

Template \LaTeX Wiki von BAzubis für BAzubis

Projektarbeit 1 (T3_2000)

im Rahmen der Prüfung zum
Master of Science (B.Sc.)

des Studienganges Informatik

an der Dualen Hochschule Baden-Württemberg Karlsruhe

von

Vorname Nachname

Abgabedatum:	01. Februar 2025
Bearbeitungszeitraum:	01.10.2024 - 31.01.2025
Matrikelnummer, Kurs:	0000000, TINF15B1
Ausbildungsfirma:	SAP SE Dietmar-Hopp-Allee 16 69190 Walldorf, Deutschland
Betreuer der Ausbildungsfirma:	B-Vorname B-Nachname
Gutachter der Dualen Hochschule:	DH-Vorname DH-Nachname

Eidesstattliche Erklärung

Ich versichere hiermit, dass ich meine Projektarbeit 1 (T3_2000) mit dem Thema:

Template \LaTeX Wiki von BAzubis für BAzubis

gemäß § 5 der “Studien- und Prüfungsordnung DHBW Technik” vom 29. September 2017 selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch nicht veröffentlicht.

Ich versichere zudem, dass die eingereichte elektronische Fassung mit der gedruckten Fassung übereinstimmt.

Karlsruhe, den July 18, 2025

Nachname, Vorname

Abstract

- English -

This is the starting point of the Abstract. For the final bachelor thesis, there must be an abstract included in your document. So, start now writing it in German and English. The abstract is a short summary with around 200 to 250 words.

Try to include in this abstract the main question of your work, the methods you used or the main results of your work.

Abstract

- Deutsch -

Dies ist der Beginn des Abstracts. Für die finale Bachelorarbeit musst du ein Abstract in deinem Dokument mit einbauen. So, schreibe es am besten jetzt in Deutsch und Englisch. Das Abstract ist eine kurze Zusammenfassung mit ca. 200 bis 250 Wörtern.

Versuche in das Abstract folgende Punkte aufzunehmen: Fragestellung der Arbeit, methodische Vorgehensweise oder die Hauptergebnisse deiner Arbeit.

Contents

Formelverzeichnis	VI
Abkürzungsverzeichnis	VII
List of Figures	VIII
List of Tables	IX
Quellcodeverzeichnis	X
1 Introduction	1
1.1 Motivation and Contribution	1
1.2 Structure	1
2 Preliminaries	2
2.1 Linear Optimization	2
2.2 Dantzig-Wolfe Decomposition	3
2.3 Graph Theory	4
2.4 Partition Refinement	5
2.5 Surprise and Entropy	6
3 Related Work	7
4 GCG	8
4.1 Detection	9
4.2 Classifiers	11
4.3 Existing Detectors	14
4.4 Example	14
5 Implementation	15
5.1 Architecture	15
5.2 Pre-Processing	15
5.3 Refinement Process	15
5.4 Post-Processing	15
5.5 Scoring	15
6 Evaluation	16
6.1 Setup	16
6.2 StrIPlib	17
6.3 Stuff	18

7	Notes	19
7.1	MIPLIB Constraint Types	19
7.2	Relaxed Constraint Types	20
7.3	Classifiers, Detectors	21
	Literaturverzeichnis	XI

Formelverzeichnis

A	mm ²	Fläche
D	mm	Werkstückdurchmesser
d_{\min}	mm	kleinster Schaftdurchmesser
L_1	mm	Länge des Werkstückes Nr. 1
	Grad	Freiwinkel
	Grad	Keilwinkel

Abkürzungsverzeichnis

GCG	Generic Column Generation
SCIP	SCIP
BaPCod	Branch-and-price Code
hMETIS	Hypergraph METIS
strIPlib	Structured Integer Program Library
MIP	Mixed Integer Programming

