Das DFG-Graduiertenkolleg GRK 2297 „Mathematische Komplexitätsreduktion“

Das DFG-Graduiertenkolleg GRK 2297 zum Thema „Mathematische Komplexitätsreduktion“ (MATHCoRE, www.mathcore.ovgu.de) wurde zum April 2017 an der Otto-von-Guericke Universität Magdeburg eingerichtet.

Die Förderung durch die Deutsche Forschungsgemeinschaft umfasst 15 Promotionsstellen und eine Postdoktorandinnenstelle. Sprecher des Graduiertenkollegs ist Prof. Dr. Sebastian Sager.

Übergreifendes thematisches Ziel des Graduiertenkollegs ist die Entwicklung und die Untersuchung von Ansätzen zur Reduktion von Komplexität auf drei Ebenen: der mathematischen Modellierung realer Probleme, der strukturellen Beschreibung mathematischer Objekte und der algorithmischen Lösung mathematischer Aufgabenstellungen. Dabei decken die beteiligten Arbeitsgruppen ein breites Spektrum von Disziplinen in den Bereichen Numerik, Stochastik, Optimierung, Systemtheorie, Kombinatorik, Algebra und Geometrie ab.

Polynome und semialgebraische Mengen, insbesondere auch konvexe Mengen wie Spektraeder und Polytope, spielen wichtige Rollen in den Forschungsarbeiten im Graduiertenkolleg. Das spiegelt sich im (links gezeigten) MATHCoRE-Logo wieder, welches aus Schnittkurven des Elliptops

$$\left\{ (x_1, x_2, x_3) \in \mathbb{R}^3 : \begin{pmatrix} 1 & x_1 & x_2 \\ x_1 & 1 & x_3 \\ x_2 & x_3 & 1 \end{pmatrix} \text{ ist positiv semidefinit} \right\}$$

mit zwei zueinander parallelen Ebenen gebildet ist.

Frank Röttger hat in diesem Frühjahr die Reihe der im Graduiertenkolleg fertig gestellten Promotionen eröffnet. Im folgenden, gemeinsam mit seinen beiden Betreuern Rainer Schwabe und Thomas Kahle verfassten Beitrag zur Geometrie optimaler Versuchspläne gibt er einen Einblick in das zwei klassische mathematische Disziplinen verbindende Gebiet seiner Dissertation, die Algebraische Statistik.

Geometrie optimaler Versuchspläne

Thomas Kahle, Frank Röttger und Rainer Schwabe

Fast jede Entscheidung ist mit Unsicherheiten behaftet und die Wissenschaft soll uns ermöglichen, Risiken zu bewerten und optimal zu handeln. Im Graduiertenkolleg „Mathematische Komplexitätsreduktion“ wird auch die Theorie der *statistischen Versuchsplanung* untersucht, die helfen soll, Experimente und Datengewinnung möglichst effizient und informativ zu gestalten.

Komplexität soll sichtbar gemacht und dann reduziert werden, wodurch wir Risikokompetenz erlangen und auch unter Unsicherheiten methodisch vorgehen können. Wir wollen die Möglichkeiten und Prinzipien zunächst in einem kleinen Beispiel einführen.

Es sollen acht Objekte $O_1, \dots, O_8$ mit unbekannten Gewichten $g_1, \dots, g_8$ mittels einer Balkenwaage vermessen werden. Dazu stehen uns kalibrierte Gegengewichte zur Verfü-

gung. Die Gegengewichte und Objekte können dabei auf beiden Seiten platziert werden. Die Balkenwaage funktioniert bekanntlich so, dass man durch Ausgleich beider Seiten feststellen kann, dass auf beiden Seiten ungefähr die gleiche Kraft wirkt, die Objekte in den beiden Schalen also ungefähr das gleiche Gewicht haben. Bei jeder Wägung machen wir einen kleinen Fehler, weil wir die Waage nicht perfekt austarieren können. Diese Fehler sind zufällig. Wir

