Combinatorial and Approximation Algorithms

Lecture Notes, Spring Term 2019

University of Zurich

Dr. Alexander Souza

Contents

Chapter 1

Introduction

1.1 Examples

We start with some examples of combinatorial optimization problems.

Example 1.1. The following problem is called the KNAPSACK problem. We are given an amount of C Euro and wish to invest it among a set of n options. Each such option i has cost c_i and profit p_i. The goal is to maximize the total profit.

Consider $C = 100$ and the following cost-profit table:

Option	Cost	Profit
1	100	150
2	1	2
3	50	55
4	50	100

Our choice of purchased options must not exceed our capital C. Thus the feasible solutions are $\{\}, \{1\}, \{2\}, \{3\}, \{4\}, \{2,3\}, \{2,4\}, \{3,4\}$. Which is the best solution? We evaluate all possibilities and find that $\{3,4\}$ give 155 altogether which maximizes our profit.

Example 1.2. Another example is a LOAD BALANCING problem: We have m machines and we have a set of n jobs that need to be done. Each job j has a processing time $p_{i,j}$ if executed by machine i. We can formulate our problem with the following mathematical program. We use the variables $x_{i,j} \in \{0,1\}$ that indicate if job j is assigned to machine i. We want to minimize the time until all jobs are finished.

$$
\begin{array}{lll}
\text{minimize} & f, & \text{“minimize finishing time } f\text{”} \\
\text{subject to} & \sum_{j=1}^{n} p_{i,j} x_{i,j} \leq f, \quad i = 1, \ldots, m & \text{“}f \text{ is largest machine time”} \\
& \sum_{i=1}^{m} x_{i,j} = 1, \qquad j = 1, \ldots, n & \text{“each job gets done”} \\
& x_{i,j} \in \{0,1\}, \qquad i = 1, \ldots, m,\ j = 1, \ldots .n & \text{“assignment”}
\end{array}
$$

1.2 Combinatorial Optimization Problems

An instance I of a *combinatorial optimization problem* (COP) can formally be defined as a tuple $I = (U, P, \text{value}, \text{extr})$ with the following meaning:

U	the *solution space* of possible outputs,
P	the *feasibility predicate*,
value	the *value function* $\text{value} : U \to \mathbb{R}$,
extr	the desired *extremum*,i.e., max or min.

extr and value together define the *objective function*. The feasibility predicate P induces a set:

$$S \quad \text{the set of } \textit{feasible solutions}\text{: } S = \{X \in U : X \text{ satisfies } P\}.$$

Our goal is to find a feasible solution where the desired extremum of value is attained. Any such solution is called an *optimum solution*, or simply an *optimum*. U and S are usually not given explicitly, but implicitly.

A central problem around combinatorial optimization is that it is often in principle possible to find an optimum solution by enumerating the set of feasible solutions, but this set mostly contains "too many" elements. This phenomenon is called *combinatorial explosion*.

Example 1.3. Let us investigate the problem in Example 1.1 in with this formalism.

$$\begin{aligned} U &= 2^{\{1,2,3,4\}}, \\ P &= \text{“total cost is at most } C\text{”, i.e., } X \in S \text{ if } \sum_{i \in X} c_i \leq C \\ S &= \{\{\}, \{1\}, \{2\}, \{3\}, \{4\}, \{2,3\}, \{2,4\}, \{3,4\}\}, \\ \text{value} &= \begin{cases} U \to \mathbb{R} \\ X \mapsto \sum_{i \in X} p_i, \end{cases} \\ \text{extr} &= \max. \end{aligned}$$

The optimum solution here is $\{3, 4\}$ with value 155.

1.3 Algorithms and Approximation

Many problems in combinatorial optimization can be solved by using an appropriate algorithm. Informally, an *algorithm* is given a (valid) input, i.e., a description of an instance of a problem and computes a solution after a finite number of "elementary steps". The number of bits used to describe an input I is called the (binary) *length* or *size* of the input and denoted $\text{size}(I)$.

Let $t : \mathbb{N} \to \mathbb{R}$ be a function. We say that an algorithm *runs* in time $O(t)$ if there is a constant α such that the algorithm uses at most $\alpha t(\text{size}(I))$ many elementary steps to compute a solution given any input I. An algorithm is called *polynomoial time* if $t : n \mapsto n^c$ for some constant c. This contrasts *exponential time* algorithms where $t : n \mapsto c^n$ for some constant $c > 1$.

Because the running times of exponential time algorithms grow rather rapidly as the input size grows, we are mostly interested in polynomial time algorithms. Of course, we desire to find an optimum solution for any given COP in polynomial time. Unfortunately

this is not always possible as many COPs are NP-hard. (It is widely believed that no polynomial time algorithm exists that solves some NP-hard COP optimally on every instance.) Thus our goal is a find "good" solutions in polynomial time.

Let I be some instance and let $S(I)$ be the set of feasible solutions of I. Let $\text{OPT}(I) = \text{extr}_{X \in S(I)} \text{value}(X)$ denote the respective optimum value of instance I. An *approximation algorithm* ALG is a polynomial time algorithm that computes some solution $X \in S(I)$ for any given instance I. The respective value obtained is denoted $\text{ALG}(I) = \text{value}(X)$. The *approximation ratio* of ALG on an instance I is defined by

$$\rho_{\text{ALG}}(I) = \frac{\text{ALG}(I)}{\text{OPT}(I)}.$$

The algorithm ALG is a *ρ-approximation* algorithm if

$$\begin{aligned} \rho_{\text{ALG}}(I) &\leq \rho \qquad \text{for any instance } I \text{ and extr} = \min, \\ \rho_{\text{ALG}}(I) &\geq \rho \qquad \text{for any instance } I \text{ and extr} = \max. \end{aligned}$$

1.4 Linear and Integer Linear Programs

A Linear Program (LP) is given by linear constraints and a linear objective function. More precisely:

Let $A = (a_{i,j})_{i=1,\ldots,m, j=1,\ldots,n} \in \mathbb{R}^{m,n}$ be a matrix and let $b = (b_i)_{i=1,\ldots,m} \in \mathbb{R}^m$ and $c = (c_j)_{j=1,\ldots,n} \in \mathbb{R}^n$ be vectors. Further let $x = (x_j)_{j=1,\ldots,n} \in \mathbb{R}^n$ be real-valued variables. Our objective function is to minimize $c^\top x$ subject to $Ax \leq b$. That is:

$$\begin{aligned} \text{minimize} \quad & \sum_{j=1}^{n} c_j x_j, & & \text{"objective function"} \\ \text{subject to} \quad & \sum_{j=1}^{n} a_{i,j} x_j \leq b_i, & i = 1, \ldots, m \quad & \text{"constraints"} \\ & x_j \in \mathbb{R}, & j = 1, \ldots, n. \quad & \text{"real values"} \end{aligned}$$

LPs can be solved in polynomial time. The Ellipsoid method is one such algorithm, but is only interesting from theoretical perspective. In practice, the Simplex algorithm is frequently used, although it is not polynomial time in the worst case.

An Integer Linear Program (ILP) is a Linear Program where each variable is allowed to take an integer value, only. That is, $x = (x_j)_{j=1,\ldots,n} \in \mathbb{Z}^n$ is an integer-vector.

$$\begin{aligned} \text{minimize} \quad & \sum_{j=1}^{n} c_j x_j, & & \text{"objective function"} \\ \text{subject to} \quad & \sum_{j=1}^{n} a_{i,j} x_j \leq b_i, & i = 1, \ldots, m \quad & \text{"constraints"} \\ & x_j \in \mathbb{Z}, & j = 1, \ldots, n. \quad & \text{"integer values"} \end{aligned}$$

Many COPs can be formulated in terms of an ILP, but solving an ILP is in general NP-hard. However, we will often replace the constraints $x_j \in \mathbb{Z}$ with $x_j \in \mathbb{R}$. This is then called an *LP-relaxation* of an ILP. Of course, an LP solution is in general not feasible for the ILP, but we can sometimes "turn" it into a feasible solution, which is not "too bad".

1.5 Randomized Algorithms

Generally speaking, *randomized algorithms* have access to (arbitrarily many) random bits, and are allowed to decide based on their outcomes. This, of course, yields that either the output and/or the running time of the algorithm are random variables. As a consequence we can not expect that the algorithm behaves exactly the same, when given the same input. (Notice that deterministic algorithms do have this property.) Why would one want to allow such indefiniteness? There are several reasons: For example, approximation algorithms are sometimes "fooled" by rather artifical counterexamples that rely on the specific strategy of the algorithm. In that perspective, randomizing strategies often yield improvements in approximation guarantee (in expectation). Furthermore, random choices sometimes yield significant speed-up of running time (in expectation), compared to worst-case running times.

There are two types of randomized algorithms: *Las Vegas* algorithms are allowed to use the random bits, but must always return correct answers. In contrast, *Monte Carlo* algorithms are allowed to return false answers, but this must not happen with "too large" probability.

If it is not desired to have a randomized algorithm, respectively a randomized construction method, then one can also try to *derandomize* a randomized algorithm. As the name suggests, this refers to the process of turning a randomized algorithm into a deterministic one. This is often achieved at the price of higher running times and/or deterioration of solution quality.

Part I

Combinatorial Algorithms

Chapter 2

Greedy Algorithms

Algorithms of the GREEDY type have in common that they construct solutions in an iterative manner and based on "locally optimal" criteria. Sometimes this strategy even yields a globally optimal solution, but mostly the final solution is non-optimal. In this chapter we give an example for optimal GREEDY algorithms (for the MINIMUM SPANNING TREE problem) and an example of a non-optimal one (for the NP-hard SET COVER problem).

2.1 Minimum Spanning Trees

Consider a telecommunications company that wants to rent a subset of an existing set of cables in a network, each of which connects two cities. The rented cables should suffice to connect all the cities and they should be as cheap as possible. This type of applications is formalized as follows:

Let G be an undirected, connected graph on the vertices $V(G) = V$ and the edges $E(G) = E$. Furthermore, let $c : E \to \mathbb{R}$ be a cost function. A graph F is a *forest* if F has no cycles. A forest F with $V(F) = V(G)$ and $E(F) \subseteq E(G)$ is a *spanning forest*. A connected forest is a *tree* and likewise a connected spanning forest is a *spanning tree*. For any vertex set C, called a *cut*, let $\delta(C) = \{e = uv \in E : u \in C, v \notin C\}$ denote the *cut-edges* over C.

Problem 2.1 MINIMUM SPANNING TREE

Instance. Undirected, connected graph G on the vertices $V(G)$ and the edges $E(G)$, cost function $c : E(G) \to \mathbb{R}$.

Task. Find a spanning tree T with cost value$(T) = \sum_{e \in E(T)} c(e)$ minimal.

See Figure 2.1 for an example. Maybe the first idea that comes to mind is to sort the edges according to non-decreasing cost and to add the edges one-by-one provided that each addition does not close a cycle. This is the invariant of the algorithm KRUSKAL.

Another idea is to maintain a minimum spanning tree of subsets of the vertex set of the graph and to extend the tree until it is a spanning tree of the graph. This is the idea behind the algorithm PRIM.

Now we will establish several structural properties of any optimal, i.e., minimum cost, spanning tree and derive the correctness of the two algorithms from those.

Theorem 2.1. *Let G be an undirected graph and $c : E \to \mathbb{R}$ be a cost function and let T be a spanning tree. Then we have*

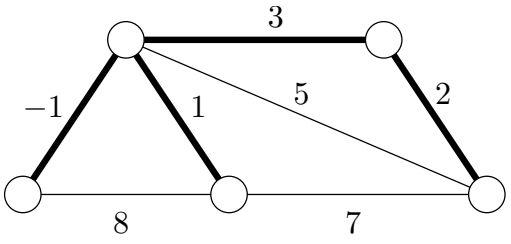


Figure 2.1: An example of an undirected, connected graph with a cost function on the edges. The bold edges indicate a minimum spanning tree in that graph.

Algorithm 2.1 KRUSKAL

Input. Undirected connected graph $G = (V, E)$, cost function $c : E \to \mathbb{R}$.

Output. Spanning tree T with minimum cost.

Step 1. Sort the edges such that $c(e_1) \leq \cdots \leq c(e_m)$ and let $T = (V, \emptyset)$.

Step 2. For $i = 1, \ldots, m$ do: If $T + e_i$ is a forest then $T = T + e_i$.

---

(1) T is a minimum spanning tree.

(2) For any $e = uv \in E(G) - E(T)$, no edge on the u-v-path in T has larger cost than e.

(3) For every $e \in E(T)$, e is a minimum cost edge over $\delta(C)$, where C is a connected component of $T - e$.

(4) We can order $E(T) = \{e_1, \ldots, e_{n-1}\}$ such that for each $i = 1, \ldots, n-1$, there is a set $X \subseteq V(G)$ such that e_i is a minimum cost edge over $\delta(X)$ and $e_j \notin \delta(X)$ for all $j = 1, \ldots, i-1$.

Proof. (1) $\Rightarrow$ (2): Suppose (2) is violated. Let $e = uv \in E(G) - E(T)$ and let e' be an edge on the u-v-path in T with $c(e') > c(e)$. Then $T' = T - e' + e$ is a spanning tree with lower cost than T.

(2) $\Rightarrow$ (3): Suppose (3) is violated. Let $e \in E(T)$, C a connected component of $T - e$ and $e' = uv \in \delta(C)$ with $c(e') < c(e)$. Observe that the u-v-path in T must contain an edge of $\delta(C)$. But the only such edge is e. So (2) is violated.

(3) $\Rightarrow$ (4): Take an arbitrary order and $X = V(C)$.

(4) $\Rightarrow$ (1): Suppose $E(T) = \{e_1, \ldots, e_{n-1}\}$ satisfies (4) and let T^* be an optimimum spanning tree such that $i = \inf\{h \in \{1, \ldots, n-1\} : e_h \notin E(T^*)\}$ is maximum. We show that $i = \infty$, i.e., $T = T^*$. Suppose not, then let $X \subseteq V(G)$ such that e_i is a minimum cost edge of $\delta(X)$ and $e_j \notin \delta(X)$ for all $j = 1, \ldots, i-1$. $T^* + e_i$ contains a cycle C. Since $e_i \in E(C) \cap \delta(X)$, at least one more edge e' (with $e' \neq e_i$) of C must belong to $\delta(X)$. Observe that $T + e_i - e'$ is a spanning tree. Since T^* is optimum $c(e_i) > c(e')$. But since $e' \in \delta(X)$, we also have $c(e') > c(e_i)$. Moreover, if $e' = e_j \in E(T)$, then $j > i$. So $c(e') = c(e_i)$ and $T^* + e_i - e'$ is another optimum spanning tree, which contradicts the maximality of i. □

Corollary 2.2. *The algorithm* KRUSKAL *computes a minimum spanning tree. It can be implemented to run in time $O(m \log m)$.*

Algorithm 2.2 PRIM

Input. Undirected connected graph $G = (V, E)$, cost function $c : E \to \mathbb{R}$.

Output. Spanning tree T with minimum cost.

Step 1. Choose $v \in V$ and let $T = (\{v\}, \emptyset)$.

Step 2. While $V(T) \neq V(G)$ do: Get minimum cost $e \in \delta(V(T))$ and let $T = T + e$.

Proof. By connectivity of G and since KRUSKAL maintains a spanning forest, it is clear that the algorithm computes a spanning tree T at termination. As it guarantees condition (2), T is optimal.

Sorting the edges requires time $O(m \log m)$. In order to decide if an edge closes a cycle in the current forest, we maintain a UNION-FIND datastructure, e.g., a BOTTOM-UP FOREST. This datastructure supports FIND operations $O(\alpha(m))$ time and UNION operations in $O(1)$ time. Here, α denotes the inverse of the Ackermann function, which grows strictly slower than log. Since there are m FIND operations and $n - 1$ UNION operations we have runningtime $O(m \log m + \alpha(m)m + n) = O(m \log m)$ in total. □

Corollary 2.3. *The algorithm* PRIM *computes a minimum spanning tree. It can be implemented to run in time* $O(m + n \log n)$.

Proof. The correctness follows from the fact that the algorithm ensures property (4).

To yield the claimed running time, it is useful to maintain a FIBONACCI HEAP priority queue. This enables EXTRACT-MIN in $O(\log n)$ amortized time and INSERT in $O(1)$ time. Since there are m INSERT and n EXTRACT-MIN operations, the claimed running time follows. □

2.2 Set Cover

The SET COVER problem this section deals with is a very simple to state – yet quite general – NP-hard combinatorial problem. It is widely applicable in sometimes unexpected ways. The problem is the following: We are given a set U (called *universe*) of n elements, a collection of sets $\mathcal{S} = \{S_1, \ldots, S_k\}$ where $S_i \subseteq U$, and a cost function $c : \mathcal{S} \to \mathbb{R}^+$. The task is to find a minimum cost subcollection $\mathcal{S}' \subseteq \mathcal{S}$ that *covers* U, i.e., such that $\cup_{S \in \mathcal{S}'} S = U$.

Example 2.4. Consider this instance: $U = \{1, 2, 3\}$, $\mathcal{S} = \{S_1, S_2, S_3\}$ with $S_1 = \{1, 2\}$, $S_2 = \{2, 3\}$, $S_3 = \{1, 2, 3\}$ and cost $c(S_1) = 10$, $c(S_2) = 50$, and $c(S_3) = 100$. These collections cover U: $\{S_1, S_2\}$, $\{S_3\}$, $\{S_1, S_3\}$, $\{S_2, S_3\}$, $\{S_1, S_2, S_3\}$. The cheapest one is $\{S_1, S_2\}$ with cost equal to 60.

For each set S, we associate a variable $x_S \in \{0, 1\}$ that indicates of we want to choose S or not. We may thus write solutions for SET COVER as a vector $x \in \{0, 1\}^k$. With this, we write SET COVER as a mathematical program.

The GREEDY algorithm follows the natural approach of iteratively choosing the most cost-effective set and remove all the covered elements until all elements are covered. Let C be the set of elements already covered at the beginning of an iteration. During this iteration define the *cost-effectiveness* of a set S as $c(S)/|S - C|$, i.e., the average cost at

Problem 2.2 SET COVER

Instance. Universe U with n elements, collection $\mathcal{S} = \{S_1, \ldots, S_k\}$, $S_i \subseteq U$, a cost function $c : \mathcal{S} \to \mathbb{R}$.

Task. Solve the problem

$$\begin{aligned} \text{minimize} \quad & \text{value}(x) = \sum_{S \in \mathcal{S}} c(S) x_S, \\ \text{subject to} \quad & \sum_{S : e \in S} x_S \geq 1 \quad e \in U, \\ & x_S \in \{0, 1\} \quad S \in \mathcal{S}. \end{aligned}$$

---

which it covers new elements. For later reference, the algorithm sets the *price* at which it covered an element equal to the cost-effectiveness of the covering set. Further recall that $H_n = \sum_{i=1}^{n} 1/i$ is called the *n-th Harmonic number* and that $\log n \leq H_n \leq \log n + 1$.

Algorithm 2.3 GREEDY

Input. Universe U with n elements, collection $\mathcal{S} = \{S_1, \ldots, S_k\}$, $S_i \subseteq U$, a cost function $c : \mathcal{S} \to \mathbb{R}$.

Output. Vector $x \in \{0, 1\}^k$

Step 1. $C = \emptyset$, $x = 0$.

Step 2. While $C \neq U$ do the following:

(a) Find the most cost-effective set in the current iteration, say S.

(b) Set $x_S = 1$ and for each $e \in S - C$ set $\text{price}(e) = c(S)/|S - C|$.

(c) $C = C \cup S$.

Step 3. Return x.

Theorem 2.5. *The* GREEDY *algorithm is an H_n-approximation algorithm for the* SET COVER *problem.*

It is an exercise to show that this bound is tight.

The following lemma is crucial for the proof of the approximation-guarantee. Number the elements of U in the order in which they were covered by the algorithm, say $e_1, \ldots, e_n$. Let x^* be an optimum solution.

Lemma 2.6. *For each $i \in \{1, \ldots, n\}$,* $\text{price}(e_i) \leq \text{value}(x^*)/(n - i + 1)$.

Proof. In any iteration, the leftover sets of the optimal solution x^* can cover the remaining elements at a cost of at most $\text{value}(x^*)$. Therefore, among these, there must be one set having cost-effectiveness of at most $\text{value}(x^*)/|U - C|$. In the iteration in which element e_i was covered, $U - C$ contained at least $n - i + 1$ elements. Since e_i was covered by the most cost-effective set in this iteration, we have that

$$\text{price}(e_i) \leq \frac{\text{value}(x^*)}{|U - C|} \leq \frac{\text{value}(x^*)}{n - i + 1}$$

which was claimed. □

Proof of Theorem 8.2. Since the cost of each set is distributed evenly among the new elements covered, the total cost of the set cover picked is

$$\text{value}(x) = \sum_{i=1}^{n} \text{price}(e_i) \leq \text{value}(x^*) H_n,$$

where we have used Lemma 8.3. □

Chapter 3

Network Flows

Flow problems are among the best-understood problems in combinatorial optimization. They are rather important because of their numerous applications.

