REVUE FRANÇAISE D'AUTOMATIQUE, D'INFORMATIQUE ET DE RECHERCHE OPÉRATIONNELLE. RECHERCHE OPÉRATIONNELLE

U. PFERSCHY

The random linear bottleneck assignment problem

Revue française d'automatique, d'informatique et de recherche opérationnelle. Recherche opérationnelle, tome 30, nº 2 (1996), p. 127-142.

<http://www.numdam.org/item?id=RO_1996__30_2_127_0>

Recherche opérationnelle/Operations Research
(vol. 30, n° 2, 1996, pp. 127-142)

THE RANDOM LINEAR BOTTLENECK ASSIGNMENT PROBLEM ([a]) (*)

by U. PFERSCHY ([1])

Communicated by Rainer E. BURKARD

Abstract. – *It is shown that the expected value of the optimal solution of an $n \times n$ linear bottleneck assignment problem with independently and identically distributed costs tends towards the infimum of the cost range as n tends to infinity. For fixed n and the uniform distribution explicit upper and lower bounds are given.*
Moreover, an algorithm with $O(n^2)$ expected running time is presented.

Keywords: Random bottleneck assignment, average case analysis, random graphs.

Résumé. – *Nous montrons que la moyenne de la solution optimale d'un problème d'affectation linéaire avec goulet d'étranglement, dans le cas de coûts indépendants identiquement distribués, tend vers l'infimum de l'intervalle des coûts lorsque n tend vers l'infini. Nous donnons explicitement les bornes supérieures et inférieures lorsque n est fixé et la distribution uniforme.*
En outre, nous présentons un algorithme en temps moyen $O(n^2)$.

Mots clés : Affectation aléatoire avec goulet d'étranglement, analyse du cas moyen, graphes aléatoires.

1. INTRODUCTION

The well-known *Linear Bottleneck Assignment Problem* (LBAP) is defined as:

$$(LBAP) \quad \min \max_{i,j} c_{ij}\, x_{ij} \tag{1}$$

$$\sum_{i=1}^{n} x_{ij} = 1, \quad 1 \le j \le n,$$

(*) Received November 1993.
([a]) This research was supported by the Fonds zur Förderung der wissenschaftlichen Forschung, Project P8971-PHY.
([1]) TU Graz, Institut für Mathematik B, Steyrergasse 30, A-8010 Graz, Autria.

$$\sum_{j=1}^{n} x_{ij} = 1, \quad 1 \le i \le n,$$
$$x_{ij} \in \{0, 1\}, \quad 1 \le i, j \le n,$$

with cost-coefficients $c_{ij} \in \mathbb{R}^+$.

It can be formulated in a graph theoretical setup as finding a perfect matching in a bipartite weighted graph $H = (S \cup T, E)$ which minimizes the maximum weight of all matching edges.

This problem is closely related to the classical *Linear Sum Assignment Problem* (*LSAP*) where (1) is replaced by

$$(LSAP) \quad \min \sum_{i,j} c_{ij}\, x_{ij}.$$

Both problems have been studied extensively in the past. For practical large-size problems many implementations are available. The $LBAP$ can be solved by a modified threshold algorithm due to Gabow and Tarjan [9] in $O(n^{5/2}\sqrt{\log n})$ time. For the $LSAP$, various $O(n^3)$ augmentation algorithms were developed, *see* e.g. Burkard, Derigs [4] or Derigs [6] and the references therein.

Moreover, various results are known about the probabilistic properties and the asymptotic behaviour of the $LSAP$ as n tends to infinity.

In 1979 Walkup [17] showed that the expected optimal value of an $LSAP$ with cost coefficients uniformly distributed between 0 and 1 is less than 3, for large n. Five years later, Karp [12] improved this bound to 2. The more general situation where the distribution function is not uniform but arbitrary has been studied by Frenk, van Houweninge and Rinnooy Kan [8] and recently more extensively by Olin [15]. The expected value of the random $LSAP$ in the asymptotic case ($n \to \infty$) has been bounded from below for the uniform (0, 1) distribution by Lazarus [14] by approximately 1.37. An improved lower bound of 1.51 and limits for general distributions are found again in the work of Olin [15].

For the $LBAP$ no explicit asymptotic investigations are known to the author. Some results from random graph theory can be applied readily to the $LBAP$ by choosing the edges of an evolutionary graph process in increasing order. (*See* e.g. Bollobás [2], Ch. VII.)

