, when some of the parts of the system do not have deterministic lifetimes, but failure distributions), which will be discussed in Chapter 5. In the Sections 4.3, 4.4, and 4.5 we present linear integer programming formulations for systems consisting of one module. One advantage of linear integer programming formulations is that a lot of research has been performed in order to find optimal or near-optimal solutions to large-scale linear integer programming problems. In the Sections 4.6 and 4.7 we develop replacement models for systems composed of several modules, which can be used in order to compute replacement schedules for the RM12 engine. Finally, in Section 4.8 we discuss how to incorporate varying conditions at the start and requirements on the condition of the engine at the end of the contract period into the optimization models.

4.2 A dynamic programming model

Dynamic programming provides a framework for decomposing certain optimization problems into a family of nested subproblems, such that the original problem can be solved by recursion. It was originally developed for optimizing sequential decision processes (to be defined below). The term *dynamic programming* was first used by Richard Bellman who wrote the first book on the subject [16]. Other general texts on dynamic programming are [17], [93], [136], [51], and [47].

In a *discrete sequential decision process* we have a finite time horizon T and consider the points in time $t = 0, 1, \dots, T$. At the time t, the process is in *state* s_t, which is assumed to depend on

- the initial state given by s_0; and
- the *decision variables* $x_0, x_1, \dots, x_{t-1}$ at the times $0, 1, \dots, t-1$.

The following properties hold for discrete sequential decision processes:

1. the contribution to the objective function between the times t and $t+1$ depends only on the state s_t and the decision x_t; and
2. the state s_{t+1} at the time $t+1$ depends only on the state s_t and the decision x_t.

The discrete sequential decision process is illustrated in Figure 4.3. The figure should be interpreted as follows. At time t the process is in state s_t and we make the decision x_t. Depending on the decision x_t the system transforms during the time between t and $t+1$ to the state s_{t+1} at time $t+1$ according to a given transformation function ϕ_t.

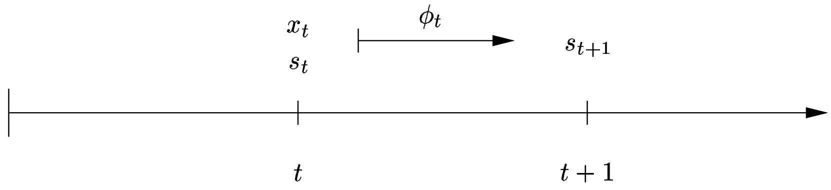

Figure 4.3: Illustration of the discrete sequential decision process.

If we denote by $g_t(s_t, x_t)$ the contribution to the cost function during the time between t and $t+1$ given the state s_t and the decision x_t, then the sequential decision process can formally be described by the optimization problem

$$\begin{aligned} z = \underset{x_0, x_1, \ldots, x_T}{\text{minimum}} \quad & \sum_{t=0}^{T} g_t(s_t, x_t) \\ \text{subject to} \quad & s_{t+1} = \phi_t(s_t, x_t), \quad t = 0, 1, \ldots, T-1, \\ & s_0 \text{ is given.} \end{aligned} \tag{4.1}$$

The domains of the states and the decision variables depend on the particular problem we consider.

To develop a recursive optimization scheme, for $k = T, T-1, \ldots, 1, 0$, let

$$\begin{aligned} z_k(s_k) = \underset{x_k, \ldots, x_T}{\text{minimum}} \quad & \sum_{t=k}^{T} g_t(s_t, x_t) \\ \text{subject to} \quad & s_{t+1} = \phi_t(s_t, x_t), \quad t = k, \ldots, T-1. \end{aligned} \tag{4.2}$$

Hence $z = z_0(s_0)$ for the given value of the initial state s_0.

Proposition 4.1 (recursion in dynamic programming) *If $z_k(s_k)$ is defined as in (4.2), then*

$$\begin{aligned} z_k(s_k) = \underset{x_k}{\text{minimum}} \quad & g_k(s_k, x_k) + z_{k+1}(s_{k+1}) \\ \text{subject to} \quad & s_{k+1} = \phi_k(s_k, x_k). \end{aligned} \tag{4.3}$$

Proof. See [93]. ∎

The recursion given in Proposition 4.1 transforms the original optimization problem (4.1) into a sequence of T subproblems. The kth subproblem (4.3) has only one decision variable and one state constraint, but there is a catch: It must be solved for *all possible values* of s_k. Therefore, the efficiency of solving (4.3) depends on the number of possible states s_k. For the replacement problem the number of values of s_k grows rapidly with the number of parts in a module, so in practice we are not able to handle more than about six parts using dynamic programming.

In the following subsections we develop a dynamic programming model for the replacement problem.

4.2.1 Assumptions and notation for the replacement problem

We consider a system that consists of N deterministic parts. The time horizon is T and the time is discretized so that we only consider the points in time $t = 0, \ldots, T$. Part i has the deterministic lifetime T_i so it can not be in service for more than T_i time steps. If a part is replaced it is always replaced by a new specimen (that is, we do not consider a stock of repaired parts).

There is a cost associated with the replacement of each group of parts. In the case of aircraft engines this cost includes, for example, purchase costs, inspection costs, fixed costs such as administration costs and transportation costs, and so on (see Section 2.2). The group costs also include the work-cost for replacing the group. This cost can be computed by solving Steiner tree problems in the graphical representation of the engine as was described in Section 3.3.

At the beginning of the planning process, that is $t = 0$, the age of each part of the system is known (that is, the parts do not necessarily have to be new at $t = 0$). At the time horizon T the system has a remaining value depending on the condition of the system, which is given in terms of the age of each part. (In the case of maintenance of aircraft engines, the maintenance contract usually states such values for the end of the contract period. Observe that if the maintenance contract instead of giving values of the conditions of the engine gives a certain "threshold condition", that is, each part must have certain "life" left, then we can set the values of the conditions that do not fulfill the threshold values to $-\infty$, and the values of the other conditions to 0.) The objective is to determine a feasible replacement scheme such that the total cost is minimized.

4.2.2 The states in the replacement problem

The state s_t of the system at the time t, is an N-vector given by the age of each specific part, that is,

$$s_t = (\tau_1^t, \ldots, \tau_N^t),$$

where τ_i^t is the age (in number of time steps) of part i. The initial state s_0 of the system is assumed to be given. If a part is new at time $t = 0$ the age is 0. However, for $t \geq 1$ the age of a specific part is always greater than or equal to 1 and less than or equal to the lifetime of the part. Hence, since we have

N parts with lifetimes $T_1, \ldots, T_N$, the number of states at each point in time $t = 1, \ldots, T$ becomes $\prod_{i=1}^{n} T_i$. From this it follows that the number of states dramatically grows with the number of parts.

4.2.3 The decision variables for the replacement problem

The decision variable x_t is an N-vector describing what to replace at time t, that is,

$$x_t = (r_1^t, \ldots, r_N^t),$$

where

$$r_i^t = \begin{cases} 1, & \text{if part } i \text{ is replaced at time } t, \\ 0, & \text{otherwise.} \end{cases}$$

The domain of x_t depends on the state s_t. For example, if part i has been in service for T_i time steps (i.e., it has reached its lifetime) at time t, then every feasible decision x_t involves the replacement of part i.

4.2.4 The cost function for the replacement problem

The cost $g_t(s_t, x_t)$ given the state s_t and the decision x_t is defined by the total cost associated with replacing the group of parts given by x_t. Any economic dependencies between the parts (see Section 3.3) are included in this cost. The cost $g_t(s_t, x_t)$ can also be defined such that it indicates that a certain replacement action x_t is not feasible given s_t by letting $g_t(s_t, x_t) = +\infty$ for each infeasible replacement action.

4.2.5 The transformation function for the replacement problem

Given the state

$$s_t = (\tau_1^t, \ldots, \tau_N^t)$$

and the replacement decision

$$x_t = (r_1^t, \ldots, r_N^t),$$

the transformation function $\phi_t(s_t, x_t)$ is given by

$$\phi_t(s_t, x_t) = (\tau_1^{t+1}, \ldots, \tau_N^{t+1}),$$

where

$$\tau_i^{t+1} = \begin{cases} \tau_i^t + 1, & \text{if } r_i^t = 0, \text{ i.e., part } i \text{ is not replaced at time } t, \\ 1, & \text{if } r_i^t = 1, \text{ i.e., part } i \text{ is replaced at time } t. \end{cases}$$

Example 4.2 (dynamic programming model for the replacement problem) Consider a system consisting of two parts ($N = 2$), where part 1 and 2 have the lifetimes 2 and 3, respectively, and assume that the time horizon is $T = 5$. Then, at each time $t = 1, \ldots, T$, the system states

$$s_t = (\tau_1^t, \tau_2^t) \in \{ (1,1), (1,2), (1,3), (2,1), (2,2), (2,3) \},$$

where τ_i^t denotes the age of part i, $i \in \{1, 2\}$.

At each point in time we can perform one of the following replacement actions:

1. replace nothing [$x_t = (0, 0)$];
2. replace part 1 [$x_t = (1, 0)$];
3. replace part 2 [$x_t = (0, 1)$]; and
4. replace part 1 and 2 [$x_t = (1, 1)$].

However, depending on the state of the system, an action may or may not be feasible. For example, if $s_t = (1, 3)$, then part 2 has reached its lifetime and must be replaced, and hence the only feasible replacement actions are $x_t = (1, 1)$ and $x_t = (0, 1)$ (i.e., replace part 1 and 2, or replace part 2 only). Depending on which replacement action that is chosen the system transforms from state $(1, 3)$ at time t to one of the states $(1, 1)$ and $(2, 1)$ at time $t+1$. Figure 4.4 illustrates the feasible replacement actions depending on the state of the system at time t.

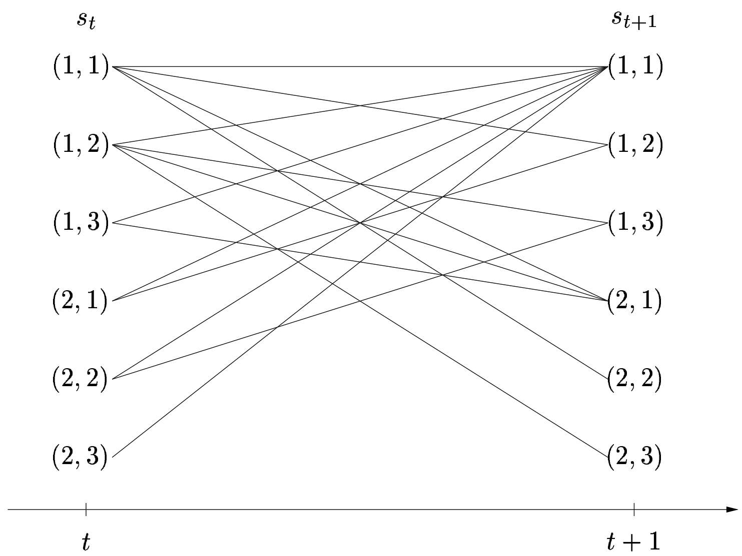

Figure 4.4: The feasible replacement actions in Example 4.2 given the state s_t at time t.

■

4.3 Model I

In this section we give a detailed description of the linear integer programming model for the deterministic replacement problem presented in [50]. All of the linear integer programming models presented in the subsequent sections are generalizations of this model.

4.3.1 Assumptions and notation

Consider a system that consists of N deterministic parts. The time horizon T is finite, and the time is discretized into time steps $t = 0, 1, \ldots, T$. At time $t = 0$ all of the parts of the system are new, and at time $t = T$ the system will be discarded. The lifetime of a new part of type i is T_i and it costs c_i. In addition to the purchase costs of the parts to be exchanged there is a fixed cost d associated with every replacement occasion, independent of how many parts that are replaced. The objective is to minimize the cost of having a working system between the times $t = 0$ and $t = T$.

Remark 4.3 (economic dependencies) The fixed cost d associated with every replacement occasion is an example of what we call an *economic dependence*. For the model presented in this section (Model I) this is the only economic dependence. The models presented in the subsequent sections will have more complex economic dependencies; namely, they consider the man-work costs associated with the replacement of each specific group of parts. ■

4.3.2 The model

In order to formulate a linear integer programming model that solves the replacement problem described in the previous subsection, for every $i = 1, \ldots, N$ and $t = 1, \ldots, T$, we introduce the following variables:

$$x_{it} = \begin{cases} 1, & \text{if part } i \text{ is to be exchanged at time } t, \\ 0, & \text{otherwise,} \end{cases}$$

$$z_t = \begin{cases} 1, & \text{if some of the parts } i = 1, \ldots, N \text{ is to be exchanged at time } t, \\ 0, & \text{otherwise.} \end{cases}$$

To force the replacement of a part before its lifetime is exceeded, some constraints are required. We need to construct constraints considering *lifetimes of the parts*, and *fixed costs*. In order to simplify notation we introduce the set $\mathcal{N} = \{1, \ldots, N\}$.

Lifetimes of the parts: Each part of the system has a fixed lifetime. At the very latest when this is reached, the part has to be exchanged. Consider part $i \in \mathcal{N}$ with the lifetime T_i. This part must be exchanged at least once every T_i time steps yielding the constraints

$$\begin{aligned}
x_{i,1} + x_{i,2} + \cdots + x_{i,T_i} &\geq 1,\\
x_{i,2} + x_{i,3} + \cdots + x_{i,T_i+1} &\geq 1,\\
&\vdots\\
x_{i,T-T_i} + x_{i,T-T_i+1} + \cdots + x_{i,T-1} &\geq 1,
\end{aligned}$$

which can be formulated as

$$\sum_{t=\ell}^{T_i+\ell-1} x_{it} \geq 1, \quad \ell = 1, \ldots, T - T_i, \quad i \in \mathcal{N}.$$

Fixed costs: Every time the replacement of some part $i \in \mathcal{N}$ is triggered, a fixed cost must be paid. That we incur the cost is indicated by the variable z_t having the value 1, leading to the constraints

$$\sum_{i \in \mathcal{N}} x_{it} \leq N z_t, \quad t = 1, \ldots, T, \tag{4.4}$$

or, mathematically tighter,

$$x_{it} \leq z_t, \quad i \in \mathcal{N}, \quad t = 1, \ldots, T. \tag{4.5}$$

Remark 4.4 (strong formulation) The model presented by Dickman et al. in [50] uses the constraints (4.4) as fixed cost constraints. However, as will be discussed further in Chapter 6, the constraints (4.5) are stronger in the sense that the linear programming relaxation of the replacement problem including these constraints has a smaller feasible set than the corresponding relaxation of the replacement problem including the constraints (4.4) instead of (4.5). The same type of constraints arises in the formulation of facility location problems; there the formulation with the constraint type (4.4) is usually called the *weak formulation* and the one with the constraint type (4.5) the *strong formulation* [139]. ■

The purchase cost of a new part of type $i \in \mathcal{N}$ is given by c_i. This cost must be paid at time t if the part is replaced. Further, if some of the parts is replaced at time t then the fixed cost d must be paid. Hence the total cost between the times 0 and T becomes

$$\sum_{t=1}^{T} \left(\sum_{i \in \mathcal{N}} c_i x_{it} + d z_t \right).$$

Using the strong formulation (4.5), we are now ready to describe a complete model for the minimization of the total cost of having a working system between the times 0 and T:

$$\begin{array}{ll} \text{minimize} & \displaystyle\sum_{t=1}^{T}\left(\sum_{i\in\mathcal{N}} c_i x_{it} + d z_t\right) \qquad \text{(Model I)} \\ \text{subject to} & \displaystyle\sum_{t=\ell}^{T_i+\ell-1} x_{it} \geq 1, \quad \ell = 1,\ldots,T-T_i, \quad i\in\mathcal{N}, \\ & x_{it} \leq z_t, \quad t=1,\ldots,T, \quad i\in\mathcal{N}, \\ & x_{it},\, z_t \in \{0,1\}, \quad t=1,\ldots,T, \quad i\in\mathcal{N}. \end{array}$$

Example 4.5 (illustration of Model I) Consider a system consisting of 2 parts, where part 1 has lifetime 2 and part 2 lifetime 3. Further, assume that the time horizon is 5. Then Model I becomes:

$$\begin{array}{ll} \text{minimize} & \displaystyle\sum_{t=1}^{5}(c_1 x_{1t} + c_2 x_{2t} + d z_t) \\ \text{subject to} & x_{11}+x_{12} \geq 1, \\ & x_{12}+x_{13} \geq 1, \\ & x_{13}+x_{14} \geq 1, \\ & x_{21}+x_{22}+x_{23} \geq 1, \\ & x_{22}+x_{23}+x_{24} \geq 1, \\ & x_{11} \leq z_1, \\ & x_{12} \leq z_2, \\ & x_{13} \leq z_3, \\ & x_{14} \leq z_4, \\ & x_{15} \leq z_5, \\ & x_{21} \leq z_1, \\ & x_{22} \leq z_2, \\ & x_{23} \leq z_3, \\ & x_{24} \leq z_4, \\ & x_{25} \leq z_5, \\ & x_{it}, z_t \in \{0,1\}, \quad i=1,2, \quad t=1,\ldots,5. \end{array}$$

Observe the band structure of the upper part of the constraint matrix. This will be utilized in Chapter 6 to establish unimodularity assuming that the z-variables are fixed. ■

4.4 Model II

We introduce more economic dependencies between the parts. Namely, assume that in order to remove a specific part it might be necessary to remove other parts. These dependencies are indicated by arcs in a graphical representation of the system as was described in Chapter 3.

4.4.1 Assumptions and notation

Assume that the system can be represented by a directed graph. Figure 4.5 illustrates such a graphical representation of a system consisting of four parts.

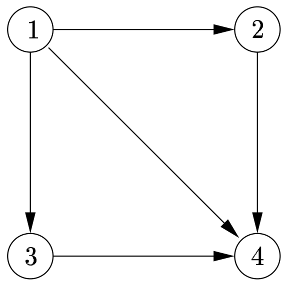

Figure 4.5: A system consisting of four parts dependent of each other.

The arcs tell, for example, that in order to remove part 4 we must remove part 1 or part 2 or part 3, and further if we choose to remove part 4 via part 2 we must also remove part 1. Node 1 is what we call the entrance node (see Chapter 3 and Remark 3.8 in particular). Since each node of the graph is a representation of a specific part we will use the words node/part interchangeably.

As for Model I we consider a system that consists of N deterministic parts. However, here we also assume that the parts are dependent of each other according to the graphical representation denoted by (V, A), where $V = \{1, \ldots, N\}$ is the node set (representing the parts) and A the set of arcs between the nodes (representing the dependencies between the parts). Each arc is represented by an ordered pair of nodes (i, j) denoting that the arc starts in node i and ends in node j. There is a work-cost associated with a removement of part i, denoted by f_i. The time horizon T is finite, and the time is discretized so that $t = 0, 1, \ldots, T$. At time $t = 0$ all of the parts are new and at time $t = T$ the system will be discarded. The lifetime of a new part of type i is T_i and it costs c_i. There is also a fixed cost d associated with every replacement occasion independent of how many parts that are replaced. The objective is to minimize the cost of keeping the system functioning between the times $t = 0$ and $t = T$.

4.4.2 The model

Consider the ith node of the graphical representation (V, A) of the system, and for $i \in V$ define the set

$$\delta(i) = \{\, j \in V \mid (j, i) \in A \,\},$$

that is, $\delta(i)$ is the set of nodes from which there are arcs directed towards i.

In order to formulate a linear integer program, for every $i = 1, \ldots, N$ and $t = 1, \ldots, T$, we introduce the following variables:

$$x_{it} = \begin{cases} 1, & \text{if part } i \text{ is to be exchanged at time } t, \\ 0, & \text{otherwise,} \end{cases}$$

$$y_{it} = \begin{cases} 1, & \text{if part } i \text{ is to be removed at time } t, \\ 0, & \text{otherwise,} \end{cases}$$

$$z_t = \begin{cases} 1, & \text{if some of the parts } i = 1, \ldots, N \text{ is to be exchanged at time } t, \\ 0, & \text{otherwise.} \end{cases}$$

We see that in addition to the variables x_{it} and z_t which were used in Model I, we also have variables y_{it} specifying if a specific part is to be removed or not at a specific point in time.

We introduce the necessary constraints.

Lifetimes of the parts: As for Model I we ensure that the parts cannot be in service for more than their respective lifetimes by the constraints

$$\sum_{t=\ell}^{T_i+\ell-1} x_{it} \geq 1, \quad \ell = 1, \ldots, T - T_i, \quad i \in \mathcal{N}.$$

Also, if a part is to be replaced it must first be removed yielding the constraints

$$x_{it} \leq y_{it}, \quad t = 1, \ldots, T, \quad i \in \mathcal{N}.$$

Work-costs: To trigger the work-cost when replacing a specific part we must introduce constraints that consider the graphical structure of the engine. Indeed, in order to remove part $i \in \mathcal{N}$ we must remove at least one of the parts in the set $\delta(i)$. This gives rise to one constraint for each node and point in time:

$$\sum_{j \in \delta(i)} y_{jt} \geq y_{it}, \quad t = 1, \ldots, T, \quad i \in \mathcal{N}. \tag{4.6}$$

Example 4.6 (illustration of the work-cost constraints in Model II) Consider the system consisting of five parts according to the graph in Figure 4.6. The work-cost constraints (4.6) then become

$$\begin{array}{llll} y_{1t} & + y_{4t} & \geq y_{2t}, & t = 1, \ldots, T, \\ y_{1t} & & \geq y_{3t}, & t = 1, \ldots, T, \\ y_{1t} + y_{2t} + y_{3t} & & \geq y_{4t}, & t = 1, \ldots, T, \\ \quad y_{2t} + y_{3t} + y_{4t} & & \geq y_{5t}, & t = 1, \ldots, T. \end{array}$$

Observe that there are no work-cost constraints corresponding to the entrance node 1. ■

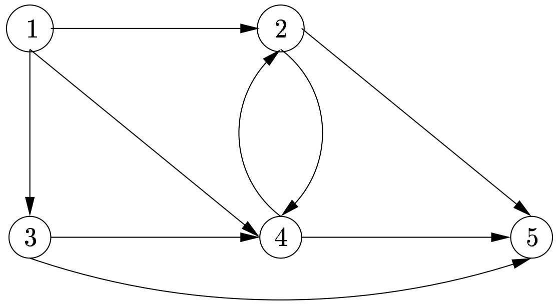

Figure 4.6: The system considered in Example 4.6

We get a stronger formulation by replacing (4.6) with

$$y_{it} \leq \underset{j \in \delta(i)}{\text{maximum}} \{y_{jt}\}, \quad i \in \mathcal{N}, \quad t = 1, \dots, T. \tag{4.7}$$

However, these constraints are not linear, but by observing that they are equivalent to the disjunction that *only one* of the constraints

$$y_{it} \leq y_{jt}, \quad j \in \delta(i),$$

have to be satisfied for each $i \in \mathcal{N}$, and by introducing the binary variables $u_{ij} \in \{0, 1\}$, $j \in \delta(i)$, $i \in \mathcal{N}$, the constraints (4.7) can be reformulated as

$$y_{it} \leq y_{jt} + (1 - u_{ij}), \quad j \in \delta(i), \quad i \in \mathcal{N}, \tag{4.8a}$$

$$\sum_{j \in \delta(i)} u_{ij} \geq 1, \quad i \in \mathcal{N}, \tag{4.8b}$$

$$u_{ij} \in \{0, 1\}, \quad j \in \delta(i), \quad i \in \mathcal{N}. \tag{4.8c}$$

Further, by introducing the variables $v_{it}^j \in \mathbb{R}_+$ for $j \in \delta(i)$, $i \in \mathcal{N}$, $t = 1, \dots, T$, we have the alternative formulation of (4.7) given by

$$v_{it}^j \leq y_{jt}, \quad j \in \delta(i), \quad i \in \mathcal{N}, \quad t = 1, \dots, T, \tag{4.9a}$$

$$\sum_{j \in \delta(i)} v_{it}^j = y_{it}, \quad i \in \mathcal{N}, \quad t = 1, \dots, T, \tag{4.9b}$$

$$v_{it}^j \geq 0, \quad j \in \delta(i), \quad i \in \mathcal{N}, \quad t = 1, \dots, T. \tag{4.9c}$$

Proposition 4.7 *A vector $y \in \{0, 1\}^{N \times T}$ fulfills* (4.7) *if and only if it fulfills* (4.9).

Proof. First suppose that $y \in \{0, 1\}^{N \times T}$ satisfies (4.7). For every $i \in \mathcal{N}$ and $t \in \{1, \dots, T\}$ such that $y_{it} = 1$ there exists a $k \in \delta(i)$ such that $y_{kt} = 1$. Let $v_{it}^k = 1$ and $v_{it}^j = 0$ for all $j \in \delta(i) \setminus \{k\}$. For every $i \in \mathcal{N}$ and $t \in \{1, \dots, T\}$

such that $y_{it} = 0$, let $v_{it}^j = 0$ for all $j \in \delta(i)$. It follows that y together with v is feasible to (4.9).

On the other hand, suppose that $y \in \{0,1\}^{N \times T}$ together with v_{it}^j for $j \in \delta(i)$, $i \in \mathcal{N}$, $t = 1, \ldots, T$, satisfies (4.9). For every $i \in \mathcal{N}$ and $t \in \{1, \ldots, T\}$ such that $y_{it} = 0$ it obviously holds that $y_{it} \leq \max_{j \in \delta(i)} \{y_{jt}\}$. For every $i \in \mathcal{N}$ and $t \in \{1, \ldots, T\}$ such that $y_{it} = 1$, according to (4.9b) it holds that $v_{it}^k > 0$ for at least one $k \in \delta(i)$ and thus, due to (4.9a), $y_{kt} > 0$, so since $y_{kt} \in \{0,1\}$ it must hold that $y_{kt} = 1$. Hence, $\max_{j \in \delta(i)} \{y_{jt}\} = 1 = y_{it}$. Therefore y satisfies (4.7). We are done. ∎

The number of variables and constraints in the formulation (4.9) is larger than in the formulation (4.8), but the advantage of the formulation (4.9) is that it contains no integer variables. Therefore, we have shown that (4.7) can be formulated as a system of *linear* constraints and *continuous* variables, which can be of great importance when developing algorithms for solving the problem. More general results on disjunctive constraints can be found in [11].

Fixed costs: As in Model I we use the fixed cost constraints

$$y_{it} \leq z_t, \quad i \in \mathcal{N}, \quad t = 1, \ldots, T.$$

Objective function: The purchase cost of a new part of type $i \in \mathcal{N}$ is c_i and the work-cost for removing a part of type $i \in \mathcal{N}$ is f_i. Also, the fixed cost that must be paid before any replacement action can occur is d. Hence the total cost becomes

$$\sum_{t=1}^{T} \left(\sum_{i \in \mathcal{N}} (c_i x_{it} + f_i y_{it}) + d z_t \right).$$

The complete model: The complete linear integer program, using the work-cost constraints (4.6), is to

$$\begin{aligned}
\text{minimize} \quad & \sum_{t=1}^{T} \left(\sum_{i \in \mathcal{N}} (c_i x_{it} + f_i y_{it}) + d z_t \right) && \text{(Model II)} \\
\text{subject to} \quad & \sum_{t=\ell}^{T_i+\ell-1} x_{it} \geq 1, \quad \ell = 1, \ldots, T - T_i, \quad i \in \mathcal{N}, \\
& x_{it} \leq y_{it}, \quad i \in \mathcal{N}, \quad t = 1, \ldots, T, \\
& \sum_{j \in \delta(i)} y_{jt} \geq y_{it}, \quad i \in \mathcal{N}, \quad t = 1, \ldots, T, \\
& y_{it} \leq z_t, \quad i \in \mathcal{N}, \quad t = 1, \ldots, T, \\
& x_{it},\, z_t,\, y_{it} \in \{0,1\}, \quad i \in \mathcal{N}, \quad t = 1, \ldots, T.
\end{aligned}$$

4.5 Model III

In Model I we only considered one economic dependence, namely the fixed cost associated with each replacement occasion. In Model II we introduced more economic dependencies, namely, we assumed that the system is possible to be represented by a graph specifying which parts that must be removed in order to remove a specific part. Further, there is a work-cost associated with the removement of each part.

In this section we generalize Model I and Model II further by enabling the introduction of more economic dependencies, namely, we introduce a removement-cost for *each specific group* of parts. These removement-costs can, for example, equal the work-costs from Model II, but they can also indicate other economic dependencies. Actually, since we consider each specific group separately the model presented below enables a great flexibility in the modelling of economic dependencies. The drawback, however, is that the number of variables grows exponentially with the number of parts. Hence, the model presented in this section is useful only when the number of parts is relatively small.

4.5.1 Assumptions and notation

Consider a system that consists of N deterministic parts. The time horizon T is finite, and the time is discretized so that $t = 0, 1, \ldots, T$. At time $t = 0$ all of the parts are new and at time $t = T$ the system will be discarded. The lifetime of a new part of type i is T_i and it costs c_i. Except for the purchase costs of the parts to be exchanged there is a fixed cost d associated with every replacement occasion independent of how many parts that are replaced.

Further, there are costs associated with the replacement of each group of parts. These costs can be the work-costs, but they can also include more general costs as, for example, inspection, administration, and cleaning costs (see Section 2.2). Observe that we can calculate the work-costs associated with the replacement of each group of parts by computing Steiner trees in the graphical representation of the system (see Section 3.3), so the model we develop here is at least as general as Model II. Let $\mathcal{K}$ be the family of all possible combinations (i.e., sets) of parts. For example, if the system consists of 3 parts, that is, $\mathcal{N} = \{1, 2, 3\}$, then

$$\mathcal{K} = \{\,\{\emptyset\}, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\,\}.$$

Denote by g_k the cost associated with the replacement of combination $k \in \mathcal{K}$. Note that $k = \{\emptyset\}$ indicates that nothing is replaced and accordingly $g_\emptyset = 0$. The total number of elements (sets) in $\mathcal{K}$ is

$$\sum_{j=0}^{N} \binom{N}{j}.$$

Obviously, the number of combinations grows rapidly with the number of parts. This "combinatorial explosion" is illustrated in Table 4.1. Therefore, the model

Table 4.1: The number of parts versus the number of elements in $\mathcal{K}$.

