Diss. ETH No. 17434

Measuring the Perceived Quality of Parameterized Multimedia Algorithms

A dissertation submitted to the
Swiss Federal Institute of Technology Zurich

for the degree of
Doctor of Sciences

presented by
Eva-Maria Schuberth
Diplom-Informatikerin, Universität Passau
born 21.07.1979
citizen of Germany

accepted on the recommendation of
Prof. Dr. Emo Welzl, examiner
Dr. Joachim Giesen, co-examiner
Prof. Dr. Klaus Simon, co-examiner

2007

To my parents

Contents

Abstract

Many multimedia algorithms have in common that their quality not only depends on objective evaluation criteria like running time or memory consumption but also on the human perception of the output. However, a systematic analysis of the perceived quality in user studies is often missing. Reasons for this shortcoming are that there are often no standardized methods to conduct and evaluate user studies and, even if there are methods, these often come to their limits as soon as the number of algorithms to be compared is large. In practical applications the number of algorithms to be compared indeed can be very large, so that straight forward evaluation methods cannot be applied anymore. A large number of algorithms especially arises in the context of parameterized algorithms where each parameter setting can be seen as an algorithm of its own. For parameterized algorithms the goal is to compare the perceived quality induced by different parameter settings. In this thesis we develop methods to measure and compare the perceived quality induced by the different parameter settings of a parameterized multimedia algorithm. We study the problem in the broader context of choice based conjoint analysis.

Choice based conjoint analysis is an instrument in market research to measure a respondent's preferences on a set of items. It deals with preferences on item sets that possess a conjoint structure. A set of items possesses a conjoint structure if it can be described in terms of attributes and attribute levels. In choice based conjoint analysis preferences are elicited in a sequence of choice tasks. In a choice task a small number of items is presented to a respondent who then is asked to choose the one that he prefers. The crucial observation is that a parameterized algorithm can be viewed as a set of items with conjoint structure where the parameters correspond to the attributes and the parameter values to the attribute levels. This allows to measure the quality of different

parameter settings of a multimedia algorithm using choice based conjoint analysis.

The results of the thesis are twofold: on the one hand we study the problem of choice based conjoint analysis and develop our own framework for eliciting respondents' preferences on a set of items with conjoint structure. On the other hand we are to our knowledge the first to apply choice based conjoint analysis to measure the perceived quality of parameterized multimedia algorithms.

The thesis is divided into two parts: in the first part we investigate the combinatorics of choice based conjoint analysis. We derive lower bounds for the number of choice tasks a respondent has to perform in order to derive a (preference) ranking of the items from his choices. There is no query strategy that can derive an item ranking from only polynomially many choice tasks (polynomial in the number of attributes and the number of levels). Therefore we take an approximation approach, i.e. we pursue the question of how many choice tasks are necessary to derive a respondent's approximate ranking of the items. Finally we develop our own framework to process data elicited in choice based conjoint studies which is based on the introduction of new modeling assumptions and the aggregation of preferences for different respondents.

In the second part of the thesis we use our framework to measure the perceived quality of algorithms from two different applications: volume visualization and gamut mapping. Both areas have in common that the output of the algorithms are images. We report on a large user study that we conducted to measure the perceived quality of the different parameter settings for an existing parameterized volume visualization algorithm. Furthermore we develop a new gamut mapping algorithm and compare it in a user study to standard reference algorithms. The user study shows that with our algorithm we are able to improve the perceived quality over the standard reference algorithms.

Zusammenfassung

Die Qualität von Multimedia Algorithmen wird meist mit Hilfe von traditionellen Bewertungskriterien wie Laufzeit oder Speicherbedarf der Algorithmen gemessen. Ein wichtiger Aspekt wird dabei oft vernachlässigt: die vom Menschen wahrgenommene Qualität des berechneten Ergebnisses. Leider wird ein systematischer Vergleich verschiedener Algorithmen im Hinblick auf die wahrgenommene Qualität nur selten durchgeführt. Für dieses Versäumnis gibt es mehrere Gründe: In vielen Gebieten gibt es keine standardisierten Methoden, um Benutzerstudien durchzuführen und auszuwerten. Selbst wenn es solche Methoden gibt, so stossen diese an ihre Grenzen sobald die Anzahl der zu vergleichenden Algorithmen gross ist. Speziell im Zusammenhang mit parametrierten Algorithmen führt dies zu Problemen. Hier will man unterschiedliche Parametereinstellungen im Hinblick auf die wahrgenommene Qualität der Ergebnisse vergleichen. In praktischen Anwendungen kann die Anzahl der Parametereinstellungen allerdings sehr gross werden, so dass einfache Methoden zur Qualitätsmessung nicht angewendet werden können. Deshalb entwickeln wir in dieser Arbeit neue Methoden, um die wahrgenommene Qualität eines parametrierten Multimedia Algorithmus für verschiedene Parametereinstellungen zu vergleichen. Dazu betrachten wir das Problem der Qualitätsmessung im breiteren Kontext der wahlbasierten Conjoint-Analyse.

Die wahlbasierte Conjoint-Analyse ist ein Instrument der Marktforschung um die Präferenzen eines Befragten bezügliche einer Menge von Produkten zu messen. Dabei besitzen die Produktmengen eine ganz spezielle Struktur: die Conjoint-Struktur. Eine Menge von Produkten besitzt eine Conjoint Struktur, wenn sie mit Hilfe von Attributen und Attribut-Ausprägungen beschrieben werden kann. In der wahlbasierten Conjoint-Analyse werden dem Befragten nacheinander kleine Mengen von Produkten gezeigt und der Befragte wird gebeten, sein bevorzugtes Produkt auszuwählen. Wir haben die wichtige Beobachtung gemacht, dass

ein parametrierter Algorithmus nichts anderes als eine Menge von Produkten mit Conjoint-Struktur ist. Die Parameter entsprechen dabei den Attributen und die Parameter-Werte den Attribut-Ausprägungen. Diese Beobachtung erlaubt es nun, die wahrgenommene Qualität von parametrierten Multimedia-Algorithmen mit Hilfe von wahlbasierter Conjoint-Analyse zu messen.

Die Ergebnisse dieser Arbeit sind zweischichtig: zum einen untersuchen wir wahlbasierte Conjoint-Analyse und entwickeln ein eigenes Framework, das es erlaubt, die Präferenzen von Befragten bezüglich einer Produktmenge mit Conjoint-Struktur zu bestimmen. Zum anderen sind wir unseres Wissens die Ersten, die wahlbasierte Conjoint Analyse auf die Messung der wahrgenommenen Qualität von parametrierten Multimedia-Algorithmen anwenden.

Diese Arbeit gliedert sich in zwei Teile: Im ersten Teil untersuchen wir die Kombinatorik von wahlbasierter Conjoint-Analyse. Wir bestimmen untere Schranken für die Anzahl von Fragen, die einem Befragten gestellt werden müssen, um von seinen Antworten eine Präferenzordnung der Produkte abzuleiten. Es gibt keine Fragestrategie, die eine solche Ordnung mit nur polynomiell vielen Fragen bestimmen könnte (polynomiell in der Anzahl der Attribute und der Attribut-Ausprägungen). Deshalb verfolgen wir einen Approximationsansatz: wir bestimmen die Anzahl der Fragen die nötig sind, um die Präferenzordnung approximativ zu bestimmen. Schliesslich entwickeln wir ein Framework, das es uns erlaubt, die wahrgenommene Qualität von parametrierten Multimedia-Algorithmen zu messen. Dieses Framework stützt sich auf zusätzliche Modellannahmen und die Aggregation von Präferenzinformation für verschiedene Befragte.

Im zweiten Teil wenden wir unser Framework an, um die wahrgenommene Qualität von Algorithmen aus den Bereichen Volumen Visualisierung und Gamut Mapping zu messen. Die beiden Gebiete verbindet, dass die Ausgabe der Algorithmen Bilder sind. Wir führen eine grosse Benutzerstudie durch, um die wahrgenommene Qualität verschiedener Parametereinstellungen eines existierenden Volumen Visualisierungsalgorithmus zu messen. Ausserdem entwickeln wir einen neuen Gamut Mapping Algorithmus und vergleichen seine Qualität mit standardisierten Referenzalgorithmen. Eine Benutzerstudie zeigt, dass unser Algorithmus deutlich bessere Bilder erzeugt als die Referenzalgorithmen.

Acknowledgments

I would like to thank all people who contributed to this work. First and foremost I would like to express my deep gratitude to my supervisor Joachim Giesen. With his unmistakable sense of discovering new and interesting research fields he made this thesis possible. He was the one to introduce me to the fascinating field of preference analysis, he was there for endless hours of discussions and always had a valuable advice for me when I needed some. I am much obliged for what I have learned from him.

I am sincerely grateful to Emo Welzl for giving me the opportunity to become a member of GREMO, for always encouraging me in my work and for always standing behind me. Moreover, I thank Klaus Simon for giving me the opportunity to work on the project about gamut mapping and for being my co-examiner.

I am also indebted to Peter Zolliker for working together with me throughout my whole PhD time and for introducing me to color science. In particular his insights into applied statistics made the chapter about modeling and aggregation possible.

I would like to thank my co-authors Joachim Giesen, Dieter Mitsche, Klaus Mueller, Miloš Stojaković, Klaus Simon, Lujin Wang, Oliver Zweifel, Daniel Zeiter and Peter Zolliker. It was a great pleasure working with them. I am also indebted to Jiří Matoušek for fruitful discussions at the GWOP workshops that in particular made the chapter about approximate sorting possible.

It was a great pleasure to share my PhD time with the current and former members of GREMO, Robert Berke, Yves Brise, Tobias Christ, Kaspar Fischer, Bernd Gärtner, Heidi Gebauer, Franziska Hefti-Widmer, Michael Hoffmann, Martin Jaggi, Shankar Ram Lakshminarayanan, Dieter Mitsche, Yoshio Okamoto, Andreas Razen, Leo Rüst, Andrea Salow, Dominik Scheder, Ingo Schurr, Shakhar Smorodinsky, Miloš Stojaković,

Marec Sulovský, Tibor Szabó, Patrick Traxler, Floris Tschurr, Elias Vicari, Uli Wagner, Frans Wessendorp and Philipp Zumstein.

My special thanks go to Philipp, Miloš and Dieter for sharing the office with me, Floris, Franziska and Andrea for resolving all kinds of administrative problems, Leo for his good advice concerning the thesis writing, Miloš, Robert and Ingo for coming with me to many concerts, Andreas and Dominik for teaching my classes when I was traveling, Michael, Marek, Yves, Martin and Andreas for giving me valuable feedback for my defense talk. Many thanks also to Ruedi Arnold for co-heading the forum for women in computer science with me.

Last but not least I am deeply grateful to my parents Rudolf and Franziska, my brother Frank and my boyfriend Ulli for their love and invaluable support they gave me during the last years and for constantly encouraging me on my way to here.

CHAPTER 1

Introduction

1.1. Motivation

The work on this thesis started with a project on image-dependent gamut mapping. Gamut mapping is a fundamental task in digital color reproduction. It is concerned with adapting an image to the limitations of a color space. For example, the set of colors that can be reproduced by a printer can be different from the set of colors contained in an image. During the printing process the non-reproducible colors have to be replaced by reproducible colors. Such a replacement strategy is called a gamut mapping algorithm. If the gamut mapping algorithm depends on the properties of the input image it is called image-dependent. The main design goal in gamut mapping is to change the color appearance of the image as little as possible during the adaptation. Therefore the evaluation of a gamut mapping algorithm should not only take traditional performance measures for algorithms like running time or memory consumption into account but also the degree to which the goal of color appearance preservation is satisfied. The latter can only be judged by humans. Thus we conducted a field study to measure the quality of a gamut mapping algorithm that we developed in comparison to standard reference algorithms.

We soon realized that it is true for many multi-media algorithms, i.e., algorithms whose output is an image, a video or an audio file, that their performance needs human evaluation. However, in practice the number of algorithms to be compared can be very large, so that straight forward quality measurement methods cannot be applied. Therefore we looked for alternative approaches to measure the perceived quality of multimedia algorithms. Another field where perceived quality and human evaluation plays a major role is market research. So it was quite natural to look into the market research literature to see if some of the methods there can be adapted to our needs in evaluating multi-media

algorithms. One first finding of our literature search was that choice tasks are very popular means to elicit preferences in a market research study. In a choice task a very small number of items (say up to five) is presented to a respondent, who then has to indicate which one of these he is most likely to buy. Since choice tasks simulate to a certain extent real buying situations they tend to provide quite reliable data. Choice based preference analysis — which got its name from the preference elicitation procedure via choice tasks — is very similar in nature to comparison based sorting, one of the most classical subjects in algorithms. Assume that the choice task is to choose only among two items, i.e., to perform paired item comparisons. If we want to assess a respondent's ranking of the items, then we have essentially to run a comparison based sorting algorithm where the comparisons are done by the respondent. This already shows one of the key problems people in market research have to face: in order to derive the ranking a lot of comparisons are necessary. If n items have to be ranked, then the information theoretic lower bound on comparisons based sorting tells us that at least $n \log n$ comparisons are needed. However, respondents often get worn out after a very small number of choice tasks and do not answer further questions faithfully anymore. This problem is mitigated in market research practice by giving to each respondent only a small number of choice tasks and aggregating information from many respondents.

The most important idea that we took from the market research literature is that sets of items, e.g., cars, detergents or laptop bags often come with a structure, namely, they can be described in terms of attributes and attribute levels. Such a structure is called a conjoint structure and preference analysis involving items with conjoint structure is called conjoint analysis. We realized that multi-media algorithms often also have a conjoint structure where the attributes are the parameters of the algorithms and the levels are the parameter values. One example of a multi-media algorithm with conjoint structure that we examined in detail is a volume visualization algorithm.

In marketing researchers often want finer information than just a ranking: instead of measuring the differences on an *ordinal* scale as for a ranking they measure the differences on an *interval* scale. To measure preferences on an interval scale completely different methods than for combinatorial (choice based) sorting algorithms need to be developed.

1.2. Contributions

Our contributions are twofold: First, we contribute to the theoretical underpinnings of choice based conjoint analysis by studying its combinatorial structure and devising new data analysis methods based on approximation, modeling and aggregation ideas. In particular we investigate a popular choice base conjoint analysis technique and raise some doubts about its reliability. Second, we use our insights into preference measurement to measure the perceived quality of two multimedia algorithms: On the one hand we develop our own framework for image-dependent gamut mapping and measure the influence of the parameter image-dependence on the quality of the result. On the other hand we measure the perceived quality induced by different parameter settings of an existing volume visualization algorithm. The outline of the thesis is as follows:

In Chapter 2 we introduce conjoint analysis in more detail. In an attempt to understand the good practical performance of a very popular volume-based approach to choice based conjoint analysis we investigate the combinatorics of choice based conjoint analysis in Chapter 3. We pursue the question of how many choice tasks a respondent needs to perform in order to derive an item ranking from his choices and compare the problem of sorting a set of items possessing a conjoint structure (conjoint structure case) with the problem of sorting a set of items without any additional structure (structureless case). We give a geometric interpretation which for both problems leads to a hyperplane arrangement. It is known (see for example [**Mat02**]) that in the structureless case the different cells in the hyperplane arrangement have the same combinatorial and geometric structure. From this observation an efficient volume-based query strategy can be derived. However, we can show that in the conjoint structure case the different cells have a different combinatorial and thus a different geometric structure. This has implications for the volume-based approach and also for commonly applied methods to validate choice based conjoint analysis algorithms. These methods rely heavily on insights from the geometric interpretation of the structureless case and try to carry these insights over to the conjoint structure case. Our findings, however, raise some doubts about the reliability and applicability of these methods.

There is no algorithm that can derive an item ranking in the conjoint structure case from only polynomially many choice tasks (polynomial in

the number of attributes and the number of levels). At the same time respondents are only willing to answer a very small number of choice tasks. Therefore we strive for alternative approaches that permit eliciting people's preferences by only confronting them with a small number of choice tasks. Basically we see three possible alternative approaches which are, in their applicability, not mutually exclusive: First, instead of heading for an exact item ranking for an individual person, one could determine an approximation of the ranking. Second, instead of finding an item ranking for every person individually, one could aggregate preference information and find an item ranking for the whole population. Third, one could introduce additional modeling assumptions. In this thesis we pursue all three approaches.

In Chapter 4 we investigate the approximation approach for the structureless case, where we are given n items without any additional (conjoint) structure. We pursue the question of how many choice tasks are necessary to derive a respondent's approximate ranking of the items when preference information is elicited from paired comparisons. We show that in order to obtain a ranking at Spearman's footrule distance $n^2/\nu(n)$ to the respondent's item ranking with any query strategy, in general at least

$$n\,(\min\{\log \nu(n), \log n\} - 6)$$

choice tasks have to be performed in the worst case.

In Chapter 5 we introduce new modeling assumptions and use the aggregation approach. We develop a new framework to derive an aggregated interval scale for the whole population for a set of items with conjoint structure. Our framework is based on Thurstone's method of comparative judgment [**Thu27**], which was designed for the case when the items have no additional structure. We extend this method to the case when items possess a conjoint structure. We can test all modeling assumptions of our framework and therefore the appropriateness of its application. In the second part of the thesis, the application part, we put our framework to work in practical applications.

In Chapter 6 we present a framework for image-dependent gamut mapping which is based on an optimization approach. We investigate the influence of image-dependence on the perceived quality of gamut mapping algorithms. We compare the developed algorithm to standard reference algorithms in a user study. We evaluate the user study by using Thurstone's method. The user study shows that image-dependence is

indeed an important parameter and that with our algorithm we are able to improve the perceived quality over the standard reference algorithms.

In Chapter 7 we measure the perceived quality of an existing parameterized volume visualization algorithm. In volume visualization one is concerned with creating a two-dimensional representation (in the form of an image) of three dimensional volume density data. Such a representation is by no means unique. The main objective of volume visualization is to produce images that allow humans to gain more insight into the data. The algorithm that we consider possesses a conjoint structure and therefore we apply the framework that we developed in Chapter 5. We describe the design of a large user study that we conducted to test the influence of the different parameters of the algorithm. The purpose of this study was twofold: On the one hand we gained valuable insights into quality perception by users in volume visualization, and on the other hand we were able to put our conjoint framework to test and learned about how to improve the user study design in general.

CHAPTER 2

Basics of Conjoint Analysis

Conjoint Analysis originated in Mathematical Psychology and is now used in many applied sciences like market research as an instrument to measure people's preferences on a set of items. Conjoint analysis has its name from the structure of the items that are considered — the conjoint structure.

Conjoint Structure. A class of items possesses a *conjoint structure* if it can be described in terms of attributes A_i and attribute levels $a_{ij} \in A_i$. That is, we consider sets of items whose profile is given as an element in $A = A_1 \times \ldots \times A_n$, where the attribute set A_i has the levels $A_i = \{a_{i1}, \ldots, a_{ik_i}\}$. An item profile $a \in A$ is just a vector $(a_{1j_1}, \ldots, a_{nj_n})$ of attribute levels $a_{ij_i} \in A_i$.

Example 2.1. A set of cars can be described by a conjoint structure with the four attributes Color, Number of Seats, Cargo Area and Price, where

- Color = {black, green, blue, red}
- Number of Seats = {5, 7}
- Cargo Area= {small, medium, large}
- Price= {20000CHF, 30000CHF, 40000CHF}

A car profile is then given as a vector in Color×Number of Seats× Cargo Area×Price, like for example (blue, 5, medium, 40000CHF)

Example 2.2. The set of all possible parameter settings of a parameterized algorithm can be described by a conjoint structure where the attribute sets are just the parameter sets and the attribute levels are the parameter values. In Chapter 7 and Chapter 6 we will investigate (parameterized) multi-media algorithms in the framework of conjoint analysis.

Conjoint analysis is concerned with the task to find a scale, i.e. a function

$$s : A \to \mathbb{R}$$

that assigns to each item a number that represents its value. Conjoint analysis techniques can be distinguished by two (not independent) parameters

(1) The elicitation procedure, i.e., the way how preference data are obtained from respondents
(2) The data processing procedure, i.e. the way how the scale is defined. The data processing procedure usually is based on modeling assumptions. Modeling assumptions are necessary since in general the elicited data tends to be very sparse and can be interpreted meaningfully only in the context of some model, which already encodes general assumptions on the structure of the preferences.

Elicitation procedure. Preference information can be elicited in many different ways. The most direct approach is to ask respondents to state their valuation of a given item, for example in terms of money. In recent years choice based conjoint analysis has become the most popular conjoint analysis technique. It got its name from the employed elicitation procedure, namely preferences are elicited in a more indirect way using a sequence of *discrete choice tasks*. A choice task consists of a small number of item profiles—typically between two and four—presented to a respondent, who has to state which one he prefers or is most likely to buy (often also a *none* choice option is included). Preference elicitation in the form of discrete choice tasks has two advantages, the cognitive burden on the respondent in each task is comparatively low, and choice tasks also simulate to a certain extent real buying situations. One drawback of data elicitation with choice tasks is, that in order to derive a scale from the choices many choice tasks are necessary in general. *Adaptive* elicitation methods deal with this problem by trying to choose the next choice task dependent on the choices in previous tasks such that the information gained in the worst case outcome of the choice task is approximately maximized. Information gain maximization in the worst case means that the minimum of the information gain among all possible outcomes of the choice task is maximized. The rules according to which the choice tasks are chosen are called the query strategy. We will measure the effectiveness of a query strategy in terms of the number of choice tasks that are performed in the query strategy in order to derive the desired scale s.

Data processing procedure. There are many different ways to analyze the obtained preference data, however, any processing procedure defines a scale that assigns a number to each item. The assigned numbers should faithfully represent the preference structure of either a single respondent or of a group of respondents. When a scale is defined for every respondent separately we call the scales individual scales whereas when only one scale is defined for a group of respondents we call the scale an aggregated scale. Furthermore, data processing procedures also differ in the type of scale they define. Stevens [**Ste46**] identified four types of scales which differ in the nature of information that can be derived from them. Each type can be characterized by a set of permissible transformations. A permissible transformation is a function $f : \mathbb{R} \to \mathbb{R}$ such that a scale s and the transformed scale $s' = f \circ s$ reveal the same nature of information about the measured data. In the context of conjoint analysis two types are of particular interest: ordinal scales and interval scales. On an *ordinal scale* the number assigned to an item represents its rank in a linear order. The permissible transformations of an ordinal scale are all strictly increasing functions. These functions preserve the underlying linear order of the items. On an ordinal scale the nominal difference between the assigned numbers has no meaning. For example one can only state that the first ranked item is preferred over the third and the 50'th ranked item, but not that the first and the third are much closer in value than the third and the 50'th ranked item.

On an *interval scale* the ratio of differences of the assigned numbers have a meaning. The permissible transformations of an interval scale are all affine functions of the form $x \to \alpha x + \beta$ where $\alpha, \beta \in \mathbb{R}$ and $\alpha > 0$, i.e. an interval scale can be translated and scaled by a positive factor. From an interval scale one can for example conclude if the first and the third or the third and 50'th ranked item are closer by looking at the ratio of the distances of their assigned numbers. If the distance between the third and 50'th ranked item turns out to be smaller than between the first and the third, then the first pair is closer then the second. But note that on an interval scale ratios of scale values themselves have no meaning since there is no *natural* zero.

Example 2.3. Temperature measured in Celsius defines an interval scale. Assume that we measure the temperature in the morning to be 10°C, in the afternoon to be 20°C and at night to be 5°C. As on a Celsius scale the zero point is arbitrarily chosen the statement that the temperature in the afternoon was twice as large as in the morning does not make

sense. However, to say that the temperature difference between afternoon and morning(10°C) is twice as large as the temperature difference between morning and night (5°C) is a valid statement. Multiplying all temperature measurements by a positive factor α and translating it by a real-value β does not change the nature of information that we can derive. For example we can make the same statement about temperature differences for temperatures given in Fahrenheit.

Modeling. Depending on a conjoint study's objectives different modeling assumptions need to be made in order to derive meaningful information from the elicited sparse data. A typical assumption that we will make throughout the thesis is the *additivity assumption* or *linear model assumption*. It is assumed that a respondent's (or a respondent group's) preference structure can be represented by an additive function, i.e. that for every attribute level a_{ij} there is a partworth $\lambda_{ij} \in \mathbb{R}$ and the scale value of an item $a = (a_{1j_1}, \ldots, a_{nj_n})$ can be computed as the sum of the partworths for all levels that describe the item:

$$s(a) = \sum_{i=1}^{n} \lambda_{ij_i}.$$

The data processing task then is to compute the partworths λ_{ij} from the choice data elicited from the respondent. The vector whose entries are the partworths is called the partworth vector. Note that an additive model really makes an assumption on the structure of the respondent's preferences. Every additive scale induces a ranking of the set of items, however, there are rankings that cannot be induced by an additive scale. Therefore the additivity assumption essentially reduces the number of item rankings that are considered. In case that the attributes A_i are continuous sets, i.e., real intervals, this assumption can be made more explicit, namely, an additive scale exists if and only if the preference structure of the respondent is such that the attributes are mutually preferentially independent, see Keeney and Raiffa [**KR93**] for details.

Part 1

Theoretical Foundations

CHAPTER 3

Combinatorics of Choice Based Conjoint Analysis

3.1. Introduction

In this chapter we investigate choice based conjoint analysis from a combinatorial point of view. We will pursue the question of how many choice tasks a respondent needs to perform in order to derive an ordinal scale from his choices. Moreover we will introduce choice based polyhedral conjoint analysis, a celebrated conjoint analysis technique developed at MIT [**THS04**] and raise some doubts about its reliability. The setup for this chapter is the following: Given is a set of items described by a conjoint structure $A = A_1 \times \ldots \times A_n$ with attributes $A_i = \{a_{i1}, \ldots, a_{im}\}$. Note that in order to keep the exposition simple we assume that every attribute has exactly m levels. The goal is to derive an ordinal scale for the items in A that represents an individual respondent's preference structure. We assume that a respondent's preference structure can be represented by an additive scale, i.e. his preferences obey the linear model. That is, the goal becomes to find a partworth value λ_{ij} for every attribute level a_{ij} such that the scale value for an item can be computed as sum of the partworths of the levels that describe this item.

As elicitation procedure we use paired comparisons, i.e., the respondent performs a sequence of choice tasks with each choice task involving two items. Each choice task can be chosen in dependency of the choices in previous choice tasks.

A conjoint structure is quite intricate once one has a closer look. To make this more explicit we will compare choice based conjoint analysis with choice based preference analysis on product sets without any structure. In the latter case the item set A is just a finite set $\{a_1, ..., a_k\}$ of items that does not possess any additional structure. We refer to that case as the structureless case as opposed to the conjoint structure case. The

structureless case can be seen a special case of the conjoint structure case where the number n of attributes is 1. The goal in the structureless case is to assign values μ_i to the items a_i which represent the underlying preference ranking of the respondent.

In this chapter we will interpret both, the conjoint structure case and the structureless case geometrically. For both cases we will derive lower bounds on the number of choice tasks that a respondent needs to perform in the worst case in order to determine an ordinal scale describing his preferences. In the structureless case the geometric interpretation leads to a hyperplane arrangement where all cells have the same geometric structure. This permits an efficient volume-based query strategy. Choice based polyhedral conjoint analysis [**THS04**] tries to carry over insights from the structureless case to the conjoint structure case. In particular it implicitly assumes that in the conjoint structure case all cells in the induced hyperplane arrangement have the same geometric structure. Moreover, even the method to validate polyhedral choice based conjoint relies heavily upon this assumption. However, we were able to show that this implicit assumption is not true, i.e. in the conjoint structure case different cells in the hyperplane arrangement can have a different combinatorial and thus a different geometric structure. These results raise serious doubts about the reliability and applicability of polyhedral choice based conjoint analysis and the validation method.

