Abstract

Transpose-and-Cache Branch-and-Bound (TCBB) has shown promise in solving large single machine quadratic penalty problems. There exist other classes of single machine job sequencing problems which are of more practical importance and which are also of considerable interest in the area of AI search. In the weighted earliness tardiness problem (WET), the best known heuristic estimate is not consistent; this is contrary to the general belief about relaxation-based heuristic. In the quadratic penalty problem involving setup times (SQP) of jobs, the evaluation function is non-order-preserving. In this paper, we present the TCBB scheme to solve these problems as well. Experiments indicate that (i) for the WET problem, the TCBB scheme is highly effective in solving large problem instances and (ii) for the SQP problem, it can solve larger instances than algorithm GREC in a given available memory.

Introduction

Best-first search algorithms like A* require substantial memory to store the generated nodes. On the other hand, depth-first search algorithm uses memory linear in the depth of the search and has very low overhead but runs slower than best-first in search graphs due to generation of duplicate nodes. The trend now is to look for variants of depth-first search which would run in reasonable time to solve problems of larger sizes.

Kaindl et al. (1995) proposed the Transpose-and-Cache Branch-and-bound (TCBB) scheme to solve large instances of single machine quadratic penalty job sequencing (QP) problem. The TCBB scheme is a variant of the depth-first branch-and-bound (DFBB) scheme. The DFBB uses a tree search space and it does not utilize the large amount of memory available. The TCBB overcomes this limitation by storing the nodes in the available memory and thus tries to avoid generation of duplicate nodes. This speeds up DFBB considerably. However when memory becomes full, TCBB continues as in DFBB and does not employ any node replacement strategies.

In the weighted earliness-tardiness (WET) problem, jobs with due dates are to be sequenced on a machine such that the sum of earliness and tardiness penalty is minimized. A job completing earlier than the due date incurs an earliness penalty (inventory carrying cost) whereas a job completing later incurs a tardiness penalty (imposed by the customer). Currently known best approach for solving this problem is dynamic programming (Ventura and Weng 1995) which runs quickly out of memory. When the penalty coefficients are unity, a branch-and-bound tree search formulation (Hoogeven, Oosterhout, and Van De Velde 1994) has been suggested. Contrary to the general belief (Pearl 1984), the proposed Lagrangian-relaxation based heuristic for the problem is non-consistent. When the heuristic is not consistent, A* graph search performs poorly since a node may be expanded more than once.

The other problem we consider in this paper is the single machine quadratic penalty problem (SQP) which includes setup times of jobs (Sen and Bagchi 1996). The presence of setup times makes the evaluation function non-order preserving (Pearl, 1984) and algorithm A* using graph search space is unsuitable for the problem since no path to a node can be discarded. Algorithm GREC has been suggested by Sen and Bagchi (1996) for finding optimal solutions using graph search space. GREC could solve 20-job problems using a hash queue of 200k nodes and runs faster than tree search, but like A*, it runs out of memory for solving large problems.

In this paper, we employ the TCBB scheme to solve optimally larger instances of the WET and the SQP problems. Instead of using g-values and f-values of nodes (Kaindl et al. 1995), the scheme is rewritten using b-values of nodes to naturally back up values obtained during the search process (Sen and Bagchi 1989). The use of backed-up values can be found in a number of algorithms (Kaindl and Kainz 1997). The use of b-values not only retains the heuristic improvement feature of the algorithm but also makes the approach applicable for the SQP problem. In addition, the scheme is enhanced with a node replacement strategy which would replace the less promising nodes (Reinefeld and Marsland 1994) when memory becomes full. Experimental results encourage the use of the TCBB scheme for the problems considered in this paper.

First, we describe the TCBB scheme and present the code for the algorithm. We then evaluate its performance in the QP problem domain. Next, we describe the WET and the SQP problems respectively, and present our
experimental findings. Concluding remarks are given at the end.

The TCBB Scheme

The TCBB scheme is based on the DFBB algorithm. The scheme starts with an upper bound on the cost of the optimal solution and examines all the paths from the start to the goal node in a depth-first manner pruning paths of costs higher than the current upper bound. DFBB never attempts to store the nodes generated except that the nodes on the current path are stored in an implicit stack. On the other hand, TCBB attempts to store the expanded nodes along with their h-values and currently known best g-values if memory is available. Hence when an already stored node is encountered while searching along a different path, TCBB can use the stored g-value and decide whether the node is to be searched again. Thus like A*, it can discard paths to a node if the paths are of higher cost. This approach introduces the graph search feature in the algorithm and the performance of the algorithm becomes comparable to that of A*. The scheme utilizes a transposition table for recognizing transpositions and for caching the best values acquired dynamically (Reinefeld and Marsland 1994).

