THE GREEDY APPROACH FOR
THE CREW PAIRING PROBLEM:
SOME CASES OF POLYNOMIAL SUBPROBLEMS

CAHIER N° 159
janvier 1999

Laurent ALFANDARI

received: October 1998.

1 LAMSADE, Université Paris-Dauphine, Place du Maréchal De Lattre de Tassigny, 75775 Paris Cedex 16, France (alfanla@lamsade.dauphine.fr).
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Résumé</td>
<td>i</td>
</tr>
<tr>
<td>Abstract</td>
<td>i</td>
</tr>
<tr>
<td>1. The crew pairing problem</td>
<td>1</td>
</tr>
<tr>
<td>2. Linear programming and column generation</td>
<td>2</td>
</tr>
<tr>
<td>3. The greedy approach</td>
<td>3</td>
</tr>
<tr>
<td>3.1 Case without constraints</td>
<td>4</td>
</tr>
<tr>
<td>3.2 Case of resource constraints</td>
<td>5</td>
</tr>
<tr>
<td>4. Conclusion</td>
<td>7</td>
</tr>
<tr>
<td>References</td>
<td>8</td>
</tr>
</tbody>
</table>
L'approche gloutonne pour le problème de construction de rotations: cas de polynomialité du sous-problème

Résumé

Nous reprenons l'algorithme glouton classique pour le Set Covering en l'appliquant au problème de construction de rotations et montrons que, dans certains cas, la résolution du sous-problème - consistant à sélectionner itérativement la rotation minimisant le ratio “coût / nombre de tâches non couvertes” - s'effectue polynomiallement en le nombre de tâches par un algorithme de plus court chemin à étiquettes multiples dans un graphe. Nous discutons également des liens qui unissent l'approche gloutonne et l'approche génération de colonnes, cette dernière pouvant également faire appel à la résolution d'un sous-problème de plus court chemin.

Mots-clés : construction de rotations, glouton, plus court chemin, génération de colonnes.

The greedy approach for the Crew Pairing Problem: some cases of polynomial subproblems

Abstract

We apply the usual greedy algorithm for the set covering to the Crew Pairing Problem and show that in some cases the subproblem - which consists of selecting the pairing minimizing a ratio “cost / number of uncovered tasks”- is solvable in polynomial time by a multi-label shortest path algorithm in a graph. We also discuss the links between the greedy approach and the column generation approach, as this latter one also often requires the solving of a shortest path subproblem.

Keywords : crew pairing, set covering, greedy, shortest path, column generation.
1 The Crew Pairing Problem

Given a set of tasks \( V = \{1, 2, \ldots, n\} \) to be covered by crews (flights by stewards and hostesses, bookings by drivers...), and given for every task \( i \in V \)
2 Linear programming and column generation

For $p \in C_A$, we define variable $x_p = 1$ if path $p$ is chosen in the solution, 0 otherwise, and parameter vector $C_p$, called "column", with components $C_{p,i} = 1$ if vertex $i \in V$ belongs to path $p$, 0 otherwise, $i = 1, \ldots, n$. The Crew Pairing Problem (CPP) is defined by the following integer linear program:

$$\min \sum_{p \in C_A} c_p x_p$$

s.e. $\forall i = 1, \ldots, n \sum_{p \in C_A} C_{p,i} x_p \geq 1$ \quad (P)

$x_p \in \{0; 1\}$, $p \in C_A$

The continuous relaxation of problem $P$ will be noted $\bar{P}$. Let us remark that exhaustive explicitation of $C_A$ is impossible for large-scale problems, as the total number of paths $|C_A|$ usually grows exponentially in $n$. The column generation approach can get round this obstacle by solving $\bar{P}$, at each step $t$, on a restricted subset of columns $C_t \subset C_A$, by an adaptation of the simplex algorithm. A convincing example of this technique can be found in [9]. The key idea is the following. Given a feasible solution $C_0 \subset C_A$ for $P$, one solves the continuous program $\bar{P}_0$, which is the restriction of $\bar{P}$ to the set of columns $C_0$:

$$\min \sum_{p \in C_0} c_p x_p$$

s.e. $\forall i = 1, \ldots, n \sum_{p \in C_0} C_{p,i} x_p \geq 1$ \quad ($\bar{P}_0$)

$x_p \in [0; 1]$, $p \in C_0$

Solving $\bar{P}_0$ produces a vector of simplex multipliers $(\pi_1^0, \pi_2^0, \ldots, \pi_n^0)$ associated to tasks $1, 2, \ldots, n$. One looks then for the column with smallest negative reduced cost among all columns of $C_A$ (so that the optimality of the procedure is guaranteed), i.e, column

$$C_{p^*} = \arg \min_{C_p \in C_A} \{ \pi_i^0 - \sum_{i=1}^n \pi_i^0 C_{p,i} \}$$

