Kapitel 1

Grundlagen der kombinatorischen Optimierung

Viele moderne Lösungsansätze für Fragestellungen aus praktischen Bereichen wie Transportlogistik oder Produktionsplanung verdanken ihre Entstehung einer stark theoretisch motivierten Grundlagenarbeit. Insbesondere beschäftigt man sich dabei mit Maximierungs- bzw. Minimierungsaufgaben, bei denen man es mit diskreten Zulässigkeitsgebieten zu tun hat: So wird eine Maschine eingeschaltet oder auch nicht, eine Verbindung existiert oder nicht, und eine Spedition kann nur ganze LKWs auf den Weg zu ihren Kunden schicken. Daraus ergibt sich in Theorie und Anwendung die Suche nach einem ganzzahligen Vektor, der ein lineares Ungleichungssystem erfüllt und gleichzeitig den Optimalwert bezüglich einer konvexen (oder sogar linearen) Funktion erreicht.

Die kombinatorische Optimierung umfaßt und vereint Aspekte aus Mathematik und Informatik. Forschungsschwerpunkte sind dabei unter anderem das Studium der Struktur des Lösungsraumes und die Entwicklung „beweisbar guter“ und in der Praxis erfolgreicher Lösungsstrategien. Im folgenden sind einige Beispiele für die Arbeit unserer Gruppe in diesen Bereichen aufgeführt.

1.1 Polyedrische Beschreibungen von Reihenfolgeplanungsproblemen

Reihenfolgeplanungsprobleme behandeln die optimale Zuweisung von Aktivitäten an knappe Ressourcen im Zeitablauf. Aktivitäten können Arbeitschritte an einem Werkstück, Instruktionen in einem Computerprogramm, Verfahrensschritte in chemischen Prozessen usw. sein. Ressourcen sind etwa die Anzahl der zur Verfügung stehenden Arbeiter oder Werkbänke, Anzahl der Prozessoren, Hauptspeicherangebot, Betriebsmittel usw. Die Zuweisung von Aktivitäten im zeitlichen Ablauf wird durch einen *Reihenfolgeplan* („Schedule“) festgelegt. In einem Reihenfolgeplan wird jeder Aktivität ein oder mehrere Zeitintervalle auf einer oder mehreren Maschinen bzw. Ressourcen zugewiesen. Ein Reihenfolgeplan ist zulässig, wenn alle Nebenbedingungen eingehalten werden. Die Optimalitätskriterien können unterschiedlicher Natur sein, wie die Minimierung der Gesamtlaufzeit, die Minimierung der Summe der individuellen Fertigstellungszeiten (Durchlaufzeit), das Einhalten vorgegebener Fertigstellungszeiten usw.

Seit den fünfziger Jahren werden Reihenfolgeplanungsprobleme im Gebiet des Operations Research untersucht. Nicht nur die große Fülle an industriellen Anwendungen (siehe z. B. Abschnitt 3.6) macht diese Fragestellungen interessant. Aus der Sicht des Wissenschaftlers bietet sich eine unüberschaubare Vielfalt leicht zu beschreibender, aber bereits für kleine Instanzen sehr schwer zu lösender Aufgaben dar. Eine theoretisch befriedigende Erschließung des gesamten Gebietes der Reihenfolgeplanung ist bisher nicht gelungen, selbst die Anpassung von Standardverfahren (Dynamische Programmierung, Branch-and-Bound usw.) zur Lösung von Spezialfällen ist nicht einfach, was die Vielzahl an Publikationen dazu belegt.

Scheduling–Probleme lassen sich auf verschiedene Weisen als Probleme mit Ganzzahligkeitsrestriktionen auffassen. Diese Ganzzahligkeitsbedingungen erschweren i. a. das Finden einer optimalen Lösung. Für die Anwendung von Verfahren der ganzzahligen Linearen Optimierung auf Scheduling–Probleme — die in jüngster Zeit verstärkt untersucht werden — ist eine problemadäquate und verfahrensgeeignete Modellierung entscheidend. Um mit Hilfe von Methoden der polyedrischen Kombinatorik ein Optimierungsproblem $\min\{\, c(L) \mid L \in \mathcal{L}\,\}$ über einer Menge $\mathcal{L}$ von Reihenfolgeplänen zu lösen, ordnet man jedem Reihenfolgeplan L einen Vektor $x(L)$ im $\mathbb{R}^n$ (für geeignetes n) zu und studiert dann die konvexe Hülle $\mathcal{P}$ der derart gebildeten Vektoren:

$$\mathcal{P} = \text{conv}\ \{\ x(L) \mid L \in \mathcal{L}\ \}.$$

Die Suche nach linearen Beschreibungen dieser Polyeder ist einer der wichtigsten Schritte für die Anwendung von Verfahren der ganzzahligen Linearen Optimierung. Vollständige lineare Beschreibungen, d.h. die Beschreibung des Polyeders nur durch lineare Ungleichungen, ermöglichen oft die polynomielle Lösbarkeit des untersuchten Problems und über die Dualitätstheorie der Linearen Programmierung die Herleitung von kombinatorischen Minimum–Maximum Aussagen. Unvollständige lineare Beschreibungen werden zur Gewinnung von Schranken in den Verfahren der ganzzahligen Linearen Optimierung benutzt.

Wir haben uns vor allem mit dem deterministischen Ein-Maschinen-Belegungsproblem unter gewissen Reihenfolgerestriktionen beschäftigt. Hier muß eine Menge $\{a, b, c, \dots\}$ von Aktivitäten auf einer Maschine ausgeführt werden. Dabei kann die Maschine zu jeder Zeit nur eine Aktivität ausführen, die Aktivitäten dürfen nicht unterbrochen werden, und zwischen einigen Aktivitäten bestehen Reihenfolgerestriktionen nach der Art: „a muß vor b ausgeführt werden“. Gesucht wird nun ein Belegungsplan – eine vollständige Anordnung der Aktivitäten –, der ein bestimmtes Zielkriterium optimiert.