N	$\lvert\mathcal{K}\rvert$
3	8
5	32
10	1,024
15	32,768
20	1,048,576

developed in this section is not very useful when the number of parts of the system is large. Observe, however, that in our aircraft application the system is partitioned into smaller subsystems (i.e., modules), and it is possible to use the model for some of them.

The objective is to minimize the cost of having a functioning system between the times $t = 0$ and $t = T$.

4.5.2 The model

We introduce a variable for each specific group $k \in \mathcal{K}$ of parts indicating whether the group is to be replaced at time t or not:

$$w_{kt} = \begin{cases} 1, & \text{if combination } k \in \mathcal{K} \text{ is replaced at time } t, \\ 0, & \text{otherwise.} \end{cases}$$

Further, as in Model I and Model II, for all $i \in \mathcal{N}$ and $t = 1, \dots, T$, we have the following variables:

$$x_{it} = \begin{cases} 1, & \text{if part } i \in \mathcal{N} \text{ is to be exchanged at time } t, \\ 0, & \text{otherwise,} \end{cases}$$

$$z_t = \begin{cases} 1, & \text{if some of the parts } i \in \mathcal{N} \text{ is to be exchanged at time } t, \\ 0, & \text{otherwise.} \end{cases}$$

We are ready to construct the necessary constraints.

Lifetimes of the parts: As for Model I we indicate by the constraints

$$\sum_{t=\ell}^{T_i+\ell-1} x_{it} \geq 1, \quad \ell = 1, \dots, T - T_i, \quad i \in \mathcal{N},$$

that the parts cannot be in service for more than their respective lifetimes.

Costs associated with the replacement of each specific group of parts: If a specific part $i \in \mathcal{N}$ is to be replaced at time step t we must trigger a replacement of some combination $k \in \mathcal{K}$ that contains part i (note that this combination can be the set that consists of part i only), that is,

$$x_{it} = \sum_{k \in \{s \in \mathcal{K} \mid s \ni i\}} w_{kt}, \quad i \in \mathcal{N}, \quad t = 1, \ldots, T.$$

Further, at each time-step it is only possible to replace *one* of the combinations of $\mathcal{K}$, yielding

$$\sum_{k \in \mathcal{K}} w_{kt} = 1, \quad t = 1, \ldots, T.$$

Observe that $w_{\emptyset,t} = 1$ implies that nothing is replaced at time step t.

Fixed costs: We use the same fixed cost constraints as in Model I, that is,

$$x_{it} \le z_t, \quad i \in \mathcal{N}, \quad t = 1, \ldots, T.$$

Objective function: The purchase cost of a new part of type $i \in \mathcal{N}$ is c_i and the fixed cost that must be paid before any replacement action can occur is d. Also, the cost associated with the replacement of combination $k \in \mathcal{K}$ is given by g_k. Hence the total cost becomes

$$\sum_{t=1}^{T} \left(\sum_{i \in \mathcal{N}} c_i x_{it} + \sum_{k \in \mathcal{K}} g_k w_{kt} + d z_t \right).$$

The complete model: We obtain the complete linear integer program:

$$\begin{aligned}
\text{minimize} \quad & \sum_{t=1}^{T} \left(\sum_{i \in \mathcal{N}} c_i x_{it} + \sum_{k \in \mathcal{K}} g_k w_{kt} + d z_t \right) && \text{(Model III)} \\
\text{subject to} \quad & \sum_{t=\ell}^{T_i+\ell-1} x_{it} \ge 1, \quad \ell = 1, \ldots, T - T_i, \quad i \in \mathcal{N}, && \text{(4.10a)} \\
& x_{it} \le z_t, \quad i \in \mathcal{N}, \quad t = 1, \ldots, T, && \text{(4.10b)} \\
& x_{it} = \sum_{k \in \{s \in \mathcal{K} \mid s \ni i\}} w_{kt}, \quad i \in \mathcal{N}, \quad t = 1, \ldots, T, && \text{(4.10c)} \\
& \sum_{k \in \mathcal{K}} w_{kt} = 1, \quad t = 1, \ldots, T, && \text{(4.10d)} \\
& x_{it}, z_t, w_{kt} \in \{0, 1\}, \quad i \in \mathcal{N}, \quad k \in \mathcal{K}, \quad t = 1, \ldots, T. && \text{(4.10e)}
\end{aligned}$$

The purchase costs c_i and the fixed cost d can be included in the costs for each specific group of parts $k \in \mathcal{K}$, that is, the total cost for replacing group k of parts is equal to

$$\tilde{g}_k = \begin{cases} g_k + \sum_{i \in k} c_i + d, & k \in \mathcal{K} \setminus \{\emptyset\}, \\ 0, & k = \{\emptyset\}. \end{cases}$$

Hence, the objective function of Model III can be written as

$$\sum_{t=1}^{T} \sum_{k \in \mathcal{K}} \tilde{g}_k w_{kt}.$$

Without loss of generality we may assume that $\tilde{g}_{k_1} \leq \tilde{g}_{k_2} + \tilde{g}_{k_3}$ for all $k_1, k_2, k_3 \in \mathcal{K}$ such that $k_1 \subseteq k_2 \cup k_3$, from which follows that the constraints (4.10d) are redundant. Further, the variables z_t can be set to 1 (since they are no longer associated with any cost), implying that the constraints (4.10b) become redundant. Then, the constraints (4.10c) are used in order to eliminate the variables x_{it} in the constraints (4.10a). Hence, we can formulate Model III as the *set covering problem* to

$$\begin{array}{ll} \text{minimize} & \displaystyle\sum_{t=1}^{T} \sum_{k \in \mathcal{K}} \tilde{g}_k w_{kt} \\ \text{subject to} & \displaystyle\sum_{t=\ell}^{T_i+\ell-1} \sum_{k \in \{s \in \mathcal{K} \mid s \ni i\}} w_{kt} \geq 1, \quad \ell = 1, \ldots, T - T_i, \quad i \in \mathcal{N}, \\ & \qquad\qquad\qquad\qquad w_{kt} \in \{0, 1\}, \quad k \in \mathcal{K}, \quad t = 1, \ldots, T. \end{array}$$

The set covering problem is an $\mathcal{NP}$-complete problem, but a lot of research has been performed that has resulted in several solution methods. Surveys on the set covering problem can be found in [59] and [115].

4.6 A replacement model for systems composed of modules with parts

We saw in Chapter 3 that aircraft engines are composed of several modules. So far we have just considered systems consisting of *one* module. In this section we present an optimization model for replacement operations in systems composed of several modules, each of which comprises a number of parts. We develop such a model by generalizing Model II. Similar modifications can be made in order to generalize Model I and Model III to module based systems. We generously allow for economic dependencies between the modules, since we consider each specific group of modules separately.

4.6.1 Assumptions and notation

We consider a system that is composed of M modules. Let $\mathcal{M} = \{1, \ldots, M\}$ be the set of modules. Module $m \in \mathcal{M}$ comprises N^m deterministic parts. Let $\mathcal{N}^m = \{1, \ldots, N^m\}$ be the set of parts in module $m \in \mathcal{M}$ (throughout the section we will use the superscript m to indicate that constants, sets and variables correspond to module $m \in \mathcal{M}$). Also, let $\mathcal{P}$ be the family of all possible combinations of modules.

Example 4.8 (illustration of a system composed of modules) The notation of the set of modules $\mathcal{M}$ and the set of parts $\mathcal{N}^m$ for $m \in \mathcal{M}$ is illustrated in Figure 4.7.

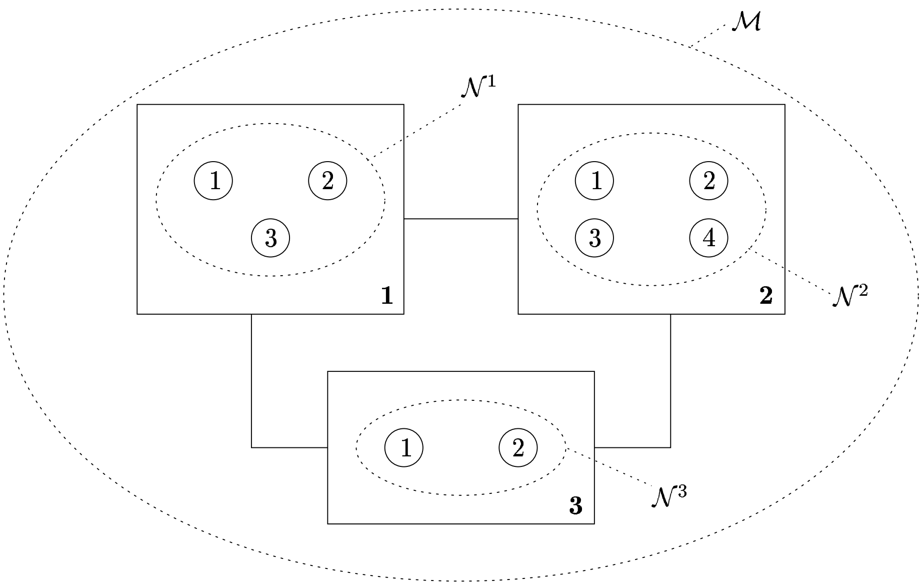

Figure 4.7: A system composed of modules.

The system consists of 3 modules, that is, $\mathcal{M} = \{1, 2, 3\}$. Module 1 comprises 3 parts, module 2 comprises 4 parts and module 3 comprises 2 parts, yielding

$$\begin{aligned} \mathcal{N}^1 &= \{1, 2, 3\}, \\ \mathcal{N}^2 &= \{1, 2, 3, 4\}, \\ \mathcal{N}^3 &= \{1, 2\}. \end{aligned}$$

The family $\mathcal{P}$ of combinations of modules equals

$$\mathcal{P} = \{\, \{\emptyset\}, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\} \,\}.$$

■

The lifetime of a new part of type $i \in \mathcal{N}^m$ is denoted by T_i^m and its purchase cost is c_i^m. In order to remove a specific part $i \in \mathcal{N}^m$ the module $m \in \mathcal{M}$ that comprises it must first be removed. Actually, some combination $p \in \mathcal{P}$ of modules such that $m \in p$ must be removed (it is of course possible to only remove module m, since $\{m\} \in \mathcal{P}$). The cost associated with the removement of combination $p \in \mathcal{P}$ of modules is denoted by d_p. (Note that the costs d_p represent the equivalent of the fixed cost d in the models presented above, where we studied systems comprising one module only.)

Further, let the assumptions and the notation introduced for Model II in Section 4.4.1 hold for each specific module of the system, that is, each module $m \in \mathcal{M}$ can be represented by a directed graph (V^m, A^m), where the set of nodes V^m equals the set of parts $\mathcal{N}^m$ and the arcs A^m indicate how the parts in module $m \in \mathcal{M}$ are connected. The work-cost associated with the removement of part $i \in \mathcal{N}^m$ is denoted by f_i^m.

The time horizon T is finite, and the time is discretized so that $t = 0, 1, \ldots, T$. At time $t = 0$ all of the parts are new and at time $t = T$ the system will be discarded. The objective is to minimize the cost of having a working system between the times $t = 0$ and $t = T$.

4.6.2 The model

Consider the ith node of the graphical representation (V^m, A^m) of the module $m \in \mathcal{M}$. Analogously to the derivation of Model II, for $i \in V^m$ we define

$$\delta^m(i) = \{\, j \in V^m \mid (j, i) \in A^m \,\},$$

that is, $\delta^m(i)$ is the set of nodes in module $m \in \mathcal{M}$ from which there are arcs in A^m directed towards $i \in V^m$.

For all $i \in \mathcal{N}^m$, $m \in \mathcal{M}$, $t = 1, \ldots, T$, and $p \in \mathcal{P}$ we introduce the following variables:

$$x_{it}^m = \begin{cases} 1, & \text{if part } i \in \mathcal{N}^m \text{ is replaced at time } t, \\ 0, & \text{otherwise,} \end{cases}$$

$$y_{it}^m = \begin{cases} 1, & \text{if part } i \in \mathcal{N}^m \text{ is removed at time } t, \\ 0, & \text{otherwise,} \end{cases}$$

$$z_{pt} = \begin{cases} 1, & \text{if module combination } p \in \mathcal{P} \text{ is removed at time } t, \\ 0, & \text{otherwise.} \end{cases}$$

We are ready to state the necessary constraints.

Lifetimes of the parts: We indicate that the parts cannot be in service for more than their respective lifetimes by the constraints

$$\sum_{t=\ell}^{T_i^m+\ell-1} x_{it}^m \geq 1, \quad \ell = 1, \ldots, T - T_i^m, \quad i \in \mathcal{N}^m, \quad m \in \mathcal{M}. \tag{4.11}$$

Also, if a part is to be replaced it must first be removed, which yields the constraints

$$x_{it}^m \leq y_{it}^m, \quad i \in \mathcal{N}^m, \quad m \in \mathcal{M}, \quad t = 1, \ldots, T. \tag{4.12}$$

Fixed costs associated with the removements of the modules: In order to remove a specific part $i \in \mathcal{N}^m$ inside module $m \in \mathcal{M}$ we must first remove some group $p \in \mathcal{P}$ of modules that contains module m, that is, $p \ni m$. We indicate this by the constraints

$$y_{it}^m \leq \sum_{p \in \{s \in \mathcal{P} | s \ni m\}} z_{pt}, \quad i \in \mathcal{N}^m, \quad m \in \mathcal{M}, \quad t = 1, \ldots, T.$$

Work-costs: To trigger the work-cost when replacing a specific group of parts we must include constraints that consider the graphical structure of each specific module of the system. Indeed, in order to remove part $i \in \mathcal{N}^m$ we must remove at least one of the parts in the set $\delta^m(i)$. This yields one constraint for each node inside each module and each point in time:

$$\sum_{j \in \delta^m(i)} y_{jt}^m \geq y_{it}^m, \quad i \in \mathcal{N}^m, \quad m \in \mathcal{M}, \quad t = 1, \ldots, T. \tag{4.13}$$

Objective function: The purchase cost of a new part of type $i \in \mathcal{N}^m$ is c_i^m and the work-cost for removing a part of type $i \in \mathcal{N}^m$ is f_i^m. Also, the fixed cost associated with the removement of the combination $p \in \mathcal{P}$ of the modules is given by d_p. Hence the total cost is

$$\sum_{t=1}^{T} \left(\sum_{m \in \mathcal{M}} \sum_{i \in \mathcal{N}^m} (c_i^m x_{it}^m + f_i^m y_{it}^m) + \sum_{p \in \mathcal{P}} d_p z_{pt} \right).$$

The complete model: We obtain the complete linear integer program:

$$\begin{aligned}
\text{minimize} \quad & \sum_{t=1}^{T} \left(\sum_{m \in \mathcal{M}} \sum_{i \in \mathcal{N}^m} (c_i^m x_{it}^m + f_i^m y_{it}^m) + \sum_{p \in \mathcal{P}} d_p z_{pt} \right) \\
\text{subject to} \quad & \sum_{t=\ell}^{T_i^m + \ell - 1} x_{it}^m \geq 1, \quad \ell = 1, \ldots, T - T_i^m, \quad i \in \mathcal{N}^m, \quad m \in \mathcal{M}, \\
& x_{it}^m \leq y_{it}^m, \quad i \in \mathcal{N}^m, \quad m \in \mathcal{M}, \quad t = 1, \ldots, T, \\
& y_{it}^m \leq \sum_{p \in \{s \in \mathcal{P} | s \ni m\}} z_{pt}, \quad \begin{cases} i \in \mathcal{N}^m, \quad m \in \mathcal{M}, \\ t = 1, \ldots, T, \end{cases} \\
& \sum_{j \in \delta^m(i)} y_{jt}^m \geq y_{it}^m, \quad i \in \mathcal{N}^m, \quad m \in \mathcal{M}, \quad t = 1, \ldots, T, \\
& x_{it}^m, z_{pt}, y_{it}^m \in \{0, 1\}, \quad \begin{cases} i \in \mathcal{N}^m, \quad m \in \mathcal{M}, \quad p \in \mathcal{P}, \\ t = 1, \ldots, T. \end{cases}
\end{aligned}$$

4.7 A replacement model suitable for systems composed of large numbers of modules

If the number of modules is large the optimization model presented in the previous section is not very useful, since the family $\mathcal{P}$ of combinations of modules becomes very large. In order to find a useful optimization model for a general module-based system (where the number of modules can be large) we cannot consider each combination of modules separately. In this section we present an optimization model based on some assumptions regarding the economic dependencies between the modules such that the combinatorial explosion is avoided.

4.7.1 Assumptions and notation

We consider a system that is composed of M modules. Let $\mathcal{M} = \{1, \dots, M\}$ be the set of modules. Module $m \in \mathcal{M}$ comprises N^m deterministic parts. Let $\mathcal{N}^m = \{1, \dots, N^m\}$ be the set of parts in module $m \in \mathcal{M}$.

The lifetime of a new part of type $i \in \mathcal{N}^m$ is denoted by T_i^m and the purchase cost is c_i^m. In order to remove a specific part $i \in \mathcal{N}^m$ the module $m \in \mathcal{M}$ that comprises it must first be removed. We assume that there is a fixed cost, denoted by d, corresponding to the removement of any module (independent of which or how many modules that are actually removed). Further, there is an individual cost, denoted by e^m, associated with the removement of each specific module $m \in \mathcal{M}$.

Each module $m \in \mathcal{M}$ can be represented by a directed graph (V^m, A^m), where the set of nodes V^m equals the set of parts $\mathcal{N}^m$ and the arcs A^m indicate how the parts inside module $m \in \mathcal{M}$ are connected. The work-cost associated with the removement of part $i \in \mathcal{N}^m$ is denoted by f_i^m.

The time horizon T is finite, and the time is discretized so that $t = 0, 1, \dots, T$. At time $t = 0$ all of the parts are new and at time $t = T$ the system will be discarded. The objective is to minimize the cost of having a functioning system between the times $t = 0$ and $t = T$.

4.7.2 The model

For all $i \in \mathcal{N}$, $m \in \mathcal{M}$, and $t = 1, \dots, T$ we introduce the following variables:

$$x_{it}^m = \begin{cases} 1, & \text{if part } i \in \mathcal{N}^m \text{ is replaced at time } t, \\ 0, & \text{otherwise,} \end{cases}$$

$$y_{it}^m = \begin{cases} 1, & \text{if part } i \in \mathcal{N}^m \text{ is removed at time } t, \\ 0, & \text{otherwise,} \end{cases}$$

$$z_t^m = \begin{cases} 1, & \text{if module } m \in \mathcal{M} \text{ is removed at time } t, \\ 0, & \text{otherwise,} \end{cases}$$

$$w_t = \begin{cases} 1, & \text{if any of the modules is removed at time } t, \\ 0, & \text{otherwise.} \end{cases}$$

Lifetimes of the parts: In order to consider the lifetimes of the parts we employ the constraints (4.11) and (4.12) from the model in Section 4.6.

Work-costs: We employ the work-cost constraints (4.13) from the model in Section 4.6.

Constraints associated with the modules: In order to replace a part $i \in \mathcal{N}^m$ in a specific module $m \in \mathcal{M}$, this module must be removed, yielding the constraints

$$x_{it}^m \leq z_t^m, \quad i \in \mathcal{N}^m, \quad m \in \mathcal{M}, \quad t = 1, \ldots, T.$$

Also, if any of the modules is removed, this is indicated by the variable w_t, which yields the constraints

$$z_t^m \leq w_t, \quad m \in \mathcal{M}, \quad t = 1, \ldots, T.$$

Objective function: The purchase cost of a new part of type $i \in \mathcal{N}^m$ is c_i^m and the work-cost for removing a part of type $i \in \mathcal{N}^m$ is f_i^m. Also, the fixed cost associated with the removement of any module is d, and the individual removement cost is e_m for each specific module $m \in \mathcal{M}$. Hence the total cost is

$$\sum_{t=1}^{T} \left(\sum_{m \in \mathcal{M}} \left(\sum_{i \in \mathcal{N}^m} (c_i^m x_{it}^m + f_i^m y_{it}^m) + e^m z_t^m \right) + d w_t \right).$$

The complete model: We obtain the complete linear integer program:

$$\begin{aligned}
\text{minimize} \quad & \sum_{t=1}^{T} \left(\sum_{m \in \mathcal{M}} \left(\sum_{i \in \mathcal{N}^m} (c_i^m x_{it}^m + f_i^m y_{it}^m) + e^m z_t^m \right) + d w_t \right) \\
\text{subject to} \quad & \sum_{t=\ell}^{T_i^m + \ell - 1} x_{it}^m \geq 1, \quad \ell = 1, \ldots, T - T_i^m, \quad i \in \mathcal{N}^m, \quad m \in \mathcal{M}, \\
& x_{it}^m \leq y_{it}^m, \quad i \in \mathcal{N}^m, \quad m \in \mathcal{M}, \quad t = 1, \ldots, T, \\
& x_{it}^m \leq z_t^m, \quad i \in \mathcal{N}^m, \quad m \in \mathcal{M}, \quad t = 1, \ldots, T, \\
& z_t^m \leq w_t, \quad m \in \mathcal{M}, \quad t = 1, \ldots, T, \\
& \sum_{j \in \delta^m(i)} y_{jt}^m \geq y_{it}^m, \quad i \in \mathcal{N}^m, \quad m \in \mathcal{M}, \quad t = 1, \ldots, T, \\
& x_{it}^m, y_{it}^m, z_t^m, w_t \in \{0, 1\}, \quad \begin{cases} i \in \mathcal{N}^m, & m \in \mathcal{M}, \\ t = 1, \ldots, T. \end{cases}
\end{aligned}$$

This model is well adapted to the structure of the RM12 engine (or, more accurately, to the data that is available for the RM12 engine; see Section 3.2.2).

4.8 Conditions at the start and at end of the contract period

The linear integer programming models presented in the Sections 4.3–4.7 assume that the parts are new at the start, $t = 0$, and that the system is discarded at the end, $t = T$. However, as discussed in Section 2.5.3 a maintenance contract may include that the condition of the engine (represented by the remaining lifetimes of the parts) at the end of the contract period must exceed some threshold value. Also, the engine must not necessarily be new at the beginning of the contract period.

In this section we show how to introduce constraints in the linear models to handle varying start conditions and condition requirements at the end of the contract period. We use the same notation as for Model I.

4.8.1 Varying start conditions

Assume that the start conditions of the engine are given by $\tilde{T}_i$ for $i \in \mathcal{N}$, that is, the remaining lifetime of part $i \in \mathcal{N}$ at time $t = 0$ is $\tilde{T}_i \leq T_i$. In the case when all of the parts are new at $t = 0$ we need not consider the possibility to replace some of the parts at time $t = 0$, but in the case of varying start conditions it can in fact be necessary to replace a part at time $t = 0$ (if $\tilde{T}_i = 0$ for some $i \in \mathcal{N}$). Now, for each part $i \in \mathcal{N}$, we must force a replacement before or at time $t = \tilde{T}_i$ which yields the constraints

$$\begin{cases} \sum_{t=1}^{\tilde{T}_i} x_{it} \geq 1, & \text{if } \tilde{T}_i \geq 1, \\ x_{i0} = 1, & \text{if } \tilde{T}_i = 0, \end{cases} \quad i \in \mathcal{N}.$$

4.8.2 Varying condition requirements at the end of the contract period

Assume that the condition requirements at the end of the contract period are given by $\bar{T}_i$ for $i \in \mathcal{N}$, that is, the remaining lifetime for part $i \in \mathcal{N}$ at time $t = T$ is required to be greater than or equal to $\bar{T}_i \leq T_i$. Then we must replace part $i \in \mathcal{N}$ at or after time $t = T - (T_i - \bar{T}_i)$ [note that if $\bar{T}_i = T_i$ the part must be replaced at time $t = T$], yielding the constraints

$$\sum_{t=T-(T_i-\bar{T}_i)}^{T} x_{it} \geq 1, \quad i \in \mathcal{N}.$$

Observe that in the dynamic programming model we could also handle *varying valuation* of the states at the end of the contract period. However, how to do this in the linear integer programming models is not straightforward and will not be treated here.

Chapter 5

Stochastic optimization models

5.1 Introduction

In the preceding chapter we developed replacement optimization models for systems consisting of deterministic parts. In real world applications it is often the case that some of the parts of the system considered are not deterministic but *stochastic* (see Definition 2.2). Because the lifetimes of the stochastic parts are represented by *distribution functions* such parts cannot be correctly represented in the deterministic models previously presented. Often, however, the stochastic parts are crucial for the total maintenance cost. Roughly 30% of all the maintenance occasions for the RM12 engine are due to a stochastic part having failed. Since the fixed cost associated with each replacement occasion is large, the stochastic parts will heavily affect the total maintenance cost. Therefore, in order to construct an optimization model that is useful in practice it is necessary to also take the stochastic parts into account.

In Section 5.2 we present the stochastic replacement problem from the point of view of aircraft maintenance. Then, in Section 5.3, we discuss some simple ideas on how to modify the linear deterministic optimization models from Chapter 4 in order to take stochastic parts into account. These methods are heuristic in the sense that they do not necessarily minimize the total expected cost for having a working engine during the given time period.

The most straightforward method for the computation of an optimal replacement scheme (that is, one that minimizes the expected total cost) for a system consisting of stochastic and deterministic parts is to use dynamic programming. We present a stochastic dynamic programming formulation of the replacement problem in Section 5.4. As we saw in Section 4.2 a main drawback of dynamic programming is that the number of states grows rapidly with the number of parts, which means that the dynamic programming model can only be used when the number of parts is small.

An alternative to stochastic dynamic programming is to use multi-stage stochastic programming models. In such models the randomness is taken into account through a finite number of scenarios, which are used in order to generalize the deterministic optimization problem that we get when the actual outcome (scenario) is given. The replacement problem is, however, not well adapted for stochastic multi-stage programming since the probability that a certain outcome occurs from one time, t, to the next, $t + 1$, depends on the replacement decision at time t. Hence the multi-stage model becomes *nonlinear* which is not desirable from a computational point of view. Instead, we only consider the first stage decision (that is, what to replace at the current replacement occasion) as stochastic, which makes it possible to formulate a *linear* two-stage stochastic model. In Section 5.5 we define the scenarios used in the two-stage model. In order to use the scenarios to formulate the two-stage model it is necessary to formulate the deterministic optimization problem that arises when it is known that a specific scenario will occur, and we discuss how to do this in Section 5.6. The two-stage stochastic model is then presented in Section 5.7.

Finally, in Section 5.8 we discuss the differences between the stochastic dynamic programming model and the two-stage stochastic model.

5.2 The aircraft engine maintenance problem

When a stochastic part fails or a deterministic part reaches its lifetime the engine must be taken to the service workshop. The failed stochastic parts and the deterministic parts that have reached their lifetimes must be replaced by new ones (or, possible, used non-failed parts). It is expensive to take the engine to the workshop and hence, when having it there, it may be motivated to replace also non-failed stochastic parts and deterministic parts that have not yet reached their respective lifetimes. Further, in order to remove a specific part often one has to remove other parts. Some parts may therefore be preventively replaced even if they are not yet very old, provided that there is no additional work-cost.

The deterministic models developed in Chapter 4 integrate both the fixed cost to take the engine to the workshop and the economic dependencies just described. Based on these models, optimal replacement schedules for the total lifetime of the engine can be computed. When stochastic parts are involved the replacement schedule cannot span the entire engine lifetime, or contract period, due to the uncertainty present in the model. Instead, all we can decide is which parts (deterministic and stochastic) that should be replaced today, when the engine is at the workshop, in order to minimize the total *expected* cost over the planning period. This decision should be based on the information we have about the different parts at the specific maintenance occasion (that is, the failure distributions of the stochastic parts and the remaining lifetimes of the deterministic parts).

Hence, an optimization model for the maintenance of an aircraft engine must be iterative, that is, a reoptimization must be performed every time the engine is taken to the service work shop for whatever reason.

5.3 Basic modifications of the deterministic optimization models

The most simple way to take a stochastic part into consideration is to use a deterministic model to compute a basic replacement scheme, and then run the engine until either a deterministic part reaches its lifetime or a stochastic part fails. If the engine is taken to the workshop because a stochastic part has failed this part is replaced and a new basic replacement scheme is computed. Otherwise the basic replacement scheme is used to decide when to take the engine to the workshop and what to replace when having it there. However, since the RM12 engine often is taken to the workshop because a stochastic part has failed, this method is not very successful for our aircraft engine application.

We can modify this method by forcing a preventive replacement of a stochastic part when the engine is at the workshop if the failure rate (i.e., hazard rate) exceeds some threshold value. The failure rate can easily be computed from the failure distribution. It is, however, not at all clear at this time what the basis for this threshold value should be.

Another way to modify the above method is to first compute a basic scheme for the deterministic parts, and then for each one of the stochastic parts compute the probability for failure until the next time the engine is taken to the workshop (according to the basic replacement scheme). Then preventive replacements are triggered for all stochastic parts whose probabilities for failure exceed a threshold value. A problem with this method arises if the basic replacement scheme suggests that there is a long time to the next maintenance occasion; then it can happen that several of the stochastic parts are replaced even if they are almost new. In this case, it would perhaps be better to not replace the stochastic parts, but to run the engine until a stochastic part fails, and then, when the stochastic part is replaced, take the opportunity to replace some of the deterministic parts as well.

Finally, a natural way to take a stochastic part into account is to consider it as deterministic by letting its lifetime be its expected one. (This is actually the special case of the two-stage stochastic model presented in Section 5.7 that arises when only one scenario is used.)

All of the modifications presented in this section are heuristic in the sense that they do not necessarily minimize the total expected cost for having a working engine during the given time period. Therefore, we will develop other methods that hopefully perform better. The dynamic programming model presented in Section 5.4 actually gives optimal solutions. The two-stage stochastic model presented in Section 5.7 is, however, again heuristic, but probably better than the heuristic methods presented here.

