Improved Evolutionary Algorithm Design for the Project Scheduling Problem Based on Runtime Analysis

Leandro L. Minku, Member, IEEE, Dirk Sudholt, Member, IEEE, and Xin Yao, Fellow, IEEE

Abstract—Several variants of evolutionary algorithms (EAs) have been applied to solve the project scheduling problem (PSP), yet their performance highly depends on design choices for the EA. It is still unclear how and why different EAs perform differently. We present the first runtime analysis for the PSP, gaining insights into the performance of EAs on the PSP in general, and on specific instance classes that are easy or hard. Our theoretical analysis has practical implications—based on it, we derive an improved EA design. This includes normalising employees’ dedication for different tasks to ensure they are not working overtime; a fitness function that requires fewer pre-defined parameters and provides a clear gradient towards feasible solutions; and an improved representation and mutation operator. Both our theoretical and empirical results show that our design is very effective. Combining the use of normalisation to a population gave the best results in our experiments, and normalisation was a key component for the practical effectiveness of the new design. Not only does our paper offer a new and effective algorithm for the PSP, it also provides a rigorous theoretical analysis to explain the efficiency of the algorithm, especially for increasingly large projects.

Index Terms—Schedule and organizational issues, evolutionary algorithms, software project scheduling, software project management, search-based software engineering, runtime analysis.

I. INTRODUCTION

SOFTWARE project scheduling is traditionally one of the major problems faced by software project managers [2]. It is a particularly demanding task [3] and, being a project planning task, is vital to many software engineering activities [4]. Its process involves several duties [3]: identify project activities; identify activity dependencies; estimate resources for activities; allocate people to activities; and create project charts. As explained by Sommerville [3], project scheduling “involves separating the total work involved in a project into separate activities and judging the time required to complete these activities. Usually, some of these activities are carried out in parallel. Project schedulers must coordinate these parallel activities and organise the work so that the workforce is used optimally.”

In this work, we concentrate on the problem of optimally allocating people (employees) to activities (tasks) using automated approaches. Similar to previous work [5], [6], [7], we will refer to this problem as the Project Scheduling Problem (PSP). Typical objectives to be optimised when making allocations are to minimise the cost and completion time of the software project. When allocating employees to tasks, employees can divide their attention among several tasks at the same time in the PSP. This is often modelled by dedication values [6] (Chang et al. [5] call it commitment quotas) representing the percentage of time an employee spends on a certain task. Employees have a maximum dedication, which must be respected to avoid overwork. The PSP also has other constraints. For instance, the team assigned to a certain task must have all the skills required for this task, and task precedence constraints among tasks must be respected. All these factors contribute to the difficulty of the PSP, making it more complex than classical scheduling problems.

The PSP is particularly challenging when the project is large. The space of possible allocations of employees to tasks is enormous, and providing an optimal allocation of employees to tasks becomes a very difficult task [4]. So, it is desirable to have automated methods to provide near optimal schedules that meet all constraints and can aid a software project manager in his/her decision of what schedule to adopt in order to minimise cost and completion time of the project. In addition, budget and schedule compression has become the norm in today’s software industry [8], [9], as a result of its intensified competitiveness due to movements such as globalization, outsourcing and open-source software [10], [11], [12], [13]. As software companies have to strive to get their jobs done faster for less money in order to succeed [8], automated scheduling tools that can provide insight into how to reduce cost and completion time become increasingly important [14]. The advantages of automated methods are:

- they help the software manager in producing schedules that meet all constraints;
- they provide a very useful insight into how to optimise objectives such as cost and completion time, as they may find solutions that no human has thought of; and
- they speed up the task of allocating employees to tasks, as they can generate schedules taking far less time than a human would if he/she had to plan the whole schedule from the beginning.
With Search-Based Software Engineering as a growing field in Software Engineering [15], evolutionary algorithms (EAs) [16] have been used to solve the PSP. Evolutionary algorithms mimic the natural evolution of species, “evolving” a population of candidate solutions, i.e. project schedules. Different EAs were adopted for different variants of the PSP, e.g. [17], [5], [6], [18], [19]. This also includes multiobjective variants of this problem [20], [7], [4], [21] as well as approaches using cooperative coevolution [22]. A comprehensive survey of related work was presented in the work by Di Penta, Harman, and Antoniol [4].

In order for search-based software engineering to be applied in the real-world successfully, we need to understand fundamentally what search algorithms can and cannot do. However, despite the existing research, it is still not well understood how design choices in EAs affect performance for the PSP, including what problem features make the PSP a challenging problem for EAs. This lack of understanding makes it hard to make appropriate design choices for EAs to solve the PSP, and it impedes the design of better EAs for this problem. Fundamental understanding of algorithms used in software engineering is essential to advance the science. Some progress was made in the well-known work by Alba and Chicano [6]. They presented an effective EA and a systematic empirical performance analysis of the impact of problem features on the performance of their EA.

Our work presents a novel perspective for designing EAs for the PSP, based on rigorous mathematical arguments. We use runtime analysis to gain insight into how EAs perform on illustrative instance classes. This sheds light on what instances are easy, and which ones are hard to solve for EAs. It also helps to make more informed design choices for the PSP. Based on that, we propose a new design that is shown to be very effective both theoretically and empirically. Our design is thoroughly analysed and shown to outperform others considering several criteria that may influence a software manager’s choice of PSP algorithm. Such an analysis is important to software engineering in practice because applying an algorithm/tool without understanding its benefits/limitations can lead to wrong conclusions.

This paper is organised as follows. Section II briefly explains the full process of software project scheduling, and how the duty of allocating employees to tasks (PSP) is integrated into it. Section III presents a formal definition of the PSP investigated in this paper. Section IV explains previous works on using search-based techniques for the PSP. Section V presents our search-based approach for solving the PSP. Section VI presents a runtime analysis to gain insights into the performance of EAs on the PSP in general, and on specific problem instances that are easy or hard. This rigorous theoretical analysis explains the efficiency of our approach. Section VII provides an experimental analysis to provide the software project manager with insight into what algorithm to choose for the PSP. It gives evidence demonstrating what algorithms are likely to behave better according to different evaluating criteria that may affect the project manager’s decision. Section VIII provides an additional discussion on the comparison of different approaches. Section IX presents the conclusions and future work.

II. THE FULL SOFTWARE PROJECT SCHEDULING PROCESS

As explained in section I, the full software project scheduling process involves several duties [3]: identify project tasks; identify task dependencies; estimate resources for tasks; allocate employees to tasks; and create project charts. This paper investigates the duty of allocating employees to tasks, which is called PSP for briefness. This section explains how the PSP is integrated into the full process of software project scheduling.

The PSP uses several pieces of information, which are determined in the prior duties of identifying tasks, identifying task dependencies, and estimating resources for tasks (estimating the effort required for the task, e.g., in person-months). The identification of tasks and their dependencies is performed by the software manager, possibly with the help of the software architect. After that, the software manager provides an estimate of the effort required to complete each of these tasks. Models such as COCOMO 81 [23] can be used as support tools in predicting the required effort based on features such as the estimated software size, development type and several cost drivers, such as required reliability, size of database, etc. Several other methods to estimate software effort exist (e.g., [24], [25], [26], [27], [28]). Additionally, the PSP also uses information about the employees, e.g., which employees are available for the project, and what their salaries and skills are.

After these prior duties are completed, one can proceed to the duty of allocating employees to tasks (PSP). This paper is concerned with the problem of allocating employees to tasks in such a way to minimise the cost and completion time of the software project. In order to use an algorithm to search for such good allocations, a method to estimate the cost and completion time of the project is necessary. Several cost and completion time estimation methods exist [29], [3], [23], [26]. A feature that is common to most of them is the use of the estimated required effort as one of the main ingredients to estimate cost and completion time. In order to guide the search towards good allocations, the method used by the PSP must consider not only the required efforts, but also the allocations of employees to tasks. Considering the allocations is very important for software projects, because its resources are mainly human resources [30]. In this paper, schedule-driven estimation [29] is used for that purpose.

After the allocation of employees to tasks is decided, project charts can be produced. For instance, Gantt charts can be created using information about the tasks and their dependencies, required efforts and allocations.

III. PROBLEM FORMULATION

This section provides a formal definition of the PSP as an optimisation problem, based on existing work [6]. It consists of the following inputs:

- a set of employees \( e_1, \ldots, e_n \) with salaries \( s_1, \ldots, s_n \), and sets of skills \( \text{skill}_{1}, \ldots, \text{skill}_{m} \) respectively,
- a set of tasks \( t_1, \ldots, t_m \) with required efforts \( \text{eff}_1, \ldots, \text{eff}_m \), and sets of required skills \( \text{req}_1, \ldots, \text{req}_m \) respectively, and
• a task precedence graph (TPG) – a directed graph with tasks as nodes and task precedence as edges.

The set of tasks and TPG can be referred to as the project to be scheduled. The goal is to assign employees to tasks in order to minimise the completion time as well as the total cost for the project (i.e. salaries paid). Employees can work on several tasks simultaneously, as indicated by their dedication to certain tasks. The dedication is the fraction of their time devoted to a particular task. The completion time is the time the project is completed. It is computed with a simple iterative algorithm (described in detail as Algorithm 1 in Section V-B) that takes into account the task precedence graph, the effort for each task, and the dedication for all employees.

The amount of dedication of an employee $e_i$ to a task $t_j$ is determined by a value $x_{i,j} \in \{0/k,1/k, \ldots, k/k\}$, where $k \in \mathbb{N}$ reflects the granularity of the solution. For $k = 1$ we only have dedications 0 and 1. For, say, $k = 10$ we have $k + 1 = 11$ dedications of $0\%, 10\%, \ldots, 100\%$. Employees can only work on a task $t_j$ if all employees working together have all the skills to do the task. More formally, we require

$$\text{req}_j \subseteq \bigcup_{i=1}^{n} \{\text{skill}_i \mid x_{i,j} > 0\}. \quad (1)$$

A maximum dedication for each employee was considered previously [6]. It reflects how much of a full-time job an employee is able to work. This can reflect both part-time jobs as well as the willingness to work overtime. Alternatively, one could introduce the productivity of an employee as a general indicator of her/his performance. The productivity of a part-time employee could then be decreased accordingly. It would also be feasible to associate different productivities with each skill so as to reflect how good an employee is at using a particular skill. This was done, for instance, in [18]. However, for the sake of simplicity we leave these aspects for future work and consider that all employees have the same maximum dedication of 1.

IV. PREVIOUS WORK ON SEARCH-BASED TECHNIQUES TO THE PSP

A. EAs for the PSP

Search-based techniques such as EAs [16] have been showing success in solving resource-constrained scheduling problems [31]. However, such resource-constrained scheduling problems do not present some features that are important in software projects, e.g., the fact that employees can divide their attention among different tasks at the same time.

The software engineering literature on PSP is more recent than the literature on resource-constraint scheduling problems. For instance, Chang et al. [5] proposed an EA for a problem formulation similar to that in Section III. One of the main differences is that they considered overwork as an extra objective to be minimised.