List of Figures

2.1	Entropy is measure of “surprise” and it increases with decreasing probability. On the left side, both colors are distributed equally, so one would be equally surprised to draw either one of the colors. On the right side, drawing a red ball from the set of elements would be really surprising, because the probability is only $\frac{1}{12}$. But because this event is very unlikely, one would <i>not expected</i> to be surprised. Therefore, the entropy is low.	6
4.1	A simplified overview of the four major stages of solving a model with GCG.	8
4.2	A simplified overview of the detection process and its detection loop.	9
4.3	Visualization of the induced tree of propagated partial decompositions.	10
4.4	Bin-Packing Model with items $\mathcal{I} = \{1, \dots, n\}$, item sizes $a_i \in \mathbb{Z}_{\geq 0}$, bins $\mathcal{J} = \{1, \dots, m\}$ and capacity C	14
6.1	The distribution of different categories of model within strIPlib. Most problems are part of “common” categories like Bin-Packing, Scheduling and Routing. The category “Others” includes e.g. Fantasy Football, Train Scheduling and different types for which only a small number of model files are available.	17

List of Tables

4.1	The structure of all 17 constraint types MIPLIB keeps track of.	13
6.1	Consumer-grade components used to run all experiments.	16
7.1	A relaxed version of the constraint types MIPLIB uses.	20

Quellcodeverzeichnis

1 Introduction

1.1 Motivation and Contribution

1.2 Structure

2 Preliminaries

2.1 Linear Optimization

2.2 Dantzig-Wolfe Decomposition

2.3 Graph Theory

2.4 Partition Refinement

2.5 Surprise and Entropy

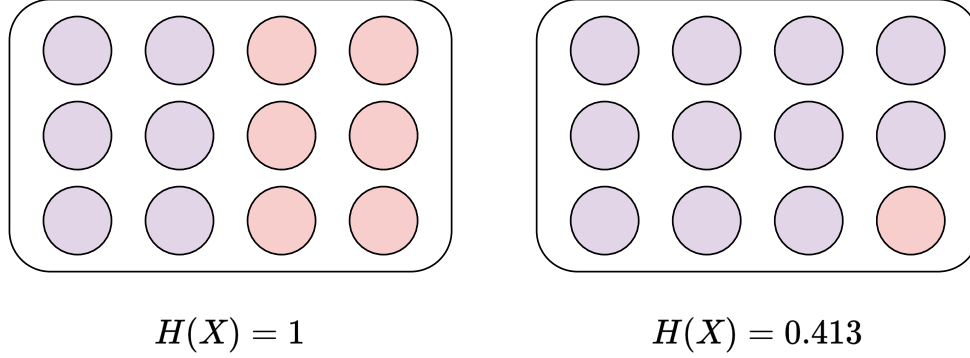


Figure 2.1: Entropy is measure of “surprise” and it increases with decreasing probability. On the left side, both colors are distributed equally, so one would be equally surprised to draw either one of the colors. On the right side, drawing a red ball from the set of elements would be really surprising, because the probability is only $\frac{1}{12}$. But because this event is very unlikely, one would *not expected* to be surprised. Therefore, the entropy is low.

The *information value* or *surprisal* of an event E is defined as

$$I(E) = \log_b \left(\frac{1}{p(E)} \right) = -\log_b (p(E)) \quad (2.1)$$

It increases as the probability of the event $p(E)$ decreases. Intuitively, if the probability is close to 1, then one wouldn’t be surprised if this event actually occurred, so the surprisal is close to 0.

The *entropy*, or *expected surprise*, $H(X)$ of a discrete random variable X which takes values in the set \mathcal{X} is defined by equation 2.2 [1].

$$H(X) = \sum_{x \in \mathcal{X}} p(x) I(X) = - \sum_{x \in \mathcal{X}} p(x) \log_b p(x) \quad (2.2)$$

where $p(x) := \mathbb{P}[X = x]$.

