

Haut Conseil de l'Évaluation de la Recherche et  
de l'Enseignement Supérieur



# DOCUMENT D'AUTOÉVALUATION

## Équipe RO



Campagne d'évaluation 2023-2024 — Vague D

## Table des matières

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>INFORMATIONS GÉNÉRALES SUR L'ÉQUIPE RO</b>   | <b>3</b>  |
| 1.1      | Les thématiques scientifiques et leurs enjeux . . . . .   | 3         |
|          | A. Optimisation black-box : algorithmes de recherche randomisés, théorie des algorithmes évolutifs, configuration et sélection d'algorithmes, méthodes AutoML . . . . . | 3         |
|          | B. Algorithmes on-line : learning-augmented algorithms, explorable uncertainty et on-line searching . . . . .   | 4         |
|          | C. Algorithmes en présence de données évolutives : robustesse, optimisation stochastique et optimisation multistage . . . . .   | 5         |
|          | D. Complexité, résolution exacte et approchée : problèmes d'ordonnancement, de clustering, choix social computationnel . . . . .  | 6         |
|          | Animation scientifique . . . . .  | 7         |
| <b>2</b> | <b>INTRODUCTION DU PORTFOLIO</b>  | <b>8</b>  |
| <b>3</b> | <b>AUTOÉVALUATION DU BILAN</b>  | <b>9</b>  |
| 3.1      | Autoévaluation de l'équipe . . . . .  | 9         |
|          | Domaine 2. Attractivité . . . . .   | 9         |
|          | Domaine 3. Production scientifique . . . . .  | 12        |
|          | Domaine 4. Inscription des activités de recherche dans la société . . . . .   | 14        |
| <b>4</b> | <b>RÉFÉRENCES BIBLIOGRAPHIQUES SIGNIFICATIVES DE RO</b>   | <b>17</b> |
| <b>A</b> | <b>ANNEXE — MEMBRES PERMANENTS AU 31/12/2022</b>  | <b>19</b> |

# 1 INFORMATIONS GÉNÉRALES SUR L'ÉQUIPE RO

**Nom de l'équipe :** Recherche Opérationnelle (RO)

**Responsable de l'équipe :** Bruno Escoffier

|   | 2017       | 2018       | 2019       | 2020       | 2021       | 2022       |
|---|------------|------------|------------|------------|------------|------------|
| PR                                      | 3          | 3          | 3          | 3          | 3          | 3          |
| MCF HDR                                 | 1          | 0          | 1          | 2          | 1          | 1          |
| MCF                                     | 4          | 4          | 3          | 2          | 2          | 3          |
| DR                                      | 1          | 1          | 1          | 1          | 1          | 1          |
| CR HDR                                  | 1          | 1          | 0          | 0          | 2          | 2          |
| CR                                      | 2          | 3          | 3          | 2          | 0          | 0          |
| <b>Total permanents</b>                 | <b>12</b>  | <b>12</b>  | <b>11</b>  | <b>10</b>  | <b>9</b>   | <b>10</b>  |
| Émérites                                | 1          | 1          | 1          | 1          | 1          | 1          |
| Doctorants                              | 4          | 7          | 8          | 9          | 7          | 9          |
| Ingénieurs CDD ou hors tutelles         | 0          | 0          | 0          | 1          | 0          | 1          |
| Post-doc, ATER, etc.                    | 0          | 1          | 0          | 1          | 3          | 5          |
| Stagiaires                              | 7          | 5          | 10         | 4          | 7          | 3          |
| <b>Total non permanents</b>             | <b>11</b>  | <b>13</b>  | <b>18</b>  | <b>15</b>  | <b>17</b>  | <b>18</b>  |
| <b>Total avec émérites</b>              | <b>24</b>  | <b>26</b>  | <b>30</b>  | <b>26</b>  | <b>27</b>  | <b>29</b>  |
| <b>Equivalent temps plein recherche</b> | <b>8.0</b> | <b>8.5</b> | <b>7.5</b> | <b>6.5</b> | <b>6.0</b> | <b>6.5</b> |

TABLE 1 – Personnels RO sur la période 2017-2022 (au 1er juillet de chaque année)

## 1.1 Les thématiques scientifiques et leurs enjeux

L'équipe RO s'intéresse principalement à la conception et à l'analyse d'algorithmes pour la résolution de problèmes complexes d'optimisation. Ces activités de recherche s'intègrent naturellement dans l'axe "Théorie et outils mathématiques pour l'informatique" du laboratoire, et dans une moindre mesure dans les axes "Intelligence artificielle et sciences des données" d'une part et "Architecture, systèmes et réseaux" d'autre part.

Les méthodes et approches développées dans l'équipe couvrent un spectre assez large (analyse de complexité, algorithmes approchés ou compétitifs, approches heuristiques, algorithmes randomisés, etc.). Un point d'intérêt particulier de l'équipe concerne l'optimisation en présence d'une *connaissance imparfaite* des données du problème à résoudre : la fonction objectif peut être inconnue (optimisation black-box), les données incertaines (optimisation robuste), les données connues au fur et à mesure (optimisation on-line, searching) ou évolutives (optimisation multistage). Les problèmes étudiés sont variés : si les problèmes d'ordonnancement, spécialité de l'équipe depuis ses débuts, restent un domaine d'expertise fort, l'équipe s'intéresse aussi à d'autres types de problèmes (problèmes de graphes par exemple).

Voici un bref descriptif des sujets d'études majeurs de l'équipe, et les avancées principales obtenues sur la période 2017-2022.

### A. Optimisation black-box : algorithmes de recherche randomisés, théorie des algorithmes évolutionnaires, configuration et sélection d'algorithmes, méthodes AutoML

De nombreux problèmes d'optimisation du monde réel, provenant aussi bien de l'industrie que du milieu académique, nécessitent des expériences physiques ou des simulations pour évaluer la qualité des solutions possibles. Par exemple, de tels problèmes d'optimisation "black-box" sont très fréquents dans les domaines de l'ingénierie, où l'impact de l'évolution des paramètres de conception est évalué à l'aide de modèles de simulation numérique. Les problèmes de type black-box ne peuvent être résolus qu'en alternant entre

- (1) la génération de nouvelles solutions candidates,
- (2) l'évaluation de leur qualité et
- (3) l'ajustement de la stratégie par laquelle les candidats suivants sont obtenus.

Un très grand nombre d'algorithmes existent, avec des forces et des faiblesses complémentaires. De plus, les approches d'optimisation black-box de pointe sont généralement accompagnées d'un ensemble de paramètres qui peuvent être utilisés pour ajuster le comportement de l'algorithme pour le problème spécifique (instance) à résoudre. Le choix des paramètres optimaux peut changer au cours du processus d'optimisation, par exemple pour permettre une transition d'un comportement d'exploration vers un comportement d'exploitation. La *sélection*,



la **configuration** et le **contrôle** de l'algorithme sont donc trois défis de méta-optimisation auxquels les utilisateurs sont confrontés lorsqu'ils doivent sélectionner une approche appropriée pour résoudre leur problème. Compte-tenu de la grande pertinence de l'optimisation black-box dans de nombreux domaines d'application, une compréhension plus approfondie des algorithmes à recommander en fonction du problème à résoudre offre un énorme potentiel pour une meilleure utilisation des ressources naturelles, humaines, financières et informatiques.

**Apports principaux de l'équipe (période 2017-2022).** Au sein de l'équipe RO, nous contribuons à l'analyse de l'optimisation black-box, tant par la recherche fondamentale que par des projets de recherche appliquée avec des partenaires académiques (ex. Institut Pasteur) et industriels (ex. Thales, Meta AI research, Honda RI).

Notamment, nous développons des assistants de sélection d'algorithmes de pointe, qui non seulement mettent à la disposition des utilisateurs les algorithmes d'optimisation black-box les plus performants, mais fournissent également des recommandations basées sur les données sur quel algorithme choisir pour quel problème [Meunier et al., 2022]. Notre approche pour améliorer les algorithmes existants hybride la recherche théorique (telle que celles résumées dans les chapitres de livre [Doerr, 2020] et [Doerr and Doerr, 2020]) avec des évaluations empiriques solides.

Pour ce dernier point, nous développons le logiciel de benchmarking IOHprofiler [Doerr et al., 2018], que nous utilisons largement pour générer, analyser et visualiser les trajectoires de recherche et les performances des algorithmes itératifs d'optimisation black-box. IOHprofiler est un atout clé dans nos activités de recherche, d'enseignement et de sensibilisation. Il devient également un outil important pour la communauté scientifique (plusieurs packages sont téléchargés plusieurs centaines de fois par mois), et dans nos partenariats industriels. Dans ce contexte, nous tirons fréquemment parti de nos connaissances théoriques pour soutenir l'analyse comparative empirique, voir [Biedenkapp et al., 2022] par exemple. De même, les observations empiriques nourrissent nos investigations théoriques, voir [Doerr et al., 2021] par exemple.

Notre vision à long terme est le développement d'approches d'optimisation black-box qui automatisent les trois défis clés énumérés ci-dessus - la sélection, la configuration et le contrôle des algorithmes - dans le processus d'optimisation [Kostovska et al., 2022].

## B. Algorithmes on-line : learning-augmented algorithms, explorability et on-line searching

Les problèmes on-line traitent de situations où l'algorithme ne connaît pas initialement les données du problème à résoudre, il les découvre au fur et à mesure. L'entrée du problème est typiquement une suite de requêtes, l'algorithme devant alors prendre une décision pour traiter chaque requête lorsqu'elle arrive. La difficulté consiste à prendre ces décisions sans avoir la connaissance des requêtes suivantes. Le prix de ce manque d'information est souvent mesuré par le rapport de compétitivité, qui compare la solution construite par l'algorithme à une solution optimale sous information parfaite.

Si l'étude de l'algorithmique on-line sous l'angle du rapport de compétitivité reste très active, la communauté s'est récemment intéressée à différents modèles où l'algorithme peut accéder à une information supplémentaire, typiquement sur les données manquantes. On peut alors s'intéresser à l'amélioration de la qualité des solutions fournies par un algorithme en fonction de la quantité et la qualité des informations reçues ("on-line algorithms with advice"). Une approche voisine est le cœur d'un domaine récent et en pleine expansion, appelé "learning augmented algorithms" : des prédictions sur les données/requêtes du problème sont fournies à l'algorithme. Ces prédictions peuvent s'avérer bonnes ou mauvaises. Le but est alors de concevoir des algorithmes qui fournissent de très bonnes solutions quand les prédictions sont correctes, sans que la qualité de la solution ne se dégrade trop vite avec l'erreur de la prédiction.

On peut aussi imaginer pouvoir accéder à l'information manquante modulo un certain coût, cadre du domaine de l'"optimisation sous incertitude explorable." Quiconque a participé à un projet industriel a constaté que l'obtention des données pour un problème d'optimisation peut être un véritable défi. Par exemple, dans un problème d'ordonnancement, on peut ne recevoir que des intervalles d'incertitude pour les paramètres de travail. On peut alors imaginer un mécanisme de requêtes permettant d'avoir accès, à un certain coût, à la valeur exacte d'un paramètre. Se pose alors le problème de poser les bonnes questions : comment minimiser le nombre de (ou le coût des) requêtes nécessaires pour calculer et certifier un planning optimal ?

Dans le domaine du **on-line searching**, un agent doit localiser un ou plusieurs éléments cachés dans un environnement qu'il découvre généralement au fur et à mesure de son exploration. L'objectif est de concevoir et d'analyser des stratégies efficaces pour localiser la cible. Des modèles avec prédictions, ou plus généralement avec information supplémentaire, sont également pertinents pour ces problèmes de recherche.