nehmen an, die Wägungen sind unabhängig und identisch verteilt, haben den Mittelwert Null und eine Varianz von σ^2 . Man kann nun einfach die Gewichte $g_1, \dots, g_8$ bestimmen, indem $O_1, \dots, O_8$ nacheinander auf der Waage platziert und mittels der Gegengewichte gewogen werden. Dazu sind acht Wägungen nötig. Wir interessieren uns für die Genauigkeit unserer Messung. Da wir mit zufälligen Fehlern modelliert haben, nehmen wir Statistik zur Hand und betrachten *Schätzer*, also Abbildungen, die die Daten des Experiments in uns interessierende Größen umrechnen. Beim direkten Wiegen sind die Schätzungen ganz einfach die Messwerte. Die Schätzer sind Zufallsvariablen, deren Varianz ein Maß für die Güte des Experiments sein soll. Beim direkten Wiegen hat der Schätzer jedes Gewichts eine Varianz von σ^2 .

Die Versuchsplanung fragt nun: *Kann man hier noch etwas verbessern?* Es leuchtet ein, dass man die Fehler ausmitteln kann, wenn man mehrfach wiegt. Daher wollen wir die Güte aber auf die Anzahl Wägungen beziehen und hier einfach als acht fixieren. In den 1940er Jahren hat sich der US-amerikanische Statistiker und Ökonom Harold Hotelling folgenden Plan zur Verringerung der Varianz ausgedacht [3]. Er nutzt seine acht Wägungen wie in Tabelle 1. Dabei gleicht er jeweils die beiden Seiten aus und bestimmt so effektiv die Differenz der Gewichte der Objekte in der linken und der rechten Schale. Wir nennen diese acht Messwerte $Y = (Y_1, \dots, Y_8)^T$.

Tabelle 1. Hotellings Wägungsplan

Wägung	Links	Rechts
1	$O_1, \dots, O_8$	Nichts
2	O_1, O_2, O_3, O_8	O_4, O_5, O_6, O_7
3	O_1, O_4, O_5, O_8	O_2, O_3, O_6, O_7
4	O_1, O_6, O_7, O_8	O_2, O_3, O_4, O_5
5	O_2, O_4, O_6, O_8	O_1, O_3, O_5, O_7
6	O_2, O_5, O_7, O_8	O_1, O_3, O_4, O_6
7	O_3, O_4, O_7, O_8	O_1, O_2, O_5, O_6
8	O_3, O_5, O_6, O_8	O_1, O_2, O_4, O_7

Wir beobachten: Jedes Objekt wird achtmal gewogen und eine kurze Rechnung zeigt, dass sich die Varianz der Schätzer verringert. Dazu beobachtet man, dass sich der Erwartungswert der Größen Y_i linear aus $g = (g_1, \dots, g_8)^T$ berechnet, jeweils versehen mit einem Messfehler $\epsilon = (\epsilon_1, \dots, \epsilon_8)^T$. Es gilt $Y = Hg + \epsilon$ wobei

$$H = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\ -1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 \end{pmatrix}$$

eine Hadamardmatrix ist. Dabei bedeutet ein Eintrag „1“ in der i -ten Zeile und j -ten Spalte von H , dass das Objekt O_j bei der i -ten Wägung in der linken Waagschale, und „-1“, dass es in der rechten Waagschale liegt. Es gilt $H^T H = 8E$

