ON THE DISTRIBUTION OF PARETO OPTIMAL SOLUTIONS IN ALTERNATIVE SPACE – THE INVESTIGATION OF MULTI OBJECTIVE PERMUTATION FLOWSHOP SCHEDULING PROBLEMS

Martin Josef Geiger

Production and Operations Management Department (510A), University of Hohenheim, Schloß Hohenheim, Osthof-Nord, 70593 Stuttgart, Germany
E-mail: mjgeiger@uni-hohenheim.de

Received 3 October 2005; accepted 4 January 2006

Abstract. The paper presents a study of the search space topology in the context of global optimization under multiple objectives. While in mono criterion problems a single global optimum has to be identified, multi objective problems require the identification of a whole set of equal. Pareto optimal alternatives. It is unclear up to now, however, whether Pareto optimal solutions appear relatively concentrated in search space, or whether their relative positions are rather distant.

This open issue is addressed to the multi objective permutation flow shop scheduling problems. Distance metrics is introduced to assess numerical evaluation of the concentration of Pareto sets. It can be seen that independent from the chosen optimality criteria, Pareto optimal alternatives appear relatively concentrated in alternative space. The result holds for an extensive range of generated problem instances for which the exact global optima are known as well as for benchmark instances taken from literature. The importance of the results can be seen in the context of metaheuristic local search approaches for which meaningful implications derive.

Keywords: Fitness landscapes, multi objective optimization, flow shop scheduling, local search, global optimization, metaheuristic.

1. Introduction

In the past decades, local search approaches have become increasingly popular for the resolution of complex continuous and combinatorial optimization problems [1]. Often classified as so called metaheuristics [2] with the most prominent examples of Simulated Annealing, Tabu Search, and Evolutionary Algorithms, these methods organize modification and improvement steps for alternatives with the ultimate goal of identifying a global optimal solution. The general way of achieving this is by means of a neighborhood, which associates with feasible solution \( x \in X \) of a particular problem a set of neighboring solutions \( nh(x) \).

Taking a minimization problem as an example, alternative \( x' \in nh(x) \) with objective function value of \( g(x') < g(x) \) then replaces \( x \) and the search continues from the improved solution. In the ideal case, global optimum \( x^* \) with \( \nexists x' \in X \mid g(x') < g(x^*) \) is found. It may, however, occur that the search terminates in solution \( \hat{x} \) with \( \nexists x' \in nh(\hat{x}) \mid g(x') < g(\hat{x}) \), so called local optimum which must not necessarily be globally optimal. Metaheuristic approaches accordingly formulate different ways of overcoming local optimality, ranging from memory structures of previously visited alternatives to probabilistic neighborhoods and acceptance criteria of inferior alternatives.

For the application of local search methods it is assumed that small modifications of alternatives lead to small changes in terms of the relevant evaluation function value. Similar to gradient methods in continuous optimization, local search approaches therefore make use of information gathered about the topology in the search space. Moreover, it is possible to interpret the progress of local search methods in analogy to a movement in a fitness landscape [3], a concept proposed in the context of theoretical biology [4]. An ex-
ample of the visualization of such a fitness landscape is given in Fig 1. The landscape as such is here induced by three components, the set of alternatives \( X \), the chosen neighborhood \( nh \), and evaluation function \( g(x) \).

It is known that the structure of the fitness landscape has a considerable influence on the performance of local search algorithms [5]. Ruggedness and the existence of many local optima complicate the search, while smooth landscapes with significant concentration of good alternatives in a subpart of the search space allow comparably easy resolution of the problem.

![Fig 1. Example of a fitness landscape](image)

It has been shown for several single objective optimization problems that they bear a search space structure with fitness distance correlation of local optima towards the global optimum [6–9]. For most multi objective optimization problems, however, corresponding studies have, apart from the first sketch in [10], not been conducted yet. In problems involving multiple objectives, vector \( G(x) = (g_1(x), ..., g_K(x)) \) of \( K \) optimality criteria is introduced to evaluate solution \( x \in X \). As the chosen objective functions are often conflicting, the resolution of these problems can be seen in the identification of a whole set of equally Pareto optimal alternatives, introduced in the definitions 1 and 2.

Definition 1 (Dominance): alternative \( x \) is said to dominate alternative \( x' \), if and only if \( g_i(x) \leq g_i(x') \) \( \forall i = 1, ..., K \) \( \land \exists i | g_i(x) < g_i(x') \). We denote the dominance of alternative \( x \) over \( x' \) with \( x \prec x' \).