**Apports principaux de l'équipe (période 2017-2022).** Présente de longue date dans le domaine de l'algorithme



mique on-line, l'équipe a participé activement au développement de ces thématiques émergentes. Elle est particulièrement pionnière dans le domaine de l'incertitude explorable, que ce soit sur des problèmes d'ordonnancement [Dürr et al., 2018] ou de graphes [Bampis et al., 2021b]. Les résultats concernent à la fois la définition de stratégies de requêtes et leur analyse en terme de rapport de compétitivité, et l'obtention de résultats négatifs montrant les limites de ce qu'il est possible de faire (cf. élément 3 du portfolio).

Après des contributions sur les problèmes on-line "with advice", la thématique des algorithmes learning augmented suscite un intérêt croissant au sein de l'équipe, avec un travail pionnier en 2020 [Angelopoulos et al., 2020], et depuis des contributions sur différents problèmes (par exemple le bin-packing [Angelopoulos et al., 2022] et l'ordonnancement [Bampis et al., 2022]), permettant en général d'obtenir les meilleurs compromis connus entre cohérence (qualité de la solution calculée lorsque les prédictions sont correctes), régularité (qualité de la solution lorsque les prédictions sont erronées mais bonnes) et robustesse (qualité de la solution quand les prédictions sont mauvaises), cf. l'élément 4 du portfolio. Enfin, des avancées ont été obtenues dans différents problèmes de on-line searching (cow-path problem, graphes, . . .), par des travaux cherchant souvent à aller au-delà de l'analyse classique du rapport de compétitivité (voir par exemple [Angelopoulos et al., 2019]).

### C. Algorithmes en présence de données évolutives : robustesse, optimisation stochastique et optimisation multistage

Il est fréquent qu'en pratique les données d'un problème soient incertaines (pas précisément connues) et/ou susceptibles d'évoluer au cours du temps. Comme indiqué précédemment, dans un problème d'ordonnancement la durée d'une tâche peut être incertaine et donnée sous la forme d'un intervalle de valeurs possibles. Dans un problème de chemin dans un graphe, le temps de traversée d'un arc peut changer avec le temps (congestion du trafic, indisponibilité de l'arc, etc.).

Dans le cas de données incertaines, deux approches sont classiquement considérées : l'*optimisation stochastique* et l'*optimisation robuste*. Dans la première, on considère une distribution de probabilités sur les valeurs possibles des données et l'on cherche alors typiquement à optimiser un critère probabiliste. En optimisation robuste, nous ne disposons pas d'informations probabilistes mais simplement d'un ensemble d'incertitudes caractérisant les valeurs possibles des données, et l'on cherche classiquement une solution qui reste de bonne qualité dans tout l'ensemble d'incertitudes.

L'*optimisation multistage*, telle qu'étudiée dans l'équipe, traite de problème où les données évoluent dans un processus temporel. Le but est de maintenir une solution réalisable au cours de ce processus temporel. Deux coûts sont généralement à prendre en compte : le coût de la solution calculée à chaque pas de temps au cours du processus, et les coûts de transition d'une solution à une autre (modifier une solution induit des coûts qui peuvent être en pratique très importants). On cherche alors à construire une séquence de solutions qui sont à la fois de bonne qualité (faible coût) et relativement stables au cours du temps. L'évolution des données peut être un processus connu à l'avance, partiellement connu (avec des prédictions fiables à court terme par exemple) ou inconnu (on est alors proche d'un problème on-line).

**Apports principaux de l'équipe (période 2017-2022).** Nous les regroupons sur trois domaines :

- ▶ en robustesse, les travaux de l'équipe sont pionniers dans l'étude de la notion d'*ancrage de décisions*, voir par exemple [Bendotti et al., 2019a]. Il s'agit de déterminer, en présence d'incertitudes sur les données, des solutions robustes en ce sens qu'on peut garantir qu'une partie de la solution ne sera pas remise en cause par l'incertitude - on peut typiquement garantir a priori la date d'exécution de certaines tâches, malgré l'incertitude qui peut peser sur la durée d'exécution d'autres tâches. Cette notion a été étudiée dans une thèse CIFRE EDF (récipiendaire de plusieurs prix). Elle ouvre plusieurs directions de recherche, qui font partie des directions futures de l'équipe.
- ▶ en optimisation stochastique, l'équipe a travaillé (dans le cadre d'une thèse CIFRE Nokia notamment) sur la minimisation de la consommation énergétique dans le cloud tout en garantissant la qualité de service, par la recherche de politiques efficaces d'allumage et d'extinction des machines virtuelles au cours du temps en fonction des statistiques du système [Tournaire et al., 2019]. Les outils mathématiques utilisés sont les Processus de Décisions Markoviens et l'apprentissage par renforcement. Nous avons également travaillé sur les propriétés structurelles de politiques optimales dans des modèles avec échéance. Un autre axe concerne la détection de cibles dans un environnement incertain dans lequel nous devons définir les actions de recherches afin de détecter la cible le plus vite possible (dans le cadre d'une thèse avec Thales DMS notamment).
- ▶ en optimisation multistage, deux directions principales ont été étudiées. La première concerne le *Unit Com-*



*mitment Problem* (UCP), problème de planification pour la résolution duquel des avancées significatives ont été obtenues dans une thèse CIFRE EDF (récipiendaire également de plusieurs prix), voir par exemple [Bendotti et al., 2019b]. Une étude sur les symétries du problème a par exemple permis d'obtenir des améliorations importantes sur la résolution pratique par rapport aux méthodes existantes. La deuxième concerne une étude systématique de la complexité et de l'approximation de problèmes multistage (problèmes de graphes, d'ordonnancement notamment) pour comprendre dans quelle mesure l'aspect multi-étapes d'un problème complexifie sa résolution (voir par exemple [Bampis et al., 2019]). Les résultats ont montré que, si certains problèmes deviennent très difficiles à résoudre dans ce cadre multistage, des algorithmes approchés peuvent être conçus pour de nombreux problèmes, notamment à l'aide de la programmation mathématique.

#### D. Complexité, résolution exacte et approchée : problèmes d'ordonnancement, de clustering, choix social computationnel

Savoir dans quelle mesure il est possible de résoudre efficacement un problème d'optimisation est une question fondamentale en algorithmique. Elle repose sur l'étude de la complexité (classique ou paramétrée) et sur la conception et l'analyse d'algorithmes (exacts ou approchés). Les travaux de l'équipe dans ces domaines ont porté principalement sur 3 familles de problèmes : les problèmes d'ordonnancement, les problèmes de clustering, et les problèmes de choix social computationnel (et quelques autres, comme des problèmes de graphes, voir par exemple [Bouquet et al., 2021]).

La résolution de problèmes d'*ordonnancement* est un sujet d'expertise de longue date de l'équipe ; c'est un vaste domaine classique de la recherche opérationnelle où, en quelques mots, il s'agit d'exécuter un certain ensemble de tâches le plus efficacement possible tout en respectant un ensemble de contraintes données (précédences, dates de disponibilité,...).

Dans un problème de *clustering* (partitionnement), on doit trouver une partition d'un ensemble de données (points d'un espace) en un certain nombre de classes, de manière à optimiser un certain critère (les deux plus classiques étant  $k$ -mean et  $k$ -median) : si ces problèmes de clustering sont étudiés depuis longtemps (partitionnement de graphes par exemple), ils ont acquis une importance considérable ces dernières années de part leur application dans des problèmes d'intelligence artificielle.

Le *choix social computationnel* étudie, sous un angle algorithmique, des problèmes de choix social où il s'agit de prendre une décision en présence de plusieurs agents aux avis souvent contradictoires. Les problèmes de ce domaine intègrent des éléments spécifiques liés au contexte du choix social, que ce soit au niveau des données (se basant généralement sur des structures de préférences), de la fonction objectif (avec par exemple des notions d'équité entre les agents), ou des propriétés que doit vérifier un algorithme ou un mécanisme (comme la non-manipulabilité). L'étude de ces problèmes est bien sûr source de collaborations avec l'équipe Décision, dont c'est la cœur de métier.

**Apports principaux de l'équipe (période 2017-2022).** Nous les regroupons sur trois domaines :

- ▶ En ordonnancement, une direction originale de l'équipe a été de prendre en compte les aspects liés à la consommation énergétique dans les problèmes d'ordonnancement (voir par exemple [Bampis et al., 2018b]). Cette direction s'inscrit dans le projet ANR Energumen, dans le but de concevoir, étudier et valider des outils pratiques pour l'allocation efficace en énergie de tâches sur plate-forme HPC à large échelle. Dans une autre direction, nous avons étudié des problèmes d'ordonnancement cycliques, avec applications au domaine temps réel (voir par exemple [Hanan and Hanzalek, 2020]). Enfin, une direction de recherche plus récente concerne l'étude de la complexité paramétrée sur des problèmes d'ordonnancement (voir par exemple [Malle et al., 2022]). Une collaboration avec l'équipe Alsoc et le laboratoire Heudisasy a été initiée sur cette thématique, qui sera reprise dans la section sur la trajectoire de l'équipe.
- ▶ L'équipe a obtenu des avancées significatives dans la résolution des problèmes du clustering (voir par exemple [Cohen-Addad et al., 2021b, Cohen-Addad et al., 2021a]), notamment concernant la résolution approchée de ce problème (avec un schéma d'approximation en temps linéaire quand l'entrée est dans un espace euclidien de dimension constante), et le calcul de représentations simplifiées de l'entrée, qui préservent la structure du problème (dans le cas particulier des espaces euclidiens par exemple a été proposé le premier algorithme déterministe pour réduire la dimension du problème de manière quasi-optimale), cf. l'élément 5 du portfolio.
- ▶ Dans le domaine du choix social computationnel, une direction originale de l'équipe concerne l'étude de problèmes d'optimisation impliquant des ressources partagées (des machines par exemple dans un problème

d'ordonnement) entre différents agents. Le but est de concevoir des mécanismes garantissant de bonnes performances dans de tels systèmes (efficacité des solutions, véracité garantie, équité, ...), voir par exemple [Pascual et al., 2018, Durand and Pascual, 2021]. Une deuxième direction concerne l'étude des préférences dites structurées, sous un angle algorithmique : des travaux récents, effectués en collaboration avec l'équipe Décision, ont permis de mieux comprendre les structures de préférences euclidiennes notamment.

### Animation scientifique

L'équipe anime un séminaire de recherche régulier<sup>1</sup>. Celui-ci alterne entre des exposés de membres extérieurs à l'équipe, et des exposés de membres de l'équipe. Chaque année les stagiaires de l'équipe sont également invités à faire un court exposé de leur sujet, lors d'une session spéciale du séminaire. Après une activité faible en 2020 due à la pandémie, le séminaire a repris une activité habituelle, avec une fréquence d'un séminaire toutes les 3 semaines en moyenne (15 exposés en 2021, 13 en 2022).

Les séminaires sont parfois suivis de réunions d'équipes, permettant de diffuser l'information et d'échanger sur la vie de l'équipe.

L'animation scientifique se base aussi sur l'invitation de nombreux chercheurs au sein de l'équipe (plus de quarante sur la période d'évaluation!).

L'équipe profite également ces dernières années d'un nombre accru de doctorants et post-doctorants, éléments très importants du dynamisme au sein d'un groupe de recherche. Ces derniers ont initié en 2021 la tenue d'un séminaire de doctorants commun aux équipes RO et Décision (une dizaine de présentations en 2022). De manière plus informelle, ils ont initié l'organisation régulière (au moins hebdomadaire) d'une session de jeux sur la pause de midi, permettant de créer du lien entre les personnes de l'équipe.

---

1. voir <https://webia.lip6.fr/~escoffier/seminaire/seminaire.html>

## 2 INTRODUCTION DU PORTFOLIO

L'équipe a constitué un portfolio constitué des 5 éléments suivants.