(E bezeichnet hier die Einheitsmatrix), also kann H einfach durch Transponieren und Division durch 8 invertiert werden: $g = 1/8(H^T Y - H^T \epsilon)$. Nun hat der transformierte Fehlerterm $1/8H^T \epsilon$ den Mittelwert 0 und die Varianz $\sigma^2/8$ – eine wesentliche Verbesserung. Der Schätzer für die Gewichte ist $\hat{g} = 1/8H^T Y$. Um also z. B. das Gewicht g_1 von O_1 zu schätzen, bestimmt man

$$\hat{g}_1 = \frac{Y_1 + Y_2 + Y_3 + Y_4 - Y_5 - Y_6 - Y_7 - Y_8}{8}.$$

Hadamardmatrizen sind nur eine der vielen Verbindungen zwischen der Versuchsplanung und Algebra sowie Kombinatorik. Viele dieser Verbindungen sind (im Gegensatz zum obigen Beispiel) nicht-linear. In den letzten Jahrzehnten wurden hier viele Fortschritte theoretischer und algorithmischer Natur gemacht. Im Graduiertenkolleg „Mathematische Komplexitätsreduktion“ in Magdeburg arbeiten Wissenschaftlerinnen und Wissenschaftler aus verschiedenen Disziplinen gemeinsam. Die Doktorarbeit von Frank Röttger [7] ist in der *Algebraischen Statistik* verortet, also dem mathematischen Gebiet, in dem nicht-lineare Algebra und Datenanalyse zusammenkommen.

Für nicht-lineare Modelle sind viele aus der linearen Theorie vertraute Eigenschaften nicht mehr gegeben. Hotellings Plan, zum Beispiel, hängt nicht von den Gewichten der Objekte ab. Egal wie schwer die einzelnen Objekte sind, Tabelle 1 bleibt immer gleich und das Ergebnis $\frac{\sigma^2}{8}$ für die Varianz immer gültig. Es stellt sich heraus, dass für nicht-lineare Modelle der optimale Versuchsplan, also das Analogon von Tabelle 1, von den zu bestimmenden Größen abhängt. Diesem Henne-Ei-Problem stellen wir uns mit geometrischen Methoden. Wir unterteilen den Parameterraum in *Optimalitätsregionen*, in denen die Abhängigkeit des optimalen Versuchsplans von den Parametern genau angegeben werden kann.

Wir illustrieren ein nicht-lineares Modell in einem Beispiel. Vor knapp 100 Jahren hat sich Ernst Zermelo mit der Bestimmung der Stärken von Schachspieler*innen beschäftigt [9]. Ihn ärgerte, dass ein Jeder-gegen-Jeden-Schachturnier nur eine Rangfolge der Spieler*innen ausgibt, aber nicht die relativen Spielstärken bestimmt. Außerdem ist bei einem Turnierabbruch keine Wertung möglich. Er wollte ein System entwickeln, in dem jede Partie neue Information über die Spielstärken liefert und ließ sich dazu von der statistischen Physik inspirieren. Dabei nutzt man ein einfaches Wahrscheinlichkeitsmodell für den Ausgang von Schachpartien. Es wird angenommen, dass m Spieler*innen je eine unbekannte Spielstärke $\pi_1, \dots, \pi_m$ haben. Die Wahrscheinlichkeit, dass dann i gegen j gewinnt, ist modelliert als $\frac{\pi_i}{\pi_i + \pi_j}$.

In seiner Arbeit bestimmt Zermelo den *Maximum Likelihood-Schätzer* (abgekürzt MLE für *maximum likelihood estimator*) für die π_i . Maximum Likelihood-Schätzung ist ein auf einem Plausibilitätsansatz beruhendes Verfahren, bei dem als Schätzung genau die Werte für π_i gewählt werden, unter denen die Wahrscheinlichkeit am größten ist, dass genau die gleichen Spielergebnisse eintreten wie beobachtet. Mathematisch sucht man das Maximum

der Likelihoodfunktion

$$L(\pi) = \prod_{i,j} \left(\frac{\pi_i}{\pi_i + \pi_j} \right)^{w_{ij}},$$

wobei w_{ij} zählt, wie oft das Ereignis „ i gewinnt gegen j “ eintritt.

Aus der Sicht der statistischen Versuchsplanung stellt sich nun die Frage, welche Spiele überhaupt gespielt werden sollten, um die maximale Information über die Spielstärken zu erhalten, also etwa, dass die Varianz des MLE möglichst klein wird.