If not specified any further, the base b of the logarithm is assumed to be 2. In chapter [these concepts will be used to define a heuristic scoring system based on constraint names.](#)

ref

3 Related Work

4 GCG

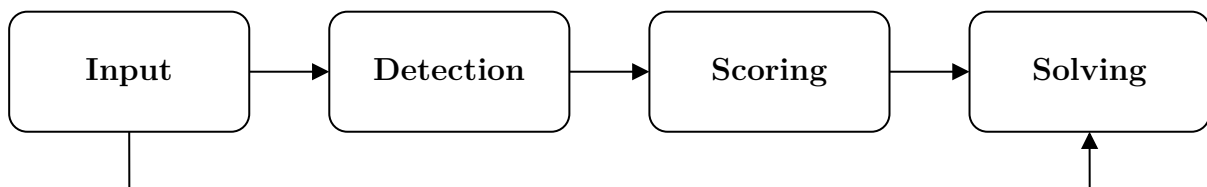


Figure 4.1: A simplified overview of the four major stages of solving a model with GCG.

In this chapter, we introduce Generic Column Generation (GCG), a decomposition solver which is based on the open-source MIP-Solver SCIP (SCIP) [2]. Readers already experienced with GCG and its capabilities may still find some details and observations interesting. For a given problem, GCG is able to perform an automatic Dantzig-Wolfe reformulation which is then solved using a branch-price-and-cut algorithm. Alternatively, GCG support a special *Benders-Mode* which reformulated the problem using Benders decomposition.

In contrast to other open-source solvers like BaPCod (Branch-and-price Code) [3] or commercial software such as *SAS* [4] which rely solely on user-provided decompositions, GCG is able to automatically detect different kinds of structures algorithmically, including but not limited to

- Single-Bordered structures
- Arrowhead structures using the third-party tool hMETIS (Hypergraph METIS) [5].
- Staircase structures

The solving process is divided into multiple consecutive stages as shown in Figure 4.1. Each stage will be explained in more detail in the following section as needed. The detection in particular aims to make GCG more accessible to a wider range of users which do not necessarily have the required theoretical background and practical experience to reformulate linear programs on their own. For more details about individual features and capabilities, we refer to the official documentation [6].

Entfernen
und auf
Kapitel
vorher
ref

Kurz
die 4
Schritte
aus Bild
erwäh-
nen und
einen
Satz

4.1 Detection

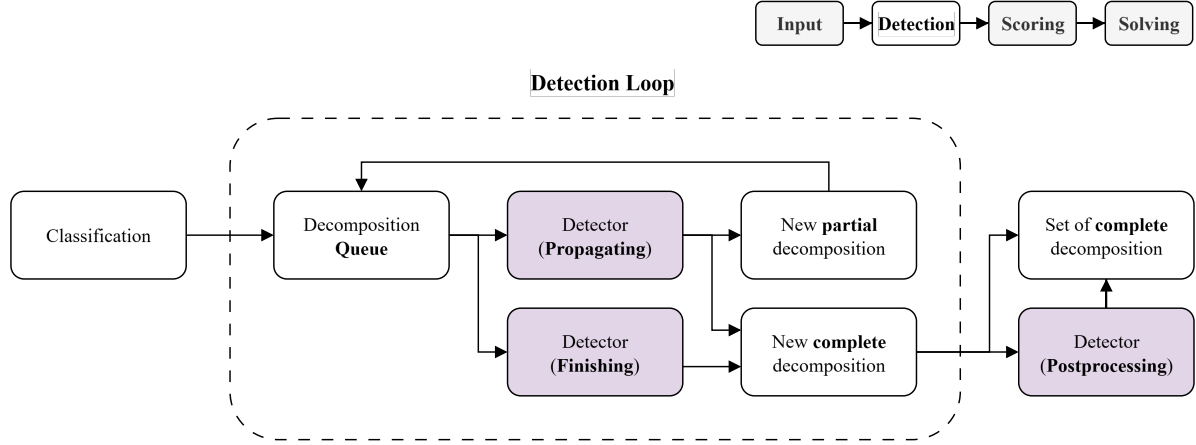


Figure 4.2: A simplified overview of the detection process and its detection loop.

