Theoretical Computer Science 259 (2001) 323–339

Theoretical Computer Science

www.elsevier.com/locate/tcs

Approximation of some NP-hard optimization problems by finite machines, in probability

Dawei Hong[a,*], Jean-Camille Birget[b,1]

[a]Department of Math. and Computer Science, Southwest State University, Marshall, MN 56258, USA
[b]Department of Computer Science, University of Nebraska, Lincoln, NE 68588, USA

Received May 1998; revised September 1999
Communicated by M. Crochemore

Abstract

We introduce a subclass of NP optimization problems which contains some NP-hard problems, e.g., bin covering and bin packing. For each problem in this subclass we prove that with probability tending to 1 (exponentially fast as the number of input items tends to infinity), the problem is approximable up to any chosen relative error bound $\varepsilon > 0$ by a deterministic finite-state machine. More precisely, let Π be a problem in our subclass of NP optimization problems, let $\varepsilon > 0$ be any chosen bound, and assume there is a fixed (but arbitrary) probability distribution for the inputs. Then there exists a finite-state machine which does the following: On an input I (random according to this probability distribution), the finite-state machine produces a feasible solution whose objective value $M(I)$ satisfies

$$P\left(\frac{|Opt(I) - M(I)|}{\max\{Opt(I), M(I)\}} \geqslant \varepsilon\right) \leqslant K\mathrm{e}^{-hn},$$

when n is large enough. Here K and h are positive constants.

Keywords: NP-optimization problems; Approximation; Probabilistic analysis of algorithms; Finite-state machines

1. Introduction

Let us first recall the standard definition of NP optimization problem [12, 4].

An *NP optimization problem* over an alphabet Σ is a four-tuple $\Pi = (\mathscr{I}, \mathscr{S}, m, opt)$ such that:

* Corresponding author.
E-mail addresses: hong@ssu.southwest.msus.edu (D. Hong), birget@cs.dal.ca (J.-C. Birget).
[1] Second author's research supported in part by NSF grant DMS-9203981.

PII: S0304-3975(00)00016-5

1. $\mathscr{I} \subseteq \Sigma^*$, the set of admissible input instances, is assumed to be recognizable in polynomial time;
2. $\mathscr{S}(I) \subseteq \Sigma^*$ is the set of all feasible solutions on input I, for every $I \in \mathscr{I}$. The relation $\{(I,s)\colon\ I \in \mathscr{I},\ s \in \mathscr{S}(I)\}$ is assumed to be decidable in deterministic polynomial time.
3. $m : \mathscr{I} \times \Sigma^* \mapsto \mathbb{R}$, the objective function, is a polynomial-time computable function.
4. $opt \in \{\max, \min\}$ indicates whether Π is a maximization or a minimization problem.

We let the inputs be finite sequences of positive rational numbers. We also assume that the values of the objective function m are positive rational numbers. Eventually we will encode rational numbers as strings.

For an NP optimization problem Π we let $opt(I)$ denote the optimum value of the objective function on input I. Let A be an algorithm which produces a feasible solution with objective value $A(I)$ on input I. We say that Π is approximated by A up to a factor ε iff for any non-empty input I we have (see [14])

$$|Opt(I) - A(I)|/max\{Opt(I), A(I)\} \leqslant \varepsilon.$$

We also call such an ε a "bound on the relative error". We say that Π is *asymptotically approximated* by A up to a factor ε iff the above relation holds for all inputs $I = (a_1, \dots, a_n)$ with n large enough.

We consider approximation properties of certain NP optimization problems in a *probabilistic* setting. We describe the inputs by sequences $(x_1, \dots, x_n)$ of independent, identically distributed ("i.i.d.") random variables x_i with values over the positive rationals. The common domain of the x_i is a probability space $(\Omega, \mathscr{B}, P)$, with underlying set Ω, σ-algebra $\mathscr{B}$, and probability measure $P : \mathscr{B} \to [0, 1]$. Note that since each x_i has only strictly positive values, the expectation $E[x_i]$ exists (allowing $+\infty$) and is not zero.

It has often been observed that the probabilistic behavior of an algorithm can be much better than the worst case behavior. Our results illustrate this again. For example, for the bin covering problem no approximation algorithm is known with arbitrarily small asymptotic approximation factor, in the worst case (see e.g. [2, 3] for background). However, we give an algorithm that has these properties with probability tending to 1 (exponentially fast) as the number of input items tends of ∞. In addition, our algorithm is just a finite-state machine (hence, it runs in real time).

The fact that any NP-hard optimization problems are approximated by finite-state machines, for any chosen relative error bound ε, is surprising. However, probability seems to be the key to reasonably good approximations by finite-state machines.

This paper is a slightly extended version of [9]. In Section 2 we define our subclass of NP optimization problems. The class is defined by a list of five axiomatic properties; the first three axioms state, in essence, that the optimum is "*semi-linear*" in terms of the input; axiom 4 is an asymptotic *positivity* constraint on the optimum, and axiom 5 is a *symmetry* condition; therefore we call this class "the *LinPosSym* subclass of NP optimization problems". The effect of the axioms is to make $Opt(x_1, \dots, x_n)$ have "probabilistic concentration". The class of NP optimization problems we end up with has a rather natural and simple definition, and contains some well-known NP-hard optimization problems (e.g., bin packing, bin covering). The main part of the paper is the probabilistic analysis, given in Section 3. Our main result is the following:

Theorem. *For any problem Π in the LinPosSym subclass of* NP *optimization problems and for any $\varepsilon > 0$, there exists a finite-state machine which on a random input $I = (x_1, \dots, x_n)$ (consisting of any number $n \geqslant 1$ of i.i.d. random variables that are positive-rational valued), produces a feasible solution whose objective value $M(I)$ satisfies*

$$P\left(\frac{|Opt(I) - M(I)|}{\max\{Opt(I), M(I)\}} \geqslant \varepsilon\right) \leqslant K\mathrm{e}^{-hn},$$

when n is large enough. Here $K > 0$ is a universal constant, and $h > 0$ is a constant depending on Π, ε, and the probability distribution of the inputs.

Discussion. It is remarkable that some NP-hard optimization problems are ε-approximable by a finite-state machine, for arbitrarily small $\varepsilon > 0$. The probability of an error exceeding ε decreases exponentially with the input size. The finite machine is deterministic. The result holds for arbitrary input distributions.

Our algorithm is very simple in outline, but the parameters "N" and $\xi_i^{(N)}$ of the algorithm are determined by a probabilistic analysis in a complicated way; the hardest part is the proof that the algorithm has the stated properties.

