Commande mixtes par apprentissage par renforcement et mode glissant avec application à un bras robotique

Amine MEBARKI, Mohamed ZERROUGUI

31 Mai 2024

SAGIP 2024

Outline

Introduction

Commande des Bras Robotiques

- **Commande Classique :**
  - Modélisation dynamique et cinématique.
  - Techniques : PID...
- **Commande Adaptatif :**
  - S'adapte aux changements des paramètres du système.
  - Techniques : Contrôle Adaptatif par Modèle de Référence (MRAC)...
- **Commande Basé sur l'Apprentissage :**
  - Utilisation des données.
  - Techniques : Apprentissage par renforcement (DDPG, PPO,...)
- **Commande Mixte :**
  - Combine des approches basées sur le modèle et l'apprentissage.

Avantages et Inconventients

Commande classisque

- Nécessite souvent un modèle assez précis de la dynamique du système.
- Possibilité d'améliorer la robustesse vis-à-vis les perturbations et variations prévisibles.
- Étude de stabilité (par Lyapunov).

Apprentissage par Renforcement

- Ne nécessite pas un modèle précis du système voir sans modèle.
- S'adapte aux changements du système (Apprentissage en temps réel).
- Risques liés à la stabilité (et sécurité) lors de l'entraînement (en temps réel).

Proposition : Conception d'une commande mixte qui combine les avantages des deux commandes tout en attenuant leurs incovenients.

Contexte

- **Objectif :**
  Garantir la stabilité et la convergence d'un système non-linéaire dont la dynamique n'est pas parfaitement connue.
- **Méthode :**
  La conception d'une loi de commande additive par mode glissant -SMC- et par renforcement -RL-.
- **Application / Système d'intérêt :**
  Les bras manipulateurs (robotiques) à plusieurs degrées de liberté.

Commande Mixte SMC-RL

Formulation du problème

Modèle Dynamique : Euler-Lagrange

Les dynamiques d'un bras manipulateur peuvent être représentées à l'aide de la formulation d'Euler-Lagrange comme suit [1] :

$$M(q)\ddot{q} + C(q,\dot{q}) + G(q) = \tau + d(t), \quad d(t) \leq \bar{d} \quad \forall \quad t \geq 0$$

On réecrit le système sous la forme :

$$\ddot{q} + H(q,\dot{q}) = B(q)\tau + f(q,t), \quad f(q,t) \leq F$$

Où :

- $H(q,\dot{q}) = M^{-1}(q)(C(q,\dot{q}) + G(q))$
- $f(q,t) = M^{-1}(q)d(t)$

0. [1] . W. Spong, S. Hutchinson, and M. Vidyasagar, Robot modeling and control. John Wiley Sons, 2020.

Formulation du problème

On prend $H(q, \dot{q}) = H_0(q, \dot{q}) + H_1(q, \dot{q})$, tel que :

- $H_0(q, \dot{q})$ représente la partie nominale (connue) du système.
- $H_1(q, \dot{q})$ représente les dynamiques non-connues/incertaines.

Nous proposons une loi de commande globale qui est constituée de deux parties :

- u_{smc}, commande par mode glissant, pour le cas nominal $H_1(q, \dot{q}) = 0$
- u_{rl}, commande issue de l'apprentissage par renforcement, pour le cas $H_1(q, \dot{q}) \neq 0$.

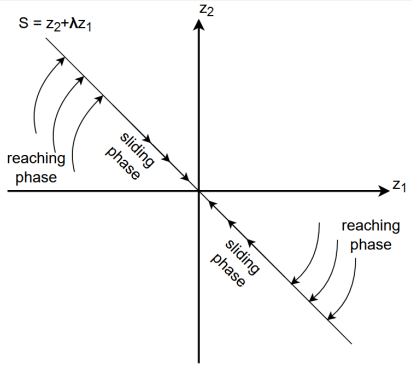
La loi de commande globale serait : $u = u_{rl} + u_{smc}$

Commande par Mode Glissant

La commande par mode glissant est une commande robuste dont l'idée est de commander un système d'ordre un à la place du système d'ordre supérieur [2].