Basispolyeder serien-paralleler Ordnungen

Motivierend für das Studium dieser Strukturen sind Reihenfolgeplanungsprobleme mit der Zielfunktion „Minimierung der (gewichteten) Summe der individuellen Fertigstellungszeiten“. Hier wird in einem Belegungsplan jede Aktivität mit ihrer Fertigstellungszeit bewertet, der Wert des Belegungsplanes ist die Summe der (gewichteten) Fertigstellungszeiten. Im einfachen Fall konstanter Prozeßzeiten von einer Einheit entspricht die Fertigstellungszeit einer Aktivität genau ihrer Position im Belegungsplan, wenn der Maschine keine Leerlaufzeiten zugewiesen werden. Die mit den Belegungsplänen korrespondierenden Vektoren der Fertigstellungszeiten sind genau die mit der Partialordnung verträglichen Permutationen.

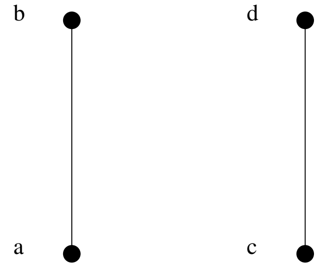

Abbildung 1.1: *Eine einfache serienparallele Ordnung*

Die vier Aktivitäten aus Beispiel 1.1 mögen die Gewichte („relative Wichtigkeit“) $w_a = 1, w_b = 4, w_c = 3$ und $w_d = 2$ und eine Laufzeit von jeweils einer Zeiteinheit haben. Eine optimaler Plan wäre dann $(\ c\, a\, b\, d\ )$ mit einem Zielfunktionswert von $3 \cdot 1 + 1 \cdot 2 + 4 \cdot 3 + 2 \cdot 4$.

Aufbauend auf früheren Arbeiten verschiedener Autoren zur konvexen Hülle der Fertigstellungszeitenvektoren haben wir durch Einführung supermodularer Funktionen auf serien–parallelen Partialordnungen erstmalig Supermodularität mit Reihenfolgerestriktionen verbunden. Sub– bzw. Supermodularität kann als diskretes Analogon der Konvexität verstanden werden. Sehr viele kombinatorische Optimierungsprobleme haben grundlegend eine sub– oder supermodulare Natur, mit dieser Erkenntnis können sie vielfach gelöst werden. Das Polytop der so erzielten Inzidenzvektoren linearer Erweiterungen stellt deshalb eine Verallgemeinerung wie auch eine Vereinheitlichung des Permutaeders von serien–parallelen Ordnungen und des Basispolyeders supermodularer Funktionen auf ungeordneten Mengen dar. Wir konnten für einige wichtige Problemklassen eine vollständige polyedrische Beschreibung finden. Durch die Erweiterung des Greedy–Algorithmus von ungeordneten Mengen auf partielle Ordnungen haben wir das Verfahren von Sidney (1975) der rekursiven Wahl rho–maximaler Ideale zur Lösung des Reihenfolgeplanungsproblems strukturell erklären können.

Mathematisch läßt sich die Situation wie folgt beschreiben:

Sei $P = (P, <)$ eine partiell geordnete Menge und g eine supermodulare Funktion auf den Idealen von P. Jeder linearen Erweiterung $L = e_1 e_2 \dots e_n$ von P ordnen wir durch $x(e_i) := g(\{e_1, \dots, e_i\}) - g(\{e_1, \dots, e_{i-1}\})$ einen Inzidenzvektor $x(L) \in \mathbb{R}^P$ zu. Die konvexe Hülle der Inzidenzvektoren aller linearen Erweiterungen von P bildet das Basispolyeder $\mathrm{B}(P, g) := \text{conv}\,\{\, x(L) \mid L \in \mathcal{L}(P)\,\}$ von P und g. Die supermodulare Funktion g heißt kompatibel mit P, wenn für alle seriell-reduzierbaren konvexen Mengen $A \oplus B$ in P und alle Ideale $I \subseteq P \setminus (A \cup B)$, für die $I \cup A$ und $I \cup A \cup B$ auch Ideale sind, der Term

$$|A|\,[g(I \cup A \cup B) - g(I \cup A)] \\ - |B|\,[g(I \cup A) - g(I)] =: f(A, B)$$

konstant (unabhängig von I) ist. Eine Teilmenge $S \subseteq P$ heißt konvex, wenn für alle Elemente $x < z$ aus S auch alle Elemente „dazwischen", d.h. $x < y < z$, in S enthalten sind. $S = A \dot\cup B$ ist seriell-reduzierbar, wenn alle Elemente von B über allen Elementen von A liegen.

Das Hauptergebnis lautet nun:

Sei P eine serien-parallele Ordnung, sei $g : I(P) \to \mathbb{R}$ supermodular und kompatibel. Das Basispolyeder $B(P, g)$ von P und g wird durch die linearen Ungleichungen

$$\begin{array}{rcl} x(I) & \geq & g(I) \\ & & \textit{für alle Ideale } I \textit{ in } P, \\ |A|x(B) - |B|x(A) & \geq & f(A, B) \\ & & \textit{für alle seriell-reduzierbaren} \\ & & \textit{konvexen Mengen } A \oplus B \\ & & \textit{in } P,\ A, B \textit{ seriell-prim}, \\ x(P) & = & g(P) \end{array}$$

vollständig beschrieben.

Rüstpolyeder partiell geordneter Mengen

Die Entwicklung und das Studium dieser Polyeder ist durch die Zielfunktion „Minimierung der Summe der ordnungsinduzierten Rüstkosten" motiviert. Folgen in einem Belegungsplan zwei Aktivitäten e und f unmittelbar aufeinander, die keiner Reihenfolgerestriktion unterliegen, oder ist f die erste Aktivität, so werden Rüstkosten bei f notwendig. Die Rüstkosten hängen dabei nur von f ab. Ist der unmittelbare Vorgänger von f im Belegungsplan ein notwendiger Vorgänger, so fallen keine Rüstkosten an. Der Wert des Belegungsplanes ist die Summe der angefallenen Rüstkosten. Bei reihenfolgeunabhängigen Prozeßzeiten und nichtnegativen Rüstkosten (Rüstzeiten) entspricht die Rüstkostenminimierung der Minimierung der maximalen Fertigstellungszeit.

In Beispiel 1.2 mögen die Aktivitäten x, b, c und d jeweils Rüstkosten eins haben, die Aktivität a habe Rüstkosten von drei. Dann hat der Reihenfolgeplan ($x\ a \mid c\ d \mid b$) insgesamt Rüstkosten drei. Der Reihenfolgeplan ($x\ c\ d \mid a\ b$) trägt Rüstkosten von vier, obwohl nur zwei anstelle von drei Rüstschritten notwendig sind.