Alba and Chicano [6] used the problem formulation explained in section 2 and proposed an EA which will be referred to as the Genetic Algorithm (GA) in this paper. Under this formulation, employees cannot exceed their maximum total dedication to tasks given by the company’s policy, but any amount of overwork within the company’s rules is allowed. Later on, Luna et al. [7] investigated a similar problem formulation in a multi-objective scenario where cost and completion time were not combined into a single fitness function. Their work also introduced a repair operator to improve finding feasible solutions. These algorithms are explained in more detail in sections IV-B and IV-C.

Di Penta et al. [4] formulated the PSP as the problem of assigning employees to teams and deciding on the order that work packages are selected to be processed in a queue system. The queue system is then used to assign work packages to teams. Different search-based techniques such as single and multi-objective EAs were investigated. The objectives are to minimise completion time and/or fragmentation, rather than completion time and cost. Fragmentation is the total amount of idle person-months in the schedule, due to the precedence constraints of the work packages to be processed.

Chang et al. [19] introduced a time-line to break down a task into smaller components of time-sliced activities. In this way, tasks can be interrupted after they start, and employees do not necessarily need to work on a certain task from its beginning until its end. Several other features were also introduced, such as distinction between an employee and a contractor, different rates of payment depending on whether there is overwork, different proficiencies of the skills, possibility of training during the project, etc. The inclusion of all these features creates a problem formulation closer to reality, but it also introduces a large number of subjective input values required by the algorithm. It is unknown to what extent the solutions provided by the EA employed in their study are sensitive to the uncertainty of these inputs.

Other examples of EAs for the PSP include the work of Kapur et al. [18], multi-objective EAs [20], [21] and cooperative coevolution algorithms [22]. A comprehensive survey of related work was presented in the work by Di Penta, Harman, and Antoniol [4].

Despite all the existing research, it is still not well understood how design choices in EAs affect performance, including what problem features make the PSP hard for EAs. This lack of understanding impedes the design of better EAs. In our work, we first tackle this issue theoretically considering the problem formulation proposed by Alba and Chicano [6] due to its popularity. A new design is then proposed to overcome problems specific to the existing algorithms designed for this problem formulation. Other problem formulations can be considered as future work.

B. GA for the PSP

Alba and Chicano [6] proposed a GA where candidate solutions for the PSP are represented as a matrix of binary values which encode the dedications $x_{i,j}$ for each employee $e_i$ and task $t_j$. The recombination operator is a 2-D single point crossover, which randomly selects a row and a column and then swaps the elements in the upper left and in the lower right quadrant of the two parents. They used bit-flip mutation, binary tournament selection of two parents for recombination, and survival selection based on elitist replacement of the worst individual.
A candidate solution is considered infeasible if there is a task with no employee associated, or the skills constraint (eq. 1) is not satisfied, or there are employees working overtime. Overtime can happen when tasks are executed in parallel, as the total dedication of employees in the candidate solution can exceed 1. Based on that, the fitness function is defined as follows:

\[ f(x) = \begin{cases} 
1/q & \text{if the solution is feasible} \\
1/(q + p) & \text{otherwise}, 
\end{cases} \]

where \( q = w_{\text{cost}} \cdot \text{cost} + w_{\text{time}} \cdot \text{time}, \)
\( p = w_{\text{penal}} + w_{\text{undt}} \cdot \text{undt} + w_{\text{reqsk}} \cdot \text{reqsk} + w_{\text{over}} \cdot \text{over}, \)
and \( w_{\text{cost}}, w_{\text{time}}, w_{\text{penal}}, w_{\text{undt}}, w_{\text{reqsk}} \) and \( w_{\text{over}} \) are pre-defined parameters, \( \text{cost} \) and \( \text{time} \) are the cost and completion time of the solution, \( \text{undt} \) is the number of tasks with no employee associated, \( \text{reqsk} \) is the number of skills still required to perform all project tasks, and \( \text{over} \) is the total amount of overwork time spent by all employees during the project.

Their fitness function penalises infeasible solutions, but whether or not it gives hints as to how to reach feasible solutions depends on the chosen values for several pre-defined parameters \( (w_{\text{penal}}, w_{\text{undt}}, w_{\text{reqsk}} \) and \( w_{\text{over}}) \). So, a wrong parameter choice could make the algorithm unable to guide the search towards feasibility. Moreover, their experimental analysis reveals that the overwork constraint can cause their GA to have very low hit rate in finding feasible solutions, i.e., the algorithm is unable to find feasible solutions in several occasions. This holds even for very simple instances where the skill constraints have been removed (every employee has all skills) and all employees have the same salary.

C. A Repair Mechanism

After the preliminary version of this paper [1] was prepared, we learned that Luna et al. [7] independently proposed a repair mechanism that facilitates the search for feasible schedules without overwork. Their repair mechanism considers the maximum total dedication of any employee at any point of time during the generated schedule:

\[ M := \max_{i,\tau} \{ e_i^{\text{work}}(\tau) \} + \varepsilon \]

where \( e_i^{\text{work}}(\tau) \) is the total dedication of employee \( i \) at time \( \tau \), and \( \varepsilon = 0.00001 \) is a small constant used to prevent floating point inaccuracies. If \( M > 1 + \varepsilon \), i.e., if there is overwork, dedications of all employees on all tasks are scaled down by dividing them by \( M \). This implies that overwork is eliminated, but this is done at the expense of increasing the execution time of all tasks by a factor of \( M \).

In many settings this approach might be suboptimal as dedications are unnecessarily reduced during the periods of the schedule where there is no overwork, as long as there is overwork in at least some period of the project.

V. OUR APPROACH

In order to overcome problems related to the hit rate and to improve solution quality, we propose an improved design, which consists of two main points. The first one is to normalise employee’s dedications (section V-A) in order to address the problem of overwork. It provides an alternative solution to the repair mechanism from [7], trying to reduce dedication values only where necessary, instead of reducing all dedication values across the board. The second one is to give a clear gradient for searching towards feasibility by introducing a new type of penalty in the evaluation of cost and completion time (section V-B). This improved design can be used with different algorithms. Our experimental analysis (section VII) uses it with the \((1+1) \) EA, RLS and Pop-EA as explained in section V-C.

A. Normalising Dedications

We alleviate the problem of overwork and hence remove a crucial obstacle in the search process of EAs by using the following approach (normalisation): if at some point of time the total dedication of an employee \( e_i \) across all active tasks is \( d_i > 1 \) then her/his dedication for all tasks at this point of time is divided by \( d_i \). This reflects a very natural way of an employee dividing her/his attention to several tasks. Note that we do not normalise “underwork”, i.e., total dedications less than 1. This would otherwise remove the possibility of balancing cost vs. completion time.

Normalisation allows for much more fine-grained schedules as employees can automatically re-scale their dedications as soon as tasks are finished or new tasks are started. For instance, assume there are two tasks \( t_1, t_2 \) suitable for employee \( e_1 \) under the problem formulation explained in section III. Assume also that the employee works on both tasks at overlapping time intervals and neither normalisation nor repair are used. If \( x_{1,1} + x_{1,2} > 1 \), the employee works overtime whenever she/he works on both \( t_1 \) and \( t_2 \) at the same time. If \( x_{1,1} + x_{1,2} \leq 1 \), on the other hand, resources are wasted when the employee works on a single task. So, regardless of the values for \( x_{1,1} \) and \( x_{1,2} \), there will always be overwork or resources wasted—unless both tasks start and finish at exactly the same time. Note that, depending on \( x_{1,1} \) and \( x_{1,2} \), there could be even both resources wasted when the employee is working on a single task and overwork when working on both tasks at the same time (whenever \( x_{1,1} < 1, x_{1,2} < 1, \) and \( x_{1,1} + x_{1,2} > 1 \)).

With normalisation the semantic of dedication values is changed slightly. For any employee working on two or more tasks, equal dedication values mean that she/he should divide her/his attention equally among all tasks. Dedications 0.5 and 1 mean that she/he should devote twice as much time for the second task as for the first task. A smallest possible example is given in Figure 1. In this example, the schedule with neither normalisation nor repair is infeasible. In the repaired schedule, the employee uses only 50% of her/his time to complete the first task. Normalisation, on the other hand, would decrease deductions of 100% towards 50% as long as two tasks are performed, and then it would automatically remove this scaling factor, so that the dedication of the first task goes back up to 100% after the other task is completed. We will come back to the comparison of normalisation and repair in section VIII.
Fig. 1. Three Gantt diagrams, highlighting the dedication of employee $e_1$ working on two tasks ($t_1$ and $t_2$) simultaneously with $x_{1,1} = 1$ and $x_{1,2} = 1$. Left: an infeasible schedule with overwork. Middle: the same schedule after repair, without normalisation. Right: the same schedule with normalisation.

B. Evaluating Costs and Completion Time

Algorithm 1 makes a schedule-driven estimation of cost and completion time by implicitly constructing a Gantt diagram as explained in the following. If the dedication matrix is the genotype during optimisation, the final schedule can be called phenotype. The phenotype for any feasible genotype is constructed iteratively (lines 8–28), and is assessed by storing the $d_{i,j}$ values from each iteration, along with the corresponding time stamps. The algorithm checks which tasks can be active at the current point of time (line 9). The normalised dedication for all suitable employees is computed for all these tasks (line 15). Then we determine the earliest point of time $t$ at which a task is finished (line 19). All finished tasks are being marked as finished (line 25), so that the next iteration can include new tasks, according to the task precedence graph. If there are tasks that have been started, but are not finished yet, their effort is updated to the remaining effort needed for completing them (line 23). This accounts for potentially piecewise evaluations of particular tasks. Along the way, all completion times and costs in all iterations are added up (lines 20–21). The total completion time and cost are the output when all tasks have been processed (line 29).

Before constructing the schedule, Algorithm 1 tests whether the genotype is infeasible in that not every task can be completed in finite time (lines 1–4). That is the case when the skills constraint (eq. 1) is not satisfied. Note that this includes the case in which there is a task with no employee designated, as tasks with no employee associated can be seen as tasks lacking all the required skills. If the genotype is infeasible, the algorithm returns very high costs and completion times (line 6): $\text{cost} = \sum_{i=1}^{n} \sum_{j=1}^{m} s_i \cdot \text{eff}_j + \text{reqsk} \cdot 2 \sum_{j=1}^{m} \text{eff}_j$, where $\text{reqsk}$ is the number of required skills missing. These values are chosen in such a way that any improvement in decreasing the number of missing skills results in a solution that has both lower costs and lower completion times:

Lemma 1: The output of Algorithm 1 is such that for every two solutions $x, y$, if $y$ has fewer missing skills than $x$ (including no missing skills) then $\text{cost}(y) < \text{cost}(x)$ and $\text{time}(y) < \text{time}(x)$.