In this paper we show that the expected value of the optimal solution of the $LBAP$ tends towards the lower end of the range of cost coefficients for any distribution function (as long as the upper end of the cost range is bounded). Moreover, we can derive functions in n as explicit upper and lower bounds for the expected solution value in the case of data distributed uniformly between 0 and 1.

Therefore, for practical problems we can expect that the gap between the optimal solution and the maximum of the row– and column-minima which is a natural lower bound computed by the usually applied heuristics, gets rather small for larger problems. Hence, the most widely used solution method of augmenting paths (*see* e.g. Derigs [5]) will most likely terminate after only a small number of augmentations, as the heuristically determined initial partial matching consisting only of edges with weights smaller or equal to the lower bound mentioned above, will be close to a perfect matching in most cases. Moreover, algorithms which reduce the given complete graph to a sparse subgraph consisting of a number of the smallest edges emanating from each vertex can be expected to yield a solution on the sparse graph which is also close to optimal for the complete graph.

Following these considerations we give an algorithm for the $LBAP$ based on the construction of a sparse subgraph with expected running time $O(n^2)$. The $LSAP$ for comparison can be solved in expected time $O(n^2 \log n)$ by an algorithm due to Karp [1].

2. ASYMPTOTIC BEHAVIOUR

To achieve the claimed convergence we first describe the probabilistic setup and then show the resulting asymptotic properties.

2.1. The probabilistic model

Let each edge cost be independently and identically distributed wih an arbitrary distribution function F. To show that the expected value of the optimal $LBAP$-solution tends towards the infimum of the possible range of the F-distributed cost coefficients we apply a constructive approach using sparse subraphs of the original graph.

Although the structure of the optimal solution of an $LBAP$ depends only on the ordering of the n^2 random cost-coefficients and only the actual solution value depends on the specific distribution function, we perform the proof for arbitrary cost distributions.

Let $G(n, d)$ be the class of bipartite digraphs with n nodes in each class S and T and outdegree d at each node. As indicated in Walkup [18] this class of graphs behaves different from the family $G^0(n, N)$ of undirected bipartite graphs with exactly N edges defined by Erdös and Rényi [7]. The existence of a perfect matching in a graph chosen uniformly from the latter family depends mainly on the existence of isolated vertices. This phenomenon is excluded in the former model of $G(n, d)$.

The canonical uniform selection of a random graph G from $G(n, d)$ can be interpreted in the following way: Starting from a graph consisting only of S and T without arcs for each node in S (and in T in turn) exactly d arcs are added by choosing as endpoints d nodes from T (resp. S) which are selected by sampling uniformly among the n possible endpoints without replacement. Thereby, each graph in $G(n, d)$ is generated with equal probability.

A perfect matching in a digraph is a subset of n arcs such that each node is either head or tail of exactly one arc. We use the following lemma established by Walkup in [18]:

LEMMA 1: *Let $Pr(n, d)$ be the probability of the existence of a perfect matching in a graph selected uniformly from the class $G(n, d)$. Then the following inequalities hold:*

$$Pr(n, 2) \geq 1 - \frac{1}{5n}$$

$$Pr(n, d) \geq 1 - \frac{1}{122}\left(\frac{d}{n}\right)^{(d+1)(d-2)} \quad for \quad d \geq 3. \quad \square$$

We now proceed in the same way as Olin [15] and Walkup [17] and define the optimal value of the $LBAP$ as Z_n, a random variable depending on the edge costs which are independently and identically distributed random variables C_{ij} with a common distribution function $F(x)$.

Our objective is to show that

$$\lim_{n \to \infty} E[Z_n] = \inf\{x | F(x) > 0\},$$

which gives a lower bound for the possible values of the edge costs.

To use Lemma 1 we have to select special subgraphs uniformly from $G(n, d)$. In order to construct a random graph in $G(n, d)$ such that the choice of the arcs adjacent to each node is independent for every node we change our model of the graph H into a directed graph following the construction of Walkup [17].

Let Y_{ij} and Z_{ij}, $i, j = 1, \ldots, n$ be independent random variables with the common distribution function

$$G(x) = Pr(Y_{ij} \leq x) = 1 - \sqrt{1 - F(x)}.$$

We set $C_{ij} = \min\{Y_{ij}, Z_{ij}\}$ and get

$$\begin{aligned} Pr(C_{ij} \leq x) &= 1 - Pr(C_{ij} > x) \\ &= 1 - Pr(Y_{ij} > x) \cdot Pr(Z_{ij} > x) \\ &= 1 - (1 - F(x)) = F(x), \end{aligned}$$

as defined above.