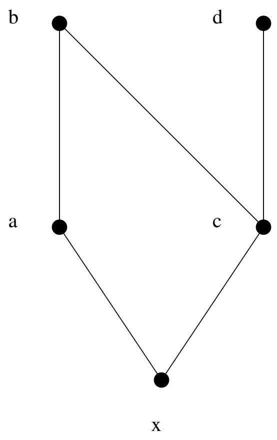

Abbildung 1.2: *Eine Partialordnung mit einem isolierten N*

Schon der Spezialfall konstanter Rüstkosten einer Einheit ist ein NP–schweres Problem, das sogenannte Sprungzahlproblem. Interessanterweise ist der Spezialfall konstanter Rüstkosten von minus eins polynomiell lösbar. Das Sprungzahlproblem wird seit fast 30 Jahren untersucht und manchmal zu Unrecht zu den irregulären Problemen gezählt, die wenig oder gar keine strukturellen Erkenntnisse zulassen. Die bekannten polyedrischen Modelle aus der Reihenfolgeplanung bilden das Rüstproblem nicht zufriedenstellend ab, da sie die Gutartigkeit bestimmter Klassen von Reihenfolgerestriktionen nicht berücksichtigen oder aber keine Lineare Optimierung ermöglichen. Wir haben hier zum ersten Mal ein allgemeines Modell für beliebige Rüstkosten aufgezeigt und dann eine vollständige polyedrische Beschreibung für serien-parallele und N-dünne Ordnungen entwickelt.

Dieser neue polyedrische Ansatz für ordnungsinduzierte Rüstprobleme kann in folgender Weise beschrieben werden:

Zu jeder linearen Erweiterung L einer partiellen Ordnung P definieren wir den 0-1-Inzidenzvektor $x(L) \in \mathbb{R}^P$ durch $x(L)_e = 1$ genau dann, wenn der unmittelbare Vorgänger des Elementes e in der linearen Erweiterung L kein Vorgänger von e in der Partialordnung P ist oder wenn e das erste Element in L ist. Diese Zuordnung ist i.a. nicht injektiv. Das Rüstpolyeder $\mathcal{P} = \text{conv}\,\{\, x(L) \mid L \in \mathcal{L}(P)\,\}$ ist die konvexe Hülle der Inzidenzvektoren aller linearen Erweiterungen von P. Hier lautet ein Hauptergebnis:

Sei P eine serien-parallele Ordnung, und seien $A_i \oplus B_i$, $i \in F$, alle maximalen bipartiten seriell-reduzierbaren *konvexen Mengen in P. Dann wird das Rüstpolyeder $\mathcal{P}$ vollständig durch die linearen Ungleichungen*

$$\begin{array}{ll} \forall\, e \in P: & 0 \le x_e \le 1\ , \\ \forall\, e \in Min(P): & x_e = 1\ , \\ \forall\, i \in F: & x(B_i) \ge |B_i| - 1\ , \\ \forall\, I \subseteq F: & x(B_I) - x(C_I) \le |B_I| - |I| \end{array}$$

beschrieben, wobei $C_i := \{\, e \in P \mid \forall\, b \in B_i\ :\ e \parallel b \,\}$ und für $I \subseteq F$ die Mengen B_I und C_I durch Vereinigung definiert sind ($B_I = \bigcup_{i \in I} B_i$, C_I entsprechend).

(Dabei heißt eine Teilmenge $S \subseteq P$ heißt bipartit, wenn sie keine drei Elemente $x < y < z$ enthält.) Die Ganzzahligkeit des Polyeders konnte mit einem „Box-TDI“-System gezeigt werden. Daneben mußte auch zu einem gegebenem 0-1-Vektor die Existenz einer passenden linearen Erweiterung nachgewiesen werden. Durch eine Erweiterung auf lokale N-Strukturen haben wir dann die vollständige lineare Beschreibung des Rüstpolyeders von N-dünnen Ordnungen hergeleitet. Im Gegensatz zu serien-parallelen Ordnungen ist hier das Rüstproblem selbst mit nichtnegativen Rüstkosten nicht mit dem Greedy-Algorithmus („Beginne mit einem beliebigen zulässigen Element und fahre mit einem teuersten Nachfolger des zuletzt gewählten Elementes fort, wenn dies möglich ist, anderenfalls wähle ein beliebiges zulässiges Element“) lösbar.

Eines der interessanten offenen Probleme ist das Erkennungsproblem für Inzidenzvektoren von linearen Erweiterungen beliebiger Partialordungen:

Existiert zu einem gegebenen $x \in \{0,1\}^P$ eine lineare Erweiterung L mit $x(L) = x$?

Interessant, schwer und bisher unbeantwortet ist auch die Frage nach der Aufnahme individueller Fertigstellungszeiten in dieses Modell, das die Minimierung der Durchlaufzeiten ermöglichen könnte.

1.2 Ganzzahlige Gitter

Viele kombinatorische Optimierungsprobleme können äquivalent als lineare Programme formuliert werden, wobei man von einer Lösung zusätzlich Ganzzahligkeit verlangt. Die Zulässigkeitsbereiche sind dann nicht mehr Polyeder in Vektorräumen wie in der Linearen Optimierung, sondern ganzzahlige Gitter. Während nicht wenige dieser Probleme, die wie z. B. Netzwerkfluß- und Zuweisungsprobleme eine spezielle Struktur besitzen, effizient lösbar sind, ist das allgemeine ganzzahlige Optimierungsproblem schon im linearen Fall beweisbar schwierig (NP-vollständig).

Bisher ist kaum erforscht, wie man an dem zum Optimierungsproblem gehörigen Gitter ablesen kann, wie schwer das ganzzahlige lineare Programm zu lösen ist. Wir haben uns dieser Frage von zwei Seiten genähert.

Gitterbasen binärer Codes

Die Menge der 0-1-Vektoren eines $GF(2)$-Vektorraumes nennt man auch einen binären Code. In kombinatorischen Optimierungsproblemen tauchen diese vor allem als Menge der Zykel (geschlossenen Kantenzüge in einem Graphen) oder als Menge von Schnitten (Kantenmengen, deren Entfernung einen unzusammenhängenden Graphen entstehen läßt) oder allgemeiner als Kreisraum (“cycle space”) und Kokreisraum (“cocycle space”) binärer Matroide auf.