For some sets of constraints $A$, the column with smallest negative reduced cost $C_{p^*}$ can be found in time polynomial in $n$, without explicitating the whole set $C_A$ of columns, by a shortest path algorithm in $G^2$, where a new cost $d^0(i,j) = d(i,j) - \pi_j^0$ is associated to arc $(i,j) \in E$. The column $C_{p^*}$ is added then to the set $C_0$ - or even better, all columns with negative reduced cost are added - forming a new set of columns $\bar{C}_A$, and the relaxed program is solved again on this new set of columns.
The columns added at step $t$ correspond to the paths with negative reduced cost computed by a shortest path algorithm in graph $G$ with valuations on arcs $d^t(i, j) = d(i, j) - \pi_j^t$. The continuous optimum is obtained as soon as, $\forall p \in C_A, c_p - \sum_{i=1}^{n} \pi_i^t C_{p,i} \geq 0$ (all reduced costs are non-negative), i.e as soon as the label of sink $b'$ is positive in the shortest path computation.

Let us remark that the above procedure is nothing but the simplex algorithm, with the particularity that the column with minimum reduced cost is not found by exhaustive enumeration of the set of columns, but by solving a polynomial optimization subproblem (shortest path). By the way, this method solves the continuous version of the problem; in case of a non-integer solution, finding an optimal integer solution from the continuous solution may require much extra time computation. We now use the greedy master-slave technique in order to construct a feasible integer solution with performance guarantee. This greedy approach shares with the column generation approach the nice property that in both cases, the “master” problem calls a “slave” subproblem, polynomial in $n$ for some sets of constraints, by implicitly exploring the set of feasible paths.

3 The greedy approach

Consider the following greedy heuristic (Greedy in short) for the problem CPP. At each step, find a path minimizing the ratio “cost of the path / number of uncovered tasks of the path”, add this
This bound is derived from the analysis of the classical greedy algorithm for the Set Covering Problem by Chvátal ([5]). Greedy is, to our knowledge, the only heuristic for CPP with worst-case performance guarantee. Baker implemented Greedy for some Crew Pairing problems applied to air transportation([4]); his results are reported in the following table.

| n   | $|C_A|$ | OPT   | Greedy | %    |
|-----|------|-------|--------|------|
| 36  | 506  | 4424  | 4424   | 0    |
| 180 | 4534 | 27803 | 28564  | 2.7  |
| 200 | 3208 | 30110 | 32095  | 6.5  |
| 335 | 5290 | 48089 | 51344  | 6.7  |
| 376 | 1245 | 74679 | 74679  | 0    |
| 426 | 2780 | 35974 | 36550  | 1.5  |
| 180 | 2275 | 28759 | 29876  | 3.8  |
| 180 | 3431 | 27936 | 33217  | 18.8 |
| 376 | 4802 | 72935 | 81979  | 12.4 |

The computational results above look convincing enough to show the practical efficiency of Greedy. Our contribution is more of theoretical order : whereas the set of feasible paths $C_A$ was a data in [4], and the complexity of Greedy was polynomial in $|C_A|$, i.e in $O(2^n)$ at worst case (the optimal path $p^*$ of definition 3 was found by exhaustive enumeration of $C_A$), we exhibit cases where the set of feasible paths $C_A$ is not known a priori, and where Greedy is computed in time polynomial in $n$, the number of tasks, by solving the slave in time polynomial in $n$.

### 3.1 Case without constraints

Let us assume that every path from $b$ to $V$ is feasible in graph $G$. We consider then the set of paths $C_{A_i} = C$. This case is not a trivial case: the rest time constraints in [9] refer to the present case via transforming the initial graph. Now we prove that the slave problem is solvable in time polynomial in $n$ by showing that, graph $G$ being acyclic, the following extension of Bellman’s sub-optimality principle holds for CPP($A_1$): if $(b_i, t_1, i_2, \ldots, i_k, j)$ is a shortest path among the paths from $b$ to $j$ containing exactly $k$ uncovered vertices, then $(b_i, i_1, i_2, \ldots, i_t)$ is a shortest path among the paths from $b$ to $i_t$ containing exactly $k - u(j)$ uncovered vertices.

Consider the following variant of the classical shortest path algorithm for acyclic graphs (cf. [7], p.49), where label $\lambda^k(j)$ is the value of a shortest path among paths from $b$ to $j$ containing exactly $k$ uncovered vertices of $V$, $k = 0, \ldots, s$, with $s = \max_{p \in C_A} |V(p)|$.

