Randomized Algorithms, Summer 2016	Lecture 7 (5 pages)

Introduction to Smoothed Analysis

Instructor: Thomas Kesselheim

We consider the knapsack problem. There are n items, each with a weight $w_i > 0$ and with a profit $p_i > 0$, and a capacity bound $W > 0$. Our task is to find a subset S of the items so as to maximize the overall profit $\sum_{i \in S} p_i$ while not exceeding the capacity bound, i.e., $\sum_{i \in S} w_i \leq W$. It is well known that the knapsack problem is NP-hard. So, unless $P = NP$, there is no polynomial-time algorithm. However, it turns out that practical instances are almost always easy to solve. Today, we will learn one of the reasons why this is true.

Henceforth, we will write the selection of items as a 0/1 vector x, where $x_i = 1$ means that item i is selected and $x_i = 0$ otherwise. Now, the objective function can easily be written as $p^T x$ and the constraint as $w^T x \leq W$. A solution is an arbitrary $x \in \{0,1\}^n$, a *feasible* solution is a solution x for which $w^T x \leq W$.

1 Nemhauser-Ullmann Algorithm

We will study an algorithm that is based on Pareto-optimal solutions.

Definition 7.1. *A solution y* dominates *another solution x if $p^T y \geq p^T x$ and $w^T y \leq w^T x$ and one of these inequalities is strict. A solution x is called* Pareto optimal *if it is not dominated by any other solution y. The* Pareto set *$\mathcal{P}$ is the set of all Pareto-optimal solutions.*

In Figure 1, you can see all possible solutions $x \in \{0,1\}^n$ represented as points in the plane. A point corresponds to a Pareto-optimal solution if there is no other solution to the top left. The next lemma shows that it is enough to restrict the attention to Pareto-optimal solutions when solving the knapsack problem.

Lemma 7.2. *There always exists an optimal solution that is also Pareto optimal.*

Proof. Among all optimal solutions let x be one that minimizes the total $w^T x$. We claim that x is Pareto-optimal. If x was dominated by some y, one of these two cases would be fulfilled

(i) $p^T y > p^T x$ and $w^T y \leq w^T x \leq W$, so x could not be optimal.

(ii) $p^T y = p^T x$ and $w^T y < w^T x \leq W$, so x would not have minimal weight.

As both cases lead to a contradiction, x cannot be dominated by y, so it is Pareto optimal. □

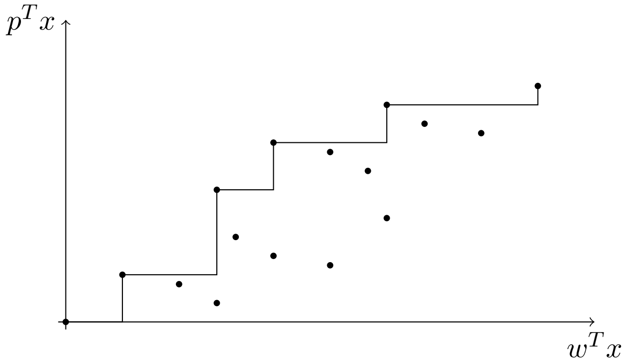

Figure 1: All possible solutions $x \in \{0,1\}^n$, represented as points in the plane. The points connected by the black line are the Pareto-optimal solutions.

This already gives us an algorithm blueprint: Enumerate all Pareto-optimal solutions $\mathcal{P}$. Among these take one that has the highest weight but at most W.

The Nemhauser-Ullmann algorithm computes the set of all Pareto-optimal solutions $\mathcal{P}$ in a smart way. If we restrict the instance to the first i items, there is a set of Pareto-optimal solutions $\mathcal{P}_i$ for this subinstance. This is the set of solutions that only use items $1, \ldots, i$ and are not dominated by another solution that only uses items $1, \ldots, i$. Note that they might not be Pareto-optimal for the full instance because, for example, item n might be extremely valuable and have no weight. This would cause $\mathcal{P}$ to contain only solutions that include item n. The set $\mathcal{P}_i$ in contrast pretends items $i + 1, \ldots, n$ do not even exist.