3.2. Geometric interpretation

Let us first of all give a geometric interpretation of choice based conjoint analysis for the conjoint structure case. A respondent performs a sequence of paired comparisons and we want to derive partworths λ_{ij} that induce his item ranking. We can assume without loss of generality that every partworth λ_{ij} is a real number in the interval $[-1, 1]$. A valid fit of the partworths is a point in the cube $[-1, 1]^{nm}$. Of course a valid fit needs not to be a good fit. A good fit would allow to predict correctly the outcome of a choice task not encountered so far with high probability. If a respondent prefers item $p = (a_{1j_1}, \ldots, a_{nj_n})$ over item $q = (a_{1l_1}, \ldots, a_{nl_n})$ in a choice task, then the parameters of the linear model—the partworths—have to satisfy the following constraint

$$\sum_{i=1}^{n} \lambda_{ij_i} \geq \sum_{i=1}^{n} \lambda_{il_i} \Leftrightarrow \sum_{i=1}^{n} \lambda_{ij_i} - \sum_{i=1}^{n} \lambda_{il_i} \geq 0.$$

This constraint can be rewritten by using the inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^{nm}$ as

$$\langle h, \lambda \rangle \geq 0,$$

where λ is the partworth vector and h is a vector in $\{-1, 0, 1\}^{nm}$ whose entries are 1 at positions ij_i and -1 at positions il_i for all $i = 1, \ldots, n$ where $ij_i \neq il_i$. All other entries in h are 0. This inequality defines a halfspace $H \subseteq \mathbb{R}^{nm}$,

$$H = \{\lambda \in \mathbb{R}^{nm} | \langle h, \lambda \rangle \geq 0\},$$

whose boundary, a hyperplane, contains the origin and has the inward pointing normal vector h. That is, all partworth vectors that are compliant with the comparison between p and q are contained in the intersection of $[-1, 1]^{nm}$ with the halfspace H. Every possible item comparison—there are only finitely many—leads to a hyperplane (bounding the corresponding halfspace). The arrangement of all these hyperplanes subdivides the cube $[-1, 1]^{nm}$ into polyhedral cells. All points in the same cell of the arrangement encode exactly the same ranking of all items with profile in $A_1 \times \ldots \times A_n$, i.e., any such point is a perfect fit for the ranking. Note however, that not all rankings of $A_1 \times \ldots \times A_n$ can be expressed by a partworth vector, i.e. in the linear model. But once we assume that a respondent's preference structure can be represented by a partworth vector, the goal in choice based conjoint analysis becomes to set up the choice tasks such that one can identify a perfect (good) partworth vector from the respondent's choices in as few choice tasks as possible. The task to identify a perfect partworth vector from choice tasks is equivalent to identifying a cell in the hyperplane arrangement. Any vector in this cell can serve as a perfect fit for the linear model. The information provided in a sequence of choice tasks is in geometric terms a nested sequence of polyhedra: The first polyhedron is just the cube $[-1, 1]^{nm}$. Each choice task provides one or more halfspaces (a one out of k choice task provides $k - 1$ halfspaces). The common intersection of $[-1, 1]^{nm}$ with all the halfspaces provided up to a given choice task represents the information gathered up to that task. Two questions arise naturally:

(1) How can one find the the cell that encodes the preference ranking efficiently, i.e., using as few choice tasks as possible.
(2) How to choose a 'good' representative point out of the unique cell once it has been identified.

Polyhedral choice based conjoint analysis, a heuristic introduced by Hauser, Toubia and Simester [**THS04**] is based on this geometric interpretation. The answers to the above questions are intertwined in this approach: a deep point inside the cell, the so called analytic center of the polyhedral cell, is considered to be a good representative point and a comparison whose corresponding hyperplane is close to the analytic center and cuts the cell into two polyhedra of almost equal volume is considered to be a good next comparison. The intuition for this approach comes from the structureless case where there exists a volume-based efficient query strategy. We will show however, that one cannot directly carry over insights from the structureless case to the conjoint structure case.

Before we investigate the structureless case and the volume based query strategy we want to make one further remark on the equivalence of finding a ranking of elements in A and finding an ordinal scale for the elements in A. Assume that the preference structure of a respondent can be represented by an additive scale s and remember that in our computational model only item comparisons count that are performed by the respondent. Then the following two problems are equivalent:

(1) Derive the respondent's ranking of the elements in A from his choices
(2) Derive a partworth vector λ that describes the respondent's preferences from his choices.

Given a partworth vector λ we can compute a scale value for every item in A by summing up the partworths for the levels that describe the item. Then we can sort the elements in A which gives us the respondent's ranking. This all does not require further comparisons of items by the respondent. Conversely, given the respondent's ranking of elements in A we can find the common intersection of the halfspaces induced by the ranking and choose an arbitrary point in this intersection as the partworth vector. This again does not require further comparisons of items in A.

3.3. The structureless case

We also have a purely geometric interpretation in the structureless case, where we want to find a value μ_i for an item a_i for $i = 1, \ldots, k$. If we again assume that the values μ_i are contained in the interval $[-1, 1]$, then any point in the cube $[-1, 1]^k$ is a valid fit as long as we do not

have any information about the ranking of the k different items. Information is gathered by item comparisons. A comparison of the i'th and the j'th item tells us which one of μ_i and μ_j is larger. In geometric terms this means, the comparison tells us on which side of the hyperplane passing through the origin and with normal $h \in \{-1, 0, 1\}^k$ the vector $s = (\mu_1, \ldots, \mu_k)$ has to lie. The normal h has entry 1 at the i'th position, -1 at the j'th position and is 0 at all other positions. There are $\binom{k}{2}$ comparisons and thus $\binom{k}{2}$ hyperplanes. The arrangement of all these hyperplanes subdivides the the cube $[-1, 1]^k$ into cells. Each cell of this arrangement corresponds to one of the $k!$ rankings of the μ_i. Hence determining the ranking of the k items from item comparisons is equivalent to determining a cell in the hyperplane arrangement. In Figure 3.3 we show two views on the subdivision of $[-1, 1]^3$ by the comparison hyperplanes for the case $k = 3$.

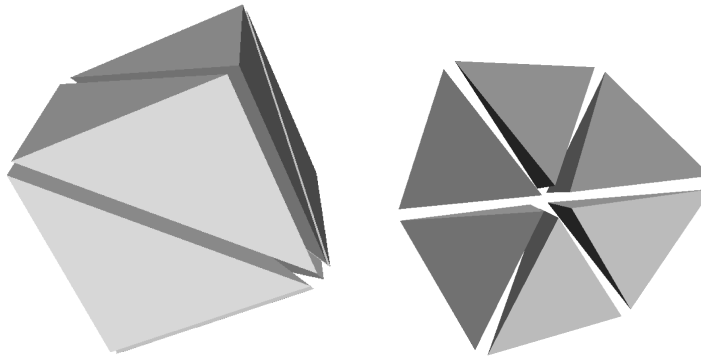

FIGURE 3.1. Subdivision of $[-1, 1]^3$ into the six cells that correspond to the six possible rankings of three items.

3.3.1. Information theoretic lower bound. We assume that we assess a respondent's preference ranking of k items from paired comparisons, i.e., from a sequence of choice tasks that each involve only two items. A query strategy to find the item ranking is nothing else but a comparison based sorting algorithm. Here we want to state the information theoretic lower bound in sorting, see [**Knu73**]. It gives an answer

to the question of how many item comparisons are needed in the worst case in order to infer an item ranking. For simplicity assume that no two items have the same value, i.e. for items p and q either p is strictly preferred over q or q is strictly preferred over p. In the information theoretic argument the notion of decision tree plays a central role. A decision tree for ranking k items represents a query strategy by describing all possible sequences of comparisons between items needed to sort the items. In the end, each of these sequences leads to a different ranking of the items. Assume that we want to rank three items p_1, p_2, p_3 according to their values μ_1, μ_2, μ_3. In Figure 3.3.1 we show one possible decision tree, i.e., query strategy for that problem.

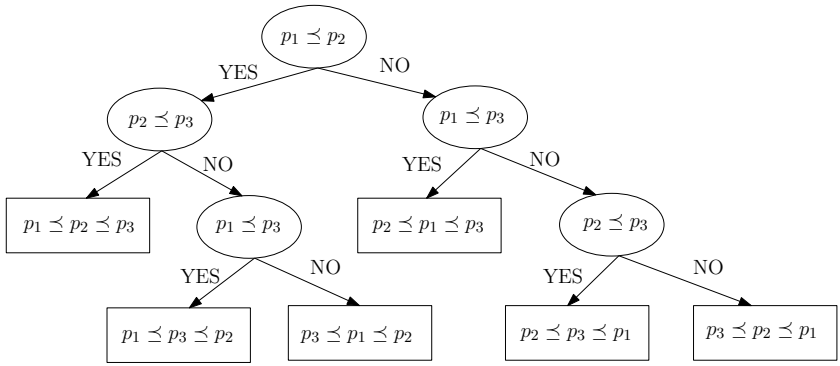

FIGURE 3.2. A decision tree for ranking three items

Every inner node—oval in Figure 3.3.1—represents an item comparison. Depending on the outcome of the comparison the subsequent comparison is determined. After a sequence of comparisons one ends up in a leaf of the tree—rectangular node in Figure 3.3.1. This leaf corresponds to the item ranking compliant with the sequence of comparison outcomes. For a particular item ranking the number of inner nodes on the path from the root to the corresponding leaf is just the number of comparisons that is needed for that particular query strategy. The height of a tree is the maximum number of inner nodes that are visited when traversing the tree from the root to its leaves. In the example of Figure 3.3.1 the height is 3. The height of a decision tree is just the number of comparisons that the query strategy needs in the worst case. We know that every decision tree for the sorting problem has $k!$ leaves corresponding to the $k!$ possible rankings. The height of any binary tree on $k!$ leaves

is at least $\log_2(k!)$. Using Stirling's approximation this number can be lower bounded by $k\log_2(\frac{k}{e})$ which is in $\Omega(k\log(k))$. Therefore any query strategy for determining a ranking of k items needs at least $k\log_2(\frac{k}{e})$ comparisons in the worst case. This bound is called the information theoretic lower bound in sorting.

In the following section we will show that the information theoretic lower bound can be reached, i.e., there is a query strategy which can determine any ranking of k items with no more than $ck\log_2(k)$ comparisons, where c is a constant independent of k. Later we will see that the same does not hold for the case that the item profiles are described by a conjoint structure. In that case there is no algorithm that can reach the information theoretic lower bound.

3.3.2. Volume cuts. There are many query strategies known that reach the information theoretic lower bound up to a constant, i.e., query strategies that always infer the ranking of k items with at most $ck\log_2(k)$ comparisons for some constant c. Here we describe such a strategy which is based on the geometric interpretation of the problem and the following two observations.

Observation 1. [**Mat02**] *Assume a respondent has performed already i comparisons. Let P_i be the set of rankings of the items that are compatible with his answers. Any query strategy that always reduces the set P_i for all i by at least a constant fraction δ reaches the information theoretic lower bound. Ideally δ would be $1/2$, but it can be shown that this is not always possible.*

Observation 2. [**Mat02**] *Every cell in the subdivision of $[-1,1]^k$ by the hyperplanes corresponding to paired item comparisons is a simplex with volume $2^k/k!$.*

In geometric terms a sequence of comparisons provides us with a nested sequence of polyhedra. The polyhedron at the beginning when no comparison was performed is just $[-1,1]^k$. See Figure 3.3.2 for an example of such a nested sequence in the case $k=3$. The volume of each polyhedron multiplied with $k!/2^k$ gives us exactly the number of rankings that are compatible with the comparisons performed so far. More importantly, any comparison whose corresponding hyperplane in the worse of the two outcomes of the comparison cuts a δ-fraction of the volume of the polyhedron, also cuts a δ-fraction of the rankings compatible with the comparisons so far. An application of the famous Brunn-Minkowski

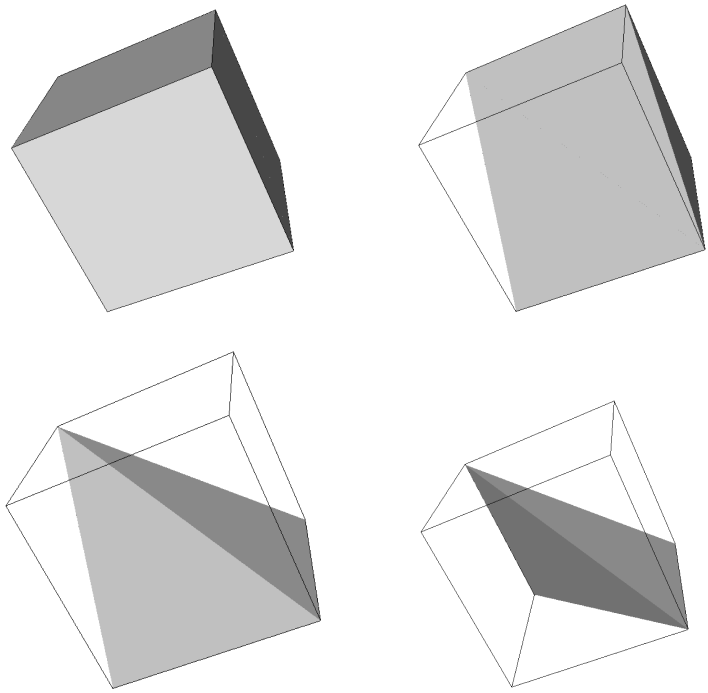

FIGURE 3.3. Nested sequence of polyhedra compliant with a sequence of comparisons.

inequality shows that any comparison of two items whose average rank in the rankings compatible with the comparisons so far differs by only 1 cuts of at least a $1/2e \sim 0.184$ fraction of the volume (and thus also of the compatible rankings). Here the rank of an item in a ranking is the number of items that are ranked below the item plus 1. It can be shown that such a comparison always exists as long as there are at least two compatible rankings. Of course we are done once there is only one compatible ranking left. See Matousek [**Mat02**] for details and a proof. At this moment we know that a good next comparison exists, but can we also find one efficiently? A heuristic to search for a good next comparison can be derived from the following observation: for any convex body in $\mathbb{R}^d$ any hyperplane that passes through the center of gravity cuts the

body into two bodies whose volume is at least a $\left(\frac{d}{d+1}\right)^d$-fraction of the volume of the original body. Of course this leaves us with the problem to find the center of gravity of a polyhedron, which is not an easy task. The analytic center as used in [**THS04**] can be seen as an approximation to the center of gravity, but no guarantees are known for the latter.

3.4. The conjoint structure case

According to the linear model each item p has a value $s(p) = \sum_{i=1}^{n} \lambda_{ij_i}$ where λ_{ij_i} is the respondent's partworth for the j_ith level of attribute i. These values induce a ranking of the items in $A_1 \times \ldots \times A_n$. Like in the unstructured case we want to determine how many item comparisons are needed to infer this ranking. Furthermore we want to study the geometric structure of the hyperplane arrangement introduced in Section 3.2. But at first we want to determine the information theoretic lower bound. Fredman has studied the information theoretic lower bound for the case of two attributes. Here we generalize his bound to the case with $n > 2$ attributes.

3.4.1. Information theoretic lower bound. As in the unstructured case any query strategy can be represented by a decision tree. Every leaf of the tree corresponds to a ranking of the items in $A_1 \times \ldots \times A_n$. For simplicity we again assume that no two items have the same value, i.e. for two items p and q either p is strictly preferred over q or q is strictly preferred over p. To determine the information theoretic lower bound one needs to determine the number of possible leaves of any decision tree, i.e., the number of possible item rankings. Let us call this number l. The height of a decision tree is lower bounded by $\log_2(l)$, which means that any query strategy to determine a ranking of $A_1 \times \ldots \times A_n$ needs at least $\log_2(l)$ item comparisons in the worst case. The total number of items in $A_1 \times \ldots \times A_n$ is m^n, but the number of possible rankings is substantially less than $m^n!$—as we would have in the unstructured case. To determine the number l of possible item rankings in the conjoint structure case we have a closer look at the hyperplane arrangement that corresponds to all comparisons of two items. This hyperplane arrangement gives us a subdivision of the cube $[-1, 1]^{nm}$ into cells and any cell corresponds to exactly one ranking of $A_1 \times \ldots \times A_n$. Hence, if we can count the number of cells we know the number of rankings and can therefore determine the information theoretic lower bound. Let us first count the number of hyperplanes: we have as many hyperplanes as we have possibilities to

choose two items p and q. We can choose two items by choosing for every attribute A_i two levels a_{ij} and a_{ik}. There are m^2 possibilities to do so for one attribute and therefore m^{2n} possibilities to choose two items. It is known that h hyperplanes partition the d-dimensional space into at most $\binom{h}{d}+\binom{h}{d-1}+\ldots+\binom{h}{0}$ many regions and thus the number of cells in the hyperplane arrangement is at most $nm\binom{m^{2n}}{nm} \leq nm^{2n^2m+1}$. Having found an upper bound for the number of cells we now can upper bound the information theoretic lower bound of the problem by the logarithm of this number. The information theoretic lower bound for ranking in the conjoint structure case is at most $(2n^2m+1)\log_2(m)+\log_2(n)$. Note again that the only thing we obtain here is an upper bound on the information theoretic lower bound and not the information lower bound itself.

3.4.2. Volume cuts. Fredman [**Fre76**] also used a different technique than employed for the information theoretic lower bound to show a lower bound on the number of necessary comparisons for the case of two attributes. He showed that any query strategy to determine a ranking obeying the linear model of the items in $A_1 \times A_2$ needs at least $(m-1)^2$ comparisons in the worst case. In particular it follows from his proof that there is a cell in the hyperplane arrangement induced by all comparisons of elements in $A_1 \times A_2$ that has at least $(m-1)^2$ facets.

In the following we consider the case of $n > 2$ attributes. We will show that in the subdivision of $[-1,1]^{nm}$ by the hyperplanes corresponding to paired item comparisons

(1) there exists a cell that has at least $(m-1)^2$ facets and that
(2) there exists a cell that has at most $m-1$ facets.

That is, in the conjoint structure case not all cells have the same combinatorial and thus geometric structure. This is in contrast to the structureless case where all cells are congruent simplices. This has important implications for volume based approaches to conjoint analysis. We will detail these implications after the proof in Section 3.5.

Note that (1) gives a lower bound of $(m-1)^2$ on the number of comparisons needed to derive a ranking of the elements in $A_1 \times \ldots \times A_n$. However, this result also follows directly from Fredman's lower bound for two attributes. To see this note that because of the linear model assumption a ranking of elements in $A_1 \times \ldots \times A_n$ for $n > 2$ uniquely determines the ranking of elements in $A_i \times A_j$ for any two attributes A_i

and A_j. If for $n > 2$ any ranking of elements in $A_1 \times \ldots \times A_n$ could be derived from less than $(m-1)^2$ comparisons this would be a contradiction to Fredman's lower bound for the case of two attributes.

Now, let us prove that there is a cell in the subdivision of $[-1,1]^{nm}$ by the hyperplanes corresponding to paired item comparisons that has at least $(m-1)^2$ facets. Suppose we have $n > 2$ attributes $A_1, \ldots, A_n$ where $A_i = \{a_{i1}, \ldots, a_{im}\}$ for $i = 1, \ldots, n$. To keep the exposition short we assume that n is even. We say that item $q \in A_1 \times A_2 \times \ldots \times A_n$ is a successor of item $p \in A_1 \times A_2 \times \ldots \times A_n$ in a ranking if the rank of q is one higher than the rank of p. An outline of the proof is as follows: we start with an item ranking ω induced by carefully chosen partworths. For every choice of $r \in \{1, \ldots, m-1\}$ and $t \in \{2, \ldots, m\}$ we modify the partworths a little bit to get another ranking ω_{rt} which differs from ω in exactly one transposition, namely in the transposition of an item p_{rt} and its successor in ω. Geometrically that means: since there are $(m-1)^2$ possible rankings ω_{rt} that differ from ω in exactly one transposition the cell corresponding to ω in the hyperplane arrangement introduced above has at least $(m-1)^2$ facets.

DEFINITION 3.1. For an item $p = (a_{1k_1}, \ldots, a_{nk_n})$ we call the sum $\sum_{i=1}^{n} k_i$ the index sum of p, the vector $(k_1, \ldots, k_n)$ the index vector of p and k_i the i-th index of p. Note that the index vector uniquely determines an item.

DEFINITION 3.2. Let ω be the ranking of elements in $A_1, \ldots, A_n$ defined by the following property: an item q is preferred over an item p in ω if and only if either the index sum of q is strictly larger than the index sum of p or the index sum of q is equal to the index sum of p and the index vector of q is lexicographically larger than the index vector of p.

LEMMA 3.3. *If we choose $\epsilon > 0$ such that $\sum_{i=1}^{n} \epsilon^i < \frac{1}{2(m-1)}$ and define the partworth for the j-th level of the i-th attribute as $\lambda_{ij} = j(1+\epsilon^i)$, then these partworths induce the ranking ω on $A_1 \times \ldots \times A_n$.*

PROOF. Let the partworths be chosen as in the lemma. Then, in the induced ranking, an item $q = (a_{1l_1}, \ldots, a_{nl_n})$ is preferred over an item $p = (a_{1k_1}, \ldots, a_{nk_n})$ if and only if

$$\sum_{i=1}^{n} l_i(1+\epsilon^i) > \sum_{i=1}^{n} k_i(1+\epsilon^i)$$

which can be rewritten as

$$\sum_{i=1}^{n} l_i - \sum_{i=1}^{n} k_i + \sum_{i=1}^{n} \epsilon^i (l_i - k_i) > 0 \tag{3.1}$$

First, we show that (3.1) is true if $\sum_{i=1}^{n} l_i > \sum_{i=1}^{n} k_i$ (*case 1*) or $\sum_{i=1}^{n} l_i = \sum_{i=1}^{n} k_i$ and $(l_1, \ldots, l_n)$ is lexicographically larger than $(k_1, \ldots, k_n)$ (*case 2*).

Case 1: Consider the case that $\sum_{i=1}^{n} l_i > \sum_{i=1}^{n} k_i$. As the values of the two sums are integral $\sum_{i=1}^{n} l_i - \sum_{i=1}^{n} k_i \geq 1$. Moreover the difference between any two indices $l_i - k_i$ can be lower bounded by $1 - m$. Therefore $\sum_{i=1}^{n} \epsilon^i (l_i - k_i) \geq -(m-1) \sum_{i=1}^{n} \epsilon^i > -1/2$, where the last inequality follows from our choice of ϵ. It follows that

$$\sum_{i=1}^{n} l_i - \sum_{i=1}^{n} k_i + \sum_{i=1}^{n} \epsilon^i (l_i - k_i) > 1 - 1/2 > 0.$$

Case 2: Consider the case that $\sum_{i=1}^{n} l_i = \sum_{i=1}^{n} k_i$ and $(l_1, \ldots, l_n)$ is lexicographically larger than $(k_1, \ldots, k_n)$.
Let $i_0 := \min\{1 \leq i \leq n : l_i - k_i \geq 1\}$. Note that for all indices i that are smaller than i_0 the difference $l_i - k_i = 0$ because $(l_1, \ldots, l_n)$ is lexicographically larger than $(k_1, \ldots, k_n)$. As $\sum_{i=1}^{n} l_i = \sum_{i=1}^{n} k_i$ it remains to show that $\sum_{i=i_0}^{n} \epsilon^i (l_i - k_i) > 0$:

$$\begin{aligned}
\sum_{i=i_0}^{n} \epsilon^i (l_i - k_i) &= \epsilon^{i_0} \underbrace{(l_{i_0} - k_{i_0})}_{\geq 1} + \sum_{i=i_0+1}^{n} \epsilon^i \underbrace{(l_i - k_i)}_{\geq -(m-1)} \\
&\geq \epsilon^{i_0} - (m-1) \sum_{i=i_0+1}^{n} \epsilon^i \\
&= \epsilon^{i_0} - \epsilon^{i_0} \underbrace{(m-1) \sum_{i=1}^{n-i_0} \epsilon^i}_{<1/2} \\
&> 0
\end{aligned}$$

where the fourth inequality follows from our choice of ϵ.

It remains to show the other direction of the equivalence, namely that (3.1) implies that $\sum_{i=1}^{n} l_i > \sum_{i=1}^{n} k_i$ or $\sum_{i=1}^{n} l_i = \sum_{i=1}^{n} k_i$ and $(l_1, \ldots, l_n)$ is lexicographically larger than $(k_1, \ldots, k_n)$. This follows immediately by the same arguments as above. □

In Figure 3.4 we illustrate the ranking ω for the case of two attributes $A = \{a_1, \ldots, a_5\}$ and $B = \{b_1, \ldots, b_5\}$ where each attribute has five levels. In the figure every item is represented by a square in a rectangular

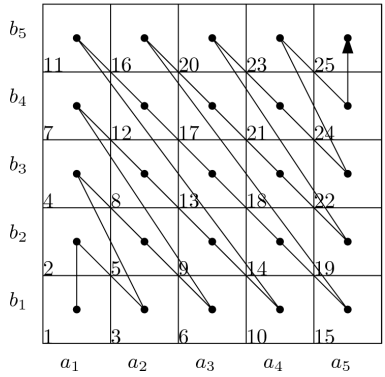

FIGURE 3.4. Ranking ω.

diagram. Altogether there are $m^2 = 25$ items/squares. One can read off the levels that describe a particular item from the column and line labeling. The lower left square represents the item described by the levels a_1 and b_1. This item is the lowest ranking item. The items are connected in increasing order by an oriented path through the diagram that represents the ranking ω, i.e., (a_1, b_2) is the second least preferred item. One can easily verify that the illustrated ranking indeed satisfies the defining property.

Remember that our goal is to define for every $r \in \{1, \ldots, m-1\}$ and every $t \in \{2, \ldots, m\}$ a ranking ω_{rt} that differs from ω in only one transposition, namely in the transposition of the item p_{rt} and its successor q_{rt} in ω. We now define the item p_{rt} and the item q_{rt}.