Unlike A* when memory becomes full, TCBB can still continue as in DFBB and can output optimal solution. Though TCBB cannot store new nodes in such situations, it can still continue to take advantage of already stored nodes. This feature of the scheme may make it more attractive than A*. In the QP problem domain, TCBB was shown to solve 60-job problems within reasonable time.

However, the TCBB scheme can be improved further. For very large problem instances when memory becomes full, use of effective memory management schemes may speed up the search process. Secondly, using g-values of nodes, a graph search algorithm is likely to maintain only the currently known least cost path to a node which in non-order-preserving cases, may lead to nonoptimal solutions. Instead, the scheme can be rewritten using b-values of nodes and arc costs below nodes. The b-value b(n) of a node n stores the cost of the currently known best path below node n. The use of b-values and arc costs not only helps in non-order-preserving cases, may lead to nonoptimal solutions. The scheme utilizes a transposition table for recognizing transpositions and for caching the best values acquired dynamically (Reinefeld and Marsland 1994).

Enhanced TCBB scheme

The enhanced version of the scheme is given in Figure 1. For ready reference, the new scheme is called as Enhanced TCBB (ETCBB). We explain below the working of the scheme in terms of b-values of nodes and arc costs.

The b-value b(n) of a node n store the cost of the currently known best path below node n. When node n is generated for the first time b(n) is set to h(n); b(n) increases whenever the cost of the currently known best path below node n through new computation. This feature of the scheme may make it more attractive than A*. In the QP problem domain, TCBB was shown to solve 60-job problems within reasonable time.

Figure 1: Algorithm ETCBB

node n (due to exploring the graph below node n) exceeds current value of b(n). As a result, if heuristic is admissible, h*(n) >= b(n) >= h(n) and hence use of b-value value helps avoiding repetitive search below n.

The algorithm uses the recursive function ETCBB to search the subgraph below the start node s. The current upper bound (we call it bound) on the cost of the subgraph below a node is passed as a parameter to the function. Let c(n,n_i) denote the cost of the arc (n,n_i). If for a successor n_i of node n, c(n,n_i) + b(n_i) < bound, ETCBB would search the subgraph below node n_i; the value (bound - c(n,n_i)) is passed as the current upper bound of the subgraph below the node n_i.

The local variable newbound in the function ETCBB stores the minimum of c(n,n_i) + b(n_i), computed amongst all successors of node n. The value of newbound updates the value of b(n) on return from the function. A newbound lower than the current upper bound indicates that a better solution path has been found while a higher newbound indicates that all the paths below node n have been pruned.
During execution, newbound also resets the value of bound if it acquires a value lower than that of bound.

When memory becomes full, the simplest alternative is to continue as in DFBB. This strategy was followed in TCBB by Kaindl et al. For large problems, this scheme would run slow since no new nodes get stored and no advantage can be taken out of the stored values of new nodes. A better idea would be to replace the non-promising stored nodes with newly generated promising ones. This decision is critical. Nodes selected for replacement should not be frequently regenerated and reexpanded. In addition, the procedure for selecting nodes for replacement should not significantly increase the overhead of the algorithm. We made special experiments to evaluate a number of node replacement strategies. We found out that TCBB would perform better if, instead of continuing as in DFBB, attempts to replace less recently used nodes when memory becomes full.

ETCBB employs a SOLVE-labelling procedure (Pearl, 1984) if the evaluation function is order preserving. ETCBB labels a node SOLVED if the least cost path below the node has been found. When a SOLVED node is encountered along a different path, ETCBB can use its b-value to update the current bound or may decide to prune the current path. On the other hand, if a non SOLVED node is encountered, ETCBB can use its b-value to prune the current path or may peep below the node to search deeper into the subgraph below the node. We have experimented with or without the SOLVE-labelling procedure. When memory is large, there is hardly any effect of the procedure. But with a lower node limit of 2k nodes for the quadratic penalty (QP) problem (Kaindl et al. 1995), the SOLVE-labelling procedure reduces number of nodes by 69% for 60-job problems, and by 72% for 62-job problems. This result is interesting. However, when the evaluation function is non-order preserving, the SOLVE-labelling procedure may not be in general applicable.