5.4 A dynamic programming model

In this section we develop a dynamic programming model for a system consisting of stochastic and deterministic parts. The model is based on the dynamic programming model presented in Section 4.2. First we develop a general stochastic dynamic programming model for a discrete stochastic sequential decision process. We then illustrate the model by considering a simple example regarding a system consisting of only one stochastic part and one deterministic part. This example is then generalized in order to describe a dynamic programming model for a general system.

5.4.1 The dynamic programming model

In Section 4.2 we defined a *discrete sequential decision process*. Here we will generalize this concept to incorporate randomness. In a *discrete stochastic sequential decision process* we have a finite time horizon T and consider the points of time $t = 0, 1, \ldots, T$. At the time t the process is in state s_t, which is assumed to depend on

- the state s_{t-1} at time $t-1$;
- the decision variable x_{t-1} at time $t-1$; and
- the stochastic variable y_{t-1} (which in turn depends on s_{t-1} and x_{t-1}).

Observe that the difference between the discrete and the stochastic discrete sequential decision process is that in the latter the state s_t depends on the stochastic variable y_{t-1}. Denote the transformation function at time t by ϕ_t, that is,

$$s_{t+1} = \phi_t(s_t, x_t, y_t).$$

Also, let $g_t(s_t, x_t, y_t)$ be the contribution to the cost function during the time between t and $t+1$. Denote by $E_\xi[f(\xi)]$ the expected value of the function f over all possible values of the stochastic variable ξ. We cannot solve the whole decision problem (that is, determine all values of $x_0, x_1, \ldots, x_T$) at time $t = 0$ since the decisions $x_1, \ldots, x_T$ will depend on the outcomes of $y_0, y_1, \ldots, y_T$. The only decision that can, and must, be made at time $t = 0$ is x_0. The decision x_1 at time $t = 1$ is then made when the outcome of y_0 is known, and so on. Therefore, the problem we want to solve is to choose x_0 at time $t = 0$ such that the expected cost of the complete decision process is minimized, that is,

$$\begin{aligned}
z = \operatorname*{minimum}_{x_0} \quad & E_{y_0}\Big[g_0(s_0, x_0, y_0) + \min_{x_1} E_{y_1}[g_1(s_1, x_1, y_1) + \cdots \\
& \qquad + \min_{x_T} E_{y_T}[g_T(s_T, x_T, y_T)] \cdots]\Big] \\
\text{subject to} \quad & s_{t+1} = \phi_t(s_t, x_t, y_t), \quad t = 0, 1, \ldots, T-1, \\
& s_0 \text{ is given.}
\end{aligned}$$

In order to develop a recursive optimization scheme, for $k = T, T-1, \ldots, 1, 0$, let

$$\begin{aligned} z_k(s_k) = \underset{x_k}{\text{minimum}} \quad & E_{y_k}\Big[g_k(s_k, x_k, y_k) + \min_{x_{k+1}} E_{y_{k+1}}[g_{k+1}(s_{k+1}, x_{k+1}, y_{k+1}) + \\ & \cdots + \min_{x_T} E_{y_T}[g_T(s_T, x_T, y_T)] \cdots]\Big] \\ \text{subject to} \quad & s_{t+1} = \phi_t(s_t, x_t, y_t), \quad t = k, \ldots, T-1. \end{aligned} \tag{5.1}$$

(Hence $z = z_0(s_0)$ for the given state s_0.) We arrive at the following recursion:

Proposition 5.1 (recursion in stochastic dynamic programming) *For $k = 0, 1, \ldots, T-1$, if $z_k(s_k)$ is defined as in (5.1), then*

$$\begin{aligned} z_k(s_k) = \underset{x_k}{\text{minimum}} \quad & E_{y_k}[g_k(s_k, x_k, y_k) + z_{k+1}(s_{k+1})] \\ \text{subject to} \quad & s_{k+1} = \phi_k(s_k, x_k, y_k). \end{aligned} \tag{5.2}$$

Proof. Follows immediately from the definition of $z_k(s_k)$. ■

Hence, if we can compute

$$z_T(s_T) = \underset{x_T}{\text{minimum}} \; E_{y_T}[g_T(s_T, x_T, y_T)]$$

for all values of s_T, then we can use the recursion scheme in Proposition 5.1 to compute x_0. Typically, $z_T(s_T)$ is given, or is at least easy to compute. In our aircraft engine application we can set $z_T(s_T) = 0$ for all states s_T if the maintenance contract does not reward any specific states of the engine (that is, if the remaining lifetimes at the end of the contract period do not have inherent values); otherwise, if the contract includes values $c(s_T)$ for each specific state s_T, then we can set $z_T(s_T) = -c(s_T)$. (The minus sign is needed since $z_T(s_T)$ is considered as a cost in the model.)

Just as for the deterministic dynamic programming model the main drawback of the stochastic dynamic programming model is that we must solve (5.2) for each possible state s_k. Hence, if the number of states is large, then the usefulness of a stochastic dynamic programming model is limited. Unfortunately, as was seen in Section 4.2 the number of states in the aircraft engine application is huge if the number of parts is large. Therefore, the stochastic dynamic programming model can only be used if the number of parts is small (from our experiments: about 6). The advantage of the dynamic programming model as compared to the two-stage model is however that it takes all the randomness in the problem into account; the two-stage model assumes that every outcome after time $t = 0$ is known. Differences between the dynamic programming model and the two-stage model are discussed further in Section 5.8.

5.4.2 A system with one stochastic and one deterministic part

We turn to our aircraft engine application. Consider a system consisting of one stochastic part and one deterministic part. The lifetime of the stochastic part is represented by a failure distribution function. Now assume that at time t the deterministic part has not reached its deterministic lifetime and has been in service for τ_d time steps and the stochastic part is non-failed and has been in service for τ_s time steps. Then, with the failure distribution function we can compute the probability, denoted by p, that the stochastic part that is currently in the system is still non-failed at time $t+1$. (The probability that the stochastic part fails between the times t and $t+1$ is then $1-p$.) In order to formulate a dynamic programming model we assume that the stochastic part only can fail at the specific times $t = 1, \ldots, T$ (and not in between). (The error from this simplification decreases as the number of time steps in the discretization increases.) This situation is illustrated in Figure 5.1.

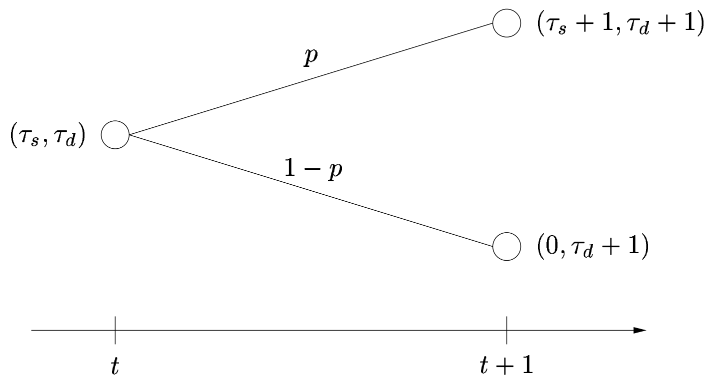

Figure 5.1: The failure probability of the stochastic part.

The numbers in the parentheses indicate the ages of the stochastic and the deterministic part, respectively. We specify by $(0, \tau_d + 1)$ that the stochastic part has failed at time $t+1$ and hence must be replaced by a new one, so that its age becomes 0.

Obviously, the probability that the stochastic part will still be non-failed at time $t+1$ depends on its age τ_s at time t. We let $p(\tau_s)$ be the probability that the stochastic part is unfailed at time step $t+1$ given that its age at time t is τ_s.

At a specific time t we can perform four replacement actions, namely,

1. replace nothing;
2. replace the stochastic part only;
3. replace the deterministic part only; and

4. replace both the stochastic and the deterministic part.

As for the dynamic programming model presented in Section 4.2 we denote a replacement action by

$$x_t = (r_s^t, r_d^t),$$

where

$$r_i^t = \begin{cases} 1, & \text{if part } i \text{ is replaced at time } t, \\ 0, & \text{otherwise.} \end{cases}$$

The above replacement actions can then be denoted by $(0,0)$, $(1,0)$, $(0,1)$, and $(1,1)$, respectively. In the dynamic programming model for deterministic parts presented in Section 4.2 the state at time $t+1$ was completely determined by the state and the replacement action at time t. However, in the case where one of the parts is stochastic, the state at time $t+1$ will depend on whether the stochastic part fails or not. For example, if the replacement action is $x_t = (0,0)$, we end up with the state (τ_s+1, τ_d+1) with probability $p(\tau_s)$ and with the state $(0, \tau_d+1)$ with probability $1-p(\tau_s)$. Figure 5.2 below illustrates the different states and the probabilities to reach them given the replacement action x_t.

The figure should be interpreted as follows: At time t the system is in state (τ_s, τ_d). Depending on the replacement action x_t, we arrive at one of the intermediate nodes. (Observe that we still are at time t.) Then the system is put into service again, and depending on whether the stochastic part fails or not we end up with one of the states at time $t+1$, with different probabilities. If each of the states s_{t+1} at time $t+1$ has a given value, denoted by $z_{t+1}(s_{t+1})$, then it is possible to compute the expected value at time t of each replacement action. For example, the expected value of the replacement action $x_t = (1,0)$ is given by

$$g_t(s_t, x_t) + p(0) \cdot z_{t+1}(1, \tau_d + 1) + [1 - p(0)] \cdot z_{t+1}(0, \tau_d + 1),$$

where $g_t(s_t, x_t)$ is the cost of the replacement action $x_t = (1,0)$ at the state $s_t = (\tau_s, \tau_d)$ at time t. (Observe that we define $g_t(s_t, x_t) = +\infty$ if the replacement action x_t is not allowed; for example, if the stochastic part has failed at time t, then it must be replaced and accordingly every allowed replacement action x_t must include a replacement of the stochastic part.) Hence, if the values of all the possible states at time $t+1$ are known, we can set the value of the state (τ_s, τ_d) at time t to the expected value of a replacement action that minimizes the expected value at time t. This valuation of the states is crucial to the stochastic dynamic programming model developed above in Section 5.4.1.

In order to apply the recursion optimization scheme given by Proposition 5.1 we must define the transformation function ϕ_t and the stochastic variables y_t for each point of time. The variables y_t denote whether the stochastic part fails or not at time $t+1$, that is,

$$y_t = \begin{cases} 1, & \text{if the stochastic part fails at time } t+1, \\ 0, & \text{otherwise.} \end{cases}$$

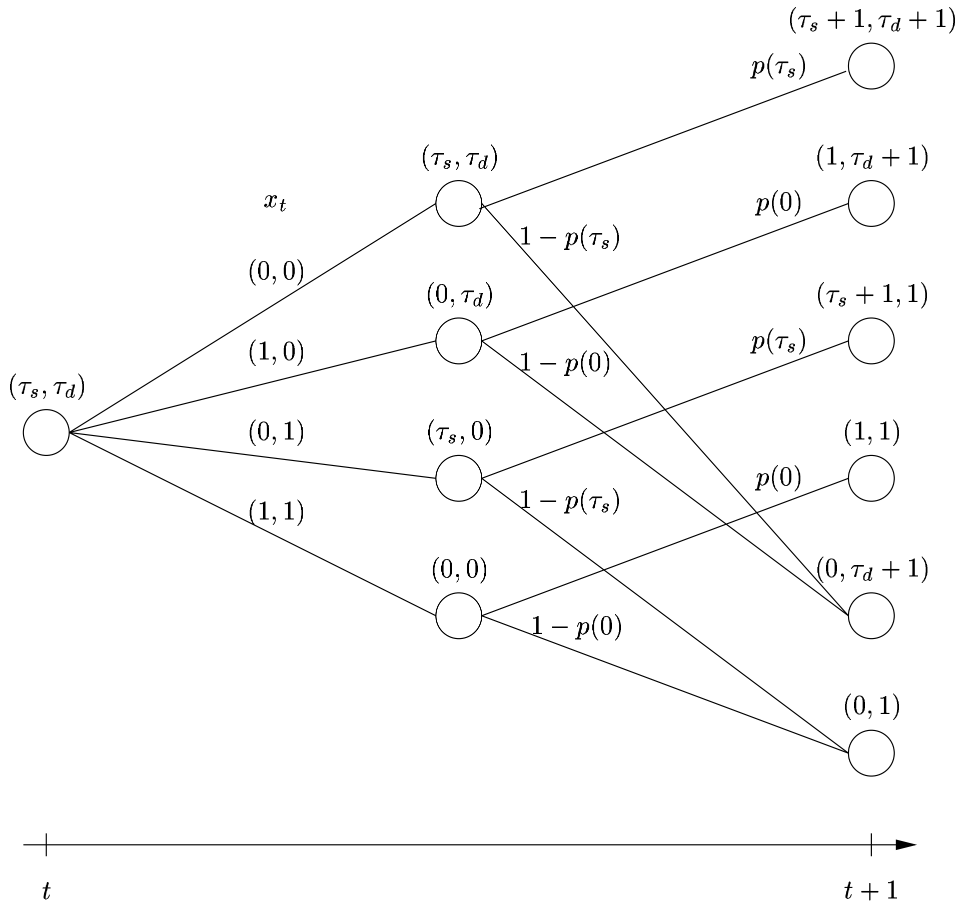

Figure 5.2: The possible states and the probabilities for reaching them.

With $s_t = (\tau_s^t, \tau_d^t)$ and $x_t = (r_s^t, r_d^t)$ we define the transformation function $\phi_t(s_t, x_t, y_t)$ by

$$\phi_t(s_t, x_t, y_t) = (\tau_s^{t+1}, \tau_d^{t+1}),$$

where

$$\tau_s^{t+1} = \begin{cases} \tau_s^t + 1, & \text{if } r_s^t = 0 \text{ and } y_t = 0, \\ 1, & \text{if } r_s^t = 1 \text{ and } y_t = 0, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$\tau_d^{t+1} = \begin{cases} \tau_d^t + 1, & \text{if } r_d^t = 0, \\ 1, & \text{otherwise.} \end{cases}$$

The recursion scheme (5.2) also requires the computation of the expected cost associated with the stochastic variable y_t. More precisely, we want to calculate $E_{y_t}[f(y_t)]$ for some given function f. In our case the value of y_t (that is, 0 or 1) occurs with a certain probability depending on the state s_t (that is, the age of the stochastic part) and the replacement action x_t, so we define

$$p(y_t, s_t, x_t) = \text{the probability for } y_t \text{ to occur given } s_t \text{ and } x_t.$$

In the recursion formula (5.2) the state s_k is given, so for a given replacement action x_k the probability that y_k occurs is known. Since the number of states y_k at each time k is finite (in our simplified case only two) we have that

$$E_{y_k}[f(y_k)] = \sum_{y_k} p(y_k, s_k, x_k) f(y_k),$$

where $\sum_{y_k}$ denotes the summation over the finite states of y_k at time k.

Hence, the recursion optimization scheme in Proposition 5.1 finally becomes

$$\begin{aligned} z_k(s_k) = \underset{x_k}{\text{minimum}} \quad & \sum_{y_k} p(y_k, s_k, x_k)[g_k(s_k, x_k) + z_{k+1}(s_{k+1})] \\ \text{subject to} \quad & s_{k+1} = \phi_k(s_k, x_k, y_k). \end{aligned}$$

5.4.3 A general system

In the previous subsection we applied the stochastic dynamic programming model presented in Section 5.4.1 on a system consisting of only one stochastic and one deterministic part. However, the approach presented there can easily be adapted to general systems consisting of N deterministic parts and M stochastic parts.

We introduce the state variable s_t as

$$s_t = (\tau^t_{s_1}, \ldots, \tau^t_{s_M}, \tau^t_{d_1}, \ldots, \tau^t_{d_N}),$$

where $\tau^t_{s_i}$ and $\tau^t_{d_j}$ are the ages of the stochastic part i and the deterministic part j, respectively, at time t.

Further, the decision variables x_t becomes

$$x_t = (r^t_{s_1}, \ldots, r^t_{s_M}, r^t_{d_1}, \ldots, r^t_{d_N}),$$

where

$$r^t_{s_i} = \begin{cases} 1, & \text{if the stochastic part } i \text{ is replaced at time } t, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$r^t_{d_j} = \begin{cases} 1, & \text{if the deterministic part } j \text{ is replaced at time } t, \\ 0, & \text{otherwise.} \end{cases}$$

The cost contribution $g_t(s_t, x_t)$ at time t is defined as

$$g_t(s_t, x_t) = \begin{cases} \text{cost to replace the parts given by } x_t, & \text{if } x_t \text{ is feasible given } s_t, \\ +\infty, & \text{otherwise.} \end{cases}$$

For example, if s_t denotes that a stochastic part has failed at time t then every replacement action x_t that do not include a replacement of that stochastic part is infeasible.

The stochastic variable y_t denotes which of the stochastic parts that will fail at time $t+1$, that is,

$$y_t = (f_1^t, \dots, f_M^t),$$

where

$$f_i^t = \begin{cases} 1, & \text{if the stochastic part } i \text{ fails at time } t+1, \\ 0, & \text{otherwise.} \end{cases}$$

The probability that y_t occurs depends on the state s_t and the replacement action x_t. Let

$$p(y_t, s_t, x_t) = \text{the probability that } y_t \text{ occurs given } s_t \text{ and } x_t.$$

Observe that the number of possible outcomes of y_t is finite (although it is large if the number M of stochastic parts is large; in fact, the number of possible outcomes is 2^M).

Finally, the transformation function ϕ_t is given by

$$\phi_t(s_t, x_t, y_t) = (\tau_{s_1}^{t+1}, \dots, \tau_{s_M}^{t+1}, \tau_{d_1}^{t+1}, \dots, \tau_{d_N}^{t+1}),$$

where

$$\tau_{s_i}^{t+1} = \begin{cases} \tau_{s_i}^t + 1, & \text{if } f_i^t = 0 \text{ and } r_{s_i}^t = 0, \\ 1, & \text{if } f_i^t = 0 \text{ and } r_{s_i}^t = 1, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$\tau_{d_j}^{t+1} = \begin{cases} \tau_{d_j}^t + 1, & \text{if } r_{d_j}^t = 0, \\ 1, & \text{otherwise.} \end{cases}$$

The recursion scheme, for $k = 0, \dots, T-1$, is then as follows:

$$\begin{aligned} z_k(s_k) = \underset{x_k}{\text{minimum}} \quad & \sum_{y_k} p(y_k, s_k, x_k)[g_k(s_k, x_k) + z_{k+1}(s_{k+1})] \\ \text{subject to} \quad & s_{k+1} = \phi_k(s_k, x_k, y_k). \end{aligned}$$

5.5 Representing the lifetime distribution by scenarios

In the stochastic two-stage model developed below we will use *scenarios* in order to discretize the failure distributions of the stochastic parts. In this section we describe what we mean by a scenario in the case of a system consisting of deterministic and stochastic parts.

5.5.1 Scenarios of stochastic parts

Briefly, a scenario is a possible future outcome. However, in order to make the concept of scenario useful in real world applications we must clearly define what we mean by *future outcome*. Also, each possible future outcome occurs with a specific probability, so an outcome always comes together with the *probability* that it occurs. In this subsection we define scenario for stochastic parts, and in the next subsection we discuss how to compute scenarios and the probability with which they occur.

Consider a system consisting of one stochastic part only. Further, assume that the time horizon T is finite and that the time is discretized so that $t = 0, 1, \ldots, T$. In principle, the stochastic part can fail anytime during the total lifetime T of the system; if it fails it is replaced by a new specimen. In order to construct scenarios that can be used in a stochastic two-stage model we assume, however, that the stochastic part only fails at the times $t = 1, \ldots, T$ (and never between two times). This is the same simplification that was made in the development of the stochastic dynamic programming model in the previous section. The maximum number of stochastic parts that will ever be needed in order to have a working system between the times 0 and T is T. Hence, it makes sense to define a scenario for the system as a sequence of T integers, where the first integer is the (deterministic!) lifetime (in number of time steps) of the stochastic part that is in the engine at the start, the second integer gives the lifetime of the stochastic part that replaces the first part when it fails (or is preventively replaced), and so on. We make the following definition:

Definition 5.2 (scenarios of stochastic parts) *A scenario, w_s, of a stochastic part, s, of a system with the time-horizon T is a sequence of T integers*

$$w_s = (T_s^1, T_s^2, \ldots, T_s^T),$$

where T_s^1 is the (deterministic) lifetime of the part of type s that is in the system at the start, T_s^2 is the lifetime of the part of type s that replaces the first part, and so on. ■

Each scenario w_s is associated with its probability $p(w_s)$, so the expected value is

$$E_{w_s}[f(w_s)] = \sum_{w_s} p(w_s) f(w_s).$$

If we have several stochastic parts in the system, Definition 5.2 can be naturally generalized. Namely, assume that we have M stochastic parts and that part i has the scenarios

$$w_{s_i} = (T^1_{s_i}, \dots, T^T_{s_i}), \quad s_i \in S_i,$$

where S_i is the set of scenarios for the stochastic part i. The set of scenarios of the whole system then becomes the collection

$$W_s = (w_{s_1}, \dots, w_{s_M}),$$

where $s = (s_1, \dots, s_M)$ and $s_i \in S_i$ for all $i = 1, \dots, M$. Observe that the number of scenarios of the system is $\prod_{i=1}^M |S_i|$, which means that the total number of scenarios increases very fast with the number of stochastic parts. The probability $p(W_s)$ that a certain scenario W_s occurs is given by $\prod_{i=1}^M p(w_{s_i})$.

Remark 5.3 (alternative scenario definition) Another, and perhaps more natural, way to define scenarios for stochastic parts is to consider each point of time separately as was made in the stochastic dynamic programming model. Consider a system consisting of one stochastic part. At each point of time we then have two possible scenarios, namely,

- the stochastic part fails; and
- the stochastic part is non-failed.

However, the probabilities that these scenarios occur depend on the age of the stochastic part which in turn depends on previous replacement actions. Hence, the probabilities for the different scenarios depend on the variables of the problem. A corresponding stochastic program for the problem will therefore be nonlinear which is not desirable from a computational point of view. On the other hand, if we use scenarios as in Definition 5.2, then it is possible to formulate a *linear* stochastic program, as we will see below. ∎

5.5.2 On the computation of scenarios

In this section we present a natural and simple way to compute the scenarios disussed above, and the probabilities with which they occur. It should be noted that the computation of scenarios can be made in several different ways, and typically the quality of the solution from a stochastic programming model will depend on the the scenarios computed. Our ambition is not to discuss how to compute the *best* scenarios for the problem (this is a hard and interesting problem in itself), but just to present a possible computational procedure.

Consider a stochastic part whose failure distribution is given by the distribution function F. Also, let f be the frequency function defined by

$$f(t) = \frac{dF(t)}{dt}.$$

In order to create scenarios for the stochastic part we must choose some lifetimes and associate them with probabilities. Assume that we want to represent the failure distribution by n lifetimes. First we partition the time interval $[0, \infty]$ into n intervals $[t_{i-1}, t_i]$, $i = 1, \ldots, n$, such that

$$\int_{t_{i-1}}^{t_i} f(t)\, dt = \frac{1}{n},$$

for all $i = 0, \ldots, n$, or, equivalently,

$$t_i = F^{-1}(i/n), \quad i = 0, \ldots, n. \tag{5.3}$$

(Note that $t_0 = 0$ and $t_n = \infty$.) Then we must choose a lifetime, $l_i \in [t_{i-1}, t_i]$, to represent the time interval $[t_{i-1}, t_i]$. We do this by taking l_i as the expected lifetime in $[t_{i-1}, t_i]$, that is,

$$l_i = n \int_{t_{i-1}}^{t_i} t f(t)\, dt. \tag{5.4}$$

We have now created n equally probable scenarios, $l_1, \ldots, l_n$. In order to construct scenarios as in Definition 5.2 we then take all vectors of the form

$$(T^1, \ldots, T^T), \tag{5.5}$$

where T^k varies over $l_1, \ldots, l_n$ for each $k = 1, \ldots, T$. The probability of each scenario becomes $1/n^T$.

Remark 5.4 (used stochastic parts) When the engine arrives at the workshop the stochastic parts are not new. The failure distributions for the parts that are currently in the engine therefore differ from the failure distributions for the new parts (that will replace the stochastic parts in the future). Hence, the lifetimes l_i for a part that is currently in the engine differs from the other and accordingly, in the scenarios of the form (5.5), the first lifetime T^1 will vary over a different set of lifetimes than T^k for $k = 2, \ldots, T$. ■

Example 5.5 (computing scenarios) Consider a stochastic part whose lifetime is represented by the (Weibull) failure distribution function

$$F(t) = 1 - \exp\left[-\left(\frac{t}{300}\right)^2\right],$$

where $\exp(x) = e^x$. Let $n = 4$. Then from the equations (5.3) and (5.4) we get the lifetimes $l_1 = 104$, $l_2 = 206$, $l_3 = 299$, and $l_4 = 452$, and the probability for each of the lifetimes is $1/4$. The failure distribution and the scenarios are illustrated in Figure 5.3. ■

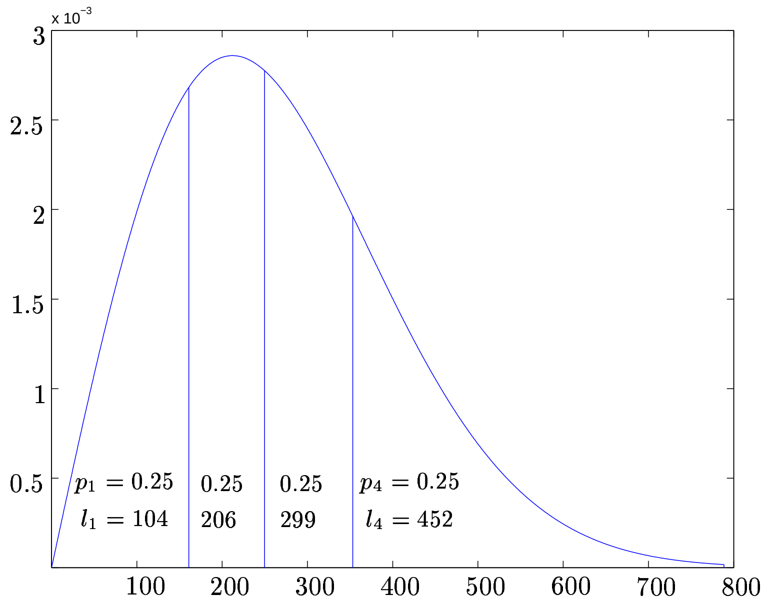

Figure 5.3: Illustration of the failure distribution and the scenarios.

5.6 Modelling varying lifetimes

Each scenario gives a sequence of lifetimes for the stochastic parts (see Definition 5.2). If we just consider one stochastic part, the first value of such a sequence is the lifetime of the part that is currently in the engine. The second value of the sequence is the lifetime of the part that will replace the stochastic part at the first replacement, the third value of the sequence is the lifetime of the part that will replace the stochastic part at the second replacement, and so on.

Therefore, in order to formulate a stochastic two stage model, we must be able to model this kind of *varying lifetimes* for each specific part. The simplest replacement problem presented in Chapter 4 had only one economic dependence, namely a fixed cost associated with taking the engine to the workshop, and the replacement problem (without varying lifetimes) was modelled by Model I in Section 4.3, that is, to

$$
\begin{aligned}
\text{minimize} \quad & \sum_{t=1}^{T} \left(\sum_{i \in \mathcal{N}} c_i x_{it} + d z_t \right) \\
\text{subject to} \quad & \sum_{t=\ell}^{\tau_i+\ell-1} x_{it} \geq 1, \quad \ell = 1, \ldots, T - T_i, \quad i \in \mathcal{N}, \\
& x_{it} \leq z_t, \quad t = 1, \ldots, T, \quad i \in \mathcal{N}, \\
& x_{it}, z_t \in \{0, 1\}, \quad t = 1, \ldots, T, \quad i \in \mathcal{N}.
\end{aligned}
$$

In this model we assumed that all parts were new at time $t = 0$ and hence we did not consider the possibility to replace some of the parts at time $t = 0$. Note, however, that when we allow varying lifetimes it might be necessary to replace parts already at time $t = 0$.

In this section we will show how to modify this model to include a part with varying lifetimes. Similarly, the other models presented in Chapter 4 can be modified to include parts with varying lifetimes. In Section 4.8 we have already discussed the case where the first lifetime is shorter than the others.

First we show how to include a part whose first lifetime differs from the other lifetimes, which are assumed to be equal. Then, we consider the more complicated case where all the lifetimes of a specific part are allowed to vary.

