Aalto University
School of Science
Master's Programme in Mathematics and Operations Research

Niko Laakkonen

# Optimization Problems in Outpatient Clinic Production Planning and Control

Master's Thesis
Espoo, February 15, 2019

Supervisor: Professor Harri Ehtamo
Advisor: Vesa Paju M.Sc. (Tech.)

Aalto University
School of Science
Master's Programme in Mathematics and Operations Research

ABSTRACT OF MASTER'S THESIS

| | | | |
|---|---|---|---|
| **Author:** | Niko Laakkonen | | |
| **Title:** Optimization Problems in Outpatient Clinic Production Planning and Control | | | |
| **Date:** | February 15, 2019 | **Pages:** | 75 |
| **Major:** | Systems and Operations Research | **Code:** | SCI3055 |
| **Supervisor:** | Professor Harri Ehtamo | | |
| **Advisor:** | Vesa Paju M.Sc. (Tech.) | | |

The outpatient clinic production planning and control is an important topic due to increasing patient volumes and expenditures. Conducting the patient appointments efficiently while ensuring quality of care requires the patient processes and the operating model of the clinic to be in good shape. Mathematical optimization of different operational decisions is a promising tool since it can quantitatively compare options.

A natural point to change the operation occurs while designing new spaces for the operation. At this point, creating and testing new operating models can be done using a simulation requiring a scheduling algorithm. This thesis targets to formulate and solve an outpatient clinic scheduling as an optimization problem. The solution algorithm should be fast but does not necessarily solve the problem to a global optimum. The schedules created by the algorithm can be used to compare different clinic setups.

The problem is formulated as an integer linear program ending up being huge by number of decision variables and constraints. Three heuristic solution algorithms are provided including genetic, first-fit-decreasing and rules based scheduling algorithms. The rules based scheduling algorithm is based on a currently used algorithm. The objective is to schedule appointments on resources while minimizing the number of resources used, staff room changes during day and appointment inbalance within the five day scheduling interval. The rules based scheduling algorithm performed best compared to the requirements supplying schedules with highest objective values out of the three algorithms.

| | |
|---|---|
| **Keywords:** | optimization, production planning and control, healthcare, mathematical programming, scheduling, outpatient clinic |
| **Language:** | English |

Aalto-yliopisto
Perustieteiden korkeakoulu
Matematiikan ja operaatiotutkimuksen maisteriohjelma

DIPLOMITYÖN TIIVISTELMÄ

| | | | |
|---|---|---|---|
| **Tekijä:** | Niko Laakkonen | | |
| **Työn nimi:** Optimointiongelmia poliklinikoiden tuotannonsuunnittelussa ja -ohjauksessa | | | |
| **Päiväys:** | 15. helmikuuta, 2019 | **Sivumäärä:** | 75 |
| **Pääaine:** | Systeemi- ja operaatiotutkimus | **Koodi:** | SCI3055 |
| **Valvoja:** | Professori Harri Ehtamo | | |
| **Ohjaaja:** | DI Vesa Paju | | |

Terveydenhuollon poliklinikoiden tuotannonsuunnittelu ja -ohjaus on tärkeä aihe kasvavien potilasmäärien ja kustannuksien vuoksi. Potilasvastaanottojen järjestäminen tehokkaasti säilyttäen korkealaatuisen hoidon vaatii hoito- ja potilasprosessien toimivuutta. Toiminnan järjestämiseen liittyvien päätösten tukeminen matemaattisella optimoinnilla tuottaa kvantitatiivisen tavan vertailla eri vaihtoehtoja.

Luonnollinen piste muuttaa terveydenhuollon yksikön toimintaa on uusien tilojen suunnittelussa. Uusien toimintamallien luomisessa ja testaamisessa voidaan hyödyntää simulointia, mikä vaatii aikataulutusalgoritmin taustalleen. Tässä työssä formuloidaan ja ratkaistaan poliklinikan aikataulutus optimointiongelmana. Aikataulutusalgoritmin tulee olla nopea ja globaalista optimiratkaisusta voidaan tinkiä. Tuotettujen aikataulujen avulla voidaan tarkastella eroja poliklinikan erilaisissa toimintamalleissa.

Poliklinikan aikataulutus kirjoitetaan lineaarisena kokonaislukuoptimoinnin tehtävänä, mikä osoittautuu valtavaksi päätösmuuttujien ja rajoitusehtojen määrällä tarkasteltuna. Ongelman ratkaisemiseen esitellään kolme heuristista algoritmia sisältäen geneettisen algoritmin, pakkausalgoritmin sekä sääntöpohjaisen aikataulutusalgoritmin. Sääntöpohjainen algoritmi perustuu käytössä olevaan algoritmiin. Ongelman tavoitteena on aikatauluttaa vastaanottoja resursseille minimoiden käytettyjen resurssien lukumäärää, henkilökunnan huonevaihtoja päivän aikana ja vastaanottojen ajoittumisen epätasapainoa käytetyllä viiden päivän aikataulutushorisontilla. Tuloksena sääntöpohjainen algoritmi suoriutui tehtävästä parhaiten ajatellen käyttäjien tarpeita kohdefunktion arvolla mitattuna.

| | |
|---|---|
| **Asiasanat:** | optimointi, tuotannonsuunnittelu, terveydenhuolto, aikataulutus, poliklinikka, avohoito |
| **Kieli:** | englanti |

# Contents

# Terms

| | |
|---|---|
| Access Policy | Chosen access method of patients for example walk-in or by appointment. Basically waiting list management. |
| Appointment Schedule | A blueprint for scheduling single patient appointments in outpatient clinic setting. |
| Appointment Scheduling | A term for the process of setting up an outpatient clinic appointment schedule. |
| Bin-Packing Problem | An optimization problem where bins/containers are filled with as many items as possible. |
| Constraint | A limitation to the decision variables in a mathematical program. |
| Consultation Session | A time interval within which patient appointments are scheduled. |
| Decision Variable | A variable for which the mathematical program seeks optimal values. |
| Deterministic | The outcome of an event is known previously by the inputs and no randomness is included. |
| Feasible Solution | A value set for decision variables which satisfies all of the constraints. |
| Fitness Function | Term for the objective function usually used in conjunction with the genetic algorithm. |
| Hard Constraint | Must be satisfied at all costs. |
| Heuristic | A solution method which does not necessarily provide optimal solution. |
| Inpatient | A patient requiring a prolonged stay at the healthcare facility, for example at the wards. |
| Master Surgical Schedule | Allocation of time between specialties in an operating theatre. Basis for single operation scheduling. |

| | |
|---|---|
| NP-hard | A problem which is not solvable with a polynomial time algorithm. These problems are hard and often require heuristic solution methods. |
| Objective Function | A mathematical quantity to be minimized or maximised in a mathematical program. |
| Offline Algorithm | Works on the whole set of objects, knowing the future objects during execution. |
| Online Algorithm | Works a single object at a time without knowing the future objects. |
| Optimal Solution | A feasible solution which also optimizes the objective function value. |
| Outpatient | A patient not requring overnight stay at the healthcare facility. |
| Patient Group | Patients with similar medical conditions. |
| Soft Constraint | Can be violated but violation will be penalized. |
| Stochastic | The outcome is determined and known only when the event happens. |

# Acknowledgements

Finishing the master's thesis was not an easy task. It included countless evenings and many weekends to write. After approximately one year, the thesis has been completed and it is time to show some appreciation for those involved in the process.

First, I would like to thank Delfoi Ltd and instructor Vesa Paju for offering this opportunity and interesting subject to write the thesis on. Also, Vesa has given inspiring guidance and great instructions when necessary. Also, other colleagues at Delfoi receive a special thanks for the learning experience on projects related to the subject.

Second, thanks for supervisor professor Harri Ehtamo for giving feedback and guidance during the thesis project.

Third, I express my deepest gratitude for Taru for being supportive and cheerful on my thesis work at home. And having patience to watch me just sit at the computer and write. Special note for concretely helping by proofreading the thesis with academic expertise.

Finally, I thank everyone else involved including family and friends for supporting and listening while having difficulties with the thesis.

Espoo, February 15, 2019

Niko Laakkonen

# Chapter 1

# Introduction

## 1.1 Background

Healthcare sector faces increasing demand and expenses while having to provide high-quality care. Resources are scarce and the demand almost always exceeds the supply requiring high resource utilization and operational efficiency. Resources are often used jointly among different units of the healthcare organization.

In addition to resource scarcity, the healthcare operation is a rather complex setting. There are multiple educational backgrounds of staff, availability of specialist equipment and fit for purpose spaces, for example consultation rooms, to consider. The specialized equipment is often costly. On top of the resources, the patient is also a factor within the service process.

The bulk of appointments in healthcare comes from different outpatient clinics. The outpatient clinic provides ambulatory care usually by appointment having fixed opening hours. The clinics sometimes specialize on a certain specialty. For example, an outpatient clinic might specialize in neurology.

To counter these hurdles, healthcare operations planning and control has been studied extensively by researchers. The wide variety of operational research methods applied in providing support for management decisions include for example mathematical programming, simulation, heuristics, queuing theory, forecasting and decision making.

In addition to the continuous operations, the operational research tools can be used while designing new spaces for a healthcare operation. The cost of building new space is high and a basic question lies in the dimensioning of the new space. To help this dimensioning problem, the healthcare operation can be simulated requiring understanding and tools for the operational planning and control.

For a review in literature, Rais and Viana [33] survey the operations research methods used in healthcare and mention forecasting of demand, location selection for centres and vehicles, capacity planning, resource scheduling, nurse scheduling, operating room scheduling and logistics. In addition operations research methods have been used in the actual healthcare practises for example in organ donation and transplant queue policies utilizing decision making tools [33].

Delfoi is a Finnish consultancy offering services in both helping day-to-day operation planning through a software and during designing and dimensioning new spaces, operating models and logistics in the healthcare sector. This study is aimed at strengthening the understanding of the quantitative side of healthcare operations planning. The optimization problems among the production control are numerous due to different precision levels of planning and case specificity.

## 1.2 Research Problem

Optimization and simulation are already used by Delfoi in healthcare evaluations but further advancement is required in optimization of the mid and short term planning and scheduling of appointments in outpatient setting. On this level and time horizon of planning, the resources are usually already described to a single person level but the scheduled appointment demand can be considered on an aggregate or single appointment level.

The goal is to build a mathematical optimization model and describe a heuristic optimization algorithm to solve the model. To achieve this, the thesis answers to the following research questions.

- What kind of optimization problems exists among outpatient clinic production and planning? What is the focus of the optimization objectives and what aspects are considered as constraints?
- Can the short term planning and scheduling of appointments be modelled as an optimization problem and can it be solved using heuristic optimization algorithms? Do these algorithms provide fair results?

Previously, Nurmi [26] has studied the planning and control of healthcare operation in his master's thesis for Delfoi. This study complements and focuses to a sub part of the process described by Nurmi [26] further increasing the understanding of Delfoi on healthcare operations planning and control.

## 1.3 Methods and Structure

To answer the research questions multiple methods will be used.

**First**, a literature review will be conducted on the healthcare operations planning and control. The review will describe the planning process and drill deeper to a few selected optimization problems among the process. The use of these tools in terms of new building planning will be discussed.

**Afterwards**, an appointment planning and scheduling problem will be modelled as a constrained optimization problem. The optimization problem will be modelled on as a general level as possible to cover organizational differences. Different possibilities for objective function will be discussed. This part forces to thoroughly think about the building of the schedule and forms the basis for the next part.

**Finally**, heuristic solution algorithms are provided for the problem. The algorithms are compared based on run times and solution quality. The main focus is on generating proper appointment plans in reasonable computing time. The goal is not to generate the best, globally optimal solution but rather a fair schedule in reasonable computation time.

# Chapter 2

# Outpatient Clinic Optimization

This section revises optimization problems in outpatient clinics by researching literature. First, mathematical optimization and outpatient clinics are briefly described. To start disassembling the uses of optimization in outpatient clinics, section 2.3 introduces different stages of service production planning. Next, sections 2.4-2.6 drill deeper to optimization problems at different stages of production planning. Finally, section 2.7 describes the use of optimization as a tool to be used during building designing instead of day-to-day operations.

## 2.1 Mathematical Optimization Problem

Mathematical optimization or programming is a field of operations research where the focus of research is in mathematical optimization models together with numerical algorithms to solve them. A mathematical program or optimization problem consists of three major parts, including decision variables, objective function and constraints. An example of a mathematical program is given in the following:

$$\begin{aligned} &\max_{\mathbf{x}} \quad f(\mathbf{x}) \\ &\text{st.} \\ &\mathbf{Ax} \leq \mathbf{b} \\ &x_i \in \{0,1\},\ \forall\, i\ . \end{aligned} \tag{2.1}$$

*Decision variables* represent the inputs of the system being optimized. In Eq. (2.1), $x_i$ are the decision variables. These inputs can be varied during the optimization phase and the goal of the optimization is to find a combination of decision variable values which yields an optimal objective function value. The combination of decision variable values is called a solution.

*Objective function* is a quantity of the system calculated from the decision variables. In Eq. (2.1) $f(\mathbf{x})$ represents the objective function which is being maximised. The value of the objective function measures how good the current combination of decision variable values is. The idea is to maximize or minimize the objective function value. Thus, setting the objective function is one of the most important things while creating the mathematical optimization problem. The basic question is: What should be optimized?

*Constraints* describe the limits of the system. In Eq. (2.1), $\mathbf{Ax} \leq \mathbf{b}$ gives the linear constraints in a matrix form. For example, a resource in scheduling problem cannot work more than one case at a time. Constraints give rise to feasible and infeasible solutions. If no feasible solution can be found, the problem is infeasible. Given that the constraints allow feasible solutions, these solutions will have to be compared using the objective function to find the optimal solution. There is usually a single objective function but the number of constraints is not limited.

Generally, the mathematical optimization techniques work on problems that can be represented using equations and the system can be quantified. The problems often include logical deduction where certain decision have different outcomes such as staff rostering, vehicle routing and scheduling.

## 2.2 Description of Outpatient Clinics

*Outpatient* (suom. avohoitopotilas) is a patient who is not admitted to stay overnight in the hospital. The other patient type is *inpatient* who is admitted to a hospital bed for an overnight stay and requires longer period of treatment. Outpatient clinics are described by stating that the clinics treat outpatients. This description is wide and there are many different kinds of outpatient clinics.

First, there are the clinics providing primary healthcare. These consist of the local healthcare centres providing municipalities the primary care required. Another type of basic care comes in the form of dentist clinics. These types of facilities provide the primary care and are often sufficient in treating the patient. The basic care could be seen as the day-to-day operation visible for most of the people.

Second, the patient might require extensive care in specialized healthcare. The specialized healthcare has also outpatient clinics specializing in a certain specialty such as neurology. The specialized healthcare outpatient clinics provide treatment by appointment. These facilities often reside in hospitals, creating a centre of specialized healthcare for treating the more severe conditions among with different inpatient wards, operating theatres and the emergency room.

Third, there are outpatient procedure centres which can give surgical treatment to patients but where the preoperative and postoperative phases are short enough for the patients to be released during the same day as admitted to the clinic. These centres can also be seen as outpatient clinics, however their operation differs from the more traditional appointment clinics by the complexity of each appointment. A procedure requires thorough planning and preparation, and the patient might spend the whole day at the center.

Additionally, the social services and other public care instances offer similar services for customers. The customers are seen on appointment and the operation is similar to that in healthcare. The trend is that these services are given at the same centres as primary healthcare and thus relevant to mention in this study. We should also note that within healthcare, there are rehabilitative services. The acute medical care is one aspect which is complemented by a mix of outpatient type services.

Apart from these general descriptions of different types of outpatient clinics, there might be clinics providing care in a mixed manner combining for example outpatient appointments and inpatient care. This study focuses on optimization of the production of outpatient appointment clinics without considering these advanced mixed care centres.

## 2.3 Stages of Outpatient Clinic Service Production Planning and Control

To understand the optimization problems in outpatient clinics, we first introduce the basics of hospital planning and control which generate suitable problems for the use of optimization tools. The frameworks introduced in this section describe a wide variety of decisions, some of which will not be studied further.

The hospital as a production facility is often characterized by a demand larger than supply [18, 40]. As a result, hospital production planning and control to focus on maximizing the utilization of the valuable resources [40]. This is already an optimizational problem where the hospital is maximizing the utilization of resources while keeping the operation feasible. The feasibility can be seen from multiple points-of-view. First, there are the hospital resources to consider. The staff should not be overburdened, while the usage of the building also has natural constraints. Second, the patient satisfaction and wait times should be kept under control meaning the service level must be satisfactory. Optimizing hospital production as a whole from this basis is not practical and should thus be divided to smaller sub-problems. This is done by hierarchical frameworks introduced next.

In Fig. 2.1, a healthcare production control framework is given, which is based on research articles by Vissers et al. [40] and De Vries et al. [13] and the visualization is given in [18, p. 157]. The planning process described by the framework starts from strategic planning with two to five years of planning horizon. The initial plan is then focused during the following four stages. The planning horizon also shortens from years to months, and finally to weeks and days. The different stages of this hierarchical process contain problems in operations research fields of optimization, forecasting and decision making.

The first stage on the framework given in Fig. 2.1 contains the strategical decision making done by the highest level of management of the hospital. The decision depict the direction of the hospital over the time horizon of following years and introduces the services provided and restrictions on resources [40]. This stage does not include production planning or control. Rather, it gives the boundaries for the healthcare facility operation. The second stage continues this strategical work and matches the patient volume roughly to resources. The patient volumes might have to be contractually agreed on with external agents, leaving the hospital responsible for realizing the service volume promised [40]. Decisions on resources are made on rough capacity and utilization rates.

