Dimentionless Policies based on the Buckingham π Theorem: Is it a good way to Generalize Numerical Results?

Alexandre Girard[1]

***Abstract*— Yes if the context, the list of variable defining the motion control problem, are dimentionnaly similar. Here we show that by modifying the problem formulation using dimentionless variables, we can re-use the optimal control law generated numerically for a specific system to a sub-space of dimentionnaly similar problems. This is demonstrated, with numerical results, for the classic motion control problem of swinging-up a torque-limited inverted pendulum. We also discuss the concept of regime, a region in the space of context variables, that can help relax the condition on dimensional similarity. It remains to be seen if this approach can also help generalizing policies for more complex high-dimentional problems.**

I. Introduction

The state-of-the-art toolbox of control engineers include many numerical algorithms and data-driven schemes. Many control approches now include a type of mathematical optimization that has no closed-form solution and that is thus solved numerically, for instance [] and []. Also, recent progress in the field of reinforcement learning have made the approach viable for generating feedback laws for some motion control problem [ETH dog robot], and we expect this list to growth quickly in the future. All in all, numerical tool are very useful and have been used to solve many hard control problem. However, they have a major drawback compared to simpler analytical approaches: parameters of the problem are not explicitly in the solution, which makes it much harder to generalize a solution and reuse it.

Analytical solutions to control problems have the nice property of allowing the solution to be adjusted to different system parameters by simply plugging the new values in the equation. Most numerical solutions act has black-boxes with respect to the parameters. For instance, if we use a reinforcement learning (RL) approach to find a good feedback law for a given task with a robot, the solution will be specific to this system. In other words, the solution, which will take the form of a black-box mapping from states to control inputs, would not be applicable with a modified rigid link twice longer for instance. With an analytical solution, we would simply update the value of the length variable in the equation, while with a RL solution we would have to re-conduct all the training implying generally multiple hours of data collection and/or computation. It would be a great asset to have the ability to adjust black-box numerical solutions with respect to some problem parameters.

[1]Alexandre Girard is with the Department of Mechanical Engineering, Universite de Sherbrooke, Qc, Canada alex.girard@usherbrooke.ca

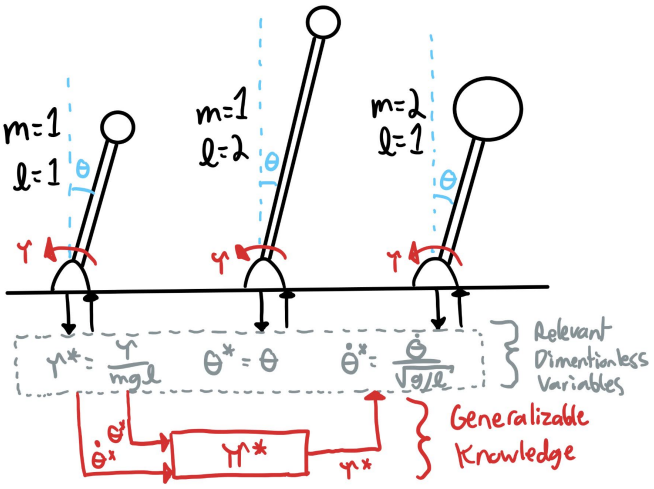


Fig. 1: Big picture

In this paper, we explore the concept of dimentionless policies, a more generic form of knowledge, has a mean to generalize numerical solutions to motion control problem. First, in section IV, we use dimensional analysis to show that ... Then in section V, this property is demonstrated for the classical motion control problem of swinging-up an inverted pendulum. Also, in section **??**, we illustrate with two exemples that the proposal dimentionnal scaling is equivalent to changing parameters in an analytical solution.

II. Background

In the field of learning, this is fall under the vast on-goin challenge of generalizing knowledge...

more specifically in robot, .. transfer learning..

ALso, representation learning...

III. Contribution

IV. Dimentionless Policies

In the following section, the Buckingham π theorem is used to develop the concept of dimentionless policies, and it is shown that multiple systems should have the same dimentionless policy if they have what we will call a similar dimentionless context.

A. Context variables in the policy mapping

A state feedback law is defined here as a mapping f, specific to a given system, from an vector space representing the state x of the dynamic system, to a vector space representing the control inputs u of the system:

$$u = f(x) \tag{1}$$

Under some assumptions, mainly a fully observable systems, an additive cost and an infinite time horizon, the optimal policy is also guarantee to be of this form [1]. Only this case is considered is considered for the following analysis.

To consider the question of how can this feedback law (a form of system specific knowledge) can be transfered in a different context, it is usefull to think about a higher dimension mapping, that we will note π, also having input arguments a vector of variables c describing the context:

$$u = \pi\left(x, c\right) \tag{2}$$

The context c is the vector of all variables that would affect what is the feedback law solution to a motion control problem. For instance, in section V, a case study is conducted considering the optimal feedback law for swinging-up an inverted pendulum. For this exemple, the context variables are the pendulum mass m, the gravitational constant g, the lenght l, but also what we will call task parameters: a parameter in the cost function q and a constraint τ_{max} on the maximum input torque, see Fig. 2. For a given state of the system, the torque solution might be different if any of the context variable change value, for instance if the pendulum is heavier, more torque limited, etc.

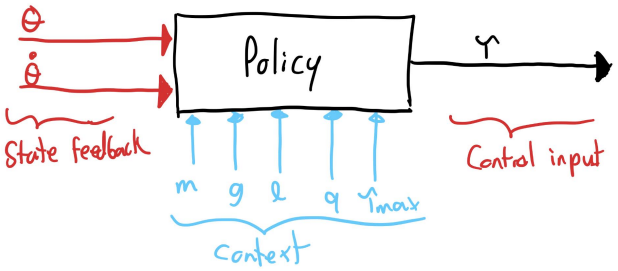


Fig. 2: For the pendulum swing-up task, the context includes 5 variables: the parameter of the system: m, g, and l, a parameter of the cost function: q and a parameter defining the constraints: τ_{max}. The optimal solution depends on all those variables.

The context must include all the variables, system parameters and task parameters, that would affect the solution to the control problem:

$$\underbrace{\begin{bmatrix} u_1 \\ \vdots \\ u_k \end{bmatrix}}_{\text{inputs}} = \pi \left(\underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}}_{\text{states}}, \underbrace{\underbrace{\begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix}}_{\text{system}}, \underbrace{\begin{bmatrix} c_{m+1} \\ \vdots \\ c_{m+l} \end{bmatrix}}_{\text{task}}}_{\text{Context } c} \right) \tag{3}$$

Then we can formalize the goal of generalizing a policy to a different context: if a good feedback policy f_a is found for a system in a context a described by variables c_a, can this knowledge help finding an equivalent good feedback policy in a different context c_b?

$$\pi\left(x, c = c_a\right) = f_a\left(x\right) \quad \Rightarrow \quad \pi\left(x, c = c_b\right) = ? \tag{4}$$

Using the Buckinham π theorem [2], we will show that if the context is dimentionnaly similar, then the dimentionless version of the policy mapping must be equal.

B. Dimentional analysis of the policy mapping

For a system with k control inputs, we can treat the augmented policy as k mappings from states and context variables to each scalar control input u_j:

$$u_j = \pi_j\left(x_1, \ldots, x_n, c_1, \ldots\ldots, c_{m+l}\right) \tag{5}$$

where eq. (5) is the jth line of the policy in vector form described by eq. (3). Then, if the state vector is defined by n variables, and the context is defined by m system parameters plus l tasks parameters, then each mapping π_j involves $1 + n + m + l$ (usually dimentionnal) variables. Here, we will assume that the policy fonction is physically meaningful, in the sense of the requirement for applying the Buckingham π theorem [Cite]. This means for exemple, that a policy that computes a force based on position and velocity measurement would be in this framework, but not of policy for playing chess for instance.