On the other hand, our result has some limitations too:

• Our algorithm assumes a fixed input distribution and a fixed ε. The construction of the machine from the distribution and from ε is not efficient. But this is not always a problem, since in many designs the error tolerance and the input distribution are pre-established, as parts of a 'design specification'.

It may be possible to make the dependence on $1/\varepsilon$ polynomial, since the "window size" N is polynomially bounded (see Eq. (3) and Theorem 3.4). But the dependence on the distribution is complicated.

• Our class of NP optimization problems contains a few natural problems; it would be desirable to find more problems in the class.

2. A subclass of NP optimization problems

In this section we define a subclass of NP-optimization problems, and we give two examples of NP-hard problems in that class. Approximation properties of this class will be studied in the next section.

For two sequences I_1 and I_2 with the same number of coordinates, we say that I_1 is *dominated* by I_2 (denoted by $I_1 \preccurlyeq I_2$) iff each coordinate in I_1 is less than or equal to the corresponding coordinate in I_2. By $I_1 \cdot I_2$ we denote the *concatenation* of the strings I_1 and I_2. We denote by $\mathbb{Q}$ the set of rational numbers, by $\mathbb{Q}_{>0}$ the set of positive rational numbers, by $\mathbf{S}^*$ the set of all finite sequences of elements of any set $\mathbf{S}$, by $\mathbb{R}$ the set of real numbers, and by $\mathbb{Z}$ the set of all integers.

Definition 2.1. We introduce the following subclass of NP optimization problems, called "the *LinPosSym* subclass of NP optimization problems". Each problem Π in this class takes finite sequences of positive rational numbers as inputs, either with bounded or unbounded values: $\mathscr{I} = \mathbb{Q}^*_{>0}$ or $\mathscr{I} = \{r \in \mathbb{Q}\colon\ 0 < r \leqslant c\}^*$ (for some rational constant c). The feasible solutions are also assumed to be strings (over $\mathbb{Q}$ or over some finite alphabet). We assume the following axioms for the *LinPosSym* subclass; the first 3 axioms state, in essence, that the optimum is "semi-linear" in terms of the input; axiom 4 is an asymptotic positivity constraint on the optimum, and axiom 5 is a symmetry condition.

(A1) Subadditivity (Superadditivity)

The empty-string input $I = ()$ is admissible, and $Opt(I) = 0$.

The concatenation of any two admissible input strings I_1 and I_2 is admissible, and the concatenation of two feasible solutions S_1 resp. S_2 on input I_1 resp. I_2 is a feasible solution on input $I_1 \cdot I_2$; the objective value of $S_1 \cdot S_2$ satisfies $m(S_1 \cdot S_2) = m(S_1) + m(S_2)$. Axiom A1 implies the following:

If Π is a minimization problem then $Opt(I_1 \cdot I_2) \leqslant Opt(I_1) + Opt(I_2)$.

If Π is a maximization problem then $Opt(I_1 \cdot I_2) \geqslant Opt(I_1) + Opt(I_2)$.

(A2) Monotonicity

For any two admissible input strings I_1 and I_2, if $I_1 \preccurlyeq I_2$ then $Opt(I_1) \leqslant Opt(I_2)$.

(A3) Restriction axiom

If $I = (a_1, \ldots, a_n)$ is an admissible input, then $I_{\hat{r}} = (a_1, \ldots, a_{r-1}, a_{r+1}, \ldots, a_n)$ (the sequence obtained by dropping a_r) is also admissible. Moreover, there is a constant $\lambda \geqslant 0$ (depending only on Π) such that

$$Opt(I_{\hat{r}}) \leqslant Opt(I) \leqslant Opt(I_{\hat{r}}) + \lambda.$$

(A4) Non-vanishing of *Opt*

For any sequence $(x_n)_{n \geqslant 0}$ of i.i.d. random variables with admissible values:

$$\liminf_{n \to \infty} \frac{1}{n} Opt(x_1, \ldots, x_n) > 0 \quad \text{a.s.}$$

(A5) Permutation invariance

For any $n > 0$, any permutation π of $\{1, \ldots, n\}$, and any admissible input $(a_1, \ldots, a_n) \in (\mathbb{Q}_{>0})^n$, the permuted input $(a_{\pi(1)}, \ldots, a_{\pi(n)})$ is admissible, and

$$Opt(a_{\pi(1)}, \ldots, a_{\pi(n)}) = Opt(a_1, \ldots, a_n).$$

Note that for every $n > 0$, $Opt(x_1, \ldots, x_n)$ is a random variable if $x_1, \ldots, x_n$ are random variables (with admissible positive rational values). This is implied by the fact that for all $n > 0$, the restriction $Opt : (\mathbb{Q}_{>0})^n \to \mathbb{Q}$ is a Borel measurable function. (The reason is trivial: Every subset of $\mathbb{Q}^n$ is a Borel set because $\mathbb{Q}^n$ is countable and because singletons are closed.)

As a consequence of axiom A3, for every input I with n coordinates,

$$Opt(I) \leqslant \lambda n. \tag{1}$$

Examples (bin covering, bin packing). Bin covering and bin packing are classical NP optimization problems whose decision versions are NP-complete (see e.g. [2, 7, 14]). For *bin covering* our algorithm represents a significant advance: no real-time algorithm was known for arbitrary ε (even for the uniform distribution). For *bin packing* the literature is extensive, but all the known good ε-approximation results were for the uniform distribution.

Proposition 2.2. *Bin covering and bin packing belong to the LinPosSym subclass of NP optimization problems.*

Proof. Super- or sub-additivity (axiom A1), monotonicity (axiom A2), and permutation invariance (axiom A5), are straightforward to check.

Axiom A3 holds with $\lambda = 1$: Indeed, when we remove an input item from a covering then, except for the one bin from which this item is removed, all bins are still covered. A similar argument applies to bin packing.

Axiom A4: For bin covering, the classical "first-fit" heuristic (see [7, 10]) satisfies: $\liminf_{n\to\infty}(1/n)\, FirstFit(x_1,\ldots,x_n) \geqslant \frac{1}{2}E[x_1] > 0$ a.s. A fortiori, this holds for the function *Opt*, since bin covering is a maximization problem.