DEFINITION 3.4. For $1 \leq r \leq m-1$ and $2 \leq t \leq m$ let $p_{rt} = (p_{rt}^{(1)}, \ldots, p_{rt}^{(n)})$ be the item whose i-th entry is given as

$$p_{rt}^{(i)} = \begin{cases} a_{ir} & \text{if } i = 1 \\ a_{im} & \text{if } 1 < i \leq \frac{n}{2} \\ a_{it} & \text{if } i = \frac{n}{2} + 1 \\ a_{i1} & \text{if } \frac{n}{2} + 1 < i \leq n \end{cases}$$

i.e. p_{rt} has the index vector $(r, m, \ldots, m, t, 1, \ldots, 1)$

The item $q_{rt} = (q_{rt}^{(1)}, \ldots, q_{rt}^{(n)})$ is the item whose i-th entry is given as

$$q_{rt}^{(i)} = \begin{cases} a_{i(r+1)} & \text{if } i = 1 \\ a_{i1} & \text{if } 1 < i \leq \frac{n}{2} \\ a_{i(t-1)} & \text{if } i = \frac{n}{2} + 1 \\ a_{im} & \text{if } \frac{n}{2} + 1 < i \leq n \end{cases}$$

i.e. q_{rt} has the index vector $(r+1, 1, \ldots, 1, t-1, m, \ldots, m)$

LEMMA 3.5. *For $1 \leq r \leq m-1$ and $2 \leq t \leq m$ the item q_{rt} is the successor of the item p_{rt} in the ranking ω.*

PROOF. Both items have the same index sum and the index vector of q_{rt} is lexicographically larger than that of p_{rt}. Therefore q_{rt} is ranked higher in ω than p_{rt}. To see that q_{rt} is the successor of p_{rt} in ω assume that there is an item x which is different from p_{rt} and q_{rt} and which is ranked higher than p_{rt} and lower than q_{rt} in ω. Then x has the same index sum as p_{rt} and q_{rt}. Furthermore the index vector of x must be lexicographically larger than that of p_{rt} and lexicographically smaller than that of q_{rt}. Therefore the first entry of the index vector of x is either r or $r+1$. Assume that it is r. Then for $1 < i < n/2$ its i-th entry is m because otherwise the index vector of x would be lexicographically smaller than that of p_{rt}. Furthermore there are positions i_1 and i_2 with $n/2 < i_1 < i_2$ such that the i_1-th entry of the index vector of x is larger than that of p_{rt} and the i_2-th entry is smaller than that of p_{rt}. However, since all entries at positions larger than $n/2 + 1$ of the index vector of p_{rt} are 1, i.e. equal to the smallest index, such an x does not exist. A similar argument can be applied for the case that the first entry of the index vector of x is $r + 1$. Therefore q_{rt} is the direct successor of p_{rt} in ω. □

DEFINITION 3.6. For an item $p = (a_{1k_1}, \ldots, a_{nk_n})$ and a partworth vector $\lambda = (\lambda_{ij})$ let $s_\lambda(p)$ denote the (scale) value of p induced by the partworths λ_{ij}, i.e.

$$s_\lambda(p) = \sum_{i=1}^{n} \lambda_{ik_i}.$$

LEMMA 3.7. *The scale value difference between q_{rt} and p_{rt} induced by the partworths λ_{ij} is given as*

$$s_\lambda(q_{rt}) - s_\lambda(p_{rt}) = \epsilon - \epsilon^{n/2+1} + (m-1)\left(\sum_{i=n/2+2}^{n} \epsilon^i - \sum_{i=2}^{n/2} \epsilon^i\right).$$

PROOF. The scale value difference between p_{rt} and q_{rt} with respect to s_λ can be written as the sum of partworth differences over all attributes.

$$\begin{aligned} s_\lambda(q_{rt}) - s_\lambda(p_{rt}) &= \underbrace{(1+\epsilon)}_{\text{attribute 1}} + \underbrace{\sum_{i=2}^{n/2}(1-m)(1+\epsilon^i)}_{\text{attributes 2 to } n/2} - \underbrace{(1+\epsilon^{n/2+1})}_{\text{attribute } n/2+1} \\ &\quad + \underbrace{\sum_{i=n/2+2}^{n}(m-1)\left(1+\epsilon^i\right)}_{\text{attributes } n/2+2 \text{ to } n} \\ &= \epsilon - \epsilon^{n/2+1} - (m-1)\left(n/2 - 1 + \sum_{i=2}^{n/2}\epsilon^i\right) \\ &\quad + (m-1)\left(n/2 - 1 \sum_{i=n/2+2}^{n}\epsilon^i\right) \\ &= \epsilon - \epsilon^{n/2+1} + (m-1)\left(\sum_{i=n/2+2}^{n}\epsilon^i - \sum_{i=2}^{n/2}\epsilon^i\right) \end{aligned}$$

□

Remember that for two indices r and t we want to modify the partworths λ_{ij} a little bit such that the induced ranking differs from ω in exactly one transposition.

DEFINITION 3.8. For $1 \leq r \leq m-1$ and $2 \leq t \leq m$ let ω_{rt} be the ranking of elements in $A_1 \times A_2 \times \ldots \times A_n$ that differs from ω in exactly one transposition, namely in the transposition of p_{rt} and its successor q_{rt} in ω.

The goal is now to find partworths that induce ω_{rt}. To this end we introduce the modification values δ_{ij}.

DEFINITION 3.9. For $1 \leq r \leq m-1$ and $2 \leq t \leq m$ we define the modification values δ_{ij} as

$$\delta_{ij} = \begin{cases} \epsilon(1-m\epsilon)+\epsilon^n/2 & \text{if } i=1 \text{ and } j=r \\ \epsilon^i(1-m\epsilon) & \text{if } 1<i<n/2 \text{ and } j=m \\ \epsilon^i(1-\epsilon) & \text{if } i=n/2 \text{ and } j=m \\ \epsilon^{i+1}(m-1) & \text{if } i=n/2+1 \text{ and } j=t \text{ or} \\ & \text{if } n/2+1<i<n \text{ and } j=1 \\ 0 & \text{else} \end{cases}$$

Furthermore, let δ denote the sum of all δ_{ij}, i.e.

$$\delta = \sum_{i=1}^{n}\sum_{j=1}^{m}\delta_{ij}.$$

Note that $\delta_{ij} \geq 0$ for all i and all j. We now define the partworths $\hat{\lambda}_{ij}$ by modifying the partworths λ_{ij}. In the remaining part of this section we will show that the ranking induced by the partworths $\hat{\lambda}_{ij}$ is exactly the ranking ω_{rt}.

DEFINITION 3.10. For $1 \leq r \leq m-1$ and $2 \leq t \leq m$ we define the modified partworth of the level a_{ij} as

$$\hat{\lambda}_{ij} = \lambda_{ij} + \delta_{ij}.$$

By $\hat{\lambda}$ we denote the partworth vector containing the modified partworths. For an item p we denote by $s_{\hat{\lambda}}(p)$ its scale value induced by the modified partworths.

Note that δ_{ij} has been chosen such that all levels that describe the item p_{rt} gets assigned a strictly larger partworth in the modification, whereas the partworths for all other levels remain unchanged. Note that when modifying the partworths the value of an item p can increase by at most δ, i.e.

$$s_{\hat{\lambda}}(p) \leq s_{\lambda}(p) + \delta.$$

In particular the scale value of p_{rt} increases exactly by the value δ. Let us now determine the value of δ.

LEMMA 3.11. *The value of δ is given as*

$$\delta = \frac{1}{2}\epsilon^n + s_{\lambda}(q_{rt}) - s_{\lambda}(p_{rt})$$

PROOF. Remember that δ is the sum of all δ_{ij}. Using Definition 3.9, δ can be written as

$$\begin{aligned}
\delta &= \frac{1}{2}\epsilon^n + \sum_{i=1}^{n/2-1} \epsilon^i(1-m\epsilon) + \epsilon^{n/2}(1-\epsilon) + \sum_{i=n/2+1}^{n-1} \epsilon^{i+1}(m-1) \\
&= \frac{1}{2}\epsilon^n + \sum_{i=1}^{n/2-1} \epsilon^i - m\epsilon \sum_{i=1}^{n/2-1} \epsilon^i + \epsilon^{n/2} - \epsilon^{n/2+1} + \sum_{i=n/2+1}^{n-1} \epsilon^{i+1}(m-1) \\
&= \frac{1}{2}\epsilon^n + \sum_{i=1}^{n/2} \epsilon^i - m\sum_{i=2}^{n/2} \epsilon^i - \epsilon^{n/2+1} + \sum_{i=n/2+2}^{n} \epsilon^i(m-1) \\
&= \frac{1}{2}\epsilon^n + \sum_{i=1}^{n/2} \epsilon^i - \sum_{i=2}^{n/2+1} \epsilon^i + (m-1)\left(\sum_{i=n/2+2}^{n} \epsilon^i - \sum_{i=2}^{n/2} \epsilon^i\right) \\
&= \frac{1}{2}\epsilon^n + \epsilon - \epsilon^{n/2+1} + (m-1)\left(\sum_{i=n/2+2}^{n} \epsilon^i - \sum_{i=2}^{n/2} \epsilon^i\right) \\
&= \frac{1}{2}\epsilon^n + s_\lambda(q_{rt}) - s_\lambda(p_{rt})
\end{aligned}$$

where the last equality follows from Lemma 3.7. □

In order to prove that the ranking induced by the partworths $\hat{\lambda}_{ij}$ is indeed the ranking ω_{rt} we will proceed in three steps:

(1) In Lemma 3.12 we show that p_{rt} is ranked higher than q_{rt} in the ranking induced by the modified partworths $\hat{\lambda}_{ij}$.
(2) In Lemma 3.13 we then show that the successor of q_{rt} in ω is still ranked higher than p_{rt} in the ranking induced by the modified partworths $\hat{\lambda}_{ij}$
(3) In Lemma 3.14 we prove that for any two items p and q where $p \neq p_{rt}$ the following holds: if q is ranked higher than p in ω then q is also ranked higher than p in the ranking induced by the modified partworths $\hat{\lambda}_{ij}$.

These three facts together imply that the ranking induced by the partworths $\hat{\lambda}_{ij}$ is indeed the ranking ω_{rt}.

LEMMA 3.12. *The item p_{rt} is ranked higher than the item q_{rt} in the ranking induced by the modified partworths $\hat{\lambda}_{ij}$. In particular it holds that*

$$s_{\hat{\lambda}}(p_{rt}) - s_{\hat{\lambda}}(q_{rt}) = \frac{1}{2}\epsilon^n > 0$$

.

PROOF. Modifying the partworths we increase the value of p_{rt} by δ, i.e., $s_{\hat{\lambda}}(p_{rt}) = s_\lambda(p_{rt}) + \delta$, whereas the value of q_{rt} remains unchanged, i.e. $s_{\hat{\lambda}}(q_{rt}) = s_\lambda(q_{rt})$. On the other hand we have shown in Lemma 3.11 that $\delta = \frac{1}{2}\epsilon^n + s_\lambda(q_{rt}) - s_\lambda(p_{rt})$. Therefore it holds that

$$\begin{aligned} s_{\hat{\lambda}}(p_{rt}) - s_{\hat{\lambda}}(q_{rt}) &= s_\lambda(p_{rt}) + \delta - s_\lambda(q_{rt}) \\ &= s_\lambda(p_{rt}) + \frac{1}{2}\epsilon^n + s_\lambda(q_{rt}) - s_\lambda(p_{rt}) - s_\lambda(q_{rt}) \\ &= \frac{1}{2}\epsilon^n \\ &> 0 \end{aligned}$$

□

LEMMA 3.13. *Let q be the successor of q_{rt} in the ranking ω. Then q is ranked higher than p_{rt} in the ranking induced by the partworths $\hat{\lambda}_{ij}$, i.e.*

$$s_{\hat{\lambda}}(q) - s_{\hat{\lambda}}(p_{rt}) > 0.$$

PROOF. Remember that the index vector of q_{rt} is given as

$$(r+1, 1, \ldots, 1, t-1, m, \ldots, m)$$

where the entry $t-1$ is at position $n/2+1$. Let q be the successor of q_{rt} in ω. Then the index vector of q is given as

$$(r+1, 1, \ldots, 1, t, m-1, m, \ldots, m)$$

where the entry t is again at position $n/2+1$. Note that q_{rt} and q have the same index sum and that the index vector of q is lexicographically larger than that of q_{rt}. Therefore q is ranked higher than q_{rt} in ω. It is easy to see that q is the successor of q_{rt} in ω by applying similar arguments as in the proof of Lemma 3.12.

The items q_{rt} and q differ only in the levels for attribute $n/2+1$ and attribute $n/2+2$. This permits to express the value of q in terms of the value of q_{rt} and the partworth differences for the two levels in which they differ:

$$
\begin{aligned}
s_{\hat{\lambda}}(q) &= s_{\hat{\lambda}}(q_{rt}) + \underbrace{1+\epsilon^{n/2+1}+\epsilon^{n/2+2}(m-1)}_{\text{partworth difference for attribute } n/2+1} \\
&\quad \underbrace{-(1+\epsilon^{n/2+2})-\epsilon^{n/2+3}(m-1)}_{\text{partworth difference for attribute } n/2+2} \\
&= s_{\hat{\lambda}}(q_{rt}) + \epsilon^{n/2+1} - \epsilon^{n/2+2} + \underbrace{(m-1)\epsilon(\epsilon^{n/2+1}-\epsilon^{n/2+2})}_{>0} \\
&> s_{\hat{\lambda}}(q_{rt}) + \epsilon^{n/2+1}(1-\epsilon) \\
&\geq s_{\hat{\lambda}}(q_{rt}) + \epsilon^{n/2+1}\left(1 - \frac{1}{2(m-1)}\right) \\
&\geq s_{\hat{\lambda}}(q_{rt}) + \frac{1}{2}\epsilon^{n/2+1} \\
&\geq s_{\hat{\lambda}}(q_{rt}) + \frac{1}{2}\epsilon^{n}.
\end{aligned}
$$

It follows that

$$s_{\hat{\lambda}}(q) - s_{\hat{\lambda}}(p_{rt}) > s_{\hat{\lambda}}(q_{rt}) + \frac{1}{2}\epsilon^n - s_{\hat{\lambda}}(p_{rt}).$$

In Lemma 3.12 we have shown that $s_{\hat{\lambda}}(q_{rt}) - s_{\hat{\lambda}}(p_{rt}) = -\frac{1}{2}\epsilon^n$. Therefore we can conclude that $s_{\hat{\lambda}}(q) - s_{\hat{\lambda}}(p_{rt}) > 0$ □

LEMMA 3.14. *For any two items p and q where $p \neq p_{rt}$ the following holds: if q is ranked higher than p in ω then q is also ranked higher than p in the ranking induced by the partworths $\hat{\lambda}_{ij}$.*

PROOF. Let $p = (a_{1k_1}, \ldots, a_{1k_n})$ and $q = (a_{1l_1}, \ldots, a_{1l_n})$ be two items in $A_1 \times \ldots \times A_n$ with $p \neq p_{rt}$ and with q being ranked higher than p in the ranking ω. Throughout the proof we denote by $\Delta_i = l_i - k_i$ the index difference at the i-th position. Note that $-(m-1) \leq \Delta_i \leq m-1$ for all $1 \leq i \leq n$.

For the partworth difference of levels a_{il_i} and a_{ik_i} with $l_i \neq k_i$ it holds

$$
\begin{aligned}
\hat{\lambda}_{il_i} - \hat{\lambda}_{ik_i} &= \lambda_{il_i} + \delta_{il_i} - \lambda_{ik_i} - \delta_{ik_i} \\
&\geq \lambda_{il_i} - \lambda_{ik_i} - \delta_{ik_i} \\
&= \Delta_i(1+\epsilon^i) - \delta_{ik_i}. \qquad (3.2)
\end{aligned}
$$

Note that δ_{ik_i} is only strictly larger than 0 if a_{ik_i} is the level that is also contained in p_{rt}.

In the following we distinguish two cases. We know that q is ranked higher than p in the ranking ω. Therefore either q has a larger index sum than p (which we investigate in *case 1*) or both items have the same index sum and the index vector of q is lexicographically larger than that of p (which we investigate in *case 2*).

Case 1: Let us first consider the case that the index sum of q is larger than the index sum of p, i.e. that $\sum_{i=1}^{n} \Delta_i \geq 1$. As $p_{rt} \neq p$ there is at least on level a_{ik_i} of item p whose partworth has not been modified, i.e. $\lambda_{ik_i} = \hat{\lambda}_{ik_i}$. This means that under the modification the overall value of p increases by at most $\delta - \min\{\delta_{ij} : \delta_{ij} > 0\}$. The minimum of all δ_{ij} with $\delta_{ij} > 0$ is $\delta_{n-1,1} = \epsilon^n(m-1) \geq \epsilon^n$. Therefore it holds that

$$s_{\hat{\lambda}}(q) - s_{\hat{\lambda}}(p) > s_{\lambda}(q) - s_{\lambda}(p) - \delta + \frac{1}{2}\epsilon^n$$

Using inequality 3.2 and summing up for all attributes we get for the scale value difference of q and p:

$$\begin{aligned}
& s_{\hat{\lambda}}(q) - s_{\hat{\lambda}}(p) \\
> \;& \sum_{i=1}^{n} \Delta_i(1+\epsilon^i) - \delta + \frac{1}{2}\epsilon^n \\
= \;& \sum_{i=1}^{n} \Delta_i(1+\epsilon^i) + \frac{1}{2}\epsilon^n \\
& - \left(\frac{1}{2}\epsilon^n + \epsilon - \epsilon^{n/2+1} + (m-1)\left(\sum_{i=n/2+2}^{n} \epsilon^i - \sum_{i=2}^{n/2} \epsilon^i \right) \right) \\
= \;& \sum_{i=1}^{n} \Delta_i(1+\epsilon^i) - \epsilon + \epsilon^{n/2+1} - (m-1)\left(\sum_{i=n/2+2}^{n} \epsilon^i - \sum_{i=2}^{n/2} \epsilon^i \right)
\end{aligned}$$

where the first equality follows from plugging in the value of δ which we have determined in the proof of Lemma 3.11. Using that

$$\sum_{i=n/2+2}^{n} \epsilon^i - \sum_{i=2}^{n/2} \epsilon^i < 0$$

we get

$$\begin{aligned} & s_{\hat{\lambda}}(q) - s_{\hat{\lambda}}(p) \\ > & \underbrace{\sum_{i=1}^{n} \Delta_i}_{\geq 1} + \sum_{i=1}^{n} \underbrace{\Delta_i \epsilon^i}_{\geq -(m-1)\epsilon^i} - \epsilon + \epsilon^{n/2+1} \\ \geq & 1 - (m-1) \sum_{i=1}^{n} \epsilon^i - \epsilon + \epsilon^{n/2+1} \\ \geq & 1 - \frac{1}{2} - \frac{1}{2} + \epsilon^{n/2+1} \\ > & 0 \end{aligned}$$

We used that $(m-1)\sum_{i=1}^{n} \epsilon^i < 1/2$. This follows from our choice of ϵ. In particular it also holds that $\epsilon < 1/2$.

Case 2: Let us now consider the case that the index sum of q equals the index sum of p and that the index vector of q is lexicographically larger than that of p. Let $i_0 := \min\{1 \leq i \leq n : \Delta_i \geq 1\}$. Note that again for all i that are smaller than i_0 it holds that $\Delta_i = 0$ because of q being lexicographically larger than p and that $\sum_{i=i_0}^{n} \Delta_i = \sum_{i=1}^{n} \Delta_i = 0$. We again distinguish different cases, now depending on the value of i_0. Let us consider first the case that $n/2 < i_0 < n$. Using Equation 3.2 and the definition of δ_{ij} (see 3.9) and summing up over all attributes starting from the i_0-th attribute, the following holds:

$$\begin{aligned} & s_{\hat{\lambda}}(q) - s_{\hat{\lambda}}(p) \\ \geq & \Delta_{i_0}(1+\epsilon^{i_0}) - \epsilon^{i_0+1}(m-1) + \sum_{i=i_0+1}^{n} \Delta_i(1+\epsilon^i) - \sum_{i=i_0+1}^{n-1} \epsilon^{i+1}(m-1) \\ = & \sum_{i=i_0}^{n} \Delta_i(1+\epsilon^i) - \epsilon^{i_0+1}(m-1) - \sum_{i=i_0+1}^{n-1} \epsilon^{i+1}(m-1) \\ = & \sum_{i=i_0}^{n} \Delta_i + \Delta_{i_0}\epsilon^{i_0} + \sum_{i=i_0+1}^{n} \Delta_i\epsilon^i - \sum_{i=i_0}^{n-1} \epsilon^{i+1}(m-1) \\ \geq & \Delta_{i_0}\epsilon^{i_0} + \sum_{i=i_0+1}^{n} \Delta_i\epsilon^i - \sum_{i=i_0+1}^{n} \epsilon^i(m-1) \\ = & \Delta_{i_0}\epsilon^{i_0} + \sum_{i=i_0+1}^{n} \epsilon^i(\Delta_i - (m-1)) \end{aligned}$$

$$\begin{aligned} &\geq \Delta_{i_0}\epsilon^{i_0} - \epsilon^{i_0} 2(m-1) \sum_{i=1}^{n-i_0} \epsilon^i \\ &> \epsilon^{i_0} - \epsilon^{i_0} \\ &= 0 \end{aligned}$$

For the case that $i_0 \leq n/2$ the same result can be shown analogously. This concludes the proof of the lemma.

□

THEOREM 3.15. *There are at least* $(m-1)^2$ *different realizable rankings of the elements in* $A_1 \times \ldots \times A_n$ *that differ from* ω *in exactly one transposition.*

PROOF. The theorem follows directly from Lemma 3.12, Lemma 3.13 and Lemma 3.14. □

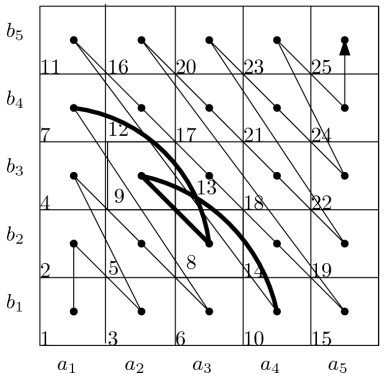

FIGURE 3.5. Ranking ω_{rt}.

In Figure 3.5 we illustrate the ranking ω_{rt} for the two attributes $A = \{a_1, \ldots, a_5\}$ and $B = \{b_1, \ldots, b_5\}$ and for $r = 2$ and $t = 3$. Observe that in ω_{rt} the item (a_2, b_3) is now preferred over (a_3, b_2).

Considering again the subdivision of $[-1, 1]^{nm}$ by the hyperplanes corresponding to paired item comparisons we have shown that the cell corresponding to the ranking ω has at least $(m-1)^2$ facets. On the other hand there are cells in this hyperplane arrangement that have less facets and therefore a different combinatorial structure. Consider for example

the ranking induced by the lexicographic ordering of the index vectors. The cell corresponding to this lexicographic ranking has at most $m-1$ many facets. Figure 3.6 illustrates that for the case of two attributes $A = \{a_1, \ldots, a_5\}$ and $B = \{b_1, \ldots, b_5\}$. It shows the two purely lexicographic rankings on $A \times B$. One can observe that only the items at the turning points of the paths are candidates for items that can be exchanged with their successor to get another linearly representable ranking.

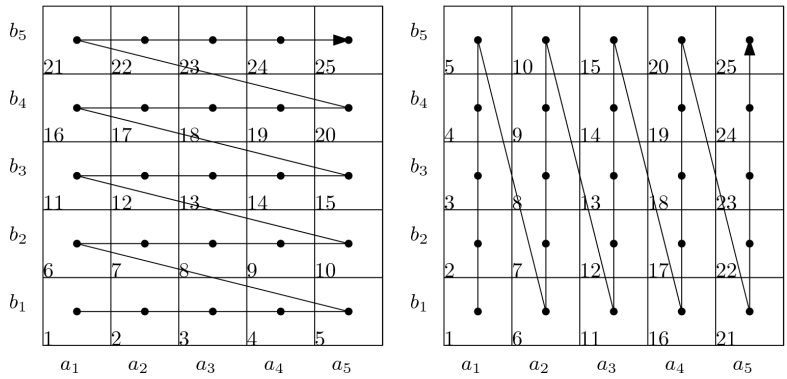

FIGURE 3.6. The two lexicographic rankings. In the left figure A is preferred over B and in the right figure B is preferred over A.

In the following we generalize this observation:

DEFINITION 3.16. Let π be the ranking on $A_1 \times \ldots \times A_n$ which is defined by the following property: An item q is preferred over an item p in π if and only if the index vector of q is lexicographically larger than that of p.

Note that π is linearly realizable, for example by choosing the partworth $\lambda_{ij} = j\epsilon^i$ where $\epsilon > 0$ is chosen such that $\sum_{i=1}^{n} \epsilon^i < 1/(m-1)$.

LEMMA 3.17. *There are at most $m-1$ linearly realizable rankings that differ from π in exactly one transposition.*

PROOF. Consider the linearly realizable ranking π_p which differs from π only in the transposition of $p = (a_{1k_1}, \ldots, a_{nk_n})$ and its successor $q = (a_{1l_1}, \ldots, a_{nl_n})$ in π. Then the set of levels describing p is disjoint

from the set of levels describing q. To see this, assume for a contradiction that there is an index i such that $a_{ik_i} = a_{il_i}$. Without loss of generality we assume that this index is 1, i.e. that $a_{1k_1} = a_{1l_1}$. We consider the item q' which is given as $q' = (b, a_{2l_2}, \ldots, a_{nl_n})$ where $b \in A_1 \setminus \{a_{1l_1}\}$. That is, q' is described by the same levels as q, except that the first level a_{1k_1} is replaced by another level b. Analogously, we define $p' = (b, a_{2k_2}, \ldots, a_{nk_n})$. For any partworth vector λ it holds: $s_\lambda(q) > s_\lambda(p)$ if and only if $s_\lambda(q') > s_\lambda(p')$. To see this note that

$$\begin{aligned} s_\lambda(q) - s_\lambda(p) &= \sum_{i=1}^{n} (\lambda_{il_i} - \lambda_{ik_i}) \\ &= \underbrace{\lambda_{1l_1} - \lambda_{1k_1}}_{=0} + \sum_{i=2}^{n} (\lambda_{il_i} - \lambda_{ik_i}) \\ &= \lambda_b - \lambda_b + \sum_{i=2}^{n} (\lambda_{il_i} - \lambda_{ik_i}) \\ &= s_\lambda(q') - s_\lambda(p') \end{aligned}$$

where λ_b is the scale value of the level b. But this means that no matter how the partworths that induce π_p are chosen, π differs from π_p in at least two transpositions, which leads to a contradiction. Therefore the set of levels describing p must be disjoint from the set of levels describing q. In the lexicographic ordering π there are exactly $m-1$ items whose set of levels is disjoint from that of their successor. These are the items whose index vector is in $\{1, \ldots, m-1\} \times \{m\} \times \ldots \times \{m\}$. Therefore there are at most $m-1$ realizable rankings that differ from π in exactly one transposition. □

3.5. Discussion

We showed that in the subdivision of $[-1,1]^{nm}$ by the hyperplanes corresponding to paired item comparisons there exists a cell that has at least $(m-1)^2$ facets and that there exist cells that have at most $m-1$ facets. This leads to the following observation:

OBSERVATION 3. *The cells in the subdivision of* $[-1,1]^{nm}$ *by the hyperplanes corresponding to paired item comparisons have different combinatorial and thus also different geometric structures.*

This observation has to be seen in contrast to Observation 2 for the structureless case where all cells are congruent simplices.

Polyhedral choice based conjoint analysis is based on the intuition from the structureless case and implicitly assumes that also in the conjoint structure case the cells have the same geometric structure. Our findings raise some doubts about the reliability of this approach. Nevertheless, it is reported in literature that polyhedral choice based conjoint performs very well with respect to certain evaluation methods. This might be explained by having a closer look at the evaluation methods themselves:

The performance of conjoint analysis techniques, in particular of polyhedral choice based conjoint analysis is often evaluated by using a simulation and not by conducting a real conjoint study. A respondent is simulated by drawing a partworth vector p uniformly at random from the cube $[-1,1]^{nm}$. The vector p simulates the respondent's true preferences. Then a conjoint study is simulated where choices are made according to p. The conjoint study results in an estimation $\hat{p}$ of the true partworth vector. Then different performance measures are applied: A commonly applied measure is the *hit rate*. For a series of choice tasks that have not been used in the conjoint study itself, the hitrate is computed as the fraction of choice tasks whose outcome can be predicted correctly by using $\hat{p}$, i.e. for which p and $\hat{p}$ lead to the same outcome. The drawback of this evaluation method is that the partworth vector p is chosen uniformly at random from the cube $[-1,1]^{nm}$. Because of Observation 3 this strategy does not uniformly sample the space of rankings compliant with a linear structure. The probability that the partworth vector p represents a ranking ω depends on the volume of the cell in the hyperplane arrangement that corresponds to ω. That is, the probability to draw a ranking which corresponds to a cell with large volume is higher than to draw a ranking corresponding to a cell with small volume. The uniform geometric sampling strategy therefore induces a combinatorial bias. On the other hand it seems likely that rankings corresponding to cells with large volume are easier to reconstruct for a volume based approach than rankings corresponding to cells with small volume. That is, polyhedral choice based conjoint is evaluated using instances on which it should perform well, anyway.

Another performance measure used to evaluate conjoint techniques in simulations is the Euclidean distance $\|p - \hat{p}\|$ between the true and the estimated partworth vector. Polyhedral choice based conjoint is just designed to make this distance small. Therefore it is not surprising that very good results are reported with respect to that measure in comparison to other conjoint techniques. However, the Euclidean distance is no

measure for the combinatorial distance between the rankings induced by p and $\hat{p}$, respectively.