Applying the Buckingham π theorem to the relashionship, tell us that if d dimensions are involved in all thoses variables, then eq. (5) can be restated into an equivalent dimentionless relationship between p dimentionless Π groups where $p \geq (1+n+m+l)-d$ [cite bukhingham pi]. Assuming d dimentions are involved in the m system parameter, and that we are in a situation where the maximum reduction is possible $p = (1 + n + m + l) - d$, we can pick d context variable $\{c_1, c_2, \ldots, c_d\}$ as the basis (the repeated variables) to scale all other variables in a dimentionless form. We will note dimentionless Π group as the base variable with an * subscript:

$$u_j^* = u_j \left[c_1\right]^{e_{1j}^u} \left[c_2\right]^{e_{2j}^u} \ldots \left[c_d\right]^{e_{dj}^u} \quad {\scriptstyle j=\{1,\ldots,k\}} \tag{6}$$

$$x_i^* = x_i \left[c_1\right]^{e_{1i}^x} \left[c_2\right]^{e_{2i}^x} \ldots \left[c_d\right]^{e_{di}^x} \quad {\scriptstyle i=\{1,\ldots,n\}} \tag{7}$$

$$c_i^* = c_i \left[c_1\right]^{e_{1i}^c} \left[c_2\right]^{e_{2i}^c} \ldots \left[c_d\right]^{e_{di}^c} \quad {\scriptstyle i=\{d+1,\ldots,m+l\}} \tag{8}$$

where exposants e_{ij} are rational numbers selected to make all equations dimentionless. Then, the Buckingham π theorem tell us that the relationship described by eq. (5) can be restated as the following relationship between dimentionsless variables:

$$u_j^* = \pi_j^*\left(x_1^*, \ldots, x_n^*, c_{d+1}^*, \ldots, c_{m+l}^*,\right) \tag{9}$$

involving d less dimentionless variables. If we apply the same procedure to all control inputs, when can then assemble the k mappings back into a vector form:

$$\underbrace{\begin{bmatrix} u_1^* \\ \vdots \\ u_k^* \end{bmatrix}}_{\text{Dimentionless feedback law } f^*} = \pi^* \Bigg(\underbrace{\begin{bmatrix} x_1^* \\ \vdots \\ x_n^* \end{bmatrix}}, \underbrace{\begin{bmatrix} c_{d+1}^* \\ \vdots \\ c_m^* \end{bmatrix}, \begin{bmatrix} c_{m+1}^* \\ \vdots \\ c_{m+l}^* \end{bmatrix}}_{\text{Dimentionless context } c^*} \Bigg) \tag{10}$$

that we will sometime write in compact form as:

$$u^* = \pi^*(x^*, c^*) \tag{11}$$

One interesting perk of this dimentional analysis, is that we can remove p variables from the context (typically p would be 2 or 3 for controlling a physical system involving time, force and length). The global problem of learning $\pi(x, c)$, i.e. the good feeback policy for all possible contexts is thus simplified in a dimentionless form. Also, an even more interesting feature, that we can use for transfering feedback laws between systems, is that a global policy $\pi(x, c)$, will have an equivalent dimentionless form for multiple context c. As illustrated at Fig. 3, the dimentionless context c^* is a lower dimentional space $(m + l - d)$ and multiple context vector c will correspond to the same dimentionless vector c^*.

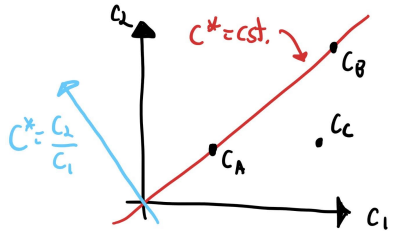


Fig. 3: Dimentionlaly similar context exemple with dimentions: $m = 1$, $l = 1$ and $d = 1$

For a given control problem, if the dimentionless context are equal, then the dimentionless feedback law should be eactly equivalent:

$$\text{if} \quad c_a^* = c_b^* \quad \text{then} \quad f_a^*(x^*) = f_b^*(x^*) \; \forall x^* \tag{12}$$

where

$$f_a^*(x^*) = \pi^*(x^*, c^* = c_a^*) \tag{13}$$
$$f_b^*(x^*) = \pi^*(x^*, c^* = c_b^*) \tag{14}$$

This is simply based on the fact that the dimentionless policy (eq. (11)) give the same output for the same inputs. This results means that the knowledge of a policy for a specific context c can actually be generalized to a sub-space of all context for which the dimentionless context c^* is equal. For instance, lets immagine we have a global policy for a shperical submarine that depends only on the velocity and the radius. In dimentionless form we would find the policy depends only on the Reynolds number, thus would be equivalent for all pair of velocity and radius that correspond to the same Reynolds number.

C. Transfering policies between contexts

In order to exploit this property, it is usefull to define transformation matrices based on scalar equations (6), (7) and (8):

$$u^* = [T_u(c)] \; u \tag{15}$$
$$x^* = [T_x(c)] \; x \tag{16}$$
$$c^* = [T_c(c)] \; c \tag{17}$$

where matrices T_u and T_x are square diagonal matrix, where each diagonal term is a multiplication of the first d context variables ($\{c_1, c_2, \ldots, c_d\}$) up to a rational power (found by applying the Buckinghan π theorem). Equations (15) and (16) are inversible (unless a context variable is equal to zero) and can be used to go back-and-forth between dimentional and dimentionless states and inputs variables. The matrix T_c however have d less row than columns and eq. (17) is not inversible: for a given context c there is only one dimentionless context c^*, however a dimentionless context c^* correspond to multiple dimentional context c.

To summarize the dimentional analysis procedure, using the transformation matrices, if a dimentional feedback law f_a for a context c_a is known:

$$f_a(x) = \pi\left(x, c = c_a\right) \tag{18}$$

its representation in dimentionless form:

$$f_a^*(x^*) = \pi\left(x^*, c^* = c_a^*\right) \tag{19}$$

can be found by scaling the input and output of f_a with T_u and T_x:

$$f_a^*(x^*) = T_u(c_a) \underbrace{f_a\left(\underbrace{T_x^{-1}(c_a)\; x^*}_{x}\right)}_{u} \tag{20}$$

and the dimentionless context values c_a^* can be found using eq. (17). Inversely, if we know a dimentionless feedback law f_b^*, matrices T_u and T_x can be used to scale it back to a specific context c_b:

$$f_b(x) = T_u^{-1}(c_b) \underbrace{f_b^*\left(\underbrace{T_x(c_b)\; x}_{x^*}\right)}_{u^*} \tag{21}$$

Thus, eq. (20) and eq. (21) can be used to take any context specific feedback law, finding its dimentionless form, and scale it back to a new context. In general, there is no guarantee that the behavior of the scaled feedback law in the new context will be similar to the behavior of the feedback law in the original context. However, if the dimentionless context are equal, then the behavior should be eactly equivalent, because the motion control problem was actually the same problem but scaled. For instance, lets suppose an optimal policy f_a is known for a specific context c_a, then applying the scaled policy:

$$f_b(x) = \left[T_u^{-1}(c_b) T_u(c_a)\right] \; f_a\left(\left[T_x^{-1}(c_a) T_x(c_b)\right] \; x\right) \tag{22}$$

on a system described by the context c_b will be equaly equivalent (up to scaling factors) if

$$c_b^* = T_c(c_b) \; c_b = T_c(c_a) \; c_a = c_a^* \tag{23}$$

In section V, we show exemples of this result with numerical solution to dimentionnaly similar pendulum swing-up problems. Furthermore, we demonstrate that in some situations, eq. (23) equality conditions can be relaxed into inequality contitions, i.e. both contexts being in the same region corresponding to a similar regime.

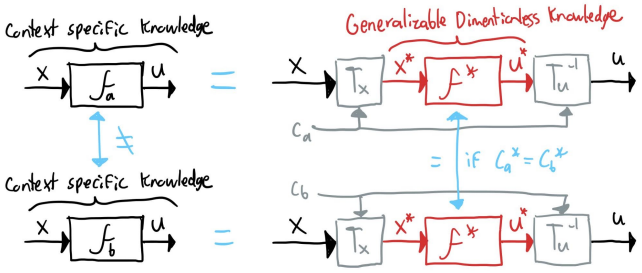


Fig. 4: Isolating the dimentionless knowledge in a policy

V. Optimal pendulum swing-up task

In this paper, we will use a version of the pendulum swing-up task as a prototype problem to test the proposed ideas of dimentionless policies. The motion control problem is defined as finding a feedback law for controling the dynamic system is described by differential equations:

$$ml^2\ddot{\theta} - mgl\sin\theta = \tau \tag{24}$$

that minimize the cost function given by:

$$J = \int (q^2\theta^2 + 0\,\dot{\theta}^2 + 1\,\tau^2)dt \tag{25}$$

subject to input constraints given by:

$$-\tau_{max} \leq \tau \leq \tau_{max} \tag{26}$$

Note that here, the cost function parameter q is included in this way to have units of torque, it was chosen to set to zero the weight on velocity for simplicity, and the weight on torque to one without loss of generality as only the relative values of weights will impact the solution.