Aus technischen Gründen geht man gerne zu logarithmischen Koordinaten über und setzt $\beta_i = \log(\pi_i)$. Dann erhält man ein *logistisches Regressionsmodell*

$$\mathbb{P}(i \text{ gewinnt gegen } j) = \frac{1}{1 + \exp(-(\beta_i - \beta_j))}.$$

Dieses Modell wird als Bradley-Terry Modell bezeichnet, nach Ralph A. Bradley und Milton E. Terry, die übrigens beide Doktoranden von Hotelling waren. Sie verwendeten dieses Modell 1952 in ihrer Arbeit [1]. Historisch verortet sich das Modell in der *discrete choice theory*, also der Theorie der diskreten Wahlmöglichkeiten. Diese geht auf Gustav Theodor Fechner zurück, welcher damit bereits um 1860 Fragestellungen in der Psychophysik untersuchte. Aktuelle Anwendungen betreffen beispielsweise die Komposition von Fruchtsaftgetränken, Wahlangebote bei Krankenversicherungen oder die Ausstattung von Transportmitteln, also Bereiche, in denen man einzelne Komponenten wählen oder abwählen kann und dann eine Präferenzordnung bestimmen möchte.

Das Bradley-Terry Modell fügt sich in die allgemeinere Klasse der in den 1970er Jahren von John Nelder und Robert Wedderburn etablierten *verallgemeinerten linearen Modelle* (abgekürzt GLM für *generalized linear models*) ein. Diese GLM sind zentral für die Modellierung von Anwendungen und umfassen auch die klassischen linearen Modelle.

In der Doktorarbeit von Frank Röttger [7] geht es unter anderem um die Analyse von nicht-linearen Modellen und Schätzern mithilfe algebraischer Statistik. Hier interessieren wir uns insbesondere dafür, wie optimale Versuchspläne für GLM mithilfe von polynomieller Algebra bestimmt werden können.

Die optimale Auswahl der Versuchseinstellungen eines Experiments ist dabei die zentrale Frage der statistischen Versuchsplanung. Erste Arbeiten zur statistischen Versuchsplanung stammen bereits aus dem 19. und frühen 20. Jahrhundert (zum Beispiel durch Kirstine Smith [8]), allerdings wird im allgemeinen das Buch *The Design of Experiments* [2] von Ronald Fisher aus dem Jahre 1935 als erstes grundlegendes Werk genannt. Fisher war zu dieser Zeit an der Rothamsted Experimental Station beschäftigt, dem ältesten noch existierenden Agrarforschungsinstitut der Welt. Die statistische Versuchsplanung ist überall dort von herausragender Bedeutung, wo Experimente aus wirtschaftlichen, ethischen oder terminlichen Gründen besonders effizient gestaltet werden müssen, also z. B. in der Agrarforschung, in der medizinischen und pharmazeutischen Entwicklung,

aber auch in der Psychologie, in der Marktforschung oder bei technischen Experimenten.

Die zentrale Idee von Fisher ist es, eine statistische *Information* zu betrachten, die heutzutage als „Fisher-Information“ bezeichnet wird. Diese Idee ging auch in die Entwicklung der Informationstheorie durch Shannon ein. Für lineare Modelle, wie im einführenden Beispiel mit der Balkenwaage, ergibt sich die Varianz des MLE als Inverse der Fisher-Information, sodass der Begriff Information eine heuristische Deutung erhält: Umso größer die Information eines Experiments ist, desto präziser ist der Schätzer. Für GLM besteht diese Beziehung zumindest asymptotisch, d. h. für große Anzahlen von Beobachtungen. Die genaue Definition der Fisher-Information lassen wir hier außen vor, sie kann in [4] gefunden werden. Interessant ist, dass es sich um eine Matrix handelt und die daraus resultierenden Optimierungsprobleme oft klassische Matrixanalysis betreffen.