3.1 Maximum Flows and Minimum Cuts

A *network* is a (simple) digraph $G = (V, A)$ where each edge has a *capacity* $c : A \to \mathbb{R}^+$ and we have two distinguished vertices, the *source* s and the *sink* t. We often write $N = (G, c, s, t)$.

For any vertex v, let $\delta^-(v)$ be the set of *incoming edges* of v, i.e., $\delta^-(v) = \{uv \in A : u \in V\}$ and $\delta^+(v)$ the set of *outgoing* edges of v, i.e., $\delta^+(v) = \{vu \in A : u \in V\}$. Let $f : A \to \mathbb{R}^+$ be any function on the edges. Define the *balance* $\mathrm{bal}_f(v)$ of vertex v with respect to f by

$$\mathrm{bal}_f(v) := \sum_{e \in \delta^+(v)} f(e) - \sum_{e \in \delta^-(v)} f(e).$$

The function f is called *conserving* at a vertex v if $\mathrm{bal}_f(v) = 0$.

The Maximum Flow problem asks to transport as many units from the source to the sink without violating the edge capacities. More precisely, a function $f : A \to \mathbb{R}^+$ is called an *s-t-flow* if:

(1) edge capacities are respected, i.e.,

$$0 \le f(e) \le c(e) \text{ for all } e \in A, \text{ and}$$

(2) f is conserving, i.e.,

$$\mathrm{bal}_f(v) = 0 \text{ for } v \in V - \{s, t\}, \quad \mathrm{bal}_f(s) \ge 0, \quad \text{and} \quad \mathrm{bal}_f(t) \le 0.$$

Its *value* is defined by $\mathrm{value}(f) = \mathrm{bal}_f(s)$. See Figure 3.1.

Problem 3.1 Maximum Flow

Instance. A network $N = (G, c, s, t)$.

Task. Find an $s - t$-flow of maximum value in N.

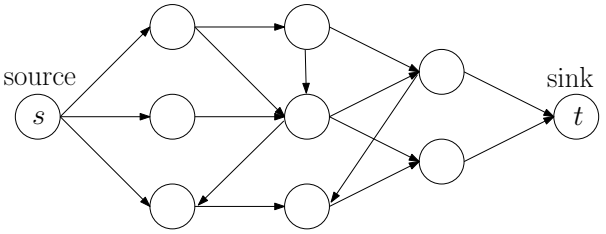


Figure 3.1: A network with source s and sink t.

We can formulate the maximum flow problem as an LP in the variables f_e for $e \in A$.

$$\begin{aligned}
\text{maximize} \quad & \sum_{e \in \delta^+(s)} f_e - \sum_{e \in \delta^-(s)} f_e, \\
\text{subject to} \quad & \sum_{e \in \delta^+(v)} f_e - \sum_{e \in \delta^-(v)} f_e = 0 \quad v \in V - \{s,t\}, \\
& f_e \leq c(e) \quad e \in A, \\
& f_e \geq 0.
\end{aligned}$$

Since the flow $f = 0$ is feasible for this LP, and the LP is obviously bounded (by $\sum_{e \in \delta^+(s)} c(e)$) we have that the MAXIMUM FLOW problem always has an optimum solution. Of course, we can solve the problem by using any algorithm for solving LPs but we are not satisfied with this – we want a combinatorial algorithm (without solving an LP) with guaranteed polynomial running time.

Let S be a subset of the vertices, called a *cut*. The induced *cut-edges* is the set of *outgoing edges* $\delta^+(S) = \{uv \in A : u \in S, v \in V - S\}$ and *incoming edges* $\delta^-(S) = \{vu \in A : u \in S, v \in V - S\}$. Define its *capacity* by $\text{cap}(S) = \sum_{e \in \delta^+(S)} c(e)$. An $s-t$-cut is a cut so that $s \in S$ and $t \in V - S$. A *minimum cut* refers to one with minimal capacity among all $s-t$-cuts. We extend the definition of *balance* also for any cut S:

$$\text{bal}_f(S) = \sum_{e \in \delta^+(S)} f(e) - \sum_{e \in \delta^-(S)} f(e).$$

The following result tells us that the value of a flow can be expressed through the incoming and outcoming flow of an arbitrary cut. Furthermore, the value of any flow (including the maximum one) is bounded from above by the capacity of any cut. We will see soon that the value of a maximum flow equals the capacity of a minimum cut.

Lemma 3.1. *For any $s-t$-cut S and any $s-t$-flow f we have that*

(1) $\text{value}(f) = \text{bal}_f(S)$,

(2) $\text{value}(f) \leq \text{cap}(S)$.

Proof. We use the flow conservation property, i.e., $\mathrm{bal}_f(v) = 0$ for all $v \in S - \{s\}$ to find

$$\begin{aligned}\mathrm{value}(f) = \mathrm{bal}_f(s) &= \sum_{v \in S} \mathrm{bal}_f(v) \\ &= \sum_{v \in S} \left(\sum_{e \in \delta^+(v)} f(e) - \sum_{e \in \delta^-(v)} f(e) \right) \\ &= \sum_{e \in \delta^+(S)} f(e) - \sum_{e \in \delta^-(S)} f(e) \\ &= \mathrm{bal}_f(S).\end{aligned}$$

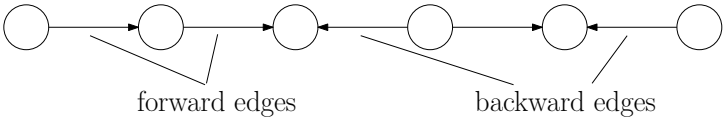
Furthermore we have $\mathrm{value}(f) \leq \sum_{e \in \delta^+(S)} c(e) = \mathrm{cap}(S)$ since $0 \leq f(e) \leq c(e)$. □

The following definitions and structural result are the basis for an algorithm. A *path* is a sequence $P = v_1, e_1, v_2, e_2, \ldots, e_k, v_{k+1}$ alternating between vertices and edges, such that $e_i = v_i v_{i+1} \in A$ or $e_i = v_{i+1} v_i \in A$ for $i = 1, \ldots, k$. The vertices v_1 and v_{k+1} are the *end-vertices* of the path. The number of edges in the path is called its *length*. A path is *simple*, if its vertices are pairwise disjoint. We will always assume that paths are simple, unless stated otherwise. A *v-w-path* P has the form $e_1 = v\cdot$ and $e_\ell = \cdot w$, i.e., it *starts* at v and *ends* at w. An edge $e = vw$ in a path is called *forward edge* if $vw \in A$; *backward edge* if $wv \in A$. (A v-v-path is called a *cycle*.)

An s-v-path P is called *f-augmenting* with respect to a flow f if

(1) $f(e) < c(e)$ for every forward edge $e \in P$,

(2) $f(e) > 0$ for every backward edge $e \in P$.



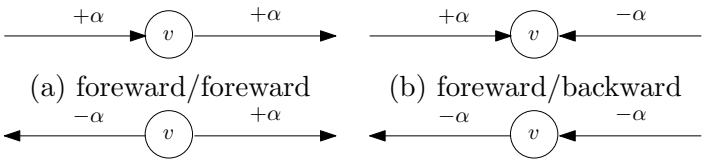
By how much can we increase the current flow value using a particular augmenting path P? Define the quantity

$$\alpha = \min\{c(e) - f(e) : e \text{ foreward edge in } P\} \cup \{f(e) : e \text{ backward edge in } P\}.$$

The following construction of a new flow f' is called *augmenting* f by α along P. Set $f'(e) = f(e) + \alpha$ if e is foreward edge in P, $f'(e) = f(e) - \alpha$ if e is backward edge in P, and $f'(e) = f(e)$ otherwise.

Observation 3.2. *The function f' defines a flow.*

Proof. By definition of the quantity α and because each edge occurs at most once in P, we have that $0 \leq f'(e) \leq c(e)$ for all $e \in A$. It remains to show that f' is flow conserving. It is clear that $\mathrm{bal}_{f'}(s) \geq \mathrm{bal}_f(s) \geq 0$ and consequently $\mathrm{bal}_{f'}(t) \leq \mathrm{bal}_f(t) \leq 0$. Consider an augmentation along edges $e_i e_{i+1}$ with $e_i = v_i v_{i+1}$ and $e_{i+1} = v_{i+1} v_{i+2}$ for $i = 1, \ldots, \ell - 1$. Call $v = v_{i+1}$ and distinguish four cases:



(a) foreward/foreward (b) foreward/backward

(c) backward/foreward (d) backward/backward

This yields the claim. □

Algorithm 3.1 FORD-FULKERSON

Input. Network $N = (G, c, s, t)$ with $c : A \to \mathbb{R}^+$.

Output. $s-t$-flow f of maximum value.

Step 1. Set $f(e) = 0$ for all $e \in A$.

Step 2. Find an f-augmenting path P. If none exists then return f.

Step 3. Compute

$$\alpha = \min\{c(e) - f(e) : e \text{ foreward edge in } P\} \cup \{f(e) : e \text{ backward edge in } P\}.$$

and augment f by α along P. Go to Step 2.

Theorem 3.3. *In a network N, the maximum value of an $s-t$-flow equals the minimum capacity of an $s-t$-cut.*

Proof. We show that an $s-t$-flow f has maximum value if and only if there is no f-augmenting path from s to t. In that case we will be able to find a minimum cut R with equal capacity.

Let there be an f-augementing path P from s to t, let α be as above and obtain f' by augmenting f by α along P. Observe that value(f') > value(f), i.e., that f is not maximal.

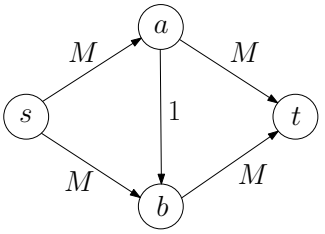
Now let there be no f-augmenting path from s to t. Consider the set S of vertices with augmenting paths from s, i.e., $S = \{v \in V : \text{there is an } f\text{-augmenting path from } s \text{ to } v\}$ and $t \notin S$. Thus S is an $s-t$-cut. By definition of augmenting paths, we must have $f(e) = c(e)$ for all $e \in \delta^+(S)$ and $f(e) = 0$ for all $e \in \delta^-(S)$. Hence, using Lemma 3.1 (1), we have value$(f) = \sum_{e\in\delta^+(S)} c(e) = \text{cap}(S)$. By Lemma 3.1 (2) f must be a maximum flow and S be a minimum cut. □

If all capacities are integers then α is an integer and the algorithm terminates after a finite number of iterations. Thus we obtain the following important consequence:

Corollary 3.4. *If the capacities of a network N are integers, then there is an integral maximum flow.*

If the capacities are not integers, then FORD-FULKERSON might not even terminate. Especially, we have not yet specified how we actually choose the augmenting paths mentioned in Step 2 of the algorithm. This must be done carefully in order to obtain a polynomial time algorithm as the following instance illustrates. It turns out that choosing shortest augmenting paths guarantee termination after a polynomial number of augmentations; see the EDMONDS-KARP algorithm.

Example 3.5. To show that FORD-FULKERSON is not a polynomial time algorithm consider the following network. Here M is a large number.



Alternatingly augmenting one unit of flow along the paths s-a-b-t and s-b-a-t requires $2M$ augmentations. This is already exponential because the (binary) input size of the graph is $O(\log M)$. In contrast the augmenting paths s-a-t and s-b-t already give a maximum flow after two augmentations.

Edmonds-Karp Algorithm

Example 3.5 suggests that it may be a good idea to always choose shortest augmenting paths, i.e., with minimum number edges. Indeed, the algorithm EDMONDS-KARP below uses this strategy and yields polynomial running time.

Algorithm 3.2 EDMONDS-KARP

Input. Network $N = (G, c, s, t)$ with $c : A \to \mathbb{R}^+$.

Output. $s-t$-flow f of maximum value.

Step 1. Set $f(e) = 0$ for all $e \in A$.

Step 2. Find a shortest f-augmenting path P w.r.t. the number of edges. If none exists then return f.

Step 3. Compute α as above and augment f by α along P. Go to Step 2.

Theorem 3.6. *The algorithm* EDMONDS-KARP *computes a maximum $s-t$-flow f in any network N with n vertices and m edges in time $O\left(nm^2\right)$.*

The following lemma is crucial for the proof of the worst-case running time. Let $f_0, f_1, f_2, \ldots$ be the flows constructed by the algorithm. Denote the shortest length of an augmenting path from s to a vertex v with respect to f_k by $x_v(k)$ and respectively from v to t by $y_v(k)$.

Lemma 3.7. *We have that*

(1) $x_v(k+1) \geq x_v(k)$ for all k and v,

(2) $y_v(k+1) \geq y_v(k)$ for all k and v.

Proof. Suppose for the sake of contradiction that (1) is violated for some pair (v, k). We may assume that $x_v(k+1)$ is minimal among the $x_w(k+1)$ for which (1) does not hold.

Let e be the last edge in a shortest augmenting path from s to v with respect to f_{k+1}. Suppose $e = uv$ is a forward edge. Hence $f_{k+1}(e) < c(e)$, $x_v(k+1) = x_u(k+1) + 1$,and $x_u(k+1) \geq x_u(k)$ by our choice of $x_v(k+1)$. Thus $x_v(k+1) \geq x_u(k) + 1$. Suppose that $f_k(e) < c(e)$ which yields $x_v(k) \leq x_u(k) + 1$ and thus $x_v(k+1) \geq x_v(k)$, a contradiction.

Hence we must have $f_k(e) = c(e)$ which implies that e was a backward edge when f_k was changed to f_{k+1}. As we used an augmenting path of shortest length we have $x_u(k) = x_v(k) + 1$ and thus $x_v(k+1) - 1 = x_u(k+1) \geq x_u(k) \geq x_v(k) + 1$. Hence $x_v(k+1) \geq x_v(k) + 2$ yields a contradiction.

Similarly when e is a backward edge. The proof of (2) is analogous to (1). □

Proof of Theorem 3.6. When we increase the flow, the augmenting path always contains a *critical* edge, i.e., an edge where the flow is either increased to meet the capacity or reduced to zero.

Let $e = uv$ be critical in the augmenting path w.r.t. f_k. This path has $x_v(k) + y_v(k) = x_u(k) + y_u(k)$ edges. If e is used the next time in an augmenting path w.r.t. f_h, say, then it must be used in the opposite direction as w.r.t. f_k.

Suppose that $e = uv$ was a forward edge w.r.t. f_k. Then $x_v(k) = x_u(k) + 1$ and $x_u(h) = x_v(h)+1$. By Lemma 3.7 $x_v(h) \geq x_v(k)$ and $y_u(h) \geq y_u(k)$. Hence $x_u(h)+y_u(h) = x_v(h) + 1 + y_u(h) \geq x_v(k) + 1 + y_u(k) \geq x_u(k) + y_u(k) + 2$. Thus the augmenting path w.r.t. f_h is at least two edges longer than the augmenting path w.r.t. f_k. Similarly if e is a backward edge.

No shortest augmenting path can contain more than $n-1$ edges and hence each edge can be critical at most $(n-1)/2$ times. As each augmenting path contains at least one critical edge, there can be at most $O(nm)$ augmentations and each one takes time $O(m)$. This yields the running time of $O(nm^2)$. □

There are further algorithms that solve the Maximum Flow problem in less time. For example the Goldberg-Tarjan algorithm runs in time $O\left(n^2\sqrt{m}\right)$; with sophisticated implementations $O\left(nm\log(n^2/m)\right)$ and $O\left(\min\{m^{1/2}, n^{2/3}\}m\log(n^2/m)\log c_{\max}\right)$ can be reached.

3.2 Minimum Cost Flows

In this section we treat a more general problem than the Maximum Flow problem, namely the Minimum Cost Flow problem. We are again given a digraph $G = (V, A)$ with edge capacities $c : A \to \mathbb{R}^+$ and in addition to that a weight function $w : A \to \mathbb{R}^+$ indicating the *cost* of an edge. Thus a *network* is denoted $N = (G, c, w, b)$.

Now we define a modified notion of a flow. For any mapping $b : V \to \mathbb{R}$ with $\sum_{v\in V} b(v) = 0$ the value $b(v)$ is called the *balance* of a vertex v. If $b(v) > 0$ then v is called a *source*, if $b(v) < 0$ a *sink*. A *b-flow* in N is a function $f : A \to \mathbb{R}$ such that

(1) $0 \leq f(e) \leq c(e)$ for all $e \in A$ and

(2) $b(v) = \mathrm{bal}_f(v) = \sum_{e\in\delta^+(v)} f(e) - \sum_{e\in\delta^-(v)} f(e)$.

A 0-flow is called a *circulation*.

The *cost* of any flow f is

$$\mathrm{value}(f) = \sum_{e\in A} f(e)w(e).$$

Now the problem is to find a b-flow with minimum cost.

The second part of our task is easy. Given a network $N = (G, c, w, b)$ with balance vector b, we can decide if a b-flow exists by solving a Maximum Flow problem: Add two vertices s and t and edges sv, vt with capacities $c(sv) = \max\{0, b(v)\}$ and $c(vt) =$

Problem 3.2 MINIMUM COST FLOW

Instance. A network $N = (G, c, w, b)$.

Task. Find an b-flow of minimum cost in N or decide that none exists.

$\max\{0, -b(v)\}$ for all $v \in V$ to N. Then any $s-t$-flow with value $\sum_{v \in V} c(sv)$ in the resulting network corresponds to a b-flow in the original network N.

The MINIMUM COST FLOW problem can be solved in polynomial time with an approach similar to the FORD-FULKERSON method. But here we augment along cycles instead of paths. Again, the choice of the augmenting cycles must be done carefully. But we omit this here and state the following theorem which refers to ORLIN's algorithm without proof.

Theorem 3.8. *There is an algorithm which solves the* MINIMUM COST FLOW *problem on any network with n vertices and m edges in time $O\left(m \log m(m + n \log n)\right)$.*

3.3 Assignment Problem

A graph $G = (V, E)$ with vertex set $V = L \cup R$ ("left" and "right") is called *bipartite* if the edge set satisfies $E \subseteq \{\ell r : \ell \in L, r \in R\}$. An *assignment* (also called a *matching*) is a subset $M \subseteq E$ such that for every $v \in V$ in the graph $H = (V, M)$ we have $\deg_H(v) \leq 1$. A matching is called *perfect* if $\deg_H(v) = 1$ for every $v \in V$. Of course, a necessary condition for the existence of a perfect matching in a bipartite graph is $|L| = |R|$.

The ASSIGNMENT Problem has numerous applications and refers to the following. We are given a bipartite graph $G = (L \cup R, E)$ and a weight function $w : E \to \mathbb{R}$. We are asked to find a subset $M \subseteq E$ with minimum total weight, i.e.,

$$\text{value}(M) = \sum_{e \in M} w(e),$$

such that M is a perfect matching or to conclude that no such matching exists.

Problem 3.3 ASSIGNMENT

Instance. Bipartite graph $G = (L \cup R, E)$ and a weight function $w : E \to \mathbb{R}$.

Task. Find perfect matching M with minimum weight $\text{value}(M) = \sum_{e \in M} w(e)$ or conclude that no such matching exists.

Theorem 3.9. *The* ASSIGNMENT *problem is a* MINIMUM COST FLOW *problem.*

Proof. Let $G = (V, E)$ be a bipartite graph with $V = L \cup R$ and $|L| = |R| = n$. Now we construct a network N for the MINIMUM COST FLOW problem. We start with the vertices V, add a vertex s and connect it with every vertex $\ell \in L$ with directed edges $s\ell$. Further add a vertex t and introduce the directed edges rt for every $r \in R$. Further add directed versions of all edges $e \in E$, i.e., a directed edge ℓr is added for every undirected edge ℓr. The capacities of all these edges is one. The weights of the $s\ell$ edges and the rt edges are zero – the weights of the ℓr edges are equal to their weights in G. Now every integral b-flow f in N with $b = (b(s), b(v_1), \ldots, b(v_n), b(t)) = (n, 0, \ldots, 0, -n)$ corresponds to a perfect matching in G with the same weight, and vice versa. □

In most applications the requirement $|L| = |R|$ is disturbing, but can usually be handeled by adding artificial vertices and edges.