La commande par mode glissant se résume par le choix de :

- La variable de glissement S.
- La commande equivalente u_{eq} ($\dot{S} = 0$).
- La commande commutante u_{sw} ($\dot{S} < 0$).



0. [2] J.-J. E. Slotine, W. Li et al., Applied nonlinear control. Prentice hall Englewood Cliffs, NJ, 1991.

Apprentissage par Renforcement

L'apprentissage par renforcement est une approche d'apprentissage automatique visant à enseigner aux agents comment résoudre des problèmes de prise de décision par essais et erreurs.

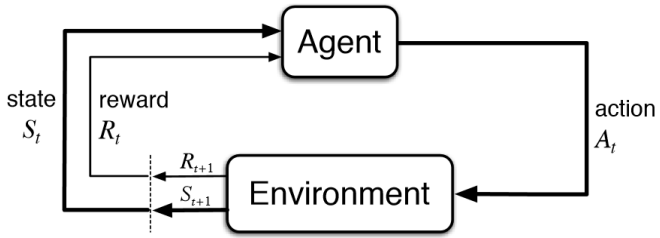


Figure – Schéma générique d'apprentissage par Renforcement [3]

0. [3] R. S. Sutton, A. G. Barto, Reinforcement learning : An introduction, MIT5 press, 2018.

Synthèse de la Loi de Commande (partie SMC)

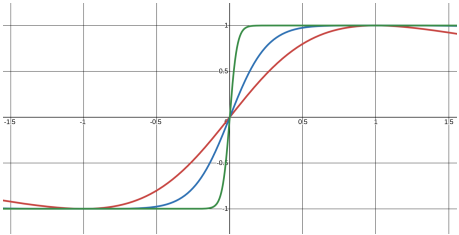
On définit :

- $e = q_d - q$, $\dot{e} = \dot{q}_d - \dot{q}$ et $S = \dot{e} + \lambda e$.
- La fonction Lyapunov $V = \frac{1}{2}S^2$.

La commande nominale serait donc : $u_{\text{smc}} = u_{\text{sw}} + u_{\text{eq}}$. Où :

- $u_{\text{eq}} = B^{-1}(q)(\lambda\dot{e} + \ddot{q}_d) + M(q)H_0(q, \dot{q})$
- $u_{sw} = B^{-1}(q)(F + K)sgn(S)$

Pour réduire le "chattering" on peut utiliser "scalar sign function"