- ▶ **Élément 1 (distinction)** : médaille de bronze du CNRS 2022 décernée à un membre de l'équipe (pour des recherches portant essentiellement sur le point A des activités décrites dans la section 1.1).
- ▶ **Élément 2 (collaboration industrielle)** : collaboration pérenne et très active avec l'entreprise EDF (pour des recherches portant essentiellement sur le point C des activités décrites dans la section 1.1).
- ▶ **Élément 3 (article scientifique)** : article intitulé "An Adversarial Model for Scheduling with Testing", par Christoph Dürr, Thomas Erlebach, Nicole Megow et Julie Meissner, publié dans la revue *Algorithmica* (2020), portant sur l'optimisation sous incertitude explorable (point B des activités décrites dans la section 1.1).
- ▶ **Élément 4 (article scientifique)** : article intitulé "Scheduling with Untrusted Predictions", par Evripidis Bampis, Kostantinos Dogeas, Alexander V. Kononov, Giorgio Lucarelli, Fanny Pascual, publié à la conférence IJCAI 2022, portant sur la prise en compte de prédictions dans la résolution de problèmes d'ordonnancement (point B des activités décrites dans la section 1.1).
- ▶ **Élément 5 (article scientifique)** : article intitulé "A new coresets framework for clustering", par Vincent Cohen-Addad, David Saupic et Chris Schwiegelshohn, publié dans les actes de la conférence Symposium on Theory of Computing (STOC) 2021, portant sur la résolution de problèmes de clustering (point D des activités décrites dans la section 1.1).

Chaque élément est détaillé dans une fiche présente dans le document du portfolio.

## 3 AUTOÉVALUATION DU BILAN

### 3.1 Autoévaluation de l'équipe

#### Domaine 2. Attractivité

**Prix et distinctions.** Un premier aspect du rayonnement scientifique de l'équipe concerne les prix et distinctions reçues pendant la période 2017-2022. Le plus notable d'entre eux est sans doute la *médaille de bronze CNRS 2022* décernée à un membre de l'équipe, pour l'excellence de ses travaux sur l'optimisation black-box (point A des thématiques décrites dans la section 1.1). Ce prix fait l'objet de l'élément 1 du portfolio de l'équipe.

Par ailleurs, un membre de l'équipe est *membre junior de l'IUF* (2019-2024), avec un projet portant sur l'optimisation multistage (point C des thématiques décrites dans la section 1.1).

L'excellence de nos doctorants a également été récompensée, avec *trois prix de thèse* de portée nationale : le prix de thèse Paul Caseau 2019 et le prix de thèse PGMO (programme Gaspard Monge pour l'optimisation) 2018 pour une thèse intitulée "Aspects combinatoires du Unit Commitment Problem" et le prix de thèse PGMO 2022 pour une thèse intitulée "Anchored solutions in robust combinatorial optimization". Egalement, un étudiant de master (co-encadré dans les équipes RO et Décision) a reçu le prix du meilleur mémoire de master de RO en 2020 (prix décerné par la société savante ROADEF).

Par ailleurs, plusieurs prix ont récompensé la qualité de nos articles scientifiques :

- ▶ un "best paper award" à la conférence SoCG 2019,
- ▶ deux "best paper awards" à la conférence GECCO (2021 et 2022),
- ▶ un "best paper award" à la conférence IEEE CEC 2021,
- ▶ un "best paper award" à la conférence EvoApplications 2021,
- ▶ le premier et le troisième prix à la compétition black-box de la conférence NeurIPS 2020,
- ▶ le prix du meilleur article étudiant à la conférence PMS 2019,
- ▶ et deux prix du meilleur article étudiant aux conférences nationales ROADEF 2019 et 2022.

**Invitations.** La qualité des recherches menées au sein de l'équipe est également attestée par les invitations de ses membres à faire des exposés dans des événements scientifiques, ou à séjourner dans des universités à l'étranger. Concernant les événements scientifiques, nous distinguons ci-dessous trois catégories : les invitations à présenter nos recherches récentes (comme les "plenary talks" de conférences/workshops), les invitations à faire des exposés tutoriaux, et les invitations à faire des exposés ou des cours dans des écoles à destination des jeunes chercheurs/chercheuses.

Concernant le premier point, nous avons été invités à faire des exposés en lien avec :

- ▶ l'optimisation black-box et les algorithmes évolutionnaires (point A des thématiques décrites dans la section 1.1) à BIOMA 2022 (Maribor, Slovenia), IEEE SSCI Students and Young Professionals workshop 2021 et 2022, MENDEL International Conference on Soft Computing 2017 (Brno, Czech Republic) et dans le workshop CoA 2022 organisé dans le cadre du GDR-IM.
- ▶ l'algorithmique on-line (dont l'incertitude explorable et/ou les learning augmented algorithms) (point B des thématiques de la section 1.1), aux conférences ACAC (Athens Colloquium on Algorithms and Complexity) 2022, FAW (Frontiers in Algorithms) 2020, MOTOR (Mathematical Optimization Theory and Operations Research) 2019, LAGOS (The Latin and American Algorithms, Graphs and Optimization Symposium) 2017,
- ▶ l'optimisation multistage (point C des thématiques de la section 1.1), à la conférence MOTOR 2020 ;
- ▶ la résolution de problèmes d'ordonnancement (point D des thématiques de la section 1.1), à la conférence ROADEF 2021 et au workshop IWDSPP (International Workshop on Dynamic Scheduling Problems) 2018.

Nous avons fait des exposés tutoriaux (invités) sur :

- ▶ l'optimisation black-box (point A des thématiques décrites dans la section 1.1) aux conférences GECCO 2017, 2018, 2019, 2020 et 2022, IEEE WCCI CEC 2020 et 2022, CEC 2019, PPSN 2018.
- ▶ l'algorithmique on-line (point B des thématiques de la section 1.1), à la conférence ROADEF 2021 ;
- ▶ la complexité et les algorithmes approchés (point D de la section 1.1), aux conférences ROADEF 2017 (algorithmes approchés) et ROADEF 2019 (complexité paramétrée).

Enfin, en lien avec la formation des jeunes chercheurs, nous avons été invités à effectuer des exposés et des cours dans des écoles sur nos domaines d'expertise. Nous pouvons citer :

- ▶ six exposés invités en lien avec l'optimisation black-box (point A des thématiques de la section 1.1) : AutoML fall school 2022 (Freiburg, Germany), SIGEVO summer school 2021 et 2022, ECOLE ITN summer school 2021, GECCO 2017 Summer School (Berlin, Germany).
- ▶ un cours sur les processus de décision markoviens (point C des thématiques de la section 1.1), à l'école d'été RESCOM 2019.
- ▶ deux exposés sur la complexité et l'approximation (point D des thématiques de la section 1.1) : école d'été sur l'Ordonnancement et l'Approximation 2019 (Metz, France) et l'école jeunes chercheurs du GDR-RO 2017 (Metz, France).

Concernant les séjours de recherche dans les universités à l'étranger, un membre de l'équipe a fait un séjour d'un an à l'Universidad de Chile (Santiago, Chili) sur la thématique de l'incertitude explorable (point B de la section 1.1), et un membre a séjourné un mois et demi à la Czech Technical University (Prague, République Tchèque) sur la thématique de l'ordonnancement (point D de la section 1.1). D'autres séjours plus courts ont bien évidemment eu lieu, nous ne les listons pas ici.

**Activités éditoriales et participation à des comités de programme de conférences.** L'équipe participe au comité éditorial de 4 revues :

- ▶ Discrete Applied Mathematics (un membre du comité éditorial)
- ▶ Evolutionary Computation (un membre du comité éditorial depuis 2019)
- ▶ ACM Transactions on Evolutionary Learning and Optimization (un membre Associate Editor, depuis 2019)
- ▶ IEEE Transactions on Evolutionary Computation (un membre Associate Editor, depuis 2022).

Des membres de l'équipe ont par ailleurs été coordinateurs/trices de numéros spéciaux de revues, en tant qu'éditeur invité. Il y en a eu 7 sur la période : deux numéros de la revue *Algorithmica* (2017 et 2019), trois numéros de la revue *Theoretical Computer Science* (2017, 2021, 2022), un numéro de la revue *Theory of Computing Systems* (2021), un numéro de la revue *IEEE Transactions on Evolutionary Computation* (2022).

Concernant les conférences, l'équipe a piloté certaines conférences en tant que PC chair, notamment dans le domaine de l'optimisation black-box (point A de la section 1.1) : FOGA 2019, PPSN 2020, ACM GECCO GECH track 2023, GECCO theory track 2017. Dans les domaines relatifs aux algorithmes on-line et approchés (points B et D de la section 1.1), un membre de l'équipe est dans le steering committee de la conférence WAOA, et a été PC chair de l'édition 2019 de la conférence.

Les membres de l'équipe sont bien sûr très régulièrement membres de comité de programme de conférences (que nous ne listons pas exhaustivement ici), sur des domaines variés. Il s'agit notamment :

- ▶ de conférences d'algorithmique et d'informatique théorique, comme ESA 2017, FOCS 2018, WAOA 2019, FCT 2021, ADT 2021, TAMC 2022, etc.
- ▶ de conférences sur l'optimisation black-box et les algorithmes évolutionnaires, comme GECCO de 2017 à 2022, FOGA de 2017 à 2021, EvoStar de 2017 à 2022, etc.
- ▶ de conférences d'intelligence artificielle, comme AutoML 2022 (Senior PC), IJCAI (tous les ans de 2019 à 2022), AAMAS 2021 et 2022, ECAI 2020, des workshops à ECML, AAI, etc.

Notons que l'engagement de l'équipe se situe aussi au niveau national, avec l'organisation de nombreuses sessions à la conférence annuelle ROADEF.

**Participation à des instances de pilotage de la recherche et d'expertise scientifique.** De part sa thématique scientifique, l'équipe est impliquée au niveau national dans le GDR-RO (nouvellement GDR-ROD) : 3 personnes sont membres de son conseil scientifique, et sont co-responsables de deux des 7 axes scientifiques du GDR (deux responsables de l'axe "Complexité, Approximation et Graphes pour la Décision et l'Optimisation", un responsable de l'axe "Décision : Modélisation, Évaluation, Incertitude"). Par ailleurs, une personne de l'équipe est membre du conseil scientifique du GIS Collex Persée. Au niveau international, une personne est membre de l'Evolutionary Computation Technical Committee de l'IEEE Computational Intelligence Society.

Concernant l'expertise scientifique, une personne de l'équipe a été membre du comité ANR CE 48 - fondements du numérique - en 2022. Plus ponctuellement, l'équipe apporte régulièrement son expertise dans l'évaluation de projets pour des instances de financement de différents pays : France (ANR), Chili (Fondecyt), République Tchèque (Research, Development and Innovation Council), Suisse (Swiss Science Foundation), Pays-Bas (Dutch Research Council (NWO)), Pologne (National Science Center (NCN)), Autriche (Austrian Science Fund (FWF)). L'équipe a également évalué des projets au niveau européen, dans les appels MSCA (Marie Skłodowska-Curie Actions), programmes ITN/DN et IF.

**Organisation d'événements scientifiques.** L'équipe est impliquée dans l'organisation d'événements scientifiques internationaux de différentes natures, notamment :

- ▶ des séminaires scientifiques, d'une durée généralement d'une semaine, rassemblant la communauté intéressée par une thématique scientifique : 2 séminaires Dagstuhl 2017, 2019 (une cinquantaine de participants chacun), deux workshops au Lorentz Center 2020, 2022 (100 et 50 participants) ;
- ▶ des conférences et workshops, comme le workshop on Benchmarking aux conférences GECCO (2019, 2020, 2021, 2022), PPSN (2020, 2022) et CEC (2020, 2021), regroupant entre 50 et 90 participants, le workshop on Analysing Algorithmic Behaviour of Optimisation Heuristics CEC 2022. Un membre de l'équipe a été Workshop chair à la conférence GECCO 2022, et organisatrice de l'EvoApplications Special Session 2022.
- ▶ des compétitions sur la résolution de problèmes : Open Optimization Competition 2020, 2021, 2022
- ▶ des événements à destination des jeunes chercheurs et chercheuses, comme l'ImAppNIO COST Training School 2017 (7 jours, 35 participants).