If we know the set $\mathcal{P}_{i-1}$, it is easy to derive candidates for $\mathcal{P}_i$. This candidate set $\mathcal{Q}_i$ contains all solutions $x \in \mathcal{P}_{i-1}$ and furthermore for every $x \in \mathcal{P}_{i-1}$ the solution that we get when adding i, i.e., x' with $x'_i = 1$, $x'_j = x_j$ for $j \neq i$. The Nemhauser-Ullmann algorithm generates this set $\mathcal{Q}_i$ and then removes solutions that dominated by other solutions in $\mathcal{Q}_i$.

> $\mathcal{P}_0 := \{0^n\}$.
> **for** $i = 1, \ldots, n$ **do**
>   generate $\mathcal{Q}_i$;
> $\mathcal{P}_i := \{x \in \mathcal{Q}_i \mid \nexists y \in \mathcal{Q}_i : y \text{ dominates } x\}$;
> **end for**
> return $x \in \mathcal{P}_n$ that maximizes $p^T x$ subject to $w^T x \leq W$;

Figure 2: Nemhauser-Ullmann algorithms

Theorem 7.3. *There is an implementation of the Nemhauser-Ullmann algorithm that performs* $\Theta(\sum_{i=0}^{n-1} |\mathcal{P}_i|)$ *operations on a unit-cost RAM.*

Proof Idea. As the focus of this lecture is not the exact implementation and analysis of this algorithm, let us only convince ourselves that even a trivial implementation only needs $O(\sum_{i=0}^{n-1} |\mathcal{P}_i|^2)$ operations. In the ith iteration of the for loop, we generate the set $\mathcal{Q}_i$ of size $2|\mathcal{P}_{i-1}|$. Filtering it to get $\mathcal{P}_i$ is trivially done in time $O(|\mathcal{Q}_i|^2)$ by comparing any pair of solutions in $\mathcal{Q}_i$ if one dominates the other. Here you can save a lot by exploiting the structure. □

This algorithm still does not run in polynomial time because there can be very many Pareto-optimal solutions. It is even possible that every single one of the 2^n solutions is Pareto optimal, e.g. if $p_i = w_i = 2^{-i}$. Therefore, a worst-case analysis cannot give us a sub-exponential bound. However, this is a pretty unnatural instance. The solution that picks items 2 to n is only slightly different in weight and in profit from the one that picks item 1. Probably, in reality instances are much nicer than what a hypothetical adversary would define as the worst-case input.

2 Stochastic Input Model

Instead of a worst-case analysis, we perform a *smoothed analysis*. The idea is to interpolate between worst-case analysis (which is too pessimistic) and average-case analysis (which usually is not able to reproduce the structure of "typical" instances). We assume that the profits p_i are defined by an adversary as in a worst-case analysis. The weights w_i are chosen from distributions that are again defined by an adversary. In more detail, the adversary defines probability density functions f_i and w_i is drawn independently according to the distribution defined by f_i.

The density functions have to fulfill the following conditions. We normalize all weights to be between 0 and 1. That is, $f_i(x) \neq 0$ only for $x \in [0, 1]$. Furthermore, the density is bounded by some parameter $\phi \geq 1$ at every point, i.e., $f(x) \leq \phi$ for all x.

Example 7.4. *One important setting that can be captured by this input model is as follows. The adversary can pick weights $0 \leq \bar{w}_i \leq 1$ but these are slightly perturbed to $w_i = \bar{w}_i + X_i$. X_i might by uniformly distributed on $[-\frac{1}{2\phi}, \frac{1}{2\phi}]$. The bigger ϕ is, the better the adversary can control the weights.*

Generally, for $\phi = 1$, the only possible distributions are uniform ones on $[0, 1]$, so we are doing an average case analysis. For $\phi \to \infty$, the adversary gets stronger and stronger and we are getting closer to a worst-case analysis. To be able to reproduce the above example $p_i = w_i = 2^{-i}$ at least approximately, ϕ needs to be in the order of 2^n.

