# Abstract: Multi-Objective Decision-Theoretic Planning

**Diederik M. Roijers** (University of Oxford; diederik.roijers@cs.ox.ac.uk)

DOI: 10.1145/3008665.3008670

Decision making is hard. It often requires reasoning about uncertain environments, partial observability and action spaces that are too large to enumerate. In such tasks decision-theoretic agents can often assist us. In most research on decision-theoretic agents, the desirability of actions and their effects is codified in a scalar reward function. However, many real-world decision problems have multiple objectives. In such cases the problem is more naturally expressed using a vector-valued reward function, leading to a *multi-objective decision problem (MODP)*.

Typically, MODPs cannot be *scalarized* to a single-objective decision problem, at it is very hard to *a priori* specificy a so-called *scalarization function* that captures the user utility for every value-vector imaginable. Instead, we provide decision support (schematically depicted in Fig. 1). In the planning phase, our algorithm produces a *coverage set (CS)*, i.e., a set of policies that covers all possible preferences between the objectives. In the selection phase, the user selects one policy from the CS. Finally this selected policy is executed.

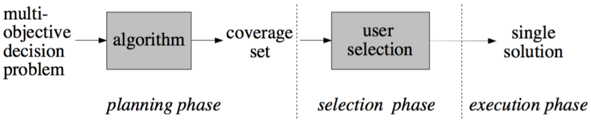

Figure 1: The decision support scenario.

We focus on decision-theoretic planning algorithms that produce a *convex coverage set (CCS)*, which is the optimal solution set when either: 1) the user utility can be expressed as a weighted sum over the values for each objective; or 2) policies can be stochastic.

We propose new methods based on two approaches to creating planning algorithms that produce an (approximate) CCS by building on existing single-objective algorithms. In the inner loop approach, we replace the summations and maximizations in the inner most loops of single-objective algorithms by cross-sums and pruning operations. In the outer loop approach, we solve a multi-objective problem as a series of scalarized, i.e., single-objective, problems.

One of our most important contributions is *optimistic linear support (OLS)* (Roijers, Whiteson, & Oliehoek, 2015a). OLS is a generic outer loop framework for multi-objective decision problems that uses single-objective solvers as subroutines. It can be applied to any MODP for which a corresponding single-objective method exists. We show that, contrary to existing methods, each intermediate result is a bounded approximation of the CCS with known bounds (even when the single-objective method used is a bounded approximate method as well) and is guaranteed to terminate in a finite number of iterations.

## Multi-Objective Coordination

The first MODP we tackle is *multi-objective coordination graphs (MO-CoGs)*. MO-CoGs are cooperative single-shot, fully observable, multi-agent decision problems. In MO-CoGs, agents must coordinate in order to find effective policies. Key to making coordination between agents efficient is exploiting loose couplings, i.e., each agents actions directly affect only a subset of the other agents. Such loose couplings are expressed by a (vector-valued) payoff function, that decomposes into a sum over local payoff functions in which only subsets of the agents participate.

We propose and compare inner loop methods and OLS-based methods. Specifically, we build upon *variable elimination (VE)* (Guestrin, Koller, & Parr, 2002) and propose *convex multi-objective variable elimination (CMOVE)* (inner loop) and *variable elimination linear support (VELS)* (OLS-based). We build on *AND/OR tree search* (Mateescu & Dechter, 2005) to propose *convex AND/OR tree search (CTS)* (inner loop) and *AND/OR tree search linear support (TSLS)* (OLS-based). We show that OLS-based methods scale better in the number of agents, both in terms of runtime and memory, while inner loop methods scale better in the number of objectives. We show experimentally that we can produce $\varepsilon$-CCSs in a fraction of the runtime that is required to produce an exact CCS.

Furthermore, we propose *variational optimistic linear support (VOLS)* (Roijers, Whiteson, Ihler, & Oliehoek, 2015) — an OLS-based method that builds on variational methods. The runtime of variational subroutines (Liu & Ihler, 2011) is not exponential in the number of agents. However, they produce only $\varepsilon$-approximate solutions. VOLS inherits the runtime and quality guarantees and can produce an $\varepsilon$-CCSs in sub-exponential runtime. We show that it is possible to reuse the reparameterized graphs produced by single-objective variational subroutines to hot-start the variational subroutines in later iterations of OLS, leading to significant speed-ups.

## Sequential Planning

The next problem settings we tackle are *multi-objective Markov decision processes (MOMDPs)* and *multi-objective partially observable Markov decision processes (MOPOMDPs)* which are single-agent sequential decision problems. Because the sequence of actions that result from executing policies in these problems affect the environment, agents have to consider both immediate and future rewards that depend on the future state of the environment.

MOMDPs are fully-observable, i.e., the agent knows at any time what the exact state of the environment is. A major challenge in MOMDPs is the size of the state and action spaces. We illustrate, using a large MOMDP called the *maintenance planning problem* (Roijers et al., 2014), that it is possible to create efficient methods using OLS and specialised single-objective subroutines, and that it is relatively easy to replace these subroutines when the state-of-the-art for the single-objective method improves.

MOPOMDPs are partially observable, which poses an important additional challenge. We propose *optimistic linear support with alpha reuse (OLSAR)* (Roijers, Whiteson, & Oliehoek, 2015b), which as far was we are aware, this is the first MOPOMDP planning method that computes the CCS and reasonably scales in the number of states of the MOPOMDP. We show how to represent the value function of MOPOMDPs in terms of $\alpha$-matrices and propose a single-objective subroutine for OLSAR called OLS-compliant Perseus (based on (Spaan & Vlassis, 2005)) that returns these $\alpha$-matrices. A key insight underlying OLSAR is that the $\alpha$-matrices produced by OCPerseus can be reused in subsequent calls to OCPerseus, greatly reducing the runtime. Our experimental results show that OLSAR greatly outperforms alternatives that do not use OLS and/or $\alpha$-matrix reuse.

## References

Guestrin, C., Koller, D., & Parr, R. (2002). Multiagent planning with factored MDPs. In *NIPS*.

Liu, Q., & Ihler, A. T. (2011). Bounding the partition function using Hölder's inequality. In *ICML* (pp. 849–856).

Mateescu, R., & Dechter, R. (2005). The relationship between AND/OR search and variable elimination. *UAI* (pp. 380–387).

Roijers, D., Scharpff, J., Spaan, M., Oliehoek, F., de Weerdt, M., & Whiteson, S. (2014). Bounded approximations for linear multi-objective planning under uncertainty. In *ICAPS* (pp. 262–270).

Roijers, D. M., Whiteson, S., Ihler, A. T., & Oliehoek, F. A. (2015). Variational multi-objective coordination. In *Malic*.

Roijers, D. M., Whiteson, S., & Oliehoek, F. A. (2015a). Computing convex coverage sets for faster multi-objective coordination. *JAIR*, *52*, 399–443.

Roijers, D. M., Whiteson, S., & Oliehoek, F. A. (2015b). Point-based planning for multi-objective POMDPs. In *IJCAI*.

Spaan, M. T. J., & Vlassis, N. (2005). Perseus: Randomized point-based value iteration for POMDPs. *JAIR*, 195–220.

**Diederik M. Roijers** did his PhD on multi-objective decision theory at the University of Amsterdam under the supervision of Shimon Whiteson and Frans A. Oliehoek. He is currently a postdoctoral researcher at the Dep. of Computer Science at the University of Oxford.