Thus, assuming there is no hidden variables and that equations (24), (25) and (26) fully describe the problem. The solution, i.e. the optimal policy for all context, should be of the form given by:

$$\underbrace{\tau}_{\text{inputs}} = \pi\left(\underbrace{\theta,\dot{\theta}}_{\text{states}}, \underbrace{\underbrace{m,g,l}_{\text{system parameters}}, \underbrace{q,\tau_{max}}_{\text{task parameters}}}_{\text{Context } c}\right) \tag{27}$$

and involving variables listed in table I.

Before, conducting the dimentional analysis, it is interesting to note that while there are 3 system paramters m, g and l, they only appear indepedently in two groups in the dynamic equation. We can thus consider only two system parameters, and for convenience mgl coresponding (the maximum static gravitationnal torque) and ω are selected, as listed at table II

A. Dimentional analysis

Here we have one control input, two state, two system parameter and two task parameter, for a total of $1 + (n = 2) + (m = 2) + (l = 2) = 7$ variables are involved. In those variables, only $d = 2$ independants dimensions (ML^2T^{-2} and T^{-1}) are present. Using $c_1 = mgl$ and $c_2 = \omega$ as the repeating variables leads to the following dimentionless groups:

TABLE I: Pendulum swing-up optimal policy variables

Variable	Description	Units	Dimensions
	Control inputs		
τ	Actuator torque	Nm	$[ML^2T^{-2}]$
	State variables		
θ	Joint angle	rad	[]
$\dot{\theta}$	Joint angular velocity	rad/sec	$[T^{-1}]$
	System parameters		
m	Pendulum mass	kg	$[M]$
g	Gravity	m/s^2	$[LT^{-2}]$
l	Pendulum lenght	m	$[L]$
	Problem parameters		
q	Weight parameter	Nm	$[ML^2T^{-2}]$
τ_{max}	Maximum torque	Nm	$[ML^2T^{-2}]$

TABLE II: Pendulum swing-up minimal system variables

	System parameters		
mgl	Maximum gravitational torque	Nm	$[ML^2T^{-2}]$
$\omega = \sqrt{\frac{g}{l}}$	Natural frequency	sec^{-1}	$[T^{-1}]$

$$\Pi_1 = \tau^* = \frac{\tau}{mgl} \qquad \frac{[ML^2T^{-2}]}{[M][LT^{-2}][L]} \tag{28}$$

$$\Pi_2 = \theta^* = \theta \qquad [] \tag{29}$$

$$\Pi_3 = \dot{\theta}^* = \frac{\dot{\theta}}{\omega} \qquad \frac{[T^{-1}]}{[T^{-1}]} \tag{30}$$

$$\Pi_4 = \tau^*_{max} = \frac{\tau_{max}}{mgl} \qquad \frac{[ML^2T^{-2}]}{[M][LT^{-2}][L]} \tag{31}$$

$$\Pi_5 = q^* = \frac{q}{mgl} \qquad \frac{[ML^2T^{-2}]}{[M][LT^{-2}][L]} \tag{32}$$

Here, all 3 torque variables (τ, q and τ_{max}) are scaled by the maximum gravitationnal torque, and the pendulum velocity variable is scaled by the pendulum natural frequency. The

transformation matrices are then

$$\tau^* = \underbrace{[1/mgl]}_{T_u} \tau \tag{33}$$

$$\begin{bmatrix}\theta^* \\ \dot{\theta}^*\end{bmatrix} = \underbrace{\begin{bmatrix}1 & 0 \\ 0 & 1/\omega\end{bmatrix}}_{T_x} \begin{bmatrix}\theta \\ \dot{\theta}\end{bmatrix} \tag{34}$$

$$\underbrace{\begin{bmatrix}q^* \\ \tau^*_{max}\end{bmatrix}}_{c^*} = \underbrace{\begin{bmatrix}0 & 0 & 1/mgl & 0 \\ 0 & 0 & 0 & 1/mgl\end{bmatrix}}_{T_c} \underbrace{\begin{bmatrix}mgl \\ \omega \\ q \\ \tau_{max}\end{bmatrix}}_{c} \tag{35}$$

According to the theorem, any policy that is only based on the variable included in our analysis can be restated as a relationship between the 5 dimentionless Π groups:

$$\tau^* = \pi^* \left(\theta, \dot{\theta}^*, q^*, \tau^*_{max}\right) \tag{36}$$

The dimentional analysis (sec. IV) told us that, for dimentionnally similar swing-up problem (which means here equal ratios q^* and τ^*_{max}) the optimal feedback laws should be equivalent in their dimentionless form. In other words, if we have an optimal policy f_a found in a specific context $c_a = [m_a, l_a, g_a, q_a, \tau_{max,a}]$, and an optimal policy f_b for a second context $c_b = [m_b, l_b, g_b, q_b, \tau_{max,b}]$. Then, both dimentionless form will be equal $f^*_a = f^*_b$ if $q^*_1 = q^*_2$ and $\tau^*_{max,1} = \tau^*_{max,2}$, what we call a *dimentionnally similar context*. Furthermore, we can thus find f_b using f_a or vice-versa unsing the scaling given by eq. (22). However, if $q^*_a \neq q^*_b$ or $\tau^*_{max,a} \neq \tau^*_{max,b}$ then f_a doesn't give us information on f_b wihtout additionnal assumptions.

B. Numerical results

Here, to demonstrate the presented ideas, a numerical algorithm (details of the methodology is presented at ...) is use to compute numerical solution to the prototype motion control problem defined by eq (24), (25) and (26). The used numerical recipe produce feedback law solution in the form of look-up table, based on a discretized grid of the state-space. The optimal (up to discretization errors) feedback laws are computed for 9 contexts listed at table III. In those 9 contexts, there is 3 sub-groups of 3 with dimentionless similar context. Also each sub-group inlcude the same 3 pendulums, illustrated at Fig. 1, a regular, a twice longer and a twice heavier. Contexts 1, 2 and 3 describe a task where the torque is limited to half the satic maximum torque. Contexts 4, 5 and 6 describe a task where the cost highly penalize applying large forces relatively to position errors. Contexts 7, 8 and 9 describe a task where the cost highly penalize position errors, relatively to applying large forces.

Figures 5 to 13 illustrate that for each sub-group with equal dimentionless context, the dimentional feedback law generated numerically, and also the resulting optimal trajectory of the system starting from rest at the bottom position, looks very similar. They are similar up to a scaling of their axis, if we neglect slight differences due to discretization errors. Furthermore, when we compute the dimentionless verion of the feedback laws f^*, using eq. (20), the dimentionless version is actually equal within the dimentionaly similar context sub-group. This was the expected results given by the dimentional analysis of section IV.

TABLE III: Pendulum swing-up problems parameters

	m	g	l	q	τ_{max}
Problems with $\tau^*_{max} = 0.5$ and $q^* = 0.1$					
Context no 1 :	1.0	10.0	1.0	1.0	5.0
Context no 2 :	1.0	10.0	2.0	2.0	10.0
Context no 3 :	2.0	10.0	1.0	2.0	10.0
Problems with $\tau^*_{max} = 1.0$ and $q^* = 0.05$					
Context no 4 :	1.0	10.0	1.0	0.5	10.0
Context no 5 :	1.0	10.0	2.0	1.0	20.0
Context no 6 :	2.0	10.0	1.0	1.0	20.0
Problems with $\tau^*_{max} = 1.0$ and $q^* = 10$					
Context no 7 :	1.0	10.0	1.0	100.0	10.0
Context no 8 :	1.0	10.0	2.0	200.0	20.0
Context no 9 :	2.0	10.0	1.0	200.0	20.0

In terms of how to use this in a practically scenario, we see that if we computed the feedback law given by Fig. 5(a), we can get the feedback law given by Fig. 6(a) or Fig. 7(a) directly by scaling the original policy with eq. (22), using the approriate context variables, without having to recompute. In some sense, we got back the ability to ajust the feedback law on the fly to new system parameters mgl or ω, as it would be the case with an anlytical solution. But this works only within the dimentionaly similar context sub-group. The feedback law given by Fig. 5(a) cannot be scalled into the feedback law given by Fig. 8(a) for instance, since τ^*_{max} and q^* are not equals.

