The knapsack problem

Formulation. Given

(i) $m \in \mathbb{N}$,

(ii) volumes $a_1, \ldots, a_m > 0$,

(iii) values $c_1, \ldots, c_m > 0$,

(iv) the total volume b,

find

$$\max \sum_{i=1}^{m} c_i x_i$$

$$\text{subject to } \sum_{i=1}^{m} a_i x_i \leqslant b, \text{ where } x_i \in \mathbb{N} \cup \{0\}.$$

Binary version: As above but here $x_i \in \{0, 1\}$.

Assumptions: $a_i, b \in \mathbb{N}$, a_i are pairwise different.

Notation. For $k \leqslant m$ and $y \in \{0, 1, \ldots, b\}$ write

$$F(k, y) = \max \left\{ \sum_{i=1}^{k} c_i x_i : \sum_{i=1}^{k} a_i x_i \leqslant y, x_i \in \mathbb{N} \right\}.$$

Recurrence formula.

$$F(k, y) = \max \left(F(k-1, y), c_k + F(k, y - a_k) \right)$$

Example

$$\max 5x_1 + 10x_2 + 12x_3 + 6x_4$$

$$\text{subject to } 4x_1 + 7x_2 + 9x_3 + 5x_4 \leqslant 15$$

$$F(k, y) = \max\left(F(k-1, y), c_k + F(k, y - a_k)\right)$$

k / y	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0	0	0	0	5	5	5	5	10	10	10	10	15	15	15	15
2	0	0	0	0	5	5	5	10	10	10	10	15	15	15	20	20
3	0	0	0	0	5	5	5	10	10	12	12	15	15	17	20	20
4	0	0	0	0	5	6	6	10	10	12	12	15	16	17	20	20

Approximating the optimal solution[1]

Naive approach. Assume $c_1 \geqslant c_2 \geqslant \dots$. Set

$$x_1 = \left\lfloor \frac{b}{a_1} \right\rfloor$$

$$x_2 = \left\lfloor \frac{b - a_1 x_1}{a_2} \right\rfloor, \quad x_3 = \left\lfloor \frac{b - a_1 x_1 - a_2 x_2}{a_3} \right\rfloor$$

and so on.

Example. Let $\theta \in \mathbb{N}$.

$$\max(\theta x_1 + (\theta - 1)x_2) \text{ subject to } \theta x_1 + x_2 \leqslant \theta.$$

The optimal solution is $(0, \alpha)$ giving the value $\theta(\theta - 1)$.
The naive solution is $(1, 0)$ giving θ.

Less naive approach

Greedy Algorithm Assume $c_1/a_1 \geqslant c_2/a_2 \geqslant \dots$. Set

$$x_1 = \left\lfloor \frac{b}{a_1} \right\rfloor$$

$$x_2 = \left\lfloor \frac{b - a_1 x_1}{a_2} \right\rfloor, \quad x_3 = \left\lfloor \frac{b - a_1 x_1 - a_2 x_2}{a_3} \right\rfloor$$

and so on.

Example. Let $\theta \in \mathbb{N}$.

$$\max(2\theta x_1 + 2(\theta - 1)x_2)$$

$$\text{subject to } \theta x_1 + (\theta - 1)x_2 \leqslant 2(\theta - 1).$$

The optimal solution is $(0, 2)$ giving the value $v_{opt} = 4(\theta - 1)$.
The greedy solution is $(1, 0)$ giving $v_{greedy} = 2\theta$.
$v_{greedy}/v_{opt} = \theta/(2\theta - 1) \to 1/2$.

[1] Here we use 15.083J Integer Programming and Combinatorial Optimization from MIT OpenCourseWare.

Asymptotic optimal solutions

Definition. Given $\varepsilon > 0$, a feasinle vector x is ε-optimal if

$$c \cdot x \geqslant (1-\varepsilon) c \cdot x^*,$$

where x^* is an optimal solution.

The above condition is equivalent to

$$\frac{c \cdot x^* - c \cdot x}{c \cdot x^*} \leqslant \varepsilon.$$

Theorem. If x is given by Greedy Axiom then x is 1/2-optimal.

Proof.

$$c \cdot x \geqslant c_1 x_1 = c_1 \left[\frac{b}{a_1}\right]$$

$$c \cdot x^* \leqslant c_1 \frac{b}{a_1} \leqslant c_1 \left(\left[\frac{b}{a_1}\right] + 1\right) \leqslant 2c_1 \left[\frac{b}{a_1}\right] \leqslant 2c \cdot x.$$

Binary knapsack

Formulation. Given

(i) $m \in \mathbb{N}$,
(ii) volumes $a_1, \ldots, a_m > 0$,
(iii) values $c_1, \ldots, c_m > 0$,
(iv) the total volume b,

find

$$\max \sum_{i=1}^{m} c_i x_i$$

$$\text{subject to } \sum_{i=1}^{m} a_i x_i \leqslant b, \text{ where } x_i \in \{0, 1\}.$$

Notation. For $k \leqslant m$ and $y \in \{0, 1, \ldots, b\}$ write

$$F(k, y) = \max \left\{ \sum_{i=1}^{k} c_i x_i : \sum_{i=1}^{k} a_i x_i \leqslant y, x_i \in \mathbb{N} \right\}.$$

Recurrence formula.

$$F(k, y) = \max \left(F(k-1, y), c_k + F(k-1, y-a_k) \right)$$

Remark. Write $c_{\max} = \max_{i \leqslant m} c_i$.
Using the formula we need $c_{\max} m^2$ operations do solve the problem

Approximating binary knapsack

Theorem. An ε-optimal solution of a binary klapsack problem can be found after $10m^3/\varepsilon$ steps.

Fix some natural number t and let $\overline{c_i}$ be c_i with the last t digits replaced by 0. Moreover,

$$\widehat{c_i} = \frac{\overline{c_i}}{10^t}.$$

Let S, S' be the optimal choices for c_i's and $\overline{c_i}$ giving the values v^* and v, respectively. Then

$$\sum_{i \in S} c_i \geqslant \sum_{i \in S'} c_i \geqslant \sum_{i \in S'} \overline{c_i} \geqslant$$
$$\geqslant \sum_{i \in S} \overline{c_i} \geqslant \sum_{i \in S} (c_i - 10^t) = \sum_{i \in S} c_i - m10^t.$$

Hence

$$\frac{v^* - v}{v^*} \leqslant \frac{m10^t}{c_{\max}}.$$

Fix $\varepsilon > 0$. If $m/c_{\max} > \varepsilon$ then we apply the exact algorithm requiring $m^2 c_{\max}$ so at most m^3/ε operations.

Otherwise, choose t such that

$$\frac{\varepsilon}{10} \leqslant \frac{m10^t}{c_{\max}} < \varepsilon.$$

Then we apply the exact algorithm for the prices $\widehat{c_i}$. Note that

$$\widehat{c_{\max}} = \frac{1}{10^t} \overline{c_{\max}} \leqslant \frac{1}{10^t} c_{\max} \leqslant 10m/\varepsilon,$$

so the number of operations we need is $m^2 \widehat{c_{\max}} \leqslant \frac{10m^3}{\varepsilon}$.