**Algorithm /S1(u)/**

1. $S \leftarrow \{b\}$;
2. $\lambda^0(b) \leftarrow 0$; $\lambda^k(b) \leftarrow \infty$, $k = 1, \ldots, s$;
3. while there exists a vertex $j \not\in S$ such that $\Gamma^{-1}(j) \subset S$ do
4. \[
S \leftarrow S \cup \{j\};
\]
5. compute, for $k = 1, \ldots, s$, label $\lambda^k(j) \leftarrow \min_{i \in \Gamma^{-1}(j)} \{\lambda^{k-u(j)}(i) + d(i,j)\}$;
6. compute $\lambda^0(j) \leftarrow \min_{i \in \Gamma^{-1}(j)} \{\lambda^0(i) + d(i,j)\}$ if $u(j) = 0$, else $\lambda^0(j) \leftarrow \infty$;
7. output path $p^*$ minimizing $\lambda^k(b')/k$ for $k = 1, \ldots, s$;

**END OF /S1(u)/**

*the value $s$ can be computed by a longest path algorithm in $G$*
Theorem 1. Algorithm S1 solves in $O(n^3)$ the slave problem associated to CPP(A1).

Proof: The complexity of algorithm S1 is in $O(ns\Delta)$, where $\Delta = \max_{j \in V \cup \{b\}} |\Gamma^{-1}(j)|$, i.e in $O(n^3)$. It remains to prove that the algorithm is optimal. As the cost of $p^*$ is $\min_k \lambda^n(b')/k$, optimality is proved if, for $j \in V \cup \{b\}$, label $\lambda^k(j)$ is indeed the value of a shortest path among paths from $b$ to $j$ containing exactly $k$ uncovered vertices. We shall first show that $\lambda^k(j)$ is a lower bound of this value, then we show this bound is achieved. Let $p = (b, i_1, i_2, \ldots, i_t, j)$ be a path from $b$ to $j$ in $G$ satisfying $\sum_{1 \leq l \leq t} u(i_l) + u(j) = k$. As $\forall j, i \in \Gamma^{-1}(j)$, $\forall k$, $\lambda^k(j) \leq \lambda^{k-u(j)}(i) + d(i, j)$, it follows that

\[
\lambda^k(j) \leq \lambda^{k-u(j)}(i_j) + d(i_j, j).
\]

\[
\lambda^{k-u(j)}(i_j) \leq \lambda^{k-u(j)-u(i_j)}(i_{t-1}) + d(i_{t-1}, i_t).
\]

\[
\lambda^{k-u(j)-u(i_j)}(i_{t-1}) \leq \lambda^{k-u(j)-u(i_{t-1})}(i_{t-2}) + d(i_{t-2}, i_{t-1}).
\]

\[
\ldots
\]

\[
\lambda^{k-u(j)-u(i_2)}(i_2) \leq \lambda^{k-u(j)-u(i_1)}(i_1) + d(i_1, i_2).
\]

\[
\lambda^{k-u(j)-u(i_1)}(i_1) \leq \lambda^{k-u(j)}(i) + d(b, i_1).
\]

Summing these $t + 1$ inequalities term by term we get

\[
\lambda^k(j) \leq \lambda^{k-u(j)}(b) + d(b, i_1) + d(i_1, i_2) + \ldots + d(i_{t-1}, i_t) + d(i_t, j).
\]

Since $u(j) + \sum_{1 \leq l \leq t} u(i_l) = k$ and $\lambda^0(b) = 0$, we obtain for every path $p$ from $b$ to $j$,

\[
\lambda^k(j) \leq c(p).
\]

The value $\lambda^k(j)$ is then a lower bound of the value of a shortest path among paths from $b$ to $j$ containing exactly $k$ uncovered vertices. Moreover, this bound is attained; we construct a path $p^*$ from $b$ to $j$ of value $\lambda^k(j)$ the following way, classical when dealing with shortest path problems: start from $j$ and, going back to the root $b$, connect to the predecessor $i$ of $j$ such that $\lambda^{k-u(j)}(i) + d(i, j)$ is minimum, then connect $i$ to the predecessor $h$ of $i$ such that $\lambda^{k-u(j)-u(i)}(h) + d(h, i)$ is minimum, and reiterate until $b$ is attained.

Let us consider now the case when all paths from $b$ to $b'$ are not feasible.