The performance of the ETCBB algorithm was found to be as good as, if not better, than the TCBB scheme for the quadratic penalty (QP) problem. Using reverse search and the same consistent (Pearl 1984) heuristic estimate function, TCBB and ETCBB were run on DEC Alphastation 250 4/266 to solve problems of different sizes. For each problem size, both the algorithms solved the same 100 randomly generated problem instances. Table 1 presents a part of our experimental results when the node limit was fixed at 16k nodes (Kaindl et al 1995). Our experiments indicated the following:

<table>
<thead>
<tr>
<th>Job</th>
<th>Node Gen. TCBB</th>
<th>Time (secs) TCBB</th>
<th>Node Gen. ETCBB</th>
<th>Time (secs) ETCBB</th>
</tr>
</thead>
<tbody>
<tr>
<td>56</td>
<td>139483</td>
<td>12.79</td>
<td>66543</td>
<td>6.94</td>
</tr>
<tr>
<td>58</td>
<td>181350</td>
<td>17.72</td>
<td>83723</td>
<td>9.35</td>
</tr>
<tr>
<td>60</td>
<td>243602</td>
<td>25.30</td>
<td>108558</td>
<td>13.06</td>
</tr>
<tr>
<td>62</td>
<td>131703</td>
<td>16.89</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: TCBB and ETCBB for the QP problem

- ETCBB ran faster than the TCBB scheme generating less number of nodes. For 60-job problems, while the node reduction factor was 2.2, the speedup factor was nearly 2 due the overhead of the algorithm. TCBB could not be run to completion for some instances of 62-job problems since it took long time to execute.
- The node generated per second by both the versions are comparable For 60-job problems, TCBB generated 9626 nodes per second whereas ETCBB generated 8312 nodes per second. Since ETCBB has a higher overhead due to its node replacement strategy, it generated less number of nodes per second than that by TCBB.
- With a node limit of 2k nodes, ETCBB took 77 seconds to solve 62-job problems generating 552790 nodes. As expected, the node generation per second dropped to 7179 as more nodes were replaced due to lower availability of memory. TCBB took long time to complete even for 50-job problem instances.

Thus ETCBB appears to be a refined version of the TCBB scheme. We now investigate its applicability to two important classes of single machine job sequencing problems. We first present the weighted earliness-tardiness problem.

**Weighted Earliness-Tardiness (WET) problem**

In the weighted earliness-tardiness problem, jobs have due dates. A job completing earlier than the due date incurs an earliness penalty whereas a job completing later incurs a tardiness penalty. The objective is to schedule the jobs in such a way that the weighted sum of earliness and tardiness penalties is minimized. For the class of problem considered in this paper, the earliness and tardiness penalties are measured as the absolute deviation of job completion times around the common due date. The weights, that is, the earliness and tardiness penalty coefficients, depend on whether a job is early or tardy but do not depend on jobs. The problem can be formulated as follows: Given the processing times \( p_j \) of \( N \) jobs, the common due date \( d \), and
the earliness and tardiness penalty coefficients \( \alpha \) and \( \beta \) respectively, the objective is to minimize \( F = \sum (\alpha E_j + \beta T_j) \) where \( E_j = \max(0,d-C_j) \) and \( T_j = \max(0,C_j-d) \) are the earliness and tardiness of job \( J_j \), and \( C_j \) denote the completion time of job \( J_j \) in a schedule.

The value of the common due date plays an important role in the complexity of the WET problem. When the due date is equal to or larger than a critical value \( d^* \), the WET problem becomes unrestricted; otherwise the problem is called restricted. The unrestricted WET problem is polynomially solvable whereas the restricted version is known to be NP-complete (Hall, Kubiak, and Sethi 1991). Other characteristics of the optimal schedule can also be found in (Hall, Kubiak, and Sethi 1991). Dynamic programming (DP) and Depth-first Branch-and-bound have been suggested so far for solving the restricted WET problem. These approaches can hardly be used in practice for solving large problem instances.

Hoogeveen, Oosterhout, and Van De Velde (1994) suggested a Lagrangian-relaxation based heuristic estimate for the problem which is the best known heuristic estimate available in the literature. Interestingly, this admissible heuristic estimate is not consistent. It relaxes a constraint \( W \leq d \) where \( W \) is the sum of the processing times of jobs completed before the due date. The heuristic is calculated as \( \max\lambda L(\lambda) = \max\lambda [\min \{ \Sigma_j (\alpha E_j + \beta T_j + \lambda(W - d)) \}] \) where \( \lambda \geq 0 \). \( L(\lambda) \) can be found by Emon’s matching algorithm (Emonon 1987) with positional weights of early and tardiness of job \( J_j \), and \( C_j \) denote the earliness and tardiness of job \( J_j \), and \( C_j \) denote the completion time of job \( J_j \) in a schedule.