5.6.1 Variations only in the first lifetime

Consider a system that comprises one part whose first lifetime differs from the other lifetimes. (The other lifetimes are, however, assumed to be equal among each other.) Assume that the first lifetime is $\tilde{T}_s$ and that the other lifetimes are T_s (s for "stochastic"). Introduce, for $t = 0, 1, \ldots, \tilde{T}_s$, the binary variables for the first replacement

$$\tilde{x}_{st} = \begin{cases} 1, & \text{if the part with the first lifetime is replaced at time } t, \\ 0, & \text{otherwise.} \end{cases}$$

The part with the first lifetime can only be replaced once during its lifetime, which gives

$$\tilde{x}_{s0} + \tilde{x}_{s1} + \cdots + \tilde{x}_{s,\tilde{T}_s} \leq 1. \tag{5.6}$$

Further, before a specimen with the lifetime T_s can be put into the engine, the specimen with the lifetime $\tilde{T}_s$ first must be replaced, which yields (the binary variables x_{st} specifies, as in Model I (see Section 4.3), whether part s is replaced or not at time t)

$$\begin{aligned} x_{s0} &= 0, \\ x_{st} &\leq \sum_{k=0}^{t-1} \tilde{x}_{kt}, \quad t = 1, \ldots, \tilde{T}_s. \end{aligned}$$

When the specimen with the lifetime $\tilde{T}_s$ has been replaced, the other parts (all of which have the lifetime T_s) can not be in service for more than T_s time steps. We get

$$\begin{aligned} \sum_{t=\ell}^{T_s+\ell-1} x_{st} + \sum_{t=\ell}^{\tilde{T}_s} \tilde{x}_{st} &\geq 1, \quad \ell = 0, \ldots, \tilde{T}_s, \\ \sum_{t=\ell}^{T+\ell-1} x_{st} &\geq 1, \quad \ell = \tilde{T}_s + 1, \ldots, T - T_s. \end{aligned}$$

Also, in order to replace part s, the fixed charge variable z_t must be active, which gives

$$\begin{aligned} \tilde{x}_{st} &\leq z_t, \quad t = 0, \ldots, \tilde{T}_s, \\ x_{st} &\leq z_t, \quad t = 0, \ldots, T. \end{aligned}$$

The modified version of Model I, where one part has varying first lifetime, becomes

$$\begin{aligned} \text{minimize} \quad & \sum_{t=0}^{T} \left(\sum_{i \in \mathcal{N}} c_i x_{it} + c_s x_{st} + d z_t \right) + \sum_{t=0}^{\tilde{T}_s} c_s \tilde{x}_{st} \qquad (5.7) \\ \text{subject to} \quad & \sum_{t=\ell}^{T_i+\ell-1} x_{it} \geq 1, \quad \ell = 0, \ldots, T - T_i, \quad i \in \mathcal{N}, \\ & x_{it} \leq z_t, \quad t = 0, \ldots, T, \quad i \in \mathcal{N}, \\ & \sum_{t=\ell}^{T_s+\ell-1} x_{st} + \sum_{t=\ell}^{\tilde{T}_s} \tilde{x}_{st} \geq 1, \quad \ell = 0, \ldots, \tilde{T}_s, \\ & \sum_{t=\ell}^{T+\ell-1} x_{st} \geq 1, \quad \ell = \tilde{T}_s + 1, \ldots, T - T_s, \\ & x_{s0} = 0, \\ & x_{st} \leq \sum_{k=0}^{t-1} \tilde{x}_{kt}, \quad t = 1, \ldots, \tilde{T}_s, \\ & \tilde{x}_{st} \leq z_t, \quad t = 0, \ldots, \tilde{T}_s, \\ & x_{st} \leq z_t, \quad t = 0, \ldots, T, \\ & x_{it}, x_{st}, \tilde{x}_{st}, z_t \in \{0, 1\}, \quad t = 0, \ldots, T, \quad i \in \mathcal{N}. \end{aligned}$$

Here, the constraint (5.6) has been excluded since it will never be optimal to replace the part with the first lifetime twice.

5.6.2 Variations in all the lifetimes

If we allow all the lifetimes of a specific part to vary the situation becomes more complex. We introduce the binary variables

$$x_{st}^r = \begin{cases} 0, & \text{if the part with the } r\text{th lifetime has not been put into the} \\ & \text{engine at time } t \text{ or before,} \\ 1, & \text{otherwise.} \end{cases}$$

The upper index r takes integer values between 1 and the number of lifetimes for the stochastic part. This number of lifetimes is given by the scenarios, and must be large enough to not be a restriction in the optimization model. This can always be achieved by letting the number of lifetimes be equal to the total number of time steps T (since we cannot replace the part more often than at every time step). We denote the total number of lifetimes by R.

Suppose that the rth lifetime is given by T_s^r. Then we must have that

$$\sum_{t=0}^{T}(x_{st}^r - x_{st}^{r+1}) \leq T_s^r, \quad r = 1, \ldots, R-1,$$
$$\sum_{t=0}^{T} x_{st}^R \leq T_s^R.$$

If a part with the rth lifetime is put into the engine at, say, time t^* then we must have that $x_t^r = 0$ for all $t < t^*$ and $x_t^r = 1$ for all $t \geq t^*$. This is enforced by the use of the constraints

$$x_{s,t+1}^r \geq x_{st}^r, \quad t = 0, \ldots, T-1, \quad r = 1, \ldots, R.$$

The part with the second lifetime can replace the part with the first lifetime (that is currently in the engine) at time $t = 0$. However, the part with the $(r+1)$th lifetime, for $r = 2, \ldots, R$, cannot be put into the engine before the part with the rth lifetime has been in the engine for at least one time step. This yields the constraints

$$x_{st}^{r+1} \leq x_{s,t-1}^r, \quad t = 1, \ldots, T, \quad r = 2, \ldots, R-1.$$

At time $t = 0$ we have $x_{s0}^1 = 1$.

If the part with the second lifetime is put into the engine at time $t = 0$ we will have

$$x_{s0}^2 = 1,$$

and otherwise $x_{s0}^2 = 0$. Further, if the part with the rth lifetime is put into the engine at time $t = 1, \ldots, T$, we will have

$$x_{st}^r - x_{s,t-1}^r = 1,$$

otherwise, $x_{st}^r - x_{s,t-1}^r = 0$. Hence the objective function becomes that to minimize

$$c_s x_{s0}^2 + \sum_{i \in \mathcal{N}} c_i x_{i0} + dz_0 + \sum_{t=1}^{T}\left(c_s \sum_{r=2}^{R}(x_{st}^r - x_{s,t-1}^r) + \sum_{i \in \mathcal{N}} c_i x_{it} + dz_t\right),$$

where c_s is the cost of the part with variable lifetime.

To connect the part with variable lifetime with the fixed cost, we must have

$$x_{s0}^2 \leq z_0$$

and

$$x^r_{st} - x^r_{s,t-1} \leq z_t, \quad t = 1, \dots, T, \quad r = 2, \dots, R.$$

The modified version of Model I with one part with variable lifetimes and the other parts with fixed lifetimes becomes

$$\begin{aligned}
\text{minimize} \quad & c_s x^2_{s0} + \sum_{i \in \mathcal{N}} c_i x_{i0} + dz_0 \qquad (5.8)\\
& + \sum_{t=1}^{T} \left(c_s \sum_{r=2}^{R} (x^r_{st} - x^r_{s,t-1}) + \sum_{i \in \mathcal{N}} c_i x_{it} + dz_t \right) \\
\text{subject to} \quad & \sum_{t=\ell}^{T_i+\ell-1} x_{it} \geq 1, \quad \ell = 0, \dots, T - T_i, \quad i \in \mathcal{N}, \\
& x_{it} \leq z_t, \quad t = 0, \dots, T, \quad i \in \mathcal{N}, \\
& \sum_{t=1}^{T} (x^r_{st} - x^{r+1}_{st}) \leq T^r_s, \quad r = 1, \dots, R-1, \\
& \sum_{t=0}^{T} x^R_{st} \leq T^R_s, \\
& x^r_{s,t+1} \geq x^r_{st}, \quad t = 0, \dots, T-1, \quad r = 1, \dots, R, \\
& x^{r+1}_{st} \leq x^r_{s,t-1}, \quad t = 1, \dots, T, \quad r = 2, \dots, R-1, \\
& x^1_{s0} = 1, \\
& x^2_{s0} \leq z_0, \\
& x^r_{st} - x^r_{s,t-1} \leq z_t, \quad t = 1, \dots, T, \quad r = 1, \dots, R, \\
& x_{it}, x^r_{st}, z_t \in \{0, 1\}, \quad t = 0, \dots, T, \quad r = 1, \dots, R, \quad i \in \mathcal{N}.
\end{aligned}$$

Remark 5.6 (numerical tests with varying lifetimes) Simple numerical tests have been made with the models (5.7) and (5.8). Both of the models were implemented in AMPL and then solved by the Branch & Bound solver in CPLEX. It took about 20 times longer to solve (5.8) than to solve (5.7) for small instances (one part with variable lifetime, 3 parts with constant lifetime and $T = 60$) even if the number of scenarios were the same. The differences were even larger when the number of parts was increased. Hence we will use the stochastic two-stage model based on (5.7) (presented in the next section) in the numerical tests in Chapter 7. ■

5.7 A stochastic two-stage model

Now we turn to the stochastic two-stage model. First we discuss a general two-stage model, and then we show how it can be adapted to the stochastic replacement problem. It should be noted that the stochastic replacement problem is in fact *not* a two-stage problem, but a multi-stage problem (since each point of time is a stage). However, every attempt to formulate a stochastic multi-stage model resulted in a non-linear model, so in order to formulate a linear stochastic model we consider the replacement problem as a two-stage problem. In Section 5.8 we will investigate the differences between the stochastic dynamic programming model (which actually solves the complete multi-stage problem with all possible scenarios) and the stochastic two-stage model.

5.7.1 A general stochastic two-stage model

Suppose that we must make decisions in a random environment. Further, suppose that the decisions can be made in two stages, that is, some decisions must be made prior to the random events, and the rest of the decisions can be made based on what actually happened. A lot of practical situations have this characteristic. A basic such problem is *the news vendor problem*, in which a news vendor goes to the publisher every morning and buys newspapers at a certain price. The vendor then walks along the streets to sell as many newspapers as possible at the selling price. The demand for newspapers varies over days according to some random variable. Any unsold newspaper can be returned to the publisher at a return price. The problem is then to decide how many newspapers to buy every morning in order to maximize the total expected income. General text books that describe stochastic two-stage models are [27], [72], [37], and [38].

Denote the first stage decision variables by x and the second stage decision variables by y. Suppose that the random events that might occur is a discrete set Ω. The probability that an event $w \in \Omega$ occurs is given by $p(w)$.

Now assume that given $w \in \Omega$ we want to solve the deterministic linear programming problem

$$\begin{array}{ll} \text{minimize} & c^{\mathrm{T}}x + d_w^{\mathrm{T}}y_w \\ \text{subject to} & Ax = b, \\ & B_w x + D_w y_w = e_w, \end{array}$$

where c, d_w, A, b, B_w, D_w, and e_w are vectors and matrices of appropriate sizes. Observe that B_w, D_w, d_w, and e_w might depend on w.

Since we do not know which event in Ω that will occur, the best we can do is to minimize the expected cost. This yields the stochastic two-stage problem to

$$\begin{array}{ll} \text{minimize} & c^{\mathrm{T}}x + \sum\limits_{w \in \Omega} p(w) d_w^{\mathrm{T}} y_w \\ \text{subject to} & Ax = b, \\ & B_w x + D_w y_w = e_w, \quad w \in \Omega. \end{array}$$

Observe that if the scenarios $w \in \Omega$ are not dependent of x, this problem is a linear program.

5.7.2 A stochastic two-stage model for the replacement problem

In the maintenance situation with one stochastic part, the first stage decision is to decide what to replace when the stochastic part has failed or a deterministic part has reached its lifetime. We will assume that this occurs at time $t = 0$, that is, the engine is taken to the workshop at time $t = 0$ independently of what is actually replaced. The second stage decision is what to replace at time $t = 1$ based on the actual outcome during the time between $t = 0$ and $t = 1$, the third stage decision is what to replace at time $t = 2$ based on the actual outcomes during the time between $t = 0$ and $t = 2$, and so on. However, in order to formulate a *linear* stochastic two-stage model we consider *all* the decisions at the times $t = 1, \ldots, T$ as two-stage decisions.

Even if this assumption is a serious simplification of reality, the two-stage model based on it makes use of the fact that the engine is at the workshop at time $t = 0$, which means that the model considers the opportunity to replace parts at time $t = 0$ at no fixed cost. This is important if the fixed costs are high, which is the case for aircraft engine maintenance. Hence the stochastic two-stage model developed in this section is particularly useful for our aircraft engine application.

We use scenarios as in Definition 5.2. Assume that the set Ω consists of all of the scenarios we consider; the scenarios can be computed according to Section 5.5.2. Given a scenario $w \in \Omega$ we then have the lifetimes of the stochastic part, that is,

$$w = (T_w^1, \ldots, T_w^T).$$

Also, the deterministic parts are not necessarily new at time $t = 0$. We assume that the remaining lifetimes of the deterministic parts that are in the engine at time $t = 0$ are given by $\tilde{T}_1, \ldots, \tilde{T}_N$. We will handle this fact by adding constraints as in Section 4.8.

The first-stage variables for the deterministic parts are all of the replacement variables at time $t = 0$, that is,

$$x_{i0}, \quad i \in \mathcal{N}.$$

The first-stage variable for the stochastic part describes the decision whether or not to replace the stochastic part at time $t = 0$. Since the only possible part for the first replacement is the specimen with the second lifetime, we get that x_{s0}^2 is the first-stage variable for the stochastic part. Note that the fixed cost variable $z_0 = 1$ since the engine is at the workshop at time $t = 0$ independent of what is replaced. Hence, z_0 can be eliminated from the model.

The second-stage variables are all of the variables at the time $t = 1, \ldots, T$, and these variables are dependent of which scenario $w \in \Omega$ that actually occurs.

We get the two-stage variables:

$$
\begin{aligned}
&x_{it}^w, \quad i \in \mathcal{N}, \quad t = 1, \ldots, T, \quad w \in \Omega,\\
&x_{s_w t}^r, \quad t = 1, \ldots, T, \quad r = 2, \ldots, T, \quad w \in \Omega,\\
&z_t^w, \quad t = 1, \ldots, T, \quad w \in \Omega.
\end{aligned}
$$

Here $x_{s_w t}^r$ is the variable for the rth item of the stochastic part at time t given the scenario $w \in \Omega$.

Now, given a scenario $w \in \Omega$, an optimal first stage replacement decision at time $t = 0$ can be computed by solving (5.8). Hence, the complete two-stage model becomes (here $R = T$, and $z_0 = 1$ has been eliminated)

$$
\begin{aligned}
\text{minimize} \quad & c_s x_{s0}^2 + \sum_{i \in \mathcal{N}} c_i x_{i0} \\
& + \sum_{w \in \Omega} p(w) \sum_{t=1}^{T} \left(c_s \sum_{r=2}^{T} (x_{s_w t}^r - x_{s_w, t-1}^r) + \sum_{i \in \mathcal{N}} c_i x_{it}^w + d z_t^w \right) \\
\text{subject to} \quad & x_{i0} + \sum_{t=1}^{\tilde{T}_i} x_{it}^w \geq 1, \quad i \in \mathcal{N}, \quad w \in \Omega, \\
& \sum_{t=\ell}^{T_i+\ell-1} x_{it}^w \geq 1, \quad \ell = 1, \ldots, T - T_i, \quad i \in \mathcal{N}, \quad w \in \Omega, \\
& x_{it}^w \leq z_t^w, \quad t = 1, \ldots, T, \quad i \in \mathcal{N}, \quad w \in \Omega, \\
& (x_{s_w 0}^1 - x_{s0}^2) + \sum_{t=1}^{T} (x_{s_w t}^1 - x_{s_w t}^2) \leq T_w^1, \quad w \in \Omega, \\
& \sum_{t=0}^{T} (x_{s_w t}^r - x_{s_w t}^{r+1}) \leq T_w^r, \quad r = 2, \ldots, T-1, \quad w \in \Omega, \\
& \sum_{t=1}^{T} x_{s_w t}^T \leq T_w^T, \quad w \in \Omega, \\
& x_{s_w 1}^2 \geq x_{s0}^2, \quad w \in \Omega, \\
& x_{s_w, t+1}^r \geq x_{s_w t}^r, \quad \begin{cases} t = 1, \ldots, T-1, \quad r = 1, \ldots, T, \\ w \in \Omega, \end{cases} \\
& x_{s_w t}^{r+1} \leq x_{s_w, t-1}^r, \quad \begin{cases} t = 1, \ldots, T, \quad r = 1, \ldots, T-1, \\ w \in \Omega, \end{cases} \\
& x_{s_w 0}^1 = 1, \quad w \in \Omega \\
& x_{s_w t}^r - x_{s_w, t-1}^r \leq z_t^w, \quad t = 1, \ldots, T, \quad r = 1, \ldots, T, \quad w \in \Omega, \\
& x_{i0}, x_{s0}^2, x_{it}^w, x_{s_w t}^r, z_t^w \in \{0, 1\}, \quad \begin{cases} t = 0, \ldots, T, \quad i \in \mathcal{N}, \\ r = 1, \ldots, T, \quad w \in \Omega. \end{cases}
\end{aligned}
$$

By allowing only the first lifetime of the stochastic part to vary, we get a computationally more tractable model. The scenarios $w \in \Omega$ then have the form $w = \tilde{T}_s^w$, where $\tilde{T}_s^w$ is the lifetime of the specimen of the stochastic part that is currently in the engine. Further, assume that the lifetime of the specimens that will be put into the engine in the future is fixed to T_s. (Typically T_s is set to the expected value of the lifetime of a new specimen of the stochastic part.) Now, given a scenario $w \in \Omega$ the problem we want to solve is given by (5.7). The first-stage variables are then $\tilde{x}_{s0}$ and x_{i0}, $i \in \mathcal{N}$. We obtain the stochastic two-stage problem to

$$
\begin{aligned}
\text{minimize} \quad & c_s\tilde{x}_{s0} + \sum_{i\in\mathcal{N}} c_i x_{i0} \\
& + \sum_{w\in\Omega} p(w)\left[\sum_{t=1}^{T}\left(\sum_{i\in\mathcal{N}} c_i x_{it}^w + c_s x_{st}^w + d z_t^w\right) + \sum_{t=1}^{\tilde{T}_s^w} c_s \tilde{x}_{st}^w\right] \qquad (5.9)\\
\text{subject to} \quad & x_{i0} + \sum_{t=1}^{\tilde{T}_i} x_{it}^w \geq 1, \quad i \in \mathcal{N}, \quad w \in \Omega,\\
& \sum_{t=\ell}^{T_i+\ell-1} x_{it}^w \geq 1, \quad \begin{cases} \ell = 1,\dots,T-T_i, \quad i \in \mathcal{N},\\ w \in \Omega,\end{cases}\\
& x_{it}^w \leq z_t^w, \quad \begin{cases} t = 1,\dots,T, \quad i \in \mathcal{N},\\ w \in \Omega,\end{cases}\\
& \tilde{x}_{s0} + \sum_{t=1}^{T_s} x_{st}^w + \sum_{t=1}^{\tilde{T}_s^w} \tilde{x}_{st}^w \geq 1, \quad w \in \Omega,\\
& \sum_{t=\ell}^{T_s+\ell-1} x_{st}^w + \sum_{t=\ell}^{\tilde{T}_s^w} \tilde{x}_{st}^w \geq 1, \quad \ell = 1,\dots,\tilde{T}_s^w, \quad w \in \Omega\\
& \sum_{t=\ell}^{T+\ell-1} x_{st}^w \geq 1, \quad \begin{cases} \ell = \tilde{T}_s^w + 1,\dots,T-T_s,\\ w \in \Omega,\end{cases}\\
& x_{st}^w \leq \tilde{x}_{s0} + \sum_{k=1}^{t-1} \tilde{x}_{sk}^w, \quad \begin{cases} t = 1,\dots,\tilde{T}_s,\\ w \in \Omega,\end{cases}\\
& \tilde{x}_{st}^w \leq z_t^w, \quad t = 1,\dots,\tilde{T}_s^w, \quad w \in \Omega,\\
& x_{st}^w \leq z_t^w, \quad t = 1,\dots,T, \quad w \in \Omega,\\
& x_{i0}, \tilde{x}_{s0}, x_{it}^w, x_{st}^w, \tilde{x}_{st}^w, z_t^w \in \{0,1\}, \quad \begin{cases} t = 0,\dots,T, \quad i \in \mathcal{N},\\ r = 1,\dots,T, \quad w \in \Omega.\end{cases}
\end{aligned}
$$

The assumption that only the first lifetime of the stochastic part is allowed to vary is motivated by the fact that the lifetime characteristics of the stochastic part in the future is more uncertain since in the future the stochastic part may be produced by other methods and perhaps with different materials, which means that the characteristics of the lifetimes change.

In Chapter 7 we will investigate how the number of scenarios in the model (5.9) affects the solution. We will compare the solution with the dynamic programming solution, which gives a lower bound on the expected total cost.

5.8 Dynamic programming versus stochastic two-stage models

The stochastic dynamic programming model finds a replacement decision x_0 at time $t = 0$ that minimizes

$$E_{y_0}\left[g_0(s_0, x_0) + \min_{x_1} E_{y_1}[g_1(s_1, x_1) + \cdots + \min_{x_T} E_{y_T}[g_T(s_T, x_T)] \cdots]\right].$$

The solution is based on the fact that between the times $t = 0$ and $t = 1$ some of the stochastic parts can fail according to the stochastic variable y_0. When the outcome of y_0 is known the replacement decision x_1 at time $t = 1$ is made, and so on.

Given the discretization of time, the stochastic dynamic programming model gives the best possible solution in the sense that the expected total cost is minimized. The drawback is that dynamic programming only can be used when the number of parts is small, so in order to handle systems with a large number of parts we must do something else. Hence we developed a linear two-stage model. This two-stage model finds a replacement decision x_0 that minimizes

$$E_w\left[c^{\mathrm{T}} x_0 + \underset{x_1,\ldots,x_T}{\text{minimum}} \sum_{t=1}^{T} c^{\mathrm{T}} x_t\right],$$

where c is the cost vector. The solution is based on the fact that after time $t = 0$ the actual outcome $w \in \Omega$ is known, that is, the lifetimes of all of the parts given in $w \in \Omega$ are known.

Therefore, the two-stage model assumes more about the future than the dynamic programming model, and hence the expected total cost when the two-stage model is used will always be at least as large as when the dynamic programming model is used. However, with the two-stage model we are at least able to solve aircraft maintenance problems (where the systems considered are large), which is not the case with the dynamic programming model.

The scenarios that are used in the dynamic programming model differ from the scenarios that are used in the two-stage model. If we use the same kind of scenarios that are used in the dynamic programming model, it is still possible to formulate a multi-stage stochastic model; such a model has the same optimal solutions as the dynamic programming model. Unfortunately, the scenarios depend on the replacement actions, which means that the multi-stage model becomes non-linear and hence hard to solve.

It is still an open question how the stochastic replacement problem should be modelled. The only practically useful model we have presented is the two-stage model, but further investigation should be made in order to properly model the stochastic replacement problem. It would be theoretically satisfactory to find a model that approximates the dynamic programming model (for example, by only considering a few of the total number of scenarios), but gives the same solution as the dynamic programming model when some parameter value (for example, the number of scenarios) becomes sufficiently large. Further, the model should be easy to solve when the parameter value (the number of scenarios) is small.

Chapter 6

On the facial structure of the replacement polytope

6.1 Introduction

In this chapter we study the structure of the set of feasible solutions to Model I presented in Section 4.3, that is, the set

$$S = \left\{ (x, z) \in \mathbb{B}^{NT} \times \mathbb{B}^{T} \;\middle|\; \sum_{t=\ell}^{T_i+\ell-1} x_{it} \geq 1, \quad \ell = 1, \ldots, T - T_i, \quad i \in \mathcal{N}; \right.$$
$$\left. x_{it} \leq z_t, \quad i \in \mathcal{N}, \quad t = 1, \ldots, T \right\}.$$

Throughout the whole chapter this set will be denoted by S. The convex hull of S will be called the *replacement polytope*.

The main goal of studying the facial structure is to be able to completely describe the convex hull of S by a finite set of linear inequalities. As will be discussed below it is then possible to solve the problem by using linear programming. Our ambition here is to take the first steps towards such a complete linear description of the replacement polytope.

In Section 6.2 we present some basic results on polyhedral combinatorics. Then in Section 6.3 we compute the dimension of the replacement polytope and investigate whether the inequalities in the original formulation in Model I define facets of the replacement polytope or not. It turns out that several of these inequalities do define facets. However, with an example we show that these "basic" inequalities do not completely define the convex hull. In Section 6.4 we derive a new facet for the example problem by using Chvátal–Gomory rounding. The chapter concludes in Section 6.5 with suggestions on further studies on the facial structure of the replacement polytope.

6.2 Polyhedral combinatorics

In this section we present some basic results on polyhedral combinatorics that will be used in the subsequent sections. General text books on polyhedral combinatorics are [92], [60], [98], [114], [115], [134], and [34].

First we will define *dimension* of a general set in $\mathbb{R}^n$. In order to do so we introduce the notation of affine set, affine combination, affine hull, affinely dependent set, and affinely independent set.

Definition 6.1 (affine sets and dimension) *Let X be a subset of $\mathbb{R}^n$.*

(a) (affine set) *The set X is an* affine set *if $\lambda x + \mu y \in X$ whenever $x, y \in X$ and $\lambda, \mu \in \mathbb{R}$ are such that $\lambda + \mu = 1$.*

(b) (affine combination) *A point $x \in \mathbb{R}^n$ is an* affine combination *of the points $x^1, \ldots, x^m \in \mathbb{R}^n$ if there exist scalars $\lambda_1, \ldots, \lambda_m$ with $\lambda_1 + \cdots + \lambda_m = 1$ such that $x = \lambda_1 x^1 + \cdots + \lambda_m x^m$.*

(c) (affine hull) *The* affine hull *of X, denoted by $\operatorname{aff} X$, is the set of all (finite) affine combinations of points of X.*

(d) (affinely dependent set) *The set X is* affinely dependent *if there exists an $x \in X$ such that $x \in \operatorname{aff}(A \setminus \{x\})$.*

(d) (affinely independent set) *If the set X is not affinely dependent it is* affinely independent.

(e) (dimension) *The* dimension *of the set X, denoted by $\dim X$, is one less than the maximum cardinality of an affinely independent set $K \subseteq X$.* ∎

A *polyhedron* in $\mathbb{R}^n$ is a set of the form

$$P = \{\, x \in \mathbb{R}^n \mid Ax \leq b \,\}, \tag{6.1}$$

where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. The *equality subsystem* $(A^=, b^=)$ of P is defined by the rows of the system $Ax \leq b$ that are fulfilled with equality for all $x \in P$. The matrix $A^=$ will be referred to as the *matrix corresponding to the equality subsystem* of P.

Proposition 6.2 (dimension of a polyhedron) *If $P \subseteq \mathbb{R}^n$ is a polyhedron, then*

$$\dim(P) + \operatorname{rank}(A^=, b^=) = n.$$

Proof. See [92, p. 87]. ∎

If $\dim P = n$ we say that P is *full-dimensional*.

We are mainly interested in the convex hull of the set S of feasible solutions to the replacement problem. We define the notion of convex hull through the notation of convex combination, and then we define the geometrical object polytope.

Definition 6.3 (convex sets and polytopes) *Let X be a subset of $\mathbb{R}^n$.*

(a) (convex set) *The set X is a* convex set *if $\lambda x + \mu y \in X$ whenever $x, y \in X$ and $\lambda, \mu \geq 0$ are such that $\lambda + \mu = 1$.*

(b) (convex combination) *A point $x \in \mathbb{R}^n$ is a* convex combination *of the points $x^1, \dots, x^m \in \mathbb{R}^n$ if there exist scalars $\lambda_1, \dots, \lambda_m \geq 0$ with $\lambda_1 + \dots + \lambda_m = 1$ such that $x = \lambda_1 x^1 + \dots + \lambda_m x^m$.*

(c) (convex hull) *The* convex hull *of the set X, denoted by* $\operatorname{conv} X$, *is the set of all (finite) convex combinations of points in X.*

(d) (polytope) *A polytope is the convex hull of a finite set of points in $\mathbb{R}^n$.* ■

In view of the definition of a polytope it is natural to call $\operatorname{conv} S$ the *replacement polytope* (or the polytope of the replacement problem).

Every polytope can be characterized as the convex hull of its extreme points.

Definition 6.4 (extreme point) *A point x in a convex set $X \subseteq \mathbb{R}^n$ is an extreme point of X if whenever $x = \lambda x^1 + (1-\lambda)x^2$, where $x^1, x^2 \in X$ and $\lambda \in (0, 1)$, then $x^1 = x^2 = x$.* ■

Proposition 6.5 *Let V be a finite set in $\mathbb{R}^n$ and let $X = \operatorname{conv} V$. Then each extreme point of X lies in V.*

Proof. See [134, p. 81]. ■

Proposition 6.6 *Every polytope is equal to the convex hull of its extreme points.*

Proof. See [34, p. 206]. ■

Another useful result is that every polytope is a polyhedron.

Proposition 6.7 *A set is a polytope if and only if it is a bounded polyhedron.*

Proof. See [134, p. 114] ■

There is an obvious relation between the dimension of a set $X \subseteq \mathbb{R}^n$ and that of $\operatorname{conv} X$.

Proposition 6.8 *Let $X \subseteq \mathbb{R}^n$, then* $\dim X = \dim(\operatorname{conv} X)$. ■

If all of the extreme points of a polyhedron are integral the polyhedron is called *integral*. We turn to a sufficient condition for a polyhedron to be integral.

Definition 6.9 (totally unimodular matrix) *A matrix is called* totally unimodular, *in short TU, if all of its square submatrices have determinant* 0, 1, *or* -1. ■

Proposition 6.10 *Let $A \in \mathbb{R}^{m \times n}$ be a totally unimodular matrix and let $b \in \mathbb{R}^m$ be integral. Then the polyhedron defined by $Ax \le b$ is integral.*

Proof. See [34, p. 221]. ■

We will use the following characterization of total unimodularity.

Proposition 6.11 (characterization of the TU property) *Let $A \in \mathbb{R}^{m \times n}$. The following statements are equivalent:*

(i) *A is TU;*

(ii) *For every $J \subseteq \{1, \dots, n\}$ there exists a partition J_1, J_2 of J such that*

$$\left| \sum_{j \in J_1} a_{ij} - \sum_{j \in J_2} a_{ij} \right| \le 1, \quad i = 1, \dots, m.$$

Proof. See [92, p. 543]. ■

We introduce the important concept of face and facet of a polyhedron.

Definition 6.12 (faces and facets of a polyhedron) *Let P be given by* (6.1).

(a) (valid inequality) *The inequality $\pi x \le \pi_0$ is called* a valid inequality *for P if it is satisfied by all points in P.*

(b) (face) *If $\pi x \le \pi_0$ is a valid inequality for P, and*

$$F = \{ x \in P \mid \pi x = \pi_0 \},$$

then F is called a face *of P, and we say that $\pi x \le \pi_0$ defines F. A face is said to be proper if $F \ne \emptyset$* and *$F \ne P$.*

(c) (facet) *A face F of P is called* a facet *of P if* $\dim F = \dim P - 1$. ■

Proposition 6.13 *If F is a facet of P, then there exists some inequality defining F.*

Proof. See [92, p. 89]. ■

We now arrive at the crucial result that every full-dimensional polyhedron can be uniquely represented by its facets.