For each $d \in \{1, \ldots, n\}$ we select an element $G_d \in G(n, d)$ in the following way: For each node $s_i \in S$ choose a set A_i of d elements arbitrarily from $\{1, \ldots, n\}$ such that $Y_{ij} \leq Y_{ik}$, $\forall j \in A_i$, $\forall k \notin A_i$ (*i.e.* d of the smallest elements of Y_{ik}, $k = 1, \ldots, n$). Add an arc in G_d from s_i to t_j, $\forall j \in A_i$. In the same way the $n \cdot d$ arcs from T to S are generated according to the values of Z_{ij}. In this way G_d is selected uniformly from $G(n, d)$ and $G_1 \subset G_2 \subset \ldots \subset G_n = K'_{n,n}$, where $K'_{n,n}$ is the complete directed bipartite graph on S and T.

For each G_d we denote the value of its maximum weighted arc by α_d^n and the number of included perfect matchings by N_d.

Each G_d induces an undirected subgraph of H by including an edge (s_i, t_j), if the arc (s_i, t_j), the arc (t_j, s_i) or both of them are in G_d. As edge cost C_{ij} in the induced subgraph we choose the minimum of Y_{ij} and Z_{ij}, which has the distribution function F as shown above.

Hence, G_n induces the complete undirected graph H and the optimal value of the $LBAP$ on G_n is equal to Z_n. Obviously, Z_n is less than or equal to α_2^n, if a perfect matching exists in G_2. Otherwise, the same inequality holds for α_3^n provided that a perfect matching exists in G_3. If no perfect matchings exist in these two sparse subgraphs, then α_n^n, the maximum weight arc of G_n, is a trivial upper bound for Z_n.

The resulting elementary inequality

$$\begin{aligned} E[Z_n] \leq\ & E[\alpha_2^n | N_2 \geq 1] \cdot Pr(N_2 \geq 1) \\ &+ E[\alpha_3^n | N_3 \geq 1, N_2 = 0] \cdot Pr(N_3 \geq 1, N_2 = 0) \\ &+ E[\alpha_n^n | N_3 = 0] \cdot Pr(N_3 = 0) \end{aligned} \tag{2}$$

will be used as a basis for the proof of our statement. In contrary to the $LSAP$ analysis, the conditioning has no effect on the expected value of α_d^n, because the are weights are independent from the structure of a graph G_d.

2.2. The main result

We are mainly interested in the analysis of $E[\alpha_2^n]$ as the probabilities in the second and third term of inequality (2) are rather small. To compute the distribution function of α_2^n we define $\alpha_2^n(x) := Pr(\alpha_2^n \leq x)$ and state

Observation 2

$$\alpha_2^n(x) = [1 - (1 - G(x))^n - n\,G(x)(1 - G(x))^{n-1}]^{2n}.$$

Proof: For each node s_i resp. t_i, $1 \leq i \leq n$, we choose according to the construction of G_2 the two smallest emanating arcs and denote their weights by $Y_{i\,(1)}$ and $Y_{i\,(2)}$ (respectively $Z_{i\,(1)}$, $Z_{i\,(2)}$).

To determine $Pr(Y_{i\,(2)} > x)$ we conclude that the second largest out of n items is greater x, if either all n items are greater than x or only one out of n possible items is smaller or equal x and all $n - 1$ other items greater than x. This yields

$$\begin{aligned} Pr(Y_{i\,(2)} \leq x) &= 1 - Pr(Y_{i\,(2)} > x) \\ &= 1 - (1 - G(x))^n - n\,G(x)(1 - G(x))^{n-1}. \end{aligned}$$

G_2 consists only of arcs with weights $Y_{i\,(1)}$, $Y_{i\,(2)}$, $Z_{i\,(1)}$ and $Z_{i\,(2)}$. Hence, the largest arc weight of G_2 is less or equal to $\max\{Y_{i\,(2)}, Z_{i\,(2)} | 1 \leq i \leq n\}$. The distribution function of this maximum over $2n$ independent and identical distributions is derived by raising $Pr(Y_{i\,(2)} \leq x)$ to the $2n$-th power. □

In order to gain transparency of the proof of our main theorem we first show that the values of α_2^n ar likely to be found very near the lower end of the distribution interval with high probability.