As mentioned in the introduction to this chapter, one integral part and distinguishing feature of GCG is its detection framework. A simplified overview of the detection currently ¹ implemented in GCG is shown in Figure 4.2. For a more detailed visualization including additional information about how pre-solving is handled we refer to the official documentation [6]. The framework consists of two major parts:

1. A **classification** step, in which a set of classifiers is partitioning the constraints (and variables) according to a certain property, producing one partition each. The goal of this step is to detect different underlying structures of the constraint matrix, which can be used during the detection loop to make more informed decisions about which constraints to assign to which block or master. Important classifiers for the remainder of this thesis are discussed in more detail in Section 4.2.
2. The **detection loop**, which consists of a set of detectors which are responsible for assigning constraints either the master or to individual blocks. In round $n + 1$ a detector receives a *partial* decomposition, that is, a decomposition in which *not all* constraints are assigned yet, from round n as input and pushes a set of newly created (partial) decompositions to a queue. In case the user did not provide a partial decomposition as input in round 0, the loop is initialized with a decomposition in which no constraint is assigned yet.

¹GCG version 3.5, as of 2025-07-18.

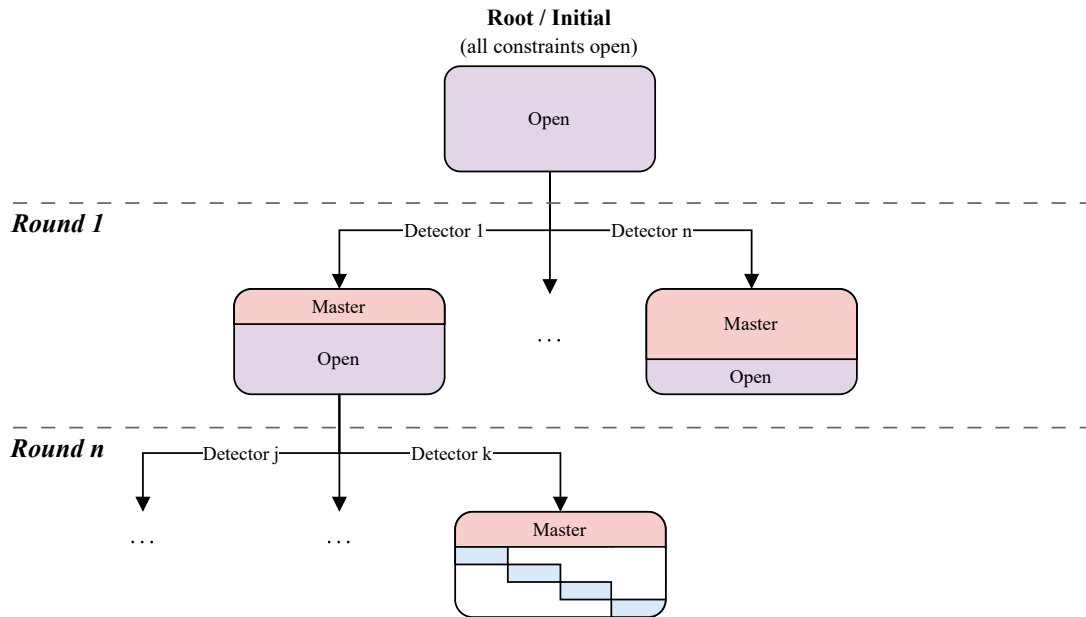


Figure 4.3: Visualization of the induced tree of propagated partial decompositions.

The concept of detecting structures in different rounds is visualized in Figure 4.3. Starting from a root decomposition in which all constraints are still unassigned or “open”, different detectors produce a set of new partial decomposition. Depending on the configuration, a detector is not allowed to work on a certain partial decomposition or its decedents twice. A very simple but concrete example of how such a tree might look like in practice can be found in Section 4.4.

Furthermore, if no detector found any new decomposition in round k , or k exceed the maximum number of rounds, the detection loop is stopped and all complete decomposition are collected, scored and exactly one is chosen for which the solving is started. The scoring and selection stage is of particular interest in practice, because the tree in Figure 4.3 might grow beyond thousand of decompositions, of which the best in terms of solving time or a different metric must be selected. Because the scoring of decompositions is not of major interest for *this* thesis, we refer to the official documentation for details [6].

