

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

Alexandre Girard¹

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 dimensionless variables, we can re-use the optimal control law generated numerically for a specific system to a sub-space of dimensionnaly similar problem. This is demonstrated, with numerical results, for the classic motion control problem of swinging-up a torque-limited inverted pendulum. **TODO:** (We also demonstrate that by leveraging this scheme when using reinforcement learning, multiple systems of various dimentions can share a data-base during the learning phase, which can be a big advantage for data efficiency.) It remains to be seen if this approach can also help generalizing policies for more complex high-dimentional problems.

I. INTRODUCTION

- a) Many numerical algorithms = black box mapping:
-Trajectory optimization -Reinforcement learning -etc.
- b) With numerical results, unlike analytical solutions, system and problem parameters are not explicitly in the solution :
- c) This makes the results "context specific" and makes its harder to generalize :

II. DIMENTIONLESS POLICIES

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

$$u = f(x) \quad (1)$$

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 transferred in a different context, it is usefull to think about a higher

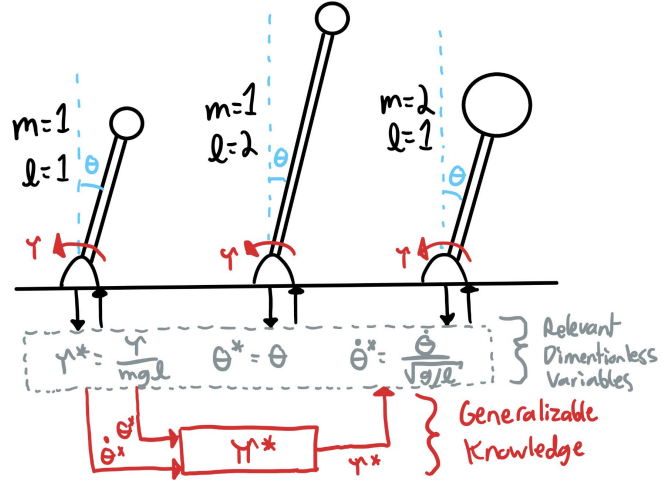


Fig. 1: Big picture

dimension mapping, that we will note π , also having input arguments a vector of variables c describing the context:

$$u = \pi(x, c) \quad (2)$$

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 III, a case study is conducted considering the optimal feedback law for swinging-up an inverted pendulum. For this example, 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.

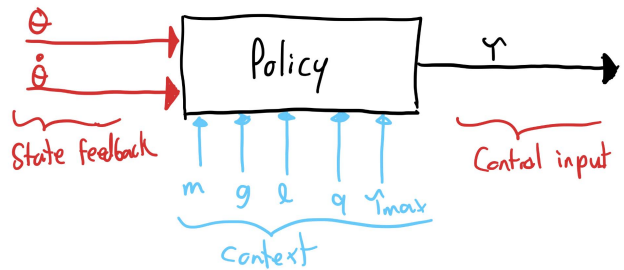


Fig. 2: For the pendulum swing-up task, the context includes 5 variables: the parameter of the system: m, g , and l , a parameter of the cost function: q and a parameter defining the constraints: τ_{max} . The optimal solution depends on all those variables.

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

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

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

Context c

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 ?

$$\pi(x, c = c_a) = f_a(x) \Rightarrow \pi(x, c = c_b) = ? \quad (4)$$

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

B. Dimensional 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 :

$$u_j = \pi_j(x_1, \dots, x_n, c_1, \dots, c_{m+l}) \quad (5)$$

where eq. (5) is the j th 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 dimensional) variables. Here, we will assume that the policy function is physically meaningful, in the sense of the requirement for applying the Buckingham π theorem [Cite]. This means for example, 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 relationship, tell us that if d dimensions are involved in all those variables, then eq. (5) can be restated into an equivalent dimensionless relationship between p dimensionless Π groups where $p \geq (1 + n + m + l) - d$ [cite buckingham pi]. Assuming d dimensions 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 $\{c_1, c_2, \dots, c_d\}$ as the basis (the repeated variables) to scale all other variables in a dimensionless form. We will note dimensionless Π group as the base variable with an *

subscript:

$$u_j^* = u_j [c_1]^{e_{1j}} [c_2]^{e_{2j}} \dots [c_d]^{e_{dj}} \quad j = \{1, \dots, k\} \quad (6)$$

$$x_i^* = x_i [c_1]^{e_{1i}} [c_2]^{e_{2i}} \dots [c_d]^{e_{di}} \quad i = \{1, \dots, n\} \quad (7)$$

$$c_i^* = c_i [c_1]^{e_{1i}} [c_2]^{e_{2i}} \dots [c_d]^{e_{di}} \quad i = \{d+1, \dots, m+l\} \quad (8)$$

