

Stabilisation d'un pendule inversé à l'aide d'un algorithme d'apprentissage automatique

Passionné par les récentes percées des technologies spatiales, le maintien à l'équilibre des véhicules spatiaux m'a toujours intrigué. Je m'en suis alors inspiré pour étudier un système plus simple, celui du pendule inversé. J'ai choisi d'améliorer les mécanismes classiques de stabilisation en y introduisant un algorithme d'apprentissage automatique.

Mon étude porte sur la comparaison entre deux boucles de régulation du pendule inversé : une première boucle de rétroaction classique, reposant sur une loi fixe de correction, et une deuxième algorithmique d'un agent apprenant par renforcement.

Le candidat atteste avoir travaillé en monôme.

Positionnement thématique (ÉTAPE 1) :

- *PHYSIQUE (Mécanique)*
- *INFORMATIQUE (Informatique pratique)*
- *MATHEMATIQUES (Autres)*

Mots-clés (ÉTAPE 1) :

Mots-clés (en français)	Mots-clés (en anglais)
<i>pendule inversé</i>	<i>Inverted pendulum</i>
<i>Stabilisation</i>	<i>Stabilization</i>
<i>Contrôle LQR</i>	<i>LQR control</i>
<i>Apprentissage automatique</i>	<i>Machine learning</i>
<i>Apprentissage par renforcement</i>	<i>Reinforcement learning</i>

Bibliographie commentée

Le problème du pendule inversé occupe historiquement une position centrale dans les sciences du contrôle et de l'automatique. Son intérêt réside dans sa capacité à concentrer, dans un système à dimension réduite, des défis théoriques majeurs tels que l'instabilité, la non-linéarité et la sensibilité aux perturbations. Ces défis font du pendule inversé un banc d'essai privilégié pour l'évaluation des méthodes de commande retrouvées actuellement dans des applications technologiques de grandes envergures, telles que le décollage des fusées, leurs recyclage, l'équilibrage des aéronefs, ou encore la stabilisation de véhicules auto-équilibrés comme les Segways et les robots bipèdes [1].

Pendant des décennies, la stabilisation du pendule inversé se réalisait à l'aide de méthodes classiques [2]. L'on trouve parmi ces dernières, la méthode de stabilisation LQR (*Linear Quadratic Regulator*). Cette technique repose sur une *boucle de rétroaction linéaire* et cherche à minimiser une fonction de coût quadratique (compromis erreur/énergie). Son atout majeur est la précision mathématique et la rapidité d'exécution [3].

Toutefois, ces méthodes classiques, basées sur des modèles analytiques et des hypothèses de validité, présentent plusieurs limites, telles que ; la fragilité face aux perturbations fortes et aux non-linéarités. A cela s'ajoute les incertitudes de mesure de la masse du pendule, des coefficients de frottement, des dimensions du système, etc.

Ces limites ont historiquement conduit à la recherche d'alternatives capables de produire un comportement stabilisant sans la nécessité d'une connaissance exhaustive de la dynamique du système. L'une de ces alternatives innovantes est l'apprentissage par renforcement. C'est un cadre qui s'inscrit dans le domaine de l'intelligence artificielle et qui s'appuie sur des approches algorithmiques d'apprentissage par interaction avec l'environnement. Dans ce contexte, un agent apprend progressivement une politique de stabilisation à partir de l'évaluation de ses actions. Les fondements théoriques de cette technique moderne sont expliqués en détail dans l'ouvrage de Sutton & Barto [4].

L'avènement du *Q-learning*, l'une des branches de l'apprentissage par renforcement, a notamment marqué une étape décisive, démontrant qu'une politique optimale peut être apprise de manière incrémentale selon des cycles d'exploitation-exploration, et sans connaissance explicite des équations du système [5],[6]. Cette capacité à apprendre par essais-erreurs ancre résolument le problème du pendule inversé dans les technologies d'intelligence artificielle modernes, offrant une alternative puissante face aux limites des approches analytiques classiques.

Problématique retenue

Comment implémenter un algorithme d'apprentissage automatique au sein du pendule inversé afin de fournir une meilleure robustesse au contrôle ?

Objectifs du TIPE du candidat

- Mettre en équation le système du pendule inversé.
- Appliquer un contrôleur LQR pour ce système.
- Implémenter l'apprentissage par renforcement (RL) pour réguler le pendule inversé.
- Comparer les performances de ces deux approches.

Références bibliographiques (ÉTAPE 1)

- [1] O. BOUBAKER : The inverted pendulum: A fundamental benchmark in control theory and robotics : *in Proc. Int. Conf. Education and E-Learning Innovations*, pp. 1–6, Jul. 2012
- [2] L. B. PRASAD, B. TYAGI, ET H. O. GUPTA : Optimal control of nonlinear inverted pendulum system using PID controller and LQR: performance analysis without and with disturbance input, : *International Journal of Automation and Computing*, vol. 11, no. 6, pp. 661–670, 2014
- [3] D. R. TYNER ET A. D. LEWIS : Geometric Jacobian linearization and LQR theory : *Journal of Geometric Mechanics*, vol. 2, no. 4, pp. 397–440, 2010
- [4] SUTTON, R. S. ET BARTO, A. G. : Reinforcement Learning: An Introduction : 2nd ed. Cambridge, MA, USA: MIT Press, 2018
- [5] S. ISRAILOV, L. FU1, J. SANCHEZ-RODRIGUEZ, F. FUSCO, G. ALLIBERT, C. RAUFASTE, ET M. ARGENTINA : Reinforcement learning approach to control an inverted pendulum: A general framework for educational purposes : *PLOS ONE*, vol.18,no.2, p. e0280071, Feb. 2023, doi: 10.1371/journal.pone.0280071
- [6] T. V. BUI : Stabilisation d'un double pendule inversé par microprocesseur : *Ph.D. dissertation*, École Nationale Supérieure des Mines de Paris, Paris, France, 1983

DOT

- [1] : Mai 2025 - Étude de la stabilisation LQR du pendule inversé et réalisation d'une première simulation Python afin d'identifier les variables d'état, la commande et les critères de stabilité.
- [2] : Juillet 2025 - Recherche documentaire sur l'apprentissage par renforcement appliqué à la commande. Identification des éléments nécessaires : état observé, action de commande, fonction de récompense et critères d'arrêt

[3] : *Septembre 2025 - Passage à une simulation Python pour tester une commande par apprentissage sur un modèle simplifié du pendule inversé. Les premiers essais ont montré une forte sensibilité au choix de la fonction récompense.*

[4] : *Janvier 2026 - Tentative d'utilisation d'une commande purement intelligente. Les résultats obtenus étant peu exploitables, la stratégie a été réorientée vers une approche combinant stabilité théorique et adaptation.*

[5] : *Février 2026 - Décision de concevoir une méthode hybride associant LQR et apprentissage par renforcement, afin de conserver une base stabilisante tout en améliorant la réponse aux perturbations.*

[6] : *Début Mai 2026 - Implémentation de l'approche hybride, réglage des paramètres d'apprentissage et tests sur différents scénarios. Le choix de la fonction récompense a joué un rôle crucial dans le succès de la simulation*

[7] : *Fin Mai 2026 - Comparaison finale des approches LQR, intelligente et hybride. Sélection des courbes les plus pertinentes, vérification des hypothèses utilisées et préparation de la présentation orale.*