A proof of this lemma is given in the appendix. This implies that any decrease in the number of required skills missing strictly decreases both components through $\text{reqsk}$. When cost and completion times are used in any reasonable single- or multi-objective fitness function, this gives a clear gradient for search algorithms towards feasibility, without the need for pre-defined parameters to guide the search towards feasibility as done in previous work [6]. It is also worth noting that this strategy deals both with solutions that are infeasible due to the skills constraint and due to tasks with no employee associated. An EA using it can have a very simple fitness function which does not need to treat these two types of infeasible solution separately as done in previous work [6].

C. New EA Designs

When designing an EA, one needs to choose the encoding (internal representation of candidate solutions), mutation (operator to create a new candidate solution by modifying an existing one), crossover (optional operator to create new candidate solutions by combining existing ones), and fitness function (procedure to evaluate the quality of a candidate solution). This section presents our design, as well as the algorithms used in our study. For general knowledge on EAs, we refer the reader to [16].

1) Encoding and Mutation: A binary encoding was used previously [6] to represent $x_{i,j}$. This however means that the
probability of turning one numerical value into another by mutation depends on the binary encoding of these two values. As a result, some mutations (e.g., mutating $7/k \equiv 01112$ towards $8/k \equiv 10002$ flipping 4 bits) are less likely than others (e.g., mutating $8/k \equiv 10002$ towards $9/k \equiv 10012$ flipping 1 bit).

We follow a different approach and work with a direct encoding of dedication values: the search space is $\{0/k, \ldots, k/k\}^{nm}$. Mutation is performed by changing each entry $x_{i,j}$ of the dedication matrix with probability $P_m = 1/(nm)$, independently from other entries. Changing an entry $x_{i,j}$ means replacing it by a value chosen uniformly at random from $\{0/k, 1/k, \ldots, k/k\} \setminus \{x_{i,j}\}$. This implies that each new value has the same chance of being created by mutation. Note that mutation has a chance of changing several dedication values at the same time, changing only a single entry, or not changing any entries at all.

2) Crossover: In this paper, the crossover operator uses two parent candidate solutions to create two offspring candidate solutions by selecting one of the following two strategies with an equal probability:

- For each employee, select its corresponding dedications to tasks from one randomly chosen parent to compose the first child, and from the other parent to compose the second child. This can be seen as exchanging “rows” of employees’ dedications if we visualise the values $x_{i,j}$ as a matrix of employees by tasks.
- For each task, select its corresponding employees’ dedications from one randomly chosen parent to compose the first child, and from the other parent to compose the second child. This can be seen as exchanging “columns” of dedications to tasks.

The above is different from the crossover operator used before [6]. The reason is that their operator, which is a 2-D single point crossover, allows part of the dedications of an employee to tasks to be inherited by one child and the remaining part by the other child. So, if a solution is good because the dedications of a certain employee are well balanced for the problem instance, crossover can break this balance by mixing these dedications with dedications of another employee. An illustrative example of that in their algorithm is the fact that, even if the parents do not contain overwork, crossover could easily generate children with overwork. A similar problem could happen to the dedications of all employees to a certain task. Our crossover operator either exchanges full rows of employees’ dedications or full columns of dedications to tasks, in an attempt to avoid breaking balances that may be important for the solutions. As it is not known a priori whether the balance in terms of rows or columns is more important, our crossover gives 50 percent of chance for exchanging rows and 50 percent for columns.

3) Fitness Function: A very simple fitness function can be used based on Algorithm 1. The fitness function (to be minimised) used in our experimental analysis is $f(x) = w_{\text{cost}} \cdot \text{cost} + w_{\text{time}} \cdot \text{time}$, where cost and time are obtained from Algorithm 1. The only pre-defined parameters in this fitness function are the weights $w_{\text{cost}}$ and $w_{\text{time}}$, which are used to adjust the desired relative importance of cost and completion time. So, it is easier to configure than other fitness functions [6]. The runtime analysis is not restricted to this fitness function.

4) Algorithms: As our main goal is to introduce runtime analysis for the PSP, we follow many previous examples in this area [32], [33], [34], [35] and use one of the simplest EAs in our theoretical studies: the (1+1) EA. It is simple because it uses a “population” of just one individual and it does not use crossover. One generation consists of mutation, and then selection checks whether the mutation has found an improvement. Algorithm 2 describes the (1+1) EA for our search space, for minimising some fitness function $f$. The encoding, mutation and fitness function used in this paper are described in sections V-C1 to V-C3.

In practice, like all EAs, it will be stopped either after a certain amount of time has elapsed, or when a satisfactory solution has been found. For a theoretical analysis it is a common practice to consider an EA as an infinite process and to measure the expected time needed to find a global optimum. Note that the expected number of components changed in one iteration equals 1. The (1+1) EA is a bare-bone version of EAs and captures their core working principles in a very clear fashion. It can also be a very effective EA, when considering the number of fitness evaluations as the performance measure. For well-known functions like ONE MAX and LEADING ONES the (1+1) EA is the best EA that only uses mutation for variation [36].

We will also investigate randomised local search (RLS). It differs from the (1+1) EA in that during mutation exactly one dedication value is changed. The entry is chosen uniformly at random. So, RLS can only make local steps, i.e., it can easily get stuck in a local optimum. The (1+1) EA has a global search operator, which allows larger jumps to be made, so that the (1+1) EA can, in principle, escape from any local optimum. However, large jumps are only made with a small probability.

From a theoretical perspective the (1+1) EA is harder to analyse than RLS. The reason is that a mutation changing multiple dedication values simultaneously may have both beneficial and detrimental effects. The mutant will be accepted in the light of detrimental mutations if the combined effect of all changes on fitness is non-negative. There are situations where a mutation is accepted, even though the mutant is “further away” from the optimum, in a sense that less dedication values coincide with the respective values in the global optimum.

In order to check whether improvements in solution quality can be obtained by using a population, our experimental analysis also considers the $(\mu + \lambda)$ population-based EA (Pop-EA) presented in Algorithm 3. It differs from the (1+1) EA in that it maintains a population of $\mu$ candidate solutions, rather than

---

**Algorithm 2** (1+1) EA for project scheduling

1: Initialise $x$.
2: repeat
3: Create $x'$ by copying $x$.
4: Apply mutation to $x'$ using probability $P_m$.
5: if $f(x') \leq f(x)$ then $x := x'$.
6: until happy
Algorithm 3 Pop-EA for project scheduling

1: Initialise population $P$ with $\mu$ candidate solutions.
2: repeat
3: Select $\lambda$ parents from $P$ using binary tournament selection.
4: for each pair of parents $x^{(1)}$ and $x^{(2)}$ do
5: With probability $P_c$, apply crossover between $x^{(1)}$ and $x^{(2)}$ to generate $x'^{(1)}$ and $x'^{(2)}$.
6: Otherwise, $x'^{(1)} \leftarrow x^{(1)}$ and $x'^{(2)} \leftarrow x^{(2)}$.
7: Apply mutation to $x'^{(1)}$ and $x'^{(2)}$ using probability $P_m$.
8: $P \leftarrow P \cup \{x'^{(1)}, x'^{(2)}\}$.
9: end for
10: Select the $\mu$ best candidate solutions from $P$ to survive for the next generation, based on the fitness function $f$.
11: until happy

A single solution. At each generation, $\lambda$ parents are selected based on binary tournaments. Crossover is applied to each pair of parents with a pre-defined probability $P_c$. Mutation is then applied to each child as in the (1+1) EA. The mutation and crossover operators used in this paper are explained in sections V-C1 and V-C2, respectively. An elitist approach that retains the best $\mu$ candidate solutions among the parents and children is used to select candidate solutions that survive for the next generation.

A summary of these algorithms is presented in Table I. This table also shows two other algorithms that will be used in section VII.

VI. Runtime Analysis

In the following, we estimate theoretically the optimisation time of the (1+1) EA and RLS, defined as the first generation in which a global optimum is found.

The only assumption we make about the fitness function $f$ is that it is Pareto-compliant in a strict sense: if $x'$ Pareto-dominates $x$ (i.e., $c(x') \leq c(x) \land \text{time}(x') \leq \text{time}(x)$) and $x$ does not Pareto-dominate $x'$ then $f(x') < f(x)$. That is, any improvement in one or both objectives also improves the fitness $f$. This is the case, for instance, when the fitness is chosen as any weighted combination of cost and completion time, and our results are independent of the choice of these weights. In the special case where all employees have the same salary, the costs are always the same. Then $f$ boils down to minimising the completion time; and as (1+1) EA and RLS only depend on the order of search points, and not on their absolute fitness values, we can then, without loss of generality, take $f(x) = \text{time}(x)$.

For details about asymptotic notation ($O$, $\Omega$, $\Theta$) used in the following, we refer the reader to [37].

A. Optimal Completion Times

The two goals of minimising costs and completion time are often conflicting. We first look at the extreme case of minimising the completion time only. In every feasible solution for each task the team has the required skills for the task. This also holds if more employees join in working on a task, regardless of their skills. The following theorem describes the optimal solutions to this special case.

Theorem 2: For every solvable PSP instance, the completion time is minimal if in the schedule all employees always work full time. Then the completion time is $1/n \sum_{j=1}^{nm} \text{eff}_j$.

If normalisation is used, a sufficient condition for minimality is that all dedication values are 1.

The proof of the first statement is straightforward as all employees will work full time until the project is completed. It is obvious that this minimises the completion time.

The second statement is not true without normalisation, as the all-1s matrix will lead to massive overwork on almost all instances. Without normalisation, the difficulty for optimisation is to find the ideal balance between different tasks, while avoiding overwork. When normalisation is used, this difficulty disappears.

B. General Time Bounds

This section presents general time bounds for arbitrary instances of the PSP and almost arbitrary EAs. The considerations herein are based on the progress EAs, and, specifically, local and global mutation operators, can make at the genotype level.

We first present a lower bound on the expected running time, in case the problem only allows for a single optimal solution. This lower bound indicates the least time we should allow for running an EA on the problem. In a more general sense, the analysis applies to the time for finding any fixed target point.

Theorem 3: Consider a PSP instance with $n$ employees and $m$ tasks, $nm > 1$, with a single global optimum (a fixed target) in the fitness function used. Consider an EA starting with a population of random solutions (of arbitrary size) and afterward creating new solutions from previously created ones by local or global mutations. The expected optimisation time (time to hit the target) of the EA, with or without normalisation, is at least $\Omega(knm \log(nm))$.

A formal proof is given in the appendix. In simple terms, the analysis shows that initially many dedication entries will differ from their optimal values. In order to find the global optimum (the target point), the EA needs sufficient time to perform a successful mutation on all these entries, where successful means that the correct value is being chosen. This takes the time claimed in the theorem.

Theorem 3 does not guarantee that an EA will find an optimum in the stated time, it just says that every EA needs at least the stated time, and maybe much longer. Depending on the problem instance, the EA may get stuck in local optima, particularly if only local mutations are being used.

In contrast, global mutations guarantee that any particular solution can be created from any other one. This means that an EA using global mutations will eventually find a global optimum, albeit this time can be exponential. A proof of the following theorem is given in the appendix.

