A Robust and Scalable Approach to Meet User Preferences in Research Project Planning

Roger Xavier Lera-Leri [a,*] and Filippo Bistaffa[a]

[a]Artificial Intelligence Research Institute (IIIA-CSIC), Barcelona, Spain

Abstract. *Research Project Planning* (RPP) is a central task routinely tackled by research institutions, which aims at planning the dedication of a set of researchers involved in a set of research projects, with the goal of meeting the preferences of each researcher as much as possible while efficiently utilizing the available budget. Despite its importance, in real-world institutions, RPP is solved manually due to the lack of automated solutions. To overcome this limitation, we put forward a flexible and scalable approach to provide robust project plans to users (e.g., the administrative staff of a research institution) by modeling the RPP as a constrained *norm approximation* problem, hence enabling the use of modern off-the-shelf optimization solvers. Results on real-world data provided by our research institution show that our approach can compute optimal plans that reflect the preferences of users (i.e., meeting the preferred dedication of researchers and efficiently spending the available budget) in a matter of seconds. Furthermore, we show that our approach can compute plans for thousands of projects and researchers within minutes, hence being able to solve RPP problem instances much larger than the ones typically encountered in an average-sized institution.

1 Introduction

Project planning is a notoriously complex endeavor that has received significant attention in multiple scientific fields [8] due to the involvement of multifaceted and challenging problems, including but not limited to scheduling, task assignment, prediction of future events, formation of well-performing teams, coordination, preference elicitation, and preference satisfaction. Along these lines, providing recommendations to guide the planning and management of early-stage projects is proven to be crucial to project success [11, 10].

Planning *academic research projects* also presents several challenging aspects. The work of researchers involved in each project must be accurately organized to ensure that the dedication required by each *work package* (i.e., a group of related tasks within the project) is satisfied. Nonetheless, such dedication is also subject to constraints of both legal (e.g., researchers cannot be assigned more than a well-defined amount of working hours) and personal nature (e.g., researchers have preferences over the distribution of their effort among the projects). Additionally, budget availability is a crucial aspect that determines the number of person-hours that can be devoted to each project, also taking into account the experience and the professional degree of each researcher. Finally, projects change over time (e.g., new projects can be approved, etc.). Therefore, accounting for uncertainty is crucial when planning real-world research projects.

* Corresponding Author. Email: rlera@iiia.csic.es.

Perhaps unsurprisingly, to the best of our knowledge, research institutions still tackle project planning by manually carrying out the above-mentioned tasks *due to the lack of automated solutions*. Against this background, in this paper, we formally define the *Research Project Planning* (RPP) problem and we put forward an automated solution that computes the optimal dedication for researchers involved in research projects and meets the preferred dedication of researchers as much as possible, while efficiently utilizing the available budget. To achieve this objective, we propose a novel *ℓ_2-norm approximation* [4] model that enables the use of off-the-shelf solvers to compute optimal solutions in a matter of seconds. Our solutions are also robust against the uncertainty inherent in project planning thanks to the proposed use of *Stochastic Robust Approximation* [4]. Our approach has been implemented as part of a planning prototype (Figure 4) that is currently being tested in our research institution, hence we believe our contributions could pave the way to a practical solution for research project planning in the real-world.

More specifically, our contributions are the following ones:

- We model RPP as an ℓ_2-norm approximation problem, discussing both a hard and a soft-constrained model. Our model allows the user to directly express preferences (i.e., the importance of meeting the specified target dedication versus the importance of efficiently spending the available budget) as model parameters.
- We evaluate our approach on real data provided by our institution, illustrating the impact of the parameters that reflect the users' preferences on the computed solutions.
- Results also show that our approach is capable of accounting for uncertainty within real-world projects by balancing the dedication of researchers in case of unexpected changes in projects' conditions. Moreover, our approach can compute optimal solutions for thousands of projects and researchers within minutes.

2 Background

We now discuss some background concepts used in the formalization of RPP. Specifically, we briefly introduce the concept of *Convex Optimization* and we then discuss a very well-known convex optimization problem part of our formalization of RPP, i.e., *Least Squares*. Finally, we introduce the *Stochastic Robust Approximation*, which allows us to deal with uncertainty within our approach.

2.1 Convex Optimization

A convex function f is a function such that

$$f\big(\alpha x + (1-\alpha)y\big) \leq \alpha f(x) + (1-\alpha) f(y), \tag{1}$$

where $0 \leq \alpha \leq 1$ and $\forall\, x, y \in \mathbb{R}^n$. A *convex optimization problem in standard form* [4, Section 4.2.1] is one of the form

$$\begin{aligned} &\text{minimize} && f_0(x), && \\ &\text{subject to} && f_i(x) \leq 0, && i = 1, 2, \ldots, m \quad (2) \\ & && a_i^T x - b_i = 0, && i = 1, 2, \ldots, p \quad (3) \end{aligned}$$

where $f_0, \ldots, f_m$ are convex functions, (2) are inequality constraints and (3) are linear equality constraints. Convex optimization problems are very appealing from a computational perspective because of their fundamental property that ensures that any locally optimal point is also (globally) optimal [4].

A well-known example of convex optimization problem is *Least Squares* (LS),[1] whose objective function is of the form

$$||Ax - b||_2^2 = x^T A^T A x - 2b^T Ax + b^T b, \tag{4}$$

where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Specifically, the *unconstrained* version of the problem has a well-known analytical solution $x = (A^T A)^{-1} A^T b$, where A^T is the transpose matrix of A. On the other hand, there is no analytical solution for the *constrained* version of the problem (i.e., with equality or inequality constraints). However, numerical methods can solve it very efficiently.