In the Bipartite Cardinality Matching problem we are given a bipartite graph $G = (V, E)$ with $V = L \cup R$, where $|L| \leq |R|$. Our task is to find a matching with maximum number of edges. We construct a network similarly as before: we add vertices s and t and the directed edges $s\ell$ and rt for all $\ell \in L$ and $r \in R$. All these edges have capacity equal to one. Any integral $s - t$-flow of value k corresponds to a matching with k edges. Thus we have to solve a Maximum Flow problem.

Chapter 4

Matchings

The theory of matchings is one of the classical topics in graph theory and combintorial optimization. The task of finding a matching occurs frequently as a subproblem of some other problem.

Let G be an undirected graph on the vertices $V(G)$ and the edges $E(G)$. A *matching* is a set $M \subseteq E(G)$ with the property that $e \cap e' = \emptyset$ for all $e \neq e' \in M$.

Problem 4.1 MATCHING

Instance. Undirected graph G on the vertices $V(G)$ and edges $E(G)$.

Task. Find a matching M in G with maximum cardinality, i.e., value$(M) = |M|$.

Let M be any matching in G. Any edge $e \in M$ is called *M-matched*, otherwise *M-free*. A vertex $v \in V(G)$ is *M-covered* if there is an edge $e = v\cdot \in M$. Otherwise v is *M-exposed*.

A matching M is called *maximal* if there is no edge e such that $M \cup \{e\}$ is a matching in G. A matching M is *maximum* if it has largest cardinality among all matchings of G. A matching covering all vertices of a graph is called *perfect*.

4.1 Maximum Matchings and Augmenting Paths

A *path* is a sequence $P = v_1, e_1, v_2, e_2, \ldots, e_k, v_{k+1}$ alternating between vertices and edges, such that $e_i = v_i v_{i+1} \in E$ for $i = 1, \ldots, k$. The vertices v_1 and v_{k+1} are the *end-vertices* of the path. The number of edges in the path is called its *length*. A path is *simple*, if its vertices are pairwise disjoint. We will always assume that paths are simple, unless stated otherwise. For any matching M, a path in which the edges alternate between matched and free edges is called *M-alternating path*. An alternating path where both end-vertices are M-exposed is called an *M-augmenting path*.

Theorem 4.1. *Let G be an undirected graph and M a matching in G. Then M is a maximum matching if and only if there is no M-augmenting path in G.*

Proof. If there is an M-augmenting path P in G, the symmetric difference $M \triangle E(P)$ is a matching and has greater cardinality than M. Thus M is not a maximum matching.

On the other hand, if there is a matching M' with $|M'| > |M|$, the symmetric difference $M \triangle M'$ is a vertex-disjoint union of alternating cycles (having even length) and paths, where at least one path must be M-augmenting. □

4.2 Blossom Algorithm

Theorem 4.1 tells us that one way to find a maximum matching is to find augmenting paths until no such exists. It is not clear how to organize the search such that a polynomial time algorithm is achieved. The central concepts behind the Blossom algorithm of Edmonds are alternating forests and blossoms. These will be introduced shortly. The main result is the following:

Theorem 4.2. *The* Blossom *algorithm computes a maximum matching in* $O(nm)$ *time.*

The main difficulty is the treatment of cycles of odd length. Observe that in any such cycle C, for any matching M, there is always at least one $(M \cap E(C))$-exposed vertex. Surprisingly, it suffices to get rid of an odd cycle by shrinking it to a single vertex.

A graph G is called *factor-critical* if $G - v$ has a perfect matching for each $v \in V(G)$. A matching M is called *near perfect*, if it covers all but one vertex. Observe that a cycle of odd length is factor critical and has a near-perfect matching.

Let G be a graph and M a matching in G. A *blossom* in G with respect to M is a factor-critical subgraph C of G with $|M \cap E(C)| = (|V(C)| - 1)/2$. The vertex of C exposed by $M \cap E(C)$ is called *base* of C.

Lemma 4.3. *Let G be a graph, M a matching in G, and C a blossom in G (with respect to M). Let $v \in V(G)$ be any M-exposed vertex and let r be the base of C. Suppose that there is an M-alternating v-r-path Q of even length with $E(Q) \cap E(C) = \emptyset$.*

Let G' and M' result from G and M by shrinking $V(C)$ to a single vertex. Then M is a maximum matching in G if and only if M' is a maximum matching in G'.

Proof. Suppose that M is not a maximum matching in G. The set $N = M \triangle E(Q)$ is a matching of the same cardinality, so it is not maximum either. By Theorem 4.1 there then exists an N-augmenting path P in G. Note that N does not cover r.

At least one of the end-vertices of P, say x does not belong to C. If P and C are disjoint, let y be the other end-vertex of P. Otherwise, let y be the first vertex on P – when traversed from x – belonging to C. Let G' result from G when shrinking $V(C)$. Let P' and N' in G' correspond to P and N in G. The end-vertices of P' are exposed by N'. Hence P' is an N' augmenting path in G'. So N' is not a maximum matching in G', and nor is M' (as it has the same cardinality).

To prove the converse, suppose that M' is not a maximum matching in G'. Let N' be a larger matching in G'. N' corresponds to a matching $\bar{N}$ in G which covers at most one vertex C in G. Since C is factor critical, $\bar{N}$ can be extended by $k = (|V(C)| - 1)/2$ many edges to a matching N in G, where

$$|N| = |\bar{N}| + k = |N'| + k > |M'| + k = |M|.$$

This shows that M is not a maximum matching in G. □

Given a graph G and a matching M in G. An *M-alternating forest* in G is a forest F in G with the following properties:

(1) $V(F)$ contains all the M-exposed vertices. Each connected component of F contains exactly one exposed vertex, called its *root*.

(2) We call a vertex $v \in V(F)$ an *outer* (*inner*) vertex, if it has even (odd) distance to the root of the connected component containing v. (In particular, the roots are outer vertices.) All inner vertices have degree 2 in F.

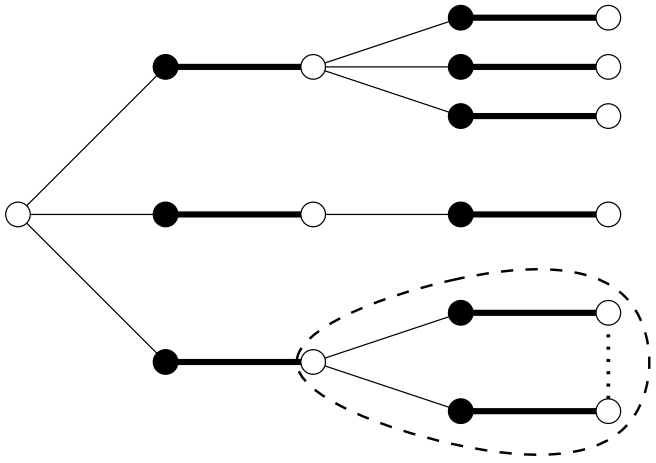


Figure 4.1: An alteranting forest (consisting of a single tree) with root r. The dotted edge forms a blossom (indicated by the dashed line) with base b. Observe that the unique path from b to r is even and alternating.

(3) For any $v \in V(F)$, the unique path from v to the root of the connected component containing v is M-alternating.

See Figure 4.1 for an illustration of an alternating forest and a blossom.

Observation 4.4. *In any alternating forest, the number of outer vertices that are not root equals the number of inner vertices.*

Proof. Each outer vertex that is not a root has exactly one neighbor which is an inner vertex and whose distance to the root is smaller. This is obviously a bijection between the outer vertices that are not a root and the inner vertices. □

The Blossom algorithm works as follows: Given some matching M, we build up an M-alternating forest F. We start with the set S of exposed vertices and no edges.

At any stage of the algorithm, we consider a neighbor v of an outer vertex u. Let $P(u)$ denote the unique path in F from u to a root. There are three interesting cases, corresponding to three operations "grow", "augment", and "shrink":

Case 1. $v \notin V(F)$. Then we add the edge uv and the matching edge $vw \in M$ to F. Thus the forest will grow.

Case 2. v is an outer vertex in a different component than u. Then we augment along the path $P(u) \cup \{uv\} \cup P(v)$.

Case 3. v is an outer vertex in the same connected component of F (with root r, say). Let ℓ be the first vertex of $P(u)$ (starting at u) which also belongs to $P(v)$. The vertex ℓ can be one of u and v. If ℓ is not a root, then it must have degree at least three. Thus it is an outer vertex then. If ℓ is a root, it is an outer vertex by definition. Therefore $C = P(u)[u, \ell] \cap \{uv\}P(v)[v, \ell]$ is a blossom with at least three vertices. We shrink C.

If none of the cases applies, all neighbors of outer vertices are inner. Then the algorithm terminates and returns the matching M found.

Proof of Theorem 4.2. We claim that M is maximum. Consider an alternating forest F that belongs to a maximum matching. Let S be the set of outer vertices and let T be the set of inner vertices of F. Denote $s = |S|$ and $t = |T|$. $G - T$ has t odd components (each outer vertex is isolated in $G - T$). Hence any matching must leave at least $s - t$ vertices uncovered. But on the other hand, the number of vertices exposed by M, i.e., the number of roots in F, is exactly $s - t$ by Observation 4.4.

For the running time observe that there are at most n augmentations taking time $O(m)$ each. Growing the forest takes time $O(1)$, which can occur at most n times between two augmentations. Shrinking a blossom C takes time $O(|V(C)|)$ which yields that the forest has $|V(C)| - 1 > 1$ vertices less. Hence the total effort for shrinkings between two augmentations is at most $O(n)$ (with a suitable implementation). Checking the other edges, that do not correspond to any of the three cases, takes total time $O(m)$ between two augmentations. □

The implementation of the Blossom algorithm is not an easy task. For this reason, we do not go into the details here.

Chapter 5

Linear Programming

Linear programs (LP) play an important role in the theory and practice of optimization problems. Many COPs can directly be formulated as LPs. Furthermore, LPs are invaluable for the design and analysis of approximation algorithms. Generally speaking, LPs are COPs with linear objective function and linear constraints, where the variables are defined on a continous domain. We will be more specific below.

5.1 Introduction

We begin our treatment of linear programming with an example of a transportation problem to illustrate how LPs can be used to formulate optimization problems.

Example 5.1. There are two brickworks w_1, w_2 and three construction sites s_1, s_2, s_3. The works produce $b_1 = 60$ and $b_2 = 30$ tons of bricks per day. The sites require $c_1 = 30$, $c_2 = 20$ and $c_3 = 40$ tons of bricks per day. The transportation costs t_{ij} per ton from work w_i to site s_j are given in the following table:

t_{ij}	s_1	s_2	s_3
w_1	40	75	50
w_2	20	50	40

Which work delivers which site in order to minimize the total transportation cost? Let us write the problem as a mathematical program. We use variables x_{ij} that tell us how much we deliver from work w_i to site s_j.

$$\begin{aligned}
\text{minimize} \quad & 40x_{11} + 75x_{12} + 50x_{13} + 20x_{21} + 50x_{22} + 40x_{23} \\
\text{subject to} \quad & x_{11} + x_{12} + x_{13} \leq 60 \\
& x_{21} + x_{22} + x_{23} \leq 30 \\
& x_{11} + x_{21} = 30 \\
& x_{12} + x_{21} = 20 \\
& x_{13} + x_{23} = 40 \\
& x_{ij} \geq 0 \quad i = 1, 2,\ j = 1, 2, 3.
\end{aligned}$$

How do we find the best x_{ij}?

The general Linear Programming task is given in Problem 5.1.

As a shorthand we shall frequently write $\max\{c^\top x : Ax \leq b\}$. We can assume that we deal with a maximization problem without loss of generality because we can treat a minimization problem if we replace c with $-c$.

Problem 5.1 LINEAR PROGRAMMING

Instance. Matrix $A \in \mathbb{R}^{m \times n}$, vectors $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$.

Task. Solve the problem

$$\begin{aligned} \text{maximze} \quad & c^\top x, \\ \text{subject to} \quad & Ax \leq b, \\ & x \in \mathbb{R}^n. \end{aligned}$$

That means answer one of the following questions.

(1) Find a vector $x \in \mathbb{R}^n$ such that $Ax \leq b$ and value$(x) = c^\top x$ is maximum, or

(2) decide that the set $P = \{x \in \mathbb{R}^n : Ax \leq b\}$ is empty, or

(3) decide that for all $\alpha \in \mathbb{R}$ there is an $x \in \mathbb{R}^n$ with $Ax \leq b$ and $c^\top x > \alpha$.

The function value$(x) = c^\top x$ is the *objective function*. A feasible x^* which maximizes value is an *optimum solution* and the value $z^* = \text{value}(x^*)$ is called *optimum value*. Any $x \in \mathbb{R}^n$ that satisfies $Ax \leq b$ is called *feasible*. The set $P = \{x \in \mathbb{R}^n : Ax \leq b\}$ is called the *feasible region*, i.e., the set of *feasible solutions*. If P is empty, then the problem is *infeasible*. If for every $\alpha \in \mathbb{R}$, there is a feasible x such that $c^\top x > \alpha$ then the problem is *unbounded*. This simply means that the maximum of the objective function does not exist.

5.2 Polyhedra

Consider the vector space $\mathbb{R}^n$. A *(linear) subspace* S of R^n is a subset of $\mathbb{R}^n$ closed under vector addition and scalar multiplication. Equivalently, S is the set of all points in $\mathbb{R}^n$ that satisfy a set of homogeneous linear equations:

$$S = \{x \in \mathbb{R}^n : Ax = 0\},$$

for some matrix $A \in \mathbb{R}^{n \times m}$. The *dimension* $\dim(S)$ is equal to the maximum number of linear independent vectors in S, i.e., $\dim(S) = n - \text{rank}(A)$. Here $\text{rank}(A)$ denotes the number of linear independent rows of A. An *affine subspace* S_b of $\mathbb{R}^n$ is the set of all points that satisfy a set of inhomogeneous linear equations:

$$S_b = \{x \in \mathbb{R}^n : Ax = b\}.$$

We have $\dim(S_b) = \dim(S)$. The *dimension* $\dim(X)$ of any subset $X \subseteq \mathbb{R}^n$ is the smallest dimension of any affine subspace which contains it.

An affine subspace of $\mathbb{R}^n$ of dimension $n - 1$ is called *hyperplane*, i.e., alternatively

$$H = \{x \in \mathbb{R}^n : a^\top x = b\},$$

for some vector $a \in \mathbb{R}^n, a \neq 0$ and scalar b. A hyperplane defines two (closed) *halfspaces*

$$H^+ = \{x \in \mathbb{R}^n : a^\top x \geq b\},$$
$$H^- = \{x \in \mathbb{R}^n : a^\top x \leq b\}.$$

As a halfspace is a convex set, the intersection of halfspaces is also convex.

A *polyhedron* in $\mathbb{R}^n$ is a set

$$P = \{x \in \mathbb{R}^n : Ax \leq b\}$$

for some matrix $A \in \mathbb{R}^{m \times n}$ and some vector $b \in \mathbb{R}^m$. A bounded polyhedron is called *polytope*.

Let $P = \{x : Ax \leq b\}$ be a non-empty polyhedron with dimension d. Let c be a vector for which $\delta := \max\{c^\top x : x \in P\} < \infty$, then

$$H_c = \{x : c^\top x = \delta\}$$

is called *supporting hyperplane* of P. A *face* of P is the intersection of P with a supporting hyperplane of P. Three types of faces are particular important, see Figure 5.1:

(1) A *facet* is a face of dimension $d - 1$,

(2) a *vertex* is a face of dimension zero (a point), and

(3) an *edge* is a face of dimension one (a line segment).

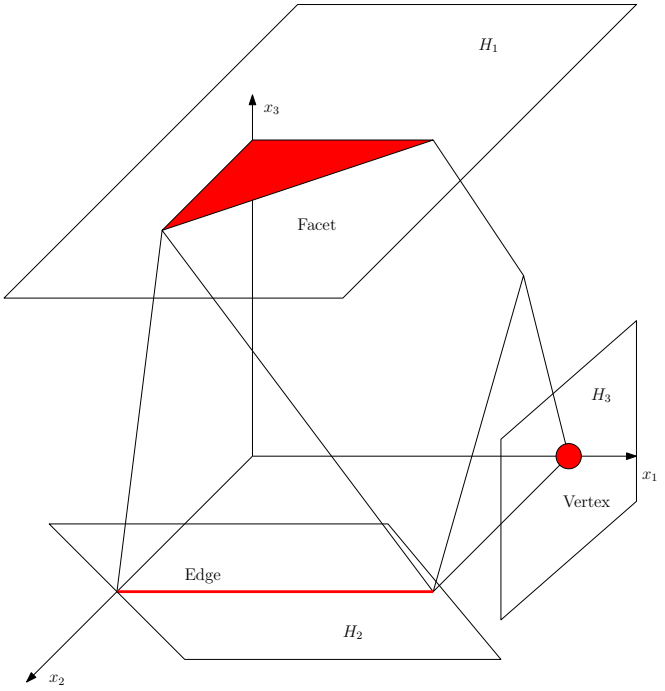


Figure 5.1: Facet, vertex, and edge.

The following lemma essentially states that a set $F \subseteq P$ is a face of a polyhedron P if and only if some of the inequalities of $Ax \leq b$ are satisfied with equality for all elements of F.

Lemma 5.2. *Let $P = \{x : Ax \leq b\}$ be a polyhedron and $F \subseteq P$. Then the following statements are equivalent:*

(1) F is a face of P.

(2) There is a vector c with $\delta := \max\{c^\top x : x \in P\} < \infty$ and $F = \{x \in P : c^\top x = \delta\}$.

(3) $F = \{x \in P : A'x = b'\} \neq \emptyset$ for some subsystem $A'x \leq b'$ of $Ax \leq b$.

As important corollaries we have:

Corollary 5.3. *If $\max\{c^\top x : x \in P\} < \infty$ for a non-empty polyhedron P and a vector c, then the set of points where the maximum is attained is a face of P.*

Corollary 5.4. *Let P be a polyhedron and F a face of P. Then F is again a polyhedron. Furthermore, a set $F' \subseteq F$ is a face of P if and only if it is a face of F.*

A important class of faces are *minimal faces*, i.e., faces that do not contain any other face. For these we have:

Lemma 5.5. *Let $P = \{x : Ax \leq b\}$ be a polyhedron. A non-empty set $F \subseteq P$ is a minimal face of P if and only if $F = \{x \in \mathbb{R}^n : A'x = b'\}$ for some subsystem of $Ax \leq b$.*

Corollary 5.3 and Lemma 5.5 already imply that Linear Programming can be solved by solving the linear *equation* system $A'x = b'$ for each subsystem $A'x \leq b'$. This approach obviously yields an exponential time algorithm. An algorithm which is more practicable (although also exponential in the worst case) is the Simplex algorithm. The algorithm is based on the following important consequence of Lemma 5.5.

Corollary 5.6. *Let $P = \{x \in \mathbb{R}^n : Ax \leq b\}$ be a polyhedron. Then all minimal faces of P have dimension $n - \operatorname{rank}(A)$. The minimal faces of polytopes are vertices.*

Thus, it suffices to search an optimum solution among the *vertices* of the polyhedron. This is what the Simplex algorithm is doing. We do not explain the algorithm in detail here, but it works as follows. Provided that the polyhedron is not empty, it finds an initial vertex. If the current vertex is not optimal, find another vertex with strictly larger objective value (pivot rule). Iterate until an optimal vertex is found or the polyhedron can be shown to be unbounded. See Figure 5.2.

The algorithm terminates after at most $\binom{m}{n}$ iterations (which is not polynomial). It was conjectured that Simplex is polynomial until Klee and Minty gave an example where the algorithm (with Bland's pivot rule) uses 2^n iterations on an LP with n variables and $2n$ constraints. It is not known if there is a pivot rule that leads to polynomial running time. Nonetheless, Simplex with Bland's pivot rule is frequently observed to terminate after few iterations when run on "practical instances".

However, there are algorithms, e.g., the Ellipsoid method and Karmakar's algorithm that solve Linear Programming in polynomial time. But these algorithms are mainly of interest from a theoretical point of view. We conclude with the statement that one can solve Linear Programming in polynomial time with "black box" algorithms.