Theorem 4: Every EA that constructs new solutions using global mutations (potentially after applying recombination and/or other operators) finds a global optimum for the
TABLE I
SUMMARY OF THE ALGORITHMS USED IN OUR STUDY.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Encoding</th>
<th>Mutation</th>
<th>Crossover</th>
<th>Parent selection</th>
<th>Survival selection</th>
<th>Population size</th>
<th>Normalisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1+1) EA</td>
<td>Direct</td>
<td>Uniform</td>
<td>–</td>
<td>–</td>
<td>Elitist</td>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>RLS</td>
<td>Direct</td>
<td>Uniform</td>
<td>1 entry</td>
<td>–</td>
<td>Elitist</td>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>Pop-EA</td>
<td>Direct</td>
<td>Uniform</td>
<td>Rows/columns</td>
<td>λ binary tournaments</td>
<td>Elitist</td>
<td>μ</td>
<td>Yes</td>
</tr>
<tr>
<td>(1+1) EA no-norm</td>
<td>Direct</td>
<td>Uniform</td>
<td>–</td>
<td>–</td>
<td>Elitist</td>
<td>1</td>
<td>No</td>
</tr>
<tr>
<td>GA [6]</td>
<td>Binary</td>
<td>Bit-flip</td>
<td>2-D single point</td>
<td>2 binary tournaments</td>
<td>Elitist</td>
<td>μ</td>
<td>No</td>
</tr>
</tbody>
</table>

1 Number of tournaments per generation.

Fig. 2. Example of a linear schedule: all tasks have to be processed sequentially.

PSP, with or without normalisation, in an expected number \((knm)^n\) of constructed solutions.

Note that this upper time bound is very crude – it is larger than the expected optimisation time of random search or exhaustive search. Many instances can be solved much faster, as we will see in section VI-D. However, Theorem 4 reflects the truth that EAs may indeed perform worse than random search. This happens when the algorithm gets trapped in a local optimum which is very dissimilar to all global optima, hence requiring many dedication values to be changed simultaneously in one mutation. This is hard to achieve for EAs. Section VI-E contains an example of a hard local optimum, where exponential time is needed to escape from it (albeit the time is smaller than the bound from Theorem 4).

C. Time to Feasibility

In order to see how efficiently our treatment of infeasible solutions guides evolution towards feasible search points, we estimate the expected time until feasibility is reached. The following theorem makes a very reasonable assumption: the total number of skills is not larger than some polynomial in \(nm\).

**Theorem 5 ([1]):** Consider RLS or the (1+1) EA, with or without normalisation, on any solvable instance where \(\sum_{i=1}^n |\text{skill}_i| \leq (nm)^\delta\) for some constant \(\delta \geq 0\). The expected time until some feasible schedule is found is \(O(nm \log(nm))\).

For a proof we refer to [1]. The upper bound is by a factor of \(k\) smaller than the lower bound from Theorem 3. This indicates that the time to feasibility is only a small fraction of the optimisation time.

D. Easy Instance Classes: Linear Schedules

Now we turn to a class of illustrative “easy” instances. Assessing how effective an EA is on easy instances is an excellent starting point for an analysis. It gives a good baseline for comparisons with other results on the running time, and most importantly we learn about the search behaviour of an EA. We need a good understanding of easy instances before we can move on to more complicated instance classes.

As easy cases we consider schedules with a “linear” structure: the task precedence graph is a chain of \(m\) vertices, or more generally any directed acyclic graph that contains such a chain. In this case all tasks have to be completed sequentially. Figure 2 gives an example. We also assume that all salaries are equal. So the problem boils down to minimising the completion time.

Note that if normalisation is switched on, it is never actually applied as at each time only one task is processed. The issue of whether normalisation is used or not is irrelevant for linear schedules. The same holds for the repair mechanism [7].

Also note that linear schedules can be optimised by minimising the completion times of all tasks individually. This property is called separability: a function is separable if it can be written as the sum of functions defined on disjoint sets of variables. This is the case for the completion time of (feasible) linear schedules as the completion time can be written as the sum of completion times for all single tasks.

We know from Theorem 2 that the optimal solution is a dedication matrix where all entries are set to 1. This is the only global optimum as otherwise increasing any dedication value \(d_{ij} < 1\) would decrease the execution time for the \(j\)-th task, while keeping the other execution times unchanged.

RLS only changes one dedication entry at a time, which implies that all tasks will be optimised separately. This leads to the following result, which was proven in [1].

**Theorem 6 ([1]):** Consider an instance with \(n\) employees with equal salaries and \(m\) tasks with arbitrary positive efforts. Let the task precedence graph contain an \(m\)-vertex chain as subgraph. Then the expected optimisation time of RLS, with or without normalisation or repair, is of order \(\Theta((knm \ln(nm)))\).

We believe that the (1+1) EA is asymptotically as efficient as RLS on linear schedules, i.e., the bounds from Theorem 6 also hold for the (1+1) EA. However, proving this is more difficult than for RLS. The reason is that global mutations can change several dedications at the same time. Mutations with beneficial and detrimental effects can be accepted if there is a net improvement in fitness. Such situations can happen because jobs with large task efforts have a larger impact on the fitness than jobs with smaller efforts. This can lead to situations where the fitness improves, but the (1+1) EA actually moves away from the global optimum. We are looking...
for a formal proof that the (1+1) EA is efficient for arbitrary task efforts.

We first prove a polynomial running time bound for the (1+1) EA on any linear schedule. We do not believe that this bound is asymptotically optimal, but it gives a polynomial upper bound for all instances. We then give better upper bounds for special cases, supporting our belief that the (1+1) EA is as efficient as RLS.

**Theorem 7**: Consider an instance with \( n \) employees with equal salaries and \( m \) tasks with arbitrary positive efforts. Let the task precedence graph contain an \( m \)-vertex chain as a subgraph. Then the expected optimisation time of the (1+1) EA, with or without normalisation or repair, is at most \( O((knm)^2) \).

A proof is given in the appendix; it uses a structural result about separable functions [38], saying that under certain conditions the expected time for the (1+1) EA optimising a separable function is bounded from above by the sum of expected times for minimising the completion times of all tasks sequentially.

If \( k = 1 \), i.e., only dedications 0 and 1 are allowed, the analysis becomes easier. It is not hard to see that then the fitness function is monotonic in the following sense: flipping only 0-bits to 1 and not flipping any 1-bit to 0 results in a strict fitness improvement. Such an operation implies that the total dedication on every task is non-decreasing, and at least one total dedication is increased strictly. By Doerr et al. [39] this yields the following.

**Theorem 8**: Consider an instance with \( n \) employees with equal salaries and \( m \) tasks with arbitrary positive efforts. Let the task precedence graph contain an \( m \)-vertex chain as a subgraph. If \( k = 1 \) then the expected optimisation time of the (1+1) EA is \( O((nm)^{3/2}) \). Additionally, if the mutation probability is changed to \( c/(nm) \) for a constant \( 0 < c < 1 \), the expected time is even bounded by \( O(nm \log(nm)) \).

With \( k = 1 \) the upper bound of \( O((nm)^{3/2}) \) is the same as \( O((knm)^{3/2}) \). Compared to the bound \( O((knm)^2) \) from Theorem 7, we have reduced the exponent of the running time. The second statement about reduced mutation probabilities indicates that the upper bound of \( O(knm \log(nm)) \) for RLS might also hold for the (1+1) EA, as we have no reason to believe that a tiny reduction of the mutation rate by an arbitrary constant factor (e.g., by 0.99) has a dramatic effect on the expected optimisation time here.

We consider one more setting which supports our belief. The following result states that whenever we have a fixed number of employees, a special case of dedication values 0 or 1, and an arbitrary number of tasks, then the expected optimisation time of the (1+1) EA is upper bounded just like it is for RLS (cf. Theorem 6). The proof uses another structural result about separable functions [38], stating that in some settings the (1+1) EA will optimise all subfunctions of a separable function in parallel.

**Theorem 9**: In the setting of Theorem 6, if additionally \( k = 1 \) (i.e., we have dedications 0 or 1) and \( n = O(1) \), then the expected optimisation time of the (1+1) EA, with or without normalisation, is bounded by \( O(knm \log(nm)) \).

The upper bound is even \( O(m \log m) \) due to our assumptions on \( k \) and \( n \), but to make a fair comparison we write it so that it matches the upper bound for RLS from Theorem 6. Again, the proof is given in the appendix.

All results above support our conjecture that the (1+1) EA is as effective as RLS for linear schedules.

### E. Difficult Instance Classes

We now look at difficult instances to get insight into what makes the PSP hard. Linear schedules are easy to solve as all tasks are processed sequentially. So we consider settings with tasks being processed in parallel. We have already seen in section V-A that without normalisation an EA struggles in finding an optimal balance between dedications for different tasks. With normalisation this problem becomes a lot easier. But we show in the following that even with normalisation, RLS and the (1+1) EA can still struggle in finding an optimal balance.

In particular, RLS can get stuck in local optima even on a very simple and tiny problem instance where costs are irrelevant. The instance contains two tasks with efforts 4 and 5, respectively. There is only one employee, so trivially we always get a fixed cost for any feasible schedule. Starting with equal dedication values of 1/2, both tasks finish at similar times. The task with higher effort takes two more time steps, throughout which the employee only works half time, see Figure 3 (a). This increases the completion time, compared to the optimal schedule where both dedications are 1, see Figure 3 (d). Any local operation either creates an infeasible schedule or increases the imbalance between the two tasks. This increases the time period at the end where the employee only works half time and it makes the schedule even worse, see Figures 3 (b), (c).

RLS has a positive probability of starting in the local optimum, and no local mutation is accepted. Thus, we get:

**Theorem 10**: There is an instance with \( k = 2 \), only \( m = 2 \) tasks and just \( n = 1 \) employee (see Figure 3) where RLS with normalisation has an infinite expected optimisation time.

This example shows that the global mutation operator used in the (1+1) EA is important in general, as it may be necessary to change several dedication values at the same time.
Also note that in the instance from Figure 3 without normalisation—with or without repair—an optimal completion time of $4 + 5 = 9$ can only be achieved if the granularity is set such that dedication values $4/9$ and $5/9$ are feasible choices. If this is not the case (i.e., if $k$ is not a multiple of 9), only completion times larger than 9 are possible. It is easy to see that for every given value of $k$ we can find a similar instance where every feasible schedule without normalisation has a strictly larger completion time than the optimal schedule with normalisation.

The above instance can also be generalised towards more than two tasks. We do this in such a way that the $(1+1)$ EA has to make a very large mutation in order to escape from a local optimum. Set $\text{eff}_1 = \text{eff}_2 = \cdots = \text{eff}_{m-1} = 3m$ and $\text{eff}_m = 3m + 1$. There are no precedence constraints. Set $k = m$ and $n = 1$, that is, there is only one employee.

Similar to the instance from Theorem 10, setting all dedication values to $1/k = 1/m$ yields a local optimum. All other solutions are worse—except for solutions where all dedications are larger than $1/m$. In order to escape from the local optimum, all dedication values need to be changed in a single mutation. The expected running time then increases exponentially in the number of tasks.