Proposition 6.14 *A full-dimensional polyhedron P has a unique (to within scalar multiplication) minimal representation by a finite set of linear inequalities. In particular, for each facet F_i of P there is an inequality $a^i x \le b_i$ (unique within scalar multiplication) representing F_i and $P = \{ x \in \mathbb{R}^n \mid a^i x \le b_i,\ i = 1, \dots, k \}$.*

Proof. See [92, p. 91]. ■

Proposition 6.7 implies that if $X \subseteq \mathbb{R}^n$ is a finite set, then the polytope $\operatorname{conv} X$ is a polyhedron. Hence, if $\operatorname{conv} X$ is full-dimensional, from Proposition 6.14 we see that if we can find all of the facet-defining inequalities of $\operatorname{conv} X$, then we have a linear description of it. Therefore, it is of interest to find facets of a given polytope. Sometimes the following characterization, which is based on the uniqueness in Proposition 6.14, is useful when proving that a certain valid inequality is a facet.

Proposition 6.15 *Let P be a full-dimensional polyhedron and let $F = \{ x \in P \mid \pi x = \pi_0 \}$ be a proper face of P. Then the following two statements are equivalent:*

(i) *F is a facet of P;*

(ii) *If $\lambda x = \lambda_0$ for all $x \in F$, then $(\lambda, \lambda_0) = \alpha(\pi, \pi_0)$ holds for some $\alpha \in \mathbb{R}$.*

Proof. See [92, pp. 91–92] ■

We close this section by remarking that if we can find a polyhedral description of $\operatorname{conv} S$, then the replacement problem can be solved by using standard linear programming algorithms. Indeed, if the Simplex method is used to find an optimal solution to the linear program $\min\{c^{\mathrm{T}}x \mid x \in \operatorname{conv} X\}$ an optimal extreme point of $\operatorname{conv} X$ will be found. From Proposition 6.5 it follows that all of the extreme points of $\operatorname{conv} X$ belong to X, so in fact we have found an optimal solution to $\min\{c^{\mathrm{T}}x \mid x \in X\}$.

6.3 The dimension and the basic facets

In this section we derive the dimension of the replacement polytope $\operatorname{conv} S$ and investigate the inequalities from the definition of S. Under natural assumptions we show that the replacement polytope is full-dimensional. Further, we show that all inequalities that are necessary in the original formulation of the replacement problem are facets of the replacement polytope.

Lemma 6.16 *The polyhedron defined by*

$$\sum_{t=\ell}^{T_i+\ell-1} x_{it} \geq 1, \quad \ell = 1, \ldots, T - T_i, \quad i \in \mathcal{N}, \tag{6.2a}$$

$$-x_{it} \geq -1, \quad i \in \mathcal{N}, \quad t = 1, \ldots, T \tag{6.2b}$$

is integral.

Proof. We derive the result by showing that the constraint matrix is totally unimodular.

First we observe that all of the elements in the constraint matrix belong to $\{-1, 0, 1\}$. Each row of the upper part of the constraint matrix defined by (6.2a) consists of ones that appear consecutively, that is, if $a_{ij} = a_{ik} = 1$ and $k > j+1$, then $a_{i\ell} = 1$ for all ℓ with $j < \ell < k$ (see Example 4.5). Further, each row of the lower part of the constraint matrix defined by (6.2b) consists of a single -1. These properties of the complete constraint matrix are closed under column deletions. Therefore it is enough to show that the assumptions in Proposition 6.11 are satisfied for the complete constraint matrix. Let $J = \{1, \ldots, NT\}$, $J_1 = \{j \in J \mid j \text{ odd}\}$, and $J_2 = J \setminus J_1$. Consider the constraint of (6.2a) defined by

$$x_{11} + x_{12} + \cdots + x_{1T_1} \geq 1.$$

Let $a \in \mathbb{R}^{NT}$ be the entries in the constraint matrix corresponding to this constraint, that is, $a_j = 1$ for $j = 1, \ldots, T_1$ and $a_j = 0$ for $j = T_1 + 1, \ldots, NT$. If T_1 is odd we get that

$$\sum_{j \in J_1} a_j - \sum_{j \in J_2} a_j = a_1 - a_2 + a_3 - a_4 + \cdots + a_{T_1} = 1,$$

and if T_1 is even we get that

$$\sum_{j \in J_1} a_j - \sum_{j \in J_2} a_j = a_1 - a_2 + a_3 - a_4 + \cdots + a_{T_1 - 1} - a_{T_1} = 0.$$

Hence

$$\left| \sum_{j \in J_1} a_j - \sum_{j \in J_2} a_j \right| \leq 1. \tag{6.3}$$

Similarly, (6.3) holds for all the rows of the constraint matrix corresponding to the constraints defined by (6.2a). Since the rows of the constraint matrix corresponding to (6.2b) only consists of a single -1, these rows obviously satisfy (6.3). Therefore we have shown that the assumptions in Proposition 6.11 are satisfied and it follows that the constraint matrix is TU.

Now, since the constraint matrix is TU and the right-hand side consists of integers it follows from Proposition 6.10 that the polyhedron it defines is integral. We are done. ■

Proposition 6.17 (dimension of the replacement polytope) *If $T_i \geq 2$ for all $i \in \mathcal{N}$, then the dimension of $\operatorname{conv} S$ is $NT + T$, that is, $\operatorname{conv} S$ is full-dimensional.*

Proof. First note that since $S \subseteq \mathbb{R}^{NT+T}$ it holds that

$$\dim(\operatorname{conv} S) \leq NT + T. \tag{6.4}$$

Then consider the set Q consisting of all $x \in \mathbb{R}^{NT}$ such that

$$\sum_{t=\ell}^{T_i+\ell-1} x_{it} \geq 1, \quad \ell = 1, \ldots, T - T_i, \quad i \in \mathcal{N},$$
$$x_{it} \leq 1, \quad i \in \mathcal{N}, \quad t = 1, \ldots, T.$$

If $x_{it} = 1$ for all $i \in \mathcal{N}$ and $t = 1, \ldots, T$, it holds that $x \in Q$, and since $T_i \geq 2$ for all $i \in \mathcal{N}$ we have that

$$\sum_{t=\ell}^{T_i+\ell-1} x_{it} > 1, \quad \ell = 1, \ldots, T - T_i, \quad i \in \mathcal{N}.$$

Further, given $i \in \mathcal{N}$ and $t \in \{1, \ldots, T\}$, since $T_i \geq 2$ for all $i \in \mathcal{N}$ we can always find a vector $x \in Q$ such that $x_{it} = 0$. Hence the rank of the matrix corresponding to the equality subsystem of Q is zero and Proposition 6.2 gives that

$$\dim Q = NT.$$

From Lemma 6.16 we have that Q is integral, and Proposition 6.6 gives that Q equals the convex hull of its extreme points. Hence from Proposition 6.8 it follows that there exists $NT + 1$ affinely independent integral vectors $y^1, \ldots, y^{NT+1} \in Q$. But this implies that the vectors

$$q^1 = \begin{pmatrix} y^1 \\ \mathbf{1} \end{pmatrix}, \ldots, q^{NT+1} = \begin{pmatrix} y^{NT+1} \\ \mathbf{1} \end{pmatrix},$$

where $\mathbf{1} \in \mathbb{R}^T$ is a vector of 1's (corresponding to the z-variables), are affinely independent vectors in S. Now, since $T_i \geq 2$, for $k = 1, \ldots, T$ there exist vectors in S of the form

$$q^{NT+1+k} = \begin{pmatrix} y^{NT+1+k} \\ \mathbf{1} - e_k \end{pmatrix},$$

where $e_k = (0, \ldots, 0, 1, 0, \ldots, 0)^{\mathrm{T}} \in \mathbb{R}^T$ is the kth unit vector. Further it holds that $q^{NT+1+k} \notin \operatorname{aff}\{q^1, \ldots, q^{NT+k}\}$, for $k = 1, \ldots, T$, so $q^1, \ldots, q^{NT+1+T}$ are affinely independent. Hence, $\dim(\operatorname{conv} S) \geq NT + T$ holds. Together with (6.4) this implies that $\dim(\operatorname{conv} S) = NT + T$. ■

Remark 6.18 The replacement polytope is not full-dimensional if $T_i \leq 1$ for some $i \in \mathcal{N}$. For example, if $T_1 = 1$, then it holds that $x_{11} = 1$ for all $(x, z) \in \operatorname{conv} S$. This means that the rank of the matrix corresponding to the equality subsystem of $\operatorname{conv} S$ is greater than or equal to one. Hence, from Proposition 6.2 it follows that $\dim(\operatorname{conv} S) \leq NT + T - 1$. However, the case that $T_i = 1$ for some $i \in \mathcal{N}$ is not very interesting in practice since it means that part i must be replaced every point of time. ■

Proposition 6.19 *If $T_i \geq 2$ for all $i \in \mathcal{N}$, then for each $i \in \mathcal{N}$ and $\ell \in \{1, \dots, T - T_i\}$ the inequality*

$$\sum_{t=\ell}^{T_i+\ell-1} x_{it} \geq 1$$

defines a facet of $\operatorname{conv} S$.

Proof. Since $T_i \geq 2$ it follows from Proposition 6.17 that $\operatorname{conv} S$ is full-dimensional. Hence we can use the uniqueness characterization in Proposition 6.15 to show the assertion. Frequently the fact that $T_i \geq 2$ for all $i \in \mathcal{N}$ will be used implicitly to motivate the existence of the points constructed.

We show that the inequality (corresponding to $i = \ell = 1$)

$$\sum_{t=1}^{T_1} x_{1t} \geq 1 \tag{6.5}$$

defines a facet of $\operatorname{conv} S$. The corresponding proofs for the other inequalities in the proposition are analogous. Let

$$F = \left\{ (x, z) \in \operatorname{conv} S \;\middle|\; \sum_{t=1}^{T_1} x_{1t} = 1 \right\},$$

where $x = (x_{11}, \dots, x_{1T}, \dots, x_{N1}, \dots, x_{NT})^{\mathrm{T}}$ and $z = (z_1, \dots, z_T)^{\mathrm{T}}$. Assume that the equality $\lambda x + \mu z = \rho$ for some $\lambda^{\mathrm{T}} \in \mathbb{R}^{NT}$, $\mu^{\mathrm{T}} \in \mathbb{R}^{T}$, and $\rho \in \mathbb{R}$, holds for all $(x, z) \in F$. This equality can be component-wise expressed as

$$\sum_{i \in \mathcal{N}} \sum_{t=1}^{T} \lambda_{it} x_{it} + \sum_{t=1}^{T} \mu_t z_t = \rho. \tag{6.6}$$

Construct a point in F according to

$$x_{11} = 1; \quad x_{1t} = 0, \quad 2 \leq t \leq T_1; \quad x_{1t} = 1, \quad T_1 + 1 \leq t \leq T, \tag{6.7a}$$
$$x_{it} = 1, \quad i \in \mathcal{N} \setminus \{1\}, \quad t = 1, \dots, T, \tag{6.7b}$$
$$z_t = 1, \quad t = 1, \dots, T. \tag{6.7c}$$

Then construct another point in F according to

$$x_{11} = 1; \quad x_{1t} = 0, \quad 2 \leq t \leq T_1; \quad x_{1t} = 1, \quad T_1 + 1 \leq t \leq T, \tag{6.8a}$$
$$x_{21} = 0; \quad x_{2t} = 1, \quad t = 2, \dots, T, \tag{6.8b}$$
$$x_{it} = 1, \quad i \in \mathcal{N} \setminus \{1, 2\}, \quad t = 1, \dots, T, \tag{6.8c}$$
$$z_t = 1, \quad t = 1, \dots, T. \tag{6.8d}$$

Now, if the points given by (6.7) and (6.8) are inserted in (6.6) we get that

$$\lambda_{11} + \sum_{t=T_1+1}^{T} \lambda_{1t} + \sum_{t=1}^{T} \left(\sum_{i \in \mathcal{N} \setminus \{1\}} \lambda_{it} + \mu_t \right) = \rho, \tag{6.9}$$

$$\lambda_{11} + \sum_{t=T_1+1}^{T} \lambda_{1t} + \sum_{t=2}^{T} \lambda_{2t} + \sum_{t=1}^{T} \left(\sum_{i \in \mathcal{N} \setminus \{1,2\}} \lambda_{it} + \mu_t \right) = \rho. \tag{6.10}$$

By subtracting (6.10) from (6.9) it follows that $\lambda_{21} = 0$. Similarly, for each $k \in \mathcal{N} \setminus \{1\}$ and $\ell \in \{1, \ldots, T\}$ the point in F defined by

$$\begin{aligned} x_{11} &= 1; \quad x_{1t} = 0, \quad 2 \le t \le T_1; \quad x_{1t} = 1, \quad T_1 + 1 \le t \le T, \\ x_{k\ell} &= 0; \quad x_{kt} = 1, \quad t \in \{1, \ldots, T\} \setminus \{\ell\}, \\ x_{it} &= 1, \quad i \in \mathcal{N} \setminus \{1, k\}, \quad t = 1, \ldots, T, \\ z_t &= 1, \quad t = 1, \ldots, T, \end{aligned}$$

together with the point given by (6.7) can be utilized to show that $\lambda_{k\ell} = 0$. Hence we have that

$$\lambda_{it} = 0, \quad i \in \mathcal{N} \setminus \{1\}, \quad t = 1, \ldots, T. \tag{6.11}$$

In the same fashion the points in F given by

$$\begin{aligned} x_{11} &= 0; \quad x_{12} = 1; \quad x_{1t} = 0, \quad 3 \le t \le T_1; \quad x_{1t} = 1, \quad T_1 + 1 \le t \le T, \\ x_{it} &= 1, \quad i \in \mathcal{N} \setminus \{1\}, \quad t = 1, \ldots, T, \\ z_t &= 1, \quad t = 1, \ldots, T, \end{aligned}$$

and

$$\begin{aligned} x_{11} &= 0; \quad x_{12} = 1; \quad x_{1t} = 0, \quad 3 \le t \le T_1, \\ x_{1\ell} &= 0; \quad x_{1t} = 1, \quad t \in \{T_1 + 1, \ldots, T\} \setminus \{\ell\}, \\ x_{it} &= 1, \quad i \in \mathcal{N} \setminus \{1\}, \quad t = 1, \ldots, T, \\ z_t &= 1, \quad t = 1, \ldots, T, \end{aligned}$$

where $\ell \in \{T_1 + 1, \ldots, T\}$, imply that $\lambda_{1\ell} = 0$. Therefore, we have that

$$\lambda_{1t} = 0, \quad t = T_1 + 1, \ldots, T. \tag{6.12}$$

Further, the points in F given by

$$\begin{aligned} x_{11} &= 1; \quad x_{1t} = 0, \quad 2 \le t \le T_1; \quad x_{1t} = 1, \quad T_1 + 1 \le t \le T, \\ x_{i1} &= 1; \quad x_{i2} = 0; \quad x_{it} = 1, \quad i \in \mathcal{N} \setminus \{1\}, \quad t = 3, \ldots, T, \\ z_t &= 1, \quad t = 1, \ldots, T, \end{aligned}$$

and

$$\begin{aligned} x_{11} &= 1; \quad x_{1t} = 0, \quad 2 \le t \le T_1; \quad x_{1t} = 1, \quad T_1 + 1 \le t \le T, \\ x_{i1} &= 1; \quad x_{i2} = 0; \quad x_{it} = 1, \quad i \in \mathcal{N} \setminus \{1\}, \quad t = 3, \dots, T, \\ z_1 &= 1; \quad z_2 = 0; \quad z_t = 1, \quad t = 3, \dots, T, \end{aligned}$$

yield that $\mu_2 = 0$. Similarly, for $\ell \in \{1, \dots, T\} \setminus \{2\}$ the points

$$\begin{aligned} x_{11} &= 0; \quad x_{12} = 1; \quad x_{1t} = 0, \quad 3 \le t \le T_1, \\ x_{1\ell} &= 0; \quad x_{1t} = 1, \quad t \in \{T_1 + 1, \dots, T\} \setminus \{\ell\}, \\ x_{i\ell} &= 0; \quad x_{it} = 1, \quad i \in \mathcal{N} \setminus \{1\}, \quad t \in \{1, \dots, T\} \setminus \{\ell\}, \\ z_t &= 1, \quad t = 1, \dots, T, \end{aligned}$$

and

$$\begin{aligned} x_{11} &= 0; \quad x_{12} = 1; \quad x_{1t} = 0, \quad 3 \le t \le T_1, \\ x_{1\ell} &= 0; \quad x_{1t} = 1, \quad t \in \{T_1 + 1, \dots, T\} \setminus \{\ell\}, \\ x_{i\ell} &= 0; \quad x_{it} = 1, \quad i \in \mathcal{N} \setminus \{1\}, \quad t \in \{1, \dots, T\} \setminus \{\ell\}, \\ z_\ell &= 0; \quad z_t = 1, \quad t \in \{1, \dots, T\} \setminus \{\ell\}, \end{aligned}$$

yield that $\mu_\ell = 0$. Hence, it follows that

$$\mu_t = 0, \quad t = 1, \dots, T. \tag{6.13}$$

Finally, for $\ell \in \{1, \dots, T_1\}$ the point in F given by

$$\begin{aligned} x_{1\ell} &= 1; \quad x_{1t} = 0, \quad t \in \{1, \dots, T_1\} \setminus \{\ell\}; \quad x_{1t} = 1, \quad T_1 + 1 \le t \le T, \\ x_{it} &= 1, \quad i \in \mathcal{N} \setminus \{1\}, \quad t = 1, \dots, T, \\ z_t &= 1, \quad t = 1, \dots, T. \end{aligned}$$

together with (6.11), (6.12) and (6.13) give that $\lambda_{1\ell} = \rho$, so

$$\lambda_{1t} = \rho, \quad t = 1, \dots, T_1.$$

Hence we have shown that the equality $\lambda x + \mu z = \rho$ has the form

$$\rho \sum_{t=1}^{T_1} x_{1t} = \rho,$$

and it follows from Proposition 6.15 (with $\alpha = \rho$) that (6.5) defines a facet of $\operatorname{conv} S$. ■

Proposition 6.20 *If $T_i \ge 2$ for all $i \in \mathcal{N}$, then for each $i \in \mathcal{N}$ and $t = 1, \dots, T$ the inequality*

$$x_{it} \le z_t$$

defines a facet of $\operatorname{conv} S$.

Proof. We show that

$$x_{11} \le z_1 \tag{6.14}$$

defines a facet of $\operatorname{conv} S$. The corresponding proofs for the other inequalities in the proposition are analogous. Let

$$F = \{\, (x, z) \in \operatorname{conv} S \mid x_{11} = z_1 \,\}.$$

We use the same notation and method as in the proof of Proposition 6.19. As in the proof of this result, we assume that $\lambda x + \mu z = \rho$ for all $(x, z) \in F$.

The points in F given by

$$x_{it} = 1, \quad i \in \mathcal{N}, \quad t = 1, \ldots, T, \tag{6.15a}$$
$$z_t = 1, \quad t = 1, \ldots, T, \tag{6.15b}$$

and

$$\begin{aligned} x_{1\ell} &= 0; \quad x_{1t} = 1, \quad t \in \{1, \ldots, T\} \setminus \{\ell\}, \\ x_{it} &= 1, \quad i \in \mathcal{N} \setminus \{1\}, \quad t = 1, \ldots, T, \\ z_t &= 1, \quad t = 1, \ldots, T, \end{aligned}$$

where $\ell \in \{2, \ldots, T\}$, imply that $\lambda_{1\ell} = 0$. Similarly, for $k \in \mathcal{N} \setminus \{1\}$ and $\ell \in \{1, \ldots, T\}$ the point

$$\begin{aligned} x_{k\ell} &= 0; \quad x_{kt} = 1, \quad t \in \{1, \ldots, T\} \setminus \{\ell\}, \\ x_{it} &= 1, \quad i \in \mathcal{N} \setminus \{k\}, \quad t = 1, \ldots, T, \\ z_t &= 1, \quad t = 1, \ldots, T, \end{aligned}$$

together with the point given by (6.15) yield that $\lambda_{k\ell} = 0$. Hence, we have shown that

$$\lambda_{1t} = 0, \quad t = 2, \ldots, T, \tag{6.16a}$$
$$\lambda_{it} = 0, \quad i \in \mathcal{N} \setminus \{1\}, \quad t = 1, \ldots, T. \tag{6.16b}$$

Further, the points in F given by (6.15) and

$$\begin{aligned} x_{i\ell} &= 0; \quad x_{it} = 1, \quad i \in \mathcal{N}, \quad t \in \{1, \ldots, T\} \setminus \{\ell\}, \\ z_\ell &= 0; \quad z_t = 1, \quad t \in \{1, \ldots, T\} \setminus \{\ell\}, \end{aligned}$$

where $\ell \in \{2, \ldots, T\}$, together with (6.16) give that $\mu_\ell = 0$, so

$$\mu_t = 0, \quad t = 2, \ldots, T. \tag{6.17}$$

Finally, the points in F given by (6.15) and

$$\begin{aligned} x_{i1} &= 0; \quad x_{it} = 1, \quad i \in \mathcal{N}, \quad t = 2, \ldots, T, \\ z_1 &= 0; \quad z_t = 1, \quad t = 2, \ldots, T, \end{aligned}$$

imply that $\lambda_{11} = -\mu_1$, which together with (6.16) and (6.17) yield that the equation $\lambda x + \mu z = \rho$ has the form

$$\alpha(x_{11} - z_1) = \rho.$$

But since $x_{11} = z_1$ for all $(x, z) \in F$ it follows that $\rho = 0$. Hence Proposition 6.15 gives that (6.14) defines a facet of $\operatorname{conv} S$. ∎

Proposition 6.21 *If $T_i \geq 2$ for all $i \in \mathcal{N}$, then for each $t = 1, \dots, T$ the inequality*

$$z_t \leq 1$$

defines a facet of $\operatorname{conv} S$.

Proof. We show that

$$z_1 \leq 1 \tag{6.18}$$

defines a facet of $\operatorname{conv} S$. The corresponding proofs for the other inequalities in the proposition are analogous. Let

$$F = \{\, (x, z) \in \operatorname{conv} S \mid z_1 = 1 \,\}.$$

We use the same notation and method as in the proof of Proposition 6.19.

The points in F given by (6.15) and

$$\begin{aligned} x_{k\ell} &= 0; \quad x_{kt} = 1, \quad t \in \{1, \dots, T\} \setminus \{\ell\}, \\ x_{it} &= 1, \quad i \in \mathcal{N} \setminus \{k\}, \quad t = 1, \dots, T, \\ z_t &= 1, \quad t = 1, \dots, T, \end{aligned}$$

where $k \in \mathcal{N}$ and $\ell = 1, \dots, T$, imply that $\lambda_{k\ell} = 0$. Therefore, it follows that

$$\lambda_{it} = 0, \quad i \in \mathcal{N}, \quad t = 1, \dots, T. \tag{6.19}$$

Further, the points in F given by (6.15) and

$$\begin{aligned} x_{i\ell} &= 0; \quad x_{it} = 1, \quad i \in \mathcal{N}, \quad t \in \{1, \dots, T\} \setminus \{\ell\}, \\ z_\ell &= 0; \quad z_t = 1, \quad t \in \{1, \dots, T\} \setminus \{\ell\}, \end{aligned}$$

where $\ell \in \{2, \dots, T\}$, together with (6.19) yield that $\mu_\ell = 0$, so

$$\mu_t = 0, \quad t = 2, \dots, T. \tag{6.20}$$

Now, from (6.19) and (6.20) it follows that the equality $\lambda x + \mu z = \rho$ has the form

$$\mu_1 z_1 = \rho,$$

but since $z_1 = 1$ for all $(x, z) \in F$ we have that $\mu_1 = \rho$. Therefore, it follows from Proposition 6.15 that (6.18) defines a facet of $\operatorname{conv} S$. ∎

Proposition 6.22 *If $T_k \geq 3$ for some $k \in \mathcal{N}$ and $T_i \geq 2$ for $i \in \mathcal{N} \setminus \{k\}$, then for each $t = 1, \ldots, T$ the inequality*

$$x_{kt} \geq 0$$

defines a facet of $\operatorname{conv} S$.

Proof. We consider the case where $k = 1$ and show that

$$x_{11} \geq 0 \tag{6.21}$$

defines a facet of $\operatorname{conv} S$. The corresponding proofs for the other inequalities in the proposition are analogous. Let

$$F = \{\, (x, z) \in \operatorname{conv} S \mid x_{11} = 0 \,\}. \tag{6.22}$$

We show that (6.21) defines a facet by constructing $NT+T$ affinely independent vectors in F.

Consider the set, denoted by Q, consisting of all $x \in \mathbb{R}^{NT}$ such that

$$x_{11} = 0, \tag{6.23a}$$

$$\sum_{\ell=1}^{T_i+\ell-1} x_{i\ell} \geq 1, \quad \ell = 1, \ldots, T - T_i, \quad i \in \mathcal{N}, \tag{6.23b}$$

$$x_{it} \leq 1, \quad i \in \mathcal{N}, \quad t = 1, \ldots, T. \tag{6.23c}$$

Since $T_1 \geq 3$ and $T_i \geq 2$ for all $i \in \mathcal{N} \setminus \{1\}$ the same arguments as in the proof of Proposition 6.17 show that the rank of the matrix corresponding to the equality subsystem of Q is one. Hence $\dim Q = NT - 1$. Then, in the same way as in the proof of Proposition 6.17, Lemma 6.16 can be used in order to guarantee the existence of $NT + T$ affinely independent vectors in F. Hence, it follows that $\dim F = NT + T - 1$, and we are done. ∎

Remark 6.23 The inequalities in Proposition 6.22 do not define facets if $T_k \leq 2$. For example, if $T_1 = 2$, then $x_{11} = 0$ implies that $x_{12} = 1$ which means that the rank of the matrix corresponding to the equality system of (6.23) is greater than or equal to 2, and it follows that $\dim F \leq NT + N - 2$, where F is defined as in (6.22), which means that F is not a facet of $\operatorname{conv} S$. ∎

By noting that $x_{it} \leq 1$ for $i \in \mathcal{N}$ and $t = 1, \ldots, T$, and $z_t \geq 0$ for $t = 1, \ldots, T$ are not necessary in order to describe the set S (since $x_{it} \leq z_t$ holds for all $i \in \mathcal{N}$ and $t = 1, \ldots, T$) we have that S equals the set

$$\sum_{t=\ell}^{T_i+\ell-1} x_{it} \geq 1, \quad \ell = 1, \ldots, T - T_i, \quad i \in \mathcal{N}, \tag{6.24a}$$

$$x_{it} \leq z_t, \quad i \in \mathcal{N}, \quad t = 1, \ldots, T, \tag{6.24b}$$

$$x_{it} \geq 0, \quad i \in \mathcal{N}, \quad t = 1, \ldots, T, \tag{6.24c}$$

$$z_t \leq 1, \quad t = 1, \ldots, T, \tag{6.24d}$$

$$x_{it},\, z_t \in \mathbb{Z}, \quad i \in \mathcal{N}, \quad t = 1, \ldots, T. \tag{6.24e}$$

From the Propositions 6.19–6.22 follows that all of the inequalities necessary in the description of S define facets of $\operatorname{conv} S$. A natural question arises: Is $\operatorname{conv} S$ completely described by the continuous relaxation of (6.24)? Unfortunately, this is not the case, which is shown by the following example.

Example 6.24 (continuous relaxation) Consider a system with $N = 2$, $T_1 = 3$, $T_2 = 4$, and $T = 5$. Then the problem to

$$\begin{aligned} \text{minimize} \quad & x_{11} + x_{12} + 2x_{13} + x_{14} + x_{21} + 100x_{22} + 100x_{23} + x_{24} \\ & + 10z_1 + 10z_2 + z_3 + 10z_4 \\ \text{subject to} \quad & (6.24), \end{aligned}$$

has the optimal solution

$$(x_{11}, x_{12}, x_{13}, x_{14}) = (0, 0, 1, 0), \tag{6.25a}$$
$$(x_{21}, x_{22}, x_{23}, x_{24}) = (1, 0, 0, 0), \tag{6.25b}$$
$$(z_1, z_2, z_3, z_4) = (1, 0, 1, 0), \tag{6.25c}$$

with objective function value 14. However, if we relax the integrality requirements, we get the optimal solution

$$(x_{11}, x_{12}, x_{13}, x_{14}) = (0.5, 0, 0.5, 0.5), \tag{6.26a}$$
$$(x_{21}, x_{22}, x_{23}, x_{24}) = (0.5, 0, 0, 0.5), \tag{6.26b}$$
$$(z_1, z_2, z_3, z_4) = (0.5, 0, 0.5, 0.5), \tag{6.26c}$$

with objective function value 13.5. Hence the convex hull of feasible solutions to (6.24) is not completely defined by the inequalities in (6.24). ■

6.4 A new class of facets: An example

Example 6.24 shows that the inequalities in (6.24) are not sufficient to describe $\operatorname{conv} S$. However, according to the Propositions 6.19–6.22 all of the inequalities in (6.24) define facets of $\operatorname{conv} S$. Since by Proposition 6.17 $\operatorname{conv} S$ is full-dimensional (under the reasonable assumption that $T_i \geq 2$ for all $i \in \mathcal{N}$) the minimal description of $\operatorname{conv} S$ is unique. Therefore, all of the inequalities in (6.24) are necessary in the description of $\operatorname{conv} S$.

