

# Résultats structurels pour les modèles de contrôle décentralisés coopératifs<sup>★</sup>

Jilles Dibangoye<sup>1</sup>, Olivier Buffet<sup>2</sup>, Olivier Simonin<sup>1</sup>

<sup>1</sup> INSA Lyon / INRIA — CITI-Inria Lab., Lyon  
jilles-steeve.dibangoye@insa-lyon.fr — <http://jilles.dibangoye.net/>  
olivier.simonin@insa-lyon.fr — <http://perso.citi.insa-lyon.fr/osimonin/>

<sup>2</sup> INRIA / Université de Lorraine / CNRS — LORIA, Nancy  
olivier.buffet@loria.fr — <http://www.loria.fr/~buffet/>

Les processus de décision markoviens partiellement observables décentralisés (Dec-POMDP) ont émergé comme le cadre standard pour la prise de décision séquentielle décentralisée [Bernstein \*et al.\* \(2002\)](#). Ce modèle général considère plusieurs agents avec des observations différentes qui coopèrent pour atteindre un objectif commun, mais ne peuvent pas communiquer entre eux. Malheureusement, sa complexité au pire cas a limité son applicabilité. Le cas à horizon fini est NEXP [Bernstein \*et al.\* \(2002\)](#).

Pour permettre une plus grande capacité de mise à l'échelle, beaucoup d'attention a été dévouée à des modèles avec des hypothèses restrictives concernant essentiellement la dynamique et les récompenses [Goldman & Zilberstein \(2004\)](#); [Becker \*et al.\* \(2004\)](#); [Nair \*et al.\* \(2005\)](#); [Melo & Veloso \(2011\)](#).

Ce travail propose une méthodologie générale, l'*analyse structurelle*, qui permet de concevoir, sans perte d'optimalité, des politiques et fonctions de valeur concises. Nous prouvons pour la première fois que les besoins en mémoire pour les politiques et les fonctions de valeurs peuvent être assymétriques, d'où des gains significatifs dans certains cas. Un autre résultat nouveau et important est la preuve que, sous de légères conditions sur les récompenses, la fonction de valeur optimale consiste en des fonctions linéaires des états cachés. Pour permettre une plus grande applicabilité, cette analyse structurelle repose sur une approche de résolution récente des Dec-POMDP qui passe par une ré-écriture sous la forme d'un MDP déterministe à espace d'état continu dont la fonction de valeur optimale est convexe et linéaire par morceaux [Dibangoye \*et al.\* \(2013\)](#).

Dans l'ensemble, il semble que l'analyse de complexité asymptotique fournisse une hiérarchie des problèmes, alors que l'analyse structurelle permette de guider la caractérisation de politiques et fonctions de valeurs (optimales) concises, ce qui peut conduire à développer une théorie et des algorithmes capables de passer à l'échelle, applicables et largement adaptables.

## Références

- BECKER R., ZILBERSTEIN S., LESSER V. R. & GOLDMAN C. V. (2004). Solving transition independent decentralized Markov decision processes. *JAIR*, **22**, 423–455.
- BERNSTEIN D. S., GIVAN R., IMMERMANN N. & ZILBERSTEIN S. (2002). The complexity of decentralized control of Markov decision processes. *Mathematics of Operations Research*, **27**(4).
- DIBANGOYE J. S., AMATO C., BUFFET O. & CHARPILLET F. (2013). Optimally solving Dec-POMDPs as continuous-state MDPs. In *IJCAI'13*.
- GOLDMAN C. V. & ZILBERSTEIN S. (2004). Decentralized control of cooperative systems : Categorization and complexity analysis. *JAIR*, **22**(1), 143–174.
- MELO F. S. & VELOSO M. M. (2011). Decentralized MDPs with sparse interactions. *AI journal*, **175**(11), 1757–1789.
- NAIR R., VARAKANTHAM P., TAMBE M. & YOKOO M. (2005). Networked distributed POMDPs : A synthesis of distributed constraint optimization and POMDPs. In *AAAI'05*.

---

★. Ceci est un résumé d'un article à paraître à IJCAI 2015.