The third stage in the framework in Fig. 2.1 is called resources planning. It matches resources and patient volumes at specialty level and uses aggregates on both demand and supply. The leading resources mentioned in the figure refer to those resources that generate further work within the hospital. For example, an operating room creates work downstream at the recovery ward, which is why the leading resources are allocated to accurate level of days or hours already at this stage [40].

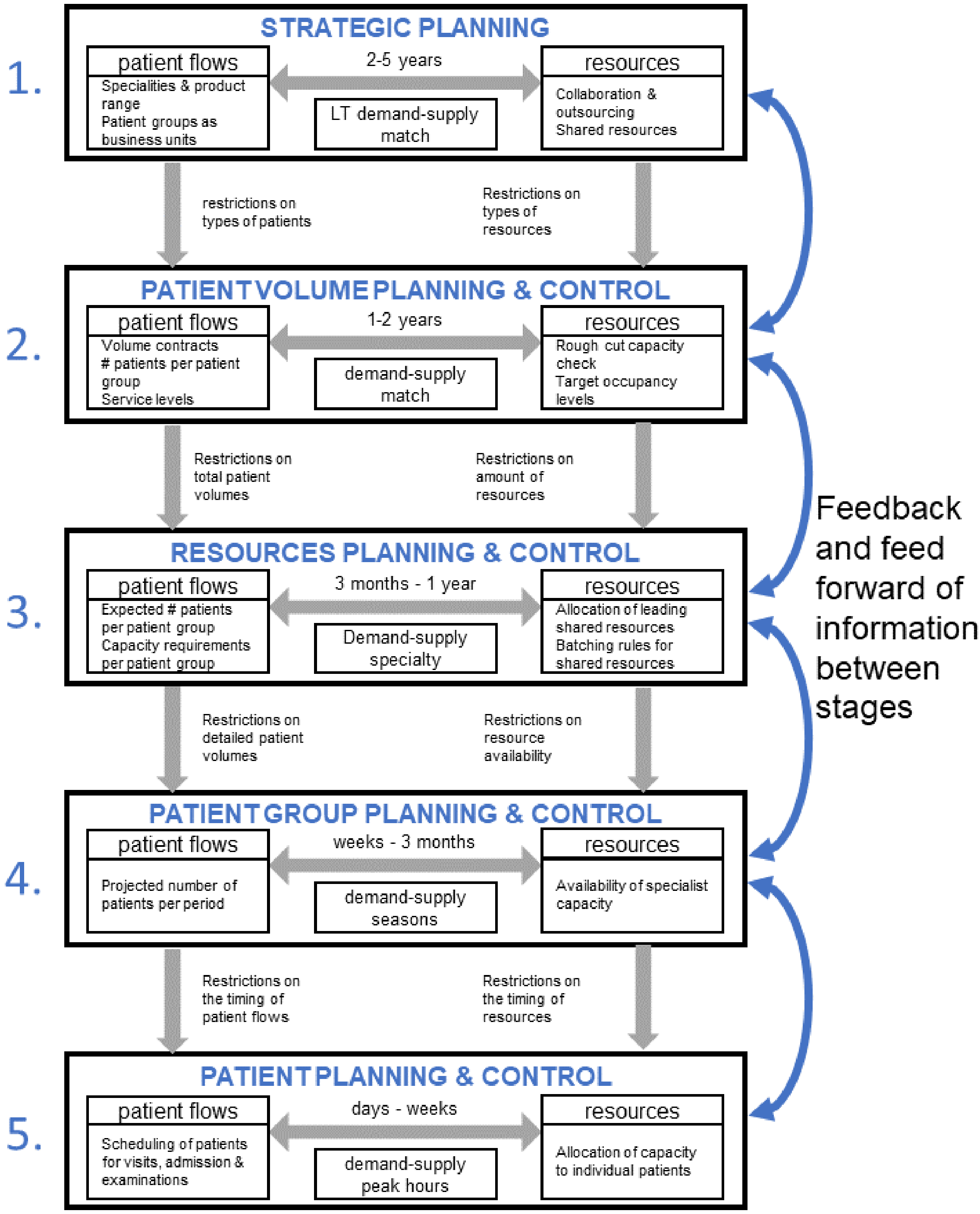



Figure 2.1: A framework for production control from [18, p. 158].

The fourth stage plans patients at a group level. At this level the patient groups are planned at an aggregate level to specific times to enable the allocation of resources at an individual level. This planning, happening at an interval of weeks to months, also makes sure that the resources are matched to seasonal changes in patient volumes. Finally, the fifth stage schedules single appointments to individual resources within the given restrictions from the planning done on the preceding four stages. The fifth stage also contains reactive elements to surprising situations.

The framework by Vissers et al. [40] shown in Fig. 2.1 depicts an important feature of the control and planning. Even though the planning is done hierarchically, there are feedback and feed forward controls to ensure smooth operation. The feedback gives precious information on the actual realized capacity used to the higher levels which enables more realistic matching of supply and demand. Feed forward steers the operation from the higher levels by giving guidelines for service level (patient waiting times) and capacity available. [40]

Another, hierarchical roadmap for planning hospital operations is given by Zonderland [46]. This framework divides the planning process to strategical, tactical and operational levels. It also follows a similar path of first creating a crude plan and then focusing it towards the day-to-day operation in steps [46]. Zonderland [46] briefly mentions also the higher level of patient division among hospitals and hospital placement as a topic of discussion. This higher strategic level is not included in the framework by Vissers et al. [40].

Further classification of healthcare planning decisions is given by Hulshof et al. [19] in a literature review. The taxonomy used divides the planning decisions to strategic, tactical and operational as in [2, 46]. Further the operational level is divided to offline and online levels. Hulshof et al. [19] give broad overview of the vast literature on ambulatory care service (includes outpatient clinics) and note that the literature mainly focuses on offline operational planning decisions on appointment scheduling. The taxonomy used by Hulshof et al. [19] is visualized in Fig. 2.2 which contains the planning decisions found on literature by their review. The division of planning to strategical, tactical and operational is also introduced in [18], where this area of interest is called resource capacity planning.

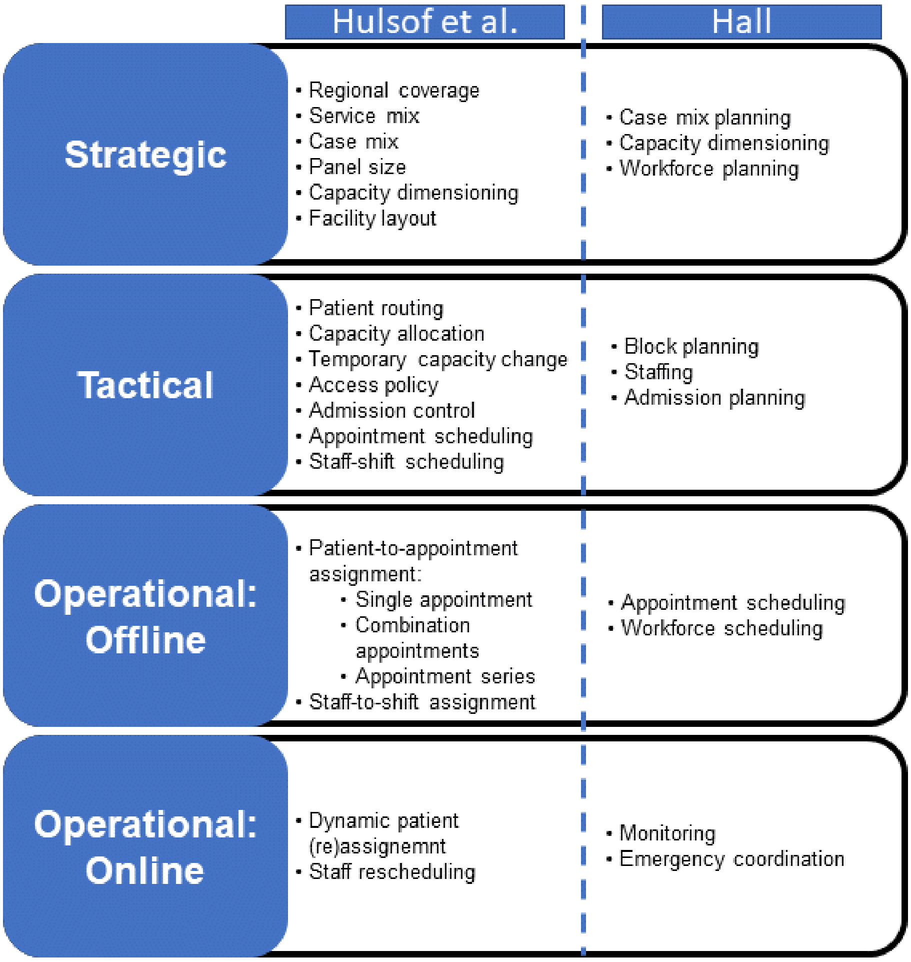



Figure 2.2: A division of healthcare planning decisions in outpatient clinics as presented by Hulshof et al. [19] and Hall [18].

Regardless of the planning framework in use, the leading thought is that first in the planning process the strategical direction of the healthcare unit is set. After that, the planning is done hierarchically starting at non-precise, aggregate level, long term plans followed by focusing these plans gradually towards a day-to-day level, where single appointments and resources are planned on a short horizon. Each of these levels has problems in the field of operations research including but not limited to optimization, decision making and forecasting. Next, selected optimization problems from the tactical and operational levels of planning are introduced more in detail.

## 2.4 Resource Planning: Shift Scheduling

A general part of resource planning is the scheduling of employees to match the required workforce indicated by forecasted demand. Early general employee scheduling problem is studied by Gloyer and McMillan [16]. Their problem includes matching demand with appropriate number of staff while considering shift regulations and personal availabilities. To match the demand, shifts are created in 15 minute time intervals. Gloyer and McMillan [16] mention nurse scheduling as a real world example of an employee scheduling problem.

The Nurse Scheduling Problem (NSP) assigns nurses to work shifts to match the demand of a healthcare operation. The goal is to ensure that the operation is viable in terms of staff while not over allocating resources. The demand is usually known as the amount of required staff over the planning horizon. Even though the name of the problem states otherwise, this employee rostering problem can obviously be used to consider other professional groups within healthcare as well. Thus, we will call it the staff scheduling problem instead. In the framework shown in Fig. 2.2, the staff scheduling problem falls under tactical and operational offline planning stages.

Building a staff scheduling problem is more or less case specific. Different organizations, countries and outpatient clinics have their own special properties and the differences lead to distinct considerations on the problem. Hall [18] gives the following questions as guidelines to shape the creation of a nurse scheduling tool.

- What input parameters should be known?
- What are the goals and outcomes of the staff scheduling problem?
- What limitations and constraints should be covered?
- What are the proposed solution methods for the problem?

Hulshof et al. [19] divide the problem to tactical and operational level. On the tactical level, the shifts are scheduled and the number of staff on each shift is matched to demand. On the operational level, the previously scheduled shifts are handed out to individual staff members. These two levels are closely related and the individual level might affect the shift scheduling on tactical level through personal preferences.

Construction of the shifts varies between studies. For example, Wong et al. [41] assign staff to certain pre-made shift patterns i.e. work a day shift on day

1. Brunner et al. [4] create the shift patterns as a part of the optimization instead of resorting to pre-made patterns. Smet et al. [36] provide a general mathematical formulation of the staff rostering problem where they mention locked shifts for certain days and resources on top of these two types of shift construction.

The *objective* of the staff scheduling problem is different in cases. Brunner et al. [4] state that most of the research papers focus on a single objective instead of multiple objective optimization. The following list mentions some objectives found in literature.

- Cost minimization ie. assigning staff to different shifts have different effects on salaries and benefits [18].
- Minimizing substitute staff and outside capacity [35].
- Staff preferences and satisfaction in shifts assigned. Wong et al. [41] collect preferences from nurses where avoiding three consecutive evening shifts gets largest weight and Smet et al. [36] mention rest time between shifts.
- Satisfaction of demand, coverage. Multiple studies mention coverage as only a soft constraint meaning that some time intervals might be understaffed [4, 36].

In addition to objectives, optimization problems include *constraints*. The constraints can be divided to *soft* and *hard constraints*. In the case of staff scheduling problem, hard constraints include labor laws, staff availability, skill types matching demand, minimum service level and locked individual shifts [4, 36, 41]. Soft constraints are often linked to the objectives of the scheduling problem. Smet et al. [36] give the objective function as a linear combination of soft constraint violations.

Burke et al. [5] state that most of the research, up to their paper, using mathematical approach to find an exact solution to the case specific staff scheduling problem resort to heuristic methods due to the complexity of the problem. The following two articles use different methods to solve their staff scheduling problem.

Brunner et al. [4] consider a flexible physician shift scheduling problem where a known demand along the planning horizon should be satisfied by assigning physicians to shifts. In this case, the flexible shift means that they are allowed to create shift patterns according to a set of rules and later the physicians are assigned to these shifts. They formulate the problem as an integer-linear-programming problem and solve this problem using column (shift) generation and branch-and-bound algorithm which in conjunction form the branch-and-price algorithm. They are able to solve instances of 18 physicians in two categories and two to six week planning horizon within reasonable run times of seconds to minutes. [4]

Wong et al. [41] create a two stage heuristic for solving the studied staff scheduling problem. They first search for a hard constraint satisfying staff schedule and then refine this by a local search algorithm. Their heuristic can solve instances of around 55 nurses for a time period of one week in run-times of 0.17 to 0.22 hours.

Concluding, the staff scheduling problem seems to be tractable using mathematical optimization.

## 2.5 Outpatient Clinic Appointment Scheduling

In the area of outpatient clinics, much of the research is focused in appointment scheduling [19]. The research questions include division of staff shifts to blocks, length of blocks, number of patients within block, sequencing of patients within block, length of patient appointment and patient overbooking [19]. These questions deal with optimizing the schedule shown in Fig. 2.3. The figure shows a single resource. For this resource, three appointment blocks have been added to a single shift including multiple patients. Cayirli and Veral [8] provide a review on outpatient appointment scheduling methods and the use of optimization methods in outpatient appointment systems is studied by Ahmadi-Javid et al. [2].

The final product of appointment scheduling is a blueprint for handing out the individual appointments for patients. This blueprint has to be made since the demand is stochastic in terms of number of realized patient appointment reservations, missed appointments and appointment lengths. The patient demand realizes one at a time and the single appointments are handed out from the blueprint in an online manner.

The rest of this section introduces different aspects considered while setting up the outpatient clinic ultimately leading to an appointment schedule.

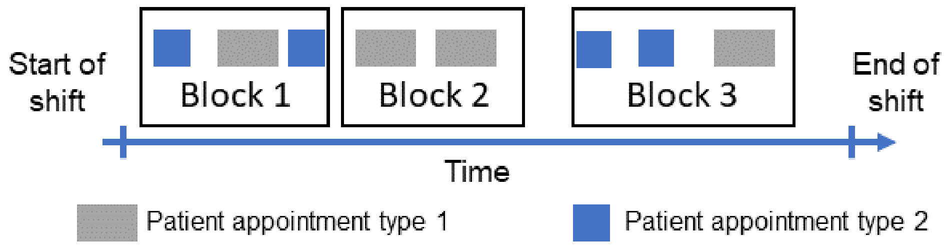



Figure 2.3: Example of appointment schedule including patient blocks and patient appointments.

### 2.5.1 Access Policy

Outpatient clinics come in many different *access policies*. Historically, clinics employ the traditional by appointment access policy where patients book appointments well in advance. Open-access scheduling clinics take same-day appointments on call [22]. Walk-in services schedule no appointments, as they serve patients walking in to the clinic for example the emergency department of an hospital. Hybrid models, incorporating elements from different access policies, are called advanced access policies and for example can leave a fraction of schedulable appointment times free for patients requesting an appointment within the next 48 hours [19]. Robinson and Chen [34] analyze the performance between these models and conclude that the open-access scheduling often outperforms the traditional policy when the no-show rate of patients is above 5 %. Obviously, this is a case specific question and should be considered in the context of the target outpatient clinic.

The advanced access policies might include specific rules for accepting walk-in patients on top of the by appointment patients. Accepting walk-ins may reduce the negative effects of no-shows [2]. Qu et al. [32] create a Markow decision process model for determining whether to accept a walk-in patient or not. They report that when walk-ins are less than 20 % of the clinic service, all walk-ins should be accepted. Seasonality in walk-ins can also be used to increase the efficiency of the practice [2]. Obviously, the walk-in patients are not present at every clinic and such hybrid access policies including walk-ins are not applicable to all.

The access policy of the outpatient clinic is a tactical or even a strategical decision taken by the management when setting up the practice. This decision can be supported by studying the effects on the lower operational level using a scenario based approach. Mathematical optimization is not directly applicable to access policy optimization but can be used to compare the scenarios induced by the chosen access policy.

### 2.5.2 Appointment Blocks

The grouping of appointments to blocks simplifies the planning and controlling of healthcare service production. These blocks are visible in Fig. 2.3. The blocks can be used to divide the time on resources between different tasks or patient groups and each appointment block can contain multiple patients with different profiles and different times.

Forming the appointment blocks can be approached from two different perspectives. First, a clinic with already locked staff shifts, the blocks should constitute to planning the time usage of the professional. Blocks should be cut by lunch breaks and switches between allocations to different patient groups. The blocks are formed from the staff perspective.

