To whom it may concern,

We would like to submit our manuscript titled "Model-based Shared Control of Data-Driven Human-Machine Systems" for consideration in the International Journal of Robotics Research. We previously submitted an earlier draft of this work as an invited paper to the IJRR RSS 2017 Special Issue. The previous submission was titled "Koopman-Based Assistance for Dynamically Shared Control of Human-Machine Systems from User Demonstration". In this new manuscript we have made profound changes to address concerns brought up during the prior review process. In particular, we have significantly modified our presentation of the theory and our interpretation of the results of our experiments. Below, we describe the changes we made in detail and how we have addressed the reviewers comments. We believe that the revisions greatly improve the quality and clarity of the paper, and hope that the improvements will be clear to the reader.

Thank you in advance for your time and consideration,

Alexander Broad, Ian Abraham, Todd Murphey, and Brenna Argall

1 Reviewer 1

In this section we respond to all of the *major comments* brought up by Reviewer 1. We also note that we have incorporated all *minor comments* in the new draft of our paper.

Reviewer 1 Comment 1

As a contribution, it is written in the paper "Development Koopman operator model-based optimal control for general, nonlinear systems identification and autonomous policy generation." However this is not really a contribution of this paper, since it applies, but does not develop Koopman operator model-based optimal control.

Response

As a general comment, we would like to mention that we have taken the viewpoint of each reviewer into consideration and therefore a common theme in our responses is that we have significantly revised the language in the paper. Importantly this includes a re-evaluation of our interpretation of the results of our human subjects studies and, relatedly, how we discuss the theory behind our methods. In response to this specific comment, we have removed the claim and have re-framed our contribution.

Reviewer 1 Comment 2

The paper lists other "viable model learning techniques": Neural Networks, Gaussian Processes and Gaussian Mixture Models. However, there is no discussion on the advantages/disadvantages of Koopman operators in comparison with other techniques.

Response

We have added a comment to briefly address this point. While we do not believe the choice of machine learning algorithm used to model the system and control dynamics would significantly impact the results of the human subjects studies, we do agree that it is important to mention the practical reasons motivating our choice.

Reviewer 1 Comment 3

"This is a particularly important aspect of our approach as the resulting shared control methodology is generalizable to any dynamic system, and specific to each user" –> But the results show repeatedly that there is not statistically significant difference between training with data from an individual, from a group of naive users or from an expert user. Therefore, it does not seem like the control methodology is really specific to each user.

Response

We have removed this comment as it no longer fits with our interpretation of the results, which now matches more closely the comments of all reviewers.

Reviewer 1 Comment 4

Were the running and terminal costs for the lunar lander system specified in the paper?

Response

The specific numbers are not in the manuscript, however, we have now open-sourced our code and made it available online (and the link is provided in the paper). Therefore, these values and all other implementation details of interest are available in the code.

Reviewer 1 Comment 5

"By requiring the user input to remain within a window of deviation from the optimal control, we ensure that the resulting input will not destabilize the system and that the user will remain otherwise unobstructed by their autonomous partner (Of course, the stabilization guarantee only holds if the model is accurate)" –> But how do we know if the model is accurate? Why are there no comparisons in the paper between predictions of the lunar lander state and of the ground truth nor between the predicted user input and the ground truth?

Response

We have removed this description of our control allocation strategy from the text and have additionally included a figure (Fig. 11) to validate the accuracy of the linear and nonlinear models explored in our studies.

Reviewer 1 Comment 6

Wouldn't it make sense to use LQR with a first order Taylor approximation of the system dynamics in order to control the lunar lander? Is the method proposed in the paper better than this simple baseline?

Response

We agree with the reviewer that LQR with a first order Taylor approximation of the system dynamics is another viable model-based control algorithm. We now include a brief discussion of alternative control methodologies, and an intuition for our implementation choice, as in our response to Comment 2 above.

Reviewer 1 Comment 7

"Initially, we hypothesized that the models may learn an individual understanding of the interaction between the system and user, however, our results seem to suggest that the system is adapting to the person, and not the other way around." –> Why would we expect the models to learn an "individual understanding of the interaction between the system and user" if the motor commands are computed using Optimal Control? The predicted user inputs play no role in the system, do they? Why can we say that "the system is adapting to the person"?