3 The Number of Pareto-Optimal Solutions

Theorem 7.5. *The expected number of Pareto-optimal solutions is at most $n^2\phi + 1$.*

Proof. For $k \in \mathbb{N}$, we let $\mathcal{F}_k$ denote the event that there are two solutions that differ in weight by at most $\frac{n}{k}$. We will split up the expected number of Pareto-optimal solutions based on whether $\mathcal{F}_k$ occurs or not as follows

$$\mathbf{E}[|\mathcal{P}|] \leq \mathbf{Pr}[\mathcal{F}_k]\, \mathbf{E}[|\mathcal{P}| \mid \mathcal{F}_k] + \mathbf{Pr}\left[\overline{\mathcal{F}_k}\right] \mathbf{E}\left[|\mathcal{P}| \;\middle|\; \overline{\mathcal{F}_k}\right] \quad .$$

We will show that $\mathcal{F}_k$ is very unlikely to occur. Therefore, we use the trivial bound $|\mathcal{P}| \leq 2^n$ for this unlikely case. Furthermore, if $\overline{\mathcal{F}_k}$, then all Pareto-optimal solutions differ in weight by at least n/k. This means that there is only a single solution of weight 0, namely the all-zero solution. All other solutions have a weight within one of the intervals $(ni/k, n(i+1)/k]$ for $i \in \{0, \ldots, k-1\}$, but each in a different interval. Let $Y_i^k = 1$ if there is a Pareto-optimal solution of weight in $(ni/k, n(i+1)/k]$, $Y_i^k = 0$ otherwise. If $\overline{\mathcal{F}_k}$, then $|\mathcal{P}| = 1 + \sum_{i=0}^{k-1} Y_i^k$.

Overall, we can bound the expectation of $\mathbf{E}[|\mathcal{P}|]$ by

$$\begin{aligned}
\mathbf{E}[|\mathcal{P}|] &\leq \mathbf{Pr}[\mathcal{F}_k]\, 2^n + \mathbf{Pr}\left[\overline{\mathcal{F}_k}\right] \mathbf{E}\left[1 + \sum_{i=0}^{k-1} Y_i^k \;\middle|\; \overline{\mathcal{F}_k}\right] \\
&= \mathbf{Pr}[\mathcal{F}_k]\, 2^n + \mathbf{Pr}\left[\overline{\mathcal{F}_k}\right] \left(1 + \sum_{i=0}^{k-1} \mathbf{E}\left[Y_i^k \;\middle|\; \overline{\mathcal{F}_k}\right]\right) \\
&= \mathbf{Pr}[\mathcal{F}_k]\, 2^n + \mathbf{Pr}\left[\overline{\mathcal{F}_k}\right] + \sum_{i=0}^{k-1} \mathbf{Pr}\left[Y_i^k = 1 \text{ and } \overline{\mathcal{F}_k}\right] \\
&\leq \mathbf{Pr}[\mathcal{F}_k]\, 2^n + 1 + \sum_{i=0}^{k-1} \mathbf{Pr}\left[Y_i^k = 1\right] \quad .
\end{aligned}$$

Below, we will show that for all $k \in \mathbb{N}$, $i \in \{0, \ldots, n-1\}$, we have

$$\mathbf{Pr}[\mathcal{F}_k] \leq 2^{2n+1}\phi\frac{n}{k} \qquad \text{and} \qquad \mathbf{Pr}\left[Y_i^k = 1\right] \leq \frac{n^2\phi}{k} \quad ,$$

which implies

$$\mathbf{E}[\mathcal{P}] \leq 2^{2n+1}\phi\frac{n}{k}2^n + 1 + k\frac{n^2\phi}{k} = \frac{2^{3n+1}\phi n}{k} + 1 + k\frac{n^2\phi}{k} \quad .$$

This bound holds for every $k \in \mathbb{N}$, so $\mathbf{E}[|\mathcal{P}|] \leq 1 + n^2\phi$. □

Lemma 7.6. *For every $\epsilon > 0$, the probability that there are two distinct solutions $x \neq y$ such that $|w^T x - w^T y| \leq \epsilon$ is at most $2^{2n+1}\phi\epsilon$.*

Proof. We fix any two solutions $x \neq y$ and show that $\mathbf{Pr}\left[w^T x - w^T y \leq \epsilon\right] \leq 2\phi\epsilon$. Because there are at most $2^n \cdot 2^n = 2^{2n}$ solutions overall, a union bound then shows the claim.