L'équipe participe également à l'organisation d'événements pour la communauté française. Nous pouvons citer par exemple la participation à l'organisation des JGA (Journées Graphes et Algorithmes) 2022 du GDR-IM (une centaine de participants, sur 3 jours), ou l'AEP (Atelier d'Evaluation de Performances) 2022 (une cinquantaine de participants, sur 2 jours).

## Référence 2. L'unité est attractive par la qualité de sa politique d'accompagnement des personnels.

Les chercheurs de l'équipe bénéficient de l'environnement de travail et de l'accompagnement mis en place au niveau du laboratoire - et plus largement aux niveaux UFR, facultaires et universitaires. Nous ne les détaillons pas ici, en nous positionnant au niveau de l'équipe.

Concernant les doctorants, nous privilégions en général le co-encadrement (profitable à la fois pour le doctorant et pour les jeunes collègues MCF et CR). Nous incitons les doctorants et les post-doctorants à faire des présentations (souvent plusieurs fois pendant la thèse pour les doctorants) à notre séminaire d'équipe. Cela leur permet d'avoir un retour sur leur travail, et fournit bien sûr un entraînement pour les présentations dans des conférences. Pour les stagiaires de l'équipe, nous organisons en général une session spéciale du séminaire où chaque stagiaire présente son sujet de stage, et les éventuels résultats obtenus. C'est à la fois formateur et utile dans leur intégration dans l'équipe. Egalement, nous sommes attentifs à la valorisation du travail de nos étudiants (doctorants notamment), comme en témoignent les différents prix et distinctions obtenus (3 prix de thèse, un prix de master, plusieurs "best paper awards", cf section 3.1 domaine 2), pour la suite de leur carrière.

Les nombreux projets financés sur la période (cf référence 3 ci-après) assurent un environnement confortable pour tous, au niveau de l'achat de matériel, de la participation à des conférences, des workshops ou des écoles.

L'attractivité des recherches menées par l'équipe permet un recrutement de doctorants et de post-doctorants de qualité, venant d'horizons variés (de France - master de Sorbonne Université, ENS, X, etc. - mais également de l'étranger (Allemagne, Grèce, Italie notamment)). Concernant les concours CR CNRS, nous attirons chaque année de bonnes candidatures (souvent auditionnées, généralement classées, parfois en liste principale mais sans affectation dans l'équipe pendant la période) ; nous accompagnons les candidats dans la constitution de leur dossier (leur projet de recherche et leur intégration dans l'équipe), et par la simulation d'auditions (au sein ou en complément du dispositif mis en place par le laboratoire).

Au niveau des permanents, plusieurs membres de l'équipe ont passé des étapes importantes de leur carrière pendant la période 2017-2022. Nous avons eu ainsi une promotion CR→DR, et deux MCF ont été promus professeur dans une autre université (un en 2017, un en 2020). Par ailleurs, 4 personnes ont passé leur d'HDR (un MCF, 2 CR, un membre associé), ce qui a permis d'augmenter le potentiel d'encadrement de l'équipe.

Enfin, un point notable de l'équipe concerne le dynamisme insufflé par les nombreux chercheurs invités. Ils et elles sont *plus d'une quarantaine* à avoir visité l'équipe pendant la période 2017-2022, provenant de divers pays (Allemagne, Australie, Espagne, Italie, Portugal, Russie, etc.). La durée typique de séjour est de l'ordre d'une semaine, mais celle de certains séjours (de l'ordre d'une dizaine) est d'au moins 3 semaines. Certaines invitations ont été financées par le dispositif des professeurs invités du laboratoire, la majorité sur des financements propres à l'équipe.

Ces invitations traduisent l'existence de nombreuses collaborations internationales de l'équipe. Parmi les principales (sur la période) nous pouvons citer :

- ▶ l'University of Coimbra (Portugal), Jozef Stefan Institute (Slovénie), Leiden University (Pays-Bas), l'University of Adelaide (Australia) et TU Munich (Allemagne) sur le point A de la section 1.1 ;
- ▶ l'Universidad de Chile (Chili), Durham University (UK), l'University of Bremen (Allemagne), York University (Canada), University of California (USA) et NTU Athens (Grèce) sur le point B de la section 1.1 ;

- ▶ le Sobolev Institute of Mathematics (Russie) sur le point C de la section 1.1 ;
- ▶ la Czech Technical University (République Tchèque), l'University of Warsaw (Pologne), l'University of Sheffield (UK) et Aarhus University (4) sur le point D de la section 1.1.

### Référence 3. L'unité est attractive par la reconnaissance de ses succès à des appels à projets compétitifs.

L'équipe RO a vu un accroissement du nombre de projets de recherche financés. Ceci récompense notamment l'effort fait concernant la variété des appels à projets sur lesquels l'équipe a candidaté, aux niveaux internationaux, nationaux et régionaux. Comme cela est illustré ci-dessous, les projets financés couvrent l'ensemble des thématiques de recherche de l'équipe décrites dans la section 1.1.

**Projets internationaux.** L'équipe a participé à 5 projets internationaux sur la période :

- ▶ 3 projets sur des thématiques relatives au point A de la section 1.1 : une action COST, dont un membre de l'équipe était vice-chair, un projet IAE PRC CNRS/RFBR avec la Russie, un projet avec l'Allemagne financé par le Franco-Bavarian University cooperation center.
- ▶ 1 projet en partenariat avec l'Allemagne, financé par l'EPSRC, sur des thématiques relatives au point B de la section 1.1.
- ▶ 1 projet PHP Polonium en partenariat avec la Pologne, sur des thématiques relatives au point D de la section 1.1.

Il est à noter qu'un membre de l'équipe a candidaté à une ERC-starting grant en 2022, évaluée A après le 2ème round mais non sélectionnée. Nous espérons que son projet sera couronné de succès dans les appels à venir.

**Projets ANR.** Les membres de l'équipe ont participé à 6 projets ANR pendant la période :

- ▶ le projet JCJC FOCAL sur les algorithmes de clustering (point D de la section 1.1),
- ▶ le projet VARIATION ("Opérateurs de variation généralisés pour les heuristiques de recherche aléatoire") relatif au point A de la section 1.1 ;
- ▶ le projet Algoridam ("Algorithmic theory of new DAta Models") liés aux points B et C de la section 1.1 ;
- ▶ trois projets relatifs au point D de la section 1.1 : Energumen, ("Energy saving in large scale distributed platforms"), Themis - ("THEory and Evidence to Measure Influence in Social structures"), et Cocorico ("Calcul, Communication, Rationalité et Incitations en Décision Collective et Coopérative").

**Autres projets nationaux et régionaux.** Les succès de l'équipe à d'autres appels nationaux et régionaux ont été variés. Nous avons eu :

- ▶ 3 projets financés par la région Île de France, dans le cadre du DIM-RFSI. Deux de ces projets ont par exemple permis de recruter un post-doctorant.
- ▶ 1 projet FUI, qui a permis le financement d'une thèse de doctorat.
- ▶ 2 projets financés par l'INS2I/le CNRS.
- ▶ 7 projets financés par la fondation mathématique Jacques Hadamard (projets PGM0). Il s'agit d'une source de financements de petite ampleur (permettant le financement de stages et de missions typiquement) mais avec des succès réguliers au niveau de l'équipe, qui permet par exemple de démarrer un projet sur une thématique naissante.
- ▶ des projets sur des appels locaux, comme un projet financé par l'institut Energy for Climate, ou un projet sur la thématique de l'ordonnancement en partenariat avec l'UTC (avec recrutement d'un post-doctorant, financement par Sorbonne Universités).

Egalement, un post-doctorant a été recruté grâce à un co-financement du programme de Marie Curie et de la région Ile-de-France.

### Référence 4. L'unité est attractive par la qualité de ses équipements et de ses compétences techniques.

Cette référence ne semble pas pertinente pour l'équipe.

## Domaine 3. Production scientifique

### RO, Évolution des publications (2017–2022)

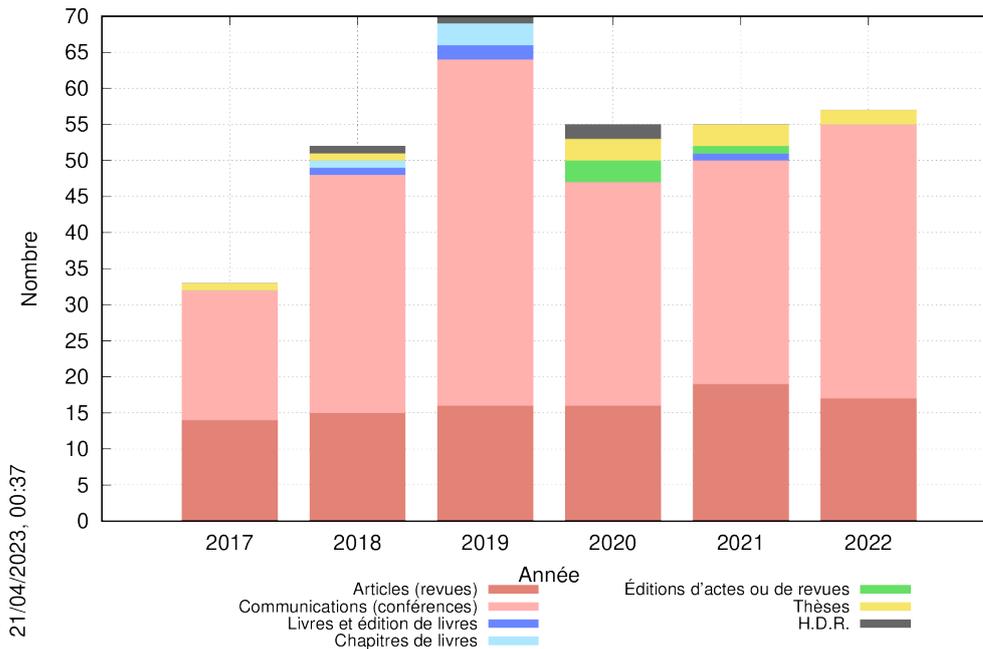


FIGURE 1 – Évolution des publications entre 2017 et 2022

|                                     | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 |
|-------------------------------------|------|------|------|------|------|------|
| <b>Articles (revues)</b>            | 1.75 | 1.76 | 2.13 | 2.46 | 3.16 | 2.61 |
| <b>Communications (conférences)</b> | 2.25 | 3.88 | 6.40 | 4.76 | 5.16 | 5.84 |

TABLE 2 – Publications par ETPR par an entre 2017 et 2022

#### Référence 1. La production scientifique de l'unité satisfait à des critères de qualité.

Les travaux de l'équipe reposent sur les fondements théoriques de l'algorithmique, sous des paradigmes et méthodes variés décrits dans la section 1.1. La production scientifique traduit la qualité des recherches menées, avec des publications dans les meilleurs conférences d'informatique théorique (mentionnons notamment 4 publications à la conférence FOCS et 3 à la conférence STOC) et d'algorithmique (par exemple 4 publications à la conférence SODA, 2 à la conférence ICALP, 3 à la conférence ESA), et dans les journaux de référence de ces domaines (citons *18 articles dans la revue Algorithmica*, et différents articles dans les revues Journal of the ACM, SIAM Journal on Computing, Journal of Computer and System Sciences, 12 articles dans la revue Theoretical Computer Science).