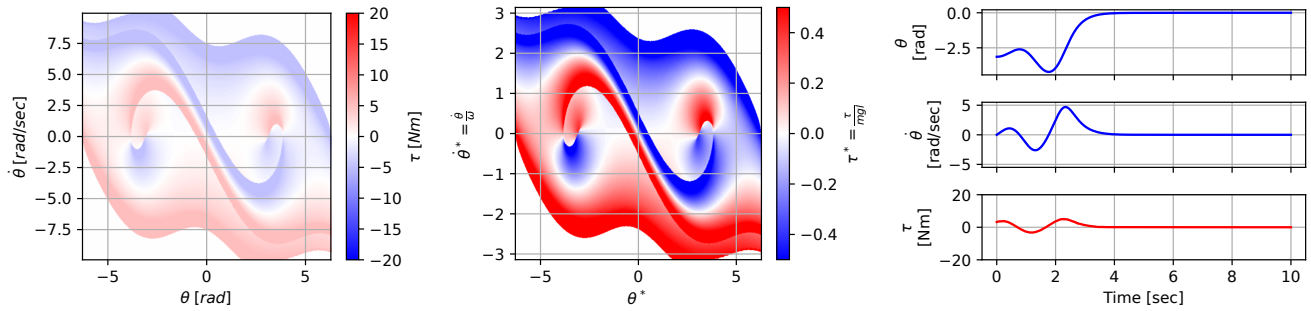
C. Trajectory solutions

Trajectory solutions are also generalizable in their dimentionless forms. Figures 5(c) to 13(c) show optimal trajectory solutions starting from the bottom position at rest. Those trajectories were computed by executing the computed optimal policies in a simulation. We can also see that within the 3 dimentionaly similar sub-groups, the optmial trajectories are actually the same but scalled. Lets assume we were trying to solve a slightly different motion control problem, instead of looking for the feedback law here lets assume we only want the trajectory solutions from the bottom position. We must think about the problem as computing 3 time-based open-loop feedback law defining trajectory for states and inputs:

$$\tau = \pi_\tau(t, m, g, l, q, \tau_{max}) \tag{37}$$

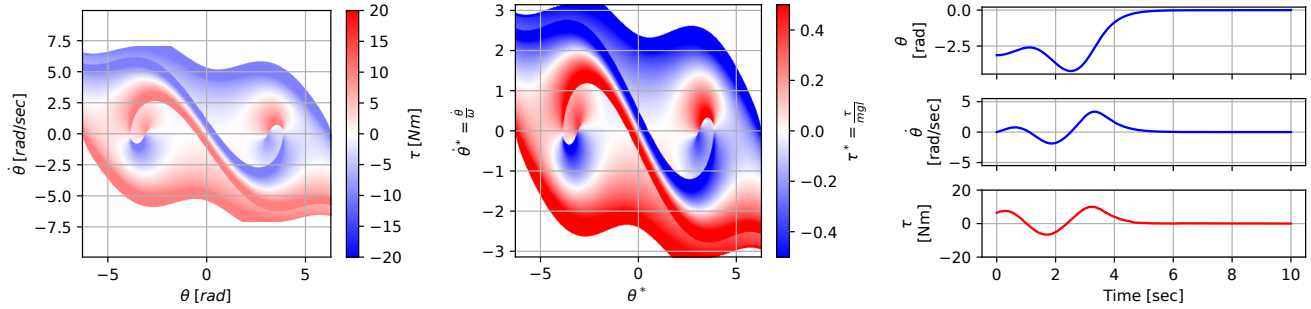
$$\theta = \pi_\theta(t, m, g, l, q, \tau_{max}) \tag{38}$$

$$\underbrace{\dot{\theta}}_{\text{trajectory}} = \pi_{\dot{\theta}}(\underbrace{t, m, g, l, q, \tau_{max}}_{\text{context}}) \tag{39}$$



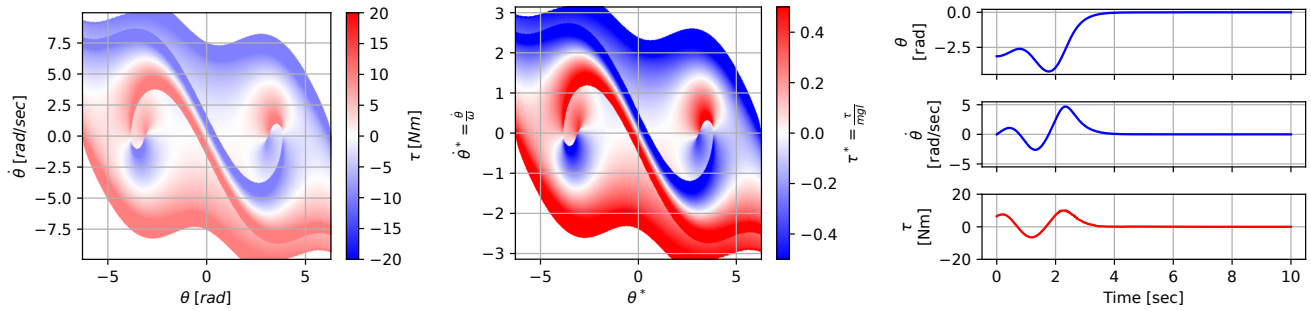
(a) Feedback law f (b) Dimentionless feedback law f^* (c) Optimal trajectory starting at $\theta = -\pi$

Fig. 5: Numerical results for context no 1



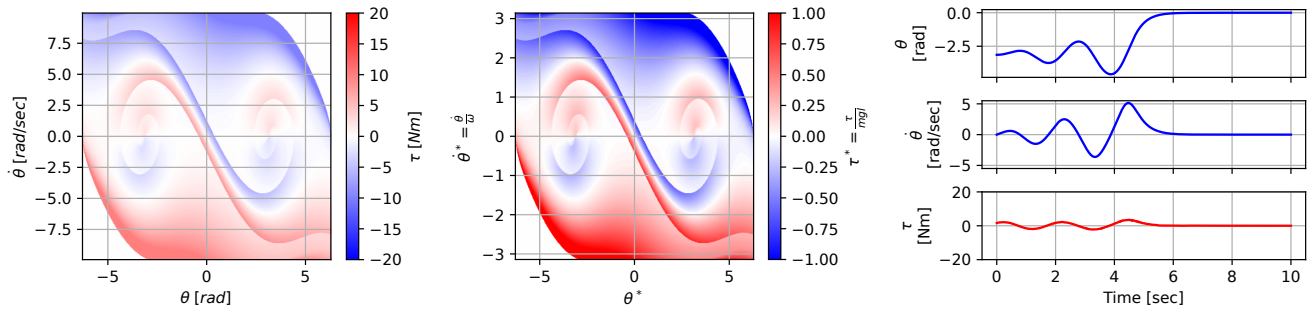
(a) Feedback law f (b) Dimentionless feedback law f^* (c) Optimal trajectory starting at $\theta = -\pi$

Fig. 6: Numerical results for context no 2



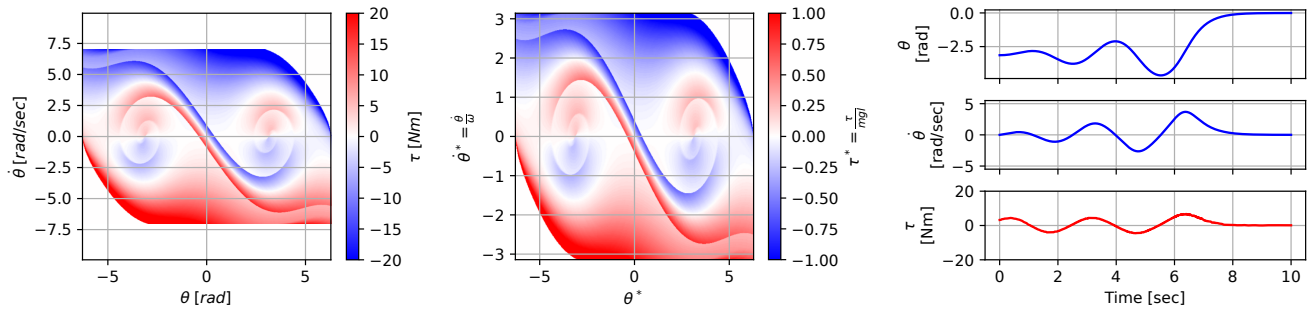
(a) Feedback law f (b) Dimentionless feedback law f^* (c) Optimal trajectory starting at $\theta = -\pi$