To conclude: we doubt that polyhedral choice based conjoint is a reliable approach to conjoint analysis. The good performance results might be due to inappropriate evaluation methods.

3.6. Conclusion

We shed some light on the geometric and combinatorial structure of polyhedral conjoint analysis. Our analysis shows that one has to be careful when transferring insights and ideas from the structureless case to the case with conjoint structure.

Recently we could improve our results [1]: it is possible to show as in this chapter that there exists a cell in the hyperplane arrangement introduced in 3.2 that has at least $(m-1)^n$ facets. In particular this means that in order to derive a ranking of the elements in $A = A_1 \times \ldots \times A_n$ at least $(m-1)^n$ paired comparisons have to be performed in the worst case. For conjoint studies arising in practice it is infeasible to ask a respondent that many questions. Respondents usually get worn out after a very small number of choice tasks and do not answer further questions faithfully anymore. The number of feasible choice tasks does not even scale with the problem size, i.e., is some sort of of "universal constant". Therefore one has to strive for alternatives in preference analysis. We see three possible alternative approaches which are, in their applicability, not mutually exclusive:

(1) Approximation. Instead of determining a respondent's exact ranking of the items one could determine only an approximate ranking. For many applications an approximation might be enough.
(2) Aggregation. Instead of determining a scale for every respondent individually one could aggregate choice information and find an aggregated scale for the whole population or for parts of the population.
(3) Modeling. One could introduce additional modeling assumptions and therefore reduce the number of free parameters.

In Chapter 4, we will investigate the approach of approximation. We will pursue the question of how many comparisons are necessary and

[1] Discussion at the 5'th GWOP with Jiří Matoušek, Miloš Stojaković, Sonja Čukić, Martin Jaggi, Joachim Giesen and Eva Schuberth

sufficient to approximately rank k items. We will restrict ourselves to the case where the items do not possess a conjoint structure.

In Chapter 5 we will combine aggregation and modeling and develop an algorithm to derive aggregated interval scales for the whole population. The algorithm is based on a new statistical model.

CHAPTER 4

Approximation

4.1. Introduction

In this chapter we are dealing with a set of n items without conjoint structure. That is, for the sake of simplicity we restrict ourselves to the structureless case here. A respondent performs a sequence of choice tasks, each involving two items and we assume that the respondent's preference structure — according to which he performs the choice tasks — can be represented by a ranking of the items. We will pursue the question of how many choice tasks are necessary and sufficient in the worst case in order to derive an approximate ranking of the n items.

In Chapter 3 we have seen that a query strategy that presents a sequence of item pairs to the respondent from which a ranking has to be derived is nothing else than a comparison based sorting algorithm and that there is no such sorting algorithm that can determine the ranking by posing less than $n \log \frac{n}{e}$ paired comparison questions in the worst case, i.e., in general $\Omega(n \log n)$ comparisons are needed (see [**Knu73**]). Here we will answer the question whether it is true that we need substantially less comparisons when we only want to know the ranking approximately.

In order to give sense to the term "approximately" we need some metric to compare rankings. Assume that we are dealing with n items. Since a ranking is a permutation of the items, this means that we need a metric on the permutation group S_n. Not all of the metrics, e.g., the Hamming distance that counts how many items are ranked differently, are meaningful for our application. For example, if in the respondent's ranking one exchanges every second item with its predecessor, then the resulting ranking has maximal Hamming distance to the original one. Nevertheless, this ranking still tells a lot about the respondent's preferences. In marketing applications Kendall's tau metric [**DG77**] is frequently used

since it seems to capture the intuitive notion of closeness of two rankings and also arises naturally in the statistics of certain random rankings [**Mal57**]. Kendall's distance of a permutation $\pi \in S_n$ to the sorted permutation is the minimum number of pairwise adjacent transpositions required to bring π into sorted order. In [**Knu73**] it is shown that this is equivalent to the number of inversions in π. Instead of working with Kendall's metric we use Spearman's footrule metric [**DG77**] which essentially is equivalent to Kendall's metric, since the two metrics are within a constant factor of each other [**DG77**]. The maximal distance between any two rankings of n items in Spearman's footrule metric is less than n^2. We show that in order to obtain a ranking at distance $n^2/\nu(n)$ to the respondent's ranking with any strategy, a respondent has in general to perform at least $n\,(\min\{\log \nu(n), \log n\} - 6)$ comparisons in the worst case. Moreover, if we allow the strategy to be randomized such that the obtained ranking is at expected distance $n^2/\nu(n)$ to the respondent's ranking, we can show that the same bound on the minimum number of comparisons holds. On the other hand, there is a deterministic strategy (algorithm), suggested in [**Cha00**], that shows that $6n \log \nu(n)$ comparisons are always sufficient.

At first glance this work seems related to work done on pre-sorting. In pre-sorting the goal is to pre-process the data such that fewer comparisons are needed afterwards to sort them. For example in [**HYY00**] it is shown that with $O(1)$ pre-processing one can save $\Theta(n)$ comparisons for Quicksort on average. Pre-processing can be seen as computing a partial order on the data that helps for a given sorting algorithm to reduce the number of necessary comparisons. The structural quantity that determines how many comparisons are needed in general to find the ranking given a partial order is the number of linear extensions of the partial order, i.e., the number of rankings consistent with the partial order. Actually, the logarithm of this number is a lower bound on the number of comparisons needed in general [**KK92**]. Here we study another structural measure, namely, the maximum diameter in the Spearman's metric of the set of rankings consistent with a partial order. Our results shows that with $o(n \log n)$ comparisons one can make this diameter asymptotically smaller than the diameter of the set of all rankings. That is not the case for the number of linear extensions which stays in $\Theta(2^{n \log n})$.

Notation. The logarithm log in this chapter is assumed to be binary, and by id we denote the identity (increasing) permutation of $[n]$.

4.2. Algorithm

The idea of the ASort algorithm is to partition the items into a sorted sequence of equal-sized bins such that the elements in each bin have smaller rank than any element in subsequent bins. This approach was suggested by Chazelle [**Cha00**] for near-sorting. The output of the algorithm is the sequence of bins. Note that we do not specify the ordering of elements inside each bin, but consider any ranking consistent with the ordering of the bins. We will show that any such ranking approximates the actual ranking of the elements in terms of Spearman's footrule metric

$$D(\pi, \mathrm{id}) = D(\pi) = \sum_{i=1}^{n} |i - \pi(i)|,$$

where $\pi(i)$ is the rank of the element of rank i in an approximate ranking, i.e., $|i - \pi(i)|$ measures deviation of the approximated rank from the actual rank. Note that for any ranking the distance in the Spearman's footrule metric to id is at most $\frac{n^2}{2}$.

Since for every i the value $|i - \pi(i)|$ is bounded by n divided by the number of bins, we see that the approximation quality depends on the number of bins.

The algorithm ASort iteratively performs a number of median searches, each time placing the median into the right position in the ranking. Here the median of n elements is defined to be the element of rank $\lfloor \frac{n+1}{2} \rfloor$.

```
ASort (B : set, m : int)
 1  B_01 := B   //   B_ij is the j'th bin in the i'th round
 2  for i := 1 to m do
 3     for j := 1 to 2^(i-1) do
 4        compute the median of B_(i-1)j
 5        B_i(2j-1) := {x ∈ B_(i-1)j | x ≤ median}
 6        B_i(2j) := {x ∈ B_(i-1)j | x > median}
 7     end for
 8  end for
 9  return B_m1, ..., B_m(2^m)
```

To compute the median in line 4 and to partition the elements in line 5 and 6 we use the deterministic algorithm by Blum et al. [**BFP+72**] that performs at most $5.73n$ comparisons in order to compute the median

of n elements and to partition them according to the median. We note that in putting the algorithm ASort to practice one may want to use a different median algorithm, like, e.g., Hoare's Find algorithm [**Hoa61**].

In the following we determine the number of comparisons the algorithm ASort needs on input B with $|B| = n$ in order to guarantee a prescribed approximation error of the actual ranking for any ranking consistent with the ordering of the bins $B_{m1}, \ldots, B_{m(2^m)}$ computed by the algorithm.

Lemma 4.1. *For every $x \in B_{ij}$, where $0 \leq i \leq m$ and $1 \leq j \leq 2^i$, it holds*

$$\sum_{k=1}^{j-1} |B_{ik}| + 1 \leq \mathrm{rank}(x) \leq \sum_{k=1}^{j} |B_{ik}|.$$

Proof. The lemma can be proven by induction on the number of rounds. By construction, the elements in B_{01} have rank at least 1 and at most $n = |B_{01}| = \sum_{k=1}^{1} |B_{0k}|$. The claim for $i = 0$ follows if we set the empty sum $\sum_{k=1}^{0} |B_{0k}|$ to 0.

Now assume that the statement holds after the $(i-1)$'th round. The algorithm partitions every bin $B_{(i-1)j}$ into two bins $B_{i(2j-1)}$ and $B_{i(2j)}$. Again by construction the elements in bin $B_{i(2j-1)}$ have rank at least

$$\sum_{k=1}^{j-1} |B_{(i-1)k}| + 1 = \sum_{k=1}^{j-1} (|B_{i(2k-1)}| + |B_{i(2k)}|) + 1 = \sum_{k=1}^{(2j-1)-1} |B_{ik}| + 1,$$

and at most

$$\sum_{k=1}^{(2j-1)-1} |B_{ik}| + |B_{i(2j-1)}| = \sum_{k=1}^{2j-1} |B_{ik}|.$$

Similarly, the elements in bin $B_{i(2j)}$ have rank at least $\sum_{k=1}^{2j-1} |B_{ik}| + 1$ and at most $\sum_{k=1}^{j} |B_{(i-1)k}| = \sum_{k=1}^{2j} |B_{ik}|$. □

Lemma 4.2. $\lfloor \frac{n}{2^i} \rfloor \leq |B_{ij}| \leq \lceil \frac{n}{2^i} \rceil$ *for $0 \leq i \leq m$ and $1 \leq j \leq 2^i$*

Proof. We prove by induction that in any round i the sizes of any two bins differ by at most 1, i.e., $\big||B_{ij}| - |B_{ik}|\big| \leq 1$ for $0 \leq i \leq m$ and $1 \leq j, k \leq 2^i$. The statement of the lemma then follows since by an averaging argument and the integrality of the bin sizes, the size of each bin must be of size either $\lceil \frac{n}{2^i} \rceil$ or $\lfloor \frac{n}{2^i} \rfloor$.

For $i = 0$ all n elements of B are in bin B_{01}. The claim for $i = 0$ follows since $\lfloor n \rfloor \leq n \leq \lceil n \rceil$.

Now assume that the statement holds for $i - 1$. Take two bins $B_{(i-1)j}$ and $B_{(i-1)k}$. We distinguish two cases.

Case 1. $B_{(i-1)j}$ and $B_{(i-1)k}$ have the same size c. If c is even, then both bins get split up into two bins each and the resulting four bins all have the same size. If c is odd, then each of the bins gets split up into two bins of sizes $\lfloor \frac{c}{2} \rfloor$ and $\lceil \frac{c}{2} \rceil$, respectively, which differ by 1.

Case 2. Without loss of generality, $|B_{(i-1)j}| = c$ and $|B_{(i-1)k}| = c + 1$. If c is even, then $B_{(i-1)j}$ gets split up into two bins both of size $\frac{c}{2}$ and $B_{(i-1)k}$ gets split up into two bins of size $\frac{c}{2}$ and $\frac{c}{2} + 1$, respectively. If c is odd, then $B_{(i-1)j}$ gets split up into two subsets of size $\frac{c+1}{2}$ and $\frac{c+1}{2} - 1$, respectively, and $B_{(i-1)k}$ gets split up into two bins of size $\frac{c+1}{2}$. In any case the bins differ in size by at most 1. □

LEMMA 4.3. *In m rounds the algorithm* ASORT *performs less than $6nm$ comparisons.*

PROOF. The algorithm by Blum et al. [**BFP+72**] needs at most $5.73n$ comparisons to find the median of n elements and to partition the elements with respect to the median. In the i'th round ASORT partitions the elements in every bin $B_{ij}, 1 \leq j \leq 2^i$ with respect to their median. Thus the i'th round needs at most

$$\sum_{j=1}^{2^i} 5.73|B_{ij}| = 5.73 \sum_{j=1}^{2^i} |B_{ij}| = 5.73n \leq 6n$$

comparisons. As the algorithm runs for m rounds the overall number of comparisons is less than $6nm$. □

THEOREM 4.4. *Let $r = \frac{n^2}{\nu(n)}$. Any ranking consistent with the ordering of the bins computed by* ASORT *in $\log \nu(n)$ rounds, i.e., with less than $6n \log \nu(n)$ comparisons, has a Spearman's footrule distance of at most r to the actual ranking of the elements from B.*

PROOF. Using the definition of Spearman's footrule metric and Lemmas 4.1 and 4.2 we can conclude that the distance of the ranking of the

elements in B to any ranking consistent with the ordering of the bins computed by ASort in m rounds can be bounded by

$$n(\lceil \frac{n}{2^m} \rceil - 1) \leq \frac{n^2}{2^m}$$

Plugging in $\log \nu(n)$ for m gives a distance less than r as claimed in the statement of the theorem. The claim for the number of comparisons follows from Lemma 4.3. □

4.3. Lower Bound

For $r > 0$, by $B_D(\mathrm{id}, r)$ we denote the ball centered at id of radius r with respect to the Spearman's footrule metric, so

$$B_D(\mathrm{id}, r) := \{\pi \in S_n : D(\pi, \mathrm{id}) \leq r\}.$$

Next we estimate the number of permutations in a ball of radius r.

Lemma 4.5.

$$\left(\frac{r}{en}\right)^n \leq |B_D(id, r)| \leq \left(\frac{2e(r+n)}{n}\right)^n.$$

Proof. Every permutation $\pi \in S_n$ is uniquely determined by the sequence $\{\pi(i) - i\}_i$. Hence, for any sequence of non-negative integers d_i, $i = 1, \ldots, n$, there are at most 2^n permutations $\pi \in S_n$ satisfying $|\pi(i) - i| = d_i$.

If $D(\pi, \mathrm{id}) \leq r$, then $\sum_i |\pi(i) - i| \leq r$. Since the number of sequences of n non-negative integers whose sum is at most r is $\binom{r+n}{n}$, we have

$$|B_D(\mathrm{id}, r)| \leq \binom{r+n}{n} 2^n \leq \left(\frac{2e(r+n)}{n}\right)^n.$$

Next, we give a lower bound on the size of $B_D(\mathrm{id}, r)$. Let $s := \lceil \frac{n^2}{r} \rceil$, and let us first assume that n is divisible by s. We divide the index set $[n]$ into s blocks of size n/s, such that for every $i \in \{1, 2, \ldots, s\}$ the ith block consists of elements $(i-1)\frac{n}{s} + 1, (i-1)\frac{n}{s} + 2, \ldots, i\frac{n}{s}$. For every s permutations $\pi_1, \pi_2, \ldots, \pi_s \in S_{n/s}$ we define the permutation $\rho \in S_n$ to be the concatenation of the permutations applied to corresponding blocks, so $\rho := \pi_1(b_1)\pi_2(b_2)\ldots\pi_s(b_s)$. Note that the distance of ρ to id with respect to Spearman's footrule metric is at most $n \cdot n/s \leq r$, since $|\rho(i) - i| \leq n/s$, for every $i \in [n]$. Obviously, for every choice

of $\pi_1, \pi_2, \ldots, \pi_s$ we get a different permutation ρ, which means that we have at least

$$\left(\left(\frac{n}{s}\right)!\right)^s \geq \left(\frac{r}{en}\right)^n$$

different permutations in $B_D(\mathrm{id}, r)$.

If n is not divisible by s, we divide $[n]$ into s blocks of size either $\lceil n/s \rceil$ or $\lfloor n/s \rfloor$, again apply an arbitrary permutation on each of them and we can obtain the same bound in an analogous fashion. □

Remark. A slightly better upper bound on $|B_D(id, r)|$ than it was given in the lemma can be obtained using a result of Knuth. In [**Knu73**] he gives an explicit formula for the number $I_n(k)$ of permutations of n elements having exactly k inversions. It can be shown that $I_n(k) \leq \binom{n+k-1}{k}$. Therefore the size of the ball $B_K(id, r)$ with respect to Kendall's tau metric K centered at id and having radius r can be upper bounded as

$$|B_K(id, r)| = \sum_{k=0}^{r} I_n(k) \leq \sum_{k=0}^{r} \binom{n+k-1}{k} = \binom{n+r}{n}.$$

Diaconis showed that for every $\pi \in S_n$

$$K(\pi) \leq D(\pi) \leq 2K(\pi),$$

where $K(\pi)$ is Kendall's distance of π to id. It follows that $B_D(id, r) \subseteq B_K(id, r)$ and therefore

$$|B_D(id, r)| \leq |B_K(id, r)| \leq \binom{n+r}{n} \leq \left(\frac{e(n+r)}{n}\right)^n,$$

which implies the upper bound that we have shown in the lemma.

Using this upper bound we now give a lower bound for the worst case running time of any comparison based approximate sorting algorithm.

THEOREM 4.6. *Let $\mathcal{A}$ be a randomized approximate sorting algorithm based on comparisons, let $\nu = \nu(n)$ be a function, and let $r = r(n) = \frac{n^2}{\nu(n)}$.*

If for every input permutation $\pi \in S_n$ the expected Spearman's footrule distance of the output to id *is at most r, then the algorithm performs at least $n\,(\min\{\log \nu, \log n\} - 6)$ comparisons in the worst case.*

PROOF. The proof of this theorem follows *Yao's Minimax Principle*, which is detailed for example in [**MR95**]. Let k be the smallest integer

such that $\mathcal{A}$ performs at most k comparisons for every input. For a contradiction, let us assume that

$$k < n\left(\min\{\log \nu, \log n\} - 6\right).$$

First, we are going to prove

$$\frac{1}{2}n! > 2^k \left(\frac{2e(2r+n)}{n}\right)^n. \tag{4.1}$$

Since $\log \nu - 6 > k/n$, we have $\frac{\nu}{2^6} > 2^{k/n}$ and since $\nu = \frac{n^2}{r}$ we get

$$\frac{n}{2e} > 2^{k/n}\frac{2e \cdot 2r}{n}. \tag{4.2}$$

On the other hand, from $\log n - 6 > k/n$ we get $\frac{n}{2^6} > 2^{k/n}$ implying

$$\frac{n}{2e} > 2^{k/n}\frac{2e \cdot n}{n}. \tag{4.3}$$

Putting (4.2) and (4.3) together, we obtain

$$\frac{n}{e} > 2^{k/n}\frac{2e(2r+n)}{n}.$$

Hence

$$\frac{1}{2}n! \geq \left(\frac{n}{e}\right)^n > 2^k \left(\frac{2e(2r+n)}{n}\right)^n,$$

proving (4.1).

We denote by R the source of random bits for $\mathcal{A}$. One can see R as the set of all infinite 0-1 sequences, and then the algorithm is given a random element of R along with the input. For a permutation $\pi \in S_n$ and $\alpha \in R$, we denote by $\mathcal{A}(\pi, \alpha)$ the output of the algorithm with input π and random bits α.

We fix $\widetilde{\alpha} \in R$ and run the algorithm for every permutation $\pi \in S_n$. Note that with the random bits fixed the algorithm is deterministic. For every comparison made by the algorithm there are two possible outcomes. We partition the set of all permutations S_n into classes such that all permutations in a class have the same outcomes of *all the comparisons* the algorithm makes. Since there is no randomness involved, we have that for every class C there exists a $\sigma \in S_n$ such that for every $\pi \in C$ we have $\mathcal{A}(\pi, \widetilde{\alpha}) = \sigma \circ \pi$, where $\circ$ is the multiplication in the permutation group S_n. In particular, this implies that the set $\{\mathcal{A}(\pi, \widetilde{\alpha}) : \pi \in C\}$ is of size $|C|$. On the other hand, since the algorithm in this setting is deterministic and the number of comparisons of the algorithm is at most k, there can be at most 2^k classes. Hence, each permutation in S_n is the

output for at most 2^k different input permutations. From Lemma 4.5 we have $|B_D(\mathrm{id}, 2r)| \leq \left(\frac{2e(2r+n)}{n}\right)^n$, and this together with (4.1) implies that at least

$$n! - 2^k \left(\frac{2e(2r+n)}{n}\right)^n > \frac{1}{2}n!$$

input permutations have output at distance to id more than $2r$.

Now, if both the random bits $\alpha \in R$ and the input permutation $\pi \in S_n$ are chosen at random, the expected distance of the output $\mathcal{A}(\pi, \alpha)$ to id is more than r. Therefore, there exists a permutation π_0 such that for a randomly chosen $\alpha \in R$ the expected distance $d_D(\mathcal{A}(\pi_0, \alpha), \mathrm{id})$ is more than r. Contradiction. □

4.4. Conclusion

We showed that any comparison based, randomized algorithm in the worst case needs at least

$$n\left(\min\{\log \nu(n), \log n\} - 6\right)$$

comparisons to approximate a given ranking of n items within expected distance $n^2/\nu(n)$. This result is complemented by an algorithm that shows that $6n(\log \nu(n))$ comparisons are always sufficient.

In particular, this means that in some cases substantially less comparisons have to be performed than for sorting exactly, provided that a sufficiently large error is allowed. That is, as long as the desired expected error is of order $n^{2-\alpha}$ for constant α one needs $\Omega(n \log n)$ comparisons, which asymptotically is not better than sorting exactly. But to achieve expected error of order $n^{2-o(1)}$ only $o(n \log n)$ comparisons are needed.

This result is interesting as a theoretical result, however, it does not help for problems arising in market research. Usually the item sets are very large in practice, especially when they are described by a conjoint structure. Then it is not even feasible to ask a respondent a linear number of choice tasks. Therefore, in the subsequent chapter we will introduce additional modeling assumption and use aggregation in order to derive a scale for the whole population.

CHAPTER 5

Modeling and Aggregation

5.1. Introduction

In this chapter we describe a method to define an interval scale for a set of items with conjoint structure from choices in paired comparisons. We will use aggregation and define an interval scale for a population of respondents. To this end we use Thurstone's method of comparative judgment [**Thu27**] — which was designed for the structureless case — and extend it to the conjoint structure case. In a nutshell our method works as follows: At first we estimate scale values for all levels of a single attribute. To this end we interpret any paired comparison as a comparison of just the two levels of the given attribute that are present in this comparison, ignoring differences in the levels of all other attributes. We can apply this method to all attributes to obtain scale values for all levels of all attributes. However, the scale values for levels of different attributes need not to be comparable yet. To make all partworths comparable we design a rescale method. It builds on the fact that for the computation of the partworths for any attribute, always the same stated preferences are used, namely, the outcomes of all paired comparisons. Finally, the scale value of an item is just the sum of the scale values (partworths) of the levels describing the item.

Like most techniques in conjoint analysis our method builds on a statistical model. Our method has the advantage that all model assumptions are testable, which is not the case with off-the-shelf conjoint analysis software like Sawtooth' software [**Sof**].

5.2. The structureless case: Thurstone's method

A good introduction into Thurstone's work is given in [**Eng00**]. Here we summarize it only briefly: Our goal is to define a meaningful scheme to assign scale values on an interval scale to k items that we label by

$1, \ldots, k$. Thurstone's intuition was that the relative frequency $F_{i \succ j}$ that i was preferred over j by the respondents is an indirect measure for the distance between i and j. He derived an interval scale from this intuition under the assumption that the scale values S_i of the items i are uncorrelated normally distributed random variables with expectations μ_i and variances $\sigma_i^2 \equiv \sigma^2$, i.e., all the variances are the same.

The idea is to assign to each item i the value μ_i, on an interval scale which still has to be defined. To do so we need to estimate all the μ_i's from the paired comparison data that we have available. It turns out that it is easier to estimate the differences $\mu_i - \mu_j$. Instead of assigning to item i the value μ_i we assign the the scale value

$$\lambda_i = \frac{1}{k} \sum_{j=1}^{k} (\mu_i - \mu_j) = \mu_i - \frac{1}{k} \sum_{j=1}^{k} \mu_j =: \mu_i - \bar{\mu}.$$

That is, we only shift the scale that assigns μ_i to item i by the value $\bar{\mu}$, i.e., as interval scales both scales are the same. By the properties of normal distributions also the differences $S_i - S_j$ are normally distributed with expectations $\mu_i - \mu_j$ and variance $2\sigma^2$. This yields

$$\begin{aligned}
P[S_i - S_j > 0] &= \frac{1}{\sqrt{4\pi\sigma^2}} \int_0^\infty e^{-\frac{(x-(\mu_i-\mu_j))^2}{4\sigma^2}} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\frac{\mu_i-\mu_j}{\sqrt{2}\sigma}}^\infty e^{-\frac{x^2}{2}} dx \\
&= 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\frac{\mu_i-\mu_j}{\sqrt{2}\sigma}} e^{-\frac{x^2}{2}} dx \\
&= 1 - \Phi\left(-\frac{\mu_i - \mu_j}{\sqrt{2}\sigma}\right) \\
&= \Phi\left(\frac{\mu_i - \mu_j}{\sqrt{2}\sigma}\right),
\end{aligned}$$

where Φ is the cumulative distribution function

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy.$$

of the standard normal distribution. Hence,

$$\mu_i - \mu_j = \left(\sqrt{2}\sigma\right) \Phi^{-1}\big(P[S_i - S_j > 0]\big).$$

We can estimate $P[S_i - S_j > 0]$ by the observed quantity $F_{i\succ j}$, i.e., the relative frequency that item i was preferred over j and thus estimate $\mu_i - \mu_j$ by $\sqrt{2}\sigma\Phi^{-1}\big(F_{i\succ j}\big) =: \widehat{\mu_{ij}}$. Note that these estimated differences in general are not consistent in the sense that in general there do not exist scale values $\lambda_i \in \mathbb{R}$ that satisfy $\lambda_i - \lambda_j = \widehat{\mu_{ij}}$ for all $i, j \in \{1, \ldots, k\}$. To define the scale values λ_i we use a least square approach, i.e. we choose the λ_i such that $\sum_{i,j}((\widehat{\mu_{ij}} - (\lambda_i - \lambda_j))^2$ is minimized. In order to get a unique solution we introduce the constraint that all scale values add up to zero. It turns out that the least squares solution for the scale values of the items can be easily determined as the column average of the matrix of estimated differences

$$\mathbf{S} = \begin{pmatrix} 0 & \widehat{\mu_{21}} & \ldots & \widehat{\mu_{k1}} \\ \widehat{\mu_{12}} & 0 & \ldots & \widehat{\mu_{k2}} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\mu_{1k}} & \widehat{\mu_{2k}} & \ldots & 0 \end{pmatrix}.$$

That is, the scale value for item i can be estimated as

$$\lambda_i = \frac{1}{k}\sum_{j=1}^{k}\widehat{\mu_{ij}} = \frac{\sqrt{2}\sigma}{k}\sum_{j\neq i}\Phi^{-1}\big(F_{i\succ j}\big).$$

The choice of σ essentially fixes the scale, but the ratio of differences of scale values is not affected by the choice of σ, i.e., any fixed choice of σ would work. A natural but arbitrary choice is $\sigma = 1$.