The solutions x and y have to differ in at least one item. Without loss of generality, assume that $x_i = 1$ and $y_i = 0$. Now fix the weights w_j for $j \neq i$ arbitrarily to values $\bar{w}_j$. The weights of x and y can now be computed as

$$w^T x = w_i + \sum_{j \neq i} \bar{w}_j x_j \qquad w^T y = \sum_{j \neq i} \bar{w}_j y_j \ .$$

Therefore

$$w^T x - w^T y = w_i + \sum_{j \neq i} \bar{w}_j x_j - \sum_{j \neq i} \bar{w}_j y_j$$

Let $\kappa = \sum_{j \neq i} \bar{w}_j x_j - \sum_{j \neq i} \bar{w}_j y_j$. Observe that this does not depend on w_i. We now have

$$|w^T x - w^T y| \leq \epsilon \quad \text{if and only if} \quad |w_i + \kappa| \leq \epsilon \quad \text{if and only if} \quad w_i \in [\kappa - \epsilon, \kappa + \epsilon] \ .$$

In terms of probabilities, this gives us

$$\mathbf{Pr}\left[|w^T x - w^T y| \leq \epsilon \;\middle|\; w_j = \bar{w}_j \text{ for } j \neq i\right] = \mathbf{Pr}\left[w_i \in [\kappa - \epsilon, \kappa + \epsilon] \mid w_j = \bar{w}_j \text{ for } j \neq i\right] \leq 2\phi\epsilon \ .$$

As this bound holds regardless of the way we condition the values w_j, it also holds without the conditioning. □

Lemma 7.7. *For every $t > 0$ and every $\epsilon > 0$, the probability that there is a Pareto-optimal solution x of weight $w^T x \in (t, t + \epsilon]$ is at most $n\phi\epsilon$.*

Proof. We first find a condition that is necessary to have a Pareto-optimal solution x of weight $w^T x \in (t, t+\epsilon]$. To this end, we keep the weights w fixed. Now, consider the knapsack problem with capacity bound t. Let x^* denote its optimal Pareto-optimal solution. We think of x^* being the winner whereas the Pareto-optimal solutions of higher profit have to have a weight that exceeds t. These are the losers. Let $x^\dagger$ be the loser of smallest weight (and smallest profit). If there is any Pareto-optimal solution of weight in $(t, t + \epsilon]$, then $x^\dagger$ must be one of them. Note that $x^\dagger$ must contain an item i that x^* does not contain. Therefore we get

$$\mathbf{Pr}\left[\exists x \in \mathcal{P} : w^T x \in (t, t + \epsilon]\right] = \mathbf{Pr}\left[\exists i : x^*_i = 0, x^\dagger_i = 1, w^T x^\dagger \in (t, t + \epsilon]\right]$$

For $i \in [n]$, let $\mathcal{E}_i$ denote the event that $x^*_i = 0, x^\dagger_i = 1, w^T x^\dagger \in (t, t + \epsilon]$. By the above considerations, we can only have a Pareto-optimal solution x of weight $w^T x \in (t, t+\epsilon]$ if one of the events $\mathcal{E}_i$ occurs. We will show that $\mathbf{Pr}\left[\mathcal{E}_i\right] \leq \phi\epsilon$ for every i. A union bound then implies

$$\mathbf{Pr}\left[\exists x \in \mathcal{P} : w^T x \in (t, t + \epsilon]\right] \leq \sum_{i=1}^{n} \mathbf{Pr}\left[\mathcal{E}_i\right] \leq n\phi\epsilon \ .$$

From now on, we keep the index i fixed and bound the probability of this particular event $\mathcal{E}_i$ to occur. Let $\mathcal{P}^{+i}$ be the set of solutions that contain item i and are not dominated by another solution that also contains item i. Analogously we denote by $\mathcal{P}^{-i}$ the solutions that do not contain item i and are not dominated by another solution that also does not contain item i. The set of Pareto-optimal solutions $\mathcal{P}$ is a subset of $\mathcal{P}^{+i} \cup \mathcal{P}^{-i}$.