3.2 Case of resource constraints

Assume that a pairing $p$ is feasible if it contains at most one lunch period and one night period, and if the total number of effective worked hours $TAT(p) - TPT(p)$ is at most 30 hours. Typically, given $q$ resources, a path is feasible if, for $h = 1, \ldots, q$, the cumulative amount of resource $h$ on this path is lower than a given threshold $m_h$. Let $r_h(e) \in \mathbb{N}$ denote the consumption of resource $h$ on arc $e$, $m = (m_1, \ldots, m_q)$, and define the following problem.

Definition 4. The problem CPP(A2) is the restriction of CPP to the set of paths $C_{A2} = \{ p \in C : \exists r_1 : E \rightarrow N, \ldots, \exists r_q : E \rightarrow N, \exists m \in \mathbb{N}^q, \sum_{e \in E(p)} r_h(e) \leq m_h, h = 1, \ldots, q \}$.

In the restricted case where all values $m_h$ are low, we propose an exact algorithm solving the slave problem, based on the following extension of Bellman's sub-optimality principle for CPP(A2): if $(b, i_1, i_2, \ldots, i_t, j)$ is a shortest path among paths from $b$ to $j$ containing $k$ uncovered vertices and for which the cumulative amount of resource $h$ is $x_h$ for $h = 1, \ldots, q$, then $(b, i_1, i_2, \ldots, i_t, j)$ is a shortest path among paths from $b$ to $j$ containing $k - u(j)$ uncovered vertices and for which the cumulative amount of each resource $h$ is $x_h - r_h(i_t, j)$.
Let us note for \((i, j) \in E, \textbf{r}_{ij} = (r_{1}(i, j), r_{2}(i, j), \ldots, r_{q}(i, j))\), and for \(x, y \in \mathbb{N}^{q}, x \leq y\) iff \(x_{h} \leq y_{h}, \forall h = 1, \ldots, q\), and \(x - y = (x_{1} - y_{1}, \ldots, x_{q} - y_{q})\). Consider now the following algorithm solving the slave problem, where label \(\lambda_{k}^{b}(j)\) is the value of a shortest path among paths \(p\) from \(b\) to \(j\) satisfying \(\sum_{\nu \in \mathcal{V}(p)} u(\nu) = k\) and \(\sum_{e \in \mathcal{E}(p)} r_{h}(e) = x_{h}, h = 1, \ldots, q\).

**Algorithm /S2(\(u\))/**

\[
\begin{aligned}
S & \leftarrow \{b\}; \\
\lambda^{0}_{b}(b) & \leftarrow 0; \\
\lambda^{k}_{x}(b) & \leftarrow \infty, k = 0, \ldots, s, x \leq m, (k, x) \neq (0,0); \\
\text{while} & \text{ there exists a vertex } j \notin S \text{ such that } \Gamma^{-1}(j) \subset S \text{ do} \\
& \quad S \leftarrow S \cup \{j\}; \\
& \quad \text{for } x \leq m \text{ do compute labels} \\
& \quad \quad \lambda^{k}_{x}(j) \leftarrow \min i \in \Gamma^{-1}(j), x \geq r_{ij} \{ \lambda^{k-u(i)}_{x-r_{ij}}(i) + d(i, j) \}^{5}, k = 1, \ldots, s; \\
& \quad \lambda^{k}_{x}(j) \leftarrow \min i \in \Gamma^{-1}(j), x \geq r_{ij} \{ \lambda^{k-u(i)}_{x-r_{ij}}(i) + d(i, j) \} \text{ if } u(j) = 0 \text{ else } \lambda^{k}_{x}(j) \leftarrow \infty; \\
& \text{end} \\
& \text{end} \\
& \text{output path } p^* \text{ minimizing } \lambda^{k}_{x}(b')/k \text{ for } k = 1, \ldots, s, x \leq m; \\
\text{end } /S2(\(u\))/
\end{aligned}
\]

\(^{5}\text{ if } \forall i \in \Gamma^{-1}(j), -(x \geq r_{ij}), \text{ then } \lambda^{k}_{x}(j) \leftarrow \infty.\)