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

$$u_j^* = \pi_j^*(x_1^*, \dots, x_n^*, c_{d+1}^*, \dots, c_{m+l}^*) \quad (9)$$

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

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

that we will sometime write in compact form as:

$$u^* = \pi^*(x^*, c^*) \quad (11)$$

One interesting perk of this dimensional 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 feedback policy for all possible contexts is thus simplified in a dimensionless form. Also, an even more interesting feature, that we can use for transferring feedback laws between systems, is that a global policy $\pi(x, c)$, will have an equivalent dimensionless form for multiple context c . As illustrated at Fig. 3, the dimensionless context c^* is a lower dimensional space ($m + l - d$) and multiple context vector c will correspond to the same dimensionless vector c^* .

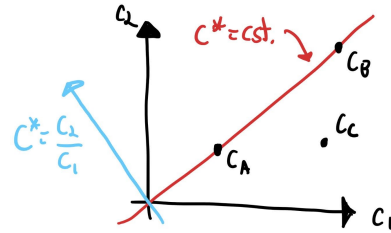


Fig. 3: Dimensional similar context example with dimensions: $m = 1$, $l = 1$ and $d = 1$

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

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

where

$$f_a^*(x^*) = \pi^*(x^*, c^* = c_a^*) \quad (13)$$

$$f_b^*(x^*) = \pi^*(x^*, c^* = c_b^*) \quad (14)$$

This is simply based on the fact that the dimensionless 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 dimensionless context c^* is equal. For instance, lets imagine we have a global policy for a shperical submarine that depends only on the velocity and the radius. In dimensionless 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. Transferring policies between contexts

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

$$u^* = [T_u(c)] u \quad (15)$$

$$x^* = [T_x(c)] x \quad (16)$$

$$c^* = [T_c(c)] c \quad (17)$$

where matrices T_u and T_x are square diagonal matrix, where each diagonal term is a multiplication of the first d context variables ($\{c_1, c_2, \dots, c_d\}$) up to a rational power (found by applying the Buckingham 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 dimensional and dimensionless 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 dimensionless context c^* , however a dimensionless context c^* correspond to multiple dimensional context c .

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

$$f_a(x) = \pi(x, c = c_a) \quad (18)$$

its representation in dimensionless form:

$$f_a^*(x^*) = \pi(x^*, c^* = c_a^*) \quad (19)$$

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

$$f_a^*(x^*) = T_u(c_a) f_a \left(\underbrace{T_x^{-1}(c_a) x^*}_x \right) \quad (20)$$

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

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

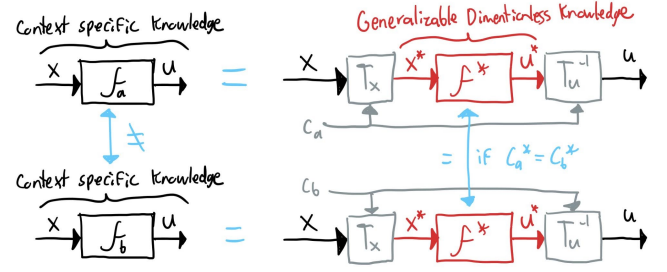


Fig. 4: Isolating the dimensionless knowledge in a policy

Thus, eq. (20) and eq. (21) can be used to take any context specific feedback law, finding its dimensionless 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 dimensionless context are equal, then the behavior should be exactly 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:

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

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

$$c_b^* = T_c(c_b) c_b = T_c(c_a) c_a = c_a^* \quad (23)$$

In section III, we show examples of this result with numerical solution to dimensionnaly 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.

III. 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 dimensionless policies. The motion control problem is defined as finding a feedback law for controlling the dynamic system is described by differential equations:

$$ml^2\ddot{\theta} - mgl \sin \theta = \tau \quad (24)$$

that minimize the cost function given by:

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

subject to input constraints given by:

$$-\tau_{max} \leq \tau \leq \tau_{max} \quad (26)$$

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:

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

Context c

and involving variables listed in table I.