More in general, a wealth of commercial and non-commercial off-the-shelf solvers are available for solving a wide range of types of convex optimization problems, such as *Linear Programming* (LP), *Quadratic Programming* (QP), *Mixed-Integer Linear Programming* (MILP) or *Mixed-Integer Quadratic Programming* (MIQP), by implementing state-of-the-art algorithms based on the simplex method or interior-point methods [4]. In this paper, we employ CPLEX.

2.2 Stochastic Robust Approximation

Let us consider an approximation problem such as LS (4). Matrix A and vector b contain the input data of our problem. We usually assume this data as certain, but in some cases, it is desirable to take into account some uncertainty or a possible variation in the data [4]. We assume that $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ are random variables. Hence, it is reasonable to minimize the expected value of the norm such that

$$\text{minimize } \mathbf{E}||Ax - b||_2^2. \tag{5}$$

This problem is the so-called *Stochastic Robust Approximation* (SRA) problem [4]. It is usually not tractable since in most cases it is very difficult to evaluate the objective or its derivatives. However, Ben-Tal and Nemirovski [2] study the cases where the convex SRA is tractable, such as LP, QP, and others. A case where the SRA problem in (5) can be solved occurs when the input data assumes only a finite number of values, i.e.,

$$\mathbf{prob}(A = A_i, b = b_i) = p_i, \quad i = 1, \ldots, k,$$

where $A_i \in \mathbb{R}^{m \times n}$, $b_i \in \mathbb{R}^m$, $1^T p = 1$, $p \geq 0$. In this case, (5) is

$$\text{minimise } p_1 ||A_1 x - b||_2^2 + \cdots + p_k ||A_k x - b||_2^2,$$

where the variables are $x \in \mathbb{R}^n$. Mulvey et al. [9] propose robust models for convex-constrained optimization problems, where the constraints have to be satisfied for each scenario $k \in \Omega$. Taking as an example the *convex optimization problem in standard form*, the robust optimization model is defined as

$$\begin{aligned} &\text{minimize} && \sum_{k \in \Omega} p_k \cdot f_{0,k}(x), && \\ &\text{subject to} && f_{i,k}(x) \leq 0, && i = 1, 2, \ldots, m \quad \text{for all } k \in \Omega \\ & && a_{i,k}^T x - b_{i,k} = 0, && i = 1, 2, \ldots, p \quad \text{for all } k \in \Omega. \quad (6) \end{aligned}$$

3 Problem Definition

In this section, we formally define the RPP problem. Intuitively, we tackle the problem of planning the dedication of a set of researchers who are involved in a set of research projects, to meet the preferred dedication of each researcher as much as possible and efficiently utilizing the available budget, while satisfying the constraints inherent in research project planning.

3.1 Basic Definitions

We consider a *planning horizon* $M = (a, \ldots, b)$ as a sequence of months. We assume that M is sufficiently large to cover the duration of all the projects.

We consider a set of researchers R, a function $€ : R \to \mathbb{R}^+$ that provides the time slot (i.e., person-hour) cost for each researcher (e.g., based on their experience within the institution) and a function $\tau : R \times M \to \mathbb{N}$ that provides the maximum number of working time slots per month for each researcher.

We consider a set of work packages $W \subseteq M \times M$, i.e., $w = (w_s, w_e) \in W$ is a couple representing the start month w_s and end month w_e of work package w. We consider a function $involved : R \times W \to \{\top, \bot\}$ that expresses whether or not a researcher is involved in a work package and a vector $d \in \mathbb{R}^{|W|}$, such that d_w measures the minimum total dedication (number of time slots) required by each work package. We consider a set of projects P as a *partition* of W, i.e., each $p \in P$ is a set of work packages and each $w \in W$ belongs to one project.

To formally pose our optimization problem, we also consider a set $X_W \subseteq R \times W \times M$ of tuples such that $(r, w, m) \in X_W$ if $involved(r, w) = \top$ and $m \in [w_s, w_e]$. We also define the set $X_P \subseteq R \times P \times M$ in the corresponding way. Finally, we consider a vector $x \in \mathbb{R}^{|X_W|}$ of $|X_W|$ real decision variables each associated to a tuple in X_W, such that $x_{r,w,m} \in [0, \tau(r, m)]$. Intuitively, $x_{r,w,m}$ represents the number of time slots that the researcher $r \in R$ devotes to work package $w \in W$ during the month $m \in M$. We assume that a one-time slot contributes the same amount of research work independently from the researcher, following the standard practice of European projects' proposals.

3.2 Optimization Objective

The main objective of our approach is to compute the optimal value of each decision variable of X_W to achieve the following objectives.

Firstly, we aim at meeting a given *target dedication* associated with each researcher $r \in R$ for each month and each project he/she is involved in. Formally, we consider a vector $t \in \mathbb{R}^{|X_P|}$ of $|X_P|$ values such that $t_{r,p,m}$ represents the target dedication of researcher $r \in R$ to project $p \in P$ during month $m \in M$. This target dedication may derive from the fact that researchers prefer certain workloads during specific periods of the year, or because the research institution might want to assign appropriate workloads to researchers to reflect their roles (e.g., principal investigator, contracted researcher).

[1] According to Boyd and Vandenberghe [4], "it is well-known that *Linear Programming* and *Least Squares* have a fairly complete theory, arise in a variety of applications, and can be solved numerically very efficiently".

Secondly, assuming that each project $p \in P$ is associated with a *budget* b_p (where $b \in \mathbb{R}^{|P|}$), our approach aims at maximizing the expenditure for contracts of the researchers involved in p without exceeding b_p. This second criterion is motivated by the fact that a research project (when granted) is allocated a given budget that usually has to be entirely spent during the duration of such a project. Henceforth, it is reasonable to assume that research institutions aim to efficiently use the budget of the project as much as possible. Notice that, for simplicity, here we do not consider other expenses such as equipment, travel costs, etc. Nonetheless, our approach can be easily adapted by defining b_p as the portion of the total budget that has been allocated to the personnel.