Second option is to form the blocks from the appointment perspective. Lee et al. [22] form the appointment blocks based on heterogeneous patient demand. They set the block length according to the single patient appointments planned within the block and consider also no-shows and random service times using log-normal distributed service times. As an end result they create a prototype block and then sequentially copy this block for an optimal schedule [22].

In the same paper by Lee et al. [22], the within block sequence of patient types is studied. They take the number of patients per type as a problem parameter and schedule these optimally within blocks. Their study includes a single server and heterogeneous, categorized patients. They create a sequence of appointments for the requested number of appointment in categories and this sequencing problem is *NP-hard* dynamic programming problem for more than two patient categories [22]. The patient sequencing is also studied by Cayirli et al. [9] and they conclude that when doctor idle time is penalized heavily, the new patients with longer service times should be set first in the patient sequence.

Another use of appointment blocks can also mean that each patient to be served within block is scheduled to arrive at the start of the block. This type of scheduling obviously leads to long patient wait times compared to a more elegant scheduling [19].

The block scheduling and other aspects of appointment scheduling are strongly interconnected.

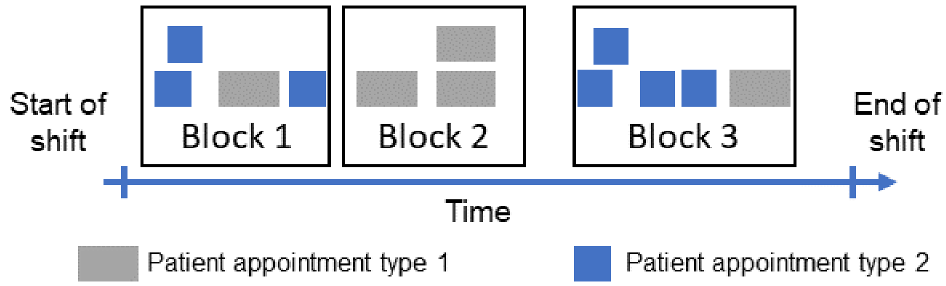



Figure 2.4: Example of appointment schedule including overbooking of patients.

### 2.5.3 Number of Patients, No-Shows and Overbooking

Performance of an outpatient clinic is often hindered by patient no-shows. The percentages of patients who miss the scheduled appointment can be over 20 % [17]. Roughly, this would mean that fifth of the capacity available is not utilized. To overcome this, overbooking models have been studied where physicians time is scheduled to multiple patients at once [22, 43]. The no-shows of patients are usually modeled using a predetermined no-show rate.

The overbooking is illustrated in Fig. 2.4. The figure includes an appointment schedule with three patients overbooked. Increasing the number of patients per consultation session balances patient waiting time and staff overtime [2]. Too many patients lead to increased patient waiting time and staff overtime. Scheduling too few patients lead to short waiting times but low resource utilization rates. Zacharias and Pinedo [43] consider overbooking by adding multiple patients to a single time slot as in Fig. 2.4.

Lee et al. [22] uses a different approach by shortening the appointment interval for each patient to allocate more patients within a block. This approach does not prepare for no-shows by assigning multiple patients to a single appointment time. Rather, the total number of patients on the appointment block is increased. In this approach, the real expected service time is then longer than the selected appointment length.

### 2.5.4 Appointment Interval, Service Time and Patient Punctuality

The length of a patient appointment or service time is stochastic. The length might vary due to patient condition or staff performance and seems to follow a log-normal distribution [9]. The appointment interval is the predetermined or planned length of an appointment. The interval does not have to correspond with the actual appointment length but giving too short appointment intervals leads to patient waiting time, while too long intervals lead to staff idle time.

In case of homogeneous patients within the appointment block and small no-show rate, a dome shaped pattern for appointment intervals has been found to be effective [14]. This means that during the start of the period, patients are planned to arrive on shorter intervals. Interval then increases towards the middle and again shortens towards the end. Erdogan and Denton [14] note that increasing no-show rate decreases the appointment intervals and finally double-booking becomes optimal.

Another factor affecting the appointment interval is the patient punctuality. Klassen and Yoogalingam [21] study the problem and report that the patient unpunctuality is a major factor in creating an efficient appointment schedule. Performance of the clinic can be increased by considering early arrivals as a tool for decreasing waiting room congestion while minimizing the impact of late arrivals. It is reported that clinics might want to shorten the appointment intervals when the standard deviation of patient unpunctuality increases and to increase the appointment interval when the mean of patients arriving early increases [21]. Zhu et al. [45] report that, if the patient unpunctuality neglected, the total cost measured in patient waiting time and doctor idle time can be as much as 30 % higher compared to considering the unpunctuality.

Review by Ahmadi-Javid et al. [2] reveals that most of the studies in the outpatient appointment scheduling assume the appointment lengths to be equal to the mean service time of the patient group. This means that the appointment intervals are uniform over the appointment block and a potential for increasing the efficiency of the appointment schedule is disregarded.

### 2.5.5 Multi-stage Appointments and Multi-Server Systems

The previous sections have mostly dealt with single appointments on which the literature is focused [2]. However, there are other appointment types including combination appointments and appointment series [19].

An example of this is an appointment at a surgical clinic. First, the patient often has a consultation session with a doctor and during this appointment a decision on surgery is made. If a surgery is required, the patient has a pre-operative discussion with a nurse during the same day. If a surgery is not required, the patient might have a consultation session with a physiotherapist to provide remedy. Thus, the patient goes through consecutive appointments called a chain or a combination appointment.

The appointment series consider cases where the patient receives treatment on multiple occasions over a period of weeks to months. These series can be incorporated to enhance the forecasted demand in the future and thus improving the planning and control process. However, when creating the appointment schedule, they can be considered as single appointments.

The chain appointments naturally create a multi-server problem. This topic is studied more extensively in the form of operating theatre scheduling which is discussed in Sec. 2.6. Also, the problem studied in Chap. 3 is multi-server and possibly multi-stage.

### 2.5.6 Optimization of Appointment Scheduling

Regardless of the methodology used or the targeted decision, the objectives behind optimizing the outpatient appointment scheduling often consist of minimizing one or many of the following criteria.

- Cost of schedule [10, 22]
  - Staff overtime [14]
  - Staff idle time [45]
- Revenue and patient comfort
  - Number of patients seen [10]
  - Patient waiting time [14, 45]

- Staff or patient preferences [42]
- Balancing of resources [23]

The objective function is usually a weighted sum of the objective criteria [2]. The main trade-off in appointment scheduling optimization is between staff idle time and patient waiting time. Since the desired trade-off is an input of the problem rather than a result, efficient frontiers can be considered [9]. This means that rather than providing a single best appointment scheduling solution, a whole frontier is provided on which none of the solutions can be improved without changing the trade-off between staff idle time and patient waiting time. This is also mentioned as a future research direction by Ahmadi-Javid et al. [2].

Considering staff preferences is important in keeping the workforce content and to discourage changing jobs. Ahmadi-Javid et al. [2] note that incorporating patient preferences and these affecting the no-show probabilities positively is a future research direction. As explained before, the no-show is a large factor in outpatient clinic performance.

The above list of objective criteria mainly targets single resources. This is due to appointment scheduling literature focusing on single-server and single appointment cases. For example, balancing criteria across multiple doctors could be taken into account when considering multi-server and multi-stage outpatient clinics.

Besides the objectives, there are the constraints limiting the appointment schedule. These often overlap with the objectives and the clear constraints, such as not exceeding number of resources, are not given.

- Staff working hours and overtime [22]
- Staff idle time [22]
- Patient waiting time [42]
- Staff or patient preferences [42]
- Demand to be met in number of patients [23]

There are clear, physical constraints to consider. Staff working hours are limited and certain demand should be met. Service levels, considered through patient waiting and access time, can be hard constrained to a certain promised level. Apart from these, the constraints of an optimization model can be used to evaluate staff or patient preferences. The constraints give the limits within which the appointment schedule should operate and the objectives guide the schedule towards a wanted outcome.

The methodologies used to reach these goals are numerous. Studies include queuing theory, computer simulation, heuristics, Markov processes and mathematical programming [19]. Optimization studies mostly consider the problem using stochastic programming [2].

Review by Cayirli and Veral [8] states that the optimization problems described by researchers often consider only a single provider instead of multiple server model. They argue that it is common practice to have a doctor specific waiting list, which on the other hand encourages a psychologically advantageous patient-doctor relationship. This means that the joint use of multiple resources to balance load is disregarded. Another consideration is that the resources are not in the focus of appointment scheduling since they are already optimized in the shift scheduling. Rather, the mindset seems to be that since the resources are already available for a specific time, they should be used optimally. The patient is often in focus and the avoidance or mitigation of no-shows is present in almost every paper.

### 2.5.7 Patient to Appointment Scheduling

After creating the (optimal) appointment schedule, the appointments have to be handed out to single patients. This is called the appointment booking problem. At this point, the demand starts to realize one appointment at a time. Scheduling these single patients is straightforward in the case of an outpatient clinic implementing a single server model. A "First come, first possible appointment" is approximately good when patients are homogenous in waiting cost and service time distribution [2]. Another policy is to order the patients by the assumed variance of their service time Mak et al. [24]. The variances should be taken in an increasing order, scheduling patients with the largest uncertainty as the last jobs within the day.

Lee et al. [22] consider filling up certain optimized patient blocks first before opening a new block when considering heterogeneous patient service times. This approach is close to bin-packing problem algorithms.

Besides handing out the planned appointment times, some studies consider building the schedule as an online problem. The appointment blueprint is built on the fly as the demand is realizing and branch of appointment scheduling describes the decisions when adding single patients to a schedule at a time. Pérez et al. [30] create a scenario based algorithm for stochastic integer program in online scheduling problem. Their algorithm optimally schedules a multi-step procedure with multiple resource requirements to an existing schedule with unknown future demand. The future is taken into account by realizing a set of future appointments (scenario) multiple times and for each of these solving an optimal scheduling of the appointments. Basically, the algorithm forecasts the future possibilities for optimal scheduling given the scheduling of the request at hand. Pérez et al. [30] report that their approach to schedule a single appointment takes multiple minutes while the benchmark algorithms take only seconds. The benchmarking algorithms are described in [29]. A bit similarly, Chakraborty et al. [10] create an on-line myopic scheduling algorithm and note that their objective is uni-modular which they use to stop when optimal number of patients for a session is reached.

Patient to appointment scheduling should not be confused to appointment scheduling of hybrid access policy clinics. These clinics should take into consideration both well in-advance calling patients and same day calling patients. For example, Chen and Robinson [12] integrate scheduling routine, pre-booked patients and scheduling call in, same day patients. Their approach is pre-calculated and does not include on-line aspects.

## 2.6 Operating Theatre Scheduling

Even though this study is mainly about outpatient clinics, operating theatre scheduling is described here. The finer precision level used in operating theatre scheduling is close to the problem studied in Chap. 3 making it relevant to the this study. An operating theatre performs surgical operations to patients. The complexity of running an operating theatre is on another level compared to a basic outpatient clinic. The operations performed often take hours and the patients require pre-operative and post-operative care. Thus, operating theatres span to a multi-stage processes, increasing the complexity and encouraging more precise planning and scheduling to seamlessly coordinate all stages. Cardoen et al. [6] provide a literature review of operating theatre planning and scheduling including both block scheduling introduced in section 2.6.1 and single surgical case scheduling introduced in section 2.6.2.

### 2.6.1 Block Scheduling

In operating theatre planning, the blueprint consisting of specialty specific time blocks for resources is called a *master surgical schedule*. This blueprint is not directly comparable to the blueprint described for outpatient clinics in Sec. 2.5. The outpatient clinic blueprint results in single, prefixed times for the patients. The master surgical schedule describes time intervals within which operations can be scheduled. Surgical cases can then be mixed to these time intervals according to their expected lengths. The master surgical scheduling process is carried out on the tactical level of hospital operation planning and control.

The main goal of creating a master surgical schedule is to divide operating room capacity between different specialties [15]. After division, all of the resources can be locked and the operations of single patients can be scheduled. The master surgical schedule also enables the check of conflicting resources and levelling the load on wards or ICU ie. the downstream on the multi-stage process [15, 18].

The demand in master surgical schedule forming is taken from the waiting lists of different specialties [1, 18]. Compared to the outpatient clinic appointment scheduling process, this approach uses known demand instead of forecasted demand. This means that either the master surgical schedule must be created close to the date of surgery or the waiting time of the patient must be long.

The goals of optimizing the master surgical schedule include

- Demand and capacity matching across specialties [44]
- Resource utilization [1]
- Resource balancing across the multi-stage process [15]

The objectives for master surgical scheduling include demand and capacity matching, resource utilization and balancing across the multi-stage process. These are on a rather generic level as the whole problem creates a preliminary schedule on aggregated demand.

The constraints in master surgical schedule creation include maximum capacity constraints, maximum parallel allocations across specialties and allocating all operating rooms a specialty each day.

Fügener et al. [15] concentrate on the downstream units when creating the master surgical schedule. They use an assignment optimization problem assigning specialties to blocks throughout days and provide exact and heuristic solution algorithms. The concentration on downstream units is important since the operating theatre cannot function if there is insufficient post-operative care available.

Zhang et al. [44] formulate the operation room block scheduling problem as a mixed integer problem. They create surgical master schedule which divides operating room capacity among different specialties as whole days. The problem is based on deterministic demand on each specialty for a single day. This demand can be postponed if necessary.

Since the demand during the creation of master surgical schedule is already based on known waiting lists, this stage of planning is closely related to the following patient scheduling stage. These two stages are integrated to a single problem for example in [1].

### 2.6.2 Patient Scheduling

The finer level of operating theatre scheduling takes place on top of the master surgical schedule. The patients from the specialty specific waiting lists are assigned to operation dates and the patient sequence within a day is resolved. The patient scheduling of operating theatre has been widely studied at the single surgical case level as is shown by the literature taxonomic classification by Hulshof et al. [19]. The operating theatre implies a major part of hospital costs but also large revenue potentials as stated in [18, 19].

The scheduling of operating theatre surgical cases differs from the outpatient clinic only by the complexity of operations. While a typical outpatient appointment might require a single professional and an appointment room fitted with chairs, table and a computer, the most complex surgical cases include multiple surgeons, nurses and costly equipment. The surgical operations also often have consequences on other parts of the hospital including but not limited to wards and the intensive care unit (ICU). A comprehensive consideration of all of the functions within the hospital should be taken in scheduling [31]. The process should be continuous to avoid patient queuing inside the operating rooms in the worst case. The length of surgeries might far exceed those of regular outpatient clinic appointments. The surgical cases also almost always include pre- and post-operative procedures.

Mathematical optimization is often employed at the operational offline scheduling level of surgical cases. The goals of optimization include utilization rate of surgical resources, staff preference, patient waiting time and staff overtime

[19]. Meskens et al. [25] provide an extensive formulation of the operation theatre schedule optimization, considering surgeons, nurses, rooms, anaesthetists, material, surgical cases and block schedules. On top of these, they have added an affinity scheme which takes into account preferences of the staff on their colleagues.

Vijayakumar et al. [39] formulate a similar day program optimization model. They argue that the program is analogous to dual-bin-packing problem and they solve the problem using a sequential heuristic algorithm based on first fit decreasing classical bin packing algorithm. Berg and Denton [3] also study the bin-packing problem as a scheduling problem in an outpatient procedure center which lies between the operating theatre and the outpatient clinic.

The surgical case scheduling can be done offline for the elective patients. In addition to this, there are the online scheduling for emergency cases. Pham and Klinkert [31] formulate the surgical case scheduling as a generalized job shop problem and minimize the makespan of completing all of the surgical cases. They include a model that can be used hierarchically to first schedule elective patients well in-advance and then second time to add new, urgent cases to previously made schedules the day before the surgery date. Pham and Klinkert [31] conclude that large instances of their problem are not tractable using general purpose mixed integer linear program solvers and heuristics should be studied.

## 2.7 Schedule Optimization as a Tool for Healthcare Building Design and Engineering

Fig. 2.2 brings up capacity dimensioning in the strategical level. Capacity decisions during this stage includes consultation rooms, staff, consultation time, equipment and waiting room capacity in the ambulatory care services. It is mentioned to be a key decision in influencing the ability of meeting demand while managing access and waiting times [19]. With regards to healthcare building designing and engineering, the number of consultation rooms, equipment and waiting room capacity is of main interest. Building and sourcing too many resources lead to high costs of building and low efficiency of operating due to low utilization rates. Too scarce resources lead to difficulties in meeting required service levels and can increase dissatisfaction of staff due to shortage of rooms.

While finding a required number of room capacity, the processes and operating models of the new healthcare facility have to be planned. This planning often takes similar steps as described in Sec. 2.3 and involves working out the guidelines for the operation planning and controlling. The key decisions should be analysed and tested prior to locking for example the room number. The design question from the operational side can be summed up to "Given this operating model, how much capacity do we need from the new facilities?"

To answer the question of capacity dimensioning, operations research methods are often employed. For example, Hulshof et al. [20] use analytical models to compare two different policies in outpatient department and room requirements considering doctor-to-patient and patient-to-doctor travel policies. They take this comparison further using simulation to evaluate the number of rooms required on both policies. Simulation is used to model situations too complex to analyze using direct calculation methods. Paju et al. [28] provide a framework for using simulation in various stages of healthcare facility planning. Their simulation framework contains a stage where the number of rooms is researched for single units. This stage can benefit from the use of optimization in creating schedules based on the future operating model and forecasted demand. Such optimization problem is described and formulated in the following Chap. 3.