Definition 2 (Pareto optimality): alternative \( x \) is said to be Pareto optimal, if and only if \( \neg \exists x' \in X | x' \prec x \). The set of all Pareto optimal alternatives is called Pareto set \( P \).

In contrast to single objective optimization problems, the structure of a whole set of globally optimal solutions is therefore of relevance for the analysis of the fitness landscape and the possible resolution by means of local search methods.

In the current paper we present a novel study of the structure of Pareto sets with application to multi objective flow shop scheduling. In section 2 the problem is introduced and a quantitative model is proposed. The analysis of Pareto sets is conducted in section 3. Metrics measuring the closeness of alternatives for permutation schedules are introduced and the results are presented for a wide range of different problem instances. Conclusions with the implications derived for the proposition of metaheuristics are given in section 4.

2. Problem statement

The permutation flow shop scheduling problem (PFSSP) can be characterized as finding an assignment of a set of jobs \( J = \{J_1, ..., J_n\} \), each of which consists of a set of operations \( J_f = \{O_{j1}, ..., O_{jn}\} \) to a set of machines \( M = \{M_1, ..., M_m\} \). The operations have to be executed subject to precedence constraints \( O_{jk} \succ O_{jk+1} \) \( \forall k = 1, ..., o_f - 1 \), enforcing that operation \( O_{jk+1} \) may only start after completion of operation \( O_{jk} \), and involve non-negative processing time \( p_{jk} \). The machine routing of PFSSP is identical for all jobs, and also the job sequence has to be the same on all machines as the passing of jobs is not permitted. Feasible schedule \( x \in X \) defines starting and completion time \( C_j \) of all operations. Its representation is possible using permutation \( \pi = \{\pi_1, ..., \pi_n\} \) of jobs which may be decoded into an active schedule [11] using the expressions (1), (2), (3) and (4).

\[
C_{\pi_1} = p_{\pi_1}, \quad (1)
\]

\[
C_{\pi_j} = C_{\pi_{j-1}} + p_{\pi_j}, \quad \forall j \in \{2, ..., n\}, \quad (2)
\]

\[
C_{\pi_k} = C_{\pi_{k-1}} + p_{\pi_k}, \quad \forall k \in \{2, ..., m\}, \quad (3)
\]

\[
C_{\pi_k} = \max\{C_{\pi_{j-1}}, C_{\pi_{k-1}}\} + p_{\pi_k}, \quad \forall j \in \{2, ..., n\}, \forall k \in \{2, ..., m\}. \quad (4)
\]

As a result, completion time \( C_j \) of jobs can be computed as \( C_j = C_{\rho_j} \), which are then used to evaluate the quality of the schedule with respect to certain optimality criteria [12]. The most prominent among them is the minimization of maximum completion time.
C_{\text{max}} = \max \{C_1, ..., C_n\}

Other measures are the minimization of the total completion time \(C_{\text{sum}} = \sum_j C_j\) which is equivalent to the minimization of the average completion time \(\bar{C} = \frac{1}{n} \sum_j C_j\) as \(n\) is constant. For problems with given due dates \(d_j\) of jobs, tardiness \(T_j = \max \{C_j - d_j, 0\}\) of jobs often plays a role. The minimization of maximum tardiness \(T_{\text{max}} = \max \{T_j\}\) and total tardiness \(T_{\text{sum}} = \sum_j T_j\) is typically considered here. The number of tardy jobs

\[U = \sum U_j\] with \(U_j = 1: T_j > 0; U_j = 0: T_j = 0\) is of relevance when e.g. compensation payments of a fixed amount of money result from tardy completion of jobs.

A machine oriented evaluation of the schedules is done involving machine idle time \(I_i\) up to the completion of the last operation on the machine and the minimization of the total idle time \(I_{\text{sum}} = \sum_i I_i\) is a possible objective.

3. The analysis of Pareto sets

3.1. Methodology

The closeness of alternatives is analyzed with respect to two distance metrics measuring the distance between two permutations.