3.3 Hard Constraints

We now specify some constraints that will allow us to reflect the restrictions inherent in real-world scenarios.

1. The total cost of the researchers working in a given project $p \in P$ does not exceed b_p. We achieve this by defining the matrix $B \in \mathbb{R}^{|P| \times |X_W|}$ where each row corresponds to a project $p \in P$ and each column to a tuple $(r, w, m) \in X_W$, such that

$$B_{p,(r,w,m)} = \begin{cases} €(r), & \text{if } w \in p, \\ 0, & \text{otherwise,} \end{cases} \tag{7}$$

and by imposing the linear constraint $Bx \leq b$.

2. The required dedication for each work package is satisfied. We achieve this by defining the matrix $D \in \{0, 1\}^{|W| \times |X_W|}$ where each row corresponds to a work package $w \in W$ and each column to a tuple $(r, w, m) \in X_W$, such that

$$D_{w,(r,w',m)} = \begin{cases} 1, & \text{if } w = w', \\ 0, & \text{otherwise,} \end{cases} \tag{8}$$

and by imposing the linear constraint $Dx \geq d$.

3. The total monthly dedication of each researcher r in a given month m is at most $\tau(r, m)$. We achieve this by defining the matrix $T \in \{0, 1\}^{(|R| \cdot |M|) \times |X_W|}$ where each row corresponds to a researcher $r' \in R$ and a month $m' \in M$ and each column to a tuple $(r, w, m) \in X_W$, such that

$$T_{(r',m'),(r,w,m)} = \begin{cases} 1, & \text{if } r' = r, m' = m, \\ 0, & \text{otherwise,} \end{cases} \tag{9}$$

and by imposing the linear constraint $Tx \leq \tau$. Here we consider τ as a vector containing all the $|R| \cdot |M|$ values of the function τ defined in Section 3.

We remark that we do not aim to specify an exhaustive list of constraints that might apply during the research planning phase of every possible research institution. Nonetheless, our framework allows one to implement restrictions that might apply in specific scenarios.[2]

4 Solving the RPP Problem

Having defined the objective and the constraints of our optimization approach, we are now ready to formally define RPP as a constrained *norm approximation problem* [4].

Thus, we define the matrix $A \in \{0, 1\}^{|X_P| \times |X_W|}$ with $|X_P|$ rows (each corresponding to a tuple $(r, p, m) \in X_P$) and $|X_W|$ columns (each corresponding to a tuple $(r', w, m') \in X_W$), such that

$$A_{(r,p,m),(r',w,m')} = \begin{cases} 1, & \text{if } w \in p, r = r', m = m', \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

By using (7), (8), (9), and (10), we formally define the optimization problem with the objectives discussed in Section 3.2 subject to the constraints discussed in Section 3.3 as

$$\begin{aligned} &\text{minimize} \;\; \|Ax - t\|_2^2 + \|Bx - b\|_2^2, \\ &\text{subject to} \;\; Bx \leq b, Dx \geq d, Tx \leq \tau. \end{aligned} \tag{11}$$

The objective function of (11) involves two components that correspond to the two objectives discussed in Section 3.2. Specifically, $\|Ax - t\|_2^2$ denotes the deviation between the computed dedication Ax and the target dedication t. Similarly, $\|Bx - b\|_2^2$ denotes the deviation in terms of budget utilization. In this formulation, Bx cannot exceed b since we also impose the linear constraint $Bx \leq b$. We discuss a formulation without this constraint in the next section.

We also remark that in (11) we directly show our norm approximation formalization as one that employs the ℓ_2 norm, i.e., the p-norm $\|x\|_p = \left(\sum_i |x_i|^p\right)^{1/p}$ for $p = 2$.[3] To be more precise, following a common practice in the optimization and machine-learning literature, we consider the square of the ℓ_2 norm, hence obtaining a constrained LS problem [4]. It is very well known that LS problems can be solved very efficiently by employing a wealth of very mature off-the-shelf solvers, further motivating our formalization of choice.

4.1 Introducing Soft Budget Constraints

In our current formalization of the RPP problem, the budget b_p of a project p cannot be exceeded in any way, as imposed by Constraint 1. However, in real-world research planning scenarios, it could be useful to relax such a constraint, thus making it a *soft constraint*. For example, funding entities could allow for small deviations from the predetermined budget via a proper justification. Furthermore, research institutions might have some funds reserved to integrate small extra expenses that cannot be covered by the budget of the project. Nonetheless, we desire to be able to control the positive values in the vector $Bx - b$, which should be penalized differently with respect to the elements with a negative value. Along these lines, we define a “soft” version of (11) in which we do not consider Constraint 1 and we assign different weights to the positive and negative components of the elements of the vector $Bx - b$. For the sake of coherence, we also assign a weight to the $\|Ax - t\|_2^2$ term, obtaining

$$\begin{aligned} \text{minimize} \;\; & \alpha \cdot \|Ax - t\|_2^2 + \beta \cdot \|\min(0, Bx - b)\|_2^2 + \\ & \gamma \cdot \|\max(0, Bx - b)\|_2^2, \\ \text{subject to} \;\; & Dx \geq d, \quad Tx \leq \tau, \end{aligned} \tag{12}$$

where the parameters $\alpha, \beta, \gamma \in \mathbb{R}^+$ are the weights and the $\min$ (resp. $\max$) function represents the *element-wise* minimum (resp. maximum) between two vectors. It should be noted that the terms $\|\min(0, Bx - b)\|_2^2$ and $\|\max(0,\; Bx - b)\|_2^2$, although of very similar form and both subject to minimization as part of the objective function, are of very different nature. On the one hand,

[2] As a matter of fact, our prototype mentioned in Section 7 includes minor additional constraints that are specific to our research institution, which we decided to leave out from the formalization discussed here for simplicity.