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	$[ML^2T^{-2}]$

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

TABLE II: Pendulum swing-up minimal system variables

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

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:

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

$$\Pi_2 = \theta^* = \theta \quad [] \quad (29)$$

$$\Pi_3 = \dot{\theta}^* = \frac{\dot{\theta}}{\omega} \frac{[T^{-1}]}{[T^{-1}]} \quad (30)$$

$$\Pi_4 = \tau_{max}^* = \frac{\tau_{max}}{mgl} \frac{[ML^2T^{-2}]}{[M][LT^{-2}][L]} \quad (31)$$

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

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

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

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

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

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:

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

The dimentional analysis (sec. II) 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 dimensionless similar context. Also each sub-group include 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 static 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.

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

Figures 5 to 13 illustrate that for each sub-group with equal dimensionless context, the dimensional 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 dimensionless version of the feedback laws f^* , using eq. (20), the dimensionless version is actually equal within the dimensionally similar context sub-group. This was the expected results given by the dimensional analysis of section II.

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 appropriate context variables, without having to recompute. In some sense, we got back the ability to adjust the feedback law on the fly to new system parameters mgl or ω , as it would be the case with an analytical solution. But this works only within the dimensionally similar context sub-group. The feedback law given by Fig. 5(a) cannot be scaled 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 dimensionless 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 dimensionally similar sub-groups, the optimal trajectories are actually the same but scaled. 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:

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

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

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

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

$$t^* = t \omega \quad [T][T^{-1}] \quad (40)$$

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

$$\tau^* = \pi_\tau^*(t^*, q^*, \tau_{max}^*) \quad (41)$$

$$\theta^* = \pi_\theta^*(t^*, q^*, \tau_{max}^*) \quad (42)$$

$$\dot{\theta}^* = \underbrace{\pi_{\dot{\theta}}^*(t^*, q^*, \tau_{max}^*)}_{\text{dim. trajectory}} \underbrace{\quad}_{c^*} \quad (43)$$

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

D. Regimes of solutions

In some situation, changing a context variable will not have 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 threshold. 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