Swisher et al. [38] depict a model for simulating a single clinic operation including staff, rooms and patients visually using discrete event simulation (DES). Their work continues in [37]. Creating such a simulation obviously requires the number of patient demand over the time horizon. This can be done using either by distributions of arriving patients [38] or by a pre-made schedule. Creation of the schedule can be done using scheduling algorithms that often include optimizational aspects. A clear example of this is by Carpenter et al. [7] where they study impacts of forecasted future demand on magnetic resonance imaging using optimization to create schedules while running different scenarios. Another example is by Liang et al. [23] who create a balanced schedule through optimization and test this using a simulation in an oncology outpatient clinic offering chemotherapy treatments. Thus, we conclude that an algorithm should mimic the expected scheduling process of the clinic when calculating the required room number. Such a scheduling algorithm is described in Chap. 4.

# Chapter 3

# Mathematical Formulation

In this chapter, a model problem will be formulated as a mathematical optimization problem, more specifically as an integer linear program. The chapter starts with describing the problem and answering to questions about what is included at what precision level and what are the desired outputs. After specifying the problem verbally, it will be converted to a mathematical form. The mathematical form is described thoroughly explaining decision variables, constraints and giving options on the objective function. Finally, the complexity of the problem is briefly analysed.

## 3.1 Problem Description

Given the literature review on Chap. 2, we can now understand the current requirements for outpatient clinic scheduling by Delfoi. The problems to be solved include multi-server clinics where each appointment might require multiple resources and might contain multiple steps. Compared to the literature studied in Chap. 2 where problems are often considered with single-server single resource problems, the problems are more complex. The multi-server model is close to the online problem studied and solved by Pérez et al. [30] but in this case the whole set of appointments is usually known in advance, making it an offline problem.

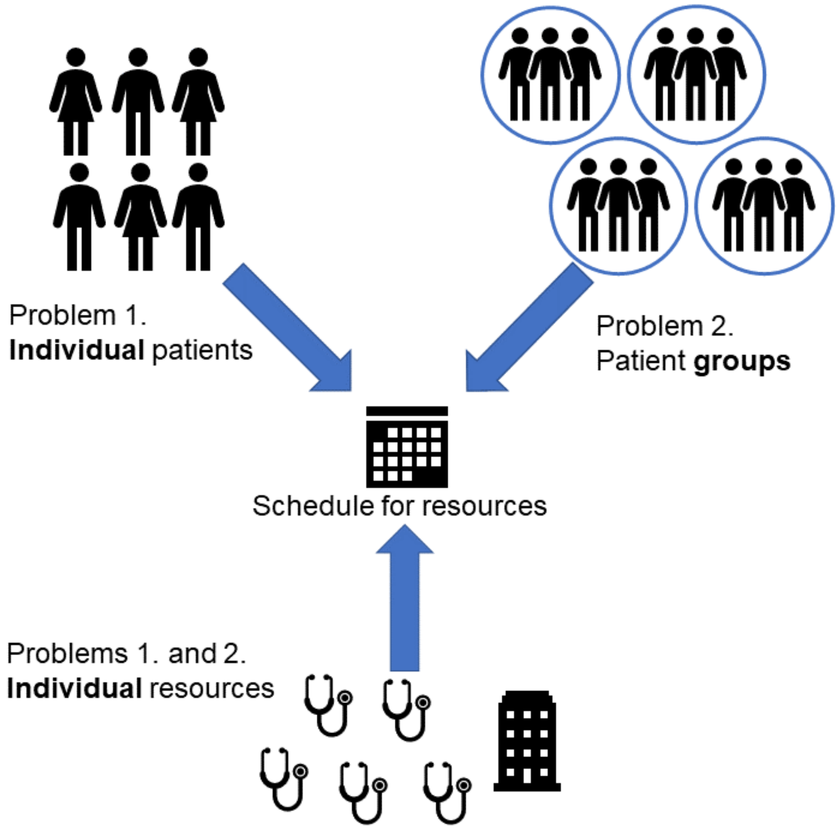



Figure 3.1: Two scheduling problems to solve.

The requirements can summed into two different problems:

Problem 1. Create a week schedule of single appointments for individual resources simulating a future operating model.

Problem 2. Create a week schedule of aggregate demand (aggregation period might vary) based on a forecast for individual resources for continuous operations planning.

These problems are also shown in Fig. 3.1. In both scheduling problems, the resources are on individual level which corresponds to both operational and tactical level in Sec. 2.3. Problem 1. corresponds to the finest level of patient planning described in Sec. 2.5.7 since single appointments are being scheduled. Problem 2. has the patient on aggregate level, namely as patient groups or blocks and corresponds to the block scheduling in Sec. 2.5.2.

The problems differ on the demand side. The aggregate demand of patient group block scheduling have considerably longer scheduled action when compared to a single appointment. There are also the stochastic considerations in outpatient clinic appointment scheduling in terms of appointments exceeding their scheduled time slot. Stochasticity is not in the scope of this study, which enables us to model the problems using only a single formulation. The key idea is that an appointment block is just a long single appointment and they are scheduled on the resources as single appointments would be.

In the following formulation, the terms *appointments* and *resources*, will be left on an abstract level. In the context of healthcare, the appointments are usually understood as patient appointments or surgical operations but in this case they can also be blocks or patient groups. There are also other than patient related work done in outpatient facilities such as back-office administrative tasks or professional meetings. The problems to be solved might contain some of this work and they should be seen as schedulable actions under the general term of appointments.

Next, the two problems are discussed in more detail.

### 3.1.1 Problem 1: Single Appointment Scheduling to Simulate Future Operating Model

The goal of problem 1 is to create a one week schedule of appointments optimally. Each appointment should be scheduled for the given resources and the lengths are known in advance. The resources are also predetermined. The main objectives for scheduling are meeting patient demand, not exceeding staff working time limits, creating a balanced day and using minimal resources.

The modelled systems include multi-server, multi-resource and multi-stage patient appointments. This high level of precision is similar to operating theatre scheduling discussed in Sec. 2.6. In the context of outpatient clinics, the largest benefit of this kind of modelling and optimization is the opportunity to simulate a future operating model. The goal is to find an optimal number of consultation rooms or an answer to the question of whether the future clinic will fit within the allocated resources. The answer is received already at finding a feasible solution to an optimization problem.

The future operating models to be compared may include different patient processes, balancing of the load along the day and lengthening the clinic opening hours towards the evening. Traditionally, the staff works in their own, named consultation rooms. Lately, the model of working has started to shift towards dynamic changing of rooms. The consultation rooms are reserved based on actual patient demand and the other tasks of specialists might be done in smaller rooms fit for making calls or dictation. All of these process changes should be tested before opening the service.

The creation of single appointment schedule for future operating model is close to simulation. The problem is often solved multiple times to compare different scenarios of the parameter settings and the operating model. After creating the schedule, the scenarios can be compared for example by the number of resources required, resource utilization rates or the resource utilization rate as a function of time depicting the day profile of the clinic.

### 3.1.2 Problem 2: Block Scheduling for Continuous Operation Planning

Considering the second problem of patient group scheduling, we have to explain the background more thoroughly. In Fig. 3.2, the planning framework of [26] is simplified to meet this study. The whole planning starts with expected and targeted patient demand. This is done on patient group level for example, there can be a total of 1000 hours of acute dental appointments per year and from this only 80 % is met due to resource constraints, coming down to 800 hours per year. This lump sum should then work as the basis of staff resourcing and shift scheduling. Optimization of this has been discussed in Sec. 2.4.

The division of demand to blocks in Fig. 3.2 is another problem to solve. This step creates blocks of variable size to fill up the targeted demand. The targeted demand should be divided to small blocks constituting of a single morning or afternoon period. The exact length of the blocks should correlate with the planned number of patients and expected length of these patient appointments. Solving this corresponds with the Sec. 2.5 resulting in the blueprint for appointment times to be handed out to individual patients.

The actual problem 2 considered in this study is the division of these preformed blocks to the resources available in the step of Block Scheduling of the planning process of Fig. 3.2. The problem is to find the optimal division of work amongst resources already fixed and considering aspects of balancing, resource preferences and attaching spaces for certain appointment blocks. This scheduling of blocks can be done for a period as long as is required noting that the longer the planning period, the more there might be staff

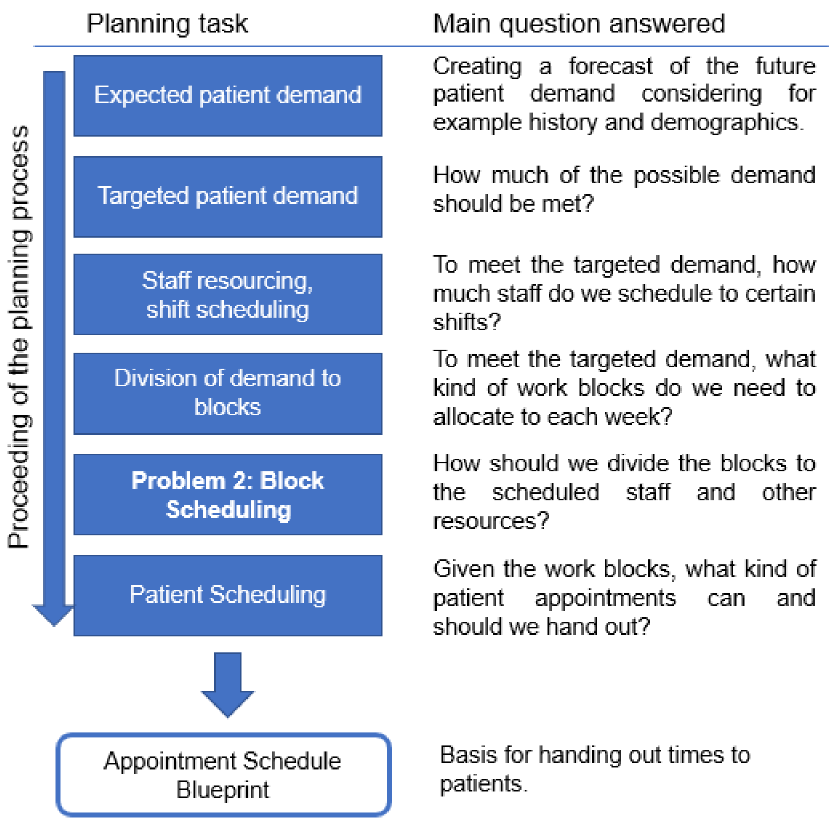



Figure 3.2: The planning framework requiring the solution of problem 2.

shift changes occurring. Thus, we argue that the staff resourcing and shift scheduling should be done for a longer period than the block scheduling.

The block schedule forms basis for appointment booking templates. Each block can incorporate a number of patient appointments and other tasks. The block scheduling problem should be solved to lock reservations on spaces and professionals. This template then provides the staff with their allocated work.

### 3.1.3 Modelling Resources in Problems 1 & 2

The healthcare operation resources are defined in detail by Hall [18, p. 5] who mentions *staff, rooms, equipment, instruments, supplies, implantable devices and organs*. For our formulation, three basic resource categories will suffice, including **staff**, **space** and **movable materials**. These categories are further divided into subcategories based on different criteria.

**Staff** is divided by different competences of the persons. Obviously, there are the educational aspects such as division between doctors and nurses or gastronomic and orthopaedic surgeons. The division of staff can also come from operational planning to divide the workload non-uniformly among the staff.

**Spaces** have differences in fixed equipment, lighting and ventilation. An operation room, for example, has much more fixed equipment than a basic reception room. The spaces also differ in size and intended function. Note that the fixed equipment and materials are seen as a part of the space and do not thus need modelling as standalone resources.

**Movable materials** can contain any resource deemed necessary. These may include movable equipment necessary for certain procedure or hospital beds, for example. The materials can also be consumable. That is, after a certain number of uses they are consumed and require replacement. From the optimization point-of-view, such consumable items should be modelled as normal equipment and the replenishment process disregarded if the study does not directly concern these items.

The number of different resource required in a healthcare operation calls for a general modelling for the different resources. The abstract resource should have as attributes the categories this resource belongs to and the availability during the optimization horizon.

### 3.1.4 Appointments

As previously discussed in Sec. 3.1, the demand side of the problems have two possibilities.

The single appointments in problem 1. represent a single action to be scheduled into resources. This action is most often a patient appointment but can also represent another action, such as an administrative task.

The aggregate demand in problem 2. contains multiple actions to be scheduled to resources in a single block. The contents of this block have similar requirements for resources and the same resources are booked for the whole length of this block. These blocks create a week schedule on high level. Since the blocks can be seen as single appointments requiring certain resources, we can combine these two problems into one optimization problem.

The requirements of resources per appointment are given as a number of resource required per resource category. This enables the scheduling of multi-resource appointments. By default, the appointments usually require at least two resources, a professional and a space.

Multi-stage (chain) appointments contain more than one scheduled part. The consecutive parts typically require different resources. The requirements might be for certain combinations of resources, for example nurse-doctor-nurse where the nurse and allocated room must be same for the first and the last part of the chain. This increases the complexity of the problem.

Each appointment can also be fixed to a certain time period. This should be converted to a resource constraint rather than introducing a time constraint on each appointment.

### 3.1.5 Optimization Time Horizon

The optimization time horizon refers to the length of the time considered by the optimization algorithm. Our two problems described in Sec. 3.1 have similar optimization time horizons of one week. However, the block scheduling in Problem 2. should be done over a month in advance while the single appointment scheduling in Problem 1. is done closer to the realization. This does not necessarily affect the formulation or solving the problems.

The optimization time horizon covers multiple days. Inevitably it contains nights and weekends, resulting in limitations in the schedules of the resources. The breaks during nights and possibly weekends will have to be taken into account while creating the mathematical problem formulation.

## 3.2 Formulation

The problem consists of allocating a set of appointments to given resources on a given interval of time. We seek to formulate the problem as mixed integer linear program.

### 3.2.1 Decision Variables

There are two possibilities for the decision variables. First possibility is to model the time horizon of the scheduling as discrete timeslots for example of five minutes. Each decision variable then tells whether some appointment is ongoing during this discrete timeslot. The discrete interval is used for example in [25, 30].

Another option is to model the starting time of each appointment as a continuous variable. Writing this out briefly results in challenges when describing which resource is allocated to what appointment. To constrain each resource to not have overlapping reservations, we need information on chronological order of the appointments. This means including binary variables for each appointment preceding every other appointment in the set of appointments. The constraints then consist of these binary variables multiplied by the starting time continuous variables, leading to non-linearities.

Additionally, the resource allocation among appointments must be described. This can be done by introducing binary variables whether a resource is allocated to an appointment or not.

Regardless of the type of decision variables, the following two categories must be described by them.

- For each appointment, a decision variable (or set of variables) describing when it happens during the optimization time horizon.
- For each resource, a complete description of the scheduled appointments during the optimization horizon.

We will be using the discrete decision variables and divide the optimization time horizon to intervals.

First, the set $I = \{1, \ldots, n_a\}$ consists of all of the individual appointments indexed from 1 to number of appointments $n_a$. Thus, each $i \in \{1, \ldots, n_a\}$ is a unique identifier for an appointment. Each appointment is an instance from a certain appointment category but the categories are not explicitly stated in the formulation.

Second, the set $J = \{1, \ldots, n_t\}$ indexes the time horizon of interest to $n_t$ discrete intervals of equal size. One can increase the accuracy of the scheduling by tightening the discretization interval $T$ which leads to increased number of intervals. An individual time interval is referred by $j \in J$.

Third, there is the set of resources indexed by $R = \{1, \ldots, n_r\}$. All different types of resources required in healthcare are indexed by the same index. Obviously sub-indices will have to be used to reference a certain resource group such as rooms or doctors.

Finally, we have a set of resource categories $C = \{1, \ldots, n_c\}$. The categories divide the resources into smaller groups. The appointments will most likely require resources from multiple categories, for example a room and a doctor.

After setting up the required indices, we can state the decision variables. We have three types of decision variables to successfully write the whole program as an integer linear program.

First, there is a decision variable which states whether an appointment $i \in I$ starts on interval $j \in J$,

$$z_{ij} \in \{0, 1\} \ . \tag{3.1}$$

Second, there are the starting values for resources,

$$x_{ijr} \in \{0, 1\} \ , \tag{3.2}$$

which define whether the appointment $i$ starts in timeslot $j$ for resource $r$. One might notice that the indexing over $r$ increases the number of decision variables tremendously. Without explicitly stating the calendar of each resource, we quickly run into non-linear functions in the constraints. This information is somewhat duplicate with the decision variables $z_{ij}$ but makes the building of the formulation simpler and more understandable.

Finally, the appointments will often span a length of more than one timeslot so we have decision variables,

$$y_{ijr} \in \{0, 1\} \ , \tag{3.3}$$

which define the hours where appointment $i$ continues in timeslot $j$ for resource $r$. Note that given this start and continue division, $x_{ijr}$ and $y_{ijr}$ should not both equal 1 for any index combination.

We note that the decision variables chosen will lead to sparse decision variable matrices.

### 3.2.2 Appointment and resource information

Next, we introduce the notation used in the optimization problem to incorporate the required information. The used notation is composed to Tab. 3.1.

For each resource, we first have categories. Each resource will belong to at least one category but can belong to multiple categories. The categories of resources are encoded using binary parameters $ct_{rc}$ which state whether resource indexed $r$ belongs to category $c$.