L'expertise de l'équipe sur ses thématiques de recherche est également visible dans ses publications dans des conférences et revues plus spécialisées. Nous pouvons citer dans cette idée des publications dans les revues de recherche opérationnelle (2 dans Mathematical Programming, 2 dans Operations Research, 8 publications dans Journal of Scheduling, 8 dans European Journal of Operations Research notamment) et dans des conférences spécialisées sur les domaines étudiés (une trentaine d'articles à la conférence de référence du domaine des algorithmes évolutionnaires GECCO, publications dans les conférences SAGT, WAOA, MFCS, SWAT, etc.).

Enfin, plusieurs thématiques de recherche de l'équipe sont en lien avec l'intelligence artificielle (auto-configuration d'algorithmes, problèmes de clustering, choix social combinatoire, learning augmented algorithms, etc.), et nous pouvons mettre en avant une production scientifique notable dans ce domaine (qui n'est pas le cœur de métier de l'équipe) avec plusieurs publications dans les actes de conférence de référence du domaine (6 à la conférence IJCAI, 5 à la conférence AAAI, 2 à la conférence AAMAS, 4 à la conférence NeurIPS notamment).

## Référence 2. La production scientifique de l'unité est proportionnée à son potentiel de recherche et correctement répartie entre ses personnels.

Avec en moyenne 2,3 publications par an et par ETPR dans des journaux (de qualité, cf point précédent), la production est riche et régulière sur la période d'évaluation (avec une légère tendance à la hausse). La mesure quantitative des publications dans des conférences est également satisfaisante (environ 4,7 par an et par ETPR) et régulière sur la période (en légère hausse, hormis un pic en 2019). Il est intéressant de noter que l'activité scientifique de l'équipe, en terme de production mais également de projets financés, couvre l'ensemble de ses thématiques - et l'ensemble de ses personnels.

Comme indiqué dans la référence 2 (accompagnement des personnes) du domaine 2, une attention est portée sur la mise en valeur des travaux des doctorants et post-doctorants. Ceux-ci participent très activement à la production scientifique de l'équipe (par exemple, sur les 17 publications revues de l'équipe en 2022, 13 ont pour co-auteur un ou plusieurs doctotant, post-doctorant et/ou stagiaire de M2).

## Référence 3. La production scientifique de l'unité respecte les principes de l'intégrité scientifique, de l'éthique et de la science ouverte. Elle est conforme aux directives applicables dans ce domaine.

L'équipe n'organise pas d'action spécifique dans ce cadre, mais applique des pratiques en phase avec l'intégrité scientifique, l'éthique et la science ouverte. Les membres de l'équipe mettent généralement une version accessible de leurs publications (le plus souvent sur les repository Arxiv ou HAL, plus rarement sur leur page web), ce qui est utile en particulier lorsque l'accès à la publication concernée est soumise à un abonnement - ce qui reste, il est vrai, souvent le cas pour nos publications (revues éditées par Springer et Elsevier par exemple). Cela dit, nous préférons le système d'abonnement à l'alternative (proposée par ces revues notamment) où les auteurs paient pour la publication (alors accessible sans frais) de leur article, procédé que nous n'avons jamais utilisé. Une attention est portée sur la qualité des supports de publication (revues ou actes de conférences), dans l'optique de privilégier la qualité à la quantité. Notons sur ce sujet qu'un nombre croissant de conférences passent sur un modèle d'actes en libre accès (y compris des conférences sélectives et reconnues)

Concernant la production de code ou de logiciel, l'équipe a principalement participé, sur la période d'évaluation, au développement du logiciel IOHprofiler<sup>2</sup>, qui est distribué sous une licence libre (3-Clause BSD License).

Egalement, le site web <http://schedulingzoo.lip6.fr>, très utilisé par la communauté d'ordonnancement (il recense les résultats de la littérature sur les principaux problèmes d'ordonnancement, avec une interface de recherche efficace) a ses sources distribuées sous GitHub sous la licence GNU. Une bibliothèque Python tryalgo, proposant une implantation Python optimisée pour résoudre plus d'une centaine de problèmes algorithmiques classiques, est distribuée sous la licence MIT. Enfin, l'équipe a participé au développement d'une ontologie (Optimization Algorithm Benchmarking Ontology<sup>3</sup>) accessible en open source.

## Domaine 4. Inscription des activités de recherche dans la société

### Référence 1. L'unité se distingue par la qualité et la quantité de ses interactions avec le monde non-académique.

La recherche opérationnelle est un domaine dont l'un des objectifs est de lier théorie et pratique. La majeure partie des activités de l'équipe RO se situe sur le versant théorique, mais l'équipe a développé sur la période 2017-2022 de nombreuses applications avec des partenaires industriels.

**EDF.** L'équipe développe une collaboration pérenne et riche avec EDF R&D. Cette collaboration fait l'objet de l'élément 2 du portfolio de l'équipe, aussi nous ne la détaillons pas ici. Mentionnons simplement qu'elle a fait l'objet sur la période de 2 thèses CIFRE, d'un contrat de collaboration, et que les recherches menées ont été récompensées par plusieurs prix comme indiqué précédemment.

**Thalès.** La collaboration avec l'entreprise Thalès s'est située sur deux plans. Le premier porte sur un problème de positionnement optimal de capteurs (pour maximiser la capacité à détecter des appareils). Un travail sur le développement de méthodes AutoML (sélection automatique d'algorithmes) a été mené dans le cadre d'une thèse CIFRE. Le deuxième point concerne également les capteurs, sur un problème de coordination de systèmes multi-capteurs aéroportés de défense. Il a également été l'objet d'une thèse CIFRE.

2. <https://iohprofiler.github.io/>

3. <https://bioportal.bioontology.org/ontologies/OPTION-ONTOLOGY>

**Nokia.** L'équipe a collaboré avec l'entreprise Nokia, au travers d'une thèse CIFRE. La problématique étudiée concernait la définition de politique de gestion de machines (allumage/extinction) dans le cloud de manière à obtenir des compromis satisfaisants entre consommation énergétique et qualité de service.

**Meta/Facebook.** La collaboration s'est articulée autour du développement d'algorithmes de type black-box pour la plateforme opensource NEVERGRAD<sup>4</sup> (avec également des aspects plus théoriques sur l'analyse de ces algorithmes).

**Honda Research Institute.** La collaboration a porté sur la conception d'algorithmes d'optimisation bayésienne pour la résolution de problèmes de haute dimension (pour des applications provenant de la mécanique des structures).

## Référence 2. L'unité développe des produits à destination du monde culturel, économique et social.

Les membres de l'équipe n'ont pas déposé de brevet, de logiciels propriétaires, ni créé de start-up sur la période. Néanmoins, au-delà des collaborations avec le monde industriel décrites ci-dessus, nous mentionnons ici deux éléments de valorisation de nos activités au niveau sociétal : des avancées dans le domaine de la santé, et notre implication dans la formation d'informaticiens.

Concernant les applications dans le domaine de la santé, l'équipe a développé :

- ▶ des recherches en collaboration avec l'Institut Pasteur. Le but était d'expliquer les cas graves d'une maladie en fonction de paramètres du patient et/ou de l'infection. La particularité de ce type de problèmes est la difficulté (le coût élevé) d'avoir des jeux de données importants, car ils nécessitent la mise en œuvre de tests complexes sur les patients. Le projet a permis de développer des méthodes d'optimisation explicables, utilisant de l'apprentissage pour régler les paramètres du problème. Cela a permis notamment un filtrage beaucoup plus efficace des informations pertinentes permettant d'expliquer le degré de sévérité de la maladie. Le projet s'est articulé autour de deux thèses, du travail d'un post-doctorant et d'un stagiaire.
- ▶ des recherches en collaboration avec une équipe de neurologues, au travers d'une thèse co-encadrée avec l'équipe Move du laboratoire. Il s'agissait de concevoir un outil d'aide à la décision pour le pronostic de patients atteints de la maladie de Charcot (évaluation de l'évolution de la maladie, ce qui détermine la prise en charge nécessaire). Le travail a mêlé des techniques de recherche opérationnelle et des techniques d'apprentissage.

Concernant la formation, l'équipe est investie, bien sûr, au niveau de l'enseignement dans les formations de Sorbonne Université et de l'Université de Nanterre, dans le MPRI (master parisien de recherche en informatique), et plus généralement sur l'enseignement de l'informatique et sur son développement. Mentionnons en particulier les éléments suivants :

- ▶ Deux membres de l'équipe ont été directeurs du département de master informatique de Sorbonne Université (un jusqu'en 2017, l'autre sur la période 2018-2020), soit une formation d'environ 800 étudiants, pour 9 puis 10 parcours de formation.
- ▶ Un membre de l'équipe a été sur toute la période directeur d'études pour les étudiants de première et deuxième année de licence MIASHS à Nanterre.
- ▶ Les membres de l'équipe ont assuré la responsabilité de formations : double-licence mathématique-informatique à Sorbonne Université (depuis 2018), troisième année du parcours MIAGE de la licence MIASHS à Nanterre (jusqu'en 2021), master MIAGE en apprentissage à Nanterre (depuis 2021).
- ▶ Un membre de l'équipe propose un entraînement d'étudiants pour la compétition de programmation et algorithmique ICPC/SWERC (2021, 2022).
- ▶ Plusieurs membres de l'équipe ont participé (montage et enseignement à Sorbonne Université) à la formation des enseignants de lycée en informatique, pour l'ouverture de l'option NSI dans les deux dernières années de lycée (intervention sur l'algorithmique, dans 2 blocs de formation, pendant 3 ans, pour une centaine d'enseignants au total).
- ▶ L'équipe a participé à la mise en place, en partenariat avec les mathématiciens, de la formation (Sorbonne Université) à l'option informatique de l'agrégation de mathématiques (montage en 2017, co-responsabilité les deux premières années).

Notons qu'également un membre de l'équipe a été *vice-présidente chargée des questions numériques de la ComUE Université Paris Lumières* jusqu'en septembre 2019.

4. <https://facebookresearch.github.io/nevergrad/>

Référence 3. L'unité partage ses connaissances avec le grand public et intervient dans des débats de société.

Les actions de l'équipe concernent essentiellement la diffusion du savoir et la sensibilisation du jeune public, notamment via les éléments suivants :

- ▶ La rédaction d'articles dans des revues à destination des lycéens ou du grand public. L'équipe a écrit 7 articles dans la revue "Tangente" (hors série 75 sur la Recherche Opérationnelle, 2021), un article dans "La Recherche" (hors-série jeux mathématiques, 2018), un article dans "The conversation" (2017), un article dans "Interstices" (2021), un article dans "Pour la science" (2022), et un article en ligne dans la collection "Images des mathématiques" du CNRS (2018).
- ▶ La participation à la fête de la science. Outre l'animation annuelle d'ateliers sur le campus Pierre et Marie Curie, un membre de l'équipe a été pendant plusieurs années coordinatrice pour le laboratoire de la participation à cette manifestation.
- ▶ Plusieurs doctorants ont participé au forum emploi-math (forum métier au centre des congrès de la Villette), à destination des lycéens et étudiants (tenue d'un stand sur la recherche opérationnelle et l'algorithmique).

Egalement, un membre de l'équipe fait partie du comité d'organisation du colloquium d'informatique de Sorbonne Université (<https://www.lip6.fr/colloquium/>).