Das von solchen Vektoren erzeugte ganzzahlige Gitter spiegelt einerseits die steigende Komplexität des zugrundeliegenden Problems wieder, andererseits vermuten wir (zusammen mit Martin Loebl und Monique Laurent), daß diese Gitter im allgemeinen folgende bemerkenswerte Eigenschaft haben:

Vermutung: *Das ganzzahlige Gitter eines binären Codes hat eine Basis aus Codewörtern.*

Dies ist ein Spezialfall einer noch allgemeineren Frage über Gitter, die von Delaunay-Polytopen erzeugt werden.

Diese Vermutung an sich widerstand bisher unseren Annäherungsversuchen. Allerdings haben wir inzwischen Gültigkeit in einigen Spezialfällen bewiesen und können außerdem nachweisen, daß kanonische Induktionsansätze, bei denen ein beliebiges Element kontrahiert oder gelöscht wird, zum Scheitern verurteilt sind.

Letzteres zeigen wir, indem wir die Äquivalenz unserer Vermutung zu einer Aussage über Untermatrizen von Hadamard-Matrizen vom Sylvester-Typ voller Zeilenlänge beweisen:

Vermutung: *Das ganzzahlige Gitter, das von den Spalten einer Untermatrix einer Sylvester-Matrix voller Zeilenlänge erzeugt wird, hat eine Basis aus Spalten.*

Die Spezialfälle, bei denen die Gültigkeit der Vermutung bekannt ist (Resultate mit A. Gallucio, M. Loebl und M. Laurent), umfassen Kreise und Schnitte in Graphen, Matroide vom Rang höchstens 4, Kreise in projektiven

Räumen, Matroide mit rekursiv monotoner Struktur, reguläre Matroide, Grafts und Einpunkterweiterungen von Matroiden ohne F_7^*-Minor.

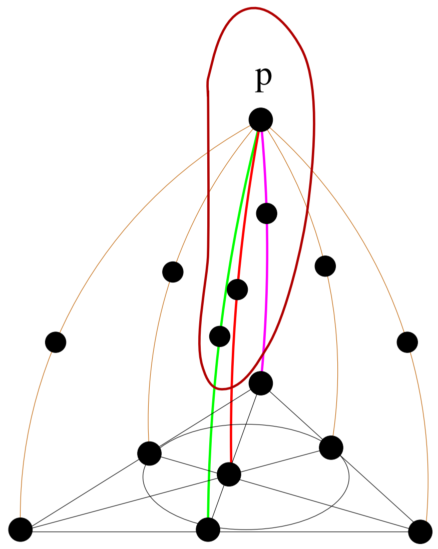

Abbildung 1.3: *Kreisgitter projektiver Räume haben eine Basis aus Codewörtern.*

Gröbnerbasen und ganzzahlige Optimierung

Conti und Traverso haben vor wenigen Jahren einen interessanten Zusammenhang zwischen ganzzahliger Optimierung und Gröbnerbasen in der kommutativen Algebra erkannt. Bei geeignet gewählter Formulierung einer Familie von beschränkten ganzzahligen Programmen läßt sich die Menge der zulässigen Lösungen mit einem Binomideal in einem Polynomring identifizieren. Legt man weiterhin eine bestimmte Monomordnung, die mit der Zielfunktion des Optimierungsproblems korrespondiert, auf dem Polynomring fest, so kann mit dem Buchberger-Algorithmus die Gröbnerbasis des Binomideals ausgerechnet werden. Der reduzierten Gröbnerbasis entspricht im Optimierungskontext die bezüglich Inklusion eindeutige minimale Testmenge, die durch folgende Eigenschaft charakterisiert ist: Eine zulässige Lösung x eines ganzzahligen Optimierungsproblems ist entweder optimal, oder es gibt ein Element t der Testmenge, so daß die Differenz $x - t$ wieder zulässig ist und bezüglich der Zielfunktion einen besseren Wert hat.

Theoretisch besteht also die Möglichkeit, durch den Einsatz algebraischer Methoden ganzzahlige Optimierungsprobleme zu lösen: Erst wird mit dem Buchberger-Algorithmus die Testmenge ermittelt. Anschließend wird eine Folge von Verbesserungsschritten „entlang“ Elementen der Testmenge ausgeführt, bis das Optimum erreicht ist.

Obwohl in dem betrachteten Fall die Gröbnerbasis nur zu Idealen mit recht spezieller Struktur berechnet werden muß, ist dieser Ansatz aufgrund der hohen Laufzeit des Buchberger Algorithmus und der potentiell gewaltigen Größe der Basis für Beispiele mit vielen Variablen sehr problematisch. Daher werden die algebraischen Methoden nicht als direkte Lösungsverfahren betrachtet. Statt dessen werden die Methoden der Computeralgebra im hier verfolgten Ansatz dazu eingesetzt, die minimalen Testmengen zu ganzzahligen Optimierungsproblemen zu berechnen, um anschließend aus den so gewonnenen Informationen Aussagen über die Struktur der Testmengen zu ermöglichen.

Zunächst wurde anhand wohlbekannter kombinatorischer Optimierungsprobleme die Struktur entsprechender Gröbnerbasen studiert. So ist es beispielsweise im Falle der Bestimmung eines Matchings maximalen Gewichts in einem vollständig bipartiten Graphen gelungen, die Testmenge des dazu dualen Problems vollständig zu beschreiben, und zwar unabhängig von der Tatsache, daß zu ihrer Berechnung aufwendige algebraische Methoden eingesetzt wurden. Dieses Resultat scheint sich auf das gewichtete b-Matching-Problem übertragen zu lassen.

Weiterhin wurde ein Vertex-Cover-Problem auf vollständigen Graphen betrachtet. Bei geeigneter Wahl der Monomordnung gilt der

Satz: Bis auf ein Element ist die reduzierte Gröbnerbasis genau die Menge aller Vektoren, die entweder negativ bzgl. der Monomordnung sind und das Gewicht um 1 erniedrigen, oder das Gewicht unverändert lassen und positiv bzgl. der Ordnung sind.

In Abbildung 1.4 sind typische Repräsentanten der minimalen Testmenge für den K_8 graphisch dargestellt. Das linke Element und Elemente vom mittleren Typ erniedrigen das Gewicht einer Knotenüberdeckung, während die Elemente vom rechten Typ lediglich für eine Umverteilung der Gewichte stehen.