To completely describe $\operatorname{conv} S$ we need however also facets other than those in (6.24). In this section we study the replacement polytope that arises in Example 6.24, that is, the convex hull of the set of all $x \in \{0,1\}^{2\times 4}$ and $z \in$

$\{0,1\}^4$ such that

$$
\begin{aligned}
&x_{11}+x_{12}+x_{13} && \geq 1, && \text{(6.27a)}\\
&x_{12}+x_{13}+x_{14} && \geq 1, && \text{(6.27b)}\\
&x_{21}+x_{22}+x_{23}+x_{24} && \geq 1, && \text{(6.27c)}\\
&x_{11} && \leq z_1, && \text{(6.27d)}\\
&x_{12} && \leq z_2, && \text{(6.27e)}\\
&x_{13} && \leq z_3, && \text{(6.27f)}\\
&x_{14} && \leq z_4, && \text{(6.27g)}\\
&x_{21} && \leq z_1, && \text{(6.27h)}\\
&x_{22} && \leq z_2, && \text{(6.27i)}\\
&x_{23} && \leq z_3, && \text{(6.27j)}\\
&x_{24} && \leq z_4. && \text{(6.27k)}
\end{aligned}
$$

We denote this set by T.

By using Chvátal–Gomory rounding (see [92, p. 210]) we construct a new valid inequality:

Proposition 6.25 *The inequality*

$$z_1 + x_{12} + x_{13} + x_{22} + x_{23} + z_4 \geq 2 \tag{6.28}$$

is a valid inequality for T.

Proof. According to (6.27d), (6.27g), (6.27h), and (6.27k) it follows from (6.27a), (6.27b), and (6.27c) that

$$
\begin{aligned}
z_1 + x_{12} + x_{13} &\geq 1, && \text{(6.29a)}\\
x_{12} + x_{13} + z_4 &\geq 1, && \text{(6.29b)}\\
z_1 + x_{22} + x_{23} + z_4 &\geq 1, && \text{(6.29c)}
\end{aligned}
$$

respectively, are valid for T. Multiplying each of the inequalities (6.29) by 1/2 and summing them result in the valid inequality

$$z_1 + x_{12} + x_{13} + \frac{1}{2}x_{22} + \frac{1}{2}x_{23} + z_4 \geq \frac{3}{2}.$$

Now, by rounding the coefficients in the left-hand side of this inequality to the nearest higher integer we get the valid inequality

$$z_1 + x_{12} + x_{13} + x_{22} + x_{23} + z_4 \geq \frac{3}{2},$$

and by observing that the left-hand side of this inequality is integer for all points in T we get that the inequality

$$z_1 + x_{12} + x_{13} + x_{22} + x_{23} + z_4 \geq 2$$

is valid for T. ■

We see that the inequality (6.28) is not satisfied by the optimal solution (6.26) to the continuous relaxation of the replacement problem in Example 6.24. In fact, if we add the inequality (6.28) to the continuous relaxation in Example 6.24 we get the optimal solution

$$\begin{aligned}(x_{11}, x_{12}, x_{13}, x_{14}) &= (0, 0, 1, 0),\\ (x_{21}, x_{22}, x_{23}, x_{24}) &= (1, 0, 0, 0),\\ (z_1, z_2, z_3, z_4) &= (1, 0, 1, 0).\end{aligned}$$

This is the solution in (6.25), that is, it is an optimal solution to the original problem! The valid inequality (6.28) in fact defines a facet of $\operatorname{conv} T$:

Proposition 6.26 *The valid inequality* (6.28) *defines a facet of* $\operatorname{conv} T$.

Proof. Since $T_i \geq 2$ for $i = 1, 2$, we have that $\operatorname{conv} T$ is full-dimensional (Proposition 6.17). Hence, as in the proof of Proposition 6.19 we can use the uniqueness characterization of the facet description (Proposition 6.15) to show the assertion.

Let

$$F = \{\, (x, z) \in \operatorname{conv} T \mid z_1 + x_{12} + x_{13} + x_{22} + x_{23} + z_4 = 2 \,\},$$

$x^i = (x_{i1}, \ldots, x_{i4})$ for $i = 1, 2$, and $z = (z_1, \ldots, z_4)$. With the same notation as in the proof of Proposition 6.19, assume that $\lambda x + \mu z = \rho$ for all $(x, z) \in F$. The points in F given by

$$\begin{aligned}(x^1; x^2; z) &= (1, 1, 0, 0;\ 1, 0, 0, 0;\ 1, 1, 0, 0),\\ (x^1; x^2; z) &= (0, 1, 0, 0;\ 1, 0, 0, 0;\ 1, 1, 0, 0),\end{aligned}$$

imply that $\lambda_{11} = 0$. Similarly, it follows that $\lambda_{14} = 0$. Further, the points in F given by

$$\begin{aligned}(x^1; x^2; z) &= (1, 0, 0, 1;\ 1, 0, 0, 1;\ 1, 0, 0, 1),\\ (x^1; x^2; z) &= (1, 0, 0, 1;\ 0, 0, 0, 1;\ 1, 0, 0, 1),\end{aligned}$$

imply that $\lambda_{21} = 0$, and similarly $\lambda_{24} = 0$. The points in F given by

$$\begin{aligned}(x^1; x^2; z) &= (1, 0, 0, 1;\ 1, 0, 0, 1;\ 1, 1, 0, 1),\\ (x^1; x^2; z) &= (1, 0, 0, 1;\ 1, 0, 0, 1;\ 1, 0, 0, 1),\end{aligned}$$

show that $\mu_2 = 0$, and in the same fashion $\mu_3 = 0$. We continue with the points in F given by

$$\begin{aligned}(x^1; x^2; z) &= (0, 1, 0, 0;\ 0, 1, 0, 0;\ 0, 1, 1, 0),\\ (x^1; x^2; z) &= (0, 0, 1, 0;\ 0, 1, 0, 0;\ 0, 1, 1, 0),\end{aligned}$$

which imply that $\lambda_{12} = \lambda_{13}$. Similarly, it can be shown that $\lambda_{22} = \lambda_{23}$. The points in F given by

$$(x^1; x^2; z) = (0,1,0,0;\ 1,0,0,0;\ 1,1,1,0),$$
$$(x^1; x^2; z) = (0,1,0,0;\ 0,0,0,1;\ 0,1,1,1),$$

together with $\lambda_{21} = \lambda_{24} = 0$ give that $\mu_1 = \mu_4$, and since

$$(x^1; x^2; z) = (1,0,0,1;\ 1,0,0,0;\ 1,0,0,1)$$

belongs to F and $\lambda_{11} = \lambda_{14} = \lambda_{21} = 0$ we have that $\mu_1 + \mu_4 = \rho$, so

$$\mu_1 = \mu_4 = \frac{\rho}{2}. \tag{6.30}$$

The point

$$(x^1; x^2; z) = (0,1,0,0;\ 1,0,0,0;\ 1,1,0,0)$$

belongs to F and since $\lambda_{21} = \mu_2 = 0$ we get that $\lambda_{12}+\mu_1 = \rho$. Hence, from (6.30) we have that $\lambda_{12} = \rho/2$, and above we showed that $\lambda_{12} = \lambda_{13}$, so $\lambda_{13} = \rho/2$ must hold. Finally, the point

$$(x^1; x^2; z) = (0,1,0,0;\ 0,1,0,0;\ 0,1,0,0)$$

belongs to F and since $\mu_2 = 0$ we have that $\lambda_{12} + \lambda_{22} = \rho$. But $\lambda_{12} = \rho/2$ so $\lambda_{22} = \rho/2$. Similarly, it follows that $\lambda_{23} = \rho/2$. Therefore we have shown that the equality $\lambda x + \mu z = \rho$ has the form

$$\rho(z_1 + x_{12} + x_{13} + x_{22} + x_{23} + z_4) = 2\rho.$$

Proposition 6.15 then gives that (6.28) is facet-defining. ∎

6.5 Conclusions

We have made an introductory study of the facial structure of the replacement polytope. We showed that the replacement polytope is full-dimensional (if the lifetimes of the parts are greater than or equal to two) and found that the necessary inequalities in the original formulation of the replacement problem are facet-defining. Unfortunately, these are not sufficient to represent the replacement polytope, as was shown by an example. By using Chvátal–Gomory rounding we showed how to find a new class of facets for the example problem. It is straightforward to generalize this procedure to an arbitrary replacement problem. However, it is still an open problem to investigate the strength of the continuous relaxation when the new class of facets is added to the replacement problem.

Chapter 7

Illustrative examples

7.1 Introduction

In this chapter we present some illustrative examples with the maintenance optimization models from the Chapters 4 and 5.

In Section 7.2 we illustrate how the fixed cost in Model I affects the resulting structure of the optimal solution. If the fixed cost is zero it is never optimal to replace any of the parts if they have not reached their respective lifetimes. The optimal solution is then to run the engine until a part reaches its lifetime, then replace just that part and put the engine into service again. In such a situation an optimization model is obviously not very useful. However, when the fixed cost is large it is important to take the opportunity to replace some of the parts that have not reached their respective lifetimes into account when the engine is taken to the workshop. It is not at all clear which parts to replace in order to minimize the total maintenance cost, and hence an optimization model is a vital tool in order to decide what to replace.

In Section 7.3 we compare the solutions from the two-stage stochastic model with the solutions from the stochastic dynamic programming model. The stochastic dynamic programming model always performs better than the two-stage stochastic model, but the two-stage model has the advantage that it can be used for systems consisting of a large number of parts. The stochastic dynamic programming model can only be used for systems where the number of parts is not greater than about 6. Hence, in our aircraft application we can not use the stochastic dynamic programming model, and this is the reason for developing the two-stage stochastic model.

The numerical tests are made with small instances (formed by systems having 4 parts) of the replacement problem, so we cannot draw any general conclusions from them. However, the main purpose is to illustrate some of the properties of the optimal solutions to the models. All of the numerical tests are made with the modelling language AMPL [56] together with the linear programming solver CPLEX [68].

7.2 Numerical tests with Model I

This section presents some numerical tests with Model I from Chapter 4. The purpose is to illustrate how the fixed cost in the model affects the structure of the optimal solution.

7.2.1 The test problem

We consider an instance of Model I with

$$\begin{array}{llll} T = 60, & N = 4, & & \\ T_1 = 13, & T_2 = 19, & T_3 = 34, & T_4 = 18, \\ c_1 = 80, & c_2 = 185, & c_3 = 160, & c_4 = 125. \end{array}$$

The data has been chosen so that the relations between the lifetimes and the costs are similar to those for the fan module of the RM12 engine. We will solve Model I for each of the values 0, 10, and 1000 of the fixed cost d. In the real maintenance situation $d = 10$ is the most reasonable value among the three.

7.2.2 Optimal solutions for different fixed costs

First we solve Model I for $d = 0$. An optimal solution is shown in Figure 7.1. The total number of replacement occasions is 11. Since the fixed cost is zero there are no advantages with replacing components before their respective lifetimes are reached.

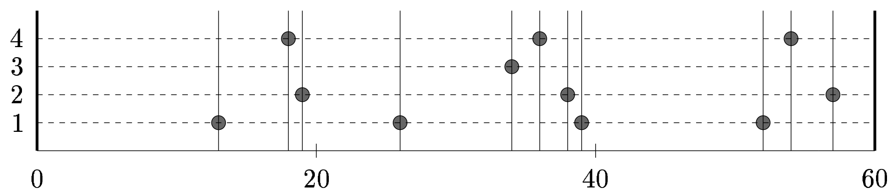

Figure 7.1: An optimal solution to the replacement problem with $d = 0$.

We next solve the problem for $d = 10$. An optimal solution is given in Figure 7.2. Compared to the case where $d = 0$ the total number of replacement occasions has decreased from 11 to 5. It is now beneficial to replace the components in larger groups. Also, the parts are replaced often even if their respective lifetimes are not reached.

Finally we solve the problem for $d = 1000$. An optimal solution is shown in Figure 7.3. Since the fixed cost is high compared to the costs of the components themselves it is very important to utilize the opportunity to replace several components at the same time. The total number of replacement occasions is 4. Actually, since $T_1 = 13$ and $T = 60$ there exists no feasible replacement scheme for which the total number of replacement occasions is less than 4!

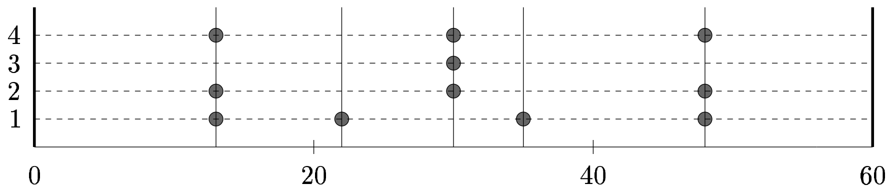

Figure 7.2: An optimal solution to the replacement problem with $d = 10$.

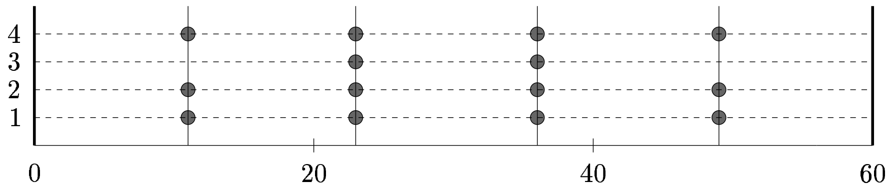

Figure 7.3: An optimal solution to the replacement problem with $d = 1000$.

7.2.3 Interpretation of the structure of the optimal solutions

Above we saw that the total number of replacement occasions decreases when the fixed cost increases; the benefits from replacing the components in groups are greater when the fixed cost is large. In Figure 7.6 the three optimal solutions from the previous subsection are illustrated. When the fixed cost increases from 0 to 10 we see that the first three replacement occasions for $d = 0$ are grouped into one for $d = 10$. The fourth replacement occasion for $d = 0$ is moved to an earlier point of time, but still it is just part 1 that is replaced. Further, the replacement occasions 5–8 for $d = 0$ are grouped into two replacement occasions for $d = 10$. The three last replacement occasions for $d = 0$ are grouped into one for $d = 10$. Similarly, when the fixed cost is increased from $d = 10$ to $d = 1000$, the replacement occasions 2–4 for $d = 10$ are grouped into two replacement occasions for $d = 1000$.

7.3 Numerical test with the stochastic two-stage model and the stochastic dynamic programming model

We numerically compare the use of the two-stage stochastic model (5.9) with the stochastic dynamic programming model presented in Section 5.4.3. In Section 7.3.1 the test data is given. Then in Section 7.3.2 the test procedure is described, and finally, the results are presented in Section 7.3.3.

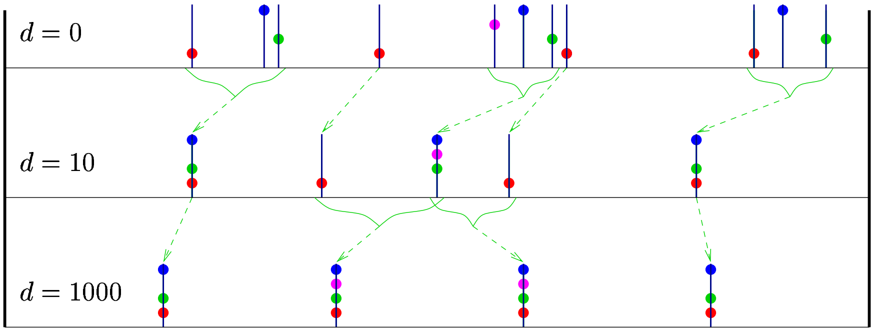

Figure 7.4: Interpretation of the differences between the optimal solutions.

7.3.1 Test data

We consider a system consisting of three deterministic parts and one stochastic part. The time horizon is $T = 60$ and the fixed cost is $d = 100$. The deterministic parts have the data

$$T_1 = 9, \quad T_2 = 13, \quad T_3 = 17,$$
$$c_1 = 30, \quad c_2 = 125, \quad c_3 = 119.$$

The lifetime of a new stochastic part is represented by the Weibull failure distribution function

$$F(t) = 1 - \exp\left[-\left(\frac{t}{\theta}\right)^{\alpha}\right], \qquad t \geq 0,$$

where $\alpha = 2$ and $\theta = 12.4$. (The Weibull distribution is very well adapted to the parts of an aircraft engine; see [124].) The expected lifetime of a new stochastic part then is 11. The cost of the stochastic part is 80. In the numerical test often the stochastic part is not new. If the age of the stochastic part is t_0, then the failure distribution function becomes

$$B(t) = \frac{F(t + t_0) - F(t_0)}{1 - F(t_0)}, \qquad t \geq 0.$$

In the stochastic two-stage model (5.9) we must create scenarios for the first lifetime, and fix the rest of the lifetimes for the stochastic part; for details see Section 5.7. A scenario tree with three possible lifetimes, T_s^1, T_s^2, and T_s^3, for the first specimen of the stochastic part and the lifetime T_s for the rest of the specimens of the stochastic part is shown in Figure 7.5.

The scenarios for the first lifetime will be computed as was described in Section 5.5.2. The rest of the lifetimes will be fixed to the expected lifetime of a new stochastic part, that is, to 11.

We represent the real world by 100 scenarios randomly picked from the failure distribution given by F, by using MATLAB [105]. The scenarios are

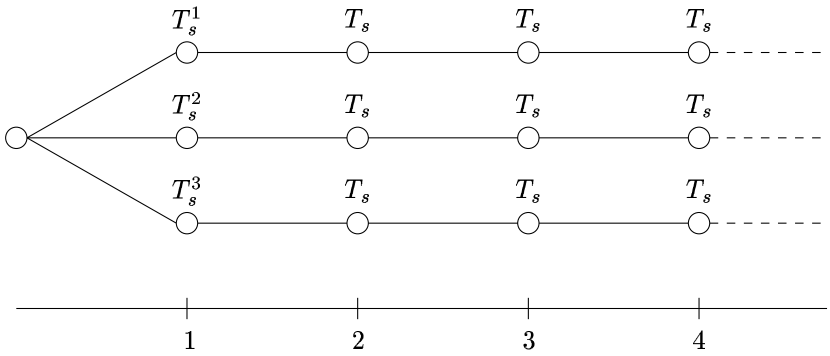

Figure 7.5: A scenario tree for the stochastic part.

of the type presented in Definition 5.2, that is, in order to create a scenario we first randomly pick a lifetime representing the first stochastic part, then we randomly pick a new lifetime representing the stochastic part that will replace the first part, and so on.

7.3.2 Description of the test

The test will be made for each of the following models:

- No scenarios: This is one of the models presented in Section 5.3. At each replacement occasion (which depends on the stochastic part as well as the deterministic parts) a new basic replacement scheme for the deterministic parts is computed. (Hence the stochastic part is only implicitly considered.)
- Two-stage model: This is the two-stage model given by (5.9), with 1, 2, 3, 4, 5, 6, 8, 12, 15, 18, and 20 scenarios, respectively. The scenarios computed are based on the real lifetime of the stochastic part at the given maintenance occasion, as was described above.
- Dynamic programming model: This is the stochastic dynamic programming model presented in Section 5.4.3.

We assume that each component of the system is new at time $t = 0$ and for each of the 100 real world scenarios we do the following (with each of the models given above):

(i) we compute the next replacement occasion by checking which of the deterministic parts that have the least life left, and compare the result with the life left for the stochastic part given by the real world scenario; if the next maintenance occasion appears after the time horizon, we stop; otherwise

(ii) we compute required input data for the given model at the given maintenance occasion;

(iii) we compute the replacement action with the given optimization model;

(iv) we go back to step (i).

With this procedure we get a cost for each of the real world scenarios for each of the models. The real world scenarios were picked at random from the failure distribution for the stochastic part; hence, we can estimate the expected cost for each of the models by summing up the cost for each of the 100 real world scenarios and then divide by 100.

7.3.3 Results

The expected costs for each of the given models are presented in Figure 7.6.

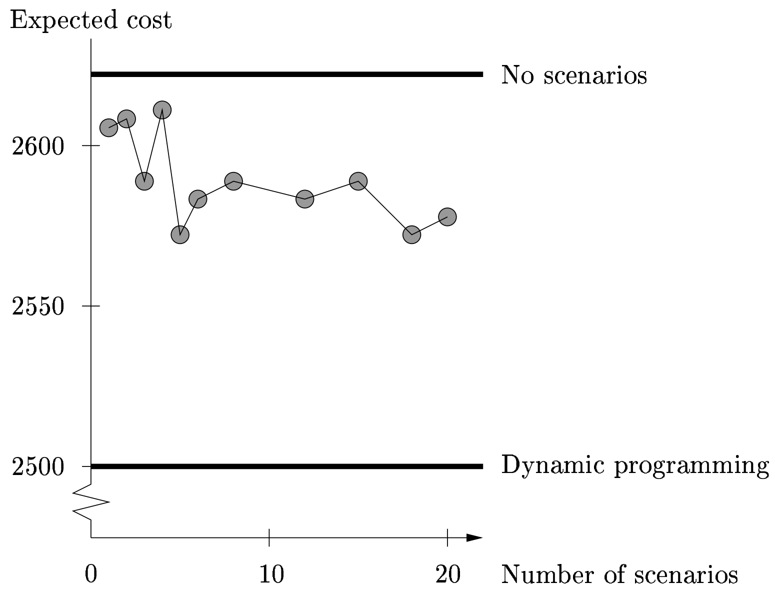

Figure 7.6: Test results with the stochastic optimization models.

We see that we always perform better with the two-stage stochastic model than with the model with no scenarios. However, the differences are not very large. This can be motivated by the fact that the system considered only consists of one stochastic part, which means that the total cost is dominated by the deterministic parts.

Further, we see that the relative benefits from using more than about 10 scenarios are small. In order to develop a computationally tractable optimization model it is important to keep the number of scenarios low, and hopefully future numerical tests can answer the question of how many scenarios are required in order to get good solutions in a realistic problem setting with many stochastic parts.

The figure also shows that the dynamic programming model always performs better than the stochastic two-stage model, which is natural according to the discussion in Section 5.8. In this test we only considered scenarios for the stochastic part with variations in the first lifetime (see Figure 7.5). We would probably get better solutions if deeper scenario trees were used. Such a scenario tree with variations in both the first and the second lifetimes is illustrated in Figure 7.7.

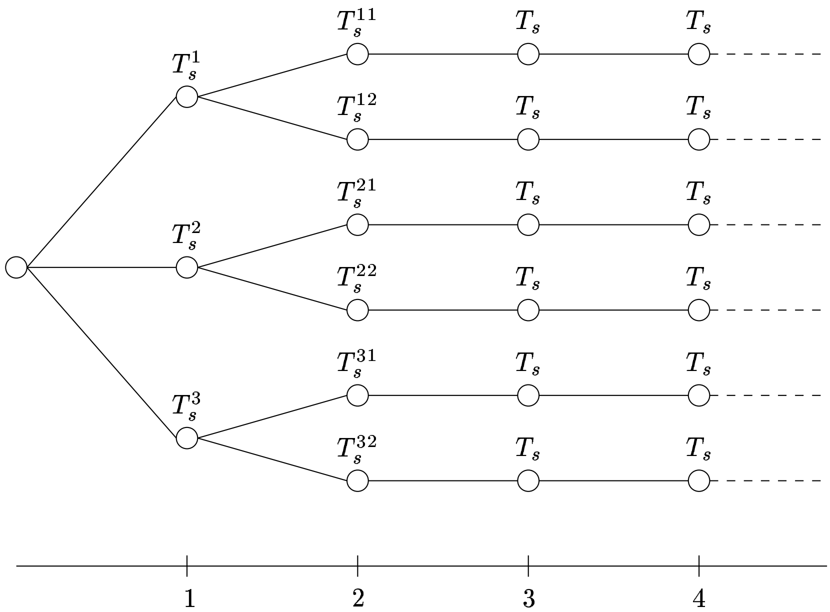

Figure 7.7: A "deep" scenario tree for the stochastic part.

Finally, we want to stress that the numerical tests presented in this section should be considered as illustrations only. In the future more numerical tests are necessary to be able to draw conclusions for a system consisting of more than just one stochastic part.

Chapter 8

Literature review

8.1 Introduction

Maintenance is a wide subject. We can speak of the maintenance of roads, software, the human body, the environment, engines, chemical processes, and more—and the literature is rich; for general reviews see [12], [103], [116], [32], [58], [8], [46], [90]. We focus here on maintenance policies for multi-component systems that might be useful when modelling the maintenance of an aircraft engine.

Many of the articles in the area of optimization of the maintenance of multi-component systems deal with systems consisting of identical stochastic components. Despite the fact that aircraft engines consist of nonidentical parts, and that some of the parts are deterministic, these models are of interest since most models for systems comprising nonidentical parts originate from them. Hence, we present models both for systems with identical and nonidentical parts.

The literature surveyed has been divided into several classes. Age and block replacement are two of the most basic maintenance policies, and much of the maintenance theory originates from them. The pure age and block replacement models and policies, as well as modifications of these, are presented in the Sections 8.2 and 8.3.

In some situations it is not possible to detect a failure without inspection. Policies incorporating inspection are investigated in Section 8.4.

In Section 8.5 condition based maintenance is discussed; condition based maintenance means that the state of the system is monitored or inspected, and when a certain threshold value is attained maintenance is performed. Note that in condition based maintenance the main issue is to compute this threshold value, while in inspection policies the focus is on deciding when to perform inspections.

In Section 8.6 we present models for "opportunistic maintenance", which we define as maintenance that can be performed at opportunities that arise randomly independent or dependent of the components of the system.

A recent theory in the maintenance literature is marginal cost analysis. This means that the difference in cost between performing maintenance directly and waiting an additional amount of time is used to make maintenance decisions. Marginal cost analysis is considered in Section 8.7.

Most of the maintenance models in the literature deal with stochastic times to failure. Nevertheless, there are situations in which data is deterministic, and this is the subject of Section 8.8.

In many maintenance models it is assumed that a spare part is always available when needed, that is, the supply of spares is infinite, and the cost of the inventory is not taken into account. In reality this is seldom the case, and in Section 8.9 models with a finite supply of spares are considered.

Sometimes there are several different components that can replace a failed component; models with choices of spares are described in Section 8.10.

When having several failed systems there is a possibility to replace failed parts from one system with non-failed parts from other systems. This kind of maintenance is usually referred to as "cannibalization policies" and is considered in Section 8.11.

In Section 8.12 we briefly mention some models that we currently think are not as important as the other models presented, but that can be potentially useful in future research.

Obviously there exist several different maintenance policies, and when choosing a maintenance policy it is interesting to see which fits best to a certain situation. This problem is considered in Section 8.13.

The main purpose of maintenance models is of course to use them in real world problems, and in Section 8.14 different applications are presented.

The chapter concludes in Section 8.15 with a discussion on how the literature surveyed is related to the maintenance optimization models developed in this thesis.

8.2 Age replacement

Under an age replacement policy a component is replaced at failure or at a specified age, whichever occurs first. The basic age replacement policy is described in [12].

Fox [57] generalizes the age replacement policy by incorporating discounting, that is, the loss incurred at a replacement decreases with time.

The discounting model presented by Fox [57] is further investigated by Ran and Rosenlund [106]. They perform a sensitivity analysis, and give some numerical examples.

Age replacement with minimal repair is discussed in [9]. Minimal repair means that the system is repaired to the condition it had just before it failed (also called "as bad as old"). Apart from optimal replacement, which minimizes the expected cost rate, the authors also investigate the improvement and deterioration of the system over time. A repairable system is said to improve (deteriorate) with time if the time between successive repairs increases (decreases).

Zheng [144] introduces an age tolerance w and an age limit T to form the interval $(T-w, T)$. He considers a system which consists of n identical units. A unit is replaced at failure or when its age exceeds T, whichever occurs first. When a unit is replaced, all the operating units with their age in the interval $(T-w, T)$ are also replaced. Optimal values of T and w are obtained so as to minimize the mean total replacement cost rate. The method is illustrated by a numerical example.

It is reasonable to assume that it will be more expensive to operate a system the older it becomes. Scheaffer [113] extends the age replacement policy with this assumption, by including in the cost function a term which increases with the time a unit is in use. A system comprising only one component is considered and for the case of exponential life distribution a detailed investigation is made.

Sheu [118] considers a system subject to shocks that arrive according to a non-homogeneous Poisson process. As shocks occur the system has two types of failures. Type I failures (minor failures) are removed by minimal repair, whereas type II failures (catastrophic failures) are removed by an unplanned replacement. The probability of a type II failure depends on the number of shocks suffered since the last replacement. The author investigates both age replacement and block replacement (see Section 8.3). A numerical example that treats a Weibull distributed system is given.

The models investigated so far consider failure times as the only failing mechanism. However, information obtained using condition monitoring devices is being used more and more in industries for maintenance scheduling. Kumar and Westberg [77] consider the problem of incorporating monitored variables when optimizing an age replacement policy. They use a proportional hazard model in which it is assumed that the hazard rate of a system consists of two multiplicative factors: the baseline hazard rate, $h_0(t)$, and generally an exponential function containing the effects of the monitored variables. Hence, the hazard rate of a system can be written as

$$h(t; z) = h_0(t)e^{z\beta},$$

where z is a row vector consisting of monitored variables, and β a column vector consisting of the corresponding regression parameters. The authors suggest a graphical method called "total time in test plotting" to find an optimal age replacement policy. A numerical example, where one wants to decide the level of pressure (z) at which a machine should be operated, is given.

Cassady et al. [29] consider an age replacement policy which maximizes the availability of an equipment. They consider a simple piece of repairable equipment that has a predetermined useful life. The successive lengths of operating periods as well as repair times are assumed to be Weibull random variables. The problem of interest is to determine an age replacement policy (T^*) that maximizes the expected proportion of time that the equipment is operating during its useful life. A method that finds an approximate value of T^* is developed, and the validity of this approximation is tested through numerical experiments.

Yang and Nachlas [141] look at the construction and usage of bivariate models. This kind of models have lifetimes which depend on two quantities, namely age and usage. Common examples are automobiles and automobile-tires for which both model year and accumulated mileage are usually included in discussions of lifetime. The authors consider cases of both stochastic functional relationships, and simple correlations between the age and usage variables. A 2-dimensional age replacement policy is presented, where the device is replaced when it fails or when it attains either age T or usage w. The objective is to maximize the availability.

Sim and Endrenyi [121] consider a device which deteriorates with the time in service, and is also exposed to Poisson distributed failures independent of the deterioration. The times to preventive maintenance have an Erlang distribution and may be, in a limiting case, completely deterministic (then this policy coincides with age replacement). The objective is to find the optimal value of the mean time (that is, the mean of the Erlang distribution) to preventive maintenance such that the availability of the device is maximized.