## 4 RÉFÉRENCES BIBLIOGRAPHIQUES SIGNIFICATIVES DE RO

- [Angelopoulos et al., 2019] Angelopoulos, S., Dürr, C., and Jin, S. (2019). Best-Of-Two-Worlds Analysis of Online Search. In *36th International Symposium on Theoretical Aspects of Computer Science (STACS 2019)*, volume 126 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 7 :1–7 :17, Berlin, Germany. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
- [Angelopoulos et al., 2020] Angelopoulos, S., Dürr, C., Kamali, S., Jin, S., and Renault, M. (2020). Online Computation with Untrusted Advice. In Vidick, T., editor, *11th Innovations in Theoretical Computer Science Conference (ITCS 2020)*, volume 151 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 52 :1–52 :15, Seattle, United States. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
- [Angelopoulos et al., 2022] Angelopoulos, S., Kamali, S., and Shadkhami, K. (2022). Online Bin Packing with Predictions. In *Thirty-First International Joint Conference on Artificial Intelligence {IJCAI-22}*, volume 36, pages 4574–4580, Vienna, Austria. International Joint Conferences on Artificial Intelligence Organization.
- [Bampis et al., 2022] Bampis, E., Dogeas, K., Kononov, A., Lucarelli, G., and Pascual, F. (2022). Scheduling with Untrusted Predictions. In *Thirty-First International Joint Conference on Artificial Intelligence {IJCAI-22}*, pages 4581–4587, Vienna, Austria. International Joint Conferences on Artificial Intelligence Organization.
- [Bampis et al., 2021b] Bampis, E., Dürr, C., Erlebach, T., Santos de Lima, M., Megow, N., and Schlöter, J. (2021b). Orienting (Hyper)graphs Under Explorable Stochastic Uncertainty. In *29th Annual European Symposium on Algorithms (ESA 2021)*, volume 204 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 10 :1–10 :18, Lisboa, Portugal. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.
- [Bampis et al., 2019] Bampis, E., Escoffier, B., Schewior, K., and Teiller, A. (2019). Online Multistage Subset Maximization Problems. In Bender, M. A., Svensson, O., and Herman, G., editors, *European Symposium on Algorithms (ESA)*, volume 144 of *Leibniz International Proceedings in Informatics (LIPIcs)*, Munich, Germany. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
- [Bampis et al., 2018b] Bampis, E., Kononov, A., Letsios, D., Lucarelli, G., and Sviridenko, M. (2018b). Energy Efficient Scheduling and Routing via Randomized Rounding. *Journal of Scheduling*, 21(1) :35–51.
- [Bendotti et al., 2019a] Bendotti, P., Chrétienne, P., Fouilhoux, P., and Pass-Lanneau, A. (2019a). The Anchor-Robust Project Scheduling Problem. *Operations Research*.
- [Bendotti et al., 2019b] Bendotti, P., Fouilhoux, P., and Rottner, C. (2019b). Sub-Symmetry-Breaking Inequalities for ILP with Structured Symmetry. In *IPCO 2019 - 20th Conference on Integer Programming and Combinatorial Optimization*, volume 11480 of *Lecture Notes in Computer Science*, pages 57–71, Ann Arbor, Michigan, United States. Springer.
- [Biedenkapp et al., 2022] Biedenkapp, A., Dang, N., Krejca, M. S., Hutter, F., and Doerr, C. (2022). Theory-inspired parameter control benchmarks for dynamic algorithm configuration. In *Proc. of Genetic and Evolutionary Computation Conference (GECCO)*, pages 766–775. ACM.
- [Bouquet et al., 2021] Bouquet, V., Delbot, F., and Picouleau, C. (2021). On the vertices belonging to all, some, none minimum dominating set. *Discrete Applied Mathematics*, 288 :9–19. 11 figures.
- [Cohen-Addad et al., 2021a] Cohen-Addad, V., Feldmann, A. E., and Saulpic, D. (2021a). Near-linear Time Approximation Schemes for Clustering in Doubling Metrics. *Journal of the ACM (JACM)*, 68(6) :1–34.
- [Cohen-Addad et al., 2021b] Cohen-Addad, V., Saulpic, D., and Schwiegelshohn, C. (2021b). A new coresets framework for clustering. In *STOC '21 : 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 169–182, Rome ( Virtual ), Italy. ACM.
- [Doerr and Doerr, 2020] Doerr, B. and Doerr, C. (2020). Theory of parameter control mechanisms for discrete black-box optimization : Provable performance gains through dynamic parameter choices. In *Theory of Evolutionary Computation : Recent Developments in Discrete Optimization*, pages 271–321. Springer. Free version available at <https://arxiv.org/abs/1804.05650>.
- [Doerr et al., 2021] Doerr, B., Doerr, C., and Lengler, J. (2021). Self-adjusting mutation rates with provably optimal success rules. *Algorithmica*, 83(10) :3108–3147. Free version available at <https://arxiv.org/abs/1902.02588>.
- [Doerr, 2020] Doerr, C. (2020). Complexity theory for black-box optimization heuristics. In *Theory of Evolutionary Computation : Recent Developments in Discrete Optimization*, pages 133–212. Springer. Free version available at <https://arxiv.org/abs/1801.02037>.

- [Doerr et al., 2018] Doerr, C., Wang, H., Ye, F., van Rijn, S., and Bäck, T. (2018). IOHprofiler : A Benchmarking and Profiling Tool for Iterative Optimization Heuristics. *CoRR*, abs/1810.05281. Available at <http://arxiv.org/abs/1810.05281>. A more up-to-date documentation of IOHprofiler is available at <https://iohprofiler.github.io/>.
- [Durand and Pascual, 2021] Durand, M. and Pascual, F. (2021). Efficiency and equity in the multi organization scheduling problem. *Theoretical Computer Science*, 864 :103–117.
- [Dürr et al., 2018] Dürr, C., Erlebach, T., Megow, N., and Meissner, J. (2018). Scheduling with Explorable Uncertainty. In *9th Innovations in Theoretical Computer Science Conference (ITCS 2018)*, volume 94 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 30 :1–30 :14, Cambridge, United States. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
- [Hanan and Hanzalek, 2020] Hanan, C. C. and Hanzalek, Z. (2020). Grouping tasks to save energy in a cyclic scheduling problem : A complexity study. *European Journal of Operational Research*, 284(2) :445–459.
- [Kostovska et al., 2022] Kostovska, A., Jankovic, A., Vermetten, D., de Nobel, J., Wang, H., Eftimov, T., and Doerr, C. (2022). Per-run algorithm selection with warm-starting using trajectory-based features. In *Parallel Problem Solving from Nature (PPSN)*, volume 13398 of *LNCS*, pages 46–60. Springer. Free version available at <https://arxiv.org/abs/2204.09483>.
- [Mallem et al., 2022] Mallem, M., Hanan, C. C., and Munier Kordon, A. (2022). Parameterized complexity of a parallel machine scheduling problem. In *International Symposium on Parameterized and Exact Computation (IPEC)*, Postdam, Germany.
- [Meunier et al., 2022] Meunier, L., Rakotoarison, H., Wong, P., Rozière, B., Rapin, J., Teytaud, O., Moreau, A., and Doerr, C. (2022). Black-box optimization revisited : Improving algorithm selection wizards through massive benchmarking. *IEEE Trans. Evol. Comput.*, 26(3) :490–500. Free version available at <https://arxiv.org/abs/2010.04542>.
- [Pascual et al., 2018] Pascual, F., Rzadca, K., and Skowron, P. (2018). Collective Schedules : Scheduling Meets Computational Social Choice. In *Seventeenth International Conference on Autonomous Agents and Multiagent Systems*, Stockholm, Sweden.
- [Tournaire et al., 2019] Tournaire, T., Castel-Taleb, H., Hyon, E., and Hoche, T. (2019). Generating optimal thresholds in a hysteresis queue : application to a cloud model. In *MASCOTS 2019 : 27th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems*, 2019 IEEE 27th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS), pages 283–294, Rennes, France. IEEE Computer Society.

## A ANNEXE — MEMBRES PERMANENTS AU 31/12/2022

La table ci dessous liste les membres permanents de l'équipe RO.

| NOM          | Prénom    | Corps     | Employeur                 |
|--------------|-----------|-----------|---------------------------|
| ANGELOPOULOS | Spyros    | CR (HDR)  | CNRS                      |
| BAMPIS       | Evipidis  | PR        | Sorbonne Université       |
| BELLITTO     | Thomas    | MCF       | Sorbonne Université       |
| DELBOT       | François  | MCF       | Université paris-Nanterre |
| DOERR        | Carola    | DR        | CNRS                      |
| DÜRR         | Christoph | DR        | CNRS                      |
| ESCOFFIER    | Bruno     | PR        | Sorbonne Université       |
| HANEN        | Claire    | PR        | Université paris-Nanterre |
| HYON         | Emmanuel  | MCF       | Université paris-Nanterre |
| PASCUAL      | Fanny     | MCF (HDR) | Sorbonne Université       |

## ÉLÉMENT DE PORTFOLIO 01



### Distinction

#### 1 DÉFINITION DE CET ÉLÉMENT

**Titre de l'élément :** Médaille de bronze 2022 du CNRS décernée à Carola Doerr.

**URL de l'élément :** <https://lip6.fr/medaille-bronze-doerr>

#### 2 MOTIVATIONS DU CHOIX DE CET ÉLÉMENT

Carola Doerr pilote au sein de l'équipe RO la recherche dans le domaine de l'optimisation black-box. Cette distinction vient récompenser une activité de recherche remarquable, non seulement par la qualité des résultats scientifiques (c.f. les différents prix obtenus, listés dans le document principal), mais aussi sur d'autres plans comme l'animation de la communauté internationale travaillant sur cette thématique, le développement de plusieurs partenariats aussi bien académiques qu'industriels, ou le succès dans de nombreux appels à projets de recherche.

Ceci participe au rayonnement de l'équipe, et plus globalement du laboratoire, avec une forte visibilité dans le domaine de l'optimisation black-box.

#### 3 PRÉSENTATION DE CET ÉLÉMENT

Les activités de l'équipe en optimisation black-box sont décrites dans le document principal de l'équipe. Pour la médaille de bronze, nous renvoyons le lecteur vers l'interview de Carola Doerr par le CNRS<sup>1</sup> : <https://lip6.fr/medaille-bronze-doerr>.

#### 4 RÉFÉRENCES BIBLIOGRAPHIQUES

---

1. Nous passons par un raccourci car l'URL est très longue

## ÉLÉMENT DE PORTFOLIO 02



# Projet ou collaboration

## 1 DÉFINITION DE CET ÉLÉMENT

**Titre de l'élément :** Collaboration industrielle avec EDF

## 2 MOTIVATIONS DU CHOIX DE CET ÉLÉMENT

L'équipe RO poursuit des recherches sur les aspects fondamentaux de l'algorithmique et de l'optimisation (conception et analyse d'algorithmes, complexité, etc.). Cependant, elle cherche également à maintenir une dimension applicative, en développant des liens avec d'autres domaines scientifiques d'une part, et avec des entreprises d'autre part. Parmi ceux-ci, la collaboration avec l'entreprise EDF - et plus spécifiquement EDF R&D - tient une place particulière au sein de l'équipe, de part la durée et la richesse des interactions. Cette collaboration est l'objet de cet élément de portfolio.

## 3 PRÉSENTATION DE CET ÉLÉMENT

La collaboration entre l'équipe RO du LIP6 et EDF R&D a débuté avant la période d'évaluation de ce rapport, au début des années 2010. Elle s'est intensifiée à partir de 2015. Sur la période 2017-2022, elle a été la source de nombreux projets, de deux thèses CIFRE, d'avancées scientifiques notables récompensées par plusieurs prix et distinctions.

Cette collaboration soutenue est grandement facilitée par un fait atypique : Pascale Bendotti, ingénieur-chercheur expert à d'EDF R&D, est membre associée du laboratoire LIP6, et ce depuis 2015. Elle a passé son habilitation à diriger des recherches en 2019 à Sorbonne Université. Ce positionnement permet un échange très régulier, essentiel pour cette collaboration dont nous mentionnons maintenant les éléments principaux.