Ziel ist es, auch bei schweren ganzzahligen Optimierungsproblemen eine – möglicherweise unvollständige – Beschreibung der Struktur der Testmenge ermitteln zu können. Diese Kenntnisse können dann in einer Verbesserungsheuristik genutzt werden.

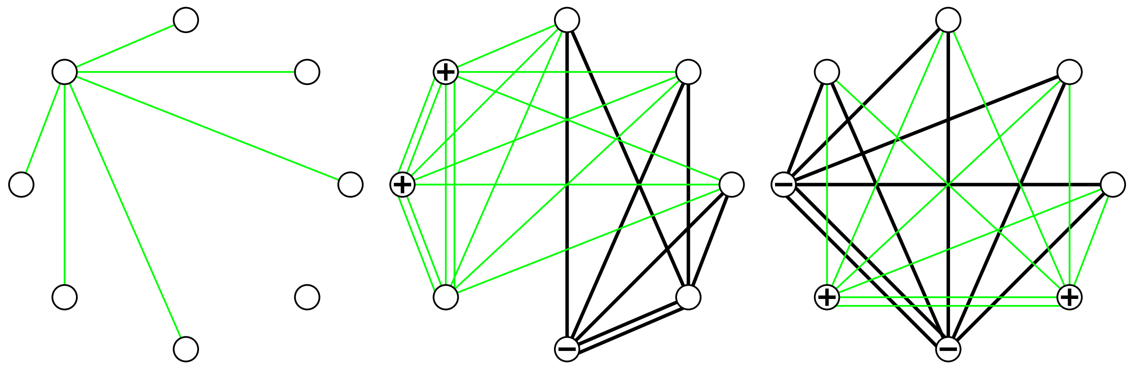

Abbildung 1.4: *Repräsentanten der drei Klassen von Elementen in der minimalen Testmenge für K_8*

1.3 Konvergenzverhalten von Simulated Annealing in endlicher Zeit

Das Simulated Annealing (SA) ist ein randomisiertes Verfahren, das in vielen praktischen Fällen gute Näherungslösungen für kombinatorische Aufgabenstellungen liefert. Diese Technik wurde Anfang der 80er Jahre entwickelt und basiert auf der Analogie zwischen dem Abkühlen geschmolzener Körper und dem Lösen großer Optimierungsprobleme. Hier wie dort ist der Erfolg des Prozesses stark abhängig von der geschickten Wahl eines Kontrollparameters, der „Temperatur". Um möglichst gute „Kühlpläne" für das Simulated Annealing erstellen zu können, ist die Untersuchung des Konvergenzverhaltens des Verfahrens eine notwendige Voraussetzung. Dieser Aufgabe haben wir uns im Rahmen eines von der Deutschen Forschungsgemeinschaft (DFG) geförderten Projektes gewidmet. Die Eignung des Simulated Annealing für praktische Anwendungen konnte am Beispiel der Erstellung von kostengünstigen Flugplänen (Abschnitt 3.2) unter Beweis gestellt werden.

Beim Simulated Annealing wird mit einer zulässigen Lösung des kombinatorischen Optimierungsproblems gestartet und eine zufällig gewählte benachbarte Lösung erzeugt. Hat diese einen besseren Zielfunktionswert, geht man zu ihr über und iteriert. Andernfalls akzeptiert man die neue Lösung nur mit einer gewissen Wahrscheinlichkeit. Diese Wahrscheinlichkeit nimmt mit steigender Iterationszahl ab (Annealing). Dieses Akzeptieren auch von schlechteren Zuständen dient dazu, nicht in lokalen Optima zu verharren, sondern zu einem globalen Optimum zu gelangen.

Das sich für Simulated Annealing in natürlicher Weise anbietende mathematische Modell ist eine inhomogene Markov-Kette. Die Verteilung der Kette konvergiert unter gewissen Voraussetzungen gegen eine vom Startzustand unabhängige Grenzverteilung. Im Falle von SA soll diese auf den Zuständen gleichverteilt konzentriert sein, die bezüglich der Zielfunktion optimale Werte haben. Um die Geschwindigkeit zu untersuchen, mit der die Verteilung der Kette gegen eine solche Gleichverteilung konvergiert, betrachten wir eine strukturelle Größe, die „Conductance", des der Markov-Kette zugrundeliegenden Transitionsgraphen.

Mit Hilfe von Abschätzungen der Conductance ist es Sinclair gelungen, perfekte Matchings in einem bipartiten Graphen gleichverteilt zu erzeugen und so mit hoher Wahrscheinlichkeit die Permanente einer Matrix zu berechnen. Dyer, Frieze und Kannan haben die Ergebnisse benutzt, um Punkte in „genügend glatten", konvexen Körpern gleichverteilt zu erzeugen, und so deren Volumen bestimmt. Schließlich konnte Gademann mit dieser Technik einen polynomiellen, randomisierten Algorithmus entwerfen, der Lineare Programme löst. Unser Ziel ist es, diese Überlegungen, die auf eine homogene Markov-Kette und auf eine Gleichverteilung ausgerichtet waren, auf den für das SA relevanten Fall der inhomogenen Markov-Kette mit der auf die Optima konzentrierten Gleichverteilung zu erweitern. Da hier die Transitionsmatrizen zeitabhängig sind, muß die steuernde Größe jetzt eine zeitabhängige Conductance sein.

Tatsächlich sind uns selbst bei naiver Abschätzung der zeitabhängigen Conductance schon einige Konvergenzbeweise gelungen. Dabei wurden unter anderem ähnliche Resultate wie die von D. Mitra, F. Romeo und A. Sangiovanni-Vincentelli bei logarithmischen Kühlplänen erzielt und teilweise erweitert. Wie nicht anders zu erwarten, liefern diese naiven Abschätzungen der Conductance nur exponentielle Laufzeitschranken, bei denen eine Näherung an die Grenzverteilung garantiert werden kann.

Weiteres Ziel ist es daher, bei schwierigen kombinatorischen Problemen genauere Aussagen über die zeitabhängige Conductance zu erhalten, um polynomielle Laufzeitschranken zu erreichen. Man wird dabei aber nur die Approximation der Optimallösung fordern können. So konnte z. B. bei einer speziellen Version des Quadratic-Assignment-Problems die hinreichende Konvergenz bei nur polynomiellem Zeitbedarf nachgewiesen werden.