LEMMA 3:

$$\lim_{n \to \infty} \alpha_2^n(x) = 1 \quad \textit{pointwise}\ \forall x\ s.t.\ G(x) \in (0, 1).$$

Proof: Using the sum representation of $\ln(1-z)$ for $|z| < 1$ and applying basic estimations we get

$$\begin{aligned}\lim_{n\to\infty} \alpha_2^n(x) &= \exp[\lim_{n\to\infty} 2n \ln[1-(1-G(x))^n - n\,G(x)(1-G(x))^{n-1}]] \\ &= \exp\left[-\lim_{n\to\infty} 2n \sum_{k=1}^{\infty} 1/k \right. \\ &\qquad \left. \times [(1-G(x))^n + n\,G(x)(1-G(x))^{n-1}]^k\right] \\ &\geq \exp\left[-\lim_{n\to\infty} 2n \sum_{k=1}^{\infty} (1-G(x))^{k(n-1)}[1+(n-1)G(x)]^k\right] \\ &\geq \exp\left[-2\lim_{n\to\infty} \sum_{k=1}^{\infty} n^{k+1}(1-G(x))^{k(n-1)}\right] \\ &\geq \exp[-2\lim_{n\to\infty} \mathcal{O}(n^2(1-G(x))^{n-1})] \\ &= \exp(0) = 1\end{aligned}$$

because $\lim_{n\to\infty} n^k z^n = 0$ for $|z| < 1$ and terms of higher order are dominated. □

After these preparations we finally state our main result:

THEOREM 4: *If* $\sup\{x|F(x) < 1\} < +\infty$ *then the optimal solution* Z_n *of a random LBAP with cost coefficients distributed according to a distribution function* F *satisfies*

$$\lim_{n\to\infty} E[Z_n] = \inf\{x|F(x) > 0\}.$$

Proof: To simplify the notation we define $a := \inf\{x|F(x) > 0\}$ and $b := \sup\{x|F(x) < 1\}$ as bounds of the cost coefficients. We first show that $\lim_{n\to\infty} E[\alpha_2^n] = a$. Note that a is also a lower bound for the cost in $G(n, d)$. With Observation 2 we get

$$E[\alpha_2^n] = b - \int_a^b \alpha_2^n(x)\,dx$$

With Lemma 3 we apply the dominated convergence principle and conclude

$$\begin{aligned}\lim_{n\to\infty} E[\alpha_2^n] &= b - \int_a^b \lim_{n\to\infty} \alpha_2^n(x)\,dx \\ &= b - \int_a^b dx \\ &= a.\end{aligned}$$

Taking into account that all edge weights are smaller than b and hence finite constants, also $E[\alpha_3^n]$ and $E[\alpha_n^n]$ are bounded by the constant b for any n. Hence, the asumptotic behaviour of inequality (2) depends only on the probabilities which can be bounded by Lemma 1. We get

$$\begin{aligned} Pr(N_2 \geq 1) = Pr(n, 2) &\geq 1 - \frac{1}{5n}, \\ Pr(N_3 \geq 1, N_2 = 0) \leq Pr(N_2 = 0) &\leq \frac{1}{5n} \\ \text{and} \quad Pr(N_3 = 0) = 1 - Pr(n, 3) &\leq \frac{81}{122} n^{-4}. \end{aligned}$$

Obviously, the first probability tends to 1, the second as well as the third probability tends towards 0. We conclude that $\lim_{n\to\infty} E[Z_n] \leq a$ and that $E[Z_n]$ is the cost of an edge distributed according to F and therefore cannot be smaller than a. These arguments also hold, if $\inf\{x|F(x) > 0\} = -\infty$. □

Remark: A well known fact from random graph theory can be seen as a consequence of Theorem 4. Let $G_{n,p}$ be the class of random graphs G with n vertices such that every edge is contained in G with probability p. We then have for every $p_0 \in (0, 1]$

$$\lim_{n\to\infty} P_r(G_{n,p_0} \text{ contains a perfect matching}) = 1.$$

An interpretation can be given by setting $F(x) = x,\ x \in [0, 1]$. According to Theorem 4 there exists an n_0 such that $E[Z_n] < p_0$ for all $n > n_0$. If we construct a subgraph of H containing all edges with weights less than $E[Z_{n_0}]$, this subgraph has a smaller expected number of edges than G_{n,p_0} and still contains a perfect matching with high probability.