### 3.1 Première thèse CIFRE

*intitulée "Aspects combinatoires du Unit Commitment Problem", réalisée par Cécile Rottner et encadrée par Pierre Fouilhoux et Pascale Bendotti, soutenue le 14 novembre 2018.*

Le problème Unit Commitment (UCP) est un problème de planification, sur un horizon de temps donné, de l'utilisation d'unités de production électrique. Le plan de production à établir repose sur une estimation de la demande sur la période, sur la capacité de production des unités, et sur un certain nombre de contraintes techniques sur ces unités de production. La difficulté du problème réside dans le fait que les plans de production à établir à chaque pas de temps sont couplés entre eux, par les contraintes de production précédentes mais également par le coût induit par le plan de production (une production stable d'une unité dans le temps génère typiquement un coût moindre qu'une utilisation avec beaucoup d'arrêts/redémarrages). La thèse a porté sur la résolution de ce problème dans le cadre de la production électrique, notamment par des méthodes de programmation linéaire en nombres entiers (PLNE). Le travail de thèse a permis en particulier d'éliminer certaines symétries du problème, débouchant sur des avancées théoriques mais également un gain pratique important par rapport aux méthodes existantes.

Cette thèse a obtenu le *prix de thèse PGMO 2018* (programme Gaspard Monge pour l'optimisation) de la fondation mathématique Jacques Hadamard et le *prix de thèse Paul Caseau 2019* (prix soutenu par l'Institut de France). Un article tiré de ces travaux de thèse a reçu le *prix du meilleur article étudiant* à la conférence nationale ROADEF'19.

### 3.2 Deuxième thèse CIFRE

*intitulée "Anchored solutions in robust combinatorial optimization", réalisée par Adèle Pass-Lanneau et encadrée par Pascale Bendotti, Philippe Chrétienne et Pierre Fouilhoux, soutenue le 16 mars 2021.*



Ce travail de thèse se situe dans le cadre de l'optimisation robuste, où l'on doit décider d'une solution alors que certaines données du problème à résoudre sont soumises à une incertitude - elles ne sont pas connues précisément, pouvant prendre leur valeur dans un ensemble donné (par exemple un intervalle pour une durée). Il s'agit alors classiquement de trouver une solution robuste, en ce sens qu'elle reste réalisable et relativement bonne quelle que soit la valeur effective des données. L'apport principal de la thèse d'Adèle Pass-Lanneau est l'étude de la notion originale d'*ancrage*, introduite lors d'une collaboration antérieure entre EDF et l'équipe RO. Il s'agit de construire une solution dans laquelle on peut garantir qu'un certain ensemble de décisions, appelées décisions ancrées, ne seront pas impactées par l'incertitude sur les données. Dans un cadre opérationnel de planification de projets par exemple, une tâche est ancrée si l'on peut garantir sa date de début, même si certaines autres tâches peuvent engendrer des retards (incertitude sur la durée des tâches). Cette notion d'ancrage a donné lieu à la définition d'un nouveau type de problèmes d'optimisation, dont la thèse a démontré les intérêts théorique et pratique.

Cette thèse a obtenu le *prix de thèse PGMO 2022* de la fondation mathématique Jacques Hadamard. Un article tiré de ces travaux de thèse a reçu le *prix du meilleur article étudiant* à la conférence nationale ROADEF'22. Un article de cette thèse a également obtenu le *prix de meilleur article étudiant* à la conférence PMS 2021.

### 3.3 D'autres collaborations.

D'autres collaborations ont également eu lieu sur la période. On peut notamment citer un projet financé par le programme PGMO en 2021 (avec notamment l'encadrement d'un stage de M2), et un contrat de collaboration en 2022 sur un sujet en lien avec l'ancrage de décisions.

### 3.4 Et pour la suite ?

Une nouvelle thèse CIFRE débute en 2023 en partenariat avec EDF R&D. La thèse porte en grande partie sur des perspectives de recherche relatives à la notion d'ancrage de décisions, ouvertes par la thèse d'Adèle Pass-Lanneau. Le partenariat se poursuit donc, au-delà du départ du principal collaborateur interne, Pierre Fouilhoux, promu professeur à l'Université Sorbonne-Paris-Nord.

## ÉLÉMENT DE PORTFOLIO 03



# Publication

## 1 DÉFINITION DE CET ÉLÉMENT

**Titre de l'élément :** An Adversarial Model for Scheduling with Testing, par Christoph Dürr, Thomas Erlebach, Nicole Megow et Julie Meißner, publié dans la revue *Algorithmica* (82) 3630–3675, 2020.

**URL de l'élément :** <https://arxiv.org/pdf/1709.02592>

## 2 MOTIVATIONS DU CHOIX DE CET ÉLÉMENT

Si vous avez pu avoir l'occasion de travailler dans un projet industriel, alors vous avez sûrement remarqué qu'une grande partie du projet consiste à obtenir les données précises. Travailler avec des estimations grossières ne peut aboutir qu'à des solutions non-optimales. Cette situation a motivé l'étude d'un modèle d'optimisation où les valeurs du problème seraient données sous forme d'un intervalle d'incertitude et l'algorithme a la possibilité d'obtenir la valeur exacte par des requêtes. Les premiers travaux dans ce modèle concernaient le nombre de requêtes minimum pour pouvoir exhiber une solution prouvée optimale. L'article que nous avons sélectionné pour l'évaluation HCERES combine le coût des requêtes avec la valeur objective de la solution produite. Il a nécessité le développement de nouvelles techniques et ouvert la porte à de nouveaux types de problèmes dans le domaine de l'ordonnancement.

## 3 PRÉSENTATION DE CET ÉLÉMENT

Cet élément est le fruit d'une collaboration entre Christoph Dürr, Thomas Erlebach, Nicole Megow et Julie Meißner. Il a d'abord été présenté sous le titre "Scheduling with Explorable Uncertainty" à la conférence *The 9th Innovations in Theoretical Computer Science Conference (ITCS)*, 2018, puis publié dans sa version complète à *Algorithmica*, (82) 3630–3675, 2020.

Pour donner une application à notre modèle, imaginez que vous ayez à envoyer  $n$  fichiers à travers une ligne de communication. Chacun des fichiers  $j$  a une taille  $u_j$ , et le temps de transmission est proportionnel à sa taille. Si vous envoyez les fichiers dans l'ordre croissant des tailles, vous minimisez le temps moyen que les fichiers doivent attendre leur fin de transmission. Maintenant vous avez la possibilité de compresser les fichiers avant leur transmission. Cette compression prend un temps, que nous avons normalisé à 1, pour simplification, et aboutira à une nouvelle taille de fichier  $p_j$ . Donc cette opération n'est intéressante que si  $1 + p_j < u_j$ . Par exemple, pour un fichier déjà compressé, vous auriez  $p_j = u_j$  et une unité de temps serait perdue pour rien. Or, à priori vous n'avez pas la possibilité de savoir si la compression est efficace avant de l'avoir faite (voir Figure 1 pour une illustration).

Un algorithme aura alors à tout moment la possibilité de compresser un fichier, ou de transmettre un fichier. Se posent alors les questions de comment mesurer la performance d'un algorithme et de trouver le meilleur algorithme. Nous adoptons le langage de l'ordonnancement pour ce problème, et appelons *tâches* les fichiers et *tester* l'opération de compression.

Une instance  $I$  du problème consiste en les durées  $u_j$  connues par l'algorithme et les durées  $p_j$ , initialement inconnues. On note  $C_j$  le temps qu'il a fallu depuis le début pour compléter l'exécution de la tâche  $j$ . La valeur objective d'un algorithme  $A$  sur l'instance  $I$  est  $A(I) = \sum C_j$ . Elle est comparée avec  $\text{OPT}(I)$ , qui est la valeur d'une solution optimale pour cette instance. Cette comparaison résulte en un *rapport de compétitivité*.

Une borne inférieure  $x$  à ce rapport est une construction qui montre qu'aucun algorithme ne peut atteindre un rapport plus petit que  $x$ . Une borne supérieure  $x$  est un algorithme avec une preuve qu'il atteint un rapport au plus  $x$ . Les résultats que nous avons obtenus sont indiqués dans le tableau 1. Ils concernent à la fois le cas d'algorithmes déterministes (dans le cas général et dans des cas particuliers) et d'algorithmes randomisés - clairement pour ce type de problème, des choix aléatoires aident pour se protéger des instances pires des cas. La borne inférieure déterministe utilise des instances structurées, où les durées  $u_j$  sont toutes identiques, ne

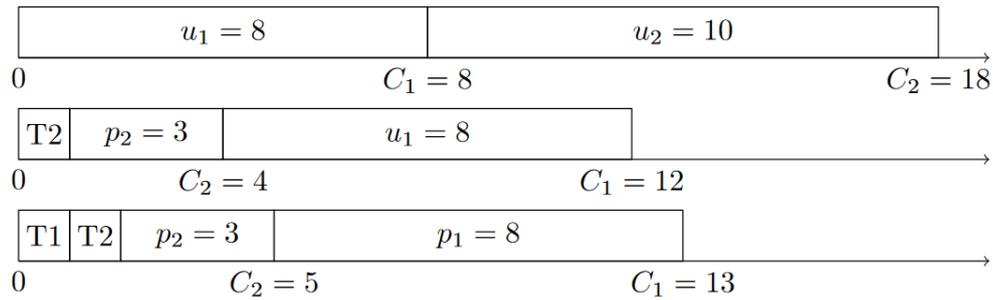


FIGURE 1 – Exemple pour 2 tâches. Le premier ordonnancement exécute les tâches dans l'ordre croissant des durées  $u_j$ , et a un coût  $8 + 18 = 26$ . Le deuxième commence par tester la tâche 2 et apprend la durée  $p_2 = 3$ . Il décide de l'exécuter avant la tâche 1, ce qui engendre le coût  $4 + 12 = 16$ . Le troisième ordonnancement teste les deux tâches et les exécute dans l'ordre croissant des durées  $p_j$ , pour finir avec un coût  $5 + 13 = 18$ . Pour cette instance le deuxième ordonnancement est optimal.

| ratio de compétitivité                 | borne inférieure | bornes supérieures |           |
|--|------------------|--------------------|-----------|
| algorithmes déterministes              | 1.8546           | 2                  | THRESHOLD |
| algorithmes randomisés                 | 1.6257           | 1.7453             | RANDOM    |
| ratio dét. quand $\forall j : u_j = p$ | 1.8546           | 1.9338*            | BEAT      |
| ...et $\forall j : p_j \in \{0, p\}$   | 1.8546           | 1.8668             | UTE       |
| ...et $p \approx 1.989$                | 1.8546           | 1.8552             | UTE       |

TABLE 1 – Nos résultats pour différents variantes du problème. La dernière colonne indique le nom de notre algorithme proposé. \* ratio asymptotique seulement, quand la taille des instances tend vers infini.

fournissant aucune possibilité à l'algorithme de distinguer les tâches, et où les durées après test sont aux extrêmes, soit 0 (meilleur bénéfice du test) soit  $p$  (aucun bénéfice du test).

D'autres travaux ont suivi, et considèrent que le temps d'exécution des tâches n'est pas influencé par le test, et que la valeur  $u_j$  initialement connue, n'est qu'une borne supérieure sur le temps d'exécution. Différents variantes de modèles ont été étudiées par exemple dans [1–3].

## 4 RÉFÉRENCES BIBLIOGRAPHIQUES

[1] Susanne Albers and Alexander Eckl. Scheduling with Testing on Multiple Identical Parallel Machines. In Anna Lubiw, Mohammad Salavatipour, and Meng He, editors, *Algorithms and Data Structures*, volume 12808, pages 29–42. Springer International Publishing, Cham, 2021. Series Title : Lecture Notes in Computer Science.

[2] Fanny Dufossé, Christoph Dürr, Noël Nadal, Denis Trystram, and Óscar C. Vásquez. Scheduling with a processing time oracle. *Applied Mathematical Modelling*, 104 :701–720, April 2022.

[3] Retsev Levi, Thomas Magnanti, and Yaron Shaposhnik. Scheduling with Testing. *Management Science*, 65(2), 2018.