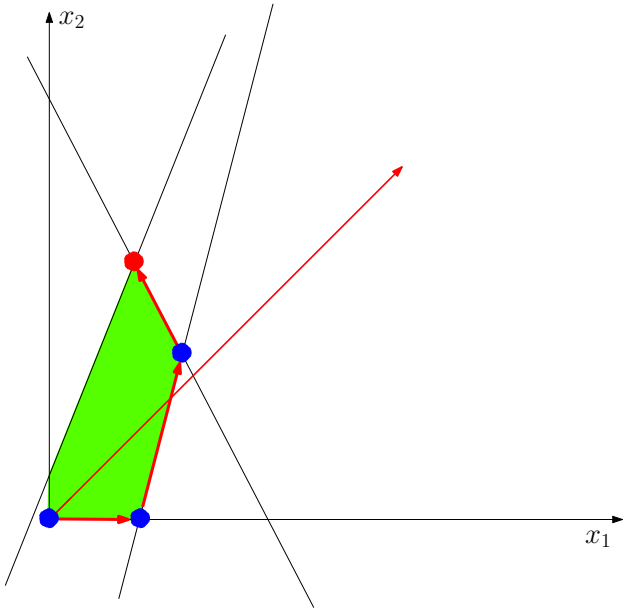


Figure 5.2: A SIMPLEX path.

5.3 Duality

Intuition behind Duality

Consider the following LP, which is illustrated in Figure 5.3

$$\begin{aligned}
\text{maximize} \quad & x_1 + x_2 && (5.1)\\
\text{subject to} \quad & 4x_1 - x_2 \leq 8 && (5.2)\\
& 2x_1 + x_2 \leq 10 && (5.3)\\
& -5x_1 + 2x_2 \leq 2 && (5.4)\\
& -x_1 \leq 0 && (5.5)\\
& -x_2 \leq 0 && (5.6)
\end{aligned}$$

and notice that this LP is in the maximization form

$$\max\{c^\top x : Ax \leq b\}.$$

Because we are dealing with a maximization problem, every feasible solution x provides the *lower bound* $c^\top x$ on the value $c^\top x^*$ of the optimum solution x^*, i.e., we know $c^\top x \leq c^\top x^*$.

Can we also obtain *upper bounds* on $c^\top x^*$? For any feasible solution x, the constraints (5.2)–(5.6) are satisfied. Now compare the objective function (5.1) with the constraint (5.3) *coefficient-by-coefficient* (where we remember that $x_1, x_2 \geq 0$ in this example):

$$\begin{array}{ccccc}
1 \cdot x_1 & + & 1 \cdot x_2 & & \\
\wedge| & & \wedge| & & \\
2 \cdot x_1 & + & 1 \cdot x_2 & \leq & 10
\end{array}$$

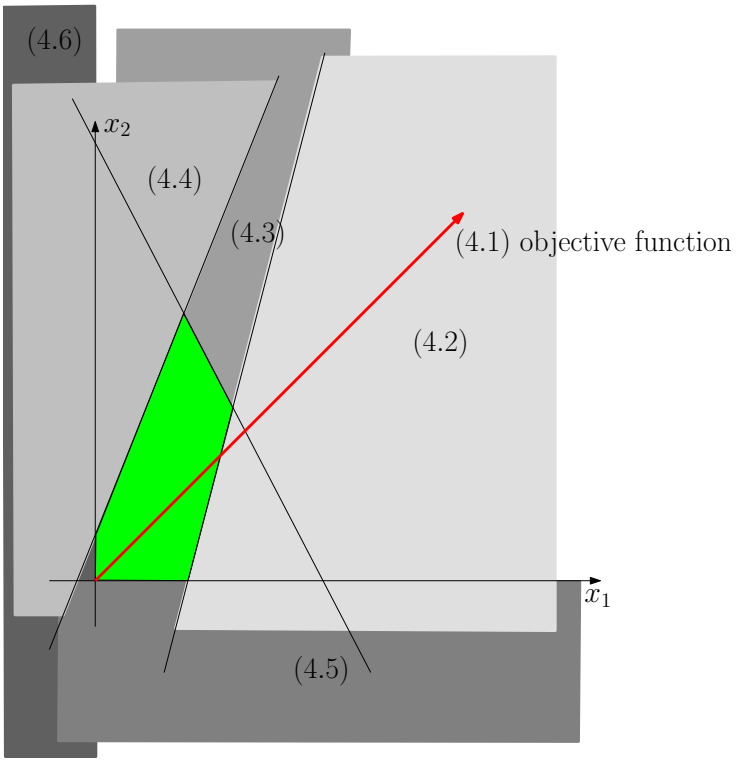


Figure 5.3: An LP.

Thus for *every* feasible solution x we have the upper bound $x_1 + x_2 \leq 10$, i.e., the optimum value can be at most 10. Can we improve on this? We could try $\frac{7}{9} \cdot (5.3) + \frac{1}{9} \cdot (5.4)$:

$$\begin{array}{ccccc} 1 \cdot x_1 & + & 1 \cdot x_2 & & \\ \wedge\mid & & \wedge\mid & & \\ (\frac{7}{9} \cdot 2 + \frac{1}{9} \cdot (-5))x_1 & + & (\frac{7}{9} \cdot 1 + \frac{1}{9} \cdot 2)x_2 & \leq \frac{7}{9} \cdot 10 + \frac{1}{9} \cdot 2 = \frac{72}{9} = 8 \end{array}$$

Hence we have $x_1 + x_2 \leq 8$ for every feasible x and thus an upper bound of 8 on the optimum value. If we look closely, our choices 7/9 and 1/9 give $\frac{7}{9} \cdot 2 + \frac{1}{9} \cdot (-5) = 1$ and $\frac{7}{9} \cdot 1 + \frac{1}{9} \cdot 2 = 1$, i.e., we have combined the coefficients of the objective function $c^\top x$ with *equality*. This is also the best bound this approach can give here.

This suggests the following genear approach for obtining upper bounds on the optimal value. Combine the constraints with non-negative multipliers $y = (y_1, y_2, y_3, y_4, y_5)$ such that each coefficient in the result equals the corresponding coefficient in the objective function, i.e., we want $y^\top A = c^\top$. We associate y_1 with (5.2), y_2 with (5.3), y_3 with (5.4), y_4 with (5.5), and y_5 with (5.6). Notice that the y_i must be non-negative because we are multiplying an *inequality* of the system $Ax \leq b$, i.e., if a multiplier y_i were negative we change the corresponding inequality from "$\leq$" to "$\geq$". Now $y_1(5.2) + y_2(5.3) + y_3(5.4) + y_4(5.5) + y_5(5.6)$ evaluates to

$$y_1(4x_1 - 1x_2) + y_2(2x_1 + x_2) + y_3(-5x_1 + 2x_2) + y_4(-x_1) + y_5(-x_2) \leq y_1 8 + y_2 10 + y_3 2 + y_4 0 + y_5 0,$$

where rearranging yields

$$(4y_1 + 2y_2 - 5y_3 - y_4)x_1 + (-y_1 + y_2 + 2y_3 - y_5)x_2 \leq 8y_1 + 10y_2 + 2y_3 + 0y_4 + 0y_5$$

and want to find values for $y_1, y_2, y_3, y_4, y_5 \geq 0$ that satisfy:

$$\begin{array}{ccccc} 1 \cdot x_1 & & + & & 1 \cdot x_2 \\ \| & & & & \| \\ (4y_1 + 2y_2 - 5y_3 - y_4)x_1 & & + & & (-y_1 + y_2 + 2y_3 - y_5)x_2 \end{array} \leq 8y_1 + 10y_2 + 2y_3 + 0y_4 + 0y_5$$

Of course, we are interested in the best choice for $y = (y_1, y_2, y_3, y_4, y_5) \geq 0$ the approach can give. This means that we want to minimize the upper bound $8y_1+10y_2+2y_3+0y_4+0y_5$. We simply write down this task as a mathematical program, which turns out to be an LP.

$$\begin{aligned} \text{minimize} \quad & 8y_1 + 10y_2 + 2y_3 + 0y_4 + 0y_5 && (5.7) \\ \text{subject to} \quad & 4y_1 + 2y_2 - 5y_3 - y_4 = 1 && (5.8) \\ & -y_1 + y_2 + 2y_3 - y_5 = 1 && (5.9) \\ & y_1, y_2, y_3, y_4, y_5 \geq 0 && (5.10) \end{aligned}$$

Further note that the new objective function is the right hand side $(8, 10, 2, 0, 0)^\top$ of the original LP and that the new right hand side is the objective function $(1, 1)^\top$ of the original LP. Thus the above LP is of the form

$$\min\{y^\top b : y^\top A = c^\top, y \geq 0\}.$$

Notice that there is a feasible solution $x = (2, 6)^\top$ for the original LP that gives $c^\top x = 8$. Further note that the multipliers $y = (0, 7/9, 1/9, 0, 0)^\top$ yield $y^\top b = 8$, i.e.,

$$c^\top x = y^\top b.$$

Hence we have a certificate that the solution $x = (2, 6)$ is indeed optimal (because we have a matching upper bound). Not surprisingly this is no exception but the principal statement of the *strong duality* theorem.

Weak and Strong Duality

Given an LP

$$P = \max\{c^\top x : Ax \leq b\}$$

called *primal*, we define the *dual*

$$D = \min\{y^\top b : y^\top A = c^\top, y \geq 0\}.$$

Lemma 5.7. *The dual of the dual of an LP is (equivalent to) the original LP.*

Now we can say that the LPs P and D are dual to each other or a *primal-dual pair*. The following forms of primal-dual pairs are standard:

$$\begin{aligned} \max\{c^\top x : Ax \leq b\} &\sim \min\{y^\top b : y^\top A = c^\top, y \geq 0\} \\ \max\{c^\top x : Ax \leq b, x \geq 0\} &\sim \min\{y^\top b : y^\top A \geq c^\top, y \geq 0\} \\ \max\{c^\top x : Ax = b, x \geq 0\} &\sim \min\{y^\top b : y^\top A \geq c^\top\} \end{aligned}$$

The following lemma is called *weak duality*.

Lemma 5.8 (Weak Duality)**.** *Let x and y be respective feasible solutions of the primal-dual pair $P = \max\{c^\top x : Ax \leq b\}$ and $D = \min\{y^\top b : y^\top A = c^\top, y \geq 0\}$. Then $c^\top x \leq y^\top b$.*

Proof. $c^\top x = (y^\top A)x = y^\top(Ax) \leq y^\top b$. □

The following *strong duality* theorem is the most important result in LP theory and the basis for a lot of algorithms for COPs.

Theorem 5.9 (Strong Duality)**.** *For any primal-dual pair* $P = \max\{c^\top x : Ax \leq b\}$ *and* $D = \min\{y^\top b : y^\top A = c^\top, y \geq 0\}$ *we have:*

(1) If P *and* D *have respective optimum solutions* x *and* y*, say, then*

$$c^\top x = y^\top b.$$

(2) If P *is unbounded, then* D *is infeasible.*

(3) If P *is infeasible, then* D *is infeasible or unbounded.*

Before we prove the theorem, we establish the *fundamental theorem of linear inequalities*. The heart of the proof actually gives a basic version of the SIMPLEX algorithm. The result also implies Farkas' Lemma.

Theorem 5.10. *Let* $a_1, \ldots, a_m, b$ *be vectors in* n*-dimensional space. Then*

either (I): $b = \sum_{i=1}^m a_i\lambda_i$ *with* $\lambda_i \geq 0$ *for* $i = 1, \ldots, m$*,*

or (II): *there is a hyperplane* $\{x : c^\top x = 0\}$*, containing* $t-1$ *linearly independent vectors from* $a_1, \ldots, a_m$ *such that* $c^\top b < 0$ *and* $c^\top a_1, \ldots, c^\top a_m \geq 0$*, where* $t = \operatorname{rank}\{a_1, \ldots, a_m, b\}$*.*

Proof. We may assume that $a_1, \ldots, a_m$ span the n-dimensional space. Clearly, (I) and (II) exclude each other as we would otherwise have the contradiction

$$0 > c^\top b = \lambda_1 c^\top a_1 + \cdots + \lambda_m c^\top a_m \geq 0.$$

To see that at least one of (I) and (II) holds, choose linearly independent $a_{i_1}, \ldots, a_{i_n}$ from $a_1, \ldots, a_m$ and set $B = \{a_{i_1}, \ldots, a_{i_n}\}$. Next apply the following iteration:

(i) Write $b = \lambda_{i_1} a_{i_1} + \cdots + \lambda_{i_n} a_{i_n}$. If $\lambda_{i_1}, \ldots, \lambda_{i_n} \geq 0$ we are in case (I).

(ii) Otherwise, choose the smallest h among $i_1, \ldots, i_n$ with $\lambda_h < 0$. Let $\{x : c^\top x = 0\}$ be the hyperplane spanned by $B - \{a_h\}$. We normalize c so that $c^\top a_h = 1$. (Hence $c^\top b = \lambda_h < 0$.)

(iii) If $c^\top a_1, \ldots, c^\top a_m \geq 0$ we are in case (II).

(iv) Otherwise, choose the smallest s such that $c^\top a_s < 0$. Then replace B by $(B - \{a_h\}) \cup \{a_s\}$. Restart the iteration anew.

We are finished if we have shown that this process terminates. Let B_k denote the set B as it is in the k-th iteration. If the process does not terminate, then $B_k = B_\ell$ for some $k < \ell$ (as there are only finitely many choices for B). Let r be the highest index for which a_r has been removed from B at the end of one of the iterations $k, k+1, \ldots, \ell-1$, say in iteration p. As $B_k = B_\ell$, we know that a_r also has been added to B in some iteration q with $k \leq q \leq \ell$. So

$$B_p \cap \{a_{r+1}, \ldots, a_m\} = B_q \cap \{a_{r+1}, \ldots, a_m\}.$$

Let $B_p = \{a_{i_1}, \ldots, a_{i_n}\}$, $b = \lambda_{i_1} a_{i_1} + \cdots + \lambda_{i_n} a_{i_n}$, and let d be the vector c found in iteration q. Then we have the contradiction

$$0 > d^\top b = d^\top(\lambda_{i_1} a_{i_1} + \cdots + \lambda_{i_n} a_{i_n}) = \lambda_{i_1} d^\top a_{i_1} + \cdots + \lambda_{i_n} d^\top a_{i_n} + \cdots + \lambda_{i_n} d^\top a_{i_n} > 0,$$

where the second inequality follows from: If $i_j < r$ then $\lambda_{i_j} \geq 0, d^\top a_{i_j} \geq 0$, if $i_j = r$ then $\lambda_{i_j} < 0, d^\top a_{i_j} < 0$, and if $i_j > r$ then $d^\top a_{i_j} = 0$. □

Lemma 5.11 (Farkas' Lemma)**.** *There is a vector*

$$\begin{cases} x & \text{with } Ax \leq b \\ x \geq 0 & \text{with } Ax \leq b \\ x \geq 0 & \text{with } Ax = b \end{cases} \quad \text{if and only if } y^\top b \geq 0 \text{ for all} \quad \begin{cases} y \geq 0 & \text{with } y^\top A = 0 \\ y \geq 0 & \text{with } y^\top A \geq 0 \\ y & \text{with } y^\top A \geq 0 \end{cases}.$$

Proof. We first show the case $x \geq 0$ with $Ax = b$ if and only if $y^\top b \geq 0$ for each y with $y^\top A \geq 0$.

Necessity is clear since $y^\top b = y^\top(Ax) \geq 0$ for all x and y with $x \geq 0$, $y^\top A \geq 0$, and $Ax = b$. For suffience, assume that there is no $x \geq 0$ with $Ax = b$. Then, by Theorem 5.10 and denoting $a_1, \ldots, a_m$ be the colums of A, there is a hyperplane $\{x : y^\top x = 0\}$ with $y^\top b < 0$ for some y with $y^\top A \geq 0$.

For the case x with $Ax \leq b$ if and only if $y^\top b \geq 0$ for each $y \geq 0$ with $y^\top A = 0$ consider $A' = [I, A, -A]$. Observe that $Ax \leq b$ has a solution x if and only if $A'x' = b$ has a solution $x' \geq 0$. Now apply what we have just proved.

For the case $x \geq 0$ with $Ax \leq b$ if and only if $y^\top b \geq 0$ for each $y \geq 0$ with $y^\top A \geq 0$ consider $A' = [I, A]$. Observe that $Ax \leq b$ has a solution $x \geq 0$ if and only if $A'x' = b$ has a solution $x' \geq 0$. Now apply what we have just proved. □

Proof of Theorem 5.9. For (1) both optima exist. Thus, if $Ax \leq b$ and $y \geq 0$, $y^\top A = c^\top$, then $c^\top x = y^\top Ax \leq y^\top b$. Now it suffices to show that there are x, y such that $Ax \leq b$, $y \geq 0$, $y^\top A = c^\top$, $c^\top x \geq y^\top b$, i.e., that

$$\text{there are } x, y \text{ such that } y \geq 0 \text{ and } \begin{pmatrix} A & 0 \\ -c^\top & b^\top \\ 0 & A^\top \\ 0 & -A^\top \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \leq \begin{pmatrix} b \\ 0 \\ c^\top \\ -c^\top \end{pmatrix}$$

By Lemma 5.11 this is equivalent to: If $u, \lambda, v, w \geq 0$ with $uA - \lambda c^\top = 0$ and $\lambda b^\top + vA^\top - wA^\top \geq 0$ then $ub + vc - wc \geq 0$.

Let u, λ, v, w satisfy this premise. If $\lambda > 0$ then $ub = \lambda^{-1}\lambda b^\top u^\top \geq \lambda^{-1}(w - v)A^\top u^\top = \lambda^{-1}\lambda(w - v)c = (w - v)c$. If $\lambda = 0$, let $Ax_0 \leq b$ and $y_0 \geq 0$, $y_0^\top A = c^\top$. (x_0, y_0 exist since P and D are not empty.) Then $ub \geq uAx_0 = 0 \geq (w - v)A^\top y_0^\top = (w - v)c$.

The claim (2) directly follows from Lemma 5.8. For (3), if D is infeasible there is nothing to show. Thus let D be feasible. From Lemma 5.11 we get: Since $Ax \leq b$ is infeasible, there is a vector $y \geq 0$ with $y^\top A = 0$ and $y^\top b < 0$. Let z be such that $z^\top A = c^\top$ and $\alpha > 0$. Then $\alpha y + z$ is feasible with objective value $\alpha y^\top b + z^\top b$, which can be made arbitrarily small since $y^\top b < 0$ and $\alpha > 0$. □

The theorem has a lot of implications but we only list two of them. The first one is called *complementary slackness* (and gives another way of proving optimality).

Corollary 5.12. *Let* $\max\{c^\top x : Ax \leq b\}$ *and* $\min\{y^\top b : y^\top A = c, y \geq 0\}$ *be a primal-dual pair and let* x *and* y *be respective feasible solutions. Then the following statements are equivalent:*

(1) x *and* y *are both optimum solutions.*

(2) $c^\top x = y^\top b$.

(3) $y^\top (b - Ax) = 0$.

Secondly, the fact that a system $Ax \leq b$ is infeasible can be proved by giving a vector $y \geq 0$ with $y^\top A = 0$ and $y^\top b < 0$ (Farkas' Lemma).

Part II

Approximation Algorithms

Chapter 6

Knapsack

This chapter is concerned with the KNAPSACK problem. This problem is of interest in its own right because it formalizes the natural problem of selecting items so that a given budget is not exceeded but profit is as large as possible. Questions like that often also arise as subproblems of other problems. Typical applications include: option-selection in finance, cutting, and packing problems.

In the KNAPSACK problem we are given a budget W and n items. Each item j comes along with a profit c_j and a weight w_j. We are asked to choose a subset of the items as to maximize total profit but the total weight not exceeding W.

Example 6.1. We are given an amount of W and we wish to buy a subset of n items and sell those later on. Each such item j has cost w_j but yields profit c_j. The goal is to maximize the total profit. Consider $W = 100$ and the following profit-weight table:

j	c_j	w_j
1	150	100
2	2	1
3	55	50
4	100	50

Our choice of purchased items must not exceed our capital W. Thus the feasible solutions are $\{1\}, \{2\}, \{3\}, \{4\}, \{2,3\}, \{2,4\}, \{3,4\}$. Which is the best solution? Evaluating all possibilities yields that $\{3,4\}$ gives 155 altogether which maximizes our profit.

Problem 6.1 KNAPSACK

Instance. Non-negative integral vectors $c \in \mathbb{N}^n, w \in \mathbb{N}^n$, and an integer W.

Task. Solve the problem

$$\begin{aligned} \text{maximize} \quad & \text{value}(x) = \sum_{j=1}^{n} c_j x_j, \\ \text{subject to} \quad & \sum_{j=1}^{n} w_j x_j \leq W, \\ & x_j \in \{0,1\} \quad j = 1, \ldots n. \end{aligned}$$

For an *item* j the quantity c_j is called its *profit*. The profit of a vector $x \in \{0,1\}^n$ is $\text{value}(x) = \sum_{j=1}^n c_j x_j$.

The number w_j is called the *weight* of item j. The *weight* of a vector $x \in \{0,1\}^n$ is given by $\text{weight}(x) = \sum_{j=1}^n w_j x_j$. In order to obtain a non-trivial problem we assume $w_j \leq W$ for all $j = 1, \ldots, n$ and $\sum_{j=1}^n w_j > W$ throughout.