Synthèse de la Loi de Commande

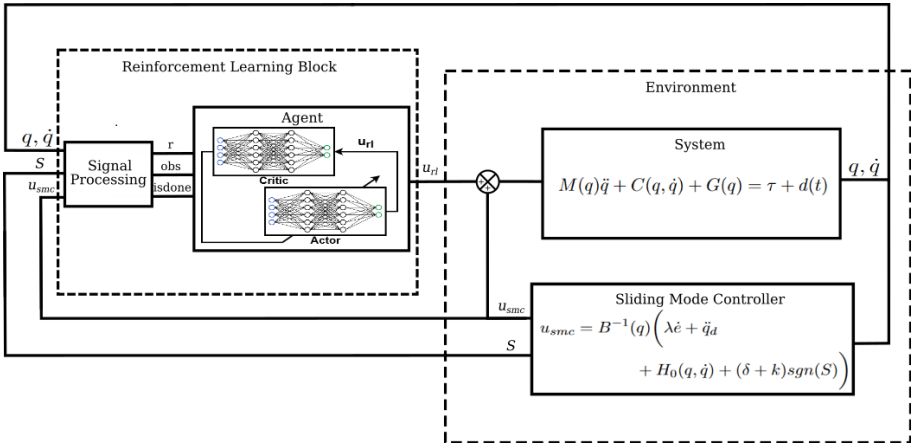


Figure – Block diagramme de la commande mixte

Approche d'entraînement

- **Algorithme :**
  - Structure utilisée : Actor-Critic.
  - Algorithme : Deep Deterministic Policy Gradient (DDPG).
- **Entraînement :**
  - Récompense basée sur la variable de glissement.
  - Pré-entraînement avant le déploiement sur le système réel.
  - L'entraînement est déclenché sur des intervalles connues.
- **Déploiement :**
  - Les valeurs de l'acteur sont copiées dans un réseau de neurones.
  - Loi d'adaptation appliquée pour le réglage fin.

Stabilité au sens Lyapunov

La politique de l'agent d'apprentissage par renforcement est de la forme :

$$u_{rl} = \pi(q, \dot{q}) = W_a^T \phi(\Psi_a(q, \dot{q})) = W_a^T \phi_a(q, \dot{q})$$

On choisi $S = \dot{e} + \lambda e$. En substituant la commande $u = u_{smc} + u_{rl}$, nous obtenons :

$$\dot{S} = H_1(q, \dot{q}) - Ksign(S) - u_{rl} + f(q, t) - Fsign(S)$$

On suppose que $H_1(q, \dot{q})$ peut être parfaitement représentée par :

$$H_1(q, \dot{q}) = W_H^T \phi(\Psi_H(q, \dot{q})) + \epsilon_H = W_H^T \phi_H(q, \dot{q}) + \epsilon_H$$

Hypothèse

Après le pré-entraînement, la sortie de l'algorithme DDPG satisfait :

$$W_H^T (\phi_H(q, \dot{q}) - \phi_a(q, \dot{q})) \leq \bar{\delta}$$

La fonction Lyapunov choisie est de la forme :

$$V(q, \dot{q}) = S^T S + tr(\Omega^T \Omega), \quad \Omega = W_H - W_a, \quad \dot{\Omega} = -\dot{W}_a$$

Stabilité au sens Lyapunov

Théorème

Sous l'hypothèse précédente, il est possible de choisir un gain approprié K qui garantit la stabilité du système tel que :

$$K \geq \bar{\delta} + \bar{\epsilon} \quad \text{et} \quad F \geq |f(q,t)|$$

$$\dot{V}(q,\dot{q}) = S^T \dot{S} + tr(\Omega^T \dot{\Omega})$$

$$\dot{V}(q,\dot{q}) = S^T \Big(W_H^T(\phi_H(q,\dot{q}) - \phi_a(q,\dot{q})) + \Omega^T \phi_a(q,\dot{q}) + \epsilon_H - Ksign(S) + f(q,t) - Fsign(S) \Big) + tr(\Omega^T(-\dot{W}_a))$$

En prennant la loi d'adaptation $\dot{W}_a = \phi_a S^T$ [4] :

$$\dot{V}(q,\dot{q}) = S^T \Big(\boxed{W_H^T(\phi_H(q,\dot{q}) - \phi_a(q,\dot{q})) + \epsilon_H - Ksign(S)} + f(q,t) - Fsign(S) \Big)$$

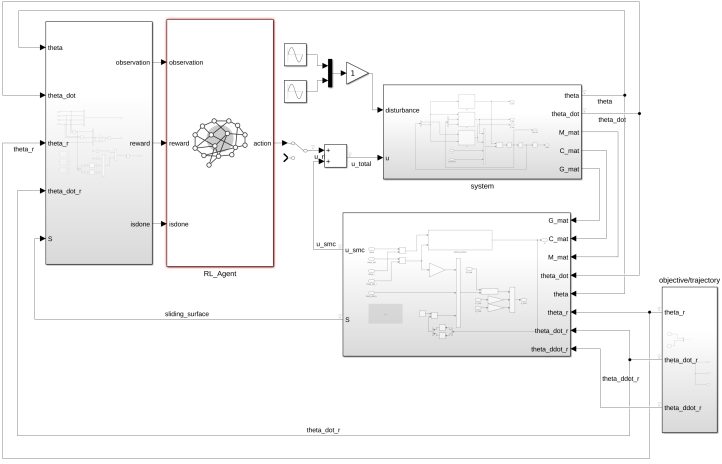
0. [4] Edoardo Vacchini et al. "Design of a deep neural network-based integral sliding mode control for nonlinear systems under fully unknown dynamics". In : IEEE Control Systems Letters(2023).