Fig. 7: Numerical results for context no 3



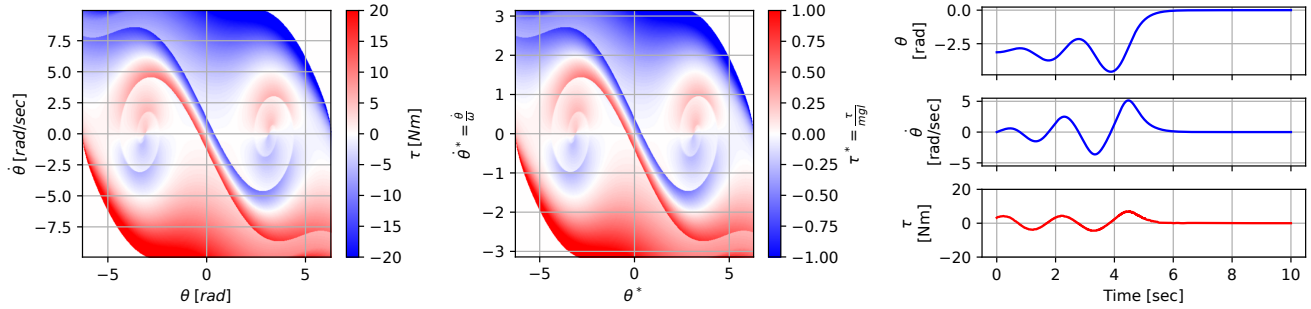
(a) Feedback law f (b) Dimentionless feedback law f^* (c) Optimal trajectory starting at $\theta = -\pi$

Fig. 8: Numerical results for context no 4



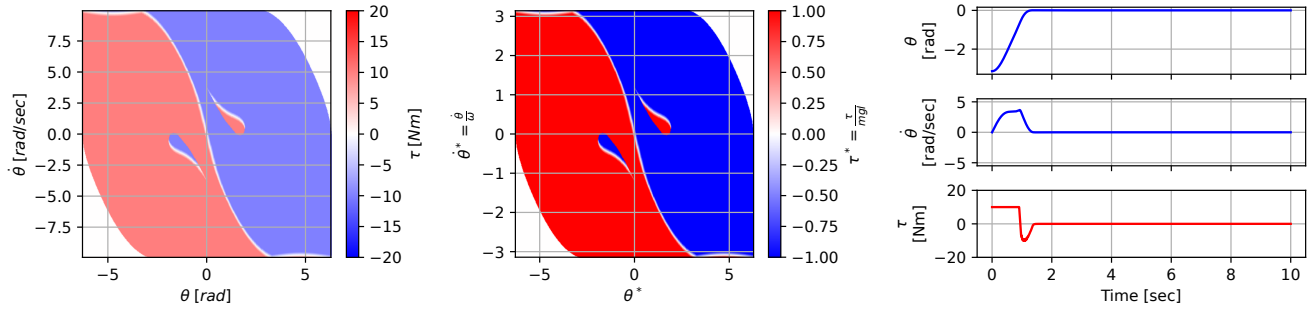
(a) Feedback law f (b) Dimentionless feedback law f^* (c) Optimal trajectory starting at $\theta = -\pi$

Fig. 9: Numerical results for context no 5



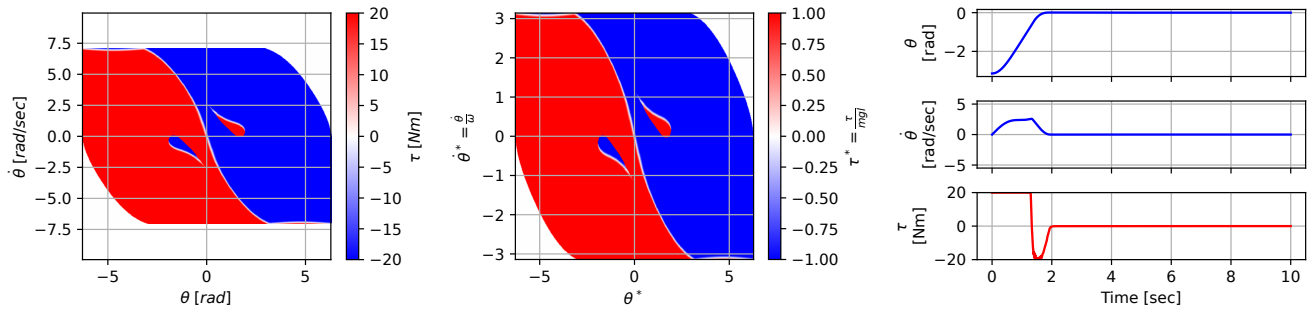
(a) Feedback law f (b) Dimentionless feedback law f^* (c) Optimal trajectory starting at $\theta = -\pi$

Fig. 10: Numerical results for context no 6



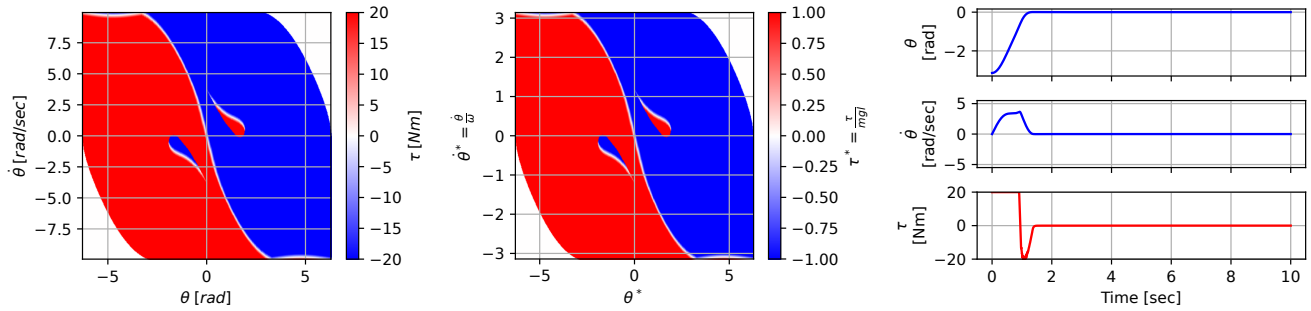
(a) Feedback law f (b) Dimentionless feedback law f^* (c) Optimal trajectory starting at $\theta = -\pi$

Fig. 11: Numerical results for context no 7



(a) Feedback law f (b) Dimentionless feedback law f^* (c) Optimal trajectory starting at $\theta = -\pi$

Fig. 12: Numerical results for context no 8



(a) Feedback law f (b) Dimentionless feedback law f^* (c) Optimal trajectory starting at $\theta = -\pi$

Fig. 13: Numerical results for context no 9

We can use the same dimentionless group as before, but with the addition of a dimentionless time:

$$t^* = t\ \omega \qquad [T][T^{-1}] \tag{40}$$

since the time t is now explicitly an argument of the global policy. Applying the same dimentional analysis, to eq. (37), eq. (38) and eq. (39), we would conclude that it can be restated as:

$$\tau^* = \pi^*_\tau(t^*, q^*, \tau^*_{max}) \tag{41}$$
$$\theta^* = \pi^*_\theta(t^*, q^*, \tau^*_{max}) \tag{42}$$
$$\underbrace{\dot{\theta}^* = \pi^*_{\dot{\theta}}(t^*}_{\text{dim. trajectory}}, \underbrace{q^*, \tau^*_{max}}_{c^*}) \tag{43}$$

Hence, as it is the case for the feedback laws, the optimal trajectories should be equivalent in their dimentionless version, if the dimentionless context is equal.

D. Regimes of solutions

In some situation, changing a context variable will not havbe any effect on the optimal policy. For instance, for the torque-limited optimal pendulum swing-up problem, augmenting τ_{max} or q while keeping the other value fixed will have little effect pass a given threeshold. For instance, if we look at the solutions for context no 4, 5 and 6, using a lot of torque is so highly penalized by the cost function that the saturation limit is not really impacting the solution (except edges cases on the boundary), hence we would expect that augmenting τ_{max} should not change the solution.