First, position based metric \(d_{\text{pos}}(x, x')\) is computed on the basis of the “inverse” permutation \(\sigma\) which expresses for each job \(\pi\) its position in \(\pi\) \([7]\), \(\sigma_x = i\). Fig 2 exemplifies this principle for given permutation.

\[
\begin{align*}
\pi & = 3 1 2 4 8 6 9 10 5 7 \\
\sigma & = 2 3 1 4 9 6 10 5 7 8
\end{align*}
\]

Fig 2. Permutation \(\pi\) and inverse permutation \(\sigma\)

The position based distance between two permutations is then computed as shown in expression (5).

\[
d_{\text{pos}}(x, x') = \sum |\sigma_j - \sigma_j'|
\]

(5)

The second distance metric \(d_{\text{prec}}(x, x')\) used in the analysis is based on the precedence of jobs in both alternatives. In the first step it is determined for each pair of jobs \(J_j, J_k \in J, j \neq k\) whether \(J_k\) is preceded by \(J_j\) in both alternatives and the number of these relations is counted. Then, this number is subtracted from \(\frac{n(n-1)}{2}\).

While the described distance metrics \(d_{\text{pos}}\) and \(d_{\text{prec}}\) measure the distance between two single alternatives, an overall measure for Pareto set has to be derived. The first measure for both metrics is given with \(Q_{\text{prec, aver}}\) and \(Q_{\text{pos, aver}}\) in expression (6) and (7).

\[
Q_{\text{prec, aver}}(P) = \frac{1}{|P|(|P|-1)} \sum_{x \in P, x' \in P} d_{\text{prec}}(x, x'),
\]

(6)

\[
Q_{\text{pos, aver}}(P) = \frac{1}{|P|(|P|-1)} \sum_{x \in P, x' \in P} d_{\text{pos}}(x, x').
\]

(7)

The proposed expressions measure the average distances of elements in \(P\) and give a general idea about the overall concentration of Pareto optimal solutions in alternative space.

A possible drawback of the average distances as described above, however, is that single elements may appear separated from a clustered majority of Pareto optimal alternatives, making measurement for the most distant element therefore necessary. Expressions (8) and (9) formulate such a measure for two distance metrics. Essentially, for each element in \(P\) the closest other element is computed and the maximum of these distances is taken as a measure of the most distant element.

\[
Q_{\text{prec, max}}(P) = \max_{x \in P} \left\{ \min_{x' \in P, x' \neq x} \left\{ d_{\text{prec}}(x, x') \right\} \right\},
\]

(8)

\[
Q_{\text{pos, max}}(P) = \max_{x \in P} \left\{ \min_{x' \in P, x' \neq x} \left\{ d_{\text{pos}}(x, x') \right\} \right\}.
\]

(9)

While \(Q_{\text{prec, aver}}, Q_{\text{pos, aver}}, Q_{\text{prec, max}}, \) and \(Q_{\text{pos, max}}\) measure the concentration of Pareto sets, the comparison with some other value is necessary to derive conclusions about whether the elements are rather close or rather distant. This can be done by comparing the values for Pareto sets with average values derived from sets of the same cardinality containing randomly generated permutations.

3.2. Problem instances

In order to analyze the distribution of Pareto optimal alternatives for multi objective flow shop scheduling problems an extensive number of problem instances with different optimality criteria combinations have been examined. Table 1 gives an overview about the different characteristics of the investigated instances with respect to well known problem classification by [13,14]. For each of the problem types from \(\gamma_1\) to \(\gamma_{12}\), 100 instances have been created following the proposal of [15]. Each of them consists of 10 jobs to be processed on 10 machines which allow the enumeration of all 10! possible schedules within reasonable time to determine true Pareto sets.

While the generated problem instances provide a first solid basis for the analysis of the problem structures, established benchmark instances known from literature are in comparison with mono criterion scheduling problems rather scarce. Basseur et al. propose in [16] a set of multi objective flow shop scheduling problems on the basis of the test
data sets of [17]. Also, in [18], a problem instance with three criteria is published, considering the maximum completion time \( C_{\text{max}} \), the average flow time \( \bar{F} \), which is here equivalent to the average completion time \( \bar{C} \), and the average tardiness \( \bar{T} \) which is equivalent to the total tardiness \( \sum T \).

Table 2 shows the problem instances taken from literature. The data sets by [4] are indicated by “Ta…”, and one by [18] is labeled with “Ba…”.

Due to the size of the search space of these problem instances, the optimal solutions are not known with absolute certainty. Instead of true Pareto sets \( P \), reference sets \( \text{refP} \) containing up to now best known solutions are therefore analyzed. The concentration measures given in the expressions (6), (7), (8) and (9) are analogously applied to the elements in \( \text{refP} \).