Unsere momentane Forschung konzentriert sich auf das NP-vollständige 3-Färbbarkeits-Problem. Dieses Problem scheint sehr geeignet zu sein, entsprechend gutes Konvergenzverhalten zu zeigen, da wir schon eine große Konvergenzgeschwindigkeit bei hohen Temperaturen zeigen konnten. Zur Zeit beschäftigen wir uns mit dem Entwurf von randomisierten Modellen, in denen alle 3-färbbaren Graphen anähernd gleichverteilt vorkommen. Hintergrund dieser Überlegungen ist der Versuch, die Komplexität dieses Problems zu verringern, um in polynomieller Laufzeit eine ausreichende Konvergenz von SA zu erhalten, ohne das Optimum nur approximieren zu müssen. Nächstes Ziel wird dann sein, geeignete Abschätzungen der zeitabhängigen Conductance auch für niedrige Temperaturen zu erhalten.

1.4 Kombinatorische Geometrie

Eine der wichtigsten Methoden der Optimierung für die Anwendung ist die Lineare Optimierung, bei der eine lineare Zielfunktion unter einer Reihe von linearen Nebenbedingungen maximiert wird. Die folgenden beiden Abschnitte behandeln Fragestellungen, die aus der Linearen Programmierung motiviert sind.

Homomorphismendualität

In den letzten Jahren haben J. Nešetřil und Koautoren Dualität als alternative Existenz gewisser Homomorphismen studiert. Zwei der Ausgangspunkte waren Graphen und Digraphen mit ihren Homomorphismen. Für Graphen zum Beispiel ist die Homomorphismen-Dualität eindeutig – und trivial:

Sei G ein Graph. Dann gibt es entweder einen Homomorphismus $K_2 \to G$, oder es gibt einen Homomorphismus $G \to K_1$, aber nicht beides.

Ein Kenner von Graphenhomomorphismen erkennt sofort die tiefe Weisheit, die in dieser Aussage steckt: Entweder ein Graph hat eine Kante, oder er hat keine. Bei Digraphen ist die Situation diffiziler. Hier definiert jeder gerichtete Baum eine Dualität.

Da in der kombinatorischen Optimierung die Dualität der linearen Programmierung in gewissem Sinne eine Art gemeinsamen Nenner fast aller bekannten Dualitätssätze bildet, haben wir die Fragestellung untersucht, inwiefern diese Dualität als Homomorphismendualität aufgefaßt werden kann, und wenn, ob sie dann eindeutig ist.

Die Dualitätstheorie der Linearen Programmierung führt auf kanonische Weise zu einem Axiomensystem für orientierte Matroide. Ausgehend von einem Vorschlag von Lovász und Schrijver haben wir daher untersucht, inwiefern sich dieser allgemeine Dualitätsbegriff in den Rahmen der Homomorphismendualität integrieren läßt. Der offensichtliche Kandidat für die Morphismen sind „strong maps“, die das kombinatorische Gegenstück zu linearen Abbildungen sind. Um zu einem sinnvollen Ansatz zu kommen, muß man zunächst eine affine Version dieser Abbildungen definieren. Allerdings ist hier die einzige existierende Homomorphismendualität wieder nur eine triviale Aussage: Entweder ein Vektorraum hat in der ersten Koordinate irgendwo einen Nicht-Nulleintrag, oder eben nicht. Dies liegt daran, daß die „strong maps“ zwar das Zusammenfassen identischer Koordinaten zulassen aber keine Koordinatenverdopplung, obwohl es sich im linearen Fall hierbei um einen Isomorphismus handelt. Trägt man dieser Tatsache Rechnung und erweitert den Morphismenbegriff, so erhält man als Resultat:

Die Dualität der Linearen Programmierung kann als Homomorphismendualität orientierter Matroide mit verallgemeinerten „strong maps“ aufgefaßt werden und ist mit diesem Morphismenbegriff eindeutig.

Polarität

Polarität, oder Ordnungsdualität von Polyedern, ist eine andere Form der Dualität als die der Linearen Programmierung. Die Unterschiede und Zusammenhänge sind noch nicht völlig verstanden.

Was die Unterschiede angeht, so konnten wir einen Beitrag leisten, der sich harmonisch in klassische Resultate aus der projektiven Geometrie einreiht.

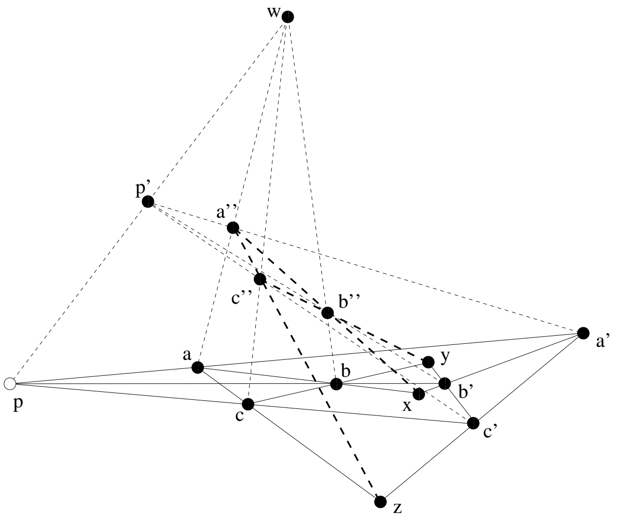

Abbildung 1.5: *Diese Koerweiterung des Non-Desargues-Matroids hat keinen Adjoint eines Adjoints*

Ein nichtlineares Matroid M vom Rang mindestens 4 kann niemals iterierte Polare beliebigen Grades haben, da diese gegen einen projektiven Raum „konvergieren“ würden, in den M eingebettet ist. Allerdings war bislang kein Beispiel eines orientierten Matroids mit Polarer bekannt, dessen Polare sich nicht polarisieren läßt. Es gelang uns, hierfür ein Beispiel zu konstruieren – siehe Abbildung 1.5. Die nicht orientierte Version dieses Beispiels ist eine Koerweiterung des Non-Desargues-Matroids mit einem koparallelen Element. Im orientierten Fall nennt man diese Erweiterung „Lawrence-Konstruktion“. Die Nähe eines solchen Beispiels zum Desarguesschen Satz war eigentlich zu erwarten, da gilt:

Eine projektive Ebene ist linear, genau dann wenn der Satz von Desargues gilt.