KNAPSACK is NP-hard which means that "most probably", there is no polynomial time optimization algorithm for it. However, in Section 6.1 we derive a simple 1/2-approximation algorithm. In Section 6.3 we can even improve on this by giving a polynomial-time $1-\varepsilon$-approximation algorithm (for every fixed $\varepsilon > 0$).

6.1 Fractional Knapsack and Greedy

A direct relaxation of KNAPSACK as an LP is often referred to as the FRACTIONAL KNAPSACK problem:

$$\begin{aligned} \text{maximize} \quad & \text{value}(x) = \sum_{j=1}^n c_j x_j, \\ \text{subject to} \quad & \sum_{j=1}^n w_j x_j \leq W, \\ & 0 \leq x_j \leq 1 \quad j = 1, \ldots, n. \end{aligned}$$

This problem is solvable in polynomial time quite easily. The proof of the observation below is left as an exercise.

Observation 6.2. *Let $c, w \in \mathbb{N}^n$ be non-negative integral vectors with*

$$\frac{c_1}{w_1} \geq \frac{c_2}{w_2} \geq \cdots \geq \frac{c_n}{w_n}$$

and let

$$k = \min \left\{ j \in \{1, \ldots, n\} : \sum_{i=1}^{j} w_i > W \right\}.$$

Then an optimum solution for the FRACTIONAL KNAPSACK *problem is given by*

$$\begin{aligned} & x_j = 1 \quad \textit{for } j = 1, \ldots, k-1, \\ & x_j = \frac{W - \sum_{i=1}^{k-1} w_i}{w_k} \quad \textit{for } j = k, \textit{ and} \\ & x_j = 0 \quad \textit{for } j = k+1, \ldots, n. \end{aligned}$$

The ratio c_j/w_j is called the *efficiency* of item j. The item number k, as defined above, is called the *break item*.

Now we turn our attention back to the original KNAPSACK problem. We may assume that the items are given in non-increasing order of efficiency. Observation 6.2 suggests the following simple algorithm: $x_j = 1$ for $j = 1, \ldots, k-1$, $x_j = 0$ for $j = k, \ldots, n$.

Unfortunately, the approximation ratio of this algorithm can be arbitrarily bad as the example below shows. The problem is that more efficient items can "block" more profitable ones.

Example 6.3. Consider the following instance, where W is a sufficiently large integer.

j	c_j	w_j	c_j/w_j
1	1	1	1
2	$W-1$	W	$1-1/W$

The algorithm chooses item 1, i.e., the solution $x = (1, 0)$ and hence value$(x) = 1$. The optimum solution is $x^* = (0, 1)$ and thus value$(x^*) = W - 1$. The approximation ratio of the algorithm is $1/(W-1)$, i.e., arbitrarily bad. However, this natural algorithm can be turned into a 1/2-approximation.

Algorithm 6.1 GREEDY

Input. Integer W, vectors $c, w \in \mathbb{N}^n$ with $w_j \leq W$, $\sum_j w_j > W$, and $c_1/w_1 \geq \cdots \geq c_n/w_n$.

Output. Vector $x \in \{0,1\}^n$ such that weight$(x) \leq W$.

Step 1. Define $k = \min\{j \in \{1, \ldots, n\} : \sum_{i=1}^{j} w_i > W\}$.

Step 2. Let x and y be the following two vectors: $x_j = 1$ for $j = 1, \ldots, k-1$, $x_j = 0$ for $j = k, \ldots, n$, and $y_j = 1$ for $j = k$, $y_j = 0$ for $j \neq k$.

Step 3. If value$(x) \geq$ value(y) return x otherwise return y.

Theorem 6.4. *The algorithm* GREEDY *is a* 1/2*-approximation for* KNAPSACK.

Proof. The value obtained by the GREEDY algorithm is equal to $\max\{\text{value}(x), \text{value}(y)\}$.

Let x^* be an optimum solution for the KNAPSACK instance. Since every solution that is feasible for the KNAPSACK instance is also feasible for the respective FRACTIONAL KNAPSACK instance we have that

$$\text{value}(x^*) \leq \text{value}(z^*),$$

where z^* is the respective optimum solution for FRACTIONAL KNAPSACK. Observe that it has the structure $z^* = (1, \ldots, 1, \alpha, 0, \ldots, 0)$, where $\alpha \in [0, 1)$ is at the break item k. The solutions x and y are $x = (1, \ldots, 1, 0, 0, \ldots, 0)$ and $y = (0, \ldots, 0, 1, 0, \ldots, 0)$.

In total we have

$$\text{value}(x^*) \leq \text{value}(z^*) = \text{value}(x) + \alpha c_k \leq \text{value}(x) + \text{value}(y) \leq 2\max\{\text{value}(x), \text{value}(y)\}$$

which implies the approximation ratio of 1/2. □

6.2 Pseudo-Polynomial Time Algorithm

Here we give a pseudo-polynomial time algorithm that solves KNAPSACK optimally by using dynamic programming. The term *pseudo-polynomial* means polynomial if the input is given in unary encoding (and thus exponential if the input is given in binary encoding).

The idea is the following: Suppose you restrict yourself to choose only among the first j items, for some integer $j \in \{0, \ldots, n\}$. So all the solutions x you consider have the form $x_i \in \{0, 1\}$ for $i = 1, \ldots, j$ and $x_i = 0$ for $i = j+1, \ldots, n$. With abuse of

notation write $x \in \{0,1\}^j 0^{n-j}$. Now the variable $m_{j,k}$ equals the minimum total weight of such a solution x with $\text{weight}(x) \leq W$ and $\text{value}(x) = k$. That is, after defining the set $W_{j,k} = \{\text{weight}(x) : \text{weight}(x) \leq W, \text{value}(x) = k, x \in \{0,1\}^j 0^{n-j}\}$ we require

$$m_{j,k} = \inf W_{j,k}.$$

(Recall that for any finite set S of integers $\inf S = \min S$ if $S \neq \emptyset$ and $\inf S = \infty$, otherwise.)

Let C be any upper bound on the optimum profit, for example $C = \sum_i c_i$. Clearly, the value of an optimum solution for KNAPSACK is the largest value $k \in \{0, \ldots, C\}$ such that $m_{n,k} < \infty$. The algorithm DYNAMIC PROGRAMMING KNAPSACK recursively computes the values for $m_{j,k}$ and then returns the optimum value for the given KNAPSACK instance. In the algorithm below, the variables $x(j,k)$ are n-dimensional vectors that store the solutions corresponding to $m_{j,k}$, i.e., with weight equal to $m_{j,k}$ and value k.

Algorithm 6.2 DYNAMIC PROGRAMMING KNAPSACK

Input. Integers W, C, vectors $w, c \in \mathbb{N}^n$.

Output. Vector $x \in \{0,1\}^n$ such that $\text{weight}(x) \leq W$.

Step 1. Set $m_{0,0} = 0$, $m_{0,k} = \infty$ for $k = 1, \ldots, C$, and $x(0,0) = 0$.

Step 2. For $j = 1, \ldots, n$ and $k = 0, \ldots, C$ do

$$m_{j,k} = \begin{cases} m_{j-1,k-c_j} + w_j & \text{if } c_j \leq k \text{ and } m_{j-1,k-c_j} + w_j \leq \min\{W, m_{j-1,k}\}, \\ m_{j-1,k} & \text{otherwise.} \end{cases}$$

If the first case applied set $x(j,k)_i = x(j-1, k-c_j)_i$ for $i \neq j$ and $x(j,k)_j = 1$. Otherwise set $x(j,k) = x(j-1,k)$.

Step 3. Determine the largest $k \in \{0, \ldots, C\}$ such that $m_{n,k} < \infty$. Return $x(n,k)$.

Theorem 6.5. *The* DYNAMIC PROGRAMMING KNAPSACK *algorithm computes the optimum value of the* KNAPSACK *instance W, $w, c \in \mathbb{N}^n$ in time $O(nC)$, where C is an arbitrary upper bound on this optimum value.*

Proof. The running time is obvious. For the correctness we prove that the values $m_{j,k}$ computed by the algorithm satisfy

$$m_{j,k} = \inf W_{j,k}$$

by induction on j. Here $W_{j,k} = \{\text{weight}(x) : \text{weight}(x) \leq W, \text{value}(x) = k, x \in \{0,1\}^j 0^{n-j}\}$ by definition.

The base case $j = 0$ is clear. For the inductive case first consider a situation when the algorithm sets

$$m_{j,k} = m_{j-1,k-c_j} + w_j,$$

i.e. we "take" the j-th item. Let $y = x(j-1, k-c_j)$ be the solution that corresponds to $m_{j-1,k-c_j}$. The solution $x = x(j,k)$ that corresponds to $m_{j,k}$ is obtained from y by setting $x_i = y_i$ for $i \neq j$ and $x_j = 1$. The value of x is $\text{value}(x) = k$. By definition of the algorithm we have $\text{weight}(x) = \text{weight}(y) + w_j = m_{j-1,k-c_j} + w_j \leq W$ and thus $x \in W_{j,k}$.

By construction of the algorithm and induction hypothesis we have $\text{weight}(x) \leq \inf W_{j-1,k}$ and $\text{weight}(x) = w_j + \inf W_{j-1,k-c_j}$. That is, the weight of x is at most the weight of any solution without the j-th item and at most the weight of any solution including the j-th item. Hence $m_{j,k} = \inf W_{j,k}$.

In the other situation, when the algorithm sets

$$m_{j,k} = m_{j-1,k},$$

then either $c_j > k$ and hence no solution with value equal to k can contain the j-th item, or $m_{j-1,k} + w_j > W$, i.e., adding the j-th item is infeasible, or $m_{j-1,k} + w_j > \inf W_{j-1,k}$, i.e., there is a solution with less weight and still value equal to k. □

6.3 Fully Polynomial-Time Approximation Scheme

Here we give a *fully polynomial time approximation scheme* (FPTAS), i.e., we show that for every fixed $\varepsilon > 0$ there is an $1 - \varepsilon$-approximation algorithm that runs in time polynomial in the input size and $1/\varepsilon$. From a complexity-theoretic point of view this is the best that can be hoped for: Assuming $\mathsf{P} \neq \mathsf{NP}$ there is no polynomial time algorithm that solves Knapsack optimally on every instance, but the FPTAS delivers solutions with arbitrarily good approximation guarantees in polynomial time. (Unfortunately not many problems admit an FPTAS.)

A common theme in constructing FPTASs is the following: First find an algorithm that solves the problem exactly (mostly using the dynamic programming paradigm). This algorithm usually has pseudo-polynomial or even exponential running time. Second construct an algorithm for "rounding" input-instances, i.e., reducing the input-size. This modification reduces the running time but may lead to inaccurate solutions.

The running time of Dynamic Programming Knapsack is $O(nC)$. If we divide we profit c_j of each item by a number t and round the result down, then this improves the running time of Dynamic Programming Knapsack by a factor of t to $O(nC/t)$ but may yield suboptimal solutions.

Algorithm 6.3 Knapsack FPTAS

Input. Integer W, vectors $w, c \in \mathbb{N}^n$, a number $\varepsilon > 0$.

Output. Vector $x \in \{0,1\}^n$ such that $\text{weight}(x) \leq W$.

Step 1. Run Greedy on the instance W, w, c and let x be the solution. If $\text{value}(x) = 0$ then return x.

Step 2. Set $t = \max\{1, \varepsilon \text{value}(x)/n\}$ and set

$$c'_j = \left\lfloor \frac{c_j}{t} \right\rfloor \quad \text{for } j = 1, \ldots, n.$$

Step 3. Set $C = 2\text{value}(x)/t$ and apply the Dynamic Programming Knapsack algorithm on the instance W, C, w, c' and let y be the solution obtained.

Step 4. If $\text{value}(x) \geq \text{value}(y)$ return x otherwise y.

Theorem 6.6. *For every fixed $\varepsilon > 0$, the* Knapsack FPTAS *algorithm is a $1-\varepsilon$-approximation algorithm with running time $O\left(n^2/\varepsilon\right)$.*

Proof. The value of the solution returned by the algorithm is equal to $\max\{\text{value}(x), \text{value}(y)\}$. Let x^* be an optimum solution for the instance W, w, c. By Theorem 6.4 we have $2\text{value}(x) \geq \text{value}(x^*)$ and hence the choice $C = 2\text{value}(x)/t$ is a legal upper bound for the optimum value of the rounded instance W, w, c'. By Theorem 6.5 y is an optimum solution for this instance and we have

$$\begin{aligned}\text{value}(y) &= \sum_{j=1}^{n} c_j y_j \geq \sum_{j=1}^{n} t c'_j y_j = t \sum_{j=1}^{n} c'_j y_j \\ &\geq t \sum_{j=1}^{n} c'_j x^*_j = \sum_{j=1}^{n} t c'_j x^*_j > \sum_{j=1}^{n} (c_j - t) x^*_j \geq \text{value}(x^*) - nt.\end{aligned}$$

If $t = 1$ then y is optimal by Theorem 6.5. Otherwise the above inequality and the choice of t yields $\text{value}(y) \geq \text{value}(x^*) - \varepsilon\text{value}(x)$ and hence

$$\text{value}(x^*) \leq \text{value}(y) + \varepsilon\text{value}(x) \leq (1+\varepsilon)\max\{\text{value}(x), \text{value}(y)\}$$

which yields the approximation guarantee $1 - \varepsilon/(1+\varepsilon)$.

The running time of Dynamic Programming Knapsack on the rounded instance is

$$O(nC) = O\left(\frac{n\text{value}(x)}{t}\right) = O\left(\frac{n^2}{\varepsilon}\right),$$

where we have used the definition of t: If $t = 1$ then $\text{value}(x) \leq n/\varepsilon$ and otherwise $t = \varepsilon\text{value}(x)/n$. This running time dominates the time needed for the other steps. □

Chapter 7

Bin Packing

Here we consider the classical Bin Packing problem: We are given a set $I = \{1, \dots, n\}$ of *items*, where item $i \in I$ has *size* $s_i \in (0, 1]$ and a set $B = \{1, \dots, n\}$ of *bins* with *capacity* one. Find an assignment $a : I \to B$ such that the number of non-empty bins is minimal. As a shorthand, we write $s(J) = \sum_{j \in J} s_j$ for any $J \subseteq I$.

7.1 Hardness of Approximation

The Bin Packing problem is NP-complete. More specifically:

Theorem 7.1. *It is* NP*-complete to decide if an instance of* Bin Packing *admits a solution with two bins.*

Proof. We reduce from Partition, which we know is NP-complete. Recall that in the Partition problem, we are given n numbers $c_1, \dots, c_n \in \mathbb{N}$ and are asked to decide if there is a set $S \subseteq \{1, \dots, n\}$ such that $\sum_{i \in S} c_i = \sum_{i \notin S} c_i$. Given a Partition instance, we create an instance for Bin Packing by setting $s_i = 2c_i / (\sum_{j=1}^{n} c_j) \in (0, 1]$ for $i = 1, \dots, n$. Obviously two bins suffice if and only if there is a $S \subseteq \{1, \dots, n\}$ such that $\sum_{i \in S} c_i = \sum_{i \notin S} c_i$. □

This allows us to derive a lower bound on the approximabilty of Bin Packing.

Corollary 7.2. *There is no ρ-approximation algorithm with $\rho < 3/2$ for* Bin Packing *unless* P = NP*.*

7.2 Heuristics

We will show that there are constant factor approximations for Bin Packing. Firstly we consider the probably most simple Next Fit algorithm, which can be shown to be 2-approximate. Secondly, we give the First Fit Decreasing algorithm and show that it is 3/2-approximate. Thus, with the above hardness result, this is best-possible, unless P = NP.

Next Fit

The Next Fit algorithm works as follows: Initially all bins are empty and we start with bin $j = 1$ and item $i = 1$. If bin j has residual capacity for item i, assign item i to bin j, i.e., $a(i) = j$, and consider item $i + 1$. Otherwise consider bin $j + 1$ and item i. Repeat until item n is assigned.

Theorem 7.3. Next Fit *is a* 2*-approximation for* Bin Packing. *The algorithm runs in* $O(n)$ *time.*

Proof. Let k be the number of non-empty bins in the assignment a found by Next Fit. Let k^* be the optimal number of bins. We show the slightly stronger statement that

$$k \leq 2 \cdot k^* - 1.$$

Firstly we observe the lower bound $k^* \geq \lceil s(I) \rceil$. Secondly, for bins $j = 1, \ldots, \lfloor k/2 \rfloor$ we have

$$\sum_{i:a(i) \in \{2j-1, 2j\}} s_i > 1.$$

Adding these inequalities we get

$$\left\lfloor \frac{k}{2} \right\rfloor < s(I).$$

Since the left hand side is an integer we have that

$$\frac{k-1}{2} \leq \left\lfloor \frac{k}{2} \right\rfloor \leq \lceil s(I) \rceil - 1.$$

This proves $k \leq 2 \cdot \lceil s(I) \rceil - 1 \leq 2 \cdot k^* - 1$ and hence the claim. □

The analysis is tight for the algorithm, which can be seen with the following instance with $2n$ items. For some $\varepsilon > 0$ let $s_{2i-1} = 2 \cdot \varepsilon$, $s_{2i} = 1 - \varepsilon$ for $i = 1, \ldots, n$.

First Fit Decreasing

The algorithm Next Fit never considers bins again that have been left behind. Thus the wasted capacity therein leaves room for improvement. A natural way is First Fit: Initially all bins are empty and we start with current number of bins $k = 0$ and item $i = 1$. Consider all bins $j = 1, \ldots, k$ and place item i in the first bin that has sufficient residual capacity, i.e., $a(i) = j$. If there is no such bin increment k and repeat until item n is assigned. One can prove that First Fit uses at most $k \leq \lceil 17/10 \cdot k^* \rceil$ many bins, where k^* is the optimal number.

There is a further natural heuristic improvement of First Fit, called First Fit Decreasing: Reorder the items such that $s_1 \geq \cdots \geq s_n$ and apply First Fit. The intuition behind considering large items first is the following: "Large" items do not fit into the same bin anyway, so we already use unavoidable bins and try to place "small" items into the residual space.

Theorem 7.4. First Fit Decreasing *is a* $3/2$*-approximation for* Bin Packing. *The algorithm runs in* $O\left(n^2\right)$ *time.*

Proof. Let k be the number of non-empty bins of the assignment a found by First Fit Decreasing and let k^* be the optimal number.

Consider bin number $j = \lceil 2/3k \rceil$. If it contains an item i with $s_i > 1/2$, then each bin $j' < j$ did not have space for item i. Thus j' was assigned an item i' with $i' < i$. As the items are considered in non-increasing order of size we have $s_{i'} \geq s_i > 1/2$. That is, there are at least j items of size larger than $1/2$. These items need to be placed in individual bins. This implies

$$k^* \geq j \geq \frac{2}{3}k.$$

Otherwise, bin j and any bin $j' > j$ does not contain an item with size larger than $1/2$. Hence the bins $j, j+1, \ldots, k$ contain at least $2(k-j)+1$ items, none of which fits into the bins $1, \ldots, j-1$. Thus we have

$$\begin{aligned} s(I) &> \min\{j-1, 2(k-j)+1\} \\ &\geq \min\{\lceil 2/3k \rceil - 1, 2(k-(2/3k+2/3))+1\} \\ &= \lceil 2/3k \rceil - 1 \end{aligned}$$

and $k^* \geq s(I) > \lceil 2/3k \rceil - 1$. This even implies

$$k^* \geq \left\lceil \frac{2}{3}k \right\rceil \geq \frac{2}{3}k$$

and hence the claim. □

7.3 Asymptotic Polynomial Time Approximation Scheme

With the hardness result that there is no approximation algorithm for Bin Packing with guarantee better than $3/2$, unless $\mathsf{P} = \mathsf{NP}$, we do not have to search for a PTAS (or even an FPTAS). However, notice that the reduction used that the optimal number of bins is "small", such as 2 or 3. It is plausible that, in "practical" instances, the optimal number k^* of bins grows as the number of items grows. Maybe we can do better for those instances.

This leads us to define: An *asymptotic polynomial time approximation scheme* (APTAS) is a familiy of algorithms, such that for any $\varepsilon > 0$ there is a number k' and a $(1+\varepsilon)$-approximation algorithm, whenever $k^* \geq k'$. For Bin Packing such a family exists. However, the involved running times are rather high, even though polynomial in n.