### Table 1. Classes of analyzed problem instances

<table>
<thead>
<tr>
<th>Problem instance type</th>
<th>No of objectives</th>
<th>Classification ( F \mid \text{prima}, d_j \mid \gamma = \ldots )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma_1 )</td>
<td>2</td>
<td>( C_{\text{max}}, T_{\text{max}} )</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>2</td>
<td>( C_{\text{max}}, C_{\text{sum}} )</td>
</tr>
<tr>
<td>( \gamma_3 )</td>
<td>2</td>
<td>( C_{\text{max}}, T_{\text{sum}} )</td>
</tr>
<tr>
<td>( \gamma_4 )</td>
<td>2</td>
<td>( T_{\text{max}}, T_{\text{sum}} )</td>
</tr>
<tr>
<td>( \gamma_5 )</td>
<td>2</td>
<td>( C_{\text{sum}}, T_{\text{max}} )</td>
</tr>
<tr>
<td>( \gamma_6 )</td>
<td>2</td>
<td>( C_{\text{sum}}, T_{\text{sum}} )</td>
</tr>
<tr>
<td>( \gamma_7 )</td>
<td>3</td>
<td>( C_{\text{max}}, T_{\text{max}}, T_{\text{sum}} )</td>
</tr>
<tr>
<td>( \gamma_8 )</td>
<td>3</td>
<td>( C_{\text{max}}, C_{\text{sum}}, T_{\text{max}} )</td>
</tr>
<tr>
<td>( \gamma_9 )</td>
<td>3</td>
<td>( C_{\text{max}}, C_{\text{sum}}, T_{\text{sum}} )</td>
</tr>
<tr>
<td>( \gamma_{10} )</td>
<td>3</td>
<td>( C_{\text{sum}}, T_{\text{max}}, T_{\text{sum}} )</td>
</tr>
<tr>
<td>( \gamma_{11} )</td>
<td>4</td>
<td>( C_{\text{max}}, C_{\text{sum}}, T_{\text{max}}, T_{\text{sum}} )</td>
</tr>
<tr>
<td>( \gamma_{12} )</td>
<td>6</td>
<td>( C_{\text{max}}, C_{\text{sum}}, T_{\text{max}}, T_{\text{sum}}, I_{\text{sum}}, U )</td>
</tr>
</tbody>
</table>

### Table 2. Problem instances taken from literature and cardinality of their reference sets

<table>
<thead>
<tr>
<th>Problem instance ( n \times m )</th>
<th>Objective functions</th>
<th>No of operations</th>
<th>No of elements in ( \text{refP} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ta 20x5 #1</td>
<td>( C_{\text{max}}, T_{\text{sum}} )</td>
<td>100</td>
<td>4</td>
</tr>
<tr>
<td>Ta 20x5 #2</td>
<td>( C_{\text{max}}, T_{\text{sum}} )</td>
<td>100</td>
<td>6</td>
</tr>
<tr>
<td>Ta 20x10 #1</td>
<td>( C_{\text{max}}, T_{\text{sum}} )</td>
<td>200</td>
<td>42</td>
</tr>
<tr>
<td>Ta 20x10 #2</td>
<td>( C_{\text{max}}, T_{\text{sum}} )</td>
<td>200</td>
<td>31</td>
</tr>
<tr>
<td>Ta 20x20</td>
<td>( C_{\text{max}}, T_{\text{sum}} )</td>
<td>400</td>
<td>40</td>
</tr>
<tr>
<td>Ta 50x5</td>
<td>( C_{\text{max}}, T_{\text{sum}} )</td>
<td>250</td>
<td>10</td>
</tr>
<tr>
<td>Ta 50x10</td>
<td>( C_{\text{max}}, T_{\text{sum}} )</td>
<td>500</td>
<td>60</td>
</tr>
<tr>
<td>Ta 50x20</td>
<td>( C_{\text{max}}, T_{\text{sum}} )</td>
<td>1000</td>
<td>90</td>
</tr>
<tr>
<td>Ta 100x10</td>
<td>( C_{\text{max}}, T_{\text{sum}} )</td>
<td>1000</td>
<td>112</td>
</tr>
<tr>
<td>Ta 100x20</td>
<td>( C_{\text{max}}, T_{\text{sum}} )</td>
<td>2000</td>
<td>193</td>
</tr>
<tr>
<td>Ba 49x15</td>
<td>( C_{\text{max}}, F, \bar{T} )</td>
<td>735</td>
<td>890</td>
</tr>
</tbody>
</table>

### 3.3. Results

The analysis of the distribution of Pareto optimal alternatives reveals for the studied problem instances significant concentration in alternative space.