5.3. Conjoint structure case

In order to define the aggregated scale in the conjoint structure case we assume that the population's scale can be represented by an additive interval scale. Let us recall that in conjoint analysis we assume that the items come from set A with conjoint structure, i.e., $A = A_1 \times \ldots \times A_n$. Thus we have choice data (from paired comparisons), where the elements in A were compared. The set A is typically fairly large and we do not have enough choice information to apply Thurstone's method directly. Instead we take a decompositional approach and first order the levels in each of the attributes $A_i, i = 1, \ldots, n$ on interval scales. For that purpose, whenever in a choice task $(a_1, \ldots, a_n) \in A$ was preferred over $(b_1, \ldots, b_n) \in A$, we consider this as a_i was preferred b_i, provided $a_i \neq b_i$. That is, we derive choice information on the attribute level from choice information on the item level. We then can apply Thurstone's method to

the choice data on attribute level to get a scale for each attribute. When applying Thurstone's method we set all variances to 1, which fixes the scales for all the attributes but does not necessarily make these scales comparable. For any attribute A_i with levels $a_{i1}, \ldots, a_{ik_i}$ let $\lambda_{i1}, \ldots, \lambda_{ik_i}$ be the scale values that we get from applying Thurstone's method on the attribute level. The next step is to aggregate these scales. We make the following assumption that extends the distribution assumption needed for Thurstone's method for a single attribute:

Assumption. For any attribute A_i the scale values $S_{i1}, \ldots, S_{ik_i}$ are all normally distributed with variance σ_{i1}^2 and expectation drawn from another normal distribution with expectation 0 and variance σ_{i2}^2. That is, we assume that the scale values for the levels of attribute A_i are drawn from a normal distribution N_i with expectation 0 and variance $\sigma_{i1}^2 + \sigma_{i2}^2$ (the convolution of the two normal distribution functions introduced before).

As when applying Thurstone's method the value σ_{i1}^2 is the same for all the S_{ij}, but not necessarily 1. Later it will be chosen such that the scales for all attributes become comparable, i.e., the scaled scale values $\sigma_{i1}\lambda_{ij}$ (our final partworths) will be comparable. Now remember that we derived the scales on the attribute level from paired comparisons on the item level. That is, all the distributions $N_i, i = 1, \ldots, n$ should describe the distribution of scale values on the item level, i.e., the item scale values should follow the N_i distributions and all these distributions should be the same, i.e.,

$$\sigma_{i1}^2 + \sigma_{i2}^2 = \sigma_{j1}^2 + \sigma_{j2}^2 \equiv 1 \quad \text{for all attributes } A_i \text{ and } A_j,$$

here the value 1 is arbitrary (we just need to choose one fixed value). Note that if we knew the values σ_{i2}, then these equalities would determine the values for the σ_{i1} (that we kept variable so far) and by that make the scales for all the attributes comparable. We can estimate the σ_{i2} from the scaled observed scale values $\sigma_{i1}\lambda_{ij}$ by

$$\sqrt{\frac{1}{k_i - 1} \sum_{j=1}^{k_i} \left(\sigma_{i1}\lambda_{ij} - \frac{1}{k_i} \sum_{i=1}^{k_i} \sigma_{i1}\lambda_{ij} \right)^2} = \sqrt{\frac{\sigma_{i1}^2}{k_i - 1} \sum_{j=1}^{k_i} \lambda_{ij}^2}.$$

Remember that we chose the scale values such that $\sum_{j=1}^{k_i} \lambda_{ij} = 0$. Using this estimate we can solve $\sigma_{i1}^2 + \sigma_{i2}^2 = 1$ for σ_{i1} to estimate σ_{i1} as

$$\sigma_{i1} = \frac{1}{\sqrt{1 + \frac{1}{k_i - 1} \sum_{j=1}^{k_i} \lambda_{ij}^2}}.$$

That is, in order to make the scales for the different attributes comparable we need to rescale the λ_{ij} that we computed with Thurstone's method (with constant variance 1) by this estimate of σ_{i1}.

Since now the scales of all the attributes are comparable we can get the scale value for an item just as the sum of the partworths of the attribute levels involved, i.e., the scale value of $(a_{1j_1}, \ldots, a_{nj_n}), a_{ij_i} \in A_i$ is given as $\sum_{i=1}^{n} \sigma_{i1} \lambda_{ij_i}$. Note also that on comparable scales each value σ_{i2} can be interpreted as a measure of how important attribute A_i is (contributes larger values to the sum). But we have to be careful, the additivity assumption does not hold for all practical applications. It only holds when the attributes are independent. Later when computing scale values on the item level we will explicitly test if the additivity assumption holds.

5.3.1. Theoretical sample size error. Let us briefly describe our error analysis. Our observed quantities are the relative frequencies $F_{i \succ j}$. We assume that any comparison of items i and j is an independent Bernoulli trial with success probability p (here "success" means that i is preferred over j). We want to estimate p by $F_{i \succ j}$. For Bernoulli trials $F_{i \succ j}$ converges to p with increasing number of repetitions of the Bernoulli trial, but here we make only a finite number m_{ij} of comparisons which procures some error. To estimate this error note that by our assumption of independent Bernoulli trials $m_{ij} F_{i \succ j}$ is drawn from binomial distribution with expectation $m_{ij} p$ and variance $m_{ij} p(1-p)$. By the central limit theorem the binomial distribution can be approximated by a normal distribution with the same expectation and variance. By rescaling with $1/m_{ij}$ we get a normal distribution with expectation p and variance $p(1-p)/m_{ij}$. We estimate this variance by

$$\frac{F_{i \succ j} \left(1 - F_{i \succ j}\right)}{m_{ij}}$$

and the error of our estimate of p by the square root of the variance, i.e., by the standard deviation

$$\sqrt{\frac{F_{i\succ j}\,(1-F_{i\succ j})}{m_{ij}}}.$$

To compute errors of our scale values we use error propagation (see Subsection 5.3.3).

5.3.2. Experimental error. We will also simulate errors by randomly dividing the respondents into two groups. For each group we can compute the scale values for all attribute levels on comparable scales as described above. So we get for each attribute level a scale value from each group. Averaging the absolute difference of these two scale values over several random groupings of the respondents provides us with an experimental error for the scale value of this attribute level. Similarly we also compute experimental errors by randomly dividing the choice tasks (paired comparisons) into two groups.

5.3.3. Error propagation. To compute errors of our scale values we use error propagation. Briefly, in error propagation we have a function $f(x_1,\ldots,x_n)$ depending on uncorrelated observable quantities $x_1,\ldots,x_n$. Assume the errors (standard deviations) of the observed quantities are $\delta_{x_1},\ldots,\delta_{x_n}$. The variation of f with respect to x_i is given by $\partial f/\partial x_i$. Since the error in x_i will only increase the error in f we weight the error in x_i with the absolute value $|\partial f/\partial x_i|$ of the variation to obtain $\delta_i = |\partial f/\partial x_i|\,\delta_{x_i}$ for the error of f incurred by the error in x_i. The total error in f is then obtained from the individual errors as $\delta_f = \sqrt{\delta_1^2+\ldots+\delta_n^2}$.

5.4. Testing the model

In our model of scale values we made two assumptions, one on the attribute level and one on the item level. The assumption on the attribute level is, that the scale values for all levels of a given attribute are uncorrelated and have the same variance, and the assumption on the item level is, that the scale value of an item is the sum of the partworths of its attribute levels.

5.4.1. Additivity assumption (linear model). Here we want to describe how to test the second assumption of our model, i.e., the additivity assumption. Let A_1 and A_2 be two attributes and let $C = A_1 \times A_2$ be the new attribute that results from combining A_1 and A_2 and let $c_1, \ldots, c_k$ be its levels. We compute scale values for the levels of C in two different ways. First, for every level $c_i = (a_{i1}, a_{i2})$ with $a_{i1} \in A_1$ and $a_{i2} \in A_2$ we add up the comparable scale values for a_{i1} and a_{i2} that we compute as described before. Let $\lambda_1, \ldots, \lambda_k$ be the resulting scale values. Second, we apply Thurstone's method directly to the combined attribute C and make the resulting scale values comparable with the scales values of all levels of attributes different from A_1 and A_2. This results in scale values $\lambda'_1, \ldots, \lambda'_k$.

If additivity holds, then we expect that $\lambda_i \approx \lambda'_i$. Our scale values are the expectations of normal distributions, i.e., the mean value parameter of such a distribution. Due to our sample size error that we compute as described in Section 5.3.1 we have that these parameters are also normally distributed themselves with expectations λ_i and λ'_i, respectively, and with variance σ_i^2 and $\sigma_i'^2$, respectively. Here σ_i and σ'_i are computed by error propagation from the errors of the observed frequencies as described in Section 5.3.3. Thus, our null hypothesis is that $\lambda_i = \lambda'_i$ for all $1 \leq i \leq k$. As test statistic we use

$$\chi^2 = \sum_{i=1}^{k} \frac{(\lambda_i - \lambda'_i)^2}{\sigma_i^2 + \sigma_i'^2} =: \sum_{i=1}^{k} y_i^2,$$

where σ_i and σ'_i are computed by error propagation from the errors of the observed frequencies. If the null hypothesis is true then the test statistic χ^2 is χ^2-distributed with $k-1$ degrees of freedom. To see this note that y_i is a realization of a standard normally distributed random variable. As $k-1$ scale values already determine the remaining scale value (the scale values have mean 0) the text statistic χ^2 has $k-1$ degrees of freedom. The null hypothesis is rejected at a significance level of α if $\chi^2 > \chi^2_{1-\alpha,k-1}$ where $\chi^2_{1-\alpha,k-1}$ is the $1-\alpha$ quantile of the χ^2-distribution with $k-1$ degrees of freedom.

5.4.2. Mosteller's test. On the attribute level we make the assumptions that the scale values are uncorrelated normally distributed with with equal variances. A test for this assumption was devised by Mosteller [**Mos51**] and is also described in [**Eng00**]. Here we only briefly review Mosteller's test, which boils down to test if our model can explain the

observed frequencies $F_{i \succ j}$. To this end we compute

$$p_{ij} = \frac{1}{2} \int_0^\infty e^{-\frac{(x-(\lambda_i-\lambda_j))^2}{4}} dx = \Phi\left(\frac{\lambda_i - \lambda_j}{\sqrt{2}}\right),$$

where we use λ_i and λ_j as computed by Thurstone's method with $\sigma = 1$. Then we transform both $F_{i \succ j}$ and p_{ij} into angles θ_{ij} and θ'_{ij}, respectively, using the arcsine transformation given by

$$\theta_{ij} = \arcsin\left(2F_{i \succ j} - 1\right) \quad \text{and} \quad \theta'_{ij} = \arcsin\left(2p_{ij} - 1\right)$$

The arcsine transformation converts binomially distributed frequencies into asymptotically normally distributed variables with variance $1/m_{ij}$, where m_{ij} is the number of comparisons of level i with level j for the given attribute. The null hypothesis is that θ_{ij} is normally distributed with expectation θ'_{ij} and variance $1/m_{ij}$ for all $i < j$. As test statistic we use

$$\chi^2 = \sum_{i<j} m_{ij}(\theta_{ij} - \theta'_{ij})^2.$$

If the null hypothesis is true then the test statistic χ^2 is χ^2-distributed with $\binom{n-1}{2}$ degrees of freedom. Thus, at level α we have to compare our test statistic to the $1 - \alpha$ quantile of the χ^2-distribution with $\binom{n-1}{2}$ degrees of freedom.

5.5. Conclusion

We have introduced a framework to derive an aggregated interval scale for a set of items with conjoint structure. The framework builds on aggregation and on additional modeling assumptions. Our framework permits to test all these modeling assumptions. This is a big advantage of our framework in comparison to off-the-shelf conjoint analysis software like Sawtooth' software [**Sof**] where modeling assumptions cannot be tested. Our framework can be applied to very large item sets which are described by a conjoint structure. In particular the introduced modeling assumptions permit to compute scale values for items that have never been compared in any choice task.

This chapter concludes Part 1. In Part 2, the application part, we apply the introduced framework to parameterized multimedia algorithms from two applications: volume visualization and gamut mapping. On the one hand this allows valuable insights in the influence of different parameters on people's quality perception in these areas. On the other hand we will

test if our framework is suited for practical problems and in particular if practical data sets fulfill the modeling assumptions.

Part 2

Applications

In Part 1 of this thesis we provided the theoretical foundations for conjoint analysis and devised a new method to derive an interval scale for a class of items with conjoint structure. Here we use this method to measure the perceived quality of parameterized multimedia algorithms. We consider a multimedia algorithm to be an algorithm whose output is an image, a video or an audio file. Many multimedia algorithms have in common that their quality not only depends on objective evaluation criteria like running time or memory consumption but also on the human perception of the output.

Here we suggest to measure the perceived quality of parameterized multimedia algorithms in the framework of conjoint analysis. The crucial observation is that a parameterized algorithm is nothing else but a class of items with conjoint structure. Every parameter can be seen as an attribute and every parameter value as an attribute level. An item then corresponds to the output of the algorithm for a certain parameter setting. The goal is to measure the perceived quality of outputs for the different parameter settings on a scale.

We use conjoint analysis to measure the perceived quality of multimedia algorithms from two different applications: gamut mapping and visualization. Both areas have in common that the output of the algorithms are images. In Chapter 6 we develop a new gamut mapping algorithm and compare it in a user study to standard reference algorithms. As the number of algorithms to be compared is small we can apply Thurstone's law of comparative judgement directly here without extending it to the case of more than one attribute. In Chapter 7 we report on a large user study that we conducted to measure the perceived quality of the different parameter settings for an existing parameterized volume visualization algorithm. For the evaluation we use the method introduced in Chapter 5.

CHAPTER 6

Image-Dependent Gamut Mapping

6.1. Introduction

A gamut is the entirety of colors that are contained in an image or that can be reproduced by a device like a monitor screen or a printer. As such a gamut is a subset of a color space. Typically the color space is three dimensional and gamuts are (finite) subsets of it. *Gamut mapping* is the process of adapting a color image to device gamut limitations, such that the original color appearance is preserved as well as possible. It is a fundamental task in digital color reproduction. Here we understand gamut mapping as a point-to-point mapping of color points from a source to a device gamut.

Example 6.1. An image can contain more and/or different colors than a printer can reproduce. Therefore the non-printable colors need to be replaced by printable colors during the printing process. A rather trivial way to do so is to replace every non-printable color by *white*. Figure 6.1 illustrates the result of this gamut mapping algorithm for a photo printer, a coated offset printer (used for example in package printing) and a newspaper printer. Note that the photo printer has the largest gamut and the newspaper printer has the smallest one. It becomes clear from Figure 6.1 that the suggested algorithm is not the best way to do gamut mapping. In Section 6.3 we will develop an algorithm that yields better results.

Gamut mapping algorithms can be classified along several dimensions. Here we are interested in one particular dimension, namely, we distinguish device-to-device and image-to-device techniques, depending on whether the source gamut is derived from the color properties of the source device or from the colors found in a specific image. In other words device-to-device mappings are defined for all color points that are potentially contained in an image whereas image-to-device algorithms

FIGURE 6.1. On top: the original image. Below: Gamut mapping of the original image where all non-printable colors are set to white for a photo printer, a coated offset printer and a newspaper printer (from left to right).

only consider color points actually contained in the specific image gamut. Device-to-device gamut mapping algorithms typically are very fast since for any image color point its mapping into the device gamut can simply be determined by a fast table look up in a pre-computed table. In contrast to that image-to-device algorithms have to determine the mapping for each image individually. This increase in complexity should lead to higher quality mappings, i.e., mappings that preserve color characteristics better than device-to-device mappings. In this chapter we want to explore the potential of image-to-device gamut mapping—also referred to as image-dependent gamut mapping. To this end we introduce a whole class of image-dependent gamut mapping algorithms and compare them in a field study to standard reference algorithms.

Our approach towards image-dependent gamut mapping is via constrained optimization: for every image we distort its colors as little as possible while preserving typical characteristics of the image. The optimization approach allows us to combine different gamut mapping concepts in a modular way. In a field study we determined the perceived visual quality induced by specific components of our approach.

6.2. Preliminaries

6.2.1. Gamut Mapping. For a detailed overview over the wide variety of gamut mapping techniques see the paper by Morovic [**Mor03**]. Here we only give a short review of the concepts most important to our work. As we already mentioned in the introduction, gamut mapping algorithms can be classified along several dimensions one of which is the device-to-device vs. image-to-device categorization. Another categorization is into global and local algorithms: in contrast to global algorithms local algorithms, like for example [**KSES05, BdEW01, ZS07**], also take the colors in the neighborhood of a pixel into account. That is, the mapping of a color point depends on where in an image the color appears. Note that local mappings are one-to-many, i.e., the same image color point can be mapped to different device color points depending on its context.

Most of todays gamut mapping algorithms are global and device-to-device [**Mor98, BF99, MS01, KIO99**], in particular in connection with ICC color management. However image-to-device algorithms can in general be expected to perform better in color rendering than device-to-device. The basic idea behind the image-to-device concept [**GWA90, MF97, KS03**] is to determine the shape and size of the source gamut by image statistics. Therefore the definition and description of a gamut boundary surface is a key issue for which a series of methods have been proposed, like for example the segment maxima boundary description [**ML00**], alpha shapes [**CL99**] and flow shapes [**GSSZ05**].

Except for clipping algorithms, most gamut mapping algorithms make use of gamut border colors to define the mapping of the in-gamut colors. The performance of these algorithms thus heavily depends on the appropriate definition of a gamut boundary surface. To get an appropriate boundary surface is not always an easy task, especially for image gamuts that in general have much more complex shapes than device gamuts. Device gamuts are typically described by convex (or close to convex) sets. Moreover some algorithms, see for example [**Cen04**], rely on a unique definition of a cusp within a hue plane of the gamut, a notion which is problematic for image gamuts. Therefore the use of most image gamut descriptions may require a subsequent smoothing of the gamut boundary surface. Alternatively additional constraints on the mapping of neighboring colors can be imposed, a method, which will be used in this work.

Here we do not consider local algorithms but it is an interesting questions open for future research if the techniques that we are going to present here can be localized.

6.2.2. Mathematical Optimization. Since we want to treat gamut mapping as an optimization problem we also shortly review the basics of mathematical optimization. A mathematical optimization problem has the general form

$$\begin{array}{ll} \text{minimize}_x & f(x) \\ \text{subject to} & c_i(x) \leq 0, \quad i = 1 \ldots n, \end{array}$$

where $f : \mathbb{R}^d \to \mathbb{R}$ is called objective function, the $c_i : \mathbb{R}^d \to \mathbb{R}$ are functions that define constraints and $x \in \mathbb{R}^d$ is the optimization variable. Note that a maximization problem obeys this general form if one multiplies its objective function by -1. A point in $\mathbb{R}^d$ that satisfies all constraints is called feasible point. A point $x^* \in \mathbb{R}^d$ is a local optimum for the problem if there exists a neighborhood $U(x^*) \subset \mathbb{R}^d$ such that $f(x^*) \leq f(x)$ for all feasible $x \in U(x^*)$. The point $x^* \in \mathbb{R}^d$ is a global optimum if $f(x^*) \leq f(x)$ for all $x \in \mathbb{R}^d$ that are feasible. In general a mathematical optimization problem neither needs to have a local nor a global minimum.

The probably best known special class of optimization problems are linear programs, where the objective function f and all constraint functions c_i are linear in the optimization variable x. One reason for the popularity of linear programs is that there are many efficient algorithms known to solve them. Another class of optimization problems that also can be solved efficiently are convex quadratic programs, where the objective function is convex and quadratic in x while the constraints are linear in x. Both, linear- and convex quadratic programs always have a global optimum, provided that the feasible region is bounded and not empty and the global optimum can be computed efficiently. This is in contrast to non-convex optimization problems, which in general cannot be solved efficiently. Already problems with about ten variables can be extremely challenging, and problems with a few hundreds of variables can already be intractable. For more details on convex optimization see for example Boyd and Vandenberghe [**BV04**].

6.3. Algorithm

6.3.1. Gamut Mapping as Optimization Problem. It seems quite natural to formulate gamut mapping as a multi-criteria optimization problem since there are several (competing) objectives to take care of [**ML01, ZS06**]:

(1) Detail preservation: image details should be preserved by the mapping.
(2) Hue preservation: the hue of a color should not be changed by the mapping.
(3) Gray axis preservation: a gray color point should be mapped onto a gray color point. In all color spaces that are usually considered for gamut mapping the gray colors lie on an axis of the color space. This is where the expression *gray axis* comes from.
(4) Continuity of the mapping: color points that are perceptually close should also stay close under the mapping.
(5) Low image distortion: a color should not be changed "too much" by the mapping.

One popular approach to multi-criteria optimization is to search for a so called Pareto optimum [**Ehr05**], i.e., a solution that is not dominated by another feasible solution. In general an optimization problem has many Pareto optima, which by definition cannot be compared to each other. The Pareto approach is usually taken if all the different criteria (objectives) are equally important (or cannot be compared to each other). In gamut mapping the objectives that we listed above are not always equally important depending on the intended application.

Therefore we take another approach towards image-dependent gamut mapping by treating different objectives differently. Preservation and continuity objectives are encoded as hard constraints, i.e., we prescribe the extent to which they are allowed to be violated by providing strict bounds, whereas low image distortion is an objective that we want to optimize under the condition that the preservation and continuity constraints are satisfied. The rationale behind this approach is that strong violations of detail preservation, hue preservation, gray axis preservation or continuity of the mapping create artifacts that harm the perceived quality drastically and thus have to be avoided by providing hard constraints.

In order to formulate the mathematical optimization problem for gamut mapping we will discuss the geometric setup of gamut mapping as we need it here. The image gamut I and the target gamut T are finite subsets (point clouds) of our working color space. We assume that the working color space is

(1) approximately hue preserving, i.e. all color points in a plane which contains the gray axis have the same hue and
(2) approximately perceptually uniform (equidistant), i.e., equal Euclidean distances in color space correspond to equal distances in visual perception.

Color spaces like OSA-UCS, hue-linearized CIELAB (see for example [**BFCE98, FE98**]) or DIN 6164 fulfill these properties. Note that the color space is one parameter of our optimization problem and that the overall quality of our mapping will depend on the extent to which it fulfills the two required properties. As the color space is approximately perceptually uniform one can treat gamut mapping as a problem in Euclidean geometry, where a point cloud that describes the image gamut has to be mapped into a shape described by the target gamut.

Let us for the moment assume that we have the following operations available:

(1) We can compute a continuous shape $\textsc{Shape}(T)$ that approximates the target gamut T. The shape should capture the geometry of the target gamut well as the quality of the overall mapping will depend on the quality of the computed shape.
(2) We can determine for every point in the working color space, if it is contained in $\textsc{Shape}(T)$ or not.
(3) We can determine the intersection points of a line with $\textsc{Shape}(T)$.

As we mentioned earlier the objective function of our optimization problem is low image distortion. Intuitively the image distortion is low if the points in the image gamut I are not displaced too much by the mapping from I into $\textsc{Shape}(T)$. In order to give sense to the term "displace too much" we need some metric on the color space. Here we want to use the Euclidean metric, which also explains our assumption that the color space has to be approximately perceptually uniform. The following optimization problem is a candidate to model low image distortion:

$$\begin{aligned} &\text{minimize}_f \quad \textstyle\sum_{x\in I} \|x - f(x)\|^2 \\ &\text{subject to} \quad f(x) \in \textsc{Shape}(T) \quad \text{for all } x \in I \end{aligned} \tag{6.1}$$

The objective function takes care of a low image distortion while the constraints just state that the solution should give a valid gamut mapping, i.e., the points from the image gamut are actually mapped into the target gamut. We optimize over all functions $f : \mathbb{R}^d \to \mathbb{R}^d$, where d is the dimension of the color space. This means that the objective function is a functional. In this form the optimization problem does not fit into the class of optimization problems as we have introduced them earlier. To make it fit we restrict the class of functions over that we optimize to a parameterized family of functions, which leaves us to optimize the parameters. In order to restrict the family of functions we exploit our second assumption on the color space, namely, that it is approximately hue preserving. This allows us to choose a focal point c in $\textsc{Shape}(T)$ and restrict the mapping of any point $x \in I$ to the ray originating in c and shooting in the direction of x. This ensures that the hue of x is preserved. Thus the family of allowed functions is given as

$$f(x) = c + \lambda_x(x - c), \qquad \lambda_x \geq 0,$$

where the ray is parameterized by λ_x. Note that the center point c is a parameter of our optimization problem. If we assume that $\textsc{Shape}(T)$ is star shaped with respect to c, i.e., that for every point $x \in \textsc{Shape}(T)$ the line segment $\overline{xc}$ is completely contained in $\textsc{Shape}(T)$, then the constraint $f(x) \in \textsc{Shape}(T)$ for $x \in I$ can be written as

$$0 \leq \lambda_x \leq \hat{\lambda}_x,$$

where $\hat{\lambda}_x$ is such that $c + \hat{\lambda}_x(x - c)$ is the intersection of the boundary of $\textsc{Shape}(T)$ with the ray originating in c and shooting in the direction of x. The restriction that $\textsc{Shape}(T)$ has to be star shaped is not really necessary. To deal with not star shaped shapes one just has to add additional constraints for each intersection point of the shape with the ray. The only reason we restrict ourselves here to star shaped shapes is to keep our exposition as simple as possible.

Since $\|x - f(x)\|^2 = (1 - \lambda_x)^2\|x - c\|^2$ our optimization problem now reads as follows:

$$\begin{array}{ll} \min_\lambda & \sum_{x\in I}(1-\lambda_x)^2\|x-c\|^2 \\ \text{subject to} & \\ & \begin{array}{rcll} \lambda_x-\hat{\lambda}_x & \leq & 0 & \text{for all } x\in I \\ -\lambda_x & \leq & 0 & \text{for all } x\in I \end{array} \end{array} \tag{6.2}$$

This is an optimization problem that falls into the class that we introduced earlier. But we are not done yet. So far we have only modeled our goals of low image distortion and hue preservation. However we do not want to neglect the other goals of gamut mapping. We take care of them by adding additional constraints to the optimization problem.

Detail preservation. To ensure detail preservation we add the constraints $c_1\|x-y\| \leq \|f(x)-f(y)\|$ for all pairs $x, y \in I$. Here $c_1 > 0$ is a constant that again is a parameter of our optimization problem. Informally speaking the added constraint enforces that points that have a large distance to each other cannot be mapped close together.

Continuity. Zolliker [**ZS06**] pointed out that the mapping should be continuous in the following sense: points that are close to each other should not be mapped too far away from each other. We take care of this goal by adding the constraints $\|f(x)-f(y)\| \leq c_2\|x-y\|$ for all pairs $x, y \in I$. Again $c_2 > c_1$ is a constant that is a parameter of our optimization problem.

Gray axis preservation. To ensure this goal it suffices to impose an additional condition on the position of the focal point: it has to be located on the gray axis. As all mappings that we allow in the optimization problem move points towards the focal point on a straight line they all map gray axis points onto gray axis points.

Combining everything we finally end up with Optimization Problem (7.3). We summarize the input and the parameters of the optimization problem in Table 1.

6.3.2. Turning the optimization problem feasible. The optimization problem for image-dependent gamut mapping that we derived in the previous section from theoretical considerations seems infeasible to

$$
\begin{aligned}
&\min_{\lambda} \quad \textstyle\sum_{x \in I} (1 - \lambda_x)^2 \|x - c\|^2 \\
&\text{subject to} \\
&\qquad\qquad \begin{aligned}
c_1\|x - y\| - \|f(x) - f(y)\| &\leq 0 \quad \forall x, y \in I, x \neq y \\
\|f(x) - f(y)\| - c_2\|x - y\| &\leq 0 \quad \forall x, y \in I, x \neq y \\
\lambda_x - \hat{\lambda}_x &\leq 0 \quad \forall x \in I \\
-\lambda_x &\leq 0 \quad \forall x \in I
\end{aligned}
\end{aligned}
\tag{6.3}
$$

Parameter	Explanation	Constraints
Image gamut I	Given as a discrete point cloud	
Target gamut T	Underlying shape with certain properties	
Working color space		1. Hue preserving 2. Equidistant
Constants $\hat{\lambda}_x$	Determined by Shape(T) which captures the geometry of discrete point cloud	
Focal point c	Point to which points are contracted	1. Interior point of Shape(T) 2. Located on gray axis
Constants c_1 and c_2	Determine the degree to which continuity and detail preservation are taken care of	$0 < c_1 < c_2$

Table 1. Input and parameters of the gamut mapping optimization problem.

solve for typical image and target gamuts. First of all the optimization problem is in a general non-convex form because the constraints are non-convex, see [**BV04**] for more information about convex functions. Second, it has far too many variables. For every point in an image gamut we have to add one variable to the optimization problem. As the number of points contained in a typical image gamut can be very large this makes also the optimization problem very large. Third, the number of

constraints is very large, it is essentially quadratic in the number of variables. Altogether these three factors lead to the practical infeasibility of the problem. In this section we show how one can modify the optimization problem to make it feasible also in practice. The modification has less variables and fewer and linear constraints. It is a convex quadratic program, which can be solved using standard methods. In general a solution to the modified problem is different from a solution to the original problem. But we will argue that a solution to the modified problem still meets our needs.