Table 3.1: Description of used notation in the optimization problem

| Notation | Description |
|---|---|
| $I$ | Set of appointments, $i \in I$ |
| $J$ | Set of timeslots, $j \in J$ |
| $R$ | Set of all resources, $r \in R$ |
| $C$ | Set of resource categories, $c \in C$ |
| $T$ | Length of discretization interval in minutes |
| $n_t$ | Number of timeslots |
| $n_r$ | Number of unique resources |
| $n_a$ | Number of appointments |
| $n_c$ | Number of resource categories |
| $n_d$ | Number of days within the optimization horizon |
| $ct_{rc}$ | Binary whether resource $r$ has competence of category $c$ |
| $s_{jr}$ | Binary whether resource $r$ has shift during timeslot $j$ |
| $d_{ic}$ | Demand of resource category $c$ for appointment $i$ |
| $l_i$ | Length of appointment $i$ in number of timeslots |
| $f_i$ | The index $f_i \in I$ of a following appointment for appointment $i \in I$. If no following, $f_i = 0$ |

Next, we have binary parameters $s_{jr}$ describing whether each resource has a shift ie. is available during discretization interval indexed $j$. This parameter locks out the shift planning level which was described in Sec. 2.4. The parameter can also be used to model pre-booked tasks such as lunches for the staff.

For each appointment indexed with $i$, we first have the demands of each resource category $c$ marked with $d_{ic}$. Second, there are the lengths of the appointments as a number of discretization intervals marked with $l_i$.

Finally, we introduce the chain appointments to the problem. The chain appointments link single appointments to each other and they need to be scheduled back-to-back on the time line. The property $f_i$ gives the index of the following appointment for each appointment $i$.

### 3.2.3 Constraints

Following our definition of decision variables in Eqs. (3.1) - (3.3) and giving the information of appointments and resources in Tab. 3.1, we can write the constraints for the optimization problem.

Each appointment can be started only once. Note that we are not hard constraining all appointments to be scheduled:

$$\sum_{j \in J} z_{ij} \leq 1 \, , \, \forall \, i \; . \tag{3.4a}$$

The appointment starts at the correct time within category for resources and the necessary amount has been allocated,

$$\sum_{r \in R} ct_{rc} x_{ijr} = z_{ij} d_{ic} \, , \, \forall \, i, j, c \; . \tag{3.4b}$$

The total amount of resources is correct for each appointment. This is required to ensure that a single resource does not count for two categories:

$$\sum_{j \in J} \left[ z_{ij} \sum_{c \in C} (d_{ic}) - \sum_{r \in R} x_{ijr} \right] = 0 \, , \, \forall \, i \; . \tag{3.4c}$$

One resource can be allocated only once given any timeslot if in shift,

$$\sum_{i \in I} x_{ijr} + y_{ijr} \leq s_{jr} \, , \, \forall \, j, r \; . \tag{3.4d}$$

The start timeslot and the continue slots are always adjacent. This also forces the correct amount of continue intervals,

$$(l_i - 1) x_{ijr} - \sum_{k=1}^{l_i} y_{i,j+k,r} \leq 0 \, , \, \forall \, i, j, r \; . \tag{3.4e}$$

The chain appointments should be back-to-back with each other,

$$z_{f_i(j+l_i)} - z_{ij} = 0 \, , \, \forall \, \{ i \in I : f_i \neq 0 \} \; . \tag{3.4f}$$

The constraints covered in Eq. 3.4 are all linear. The binary constraints of the decision variables are given in Eqs. (3.1) - (3.3). We also notice that (3.4e) forcing the correct amount of timeslots for an appointment has to be taken over all combinations of $i, j$ and $r$. This increases the number of constraints tremendously. The constraints given here are rather general. The problem could be easily constrained further by introducing case specific rules. Any additional constraint will complicate the problem further and makes solving more difficult.

### 3.2.4 Objective Function

Finally, after setting the problem otherwise up, we need the objective of optimization. The decision variables and constraints give rise to the set of feasible solutions, but the optimality criteria enables us to compare these solutions. As previously discussed in Chap. 2, there are multitude of criteria to optimize. Thus, we give a few here which are central to account for in the problem described in Sec. 3.1.

The final objective function used in this study will be a weighted sum (3.5) of the performance criteria stated in (3.6) which has been the approach for most of the studies concerning multiple criteria [2]. This makes the program a multi-objective optimization problem. Many of the articles referenced earlier also employ a multi-objective objective function [4, 25],

$$\min_{\mathbf{X},\mathbf{Y}} \sum_{criteria} w_{crit} f_{crit}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \ . \tag{3.5}$$

Different criteria $f_{crit}$ for the objective function will be given next in Eq. (3.6). The criteria do not have to be all considered. Suitable criteria for the situation can be chosen.

Weighted profile of start hours to encourage, for example, the use of late time within day. Weights given as parameters $\beta_{ij}$,

$$\sum_{i \in I} \sum_{j \in J} \beta_{ij} z_{ij}. \tag{3.6a}$$

Overall resource utilization rate with single resource specific weights $\alpha_r$,

$$\frac{1}{n_r} \sum_{r \in R} \left[ \alpha_r \frac{\sum_{j \in J} \sum_{i \in I} x_{ijr} + y_{ijr}}{\sum_{j \in J} s_{jr}} \right] \ . \tag{3.6b}$$

Appointment spread within and between day can be balanced by introducing a target level $t_j$ for all of the time intervals $j$. This term should be minimized and is close to the variance of ongoing appointments at $j$. To balance a certain category of appointments, one should consider only a subset of $I$ in this criteria:

$$\sum_j \left[ \left( \sum_{i \in I} \frac{\sum_{r \in R} (x_{ijr} + y_{ijr})}{\sum_{c \in C} d_{ic}} \right) - t_j \right]^2 . \tag{3.6c}$$

Maximize the number of scheduled appointments:

$$\sum_{i \in I} \sum_{j \in J} z_{ij}. \tag{3.6d}$$

Minimize the maximal number of resources in use by category $c$:

$$\max_{j \in J} \sum_{r \in R} ct_{rc} \sum_{i \in I} x_{ijr} + y_{ijr} . \tag{3.6e}$$

Maximize the weighted profile of appointments and time slots over resources to model staff preferences. Larger weight implies larger preference. For indifference, set $\eta_{ijr} = 0 \; \forall \; r \in R$,

$$\sum_{i \in I} \sum_{j \in J} \sum_{r \in R} \eta_{ijr} (x_{ijr} + y_{ijr}) . \tag{3.6f}$$

The objective terms in Eq. (3.6) have both linear and non-linear criteria. Generally, non-linear objectives are harder to solve. Non-linear functions can often be modelled as linear by adding terms to the objective function and additional constraints. This again increases the problem size. The heuristic algorithms discussed in the following Chap. 4 are indifferent to the concepts of linearity and non-linearity. Thus, we do not provide a mathematical description of linearising the objective function.

One difficult question is finding suitable weights $w_{crit}$ for the different criteria in the objective function in Eq. (3.5). There is no guarantee that the terms selected have similar ranges compared to each other. Thus, a normalization is required and afterwards the weights have to be set to describe the wanted set up of the schedule created. The process of assigning weights should be based on the relative importance of the criteria.

In this case, the multi-objective optimization problem is turned into single objective optimization problem using the weighted sum method. Another approach would be to consider the criteria separately using an effective frontier and dominance criterion as discussed previously in Sec. 2.5.6 and considered by Cayirli et al. [9].

## 3.3 Analysis of Formulation Complexity

To understand the complexity of the problem, Tab. 3.2 shows the total number of decision variables and constraints for different number of schedulable appointments. The calculated values are based on the constraints and decision variable described in the previous sections 3.2.1 and 3.2.3.

For example of Problem 1., Oulu University Hospital had 537 thousand specialized care outpatient appointments in 2017 [27]. Dividing this up to single weeks and single outpatient clinics, we end up with numbers ranging from hundreds to thousands of schedulable appointments. Within the same outpatient clinic, these appointments might use shared resources and have to be thus considered on the same scheduling instance. By the complexity table, this would rack up to 10 million constraints and decision variables on the 500 appointments being scheduled. The problem 2 with aggregate demand has obviously less blocks to schedule. A harsh estimate on the number of blocks would be 50 half day blocks thus requiring around 25 resources to schedule the whole week. Already at 50 appointments, the number of decision variables and constraints rise up to one million with only 20 resources as in Tab. 3.2. Due to this complexity, we turn to heuristic algorithms in Chap. 4.

Table 3.2: Number of decision variables and constraints on different number of appointments $i$ and fixed number of resources $r = 20$ and discretization intervals $j = 200$.

| $i$ | Total Decision Variables | Total Constraints |
|---|---:|---:|
| 10 | 202000 | 220020 |
| 20 | 404000 | 420040 |
| 50 | 1010000 | 1020100 |
| 100 | 2020000 | 2020200 |
| 300 | 6060000 | 6020600 |
| 500 | 10100000 | 10021000 |

# Chapter 4

# Solution Algorithms

This chapter presents three possible algorithms for solving the problem described in Chap. 3. First, a well-known optimization heuristic, *genetic algorithm*, is applied to the problem. Second, a bin-packing problem based algorithm is given which solves for a minimal number of resources used. Third, a more general sequential, online scheduling rules based algorithm is described. The third algorithm has previously been developed by Delfoi and aims to solve a problem with multiple objectives.

## 4.1 Algorithm Requirements from Users

The scheduling problem described in Chap. 3 is usually faced by specialists in production planning not in mathematical programming. Thus, the use of the algorithm should be as easy as possible. The users are capable of providing input data and parameters in the required form and do not require guidance apart from templates. The calculation from data for the required demand inputs and the description of resources i.e. input data management is left outside the scope of this study.

Since in Sec. 3.3 the problem was found to be very complex, we are looking for a heuristic solution procedure. The heuristic algorithm should be fast enough to provide schedules with upper limit of half hour in problem 1 and preferably in few minutes in problem 2.

The algorithm does not have to provide a global optimum. The real world operation will have variation in terms of sick-leaves and patient no-shows. The optimized schedule provides a starting point for the operation planner and is subject to change. When using the algorithm as a simulation tool, too tight schedules are not wanted either, since for example achieving a 100 % utilization rate is not feasible in a real world operation.

One requirement is also that the algorithm (and formulation) can cope with infeasible problems. That means that all of the appointments might not fit to the planning period with the given resources.

The final results of the algorithm have to be made available visually and in the forms of different reports calculated from the created schedule. The visualization and calculation of such reports are considered trivial after creating the schedule and thus left outside the scope of this study.

## 4.2 Genetic Algorithm

### 4.2.1 The Evolutionary Process

Genetic algorithms mimic the evolution process of nature. The algorithms work on a population or pool of solutions which are used to create offspring through genetic operators. Offspring form the next generation of solutions. An important factor is the fitness function which measures the fitness of each solution. More fit solutions are more likely to participate in creating the offspring and thus passing on good qualities to the next generation. This should increase the average fitness of the next generation.

The basic structure of the algorithm is shown in Fig. 4.1. The genetic algorithm was applied by representing the solution by the start times of each appointment. This corresponds to $z_{ij}$ in the formulation in Chap. 3. Instead of using binary variables, the start was given as the timeslot index $j$. Thus, the newly formed $z_i \in 1, ..., n_t$ where $n_t$ is the number of timeslots. Each solution or genome is then a vector $\mathbf{z}^k$ of length $n_a$ corresponding to start times of the $n_a$ appointments. We index the solutions by $k$. Based on this vector resources are allocated to the appointments. The allocation could also be encoded to the genome but was left outside.

The first step is to create a population of initial solutions. The population size $n_k$ is a parameter of the algorithm. In this case, random starting times were picked and then these solutions were repaired by re-timing appointments left unallocated. This way, a diverse starting population could be created.

Then, new generations are formed by using the genetic operators until the maximum number of generations is reached. The following section discusses the genetic operators in detail. Each of the generated genomes are evaluated for fitness. Based on the fitness, the genomes are selected for reproduction.

A multi-objective fitness function was used. The fitness function is a combination of four different criteria $f_x$ in this case. The criteria include:

- $f_a$, proportion of appointments successfully scheduled
- $f_r$, proportion of rooms used
- $f_s$, proportion of staff who change rooms less than two times on day
- $f_b$, appointment balancing between days for each appointment category

The fitness function $f^k$ is then calculated as a weighted sum of the criteria given above for completely feasible schedules:

$$f^k = f_a^k + I_{f_a^k=1} \left( w_r f_r^k + w_s f_s^k + w_b f_b^k \right) \ , \tag{4.1}$$

where $I_{f_a=1}$ is an indicator function for whether all of the appointments have been successfully scheduled. The fitness was directly used to keep $n_{keep}$ top fitness genomes as is for the next generation.

The evolutionary process should lead to increasing average fitness. The implementation done in this study leaves ample room for improvement. The process and operators used represent the genetic algorithm in basic form. There are multiple of methods to boost the performance of the genetic algorithm. These include, for example, adaptive operator selection based on fitness of the produced offspring.

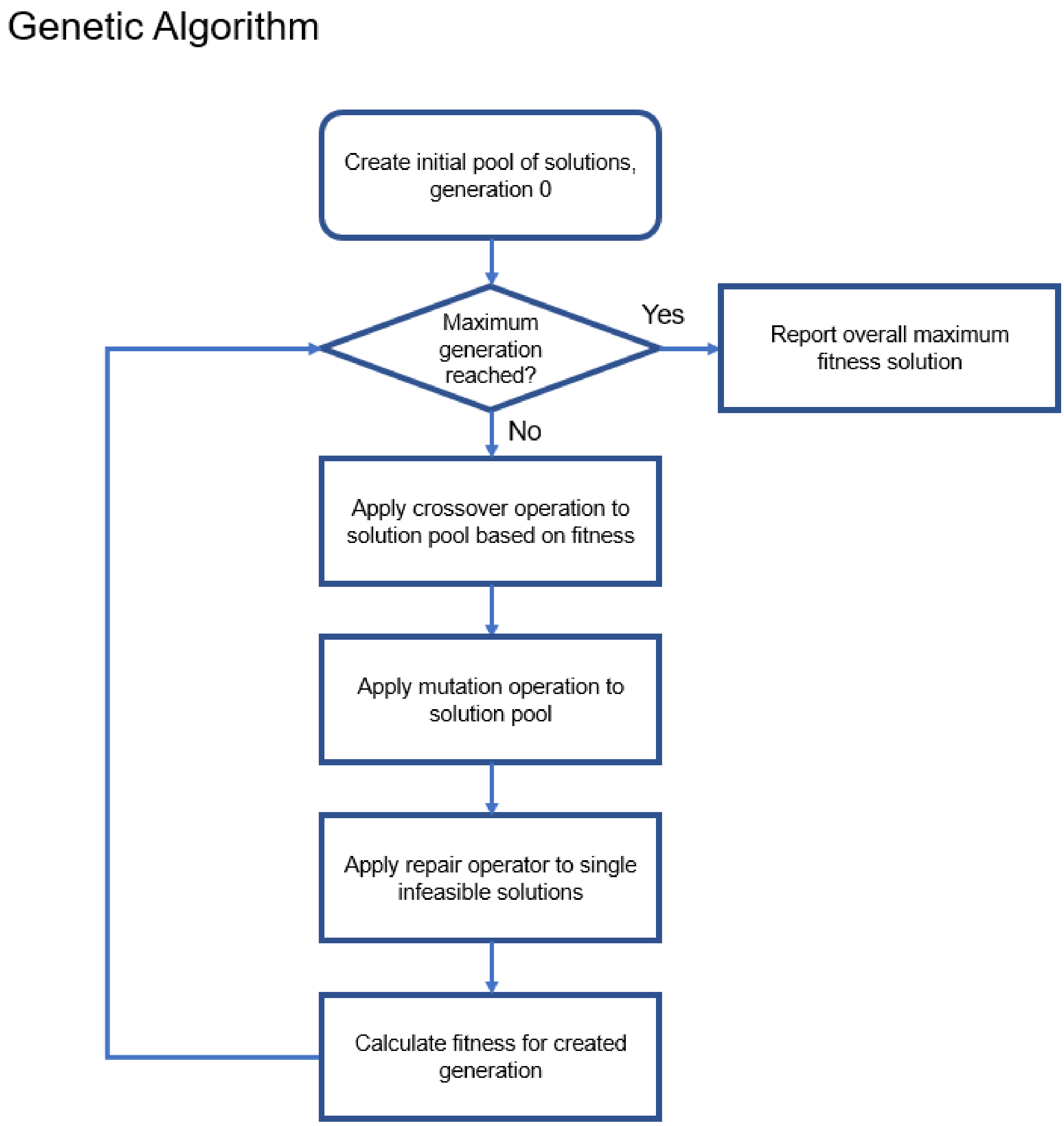



Figure 4.1: A diagram of a genetic algorithm for solving the scheduling problem.

### 4.2.2 Genetic Operators

The evolution from generation to the next generation is done by genetic operators. Basic operators include selection, cross-over and mutation operators. Another operator which is employed is a repair operator. An example process is shown in Fig. 4.2.

The selection and cross-over operator works by first selecting most fit genomes ranked using the fitness function given in Eq. (4.1). The fitness proportionate selection was used. In this method, a genome $\mathbf{z}^k$ is chosen by a probability of,

$$p(\mathbf{z}^k) = \frac{f^k}{\sum\limits_{h=1}^{n_k} f^h} \ ,$$

which is the fitness normalized to a discrete probability distribution.