An arc in the search space represents the scheduling of a job and its cost may be taken as the incremental cost of scheduling the job. Therefore, the cost of the two successor arcs below a node would be \( \alpha \) times the earliness of the job if the job is scheduled to complete before the due date, and would be \( \beta \) times the tardiness of the job if the job is scheduled to complete after the due date. In this representation, there can be almost \( d+pN \) nodes at any level. As a result, the total number of nodes in the graph will be \( O(N(d+p_s)) \).

Successors are ordered in ETCBB following a heuristic rule. At a node \( n \), a job would be scheduled before due date or after due date. Let \( n_{\text{left}} \) and \( n_{\text{right}} \) denote the unscheduled interval before and after due date. Heuristic or after due date. Let \( n_{\text{left}} \) and \( n_{\text{right}} \) denote the unscheduled interval before and after due date. Heuristic estimates are computed following the Lagrangian-relaxation based heuristic suggested by Hoogeveen, Oosterhout and Van De Velde described above.

**Experimental results**

In our experiments, we fixed the memory as 16k nodes, and tested ETCBB, DP, DFBB and A* on DEC Alphastation 250/4266 for problems with different due dates and with different ratios of \( \alpha \) and \( \beta \). Due date was fixed at \( d=kd^* \), \( 0 < k < 1 \) to solve restricted instances only. The problem criticality increases as \( d \) approaches \( d^* \), and for a fixed \( k \), it also increases as the ratio of \( \alpha \) and \( \beta \) decreases (Bagchi, Sullivan, and Chang 1987). Table 2 reports our findings for ETCBB and Table 3 for other algorithms. Processing times were generated randomly from a uniform distribution in the range 1 to 99. Results were based on the average of 100 randomly problem

<table>
<thead>
<tr>
<th>Job size</th>
<th>( \alpha : \beta )</th>
<th>Node limit = 16k nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Node</td>
<td>Time (secs)</td>
</tr>
<tr>
<td></td>
<td>Gen</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>5:1</td>
<td>846</td>
</tr>
<tr>
<td></td>
<td>1:5</td>
<td>682</td>
</tr>
<tr>
<td>500</td>
<td>5:1</td>
<td>1570</td>
</tr>
<tr>
<td></td>
<td>1:5</td>
<td>1677</td>
</tr>
<tr>
<td>1000</td>
<td>5:1</td>
<td>2764</td>
</tr>
<tr>
<td></td>
<td>1:5</td>
<td>3301</td>
</tr>
<tr>
<td>1500</td>
<td>5:1</td>
<td>3082</td>
</tr>
<tr>
<td></td>
<td>1:5</td>
<td>4591</td>
</tr>
<tr>
<td>2000</td>
<td>5:1</td>
<td>5540</td>
</tr>
<tr>
<td></td>
<td>1:5</td>
<td>6815</td>
</tr>
</tbody>
</table>

Table 2: ETCBB for the WET problem
instances. Our observations are as follows:

- ETCBB could readily solve 2000-job problems; for the most restricted case with \( d=0.9d^* \) and with \( \alpha = 1.5 \), it took around 23 seconds to solve 2000-job problems.

- DP, A* and DFBB performed poorly in this domain. For a node limit of 16k nodes, DP ran out of memory for 20-job problems. A* with heuristic modification (Mero 1984) could solve upto 100-job problems. Other variants of A* with tie resolution in favour of lower or higher g-valued nodes indicated a similar trend. DFBB took 8.28 seconds to solve 100-job problem instances for the most restricted case generating 294463 nodes (Table 3). We could not run DFBB to completion for 200-job problem instances since it took long time to execute.

- ETCBB generated 8663 nodes for the most restricted problem which is smaller than the 16k node limit. Hence the node replacement strategy has no role to play. The TCBB version would have performed in a similar way. We have not experimented with the TCBB version because we expect it to perform similarly and secondly, our objective is to show the applicability of the scheme to this domain, not to compare the relative performance of the two versions. Since no nodes were replaced, this remarkable performance of the scheme can thus be attributed to dynamic heuristic improvements incorporated in the algorithm.

We now present the SQP problem and show that ETCBB can optimally solve larger instances than GREC in a given available memory.