3. UPPER AND LOWER BOUNDS

To illustrate the behaviour of the expected optimal solution value, we state upper and lower bounds of $E[Z_n]$ as functions of n. They can give a clearer picture about the actual situation in a given application than the asymptotic result. As we mentioned above, the optimal assignment of an $LBAP$ depends only on the ordering of the cost-coefficients. In this section we restrict ourselves to costs distributed uniformly between 0 and 1, *i.e.* we set the distribution function $F(x) = x$ for $x \in [0, 1]$ as any other distribution can be transformed into a uniform distribution without changing the order of the random costs.

LEMMA 5: *For* $n > 78$

$$E[Z_n] < 1 - \left[\frac{2}{n(n+2)}\right]^{2/n} \frac{n}{n+2} + \frac{123}{610\,n}.$$

Proof: In (2) we replace $Pr(N_2 \geq 1)$ by 1 and get by bounding $E[\alpha_3^n]$ and $E[\alpha_n^n]$ by 1

$$E[Z_n] \leq E[\alpha_2^n] + Pr(N_2 = 0) + Pr(N_3 = 0).$$

With Lemma 1 the two probabilities can be bounded by

$$Pr(N_2 = 0) + Pr(N_3 = 0) \leq \frac{1}{5\,n} + \frac{81}{122}\,n^{-4} \leq \frac{123}{610\,n}.$$

Observation 2 yields

$$E[\alpha_2^n] = 1 - \int_0^1 [1 - (1-x)^{n/2} - n(1-(1-x)^{1/2})(1-x)^{(n-1)/2}]^{2n}\,dx.$$

Applying the Bernoulli inequality and taking $t \in (0, 1)$ we get

$$\begin{aligned} E[\alpha_2^n] &< 1 - \int_{1-t}^1 [1 - 2n(1-x)^{n/2} \\ &\qquad - 2n^2(1-(1-x)^{1/2})(1-x)^{(n-1)/2}]\,dx \\ &= 1 - t - \frac{4n}{n+2}(1-n)(1-x)^{(n+2)/2}\Big|_{1-t}^{1} \\ &\quad - \frac{4n^2}{n+1}(1-x)^{(n+1)/2}\Big|_{1-t}^{1} \\ &= 1 - t + \frac{4n}{n+2}(1-n)\,t^{(n+2)/2} + \frac{4n^2}{n+1}\,t^{(n+1)/2} \\ &= 1 - t + n\,t^{(n+2)/2}\left(\frac{4-4n}{n+2} + \frac{4n}{n+1}\,t^{-1/2}\right). \end{aligned}$$

The last factor is smaller than 1 if

$$t^{1/2} \geq \frac{4n^2 + 8n}{5n^2 + 3n - 2}. \tag{3}$$

If this condition holds we have

$$E[\alpha_2^n] < 1 - t + n\, t^{(n+2)/2}.$$

To minimize this expression we set $t = \left(\frac{2}{n(n+2)}\right)^{2/n}$. Thereby condition (3) is fulfilled for $n > 78$. Substituting this choice of t (which is always less than 1) in the last bound of $E[\alpha_2^n]$ completes the proof. □

THEOREM 6: *Let $B(x, y)$ be the Beta function. Then*

$$E[Z_n] \geq 1 - n\, B\left(n,\, 1 + \frac{1}{n}\right) = \frac{\ln n + 0.5749\cdots}{n} + \mathcal{O}\left(\frac{\ln^2 n}{n^2}\right).$$

Proof: We use a natural lower bound of $E[Z_n]$, namely the maximum of the minimal edge weights of each matrix row *i.e.*

$$L_n = \max_{1 \leq i \leq n} \{\min\{c_{ij} | 1 \leq j \leq n\}\}.$$

As each row minimum is distributed with $1 - (1 - x)^n$ for $x \in [0, 1]$, the distribution function of L_n is $[1 - (1 - x)^n]^n$, $x \in [0, 1]$. Hence a lower bound is

$$E[Z_n] \geq 1 - \int_0^1 [1 - (1 - x)^n]^n\, dx.$$

Writing the integral as a binomial sum we get

$$\begin{aligned} E[Z_n] &\geq 1 - \sum_{k=0}^{n} \binom{n}{k} (-1)^k \int_0^1 (1 - x)^{nk}\, dx \\ &= 1 - \sum_{k=0}^{n} \binom{n}{k} (-1)^k \frac{1}{nk + 1}. \end{aligned}$$