Grammatik
Wort-
wahl

4.2 Classifiers

4.2.1 Name Classifiers

4.2.2 Numeric Classifiers

In contrast to a classifier based on properties the lassifiers are solely based on

4.2.3 Type Classifiers

Classifiers based

SCIP Types

When using GCG as a library, the type of a variable or constraint can be retrieved via. `SCIPconsGetType(cons)` or `SCIPvarGetType(cons)` respectively. The former function is not provided by SCIP itself, but is implemented in GCG instead. The implementation compares the name of the handler the constraint is assigned to and compares it to a known list of constraint handlers. The list of supported handlers includes *Knapsack*, *Set Partitioning*, *Set Covering*, *Set Packing*, *Varbound* and *General*, in case no special structure was detected. Variables can be classified as *Integer*, *Binary* or *Continuous* ².

the

Check
List

The clear downside of this classification is its important precondition. In order to use this feature properly and retrieve a meaningful type via. the two methods, pre-solving must have been executed prior to detection. When GCG reads the problem as e.g. an `.lp` file, all constraints are added as linear constraints to the underlying SCIP model. These constraints are usually “upgraded” if possible, that is, their structure is analyzed and assigned to the correct constraint handler during pre-solving. This is done to take advantage of properties only possessed by certain types of constraints, e.g. a solution to a set of Knapsack constraints *can* be computed more efficiently by using an algorithm based on dynamic programming. For more detailed information we refer to the official documentation [7].

Preliminary testing showed that it is not trivial to configure the pre-processing in such a way that *only* the upgrade mechanism is triggered and variables and constraints remain unchanged.

Add test
config to
appendix

²There are more types of variables in newer versions of SCIP such as *Implicit Integer*, but these three basic types are sufficient for the purpose of this discussion.

MIPLIB Constraint Types

Nr.	Type	Linear Constraint	Notes
1	Empty	\emptyset	-
2	Free	$-\infty \leq x \leq \infty$	No finite side.
3	Singleton	$a \leq x \leq b$	-
4	Aggregation	$ax + by = c$	-
5	Precedence	$ax - ay \leq b$	x, y have same type.
6	Variable Bound	$ax + by \leq c$	$x \in \{0, 1\}$
7	Set Partitioning	$\sum 1x_i = 1$	$\forall i : x_i \in \{0, 1\}$
8	Set Packing	$\sum 1x_i \leq 1$	$\forall i : x_i \in \{0, 1\}$
9	Set Covering	$\sum 1x_i \geq 1$	$\forall i : x_i \in \{0, 1\}$
10	Cardinality	$\sum 1x_i = b$	$\forall i : x_i \in \{0, 1\}, b \in \mathbb{N}_{\geq 2}$
11	Invariant Knapsack	$\sum 1x_i \leq b$	$\forall i : x_i \in \{0, 1\}, b \in \mathbb{N}_{\geq 2}$
12	Equation Knapsack	$\sum a_i x_i = 1$	$\forall i : x_i \in \{0, 1\}, b \in \mathbb{N}_{\geq 2}$
13	Bin Packing	$\sum a_i x_i + ay \leq a$	$\forall i : x_i, y \in \{0, 1\}, b \in \mathbb{N}_{\geq 2}$
14	Knapsack	$\sum a_i x_i \leq b$	$\forall i : x_i \in \{0, 1\}, b \in \mathbb{N}_{\geq 2}$
15	Integer Knapsack	$\sum a_i x_i \leq b$	$\forall i : x_i \in \mathbb{Z}, b \in \mathbb{N}$
16	Mixed Binary	$\sum a_i x_i + \sum p_j s_j \{ \leq, = \} b$	$\forall i : x_i \in \{0, 1\}, \forall j : s_j \text{ continuous}$
17	General Linear	$\sum a_i x_i \{ \leq, \geq, = \} b$	No special structure.

Table 4.1: The structure of all 17 constraint types MIPLIB keeps track of.