**Sequence-dependent Quadratic penalty problem (SQP)**

The SQP problem is similar in formulation to the QP problem except that jobs may have setup times which are sequence-dependent. Setup time \( s_{ij} \) for job \( J_j \) when it immediately follows job \( J_i \) is said to be sequence-dependent if the setup time is dependent on the preceding job \( J_i \). The presence of setup times makes the problem extremely difficult to solve and has been shown to be NP-complete in (Rinooy Kan 1976).

Sen and Bagchi (1996) have shown that due to the presence of sequence-dependent setup times, evaluation function becomes non-order-preserving. As a result, no path can be discarded and A*-based graph search can fail to output optimal solutions. Tree search methods tend to be inefficient because many duplicate nodes get generated. They have suggested a graph search algorithm GREC to optimally solve this problem. GREC has been reported to solve 20-job problems using a hash queue of 200k nodes.

We have applied ETCBB to solve this problem optimally. Nodes were represented as pair \((V, J_i)\) where \( V \) is the of jobs scheduled and \( J_i \) is the last job (Sen and Bagchi 1996). Computation of heuristic estimate and successor ordering are followed as in GREC. Since no g-value of nodes are maintained, no paths need to be discarded. When a node \( n=(V, J_i) \) first enters the search graph, the completion time \( t_i \) of node \( n \)’s last job is stored at node \( n \) as a parameter \( T \). At a subsequent instant when \( n \) is reached along a different path with \( t'_i \) being the completion time of job \( J_i \), \( T \) is reset to \( t'_i \). The b-value \( b(n) \) at the node \( n \) depends on the incoming path and thus it depends on \( T \). The value of \( b(n) \) along the path of downward movement can be expressed as:

\[
b(n) = AT + BT + C
\]

where A, B and C are parameters dependent on node \( n \) but independent of \( T \). Let \( Q' \) be the remaining jobs to be processed at \( n \). Then

\[
A = \sum_{j} \left[ \beta_{kj} | J_k \text{ is in } Q' \right]
\]

\[
B = \sum_{j} \left[ \beta_{kj} t'_i | J_k \text{ is in } Q' \right]
\]

and C can be viewed as the b-value at node \( n \) with node \( n \) as the origin (i.e. with \( T = 0 \)). The job completion times \( t'_i \) in B also take node \( n \) as origin. A, B, and C must be stored.
at each node. Table 4 shows our experimental results and compares the performance of ETCBB with GREC. We can observe the following:

- With 16k node limit, while GREC ran out of space for 16-job problems, ETCBB could solve 22-job problems within 100 seconds. However, GREC ran faster than ETCBB. In (Sen and Bagchi 1996), GREC was shown to solve 20-job problems but using hash queue of size around 200k nodes. We

<table>
<thead>
<tr>
<th>Node limit = 200k nodes</th>
<th>G0</th>
<th>G1</th>
<th>Time (secs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Job size</td>
<td>Gen</td>
<td>Time (secs)</td>
<td>Gen</td>
</tr>
<tr>
<td>20</td>
<td>36646</td>
<td>6.13</td>
<td>151741</td>
</tr>
<tr>
<td>22</td>
<td>160041</td>
<td>32.57</td>
<td></td>
</tr>
</tbody>
</table>

Table 5: GREC and ETCBB for the SQP problem with 200k node limit

ran both ETCBB and GREC with 200k node limit on a faster computer SUN Enterprise 5500. The results are given in Table 5. GREC ran faster than ETCBB for 20-job problems but ran out of “overflow” area for 22-job problems.

- As in the QP problem, the node generation per second is higher in GREC than ETCBB due to higher overhead of node replacement strategy in ETCBB. Since both the algorithms cache nodes, the better performance of ETCBB is due to its memory management scheme.

**Conclusion**

This paper shows that the Transpose-and-Cache Branch-and-Bound scheme has a future in solving large instances of problems with natural cut-off bound. Specifically, in some single machine job sequencing problems, the scheme has advanced the existing state of solutions. We hope that this paper would encourage future research in the enhancement and application of the scheme.

Many questions remain. The problems with setup times are difficult and needs special attention. Also, how would the TCBB scheme perform in other domains like project scheduling? Future research should address these issues.

**Acknowledgments.** This research was partly supported by the Indian Institute of Management Calcutta where both the authors were employed. The authors are grateful to Herman Kaindl for his invaluable suggestions. The authors also thank Gerhard Kainz for sending us the TCBB code.

**References**