**Theorem 11:** For every $m \in \mathbb{N}$ there is an instance with $m$ tasks and one employee and an initial search point for the $(1+1)$ EA with normalisation such that its expected optimisation time from there is at least $m^m$.

A proof is given in the appendix.

Note that Theorem 11 does not make a claim about the expected optimisation time with uniform initialisation as the probability of getting stuck in the local optimum stated in the proof might be exponentially small. However, Theorem 11 proves the existence of local optima which are very hard to overcome.

Even though normalisation makes it easier to balance dedications, there is still a risk of non-optimal equilibria between dedications for tasks processed in parallel. This can present a major obstacle for EAs as many dedications might need to be changed in a single mutation.

Finally, note that the instance from Theorem 11 can easily be modified by adding further tasks that have to be processed sequentially, and after all existing tasks. Adding these new tasks does change the number of tasks, but it does not significantly affect the expected optimisation time. This means that for every given value of $m$ we can mix characteristics of our “hard” and “easy” instances. By deciding how many tasks shall belong to the easy or hard part, we can create instances of tunable difficulty. This shows that there are not only easy and hard instances, but there is a fine scale of expected optimisation times that can be observed for EAs.

**VII. Experimental Analysis**

Faced with several different possible choices of algorithms that can be used for the PSP, it is desirable to provide the software project manager with insight into what algorithm to choose. This insight should be followed by evidence demonstrating what algorithms are likely to behave better according to different evaluation criteria that may affect the project manager’s decision. With this as the main objective, this section presents an empirical analysis of different algorithms based on the hit rate (number of runs in which a feasible solution was found), solution quality and convergence time. As a secondary objective, this section also provides information that can be used by other researchers trying to improve these algorithms.

The algorithms included in the analysis are our three algorithms with normalisation ($(1+1)$ EA, RLS and Pop-EA), and two algorithms without normalisation ($(1+1)$ EA no-norm and a state-of-the-art GA [6])\(^1\), as shown in Table I. Preliminary results using $(1+1)$ EA have been presented in [1]. We will refer to the algorithms with normalisation simply as $(1+1)$ EA, RLS and Pop-EA. When referring to the $(1+1)$ EA without normalisation, we will explicitly state that or use the term “no-norm”.

The $(1+1)$ EA without normalisation works similarly to our $(1+1)$ EA, but with an extra condition on the fitness function to consider overwork. In this case, the fitness is $f(x) = w_{\text{press}} + \text{over}$ if the skills constraint (eq. 1) is satisfied but there is overwork. The value over is the total amount of over-time spent by all employees during the project [6] and $w_{\text{press}} = w_{\text{cost}} \cdot 2 \sum_{i=1}^{n} \sum_{j=1}^{m} s_{i j} \cdot \text{eff}_j + w_{\text{time}} \cdot 2k \sum_{j=1}^{m} \text{eff}_j$.

This section is further divided as follows: section VII-A presents the data sets used in the experiments; section VII-B presents the parameters; section VII-C presents the results in terms of hit rate; section VII-D in terms of solution quality; and section VII-E in terms of convergence time.

**A. Data Sets**

In order to make a fair comparison against the GA, we used the same 48 instances (benchmarks 1-5) of the PSP as before [6]. Benchmarks 4–5 can be found at http://tracer.lcc.uma.es/problems/psp/generator.html. Benchmarks 1–3 were made available to us by Alba and Chicano. As the number of problem instances is high, we avoid biasing the conclusions and remove the possibility of hand-tuning the algorithm to a particular problem instance. Moreover, we can verify how the algorithms are affected by problem features such as the number of employees, number of tasks, number of employees’ skills and number of project demanded skills.

Benchmarks 1-3 were used to analyse the effect of varying each of the three problem features (number of employees, number of tasks, and number of employees’ skills) while maintaining the other features fixed. These data sets use the same salary for all employees ($10,000), so that, given a project, the cost of all solutions for this project is always the same. In this way, the ideal cost per unit of time is known and it is possible to evaluate how close a given solution is to the optimum in terms of completion time.

Benchmark 1 is composed of 4 instances varying the number of employees among 5, 10, 15 and 20. Benchmark 2 is composed of 3 instances varying the number of tasks among 10, 20 and 30. Benchmark 3 is composed of 5 instances

\(^1\)Our implementation of the $(1+1)$ EA, RLS and Pop-EA were based on the Opt4J framework [40].
varying the number of employees’ skills among 2, 4, 6, 8 and 10 skills, which are randomly selected from a set of 10 project skills. Each task requires five different skills in this benchmark. In benchmarks 1 and 2, all employees have all necessary skills, i.e., the skills constraint (eq. 1) is always satisfied. Instances within each of the benchmarks 1 and 3 represent the same project to be developed (i.e., they have the same tasks and TPG) with the number of tasks fixed to 10. Each instance of benchmark 2 represents a different project, as the number of tasks is different. In benchmarks 2 and 3, the number of employees is fixed to 5.

Benchmarks 4 and 5 are composed of instances that represent different projects and each employee has a different salary. Each benchmark is composed of 18 instances which vary all the previous problem features. The number of employees can be 5, 10 or 15 and the number of tasks can be 10, 20 or 30. In benchmark 4, the total number of project skills is 10, and two ranges were considered separately for the number of employees’ skills: 4-5 and 6-7. In benchmark 5 the number of project demanded skills can be 5 or 10, and the number of skills per task and employee is in the range 2-3.

B. Experimental Setup

For a fair comparison between our algorithms and the GA, we have used the following parameters, which correspond to the parameters used previously in the literature [6]:

- Parameters for all our algorithms:
  - constant for the granularity of the solution $k = 7$;
  - $w_{\text{cost}} = 10^{-6}$;
  - $w_{\text{time}} = 10^{-1}$;
  - number of independent runs per problem instance 100.

- Parameters specific to the non-population-based algorithms:
  - number of generations 5064 (= number of fitness evaluations considering the initial population of size 64 used previously [6]).

- Parameters specific to the Pop-EA:
  - size of the population $\mu = 64$, as in [6];
  - number of children generated at each generation $\lambda = 64$;
  - number of generations 79 (this is the rounded number of fitness evaluations in our experiments with the non-population-based algorithms divided by the population size ($79.125 \approx 79$), giving a total of 5056 fitness evaluations).

The GA [6] always applies crossover, whereas our Pop-EA uses a probability of crossover. This probability was set to 0.75, which is the middle of the crossover rate interval [0.6, 0.9] suggested by Eiben and Smith [16]. The mutation operator provided by the framework Opt4J selects any value in $\{0/k, \ldots, k/k\}$ uniformly at random. This does not exclude the original value, hence each deduction value is only changed with probability $1/(nm) \cdot k/(k + 1)$ in our experiments with (1+1) EA and Pop-EA.

C. Hit Rate

An algorithm designed for the PSP should ideally provide feasible allocations of employees to tasks. Algorithms unable to provide feasible solutions are not useful for the software project manager. With that in mind, this section analyses the hit rate (Table II).

Our three algorithms with normalisation always achieved hit rate 100, i.e., all runs always found a feasible solution. The algorithms without normalisation frequently presented much lower hit rates, sometimes even a hit rate of zero, i.e., even after running the algorithms 100 times, no feasible solution was found. In order to produce a boundary to capture the unknown population\(^2\) of hit rates, we calculated the modified (adjusted) Wald binomial confidence interval [41] with 95% of confidence for each hit rate. For hit rate of 100, the interval is [96.83, 100.00]. The upper limits of the confidence intervals for the hit rates of 90 or less shown in Table II are all lower than the lower limit of the interval for hit rates of 100, meaning that the differences between hit rates of 90 or less and the hit rate of the algorithms with normalisation are statistically significant.

As the (1+1) EA without normalisation presented in general much lower hit rates than the (1+1) EA with normalisation, normalisation plays an important role in improving the algorithm’s performance. In addition, as the hit rate of the (1+1) EA is the same as the hit rate of RLS and Pop-EA, the choice among these three algorithms with normalisation did not alter the results in terms of hit rate for these problem instances.

Another interesting observation is that our algorithms with normalisation always managed to achieve hit rate of 100 independent of the problem features. The previous study [6] revealed that the GA’s hit rate varied depending on the problem features. Our experiments using (1+1) EA without normalisation also presented different hit rates for different instances. So, normalisation also plays an important role in making the hit rate of the algorithm less dependent on the problem features. In other words, it helps to improve robustness of the algorithm to different problem instances. As the hit rate was the same for all algorithms with normalisation, they all presented similar robustness.

In summary, this section shows that the algorithms with normalisation obtained the perfect hit rate, for all tested problem instances.

D. Solution Quality

Another criterion that should be considered by a software manager when choosing a PSP algorithm is the quality of the solutions provided, i.e., how good the cost and completion time of the solutions is. In this section, we analyse the quality of the feasible solutions in terms of fitness, cost and completion time.

\(^2\)The term population in this sentence refers to the statistical meaning of population, against the known sample of hit rates.
TABLE II
HIT RATE OUT OF 100 RUNS FOR THE (1+1) EA WITHOUT NORMALISATION. AND AN EXISTING GA (OBTAINED FROM [6]). THE HIT RATE FOR THE ALGORITHMS WITH NORMALISATION WAS ALWAYS 100.

(a) Benchmarks 1–3

<table>
<thead>
<tr>
<th>Employees</th>
<th>Benchmark 1</th>
<th>Benchmark 2</th>
<th>Benchmark 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>(1+1) EA no-norm</td>
<td>97</td>
<td>87</td>
</tr>
<tr>
<td>10</td>
<td>(1+1) EA no-norm</td>
<td>100</td>
<td>65</td>
</tr>
<tr>
<td>15</td>
<td>(1+1) EA no-norm</td>
<td>97</td>
<td>49</td>
</tr>
<tr>
<td>20</td>
<td>(1+1) EA no-norm</td>
<td>96</td>
<td>51</td>
</tr>
</tbody>
</table>

(b) Benchmarks 4–5

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Benchmark 4</th>
<th>Benchmark 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>(1+1) EA no-norm</td>
<td>2.089</td>
</tr>
<tr>
<td>20</td>
<td>(1+1) EA no-norm</td>
<td>0.217</td>
</tr>
<tr>
<td>30</td>
<td>(1+1) EA no-norm</td>
<td>0.000</td>
</tr>
</tbody>
</table>

TABLE III
AVERAGE RANKING OF ALGORITHMS ACCORDING TO FITNESS ACROSS PROBLEM INSTANCES. SMALLER RANKING VALUES ARE BETTER RANKINGS.

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Average Rank</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pop-EA</td>
<td>1.875</td>
<td>0.5708</td>
</tr>
<tr>
<td>(1+1) EA</td>
<td>2.0833</td>
<td>0.3472</td>
</tr>
<tr>
<td>RLS</td>
<td>2.7292</td>
<td>0.6058</td>
</tr>
<tr>
<td>(1+1) EA no-norm</td>
<td>4.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

1) Ranking of Approaches According to Fitness: The fitness function used in this paper is a measure of the solution quality given the relative importance of cost and completion time selected by a project manager. Following Demšar’s recommendation for comparing two algorithms over multiple data sets [42], we used the Friedman test to compare the average ranking of the (1+1) EA, RLS, Pop-EA and the (1+1) EA without normalisation according to their fitness across problem instances. The average ranking is shown in Table III. The GA has not been ranked here because we do not have its corresponding fitness values from [6]. The Friedman test detected statistically significant difference in the average ranks ($F_D = 246.30 > F(3, 141) = 2.67, p$-value less than 0.0001).