Number of variables. In order to reduce the number of variables we modify the optimization problem such that only the points in the image gamut boundary are taken into account. The idea is that once we have found a good mapping for the boundary points we can easily extend this mapping to interior points of the image gamut. That is, the optimization problem essentially stays the same only the number of variables is reduced by discarding all points that do not belong to the boundary of the image gamut. The modification needs an operator $\textsc{Boundary}(I)$ that provides us with a finite set of points which are considered to be located on the boundary of the shape that underlies the point set I. We assume that for every pair of points $p_1, p_2 \in \textsc{Boundary}(I)$ the angle between the rays shooting from the focal point $c \in \textsc{Shape}(T)$ to p_1 and p_2, respectively, is not equal to zero. A solution to the modified optimization problem gives for every image gamut boundary point $x \in \textsc{Boundary}(I)$, a parameter λ_x which determines its displacement as $f : x \mapsto c + \lambda_x(x - c)$.

To extend the mapping from the boundary points to the interior points, i.e., the points of I not contained in $\textsc{Boundary}(I)$ we make the assumption that the shape that underlies the target gamut T is star shaped with respect to the focal point c and proceed as described in the following, see also Figure 6.2. For $x \notin \textsc{Boundary}(I)$ determine the point $y \in \textsc{Boundary}(I)$ that lies closest to the ray originating in the focal point and shooting in the direction of x. Closeness can be measured here in terms of the angle between two rays. Project the point x onto the ray from c through y. Let x' be the projection. Now compute a displacement of x' on the ray from c through y. If x' is not contained in the segment $\overline{cy}$, then the displacement of x' is chosen as $f : x' \mapsto f(y) = c + \lambda_y(y - c)$,

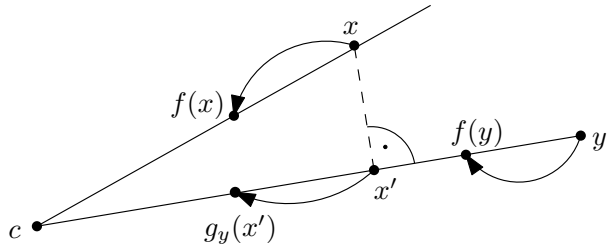

FIGURE 6.2. Extending the mapping from boundary points to interior points.

where λ_y is determined by a solution to the modified optimization problem. In that case we set $\lambda_x = \lambda_y \|y-c\|/\|x'-c\|$. Otherwise the displacement of x' is computed via some compression function $g_y : \overline{cy} \to \overline{cf(y)}$, which maps the segment $\overline{cy}$ onto the segment $\overline{cf(y)}$. The compression function g_y is another parameter of our optimization problem, examples include linear and non-linear compression. The displacement of x' is now given as $x' \mapsto g_y(x')$. Define λ_x as $\|c-g_y(x')\|/\|c-x'\|$, i.e., as the quotient of the lengths of the segments $\overline{cg_y(x')}$ and $\overline{cx'}$. The displacement of x is finally given as

$$f : x \mapsto c + \lambda_x(x-c)$$

Number of constraints. Reducing the number of variables also decreases the number of constraints. To reduce this number even further we only retain the detail and continuity constraints for those points $x, y \in \text{BOUNDARY}(I)$ that are close to each other. Again we measure closeness by the angle between the rays shooting from the focal point c to x and y, respectively. More specifically we retain only constraints between those points $x, y \in \text{BOUNDARY}(I)$, where the angle between the corresponding rays is smaller than some value α, where α becomes another parameter of the optimization problem. Note that the number of constraints increases with increasing value of α.

Linearization. Finally we linearize the continuity and detail constraints. Let $x, y \in \text{BOUNDARY}(I)$ be two points that are close to each other in terms of the angle between the corresponding rays. Without loss of generality we assume that y is closer to c in Euclidean distance than x, i.e., $\|x-c\| \geq \|y-c\|$. Since x and y are close to each other we measure

distances along the rays shooting from c to x and y respectively and approximate the expression

$$\|x-y\| \quad \text{by} \quad \|x-c\|-\|y-c\|$$

and the expression

$$\|f(x)-f(y)\| \quad \text{by} \quad \big|\|f(x)-c\|-\|f(y)-c\|\big|.$$

The modified continuity and detail constraints are now written as follows

$$\begin{aligned} c_1(\|x-c\|-\|y-c\|) &\leq \Big|\|f(x)-c\|-\|f(y)-c\|\Big| \\ &\leq c_2(\|x-c\|-\|y-c\|), \end{aligned}$$

which we further modify as follows

$$\begin{aligned} c_1(\|x-c\|-\|y-c\|) &\leq \|f(x)-c\|-\|f(y)-c\| \\ &\leq c_2(\|x-c\|-\|y-c\|). \end{aligned}$$

Note that the last modification implicitly introduces the following new monotonicity constraint

$$\|x-c\| \geq \|y-c\| \quad \Rightarrow \quad \|f(x)-c\| \geq \|f(y)-c\|,$$

i.e., as mentioned earlier we only introduce the modified constraints if $\|x-c\| \geq \|y-c\|$. Using $\|f(x)-c\|-\|f(y)-c\| = \lambda_x\|(x-c)\|-\lambda_y\|(y-c)\|$ the modified continuity and detail constraints become

$$\begin{aligned} c_1(\|x-c\|-\|y-c\|) &\leq \lambda_x\|(x-c)\|-\lambda_y\|(y-c)\| \\ &\leq c_2(\|x-c\|-\|y-c\|) \end{aligned}$$

which can be rewritten as

$$(c_1-\lambda_x)\|x-c\|-(c_1-\lambda_y)\|y-c\|) \leq 0$$

and

$$(c_2-\lambda_y)\|y-c\|-(c_2-\lambda_x)\|x-c\|) \leq 0.$$

Note that both constraints are now linear functions in λ_x as well as in λ_y.

Combining all the modifications leads to the Optimization Problem (7.4).

Using the variable substitution $\mu_x = 1-\lambda_x$ one can see that this problem is a convex quadratic program, which can be solved using standard methods. The modifications introduce three new parameters into our optimization problem: the operator $\textsc{Boundary}(I)$, which determines the boundary points of the discrete point cloud I, the angle α which determines the size of the neighborhood of a boundary point and the

$\min_\lambda \sum_{x \in \text{BOUNDARY}(I)} (1 - \lambda_x)^2 \|x - c\|^2$

subject to

$$
\begin{aligned}
(c_1 - \lambda_x)\|x - c\| - (c_1 - \lambda_y)\|y - c\| \leq 0 \quad & \forall x, y \in \text{BOUNDARY}(I), \\
& \|x - c\| \geq \|y - c\|, \\
& \angle(c - x, c - y) < \alpha \\
& \text{and } x \neq y \\
(c_2 - \lambda_y)\|y - c\| - (c_2 - \lambda_x)\|x - c\| \leq 0 \quad & \forall x, y \in \text{BOUNDARY}(I), \\
& \|x - c\| \geq \|y - c\|, \\
& \angle(c - x, c - y) < \alpha \\
& \text{and } x \neq y \\
\lambda_x - \hat{\lambda}_x \leq 0 \quad & \forall x \in \text{BOUNDARY}(I) \\
-\lambda_x \leq 0 \quad & \forall x \in \text{BOUNDARY}(I)
\end{aligned}
\tag{6.4}
$$

compression function g, which we use to extend the displacement mapping from boundary to interior points. We also have to assume now that the shape underlying the target gamut is star shaped with respect to the focal point. In Table 2 we summarize the input and the parameters of the modified optimization problem.

6.3.3. Parameterizations. The optimization approach towards image-dependent gamut mapping as described above is general in the sense that it contains several unspecified parameters. In the following we will describe choices for these parameters that we used in our experiments. We tried several options for determining the image gamut boundary points, i.e., for the operator BOUNDARY(I), and the compression function g.

6.3.3.1. BOUNDARY*(I)*. The problem to determine the boundary points of an image gamut can be solved by associating a shape with the finite point set $I \subset \mathbb{R}^3$ and by defining the boundary points of I as a appropriate, finite set of points that lie on the boundary of the shape. Frequently encountered objectives for the associated shape are: it should capture the "geometry" of the point set as well as possible and it should be efficiently computable. Here we describe two different shapes that can be

PARAMETER	CONSTRAINTS
Image gamut I	
Target gamut T	Underlying shape star shaped with respect to the focal point
Working Color Space	1. Hue preserving 2. Perceptually uniform
Constants $\hat{\lambda}_x$	Determined by SHAPE(T), which should capture the geometry of the discrete point cloud T well
Operator BOUNDARY(I)	1. Should capture the boundary shape of the point cloud I well 2. $\angle(p_1 - c, p_2 - c) \neq 0$ for all $p_1, p_2 \in$ BOUNDARY(I) with $p_1 \neq p_2$
Compression function g	
Focal point	1. Interior point of SHAPE(T) 2. Located on gray axis
Constants c_1 and c_2	$0 < c_1 < c_2$
Angle α	$0 < \alpha$

TABLE 2. Input and parameters of the feasible gamut mapping optimization problem.

associated with a point cloud. Both shapes have continuous boundaries. In order to have a finite representation for BOUNDARY(I) we choose a finite sample from these continuous boundaries.

Convex Hull. The convex hull of a finite point set $I \subset \mathbb{R}^3$ is the set of all convex combinations of elements in I, i.e., the set

$$conv(I) := \left\{ \sum_{p \in I} \alpha_p p \,\middle|\, \alpha_p \geq 0 \text{ for all } p \in I \text{ and } \sum_{p \in I} \alpha_p = 1 \right\}.$$

Since the line segment defined by any two points in $conv(I)$ is contained in $conv(I)$ the set $conv(I)$ is star-shaped with respect to any of its points, i.e., any point in $conv(I)$ can be connected to any other point in $conv(I)$ by a line segment that itself is a subset of $conv(I)$. The convex hull of n points in $\mathbb{R}^3$ and its boundary can be efficiently computed in time

$O(n \log n)$, see for example [**Ski98**]. The drawback of working with the convex hull is that in general it does not fit the point cloud (image gamut) tightly. To mitigate the latter problem alpha shapes [**CL99, EM94**] have been used in the context of gamut mapping. Since we did not use alpha shapes we skip the exact definition here—roughly speaking alpha shapes allow some non-convexity controlled by a real parameter.

Star Shape. Besides the convex hull we use another shape that we simply call star shape.

Let again I be a finite set of points in $\mathbb{R}^3$. Additionally we assume now that we are given a center point $c \in \mathbb{R}^3$. The star of I with respect to c is the set of all line segments connecting I to c, i.e.,

$$star_c(I) = \{\alpha p + (1 - \alpha)c \,|\, p \in I, 0 \leq \alpha \leq 1\}.$$

By definition the $star_c(I)$ is star shaped with respect to c, in fact it is the smallest subset of $\mathbb{R}^3$ that contains I and is star-shaped with respect to c. The star of I can be seen as a special case of gamut description with the segment maxima method for infinitesimal small angles [**Mor98**].

In general almost all points of I will be in the boundary of $star_c(I)$. Since our goal was to work only with a small set of boundary points we modify the definition of a star in order to get a shape that better suits our purposes. Therefore we replace the line segments $\overline{cp}, p \in I$ of $star_c(I)$ by cylinders symmetrical to the line segments whose two ends we replace by circular cones with focal points c and p, respectively, see Figure 6.3. This construction effectively (depending on the width of the cylinders and the size of the cones) reduces the the number of boundary points in $star_c(I)$. We call the modified star of I the star shape of I.

FIGURE 6.3. Modified line segments in the definition of the star shape.

6.3.3.2. *Compression Functions.* We solve the optimization problem only for image gamut boundary points. To compute the mapping also for interior points we use a compression function g described in Section 6.3.2. For every boundary point $y \in \text{BOUNDARY}(I)$ the compression function

g_y maps the line segment connecting c and y to the line segment connecting c and $f(y)$, i.e. $g_y : \overline{cy} \to \overline{cf(y)}$. Remember that we get λ_y and thus the point $f(y) = c + \lambda_y(y - c)$ by solving the optimization problem. We use two different types of compression functions, namely linear and non-linear compression.

Linear compression. For a boundary point $y \in \text{BOUNDARY}(I)$ the linear compression function is given as

$$g_y(x) = c + \lambda_y(x - c).$$

In other words: $\lambda_x = \lambda_y$ for all points x on the line segment connecting c and y. But note that only λ_y is obtained from the solution to the optimization problem.

Non-linear compression. For a boundary point $y \in \text{BOUNDARY}(I)$ we use the non-linear compression function

$$g_y(x) = c + \left(\frac{\omega}{q} \left(\lambda_y \tanh \left(\lambda_y^{-1} \tanh^{-1}(q) \right) \right) + (1 - \omega)\lambda_y \right) (x - c),$$

where $\omega \in [0, 1]$ is a parameter that controls the non-linearity and $q = \|x - c\| / \|y - c\|$. Note that for $\omega = 0$, we get the linear compression function as a special case. For $\omega > 0$ the factor λ_x for a point x on the line segment $\overline{cy}$ depends on the distance from x to c and therefore differs with varying x. Figure 6.4 shows how the distance of a mapped point from the focal point, i.e., $\|f(x) - c\|$, depends on the distance of the original point to the focal point, i.e. $\|x - c\|$, for different parameters ω and for a clipping function. The clipping function leaves all points that are closer to c than $f(y)$ untouched and clips all points that a further away from c than $f(y)$ onto the point $f(y)$. For $\omega > 0$ the non-linear compression function compresses colors near the focal point less than colors that are further away.

6.3.3.3. *Combinations.* The general optimization approach permits different parameter settings. We did extensive studies for four of these settings where we altered the operator BOUNDARY(I) and the compression function g. Note that we do not test all possible parameter settings but make a pre-selection of reasonable parameter settings. Specifically these are the following:

(1) **LOptConv.** Here we use the convex hull to determine BOUNDARY(I) and linear compression to map the interior points.

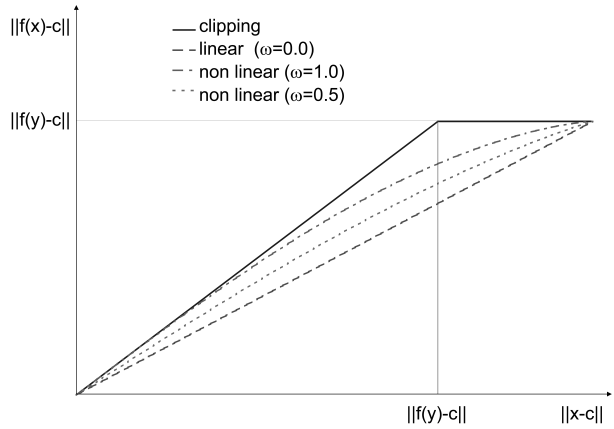

FIGURE 6.4. Comparison of the non-linear compression function for different parameters ω and the clipping function.

(2) **NOptConv.** Non-linear compression is used instead of linear compression. For the determination of BOUNDARY(I) again the convex hull is used.
(3) **LOptStar.** Here we use the star shape to determine BOUNDARY(I) and linear compression to map the interior points.
(4) **NOptStar.** Non-linear compression is used instead of linear compression. For the determination of BOUNDARY(I) again the star shape is used.

For the remaining parameters we fix the values as follows: our working color space is CIELAB, although CIELAB has a poor hue linearity particularly in the blue region, see [**BFCE98, FE98**]. We used CIELAB only for the reason that we had software ready for processing in CIELAB. It is straight forward to solve the optimization problem also for another color space. For a more sophisticated implementation the use of a color space that preserves hue better, like for example the hue-corrected CIELAB [**BFCE98**] may be more appropriate. In order to determine the values for the constants $\hat{\lambda}_x$ we use the description of the target gamut in the ICC profiles. For a point $x \in \text{BOUNDARY}(I)$ we determine $\hat{\lambda}_x$ by intersecting the target gamut boundary with a ray originating in c and shooting in the direction of x. We fix the remaining

parameters to values that turned out to perform well: $c_1 = 0.1$, $c_2 = 1.0$, $c = (57, 0, 0)$, $\alpha = 2°$.

6.3.4. Implementation. We implemented a small test environment for image-dependent gamut mapping using Java 5. As an optimization tool we used the software MOSEK [**mos**]. The test environment has a graphical user interface that makes it easy to experiment with different parameter configurations for the gamut mapping. Additionally it permits a visualization of the shape of image and target gamuts.

In our experimental implementation we quantized for efficiency and and legacy software reasons the image gamut to a grid in CIELAB space. The quantization was done as follows: the L^*-axis of the CIELAB space is discretized to 101 equal distant values and the a^* and b^* axes are both discretized to 256 equal distant values each. The maximal error we make by quantizing the continuous CIELAB values is $\frac{\sqrt{3}}{2}$ which is negligible, as under normal conditions most humans cannot perceive it.

6.3.5. Qualitative results. Here we show the effect of using different source gamut descriptions. Figure 6.5 shows three different source gamut descriptions for the FRUIT image and Figure 6.6 shows the mapping results.

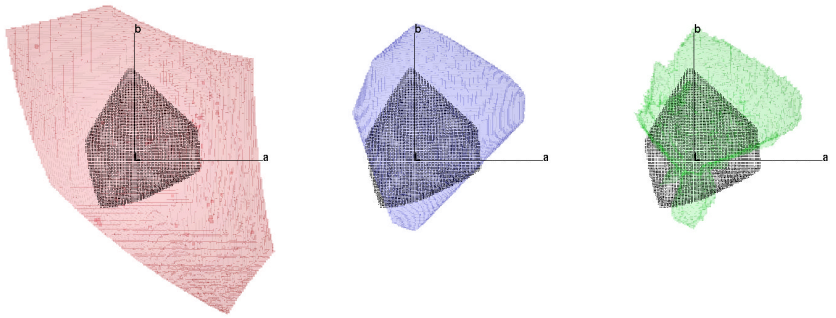

FIGURE 6.5. Full device gamut (left), convex hull of image gamut (middle) and star shape of image gamut (right) of FRUIT image, each compared to target gamut (dark)

FIGURE 6.6. FRUIT image mapped with linear compression using device gamut, (left), convex hull of the image gamut (middle) and star shape of the image gamut (right).

The three different source gamut descriptions are (1) the device gamut, i.e., the space of all colors that potentially could be contained in an image, (2) the convex hull of the color points that are actually contained in the image and (3) the star shape of the color points contained in the image. For the mapping linear compression is used in all three cases. Major visual improvements can be seen in the brown colors for the basket and in the color of the apple.

Figure 6.7 shows the star shape of the image gamut of the FRUIT image before and after the mapping. An appropriate choice of the constants c_1 and c_2 for the detail preservation and continuity constraint, respectively, ensures, that the basic shape of the gamut surface is conserved. Note that comparing the geometry of the original and the mapped gamut surface is a visual sanity check of the extent to which the mapped image could suffer from continuity artifacts or detail loss.

6.4. Psycho-visual Tests

We conducted a user study to assess the performance of our optimization based image-dependent gamut mapping algorithm in terms of perceived quality. For our experiment we have chosen the sRGB to newspaper printing work-flow, since it has an especially small destination gamut and we expect therefore a maximum effect in applying different mapping algorithms. We made a pre-selection of suitable parameter settings in Section 6.3.3.3 and obtained four different incarnations of our parameterized gamut mapping algorithm. In the field study we compared these

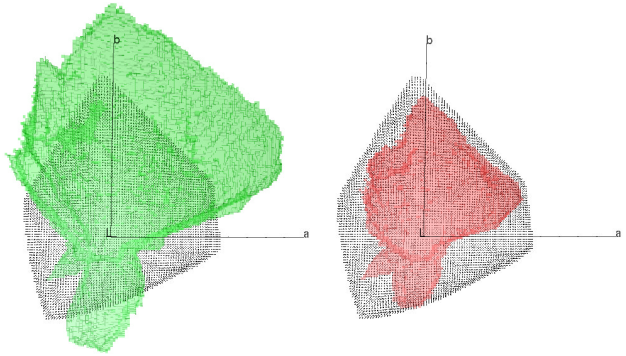

FIGURE 6.7. Mapping of the star shape of the image gamut of the FRUIT image into the target gamut. Note that the shape of the image gamut is partially retained due to the detail preservation and continuity constraints.

four algorithms to standard reference algorithms which are not image-dependent. For the evaluation we were able to directly apply Thurstone's method which we introduced in Chapter 5. This was possible because the number of algorithms to be compared is relatively small. Before we report on the data analysis however, we first give a short description of the used reference algorithms and the experimental setup.

6.4.1. Reference gamut mapping algorithms. As reference algorithms we used the following two algorithms that were proposed as reference algorithms in the CIE-guidelines [**Cen04**]:

(1) **SGCK.** This algorithm is one of the two gamut mapping algorithms recommended by CIE to be used as reference algorithm in psycho-visual tests. The details of this algorithm are described in the CIE-guidelines [**Cen04**]. This algorithm is a combination of GCUSP [**Mor98**] and the sigmoid lightness mapping and cusp knee scaling proposed by Braun and Fairchild [**BF99**].
(2) **Hue preserving minimal ΔE (HPMinDE).** The details of this algorithm are described in the CIE-guidelines [**Cen04**]. It keeps colors inside the target gamut unchanged and maps outside colors to the target gamut border. This is done by minimum distance clipping within planes of constant hue.

We also compared our optimization based image-dependent algorithms to their non-image-dependent counterparts, namely:

(3) **LDev**, linear compression for the device gamut.
(4) **NDev**, non-linear compression for the device gamut.

6.4.2. Experimental set-up. The test procedure for the psycho-visual test was done following the CIE-guidelines [**Cen04**]. For data elicitation we used paired comparisons: the original image and two mappings thereof were shown simultaneously on a monitor screen, the original in the middle of the upper half of the screen and the two mappings below the original side by side. All images had a constant height of 10.5 cm. The observing person had to select the mapped image which he or she judged to be the better representation of the original image. If both mappings were judged to have equal quality, the original image had to be selected.

All paired comparisons have been performed on calibrated LCD screens (EIZO cg220 and EIZO cg210). The background of the screen was set to neutral gray. Behind the screen and in the back of the observer dark gray and black paper backgrounds, respectively, were used. The illuminance of the surrounding (with monitor switched off) was well below the recommended upper limit of 35 lx. On the screen no reflections of the illumination were visible to the observer.

Two types of test images were used: first, 8 traditional test images, four of them ISO test images and the SKI image recommended by the CIE guidelines (see Fig. 6.8) and second, a set of 62 images from a newspaper agency [**KEY03**]. In the analysis we consider the frequency that an algorithm was preferred over the other, independent of the image that was mapped by the algorithms, i.e., all images contribute to the same study. Note that for the volume visualization study in the following Chapter we take a different approach: there each one of two different visualization objects yields its own conjoint study.

The participants of our study were students or staff of our institutes. Every participant had passed the "Ishihara test" [**Ish62**]. The participants were instructed by means of a pre-test where twelve image pairs had to be compared. In total 42 persons, made a total of 3920 paired comparisons. For each of the 28 pairs of algorithms all 70 test images were shown twice during the test. This results in 140 comparisons of every algorithm pair. A single observer made 80-160 comparisons on 10-20 different images (each image was presented for 8 different algorithm

FIGURE 6.8. ISO test images plus the SKI image as recommended by CIE. The captions of the images are used as reference in the paper.

pairs). For each respondent the pairs were presented in random order. Moreover the positions of the images were determined randomly in order to eliminate effects of preference for the left or the right position.

6.4.3. Data Analysis. To analyze the elicited data we used Thurstone's method which we introduced in Chapter 5. Altogether there are eight gamut mapping algorithms that we compared to each other. Table 3 shows the relative frequencies $F_{i \succ j}$ that algorithm i was preferred over algorithm j. If algorithm i and algorithm j were judged to be equivalent we counted that as half a comparison towards i and half a comparison towards j.

Applying Thurstone's method yields a scale value for every algorithm. These are shown in Figure 6.9. The interval on top of each bar shows the computed theoretical error.

6.5. Discussion

The performance of **LDev** (linear compression) and **HPMinDE** (clipping) is poorer than that of the more sophisticated algorithms **SGCK**

	HPMinDE	SGCK	LDev	LOptConv	LOptStar	NDev	NOptConv	NOptStar
HPMinDE	-	0.53	0.69	0.56	0.64	0.62	0.82	0.85
SGCK	0.47	-	0.30	0.45	0.62	0.66	0.68	0.79
LDev	0.31	0.70	-	0.72	0.89	0.84	0.78	0.88
LOptConv	0.44	0.55	0.28	-	0.76	0.66	0.78	0.79
LOptStar	0.36	0.38	0.11	0.24	-	0.37	0.72	0.70
NDev	0.38	0.34	0.16	0.34	0.63	-	0.65	0.66
NOptConv	0.18	0.32	0.22	0.22	0.28	0.35	-	0.58
NOptStar	0.15	0.21	0.12	0.21	0.30	0.34	0.42	-

TABLE 3. Relative frequency matrix. The numbers have to be read as the relative number of observations where the row algorithm was preferred over the column algorithm.

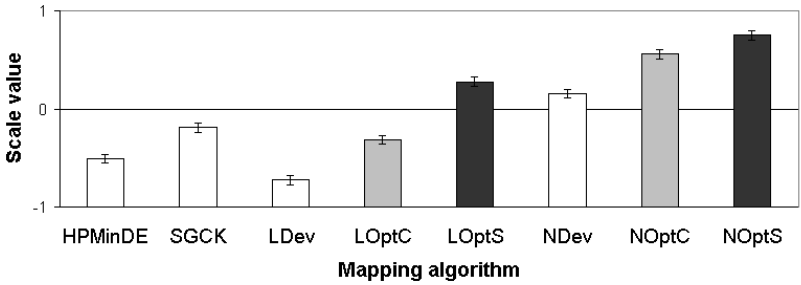

FIGURE 6.9. Scale values for the algorithms: algorithms using device gamuts are in light color, algorithms using the convex hull of the image gamut are in gray color and algorithms using the star shape of the image gamut are in dark color.

and **NDev**. The reason why clipping is rated so poorly in our application is that we use a particularly small target gamut, the newspaper

gamut. This leads to a very high loss of detail when clipping colors onto the target gamut boundary. The most interesting result is the gain in scale value of any algorithm if image gamuts are used instead of device gamuts. Both linear compression and non-linear compression show a significant gain if the device gamut is replaced by the convex hull of the image gamut. A further gain is reached when the star shape of the image gamut is used. This is more pronounced for linear compression than for non-linear compression, which is not surprising, as non-linear compression already has a high preference level and its potential for improvement is smaller than that of linear compression. However, the use of image gamuts together with a not so sophisticated algorithm (like linear compression) can not compensate for using a more sophisticated algorithm (like non-linear compression) together with device gamuts. **NDev** performs better or equivalently to linear compression with image gamuts.

Considering Figure 6.5 it is surprising that using the star shape instead of the convex hull yields such a high gain in scale value for both, linear and non-linear compression, as the volumes of the star shape and the convex hull on first sight appear to be of similar size. Note however that the 2-dimensional projection of a 3-dimensional shape may be misleading: the volume of the star shape is actually much smaller than the volume of the convex hull (it has indentations which are not visible in Figure 6.5). This is in line with the high gain in scale value when using the star shape instead of the convex hull.