Zur Illustration betrachten wir nun erneut das Bradley-Terry Modell für m Schachspieler*innen. Dann gibt es $\binom{m}{2}$ mögliche Spielansetzungen. Ein Versuchsplan besteht darin, eine vorher festgelegte Anzahl zu spielender Partien auf die möglichen Paarungen aufzuteilen. Ein *Versuchsplan* kann dann verstanden werden als eine Abbildung ξ , die jeder Paarung (i, j) einen nicht-negativen Wert $\xi_{ij} = \xi(i, j) \geq 0$ zuweist, so dass $\sum_{i,j} \xi_{ij} = 1$. Wir interpretieren ξ_{ij} als den Anteil der Partien, die zwischen i und j gespielt werden, an der Gesamtanzahl aller Partien, wobei wir diskrete Effekte bei wenigen Partien nicht berücksichtigen. Dieser Ansatz geht auf eine Arbeit von Kiefer zurück [5], und solche Versuchspläne werden auch als approximativ bezeichnet. Ihr Vorteil ist, dass reelle Variablen einfacher optimiert werden können. Um aus einem approximativen Versuchsplan ein Schachturnier mit endlich vielen Partien zu erzeugen, muss also gerundet werden. Die dabei auftretenden Verzerrungen lassen wir hier auch außen vor.

Abbildung 1. Graphische Darstellung des Versuchsplans ξ für $m = 4$

Ein *optimaler* Versuchsplan ξ^* ist dann ein approximativer Versuchsplan, welcher die Fisher-Information geeignet maximiert. Die Fisher-Information ist eine (Hesse'sche) Matrix, weshalb zur Optimierung ein sogenanntes Kriterium gewählt wird, welches maximiert wird. In Übereinstimmung mit gängigen Ansätzen in der Literatur verwenden wir das D -Kriterium – die Determinante der Fisher-Informationsmatrix. Diese Wahl ist nicht vollkommen willkürlich, denn die Maximierung der Determinante der Fisher-Information und damit die Minimierung der Determinante der (asymptotischen) Varianz des MLE ist äquivalent zur Minimierung des Volumens des zugehörigen asymptotischen

Abbildung 2. Optimalitätsregionen für $m = 4$

Konfidenzellipsoids, d. h. eines Vertrauensbereichs für die unbekannten relativen Spielstärken.

Praktischerweise ist die Determinante eine konkave Funktion. Dies macht sich der Äquivalenzsatz von Kiefer-Wolfowitz [6] zunutze, welcher die Optimalität eines Versuchsplan mit Hilfe notwendiger Bedingungen erster Ordnung charakterisiert. Im Prinzip funktioniert das bekannte Optimierungsverfahren „Ableitung Null setzen“. Wegen der Konkavität ist das dann auch hinreichend. Aus diesen Bedingungen ergeben sich Gleichungen und Ungleichungen in den Parametern.

Alles zusammengefasst wollen wir also sehen, wie die Lösung (der optimale Versuchsplan) eines Optimierungsproblems (Information maximieren!) von den unbekannten Parametern (Spielstärken) $\beta_1, \dots, \beta_m$ abhängt. Dabei ändern sich sowohl die Zielfunktion als auch der Optimierungsbereich in Abhängigkeit von den β_i . Es zeigt sich aber, dass die Bedingungen von Kiefer und Wolfowitz (nach einer weiteren Umformulierung) *polynomielle* Ungleichungen und Gleichungen sind. Damit können wir Algebra in Stellung bringen. Hieraus ergibt sich die geometrische Einsicht:

Es gibt durch Polynome beschriebene Optimalitätsregionen im Parameterbereich, sodass in jeder Region der optimale Versuchsplan ξ durch eine Formel ausgedrückt werden kann, die die Parameter β in die optimalen Gewichte ξ_{ij} umrechnet.