In order to get a more fine-grained classification based on inferred constraint types,

4.2.4 Detectors

4.3 Existing Detectors

4.4 Example

$$\begin{array}{ll}
 \min & \sum_{j=1}^m y_j \\
 \text{s.t.} & \sum_{j=1}^m x_{ij} = 1 \quad \forall i \in \mathcal{I} \\
 & \sum_{i=1}^n a_i x_{ij} \leq C y_j \quad \forall j \in \mathcal{J} \\
 & x_{ij} \in \{0, 1\} \quad \forall i \in \mathcal{I}, \forall j \in \mathcal{J} \\
 & y_j \in \{0, 1\} \quad \forall j \in \mathcal{J}
 \end{array}$$

Figure 4.4: Bin-Packing Model with items $\mathcal{I} = \{1, \dots, n\}$, item sizes $a_i \in \mathbb{Z}_{\geq 0}$, bins $\mathcal{J} = \{1, \dots, m\}$ and capacity C .

In order to illustrate the detection

5 Implementation

5.1 Architecture

5.2 Pre-Processing

5.3 Refinement Process

5.4 Post-Processing

5.5 Scoring

6 Evaluation

6.1 Setup

Type	Name	Metric
CPU	AMD Ryzen 3700X	3.8 GHz
RAM	-	16GB

Table 6.1: Consumer-grade components used to run all experiments.

All experiences were run on a system with components as specified in Table 6.1.

6.2 StrIPLib

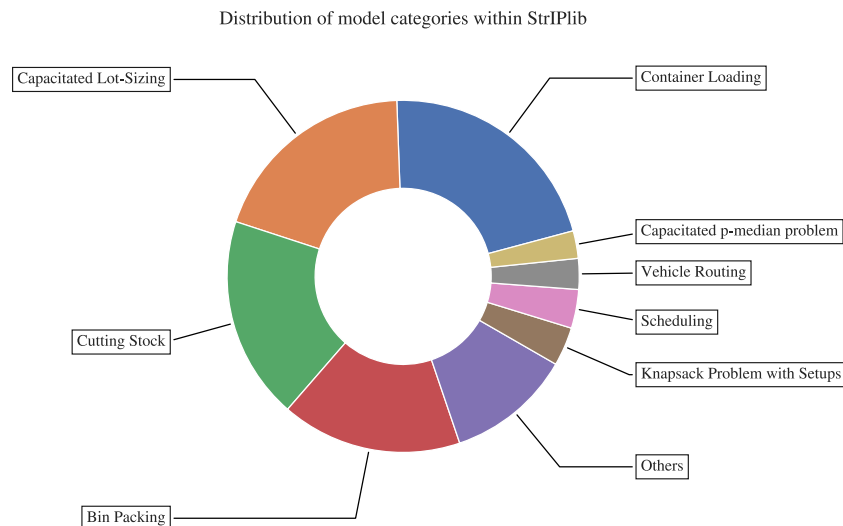


Figure 6.1: The distribution of different categories of model within strIPLib. Most problems are part of “common” categories like Bin-Packing, Scheduling and Routing. The category “Others” includes e.g. Fantasy Football, Train Scheduling and different types for which only a small number of model files are available.

The Structured Integer Program Library (strIPLib) is a collection of over 21000 mixed-integer programs with an exploitable structure such as Block-Diagonal and Staircase . All instances are assign to exactly one of the 33 main categories as highlighted in Figure 6.1. Each categories is further sub-divided into a number of smaller sub-categories, because each kind of problem can be modeled (e.g. three-index vs. four-index) *or* decomposed in a variety of different ways.

The number of available instances per category ranges from as low as 2 for Binary/Ternary Code Construction and up to ≈ 4700 for Container Loading, which makes the data-set in-balanced with respect to available models per main category. This is only of theoretical concern and is further discussed in section . Furthermore, the largest four categories account for $\approx 80\%$ of the total instance count with the remaining 20% distributed across 29 categories. One singular category “Haplotype Inference” with 40 instances is excluded from all tests, because the problem files are not readable by either GCG or SCIP. This behavior can be traced back to the used variable names in these models, which all contain the special character “^”.

