Learning Optimal Control

Kiran Kumar Sathyanarayanan   
*Masters Automation and Robotics*  
*Techniche Universität Dortmund*Dortmund, Germany  
line 5: email address or ORCID

Rajesh Pushparaj  
*Masters Automation and Robotics*  
*Techniche Universität Dortmund*Dortmund, Germany  
rajesh.pushparaj@tu-dortmund.de Prathap Amudha Ramesh Babu  
*Masters Automation and Robotics*  
*Techniche Universität Dortmund*Dortmund, Germany  
line 5: email address or ORCID

Sriram Saravanan  
*Masters Automation and Robotics*  
*Techniche Universität Dortmund*line 4: City, Country  
line 5: email address or ORCID

*Abstract*—Double inverted pendulum is one of the most popular unstable and chaotic system, it has two joints and two rods attached to a moving cart. In this project we try to imitate the Model predictive controller with Feedforward Neural Network that makes the double inverted pendulum to stand upright and makes it stay in that position. And we also use Reinforcement learning to make a Neural Network that learns to make the pendulum stand upright using reward and action, then we compare the training complexity and performance of the controllers to determine the best one in terms efficiency.

Keywords—component, formatting, style, styling, insert (key words)

# Introduction (*Heading 1*)

Controlling unstable and chaotic systems has lot of applications in Automation, Robotics and Process Control. One such classic chaotic system is double inverted pendulum, it has two rods with two joints attached to a moving cart. The task is to make the pendulum stand upright without falling, we can model predictive controller(MPC) to achieve this task. Model predictive control is an advanced method of process control which uses the model of the system and initial state to predict the future of the systems using optimization algorithms. In this project we try to imitate the model predictive control by collecting dataset for different initial states and using the MPC to predict the force given to the cart for various state of the pendulum.

In imitation learning we use supervised learning to create a feedforward neural network and train it with datasets generated using the MPC controller. And after training we export the model and simulate the model with the same initial condition given for generating the dataset with MPC. And the exported model predicts the force similar to MPC does.

In the second part of the project we try to achieve the task using reinforcement learning(RL). Where the training of the neural network is done using reward and action. The process take the model as an environment and uses an agent which tries to achieve the desired task and a feedback is given to the agent using a critic which uses reward function to calculate the reward for each episode of training. After training the neural network(NN) is exported similarly as done in imitation learning and simulated to the initial state for which the NN is trained.

Finally we compare the training complexity and performance of the both imitation learning NN and the RL based NN model to determine the best model in terms efficiency for this double inverted pendulum problem.

The process of training NN model for each case is explained in detail in the upcoming sections.

# imitation learning

In imitation learning we use the MPC to control the pendulum making it stand upright and record the states and corresponding force in cart. We create a linear regression model for the recorded data and use it to control the pendulum thus imitating the MPC’s predictions. The process contains the following steps which are explained below

## Data generation

First, we start collecting the datapoints for various initial states of the pendulum using the MPC controller. In our system we have the following six states

*x, θ*1*, θ*2*, x*˙ *, θ*˙1*, θ*˙2

*x* will be the position of the cart, *θ*1 and *θ*2 are the angle of the first and second pendulum, *x*˙ is the cart velocity and *θ*˙1 *a*nd *θ*˙2 are the angular velocity of the pendulums.

For various initial states we try to make the pendulum stand upright using the MPC controller and the data is collected as pair of states and force to the cart. In order to improve the dataset for each initial state, MPC controller is ran for 4 times for each initial states.

The data files are converted to .xls format for easy access using the pandas library. The datasets are then checked for duplicates and null entries and these data is filtered. Then the data is scaled in order to improve the learning rate of the training.

Then the data is split into two sets test and training randomly. The train set is used for training the model and the test data is for evaluating the model after training. We have split the data into 80% of train data and 20% of test data. Now, this data is used for training the model.

## Neural Network Architecture

Next step is to create a neural network for learning from the datasets. The architecture of the neural network is as follows,

Input layer should have 6 nodes for the 6 states for input. Output layer will be one since we have only one output force for the cart. There are 6 hidden layers with 64, 64, 128, 128, 256, 256 nodes in each hidden layer.

We have no activation function for the input and the output layer and all the other layers have *tanh* as the activation function. *Adam* is used as the optimization algorithm with the learning rate 0.01. Mean Squared Error is used as the loss function.

We use keras framework from tensorflow with python for building and training the model. The model is initially compiled with random weights.

## Model building and simulation

For the training the model training is ran for 1000 epochs and the callback function is used to save the best weights throughout the training process.

The training is done till the loss function reaches 0.2 and the model and the weights are exported for simulation.

The model and weights exported from the training is used to predict the force of the cart for the states. The initial state is taken from one of the initial state used for generating the datasets. And the NN model is used in place of the MPC to predict the force. The results are rendered in ImageMagick tool.

# reinforcement learning

Reinforcement learning is the training of machine learning models to make a sequence of decisions. RL has the following components agents, environments, states, actions and rewards.