Reineke et al. [107] consider the problem of determining the appropriate age replacement policy for a complex system that has high availability requirements. The system under study consists of five unique and independent functional subsystems connected in the bridge structure shown in Figure 8.1.

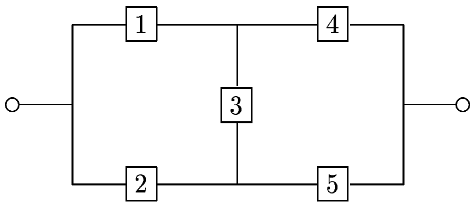

Figure 8.1: Bridge structure from [107].

Each subsystem has two failure modes which are modelled as independent competing risks. A wear-out mode and a chance failure mode are modelled for each subsystem. The repair time and cost of a subsystem is dependent on the failure mode. If the system is still operating at the age replacement time T, then the working subsystems are preventively maintained and the failed subsystems are repaired. The authors derive expressions for finding the optimal T such that the average cost over an infinite time is minimized or the availability maximized. An analytic methodology is presented for performing a tradeoff analysis between the expected system cost accumulation rate and the limiting system availability. The method is illustrated by an example where the wear-out failure modes are modelled by Weibull distributions, and the chance to failure modes characterized by exponential distributions.

Usually the optimal replacement rule in an age replacement policy is determined by minimizing the expected long-run cost per unit time. In this case the function to be minimized has the form

$$C(t) = \frac{A(t)}{B(t)}, \tag{8.1}$$

where $A, B : \mathbb{R}_+ \to \mathbb{R}$. Generally it is hard to find the optimal replacement age, so often special cases are considered (e.g., exponentially distributed life times). However, Aven and Bergman [7] present a general solution scheme for the general case. They minimize (8.1) by minimizing the λ-function

$$C_\lambda(t) = A(t) - \lambda B(t).$$

They show that there always exists a $\lambda = \lambda^*$ such that $t = T_{\lambda^*}$ minimizes (8.1), and give an iterative method to find it.

8.3 Block replacement

Under a block replacement policy the components of a system are replaced at failure or at fixed times kT ($k = 1, 2, \ldots$), whichever occurs first. The basic block replacement policy is described in [12].

The main drawback of the block replacement policy is that at planned replacement times practically new items might be replaced. Berg [23] presents a modified block replacement method where failed items are still replaced after failure, but items possessing age $b \in [0, T)$ or less at scheduled block replacement points $T, 2T, 3T, \ldots$ are not replaced by new items but are instead permitted to remain in service. The objective is to find b and T such that the expected cost per unit time per item taken over an infinite time horizon is minimized.

A reasonable assumption is that it costs more to run a unit the older it becomes, and in [25] the modified block replacement policy from [23] is extended to cover this assumption.

Archibald and Dekker [5] extend the modified block replacement policy presented in [23] in two ways. They consider (i) a discrete time framework which allows the use of any discrete lifetime distribution, and (ii) multi-component systems. The case when the system consists of identical components is treated in detail, and an example with a Weibull lifetime distribution is presented. The authors outline how to extend the model to multi-component systems with non-identical components.

Another modification of the block replacement policy is presented by Tango [126]. He suggests the utilization of used items in the following policy:

1. Exchange operating items for new ones at times kT ($k = 1, 2, \ldots$).
2. If items fail in $[(k-1)T, kT - v)$, they are replaced by new items, but if they fail in $[kT - v, kT)$, they are replaced by used items, where $0 \leq v \leq T$.

If $v = 0$ the above policy coincides with the pure block replacement policy. A comparison with the pure block replacement policy is made for the Erlang distribution with two stages.

Abdel-Hameed [1] considers a block replacement model with minimal repair. This means that, at failure, the device is either restored to its condition prior to failure (minimal repair) or replaced (unplanned replacement).

Under a periodic maintenance policy a system undergoes repair and preventive maintenance at fixed times kT $(k = 1, 2, \ldots)$. This policy can be seen as a variation of block replacement. Chan and Shaw [30] consider a periodic maintenance policy for a system whose failure rate depends on the age and on the number of preventive maintenance occasions until that age. It is assumed that the failure times are Weibull distributed. The availability is modelled as a random variable, and the objective is to find T such that the probability of a specified availability is maximized. A numerical example is given to illustrate the design criteria.

In many of the existing maintenance models it is assumed that the time to failure distributions are exponential. This means that the renewal process becomes stationary and analytical results can be derived. However, often the times to failure are not stationary and in this case special methods can be useful. In [95] a block replacement policy is studied for a multi-unit system with its components' failure pattern modelled as non-stationary stochastic point processes under economic and availability dependency criteria. A numerical example is given where the times to failure follow Gamma distributions. By simulation the optimal time to replacement of the system is derived as a compromise of availability and cost.

Bahrami-G. et al. [10] consider the problem of finding the optimal block replacement policy where the objective is to minimize the total down-time per unit time. The function to be minimized becomes hard to evaluate for the general case. A common approach is to assume a "nice" failure time distribution and then solve the problem. Here the authors derive a simplified equation and then use the Newton-Rhapson algorithm to solve it. Numerical examples with normal distributed times to failure are presented.

Ait-Kadi and Cléroux [2] introduce a block replacement policy where at failure the item is either replaced by a new or a used item or remains inactive until the next planned replacement. The policy is defined in the following way:

1. Preventive replacements by new items are made at times kT, $k = 1, 2, \ldots$, independently of the item's failure history.
2. If a failure occurs in the time interval $[(k-1)T, kT - \delta_1)$, $k = 1, 2, \ldots$, $0 \leq \delta_1 \leq T < \infty$, then the item is replaced by a new one.
3. If a failure occurs in one time interval $[kT - \delta_1, kT - \delta_2)$, $k = 1, 2, \ldots$, $0 \leq \delta_2 \leq \delta_1 \leq T < \infty$, then the item is replaced by a used one.
4. If a failure occurs in the time interval $[kT - \delta_2, kT)$, the item remains inactive or works less efficiently until the next planned replacement at time kT.

Each item which has been removed in a planned replacement after attaining age T and which is still working will be considered as a used item. It is also assumed that a used item costs less than a new one, and an item cannot be installed more than twice. The optimal policy $(T^*, \delta_1^*, \delta_2^*)$ is the one which minimizes the average cost per time unit over an infinite time span. The problem is a nonlinear optimization problem with linear constraints and is solved by using a generalized reduced gradient algorithm. Numerical results for Gamma distributed and Weibull distributed times to failure are given.

8.4 Inspection policies

Sometimes failed components can be detected and replaced only by inspection. There is a cost related to the time a component is not operative. Under an inspection policy the objective is to find the inspection schedule that minimizes the expected average cost.

An inspection policy for a system of n units working in parallel and with identical failure distributions is considered in [4]. It is assumed that the lifetime distributions are known to be exponential, but with unknown expected lifetimes. The suggested inspection policy is sequentially adaptive in the sense that the first inspection epoch is determined arbitrarily. At the time of the first inspection, the parameter of the lifetime distribution is estimated based on the occurred number of failures. On the basis of that estimate, the next inspection epoch is determined, and so on. It is assumed that failures can be discovered only through inspection, and when inspection takes place all units are inspected and failing units are replaced by new ones. The optimal inspection policy is defined to be the policy which is determined by the time interval that minimizes the average cost per unit of time between inspections.

The inspection problem also arises in [6]. Assaf and Shanthikumar consider N machines which are subject to random failures. The times to failure are independent and have the same exponential distribution. The repair cost is assumed to have both a constant term reflecting the overhead cost of repair and a cost of repair per machine. Another cost is the cost incurred due to failing machines. This cost is the same for all machines and proportional to the elapsed time between the failure of a machine and its time of repair. The main results are (1) a necessary and sufficient condition for it to be optimal to never inspect or repair, and (2) a characterization of the optimal control threshold.

Vaurio [131] considers a preventive maintenance policy with periodic inspections. This kind of policy is of interest, for example, in non-monitored production lines. Inspections are performed at intervals T, and failures are only detected by inspection. It is assumed that the unit is repaired immediately when a failure is detected, and that a repair makes the unit as good as new. The unit is preventively renewed after M inspection intervals, or at failure, whichever occurs first. The objective is to find T and M such that the total expected cost rate is minimized. A numerical example is given to illustrate the method.

Sometimes when a system fails it may not be obvious which components are at fault. Butler and Lieberman [28] investigate different inspection policies for fault location. It turns out that determining the optimal inspection sequence is impractical, so heuristic procedures must be used.

8.5 Condition based maintenance

Maintenance policies such as age and block replacement are examples of scheduled maintenance policies. These policies are easy to implement since they have a clear structure. Nevertheless, often condition based maintenance can be a better and more cost effective type of maintenance. Under a condition based maintenance policy a technical state of the system is monitored or inspected, and when a specific threshold value is reached the system is replaced or preventive maintenance is performed.

Park [99] derives the optimal wear-limit for preventive replacement for an item with wear-dependent failure rate, by minimizing the long-run total mean cost rate. The optimal strategy has the same form as the age replacement policy.

In [100] the optimal wear-limit of [99] is extended to regard several risk factors that are monitored continuously or periodically.

Legát et al. [83] consider condition based maintenance with both infinite and finite-time horizons. For the case of a finite-time horizon a simple correction to the infinite-time solution is given that provides a good approximation to the exact solution.

Abdel-Hameed [1] looks at a device that is subject to deterioration, and the deterioration level is only monitored periodically at fixed times. Further he assumes that a failure is detected only by inspection and that the device can be replaced before or at failure. The decision to replace the device before failure is detected depends on the deterioration level at the inspection time. The objective is to find the optimal replacement time that yields the smallest possible long run average cost per a unit of time.

Özekici [150] considers systems that consist of stochastically and economically dependent components. The states of the components are inspected periodically, and the inspector decides which components must be replaced based on observation. The effect of the dependencies on periodic replacement policies is discussed, and theorems that characterize optimal replacement policies are given.

Berenguer et al. [19] consider a system which deteriorates stochastically. The state of the system is a continuous random variable and can be observed only by inspection. The problem is modelled by a semi-Markov decision process, and the objective is to determine at each inspection epoch whether a preventive maintenance is necessary and when the next inspection should be done. The continuous feature of the model allows a more theoretical analysis than in the case of discrete models. A numerical method that arises from analytical properties is proposed and numerical experiments are carried out.

To successfully implement condition based maintenance one must be capable of detecting incipient failures prior to their occurrence. This problem is considered by Yang and Lin [142]. They employ the Petri net modelling method coupled with parameter trend and fault tree analysis to perform early failure detection and isolation for preventive maintenance. An example is given that illustrates the approach.

The principles of condition-based preventive maintenance and when it can successfully be implemented are discussed by Mann et al. [71]. A comparison between the age replacement policy and the condition-based policy is made.

A general discussion about how to implement and use condition-based maintenance is also made in [130].

8.6 Opportunistic maintenance

Opportunistic maintenance refers to the situation in which preventive maintenance is carried out at opportunities. In the literature sometimes it is assumed that these opportunities arise independently of the failure process, and sometimes the opportunities are by definition equal to failure epochs of individual components. In the last case, due to economies of scale (for example, fixed costs at each maintenance occasion independent of what is replaced), the unpleasant event of a failing component is at the same time considered as an opportunity for the preventive maintenance of other components. This situation is typical for the maintenance of aircraft engines.

Berg [20] considers a machine that consists of two nonidentical stochastically failing units. He investigates first a policy under which at failure of either of the two units, the unfailed unit is also replaced if its age exceeds a predetermined control limit, and then he generalizes this procedure in the way that every unit is also replaced when it reaches a predetermined critical age. The special case of Erlang distributed times to failure is investigated in detail.

Ouali et al. [97] consider two opportunistic maintenance strategies for a system comprising n nonidentical components. It is assumed that every component i has its own optimal and fixed age policy T_i^*, that is, component i is replaced upon failure or when its age reaches T_i^*, whichever occurs first. The two strategies are defined as follows:

1. Opportunistic maintenance activities are allowed on a nonfailed component j $(j \neq i)$ if the difference between the expected preventive time T_j^* of component j and the failure instant of component i is less than the threshold δ.

2. Same as strategy 1 with the extension that opportunistic activities are also allowed on a component j when component i is preventively replaced at T_i^* and $T_j^* - T_i^* \leq \delta$.

The objective is to find the threshold δ such that the total average cost per unit time over an infinite operating time is minimized. The authors outline how to use simulation to find optimal strategies.

In [61] a group preventive replacement policy problem is formulated. The system considered consists of m identical elements working independently under the same conditions. At time $t_0 = 0$ every element is new; at time T the whole system is replaced by a new one. During the time interval $[0, T]$, if one element fails it has to be replaced immediately by a new one. If this happens the repairman can replace any number of working elements that he wants. The only economic dependence considered is that a fixed cost is incurred whatever the number of elements replaced. The objective is to find which components to replace at each time-step to minimize the cost for running the system during the given time-period. The authors give both a continuous and a discrete time formulation, and use the theory of optimal control of jump processes to obtain a dynamic programming equation. For the discrete time version of the model, numerical computations of optimal and suboptimal strategies of group preventive policies are performed.

In [82] the dynamic programming model in [61] is extended to systems consisting of m components with nonidentical independent lifetime distributions characterized by discrete nondecreasing failure rates. A numerical illustration with a 4-component system similar to a modular jet engine is given, and a class of suboptimal strategies is discussed.

Vergin and Scriabin [132] present dynamic programming models for determining optimal policies for equipment comprising two or three identical components with economic dependence. The failure rates of the components increase with age. Both opportunistic replacement (i.e., it is cheaper to exchange several parts jointly than separately) and preventive replacement (i.e., there is a cost of a breakdown) are considered. The authors show that the optimal policy is close to an (n, N) policy in which a component undergoes preventive replacement if it has operated for N periods and undergoes opportunistic joint replacement if it has operated for n periods if either another component fails or another component reaches it preventive replacement age. This (n, N) policy is suggested to be used for systems consisting of more than three components since the dynamical programming model then would become very complex.

Many maintenance models consider the grouping of maintenance activities on a long-term basis with an infinite horizon. This makes it very difficult to incorporate short-term circumstances such as opportunities or a varying use of components. Wildeman, Dekker, and Smit [137] propose a rolling-horizon approach that takes a long-term tentative plan as a basis for a subsequent adaption according to information that becomes available in the short term. Their approach consists of five phases and is made generally such that it can be applied to many different preventive maintenance optimization models. In fact, it is possible to combine, for example, activities modelled according to a minimal-repair model with activities modelled according to a block replacement model. The authors develop a dynamic programming algorithm and illustrate their method with an example.

Savic et al. [110] look at a system that consists of several nonidentical parts, which is operated until failure of one of its components. When the failed component is replaced there is an opportunity to replace other components as well. Group replacement assumes that when any component is replaced, other components belonging to the same group will also be replaced. The problem of the selection of the optimal opportunity-based maintenance policy can be formulated as a partition problem. Namely, given a finite set O of components in a system, is it possible to divide them into n exclusive groups $O_i \in O$ such that the total maintenance cost is minimized? The authors develop two genetic algorithms for finding an optimal partition of O into such groups. An example of a 20-part system illustrates the algorithm. In [111] a further analysis of the genetic algorithm is considered, and a real-size system comprising 250 components is regarded.

When considering the age replacement policy often unrealistic assumptions about the distributions of unit lifetimes are made (e.g., exponential lifetimes) to be able to find an optimal policy. To solve the problem with general lifetime distributions Zheng and Fard [145] develop a hazard rate tolerance policy, which means that a unit is replaced at failure or when its hazard (failure) rate exceeds a limit L, whichever occurs first. The policy is also opportunistic in the sense that when a unit is replaced because its hazard rate reaches L, all the operating units with their hazard rate falling in the interval $(L-u, L)$ are also replaced. Optimal L and u are obtained to minimize the average total replacement cost rate. A numerical example is given that considers a system composed of nonidentical components.

An opportunity-based age replacement policy is presented by Dekker and Dijkstra [42]. A component can only be replaced preventively at an opportunity, contrary to a failure, at which the component is directly replaced. The opportunities arise according to a Poisson process, independently of failures of the component. A component is preventively replaced at an opportunity if its age has passed a control limit, and the objective is to find the control limit that minimizes the average cost.

An approximate method for the opportunity-based age replacement policy in [42] is proposed by Sherwin [117].

Jhang and Sheu [70] propose an opportunity-based age replacement policy with minimal repair. They assume that a system has two types of failures. Type I failures are removed by minimal repairs, whereas type II failures are removed by replacements. Type I and type II failures are age-dependent. A system is replaced at type II failure or at the first opportunity after age T, whichever occurs first. The cost of the minimal repair of the system at a specific age depends on a random part and a deterministic part. The opportunities arise according to a Poisson process, independent of failures of the component. The optimal T^* which would minimize the cost rate is discussed, and an algorithm for solving a simple special case is given.

8.7 Marginal cost analysis

In marginal cost analysis the maintenance action depends on the cost to perform the maintenance action immediately compared to waiting an additional time; "additional time" can be defined in several ways.

Berg [21] computes the marginal costs for age and block replacement, respectively. By marginal cost he means the difference, per unit time, between the cost of an age (or block) replacement now and the expected costs associated with waiting an additional short time. The marginal cost is then used to draw conclusions about the existence and uniqueness of optimal solutions to the age and block replacement policies, respectively. This procedure can also be adopted to other preventive replacement policies than age and block replacement.

The marginal cost approach from [21] is generalized by Dekker and Smeitink [44]. In their model a component is replaced upon failure and can only be replaced preventively at maintenance opportunities. In a production system a maintenance opportunity can occur for a variety of reasons, for example, breakdowns of essential units. The occurrence of opportunities is described by a renewal process. The authors first derive a model for a system comprising only one unit, and then an extension to the multi-component case is made. An exact optimization algorithm for the case of K_2-distributed times between opportunities is presented. This algorithm can also be used as an approximative method in the case of other times between opportunities distributions, and its performance is checked by simulation.

The standard approach in age and block replacement policies is to derive an expression for the expected average cost rate and then minimize this expression. However, with more complex models, for example, replacement models with repair, the use of this approach results in quite laborious mathematics. Berg [22] uses marginal cost analysis to overcome this problem for a replacement model with minimal repair.

In many cases preventive maintenance is more economic when applied to an entire group of components than to individual ones, and the problem then is to decide when to maintain the entire group (this is often called a group replacement policy). Dekker and Roelvink [43] present a group age replacement criterion based on marginal cost considerations. The marginal costs are interpreted as the extra costs caused by deferring preventive replacement for an additional time unit. The criteria force a replacement if the marginal costs exceed the minimum average costs. For the marginal costs, formulas in the component ages are derived, whereas the minimum average costs are approximated by the block replacement model. The authors investigate the performance of the criteria by means of discrete time Markov decision chains for two components, and by simulation for multiple identical components.

8.8 Deterministic replacement models

In the above maintenance models the times to failure are stochastic and have known distributions. In a deterministic replacement model, however, all data is deterministic. For an aircraft engine all of the safety-critical parts can be considered as deterministic since they have estimated fixed lifetimes (see Section 2.2).

In [15] the replacement of one item which deteriorates and becomes obsolete with time is considered. The time horizon is finite and all data is assumed to be deterministic. This problem is normally solved through the use of dynamic programming, but the authors formulate the problem as a $0-1$ linear programming problem. A numerical example is used to illustrate the model.

Dickman et al. [50] present a deterministic mixed integer linear programming model for a multi-unit system with one economic dependence; a fixed cost is incurred when the system is taken to the workshop independent of the number of parts that are replaced. (Actually, this mixed integer linear programming model is crucial for the development of the deterministic optimization models in Chapter 4.)

8.9 Models with a finite supply of spares

In the maintenance models in the literature it is often assumed that the supply of spare parts is infinite, and that there is no inventory cost. In reality this is not the case. In this section we present some of the articles that take the inventory into account. The research in this area is lively and more articles can be found in, for example, [32].

Nakagawa and Osaki [91] extend the age replacement model by assuming that the system (one-unit system) is supported by a buffer which stores only one spare unit for replacement, and that whenever each replacement starts, one new unit is ordered and then arrives at the buffer in a random delivery time. The unit is replaced upon failure or at the age-limit T, whichever occurs first, whenever the spare is available. Otherwise one has to wait until the new unit arrives. The objective is to minimize the long-run total average cost per unit time, and numerical experiments are presented to illustrate how to determine the optimal age-limits.

Liao and Yuan [84] extend the policy of [91] by considering the possibility that the new unit after arrival may not be acceptable in quality, that is, it should be rejected. They assume that first, whenever each replacement takes place, a new unit is ordered, then delivered, and finally tested after its arrival in a totally deterministic or random delivery time, and second, that the unit, when acceptable, will enter the buffer as a spare.

Derman, Lieberman, and Ross [48] investigate a special version of the replacement problem. They consider a system with one vital component for which there are n spares. Whenever the vital component fails, the system fails. The idea is to judiciously replace the component in use with an available spare in order to extend the life of the system. Once a component has been removed it cannot be used again. The objective is to determine the schedule that maximizes the expected life of the system. This problem can be generalized into a situation in which there are several components in the system, and the authors discuss a special case when the system consists of two exponentially distributed components.

Simpson [122] considers an inventory system with two inventories: the serviceable inventory and the repairable inventory. In a spare-part repairable inventory system there is complete dependence between the demand for serviceables and the return of repairables as all failing units are immediately replaced by a serviceable unit. The serviceable inventory is depleted by the demand from the maintenance model and is replenished by repairing units available in the repairable inventory and/or by purchasing new items. The repairable inventory is depleted by repairing units for serviceable and/or by junking and is replenished by failed replaced units. The author investigates the structure of the optimal solution to this inventory problem.

Chelbi and Aït-Kadi [31] discuss a periodic block replacement strategy taking into account the state of a spare parts inventory. In order to precise the modelling assumptions and establish basic conditions, a system made up of only one component is considered. The objective is to find the replacement period, T, the replenishment cycle, $R = kT$ for some $k \in \{1, 2, \ldots\}$, and the ordering point such that the total expected cost per time unit over an infinite span is minimized. An algorithm for solving the problem is proposed and numerical experiments are performed.

Zohrul-Kabir and Al-Olayan present in [147] and [148] a policy for the joint optimization of age replacement and spare provisioning. It combines the age replacement policy with a continuous review (s, S) type inventory policy, where s is the stock reorder level and S is the maximum stock level. The operating principle for this policy can be described as follows: "An order for $(S - s)$ spare units is placed when the inventory level falls to s. The operating unit is replaced preventively at T provided a spare is available. Otherwise the unit is replaced as soon as the stock is replenished. If the operating unit fails before T, the failed unit will be replaced as soon as a spare is available." The order lead time is considered to be randomly distributed, and the objective is to find T, s, and S such that the expected total cost per unit time is minimized. Because of the complexity in formulating a mathematical model for multi-unit situations, the authors develop a simulation model to determine a near-optimal policy. Results from different case problems are presented and a comparison is made to the case when age replacement and inventory policies are optimized separately.

Often it is preferable to use a general purpose simulation language when performing simulations. In [149] the construction of such a simulation language for the policy for joint optimization of age replacement and spare provisioning in [148] is described.

Sarker and Haque [109] investigate the same policy as in [147] and [148], but they consider block replacement instead of age replacement. A simulation model is developed and numerical experiments are conducted.

8.10 Models with multiple choice of spares

Often failed components are replaced by new ones, but in some cases there are multiple choices of spares. Then the question arises which of the spares should be chosen to replace a certain component. This is the subject of the present section.

Some systems consist of a framework which is virtually independent of its components, for example, a car where the components are the tires, battery, and ignition system, and the framework is the bodywork and chassis of the car. Such systems are investigated by Thomas [129]. For simplicity systems with only one component are considered, though the results apply in more general situations. The replacement policy for such a system depends on two interrelated decisions:

1. If the component fails and there is a choice of replacement with different costs and lifetime distributions which should be chosen?

2. Since the system can be replaced with a new one, including a new component, is it worth replacing the whole system, when just the component has failed?

Thomas proves the existence of optimal answers to these questions, though for the first question exponential lifetimes are assumed. However, how to compute the optimal answers is not investigated.

8.11 Cannibalization policies

When two or more systems are inoperative because of a different failed part in each, a common maintenance practice is to take the required part from one machine to restore the other. This policy is often called cannibalization.

In [55] and [74] different cannibalization policies are compared by simulation, but no analytic method is developed for finding the optimal cannibalization policy.

8.12 Other models

In this section we briefly mention some models other than those presented above. We have collected them here since we think that they are not as important as the models presented above, but believe that they may be potentially useful in future research.

In most of the maintenance models it is assumed that the components of the system are statistically *independent*. Models considering statistically *dependent* parts are found in [3], [127], [96].

In a k-out-of-n system failures of components are allowed. However, in order to have a working system at least k components (of the total number, n) have to be operative. Models for k-out-of-n systems are discussed in [102], [123], [119].

The above maintenance models always assume that the costs of, for example, new items or repairs in the future are known, as well as the distributions of times to failures, time horizons, and other data. In reality we only know this data with some uncertainty, and models that take this into account are treated in [88], [120], [101], [104], [151], [94], [143], [49], [135], [75].

Traditional maintenance models assume that the system after maintenance is either "as good as new" (replacement) or "as bad as old" (minimal repair). A more realistic assumption is that the system after maintenance lies somewhere in between “as good as new” and “as bad as old”. This kind of maintenance is usually called imperfect repair, and is discussed in [85], [89], [18], [86], [108].

8.13 Finding the best maintenance policy

As we have seen above there are several maintenance policies that can be used, and a policy that is the best under certain circumstances might not be the best in another situation. In this section we consider the problem of finding the best policy.

In [24] a comparison of age, block, and failure replacement policies is made. Let a and b be the planned replacement costs for age and block replacement, respectively. It can be shown that if $a = b$, then age replacement is preferable to block replacement, and if a and b are sufficiently large then failure replacement is the best policy. However, if $b < a$ then the question arises which policy should be chosen. The authors suggest a method for making this choice and show how it varies with a and b.

Dekker [39] presents a framework which covers several optimization models, and develops a uniform analysis for these models. From this analysis penalty functions are derived which can act as priority criterion functions, and also serve as basic elements in a method for determining optimal combinations of activities, and in maintenance planning.

When choosing a maintenance strategy of a complex system one is often interested in finding an optimal combination of several maintenance alternatives (i.e., use different maintenance policies on different parts). Bevilacqua and Braglia [26] investigate this problem and use an application of the analytic hierarchy process (AHP) for selecting the best maintenance strategy for an Italian oil refinery. Five possible alternatives are defined: preventive, predictive, condition-based, corrective, and opportunistic maintenance. With the AHP technique several aspects which characterize each of the mentioned strategies are arranged in a hierarchic structure and evaluated using a series of pairwise judgments. It turns out that the maintenance management of the oil refinery is satisfied with the AHP methodology, and the authors draw the conclusion that AHP can enhance and improve the understanding of the dynamics of a similar complex problem and represents an effective approach to arrive at decisions.

Lam and Yeh [80] consider continuous time Markov deteriorating systems, where the degree of deterioration (except failure) of the system is known only through inspection. Iterative algorithms are developed to derive the optimal maintenance policy and the corresponding cost value for each of the following five strategies:

1. *Failure replacement.* No inspection is performed. The system is replaced only when it is in the failed state.

2. *Age replacement.* The system is replaced at age t or when it fails, whichever occurs first.

3. *Sequential inspection.* If the system state is i, then (a) the system is replaced immediately, or (b) the system is scheduled for inspection at t_i later. The set $\{t_i\}_0^n$ and the criteria for decision (replacement or continuing operation) must be specified initially.
4. *Periodic inspection.* This is a special case of sequential inspection wherein t_i is the same for all i.
5. *Continuous inspection.* The system is inspected continuously. When the system enters state i, then (a) the system is replaced immediately, or (b) the system is allowed to continue operating. The criteria for the decision must be specified initially.

The optimal policy minimizes the mean long run cost rate, and when the optimal policy has been computed for each strategy of the above five strategies, then the maintenance policy can be chosen.

8.14 Applications of maintenance optimization models

Finding optimal maintenance policies clearly is in the area of applied mathematics, and in this section we present some real world applications as well as problems that arise when implementing maintenance policies.

Hopp and Kuo [63] look at the maintenance of an aircraft engine. They divide the parts into non-safety-critical and safety-critical parts. The non-safety-critical parts will not fail, but their performance loss cost will increase with age. The safety-critical parts have lifetime distributions, and when a safety-critical part fails it destroys the whole system. However, there are no performance loss costs associated with the safety-critical parts. The authors draw the conclusion that optimal policies are likely to be extremely difficult to compute and, because their form is complex, very difficult to communicate and use in practice. Therefore, heuristics are suggested for the case of a system with non-safety-critical components only, and for the case of a system with one safety-critical component and multiple non-safety-critical components. Lower bounds are computed to evaluate the performance of the heuristics.

In [64] the maintenance of the compressor of an aircraft engine is considered. It is assumed that fatigue crack is the underlying failure mechanism, and the crack growth is due to the number of "shocks" monitored by sensors. The true crack size is unobservable unless some maintenance task is performed on it. The observed information about the crack growth process is the crack size found at the most recent inspection/replacement and the number of shocks experienced since then. At the beginning of each flight it is decided, based on the observed state and the number of shocks to be incurred during the flight, whether or not to schedule an inspection at the end of the current flight. After inspection the true crack size will become known, and it must be decided whether a blade replacement is needed or not. A dynamic programming recursion for the problem is developed. The authors point out that a general policy from a complex dynamic program can be difficult to communicate, and therefore it is useful to characterize the optimal policy as having some kind of simple structured form. This turns out to be possible for the compressor maintenance problem.

A deeper treatment of the theory and the applications presented above in [63] and [64] can be found in [79].