[3] Other values of p could also be considered, resulting in different types of norm approximation problems [4]. An in-depth analysis of the impact of the parameter p on our approach is beyond the scope of this paper.

$\|\min(0, Bx - b)\|_2^2$ is being minimized as it reflects our second optimization criterion discussed in Section 3.2. On the other hand, $\|\max(0, Bx - b)\|_2^2$ is being minimized as it is effectively implementing a soft constraint taking the place of Constraint 1. Along these lines, we can characterize the semantics of the parameters α, β, and γ as follows. On the one hand, α represents the importance of meeting the target dedication t, whereas β is the importance of spending the budget b as efficiently as possible. Finally, γ is the importance of exceeding the available budget as little as possible. By setting $\gamma = \infty$ we obtain a formulation equivalent to the hard-constrained one in (11). The rest of this section discusses the computational complexity of (12). Specifically, we show that the soft-constrained problem in (12) is still a constrained LS problem, hence maintaining the same desirable properties of the computational efficiency of (11).

Proposition 4.1. *The optimization problem in* (12) *can be posed as the constrained LS problem*

$$\begin{aligned} \textit{minimize} \quad & \alpha \cdot \|Ax - t\|_2^2 + \beta \cdot \|v\|_2^2 + \gamma \cdot \|u\|_2^2, \\ \textit{subject to} \quad & Dx \geq d,\, Tx \leq \tau, Bx - b \leq u, \\ & Bx - b \geq v, u \geq 0, v \leq 0, \end{aligned} \tag{13}$$

where $u, v \in \mathbb{R}^{|P|}$ *are additional auxiliary variables.*

Proof. The proof is in the Supplementary Material.[4] □

4.2 Introducing Uncertainty with Stochastic Instances

In the above-defined formalization of the problem (13), we assumed that the data is fully known a priori, i.e., matrices A, D, T and B and vectors t, d, τ and b are fixed. However, in real-world research planning scenarios, the data and conditions of the plan can change throughout the planning horizon. For instance, a new project may be assigned to the research institution, new researchers may start working on ongoing projects, the requirements of the project may change, etc. Hence, it is crucial to handle such an uncertainty.

Usually, when an unexpected change in the middle of the duration of a research project happens, a straightforward strategy is to recompute the planning for the rest of the project. To implement such a strategy, we can compute the planning for the following months considering the dedication and the budget already spent on the work packages. Unfortunately, this simple strategy might not work in some situations. For instance, consider the case when new projects are assigned to the research institution. In such cases, researchers' time distribution may adapt to cover the dedication of ongoing and new projects. This may probably cause an abrupt increase in researchers' efforts to satisfy the dedication of the projects in the following months. Furthermore, in more extreme cases, the dedication of the projects may not be possible to satisfy without the involvement of more researchers. Along these lines, we now discuss how we account for uncertainty thanks to a technique called SRA [4].

First, we start by considering a set of possible scenarios $\Omega = \{\omega_0, \omega_1, \ldots, \omega_K\}$, where ω_0 is the scenario with the current data, and $\omega_1, \ldots, \omega_K$ are possible variations of ω_0. A^k, D^k, T^k and B^k for all $k = 0, 1, \ldots, K$ are the matrices for the particular scenarios, using the same notation for the vectors. By using such a general methodology, the new scenario can cope with variations of any kind: additions of new projects, changes in preferences or budget, or the involvement of new researchers are some examples.

[4] https://filippobistaffa.github.io/files/PAIS_2024_Suppl_Material.pdf.

Then, by following the methodology introduced in Section 2.2 [4, 9], we aim to minimize the expected value of the objective function, which represents the weighted mean of the objective functions of each scenario. Thus, we minimize the expected value of (13):

$$\begin{aligned} \text{minimize} \quad & p^T \cdot \big(\alpha \cdot \mathbf{A} + \beta \cdot \mathbf{B} + \gamma \cdot \mathbf{\Gamma}\big), \\ \text{subject to} \quad & \|A^k x - t^k\|_2^2 \leq \mathbf{A}^k,\ \|v^k\|_2^2 \leq \mathbf{B}^k,\ \|u^k\|_2^2 \leq \mathbf{\Gamma}^k, \\ & B^k x - b^k \leq u^k,\ B^k x - b^k \geq v^k, \\ & u^k \geq 0,\ v^k \leq 0, \qquad \forall\, k = 0, 1, \ldots, K, \end{aligned} \tag{14}$$

where $p \in [0,1]^{K+1}$ is a vector with the probability that a given scenario occurs and $\mathbf{A}, \mathbf{B}, \mathbf{\Gamma} \in \mathbb{R}^{+K+1}$ are auxiliary variables.