RL models are given an environment, a set of actions they can perform, and a goal or a reward they must pursue. Agents act on the environment and try to maximize the cumulative reward so that it will learn to do the given task.

In this project we use the double pendulum environment and make the agent to learn to make the pendulum to stand upright by taking actions with the given feedback. The entire process in explained in detail as follows,

## Double pendulum environment

First, we should have the environment of our system for the agent to act upon and learn. We used *gym* library from OpenAI to generate the environment for the double inverted pendulum.

A *gym* environment should have 4 important elements init, step, reset and render.

Init – It contains the action space, observation space and the environment parameters such as mass of the cart and poles, length of the poles, etc.

Step – It computes the rewards and new states based on the input state using the model of the pendulum.

Reset – It reset the state of the pendulum to the initial position after each episode. Episode is one learning cycle in RL.

Render – It is used to visualize the learning process.

Init function contains the system parameters, angle at which the episode has to fail, and initialize the observation space. Step function has the model of the system itself, cost function for reward. Here, the potential energy is used to calculate the reward. And it computes the new state based on the input and action. Reset function brings the pendulum back to initial state for the new episode.

## DDPG Algorithm

Deep Deterministic Policy Gradient is a model free off-policy RL algorithm for learning continuous action. Since our system is a continuous system DDPG is the best option for this task. It has experience replay and slow-learning target networks.

Experience reply enables reinforcement learning agents to memorize and reuse past experiences, just as humans replay memories for the situation at hand.

There are two NN Actor and critic, Actor takes the state as input and outputs the best action based on the value function and the critic takes the state and reward as the input and predicts the best set of action which yields the maximum reward using the Q-function.

Actor NN has the following architecture, Input layers with 6 nodes, 2 hidden layers with 256 nodes and 1 output layer. Hidden layers has *ReLu* as activation function and output layer has *tanh.* Adam optimization algorithm with 0.0005 learning rate is used.

Critic NN has the 2 NN for state and reward and finally concatenated to one NN. State NN has the architecture, 6 nodes in the input layer, 1 hidden layer with 16 nodes and output layer with 32 nodes. Reward NN has the architecture of 1 node in the input layer and 32 nodes in the output layer. Both have *ReLu*  activation function in the hidden layer and the output layer has *tanh*.

Actor will have less learning rate compared to critic since actor should choose the action trajectory based on the critic’s Q-function value.

## Equations

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

Number equations consecutively. Equation numbers, within parentheses, are to position flush right, as in (1), using a right tab stop. To make your equations more compact, you may use the solidus ( / ), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

*a**b* 

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use “(1)”, not “Eq. (1)” or “equation (1)”, except at the beginning of a sentence: “Equation (1) is . . .”

## Some Common Mistakes

* The word “data” is plural, not singular.
* The subscript for the permeability of vacuum **0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter “o”.
* In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
* A graph within a graph is an “inset”, not an “insert”. The word alternatively is preferred to the word “alternately” (unless you really mean something that alternates).
* Do not use the word “essentially” to mean “approximately” or “effectively”.
* In your paper title, if the words “that uses” can accurately replace the word “using”, capitalize the “u”; if not, keep using lower-cased.
* Be aware of the different meanings of the homophones “affect” and “effect”, “complement” and “compliment”, “discreet” and “discrete”, “principal” and “principle”.
* Do not confuse “imply” and “infer”.
* The prefix “non” is not a word; it should be joined to the word it modifies, usually without a hyphen.
* There is no period after the “et” in the Latin abbreviation “et al.”.
* The abbreviation “i.e.” means “that is”, and the abbreviation “e.g.” means “for example”.

An excellent style manual for science writers is [7].

# Using the Template

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

## Authors and Affiliations

**The template is designed for, but not limited to, six authors.** A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

### For papers with more than six authors: Add author names horizontally, moving to a third row if needed for more than 8 authors.

### For papers with less than six authors: To change the default, adjust the template as follows.

#### Selection: Highlight all author and affiliation lines.

#### Change number of columns: Select the Columns icon from the MS Word Standard toolbar and then select the correct number of columns from the selection palette.

#### Deletion: Delete the author and affiliation lines for the extra authors.

## Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is “Heading 5”. Use “figure caption” for your Figure captions, and “table head” for your table title. Run-in heads, such as “Abstract”, will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles named “Heading 1”, “Heading 2”, “Heading 3”, and “Heading 4” are prescribed.

## Figures and Tables

#### Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 1”, even at the beginning of a sentence.

1. Table Type Styles

| Table Head | Table Column Head | | |
| --- | --- | --- | --- |
| Table column subhead | Subhead | Subhead |
| copy | More table copya |  |  |

1. Sample of a Table footnote. (*Table footnote*)
2. Example of a figure caption. (*figure caption*)

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

##### Acknowledgment *(Heading 5)*

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

##### References

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

1. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*
2. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
3. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
4. K. Elissa, “Title of paper if known,” unpublished.
5. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
6. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
7. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

**IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.**