Fig. 14 and 15 show a slice (to allow visualization) of the optimal policy solution, for various contexts. Fig. 14 illustrate how changing τ^*_{max} while keeping q^* fixed. We can see that when $\tau^*_{max} < 0.3$ the policy is almost always on the min-max allowable values, this behavior is often called *bang-bang*. At the other extreme when $\tau^*_{max} < 2.5$ the policy solution is continuous and almost never affected by the saturation. Fig. 15 illustrate how changing q^* while keeping τ^*_{max} fixed. We can see that when $q^* < 0.1$ the optimal policy solution does not reach the min-max saturation, while when $q^* > 1.0$, is almost always on the min-max allowable values. We can note that is the limit of very large q^* value, the motion control problem would then to be equivalent to

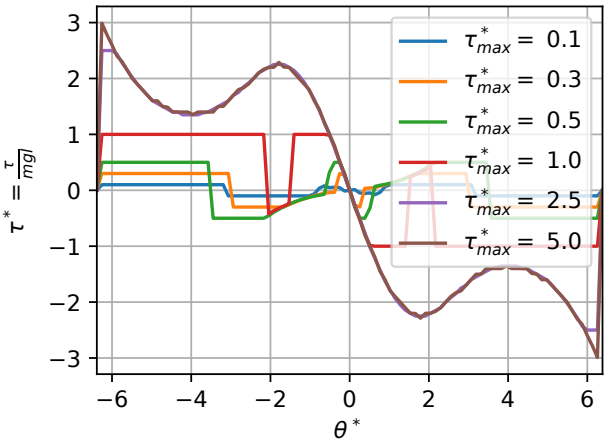


Fig. 14: Optimal dimentionless policy for various contexts $\tau^* = \pi^*(\theta^*, \dot{\theta}^* = 0, q^* = 0.5, \tau^*_{max} = [0.1, ..., 5.0])$

the minimum time problem, for which we know the solution is a bang-bang type policy.

We can see that for extreme context values, we have two type of behaviors, illustrated as region in the dimentionless context space at Fig. 16. Those regions are best caracterized by a ratio of q^* and τ^*_{max}, a new dimentionless value that we will define as:

$$R^* = \frac{\tau^*_{max}}{q^*} = \frac{\tau_{max}}{q} \tag{44}$$

the ratio of the maximum torque saturation τ_{max} over the weight in the cost function q (both have units of torque). When the value of $R^* \approx 1$, the policy solution is continuous in some region, and on the min-max value in some other region of the state-space, a behavior we will call the transition regime. When the value of $R^* \ll 1$, the constraint on torque drives the solution to have have a bang-bang type behavior. In this region, that we would approximate here based on our sensitivity analysis to $R^* \ll 0.1$, the global policy is only a function of τ^*_{max}:

$$\pi^*(\theta^*, \dot{\theta}^*, q^*, \tau^*_{max}) \approx \pi^*(\theta^*, \dot{\theta}^*, \tau^*_{max})\ if\ R^* \ll 1 \tag{45}$$

The value of q^* is not affecting the solution. On the other hang, when the value of $R^* \gg 1$, the policy solution is

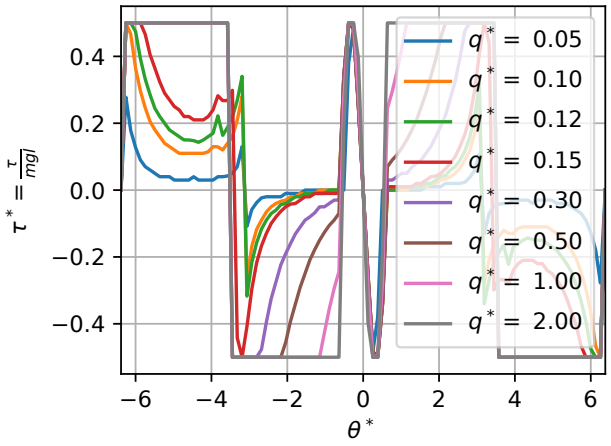


Fig. 15: Optimal dimentionless policy for various contexts $\tau^* = \pi^*(\theta^*, \dot{\theta}^* = 0, q^* = [0.05, ..., 2.0], \tau^*_{max} = 0.5)$

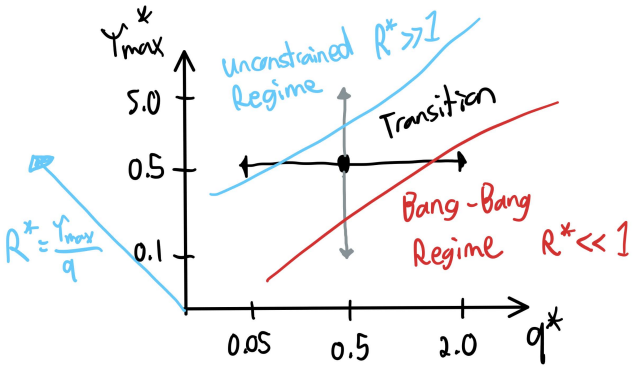


Fig. 16: Regime zones

unconstrained. In this region, that we would approximate here based on our sensitivity analysis to $R^* \gg 10$, the global policy is only a function of q^* since the constraint is so far away:

$$\pi^*(\theta^*, \dot{\theta}^*, q^*, \tau^*_{max}) \approx \pi^*(\theta^*, \dot{\theta}^*, q^*) \; if \; R^* \gg 1 \tag{46}$$

The concept of regime is often leveraged in fluid mechanics, it allow to generalize results between situation where the relevant dimentionless number are not exactly matched. For instance, when the Mach number is small $Ma < 0.3$, we can generaly assume to be in an incompressible regime where various speed of sound parameters (for instance having $Ma = 0.1$ or $Ma = 0.3$) would not really change the behavior. Here, for the purpose of transfering policy solution between different system or context, it means that the condition of having the same exact dimentionless context variables matched, the condition can be relaxed to an inequality that correspond to a regime. For instance, for two motion control problems, if both context are in the unconstrained regime, it is sufficient to match only q^* to have equivalent dimentionless policies. More formally, from eq. (46), we can say that:

$$f^*_a(x^*) = f^*_b(x^*) \; \forall \; x^* \tag{47}$$

$$\text{if} \quad q^*_a = q^*_b \; , \quad R^*_a \gg 1 \quad \text{and} \quad R^*_b \gg 1 \tag{48}$$

Also, for two motion control problems, if both context are in the bang-bang regime, it is sufficient to match only τ^*_{max} to have equivalent dimentionless policies. More formally, from eq. (45), we can say that:

$$f^*_a(x^*) = f^*_b(x^*) \; \forall \; x^* \tag{49}$$

$$\text{if} \quad \tau^*_{max,a} = \tau^*_{max,b} \; , \; R^*_a \ll 1 \text{ and } R^*_b \ll 1 \tag{50}$$

Another point of view, is that assuming we are in one of those regime means we could have removed one variable from the context from the start of the dimentional analysis. All in all, the impact of having such regimes indentified is that we can increase the sub-space of contexts for which the dimentionless version of the policy should be equivalent, leading to a potentially larger pools of systems that can share learned policy or numerical results.

E. Methodology

We obtained the optimal feedback law by using the basic dynamic programming algorithm [1] on a discretized version of the continuous system. The approach is almost equivalent to the value iteration algorithm, sometime refer to as model-based reinforcement learning [3], with the exception that here the total number of iteration steps was fixed (corresponding to a very long time horizon), instead of stopping the iteration after reaching a convergence criterion. This approach was chosen to have consistent results across all contexts, that lead to wide range of order-of-magnitude cost-to-go solution J. The selected discretization parameters are as follow: the time step is 0.025 sec, the state space is discretized into an even 501 x 501 grid and the continuous torque is discretized into 101 discrete control options. Special out-of-bound and on-target termination states are included to guarantee convergence [1]. The source code is available here: (TODO make a special branch) and this google colab page allow reproducing the results: `https://colab.research.google.com/drive/1kf3apyHlf5t7XzJ3uVM8mgDsneVK_63r?usp=sharing`.

1) Additionnal dimentionless parameters for the solver: Using dynamic programming for solving the optimal policy numerically required setting additional parameters that define the domain. Although those parameter should not affect the optimal policy far away from the boundaries, here a dimensionless version of those parameters was kept fixed in all the experiments:

$$\theta^*_{max} = \theta_{max} = 2\pi \tag{51}$$

$$\dot{\theta}^*_{max} = \frac{\dot{\theta}_{max}}{\omega} = \pi \tag{52}$$

$$t^*_f = t_f \; \omega = 10 \times 2\pi \tag{53}$$

θ_{max} is the range of angle for witch the optimal policy is solved, here set at one full revolution. $\dot{\theta}_{max}$ is the range

of angular velocity for witch the optimal policy is solved, here the dimentionless ratio scaled with the natural frequency is set at 2. t_f is the time horizon, here its associated dimentionless ratio is fixed to always corespond to 10 periods of the pendulum using the natural frequency.