For bin packing, $Opt(x_1,\ldots,x_n) \geqslant \sum_{i=1}^{n} x_i$. By the strong law of large numbers, $\lim_{n\to\infty} \sum_{i=1}^{n} x_i/n = E[x_1]$ a.s., and we know that $E[x_1] > 0$. □

3. Approximation with high probability

Let Π be a problem in the *LinPosSym* subclass of NP optimization problems (Definition 2.1). Recall that we represent the inputs by sequences of i.i.d. random variables $(x_1,\ldots,x_n)$, $n > 0$, defined on a probability space $(\Omega, \mathscr{B}, P)$, with values in $\mathbb{Q}_{>0}$. Let us fix a probability distribution $F : \mathbb{Q}_{>0} \to [0,1]$ for x_i (common to all x_i).

3.1. Probability concentration

Theorem 3.1. *If Π belongs to the LinPosSym subclass of* NP *optimization problems then there is a strictly positive constant μ* (*depending on Π and on the distribution*) *such that*

$$\lim_{n\to\infty} \frac{1}{n} Opt(x_1,\ldots,x_n) = \mu \quad a.s.$$

Moreover, for any $\varepsilon_1 > 0$ *we have*

$$\text{for all } n \text{ large enough,} \quad P\left(\left|\frac{1}{n}Opt(x_1,\ldots,x_n) - \mu\right| \leqslant \varepsilon_1\right) \geqslant 1 - 2\exp\left(-\frac{\varepsilon_1^2 n}{8\lambda^2}\right),$$

$$\text{and for all } n > 0, \quad P\left(\left|\frac{1}{n}Opt(x_1,\ldots,x_n) - \mu\right| \leqslant \varepsilon_1\right) \geqslant 1 - \alpha\exp\left(-\frac{\varepsilon_1^2 n}{8\lambda^2}\right),$$

where $\alpha > 0$ *is a constant depending on the problem and the distribution. Here* λ *is as in axiom* A3 *of Definition* 2.1.

Proof. By applying Theorem 4.2 (in the appendix) to the super- (or sub-) additive process $\{Opt(x_{s+1},\ldots,x_t)\colon s,t \in \mathbb{Z},\ s<t\}$ we have $\lim_{n\to\infty}(1/n)Opt(x_1,\ldots,x_n) = \mu$ a.s., for some constant μ. By axiom A4, μ is strictly positive. This proves the limit result.

The above limit result implies $\lim_{n\to\infty}(1/n)E[Opt(x_1,\ldots,x_n)] = \mu$. So for all n large enough,

$$\left|\frac{1}{n}E[Opt(x_1,\ldots,x_n)] - \mu\right| \leqslant \frac{\varepsilon_1}{2}. \tag{2}$$

Now for any $n > 0$ we construct a *martingale* $X_0, X_1, \ldots, X_n$ as follows: For $i = 1,\ldots,n$,

$$X_i = E[Opt(x_1,\ldots,x_n) \mid x_1,\ldots,x_i],$$

and

$$X_0 = E[Opt(x_1,\ldots,x_n)].$$

Hence, by classical properties of the conditional expectation,

$$X_n = Opt(x_1,\ldots,x_n).$$

For X_{i-1} with $1 \leqslant i \leqslant n+1$ we also have the following explicit formula:

$$X_{i-1}(\omega) = \int_{v_i,\ldots,v_n \in \Omega} Opt(x_1(\omega),\ldots,x_{i-1}(\omega),x_i(v_i),\ldots,x_n(v_n))\, \mathrm{d}F(x_i(v_i))\cdots \mathrm{d}F(x_n(v_n)).$$

Then by classical properties of integrals,

$$\begin{aligned}|X_{i-1} - X_i| &\leqslant \int_{v_i,v_{i+1},\ldots,v_n \in \Omega} |Opt(x_1(\omega),\ldots,x_{i-1}(\omega),x_i(v_i),x_{i+1}(v_{i+1}),\ldots,x_n(v_n)) \\ &\quad - Opt(x_1(\omega),\ldots,x_{i-1}(\omega),x_i(\omega),x_{i+1}(v_{i+1}),\ldots,x_n(v_n))| \\ &\quad \mathrm{d}F(x_i(v_i))\cdots \mathrm{d}F(x_n(v_n)).\end{aligned}$$

Moreover,

$$\begin{aligned} &Opt(\ldots,a_{i-1},b_i,b_{i+1},\ldots) - Opt(\ldots,a_{i-1},a_i,b_{i+1},\ldots) \\ &\quad = Opt(\ldots,a_{i-1},b_i,b_{i+1},\ldots) - Opt(\ldots,a_{i-1},b_{i+1},\ldots) \\ &\qquad - (Opt(\ldots,a_{i-1},a_i,b_{i+1},\ldots) - Opt(\ldots,a_{i-1},b_{i+1},\ldots)) \\ &\quad \leqslant \lambda, \end{aligned}$$

where the last inequality follows from axiom A3 and the fact that $Opt(\ldots)$ is non-negative. Hence we have for $i = 1,\ldots,n$:

$$|X_{i-1} - X_i| \leqslant \lambda.$$

Letting $t = (\varepsilon_1 n)/2$ and $C_i \leqslant \lambda$ $(1 \leqslant i \leqslant n)$ in Azuma's lemma (see the appendix) we obtain

$$P\left(|Opt(x_1,\ldots,x_n) - E[Opt(x_1,\ldots,x_n)]| \leqslant \frac{1}{2}\varepsilon_1 n\right) \geqslant 1 - 2\exp\left(-\frac{\varepsilon_1^2 n}{8\lambda^2}\right).$$

The theorem now follows by (2), for all large enough n.

From this we can then derive the "for all n" result. Let us denote $p_n = P(|(1/n)Opt(x_1,\ldots,x_n) - \mu| \leqslant \varepsilon_1)$. We just proved that there exists n_0 (depending on the problem and the distribution) such that for all ε_1 and for all $n > n_0$,

$$p_n \geqslant 1 - 2\exp\left(-\frac{\varepsilon_1^2 n}{8\lambda^2}\right).$$

And for $n \leqslant n_0$, we can write $p_n \geqslant 1 - \alpha_n \exp(-\varepsilon_1^2 n/8\lambda^2) = 0$, where $\alpha_n = \exp(-\varepsilon_1^2 n/8\lambda^2)$. Now if we pick $\alpha = \max\{2, \alpha_1, \ldots, \alpha_{n_0}\}$ we will have $p_n \geqslant 1 - \alpha\exp(-\varepsilon_1^2 n/8\lambda^2)$ for *all* $n > 0$. □

Corollary 3.2. *If Π belongs to the LinPosSym subclass of NP optimization problems then there is a constant $\beta > 0$ (depending on Π and on the probability distribution) such that for all $n > 0$:*

$$\left|\frac{1}{n}E[Opt(x_1,\ldots,x_n)] - \mu\right| \leqslant \frac{\beta}{\sqrt{n}}.$$

The constant μ is the same as in Theorem 3.1.