Now, we need to deal with the constraints of the formalization. Mulvey et al. [9] enforce as hard constraints all constraints in every scenario, i.e., $\forall\ \omega_k \in \Omega$. We choose to enforce only as hard constraints those that belong to the real scenario ω_0. Constraints belonging to the scenario with variations in the data ω_k for $k = 1, \ldots, K$ are encoded as soft constraints, i.e., preferences that should be fulfilled as much as possible. We impose the hard constraints on ω_0 because it is the "real" scenario, while we impose soft constraints on the hypothetical scenarios because imposing hard constraints might abruptly change the dedication distribution, yielding infeasible problems. Hence, we add the following set of constraints to (14):

$$\begin{aligned} & D^k x \geq d^k - l^k,\ T^k x \leq \tau^k + h^k, \quad \forall\, k = 0, 1, \ldots, K, \\ & l^k, h^k \geq 0, \quad \forall\, k = 1, \ldots, K, \qquad l^0, h^0 = 0, \end{aligned} \tag{15}$$

where $l, h \in \mathbb{R}^{+K+1}$ are variables indicating dissatisfaction with the hypothetical scenarios. We set l^0, h^0 to 0, effectively imposing a hard constraint for the "real" scenario. By adding (15) to (14), we obtain:

$$\begin{aligned} \text{minimize} \quad & p^T \cdot \big(\alpha \cdot \mathbf{A} + \beta \cdot \mathbf{B} + \gamma \cdot \mathbf{\Gamma} + \delta \cdot l + \rho \cdot h\big), \\ \text{subject to} \quad & \|A^k x - t^k\|_2^2 \leq \mathbf{A}^k,\ \|v^k\|_2^2 \leq \mathbf{B}^k,\ \|u^k\|_2^2 \leq \mathbf{\Gamma}^k, \\ & B^k x - b^k \leq u^k,\ B^k x - b^k \geq v^k,\ u^k \geq 0,\ v^k \leq 0, \\ & D^k x \geq d^k - l^k,\ T^k x \leq \tau^k + h^k, \quad \forall\, k = 0, 1, \ldots, K, \\ & l^k, h^k \geq 0, \quad \forall\, k = 1, \ldots, K, \qquad l^0, h^0 = 0. \end{aligned} \tag{16}$$

Note that we add the terms $\delta \cdot l$ and $\rho \cdot h$ in the objective function as penalty terms for violating the soft constraints of the hypothetical scenarios. Such terms are weighted with $\delta, \rho \in \mathbb{R}^+$ parameters.

5 Experimental Analysis

The main objective of our experimental evaluation is to assess the impact of the parameters α, β, and γ on the solutions computed by our approach, to guide the final user (e.g., research institutions) while setting these values to obtain the desired behavior.

In our tests, we consider real-world data of the research projects currently active in our research institution. Specifically, from the projects' proposals we collect data about the set of researchers R, the set of projects P, the set of work packages W, and the required dedication d_w for each $w \in W$. Overall, our dataset involves 6 projects and 27 work packages, whose dedication must be allocated during a planning horizon M of 67 months.

We consider two types of researchers for the set R: *permanent researchers*, who are characterized by a higher degree of experience and a higher salary, and *contracted researchers*, whose contract is finite and whose salary is usually lower than senior colleagues.

Our dataset also includes the budget b_p for each project $p \in P$ and the target dedication vector t. As expected, the budget and the target dedication specified in the proposal of each project were determined to be "just enough" to cover the dedication required by each work package, as a result of a manual planning phase. Nonetheless, since we are proposing and evaluating a tool for project planning, we are interested in testing it in situations where the budget is not so well calibrated, in order to measure its impact on the solution computed by our approach (also in relation to the above-mentioned parameters). To this end, in addition to the original budget data, which we employ in the so-called *sufficient budget* scenario, we derive two additional scenarios from our real-world data: the *abundant budget* scenario where the original budget of each project is increased by 5% and the *scarce budget* scenario with a budget reduced by 5% for each project. In all these scenarios we maintain the original target dedication vector t to better assess the impact of the budget variation. Since our approach relies on a LS model that can be solved very efficiently (see Section 4), we do not show results about the runtime, which is negligible (< 1 second) for our data. In the Supplementary Material, we report a scalability evaluation on large-scale artificial instances.

Our implementation[5] uses the CVXPY v1.2.0 library to model the constrained LS models in (13) and (16), and we use the IBM CPLEX v22.1.0.0 solver to solve such models. We run our experiments on a machine with a 12-core 3.70GHz CPU and 64GB of RAM.

5.1 Discussion of the Results

To analyze our solutions quantitatively we will consider the following metrics. First, we define the *budget residual* as the mean percentage deviation between the cost computed by our approach and the budget available for each project. More formally,

$$\text{Budget residual} = \frac{100}{|P|} \sum_{p \in P} \left| \frac{\min((Bx - b)_p, 0)}{b_p} \right| = \frac{100}{|P|} \sum_{p \in P} \left| \frac{v_p}{b_p} \right|.$$

Such a metric is a variation of *Mean Absolute Percentage Error* (MAPE) [1], which is always positive, and in this specific case is bounded within [0,100]. Intuitively, a budget residual close to 0% indicates that we are efficiently using the whole budget. On the other hand, a budget residual close to 100%, indicating a minimal expenditure of the budget, is difficult to achieve in practice, since such a budget is required to guarantee a certain dedication of the researchers.

Similarly, we consider the *budget excess* defined as

$$\text{Budget excess} = \frac{100}{|P|} \sum_{p \in P} \left| \frac{\max((Bx - b)_p, 0)}{b_p} \right| = \frac{100}{|P|} \sum_{p \in P} \left| \frac{u_p}{b_p} \right|.$$

This metric gives information about the mean percentage of the excess between the computed cost and the budget for each project. While the budget excess is always positive, it is not upper-bounded, since there is no constraint bounding the exceeding in the soft-constrained formalization in (13). In the case of the hard-constrained formalization in (11), the budget excess is always 0%. Moreover, we consider the *dedication residual* defined as

$$\text{Dedication residual} = \frac{100}{|W|} \sum_{w \in W} \left| \frac{(Dx - d)_w}{d_w} \right|$$

to represent the relative percentage difference between the dedication corresponding to the computed solution and the required dedication of each work package. A dedication residual close to 0% indicates that the dedication assigned to work packages is strictly the required one. Finally, we define the *target deviation* as the mean percentage deviation from the target dedication. Formally,

$$\text{Target deviation} = \frac{100}{|X_P|} \sum_{i=1}^{|X_P|} \left| \frac{(Ax - t)_i}{\max(t_i) - \min(t_i)} \right|.$$