VI. Closed-form parametric policies

To better understand the concept of a dimentionless policy, here we apply the Buckingham π theorem on well-known closed form solution to classical motion control problems.

A. Dimentionless Computed torque

The computed torque feedback law is a model-based policy (assuming no torque limits here), that is the solution to the motion control problem of making a mechanical system converging on a desired trajectory, with a specified 2nd order exponential time profile defined by

$$0 = (\ddot{\theta}_d - \ddot{\theta}) + 2\omega_d\zeta(\dot{\theta}_d - \dot{\theta}) + \omega_d^2(\theta - \theta) \tag{54}$$

For the specific case of the pendulum-swing up, the desired trajectory is simply the up-right position ($\ddot{\theta}_d = \dot{\theta}_d = \theta_d = 0$), leaving only two parameter defining the tasks ω_d and ζ. Then, the computed torque control law takes this form:

$$\tau = mgl\sin\theta - 2ml^2\omega_d\zeta\dot{\theta} - ml^2\omega_d^2\theta \tag{55}$$

where the only parameters are the system parameters and two variables caracterizing the convergence speed. Hence, the dimentional global policy is a function of those variables:

$$\underbrace{\tau}_{\text{inputs}} = \pi_{ct}\left(\underbrace{\theta,\dot{\theta}}_{\text{states}}, \underbrace{m,g,l}_{\text{system parameters}}, \underbrace{\omega_d,\zeta}_{\text{task parameters}}\right) \tag{56}$$

having the dimension presented at table 56.

Here, 7 variables are involved and only $p = 2$ independants dimensions (ML^2T^{-2} and T^{-1}), and normally 5 dimentionless groups can be formed:

$$p = (1 + (n = 2) + (m = 2) + (l = 2)) - (p = 2) = 5 \tag{57}$$

Using mgl and ω, the system parameters, as the repeating variables lead to the following dimentionless groups:

$$\Pi_1 = \tau^* = \frac{\tau}{mgl} \qquad \frac{[ML^2T^{-2}]}{[M][LT^{-2}][L]} \tag{58}$$

$$\Pi_2 = \theta^* = \theta \qquad [-] \tag{59}$$

$$\Pi_3 = \dot{\theta}^* = \frac{\dot{\theta}}{\omega} \qquad \frac{[T^{-1}]}{[T^{-1}]} \tag{60}$$

$$\Pi_4 = \omega_d^* = \frac{\omega_d}{\omega} \qquad \frac{[T^{-1}]}{[T^{-1}]} \tag{61}$$

$$\Pi_5 = \zeta^* = \zeta \qquad [] \tag{62}$$

Then, applying the Buckingham π theorem tell us that the computed torque policy can be restated as a relation between the dimensionless variable:

$$\tau^* = \pi_{ct}^*\left(\theta, \dot{\theta}^*, \omega_d^*, \zeta^*\right) \tag{63}$$

TABLE IV: Computed torque variables

Variable	Description	Units	Dimensions
Control inputs			
τ	Actuator torque	Nm	$[ML^2T^{-2}]$
State variables			
θ	Joint angle	rad	[]
$\dot{\theta}$	Joint angular velocity	rad/sec	$[T^{-1}]$
System parameters			
mgl	Maximum gravitational torque	Nm	$[ML^2T^{-2}]$
$\omega = \sqrt{\frac{g}{l}}$	Natural frequency	sec^{-1}	$[T^{-1}]$
Task parameters			
ω_d	Desired closed-loop frequency	sec^{-1}	$[T^{-1}]$
ζ	Desired closed-loop damping	$-$	[ ]

Here we can confirm directly since we have an analytical solution: applying eq. (20) to the computed torque feedback law given by eq. (55), leads to the dimentionless form following form:

$$\tau^* = \left[\frac{1}{mgl}\right]\left(mgl\sin\theta - 2ml^2\omega_d\zeta\left(\omega\dot{\theta}^*\right) - ml^2\omega_d^2\theta\right) \tag{64}$$

$$\tau^* = \sin\theta^* - 2\omega_d^*\zeta\dot{\theta}^* - (\omega_d^*)^2\theta^* \tag{65}$$

confirming the structure predicted by equation eq. (63).

Also, we will use this exemple analytical solution to show that, when the dimentionless context is equal, scaling a policy using eq. (22), is equivalent to pluging new values of the system parameters. To simplify the following development we introduce new variables groups:

$$G_a = m_a g_a l_a \quad H_a = m_a l_a^2 \quad \omega_a = \sqrt{G_a/H_a} \tag{66}$$

$$G_b = m_b g_b l_b \quad H_b = m_b l_b^2 \quad \omega_b = \sqrt{G_b/H_b} \tag{67}$$

where indices a and b refer to a context instance. By just substituting the variable in the analytical policy solution given by eq (55), we can get two instance of the feedback law:

$$f_a = G_a\sin\theta - 2H_a(\omega_d\zeta)_a\dot{\theta} - H_a(\omega_d^2)_a\theta \tag{68}$$

$$f_b = G_b\sin\theta - 2H_b(\omega_d\zeta)_b\dot{\theta} - H_b(\omega_d^2)_b\theta \tag{69}$$

Now, lets try to get to f_b from f_a using eq. (22):

$$f_b = \left[\frac{G_b}{G_a}\right]\left[G_a\sin\theta - 2H_a(\omega_d\zeta)_a\left[\frac{\omega_a}{\omega_b}\right]\dot{\theta} - H_a(\omega_d^2)_a\theta\right] \tag{70}$$

Then, distributing $1/G_a$ leads to:

$$f_b = G_b \left[\sin\theta - 2\frac{(\omega_d)_a}{\omega_a}\zeta_a \frac{\dot{\theta}}{\omega_b} - \left(\frac{(\omega_d)_a}{\omega_a}\right)^2 \theta \right] \tag{71}$$

where the this intermediate step illustrate the dimentionless generic form in the brackets. Finaly, by distributing the G_b factor we can get:

$$f_b = G_b \sin\theta - 2H_b \left(\frac{\omega_b}{\omega_a}(\omega_d)_a\right)\zeta_a\dot{\theta} - H_b\left(\frac{\omega_b}{\omega_a}(\omega_d)_a\right)^2\theta \tag{72}$$

which will be exactly equivalent to eq. (69) (i.e. be equivalent to changing the value of context variable in eq. (68) from the a to the b version) if:

$$\frac{\omega_b}{\omega_a}(\omega_d)_a = (\omega_d)_b \quad \text{and} \quad \zeta_a = \zeta_b \tag{73}$$

which is the condition of having equal dimentionless context $c_a^* = c_b^*$ for this motion control problem:

$$\frac{(\omega_d)_a}{\omega_a} = \omega_a^* = \omega_b^* = \frac{(\omega_d)_b}{\omega_b} \quad \text{and} \quad \zeta_a^* = \zeta_b^* \tag{74}$$

This example illustrates that applying the scaling of eq. (22) based on the dimensional analysis framework, is equivalent to changing the context variables in an analytical solution, when dimentionless context variables are equal.