Given the established selection operator, two genomes are selected to work as parents. These two parents then produce a single offspring by a cross-over operation. In this case a three-point cross-over was used. As illustrated in Fig. 4.2, the cross-over operator alternates between the two parents taking gene sections between the points for the offspring.

The mutation operator is applied to random genomes and random genes are altered. This mutation leads for the algorithm to explore completely new solutions. In this case, the mutation was used to sample a completely new start time for the selected genes. For each genome, a mutation occurs with probability $p_m$ and the size of the mutation in number of genes was fixed. The mutation probability $p_m$ is a parameter of the algorithm. Too high mutation rate leads to divergence of the algorithm while too low rate leads to slow convergence since the solution space is not searched efficiently. The mutation operator is not dependent on the fitness of the genome.

After creating an offspring using selection, cross-over and mutation, the appointment start times might not constitute to a feasible schedule where all appointments can be completed using the given resources. This is tackled by repairing the infeasible offspring by finding a new feasible start time for each of the broken genes. This search is used to boost the algorithm for working on completely feasible schedules instead of spending time on trying to allocate all of the appointments with resources.

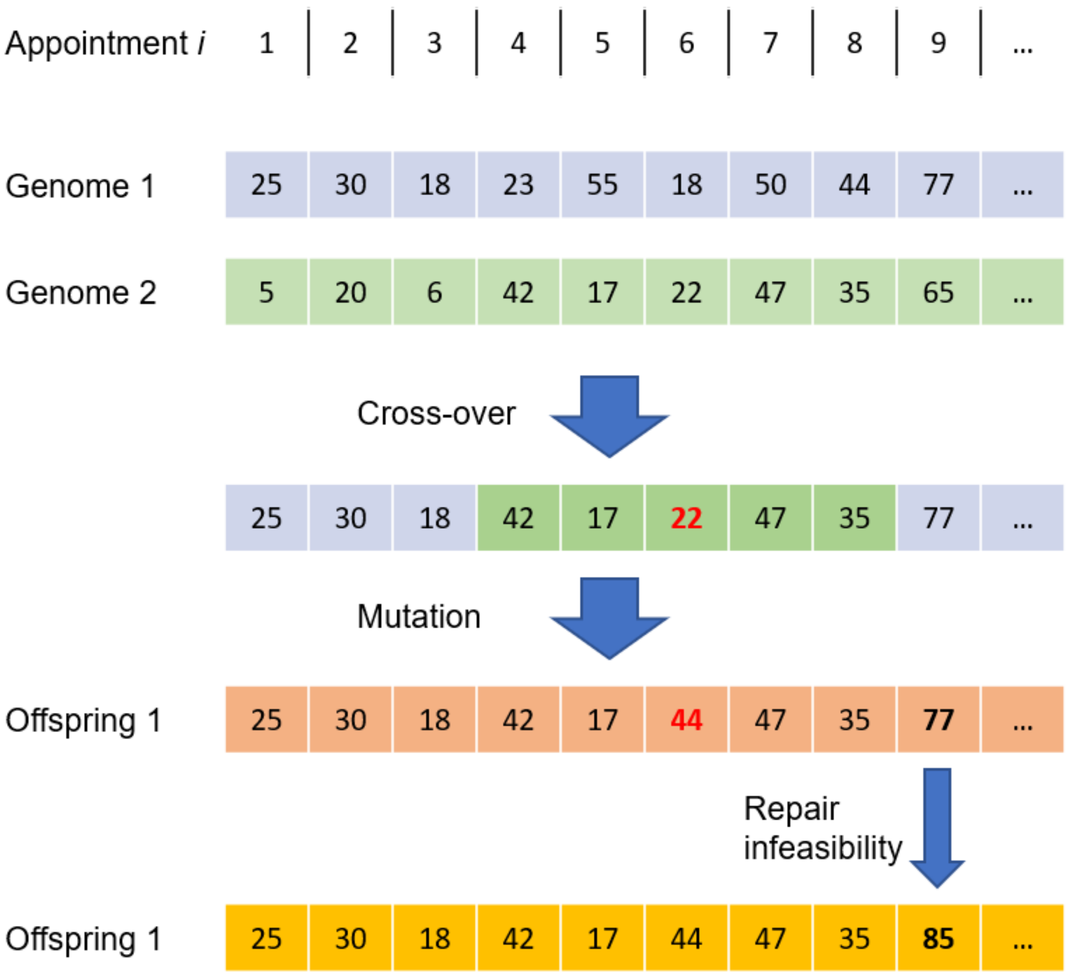



Figure 4.2: An example of generating offspring given two parent genomes using different genetic operators. Each gene in genome represent a starting time $z_i$ for appointment $i$.

## 4.3 First Fit Decreasing Algorithm

### 4.3.1 The Bin-Packing Problem and Connection to the Outpatient Clinic Scheduling

The second algorithm given is a first fit decreasing algorithm. The background of this algorithm is in the bin-packing problems. The goal of a bin-packing problems is to fill bins with pre-determined items while minimizing the number of bins used. Each item and bin has a size, and the idea is to find an optimal packing of each bin to maximize the packing percentage of the bins to be able to use as few bins as possible.

Considering the bin-packing problem and the outpatient clinic scheduling problem, we see that there are similarities. The resources can be seen as the bins and the appointments as the items. Bin size is the working hours of the resources and the item size is the length of the appointment. The analogy is quite complete though. The appointments might require multiple resources, thus leading to an item having to be in multiple bins at a time. One solution would be to constrict the formulation to a single resources only and combining resources to "working combinations" to model the multi-resource nature of appointments. However, this approach would not be able to cope with more complex patient processes. Additionally, we have the scheduling time to consider which is not present in the bin-packing problem.

Instead, we consider that each resource forms a bin and an appointment might be scheduled to multiple bins simultaneously. Now we can formulate the problem so that each resource is open if they have at least one scheduled appointment and otherwise still closed. The first and foremost goal of the problem at hand is to create a feasible schedule while minimizing the number of resources used. When requiring multiple resources, the resource combination is open if all of it's members have been opened.

In terms of healthcare, Vijayakumar et al. [39] formulate an operating theatre surgery scheduling using a dual bin-packing formulation. A dual problem of bin-packing problem is to find out how many items can be fitted to a pre-determined number of bins [39]. Due to problem complexity, the dual bin-packing problem can be solved by employing a sequential first-fit heuristic algorithm. This heuristic sorts the items based on a criteria and assigns the sorted list of items to the first bin available [39].

### 4.3.2 Heuristic Solution algorithm

The basic heuristic methods for bin-packing problems include first-fit algorithms. In this class of algorithms, the items are considered one at a time and fitted to the first bin where they fit. If an item does not fit to any of the existing bins, a new bin is opened. This idea can be used on the healthcare scheduling.

In Fig. 4.3, a first-fit-decreasing algorithm is given for the scheduling problem. First, the appointments are sorted in terms of total workload on resources. This means that the appointment length is multiplied by the number of resources required and sorted on decreasing workload order. The larger appointments are then fitted first since they are harder to schedule than shorted appointments with possibly less resources to consider.

After sorting the appointments, the appointments are scheduled one at a time. The appointment is fitted to first possible resource combination which has been opened. If no opened combination can be found, we select a combination which opens as few new resources as possible.

The first-fit-decreasing algorithm selects the first timeslot available. This should lead to tightly packed schedules. Scheduling to first possible time and resource combination will probably lead to poor solutions with regards to other objectives than the resource usage. Thus, this algorithm is considered as a benchmark since it should be fast. The idea behind this algorithm is extended in the following section considering the rules based scheduling algorithm.

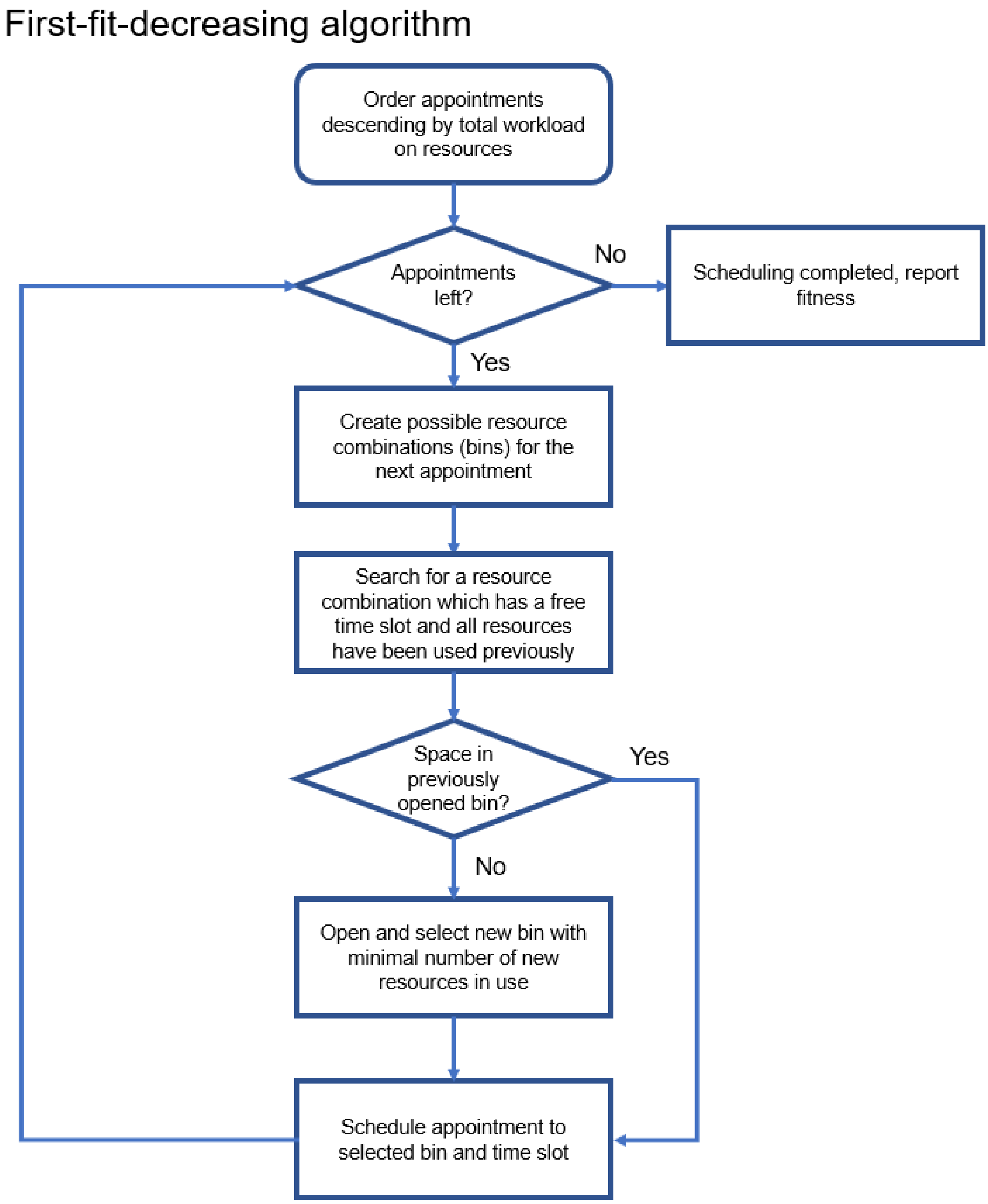



Figure 4.3: A heuristic sequential scheduling algorithm for creating appointment schedules.

## 4.4 Rules Based Sequential Scheduling

### 4.4.1 Sequential Scheduling

Following the idea of the First-Fit-Decreasing algorithm, Fig. 4.4 describes a heuristic procedure for schedule optimization. This kind of an algorithm has been previously developed by Delfoi. The heuristic employs a sequential scheme where appointments are scheduled one at a time and locked afterwards. The whole set of patient appointments and other tasks are known prior to scheduling and the service times are deterministic. For each appointment, the required resources are also given as inputs which corresponds to $d_{ic}$ in the problem formulation in Chap. 3.

Working on a single appointment at a time, the algorithm finds resource combinations and time intervals for the appointment. Alternatives of resources combinations and time intervals are evaluated based on a set of rules. The rules are used to guide the solution towards a better schedule while the first-fit-decreasing algorithm only picks the first possible already opened combination of resources. This increases the complexity of the algorithm, but should result in better solution evaluated with the selected objective function in Eq. (4.1). After finding the best alternative, the appointment is scheduled on the resources. The rules used to evaluate and compare different combinations are discussed in the next section.

The sequential nature of this algorithm makes it an online algorithm. Sequential scheduling in outpatient clinic setting has been studied by Chakraborty et al. [11] and Yan et al. [42]. They both consider a myopic scheduling where each appointment is scheduled independent of the following appointments. They both report that their objective functions of total expected revenue for scheduled appointments is unimodal, leading to a stopping rule. Once the new appointment decreases the total objective value, the scheduler should not add any further appointments to this considered time interval. These studies strengthen our incentive to study a sequential algorithm. The problems solved in this case include multi-server environments and thus are not directly comparable to [11, 42] which concentrate on a single-server models.

The algorithm shown in Fig. 4.4 employs myopic scheduling. Appointments are scheduled based on information of prior scheduled appointments and the current appointment only. Since we have complete information on all of the following appointments also, this policy disregards using that information for optimal assignment. This will most likely lead to a suboptimal solution.

To provide an optimal schedule, one should lock the single appointment only after considering the cost-to-go for this scheduling decision. That would mean iterating, explicitly or implicitly, the following scheduling decisions from this stage to the end and calculating the optimal cost-to-go. This is a dynamic programming issue and is left outside the scope of this study. Pérez et al. [30] study this type of online scheduling problem with consideration in expected demand arriving after scheduling the current appointment.

### 4.4.2 Scheduling Rules

The scheduling rules in the context of this algorithm are quantities that can be calculated for the scheduling decision about to be made. Each resource combination and time interval can be evaluated on how they affect the objective function given in Eq. (4.1). Instead of calculating the objective over the complete schedule, we are calculating the objective for a single scheduling of an appointment. These rules rely only on the information of the prior scheduled appointments.

The rules used to guide the scheduling towards a good solution were in form of questions:

- Are new resources required?
- Does this time slot unbalance the appointments between days?
- Does this time slot and room result in excess room changes for the staff?

and the evaluation is returned as a binary number. The rules correspond to the components of the selected fitness function in Eq. (4.1) and the basic idea is that if we avoid scheduling decisions which lead to negative effect in the fitness function, we should end up with a good solution. The final decision on which appointment to schedule next is done by comparing the calculated criteria using pre-assigned weights.

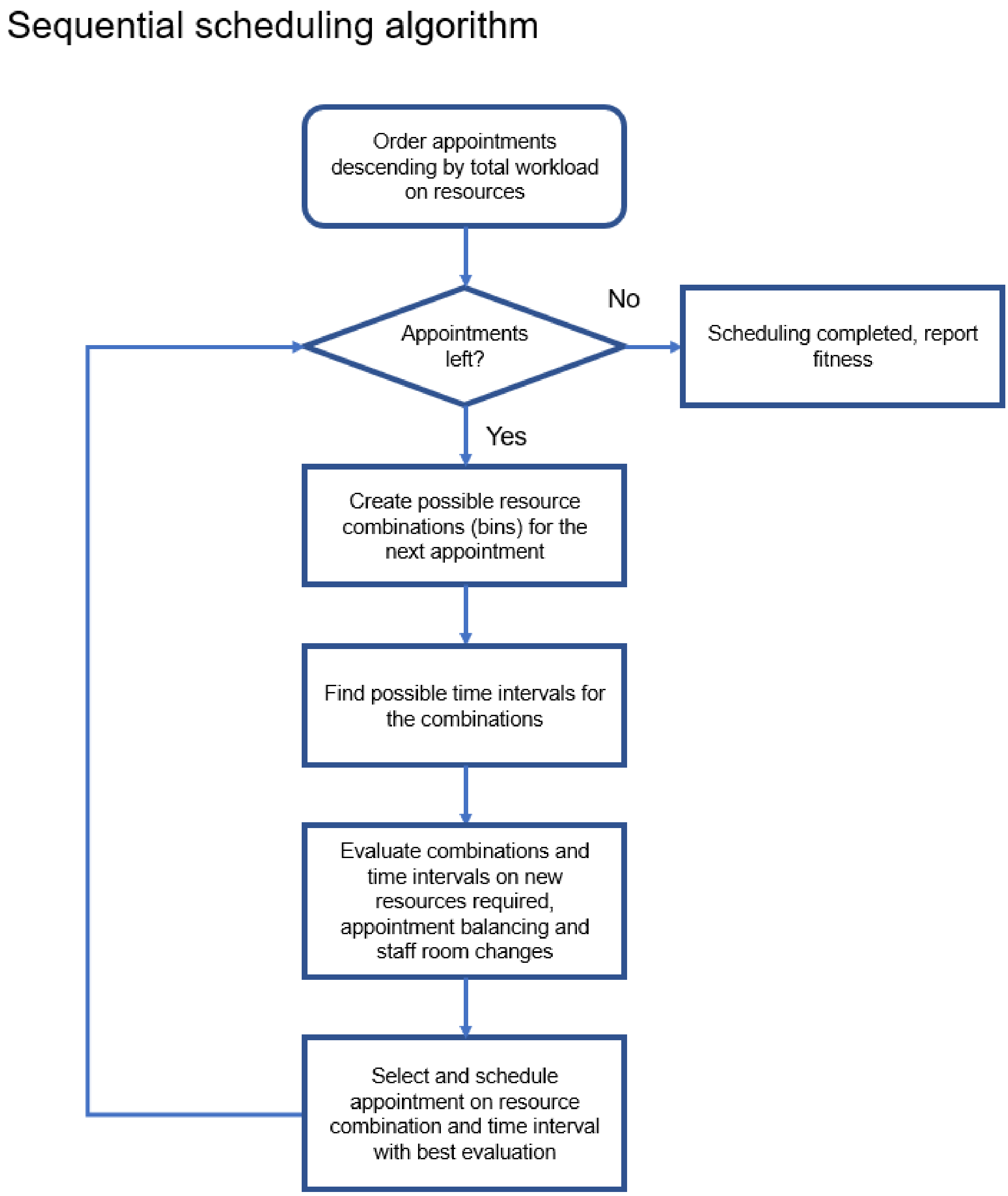



Figure 4.4: A heuristic sequential scheduling algorithm for creating appointment schedules.