Using an identity from complex analysis which can be verified by computing the residues (*cf.* e.g. [13], Ch. VI) we further have

$$\mathcal{I} := \sum_{k=0}^{n} \binom{n}{k} (-1)^k \frac{1}{nk+1} = \frac{1}{2\pi i} \oint_C f(z)\, dz, \tag{4}$$

with

$$f(z) := \frac{(-1)^n\, n!}{(nz+1)\, z\,(z-1)\cdots(z-n)}\cdot$$

and integrating along a path C enclosing the poles $0, 1, \cdots, n$. Extending the path around the singularity $z = -1/n$ to $\overline{C}$ we have to subtract the residue at the added pole to retain the equality.

$$\mathcal{I} = \frac{1}{2\pi i} \oint_{\overline{C}} f(z)\, dz - \mathrm{Res}\,[f(z),\, z = -1/n] \tag{5}$$

By extending $\overline{C}$ further to a circle with radius R the integral in (5) tends to 0 as R tends to infinity, because no other singularities exist and $f(z)$ can be bounded on the circle by $|f(z)| < c/R^{n+2}$ with a constant c. Evaluation of the residue at $z = -1/n$ yields

$$\mathcal{I} = \frac{n!\, n^n}{(n+1)(2n+1)\cdots(n^2+1)} = n!\, n^n \prod_{k=1}^{n} \frac{1}{kn+1},$$

and expressing the finite product by the Gamma function

$$\mathcal{I} = \frac{\Gamma(n+1)\,\Gamma\left(1+\dfrac{1}{n}\right)}{\Gamma\left(n+1+\dfrac{1}{n}\right)}.$$

Applying $B(x, y) = \dfrac{\Gamma(x)\,\Gamma(y)}{\Gamma(x+y)}$ the Gamma function can be replaced by the Beta function which proves the first bound.

For an exact estimation of the behaviour of this bound we use formula 6.1.35 in [1]

$$\Gamma\left(1+\frac{1}{n}\right) = 1 - 0.5749\ldots\frac{1}{n} + \mathcal{O}\left(\frac{1}{n^2}\right)$$

and apply formula 6.1.47 from [1]:

$$\frac{\Gamma(n+1)}{\Gamma\left(n+1+\frac{1}{n}\right)} = n^{-1/n}\left(1-\frac{1}{2n^2}+\mathcal{O}\left(\frac{1}{n^3}\right)\right)$$

$$= \exp-\left(-\frac{1}{n}\ln n\right)\left[1+\mathcal{O}\left(\frac{1}{n^2}\right)\right]$$

$$= \left[1-\frac{\ln n}{n}+\mathcal{O}\left(\frac{\ln^2 n}{n^2}\right)\right]\left[1+\mathcal{O}\left(\frac{1}{n^2}\right)\right]$$

$$= 1-\frac{\ln n}{n}+\mathcal{O}\left(\frac{\ln^2 n}{n^2}\right)$$

Putting things together the claimed bound follows. □

Table 1 illustrates the bounds and their differences for special values of n. It can be seen by the differences that a reasonable interval for $E[Z_n]$ is given even for smaller values of n.

TABLE 1

Upper and lower bounds for the expected value of the optimal solution of a random LBAP of size n, an estimation of the lower bound and the gap between upper and lower bound.

n	Upper bound	Lower bound	$(\ln n + 0.5\ldots)/n$	Upper – Lower
80	0.2057	0.0601	0.0620	0.1456
100	0.1755	0.0505	0.0518	0.1250
150	0.1301	0.0366	0.0372	0.0935
200	0.1044	0.0289	0.0294	0.0755
300	0.0758	0.0207	0.0209	0.0551
400	0.0601	0.0163	0.0164	0.0438
500	0.0501	0.0135	0.0136	0.0366
600	0.0431	0.0117	0.0116	0.0314
800	0.0339	0.0090	0.0091	0.0249
1000	0.0281	0.0075	0.0075	0.0206
1500	0.0198	0.0052	0.0052	0.0146
2000	0.0155	0.0041	0.0041	0.0114

4. SOLVING THE *LBAP* IN LINEAR EXPECTED TIME

We give a straight forward algorithm using a simple sparse subgraph $\tilde{H}$ of H which contains a matching with high probability. Therefore, we assume that all elements of the cost matrix are drawn independently from the same arbitrary distribution. (In fact it is sufficient, if the ordering of any subset of cost coefficients generates a random permutation of their original indices.)