Um das zu illustrieren, analysieren wir ein fiktives Schachturnier mit den vier Spieler*innen:

1. Markus, ELO: 1872
2. Angelika, ELO: 1415
3. Tamara, ELO: 1021
4. Heinz, ELO: 803

Die ELO-Zahlen sind Vorabschätzungen der Parameter, Markus ist also bisher ein viel stärkerer Spieler als Heinz. Für den heutigen Turnieraussgang ist das aber zunächst irrelevant. Wir wollen im Turnier möglichst viel neue Information gewinnen.

Bei vier Spieler*innen sind insgesamt sechs verschiedene Paarungen möglich. Diese sind durch Linien im Graphen der Abbildung 2 dargestellt. Die Knoten bezeichnen dort die Spieler*innen. In einem Jeder-gegen-Jeden-Zwei-Mal-Turnier würden somit insgesamt 12 Partien gespielt werden. Dabei kommt es auch zweimal zur „uninteressantesten“ Paarung Markus gegen Heinz, von der schon heuristisch kaum neue Information zu erwarten ist. Sollte Heinz gewinnen, hätten wir zwar viel neue Information, aber das ist eben sehr unwahrscheinlich.

Die statistische Versuchsplanung fragt nun, welche 12 Partien angesetzt werden sollten, um möglichst viel über die aktuellen Spielstärken zu lernen. Wir wissen bereits, dass der optimale Versuchsplan von den wirklichen Spielstärken abhängt, aber glücklicherweise haben wir mit den ELO-Zahlen Vorabinformation. Das ist auch in vielen anderen Situationen der Fall, oder aber man führt zunächst einen Pilotversuch durch, um eine grobe Voreinschätzung für die Größen der Parameter zu bekommen. Solche Information kann dann mit geometrischer Information über die Optimalitätsregionen kombiniert werden, um einen passenden Versuchsplan zu wählen.

In unserer Arbeit [4] haben wir unter anderem den Fall $m = 4$ im Detail untersucht und beschrieben, für welche Parameter welcher Versuchsplan optimal ist. Da Spielstärken relativ sind, können sie auf eine*n der Spieler*innen bezogen werden, und es ist nur ein drei-dimensionaler Parameterraum nötig. Dieser ist in Abbildung 2 in Optimalitätsregionen zerlegt, wobei der Übersichtlichkeit halber nicht alle Regionen dargestellt sind. Die fehlenden Regionen sind aber symmetrisch zu den dargestellten Regionen. Dabei gibt es vier Typen von Regionen, welche durch unterschiedliche Farben gekennzeichnet sind.

In der orangen Region sind Versuchspläne mit sechs positiven ξ_{ij} optimal. Diese Region umschließt den Ursprung $\beta = 0$, der die Situation beschreibt, dass alle Spieler*innen gleich stark sind. Das bedeutet: Sind die Spielstärken sehr ähnlich, dann macht es Sinn, alle möglichen Paarungen

anzusetzen. Allerdings können sich deren Häufigkeiten unterscheiden. In der grünen Region werden nur fünf von sechs möglichen Paarungen angesetzt. Hier wird auf die Paarung zwischen der besten und schlechtesten Spieler*in verzichtet. In der blauen Region sind nur noch vier und in der roten Region die minimale Anzahl von drei Paarungen vorgesehen. Dabei spielt bei vier verschiedenen Paarungen entweder die beste Spieler*in nur gegen die zweitbeste oder die schlechteste Spieler*in gegen den zweitschlechtesten. Im Fall von nur drei verschiedenen Paarungen treten nur Spieler*innen mit benachbarter Spielstärke gegeneinander an. Man kann es sich dadurch leicht klar machen, dass bei m Spieler*innen mindestens $m - 1$ verschiedene Paarungen angesetzt werden müssen. Ist der Graph in Abbildung 1 nämlich nicht zusammenhängend, gibt es mehrere Gruppen von Spieler*innen, die durch das Turnier gar nicht miteinander verglichen werden. Man zeigt leicht, dass die Determinante der Information von solchen Plänen einfach null ist.