Proof. In the second inequality of Theorem 3.1, let us take ε_1 to be of the form $s/\sqrt{n}$; the inequality becomes then $P(|(1/n)Opt(x_1,\ldots,x_n) - \mu| \leqslant s/\sqrt{n}) \geqslant 1 - \alpha\exp(-s^2/8\lambda^2)$. Equivalently, $P(|(1/n)Opt(x_1,\ldots,x_n) - \mu| > s/\sqrt{n}) < \exp(-s^2/8\lambda^2)$.

Therefore by the definition of expectation,

$$\left|\frac{1}{n}E[Opt(x_1,\ldots,x_n)] - \mu\right| = \left|\frac{1}{n}\int_\Omega Opt(x_1,\ldots,x_n)\,\mathrm{d}F - \mu\right|.$$

We partition the probability space Ω into the subsets

$$X_s = \left\{ \omega \in \Omega \colon \frac{s-1}{\sqrt{n}} \leqslant \left| \frac{1}{n} Opt(x_1(\omega),\ldots,x_n(\omega)) - \mu \right| < \frac{s}{\sqrt{n}} \right\},$$

where s ranges over all positive integers. Let us break up the integral according to this partition, and move the absolute values inside the integrals; then

$$\begin{aligned} \left| \frac{1}{n} E[Opt(x_1,\ldots,x_n)] - \mu \right| &\leqslant \sum_{s=1}^{\infty} \int_{X_s} \left| \frac{1}{n} Opt(x_1,\ldots,x_n) - \mu \right| \mathrm{d}F \\ &< \sum_{s=1}^{\infty} \frac{s}{\sqrt{n}} \alpha \exp\left(-\frac{(s-1)^2}{8\lambda^2} \right) \\ &= \frac{\alpha}{\sqrt{n}} \sum_{s=1}^{\infty} s \exp\left(-\frac{(s-1)^2}{8\lambda^2} \right). \end{aligned}$$

The latter sum converges and evaluates to a positive constant. This yields the upper bound $\beta/\sqrt{n}$ for some constant $\beta > 0$. □

3.2. An approximation algorithm

Given any problem Π in the *LinPosSym* subclass of NP optimization problems, and given an approximation factor $\varepsilon > 0$ and a probability distribution F, we present a real-time *deterministic* algorithm that produces an ε-approximation for Π, with high probability. This algorithm takes a sequence of positive rational numbers as input, but only looks at an input sequence through a "window" of size N. The window size N is a constant (in terms of the inputs), depending on Π, F, and ε, and satisfying $N = \mathrm{O}(1/\varepsilon^2)$. In the next subsection we will encode the rational numbers as strings over a fixed alphabet, and turn the algorithm into a finite-state machine.

Our approximation algorithm $\mathscr{A}$ has the following specification:
Input of $\mathscr{A}$: a sequence $I = (a_1,\ldots,a_n) \in \mathscr{I}$;
Output of $\mathscr{A}$: a feasible solution $S \in \mathscr{S}(I)$ of Π, whose objective value $m(I,S) = \mathscr{A}(I)$ satisfies the probabilistic inequality in Theorem 3.4 below.

Preprocessing (independent of the input I):
Pick N to be the smallest *square* integer so that

$$N > \frac{1}{\varepsilon^2} \left(\frac{2}{\mu} (\beta + 3\lambda\kappa) \right)^2,$$

where the constant $\mu, \beta, \lambda, \kappa$ are given, respectively, in Theorem 3.1, Corollary 3.2, axiom A3, and the Kolmogorov–Smirnov statistics (4.5). This choice of N will be justified at the end of the proof of Theorem 3.4. The number N is called *the window size*, due to its role in the algorithm. Since N is chosen to be minimal subject to the above inequality, we have

$$N = \mathrm{O}\left(\frac{1}{\varepsilon^2} \right). \tag{3}$$

Let q_F be the quantile transformation of the distribution F (see the appendix). For all i $(1 \leqslant i \leqslant N)$ we define

$$\xi_i^{(N)} = q_F\left(\frac{i}{N}\right).$$

Let us consider the set $\{\xi_1^{(N)}, \ldots, \xi_N^{(N)}\}$ and consider all sequences of length N consisting of elements from this set (so there are $\leqslant N^N$ such sequences). For every such sequence Ξ, used as an input of problem Π, we pick an *optimal* feasible solution $\sigma(\Xi) \in \mathscr{S}(\Xi)$; its optimal objective value is $Opt(\Xi)$. We arrange the results of this preprocessing into a table $\mathscr{T}$ which, for every Ξ gives $\sigma(\Xi)$ and $Opt(\Xi)$.

This completes the preprocessing.

Lemma 3.3. *Each* $\xi_i^{(N)}$ $(1 \leqslant i \leqslant N)$ *is a positive rational number* (*except perhaps* $\xi_N^{(N)}$ *which could be* $+\infty$ *if* x_i *is unbounded*)*. Moreover,* $\xi_1^{(N)} \leqslant \cdots \leqslant \xi_N^{(N)}$.

Proof. Note that q_F *is positive-rational valued*, since F is the distribution of a random variable (namely x_i) which is positive-rational valued (recall that "min" is used in the definition of the quantile transformation). Also, if the random variable x_i is bounded by c (i.e., $\forall \omega$, $x_i(\omega) \leqslant c$) then q_F is also bounded by c. Also, $\xi_i^{(N)} \leqslant c$ if $x_i \leqslant c$. □

Remark. Recall that everything we do is for a fixed distribution F and a fixed ε. We do not have an algorithm that finds N, the $\xi_i^{(N)}$'s, and the algorithm $\mathscr{A}$, from ε and F.