The expected running time of this algorithm is linear in the number of edges.

Algorithm

1. Choose the $2n \log n$ cheapest edges of H and denote them by $\tilde{E}$(*i.e.* $\max\{c_{ij} | (i, j) \in \tilde{E}\} \leq \min\{c_{ij} | (i, j) \in E \backslash \tilde{E}\}$).
   Let $\tilde{H} = (S \cup T, \tilde{E})$.
2. Solve an $LBAP$ on the sparse subgraph $\tilde{H}$.
   If a perfect matching in $\tilde{H}$ is thereby detected then **stop**.
   Otherwise **goto 3**.
3. {Executed only with low probability}.
   Solve the $LBAP$ on the complete graph H.
   stop.

Analysis

Step 1 can be performed in $O(n^2)$ time using a version of the linear median algorithm to find the $2n \log n$-smallest edge.

To solve the $LBAP$ on a graph with $2n \log n$ edges in Step 2 we employ the method proposed by Gabow and Tarjan [9]. This algorithm starts like the classical threshold algorithm but stops before reaching a perfect matching which is then constructed by computing augmenting paths. It takes $O(m\sqrt{n \log n})$ time, where m denotes the number of edges in the graph. This yields an $O((n \log n)^{3/2})$ time bound for Step 2.

To analyze the expected running time of Step 3 we distinguish two cases:

Case I: $\tilde{H}$ contains isolated vertices

As the $2n \log n$ edges of $\tilde{E}$ are a random selection of the total edge set, we have an evolutionary random graph process, which is the consecutive insertion of random edges into a graph consisting only of the vertex set at the beginning.

It is known that after randomly inserting $n \log n$ edges in a general graph the expected value of the number of isolated vertices is $E[X] \sim 1/n$ (*see* Palmer [16], (3.1.5) and Theorem 3.1.1). Slightly modifying the arguments and bounding techniques used in [16] the same result can be attained for bipartite graphs. (Note that our bipartite graph has $2n$ vertices. Hence, to be precise we have to take $2n \log n$ edges.)

Markov's inequality

$$Pr(X \geq t) \leq \frac{E[X]}{t},$$

which holds for every random variable $X \geq 0$ and every $t > 0$, guarantees that the probability of the occurrence of an isolated node in $\tilde{H}$ is less than or equal $1/n$. (Let X be the number of isolated vertices and $t = 1$).

Hence the probability of the occurrence of Case I is less than $1/n$. Obviously, we will not have a matching of size n in this case and therefore always have to perform Step 3 which takes $O(n^{5/2}(\log n)^{1/2})$ time (*see* [9]).

The expected running time of Case I is $O(n^{3/2}(\log n)^{1/2})$.

Case II: $\tilde{H}$ contains no isolated vertices

In this case we have a bipartite graph generated by a random graph process which has no isolated vertices. The probability that such a graph does not contain a complete matching is bounded in Bollobás and Thomason [3], Theorem 5, by

$$\sum_{k=2}^{n_1} (e \log n)^{3k} n^{1-k+k^2/n},$$

where $n_1 = \lfloor (n+1)/2 \rfloor$. Lemma 7 below bounds this sum with $O(1/\sqrt{n} \log n)$.

Therefore, the probability of the execution of Step 3 is less than $O(1/\sqrt{n} \log n)$. Using the Gabow and Tarjan algorithm Step 3 takes again $O(n^{5/2}(\log n)^{1/2})$ time. Hence the expected running time of Case II is $O(n^2)$.

The total expected running time of the algorithm is dominated by $O(n^2)$ which is linear in the number of edges of the complete graph. Unless the underlying graph has some known special structure every edge cost may influence the optimal solution and thus has to be inspected, which makes $O(n^2)$ steps necessary in every algorithm.

Remark: The algorithm provides not only a good bound in terms of complexity, but is also very efficient and easy to use. For practical application we would suggest to use a specialized algorithm for sparse graphs in Step 2. Possibilities to use the effort spent on Step 2 in the solution of Step 3 should be exploited.