Theorem 7.5. *For any $0 < \varepsilon \leq 1/2$ there is an algorithm that runs in time polynomial in n and finds an assignment having at most $k \leq (1+\varepsilon) \cdot k^* + 1$ many bins.*

Lemma 7.6. *Let $\varepsilon > 0$ and $d \in \mathbb{N}$ be constants. For any instance of* Bin Packing *where $s_i \geq \varepsilon$ and $|\{s_1, \ldots, s_n\}| \leq d$, there is a polynomial time algorithm that solves it optimally.*

Proof. The number of items in a bin is bounded by $m := \lfloor 1/\varepsilon \rfloor$. Therefore, the number of different assignments for one bin is bounded by $r = \binom{m+d}{m}$, which is a (large) constant. There are at most n bins used and therefore, the number of feasible assignments is bounded by $p = \binom{n+r}{r}$. This is a polynomial in n. Thus we can enumerate all assignments and choose the best one to give an optimum solution. □

Lemma 7.7. *Let $\varepsilon > 0$ be a constant. For any instance of* Bin Packing *where $s_i \geq \varepsilon$, there is a $(1+\varepsilon)$-approximation algorithm.*

Proof. Let I be the given instance. Sort the n items by increasing size and partition them into $g = \lceil 1/\varepsilon^2 \rceil$ many groups each having at most $q = \lfloor n\varepsilon^2 \rfloor$ many items. Notice that two groups may contain items of the same size.

Construct an instance J by rounding up the size of each item to the size of the largest item in its group. Instance J has at most g many different item sizes. Therefore, we can find an optimal assignment for J by invoking Lemma 7.6. This is clearly a feasible assignment for the original item sizes.

Now we show that $k^*(J) \leq (1+\varepsilon)k^*(I)$: We construct another instance J' by rounding down the size of each item to the smallest item size in its group. Clearly $k^*(J') \leq k^*(J)$.

The crucial observation is that an assignment for instance J' yields an assignment for all but the largest q items of the instance J. Therefore

$$k^*(J) \leq k^*(J') + q \leq k^*(I) + q.$$

To finalize the proof, since each item has size at least ε, we have $k^*(I) \geq n \cdot \varepsilon$ and $q = \lfloor n\varepsilon^2 \rfloor \leq \varepsilon \cdot k^*(I)$. Hence

$$k^*(J) \leq (1 + \varepsilon) \cdot k^*(I)$$

and the claim is established. □

Proof of Theorem 7.5. Let I denote the given instance and I' the instance after discarding the items with size less than ε from I. We can invoke Lemma 7.7 and find an assignment which uses at most $k(I') \leq (1 + \varepsilon) \cdot k^*(I')$ many bins. By using First Fit, we assign the items with sizes less than ε into the solution found for instance I'. We use additional bins if an item does not fit into any of the bins used so far.

If no additional bins are needed, then our assignment uses $k(I) \leq (1 + \varepsilon) \cdot k^*(I') \leq (1 + \varepsilon) \cdot k^*(I)$ many bins. Otherwise, all but the last bin have residual capacity less than ε. Thus $s(I) \geq (k(I) - 1)(1 - \varepsilon)$, which is a lower bound for $k^*(I)$. Thus we have

$$k(I) \leq \frac{k^*(I)}{1 - \varepsilon} + 1 \leq (1 + 2\varepsilon) \cdot k^*(I) + 1,$$

where we have used $0 < \varepsilon \leq 1/2$. □

Chapter 8

Set Cover

The Set Cover problem this chapter deals with is again a very simple to state – yet quite general – NP-hard combinatorial problem. It is widely applicable in sometimes unexpected ways. The problem is the following: We are given a set U (called *universe*) of n elements, a collection of sets $\mathcal{S} = \{S_1, \ldots, S_k\}$ where $S_i \subseteq U$, and a cost function $c : \mathcal{S} \to \mathbb{R}^+$. The task is to find a minimum cost subcollection $\mathcal{S}' \subseteq \mathcal{S}$ that *covers* U, i.e., such that $\cup_{S \in \mathcal{S}'} S = U$.

Example 8.1. Consider this instance: $U = \{1, 2, 3\}$, $\mathcal{S} = \{S_1, S_2, S_3\}$ with $S_1 = \{1, 2\}$, $S_2 = \{2, 3\}$, $S_3 = \{1, 2, 3\}$ and cost $c(S_1) = 10$, $c(S_2) = 50$, and $c(S_3) = 100$. These collections cover U: $\{S_1, S_2\}$, $\{S_3\}$, $\{S_1, S_3\}$, $\{S_2, S_3\}$, $\{S_1, S_2, S_3\}$. The cheapest one is $\{S_1, S_2\}$ with cost equal to 60.

For each set S, we associate a variable $x_S \in \{0, 1\}$ that indicates of we want to choose S or not. We may thus write solutions for Set Cover as a vector $x \in \{0, 1\}^k$. With this, we write Set Cover as a mathematical program.

Problem 8.1 Set Cover

Instance. Universe U with n elements, collection $\mathcal{S} = \{S_1, \ldots, S_k\}$, $S_i \subseteq U$, a cost function $c : \mathcal{S} \to \mathbb{R}$.

Task. Solve the problem

$$\begin{aligned} \text{minimize} \quad & \text{value}(x) = \sum_{S \in \mathcal{S}} c(S) x_S, \\ \text{subject to} \quad & \sum_{S : e \in S} x_S \geq 1 \quad e \in U, \\ & x_S \in \{0, 1\} \quad S \in \mathcal{S}. \end{aligned}$$

Define the *frequency* of an element to be the number of sets it is contained in. Let f denote the frequency of the most frequent element. In this chapter we present several algorithms that either achieve approximation ratio $O(\log n)$ or f. Why are we interested in a variety algorithms? Is one algorithm not sufficient? Yes, but here the focus is on the *techniques* that yield these algorithms.

8.1 Greedy Algorithm

The GREEDY algorithm follows the natural approach of iteratively choosing the most cost-effective set and remove all the covered elements until all elements are covered. Let C be the set of elements already covered at the beginning of an iteration. During this iteration define the *cost-effectiveness* of a set S as $c(S)/|S - C|$, i.e., the average cost at which it covers new elements. For later reference, the algorithm sets the *price* at which it covered an element equal to the cost-effectiveness of the covering set. Further recall that $H_n = \sum_{i=1}^{n} 1/i$ is called the *n-th Harmonic number* and that $\log n \leq H_n \leq \log n + 1$.

Algorithm 8.1 GREEDY

Input. Universe U with n elements, collection $\mathcal{S} = \{S_1, \ldots, S_k\}$, $S_i \subseteq U$, a cost function $c : \mathcal{S} \to \mathbb{R}$.

Output. Vector $x \in \{0, 1\}^k$

Step 1. $C = \emptyset$, $x = 0$.

Step 2. While $C \neq U$ do the following:

(a) Find the most cost-effective set in the current iteration, say S.

(b) Set $x_S = 1$ and for each $e \in S - C$ set $\text{price}(e) = c(S)/|S - C|$.

(c) $C = C \cup S$.

Step 3. Return x.

Theorem 8.2. *The* GREEDY *algorithm is an H_n-approximation algorithm for the* SET COVER *problem.*

It is an exercise to show that this bound is tight.

Direct Analysis

The following lemma is crucial for the proof of the approximation-guarantee. Number the elements of U in the order in which they were covered by the algorithm, say $e_1, \ldots, e_n$. Let x^* be an optimum solution.

Lemma 8.3. *For each $i \in \{1, \ldots, n\}$,* $\text{price}(e_i) \leq \text{value}(x^*)/(n - i + 1)$.

Proof. In any iteration, the leftover sets of the optimal solution x^* can cover the remaining elements at a cost of at most $\text{value}(x^*)$. Therefore, among these, there must be one set having cost-effectiveness of at most $\text{value}(x^*)/|U - C|$. In the iteration in which element e_i was covered, $U - C$ contained at least $n - i + 1$ elements. Since e_i was covered by the most cost-effective set in this iteration, we have that

$$\text{price}(e_i) \leq \frac{\text{value}(x^*)}{|U - C|} \leq \frac{\text{value}(x^*)}{n - i + 1}$$

which was claimed. □

Proof of Theorem 8.2. Since the cost of each set is distributed evenly among the new elements covered, the total cost of the set cover picked is

$$\text{value}(x) = \sum_{i=1}^{n} \text{price}(e_i) \leq \text{value}(x^*)H_n,$$

where we have used Lemma 8.3. □

Dual-Fitting Analysis

Here we will give an alternative analysis of the GREEDY algorithm for SET COVER. We will use the *dual fitting* method, which is quite general and helps to analyze a broad variety of combinatorial algorithms.

For sake of exposition we consider a minimization problem, but the technique works similarly for maximization. Consider an algorithm ALG which does the following:

(1) Let (P) be an integer programming formulation of the problem of interest. We are interested in its optimal solution x^*, respectively its objective value $\text{value}(x^*)$. Let (D) be the dual of a linear programming relaxation of (P).

(2) The algorithm ALG computes a feasible solution x for (P) and a "solution" y for (D), where we allow that y is *infeasible* for (D). But the algorithm has to ensure that

$$\text{value}(x) \leq \overline{\text{value}}(y),$$

where value is the objective function of (P) and $\overline{\text{value}}$ is the objective function of (D).

(3) Now divide the entries of y by a certain quantity α until $y' = y/\alpha$ *is* feasible for (D). (The method of dual fitting is applicable only if this property can be ensured.) Then $\overline{\text{value}}(y')$ is a lower bound for $\text{value}(x^*)$ by weak duality, i.e.,

$$\overline{\text{value}}(y') \leq \text{value}(x^*)$$

by Lemma 5.8.

(4) Putting these things together, we obtain the approximation guarantee of α by

$$\text{value}(x) \leq \overline{\text{value}}(y) = \overline{\text{value}}(\alpha y') = \alpha\overline{\text{value}}(y') \leq \alpha\text{value}(x^*).$$

Now we apply this recipe to SET COVER and consider the GREEDY algorithm. For property (1) we use our usual formulation

$$\begin{aligned} \text{minimize} \quad & \sum_{S \in \mathcal{S}} c(S)x_S, \\ \text{subject to} \quad & \sum_{S: e \in S} x_S \geq 1 \quad e \in U, \\ & x_S \in \{0,1\} \quad S \in \mathcal{S}. \end{aligned} \tag{P}$$

When we relax the constraints $x_S \in \{0,1\}$ to $0 \leq x_S \leq 1$ and dualize the corresponding linear program we find

$$\begin{aligned} \text{maximize} \quad & \sum_{e \in U} y_e, \\ \text{subject to} \quad & \sum_{e \in S} y_e \leq c(S) \quad S \in \mathcal{S}, \\ & y_e \geq 0. \end{aligned} \tag{D}$$

This dual can be derived purely mechanically (by applying the primal-dual-definition and rewriting constraints if needed), but this program also has an intuitive interpretation. The constraints of (D) state that we want to "pack stuff" into each set S such that the cost $c(S)$ of each set is not exceeded, i.e., the sets are not overpacked. We seek to maximize the total amount packed.

How about property (2)? The algorithm Greedy computes a certain feasible solution x for (P), i.e., a solution $x_S = 1$ if the algorithm picks set S and $x_S = 0$ otherwise. What about the vector y? Define the following vector: For each $e \in U$ set $y_e = \text{price}(e)$, where $\text{price}(e)$ is the value computed during the execution of the algorithm.

By construction of the algorithm we have

$$\text{value}(x) = \sum_{S \in \mathcal{S}} c(S) x_S = \sum_{e \in U} \text{price}(e) = \sum_{e \in U} y_e = \overline{\text{value}}(y),$$

i.e., Greedy satisfies property (2) of the dual fitting method (even with equality).

For property (3) the following result is useful.

Lemma 8.4. *For every $S \in \mathcal{S}$ we have that*

$$\sum_{e \in S} y_e \leq H_n c(S).$$

Proof. Let $S \in \mathcal{S}$ with, say, m elements. Consider these in the ordering the algorithm covered them, say, $e_1, \ldots, e_m$. At the iteration when e_i gets covered S contains $m - i + 1$ uncovered elements. Since Greedy chooses the most cost-effective set we have that

$$\text{price}(e_i) \leq \frac{c(S)}{m - i + 1},$$

i.e., the cost-effectiveness of the set the algorithm chooses can only be smaller than the cost-effectiveness of S. (Be aware that "smaller" is "better" here.)

Summing over all elements gives

$$\sum_{e \in S} y_e = \sum_{i=1}^{m} \text{price}(e_i) \leq c(S) \sum_{i=1}^{m} \frac{1}{m - i + 1} = c(S) H_m \leq c(S) H_n$$

as claimed. □

Now we are in position to finalize the dual-fitting analysis using property (4).

Proof of Theorem 8.2. Define the vector $y' = y/H_n$, where y is defined above. Observe that for each set $S \in \mathcal{S}$ we have

$$\sum_{e \in S} y'_e = \sum_{e \in S} \frac{y_e}{H_n} = \frac{1}{H_n} \sum_{e \in S} y_e \leq c(S)$$

using Lemma 8.4. That means y' is feasible for (D). Using the property (4) of the dual fitting method proves the approximation guarantee of at most H_n. □

8.2 Primal-Dual Algorithm

The primal-dual schema introduced here is the method of choice for designing approximation algorithms because it often gives algorithms with good approximation guarantees and good running times. After introducing the ideas behind the method, we will use it to design a simple factor f algorithm, where f is the frequency of the most frequent element.

The general idea is to work with an LP-relaxation of an NP-hard problem and its dual. Then the algorithm iteratively changes a primal and a dual solution until the relaxed primal-dual complementary slackness conditions are satisfied.

Primal-Dual Schema

Consider the following primal program:

$$\begin{aligned} \text{minimize} \quad & \text{value}(x) = \sum_{j=1}^{n} c_j x_j, \\ \text{subject to} \quad & \sum_{j=1}^{n} a_{ij} x_j \geq b_i \quad i = 1, \ldots, m, \\ & x_j \geq 0 \quad j = 1, \ldots, n. \end{aligned}$$

The dual program is:

$$\begin{aligned} \text{maximize} \quad & \overline{\text{value}}(y) = \sum_{i=1}^{m} b_i y_i, \\ \text{subject to} \quad & \sum_{i=1}^{m} a_{ij} y_i \leq c_j \quad j = 1, \ldots, n, \\ & y_i \geq 0 \quad i = 1, \ldots, m. \end{aligned}$$

Most known approximation algorithms using the primal-dual schema run by ensuring one set of conditions and suitably relaxing the other. We will capture both situations by relaxing both conditions. If primal conditions are to be ensured, we set $\alpha = 1$ below, and if dual conditions are to be ensured, we set $\beta = 1$.

Primal Complementary Slackness Conditions. Let $\alpha \geq 1$. For each $1 \leq j \leq n$:

$$\text{either} \quad x_j = 0 \quad \text{or} \quad c_j/\alpha \leq \sum_{i=1}^{m} a_{ij} y_i \leq c_j.$$

Dual Complementary Slackness Conditions. Let $\beta \geq 1$. For each $1 \leq i \leq m$:

$$\text{either} \quad y_i = 0 \quad \text{or} \quad b_i \leq \sum_{j=1}^{n} a_{ij} x_j \leq \beta b_i.$$

Lemma 8.5. *If x and y are primal and dual feasible solutions respectively satisfying the complementary slackness conditions stated above, then*

$$\text{value}(x) \leq \alpha\beta\overline{\text{value}}(y).$$

Proof. We calculate directly using the slackness conditions and obtain

$$\begin{aligned}\text{value}(x) &= \sum_{j=1}^{n} c_j x_j \leq \alpha \sum_{j=1}^{n} \left(\sum_{i=1}^{m} a_{ij} y_i \right) x_j \\ &= \alpha \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j \right) y_i \leq \alpha\beta \sum_{i=1}^{m} b_i y_i = \overline{\text{value}}(y)\end{aligned}$$

which was claimed. □

The algorithm starts with a primal infeasible solution and a dual feasible solution; usually these are $x = 0$ and $y = 0$ initially. It iteratively improves the feasibility of the primal solution and the optimality of the dual solution ensuring that in the end a primal feasible solution is obtained and all conditions stated above, with a suitable choice for α and β, are satisfied. The primal solution is always extended integrally, thus ensuring that the final solution is integral. The improvements to the primal and the dual go hand-in-hand: the current primal solution is used to determine the improvement to the dual, and vice versa. Finally, the cost of the dual solution is used as a lower bound on the optimum value, and by Lemma 8.5, the approximation guarantee of the algorithm is $\alpha\beta$.

Primal-Dual Algorithm

Here we derive a factor f approximation algorithm for Set Cover using the primal-dual schema. For this algorithm we will choose $\alpha = 1$ and $\beta = f$. We will work with the following primal LP for Set Cover

$$\begin{aligned} \text{minimize} \quad & \text{value}(x) = \sum_{S \in \mathcal{S}} c(S) x_S, \\ \text{subject to} \quad & \sum_{S: e \in S} x_S \geq 1 \quad e \in U, \\ & x_S \geq 0 \quad S \in \mathcal{S}. \end{aligned}$$

and its dual

$$\begin{aligned} \text{maximize} \quad & \overline{\text{value}}(y) = \sum_{e \in U} y_e, \\ \text{subject to} \quad & \sum_{e \in S} y_e \leq c(S) \quad S \in \mathcal{S}, \\ & y_e \geq 0 \quad e \in U. \end{aligned}$$

For these LPs the primal and dual complementary slackness conditions are:

Primal Complementary Slackness Conditions. For each $S \in \mathcal{S}$:

$$\text{either} \quad x_S = 0 \quad \text{or} \quad \sum_{e \in S} y_e = c(S).$$

A set S will be said to be *tight* if $\sum_{e \in S} y_e = c(S)$. So, this condition states that: "Pick only tight sets into the cover."

Dual Complementary Slackness Conditions. For each $e \in U$:

$$\text{either} \quad y_e = 0 \quad \text{or} \quad \sum_{S:e \in S} x_S \leq f.$$

Since we will find a 0/1 solution for x, these conditions are equivalent to: "Each element having non-zero dual value can be covered at most f times." Since each element is in at most f sets, this condition is trivially satisfied for all elements.

These conditions suggest the following algorithm:

Algorithm 8.2 PRIMAL-DUAL SET COVER

Input. Universe U with n elements, collection $\mathcal{S} = \{S_1, \ldots, S_k\}$, $S_i \subseteq U$, a cost function $c : \mathcal{S} \to \mathbb{R}$.

Output. Vector $x \in \{0,1\}^k$

Step 1. $x = 0$, $y = 0$. Declare all elements uncovered.

Step 2. Unless all elements are covered, do:

(a) Pick an uncovered element, say e, and raise y_e until some set goes tight.

(b) Pick all tight sets S in the cover, i.e., set $x_S = 1$.

(c) Declare all the elements occuring in these sets as covered.

Step 3. Return x.

Theorem 8.6. *The algorithm* PRIMAL-DUAL SET COVER *is a f-approximation algorithm for* SET COVER.

Proof. At the end of the algorithm, there will be no uncovered elements. Further no dual constraint is violated since we pick only tight sets S into the cover and no element $e \in S$ will later on be a candidate for increasing y_e. Thus, the primal and dual solutions will both be feasible. Since they satisfy the primal and dual complementary slackness conditions with $\alpha = 1$ and $\beta = f$, by Lemma 8.5, the approximation guarantee is f. □

Example 8.7. A tight example for this algorithm is provided by the following set system. The universe is $U = \{e_1, \ldots, e_{n+1}\}$ and $\mathcal{S}$ consists of $n-1$ sets $\{e_1, e_n\}, \ldots, \{e_{n-1}, e_n\}$ of cost 1 and one set $\{e_1, \ldots, e_{n+1}\}$ of cost $1 + \varepsilon$ for some small $\varepsilon > 0$. Since e_n appears in all n sets, this system has $f = n$.

Suppose the algorithm raises y_{e_n} in the first iteration. When y_{e_n} is raised to 1, all sets $\{e_i, e_n\}$, $i = 1, \ldots, n-1$ go tight. They are all picked in the cover, thus covering the elements $e_1, \ldots, e_n$. In the second iteration $y_{e_{n+1}}$ is raised to ε and the set $\{e_1, \ldots, e_{n+1}\}$ goes tight. The resulting set cover has cost $n + \varepsilon$, whereas the optimum cover has cost $1 + \varepsilon$.

8.3 LP-Rounding Algorithms

The central idea behind algorithms that make use of the *LP-rounding* technique is as follows: Suppose you have an LP-relaxation of a certain NP-hard problem. Then you can solve this optimally and try to "round" the optimal fractional solution to an integral one.