## ÉLÉMENT DE PORTFOLIO 04



# Publication

## 1 DÉFINITION DE CET ÉLÉMENT

**Titre de l'élément :** Scheduling with Untrusted Predictions, par Evripidis Bampis, Kostantinos Dogeas, Alexander V. Kononov, Giorgio Lucarelli et Fanny Pascual, présenté à la conférence IJCAI 2022 : 4581-4587.

**URL de l'élément :** <https://www.ijcai.org/proceedings/2022/636>

## 2 MOTIVATIONS DU CHOIX DE CET ÉLÉMENT

Au cours des dernières décennies, la conception et l'analyse des algorithmes se sont basées sur l'analyse du pire cas, où la performance d'un algorithme est caractérisée par sa pire performance sur toutes les instances d'une taille donnée. Cette approche a donné lieu au développement d'algorithmes élégants et mathématiquement cohérents pour de nombreux problèmes de calcul et a permis le développement de la théorie de la complexité. L'analyse du pire des cas a été utilisée pour mesurer le temps de calcul ou la qualité de la solution des algorithmes.

Cependant, pour de nombreux problèmes, l'analyse du pire cas ne permet pas de prédire les performances des algorithmes en pratique, les exemples les plus célèbres étant la méthode du simplexe pour la programmation linéaire qui prend un temps exponentiel dans le pire des cas, mais qui est très efficace en pratique, ou le problème du remplacement du cache où l'analyse du pire cas n'est pas capable de différencier les algorithmes. Cet échec est dû à des hypothèses trop pessimistes concernant l'incertitude sur l'entrée. En effet, dans de nombreux scénarios, l'entrée est loin d'être le pire cas et présente certaines caractéristiques prévisibles. La disponibilité et l'accessibilité des données pour de nombreuses applications, ainsi que les progrès considérables de l'apprentissage automatique et de l'intelligence artificielle au cours des dernières décennies, permettent d'espérer de bonnes prédictions de l'entrée. Cependant, il n'existe aucune garantie formelle sur la qualité d'un prédicteur (d'apprentissage automatique).

Depuis 2018, une nouvelle thématique a émergé dans le domaine de l'algorithmique, l'algorithmique avec prédictions. Elle vise à exploiter les données pour traiter l'incertitude (d'entrée) afin de concevoir des algorithmes qui, lorsque la précision des prédictions est bonne, offrent une performance (garantie) proche de celle d'un algorithme qui a un accès complet à l'entrée, et en même temps, lorsque les prédictions sont fausses, offrent une performance qui n'est pas beaucoup plus mauvaise que celle d'un algorithme sans accès aux prédictions. Cette nouvelle thématique a déjà donné lieu à des nombreux travaux (voir le site [3] et l'état de l'art récent [6]). C'est dans cette thématique que ce travail s'inscrit.

## 3 PRÉSENTATION DE CET ÉLÉMENT

Dans ce travail, nous considérons un problème d'ordonnancement non-clairvoyant sur une ou plusieurs machines. On nous donne un ensemble de tâches dont les durées réelles ne sont pas connues à l'avance. Cependant, une valeur prédite de la durée est donnée pour chaque tâche. Chaque tâche est également caractérisée par une date d'arrivée. Dans cet article, la préemption et la migration sont autorisées sans coût supplémentaire.

En d'autres termes, les tâches peuvent être interrompues et poursuivre leur exécution à un moment ultérieur (préemption) et éventuellement sur une autre machine (migration). Chaque machine peut exécuter au maximum une tâche à la fois. Dans la première version, nous considérons l'ordonnancement sur une seule machine. Les tâches arrivent au fil du temps, et l'algorithme n'a aucune connaissance préalable de l'existence d'une tâche. Dans la deuxième version, nous considérons l'ordonnancement sur plusieurs machines identiques. Dans ce cas, nous supposons que tous les temps d'arrivées de tâches sont nuls et que l'algorithme connaît le nombre total de tâches ainsi que la prédiction de leurs durées.

L'objectif dans les deux problèmes est de minimiser la somme des temps de complétude. Nous proposons des algorithmes qui sont à la fois cohérents (optimaux quand les prédictions sont correctes) et robustes (de "bonne qualité" lorsque les prédictions sont mauvaises). Ce travail a été motivé par deux questions ouvertes par Purohit

et al. (NeurIPS 2018) dans leur papier fondateur de l'étude de l'ordonnancement non-clairvoyant dans le cadre de l'algorithmique avec prédictions.

L'article dont il est question dans cet élément de portfolio [2] est référencé dans le site qui rassemble les papiers du domaine (voir [3]) et, bien que très récent, il commence à être cité (voir par exemple [1, 4, 5]).

## 4 RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Eric Balkanski, Vasilis Gkatzelis, and Xizhi Tan. Strategyproof Scheduling with Predictions. In Yael Tauman Kalai, editor, *14th Innovations in Theoretical Computer Science Conference (ITCS 2023)*, volume 251 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 11 :1–11 :22, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.
- [2] Evripidis Bampis, Konstantinos Dogeas, Alexander V. Kononov, Giorgio Lucarelli, and Fanny Pascual. Scheduling with untrusted predictions. In Luc De Raedt, editor, *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022*, pages 4581–4587. ijcai.org, 2022.
- [3] ALPS contributors. Algorithms with predictions paper-tracker. 2023.
- [4] Alexandra Lassota, Alexander Lindermayr, Nicole Megow, and Jens Schlöter. Minimalistic predictions to schedule jobs with online precedence constraints. *CoRR*, abs/2301.12863, 2023.
- [5] Russell Lee, Bo Sun, John C. S. Lui, and Mohammad H. Hajiesmaili. Pareto-optimal learning-augmented algorithms for online k-search problems. *CoRR*, abs/2211.06567, 2022.
- [6] Michael Mitzenmacher and Sergei Vassilvitskii. Algorithms with predictions. In *Beyond Worst Case Analysis*, pages 646–662. Cambridge University Press, 2021.

## ÉLÉMENT DE PORTFOLIO 05



# Publication

## 1 DÉFINITION DE CET ÉLÉMENT

**Titre de l'élément :** A new coresets framework for clustering, par Vincent Cohen-Addad, David Saulpic et Chris Schwiegelshohn, article présenté à la conférence *Symposium on Theory of Computing (STOC)* 2021.

**URL de l'élément :** <https://hal.sorbonne-universite.fr/hal-03505350>

## 2 MOTIVATIONS DU CHOIX DE CET ÉLÉMENT

L'étude des problèmes de clustering a connu ces dernières années un regain d'intérêt très fort par la communauté de l'informatique théorique, de part notamment son importance en intelligence artificielle. Une approche féconde dans la résolution de ces problèmes consiste à construire des *coresets*. Intuitivement, cela consiste à remplacer l'ensemble de points initiaux (que l'on doit partitionner) par un ensemble de points, idéalement beaucoup plus petit, équivalent en terme de résolution (cf le prochain paragraphe pour une définition précise). L'obtention de *coresets* petits permet donc de condenser la donnée du problème - et d'en améliorer la résolution.

Dans cet article, nous présentons une nouvelle technique générique pour construire des *coresets*. Cette technique tranche avec les précédentes, pour plusieurs raisons. D'abord, elle permet de retrouver, et même d'améliorer, les précédents résultats connus, de façon particulièrement simple et concise. Ensuite, en terme de techniques, notre algorithme de construction commence par partitionner l'entrée en groupes qui ont des propriétés fortes, que l'on exploite ensuite. Les précédentes analyses de l'état de l'art [1, 3, 5, 8, 9] n'essayaient pas de tirer profit d'une quelconque structure. Notre façon de partitionner l'entrée a ensuite essaimée, et plusieurs nouveaux résultats s'en inspirent [2, 6, 7]. Nous appliquons notre technique pour calculer des *coresets* pour les problèmes de clustering  $k$ -median et  $k$ -means quand l'espace métrique d'entrée vérifie différentes propriétés : par exemple, quand c'est un espace euclidien  $(\mathbb{R}^d, \ell_2)$ , quand la métrique est induite par un graphe de *treewidth* bornée ou un graphe provenant d'une famille excluant un mineur.

Par ailleurs, nous avons montré par la suite [4] que certains des résultats obtenus dans l'article sont optimaux : en plus d'être assez générale et de s'appliquer à de nombreux espaces métriques différents, notre technique est donc aussi très précise.

## 3 PRÉSENTATION DE CET ÉLÉMENT

Cet article présente donc une construction de *coresets* pour les problèmes de clustering  $k$ -median et  $k$ -means. Pour simplifier, concentrons-nous sur  $k$ -median dans l'espace  $(\mathbb{R}^d, \ell_2)$ . Un *coreset* est une façon de compresser l'entrée tout en préservant la fonction de coût de  $k$ -median : plus précisément, étant donné  $\varepsilon > 0$ , un  $\varepsilon$ -coreset pour  $P$  est un ensemble  $\Omega$  tel que, pour tout ensemble de  $k$  centres  $S$ ,

$$\text{cost}(\Omega, S) = (1 \pm \varepsilon)\text{cost}(P, S).$$

Pour contruire un tel  $\Omega$ , la technique la plus répandue est de construire une distribution (bien choisie) de probabilité  $\pi$  sur  $P$ , de tirer l'ensemble  $\Omega$  selon cette distribution et d'utiliser des bornes de concentrations pour montrer qu'en choisissant  $\Omega$  suffisamment grand, on obtient la garantie de *coreset*. La distribution communément analysée est appelée *sensitivity sampling* ou *importance sampling* : la probabilité de tirer un point est proportionnel à sa sensibilité, qui est sa contribution relative maximum dans n'importe quelle solution  $S$ .

Cette distribution est complexe à approximer et à analyser. Nous proposons à la place de décomposer l'entrée en groupes, tel que dans chaque groupe les points sont essentiellement équivalents : utiliser une distribution uniforme dans chaque groupe permet de construire un *coreset* (essentiellement). Cela permet de simplifier grandement les outils conceptuels utilisés par notre analyse, mais aussi de construire des *coresets* de taille optimale, comme mentionné précédemment.

## 4 RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Daniel Baker, Vladimir Braverman, Lingxiao Huang, Shaofeng H. C. Jiang, Robert Krauthgamer, and Xuan Wu. Coresets for clustering in graphs of bounded treewidth, 2020.
- [2] Vladimir Braverman, Vincent Cohen-Addad, Shaofeng H.-C. Jiang, Robert Krauthgamer, Chris Schwiegelshohn, Mads Bech Tofttrup, and Xuan Wu. The power of uniform sampling for coresets. In *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS*. IEEE, 2022.
- [3] Vladimir Braverman, Shaofeng H.-C. Jiang, Robert Krauthgamer, and Xuan Wu. Coresets for clustering in excluded-minor graphs and beyond. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA*. SIAM, 2021.
- [4] Vincent Cohen-Addad, Kasper Green Larsen, David Saulpic, and Chris Schwiegelshohn. Towards optimal lower bounds for k-median and k-means coresets. In *Proceedings of the 54th ACM Symposium on Theory of Computing STOC*, 2022.
- [5] Dan Feldman and Michael Langberg. A unified framework for approximating and clustering data. In *Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011*,, 2011.
- [6] Lingxiao Huang, Shaofeng H.-C. Jiang, Jian Li, and Xuan Wu. Coresets for clustering with general assignment constraints. *CoRR*, 2023.
- [7] Lingxiao Huang, Jian Li, and Xuan Wu. Towards optimal coreset construction for  $(k, z)$ -clustering : Breaking the quadratic dependency on  $k$ . *CoRR*, abs/2211.11923, 2022.
- [8] Lingxiao Huang and Nisheeth K. Vishnoi. Coresets for clustering in euclidean spaces : importance sampling is nearly optimal. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC*, 2020.
- [9] Christian Sohler and David P. Woodruff. Strong coresets for k-median and subspace approximation : Goodbye dimension. In *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS*, 2018.