Response

We have removed this text from the new manuscript. The aspect of the user that is captured by the modeling process stems from the control input provided during the data collection and subsequent model learning process. During the data collection phase, we capture only the control signal provided by the user (i.e. no input is captured from the autonomy) and therefore the predictive nature of the Koopman operator is learned from the user's response to the state and environment. During shared control the autonomously generated signals are only used to block inputs from the human that are deemed suboptimal, they are never executed by the system itself.

Reviewer 1 Comment 8

"The experimental results of the linear study demonstrate that we are (1) successfully able to learn a model of the system dynamics and user interaction" –> Was this really demonstrated? There was no comparison between predictions of the lander state and ground truth nor of predictions of the user interaction and ground truth.

Response

As mentioned in response to Comment 5, we have now included a figure (Fig. 11) to demonstrate the accuracy of the learned models in predicting the future state of the system.

Reviewer 1 Comment 9

In Figure 6, (b) and (c) show no statistically significant difference between "Individual", "General" and "Expert". The caption says, however, that there is.

Response

We have fixed this error in the new manuscript.

2 Reviewer 2

In this section we address the concerns of Reviewer 2.

Reviewer 2 Comment 1

This paper proposes a new method for learning-based shared autonomy based on a finite-dimensional approximation to the Koopman operator. The authors state that their method learns the joint dynamics of the human-robot system, which would allow them to tailor the learning in a "user-specific" way. The authors present experimental results on a video-game-like simulator, in which it can be seen that using their shared-autonomy framework is comparatively better than giving users direct control of the simulated system; however, no significant differences are found when applying controllers based on models trained on different users.
Unfortunately, the method proposed in this paper does not in fact learn the joint human-robot dynamics as the authors seem to intend. The Koopman-based model is only learning a mapping from state and input to the next state (it is also implicitly learning a prediction of the next user input but this gets discarded). This amounts exactly to learning the physical dynamics of the system f(x,u). Therefore, regardless of what user's data are taken to train the model, the ground truth that is being learned is exactly the same function. (The only difference one could hope to find would be in the distribution of training data differing between users; however, the authors are using data from the practice rounds, in which users control the system directly, without the shared-autonomy interventions, and therefore the data distribution is not from the closed-loop human-automation setting that the model is applied to.)

Response

As mentioned in our response to Reviewer 1, we have taken the comments and viewpoints of each reviewer into consideration and have significantly revised the language in the paper, including our interpretation of the results. With regards to the comments above, we have removed this language from the paper and have clarified our interpretation of the modeling process. To be clear, as the reviewer mentions, the Koopman operator *does* predict the user's commands at each timestep. For this reason, each model does in fact represent information about the system and the human operator that provided data during the training process – namely the person's control response to the state of the system and/or the environment. Additionally, we note that the distribution of data provided by each user is unique as we do not interfere with the user's controls during the data collection process (with the exception of the online learning scenario, which is described in detail in the paper). However, as the reviewer mentions, under our shared control paradigm, only the learned system and control dynamics matrices are used to generate the autonomous input and we therefore re-word our description of the interaction between the human and autonomy in this work.

Reviewer 2 Comment 2

In any case, the physics model seems simple enough that the dynamics are being quite accurately learned in all cases, which as the authors report results in essentially the same learned dynamics across users (the authors consider this a "striking similarity", but there is nothing striking about it, since the ground-truth dynamics are exactly the same in all cases). As a result of this, all resulting optimal controllers will behave in essentially the same way, which also explains the fact that there are no significant differences between shared-autonomy controllers trained on different users.
The above is a critical issue affecting the novelty of this paper's contribution. Since in reality nothing more than the physical dynamical model is being learned from the data, this method fits exactly in the standard pipeline of system identification followed by optimal control on the learned model.