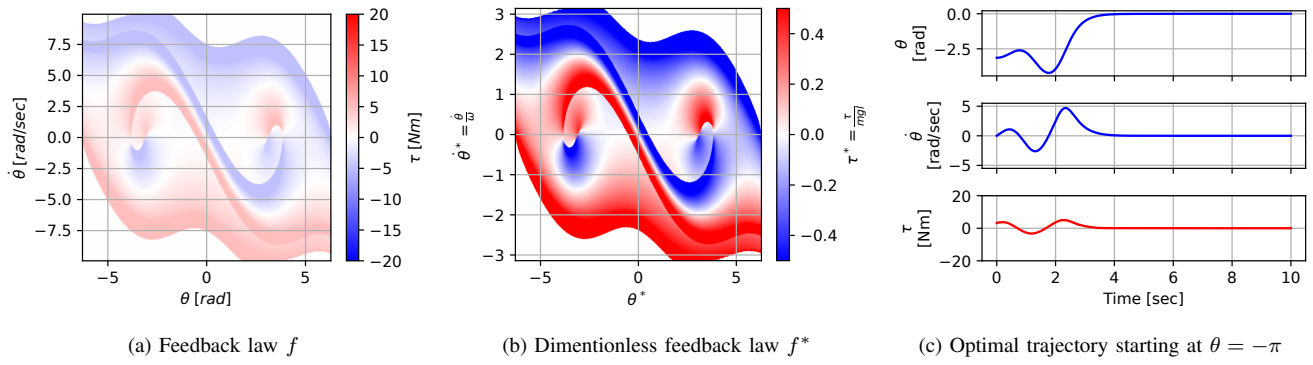


Fig. 5: Numerical results for context no 1

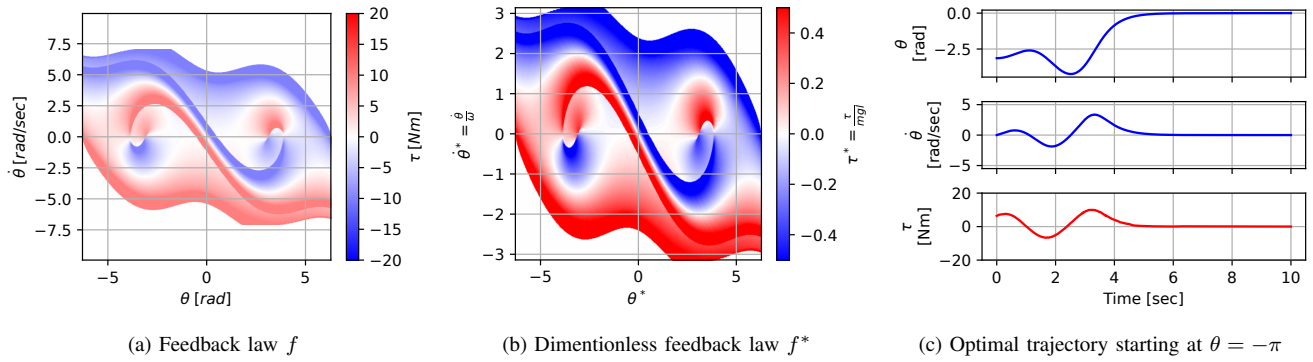


Fig. 6: Numerical results for context no 2

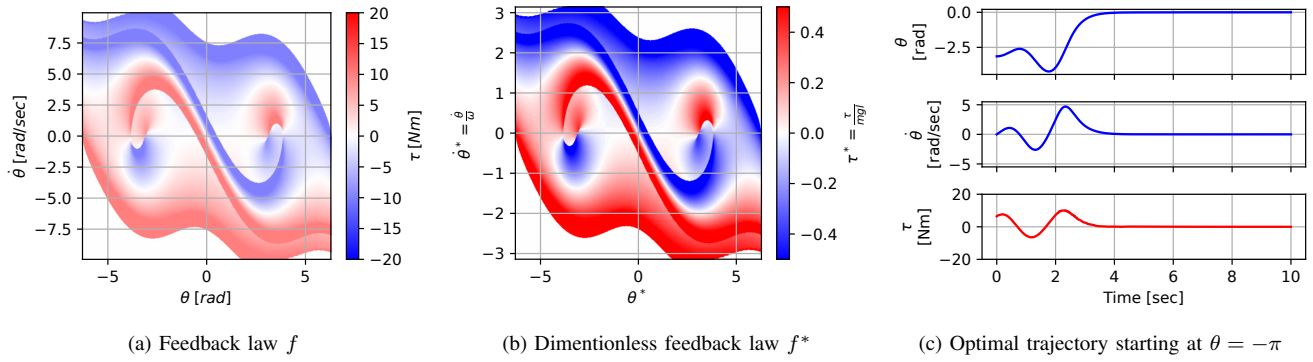


Fig. 7: Numerical results for context no 3

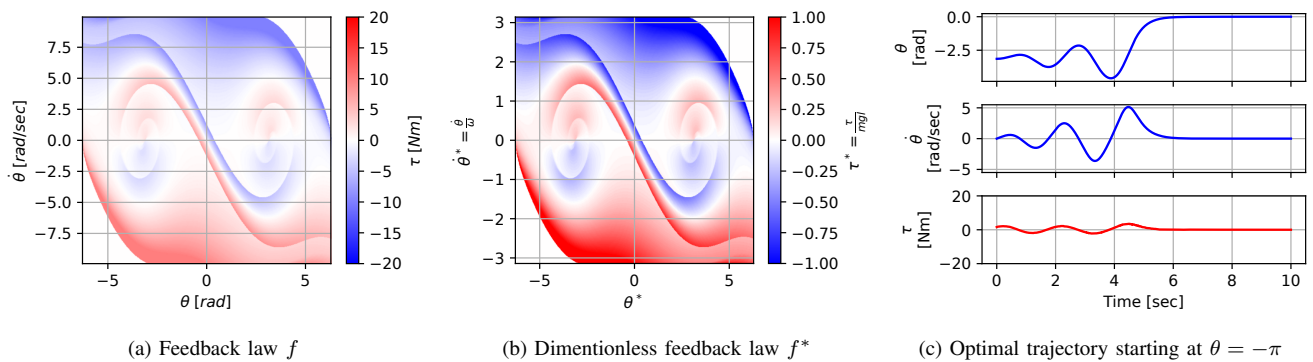


Fig. 8: Numerical results for context no 4

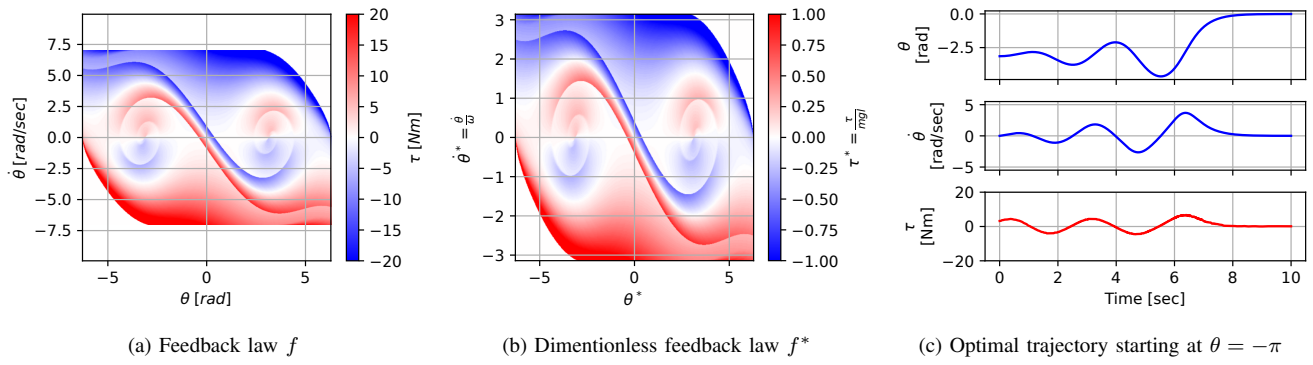


Fig. 9: Numerical results for context no 5

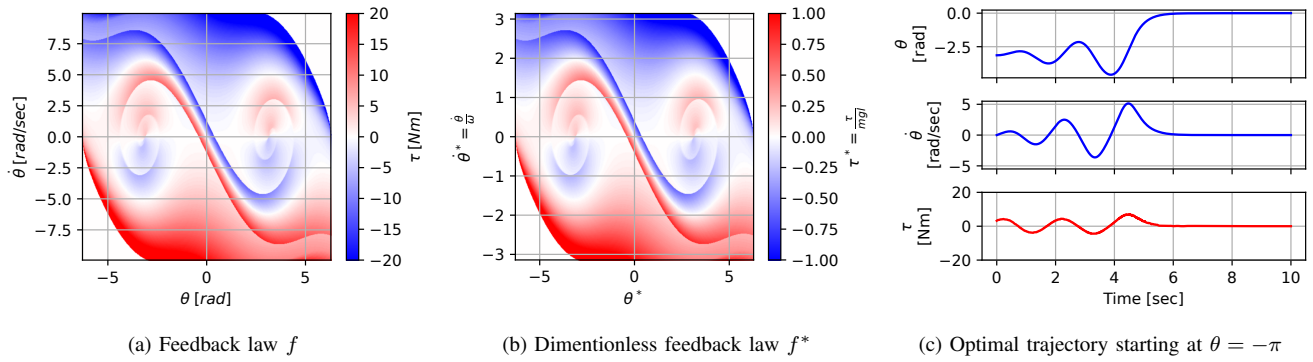


Fig. 10: Numerical results for context no 6

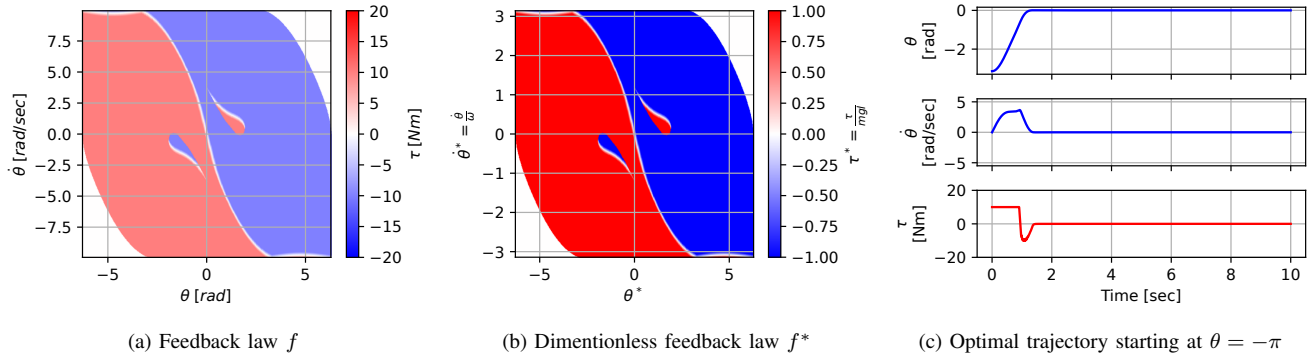


Fig. 11: Numerical results for context no 7

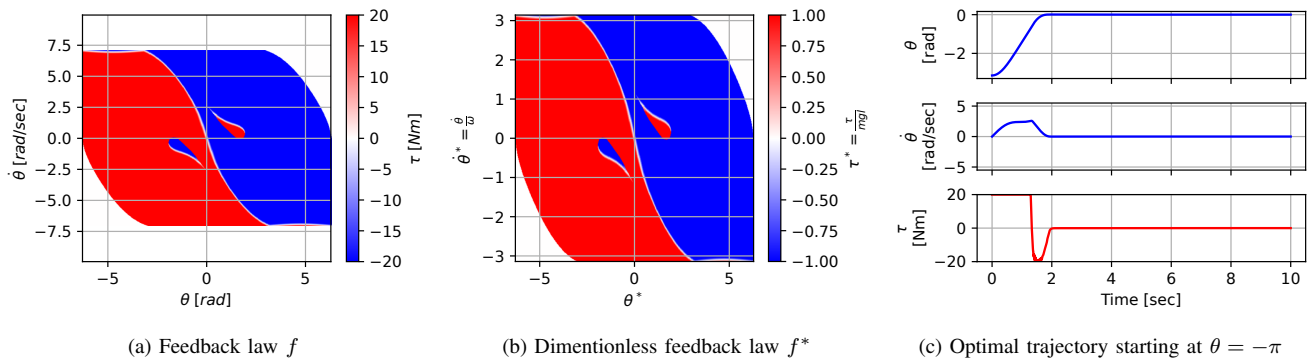


Fig. 12: Numerical results for context no 8

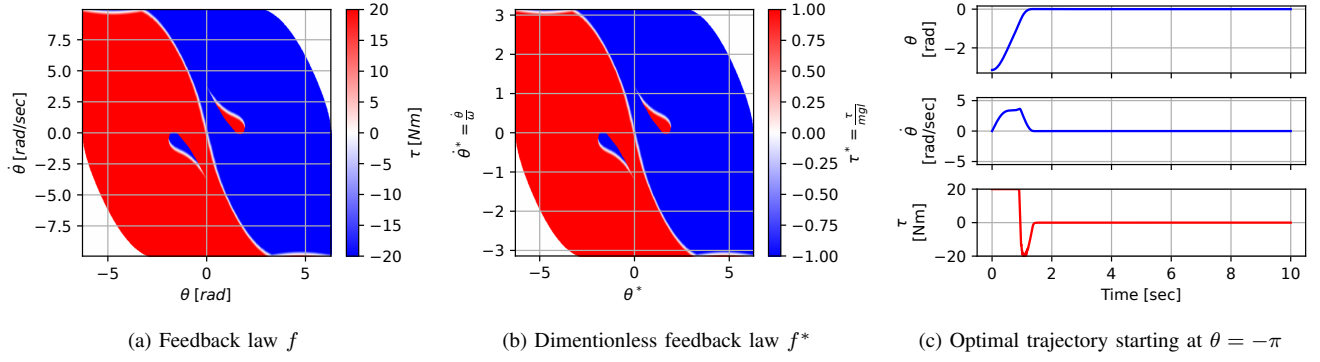


Fig. 13: Numerical results for context no 9

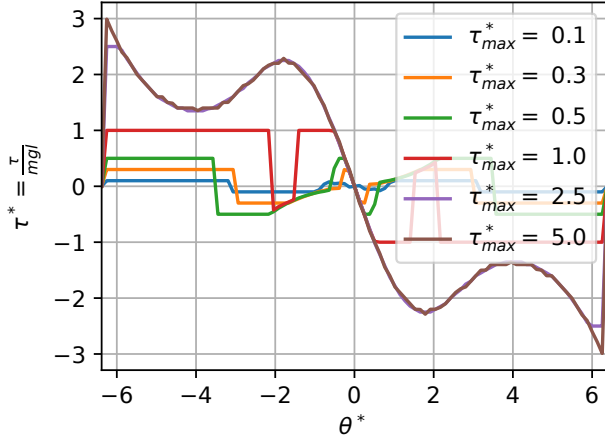


