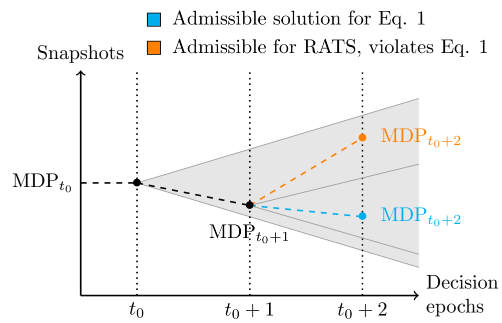
We thank the reviewers for their thoughtful reviews and comments. We intend to include the answers below in the paper.

Clarifying statements. Thank you (reviewer #4) for pointing out these weaknesses that can easily be fixed. We believe we attempted to lay bridges between various fields (tree search, robust MDPs, non-stationary planning...) which could explain the difference in clarity appreciation between reviewers. We elaborate on specific clarifications below.

When does Property 1 hold? There may be a misunderstanding here; Property 1 is not an assumption. It is a direct consequence of the LC-NSMDP definition and thus always holds.

Section 5's relaxation vs. full problem. In the full problem (Equation 1), MDP_t belongs to a set of MDPs defined (constrained) by MDP_{t-1}. This is a stronger requirement than the one used by RATS, which only constrains MDP_t to belong to the "cone" of MDPs originating from MDP_{t_0} (see figure). There currently is no generic bound on the optimality gap between the relaxation and the full problem. Although this is an interesting problem, this relaxation was only introduced to allow the bottom-up minimax method and we believe further research should rather investigate alternative robust algorithms that lift this relaxation.



Wasserstein metric (WM) vs. total variation (TV). Indeed, TV is a legitimate measure in *non-metric* state spaces. However, many discrete state spaces (as the ones used in the experiments) still exhibit a metric between states and WM computes distances that depend on that metric while TV does not. Not taking that metric into account would yield worst case snapshots with little physical meaning (think of a path planning task where the worst case snapshot transforms the outcome distribution of a "turn right" transition from a street in Montreal to a street in Vancouver).

On the experimental section. The purpose of this work was really to lay ground for a principled approach to NSMDP planning. We believe the main contribution of this section is the comparison between robust and non-robust policies on a variety of demonstrative scenarios. Although it is an important area of future work, the goal to scale up seemed like a drift from this main goal. On that topic, we argue that RATS aims at the same result as Robust Dynamic Programming (RDP, Iyengar 2005) and, for the benchmarks reported, yields the same policy. Furthermore, the comparison with other related approaches (RDP seeming the most appropriate, because non-stationary planning assumes full model knowledge) would be relevant in a context where one wants to scale to larger problems, including those where function approximation is needed. In particular, such a comparison should highlight that RDP does DP —it is offline and plagued by the curse of dimensionality— while RATS does tree search from the current state and snapshot —it is online and can take advantage of heuristic search. Scaling up, as in many game-theoretic approaches, likely requires a combination of online search, function approximation and relevant heuristics, and we believe this topic deserves another paper. We agree nonetheless that the question comes naturally and intend to add the present paragraph to the experimental section.

Extension to continuous time processes. The extension of our work to *deterministic* continuous durations is straightforward. We chose to present our work in the case where decision epochs and epoch indices coincide for clarity but will include an additional paragraph in the appendix to lift this ambiguity. Considering *stochastic* continuous durations can be done in a number of ways. Semi-MDPs (and their extensions, *e.g.* generalized SMDPs) assume such durations but in a time-independent, stationary setting. Having both time-dependency (non-stationarity) and stochastic continuous durations introduces an additional layer of complexity and the extension of our results in that case might be feasible but seems non-trivial.

Bounded parameters MDPs. We omitted the reference because it is a precursor of the Robust MDP literature, but we definitely acknowledge the link with the Δ_t set (and the rectangularity assumption in RDP) so will try to include it back.

Bounded evolution vs. bounded change rates. Bounding the absolute evolution is an interesting extension that sits between our model and Robust MDPs. We could include it in Equation 6 by adding a *constant* upper bound on the distance to MDP_{t_0}. Assuming bounded evolution amounts to say "the model cannot evolve far from the current snapshot" while bounded change rate says "the model can evolve arbitrarily far from the current snapshot but slowly". In practice, both assumptions make sense for real world instances. Thank you for this very relevant comment, this will be a valuable addition in the paper.

Zero-shot planning. We agree this terminology is confusing (we meant "without data acquisition" as often in zero-shot learning) and intend to remove it altogether.

Practical problem instances. This research was inspired by path planning problems in autonomous glider soaring and road traffic management. These examples are mentioned in Sections 2 and 4 but could be present since Section 1.