As the Friedman test detected statistically significant difference, the Nemenyi post-hoc test [42] was then used to compare the average rank of each algorithm against each other. In this way, it is possible to know which of the algorithms actually performed differently or similarly to each other. The test detected statistically significant differences between all algorithms at the level of $\alpha = 0.05$ except between (1+1) EA and RLS (see Table IV). These results together with the average rankings show that Pop-EA obtained the best average rank across problem instances, (1+1) EA without normalisation obtained the worst, and (1+1) EA and RLS obtained similar average ranks to each other.

It is worth noting that the rank obtained by the (1+1) EA without normalisation was always the worst, for all problem instances (the standard deviation of its ranking is zero in Table III). Normalisation was able to improve fitness, in such a way that our (1+1) EA achieved better ranking than the (1+1) EA without normalisation. So, normalisation played an important role to improve fitness. The use of a population through Pop-EA improved the fitness further.

Pop-EA obtained the best average ranking across problem instances. Out of 48 problem instances there were only five instances where Pop-EA was not the best ranked algorithm:

- benchmark 4’s instance with 10 tasks, 6–7 employees’ skills, 10 employees;
- benchmark 4’s instance with 10 tasks, 4–5 employees’ skills, 5 employees;
- benchmark 5’s instance with 10 tasks, 5 project skills, 10 employees;
- benchmark 5’s instance with 20 tasks, 10 project skills, 5 employees;
- benchmark 5’s instance with 20 tasks, 10 project skills, 15 employees.

For these problem instances, either the (1+1) EA or RLS obtained the best rank. These instances have varied number of tasks, employees’ skills, project skills and employees. It was not possible to identify a pattern that reveals for what type of instances the Pop-EA was not the best ranked algorithm. However, we show in section VII-D2 that the risk of using Pop-EA instead of (1+1) EA or RLS is very small.

In summary, this section shows that normalisation plays an important role in improving the (1+1) EA’s fitness, and that the use of Pop-EA was able to improve it further, whereas...
RLS obtained similar fitness to the (1+1) EA. Pop-EA usually obtained the best fitness.

2) Practical Effect of the Differences in Fitness: In addition to the statistical significance of the difference in fitness, it is also important to analyse the practical differences in solution quality when choosing an algorithm. For example, a project manager could still opt for using an algorithm likely to produce worse solution quality than the best one if the difference in quality between these two algorithms is unlikely to be high. A possible reason for a software manager to opt for a worse algorithm would be if he/she has easier access to the implementation of this algorithm or the algorithm is far more efficient to execute than others. If the software manager can use the algorithm most likely to be the best, he/she should still be aware of the risk incurred in this choice. For example, if an algorithm is likely to produce the best solution quality in most cases, and in a few cases the solution quality is likely to be just slightly worse than other algorithms, then the risk incurred in opting for this “best” algorithm is very small. Thus, it is reasonable for a software manager to always opt for it. On the other hand, if this algorithm is likely to perform much worse than another in a few cases, the risk in always choosing it is high.

We define the practical effect of differences in solution quality for a certain problem instance $i$ as:

$$PE_i = \frac{|\text{measure}(A) - \text{measure}(B)|}{\text{measure}(A)}$$

where $A$ is the algorithm with the best fitness for the problem instance $i$, $B$ is an alternative algorithm, and measure is either the average cost or completion time. This is a measure of how large the difference in terms of cost/completion time is in relation to the cost/completion time of the best fit solution.

Table V presents the minimum, maximum, average and standard deviation of the practical effect in terms of cost and completion time across problem instances. We can observe that the practical effect in terms of cost and completion time considering only the set of algorithms with normalisation is very low. For both cost and completion time, the maximum practical effect is considerably less than one percent. In contrast, the practical effects between the best ranked algorithm with normalisation and the (1+1) EA without normalisation are much larger.

For illustration purposes, Table V also shows the actual maximum difference between the best and the worst fitness ranked algorithms in cost and time. Considering that the salary of an employee is around $10,000 per unit of time, let’s regard units of time as months. We can see that the maximum difference in cost between the best and the worst fitness ranked algorithms with normalisation is only approximately $173$, whereas the difference considering the (1+1) EA without normalisation is of approximately $57,521$. In terms of completion time, the differences are about 1 day and a half, and more than 7 years, respectively. This further illustrates the practical effect of choosing between the algorithms with and without normalisation.

In summary, this section shows that it is important to opt for algorithms with normalisation in terms of practical effect of solution quality, and that the risk incurred in choosing one or another algorithm with normalisation is not high. For instance, a software manager could opt for the Pop-EA, which has the best average rank in terms of fitness and is very unlikely to perform much worse than the other algorithms.

3) Variance of Cost and Completion Time: The variance in cost and completion time is another factor that should be taken into account when choosing an algorithm. If the solution quality when running an algorithm with different random seeds varies a lot, it would be difficult for the software manager to know whether a certain run is likely to have produced good or bad solutions.

The algorithms with normalisation obtained very low variances in cost and completion time across solutions. Table VI shows the minimum and maximum standard deviation divided by the average. Statistics regarding the GA were not available from [6]. The standard deviation in cost was never more than 0.04 percent of the average cost and the standard deviation in completion time was never more than 0.60 percent of the average completion time of the project for the algorithms with normalisation. When normalisation was not used in the (1+1) EA, the variance in cost was from 0.25–1.90 percent of the total cost and the variance in completion time was 4.65–52.19 percent of the total completion time whenever there were more than three feasible solutions. So, considering the criterion variance, it would be good for a software manager to opt for algorithms with normalisation.

The reason for the low variances is because normalisation made it easier to find near optimal solutions, as shown in section VII-D4. As it was easier to find near optimal solutions, there were less cases where the solution deviated considerably from the optimum, and thus the variances were lower.

4) Distance to the Optimal Cost and Completion Time: No matter whether a certain algorithm produces solutions of better or worse quality than another, if all these solutions are far away from the best cost and completion time, i.e., if the cost and completion time of the solutions are very poor, then these algorithms are not useful. As we know the optimal cost and completion time for benchmarks 1–3, it is possible to evaluate how close the solutions found by the algorithms are to these optima. Tables VII and VIII show some values descriptive of the solution quality for benchmarks 1–3. The cost for all solutions of benchmarks 1, 3 and the first instance of benchmark 2 were $980,000. The cost for the second and third instances of benchmark 2 were $2,600,000 and $2,700,000, respectively.

The cost per unit of time ($\text{cost/time}$) for benchmarks 2 and 3 is optimal at $50,000 (all employees working full time), as there are 5 employees and all have the same salary. Each problem instance of benchmark 1 has the optimal $\text{cost/time}$ increased by $50,000 in relation to the previous instance, as the number of employees is increased by 5. As we can see from Table VII, all our algorithms with normalisation obtained near-optimal solutions for these benchmarks, and our Pop-EA obtained solutions even closer to the optimum than the other algorithms with normalisation. The algorithms without normalisation, in contrast, obtained solutions further away from the optimum.
TABLE V
STATISTICS OF THE PRACTICAL EFFECT (EQUATION 2) OF DIFFERENCES IN SOLUTION QUALITY ACROSS PROBLEM INSTANCES. THE ACTUAL MAXIMUM DIFFERENCE IN UNITS OF COST AND TIME IS ALSO SHOWN. STATISTICS REGARDING THE DIFFERENCE IN COST DO NOT CONSIDER PROBLEM INSTANCES WITH FIXED COST (BENCHMARKS 1−3), AND ONLY FEASIBLE SOLUTIONS ARE CONSIDERED.

<table>
<thead>
<tr>
<th>Best against worst fitness ranked algorithm with normalisation</th>
<th>Best fitness ranked algorithm against (1+1) EA without normalisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost</td>
<td>Duration</td>
</tr>
<tr>
<td>Min PE</td>
<td>0.0000%</td>
</tr>
<tr>
<td>Max PE</td>
<td>0.0238%</td>
</tr>
<tr>
<td>Ave PE</td>
<td>0.0025%</td>
</tr>
<tr>
<td>Std Dev PE</td>
<td>0.0043%</td>
</tr>
<tr>
<td>Max difference in units of time and cost</td>
<td>172.8333</td>
</tr>
</tbody>
</table>

TABLE VI
MINIMUM AND MAXIMUM VALUES OF STANDARD DEVIATION IN COST AND COMPLETION TIME ACROSS DIFFERENT SOLUTIONS FOR THE SAME PROBLEM INSTANCE DIVIDED BY AVERAGE. STATISTICS REGARDING MINIMUM STANDARD DEVIATION OF COST DO NOT CONSIDER THE PROBLEM INSTANCES WITH FIXED COST (BENCHMARKS 1−3), AND ONLY STATISTICS FOR PROBLEM INSTANCES WITH AT LEAST THREE FEASIBLE SOLUTIONS ARE CONSIDERED.

<table>
<thead>
<tr>
<th>Pop-EA</th>
<th>(1+1) EA</th>
<th>RLS</th>
<th>(1+1) EA no-norm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min sdev / avg cost</td>
<td>0.0000%</td>
<td>0.0010%</td>
<td>0.0059%</td>
</tr>
<tr>
<td>Max sdev / avg cost</td>
<td>0.0148%</td>
<td>0.0344%</td>
<td>0.0378%</td>
</tr>
<tr>
<td>Min sdev / avg completion time</td>
<td>0.0002%</td>
<td>0.0183%</td>
<td>0.0233%</td>
</tr>
<tr>
<td>Max sdev / avg completion time</td>
<td>0.2768%</td>
<td>0.1834%</td>
<td>0.5850%</td>
</tr>
</tbody>
</table>

TABLE VII
QUALITY OF FEASIBLE SOLUTION MEASURED BY THE COST DIVIDED BY THE COMPLETION TIME. THE AVERAGES WERE CALCULATED CONSIDERING ONLY THE RUNS IN WHICH A FEASIBLE SOLUTION WAS FOUND. THE BEST VALUES ARE IN BOLD. n, m and sk are the number of employees, tasks and employees’ skills. Stdev is the standard deviation of the average values reported in the table.

(a) Benchmark 1: avg. cost/time. The optimal value is n · 10,000.