The target deviation can also be seen as a variation of the MAPE. We divided the difference between the computed and the target dedication for the range of the target instead of the target dedication. This variation avoids big target deviation values when having small differences between the computed distribution and the target dedication for values of $t_i \to 0$. A target deviation close to 0% indicates that the computed plan is close to the researchers' target preferences. Finally, since our dataset includes permanent and contracted researchers, we also measure the *contracted researchers' dedication*, i.e., the percentage of time slots assigned to contracted researchers with respect to the total. Our objective is to assess the impact of α, β, γ in the three budget scenarios (i.e., *scarce*, *sufficient*, and *abundant*). More specifically, in all our experiments we consider $\alpha \in \{0.25, 0.50, 0.75, 1.00\}$ and $\beta \in \{0, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}\}$. We consider different values for α and β since they are associated with components with different units of measurement in our objective function (13), i.e., $\|Ax - t\|_2^2$ and $\|v\|_2^2$, respectively. As for the parameter γ, we consider $\gamma \in \{0, 10^{-5}, \infty\}$ in order to account for an *unconstrained* scenario ($\gamma = 0$), a *soft-constrained* scenario ($\gamma = 10^{-5}$), and a *hard-constrained* scenario ($\gamma = \infty$).

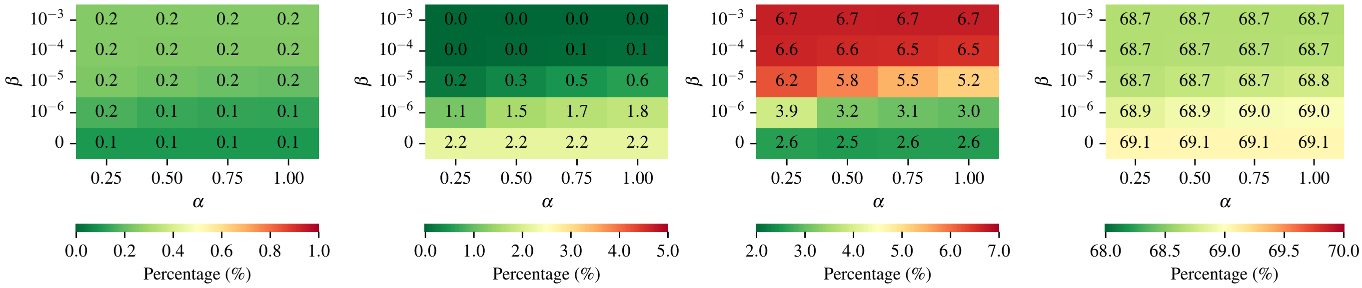
5.1.1 Sufficient Budget Scenario

Figure 1 shows the results for the *sufficient budget* scenario. Overall, we observe that, for the three different values of γ parameter, the solutions computed by our approach do not vary by a significant amount, therefore we only report results for $\gamma = \infty$, which accounts for the *hard-constrained* formulation. This behavior is expected, since in this scenario we consider a budget that has been carefully calibrated to be sufficient for satisfying the dedication required by the work packages (see Section 5). Indeed, in this scenario the resulting target deviation is negligible (the maximum displacement is 0.25%), i.e., our approach meets researchers' dedication preferences quite precisely. Nonetheless, we observe a variation when changing β, i.e., the parameter that controls the importance of efficiently allocating the budget. Indeed, in Figure 1b we notice that when increasing β we achieve a budget residual that progressively approaches 0 (i.e., all the budget is efficiently spent). This is achieved by assigning a slightly larger dedication (upper rows of Figure 1c) of up to 6.7%. Regarding the dedication of contracted researchers, there is not a significant variation for the different configurations of α and β. Finally, we observe that the budget excess is always 0% since the budget is sufficient to satisfy the required dedication in this case.

5.1.2 Abundant Budget Scenario

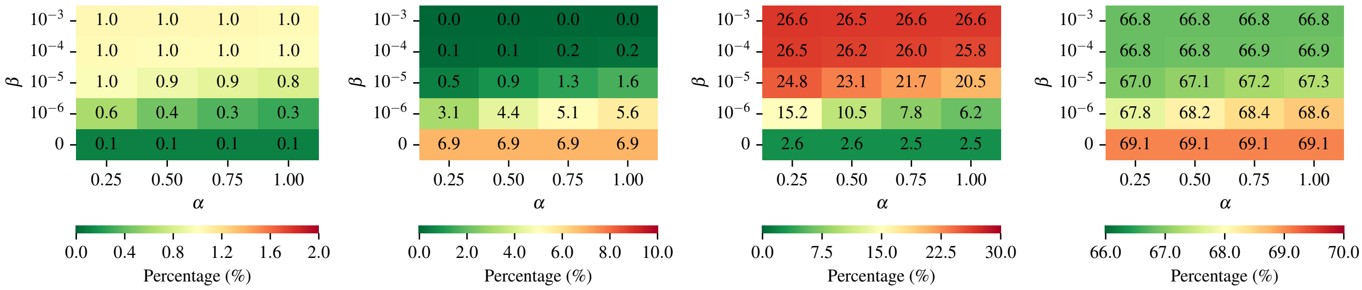
In Figure 2 we report the results for the *abundant budget* scenario. Here we consider a situation in which the budget is not a limiting factor, hence we observe that varying γ (i.e., the parameter that controls the importance of not surpassing the available budget) does not impact the behavior of the algorithm. Hence, similarly to the previous case we only report results for $\gamma = \infty$. On the other hand, in this scenario, we observe a significant impact of the parameters α and

[5] Online at https://github.com/RogerXLera/ResearchProjectPlanning.