The algorithm $\mathscr{A}$ will be based on the step function ξ from $\mathbb{Q}_{>0}$ to $\{\xi_1^{(N)}, \ldots, \xi_{N-1}^{(N)}, \xi_N^{(N)}\}$, defined by

$$\xi(x) = \min\{\xi_i^{(N)}\colon\ x \leqslant \xi_i^{(N)},\ 1 \leqslant i \leqslant N\}.$$

Algorithm $\mathscr{A}$

for $k := 0$ to $\lfloor \frac{n}{N} \rfloor - 1$ do
  begin
  $\Xi_k := (\xi(x_{kN+1}), \ldots, \xi(x_{kN+i}), \ldots, \xi(x_{kN+N}))$;
  in the table $\mathscr{T}$, look up $\sigma(\Xi_k)$ and output it (by concatenating $\sigma(\Xi_k)$ on the right of the output already produced);
  end

The *intuition* for this algorithm is simple: We break the random input $(x_1, \ldots, x_n)$ into $\lfloor n/N \rfloor$ successive (non-overlapping) segments $I_k = (x_{kN+1}, \ldots, x_{kN+N})$, for $k = 0, 1, \ldots, \lfloor n/N \rfloor - 1$. We discard the remainder $(x_{\lfloor n/N \rfloor N+1}, \ldots, x_n)$ since it has length $< N$ which is asymptotically negligible. For every input segment I_k (which consists on N input numbers, where N is the constant window size chosen in Preprocessing), we find a segment Ξ_k by applying the function ξ to I_k. By the Kolmogorov–Smirnov statistics (Theorem 4.5, especially (7)), we expect I_k and Ξ_k to have closely related *Opt* values, with high probability. The optimal solutions $\sigma(\Xi_k)$ for all the Ξ_k's have been precomputed, independently of the input (what depends on the input are the Ξ_k's that are actually picked, and their order). By axiom A1 we can concatenate feasible solutions. The next theorem shows that the resulting total solution is indeed close to an optimum, asymptotically, with high probability.

Digression. It is interesting (although this is not needed here) that I_k and $\xi(I_k)$ have related distributions. We define the following discrete probability distribution F_N concentrated on $\{\xi_1^{(N)},\ldots,\xi_{N-1}^{(N)},\xi_N^{(N)}\}$:

$$F_N(t)=\begin{cases} F(\xi_1^{(N)}) & \text{if } t\leqslant\xi_1^{(N)},\\ F(\xi_i^{(N)}) & \text{if } \xi_{i-1}^{(N)}<t\leqslant\xi_i^{(N)},\ 1<i\leqslant N.\end{cases}$$

If $X=(x_1,\ldots,x_N)$ is a sequence of i.i.d. random variables with distribution F, then $\xi(X)=(\xi(x_1),\ldots,\xi(x_N))$ is a sequence of i.i.d. random variables with distribution F_N. Conversely, if Ξ is a sequence of N i.i.d. random variables over $\{\xi_1^{(N)},\ldots,\xi_{N-1}^{(N)},\xi_N^{(N)}\}$ with distribution F_N, then Ξ is almost surely equal to $\xi(X)$ for some sequence X of N i.i.d. random variables with distribution F. [End of Digression.]

Theorem 3.4 (Correctness of $\mathscr{A}$). *Let Π be in the LinPosSym subclass of* NP *optimization problems, and let $I=(x_1,\ldots,x_n)$ be a random input, consisting of i.i.d. random variables with positive-rational values and with distribution F. Then for any $\varepsilon>0$, the algorithm $\mathscr{A}$ produces a feasible solution whose objective value $\mathscr{A}(I)$ satisfies*

$$P\left(\frac{|Opt(I)-\mathscr{A}(I)|}{\max\{Opt(I),\mathscr{A}(I)\}}\geqslant\varepsilon\right)\leqslant K\mathrm{e}^{-hn},$$

when n is large enough. Here $K>0$ is a universal constant, and $h>0$ is a constant depending on Π,ε and F.

The window size N in the algorithm satisfies $N<\gamma/\varepsilon^2$ where γ is a positive constant (which depends on the problem and the distribution).

Proof. We prove the theorem in the case where Π is a maximization problem; the proof is similar for minimization problems.

As we saw in the intuitive motivation of $\mathscr{A}$, we can assume for simplicity that n is divisible by N (this has no effect asymptotically). Accordingly, we write n/N instead of $\lfloor n/N\rfloor$. We apply the Hoeffding inequality (9) (in the appendix), with $X_k:=Opt(\Xi_k)$ $(1\leqslant k\leqslant n/N)$, where Ξ_k is as in the algorithm. Since different Ξ_k's are obtained in the algorithm by applying the function ξ to non-overlapping I_k's, it follows that the $Opt(\Xi_k)$'s are i.i.d. random variables. In the Hoeffding inequality we let X be $Opt(\Xi_1)$. Since $0<Opt(\Xi_k)\leqslant\lambda N$ (as a consequence of axiom A3, see inequality (1)), we let $n:=n/N$, $t:=N\varepsilon_2$ and $c_{\mathrm{H}}:=\lambda N$ in the Hoeffding inequality. Note that by Axiom A1, $\sum_{k=1}^{n/N}Opt(\Xi_k)=\mathscr{A}(I)$. Then (9) yields for any $\varepsilon_2>0$,

$$P\left(\frac{n}{N}\cdot(E[Opt(\Xi_1)]-N\varepsilon_2)\leqslant\mathscr{A}(I)\right)\geqslant 1-\exp\left(-\frac{2\varepsilon_2^2 n}{N\lambda^2}\right). \tag{4}$$

Let us now estimate $E[Opt(\Xi_1)]$ and compare it with $E[Opt(I_1)]$, where $I_1=(x_1,\dots,x_N)$. Suppose that by sorting the elements of I_1 in nondecreasing order we have $x_{i_1}\leqslant\cdots\leqslant x_{i_N}$ (order statistics). Then by the Kolmogorov–Smirnov statistics (7) we have for any s (such that $1\leqslant s\leqslant\sqrt{N}+1$):

$$P(E_1\cap E_2)\geqslant 1-\kappa \mathrm{e}^{-2(s-1)^2},$$

where the event E_1 and E_2 are defined by

$$E_1=\{\omega\in\Omega\colon (x_{i_{(s-1)\sqrt{N}+1}}(\omega),\dots,x_{i_N}(\omega))\succcurlyeq(\xi_1^{(N)}(\omega),\dots,\xi^{(N)}_{N-(s-1)\sqrt{N}}(\omega))\},$$

$$E_2=\{\omega\in\Omega\colon (\xi^{(N)}_{(s-1)\sqrt{N}+1}(\omega),\dots,\xi_N^{(N)}(\omega))\succcurlyeq(x_{i_1}(\omega),\dots,x_{i_{N-(s-1)\sqrt{N}}}(\omega))\}.$$

Recall that N is a square.

For all s $(1\leqslant s\leqslant\sqrt{N}+1)$, consider the event

$$F_s=\{\omega\in\Omega\colon |Opt(I_1(\omega))-Opt(\Xi_1(\omega))|\leqslant 2\lambda s\sqrt{N}\}.$$

Claim. $E_1\cap E_2\subseteq F_{s-1}$.