Fig. 14: Optimal dimensionless policy for various contexts $\tau^* = \pi^*(\theta^*, \dot{\theta}^* = 0, q^* = 0.5, \tau_{max}^* = [0.1, \dots, 5.0])$

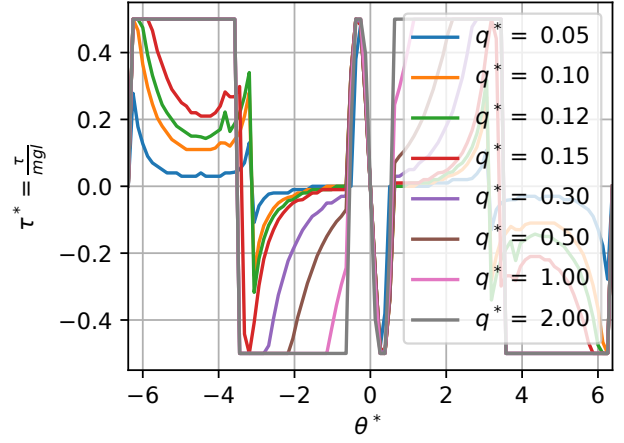


Fig. 15: Optimal dimensionless policy for various contexts $\tau^* = \pi^*(\theta^*, \dot{\theta}^* = 0, q^* = [0.05, \dots, 2.0], \tau_{max}^* = 0.5)$

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 dimensionless context space at Fig. 16. Those regions are best characterized by a ratio of q^* and τ_{max}^* , a new dimensionless value that we will define as:

$$R^* = \frac{\tau_{max}^*}{q^*} = \frac{\tau_{max}}{q} \quad (44)$$

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 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}^* :

$$\pi^*(\theta^*, \dot{\theta}^*, q^*, \tau_{max}^*) \approx \pi^*(\theta^*, \dot{\theta}^*, \tau_{max}^*) \text{ if } R^* \ll 1 \quad (45)$$

