Offre de Stage

- **Titre :** Information discovery in combinatorial optimization under uncertainty
- **Localisation :** Institut de Mathematiques de Bordeaux (IMB), Talence (Bordeaux) or Laboratoire d'Informatique, de Robotique et de Microelectronique de Montpellier (LIRMM) - Montpellier, depending on the student's preference.
- **Thème :** Stochastic optimization
- **Mots clefs :** stochastic optimization, combinatorial optimization, decomposition methods
- **Les noms, prénoms et courriels des encadrants**

  Co-encadrant Ayşe Nur Arslan, ayse-nur.arslan@inria.fr

  Co-encadrent Michael Poss, michael.poss@lirmm.fr

- **Profil recherché :** Second year masters' or 3rd year engineering student in Operations Research or related fields

  Required skills :

  Mixed-integer linear programming

  C++ or Python/Julia (the internship will be done in Julia but no priori knowledge is required)

  Knowledge in one of the following fields is appreciated :

  Optimization under uncertainty approaches

  Decomposition methods in MILP
- The internship may lead to a thesis offer within one of the projects of the Inria-Bordeaux research team EDGE or the MAORE team of LIRMM-Montpellier.

Contexte scientifique

This internship concerns the study of combinatorial optimization problems under uncertainty integrating information discovery. Combinatorial optimization is the study of problems involving only binary variables and includes problems such as the knapsack, traveling salesman, and spanning tree problems. Here, we we would like to study these problems under the presence of uncertainty, that is not all parameters of the problem are known with uncertainty. Further, we would like to incorporate the paradigm of information discovery to these problems, that is, the possibility of querying a subset of uncertain parameters before optimization is performed. We next formalize these concepts.

In this internship, we will particularly be interested in combinatorial optimization problems defined on a graph $G = (V, A)$ where V is the set of vertices and A is the set of arcs. Mathematically, we will write these problems as :

$$\min_{x \in X} \; c^\top x \qquad \text{(Graph-Det)}$$

where set X defines the structure of solutions. This structure may represent a flow, matching, a cycle cover etc. Depending on the structure of the set X, the problem can be polynomially solvable (e.g., min-cost flow problems) or NP-Hard (e.g., coloring).

We will focus our attention on graph problems under uncertainty where the existence of the arcs of the graph is not known with certainty. To this end, assume that each arc exists or does not exist independently and let $\tilde{\xi}_a$ be a random parameter that follows a Bernoulli distribution with probability of non-existence p_a for $a \in A$ governing this uncertainty. Let further $\tilde{\xi} = (\tilde{\xi}_a)_{a \in A}$. With this notation, we may define a function $f(\xi, x)$ which given a realization ξ of the random vector $\tilde{\xi}$ and a solution $x \in X$, returns the cost of this solution. We now define an optimization problem where the expected value of this cost is optimized :

$$\min_{x \in X} \ \mathbb{E}_{\tilde{\xi}}[f(\xi, x)]. \tag{Graph-Stoc}$$

The difficulty of this problem will stem from the evaluation of the expectation term which involves an exponential number of realizations. This difficulty can be circumvented by use of sampling techniques (such as sample average approximation, see [2]).

We now extend (Graph-Stoc) by adding the possibility of information discovery (see [4]). To this end, assume that the decision maker can query a subset of arcs $I \subset A$ (subject to a budget), and discover, as a result of his query, the realization $(\xi_a)_{a \in I}$ of a subset of random parameters $\tilde{\xi}$. He then solves a restriction of (Graph-Det) with this knowledge, that is, integrating the existence/non-existence of arcs in I, and using only those arcs that are in I. The decision maker then wants to determine how to choose the subset I optimally so that he can minimize the expected value of the solution integrating the realizations of parameters in I. Formally, we write, using decision variables $w \in \{0,1\}^{|A|}$ for the selection of arcs to query :

$$\min_{w \in W \subseteq \{0,1\}^{|A|}} \quad \mathbb{E}_{\tilde{\xi}} \left[\min_{x \in X(\xi \circ w)} \quad c^\top x \right] \tag{Graph-Stoc-ID}$$

where the expectation is with respect to the parameters that are chosen to be observed. We remark that the set X is now restricted by the product $\xi \circ w$, in order to integrate the discovered information. This restriction will manifest itself in the form of constraints of type $x \circ w \leq U(\xi \circ w)$ restricting the use of arcs that do not exist based on the discovered information, with parameter U being the upper bound of variable x in (Graph-Det).

One of the interesting applications of the proposed framework is the kidney exchange problem (KEP) where the set of vertices V represents the set of patient-donor pairs, and the set of arcs A represents the compatibility information between the donors and patients of different pairs. The goal of the kidney exchange problem is to find a maximum profit K-cycle packing of G (for given K), where a $K-$cycle is a cycle whose size is smaller than K. The K-cycle packing problem is known to be NP-Hard for $3 \leq K < +\infty$, however, it is well-solved using state-of-the-art decomposition algorithms (see [1]). The uncertainty in the kidney exchange problem stems from the fact that the compatibility information integrated to the graph described above is subject to additional medical tests. It is therefore only by performing these costly medical tests that one can discover whether a given transplant operation can go through or not. On the other hand, for a given donor-patient pair the probability that the test gives a favorable outcome can be estimated. It is therefore appropriate to define an optimization problem in order to find the donor-patient pairs to test so that the expected value of an exchange solution constructed with the tested pairs is maximized. A solution constructed in this way is not subject to failures due to incompatibility since all pairs have been tested prior to optimization, as such it is robust and does not require multiple phases of testing and optimization.

Objectifs

The goal of this internship is to study problems of form (Graph-Stoc-ID) from a methodological and practical perspective. This will involve development of Benders' decomposition schemes (potentially coupled with Dantzig-Wolfe reformulation and column generation algorithms). From the practical perspective, we will explore the effect of information discovery on the quality of constructed solutions. The kidney exchange problem will serve as a inspiration for our developments as its (Graph-Stoc-ID) version has already been studied in the literature (see [3]). Additionally, realistic instances for this problem are available online and are frequently used in the literature.

Références

[1] J. Omer, A. N. Arslan, and F. Yan. KidneyExchange.jl : A Julia package for solving the kidney exchange problem with branch-and-price. Preprint : https ://hal.inria.fr/hal-03830810, Oct. 2022.

[2] A. Shapiro. Monte carlo sampling approach to stochastic programming. In *ESAIM : proceedings*, volume 13, pages 65–73. EDP Sciences, 2003.

[3] B. Smeulders, V. Bartier, Y. Crama, and F. C. Spieksma. Recourse in kidney exchange programs. *INFORMS Journal on Computing*, 34(2) :1191–1206, 2022.

[4] P. Vayanos, D. Kuhn, and B. Rustem. Decision rules for information discovery in multi-stage stochastic programming. In *2011 50th IEEE Conference on Decision and Control and European Control Conference*, pages 7368–7373. IEEE, 2011.