In the following we will elaborate on how one could improve the quality of the optimization algorithms even further. The choice of only one focal point to which we map colors simplifies our analysis, however it is not necessary. Our optimization approach permits to choose different focal points for different colors, for example in dependency of the hue, which gives us more degrees of freedom in the optimization problem. Then we only have to ensure that any ray along which we map a color intersects the target gamut boundary exactly once, i.e., we have to choose the boundary description appropriately. Furthermore in our implementation we used the CIELAB color space. We would expect an improvement of our results when using a color space that has a better hue preservation, like for example the hue corrected CIELAB.

Our optimization approach also can be modified in order to obtain a local algorithm that preserves spatial variations, see also [**KSES05, BdEW01**] for work in that direction. However this is not completely straight forward. The biggest challenge is that the obtained optimization

problems usually get very large and thus are not easily turned feasible in practice.

6.6. Conclusion

We have applied Thurstone's method to explore the potential of image-dependent gamut mapping. To this end we have developed a new parameterized image-dependent gamut mapping algorithm based on optimization. We did not measure the perceived quality of every possible parameter setting, but made a pre-selection of four different incarnations of the algorithm that seemed reasonable to us. We compared these to standard reference algorithms. The advantage of this procedure is that we had to compare only a small number of algorithms in the user study instead of comparing all possible parameter settings. Therefore substantially less respondents were necessary.

Our starting point was the time-quality tradeoff one faces in gamut mapping: in comparison to image-independent methods image-dependent algorithms have a worse running time, however the perceived quality of the mapping should improve. The user study permits to quantify this tradeoff. It shows that with the suggested algorithm we can improve the perceived quality significantly over standard reference algorithms.

CHAPTER 7

Volume Visualization

In the previous chapter we developed our own parameterized gamut mapping algorithm and investigated the importance of image-dependence for the perceived quality of a gamut mapping algorithm. Here we consider an existing volume visualization algorithm. We measure the perceived quality induced by different parameter settings using the extension of Thurstone's method introduced in Chapter 5.

7.1. Introduction

Volume visualization is a fast growing field in computer graphics and data visualization. It is concerned with creating a two dimensional image of three-dimensional volume data, which most commonly is acquired by a 3D-scanner. Every volume element is directly projected into the viewplane of the observer. To this end a transfer function is designed which assigns to every volume element a color and a opacity value. Then, in the ray casting step, rays are shot from the position of the observer into the data and are traversed with a certain step size to sample data from the ray. From the sampled data finally the projection into the image plane is computed. The main purpose of volume visualization is to produce images that allow users to gain more insight into the data. Therefore the quality of visualization algorithms should be evaluated on how well they serve this purpose. Such an evaluation needs the involvement of humans.

In this chapter we apply the conjoint analysis technique from Chapter 5 to derive a scale for the different parameter settings of an existing parameterized volume visualization algorithm. We report on a large user study that we conducted in order to determine parameter settings that fit two important visualization purposes: visual aesthetics and conveyance of detail. By doing so we test our conjoint analysis method in practice and we also gain further insights for volume visualization. For example it is

possible to determine the relative importance of the algorithm's parameters. It is also possible to decide if improving on one parameter's level significantly increases the perceived quality of the algorithm, i.e., it could turn out that decreasing the algorithms step size below some level does not influence the quality of resulting renderings anymore, thus allowing to optimize the algorithm's speed without sacrificing on the perceived quality. Finally, we can also study how age, gender or color deficiencies affect users' preferences.

In the following we first discuss the setup of the user study in which we elicited the preference data. Then we introduce the parameterized visualization algorithm that we consider. Finally we present and discuss the results of our evaluation.

7.2. Data Collection

As elicitation procedure we chose choice based conjoint analysis, where each choice task was a paired comparison between two renderings, i.e., between two parameter settings. Note that the cognitive burden increases when the number of items from which to choose increases. Higher cognitive burden usually results in poorer data quality. We decided to use choice tasks with the least cognitive burden, namely paired comparisons.

For our study we used two data sets. The first data set FOOT is meant to cover the medical application domain, whereas the second data set ENGINE covers the engineering application area. See Figure 7.5 and Figure 7.7 for various images rendered for these two data sets. For the FOOT data set we had 2250 different parameter settings resulting in 2250 different renderings (images) and for the ENGINE data set we had 2700 different parameter settings. Perceived quality itself can be measured along different directions. We made this more explicit by asking two different questions: *Which image do you like best?* and *Which image shows more detail?* We will later refer to the first question as AESTHETICS and the second as DETAIL. Note that the second question is more specific than the first, which is fairly general.

Each combination of data set and a question is considered as a different conjoint study, i.e., we conducted the four different conjoint studies [ENGINE, AESTHETICS], [ENGINE, DETAIL], [FOOT, AESTHETICS] and [FOOT, DETAIL].

We elicited data for our conjoint studies from visitors at an exhibition that took place to celebrate the 25th anniversary of the computer science department at ETH Zürich. Our survey took place in a room at the exhibition that was darkened using light-impermeable black curtains. The room had six work places each having a computer with mouse and LCD screen with resolution 1280x1024 pixels. During the survey the light in the room was switched off. 786 visitors of the exhibition participated in our study. From them we collected the following data: age, gender and color deficient (yes or no). To test for color blindness we used the Ishihara test [**Ish62**].

Every respondent took part in exactly two of our conjoint studies, i.e., a respondent always had to answer the same randomly chosen question on both data sets. That is, each respondent participated in one study for each data set FOOT and ENGINE, respectively. We did not conduct the two studies one after the other, but interleaved them: alternately a respondent was shown ENGINE image pairs and FOOT image pairs to choose from, altogether 20 pairs for each data set. Our motivation for this procedure was that in conjoint analysis surveys respondents often get bored after only a couple of questions. So we tried to make the survey a little more exciting by alternating the data sets. The image pairs for the comparisons were determined as follows: the first image was drawn uniformly at random from the set of all images. The second image was then drawn uniformly at random from the set of images having for each parameter a different value than the first image. The images were presented side by side on the screen with a black stripe separating them. The background of the screen was set to black. All images had a resolution of 512x512 pixels. Respondents chose an image by clicking on it with the mouse. After a click the next image pair was shown. Typically the respondents needed three to five minutes to complete the survey.

7.3. Algorithm

We consider a parameterized volume visualization algorithm where the visualization of an object can be described in terms of the parameters COLORMAP, RENDERING, VIEWPOINT, RESOLUTION, STEP SIZE and BACKGROUND. The parameter COLORMAP has three values which correspond to different color maps that are applied for transfer function design. For all transfer functions, the opacity value has been set to always reveal most of the object's structures, in order to suppress 'occlusion' to

act as an independent variable. The parameter RENDERING describes the applied rendering mode and has five values: DVR (Direct Volume Rendering), DVRNS (Direct Volume Rendering with No Shading, just compositing), DVRGM (Direct Volume Rendering with Gradient Modulation to highlight surfaces), XRAY (Colored X-Ray) and MIP (Colored Maximum Intensity Projection). The parameter VIEWPOINT has six values for ENGINE and five values for FOOT. It describes the viewpoint under which the observer sees the object. Different viewpoints are chosen in such a way that most structures are always kept visible, again to prevent 'occlusion' to play a significant role in the study. The parameter RESOLUTION describes the screen resolution used for rendering and has two values: rendering at the resolution of the dataset and twice that. STEP SIZE is the ray traversal increment (measured in voxel size), which has three values, 0.2, 0.5 and 1.0. Finally the parameter BACKGROUND describes the color of the background and has five values: BLACK, WHITE, DARK GREEN, DARK BLUE and YELLOW. Combining these parameters results in 2700 ENGINE images and in 2250 FOOT images. The image size is always 512×512 (the image rendered at reduced resolution, that is, at volume resolution, was scaled up with bilinear filtering). The ENGINE data size is $256 \times 256 \times 256$, and the FOOT data size is $154 \times 263 \times 222$.

7.4. Results

In this section we report on how we applied our data analysis method that we described in Chapter 5 to obtain meaningful scale values for our four conjoint studies. All subsequent results refer to respondents that are more than 10 years old[1] and have passed the Ishihara test for color blindness. Among all respondents fulfilling these two criteria there were 317 respondents participating in the two studies with test question DETAIL and 366 respondents participating in the other two studies with test question AESTHETICS.

In a first step we computed scale values using the method described in Section 5.3. These scale values need not to be meaningful since model assumptions that underlie these computations might not be met in our studies. Hence we discuss in the following how to obtain meaningful scale values from the initially computed ones.

[1]We found no significant differences between respondents younger and older than 17, respectively.

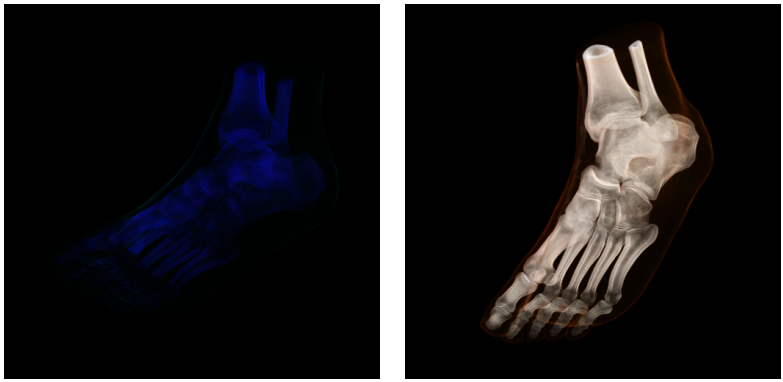

FIGURE 7.1. The images with the highest scale values for the studies [FOOT,AESTHETICS] (left) and [FOOT,DETAILS] (right) before taking care of parameter dependencies.

7.4.1. Testing additivity. As pointed out earlier if the linear model assumption holds, then the scale values for different parameters are comparable and we can determine the scale value of an image (rendering for a specific choice of parameter values) by adding up the scale values for the parameter values used to render the image. However, the top ranked image that we get this way for the FOOT data set and AESTHETICS question does not look like a reasonable first choice, see Figure 7.1. The reason is not surprising: the parameters COLOR and BACKGROUND are not independent of each other for this study.

We tested all pairs of parameters on interdependencies for all four studies using the additivity test described in Section 5.4.1. Table 1 summarizes the result of this test for all combinations of parameters.

Based on the additivity test results we decided to to combine the parameters RENDERING and STEPSIZE for both data sets [FOOT,ENGINE] and both questions [AESTHETICS,DETAILS]. For the FOOT data set and both questions we also combine the parameters COLORMAP and BACKGROUND[2]. That is, we compute new scale values for the combined parameters and use them to replace the scale values for the original parameters.

[2]Note that though the top ranked image for the study [FOOT, DETAIL] looks reasonable, see Figure 7.1 it turns out that we have to combine the two color parameters also for this study.

FOOT		C	R1	V	R2	S	B
Colormap	C	*	4				1
Rendering	R1	2	*	5		3	2
Viewpoint	V			*			
Resolution	R2				*		
StepSize	S		4			*	6
Background	B	1	3				*
ENGINE		C	R1	V	R2	S	B
Colormap	C	*					
Rendering	R1		*			1	
Viewpoint	V			*			
Resolution	R2				*		
StepSize	S		1			*	
Background	B		2				*

TABLE 1. Test of additivity assumption for pairs of parameters with significance level $\alpha = 0.01$. The numbers denote the rank order of relevance, i.e., smaller values indicate more relevant dependencies. The values below the diagonal are for the AESTHETICS question and the values above the diagonal are for the DETAILS question.

This already gives our final scale values that we summarize in Table 4 to Table 7. Figure 7.4 to Figure 7.7 show the top twelve renderings for each of the four studies.

7.4.2. Mosteller's test. We also tested our model assumptions on the parameter (attribute) level using Mosteller's test, see Section 5.4.2, for all parameters (including the combined ones). With a few exceptions all parameters passed the test at the $\alpha = 0.01$ significance level. All exceptions concerned the RENDERING parameter. Possible reasons are unequal variances of the distributions of the scale values for different levels, inappropriateness of a one-dimensional scale or an underestimation of the error.

To further investigate the last point, underestimation of the error, we compared the computed sample size error with the two experimental

errors described in Section 5.3.2. In Figure 7.2 we show a comparison between the two experimental errors and the computed error. All computed errors are within 15% of the experimental errors, except for the parameter RENDERING which shows an underestimation of up to 40%. This finding also puts the results on the additivity tests involving the RENDERING parameter into a new perspective. Some of the detected interdependencies in Table 1 are not significant anymore if the error estimates for RENDERING are adjusted.

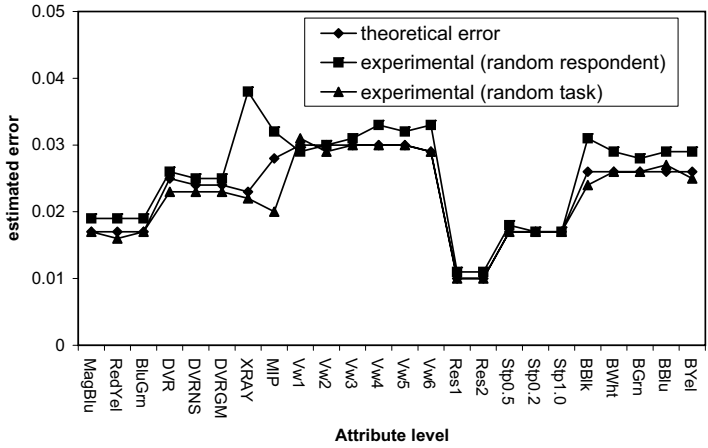

FIGURE 7.2. Three types of error estimation for the scale values for the [ENGINE, AESTHETICS] study.

7.5. Discussion

We now demonstrate how the computed scale values can be used to gain insight into several (related) questions like: What is the relative importance of the parameters? For one parameter, what are the most preferred levels? Do the preferences depend on age, gender or color deficiencies? Do the preferences depend on the data set or the question asked? What preferential dependencies are there among the different parameters?

		Aesthetics	
Foot	1.	Rendering-StepSize	(0.31)
	2.	Colormap-Background	(0.3)
	3.	Viewpoint	(0.14)
	4.	Resolution	(0.05)
Engine	1.	Rendering-StepSize	(0.56)
	2.	Background	(0.19)
	3.	Resolution	(0.12)
	4.	Viewpoint	(0.09)
	5.	Colormap	(0.05)

Table 2. Rank order of the parameters used in the two Aesthetics studies. The rank order is derived from estimated variances (shown in brackets).

7.5.1. Relative importance of parameters. As we pointed out at the end of Section 5.3 the standard deviation σ_{i2} for attribute A_i can be interpreted as the relative importance of attribute A_i. In our setting the attributes are the parameters of the visualization algorithm. Using the estimated standard deviation we get the rank ordering of the parameters as shown in Table 2 and Table 3. From these results it is safe to conclude that overall the rendering mode (combined parameter Rendering-StepSize) is the most important parameter. A second important parameter is the color scheme used (or the background), although this finding is not as pronounced. The viewpoint is somewhat important (mostly for the foot), while the resolution is somewhat important for the engine. The other parameters are relatively unimportant, at least at the levels we have measured.

7.5.2. Most preferred levels. The results of Tables 2 and 3 as well as Figure 5 reveal useful information. We observe that the algorithms Xray and MIP are not considered useful by our respondents (but note that these were non-expert viewers – doctors can see a lot more in those renderings). The DVRGM algorithm performs (slightly) better than DVR, which performs better than DVRNS.

There is also a clear preference for achromatic backgrounds. Only blue is also found to be somewhat useful, possibly because blue in human

		Detail	
Foot	1.	Rendering-StepSize	(0.52)
	2.	Colormap-Background	(0.35)
	3.	Viewpoint	(0.12)
	4.	Resolution	(0.08)
Engine	1.	Rendering-StepSize	(0.77)
	2.	Resolution	(0.09)
	3.	Viewpoint	(0.08)
	4.	Background	(0.05)
	5.	Colormap	(0.01)

Table 3. Rank order of the parameters used in our two Detail studies. The rank order is derived from estimated variances (shown in brackets).

perception creates depth and forms a good background for this reason. Highly saturated backgrounds are generally disliked. Interestingly, there are also differences between the two achromatic backgrounds: a black background is considered more aesthetic, but white seems to show detail better. This is particularly true for the engine which is overall a more complex dataset. It is most likely also an object less familiar to the respondents. Therefore they require more detail and a higher resolution is also more important (than for the less complex foot).

For the engine, the colormap applied does not seem to matter as much, but for the detail views, the foot (bone) is strongly preferred to be seen in a color resembling that of bright bone (skin grey). This indicates that for object inspection, viewers like to see objects in colors that are most natural and at the same time bright (when such a color is generally agreed on), but for objects less defined in that respect the color choice is a matter of taste (as is the case for the engine), as long as they are bright and define contrast well. In the aesthetics category viewers still preferred a natural color (for the foot), but the brightness condition was no longer so important (by definition of the task criterion).

An interesting observation can also be made with respect to the viewpoint. A common feature is that viewers prefer to see objects at oblique angles, which generally gives objects a more three-dimensional appearance and also reveals more features (such views are also used for product

advertisements). But the engine was general preferred to be situated as standing on a surface — the 'flying' views where the engine was rotated at an arbitrary angle were rated low. On the other hand, the foot was acceptable at most orientations. We believe that the 'flying' engine was deemed unrealistic, and perhaps even dangerous and therefore unappealing, while a foot is seen at general orientation commonly in real life (just not as a bone).

7.5.3. Dependency on the respondent. We observed that the experimental error, see Section 5.3.2 is larger when dividing respondents into different sets than when dividing choice tasks into different sets. This indicates that although the respondents answered only 20 choice tasks for each data set, we can already detect a dependency on the individual's preferences, i.e., preferences are not homogeneous over the population.

We also analyzed preferential differences between different sub-groups male vs. female and young vs. old, respectively) of our population respondents [3]:

We only found significant differences between male and female respondents for the COLORMAP parameter in the [FOOT,AESTHETICS] study: female respondents most prefer BLUECYAN (scale value: 0.07(3)[4]), which is also liked by the male respondents (0.07(2)) but not as much as SKINGRAY (0.99(2)), which is the least preferred color of the females (-0.04(3)). Magenta is least preferred by the males (-0.12(2)), whereas females (-0.03(3)) prefer it over SKINGRAY.

In general we found no significant differences between the two age classes 17 years or younger (teenagers) and older than 17 years (adults). We only found two exceptions concerning the AESTHETICS question. For adults the preferences within the RENDERING parameter are more pronounced than for teenagers, though the ranking order of the individual levels remained the same. On the other hand teenagers tend to have more pronounced preferences concerning the background color, again with basically the same order on the individual levels.

[3] We collected preference data from 37 persons showing color deficiencies, but the sample size was not sufficient to detect significant differences to the rest of the population.
[4] Numbers in parenthesis show the estimated standard deviation in units of the last shown digit.

These findings have interesting consequences if one wants to personalize visualization systems: it seems hard to do so based on socio-demographic data (as age and gender) only.

7.5.4. Dependency on the data set. Preferences obtained for the FOOT dataset differ significantly from preferences for ENGINE dataset. This difference is most pronounced for the combined parameter RENDERING-STEPSIZE, which is much more important for the ENGINE dataset for both questions.

Note, that the parameters VIEWPOINT and COLOR can not be compared directly for the two datasets, because different colors and viewpoints were used as parameter levels.

7.5.5. Dependency on the question. The observed preferences in the DETAIL studies are significantly different from the preferences in the AESTHETICS studies. The question about detail separates the preferences for different parameter values better. This means that there is more mutual consent in the test population about detail. We believe this is due to the fact that the question about detail is more specific, and less subject to personal taste. The question about details separates the preferences on the ENGINE data set into two distinct preference classes (DVRXX against XRAY/MIP). This separation does not show in the [ENGINE,AESTHETICS] study. We visualize this by plotting the sorted list of scale values for all images, see Figure 7.3.

7.5.6. Parameter interdependence. As discussed earlier, our additivity test shows that the independence assumption is not fulfilled for the parameters COLORMAP and BACKGROUND for the FOOT data set. This finding seems very reasonable since similar object and background color certainly should have a negative impact on the perceived image appearance. Furthermore details are better visible if the contrast between foreground and background color is high.

The additivity test also shows that the parameters RENDERING and STEPSIZE are not independent. The observed interdependency is less intuitive than the one between COLORMAP and BACKGROUND, but can be explained also. The scale values for the combined parameter show that the changes in STEPSIZE do not induce the same magnitude of change for the scale values of the different RENDERING levels. In particular for XRAY and MIP levels the changes in STEPSIZE seem to have no or only marginal influence. This can be due to the fact that MIP and XRAY

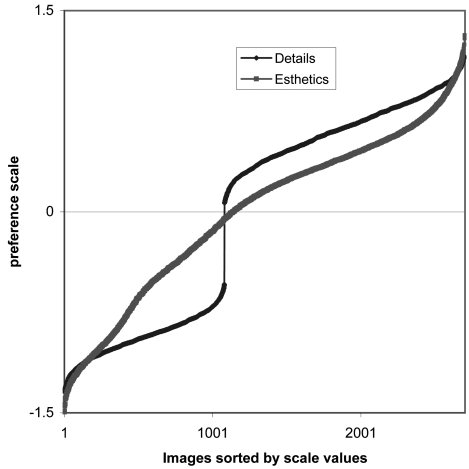

FIGURE 7.3. Sorted list of scale values for the [ENGINE, DETAILS] and [ENGINE, AESTHETICS] studies. For the first study a clear separation is visible.

algorithms lack coherency in structure and are mostly used for quick survey modalities, but not for careful diagnosis. Our study seems to show that the visual system cannot detect errors or even inconsistencies, and thus viewers do not become aware of possible errors.

7.6. Tables and Figures

Engine		Aesthetics	Details
Colormap	MagentaBlue	-0.061(17)	-0.006(18)
	RedYellow	0.065(17)	-0.001(18)
	BlueGreen	-0.004(17)	0.007(18)
Background	Black	0.378(26)	0.049(28)
	White	-0.034(26)	0.078(28)
	Green	-0.162(26)	-0.018(28)
	Blue	-0.063(26)	-0.062(28)
	Yellow	-0.120(26)	-0.046(28)
Rendering	DVR	0.514(25)	0.719(28)
	DVRNS	0.353(24)	0.530(25)
	DVRGM	0.385(24)	0.629(26)
	XRAY	-0.305(23)	-1.005(31)
	MIP	-0.947(28)	-0.872(28)
StepSize	0.5	0.028(17)	0.026(18)
	0.2	0.051(17)	0.066(18)
	1.0	-0.078(17)	-0.093(18)
Viewpoint	side-front	0.132(30)	0.118(32)
	side-back	0.052(30)	0.071(33)
	side-top	0.060(30)	0.027(32)
	side-down	-0.120(30)	-0.041(32)
	front	-0.007(30)	-0.073(32)
	side	-0.117(29)	-0.101(32)
Resolution	high	0.115(10)	0.091(11)
	low	-0.115(10)	-0.091(11)

Table 4. Scale values for all parameter levels of the two Engine conjoint studies. Numbers in parenthesis show the estimated standard deviation in units of the last shown digit.

Engine		Aesthetics	Details
Rendering	DVR, 0.5	0.60(5)	0.81(7)
-StepSize	DVR, 0.2	0.51(5)	0.86(6)
	DVR, 1.0	0.41(5)	0.49(5)
	DVRNS, 0.5	0.18(4)	0.41(5)
	DVRNS, 0.2	0.44(4)	0.64(5)
	DVRNS, 1.0	0.40(4)	0.49(5)
	DVRGM, 0.5	0.63(5)	0.85(6)
	DVRGM, 0.2	0.48(5)	0.71(5)
	DVRGM, 1.0	0.07(4)	0.32(4)
	XRAY, 0.5	-0.29(4)	-0.95(5)
	XRAY, 0.2	-0.32(4)	-1.00(6)
	XRAY, 1.0	-0.29(4)	-1.05(8)
	MIP, 0.5	-0.89(5)	-0.89(5)
	MIP, 0.2	-0.93(5)	-0.86(5)
	MIP, 1.0	-1.03(6)	-0.83(6)
Colormap	MagBlu-BBlk	0.29(5)	-0.05(5)
-Background	MagBlu-BWht	-0.12(5)	0.05(5)
	MagBlu-BGrn	-0.22(5)	0.07(5)
	MagBlu-BBlu	-0.10(5)	-0.11(5)
	MagBlu-BYel	-0.17(5)	0.01(5)
	RedYel-BBlk	0.44(5)	0.20(5)
	RedYel-BWht	-0.08(5)	0.11(5)
	RedYel-BGrn	-0.06(5)	-0.13(5)
	RedYel-BBlu	0.07(5)	-0.01(5)
	RedYel-BYel	-0.04(5)	-0.17(5)
	BluGrn-BBlk	0.40(5)	-0.00(5)
	BluGrn-BWht	0.10(5)	0.06(5)
	BluGrn-BGrn	-0.20(5)	0.01(5)
	BluGrn-BBlu	-0.17(5)	-0.06(5)
	BluGrn-BYel	-0.15(5)	0.02(5)

Table 5. Scale values for all combined parameter levels of the two Engine conjoint studies. Numbers in parenthesis show the estimated standard deviation in units of the last shown digit.

Foot		Aesthetics	Details
Colormap	SkinGrey	0.039(17)	0.146(18)
	BlueCyan	0.070(17)	-0.079(18)
	Magenta	-0.109(17)	-0.067(18)
Background	Black	0.419(27)	0.246(28)
	White	-0.063(26)	0.047(27)
	Green	-0.105(26)	-0.097(28)
	Blue	0.007(26)	-0.103(28)
	Yellow	-0.258(26)	-0.093(28)
Rendering	DVR	0.095(26)	0.361(27)
	DVRNS	-0.001(25)	0.058(26)
	DVRGM	0.484(26)	0.561(27)
	XRAY	-0.308(26)	-0.758(28)
	MIP	-0.270(25)	-0.223(26)
StepSize	0.5	0.035(17)	0.065(18)
	0.2	0.061(17)	0.038(18)
	1.0	-0.096(17)	-0.103(18)
Viewpoint	side-60	0.126(26)	0.174(28)
	top-90	-0.158(26)	-0.118(28)
	top-0	-0.133(26)	-0.151(28)
	side-30	0.208(27)	0.098(28)
	top-135	-0.044(26)	-0.003(28)
Resolution	high	0.045(10)	0.080(11)
	low	-0.045(10)	-0.080(11)

Table 6. Scale values for all parameter levels of the two Foot studies. Numbers in parenthesis show the estimated standard deviation in units of the last shown digit.

Foot		Aesthetics	Details
Rendering	DVR, 0.5	0.00(5)	0.17(5)
-StepSize	DVR, 0.2	0.29(5)	0.64(5)
	DVR, 1.0	0.02(5)	0.26(5)
	DVRNS, 0.5	0.08(5)	0.17(5)
	DVRNS, 0.2	-0.01(5)	-0.03(5)
	DVRNS, 1.0	-0.09(5)	0.02(5)
	DVRGM, 0.5	0.67(5)	1.07(6)
	DVRGM, 0.2	0.60(5)	0.64(5)
	DVRGM, 1.0	0.16(5)	0.04(5)
	XRAY, 0.5	-0.31(5)	-0.80(5)
	XRAY, 0.2	-0.36(5)	-0.77(5)
	XRAY, 1.0	-0.25(5)	-0.73(6)
	MIP, 0.5	-0.24(5)	-0.26(5)
	MIP, 0.2	-0.26(5)	-0.23(5)
	MIP, 1.0	-0.31(5)	-0.19(5)
Colormap	SkinGray-BBlk	0.73(5)	0.74(6)
-Background	SkinGray, Wht	-0.29(5)	-0.30(5)
	SkinGray, Grn	-0.11(5)	0.11(5)
	SkinGray, Blu	0.24(5)	0.47(5)
	SkinGray, Yel	-0.40(5)	-0.24(5)
	BluCya, Blk	0.30(5)	-0.13(5)
	BluCya, Wht	0.26(5)	0.36(5)
	BluCya, Grn	0.02(5)	-0.11(5)
	BluCya, Blu	-0.26(5)	-0.75(6)
	BluCya, Yel	0.04(5)	0.17(5)
	Mag, Blk	0.20(5)	0.16(5)
	Mag, Wht	-0.14(5)	0.08(5)
	Mag, Grn	-0.19(5)	-0.29(5)
	Mag, Blu	0.03(5)	-0.07(5)
	Mag, Yel	-0.42(5)	-0.19(5)

Table 7. Scale values for the combined parameters of the two Foot studies. Numbers in parenthesis show the estimated standard deviation in units of the last shown digit.