Eine mathematisch interessante Struktur entsteht in den roten Regionen bzw. für Versuchspläne mit der minimalen Anzahl von verschiedenen angesetzten Partien. Hier konnten wir für eine beliebige Anzahl von Spieler*innen zeigen, dass diese Versuchspläne nur dann optimal sein können, wenn der zugehörige Graph ein Pfad ist, und auch die Optimalitätsregionen konkret angeben. Es gilt dabei, dass Versuchspläne mit einer minimalen Anzahl von genutzten Paarungen die durchgeführten Partien immer gleich oft ansetzen. In unserem Beispiel betragen diese Gewichte also $1/3$. Dieser Versuchsplan kann praktisch so umgesetzt werden, dass jede der drei Paarungen im Turnier viermal angesetzt wird.

Diese Versuchspläne werden als saturierte Versuchspläne bezeichnet. Für das Beispiel mit vier Spieler*innen gibt es genau $\binom{6}{3} = 20$ saturierte Versuchspläne, doch nur 12 können optimal sein. Diese sind genau die saturierten Versuchspläne, deren graphische Darstellung einem Pfad entspricht. Ein Beispiel ist der Graph in Abbildung 3.

Unser fiktives Schachturnier ist am informativsten, wenn jeweils Markus und Angelika, Angelika und Tamara, sowie Tamara und Heinz vier Partien spielen. Auf dieses Ergebnis wäre man auch intuitiv gekommen, aber die Geometrie der Versuchsplanung liefert genaue Parameterwerte, ab denen ein anderer Versuchsplan informativer ist. Wären die Spielstärken sehr eng beieinander, wäre ein Jeder-gegen-Jeden-Turnier optimal. Die Theorie bestätigt also unter anderem die Intuition, dass eine Partie zwischen zwei möglichst gleich starken Spieler*innen informativer ist, als eine, deren Ergebnis von vornherein klar ist.

In unserem Fall hatten wir sehr gute Vorabinformation. Was aber, wenn die Vorabinformation schlecht ist? In diesem Fall stellt sich die Frage, welchen Fehler man macht, wenn man einen nicht-optimalen Versuchsplan verwendet. Dazu betrachtet man die relative Qualität von Versuchsplänen, die *Effizienz*. Sie bemisst den relativen Informationsverlust eines Versuchsplans im Vergleich mit dem jeweiligen optimalen Versuchsplan. Damit kann man abschätzen, ob es sich eher lohnt, den Versuchsplan zu verändern, oder noch in weitere Datenpunkte zu investieren. Beispielsweise bedeutet eine Effizienz von $1/2$, dass mit diesem Versuchs-

Abbildung 3. Graphische Darstellung eines optimalen saturierten Versuchsplans für das fiktive Schachturnier

Abbildung 4. Effizienz des konstanten Versuchsplans. Die vertikalen Trennlinien markieren die Übergänge zwischen den verschiedenen Optimalitätsregionen.

plan doppelt so viele Beobachtungen benötigt werden, um die gleiche Information wie mit dem optimalen Versuchsplan zu gewinnen. Für das Bradley–Terry-Modell haben wir die Effizienz des konstanten Versuchsplans, also des Jeder-gegen-Jeden-gleich-oft-Turniers mit dem optimalen Versuchsplan entlang einer Geraden im Parameterraum berechnet, die alle vier Regionen in Abbildung 2 trifft. Die Effizienz fällt asymptotisch gegen $1/2$, d. h. führt man bei sehr unterschiedlichen Spielstärken ein Jeder-gegen-Jeden-Turnier durch, so braucht man fast doppelt so viele Beobachtungen wie mit dem optimalen Versuchsplan. Dies stimmt überein mit den drei von sechs möglichen Paarungen, die in Abbildung 3 ausgewählt wurden. Die anderen drei Paarungen haben in diesem Fall nur sehr wenig Information. Für größeres m fällt der Unterschied größer aus.