Here we derive a factor f approximation algorithm for Set Cover but this time by rounding the fractional solution of an LP to an integral solution (instead of the primal-dual schema). We consider our usual LP relaxation for Set Cover

$$\begin{aligned} \text{minimize} \quad & \text{value}(x) = \sum_{s \in \mathcal{S}} c(S) x_S, \\ \text{subject to} \quad & \sum_{S: e \in S} x_S \geq 1 \quad e \in U, \\ & x_S \geq 0 \quad S \in \mathcal{S}. \end{aligned}$$

Simple Rounding Algorithm

The idea of the algorithm below is to include those sets S into the cover for which the corresponding value z_S in the optimal solution z of the LP is "large enough".

Algorithm 8.3 Simple Rounding Set Cover

Input. Universe U with n elements, collection $\mathcal{S} = \{S_1, \ldots, S_k\}$, $S_i \subseteq U$, a cost function $c : \mathcal{S} \to \mathbb{R}$.

Output. Vector $x \in \{0, 1\}^k$

Step 1. Set $x = 0$, solve the LP relaxation below, and call the optimal solution z.

$$\begin{aligned} \text{minimize} \quad & \text{value}(x) = \sum_{S \in \mathcal{S}} c(S) x_S, \\ \text{subject to} \quad & \sum_{S: e \in S} x_S \geq 1 \quad e \in U, \\ & x_S \geq 0 \quad S \in \mathcal{S}. \end{aligned}$$

Step 2. For each set S set $x_S = 1$ if $z_S \geq 1/f$.

Step 3. Return x.

Theorem 8.8. *The algorithm* Simple Rounding Set Cover *is an f-approximation algorithm for* Set Cover.

Proof. Let x be the solution returned by the algorithm and z be the optimal solution of the LP. Consider an arbitrary element $e \in U$. Since e is in at most f sets, one of these sets must be picked to the extent of at least $1/f$ in the fractional solution z. If this were not the case then $\sum_{S: e \in S} z_S < \sum_{S: e \in S} 1/f \leq f \cdot 1/f = 1$ yields a contradiction to the feasibility of z. Thus e is covered due to the definition of the algorithm and x is hence a feasible cover. We further have $x_S \leq f z_S$ and thus

$$\text{value}(x) \leq f \text{value}(z) \leq f \text{value}(x^*)$$

where x^* is an optimal solution for the Set Cover problem. □

Randomized Rounding

Another natural idea for rounding fractional solutions is to use randomization: For example, for the above relaxation, observe that the values z_S are between zero and one. We may thus interpret these values as probabilities for choosing a certain set S.

Here is the idea of the following algorithm: Solve the LP-relaxation optimally and call the solution z. With probability z_S include the set S into the cover.

This basic procedure yields a vector x with expected value equal to the optimal fractional solution value but might not cover all the elements. We thus repeat the procedure "sufficiently many" times and include a set into our cover if it was included in any of the iterations. We will show that $O(\log n)$ many iterations suffice yielding an $O(\log n)$-approximation algorithm.

Algorithm 8.4 Randomized Rounding Set Cover

Input. Universe U with n elements, collection $\mathcal{S} = \{S_1, \ldots, S_k\}$, $S_i \subseteq U$, a cost function $c : \mathcal{S} \to \mathbb{R}$.

Output. Vector $x \in \{0,1\}^k$

Step 1. Set $x = 0$, solve the LP relaxation below, and call the optimal solution z.

$$\begin{aligned} \text{minimize} \quad & \text{value}(x) = \sum_{S \in \mathcal{S}} c(S) x_S, \\ \text{subject to} \quad & \sum_{S : e \in S} x_S \geq 1 \quad e \in U, \\ & x_S \geq 0 \quad S \in \mathcal{S}. \end{aligned}$$

Step 2. Repeat $\lceil 3 \log n \rceil$ times: For each set S set $x_S = 1$ with probability z_S.

Step 3. Return x.

Theorem 8.9. *With probability at least $1 - 1/n^2$ the algorithm* Randomized Rounding Set Cover *returns a feasible solution, which is expected $\lceil 3 \log n \rceil$-approximate for* Set Cover.

Proof. Let z be an optimal solution for the LP. We estimate the probability that an element $e \in U$ is covered in one iteration in Step 2. Let e be contained in m sets and let $z_1, \ldots, z_m$ be the probabilities given in the solution z. Since e is fractionally covered we have $z_1 + \cdots + z_m \geq 1$. With easy but tedious calculus we see that – under this condition – the probability for e being covered is minimized when the z_i are all equal, i.e., $z_1 = \cdots = z_m = 1/m$:

$$\Pr[x_S = 1] = 1 - (1 - z_1) \cdot \cdots \cdot (1 - z_m) \geq 1 - \left(1 - \frac{1}{m}\right)^m \geq 1 - \frac{1}{e}.$$

Each element is covered with probability at least $1 - 1/e$. But maybe we have not covered all elements after $\lceil 3 \log n \rceil$ iterations. The probability that the element e is not covered at the end of the algorithm, i.e., after $\lceil 3 \log n \rceil$ iterations is

$$\Pr[e \text{ is not covered}] \leq \left(\frac{1}{e}\right)^{\lceil 3 \log n \rceil} \leq \frac{1}{n^3}.$$

Thus the probability that there is an uncovered element is at most

$$\sum_{e \in U} \Pr\left[e \text{ is not covered}\right] \leq n \cdot \frac{1}{n^3} \leq \frac{1}{n^2}.$$

Hence the retured solution x is feasible with probability at least $1 - 1/n^2$.

Consider a single iteration in Step 2 and let $y \in \{0,1\}^k$ be the vector that indicates which sets are included in this particular iteration. For each set S let $y_S = 1$ with probability z_S. Then we have

$$\mathbb{E}\left[\text{value}(y)\right] = \sum_{S \in \mathcal{S}} \mathbb{E}\left[c(S) y_S\right] = \sum_{S \in \mathcal{S}} c(S) \Pr\left[y_S = 1\right] = \sum_{S \in \mathcal{S}} c(S) z_S = \text{value}(z).$$

Now we consider all iterations in Step 2 and clearly have

$$\mathbb{E}\left[\text{value}(x)\right] \leq \lceil 3 \log n \rceil \cdot \mathbb{E}\left[\text{value}(y)\right] \leq \lceil 3 \log n \rceil \cdot \text{value}(z) \leq \lceil 3 \log n \rceil \cdot \text{value}(x^*),$$

where x^* is an optimal solution for Set Cover. So, the algorithm returns a feasible solution, with probability at least $1 - 1/n^2$, whose expected value is $\lceil 3 \log n \rceil$-approximate. □

The proof above shows that the algorithm is a $\lceil 3 \log n \rceil$-approximation in expectation. But we can actually state that the approximation ratio is $4 \cdot \lceil 3 \log n \rceil$ with probability around 3/4. Use Markov's inequality $\Pr\left[X > t\right] \leq \mathbb{E}\left[X\right] / t$ to show

$$\Pr\left[\text{value}(x) > 4 \cdot \lceil 3 \log n \rceil \cdot \text{value}(z)\right] \leq \frac{\mathbb{E}\left[\text{value}(x)\right]}{4 \cdot \lceil 3 \log n \rceil \cdot \text{value}(z)} \leq \frac{1}{4}$$

The probability that either not all elements are covered or the obtained solution has value larger than $4 \cdot \lceil 3 \log n \rceil$ times the optimal value is at most $1/n^2 + 1/4 \leq 1/2$ for all $n \geq 2$. Thus we have to run the whole algorithm at most two times in expectation to actually get a $4 \cdot \lceil 3 \log n \rceil$-approximate solution.

Chapter 9

Makespan Scheduling

In this chapter, we consider the classical Makespan Scheduling problem. We are given m machines for scheduling, indexed by the set $M = \{1, \ldots, m\}$. There are furthermore given n jobs, indexed by the set $J = \{1, \ldots, n\}$, where job j takes $p_{i,j}$ units of time if scheduled on machine i. Let J_i be the set of jobs scheduled on machine i. Then $\ell_i = \sum_{j \in J_i} p_{i,j}$ is the *load* of machine i. The maximum load $\ell_{\max} = c_{\max} = \max_{i \in M} \ell_i$ is called the *makespan* of the schedule.

The problem is NP-hard, even if there are only two identical machines. However, we will derive several constant factor approximations and a PTAS for identical machines and a 2-approximation for the general case.

9.1 Identical Machines

In the special case of *identical machines*, we have that $p_{i,j} = p_j$ for all $i \in M$ and all $j \in J$. Here p_j is called the *length* of job j.

List Scheduling

As a warm-up we consider the following two heuristics for Makespan Scheduling. The List Scheduling algorithm works as follows: Determine any ordering of the job set J, stored in a list L. Starting with all machines empty, determine the machine i with the currently least load and schedule the respective next job j in L on i. The load of i before the assignment of j is called the *starting time* s_j of job j and the load of i after the assignment is called the *completion time* c_j of job j. In the Sorted List Scheduling algorithm we execute List Scheduling, where the list L consists of the jobs in decreasing order of length.

Theorem 9.1. *The* List Scheduling *algorithm is a* 2*-approximation for* Makespan Scheduling *on identical machines.*

Proof. Let T^* be the optimal makespan of the given instance. We show that $s_j \leq T^*$ for all $j \in J$. This implies $c_j = s_j + p_j \leq T^* + p_j \leq 2 \cdot T^*$ for all $j \in J$, since we clearly must have $T^* \geq p_j$ for all $j \in J$.

Assume that $s_j > T^*$ for some $j \in J$. Then we have that the load *before* the assignment of j is $\ell_i > T^*$ for all $i \in M$. Thus the jobs $J' \subseteq J$ scheduled before j by the algorithm have total length $\sum_{j' \in J'} p_{j'} > m \cdot T^*$. On the other hand, since the optimum solution schedules all jobs J until time T^* we have $\sum_{j \in J} p_j \leq m \cdot T^*$. A contradiction and the List Scheduling algorithm must start all jobs not later than time T^*. □

Here we show that Sorted List Scheduling is a 3/2-approximation, but one can actually prove that the algorithm is a 4/3-approximation.

Theorem 9.2. *The* Sorted List Scheduling *algorithm is a* 3/2*-approximation for* Makespan Scheduling *on identical machines.*

Proof. Let T^* be the optimal makespan of the given instance. Partition the jobs $J_L = \{j \in J : p_j > T^*/2\}$ and $J_S = J - J_L$, called *large* and *small* jobs. Notice that there can be at most m large jobs: Assume that there are more than m such jobs. Then, in any schedule, including the optimal one, there must be at least two such jobs scheduled on some machine. Since the length of a large job is more than $T^*/2$, this contradicts that T^* is the optimal makespan.

Since there are at most m large jobs and the algorithm schedules those first and hence on individual machines, we have that each large job completes not later than T^*, i.e., $c_j \leq T^*$ for all $j \in J_L$. Thus, if a job completes later than T^* it must be a small job having length at most $T^*/2$. Since each job starts not later than T^* we have $c_j \leq T^* + p_j \leq 3/2 \cdot T^*$ for every small job $j \in J_S$. □

Polynomial Time Approximation Scheme

In this section we give a *polynomial time approximation scheme (PTAS)* for Makespan Scheduling on identical machines. This means, for any error parameter $\varepsilon > 0$, there is an algorithm which determines a $(1+\varepsilon)$-approximate solution with running time polynomial in the input size, but arbitary in $1/\varepsilon$. We will give a PTAS with running time $O\left(n^{2k} \cdot \lceil \log_2 1/\varepsilon \rceil\right)$, where $k = \lceil \log_{1+\varepsilon} 1/\varepsilon \rceil$.

The two main ingredients of the algorithm are these:

(1) Firstly, assume that we are given the optimal makespan T^* at the outset. Then we can try to construct a schedule with makespan at most $(1+\varepsilon) \cdot T^*$. But how do we determine the number T^*? It turns out that we can perform *binary search* in an interval $[\alpha, \beta]$, where α is any lower bound on T^* and β any upper bound on T^*. This binary search will enable us to eventually find a number B, which is within $(1+\varepsilon)$ times T^* and where the number of binary search iterations depends on the error parameter ε.

(2) Secondly, assume that the number of *distinct* values of job lengths is a constant k, say. Then we can determine all configurations of jobs that do not violate a load bound of t if scheduled on a single machine. This is the basis of a dynamic programming scheme to determine a schedule on m machines. Of course, this approach involves rounding the original job lengths to constantly many values, which introduces some error. The error can be controled by adjusting the constant k of distinct job lengths at the expense of running time and space requirement for the dynamic programming table.

Consider the instance J with jobs of lengths $p_1, \ldots, p_n$. Let t be a parameter and let $m(J,t)$ be the smallest number of machines required to schedule the jobs J having makespan at most t. With this definition, the minimum makespan T^* is given by $T^* = \min\{t : m(J,t) \leq m\}$. We will later perform binary search on the parameter t, in the interval $[\alpha, \beta]$, where $\alpha = \max\{\max_{j \in J} p_j, 1/m \cdot \sum_{j \in J} p_j\}$ and $\beta = 2 \cdot \alpha$. Notice that α is a lower bound on T^* and β an upper bound on T^*.

Dynamic Programming. Assume for now that $|\{p_1, \ldots, p_n\}| = k$, i.e., there are k distinct job lengths. Fix an ordering of the jobs lengths. Then a k-tuple $(i_1, \ldots, i_k)$ describes for any $\ell \in \{1, \ldots, k\}$ the number i_ℓ of jobs having the respective length. For any k-tuple $(i_1, \ldots, i_k)$ let $m(i_1, \ldots, i_k, t)$ be the smallest number of machines needed to schedule these jobs having makespan at most t. For a given parameter t and an instance $(n_1, \ldots, n_k)$ with $\sum_{\ell=1}^{k} n_\ell = n$, we first compute the set Q of all k-tuples $(q_1, \ldots, q_k)$ such that $m(q_1, \ldots, q_k, t) = 1$, $0 \leq q_\ell \leq n_\ell$ for $\ell = 1, \ldots, k$, i.e., all sets of jobs that can be scheduled on a single machine with makespan at most t. Clearly, Q contains at most $O\left(n^k\right)$ elements. Having these numbers computed, we determine the entries $m(i_1, \ldots, i_k, t)$ for every $(i_1, \ldots, i_k) \in \{0, \ldots, n_1\} \times \cdots \times \{0, \ldots, n_k\}$ of a k-dimensional table as follows: The table is initialized by setting $m(q, t) = 1$ for every $q \in Q$. Then we use the following recurrence to compute the remaining entries:

$$m(i_1, \ldots, i_k, t) = 1 + \min_{q \in Q} m(i_1 - q_1, \ldots, i_k - q_k, t).$$

Computing each entry takes $O\left(n^k\right)$ time. Thus the entire table can be computed in $O\left(n^{2k}\right)$ time, thereby determining $m(n_1, \ldots, n_k, t)$ in polynomial time provided that k is a constant.

Rounding. Let $\varepsilon > 0$ be an error parameter and let $t \in [\alpha, \beta]$ as defined above. We say that a job j is small if $p_j < \varepsilon \cdot t$. Small jobs are removed from the instance for now. The rest of the job lengths are rounded down as follows: If a job j has length $p_j \in [t \cdot \varepsilon \cdot (1+\varepsilon)^i, t \cdot \varepsilon \cdot (1+\varepsilon)^{i+1})$ for $i \geq 0$, it is replaced by $p'_j = t \cdot \varepsilon \cdot (1+\varepsilon)^i$. Thus there can be at most $k = \lceil \log_{1+\varepsilon} 1/\varepsilon \rceil$ many distinct job lengths. Now we invoke the above dynamic programming scheme and determine a the optimal number of machines for scheduling these jobs if the makespan is at most t. Since the rounding reduces the length of each job by a factor of at most $(1+\varepsilon)$, the computed schedule has makespan at most $(1+\varepsilon) \cdot t$ when considering the original job lengths. Now we schedule the small jobs greedily in leftover space and open new machines if needed. Clearly, whenever a new machine is opened, all previous machines must be loaded to an extent of at least t. Denote by $a(J, t, \varepsilon)$ the number of machines used by this algorithm. Recall that the makespan is at most $(1+\varepsilon) \cdot t$.

Lemma 9.3. *We have that $a(J, t, \varepsilon) \leq m(J, t)$.*

Proof. If the algorithm does not open any new machines for small jobs, then the assertion clearly holds since the rounded down jobs have been scheduled optimally with makespan t. In the other case, all but the last machine are loaded to the extent of t. Hence, the optimal schedule of J having makespan t must also use at least $a(J, t, \varepsilon)$ machines. □

Corollary 9.4. *We have that $T = \min\{t : a(J, t, \varepsilon) \leq m\} \leq \min\{t : m(J, t) \leq m\} = T^*$.*

Binary Search. If T could be determined with no additional error during the binary search, then clearly we could use the above algorithm to obtain a schedule with makespan at most $(1+\varepsilon) \cdot T^*$. Next, we will specify the details of the binary search and show how to control the error it introduces. The binary search is performed in the interval $[\alpha, \beta]$ as defined above. Thus, the length of the available interval is $\beta - \alpha = \alpha$ at the start of the search and it reduces by a factor of two in each iteration. We continue the search until it drops to a length of at most $\varepsilon \cdot \alpha$. This will require $\lceil \log_2 1/\varepsilon \rceil$ many iterations. Let B be the right endpoint of the interval $[A, B]$ we terminate with.

Lemma 9.5. *We have that* $B \leq (1+\varepsilon) \cdot T^*$.

Proof. Clearly $T = \min\{t : a(J,t,\varepsilon) \leq m\}$ must be in the interval $[B - \varepsilon \cdot \alpha, B]$. Hence

$$B \leq T + \varepsilon \cdot \alpha \leq (1+\varepsilon) \cdot T^*,$$

where we have used $T^* \geq \alpha$ and Corollary 9.4. □

For $t \leq B$ this directly gives the result we wanted to show:

Theorem 9.6. *For any* $0 < \varepsilon \leq 1$ *the algorithm produduces a schedule with makespan at most* $(1+\varepsilon)^2 \cdot T^* \leq (1+3\varepsilon) \cdot T^*$ *within running time* $O\left(n^{2k} \cdot \lceil \log_2 1/\varepsilon \rceil\right)$, *where* $k = \lceil \log_{1+\varepsilon} 1/\varepsilon \rceil$.

9.2 Unrelated Machines

Here we give a 2-approximation algorithm for Makespan Scheduling on *unrelated machines*, which means that job j takes time $p_{i,j}$ if scheduled on machine i. The algorithm is based on a suitable LP-formulation and a procedure for rounding the LP.

An obvious integer program for this problem is the following: Let $x_{i,j}$ be a variable indicating if job j is assigned to machine i. The objective is to minimize the makespan. The first set of constraints ensures that each job is scheduled on one of the machines and the second ensures that each machine has a load of at most t.

$$\begin{aligned} \text{minimize} \quad & t \\ \text{subject to} \quad & \sum_{i \in M} x_{i,j} = 1 \quad j \in J, \\ & \sum_{j \in J} p_{i,j} x_{i,j} \leq t, \quad i \in M, \\ & x_{i,j} \in \{0,1\}. \end{aligned}$$

If we relax the constraints $x_{i,j} \in \{0,1\}$ to $x_{i,j} \in [0,1]$, it turns out that this formulation has unbounded integrality gap. (It is left as an exercise to show this.) The main cause of the problem is an "unfair" advantage of the LP-relaxation: If $p_{i,j} > t$, then we must have $x_{i,j} = 0$ in any feasible integer solution, but we might have $x_{i,j} > 0$ in feasible fractional solutions. However, we can not formulate the statement "if $p_{i,j} > t$ then $x_{i,j} = 0$" in terms of linear constraints.

Parametric Pruning. We will make use of a technique called *parametric pruning* to overcome this difficulty. Let the parameter t be a "guess" of a lower bound for the actual makespan T^*. Of course, we will do binary search on t in order to determine a suitable value in an outside loop. However, having a value for t fixed, we are now able to enforce constraints $x_{i,j} = 0$ for all machine-job pairs i, j for which $p_{i,j} > t$. Define $S_t = \{(i,j) : p_{i,j} \leq t\}$. We now define a family LP(t) of linear programs, one for each value of the parameter t. LP(t) uses the variables $x_{i,j}$ for which $(i,j) \in S_t$ and asks if there is a feasible

solution using the restricted assignment possibilities, only.

$$\begin{aligned} \text{minimize} \quad & 0 && (\text{LP}(t)) \\ \text{subject to} \quad & \sum_{i:(i,j)\in S_t} x_{i,j} = 1 \quad j \in J, \\ & \sum_{j:(i,j)\in S_t} p_{i,j} x_{i,j} \leq t, \quad i \in M, \\ & x_{i,j} \geq 0 \quad (i,j) \in S_t. \end{aligned}$$

Extreme Point Solutions. With a binary search, we find the smallest value for t such that LP(t) has a feasible solution. Let T be this value and observe that $T^* \geq T$, i.e., the actual makespan is bounded from below by T. Our algorithm will "round" an extreme point solution of LP(T) to yield a schedule with makespan at most $2 \cdot T^*$. Extreme point solutions to LP(T) have several useful properties.