Proof. For all $\omega\in E_1\cap E_2$ and all k with $2(s-1)\sqrt{N}<k\leqslant N$,

$$x_{i_{k-2(s-1)\sqrt{N}}}(\omega)\leqslant\xi^{(N)}_{k-(s-1)\sqrt{N}}(\omega)\leqslant x_{i_k}(\omega). \tag{5}$$

Applying ξ to $x_{i_k}(\omega)$ in the above inequality yields (by the definition of ξ)

$$x_{i_{k-2(s-1)\sqrt{N}}}(\omega)\leqslant\xi^{(N)}_{k-(s-1)\sqrt{N}}(\omega)\leqslant \xi(x_{i_k}(\omega))$$

for $2(s-1)\sqrt{N}<k\leqslant N$.

By axiom A5 (on permutations), this implies that the subsequence $I_1'(\omega)$ of $I_1(\omega)$ consisting of the numbers $x_{i_{k-2(s-1)\sqrt{N}}}(\omega)$ with $2(s-1)\sqrt{N}<k\leqslant N$, is dominated by the appropriately permuted subsequence $\Xi_1'(\omega)$ of $\Xi_1(\omega)$ consisting of the numbers $\xi(x_{i_k}(\omega))$ (with $2(s-1)\sqrt{N}<K\leqslant N$).

Hence by axioms A2 and A3,

$$Opt(I_1(\omega))-2(s-1)\sqrt{N}\lambda\leqslant Opt(I_1'(\omega))\leqslant Opt(\Xi_1'(\omega))\leqslant Opt(\Xi_1(\omega)).$$

Hence, for all $\omega\in E_1\cap E_2$,

$$Opt(I_1(\omega))-2(s-1)\sqrt{N}\lambda\leqslant Opt(\Xi_1(\omega)).$$

Applying ξ to $x_{i_{k-2(s-1)\sqrt{N}}}(\omega)$ in (5), by a similar argument we have for all $\omega\in E_1\cap E_2$,

$$Opt(\Xi_1(\omega))-2(s-1)\sqrt{N}\lambda\leqslant Opt(I_1(\omega)).$$

The Claim is equivalent to the above two inequalities. □

By the Claim and the Kolmogorov–Smirnov inequality (7) we have for any s such that $1\leqslant s\leqslant\sqrt{N}+1$:

$$P(F_s-F_{s-1})\leqslant P(\overline{F_{s-1}})\leqslant P(\overline{E_1\cap E_2})\leqslant\kappa\mathrm{e}^{-2(s-1)^2}.$$

Recall the following fact involving conditional expectations:

$$E[X] = \sum_{s=1}^{m} E[X|B_s] \cdot P(B_s)$$

if $\{B_s\colon 1 \leqslant s \leqslant m\}$ is any partition of the probability space Ω. Hence,

$$|E[Opt(I_1)] - E[Opt(\Xi_1)]| \leqslant \sum_{s=1}^{\sqrt{N}+1} E[|Opt(I_1) - Opt(\Xi_1)||F_s - F_{s-1}] \cdot P(F_s - F_{s-1}).$$

But by the definition F_s we have $E[|Opt(I_1) - Opt(\Xi_1)||F_s - F_{s-1}] \leqslant 2\lambda s\sqrt{N}$. Therefore

$$|E[Opt(I_1)] - E[Opt(\Xi_1)]| \leqslant 2\lambda\sqrt{N} \sum_{s=1}^{\sqrt{N}+1} sP(F_s - F_{s-1}) \leqslant 2\kappa\lambda\sqrt{N} \sum_{s=1}^{\infty} s\mathrm{e}^{-2(s-1)^2} = L\sqrt{N}$$

for some constant L (depending on Π only), such that $L \leqslant 2\kappa\lambda \cdot 1.272 \leqslant 3\kappa\lambda$. (The infinite sum evaluates to $\sum_{s=1}^{\infty} s\exp(-2(s-1)^2) = 1.271 \pm 0.001$.)

Combining this and (4) we obtain for any $\varepsilon_2 > 0$,

$$P\left(n\left(\frac{E[Opt(I_1)]}{N} - \frac{L}{\sqrt{N}} - \varepsilon_2\right) \leqslant \mathscr{A}(I)\right) \geqslant 1 - \exp\left(-\frac{2\varepsilon_2^2 n}{N\lambda^2}\right).$$

By Corollary 3.2 $1/NE[Opt(I_1)] \geqslant \mu - \beta/\sqrt{N}$. Hence,

$$P(H_1 \cdot n \leqslant \mathscr{A}(I)) \geqslant 1 - \exp\left(-\frac{2\varepsilon_2^2 n}{N\lambda^2}\right) \quad \text{where } H_1 = \mu - \frac{\beta + L}{\sqrt{N}} - \varepsilon_2.$$

By this and Theorem 3.1 we have for any $\varepsilon_1, \varepsilon_2 > 0$,

$$P\left(1 - H_2 \leqslant \frac{\mathscr{A}(I)}{Opt(I)} \leqslant 1\right) \geqslant 1 - 2\exp\left(-\frac{\varepsilon_1^2 n}{8\lambda^2}\right) - \exp\left(-\frac{2\varepsilon_2^2 n}{N\lambda^2}\right),$$

where

$$H_2 = \frac{1}{\sqrt{N}} \frac{\beta + L}{\mu + \varepsilon_1} + \frac{\varepsilon_1 + \varepsilon_2}{\mu + \varepsilon_1}.$$

For any $\varepsilon > 0$, since $\mu > 0$, we may take ε_1 and ε_2 small enough so that the second term of H_2 is less than $\varepsilon/2$. Then we take N large enough so that the first term of H_2 is less than $\varepsilon/2$, i.e., we take N as in Preprocessing; so the constant γ in the Theorem can be chosen as $\gamma = (1 + 2/\mu(\beta + 3\lambda\kappa))^2$. Now $H_2 < \varepsilon$, and the theorem follows. □

3.3. *The finite-state machine*

We will use deterministic finite-state machines with output (or Mealy machines) in their most classical sense; see e.g. [11] for a definition.