Um interdisziplinäre Ergebnisse wie die hier vorgestellten zu erreichen, ist ein reger Austausch zwischen verschiedenen Gebieten der Mathematik nötig. Im DFG Graduiertenkolleg 2297 „Mathematische Komplexitätsreduktion“ in Magdeburg arbeiten Mathematiker*innen mit verschiedenen Hintergründen gemeinsam an spannenden Problemen aus Anwendungen. Unser Beispiel hier berührt Statistik, Optimierung und auch reelle algebraische Geometrie. Das diverse Magdeburger Forschungsumfeld stellte sich dabei als optimaler Standort für diese Arbeit heraus.

Literatur

- [1] R. A. Bradley and M. E. Terry. Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika*, 39:324–345, 1952.

- [2] R. A. Fisher. *The Design of Experiments*. Oliver and Boyd, 1935.
- [3] H. Hotelling. Some improvements in weighing and other experimental techniques. *Ann. Math. Statist.*, 15(3):297–306, 09 1944.
- [4] T. Kahle, F. Röttger, and R. Schwabe. The semi-algebraic geometry of optimal designs for the Bradley–Terry model. arXiv:1901.02375, Jan 2019.
- [5] J. Kiefer. Optimum experimental designs. *Journal of the Royal Statistical Society: Series B (Methodological)*, 21(2):272–304, 1959.
- [6] J. Kiefer and J. Wolfowitz. The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 12:363–366, 1960.
- [7] F. Röttger. *Geometry of Optimal Design and Limit Theorems*. Dissertation. Otto-von-Guericke Universität Magdeburg, 2020.
- [8] K. Smith. On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations. *Biometrika*, 12(1/2):1–85, 1918.
- [9] E. Zermelo. Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 29:436–460, 1929.

Prof. Dr. Thomas Kahle, Otto-von-Guericke-Universität Magdeburg,
FMA-IAG, Universitätsplatz 2, 39106 Magdeburg
thomas.kahle@ovgu.de

Dr. Frank Röttger, Max-Planck-Institut für Mathematik
in den Naturwissenschaften, Inselstraße 22, 04103 Leipzig
frank.roettger@mis.mpg.de

Prof. Dr. Rainer Schwabe, Otto-von-Guericke-Universität Magdeburg,
FMA-IMST, Universitätsplatz 2 39106 Magdeburg
rainer.schwabe@ovgu.de

Thomas Kahle ist Professor für Algebra an der Otto-von-Guericke Universität in Magdeburg. Nach der Promotion in Leipzig war er als Postdoc unter anderem in Stockholm, Cambridge und Berkeley. Er beschäftigt sich mit kommutativer Algebra, Kombinatorik, algebraischer Statistik und mathematischer Biologie. Zusammen mit Petra Schwer sendet er den Podcast „Pi ist genau 3“.

Frank Röttger ist wissenschaftlicher Mitarbeiter am Max-Planck-Institut für Mathematik in den Naturwissenschaften in Leipzig. Nach dem Studium an der Universität Münster promovierte er von 2017 bis 2020 als Fellow des Graduiertenkollegs „Mathematische Komplexitätsreduktion“ bei Thomas Kahle und Rainer Schwabe an der Otto-von-Guericke-Universität Magdeburg. Seine Forschungsinteressen liegen in der Schnittstelle von Algebra, Statistik und Wahrscheinlichkeitstheorie, unter anderem in der Versuchsplanung und der probabilistischen Kombinatorik.

Rainer Schwabe ist Professor im Ruhestand für Statistik an der Otto-von-Guericke-Universität in Magdeburg. Sein Forschungsgebiet ist optimale und adaptive Versuchsplanung, insbesondere für Anwendungen in Medizin und Psychologie. Nach Studium und Forschungsstellen an der Freien Universität Berlin war er als Medizinischer Biometriker am Universitätsklinikum in Tübingen praktisch tätig. Danach war er maßgeblich treibende Kraft bei der Einrichtung von Statistikstudiengängen und dem Ausbau statistischer Beratung in Magdeburg.