ref

ref

Models
ohne
Namen
noch
ergänzen
(Prob-
lem-
beschrei

6.3 Stuff

7 Notes

7.1 MIPLIB Constraint Types

7.2 Relaxed Constraint Types

Nr.	Type	Linear Constraint	Notes
1	Empty	\emptyset	-
2	Free	$-\infty \leq x \leq \infty$	No finite side.
6	Variable Bound	$ax + by \leq c$	$x \in \{0, 1\}$
7	Set Partitioning	$\sum 1x_i = 1$	$\forall i : x_i \in \{0, 1\}$
8	Set Packing	$\sum 1x_i \leq 1$	$\forall i : x_i \in \{0, 1\}$
9	Set Covering	$\sum 1x_i \geq 1$	$\forall i : x_i \in \{0, 1\}$
10	Cardinality	$\sum 1x_i = b$	$\forall i : x_i \in \{0, 1\}, b \in \mathbb{N}_{\geq 2}$
10	Cardinality	$\sum 1x_i = b$	$\forall i : x_i \in \{0, 1\}, b \in \mathbb{N}_{\geq 2}$
11	Invariant Knapsack	$\sum 1x_i \leq b$	$\forall i : x_i \in \{0, 1\}, b \in \mathbb{N}_{\geq 2}$
12	Equation Knapsack	$\sum a_i x_i = 1$	$\forall i : x_i \in \{0, 1\}, b \in \mathbb{N}_{\geq 2}$
13	Bin Packing	$\sum a_i x_i + ay \leq a$	$\forall i : x_i, y \in \{0, 1\}, b \in \mathbb{N}_{\geq 2}$
14	Knapsack	$\sum a_i x_i \leq b$	$\forall i : x_i \in \{0, 1\}, b \in \mathbb{N}_{\geq 2}$
15	Integer Knapsack	$\sum a_i x_i \leq b$	$\forall i : x_i \in \mathbb{Z}, b \in \mathbb{N}$
16	Mixed Binary	$\sum a_i x_i + \sum p_j s_j \{ \leq, = \} b$	$\forall i : x_i \in \{0, 1\}, \forall j : s_j \text{ continuous}$
17	Mixed	$\sum a_i x_i \{ \leq, \geq \} b$	No special structure.
17	Mixed Equation	$\sum a_i x_i \{ = \} b$	No special structure.

Table 7.1: A relaxed version of the constraint types MIPLIB uses.

7.3 Classifiers, Detectors

Classifiers:

- VarTypes, ConsTypes (SCIP)
- MIPLIB Cons
- Name based / Levenstein
- Non-Zeroes
- Objective Values / Sign

Detectors:

- Cons Class, Var Class
- Connected Base
- HMETIS
- Dense Master Cons
- Detectors for single constraint types
- Staircase Heur
- Greedy
- Post Process

Literaturverzeichnis

- [1] Cover, T. M./ Thomas, J. A. *Elements of Information Theory*. 2nd ed. Hoboken, N.J: Wiley-Interscience, 2006.
- [2] “Experiments with a Generic Dantzig-Wolfe Decomposition for Integer Programs”. In: Gamrath, G./ Lübbecke, M. E. *Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 239–252. URL: http://link.springer.com/10.1007/978-3-642-13193-6_21 (visited on 07/08/2025).
- [3] Sadykov, R./ Vanderbeck, F. *BaPCod - a Generic Branch-and-Price Code*. Technical Report. Inria Bordeaux Sud-Ouest, 11/2021. URL: <https://inria.hal.science/hal-03340548>.
- [4] *SAS: Data and AI Solutions*. URL: https://www.sas.com/en_us/home.html (visited on 07/08/2025).
- [5] Karypis, G. et al. “Multilevel Hypergraph Partitioning: Application in VLSI Domain”. In: *Proceedings of the 34th Annual Conference on Design Automation Conference - DAC '97*. The 34th Annual Conference. Anaheim, California, United States: ACM Press, 1997, pp. 526–529. URL: <http://portal.acm.org/citation.cfm?doid=266021.266273> (visited on 07/08/2025).
- [6] *GCG*. URL: <https://gcg.or.rwth-aachen.de/> (visited on 07/08/2025).
- [7] *SCIP Doxygen Documentation: Overview*. URL: <https://www.scipopt.org/doc/html/> (visited on 07/08/2025).