Theorem 3.5. *Let Π be a problem in the LinPosSym class of* NP *optimization problems and let $I = (x_1, \ldots, x_n)$ be a random input, consisting of i.i.d. random variables with positive rational values, and with distribution F. Rational numbers are coded as strings over a finite alphabet. Let $\varepsilon > 0$ be fixed.*

Then there exists a finite-state machine *$\mathscr{M}$ which on input I produces a feasible solution whose objective value $\mathscr{M}(I)$ satisfies*

$$P\left(\frac{|Opt(I) - \mathscr{M}(I)|}{\max\{Opt(I), \mathscr{M}(I)\}} \geqslant \varepsilon\right) \leqslant K\mathrm{e}^{-hn}$$

when n is large enough. Here $K > 0$ is a universal constant and $h > 0$ is a constant depending on Π, ε and F.

Proof. Since ε, F are fixed, N is fixed, and the table $\mathscr{T}$ is fixed and finite. So the algorithm $\mathscr{A}$ looks like a finite-state machine (i.e., a Mealy machine), except that so far, input sequences consisted of arbitrary positive rational numbers. The inputs of an automaton have to be strings over a finite alphabet.

To encode the inputs over a finite alphabet we proceed as follows. First, we represent every positive rational number in the form $b/c + a$ with $a, b, c \in \mathbb{N}$, and $b < c$, and where a, b, c are written in *reverse binary*: the least significant bit is at the left and is read first. Second, the pair of numbers b, c is represented as two "parallel bit streams" (i.e., two bit strings of equal length, possibly with leading zeros, lined up bit by bit) over the letters $\binom{0}{0}, \binom{0}{1}, \binom{1}{0}, \binom{1}{1}$. For example, the string $\binom{1}{1}\binom{0}{1}\binom{1}{0}\binom{0}{0}\binom{0}{1} + 0001$ represents the number $\frac{5}{19} + 8$. The total *alphabet of* $\mathscr{M}$ is $\{0, 1, +, \binom{0}{0}, \binom{0}{1}, \binom{1}{0}, \binom{1}{1}, \#\}$, where # serves as a separator between input items.

For an input item $b/c + a$ (represented as a string), $\mathscr{M}$ simultaneously compares this item with the N *fixed* numbers $\xi_i^{(N)}$ $(1 \leqslant i \leqslant N)$, and thus computes Ξ_k from the kth input subsegment I_k (where $0 \leqslant k < \lfloor n/N \rfloor$). A finite automaton can compare a variable rational number $b/c + a$ (represented as a string in the above way), with a fixed number $\xi_i^{(N)}$ as follows.

The finite automaton first compares the integral part a of $b/c + a$ with the (fixed) integral part of $\xi_i^{(N)} = \beta_i/\gamma_i + \alpha_i$ (recall that each $\xi_i^{(N)}$ is indeed rational). This is done as follows: We assume that each $\xi_i^{(N)}$ is stored in the finite memory of $\mathscr{M}$. While the integral part a of the input item is being read, $\mathscr{M}$ generates α_i and (based on these two parallel bit streams) $\mathscr{M}$ compares a and α_i. A classical three-state automaton can compare integers represented in reverse binary (see Fig. 1).

If the integral parts are equal, the fractional part b/c of the input item is compared next with the (fixed) fractional part β_i/γ_i of $\xi_i^{(N)}$. This is done as follows:

The finite automaton can multiply a variable number (represented in reverse binary) by a fixed number, using a construction similar to the classical binary adder. Thus, given b/c the automaton computes $\gamma_i b/\beta_i c$ (all fractions represented as parallel bit streams in reverse binary). At the same time, the automaton compares $\beta_i c$ and $\gamma_i b$ (using the three-state comparator of Fig. 1). $\square$

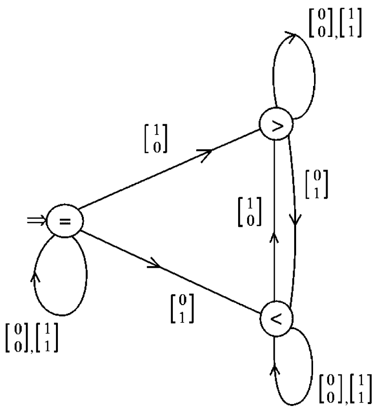

Fig. 1.

Acknowledgements

Both authors would like to thank David A. Klarner for his kind encouragement.

4. Appendix A. Probability theory

A.1. Kingman's theorems on subadditive processes

A *subadditive process* is a family of real random variables $\{X_{s,t}\colon\ s,t \in \mathbb{Z},\ s<t\}$ such that:

(1) For all $s<t<u$, $X_{s,u} \leqslant X_{s,t} + X_{t,u}$;
(2) The distribution of $X_{s,t}$ is completely determined by $t-s$;
(3) $X_{0,t}$ has finite expectation and there is a constant α_0 such that for all $t \geqslant 1$, $E[X_{0,t}] \geqslant \alpha_0 t$.

By reversing inequalities about X or $E[X]$ in (1) and (3) we obtain a *superadditive process* (see [13]).

Proposition A.1 (Kingman [13]). *If* $\{X_{s,t}\colon\ s,t \in \mathbb{Z},\ s<t\}$ *is a subadditive* (*or superadditive*) *process then there is a constant* $\rho \neq \infty$ *such that*

$$\lim_{t\to\infty} \frac{E[X_{0,t}]}{t} = \rho.$$

Theorem A.2 (Kingman [13]). *If* $\{X_{s,t}\colon\ s,t \in \mathbb{Z},\ s<t\}$ *be a subadditive or superadditive process, and let* ρ *be the same as in Proposition* A.1. *The for all* s,

$$\lim_{t\to\infty} \frac{X_{s,t}}{t-s} = \rho \quad a.s.$$

A.2. Azuma's martingale lemma

Theorem A.3 (Azuma [1]). *Suppose the sequence* $X_0, X_1, \ldots, X_n$ *is a martingale. For each* i $(1 \leqslant i \leqslant n)$, *let* $C_i = \sup\{|X_i(\omega) - X_{i-1}(\omega)|:\ \omega \in \Omega\}$ *and let* $C = \sum_{i=1}^{n} C_i^2$. *Then for all* $t > 0$,

$$P(|X_n - X_0| > t) \leqslant 2 \exp\left(-\frac{t^2}{2C}\right).$$

A.3. Quantile transformation

The *quantile transformation* q_F of a probability distribution F over $\mathbb{R}$ (see [6]) is the function $[0,1] \to \mathbb{R} \cup \{-\infty, +\infty\}$, defined by

$$q_F(z) = \min\{t \in \mathbb{R} \cup \{-\infty\}:\ F(t) \geqslant z\}.$$

This definition uses min (instead of inf) because F is continuous from the left.