B. Dimentionless Linear Quatratic Reglator (LQR) solution

Here we analyse a simplified motion control problem with the LQR framework. A linearized verison of the equation of motion is used:

$$ml^2\ddot{\theta} - mgl\theta = \tau \tag{75}$$

Also, the same cost function that was used in section V is used to formulate the optimal control problem:

$$J = \int (q^2\theta^2 + 0\,\dot{\theta}^2 + 1\,\tau^2)dt \tag{76}$$

However, here no constraints on the torque are included in the problem. With this problem definition, the same variable as in section V are presents, except the torque limit, see table V. The global policy solution should then have the form:

$$\underbrace{\tau}_{\text{inputs}} = \pi\left(\underbrace{\theta,\dot{\theta}}_{\text{states}}, \underbrace{m,g,l}_{\text{system parameters}}, \underbrace{q}_{\text{task parameters}}\right) \tag{77}$$

We can thus select the same dimentionless group as before, and conclude that eq. (77) can be restated under this form:

$$\tau^* = \pi^*\left(\theta, \dot{\theta}^*, q^*\right) \tag{78}$$

For this motion control problem, an analytical solution exist (see appendix A), and the policy is

$$\tau = \left[mgl + \sqrt{(mgl)^2 + q^2}\right]\theta \tag{79}$$

$$+ \left[\sqrt{2ml^2}\sqrt{mgl + \sqrt{(mgl)^2 + q^2}}\right]\dot{\theta} \tag{80}$$

TABLE V: Pendulum swing-up optimal policy variables

Variable	Description	Units	Dimensions
Control inputs			
τ	Actuator torque	Nm	$[ML^2T^{-2}]$
State variables			
θ	Joint angle	rad	[]
$\dot{\theta}$	Joint angular velocity	rad/sec	$[T^{-1}]$
System parameters			
mgl	Maximum gravitational torque	Nm	$[ML^2T^{-2}]$
$\omega = \sqrt{\frac{g}{l}}$	Natural frequency	sec^{-1}	$[T^{-1}]$
Task parameters			
q	Weight parameter	Nm	$[ML^2T^{-2}]$

Applying eq. (20) to this feedback law given leads to the dimentionless form following form, using again $G = mgl$ and $H = ml^2$ for shortness:

$$\tau^* = \left[\frac{1}{G}\right]\left[G + \sqrt{G^2 + q^2}\right]\theta \tag{81}$$

$$+ \left[\frac{1}{G}\right]\left[\sqrt{2H(G + \sqrt{G^2 + q^2})}\right]\left[\omega\dot{\theta}^*\right] \tag{82}$$

$$\tau^* = \left[1 + \sqrt{\frac{G^2 + q^2}{G^2}}\right]\theta + \left[\sqrt{\frac{2H\omega^2}{G}\frac{G + \sqrt{G^2 + q^2}}{G}}\right]\dot{\theta}^* \tag{83}$$

$$\tau^* = \left[1 + \sqrt{1 + (q^*)^2}\right]\theta + \left[\sqrt{2}\sqrt{1 + \sqrt{1 + (q^*)^2}}\right]\dot{\theta}^* \tag{84}$$

With the final dimentionless form only a function of states and the dimentionless cost parameter, as predicted by the dimentional analysis (eq. (78)).

We can also use this second example to show that scalling the policy with eq. (22) will be equivalent to substituting new context variables. Let again say we have two context, labelled a and b, that we can use the global policy solution to have the two dimensional version:

$$f_a = \left[G_a + \sqrt{G_a^2 + q_a^2}\right]\theta + \left[\sqrt{2H_a(G_a + \sqrt{G_a^2 + q_a^2})}\right]\dot{\theta} \tag{85}$$

$$f_b = \left[G_b + \sqrt{G_b^2 + q_b^2}\right]\theta + \left[\sqrt{2H_b(G_b + \sqrt{G_b^2 + q_b^2})}\right]\dot{\theta} \tag{86}$$

We can try to find f_b by scaling f_a using eq. (22):

$$f_b = \left[\frac{G_b}{G_a}\right] f_a\left(\theta, \left[\frac{\omega_a}{\omega_b}\right]\dot{\theta}\right) \tag{87}$$

$$f_b = G_b\left[1+\sqrt{1+(q_a^*)^2}\right]\theta \tag{88}$$

$$+ G_b\left[\sqrt{2}\sqrt{1+\sqrt{1+(q_a^*)^2}}\right]\frac{\dot{\theta}}{\omega_b} \tag{89}$$

where $q_a^* = q_a/G_a$ is the dimentionless version of the cost parameter. Then, we can distribute G_b:

$$f_b = \left[G_b+\sqrt{G_b^2+(G_b q_a^*)^2}\right]\theta \tag{90}$$

$$+ \left[\sqrt{\frac{2G_b}{\omega_b^2}}\sqrt{G_b+\sqrt{G_b^2+(G_b q_a^*)^2}}\right]\dot{\theta} \tag{91}$$

$$f_b = \left[G_b+\sqrt{G_b^2+(G_b q_a^*)^2}\right]\theta \tag{92}$$

$$+ \left[\sqrt{2H_b}\sqrt{G_b+\sqrt{G_b^2+(G_b q_a^*)^2}}\right]\dot{\theta} \tag{93}$$

which is equivalent to eq. (86), if

$$G_b q_a^* = q_b \quad \text{or equivalently} \quad q_a^* = q_b^* \tag{94}$$

which is the condition of having equal dimentionless context $c_a^* = c_b^*$ for this motion control problem.

VII. Conclusion

The concept of dimentionless context is powerful, in the sense that it shows how to transfer a control policy to diferent context where the results should be exactly equivalent. However, it is limited because if the dimentionless context c^* is not exactly equal, then nothing can be deduced regarding if a policy is transferable. Furthermore, the challenge of leveraging this idea is to include all meaningful context variables. If a meaningful variable (in the sense that the policy would be different if its value is changed) is omitted from the context vector c in the dimentional analysis, then the dimentional analysis results might be wrong. On the other hand, if we include too many variables to fully describe a context, then dimentionnaly similar context space will probably be so specific it won't be pratical to use for transfering policy between systems. Henseforth, finding the appropriate parametrization of the context will be critical in order to leverage this principle for sharing policy between similar system,

Appendix

A. LQR analytic solution

In this section, we show that the policy given by eq. (80), is optimal with respect to the LQR problem defined in section VI-B.

We can write the equation of motino given by eq. (75) in state-space form, using $G = mgl$ and $H = ml^2$, as :

$$\frac{d}{dt}\begin{bmatrix}\theta\\ \dot{\theta}\end{bmatrix} = \underbrace{\begin{bmatrix}0 & 1\\ G/H & 0\end{bmatrix}}_{A}\underbrace{\begin{bmatrix}\theta\\ \dot{\theta}\end{bmatrix}}_{x} + \underbrace{\begin{bmatrix}0\\ 1/H\end{bmatrix}}_{B}\underbrace{[\tau]}_{u} \tag{95}$$

Then, adapting a solution from [4], if we parameterize the weight matrix of the cost function this way:

$$J = \int_0^\infty \left(x^T \underbrace{\begin{bmatrix}a(a-2G) & 0\\ 0 & b^2-2aH\end{bmatrix}}_{Q} x + u^T \underbrace{[1]}_{R} u \right) dt \tag{96}$$

the optimal cost-to-go is given by

$$J = x^T \underbrace{\begin{bmatrix}b(a-G) & aH\\ aH & bH\end{bmatrix}}_{S} x \tag{97}$$

and the optimal feedback policy is

$$u = -\underbrace{\left[R^{-1}B^T S\right]}_{K} x = -\underbrace{\begin{bmatrix}a & b\end{bmatrix}}_{K} x \tag{98}$$

This solution can by verifyied by subtituting matrices into the algebraic Riccati equation given by:

$$0 = SA + A^T S - SBR^{-1}B^T S + Q \tag{99}$$

as all this fit into the continuous infinite horizon classical LQR result [1]. Then, we can see that the cost function defined in section VI-B, is a special case where $Q_{11} = q^2$ and $Q_{22} = 0$, leading to:

$$q^2 = a(a-2G) \tag{100}$$

$$0 = b^2 - 2aH \tag{101}$$

Solving for a and b, and keeping the positive solution, leads to

$$a = G + \sqrt{G^2+q^2} \tag{102}$$

$$b = \sqrt{2aH} = \sqrt{2H(G+\sqrt{G^2+q^2})} \tag{103}$$

that when substituted into eq. (98) is equal to the policy given by eq. (80).

References

[1] D. P. Bertsekas, *Dynamic Programming and Optimal Control: Approximate Dynamic Programming*, Nashua, NH, 2012.

[2] M. E. Buckingham, "On Physically Similar Systems; Illustrations of the Use of Dimensional Equations," *Physical Review*, Oct. 1914, publisher: American Physical Society (APS). [Online]. Available: https://www.scienceopen.com/document?vid=805fe995-1849-413a-b228-3fe616732290

[3] R. S. Sutton and A. G. Barto, *Reinforcement Learning, second edition: An Introduction*, second edition ed. Cambridge, Massachusetts: Bradford Books, Nov. 2018.

[4] B. HANKS and R. SKELTON, "Closed-form solutions for linear regulator-design of mechanical systems including optimal weighting matrix selection," National Aeronautics and Space Administration, NASA Technical Memorandum 104052, Jan. 1991, _eprint: https://arc.aiaa.org/doi/pdf/10.2514/6.1991-1117. [Online]. Available: https://arc.aiaa.org/doi/abs/10.2514/6.1991-1117