Die Existenz einer orientierten Polaren zeigen wir mittels eines neuen Konstruktionsverfahrens, das unter gewissen Voraussetzungen Elemente außerhalb einer Hyperebene verdoppelt.

1.5 Gradbeschränkte Netzwerke

In den Bereichen Netzwerkdesign und VLSI steht man häufig vor der Aufgabe, ein System von Verbindungen mit möglichst günstigen Gesamtkosten zu erstellen; aus Stabilitäts- und Kapazitätsgründen existieren dabei aber Obergrenzen für die Belastung der einzelnen Knoten.

Mathematisch gesprochen steht man vor folgendem Problem: Gegeben ein Graph G mit Kosten $w(e)$ für jede Kante e sowie einer Kapazitätsschranke $d(v)$ für jeden Knoten v. Gesucht ist ein zusammenhängender Teilgraph mit kleinstmöglichen Gesamtkosten, in dem kein Knoten höheren Grad als $d(v)$ hat. Da sich das bekannte Traveling-Salesman-Problem als eine solche Fragestellung formulieren läßt, ist nicht damit zu rechnen, daß es einen exakten Algorithmus gibt, der für jede Probleminstanz in polynomialer Laufzeit eine optimale Lösung findet. Aus diesem Grunde ist man daran interessiert, Methoden zu entwickeln, die in kurzer Zeit möglichst gute Approximationen von Optimallösungen liefern. Ein Ansatz dafür besteht darin, von einem zusammenhängenden Netzwerk kleinstmöglichen Gesamtgewichts auszugehen und diesen minimalen aufspannenden Baum T dann so zu modifizieren, daß die Gradbeschränkungen der einzelnen Knoten eingehalten werden. Es gelang uns, mit diesen Ansatz eine Netzwerkflußmethode zu entwickeln, die in verschiedener Hinsicht optimal ist: Unser Algorithmus kann im allgemeinen von keiner Methode unterboten werden, die nur Informationen über Topologie und Kantengewichte eines minimalen aufspannenden Baumes sowie die Dreiecksungleichung benutzt. Wir können garantieren, daß die Kosten des konstruierten Baumes diejenigen des minimalen aufspannenden Baumes T um nicht mehr als einen Faktor von

$$2 - \min\left\{\frac{d(v)-2}{\deg_T(v)-2} \mid \deg_T(v) > 2\right\}$$

übersteigt, vorausgesetzt, daß $d(v) \geq 2$ für alle Knoten v gilt. (Dabei ist $\deg_T(v)$ der Grad von v im minimalen aufspannenden Baum T.) Dies ist der bestmögliche derartige Faktor. Und schließlich hat eine Variante unserer Methode lineare, also optimale Laufzeit und garantiert gleichzeitig diesen Approximationsfaktor.

Aus unseren Ergebnissen lassen sich zudem Aussagen über Spezialfälle des Problemes ableiten, in denen die Kostenfunktion durch geometrische Abstände der Netzknoten gegeben ist. Insbesondere für den Fall von L_1-Abständen in der Ebene (die sogenannte Manhattan-Metrik, die vor allem für VLSI-Probleme von entscheidender Bedeutung ist) folgt, daß der obengenannte Quotient den Wert von 1,5 nicht übersteigen kann, wenn die Gradbeschränkungen der Knoten $d(v) = 3$ betragen. Dies ist zur Zeit der beste bekannte Wert. An dieser Stelle gibt es noch eine Reihe von offenen Fragen, da nicht klar ist, ob für $d(v) = 3$

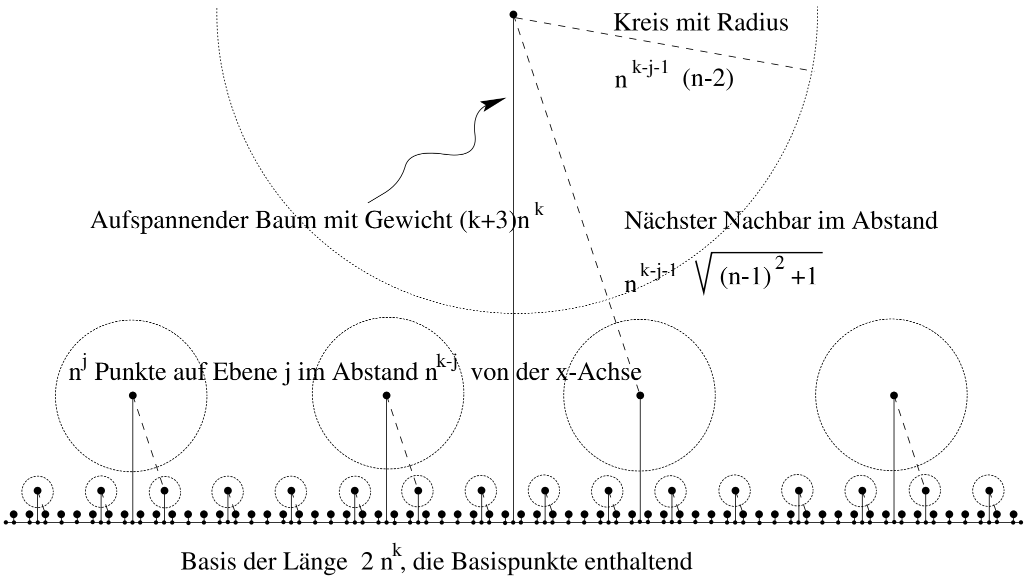

Abbildung 1.6: *Eine Familie von Punktmengen mit großer Tourlänge*

tatsächlich größere Quotienten als $\frac{3+\sqrt{2}}{4} = 1,07\ldots$ auftreten können. Es gelang uns aber, den seit einiger Zeit offenen Fall von $d(v) = 2$ zu lösen, indem wir eine Klasse von geometrischen Beispielen konstruierten, für die der genannte Quotient beliebig nahe an 2 kommen kann.

1.6 Sichtbarkeitsdarstellung von Graphen

Seit vielen Jahrzehnten beschäftigen sich Graphentheoretiker mit Eigenschaften von diskreten mathematischen Strukturen. Mit zunehmendem Einsatz von Computern für Aufgaben in der Visualisierung von Daten ist das Interesse an algorithmischen Verfahren zur Darstellung von Graphen („Graph Drawing") immer größer geworden.