<table>
<thead>
<tr>
<th>n</th>
<th>Pop-EA</th>
<th>(1+1) EA</th>
<th>RLS</th>
<th>(1+1) EA no-norm</th>
<th>GA [6]</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>49,998</td>
<td>49,978</td>
<td>49,901</td>
<td>36,581</td>
<td>44,790</td>
</tr>
<tr>
<td>10</td>
<td>99,981</td>
<td>99,940</td>
<td>99,828</td>
<td>75,559</td>
<td>86,957</td>
</tr>
<tr>
<td>15</td>
<td>149,964</td>
<td>149,900</td>
<td>149,660</td>
<td>114,754</td>
<td>126,779</td>
</tr>
<tr>
<td>20</td>
<td>199,919</td>
<td>199,830</td>
<td>199,614</td>
<td>151,235</td>
<td>166,667</td>
</tr>
</tbody>
</table>

(b) Benchmark 2: avg. cost/time. The optimal value is 50,000.

<table>
<thead>
<tr>
<th>n</th>
<th>Pop-EA</th>
<th>(1+1) EA</th>
<th>RLS</th>
<th>(1+1) EA no-norm</th>
<th>GA [6]</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>49,998</td>
<td>49,978</td>
<td>49,901</td>
<td>36,581</td>
<td>44,790</td>
</tr>
<tr>
<td>10</td>
<td>99,981</td>
<td>99,940</td>
<td>99,828</td>
<td>75,559</td>
<td>86,957</td>
</tr>
<tr>
<td>15</td>
<td>149,964</td>
<td>149,900</td>
<td>149,660</td>
<td>114,754</td>
<td>126,779</td>
</tr>
<tr>
<td>20</td>
<td>199,919</td>
<td>199,830</td>
<td>199,614</td>
<td>151,235</td>
<td>166,667</td>
</tr>
</tbody>
</table>

(c) Benchmark 3: avg. cost/time. The optimal value is 50,000.

<table>
<thead>
<tr>
<th>sk</th>
<th>Pop-EA</th>
<th>(1+1) EA</th>
<th>RLS</th>
<th>(1+1) EA no-norm</th>
<th>GA [6]</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>49,998</td>
<td>49,983</td>
<td>49,925</td>
<td>–</td>
<td>45,230</td>
</tr>
<tr>
<td>4</td>
<td>49,996</td>
<td>49,983</td>
<td>49,888</td>
<td>–</td>
<td>45,069</td>
</tr>
<tr>
<td>6</td>
<td>49,998</td>
<td>49,980</td>
<td>49,894</td>
<td>38,762</td>
<td>44,651</td>
</tr>
<tr>
<td>8</td>
<td>49,998</td>
<td>49,979</td>
<td>49,905</td>
<td>36,518</td>
<td>44,617</td>
</tr>
<tr>
<td>10</td>
<td>49,998</td>
<td>49,978</td>
<td>49,901</td>
<td>37,182</td>
<td>44,427</td>
</tr>
</tbody>
</table>

Stdev | 8.34 | 6.43 | 34.09 | 10,234.18 | – |

It is worth noting that these results corroborate the analysis presented in section VII-D1, which show that normalisation plays a key role in improving the solution quality obtained by the (1+1) EA, and that Pop-EA manages to improve it further through the use of a population. They also corroborate the results in section VII-D2, which show that the practical effect of using an algorithm with normalisation instead of without normalisation was large, whereas choosing among the three algorithms with normalisation did not change solution quality as much.

The results also show that the solution quality of our algorithms with normalisation were less affected by variations in the number of employees, tasks and employees’ skills than the other algorithms. This is shown by their more similar product n · time across instances of benchmark 1 (lower standard deviation in Table VIII), and by their more similar cost/time across instances of benchmarks 2 and 3 (lower standard deviation in Tables VII(b) and VII(c)). These values were more different across instances of a same benchmark for the algorithms without normalisation.

In summary, this section shows that the solutions found by the algorithms with normalisation for benchmarks 1–3 were near-optimal, whereas the solutions found by the (1+1) EA without normalisation and by the GA were further from the optimal cost and completion time.

E. Evaluation of Convergence Time

This section complements the analysis of how much each algorithm is influenced by the features of the problem instances such as the number of employees and tasks. If the
solution quality varies a lot depending on the features of the problem instance, then the choice for an algorithm is more difficult and incurs more risk. Sections VII-C and VII-D show that the hit rate and the quality of the solutions produced by algorithms with normalisation were much less affected by problem features than the ones produced by the algorithms without normalisation. However, even if the hit rate and the solution quality obtained by a certain algorithm for different problem instances are not affected by different features of these instances, that does not necessarily mean that the algorithm is not affected by such features.

An analysis of convergence time can show to what extent the algorithms were affected by problem features, providing useful insight into whether the algorithm would still be able to provide good solutions if the problem grew a bit more. Fast convergence to good solutions means that if the problem was a bit larger, the algorithm may still be able to find good solutions within the budget of fitness evaluations.

We analyse the convergence plots of the algorithms with the best and the worst fitness ranking: Pop-EA and (1+1) EA without normalisation, respectively (section VII-D1). The plots were generated for each problem instance and use the inverse of the average fitness across multiple runs, rather than the fitness itself. This allows extra visual comparison of the plots of the Pop-EA against the plots of the GA published in [6]. These plots are comparable because the inverse of the fitness function of the Pop-EA is the same as the fitness function of the GA when the solutions are feasible, and Pop-EA obtained feasible solutions in 100% of the runs.

Examples of representative plots are shown in Figure 4. Pop-EA frequently starts with solutions of considerably good fitness, whereas the algorithms without normalisation (both (1+1) EA without normalisation and the GA) typically start with solutions of very low quality (1/fitness close to zero). For example, Pop-EA’s inverse of fitness in Figures 4(a), 4(d) and 4(e) starts higher than the (1+1) EA no-norm’s in Figures 4(b), 4(c) and 4(f), respectively. This is because normalisation eases the overwork constraint, allowing for feasible solutions to be found very quickly, whereas the algorithms without normalisation need longer time to find feasible solutions.

In addition to frequently starting off with solutions of better quality, the Pop-EA also converges very quickly, usually within 1280 fitness evaluations (20 generations). While algorithms without normalisation struggled to find feasible solutions, Pop-EA was able not only to quickly find feasible solutions, but also to converge faster to solutions of better quality. This occurs both for benchmarks 1–3, for which it is known that Pop-EA found near-optimal solutions (section VII-D4), and for benchmarks 4 and 5. Even though we do not know the optimal cost and duration for benchmarks 4 and 5, the fact that convergence was faster for all benchmarks and the better fitness achieved by the Pop-EA in comparison to (1+1) EA without normalisation suggest that this fast convergence has led to solutions of good, possibly near-optimal, quality, rather than being a premature convergence to poor solutions.

The convergence time usually did not present considerable changes for instances with different numbers of employees, employees’ skills and project skills. For example, the convergence time is similar for different numbers of employees in Figure 4(a). This confirms that Pop-EA was not much affected by changes in these features for these problem instances. Changes in the number of tasks affected the convergence time a bit more, with a tendency of smaller numbers of tasks taking longer time to converge. This behaviour is exemplified by Figure 4(d) vs 4(e) and Figure 4(g) vs 4(h). This is a somewhat surprising behaviour, as one would expect instances with more tasks to be more difficult and the algorithm to take longer to converge. However, as the difference between the final and optimal solution quality from 10 to 20 and 30 tasks in benchmark 2 (Table VII(b)) does not decrease monotonically, it is possible that the difference in difficulty of these instances is due to random factors involved in the instance generation. Note that instances with different numbers of tasks have different randomly generated TPGs, which may considerably change the difficulty of the problem.

In summary, this section shows that Pop-EA managed to quickly find feasible solutions; it presented generally lower convergence time than the (1+1) EA without normalisation and the GA; and its convergence time was not much affected by features such as numbers of employees, employees’ skills and project skills.

VIII. Discussion

Our theoretical and empirical results have confirmed that normalisation is very effective in avoiding overwork, and that EAs with normalisation quickly and consistently evolve schedules of good quality. However, one question remains open: is normalisation better than repair? More precisely, is the completion time with normalisation always better than with repair?

It is tempting to think that the answer to the latter is yes. After all, repair divides all tasks by the maximum overwork, across the whole schedule, whereas normalisation only divides dedications where necessary. Intuitively, since normalisation divides dedications by less, normalised schedules should be no longer than repaired ones. If \( x \) is a feasible schedule, \( \text{time}(\text{repair}(x)) \) denotes the completion time of the repaired schedule \( x \), and \( \text{time}(\text{normalise}(x)) \) denotes the completion time of the normalised schedule \( x \), we might think that for all instances and all feasible \( x \)

\[
\text{time}(\text{normalise}(x)) \leq \text{time}(\text{repair}(x)).
\]

However, it turns out that this is not true for all schedules. Figure 5 gives an example where, for a particular configuration of dedication values, repair leads to a shorter completion time than normalisation. In this example repair doubles all execution times. Compared to this schedule, normalisation allows \( e_1 \) to finish his/her job earlier. This means that, because of the precedence constraints, \( e_2 \) starts to work on three tasks simultaneously. The bottom two tasks \( e_2 \) is working on therefore finish later, and this postpones the time \( e_3 \) is allowed to start on her/his tasks. Overall, the completion time with repair is 8, and the completion time with normalisation is 9.
Fig. 4. Convergence plots.

Fig. 5. An example of a schedule where repair leads to shorter completion times than normalisation. Dedication values of three employees $e_1, e_2, e_3$ are shown; all dedications not listed are 0. Left: an infeasible schedule with overwork. Middle: the same schedule after repair. Right: the same schedule with normalisation.
Theorem 12: There exists a PSP instance and a feasible solution \( x \) such that
\[
\text{time(normalise}(x)) > \text{time(repair}(x))
\]
So we cannot claim that normalisation is always better than repair, with regard to the completion time only.

However, what can be said is that the optimal completion time with normalisation is never more than the optimal completion time with repair. This is because the all-ones dedication matrix has every employee always working full time, leading to a minimal completion time (cf. Theorem 2).

Theorem 13: For every solvable PSP instance the shortest completion time of any feasible schedule with normalisation is no larger than the shortest completion time of any feasible schedule with repair. Formally, if \( X \) is the set of all feasible schedules,
\[
\min_{x \in X} \{\text{time(normalise}(x))\} \leq \min_{x \in X} \{\text{time(repair}(x))\}
\]
The disadvantage of using normalisation is that schedules become more complex as the dedication for particular jobs may change as other jobs are started or finished.

IX. CONCLUSIONS

Research in search-based software engineering has primarily been experimental so far. Few theoretical studies exist. This paper makes a concerted effort, through both theoretical analysis and experimental studies, in understanding more deeply why and when an EA works (or fails) on some PSP problems. The theoretical analysis has inspired new designs of EAs, which have been shown to perform better than existing work.

The new design includes normalisation of dedication values, a tailored mutation operator, and fitness functions with a strong gradient towards feasible solutions. Normalisation removes the problem of overwork and allows an EA to focus on the solution quality. It facilitates finding the right balance between dedication values for different tasks and allows employees to adapt their workload whenever other tasks are started or finished. This is an advantage over the repair mechanism [7] which decreases dedications uniformly for all tasks and at all times.