Response

We believe the reviewer's conclusion stems from a miscommunication and therefore again mention that we have made significant changes to the language and presentation of the our manuscript. In particular, we note that the uniqueness in our work stems from the fact that the model is learned from demonstration, and is used in a shared control paradigm for an *a priori* unknown system. This is something that, to the best of our knowledge, has never been researched before and has many potential points of failure. We describe the theory of this idea and provide the results of two separate experimental trials to validate our idea.

Reviewer 2 Comment 3

Further, while the authors state that their approach departs from reinforcement learning (RL), they seem to only be considering model-free RL (direct policy learning). In model-based RL it is standard to learn a dynamics model of the system and subsequently optimize a control policy based on the learned model (this can be done through a wide range of techniques from value iteration to MPC). Therefore the authors do not correctly place their work in context: this work fits perfectly in the model-based RL framework.

Response

We have clarified this point further in the related work section. As part of this effort we have reduced the text in this section to highlight the most closely related work and describe how our technique fits into the Optimal Control, Reinforcement Learning, and Shared Control landscapes.

Reviewer 2 Comment 4

The authors also claim that an advantage of their proposed method is that it does not depend on hand-engineered features, yet their finite approximation of the Koopman operator relies exactly on these hand-engineered features. Using linear weightings of nonlinear state and input features is an extremely common practice in learning (for example, it is a standard form for reward functions in inverse reinforcement learning literature). While there is nothing wrong with this, the advantages of the authors' learning method with respect to any other feature-based approaches is unclear.

Response

We have removed this text from the new manuscript and, as described in our response to Reviewer 1, have incorporated additional information about how our approach relates to other learning methods.

Reviewer 2 Comment 5

Finally, the authors make an argument about the importance of stability in shared-autonomy controllers. However, other than some vague (and in fact incorrect) claims about their ability to give stability guarantees on their proposed control merging scheme, no such guarantees are presented. In fact, the authors propose a simple merging strategy (MDA) that applies a zero control action when the user input disagrees excessively with the computed optimal control action. This strategy can easily be seen both non-stabilizing and unsafe in the lunar lander experiment proposed by the authors: if the user keeps introducing an input in the "wrong direction", the lunar lander will simply not exert any thruster action, which eventually will be guaranteed to make the vehicle spin out of control and/or collide with the ground.

Response

We have re-worked our discussion of the impact of our control allocation algorithm on the stability of the shared control system. In particular, we now focus on the fact that we improve stability without claiming any guarantees of global stability. As the reviewer mentioned, there are scenarios which do not result in a stable shared control system (e.g. in the case of an adversarial human operator who always tries to provide input in the opposite half plane as the optimal control algorithm).

Reviewer 2 Comment 6

Overall, the main issue with this paper is that the learning algorithm is producing an estimate of the physics model alone. Without the claimed learning of the joint human-automation dynamics, the proposed technique is a particular form of feature-based model-based RL applied to shared autonomy. The results, consistently with this, only show that shared-autonomy control can be useful in comparison to giving the user direct control of the physical platform, which is a well-established fact in the shared-autonomy and human-robot interaction literature. All results relating to individualized learning are negative, not because such learning would not potentially be useful but simply because unfortunately it is not taking place in this case.

Response

This overview relates back to much of what we have covered already, so we simply restate the fact that we have made significant changes to the presentation of our work and importantly our interpretation of the results.

3 Associate Editor

Reviewer 3 Comment 1

In light to the reviews and in my own reading the paper, I suggest making clearer the stated contribution and benefits of their method over alternative model-based RL approaches to shared-autonomy control. Additionally showing quantitative comparisons of the Koopman operator learned dynamics to other more common methods and/or the ground truth system dynamics can help the readers to better understand this work in the broader RL and robotics context.

Response

As with Reviewers 1 and 2, we now provide a quick response to the comments provided by the associate editor. We believe the new draft of our paper is significantly more clear with respect to our contributions. We have also made large changes to our presentation of the theory and our interpretation of the results. Finally, we have included an additional figure as a quantitative comparison of the learned dynamics to the ground truth and improve our discussion of how our model-based shared control algorithm fits into the landscape of prior work. In conclusion, we believe the new draft of our manuscript address all concerns brought up during the previous review process. We are very grateful for the feedback as we believe the new draft is significantly improved and more clearly represents our contributions.