Wie läßt sich ein Graph so zeichnen, daß die Zeichnung möglichst einfach zu verstehen ist? Naturgemäß ist dies keine scharf definierte Aufgabenstellung. Betrachtet man aber Abbildung 1.7, in der die Filestruktur eines Systems zur Prognose der Bestandsentwicklung der Landesbausparkassen dargestellt ist, so wird deutlich, wie wichtig es ist, komplexe Graphen so darzustellen, daß die wichtigen Informationen leicht ablesbar sind.

Angesichts der Vielfältigkeit darzustellender Informationen ist es nicht nur wichtig, Algorithmen zu untersuchen, sondern auch von Bedeutung, neue Konzepte für die Darstellung von Graphen zu entwickeln. Ein solches Konzept ist die sogenannte „Sichtbarkeitsdarstellung" von Graphen. Dabei werden Knoten durch Objekte im Raum repräsentiert, die miteinander durch eine Kante verbunden sind, wenn sie sich auf wohldefinierte Weise „sehen" können. Je nach Art der Objekte und Definition der Sichtbarkeit können recht unterschiedliche Graphen dargestellt werden. Aufgrund ihrer einfachen Struktur interessiert man sich besonders für Objekte wie achsenparallele Quadrate, Rechtecke, Kreise, Würfel und Quader. Interessante Versionen der Sichtbarkeit von zwei Objekten sind dadurch gegeben, daß es eine z-achsenparallele oder allgemein achsenparallele Verbindung eines Paares von jeweils inneren Punkten gibt, die keine anderen Objekte berührt. Abbildung 1.7 zeigt die Struktur eines komplizierten Softwarepaketes und eine Aufsicht auf eine Repräsentierung; dabei entspricht die Größe der verwendeten Rechtecke ungefähr der Wichtigkeit der dargestellten Files. Die (gerichteten) Kanten sind (von unten nach oben) durch Überlappung zweier Rechtecke repräsentiert. Input-Files sind schwarz dargestellt, Output-Files weiß.

Welche Graphen lassen sich auf diese Weise repräsentieren? Bislang gibt es keine algorithmischen Methoden zur Berechnung einer Sichtbarkeitsdarstellung. Es lassen sich aber eine Reihe von mathematischen Aussagen darüber treffen, wo die Möglichkeiten solcher Darstellungen an ihre Grenzen stoßen. Mit Hilfe von kombinatorischen Methoden konnten wir obere und untere Schranken für die Größe von vollständigen Graphen mit Sichtbarkeitsdarstellungen finden.

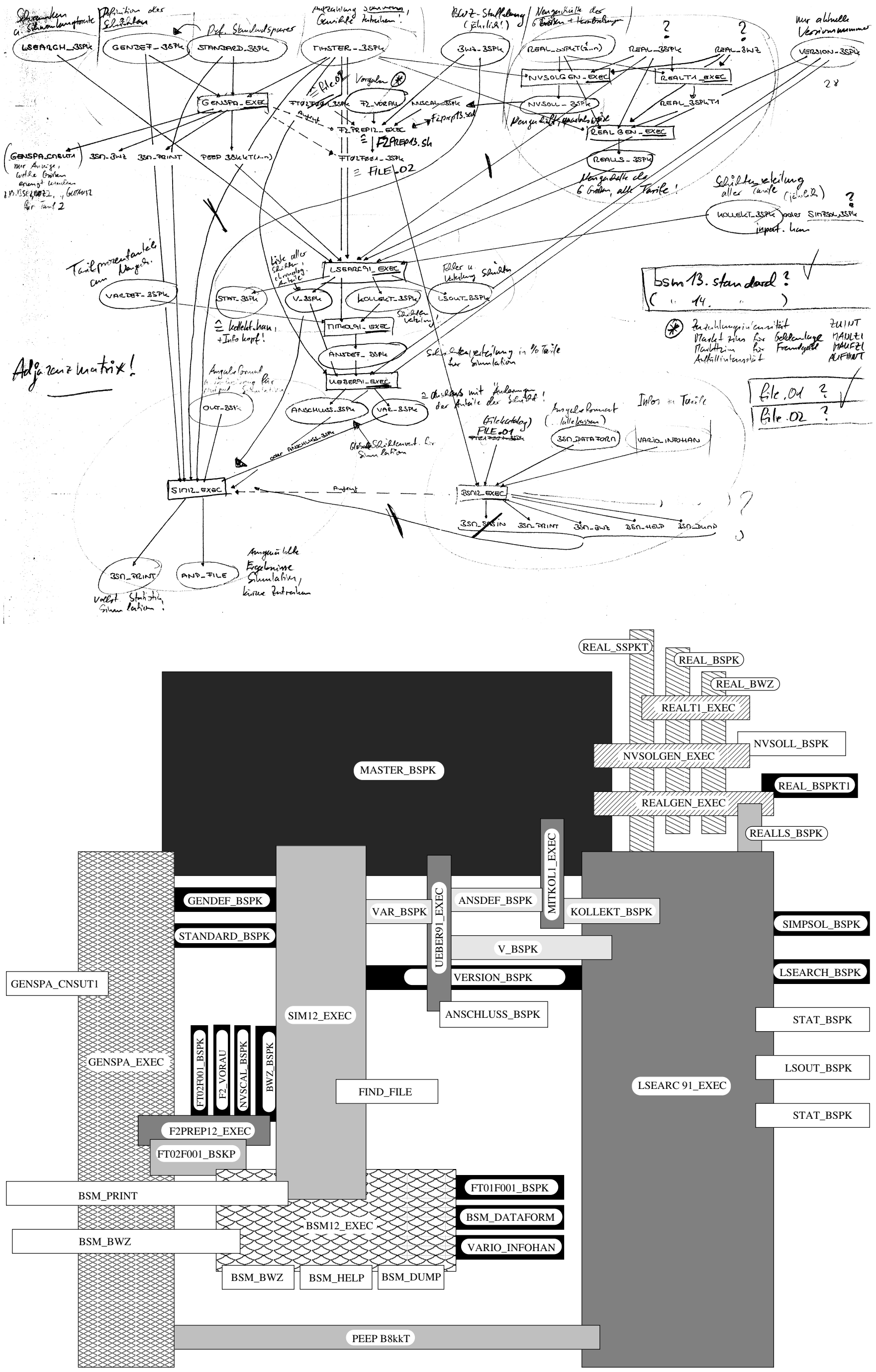

Abbildung 1.7: Ein kompliziertes Softwarepaket: Struktur und Sichtbarkeitsdarstellung