We have derived general upper and lower bounds for the expected optimisation time of broad classes of EAs. More specific results were presented for RLS and the (1+1) EA, including a general upper bound on the time a feasible solution is reached. For linear schedules both the (1+1) EA and RLS are effective. However, despite using normalisation they still struggle to escape from local optima where many dedication values form an equilibrium.

Our empirical study is based on comparisons among three algorithms with normalisation ((1+1) EA, RLS and Pop-EA) and two algorithms without normalisation ((1+1) EA norm and a state-of-the-art GA). Our analyses confirm that normalisation is very effective in improving solutions in terms of several criteria (hit rate, solution quality and convergence time) that may influence a software manager’s decision in adopting a certain algorithm. Our results show that algorithms with normalisation should be favored over algorithms without normalisation when considering these criteria, and that normalisation makes the EAs more robust to different problem instances. The practical effect of the differences in solution quality between algorithms with and without normalisation is high, whereas the differences among the algorithms with normalisation are small, with the Pop-EA achieving the best results.

Future work includes experimental analysis of the runtime or generation-to-success distributions [43], [44], empirical comparison against the repair operator, case studies using real-world software projects, and investigation of different problem formulations (e.g., when different combinations of dedications affect the productivity). The solution quality in single-objective EAs such as the ones investigated in this work is inherent to the fitness, which considers the relative importance between cost and completion time based on their weights. We shall propose the use of other weights for the fitness function and the extension of our design to multi-objective formulations of the problem.

Acknowledgments:: The authors thank Enrique Alba for providing instances from [6], Francisco Chicano for pointing us to [7], Rami Bahsoon for the fruitful discussions, and the Associate Editor and anonymous reviewers for their constructive comments. This research was supported by EPSRC grant EP/J017515/1. Xin Yao was supported by a Royal Society Wolfson Research Merit Award.

APPENDIX

This appendix contains proofs omitted from the main part.

Proof of Lemma 1: If \( y \) has missing skills then the claim is obvious as then both cost and completion times are proportional to the number of missing skills, reqsk.

Now assume that \( y \) has no missing skills and \( x \) has reqsk missing skills. The completion time for any feasible schedule is bounded by \( k \sum_{j=1}^{m} \text{eff} j \). This corresponds to a schedule where all tasks are processed sequentially and for each task there is only one employee working on it, with the minimum dedication of \( 1/k \). Paying all employees their full salary during the time needed for one employee to work full time yields an upper bound of \( \sum_{i=1}^{n} \sum_{j=1}^{m} s_i \text{eff} j \) on the costs. Those upper bounds for completion times and costs are lower than all penalties for infeasible solutions, hence we have
\[
\text{time}(y) \leq k \sum_{j=1}^{m} \text{eff} j \leq \text{reqsk} \cdot 2k \sum_{j=1}^{m} \text{eff} j = \text{time}(x)
\]
as well as
\[
\text{cost}(y) \leq \sum_{i=1}^{n} \sum_{j=1}^{m} s_i \text{eff} j \leq \text{reqsk} \cdot 2 \sum_{i=1}^{n} \sum_{j=1}^{m} s_i \text{eff} j = \text{cost}(x).
\]

Proof of Theorem 3: Let \( \mu \) be the number of solutions in the initial population. For very large populations, \( \mu > knm \log(nm) \), we argue as follows: the probability that a uniform random solution hits the target point is \((k + 1)^{−nm}\). Hence, the probability that one of the first \( knm \log(nm) \) constructed solutions hits the target point is at most \( knm \log(nm) \cdot (k + 1)^{−nm} \). Hence, with probability at...
least $\Omega(1)$ this does not happen, and this gives a lower bound of $\Omega(1) \cdot knm \log(nm)$ on the expectation as claimed. In the following, we assume $\mu \leq knm \log(nm)$.

Call an entry of the dedication matrix bad if it disagrees with the global optimum. Observe that in order to find the optimum it is necessary to change all bad entries at least once in a mutation. For an initial solution created uniformly at random, each entry is bad at initialisation with probability $k^{-1}(k + 1)$. The expected number of bad initial entries is $knm/(k+1)$. By classical Chernoff bounds [45], with probability $1 - e^{-\Omega(nm)}$ the initial number of bad entries is at least $\frac{nm}{3}$. The probability that all of these will have at least one mutation is $(1 - 1/(knm))^t$. The probability that all of these will have at least one mutation is at least $\frac{nm}{3}$.

Since both local and global mutations treat all matrix entries in the same fashion, we can assume without loss of generality that all initial solutions share the same $\frac{nm}{3}$ entries in the matrix which are bad. Then a necessary condition for finding the target is that each of these entries is turned good at least once in at least one individual.

Let $t := (knm - 1)\ln(nm/3)$. The probability of not changing a particular entry in $t$ mutations is $(1 - 1/(knm))^t$. The probability that there is a bad entry which is never turned good during $t$ mutations is at least

$$1 - \left(1 - \left(1 - \frac{1}{knm}\right)^t\right)^{nm/3} \geq 1 - \left(1 - e^{-\ln(nm/3)}\right)^{nm/3} \geq 1 - \left(1 - \frac{3}{nm}\right)^{nm/3} \geq 1 - e^{-1},$$

where in the last step we have used $(1 - 1/x)^x \leq e^{-1}$ for $x \geq 1$. Using the union bound for the initialisation, with probability at least $1 - e^{-\Omega(nm)} = \Omega(1)$ the EA has not found an optimum after $t = \Omega(knm \cdot \ln(nm))$ steps. This establishes the bound $\Omega(1) \cdot t = \Omega(knm \cdot \log(nm))$.

**Proof of Theorem 4:** With every constructed solution, global mutation is applied to a parent or an intermediate solution resulting from recombination or other operators. Call this solution $x$ and fix a global optimum $x^*$. If $x$ and $x^*$ differ in $\ell$ dedication entries, the probability of global mutation turning $x$ into $x^*$ is exactly

$$\left(\frac{1}{knm}\right)^\ell \cdot \left(1 - \frac{1}{nm}\right)^{nm-\ell} \geq \left(\frac{1}{knm}\right)^{nm},$$

as the probability of mutating a wrong entry of the dedication matrix and selecting the right one is $1/(nm) \cdot 1/k$. The expected waiting time for an event that happens with probability at least $(knm)^{-nm}$ in each time step is at most the inverse probability, $(knm)^{nm}$.

In order to prove Theorem 7, we use the following general result about separable functions (Theorem 12 in [38], partly extended as described in Corollary 13 [38]).

**Theorem 14 (Doerr, Sudholt, Witt [38]):** Let $f = \sum_{i=1}^{m} w_i f_i$ where $w_i \in \mathbb{Z}$ and all $f_i$ are defined on pairwise disjoint sets of variables. For each $1 \leq i \leq n$, let $f_i$ fulfill the following assumptions. There are $\ell_i \in \mathbb{N}$ and $0 = a_{i0} < a_{i1} < \ldots < a_{i\ell_i} = \max \{f_i(x) \mid x \in S\}$. For all $1 \leq j \leq \ell_i$, let $a_{ij} := a_{i\ell_i} - a_{ij-1}$. Assume that $a_{ij} \geq a_{ij+1}$ for all $1 \leq j \leq \ell_i - 1$. Let $A_{ij} := \{x \in S \mid a_{ij-1} \leq f_i(x) < a_{ij}\}$ for all $1 \leq j \leq \ell_i + 1$, where $a_{\ell_i+1} := \infty$. Assume that for all $1 \leq j \leq \ell_i$ and all $x \in A_{ij}$, the $(1+1)$ EA optimising $f_i$ with current search point equal to $x$ with probability $p$ in one iteration finds a search point $y$ with $f_i(y) \geq f_i(x) + a_{ij}$.

Then the expected optimisation time of the $(1+1)$ EA on $f$ is at most $e(1/p) \sum_{i=1}^{m} \ell_i$.

**Proof of Theorem 7:** The time for reaching feasibility is smaller than the time claimed, according to Theorem 5, so we only need to consider the expected time after a feasible schedule has been found.

For linear schedules with a feasible solution $x$ the completion time is a separable function in a sense defined above:

$$\text{time}(x) = \sum_{i=1}^{m} \text{eff}_i \cdot \text{time}_i(x),$$

where $\text{time}_i(x) = 1/\sum_{j=1}^{n} x_{ij}$ is the unweighted completion time for the $i$-th task. Note that the optimal value of $\text{time}_i(x)$ is $1/n$, when all employees have dedication 1 on task $i$. Since Theorem 14 is defined for maximisation with the minimum fitness value for each subfunction being at 0, an equivalent problem is maximising $f(x) = \sum_{i=1}^{m} \text{eff}_i \cdot f_i(x)$ where $f_i(x) = k - \text{time}_i(x)$. Note that all $f_i$ concern mutually disjoint sets of dedication values, and $f_i$ attains its minimum at 0, when only one employee has a positive dedication of $1/k$.

In the notation of Theorem 14 for each $1 \leq i \leq m$ we have $\ell_i := nk$ different $f_i$-values $a_{i0} < a_{i1} < \ldots < a_{i(nk-1)}$ with $a_{ij} := k - \frac{1}{(j+1)/k} = \frac{k \cdot j}{j + 1}$.

The differences between subsequent $a_{ij}$-values are decreasing: for $d_{ij} := a_{ij} - a_{ij-1}$ and all $1 \leq j < n$ we have

$$d_{ij} = k \left(\frac{j}{j+1} - \frac{j-1}{j}\right) = k \frac{\frac{1}{j+1}}{\frac{j}{j+1} + 1} = d_{ij+1}.$$

Now, if the $(1+1)$ EA was optimising $f_i$ as fitness function, it could increase the current $f_i$-value by increasing the dedication of any employee that does not yet have dedication 1. This happens with probability at least $1/(knm) \cdot (1 - 1/(nm))^{nm-1} \geq 1/(eknm)$. By Theorem 14, the expected optimisation time of the $(1+1)$ EA on $f$ is therefore at most $e \cdot (eknm) \cdot \sum_{i=1}^{m} nk = \Omega((knm)^2)$.

For the proof of Theorem 9 we use the following theorem on separable pseudo-Boolean functions, i.e., a separable function defined on the space $\{0, 1\}^n$. This is the case here as we assume $k = 1$.

The following theorem states that, roughly speaking, under certain conditions the $(1+1)$ EA can optimise subfunctions of a separable function in parallel, with a mild overhead of a logarithmic factor.

**Theorem 15 (Doerr, Sudholt, Witt [38]):** Let $f = \sum_{i=1}^{k} w_i f_i$ where $w_i \in \mathbb{Z}$ and each $f_i : \{0, 1\}^\ell \to \{0, 1, \ldots, r\}$ attains
implies $f_{2})$ time until the completion time for this task is minimised, reflecting the inverse of the total dedication of all employees. If the (1+1) EA starts with all dedication values set to 1, we cannot directly apply Theorem 15 since $w_i$ is written as a function of the time. This amounts to a total idle time of at least $3m/2$. As the total workload that needs to be completed is $3m^2$, the completion time is at least $3m^2 + 3m^2 > 3m^2 + m$. This proves the claim for the modified and the original instance and the proof is complete.

**References**