Now, we also fix all weights except w_i. In this conditioned probability space, the weight of all solutions not containing i are fixed. So, the set $\mathcal{P}^{+i}$ is not random anymore. The weights of solutions containing i are of course not fixed yet but they all change by the same amount depending on w_i. Therefore, the set $\mathcal{P}^{-i}$ is also fixed.

Let x^{**} be the solution in $\mathcal{P}^{-i}$ of maximum profit and weight at most t. This is our candidate for x^*. Furthermore, let $x^{\dagger\dagger}$ be the solution in $\mathcal{P}^{+i}$ of minimum profit above $p^T x^{**}$. We note

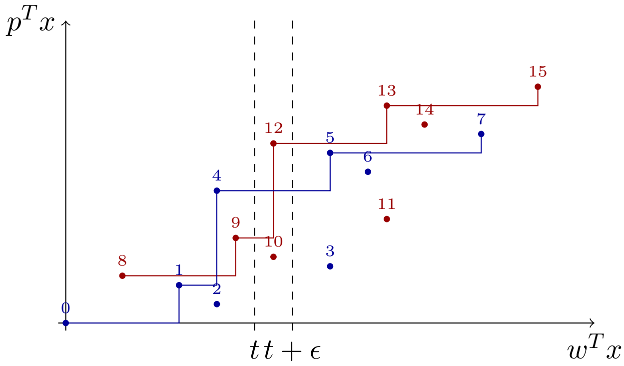

Figure 3: Points are drawn in red if the respective solution contains item i, otherwise blue. The red line connects the solutions in $\mathcal{P}^{+i}$, the blue line the ones in $\mathcal{P}^{-i}$. Changing the weight w_i corresponds to a shift of the horizontal coordinate of every red point by the same amount. The event $\mathcal{E}_i$ only occurs if the weight of solution 12 falls in the interval $(t, t+\epsilon]$. If solution 12's weight is smaller then the optimal solution of weight at most t contains item i. If solution 12's weight is higher, there is no Pareto-optimal solution in the weight interval.

that the event $\mathcal{E}_i$ can only occur if $w^T x^{\dagger\dagger} \in (t, t+\epsilon]$. This is for the following reasons. If $w^T x^{\dagger\dagger} \leq t$, then x^{**} cannot be the optimal solution of weight at most t because $x^{\dagger\dagger}$ is better. If $w^T x^{\dagger\dagger} > t + \epsilon$, no solution that contains i but is not dominated by x^{**} can have a weight in $(t, t+\epsilon]$.

Like in the proof of Lemma 7.6, we can bound the probability of $w^T x^{\dagger\dagger} \in (t, t+\epsilon]$ by using $w^T x^{\dagger\dagger} = w_i + \kappa$ for $\kappa = \sum_{j \neq i} \bar{w}_j x_j^{\dagger\dagger}$. Overall, we get $\mathbf{Pr}\left[\mathcal{E}_i\right] \leq \mathbf{Pr}\left[w^T x^{\dagger\dagger} \in (t, t+\epsilon]\right] \leq \phi\epsilon$. $\square$

4 Further Reading

- Heiko Rglin's lecture notes (Chapter 3): `http://www.roeglin.org/teaching/Skripte/ProbabilisticAnalysis.pdf`
- The results are a special case of this paper:
  Ren Beier, Berthold Vcking: Typical Properties of Winners and Losers in Discrete Optimization. SIAM J. Comput. 35(4): 855-881 (2006)