Proposition A.4 (Gaenssler [6]). *Let* x_i $(1 \leqslant i \leqslant k)$ *be i.i.d. real-valued random variables with distribution* F. *Then there exist uniformly distributed i.i.d. random variables* u_i $(1 \leqslant i \leqslant k)$ *over* $[0,1]$ *such that for all* i, $x_i = q_F(u_i)$ *a.s.*

A.4. Kolmogorov–Smirnov statistics

Let $(u_{(1)}, \ldots, u_{(n)})$ be the order statistics of a sequence $(u_1, \ldots, u_n)$ of i.i.d. random variables that are uniformly distributed over $[0,1]$. The following is a direct consequence of a famous result, called the Kolmogorov–Smirnov statistics – see e.g. [5]. (Indeed, it is not hard to show that the expression $\max_{x \in \mathbb{R}} |S_n(x) - F(x)|$ in Chapter 10 of [5], where $F(x)$ is the uniform distribution on $[0,1]$, is equal to the expression $\max_{1 \leqslant i \leqslant n} |u_{(i)} - i/n|$ in the Theorem below.)

Theorem A.5. *There is a constant* $\kappa > 0$ *such that for all* $s \geqslant 0$,

$$P\left(\max_{1 \leqslant i \leqslant n} \left\{\left|u_{(i)} - \frac{i}{n}\right|\right\} \geqslant \frac{s}{\sqrt{n}}\right) \leqslant \kappa \mathrm{e}^{-2s^2}.$$

Applications of the Kolmogorov–Smirnov statistics:

Consider the two events

$$G_n^{(1)} = \bigcap_{i=1}^{n} \left\{ q_F\left(\max\left\{0, \frac{i}{n} - \frac{s}{\sqrt{n}}\right\}\right) \leqslant q_F(u_{(i)}) \right\},$$

$$G_n^{(2)} = \bigcap_{i=1}^{n} \left\{ q_F(u_{(i)}) \leqslant q_F\left(\min\left\{1, \frac{i}{n} + \frac{s}{\sqrt{n}}\right\}\right) \right\}.$$

Then by Theorem A.5 it follows that for any $s > 0$,

$$P(G_n^{(1)} \cap G_n^{(2)} \geqslant 1 - \kappa \mathrm{e}^{-2s^2}, \tag{A.1}$$

where κ is the universal constant from Theorem A.5.

For all $n \geqslant 1$ and $1 \leqslant i \leqslant n$, we define $\xi_i^{(n)} = q_F(i/n)$. Now, consider the two events

$$H_n^{(1)} = \bigcap_{i=s\sqrt{n}+1}^{n} \{q_F(u_{(i)}) \geqslant \xi_{i-s\sqrt{n}}^{(n)}\},$$

$$H_n^{(2)} = \bigcap_{i=s\sqrt{n}+1}^{n} \{\xi_i^{(n)} \geqslant q_F(u_{(i-s\sqrt{n})})\}.$$

Then (A.1) implies

$$P(H_n^{(1)} \cap H_n^{(2)}) \geqslant 1 - \kappa e^{-2s^2}. \tag{A.2}$$

A.5. Hoeffding inequalities

Theorem A.6 (Hoeffding [8]). *Let $X_i, \ldots, X_n$ be i.i.d. random variables, and assume $0 \leqslant X_i \leqslant c_{\mathrm{H}}$ $(i = 1, \ldots, n)$ for some constant $c_{\mathrm{H}} > 0$. Let X be another random variable with the same distribution as $X_1, \ldots, X_n$. Then for all $t > 0$,*

$$P\left(\sum_{i=1}^{n} X_i - nE[X] \leqslant t\right) \geqslant 1 - \exp\left(-\frac{2t^2}{nc_{\mathrm{H}}^2}\right). \tag{A.3}$$

Replacing X by $2E[X] - X$ and X_i by $2E[X_i] - X_i$ in (8) yields

$$P\left(nE[X] - \sum_{i=1}^{n} X_i \leqslant t\right) \geqslant 1 - \exp\left(-\frac{2t^2}{nc_{\mathrm{H}}^2}\right). \tag{A.4}$$

References

[1] K. Azuma, Weighted sums of certain dependent variables, Tohoku Math. J. 3 (1965) 357–367.

[2] E.G. Coffman Jr., M.R. Garey, D.S. Johnson, Approximation algorithms for bin packing: a survey, in: D. Hochbaum (Ed.), Approximation Algorithms for NP-hard Problems, PWS Publishing Co., Boston, MA, 1997, pp. 46–93.

[3] E.G. Coffman Jr., G.S. Lueker, Probabilistic Analysis of Packing and Partitioning Algorithms, Wiley, New York, 1991.

[4] P. Crescenzi, A. Panconesi, Completeness in approximation classes, Inform. and Comput. 93 (1991) 241–262.

[5] M. Fisz, Probability Theory and Mathematical Statistics (Translation of Rachunek prawdopodobienstwa i statystyka matematyczna, by R. Bartoszynski), Krieger Pub, Malabar, FL, 1980.

[6] P. Gaenssler, Empirical Processes, Lecture Note – Monograph Series, Vol. 3, Institute of Mathematical Statistics, Hayward, CA, 1983.

[7] S. Han, D. Hong, J.Y.-T. Leung, Probabilistic analysis of a bin covering algorithm, Oper. Res. Lett. 18 (1996) 193–199.

[8] W. Hoeffding, Probability inequalities for sums of bounded random variables, J. Amer. Statist. Assoc. 58 (1965) 13–30.

[9] D. Hong, J.C. Birget, Probabilistic approximation of some NP optimization problems by finite-state machines, in: J. Rolim (Ed.), Randomization and Approximation Techniques in Computer Science (RANDOM'97), Lecture Notes in Computer Science, Vol. 1269, Springer, Berlin, July, 1997, pp. 151–164. RANDOM'97.

[10] D. Hong, J.Y-T. Leung, Probabilistic analysis of k-dimensional packing algorithms, Inform. Process. Lett. 55 (1995) 17–24.

[11] J. Hopcroft, J. Ullman, Introduction to Automata Theory, Languages, and Computation, Addison-Wesley, Reading, MA, 1979.

[12] D.S. Johnson, Approximation algorithms for combinatorial problems, J. Comput. System Sci. 9 (1974) 256–278.
[13] J.F.C. Kingman, Subadditive processes, Lecture Notes in Mathematics, Vol. 539, Springer, Berlin, 1976, pp. 168–222.
[14] C.H. Papadimitriou, Computational Complexity, Addison-Wesley, Reading, MA, 1994.