Lemma 9.7. *Any extreme point solution to* LP(T) *has at most* $n + m$ *many non-zero variables.*

Proof. Let $r = |S_T|$ represent the number of variables on which LP(T) is defined. Recall that a feasible solution is an extreme point solution to LP(T) if and only if it sets r many linearly independent constraints to equality. Of these r linearly independent constraints, at least $r - (n + m)$ must be chosen from the third set of constraints, i.e., of the form "$x_{i,j} \geq 0$". The corresponding variables are set to zero. So, any extreme point solution has at most $n + m$ many non-zero variables. □

Let x be an extreme point solution to LP(T). We will say that job j is *integrally set* if $x_{i,j} \in \{0, 1\}$ for all machines i. Otherwise, i.e., $x_{i,j} \in (0, 1)$ for some machine i, job j is said to be *fractionally set*.

Corollary 9.8. *Any extreme point solution to* LP(T) *must set at least* $n - m$ *many jobs integrally.*

Proof. Let x be an extreme point solution to LP(T) and let α and β be the number of jobs that are integrally and fractionally set by x, respectively. Each job of the latter kind is assigned to at least 2 machines and therefore results in at least 2 non-zero entries in x. Hence we get $\alpha + \beta = n$ and $\alpha + 2\beta \leq n + m$. Therefore $\beta \leq m$ and $\alpha \geq n - m$. □

Algorithm. The algorithm starts by computing the range in which it finds the right value for T. For this it constructs a greedy schedule, in which each job is assigned to the machine on which it has the smallest length. Let α be the makespan of this schedule. Then the range is $[\alpha/m, \alpha]$ (and it is an exercise to show that α/m is indeed a lower bound on T^*).

The LP-rounding algorithm is based on several interesting properties of extreme point solutions of LP(T), which we establish now. For any extreme point solution x for LP(T) define a bipartite graph $G = (M \cup J, E)$ such that $(i, j) \in E$ if and only if $x_{i,j} > 0$. Let $F \subseteq J$ be the fractionally set jobs in x and let H be the subgraph of G induced by the vertex set $M \cup F$. Clearly $(i, j) \in E(H)$ if $0 < x_{i,j} < 1$. A matching in H is called *perfect* if it matches every job $j \in F$. We will show and use that the graph H admits perfect matchings.

We say that a connected graph on a vertex set V is a *pseudo tree* if it has at most $|V|$ many edges. Since the graph is connected, it must have at least $|V| - 1$ many edges. So,

Algorithm 9.1 SCHEDULE UNRELATED

Input. J, M, $p_{i,j}$ for all $i \in M$ and $j \in J$

Output. $x_{i,j} \in \{0,1\}$ for all $i \in M$ and $j \in J$

Step 1. By binary search in $[\alpha/m, \alpha]$ compute smallest value T of the parameter t such that LP(t) has a feasible solution.

Step 2. Let x be an extreme point solution for LP(T).

Step 3. Construct graph H and find perfect matching P.

Step 4. Round in x all fractionally set jobs according to the matching P.

---

it is either a tree or a tree with an additional single edge (closing exactly one cycle). A graph is a *pseudo forrest* if each of its connected components is a pseudo tree.

Lemma 9.9. *We have that G is a pseudo forrest.*

Proof. We will show that the number of edges in each connected component of G is bounded by the number of vertices in it. Hence, each connected component is a pseudo tree.

Consider a connected component G_c. Restrict LP(T) and the extreme point solution x to the jobs and machines of G_c, only, to obtain LP$_c$(T) and x_c. Let $x_{\bar{c}}$ represent the rest of x. The important observation is that x_c must be an extreme point solution for LP$_c$(T). Suppose that this is not the case. Then, x_c is a convex combination of two feasible solutions to LP$_c$(T). Each of these, together with $x_{\bar{c}}$ form a feasible solution for LP(T). Therefore x is a convex combination of two feasible solutions to LP(T). But this contradicts the fact that x is an extreme point solution. With Lemma 9.7 G_c is a pseudo tree. □

Lemma 9.10. *Graph H has a perfect matching P.*

Proof. Each job that is integrally set in x has exactly one edge incident at it in G. Remove these jobs together with their incident edges from G. The resulting graph is clearly H. Since an equal number of edges and vertices have been removed from the pseudo forrest G, H is also a pseudo forrest.

In H, each job has a degree of at least two. So, all leaves in H must be machines. Keep matching a leaf with the job it is incident to and remove them both from the graph. (At each stage all leaves must be machines.) In the end we will be left with even cycles (since we started with a bipartite pseudo forrest.) Match alternating edges of each cycle. This gives a perfect matching P. □

Theorem 9.11. *Algorithm* SCHEDULE UNRELATED *is a 2-approximation for* MAKESPAN SCHEDULING *on unrelated machines.*

Proof. Clearly $T \leq T^*$ since LP(T^*) has a feasible solution. The extreme point solution x to LP(T) has a fractional makespan of at most T. Therefore, the restriction of x to integrally set jobs has an integral makespan of at most T. Each edge (i,j) of H satisfies $p_{i,j} \leq T$. The perfect matching found in H schedules at most one extra job on each machine. Hence, the total makespan is at most $2 \cdot T \leq 2 \cdot T^*$ as claimed. The algorithm clearly runs in polynomial time. □

It is an exercise to show that the analysis is tight for the algorithm.

Chapter 10

Satisfiability

The SATISFYABILITY problem asks if a certain given Boolean formula has a satisfying assignment, i.e., one that makes the whole formula evaluate to true. There is a related optimization problem called MAXIMUM SATISFIABILITY. The goal of this chapter is to develop a deterministic 3/4-approximation algorithm. We first give a corresponding randomized algorithm which will then be derandomized.

We are given the Boolean *variables* $X = \{x_1, \ldots, x_n\}$, where each $x_i \in \{0, 1\}$. A *literal* ℓ_i of the variable x_i is either x_i itself, called a *positive* literal, or its negation $\bar{x}_i$ with truth value $1 - x_i$, called a *negative* literal. A *clause* is a disjunction $C = (\ell_1 \vee \cdots \vee \ell_k)$ of literals ℓ_j of X; their number k is called the *size* of C. For a clause C let S_C^+ denote the set of its positive literals; similarly S_C^- the set of its negative literals. Let $\mathcal{C}$ denote the set of clauses. A Boolean formula in *conjunctive form* is a conjunction of clauses $F = C_1 \wedge \cdots \wedge C_m$. Each vector $x \in \{0, 1\}^n$ is called a *truth assignment*. For any clause C and any such assignment x we say that x *satisfies* C if at least one of the literals of C evaluates to 1.

The problem MAXIMUM SATISFIABILITY is the following: We are given a formula F in conjunctive form and for each clause C a weight w_C, i.e., a weight function $w : \mathcal{C} \to \mathbb{N}$. The objective is to find a truth assignment $x \in \{0, 1\}^n$ that maximizes the total weight of the satisfied clauses. As an important special case: If we set all weights w_C equal to one, then we seek to maximize the number of satisfied clauses.

Now we introduce for each clause C a variable $z_C \in \{0, 1\}$ which takes the value one if and only if C is satisfied under a certain truth assignment x. Now we can formulate this problem as a mathematical program as follows:

Problem 10.1 MAXIMUM SATISFIABILITY

Instance. Formula $F = C_1 \wedge \cdots \wedge C_m$ with m clauses over the n Boolean variables $X = \{x_1, \ldots, x_n\}$. A weight function $w : \mathcal{C} \to \mathbb{N}$.

Task. Solve the problem

$$\begin{aligned} \text{maximize} \quad & \text{value}(z) = \sum_{C \in \mathcal{C}} w_C z_C, \\ \text{subject to} \quad & \sum_{i \in S_C^+} x_i + \sum_{i \in S_C^-} (1 - x_i) \geq z_C \quad C \in \mathcal{C}, \\ & z_C \in \{0, 1\} \quad C \in \mathcal{C}, \\ & x_i \in \{0, 1\} \quad i = 1, \ldots, n. \end{aligned}$$

The algorithm we aim for is a combination of two algorithms. One works better for small clauses, the other for large clauses. Both are initially randomized but can be *derandomized* using the method of conditional expectation, i.e., the final algorithm is deterministic.

10.1 Randomized Algorithm

For each variable x_i we define the random variable X_i that takes the value one with a certain probability p_i and zero otherwise. This induces, for each clause C, a random variable Z_C that takes the value one if C is satisfied under a (random) assignment and zero otherwise.

Algorithm for Large Clauses

Consider this algorithm Randomized Large: For each variable x_i with $i = 1, \ldots, n$, set $X_i = 1$ independently with probability $1/2$ and $X_i = 0$ otherwise. Output $X = (X_1, \ldots, X_n)$.

Define the quantity

$$\alpha_k = 1 - 2^{-k}.$$

Lemma 10.1. *Let C be a clause. If $\text{size}(C) = k$ then*

$$\mathbb{E}[Z_C] = \alpha_k.$$

Proof. A clause C is not satisfied, i.e., $Z_C = 0$ if and only if all its literals are set to zero. By independence, the probability of this event is exactly 2^{-k} and thus

$$\mathbb{E}[Z_C] = 1 \cdot \Pr[Z_C = 1] + 0 \cdot \Pr[Z_C = 0] = 1 - 2^{-k} = \alpha_k$$

which was claimed. □

Theorem 10.2. *In expectation, the algorithm* Randomized Large *is a $1/2$-approximation algorithm for* Maximum Satisfiability.

Proof. By linearity of expectation, Lemma 10.1, and $\text{size}(C) \geq 1$ we have

$$\mathbb{E}[\text{value}(Z)] = \sum_{C \in \mathcal{C}} w_C \mathbb{E}[Z_C] = \sum_{C \in \mathcal{C}} w_C \alpha_{\text{size}(C)} \geq \frac{1}{2} \sum_{C \in \mathcal{C}} w_C \geq \frac{1}{2} \text{value}(z^*)$$

where (x^*, z^*) is an optimal solution for Maximum Satisfiability. We have used the obvious bound $\text{value}(z^*) \leq \sum_{C \in \mathcal{C}} w_C$. □

Algorithm for Small Clauses

Maybe the most natural linear programming relaxation of the problem is:

$$\begin{aligned}
\text{maximize} \quad & \text{value}(z) = \sum_{C \in \mathcal{C}} w_C z_C, \\
\text{subject to} \quad & \sum_{i \in S_C^+} x_i + \sum_{i \in S_C^-} (1 - x_i) \geq z_C \quad C \in \mathcal{C}, \\
& 0 \leq z_C \leq 1 \quad C \in \mathcal{C} \\
& 0 \leq x_i \leq 1 \quad i = 1, \ldots, n.
\end{aligned}$$

In the sequel let $(\bar{x}, \bar{z})$ denote an optimum solution for this LP.

Consider this algorithm RANDOMIZED SMALL: Determine $(\bar{x}, \bar{z})$. For each variable x_i with $i = 1, \ldots, n$, set $X_i = 1$ independently with probability $\bar{x}_i$ and $X_i = 0$ otherwise. Output $X = (X_1, \ldots, X_n)$.

Define the quantity

$$\beta_k = 1 - \left(1 - \frac{1}{k}\right)^k.$$

Lemma 10.3. *Let C be a clause. If* $\text{size}(C) = k$ *then*

$$\mathbb{E}[Z_C] = \beta_k \bar{z}_C.$$

Proof. We may assume that the clause C has the form $C = (x_1 \vee \cdots \vee x_k)$; otherwise rename the variables and rewrite the LP.

The clause C is satisfied if $x_1, \ldots, x_k$ are not all set to zero. The probability of this event is

$$\begin{aligned} 1 - \Pi_{i=1}^k (1 - \bar{x}_i) &\geq 1 - \left(\frac{\sum_{i=1}^k (1 - \bar{x}_i)}{k}\right)^k \\ &= 1 - \left(1 - \frac{\sum_{i=1}^k \bar{x}_i}{k}\right)^k \\ &\geq 1 - \left(1 - \frac{\bar{z}_C}{k}\right)^k. \end{aligned}$$

Above we firstly have used the arithmetic-geometric mean inequality, which states that for non-negative numbers $a_1, \ldots, a_k$ we have

$$\frac{a_1 + \cdots + a_k}{k} \geq \sqrt[k]{a_1 \cdot \cdots \cdot a_k}.$$

Secondly the LP guarantees the inequality $\bar{x}_1 + \cdots + \bar{x}_k \geq \bar{z}_C$.

Now define the function $g(t) = 1 - (1 - t/k)^k$. This function is concave with $g(0) = 0$ and $g(1) = 1 - (1 - 1/k)^k$ which yields that we can bound

$$g(t) \geq t(1 - (1 - 1/k)^k) = t\beta_k$$

for all $t \in [0, 1]$.

Therefore

$$\Pr[Z_C = 1] \geq 1 - \left(1 - \frac{\bar{z}_C}{k}\right)^k \geq \beta_k \bar{z}_C$$

and the claim follows. □

Theorem 10.4. *In expectation, the algorithm* RANDOMIZED SMALL *is a $1-1/e$-approximation algorithm for* MAXIMUM SATISFIABILITY.

Proof. The function β_k is decreasing with k. Therefore if all clauses are of size at most k, then by Lemma 10.3

$$\mathbb{E}[\text{value}(Z)] = \sum_{C \in \mathcal{C}} w_C \mathbb{E}[Z_C] \geq \beta_k \sum_{C \in \mathcal{C}} w_C \bar{z}_C = \beta_k \text{value}(\bar{z}) \geq \beta_k \text{value}(z^*),$$

where (x^*, z^*) is an optimal solution for MAXIMUM SATISFIABILITY. The claim follows since $(1 - 1/k)^k < 1/e$ for all $k \in \mathbb{N}$. □

3/4-Approximation Algorithm

Consider the algorithm Randomized Combine: With probability 1/2 run Randomized Large otherwise run Randomized Small.

Lemma 10.5. *Let C be a clause, then*

$$\mathbb{E}[Z_C] \geq \frac{3\bar{z}_C}{4}.$$

Proof. Let the random variable B take the value zero if the first algorithm is run, one otherwise. For a clause C let $\text{size}(C) = k$. By Lemma 10.1 and $\bar{z}_C \leq 1$

$$\mathbb{E}[Z_C \mid B = 0] = \alpha_k \geq \alpha_k \bar{z}_C.$$

and by Lemma 10.1

$$\mathbb{E}[Z_C \mid B = 1] \geq \beta_k \bar{z}_C.$$

Combining we have

$$\mathbb{E}[Z_C] = \mathbb{E}[Z_C \mid B = 0]\Pr[B = 0] + \mathbb{E}[Z_C \mid B = 1]\Pr[B = 1] \geq \frac{\bar{z}_C}{2}(\alpha_k + \beta_k).$$

Inspection shows that $\alpha_k + \beta_k \geq 3/2$ for all $k \in \mathbb{N}$. □

Theorem 10.6. *In expectation, the algorithm* Randomized Combine *is a* 3/4*-approximation algorithm for* Maximum Satisfiability.

Proof. This follows from Lemma 10.5 and linearity of expectation. □

10.2 Derandomization

The notion of *derandomization* refers to "turning" a randomized algorithm into a deterministic one (possibly at the cost of additional running time or deterioration of approximation guarantee). One of the several available techniques is the method of *conditional expectation*.

We are given a Boolean formula $F = C_1 \wedge \cdots \wedge C_m$ in conjunctive form over the variables $X = \{x_1, \ldots, x_n\}$. Suppose we set $x_1 = 0$, then we get a formula F_0 over the variables $x_2, \ldots, x_n$ after simplification; if we set $x_1 = 1$ then we get a formula F_1.

Example 10.7. Let $F = (x_1 \vee x_2) \wedge (\bar{x}_1 \vee x_3) \wedge (x_1 \vee \bar{x}_4)$ where $X = \{x_1, \ldots, x_4\}$.

$$\begin{aligned} x_1 = 0: &\quad F_0 = (x_2) \wedge (x_4) \\ x_1 = 1: &\quad F_1 = (x_3) \end{aligned}$$

Applying this recursively, we obtain the tree $T(F)$ depicted in Figure 10.1. The tree $T(F)$ is a complete binary tree with height $n+1$ and $2^{n+1}-1$ vertices. Each vertex at level i corresponds to a setting for the Boolean variables $x_1, \ldots, x_i$. We label the vertices of $T(F)$ with their respective conditional expectations as follows. Let $X_1 = a_1, \ldots, X_i = a_i \in \{0, 1\}$ be the outcome of a truth assignment for the variables $x_1, \ldots, x_i$. The vertex corresponding to this assignment will be labeled

$$\mathbb{E}[\text{value}(Z) \mid X_1 = a_1, \ldots, X_i = a_i].$$

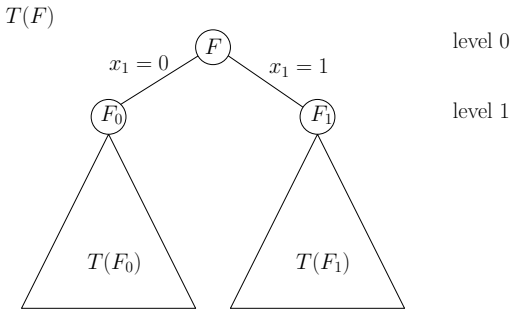


Figure 10.1: Derandomization tree for a formula F.

If $i = n$, then this conditional expectation is simply the total weight of clauses satisfied by the truth assignment $x_1 = a_1, \ldots, x_n = a_n$.

The goal of the remainder of the section is to show that we can find deterministically in polynomial time a path from the root of $T(F)$ to a leaf such that the conditional expectations of the vertices on that path are at least as large as $\mathbb{E}[\text{value}(Z)]$. Obviously, this property yields the desired: We can construct determistically a solution which is at least as good as the one of the randomized algorithm in expectation.

Lemma 10.8. *The conditional expectation*

$$\mathbb{E}[\text{value}(Z) \mid X_1 = a_1, \ldots, X_i = a_i]$$

of any vetex in $T(F)$ can be computed in polynomial time.

Proof. Consider a vertex $X_1 = a_1, \ldots, X_i = a_i$. Let F' be the Boolean formula obtained from F by setting $x_1, \ldots, x_i$ accordingly. F' is in the variables $x_{i+1}, \ldots, x_n$.

Clearly, by linearity of expectation, the expected weight of any clause of F' under any random truth assignment to the variables $x_{i+1}, \ldots, x_n$ can be computed in polynomial time. Adding to this the total weight of clauses satisfied by $x_1, \ldots, x_i$ gives the answer. □

Theorem 10.9. *We can compute in polynomial time a path from the root to a leaf in $T(F)$ such that the conditional expectation of each vertex on this path is at least $\mathbb{E}[\text{value}(Z)]$.*

Proof. Consider the conditional expectation at a certain vertex $X_1 = a_1, \ldots, X_i = a_i$ for setting the next variable X_{i+1}. We have that

$$\begin{aligned}
&\mathbb{E}[\text{value}(Z) \mid X_1 = a_1, \ldots, X_i = a_i] \\
&\quad = \mathbb{E}[\text{value}(Z) \mid X_1 = a_1, \ldots, X_i = a_i, X_{i+1} = 0] \Pr[X_{i+1} = 0] \\
&\qquad + \mathbb{E}[\text{value}(Z) \mid X_1 = a_1, \ldots, X_i = a_i, X_{i+1} = 1] \Pr[X_{i+1} = 1].
\end{aligned}$$

We show that the two conditional expectations with X_{i+1} can *not* be both strictly smaller than $\mathbb{E}[\text{value}(Z) \mid X_1 = a_1, \ldots, X_i = a_i]$. Assume the contrary, then we have

$$\begin{aligned}
&\mathbb{E}[\text{value}(Z) \mid X_1 = a_1, \ldots, X_i = a_i] \\
&\qquad < \mathbb{E}[\text{value}(Z) \mid X_1 = a_1, \ldots, X_i = a_i] (\Pr[X_{i+1} = 0] + \Pr[X_{i+1} = 1])
\end{aligned}$$

which is a contradiction since $\Pr[X_{i+1} = 0] + \Pr[X_{i+1} = 1] = 1$.

This yields the existence of such a path can by Lemma 10.8 it can be computed in polynomial time. □

The derandomized version of a randomized algorithm now simply executes these proofs with the probability distribution as given by the randomized algorithm.