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

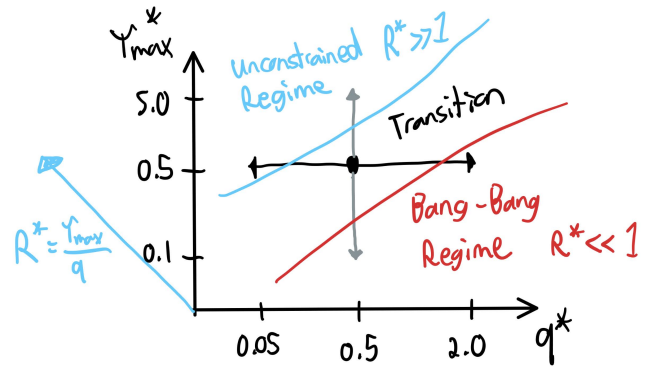


Fig. 16: Regime zones

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:

$$\pi^*(\theta^*, \dot{\theta}^*, q^*, \tau_{max}^*) \approx \pi^*(\theta^*, \dot{\theta}^*, q^*) \text{ if } R^* \gg 1 \quad (46)$$

The concept of regime is often leveraged in fluid mechanics, it allow to generalize results between situation where the relevant dimensionless number are not exactly matched.

For instance, when the Mach number is small $Ma < 0.3$, we can generally 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 transferring policy solution between different system or context, it means that the condition of having the same exact dimensionless 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 dimensionless policies. More formally, from eq. (46), we can say that:

$$f_a^*(x^*) = f_b^*(x^*) \forall x^* \quad (47)$$

$$\text{if } q_a^* = q_b^*, \quad R_a^* \gg 1 \quad \text{and} \quad R_b^* \gg 1 \quad (48)$$

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 dimensionless policies. More formally, from eq. (45), we can say that:

$$f_a^*(x^*) = f_b^*(x^*) \forall x^* \quad (49)$$

$$\text{if } \tau_{max,a}^* = \tau_{max,b}^*, \quad R_a^* \ll 1 \quad \text{and} \quad R_b^* \ll 1 \quad (50)$$

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 dimensional analysis. All in all, the impact of having such regimes indentified is that we can increase the sub-space of contexts for which the dimensionless 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/1kf3apyH1f5t7XzJ3uVM8mgDsneVK_63r?usp=sharing.