## 4.5 Comparison of the Algorithms

Comparing the three algorithms discussed in the previous sections, we first note that the genetic algorithm works on the whole appointment set while the two other algorithms only consider a single appointment at a time. Thus, the genetic algorithm is an offline algorithm while the first-fit-decreasing and rules based scheduling are online algorithms by nature.

Another clear difference is that the genetic algorithm creates multiple schedules, number of generations multiplied by the population size to be exact. Again, the other two algorithms are similar and create only a single schedule. From this perspective, the genetic algorithm tries to modify the solution towards an optimal one while the other two create an initial schedule as good as possible.

Comparing the first-fit-decreasing and the rules based scheduling algorithms, we note that they are similar and the rules based scheduling can be seen as an extension of the first-fit algorithm. The rules based scheduling employs multiple rules while the first-fit-decreasing algorithm only uses the single rule of finding an allocation of resources which does not open new resources.

# Chapter 5

# Test Results

This chapter compares the three algorithms presented in Chap. 4. Three different test problems are solved by each of the algorithm and results compared in terms of objective function values and run times.

## 5.1 Test Problems

To compare the algorithms and validate that they can provide solutions for the problem, they were tested using three different test problems. The test problems were for a five day period. Two smaller problems were selected with tight and excess resourcing compared to the number of appointments. Additionally one larger problem was tested. The problem sizes are given in Tab. 5.1.

Test problems TP 1 and TP 2 had a hundred appointments while the larger TP 3 had a 1000 appointments. The appointment sets in TP 1 and TP 2 were the same. On the resource side, all problems had enough resources to accommodate all of the appointments. TP 1 had more resources than TP 2 for the same appointments.

The types of appointments used are single or two professional appointments. Additionally, a single room is required for each appointment. Chain appointments or other more complex appointment types were not considered.

Considering these problems, the larger is close to a large outpatient clinic while the smaller ones could be the size of a specific specialty outpatient clinic when considering Problem 1 from Chap. 3 where a single appointment level

Table 5.1: Dimensions of the three test problems.

| | TP 1 | TP 2 | TP 3 |
|---|---|---|---|
| Appointments | 100 | 100 | 1000 |
| Appointment Categories | 4 | 4 | 10 |
| Average appointment length (min) | 95 | 95 | 74 |
| Resources required for appointment | 2-3 | 2-3 | 2-3 |
| Rooms | 10 | 6 | 48 |
| Doctors | 10 | 6 | 48 |
| Nurses | 4 | 2 | 23 |

schedule has to be made. The problems cannot be seen hard or exceptional on the real cases encountered and thus the algorithms must be able to perform on these.

## 5.2 Solutions and Run Times

The test problems were run using all three algorithms. The genetic algorithm was run with 100 generations and population size of 50. The parameters of the genetic algorithm were experimented with and suitable values found.

The resulting objective values are given in Tab. 5.2 where the objective values correspond to the objective function given in Eq. (4.1). Note that larger values are better, and thus the objective values represent proportions of rooms not used and staff not changing rooms excessively.

The first result is that the rules based scheduling algorithm has the highest total objective on all of the test problems. The genetic algorithm yields second best result on TP 1 and TP 3 but loses to the first-fit-decreasing algorithm in TP 2.

Going through the components of the objectives, we note that all of the algorithms are able to form complete schedules with the apps scheduled $f_a$ yielding values of 1 on all of the test problems. On the rooms used $f_r$, we note that the genetic algorithm has the lowest value on all of the test problems, thus using the most rooms. First-Fit-Decreasing and the rules based scheduling yield similar results on TP 1 and TP 2. Comparing with the number of rooms in Tab. 5.1, these two algorithms pack the appointments to 5 rooms out of 10 and 6 available on the two problems being indifferent on the amount of resources available.

Table 5.2: Objective values of the final solutions. Larger values indicate better solutions.

| | Genetic | First-Fit-Decreasing | Rules based scheduling |
|---|---|---|---|
| **TP 1** | | | |
| Total objective $f$ | **1.66** | **1.49** | **1.81** |
| Apps scheduled $f_a$ | 1.00 | 1.00 | 1.00 |
| Rooms used $f_r$ | 0.30 | 0.50 | 0.50 |
| Room changes $f_s$ | 0.78 | 0.40 | 1.00 |
| App balance $f_b$ | 0.79 | 0.64 | 0.76 |
| **TP 2** | | | |
| Total objective $f$ | **1.33** | **1.40** | **1.44** |
| Apps scheduled $f_a$ | 1.00 | 1.00 | 1.00 |
| Rooms used $f_r$ | 0.00 | 0.17 | 0.17 |
| Room changes $f_s$ | 0.30 | 0.40 | 0.425 |
| App balance $f_b$ | 0.76 | 0.64 | 0.76 |
| **TP 3** | | | |
| Total objective $f$ | **1.24** | **1.08** | **1.58** |
| Apps scheduled $f_a$ | 1.00 | 1.00 | 1.00 |
| Rooms used $f_r$ | 0.06 | 0.31 | 0.21 |
| Room changes $f_s$ | 0.00 | 0.00 | 0.61 |
| App balance $f_b$ | 0.91 | 0.02 | 0.91 |

Continuing with the room changes $f_s$, we note that the rules based scheduling has largest values on all of the three test problems. The first-fit-decreasing algorithm does not actively consider room changes and thus falls behind. Still, the scheduling order of the first-fit-decreasing leads to it beating the genetic algorithm on two of the problems. Finally, the appointment balancing over days $f_b$ has the genetic algorithm and the rules based scheduling showing larger values than the first-fit-decreasing algorithm, especially in the largest problem TP 3.

Tab. 5.3 gives the run times of each of the algorithms on the three test problems. We instantly notice that the genetic algorithm has run times almost two orders slower than the first-fit-decreasing algorithm which runs the fastest. Considering that the genetic algorithm did not beat the rules based scheduling on the objective values, the longer run time is unacceptable. Comparing with the algorithm requirements given in Chap. 4, the run times of the rules based scheduling are in-line with the expectations.

Table 5.3: Run times of the three algorithms on the three test problems. Value in parenthesis tells the average generation run time for the genetic algorithm.

| | TP 1 | TP 2 | TP 3 |
|---|---|---|---|
| Genetic | 68 s (0.6) | 168 s (1.6) | 1546 s (11.2) |
| First-Fit-Decreasing | 0.4 s | 0.3 s | 37 s |
| Rules based scheduling | 2.1 s | 0.7 s | 269 s |

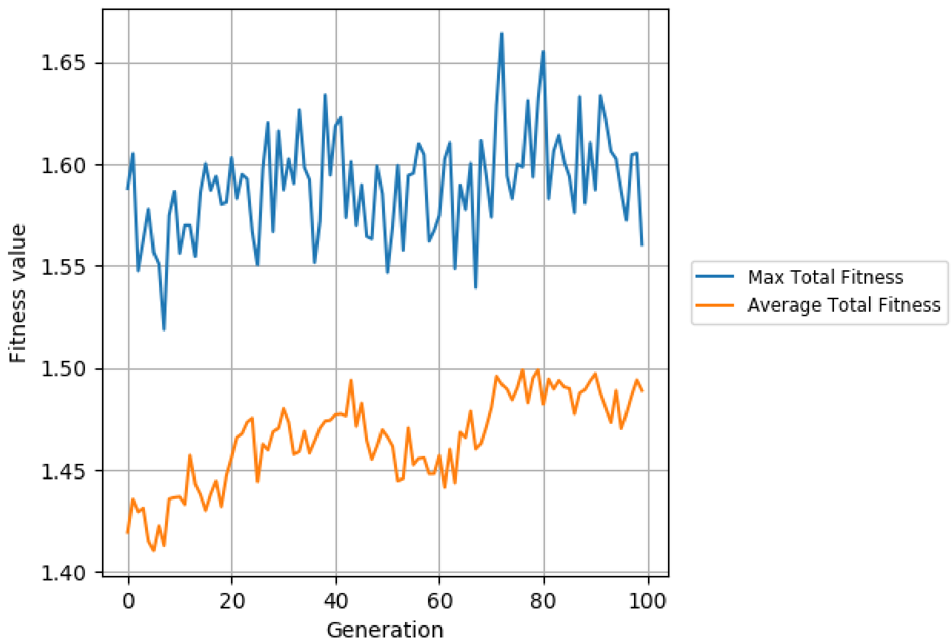
The genetic algorithm is an evolutionary process and thus should converge towards an optimal solution. During the test runs, 100 generations was used. This resulted in the evolution of the fitness as shown in Figs. 5.1, 5.2 and 5.3 for each of the test problems. In the figures, the above plot contains the maximum and average total fitness as a function of the generation. The lower plot contains the average fitness function components as a function of the generation.

First, we notice that the maximum total fitness has quite a bit of variance. This happens since we let the best genome mutate also. The average total fitnesses have less variance but some exists since the allocation of resources has randomness built to it. Preferably, the fitness curves should be monotonously increasing due to the evolution proceeding. In all of the test problems we note that the average total fitness does increase.

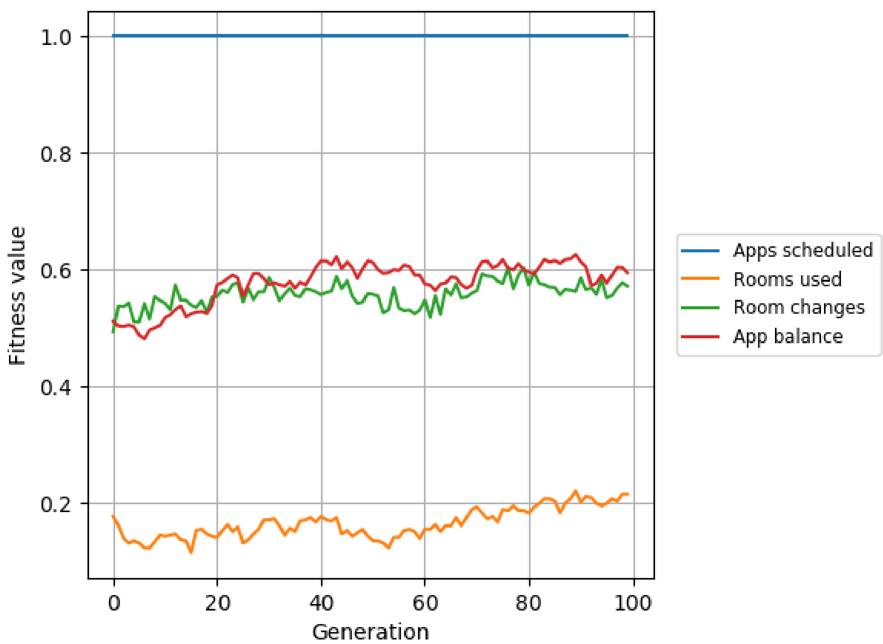
Considering the fitness components, we note that the apps scheduled fitness is 1 for TP 1 in Fig. 5.1 and 5.3 meaning that all appointments get successfully scheduled. In TP 2 Fig. 5.2, we see that the proportion of scheduled appointments slowly increases. The algorithm is thus converging towards a solution with all appointments scheduled.

It is difficult to judge whether the algorithm has converged from these figures. Comparing to the other algorithms, the genetic algorithm was slow and there is still clearly room for improvement. For the genetic algorithm to compete with the rules based scheduling algorithm, the convergence speed should improve drastically.

Concluding with the test results, we state that the rules based scheduling performed best considering the results and the run-times. The first-fit-decreasing algorithm was fast but did not provide good results. The genetic algorithm provided moderate results but the run times were too long.

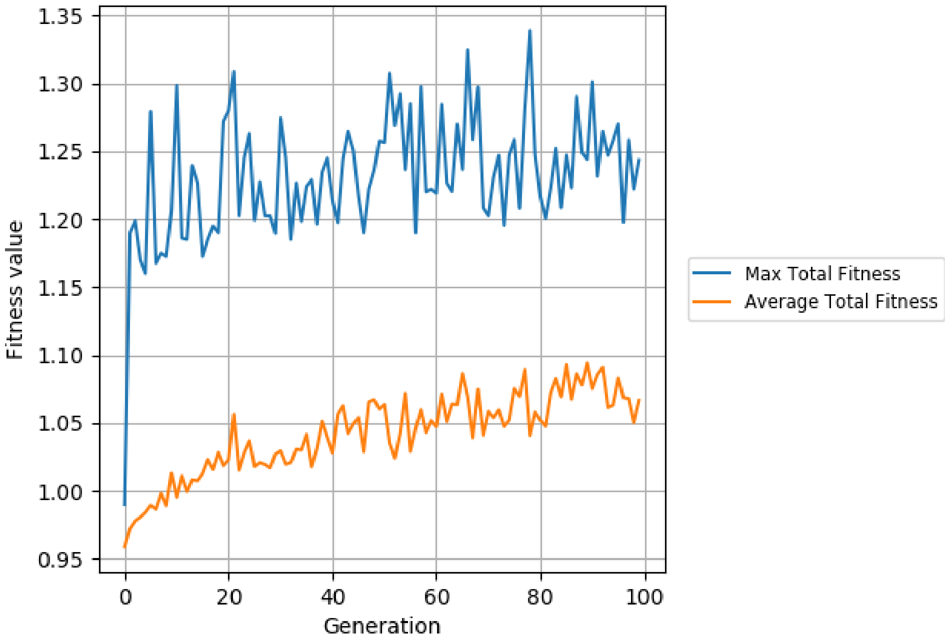



(a) Fitness Convergence

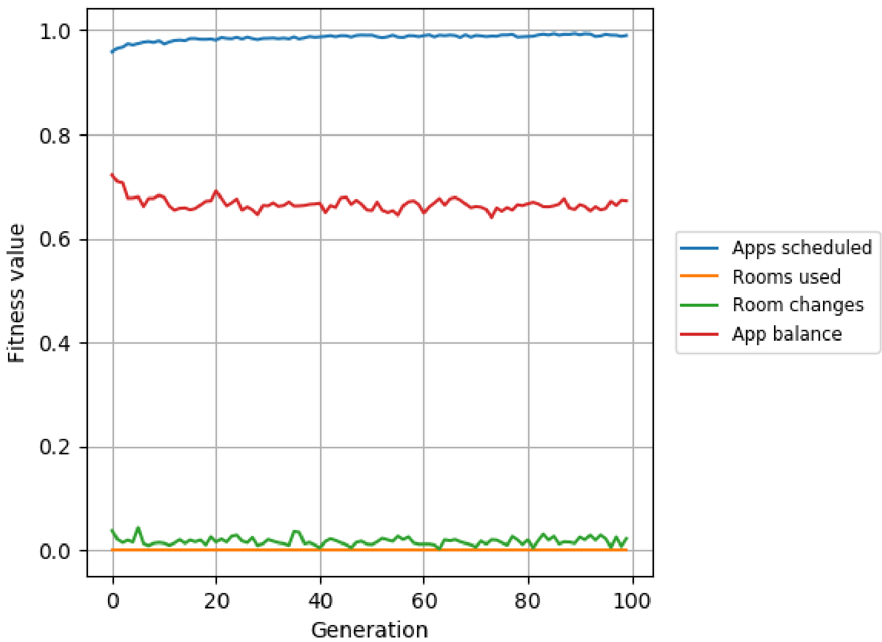



(b) Fitness components

Figure 5.1: Genetic algorithm total fitness convergence and fitness components on test problem TP1.