(a) Target residual (b) Budget residual (c) Dedication residual (d) Contracted dedication

Figure 1: Results in the *sufficient budget* scenario for $\gamma = \infty$ (best viewed in colors).



(a) Target residual (b) Budget residual (c) Dedication residual (d) Contracted dedication

Figure 2: Results in the *abundant budget* scenario for $\gamma = \infty$ (best viewed in colors).

β. First, we notice that for $\beta = 0$ the approach computes a solution equivalent to the *sufficient budget* scenario. In fact, for such parameter configuration, the target deviation (0.10%) and the contracted dedication (69.1%) are equivalent in both scenarios (bottom row of Figures 2a and 2d). In addition, we notice that the budget residual for $\beta = 0$ varies between the abundant and the sufficient budget scenario (bottom row of Figure 2b). This increase is due to the fact that the available budget (b) has increased by $+5\%$ with respect to the sufficient budget scenario, but the computed cost of the researchers (Bx) is indeed equivalent to the original for $\beta = 0$. This behavior is due to the absence of any incentive to spend all the budget, thus the algorithm aims to meet the target dedication as much as possible (bottom row of Figure 2a). On the other hand, when increasing β the algorithm gives more importance to spending the available budget (upper rows of Figure 2b). This is achieved by assigning a larger overall dedication to the researchers (upper rows of Figure 2c), but also by employing more "expensive" (permanent) researchers, since contracted researchers' dedication decreases when increasing β (Figure 2d). By doing so, we observe that the target deviation also gets affected, although the deviation increases only by 1% at most (Figure 2a). Also, for the intermediate values of β ($\beta = 10^{-4}, 10^{-5}, 10^{-6}$), when we increase α (i.e., the parameter that controls the importance of meeting the target dedication) we obtain a less efficient budget spending but a dedication that closely meets the target one. Again, we do not show the budget excess because it is always 0% due to the same reasons discussed in the sufficient budget scenario.

5.1.3 *Scarce Budget Scenario*

Figure 3 shows the results for the *scarce budget* scenario. We observe that budget scarcity has three important consequences. First, as expected γ has now a significant impact on the results, since the budget is not sufficient to cover the required dedication, and hence, the budget constraint in our formulation has to be violated, i.e., $\|\max(0, Bx - b)\|_2^2 > 0$. Second, since the available budget will always be entirely allocated (budget residual $\approx 0\%$), β loses any impact on the considered measures. Third, in this scenario, we observe that the dedication surplus is negligible for all parameter configurations (since the algorithm will always allocate the minimum amount of dedication given the budget scarcity). Finally, α has a small impact on the metrics defined for the scarce budget scenario. Given the above-mentioned discussion, we only report results for different values of γ disregarding the values of α and β.

For $\gamma = 0$ (unconstrained case), we correctly obtain a behavior equivalent to the *sufficient budget* case, similarly to what we observed in the *abundant* scenario for $\beta = 0$. This is due to the absence of any constraint that limits the spending of the budget, hence the algorithm is free to allocate any amount of money to meet the target dedication as much as possible. More specifically, we obtain the same target deviation (0.10%) and contracted researchers' dedication (69.1%) as in the other scenarios. Again, the budget excess is almost ($\sim 3.5\%$) due to the reduced budget, although the computed cost is the same as in the other scenarios. For $\gamma = 10^{-5}$ (soft-constrained case), the budget excess is reduced down to 1.85%.

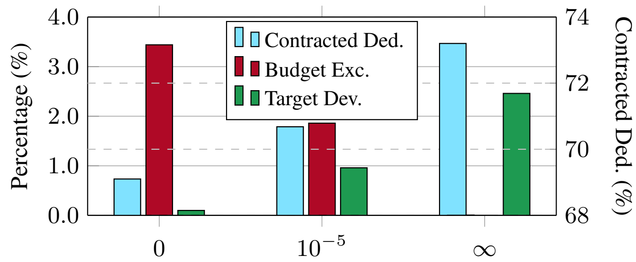


Figure 3: Target deviation, budget excess, and the contracted researchers' dedication, for different γ and $\alpha = 0.25$, $\beta = 0$ in the *scarce budget* scenario (best viewed in colors).

This is achieved by increasing the dedication of contracted researchers, which is reflected in the contracted researchers' dedication and the target deviation by an increase up to 70.7% and 0.95% respectively. Finally, for $\gamma = \infty$ (hard-constrained case), the algorithm can compute a plan without exceeding the budget, by assigning more dedication to contracted researchers (up to 73.2%). In general, our algorithm opts for higher contracted researchers' dedications when forced to allocate a lower budget, when setting a greater γ.

Overall, our results demonstrate that α, β, and γ have a significant impact, depending on the budget availability in relation to the required and target dedication.

5.2 *Experiments with Uncertainty*

We now proceed to evaluate the impact of considering uncertainty on our approach. Following the methodology in Section 4.2, we generate the hypothetical scenarios in the set $\Omega = \{\omega_0, \omega_1, \ldots, \omega_K\}$, i.e., the variations of the initial data ω_0, with $K = 10$. Such hypothetical scenarios represent the possibility of having unexpected events that change the initial projects' conditions. Along these lines, each hypothetical scenario ω_k has the same initial projects as ω_0 plus the hypothetical new projects generated via the same methodology of our scalability evaluation described in the Supplementary Material.

For simplicity, among all possible variations (e.g., the assignment of a new project, the renouncement or the involvement of a new researcher, variations in the project budget or deadline), we consider one of the most common situations in real-world project planning, i.e., the addition of new projects during the planning horizon. The number of additional projects is uniformly chosen between 1 and 3.

To compare with our approach without uncertainty, we run the experiments for the same values of α and β discussed in Section 5.1.