1) *Additionnal dimensionless 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:

$$\theta_{max}^* = \theta_{max} = 2\pi \quad (51)$$

$$\dot{\theta}_{max}^* = \frac{\dot{\theta}_{max}}{\omega} = \pi \quad (52)$$

$$t_f^* = t_f \omega = 10 \times 2\pi \quad (53)$$

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 dimensionless ratio scaled with the natural frequency is set at 2. t_f is the time horizon, here its associated dimensionless ratio is fixed to always corespond to 10 periods of the pendulum using the natural frequency.

IV. CLOSED-FORM PARAMETRIC POLICIES

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

A. Dimensionless 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

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

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

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

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

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

having the dimension presented at table 56.

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

$$p = (1 + (n = 2) + (m = 2) + (l = 2)) - (p = 2) = 5 \quad (57)$$

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

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

$$\Pi_2 = \theta^* = \theta \quad [-] \quad (59)$$

$$\Pi_3 = \dot{\theta}^* = \frac{\dot{\theta}}{\omega} \quad \frac{[T^{-1}]}{[T^{-1}]} \quad (60)$$

$$\Pi_4 = \omega_d^* = \frac{\omega_d}{\omega} \quad \frac{[T^{-1}]}{[T^{-1}]} \quad (61)$$

$$\Pi_5 = \zeta^* = \zeta \quad [] \quad (62)$$

Then, applying the Buckingham π theorem tell us that

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

Here we can confirm directly since we have an analytical solution: dividing eq (55) by mgl and making substitution based on the dimensionless group definition, leads to :

$$\tau^* = \sin \theta - 2\omega_d^*\zeta^*\dot{\theta}^* - (\omega_d^*)^2\theta \quad (64)$$

B. Dimentionless Linear Quatratic Reglator (LQR) solution

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

$$ml^2\ddot{\theta} - mgl\theta = \tau \quad (65)$$

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

$$J = \int (q^2\theta^2 + 0\dot{\theta}^2 + 1\tau^2)dt \quad (66)$$

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

$$\underbrace{\tau}_{\text{inputs}} = \pi \left(\underbrace{\theta, \dot{\theta}}_{\text{states}}, \underbrace{m, g, l}_{\text{system parameters}}, \underbrace{q}_{\text{task parameters}} \right) \quad (67)$$

We can thus select the same dimentionless group as before,

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	$[ML^2T^{-2}]$

and conclude that eq. (67) can be restated under this form:

$$\tau^* = \pi^* \left(\theta, \dot{\theta}^*, q^* \right) \quad (68)$$

For this motion control problem, an analytical solution exist (for the development, see alex screen shot 4 july 2023), and the policy is

$$\tau = [mgl]\theta + \left[\sqrt{(mgl)^2 + q^2} \right] \dot{\theta} + \quad (69)$$

$$\left[\sqrt{2ml^2} \sqrt{mgl + \sqrt{(mgl)^2 + q^2}} \right] \dot{\theta}^* \quad (70)$$

It is possible to, by this and that., arrive to this dimentionless form:

$$\tau^* = \left[1 + \sqrt{1 + (q^*)^2} \right] \theta + \left[\sqrt{2} \sqrt{1 + \sqrt{1 + (q^*)^2}} \right] \dot{\theta}^* \quad (71)$$

That correspond to the form of eq. (68) predicted by the dimentional analysis.

I am not sure where I am going with this..

V. CONCLUSION

The concept of dimensionless context is powerful, in the sense that it shows how to transfer a control policy to different context where the results should be exactly equivalent. However, it is limited because if the dimensionless 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 dimensional analysis, then the dimensional analysis results might be wrong. On the other hand, if we include too many variables to fully describe a context, then dimensionally similar context space will probably be so specific it won't be practical to use for transferring policy between systems. Henceforth, finding the appropriate parametrization of the context will be critical in order to leverage this principle for sharing policy between similar system,

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.