Simulation et Résultats

Simulation et Résultats

Simulation réalisée à l'aide de MATLAB Simulink.

Objectif de commande : converger vers un point cible fixe $q = [0, 0]$ et $\dot{q} = [0, 0]$.



Simulation et Résultats (SMC seul)

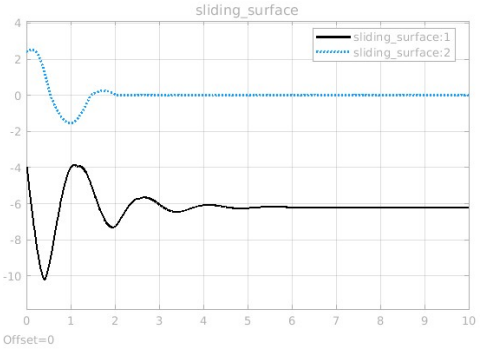


Figure – Surface de Glissement

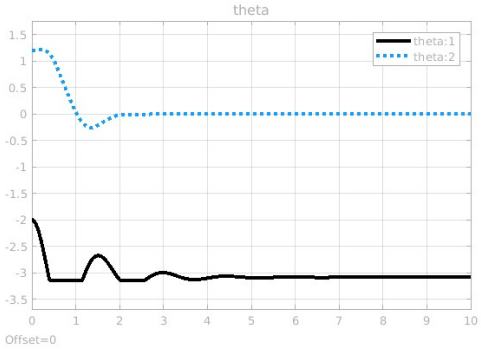


Figure – Positions des Joints

Simulation et Résultats (Commande Mixte)

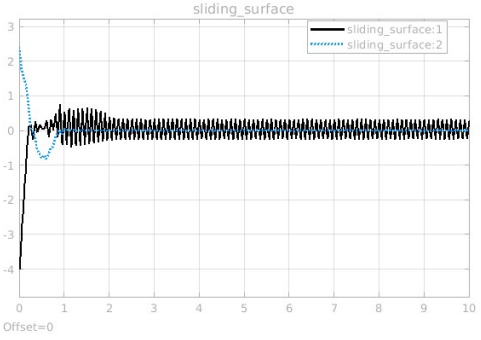


Figure – Surface de Glissement

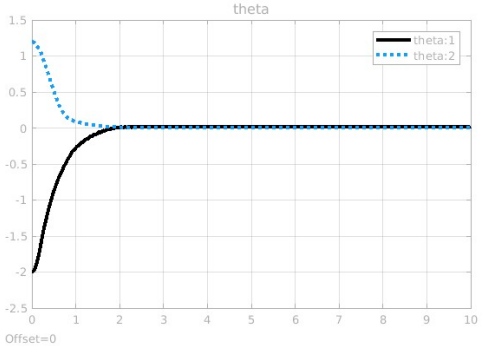


Figure – Positions des Joints

Conclusions et Perspectives

Conclusions et Perspectives

- Conslusions :
  - Méthodologie pour gérer les dynamiques inconnues par l'intégration des méthodes de commande non linéaire robuste et d'apprentissage.
  - Étude et proposition des conditions de stabilité basées sur une fonction de Lyapunov mixte.
- Perspectives :
  - Évaluer l'efficacité de la méthode proposée dans des applications réelles.
  - La conception d'un algorithme (actor-critic) assurant la stabilité du système sans nécessiter un réseau de neurones supplémentaire.