5.2.1 *Summary of the Results with Uncertainty*

For the sake of brevity, we just report a summary of the results for the sufficient budget scenario for $\gamma = \infty$, i.e., the hard-constrained formalization. We report the full heatmap plots corresponding to these results in the Supplementary Material.

Our results show that α and β have a significant impact on the target deviation. Indeed, by comparing the results of the approaches with and without uncertainty, we observe that the maximum value of the target deviation for the uncertainty approach is $12.8 \pm 1.2\%$, while for the formalization without uncertainty is 0.25% (2.5% for the scarce budget scenario). This result is expected for the uncertainty approach, as the algorithm pursues a trade-off between satisfying the preferences of the researchers and saving dedication for the projects of the hypothetical scenarios. Moreover, as we increase α, we reduce the target deviation, and as we increase β, we reduce the budget residual. This is similar to the case without uncertainty, as expected. We also observe a significant reduction in the dedication residual compared to the approach without uncertainty. More specifically, the maximum value of the dedication residual for the approach considering uncertainty is 2.4%, while for the approach not considering uncertainty is 6.7%. Finally, we observe a significant drop in the contracted researchers' dedication compared to the approach without uncertainty (48.2% vs 69.1%). This difference can be explained as follows. As we discussed in Section 5.1, contracted researchers usually execute most of the workload compared to permanent researchers. Hence, to account for the possible addition of new projects, our approach balances the dedication of permanent and contracted researchers. This allows contracted researchers to be involved in new projects by reducing their dedication to the current projects and increasing the dedication of the permanent researchers.

Overall, we conclude that our approach is capable of accounting for uncertainty by balancing the dedication of researchers.

6 Related Work on RCPSP

To the best of our knowledge, the RPP problem has never been studied before in the *planning* literature. Nonetheless, some real-world applications that share similarities with the RPP problem have been studied under the name of *Resource-Constrained Project Scheduling Problem* (RCPSP). RCPSP is a well-known NP-hard problem [3] that consists of activities that must be scheduled to meet precedence and resource constraints to minimize the project's makespan [6, 7].

Despite the apparent similarities, RCPSP and RPP are formally different problems due to their distinct optimization goals. Indeed, the ultimate goal of RPP is *not* the minimization of the makespan, in contrast with RCPSP. Instead, RPP aims at minimizing the deviations between the computed and the desired dedications of researchers, on one hand, and between the utilized budged and the available one, on the other hand (see (11)). As a matter of fact, minimizing the makespan in the case of RPP is impossible, since work packages have fixed start and end dates. These reasons prevent us from employing any of the above-mentioned approaches, thus motivating our proposal of a novel approach for RPP.

7 Conclusions & Future Work

We considered the RPP problem, a central task routinely tackled manually by research institutions when planning the dedication of researchers involved in research projects. To alleviate this burden for the administrative staff, we put forward an automated approach to compute plans that can account for uncertainty and we evaluated it on a real-world dataset provided by our research institution. Results showed that our approach can compute optimal solutions that meet the preferences of the user in real-time, including instances that are larger than the ones managed by an average-sized research institution. Our approach has been implemented as part of a prototype that is being tested in our institution (Figure 4). As future research directions, we plan to test our prototype in other institutions with different datasets. In addition, we plan to provide contrastive explanations to the users to make our solutions easier to understand and trustworthy.

Acknowledgements

The authors were supported by the research projects ACISUD (PID2022-136787NB-I00) and Yoma OR (OPE02570).

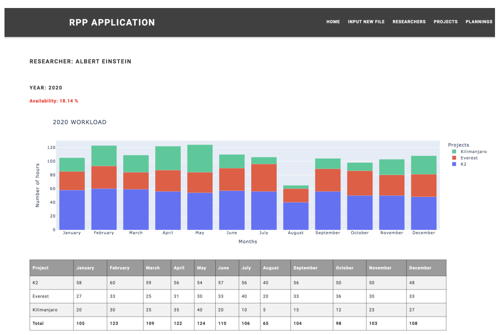


Figure 4: Screenshot of our prototype (best viewed in colors).

References

[1] J. S. Armstrong and L.-R. Forecasting. From crystal ball to computer. *New York ua*, 348, 1985.

[2] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998.

[3] J. Blazewicz, J. K. Lenstra, and A. R. Kan. Scheduling subject to resource constraints: classification and complexity. *Discrete Applied Mathematics*, 5(1):11–24, 1983.

[4] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, U.K., 2004.

[5] A. Drexl, R. Nissen, J. H. Patterson, and F. Salewski. Progen/πx–an instance generator for resource-constrained project scheduling problems with partially renewable resources and further extensions. *European Journal of Operational Research*, 125(1):59–72, 2000.

[6] S. Hartmann and D. Briskorn. A survey of variants and extensions of the resource-constrained project scheduling problem. *European Journal of Operational Research*, 207:1–14, 11 2010.

[7] S. Hartmann and D. Briskorn. An updated survey of variants and extensions of the resource-constrained project scheduling problem. *European Journal of Operational Research*, 297(1), 02 2022.

[8] H. Kerzner. *Project Management: a Systems Approach to Planning, Scheduling, and Controlling*. John Wiley & Sons, 2017.

[9] J. M. Mulvey, R. J. Vanderbei, and S. A. Zenios. Robust optimization of large-scale systems. *Operations Research*, 43(2):264–281, 1995.

[10] C. Pérez, L. Climent, G. Nicoló, A. Arbelaez, and M. A. Salido. A hybrid metaheuristic with learning for a real supply chain scheduling problem. *Engineering Applications of Artificial Intelligence*, 126:107188, 2023.

[11] L. Raymond and F. Bergeron. Project management information systems: An empirical study of their impact on project managers and project success. *International Journal of Project Management*, 26(2): 213–220, 2008.