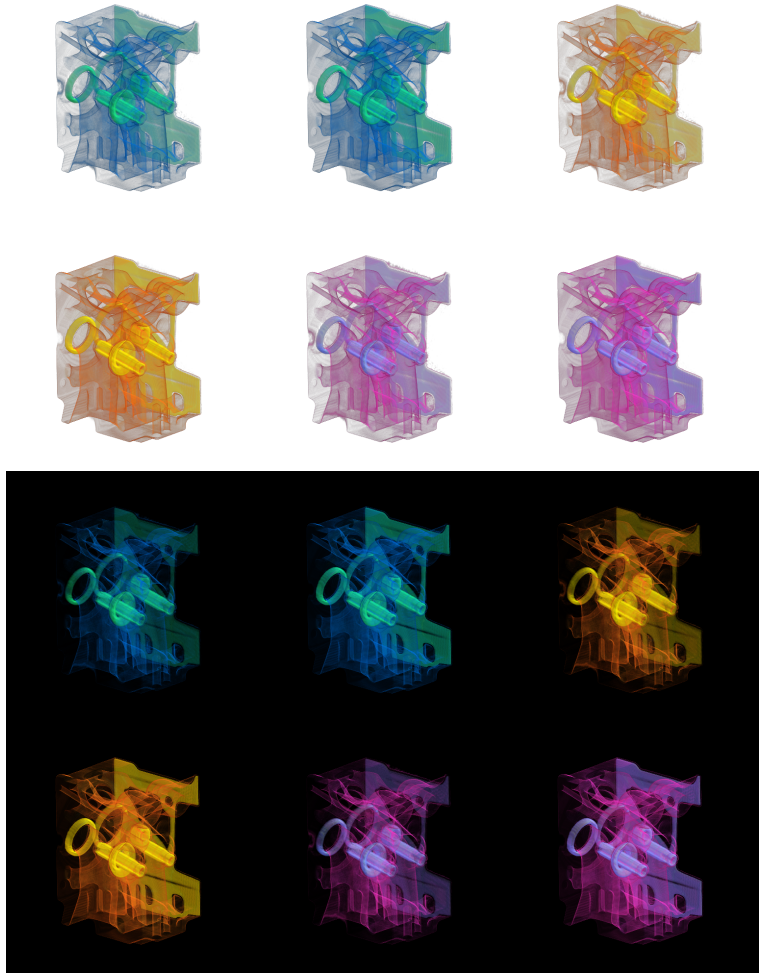

FIGURE 7.4. Best 12 renderings (ranking decreasing from left to right and from top to bottom) for the [ENGINE, DETAIL] study

.

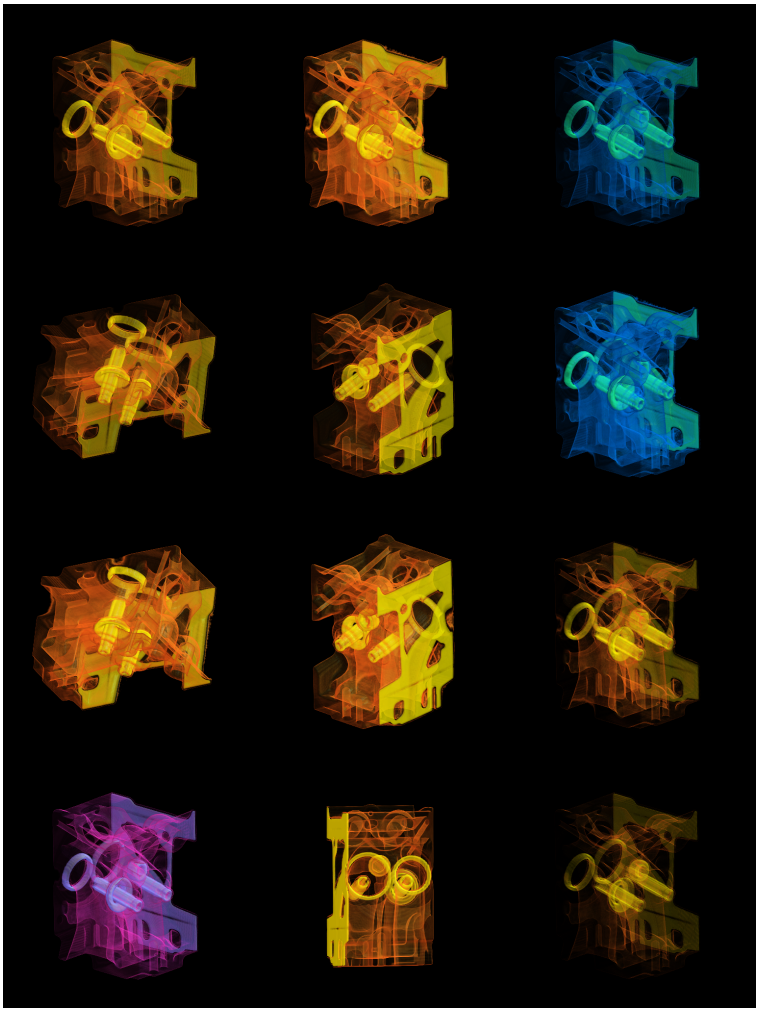

FIGURE 7.5. Best 12 renderings (ranking decreasing from left to right and from top to bottom) for the [ENGINE, AESTHETICS] study

.

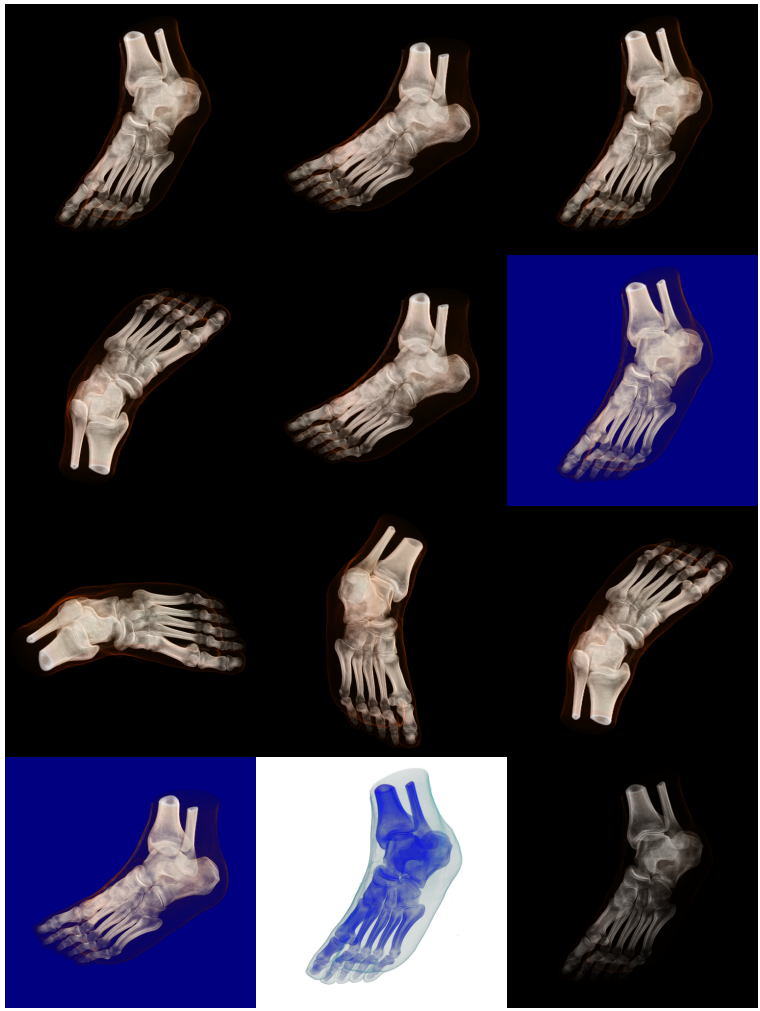

FIGURE 7.6. Best 12 renderings (ranking decreasing from left to right and from top to bottom) for the [FOOT, DETAIL] study

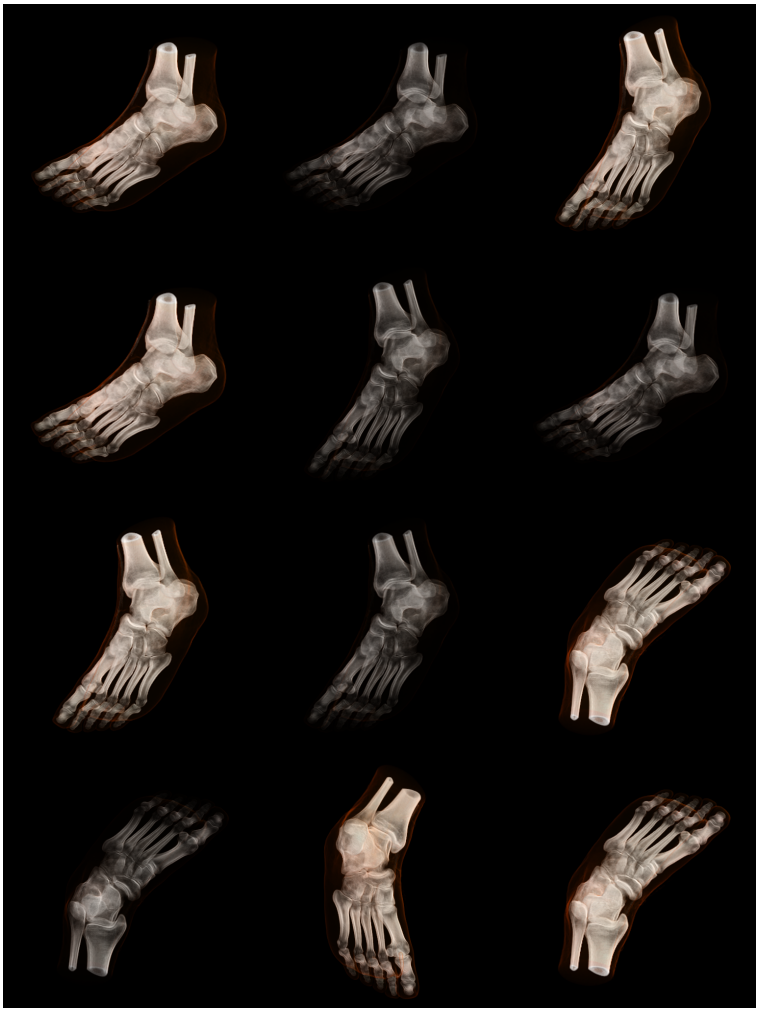

FIGURE 7.7. Best 12 renderings (ranking decreasing from left to right and from top to bottom) for the [FOOT, ESTHETICS] study

.

7.7. Conclusion

We have demonstrated that the conjoint analysis method that we have derived in Chapter 5 is a useful and efficient tool to determine the individual and grouped influence of a large set of parameters on human perception in volume visualization. The user study gave us insights into quality perception by users in volume visualization, but we also learned about how to improve the user study design in general.

7.7.1. Volume Visualization. We verified a few known results in volume visualization, such as the effect of rendering fidelity, but we also teased out some lesser known but important results, such as preferred object orientations, color schemes, and the relationship of step size and rendering mode. Our method assigns scale values to parameter levels, which allows to determine the relative importance of the parameters and their levels. The scale values can be used to rank all the different parameter settings but also provides insights beyond that: we were able to conclude from the computed scale values that preferences depend on the individual, which in itself is not so surprising, but we also found that one cannot predict an individual's preferences from the sociodemographic data available to us (age and gender). Another interesting finding is that our conjoint analysis method can help to resolve tradeoff decisions. For example our data show that for the DVRGM algorithm it is not necessary to go down to a step size of 0.2—step size 0.5 even gives perceptually better results. That is, it is not worth to invest the extra time needed for a smaller step size (time-quality tradeoff). If time is of the essence, then one can also lower the resolution since our data show that the influence of the resolution on the perceived quality is low. In the future our conjoint analysis method could be used to test various algorithms devised for automated viewpoint selection, transfer function design, and others.

7.7.2. Study design. During the evaluation of the volume visualization study it turned out that one could improve our strategy to determine the choice tasks. Remember that the image pairs presented in a choice task were chosen as follows: the first image was drawn uniformly at random from the set of all images. The second image was then drawn uniformly at random from the set of images having for each parameter a different level than the first image. This strategy leads to problems —

caused by 'missing' preference data — in case that at least one of the following two things happen:

(1) There are at least two parameters for which the additivity assumption does not hold, i.e. the two parameters have to be combined.
(2) There is a large difference in the relative importance of different parameters, for example there is one dominating parameter which is much more important than the others.

In our study both things happened and in the following we illustrate which problems are involved:

First, we found two pairs of parameters for which the additivity assumption does not hold and which therefore have to be combined. Because of our strategy to choose the choice tasks there are pairs of levels of the combined parameters that have never been compared by any respondent. To see this, assume we combine parameters A and B and obtain the new parameter $C := A \times B$. Then level $c_1 = (a_1, b_1) \in C$ and level $c_2 = (a_2, b_2) \in C$ have never been compared by any respondent whenever $a_1 = a_2$ or $b_1 = b_2$. For these pairs of levels the relative frequency that c_1 was preferred over c_2 is not defined and therefore our evaluation method cannot be applied directly. In the presented study we got around this problem by estimating the missing relative frequencies using the observed ones. However, we suggest for future studies to avoid this problem from the beginning.

Second, the parameter RENDERING dominates the other parameters, i.e. it has a much higher relative importance than the other parameters. Probably this is the reason why Mosteller's test fails for RENDERING. We suppose that in many choice tasks an image is selected because it has a 'better' level for RENDERING, independent of the levels for the remaining parameters. Two images in a choice task have always a different level for RENDERING because of our strategy to choose the choice tasks. That is, the levels for RENDERING will most of the time determine the decision in a choice task. One might want to investigate the influence of the remaining parameters closer by keeping the level for RENDERING fixed and thus blinding out its influence. However, the data that we elicited does not contain such information as there is no choice task where the level for RENDERING is the same for both images.

To avoid these problems in future studies one could change the query strategy such that a choice task is drawn uniformly at random from all item (image) pairs. Alternatively one could conduct pre-studies to

determine the parameters that have to be combined and to learn about the relative importance of parameters. Then, in the main study one can directly work with the combined parameters. Furthermore, if it turns out for example that there is a dominating parameter, then one might want to conduct an additional study where the level of this parameter is kept fixed.

CHAPTER 8

Conclusion

We started this thesis with developing an image-dependent gamut mapping algorithm. Since the output of a gamut mapping algorithm needs human evaluation we conducted a field study to test the algorithm. We evaluated the field study using Thurstone's law of comparative judgment. Soon we realized that it is true for many algorithms whose output is an image, a video or an audio file, that their quality should be evaluated by humans. However, when these algorithms are parameterized and allow a large number of different parameter settings, Thurstone's method cannot be applied anymore and new methods have to be found. Therefore this thesis is about how to measure the perceived quality of parameterized multimedia algorithms. The observation that a parameterized algorithm is nothing else but a set of items with conjoint structure motivated us to investigate this problem in the broader context of choice based conjoint analysis.

In the first part of the thesis we gained some theoretical insights into conjoint analysis. Triggered by a celebrated volume-based approach we started to study the combinatorics of choice based conjoint analysis. Assuming a linear model we compared the problem of finding a ranking of items that possess a conjoint structure with the problem of finding a ranking of items without any structure and gave a geometric interpretation for both cases. It turned out that insights from the structureless case cannot be carried over to the conjoint structure case. These findings raise doubts about the reliability and applicability of volume-based approaches to conjoint analysis.

Moreover it turned out, that there is no algorithm that can derive an item ranking in the conjoint structure case from only polynomially many choice tasks (polynomial in the number of attributes and the number of levels). Since it is not feasible to ask respondents many questions in a field study we strove for alternative approaches. First we investigated

the question of how many item comparisons are necessary in order to derive an approximate ranking of the items. Second, we introduced new modeling assumptions and used aggregation in order to build a framework for the evaluation of choice based conjoint studies. The framework permits to test all modeling assumptions. This is a big advantage of our framework in comparison to off-the-shelf conjoint analysis software like Sawtooth' software [**Sof**] where modeling assumptions cannot be tested. Our framework can be applied to very large item sets which possess a conjoint structure. In particular the introduced modeling assumptions permit to compute scale values for items that have never been compared in any choice task.

In the second part of the thesis we put our framework to work in practical applications. First, we presented a new image-dependent gamut mapping algorithm. The gamut mapping algorithm is based on an optimization approach and involves the determination of an image gamut's boundary. Let us remark at this point that in the course of the project we also studied in detail different methods for gamut boundary determination. In particular we used flow shapes and kernel methods to compute the boundary of an image gamut. It turned out that these approaches are suited only to a certain degree for boundary computation in practice. Details about these approaches can be found in [**GSSZ05**] and [**GSSZ06**]. We used Thurstone's method to evaluate a field study that we conducted to test the quality of the gamut mapping algorithm. The user study showed that image-dependence is an important parameter and that with our algorithm we are able to improve the perceived quality over standard reference algorithms.

Finally we measured the perceived quality of an existing volume visualization algorithm. The algorithm has quite large number of different parameter settings that we want to compare. This large number renders Thurstone's method not applicable to this task, but fortunately the algorithm in terms of its parameters has a conjoint structure. Therefore we were able to apply our framework that we developed in Chapter 5. We described the design of a large user study that we had conducted to test the influence of the different parameters of the algorithm. Our insights were twofold: On the one hand we gained valuable insights into quality perception by users in volume visualization. We verified a few known results, such as the effect of rendering fidelity, but we also teased out some lesser known but important results, such as preferred object orientations, color schemes, and the relationship of step size and rendering

mode. On the other hand we were able to put our conjoint framework to test and learned about how to improve the user study design in general.

Our two applications — gamut mapping and volume visualization — have in common that the output of the algorithms are images. However, as the title of the thesis suggests our framework for choice based conjoint analysis should be suitable to measure the perceived quality of all kinds of parameterized multimedia algorithms, like for example algorithms whose output is an audio file or a video. It would be interesting for future research to conduct field studies in theses areas.

Bibliography

[BdEW01] R. Bala, R. deQueiroz, R. Eschbach, and W. Wu, *Gamut mapping to preserve spatial luminance variations*, Journal of Imaging Science and Technology **45** (2001), no. 5, 436–443. 6.2.1, 6.5

[BF99] G. J. Braun and M. D. Fairchild, *Image lightness rescaling using sigmoidal contrast enhancement functions*, Journal of Electronic Imaging **8** (1999), no. 4, 380–393. 6.2.1, 6.4.1

[BFCE98] G.J. Braun, M. D. Fairchild, C. F. Carlson, and F. Ebner, *Color gamut mapping in a hue-linearized cielab color space*, Proceedings of the 6th IS&T/SID Color Imaging Conference (1998), 163. 6.3.1, 6.3.3.3

[BFP+72] M. Blum, R. W. Floyd, V. Pratt, R. L. Rivest, and R. E. Tarjan, *Linear time bounds for median computations*, STOC '72: Proceedings of the fourth annual ACM symposium on Theory of computing, ACM Press, 1972, pp. 119–124. 4.2, 4.2

[BV04] Stephen Boyd and Lieven Vandenberghe, *Convex optimization*, Cambridge University Press, 2004. 6.2.2, 6.3.2

[Cen04] Central Bureau of the CIE, Vienna, *Cie publication 156: Guidelines for the evaluation of gamut mapping algorithms*, 2004. 6.2.1, 6.4.1, 6.4.2

[Cha00] B. Chazelle, *The soft heap: An approximate priority queue with optimal error rate*, Journal of the ACM **47** (2000), no. 6, 1012–1027. 4.1, 4.2

[CL99] T. J. Cholewo and S. Love, *Gamut boundary determination using alpha-shapes*, 7th Color Imaging Conference (Scottsdale, AR), vol. 7, IS&T/SID, 1999, pp. 200–204. 6.2.1, 6.3.3.1

[DG77] P. Diaconis and R. L. Graham, *Spearman's footrule as a measure of disarray*, Journal of the Royal Statistical Society **39(2)** (1977), 262–268. 4.1

[Ehr05] Matthias Ehrgott, *Multicriteria optimization*, second ed., Springer, 2005. 6.3.1

[EM94] Herbert Edelsbrunner and Ernst P. Mücke, *Three-dimensional alpha shapes*, ACM Transactions on Graphics **13** (1994), no. 1, 43–72. 6.3.3.1

[Eng00] Peter G. Engeldrum, *Psychometric scaling, a toolkit for imaging systems development*, Imcotek Press, Winchester MA, USA, 2000. 5.2, 5.4.2

[FE98] Mark D. Fairchild Fritz Ebner, *Finding constant hue surfaces in color space*, Color Imaging: Device-Independent Color, Color Hardcopy, and Graphic Arts III (G. B. Beretta and R. Eschbach, eds.), vol. 3300, 1998, pp. 107–117. 6.3.1, 6.3.3.3

[Fre76] M. L. Fredman, *How good is the information theory bound in sorting*, Theoretical Computer Science **1** (1976), 355–361. 3.4.2

[GSSZ05] J. Giesen, E. Schuberth, K. Simon, and P. Zolliker, *Toward image-dependent gamut mapping: fast and accurate gamut boundary determination*, Color Imaging X: Processing, Hardcopy and Applications (San Jose, CA) (Reiner Eschbach and Gabriel Marcu, eds.), vol. 5667, 2005, pp. 201–210. 6.2.1, 8

[GSSZ06] ———, *A kernel approach to gamut boundary determination.*, Proceedings of the 14th European Signal Processing Conference (EUSIPCO) (2006). 8

[GWA90] R. S. Gentile, E. Walowitt, and J. P. Allebach, *A comparison of techniques for color gamut mismatch compensation*, Journal of Imaging Technologies **16** (1990), 176–181. 6.2.1

[Hoa61] C. A. R. Hoare, *Algorithm 65: find*, Commun. ACM **4** (1961), no. 7, 321–322. 4.2

[HYY00] H. K. Hwang, B. Y. Yang, and Y. N. Yeh, *Presorting algorithms: an average-case point of view*, Theoretical Computer Science **242(1-2)** (2000), 29–40. 4.1

[Ish62] S. Ishihara, *Ishihara's tests for colour blindness*, concise ed., Tokyo, Isshinkai, 1962. 6.4.2, 7.2

[KEY03] KEYSTONE, *"jahresrueckblick 2003", collection of 268 press photographs*, 2003. 6.4.2

[KIO99] N. Katoh, M. Ito, and S. Ohno, *Three-dimensional gamut mapping using various color difference formulae and color spaces*, Journal of Electronic Imaging **8** (1999), no. 4, 365–379. 6.2.1

[KK92] J. Kahn and J. H. Kim, *Entropy and sorting*, STOC '92: Proceedings of the twenty-fourth annual ACM symposium on Theory of computing, ACM Press, 1992, pp. 178–187. 4.1

[Knu73] D. E. Knuth, *The art of computer programming*, vol. 3, Addison Wesley, 1973. 3.3.1, 4.1, 4.3

[KR93] R. L. Keeney and H. Raiffa, *Decisions with multiple objectives: Preferences and value trade-offs*, Cambridge University Press, Cambridge, 1993. 2

[KS03] H. Kotera and R. Saito, *Compact description of 3-d image gamut by r-image method*, Journal of Electronic Imaging **12** (2003), no. 4, 660–668. 6.2.1

[KSES05] R. Kimmel, D. Shaked, M. Elad, and I. Sobel, *Space-dependent color gamut mapping: a variational approach*, IEEE Transactions on Image Processing **14** (2005), no. 6, 796–803. 6.2.1, 6.5

[Mal57] C. L. Mallows, *Non-null ranking models*, Biometrica **44** (1957), 114–130. 4.1

[Mat02] J. Matoušek, *Lectures on discrete geometry*, Graduate Texts in Mathematics, vol. 212, Springer-Verlag, New York, 2002. 1.2, 1, 2, 3.3.2

[MF97] E. D. Montag and M. D. Fairchild, *Psychophysical evaluation of gamut mapping techniques using simple rendered images and artificial gamut boundaries*, IEEE Transactions on Image Processing **6** (1997), no. 7, 977–989. 6.2.1

[ML00] J. Morovic and M. R. Luo, *Calculating medium and image gamut boundaries for gamut mapping*, Color Research and Applications **25** (2000), no. 6, 394–401. 6.2.1

[ML01] ———, *The fundamentals of gamut mapping: A survey*, Journal of Imaging Science and Technology **45** (2001), no. 3, 283–290. 6.3.1

[Mor98] J. Morovic, *To develop a universal gamut mapping algorithm*, Ph.D. thesis, University of Derby, UK, 1998. 6.2.1, 6.3.3.1, 6.4.1

[Mor03] ———, Digital Color Imaging (G. Sharma, ed.), The Electrical Engineering and Applied Signal Processing Series, CRC Press LLC, 2003, pp. 639–685. 6.2.1

[mos] *Mosek aps optimization software.*, http://www.mosek.com. 6.3.4

[Mos51] F. Mosteller, *Remarks on the method of paired comparisons*, Psychometrika **16** (1951), 207. 5.4.2

[MR95] R. Motwani and P. Raghavan, *Randomized algorithms*, Cambridge University Press, New York, 1995. 4.3

[MS01] J. Morovic and P. L. Sun, *Non-iterative minimum de gamut clipping*, Ninth Color Imaging Conference, vol. 9, The Society for Imaging Science and Technology, 2001, pp. 251–256. 6.2.1

[Ski98] Steven S. Skiena, *The algorithm design manual*, Springer-Verlag, 1998. 6.3.3.1

[Sof] Sawtooth Software, `www.sawtoothsoftware.com`. 5.1, 5.5, 8

[Ste46] S.S. Stevens, *On the theory of scales of measurement.*, Science **103** (1946), 677–680. 2

[THS04] O. Toubia, J. R. Hauser, and D. I. Simester, *Polyhedral methods for adaptive choice-based conjoint analysis*, Journal of Marketing Research **41(1)** (2004), 116–131. 3.1, 3.2, 3.3.2

[Thu27] L Thurstone, *A law of comparative judgement*, Psychological Review, 1927, pp. 273–286. 1.2, 5.1

[ZS06] P. Zolliker and K. Simon, *Continuity of gamut mapping algorithms*, Journal of Electronic Imaging **15** (2006), no. 1, 13004. 6.3.1, 6.3.1

[ZS07] ———, *Retaining local image information in gamut mapping algorithms*, IEEE Transactions on Image Processing **16** (2007), no. 3, 664–672. 6.2.1

Curriculum Vitae

Eva Schuberth

born on July 21, 1979 in Pfarrkirchen, Germany

1989 - 1998	Wilhelm-Diess Gymnasium in Pocking, Germany Degree: Abitur
1998 - 2004	Studies of Computer Science at University of Passau, Germany Degree: Diplom-Informatikerin
2001 - 2002	Bayerische Elite-Akademie, München, Germany
since June 2004	Ph.D. student at ETH Zurich, Switzerland