LEMMA 7: *Let* $a_k := (e \log n)^{3k}\, n^{1-k+k^2/n}$. *Then*

$$\Sigma := \sum_{k=2}^{n_1} a_k = O\left(\frac{1}{\sqrt{n}\,\log n}\right).$$

Proof: Σ can be written as $a_2 + \sum_{k=3}^{n_1} a_k$. We observe that for every n

$$\ln a_k = \frac{\ln n}{n} k^2 + [3 \ln (e \log n) - \ln n] k + \ln n$$

is a convex function in k for $3 \leq k \leq n_1$. Hence its maxima are attained at the boundary of the interval and we get a rough estimation by $a_k \leq \max\{a_3, a_{n_1}\} \leq a_3 + a_{n_1}$. For simplicity we consider the case of n even. Thereby we have the inequality

$$\begin{aligned}\Sigma &< a_2 + n\,(a_3 + a_{n_1})\\ &= (e \log n)^6\, n^{-1+4/n} + (e \log n)^9\, n^{-1+9/n} + (e \log n)^{3n/2}\, n^{2-n/4}\end{aligned}$$

Each of these three terms is bounded by

$$O\,(1/\sqrt{n\,\log n}) \quad \text{as} \lim_{n\to\infty} n^{1/n} = 1. \quad \square$$

ACKNOWLEDGEMENT

The author would like to thank Rainer E. Burkard, Wolfgang Ring, Günter Rote and Gerhard J. Woeginger for valuable suggestions and two referees who helped to improve the paper considerably.

REFERENCES

1. M. ABRAMOWITZ and I. A. STEGUN, *Handbook of Mathematical functions*, Dover Publications, New York, 1965.
2. B. BOLLOBÁS, *Random Graphs*, Academic Press, 1985.
3. B. BOLLOBÁS and A. THOMASON, Random graphs of small order, Random Graphs'83, *Annals of Discrete Math.*, 1985, *28*, pp. 47-97.
4. R. E. BURKARD and U. DERIGS, Assignment and Matching Problems: Solution Methods with FORTRAN-Programs, *Springer Lecture Notes in Economics and Mathematical Systems*, 1980, *184*.
5. U. DERIGS, The shortest augmenting path method for solving assignment problems, *Annals of Operations Research*, 1985, *4*, pp. 57-102.

6. U. Derigs, Programming in networks and graphs, *Springer Lectures Notes in Economics and Mathematical Systems*, 1988, *300*.
7. P. Erdös and A. Rényi, On random matrices, *Publ. Math. Inst. Hungar. Acad. Sci.*, 1964, *8*, pp. 455-461.
8. J. B. G. Frenk, M. Van Houweninge and A. H. G. Rinnooy Kan, *Order statistics and the linear assignment problem*, Report 8609/A, Econometric Institute, Erasmus University, Rotterdam, The Netherlands, 1986.
9. H. N. Gabow and R. E. Tarjan, Algorithms for two bottleneck optimization problems, *J. of Algorithms*, 1988, *9*, pp. 411-417.
10. J. E. Hopcroft and R. M. Karp, An $n^{5/2}$ algorithm for maximum matchings in bipartite graphs, *SIAM J. Comput.*, 1973, 2, pp. 225-231.
11. R. M. Karp, An algorithm to solve the $m \times n$ assignment problem in expected time $O(mn \log n)$, *Networks*, 1980, *10*, pp. 143-152.
12. R. M. Karp, *An upper bound on the expected cost of an optimal assignment*, Technical report, Computer Sc. Div., Univ. of California, Berkeley, 1984.
13. S. Lang, *Complex Analysis*, Springer, 1985.
14. A. J. Lazarus, The assignment problem with uniform (0, 1) cost matrix. Master's thesis, Department of Mathematics, Princeton University, 1979.
15. B. Olin, Asymptotic properties of random assignment problems. PhD-thesis, Division of Optimization and Systems Theory, Department of Mathematics, Royal Institute of Technology, Stockholm, 1992.
16. E. M. Palmer, *Graphical Evolution*, J. Wiley & Sons, 1985.
17. D. W. Walkup, On the expected value of a random assignment problem, *SIAM J. Comput.*, 1979, *8*, pp. 440-442.
18. D. W. Walkup, Matchings in random regular bipartite digraphs, *Discrete Mathematics*, 1980, *31*, pp. 59-64.