**Theorem 2.** Algorithm S2 solves the slave problem for CPP(A2) in \(O(n^{3} \prod_{h=1}^{q}(m_{h} + 1))\).

**Proof:** The complexity of S2 is in \(O(ns \Delta \prod_{1 \leq h \leq q}(m_{h} + 1))\), i.e., in \(O(n^{3} \prod_{1 \leq h \leq q}(m_{h} + 1))\). Now, the slave problem is solved if \(\lambda^{k}_{x}(j)\) is indeed the value of a shortest path among paths \(p\) from \(b\) to \(j\) satisfying \(\sum_{\nu \in \mathcal{V}(p)} u(\nu) = k\) and \(\sum_{e \in \mathcal{E}(p)} r_{h}(e) = x_{h}, h = 1, \ldots, q\) (then a path minimizing \(\lambda^{k}_{x}(b')/k\) for \(k = 1, \ldots, s, x \leq m\) is indeed an optimal path). Let \(p = (b, \ell_{1}, \ell_{2}, \ldots, \ell_{t}, j)\) be a path satisfying the above conditions. As \(\forall j, \forall i \in \Gamma^{-1}(j), \forall k, \lambda^{k}_{x}(j) \leq \lambda^{k-u(i)}_{x-r_{ij}}(i) + d(i, j)\), we get

\[
\lambda^{k}_{x}(j) \leq \lambda^{k-u(i)}_{x-r_{ij}}(\ell_{t}) + d(\ell_{t}, j)
\]

\[
\lambda^{k-u(i)}_{x-r_{ij}}(\ell_{t}) \leq \lambda^{k-u(i)-u(i_{t})}_{x-r_{ij}-r_{i_{t-1}i_{t}}}(\ell_{t-1}) + d(\ell_{t-1}, \ell_{t})
\]

\[
\ldots
\]

\[
\lambda^{k-u(i_{1})}_{x-r_{ij}}(\ell_{1}) \leq \lambda^{k-u(i_{1})}_{x-r_{ij}r_{i_{1}}}(b) + d(b, i_{1})
\]

By summing these \(t + 1\) inequalities we obtain

\[
\lambda^{k}_{x}(j) \leq \lambda^{k-u(i_{1})-\ldots-u(i_{t})}_{x-r_{i_{1}i_{2}}-\ldots-r_{i_{t}i_{t+1}}}(b) + d(b, i_{1}) + d(i_{1}, i_{2}) + \ldots + d(i_{t}, j)
\]

Since \(u(i_{1}) + \ldots + u(i_{t}) + u(j) = k, r_{b_{1}i_{1}} + r_{i_{1}i_{2}} + \ldots + r_{i_{t}i_{t+1}} = x, \) and \(\lambda^{0}_{b}(b) = 0,\) the following inequality finally holds:

\[
\lambda^{k}_{x}(j) \leq c(p)
\]

Hence we showed that \(\lambda^{k}_{x}(j)\) is a lower bound of the value of the shortest path mentioned above, and this bound is attained by constructing the path as previously.
The complexity of S2 is polynomial if all values \( m_n \) are polynomially bounded, the algorithm running in \( O(n^3) \) if these values are fixed constants independent of \( n \). The applicability of algorithm S2 for resource constraints is thus restricted to “easy” resources (such as the number of stops, breaks, or nights spent out of the base). The case of time resources is more intricate. For instance, assume that the total effective worked time of a pairing is constrained by an upper bound, i.e., \( \exists m \in \mathbb{N}, p \in C_A \implies TAT(p) - TPT(p) \leq m \); as the minute is often chosen as time unit, \( m \) can be quite a high value for weekly pairings. On the opposite, the constraint \( TAT(p) \leq m \) is not an obstacle as \( TAT(p) \) only depends of the first task and the last task of the pairing (see [9]). Generally speaking, the Shortest Path Problem with resource constraints is NP-hard ([6], ND30).

Finally, let us remark that algorithm S2 can be easily modified in order to solve, in a column generation approach for CPP(A), the subproblem which consists of finding at step \( t \) the path with smallest reduced cost; simply compute label \( \lambda_x(j) \leftarrow \min_{i \in \Gamma^{-1}(j), x \geq r_j} \{ \lambda_x - r_j(i) + d(i, j) - \pi^f_j \} \) in the while loop, and finally output path of value \( \min_{x \leq m} \lambda_x(b') \).

4 Conclusion

The classical greedy algorithm for the set covering problem has showed great efficiency in crew pairing contexts([4]), as well as interesting complementarity with the column generation approach, since the greedy provides a good integer solution that can initialize a simplex computation. We also showed the similarity of the two approaches by exhibiting some simple sets of constraints that make the subproblem solvable in polynomial time, in both cases, by a shortest path algorithm in a graph, avoiding exhaustive explicitation of the feasible pairings. Furthermore, we conjecture that, given a set of constraints \( A \) on the feasible paths, the subproblem in the column generation approach is solvable in polynomial time if and only if the slave problem in the greedy approach is solvable in polynomial time for CPP(A). This greedy technique can be applied in fact to a larger class of covering and partitioning problems called “master-slave tractable” problems, by exploiting an unusual reduction to the set covering problem (see [2], [10] for more details).
References