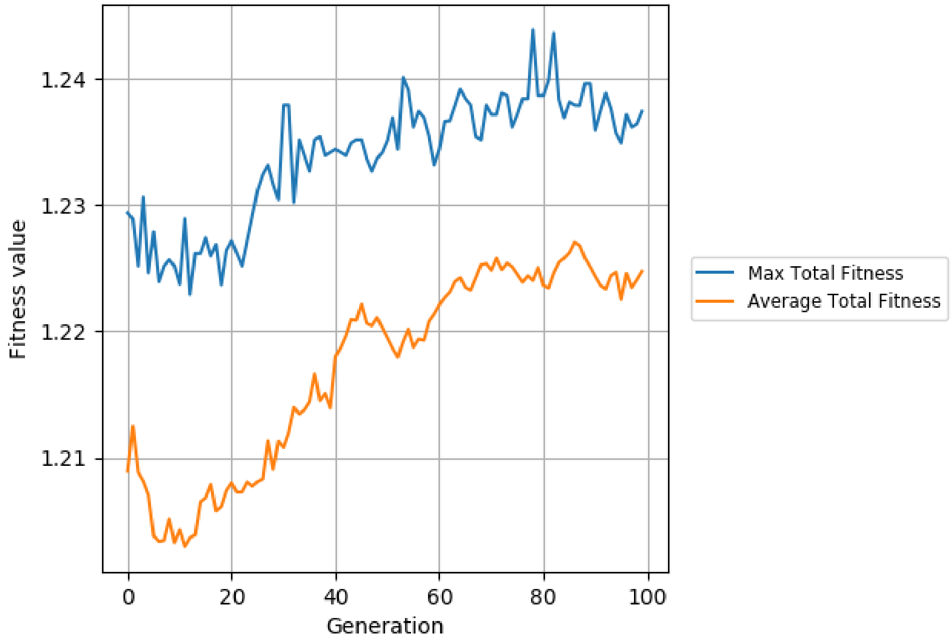



(a) Fitness Convergence

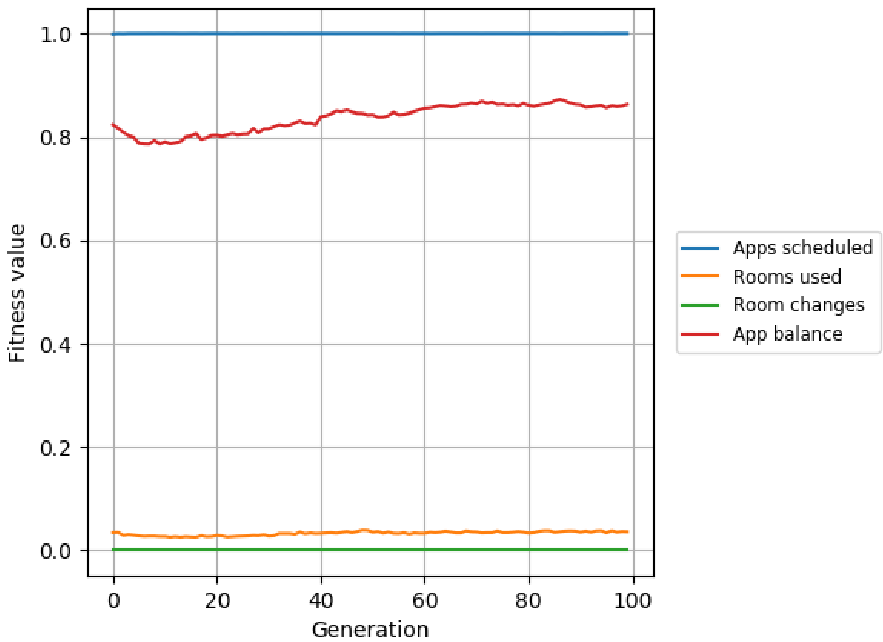



(b) Fitness components

Figure 5.2: Genetic algorithm total fitness convergence and fitness components on test problem TP2.




(a) Fitness Convergence




(b) Fitness components

Figure 5.3: Genetic algorithm total fitness convergence and fitness components on test problem TP3.

# Chapter 6

# Conclusions

This thesis studied optimization problems within outpatient clinic production planning and control. The purpose was to present optimization problems and to formulate and solve an outpatient clinic scheduling problem. Literature review was conducted to present the problems and problem formulation was given as a integer linear program. The scheduling problem was solved with three heuristic algorithms and the performance compared.

Starting with the literature review on the optimization problems, the hierarchical planning framework was first introduced. The planning process is continuous and contains stages looking at different time horizons. Each stage increases the precision of the plan and each stage contains optimization problems of their own.

With regards to outpatient clinic appointment scheduling, it was found that there are multiple decisions to make. These decisions include but are not limited to, access policy, dealing with patient variability, appointment length and number of patients to serve. Finding an optimal combination of these decisions will lead to setting up the outpatient clinic and forming the appointment blueprint. Regardless of the decisions considered, there were two main categories of objectives in the optimization problems studied: cost based and comfort based. These categories also formed the major trade-off to consider. Many of the studies considered the stochastic nature of the operation but concentrated on optimizing a single server model.

The optimization literature on outpatient clinics seemed to focus on maximizing the revenue from the operation while keeping the patient waiting times and staff overtime low. This approach is a bit inadequate for Finland where the healthcare is public. Based on research as well as practical lessons learned, there are other aspects to consider such as administrative tasks or patient phone calls which the staff has to also do. Especially within specialized care in larger hospitals, the doctors have other tasks including for example research, teaching and visits to the wards which constitute a major part of the time available on these resources. Taking these into account should be done regardless of the fact that it increases the problem complexity.

Most studies in outpatient clinic appointment scheduling consider only a single-server model and create a one-day or even a few appointments long schedules. The idea is that these blocks can be then duplicated and schedule created. However, there are more complex patient processes involving multiple resources and spanning a longer time. The care processes are constantly evolving and the tools have to take into account complex processes also. There is efficiency to be found in properly setting up such an operation.

Proceeding to formulating the scheduling problem, we first described that there are two problems that should be solved differing on the demand side. The first, problem 1, had single appointments while problem 2 had demand as appointment blocks. These problems were formulated into a single integer linear program considering all resources on individual level. Surprisingly, the objective of optimization was the hardest part to formulate. It included linear and non-linear criteria to choose from. Keeping up with the linearity of the program made the problem size to expand tremendously, resulting in millions of decision variable and constraints. Due to this complexity, heuristic algorithms were tested.

Finally, three different algorithms were given for solving the problem. These included a genetic, first-fit-decreasing and rules based scheduling algorithms. Out of the three, only the genetic algorithm considers multiple schedules while the other two algorithms create the schedule in a single pass. Also, the genetic algorithm considers the whole schedule at a time while the other two schedule a single appointment at a time. These differences lead to large difference in run times where the genetic algorithm performed poorly. The rules based scheduling algorithm performed best in terms of the objective value.

Based on the results, the rules based scheduling algorithm was the tool of choice. Considering the algorithm, it is understandable and the schedule is built almost as a simulation of handing out appointments for patients. Considering this, the algorithm is closer to simulation than optimization. Since the main use is to test new operating models or resourcing decisions, this kind of combination of simulation and optimization is acceptable even though a global optimum is disregarded. The algorithm leaves ample room for modifications or bringing in new schedule creation rules. The algorithm filled all of the user requirements given in Sec. 4.1.

For further research, there are interesting questions to answer. First, this thesis studied a tool for creating appointment schedules given a deterministic demand. Including stochasticity is an obvious direction and could be implemented by first scheduling using the algorithm presented here and then simulating the created schedule using a simulation software with randomness added to the appointments. This idea could also be appended with walk-in patients to see how waiting times of patients evolve. Further developing the algorithm would be beneficial.

Second, a complete case study on the planning and scheduling of a real organization could be conducted. This study gave tools but only on a theoretical level. What is the current state of the operation and could it be improved by using optimization? Is it even plausible to apply for example the overbooking strategies given in Chap. 2? How can one implement these changes? Can measurable benefits be achieved?

Third, investigating the scheduling of patient chain of care in whole. In this study, the patient was investigated through a single appointment. But often after a diagnosis, a patient requires a chain of care including multiple departments for doctor appointments or visits to the laboratory for tests. This information could be used to better forecast the future demand over department borders to balance load or for the use of capacity dimensioning. From patient point-of-view, the lead time of the care could be optimized. This comes close to the forecasting of demand which is also an important question outside the scope of this study.

Finally, finding or developing concrete tools for the other optimization problems given in the literature review in Chap. 2 including for example the nurse scheduling problem. Strategical level planning problems were outside the scope of this study. Optimization problems including between hospital division of patients and hospital placement problems could bring insight and tools for the hot topic of healthcare reform in Finland. There are probably existing studies on this subject.

# Bibliography

[1] Alessandro Agnetis, Alberto Coppi, Matteo Corsini, Gabriella Dellino, Carlo Meloni, and Marco Pranzo. A decomposition approach for the combined master surgical schedule and surgical case assignment problems. *Health care management science*, 17(1):49–59, 2014.

[2] Amir Ahmadi-Javid, Zahra Jalali, and Kenneth Klassen. Outpatient appointment systems in healthcare: A review of optimization studies. *European Journal of Operational Research*, 258(1):3–34, 2017.

[3] Bjorn Berg and Brian Denton. Fast approximation methods for online scheduling of outpatient procedure centers. *INFORMS Journal on Computing*, 29(4):631–644, 2017.

[4] Jens Brunner, Jonathan Bard, and Rainer Kolisch. Midterm scheduling of physicians with flexible shifts using branch and price. *Iie Transactions*, 43(2):84–109, 2010.

[5] Edmund Burke, Patrick De Causmaecker, Greet Berghe, and Hendrik Van Landeghem. The state of the art of nurse rostering. *Journal of scheduling*, 7(6):441–499, 2004.

[6] Brecht Cardoen, Erik Demeulemeester, and Jeroen Beliën. Operating room planning and scheduling: A literature review. *European journal of operational research*, 201(3):921–932, 2010.

[7] Adam Carpenter, Lawrence Leemis, Alan Papir, David Phillips, and Grace Phillips. Managing magnetic resonance imaging machines: support tools for scheduling and planning. *Health care management science*, 14(2):158–173, 2011.

[8] Tugba Cayirli and Emre Veral. Outpatient scheduling in health care: a review of literature. *Production and operations management*, 12(4):519–549, 2003.

[9] Tugba Cayirli, Emre Veral, and Harry Rosen. Designing appointment

scheduling systems for ambulatory care services. *Health care management science*, 9(1):47–58, 2006.

[10] Santanu Chakraborty, Kumar Muthuraman, and Mark Lawley. Sequential clinical scheduling with patient no-shows and general service time distributions. *IIE Transactions*, 42(5):354–366, 2010.

[11] Santanu Chakraborty, Kumar Muthuraman, and Mark Lawley. Sequential clinical scheduling with patient no-show: The impact of pre-defined slot structures. *Socio-Economic Planning Sciences*, 47(3):205–219, 2013.

[12] Rachel Chen and Lawrence Robinson. Sequencing and scheduling appointments with potential call-in patients. *Production and Operations Management*, 23(9):1522–1538, 2014.

[13] Guus De Vries, J. Will Bertrand, and Jan Vissers. Design requirements for health care production control systems. *Production Planning & Control*, 10(6):559–569, 1999.

[14] S. Ayca Erdogan and Brian Denton. Dynamic appointment scheduling of a stochastic server with uncertain demand. *INFORMS Journal on Computing*, 25(1):116–132, 2013.

[15] Andreas Fügener, Erwin Hans, Rainer Kolisch, Nikky Kortbeek, and Peter Vanberkel. Master surgery scheduling with consideration of multiple downstream units. *European journal of operational research*, 239 (1):227–236, 2014.

[16] Fred Gloyer and Claude McMillan. The general employee scheduling problem: An integration of ms and ai. *Computers & Operations Research*, 13:563–573, 1986.

[17] Diwakar Gupta and Wen-Ya Wang. Patient appointments in ambulatory care. In Randolph Hall, editor, *Handbook of healthcare system scheduling*, volume 168 of *International series in operations research & management science*, chapter 4, pages 65–104. Springer Science+Business Media, New York, 2012.

[18] Randolph Hall. *Handbook of healthcare system scheduling*. International series in operations research & management science, v. 168. Springer Science+Business Media, New York, 2012.

[19] Peter Hulshof, Nikky Kortbeek, Richard Boucherie, Erwin Hans, and Piet Bakker. Taxonomic classification of planning decisions in health care: a structured review of the state of the art in or/ms. *Health systems*, 1(2):129–175, 2012.

[20] Peter Hulshof, Peter Vanberkel, Richard Boucherie, Erwin Hans, Mark van Houdenhoven, and Jan-Kees van Ommeren. Analytical models to determine room requirements in outpatient clinics. *OR spectrum*, 34(2): 391–405, 2012.

[21] Kenneth Klassen and Reena Yoogalingam. Strategies for appointment policy design with patient unpunctuality. *Decision Sciences*, 45(5):881–911, 2014.

[22] Seung Lee, Gregory Heim, Chelliah Sriskandarajah, and Yunxia Zhu. Outpatient appointment block scheduling under patient heterogeneity and patient no-shows. *Production and Operations Management*, 27(1): 28–48, 2018.

[23] Bohui Liang, Ayten Turkcan, Mehmet Erkan Ceyhan, and Keith Stuart. Improvement of chemotherapy patient flow and scheduling in an outpatient oncology clinic. *International Journal of Production Research*, 53 (24):7177–7190, 2015.

[24] Ho-Yin Mak, Ying Rong, and Jiawei Zhang. Appointment scheduling with limited distributional information. *Management Science*, 61(2): 316–334, 2014.

[25] Nadine Meskens, David Duvivier, and Arnauld Hanset. Multi-objective operating room scheduling considering desiderata of the surgical team. *Decision Support Systems*, 55(2):650–659, 2013.

[26] Tomi Nurmi. Perusterveydenhuollon toiminnansuunnittelu ja -ohjaus. 2016.

[27] Oulu University Hospital (OYS). Seurantatietoa 7/2018 tilastot. Available online: https://www.ppshp.fi/Sairaanhoitopiiri/Toiminta-ja-talous/Tilastot/Pages/default.aspx. Accessed 14.9.2018.

[28] Vesa Paju, Eeva Ketola, Kristiina Kariniemi-Örmälä, Juha Halonen, and Rashmi Werning. Using simulation in various stages of hospital planning. In Ira Verma and Laura Nenonen, editors, *ARCH 14 - International Conference on Research on Health Care Architecture - November 19-21, 2014, Espoo, Finland - Conference Proceedings*, volume 6 of *Aalto University publication series ART + DESIGN + ARCHITECTURE*, pages 281–295. Aalto University, Espoo, 2015.

[29] Eduardo Pérez, Lewis Ntaimo, Wilbert Wilhelm, Carla Bailey, and Peter McCormack. Patient and resource scheduling of multi-step medical

procedures in nuclear medicine. *IIE Transactions on Healthcare Systems Engineering*, 1(3):168–184, 2011.

[30] Eduardo Pérez, Lewis Ntaimo, César Malavé, Carla Bailey, and Peter McCormack. Stochastic online appointment scheduling of multi-step sequential procedures in nuclear medicine. *Health care management science*, 16(4):281–299, 2013.

[31] Dinh-Nguyen Pham and Andreas Klinkert. Surgical case scheduling as a generalized job shop scheduling problem. *European Journal of Operational Research*, 185(3):1011–1025, 2008.

[32] Xiuli Qu, Yidong Peng, Jing Shi, and Linda LaGanga. An mdp model for walk-in patient admission management in primary care clinics. *International Journal of Production Economics*, 168:303–320, 2015.

[33] Abdur Rais and Ana Viana. Operations research in healthcare: a survey. *International transactions in operational research*, 18(1):1–31, 2010.

[34] Lawrence Robinson and Rachel Chen. A comparison of traditional and open-access policies for appointment scheduling. *Manufacturing & Service Operations Management*, 12(2):330–346, 2010.

[35] Elina Rönnberg and Torbjörn Larsson. Automating the self-scheduling process of nurses in swedish healthcare: a pilot study. *Health Care Management Science*, 13(1):35–53, 2010.

[36] Pieter Smet, Patrick De Causmaecker, Burak Bilgin, and Greet Vanden Berghe. Nurse rostering: A complex example of personnel scheduling with perspectives. In A. Sima Uyar, Ender Ozcan, and Neil Urquhart, editors, *Automated Scheduling and Planning*, volume 505 of *Studies in computational intelligence*, pages 129–154. Springer-Verlag, Berlin Heidelberg, 2013.

[37] James Swisher and Sheldon Jacobson. Evaluating the design of a family practice healthcare clinic using discrete-event simulation. *Health Care Management Science*, 5(2):75–88, 2002.

[38] James R Swisher, Sheldon H Jacobson, J Brian Jun, and Osman Balci. Modeling and analyzing a physician clinic environment using discrete-event (visual) simulation. *Computers & operations research*, 28(2):105–125, 2001.

[39] Bharathwaj Vijayakumar, Pratik Parikh, Rosalyn Scott, April Barnes, and Jennie Gallimore. A dual bin-packing approach to scheduling surgi-

cal cases at a publicly-funded hospital. *European Journal of Operational Research*, 224(3):583–591, 2013.

[40] Jan Vissers, J. Will Bertrand, and Guus De Vries. A framework for production control in health care organizations. *Production Planning & Control*, 12(6):591–604, 2001.

[41] T. Wong, M. Xu, and K. Chin. A two-stage heuristic approach for nurse scheduling problem: A case study in an emergency department. *Computers & Operations Research*, 51:99–110, 2014.

[42] Chongjun Yan, Jiafu Tang, Bowen Jiang, and Richard Fung. Sequential appointment scheduling considering patient choice and service fairness. *International Journal of Production Research*, 53(24):7376–7395, 2015.

[43] Christos Zacharias and Michael Pinedo. Appointment scheduling with no-shows and overbooking. *Production and Operations Management*, 23 (5):788–801, 2014.

[44] Bo Zhang, Pavankumar Murali, M. Dessouky, and David Belson. A mixed integer programming approach for allocating operating room capacity. *Journal of the Operational Research Society*, 60(5):663–673, 2009.

[45] Han Zhu, Youhua Chen, Eman Leung, and Xing Liu. Outpatient appointment scheduling with unpunctual patients. *International Journal of Production Research*, 56(5):1982–2002, 2018.

[46] Maartje E. Zonderland. *Appointment planning in outpatient clinics and diagnostic facilities*. SpringerBriefs in health care management and economics. Springer, New York, 2014.