An application to compressor units on jet engines is given by Epstein and Wilamowsky [54]. Specifically, B is a component (disk) of a system (compressor unit) with the following properties:

1. The system is known to fail exponentially (not due to B) at which time it is brought to the maintenance base at a specific cost.

2. B never fails but has a life-limit of Y hours. After Y hours, the machine must be brought to the maintenance base to exchange B.

It is shown that the optimal strategy for B is given by an (X_1, X_2)-rule such that:

1. If, when the system fails, B is accessible, replace B if it has less than X_2 hours of remaining life.

2. If, when the system fails, B is not accessible, replace B if it has less than X_1 hours of remaining life ($X_1 \leq X_2$).

3. Under all conditions replace B after Y hours of usage.

X_1 and X_2 are computed by minimizing the expected cost per expected life for B. In the case of several disks this policy is not optimal, but the authors develop a model for a system of two disks which can be extended to several disks.

Holland and McLean [62] describe a practical procedure to obtain approximate optimal replacement policies for homogeneous pieces of equipment. They suggest a method for fitting the Weibull distribution to observed data and from the estimated distribution parameters a replacement policy is found.

Crocker and Kumar [35] introduce the concepts of hard life and soft life. Hard life is defined as the age of the component at which it has to be replaced. Soft life is the age of the component after which it will be rejected the next time the engine is recovered. The method can be considered as a modified age replacement policy. The authors use Monte Carlo simulation to find the optimal value of the soft life and hard life for a very simplified aircraft engine, namely one that consists of one part only.

Kumar et al. [78] look at the concept of evolutionary maintenance, which means that the maintenance and inspection schedule is adjusted after every service activity. The procedure requires knowledge of the time to failure distribution of each module of the engine and the corresponding hard lives (i.e., the age at which a component will be replaced preventively) of any of its components. After every maintenance and inspection activity, the hard life and preventive maintenance schedule are revised by adjusting the maintenance interval without increasing the risk of failure. A simple example, which considers a high pressure turbine whose time to failure follows a Weibull distribution, is given which clearly shows how the hard life can be extended.

Jardine et al. [69] discuss work completed at Cardinal River Coals in Canada to improve the existing oil analysis condition monitoring program being undertaken for wheel motors. The proportional hazards model approach is used to identify the key condition variables relating to failure. Those key variables are then incorporated into a decision model that provides an optimal recommendation on whether to continue operating a wheel motor or remove it for overhaul on the basis of data obtained from an oil sample. It turns out that the overhaul costs were reduced by 20–30 percent.

Because of the very high lost production cost in chemical plants, maintenance and reliability play a crucial role. In [125] a general framework for preventive maintenance optimization in chemical process operations is developed. How to apply policies, like age and block replacement, is discussed. The authors suggest a combination of a genetic algorithm and Monte Carlo simulation to optimize different objective functions.

Zohrul-Kabir [146] evaluates the overhaul/replacement policy for a fleet of buses. The study is focused on evaluating the effectiveness of the existing overhaul routine and determining the overhaul cost limit. A model for the optimal cost limit is derived, and the existing data is analyzed to get input to the model. Numerical results and a small sensitivity analysis are presented.

Dekker discusses in [40] and [41] the practical role of optimization models. The focus is on applications yielding advice to management concerning the maintenance on existing systems. One main problem is that there is quite a diversity in maintenance problems and most of them are not repetitive, which means that standard models most often can not be used. Furthermore, in many areas one complains about the gap between theory and practice, for example:

- Maintenance optimization models are difficult to understand and to interpret.
- Many papers have been written for mathematical purposes only.
- Although many good ideas have been developed in industry, only a small amount has appeared in scientific literature.
- Despite the multitude of models there is little knowledge of which models are suited for which practical problems.
- The existing models often focus on planned maintenance, but in practice one is more often interested in condition-based maintenance.

Scarf [112] thinks that the connection between the mathematical maintenance models and real maintenance-related problems is too weak. He points out that what in the literature is called applications and case studies appear to have been motivated by the need to find an application for a particular model, rather than by the solution of the problem of interest to engineers and managers.

8.15 Conclusions

The foregoing survey describes the literature related to optimization maintenance and replacement models for multi-unit systems that can be useful when modelling the maintenance of an aircraft engine. The main part of the literature assumes that

- the systems consist of stochastic parts only;
- the time horizon is infinite; and
- a policy is used to find a replacement scheme.

From the literature it turns out that it is extremely hard to find an optimal replacement schedule when the number of parts is larger than 2, and hence different replacement policies are developed. Such policies reduce the complexity of the problems, but the solutions found are most often not optimal. Also, the literature points out that the case of a finite time horizon is even harder than the infinite time horizon case.

In our aircraft application the time horizon is finite and the number of parts is large (about 50), so if all of the parts had been stochastic it would have been necessary to use replacement policies. However, about 75% of the components considered in an aircraft engine are deterministic (that is, they have estimated fixed lifetimes), so our problem is more structured than the completely stochastic systems considered in the literature. Therefore it might be possible to find optimal solutions to the replacement problem and avoid the use of policies. This is the subject of this thesis.

Our idea is to model the deterministic system with an integer linear programming model based on the model presented in [50], and then to use this model to formulate a two-stage stochastic model for the system with both deterministic and stochastic parts. This approach has not been found in the literature. In future research it would be interesting to develop appropriate replacement policies for systems consisting of both deterministic and stochastic parts and compare them with the two-stage model developed in this thesis.

Bibliography

[1] M. Abdel-Hameed, *Inspection, maintenance and replacement models*, Computers and Operations Research, 22 (1995), pp. 435–441.

[2] D. Ait-Kadi and R. Cléroux, *Optimal block replacement policies with multiple choice of failure*, Naval Research Logistics, 35 (1988), pp. 99–110.

[3] S. L. Albin and S. Chao, *Preventive replacement in systems with dependent components*, IEEE Transactions on Reliability, 41 (1992), pp. 230–238.

[4] D. Anbar, *An asymptotically optimal inspection policy*, Naval Research Logistics Quarterly, 23 (1976), pp. 211–218.

[5] T. W. Archibald and R. Dekker, *Modified block-replacement for multiple-component systems*, IEEE Transactions on Reliability, 45 (1996), pp. 75–83.

[6] D. Assaf and J. G. Shanthikumar, *Optimal group maintenance policies with continuous and periodic inspections*, Management Science Journal, 33 (1987), pp. 1440–1452.

[7] T. Aven and B. Bergman, *Optimal replacement times - a general setup*, Journal of Applied Probabiliy, 23 (1986), pp. 432–442.

[8] T. Aven and R. Dekker, *A useful framework for optimal replacement models*, Reliability Engineering and System Safety, 58 (1997), pp. 61–67.

[9] I. Bagai and K. Jain, *Improvement, deterioration, and optimal replacement under age-replacement with minimal repair*, IEEE Transactions on Reliability, 43 (1994), pp. 156–162.

[10] K. Bahrami-G., J. W. H. Price, and J. Mathew, *The constant-interval replacement model for preventive maintenance: A new perspective*, International Journal of Quality and Reliability Management, 17 (2000), pp. 822–838.

[11] E. Balas, *Disjunctive programming*, Annals of Discrete Mathematics, 5 (1979), pp. 3–51.

[12] R. BARLOW AND F. PROSCHAN, *Mathematical Theory of Reliability*, Wiley, New York, 1965.

[13] J. BEASLEY, *An algorithm for the steiner problem in graphs*, Networks, 14 (1984), pp. 147–159.

[14] ——, *An sst-based algorithm for the steiner problem in graphs*, Networks, 19 (1989), pp. 1–16.

[15] C. R. BECTOR, O. HAWALESHKA, AND A. GILL, *A $0-1$ linear programming formulation for solving an equipment replacement problem*, International Journal of Systems Science, 23 (1992), pp. 2393–2399.

[16] R. BELLMAN, *Dynamic programming*, Princeton University Press, 1957.

[17] R. E. BELLMAN AND S. E. DREYFUS, *Applied Dynamic Programming*, Princeton University Press, 1962.

[18] M. BEN-DAYA AND A. S. ALGHAMDI, *On an imperfect preventive maintenance model*, International Journal of Quality and Reliability Management, 17 (2000), pp. 661–670.

[19] C. BERENGUER, C. CHU, AND A. GRALL, *Inspection and maintenance planning: an application of semi-markov decision processes*, Journal of Intelligent Manufacturing, 8 (1997), pp. 467–476.

[20] M. BERG, *General trigger-off replacement procedures for two-unit systems*, Naval Research Logistics Quarterly, 25 (1978), pp. 15–29.

[21] ——, *A marginal cost analysis for preventive replacement policies*, European Journal of Operational Research, 4 (1980), pp. 136–142.

[22] ——, *The marginal cost analysis and its application to repair and replacement policies*, European Journal of Operational Research, 82 (1995), pp. 214–224.

[23] M. BERG AND B. EPSTEIN, *A modified block replacement policy*, Naval Research Logistics Quarterly, 23 (1976), pp. 15–24.

[24] ——, *Comparison of age, block, and failure replacement policies*, IEEE Transactions on Reliability, 27 (1978), pp. 25–30.

[25] ——, *A note on a modified block replacement policy for units with increasing marginal running costs*, Naval Research Logistics Quarterly, 26 (1979), pp. 157–160.

[26] M. BEVILACQUA AND M. BRAGLIA, *The analytic hierarchy process applied to maintenance strategy selection*, Reliability Engineering and System Safety, 70 (2000), pp. 71–83.

[27] J. R. Birge and F. Louveaux, *Introduction to stochastic programming*, Springer, 1997.

[28] D. A. Butler and G. J. Lieberman, *Inspection policies for fault location*, Operations Research, 32 (1984), pp. 566–574.

[29] C. R. Cassady, L. M. Maillart, R. O. Bowden, and B. K. Smith, *Characterization of optimal age-replacement policies*, Proceedings Annual Reliability and Maintainability Symposium, (1998), pp. 170–175.

[30] J.-K. Chan and L. Shaw, *Modeling repairable systems with failure rates that depend on age and maintenance*, IEEE Transactions on Reliability, 42 (1993), pp. 566–571.

[31] A. Chelbi and D. Aït-Kadi, *Spare provisioning strategy for preventively replaced systems subjected to random failure*, International Journal of Production Economics, 74 (2001), pp. 183–189.

[32] D. I. Cho and M. Parlar, *A survey of maintenance models for multi-unit systems*, European Journal of Operational Research, 51 (1991), pp. 1–23.

[33] S. Chopra, E. Gorres, and M. R. Rao, *Solving a steiner tree problem on a graph using branch and cut*, ORSA J. Comput., 4 (1992), pp. 320–335.

[34] W. J. Cook, W. H. Cunningham, W. R. Pulleyblank, and A. Schrijver, *Combinatorial optimization*, Wiley, 1998.

[35] J. Crocker and U. D. Kumar, *Age-related maintenance versus reliability centred maintenance: a case study on aero-engines*, Reliability Engineering and System Safety, 67 (2000), pp. 113–118.

[36] K. Dagdelen, *Optimum Multi Period Open Pit Mine Production*, PhD thesis, Colorado School of Mines, 1985.

[37] G. B. Dantzig, *Linear programming and extensions*, Princeton University Press, 1963.

[38] G. B. Dantzig and M. N. Tapa, *Linear programming 2: Theory and extensions*, Springer, 2003.

[39] R. Dekker, *Integrating optimisation, priority setting, planning and combining of maintenance activities*, European Journal of Operational Research, 82 (1995), pp. 225–240.

[40] ——, *On the use of operations research models for maintenance decision making*, Microelectronics and Reliability, 35 (1995), pp. 1321–1331.

[41] ——, *Applications of maintenance optimization models: A review and analysis*, Reliability Engineering and System Safety, 51 (1996), pp. 229–240.

[42] R. Dekker and M. Dijkstra, *Opportunity-based age replacement: Exponential distributed times between opportunities*, Naval Research Logistics, 39 (1992), pp. 175–190.

[43] R. Dekker and I. F. K. Roelvink, *Marginal cost criteria for preventive replacement of a group of components*, European Journal of Operational Research, 84 (1995), pp. 467–480.

[44] R. Dekker and E. Smeitink, *Opportunity based block replacement*, European Journal of Operational Research, 53 (1991), pp. 46–62.

[45] R. Dekker, F. van der Duyn Schouten, and R. Wildeman, *A review of multi-component maintenance models with economic dependencies*, Mathematical Methods of Operations Research, 45 (1997), pp. 411–435.

[46] R. Dekker, R. E. Wildeman, and F. A. v. Duyn-Schouten, *A review of multi-component maintenance models with economic dependence*, Mathematical Methods of Operations Research, 45 (1997), pp. 411–435.

[47] E. V. Denardo, *Dynamic Programming Models and Applications*, Prentice-Hall, 1982.

[48] C. Derman, G. J. Lieberman, and S. M. Ross, *On the use of replacements to extend system life*, Operations Research, 32 (1984), pp. 616–627.

[49] C. Derman and D. R. Smith, *Renewal decision problem-random horizon*, Mathematics of Operations Research, 4 (1979), pp. 225–232.

[50] B. Dickman, S. Epstein, and Y. Wilamowsky, *A mixed integer linear programming formulation for multi-component deterministic opportunistic replacement*, The Journal of the Operational Research Society of India, 28 (1991), pp. 165–175.

[51] S. E. Dreyfus and A. M. Law, *The Art and Theory of Dynamic Programming*, Academic Press, 1977.

[52] S. E. Dreyfus and R. A. Wagner, *The steiner problem in graphs*, Networks, 1 (1971), pp. 195–207.

[53] D. B. Edmunds, *Modular engine maintenance concept considerations for aircraft turbine engines*, Aircraft Engineering, 50 (1978), pp. 14–17.

[54] S. Epstein and Y. Wilamowsky, *A replacement schedule for multicomponent life-limited parts*, Naval Research Logistics Quarterly, 29 (1982), pp. 685–692.

[55] W. W. Fisher and J. J. Brennan, *The performance of cannibalization policies in a maintenance system with spares, repair and resource constraints*, Naval Research Logistics Quarterly, 33 (1986), pp. 1–15.

[56] R. Fourer, D. M. Gay, and B. W. Kernighan, *AMPL: A modeling language for mathematical programming*, Boyd & Fraser, 1993.

[57] B. Fox, *Age replacement with discounting*, Operations Research, 14 (1966), pp. 533–537.

[58] J. B. G. Frenk, R. Dekker, and M. J. Kleijn, *A unified treatment of single component replacement models*, tech. rep., Econometric Institute, Erasmus University Rotterdam, The Netherlands, 1996.

[59] R. Garfinkel and G. L. Nemhauser, *Perspective on optimization: A collection of expository articles*, Addison-Wesley, 1972, ch. Optimal set covering: a survey, pp. 164–193.

[60] R. S. Garfinkel and G. L. Nemhauser, *Integer programming*, Wiley, 1972.

[61] A. Haurie and P. L'Ecuyer, *A stochastic control approach to group preventive replacement in a multicomponent system*, IEEE Transactions on Automatic Control, 27 (1982), pp. 387–393.

[62] C. W. Holland and R. A. McLean, *Applications of replacement theory*, AIIE Transactions, 7 (1975), pp. 42–47.

[63] W. J. Hopp and Y.-L. Kuo, *Heuristics for multicomponent joint replacement: Application to aircraft engine maintenance*, Naval Research Logistics, 45 (1998), pp. 435–458.

[64] ——, *An optimal structured policy for maintenance of partially observable aircraft engine components*, Naval Research Logistics, 45 (1998), pp. 335–352.

[65] H. Hummels, *Safety and aircraft maintenance*, International Journal of Value-Based Management, 10 (1997), pp. 127–146.

[66] F. K. Hwang and D. S. Richards, *Steiner tree problems*, Networks, 22 (1992), pp. 55–89.

[67] F. K. Hwang, D. S. Richards, and P. Winter, *The steiner tree problem*, Annals of Discrete Mathematics, 53 (1992).

[68] ILOG, *ILOG CPLEX 6.5: Reference manual*, 1999.

[69] A. K. S. Jardine, D. Banjevic, M. Wiseman, S. Buck, and T. Joseph, *Optimizing a mine haul truck wheel motors' condition monitoring program*, Journal of Quality in Maintenance Engineering, 7 (2001), pp. 286–301.

[70] J. P. JHANG AND S. H. SHEU, *Opportunity-based age replacement policy with minimal repair*, Reliability Engineering and System Safety, 64 (1999), pp. 339–344.

[71] L. M. JR, A. SAXENA, AND G. M. KNAPP, *Statistical-based or condition-based preventive maintenance?*, Journal of Quality in Maintenance Engineering, 1 (1995), pp. 46–59.

[72] P. KALL AND S. W. WALLACE, *Stochastic programming*, Wiley, 1994.

[73] R. M. KARP, *Reducibility among combinatorial problems*, Complexity of Computer Computations, (1972), pp. 85–103.

[74] D. KHALIFA, M. HOTTENSTEIN, AND S. AGGARWAL, *Cannibalization policies for multistate systems*, Operations Research, 25 (1977), pp. 1032–1039.

[75] K. A. H. KOBBACY, D. F. PERCY, AND B. B. FAWZI, *Small data sets and preventive maintenance modelling*, Journal of Quality in Maintenance Engineering, 3 (1997), pp. 136–142.

[76] T. KOCH AND A. MARTIN, *Solving steiner tree problems in graphs to optimality*, Networks, 32 (1998), pp. 207–232.

[77] D. KUMAR AND U. WESTBERG, *Maintenance scheduling under age replacement policy using proportional hazards model and TTT-plotting*, European Journal of Operational Research, 99 (1997), pp. 507–515.

[78] U. D. KUMAR, J. CROCKER, AND J. KNEZEVIC, *Evolutionary maintenance for aircraft engines*, Reliability and Maintainability Symposium, (1999), pp. 62–68.

[79] Y. L. KUO, *Scheduling Aircraft Engine Maintenance: Modeling and Optimization*, PhD thesis, Northwestern University, Evanston, 1995.

[80] C. T. LAM AND R. H. YEH, *Optimal maintenance-policies for deteriorating systems under variorus maintenance strategies*, IEEE Transactions on Reliability, 43 (1994), pp. 423–430.

[81] E. L. LAWLER, *Combinatorial optimization: Networks and matroids*, Holt, Rinehart and Winston, 1976.

[82] P. L'ECUYER AND A. HAURIE, *Preventive replacement for multicomponent systems: An opportunistic discrete-time dynamic programming model*, IEEE Transactions on Reliability, 32 (1983), pp. 117–118.

[83] V. LEGÁT, A. H. ŽALUDOVÁ, V. ČERVENKA, AND V. JURČA, *Contribution to optimization of preventive replacement*, Reliability Engineering and System Safety, 51 (1996), pp. 259–266.

[84] W.-P. LIAO AND J. YUAN, *Optimal replacement for a one-unit system subject to delivery and test*, Journal of Quality in Maintenance Engineering, 4 (1998), pp. 51–65.

[85] D. LIN, M. J. ZUO, AND R. C. M. YAM, *Sequential imperfect preventive maintenance models with two categories of failure modes*, Naval Research Logistics, 48 (2001), pp. 172–183.

[86] C. E. LOVE, M. A. ZITRON, AND Z. G. ZHANG, *An smdp approach to optimal repair/replacement decisions for systems experiencing imperfect repairs*, Journal of Quality in Maintenance Engineering, 4 (1998), pp. 131–149.

[87] N. MACULAN, *The steiner problem in graphs*, Ann. Discr. Math., 31 (1987), pp. 185–212.

[88] T. A. MAZZUCHI AND R. SOYER, *A bayesian perspective on some replacement strategies*, Reliability Engineering and System Safety, 51 (1996), pp. 295–303.

[89] A. MONGA, M. J. ZUO, AND R. TOOGOOD, *System design with deteriorative components for minimal life cycle costs*, IEEE International Conference on Systems, Man, and Cybernetics, (1995), pp. 1843–1848.

[90] S. MORARU, F. SISAK, AND I. POPESCU, *Overview of preventive maintenance policies for systems, generally, and for digital systems, particularly*, Proceedings of the 6th International Conference on Optimization of Electrical and Electronic Equipments, 3 (1998), pp. 801–806.

[91] T. NAKAGAWA AND S. OSAKI, *Optimum replacement policies with delay*, Journal of Applied Probability, 11 (1974), pp. 102–110.

[92] G. L. NEMHAUSER AND L. A. WOLSEY, *Integer and combinatorial optimization*, Wiley, 1988.

[93] L. NEMHAUSER, *Introduction to Dynamic Programming*, Wiley, 1966.

[94] J. M. V. NOORTWIJK, *Optimal maintenance decisions on the basis of uncertain failure probabilities*, Journal of Quality in Maintenance Engineering, 6 (2000), pp. 113–122.

[95] H. C. NORIEGA AND P. F. F. E MELO, *A non-stationary point-process maintenance-model for multi-unit systems*, Proceedings Annual Reliability and Maintainability Symposium, (1998), pp. 189–194.

[96] O. G. OKOGBAA AND X. PENG, *Time series intervention analysis for preventive/predictive maintenance management of multi-unit system*, IEEE International Conference on Systems, Man, and Cybernetics, 5 (1998), pp. 4659–4664.

[97] M.-S. Ouali, D. Ait-Kadi, and A. Gharbi, *A simulation model for opportunistic maintenance strategies*, IEEE International 7th Conference on Emerging Technologies and Factory Automation, 1 (1999), pp. 703–708.

[98] C. H. Papadimitriou and K. Stieglitz, *Combinatorial optimization: Algorithms and complexity*, Prentice-Hall, 1982.

[99] K. S. Park, *Optimal wear-limit replacement with wear-dependent failures*, IEEE Transactions on Reliability, 37 (1988), pp. 293–294.

[100] ——, *Condition-based predictive maintenance by multiple logistic function*, IEEE Transactions on Reliability, 42 (1993), pp. 556–560.

[101] D. F. Percy and K. A. H. Kobbacy, *Preventive maintenance modelling: A bayesian perspective*, Journal of Quality in Maintenance Engineering, 2 (1996), pp. 15–24.

[102] H. Pham and H. Wang, *Optimal* (τ, t) *opportunistic maintenance of a k-out-of-n:g system with imperfect pm and partial failure*, Naval Research Logistics, 47 (2000), pp. 223–239.

[103] W. P. Pierskalla and J. A. Voelker, *A survey of maintenance models: The control and surveillance of deteriorating systems*, Naval Research Logistics Quarterly, 23 (1979), pp. 353–388.

[104] E. Popova and H.-C. Wu, *Renewal reward processes with fuzzy rewards and their applications to t-age replacement policies*, European Journal of Operational Research, 117 (1999), pp. 606–617.

[105] E. Pärt-Enander and A. Sjöberg, *Användarhandledning för MATLAB 5*, Uppsala Universitet, 1998.

[106] A. Ran and S. I. Rosenlund, *Age replacement with discounting for a continuous maintenance cost model*, Technometrics, 18 (1976), pp. 459–465.

[107] D. M. Reineke, W. P. Murdock, E. A. Pohl, and I. Rehmert, *Improving availability and cost performance for complex systems with preventive maintenance*, Proceedings Annual Reliability and Maintainability Symposium, (1999), pp. 383–388.

[108] B. R. Sarker and J. Yu, *A balanced maintenance schedule for a failure-prone system*, International Journal of Quality and Reliability Management, 12 (1995), pp. 183–191.

[109] R. Sarker and A. Haque, *Optimization of maintenance and spare provisioning policy using simulation*, Applied Mathematical Modelling, 24 (2000), pp. 751–760.

[110] D. A. SAVIC, G. A. WALTERS, AND J. KNEZEVIC, *Optimal, opportunistic maintenance policy using genetic algorithms, 1: Formulation*, Journal of Quality in Maintenance Engineering, 1 (1995), pp. 34–49.

[111] ——, *Optimal, opportunistic maintenance policy using genetic algorithms, 2: Analysis*, Journal of Quality in Maintenance Engineering, 1 (1995), pp. 25–34.

[112] P. A. SCARF, *On the application of mathematical models in maintenance*, European Journal of Operational Research, 99 (1997), pp. 493–506.

[113] R. L. SCHEAFFER, *Optimum age replacement with an increasing cost factor*, Technometrics, 13 (1971), pp. 139–144.

[114] A. SCHRIJVER, *Theory of linear and integer programming*, Wiley, 1986.

[115] ——, *Combinatorial optimization: Polyhedra and efficiency*, Springer, 2003.

[116] Y. SHERIF AND M. SMITH, *Optimal maintenance models for systems subject to failure - a review*, Naval Research Logistics Quarterly, 28 (1981), pp. 47–74.

[117] D. J. SHERWIN, *Age-based opportunity maintenance*, Journal of Quality in Maintenance Engineering, 5 (1999), pp. 221–235.

[118] S.-H. SHEU, *A generalized age and block replacement of a system subject to shocks*, European Journal of Operational Research, 108 (1998), pp. 345–362.

[119] S.-H. SHEU AND J.-P. JHANG, *A generalized group maintenance policy*, European Journal of Operational Research, 96 (1996), pp. 232–247.

[120] S.-H. SHEU, R. H. YEH, Y.-B. LIN, AND M.-G. JUANG, *A bayesian perspective on age-replacement with minimal repair*, Reliability Engineering and System Safety, 65 (1999), pp. 55–64.

[121] S. H. SIM AND J. ENDRENYI, *Optimal preventive maintenance with repair*, IEEE Transactions on Reliability, 37 (1988), pp. 92–96.

[122] V. P. SIMPSON, *Optimum solution structure for a repairable inventory problem*, Operations Research, 26 (1978), pp. 270–281.

[123] M. A. J. SMITH AND R. DEKKER, *Preventive maintenance in a* 1 *out of* n *system: the uptime, downtime and costs*, European Journal of Operational Research, 99 (1997), pp. 565–583.

[124] J. SVENSSON, *Two statistical models used on aircraft engine data modeling times between repairs*. Department of Mathematical Statistics at Chalmers University of Technology, 2004.

[125] J. S. TAN AND M. A. KRAMER, *A general framework for preventive maintenance optimization in chemical process operations*, Computers and Chemical Engineering, 21 (1997), pp. 1451–1469.

[126] T. TANGO, *Extended block replacement policy with used items*, Journal of Applied Probability, 15 (1978), pp. 560–572.

[127] H. M. TAYLOR AND B. E. RODRIGUEZ, *Optimal replacement for fault-tolerant systems*, Naval Research Logistics Quarterly, 33 (1986), pp. 747–757.

[128] J. THEANDER, *Förstudie till examensarbete*, tech. rep., Volvo Aero Corporation, 1997.

[129] L. C. THOMAS, *Replacement of systems and components in renewal decision problems*, Operations Research, 33 (1985), pp. 404–411.

[130] A. H. C. TSANG, *Condition-based maintenance: Tools and decision making*, Journal of Quality in Maintenance Engineering, 1 (1995), pp. 3–17.

[131] J. K. VAURIO, *Availability and cost functions for periodically inspected preventively maintained units*, Reliability Engineering and System Safety, 63 (1999), pp. 133–140.

[132] R. C. VERGIN AND M. SCRIABIN, *Maintenance scheduling for multicomponent equipment*, AIIE Transactions, 9 (1977), pp. 297–305.

[133] H. M. WAGNER, *Principles of operations research with applications to managerial decisions*, Prentice-Hall, 1975.

[134] R. WEBSTER, *Convexity*, Oxford University Press, 1994.

[135] C. E. WELLS AND J. L. BRYANT, *Optimal preventive maintenance policies for systems with missions of random duration*, IIE Transactions Industrial Engineering Research and Development, 17 (1985), pp. 338–345.

[136] D. J. WHITE, *Dynamic Programming*, Holden-Day, 1969.

[137] R. E. WILDEMAN, R. DEKKER, AND A. C. J. M. SMIT, *A dynamic policy for grouping maintenance activities*, European Journal of Operational Research, 99 (1997), pp. 530–551.

[138] P. WINTER, *Steiner problems in networks: A survey*, Networks, 17 (1987), pp. 129–167.

[139] L. A. WOLSEY, *Integer Programming*, Wiley, 1998.

[140] R. T. WONG, *A dual ascent approach for steiner tree problems on a directed graph*, Mathematical Programming, 28 (1984), pp. 271–287.

[141] S.-C. YANG AND J. A. NACHLAS, *Bivariate reliability and availability modeling*, IEEE Transactions on Reliability, 50 (2001), pp. 26–35.

[142] S. K. Yang and T. S. Liu, *A petri net approach to early failure detection and isolation for preventive maintenance*, Quality and Reliability Engineering International, 14 (1998), pp. 319–330.

[143] W. Y. Yun and C. H. Choi, *Optimal replacement intervals with random time horizon*, Journal of Quality in Maintenance Engineering, 6 (2000), pp. 269–274.

[144] X. Zheng, *All opportunity-triggered replacement policy for multiple-unit system*, IEEE Transactions on Reliability, 44 (1995), pp. 648–652.

[145] X. Zheng and N. Fard, *Hazard-rate tolerance method for an opportunistic-replacement policy*, IEEE Transactions on Reliability, 41 (1992), pp. 13–20.

[146] A. B. M. Zohrul-Kabir, *Evaluation of overhaul/replacement policy for a fleet of buses*, Journal of Quality in Maintenance Engineering, 2 (1996), pp. 49–59.

[147] A. B. M. Zohrul-Kabir and A. S. Al-Olayan, *Joint optimization of age replacement and continuous review spare provisioning policy*, International Journal of Operations and Production Management, 14 (1994), pp. 53–69.

[148] ——, *A stocking policy for spare part provisioning under age based preventive replacement*, European Journal of Operational Research, 90 (1996), pp. 171–181.

[149] A. B. M. Zohrul-Kabir and S. H. A. Farrash, *Simulation of an integrated age replacement and spare provisioning policy using slam*, Reliability Engineering and System Safety, 52 (1996), pp. 129–138.

[150] S. Özekici, *Optimal periodic replacement of multicomponent reliability system*, Operations Research, 36 (1988), pp. 542–552.

[151] ——, *Optimal maintenance policies in random environments*, European Journal of Operational Research, 82 (1995), pp. 283–294.