Figures 3 and 4 plot the values of Pareto sets for 1200 generated problem instances as described in Table 1 and give the comparison with the values of random permutation sets of the same cardinality. It can be seen that all Pareto sets are for both the precedence based and the position based metric significantly concentrated in alternative space. While with increasing cardinality the average distances quite naturally increase, the values stay well below the benchmark of...
random permutations. In brief, this demonstrates that Pareto optimal alternatives in average share significant similarities in terms of their characteristics in alternative space.

The results of the analysis of the maximum distances are given in Figures 5 and 6. Again, the values of Pareto sets are plotted against random permutation sets of the same cardinality. It can be seen that with increasing cardinality of the sets the maximum distance of the elements decreases. This is to be expected as the chance of having solution $x'$ close to solution $x$ increases with the number of elements.

Only in a very few cases the most distant element in $P$ appears more distant from the other elements in comparison with the random permutations. The vast majority of Pareto sets, however, shows significant concentration also with respect to its most distant element.

![Fig 5. Most separated elements of Pareto sets for precedence based metric](image)

![Fig 6. Most separated elements of Pareto sets for position base metric](image)

While Figures 3, 4, 5 and 6 plot the values of all Pareto sets, Table 3 gives the average values of $Q_{\text{prec,aver}}$, $Q_{\text{pos,aver}}$, $Q_{\text{prec,max}}$ and $Q_{\text{pos,max}}$ separated with respect to the number of considered objectives. It can be seen that the average distances increase with increasing number of objectives while the maximum distances tend to decrease. This result is in conjunction with the results above as the number of elements in $P$ increases with the number of objectives.

The analysis of the multi objective flow shop scheduling problems taken from literature confirms the achieved results for the generated instances. Table 4 gives the values obtained for the reference sets containing the best alternatives known so far. The values of the random permutation sets of the same cardinality are given for comparison reasons in brackets.

![Table 4. Concentration of reference sets for the benchmark problem instances from literature](table)

Also for the investigated benchmark instances, the reference sets of potential Pareto optimal alternatives show high concentration in alternative space. This can be seen for the average distances as well as for the most separated element.

Moreover, the concentration increases with the increase of problem size. This effect is especially strong for “Ta...” instances with two objectives. While the overall result also holds for “Ba 49×15” instance of [18] with three objectives, the reference set consists in this case, however, of comparably more diverse alternatives.
4. Conclusions

The investigation of the distribution of Pareto optimal alternatives in alternative space has been presented. The analysis has been carried out using the example of the multi objective flow shop scheduling problem. After an introduction in section 1, this problem has been explained in section 2 and the most important optimality criteria have been presented.

The analysis of the concentration of optimal alternatives has been conducted involving four concentration measures on the basis of two distance metrics for permutation schedules. It was possible to observe that for an extensive set of problem instances involving twelve different optimality criteria definitions, as well as for a set of problem instances taken from literature, Pareto optimal solutions appear relatively concentrated in alternative space. In terms of the characteristics of their permutations, they share significant similarities which tend to grow with the increase of problem size.

The implications of this study are twofold.

First, the results help to explain the underlying principles of the successful application of local search approaches to the considered problem class of multi objective flow shop scheduling [19, 20]. Obviously, these problems bear a structure that may be exploited during the resolution using metaheuristics as global optima are concentrated in alternative space.

Second, an implication of the outcome of the study lies in its potential to help defining more effective resolution approaches for the studied problem class. Given concentration of optimal solutions as observed in the analysis, a possible focus of local search should lie on the intensification of search rather than on diversification.

5. Acknowledgements

The author would like to thank Matthieu Basseur for providing multi objective flow shop scheduling benchmark instances on the basis of [19] and their corresponding best known alternatives on http://www.lifl.fr/~basseur/benchs_matth.html, and an anonymous referee for providing helpful suggestions.

References


Pagrindiniai žodžiai: landšafto profilis, daugiatikslei optimizacija, planavimas, vietinis tyrimas, globalinė optimizacija, metaeuristika.