非線形最適レギュレータにおけるモデルフリー再設計のための二段階強化学習

Two-step reinforcement learning for model-free redesign of nonlinear optimal regulator

堀研究室　学籍番号：82213188　南 芽衣

1. 研究背景・目的

多くの実用的な制御システムには閉ループシステムが含まれる．しかし，閉ループシステムのモデリングは困難であり，精度の良いモデルを利用できない場合が多く存在する．このとき，閉ループシステムの入力や出力の測定値のみに基づいて（モデルフリー），制御器を再設計もしくはチューニングすることが求められる．この方法として，強化学習[1](RL) を用いたアプローチが近年活発に研究されている[2]．強化学習では試行錯誤を通じて最適な制御器を完全自律的に学習する手法である．しかし強化学習は学習過程において，性能の悪い制御器を用いた試行を何度も繰り返すため，制御対象の損耗を引き起こす．そこで，学習過程において閉ループシステムの性能レベルをある程度維持でき，必要な試行回数を削減できる学習手法の開発が望まれている．

そこで本稿では，二段階の設計によって従来の強化学習[1] に比べて過渡学習性能を向上させた，非線形最適レギュレータを設計するための完全モデルフリーアプローチを提案する．

2. 二段階強化学習

状態 $\boldsymbol{x}_k \in \mathbb{R}^n$，入力 $\boldsymbol{u}_k \in \mathbb{R}^m$ に関するコスト関数

$$\sum_{k=0}^{\infty} \gamma^k (\boldsymbol{x}_k^\top Q \boldsymbol{x}_k + \boldsymbol{u}_k^\top R \boldsymbol{u}_k) \tag{1}$$

を最小にするような制御器を設計する問題を考える．ただし，$Q \in \mathbb{R}^{n\times n}, R \in \mathbb{R}^{m\times m}$ はそれぞれ半正定値，正定値行列であり，$\gamma \in (0,1)$ は定数である．この問題に対し，Fig. 1 で表される制御器の設計を 2 つのステップに分けて行う．

Step 1 ではまず，既知の安定なフィードバックゲイン K^{init} を用い，探索項を印加しながらシステムを動かすことで入力と状態の時系列データを取得する．その後，ベルマン方程式から導出される関係式

$$F^j[\mathrm{vec}(G_1^{j+1})^\top, \mathrm{vec}(G_2^{j+1})^\top, \mathrm{vec}(G_3^{j+1})^\top]^\top = \boldsymbol{h}^j \tag{2}$$

によって G_i^{j+1}, $i = 1,2,3$ を更新し，$K^{j+1} = -(R + G_3^{j+1})^{-1} G_2^{j+1}$ に代入することで K^j の更新を行う．ただし，$\boldsymbol{h}^j$ は各時刻における二次コストに関するベクトル，F^j は入力と状態に関する行列，G_i^{j+1} はコスト関数に関する行列であり，F^j, $\boldsymbol{h}^j$ は取得した時系列データと K^j から計算される．$||K^j - K^{j-1}||$ が閾値以下になったら更新を終了し，このときの K^j を K^{AC} とする．更新される K^j は，準最適線形 LQR 制御則に収束する．

Step 2 では，Step 1 で設計された線形制御器と並列に繋げた非線形制御則 $\boldsymbol{\mu}$ を強化学習の一種である Actor-Critic 法[1] によって設計する．線形制御則 K^{AC} が存在することで，学習中にシステムが不安定になることを防ぎ，損耗を防ぐことができる．この二段階の設計手法による制御則は，線形制御則だけでは実現できない性能を達成する．

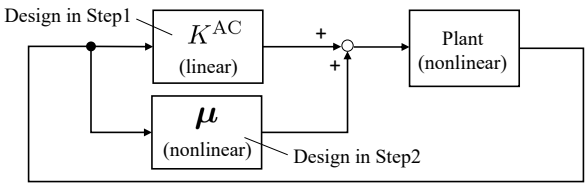

Fig. 1: Structure of the proposed method

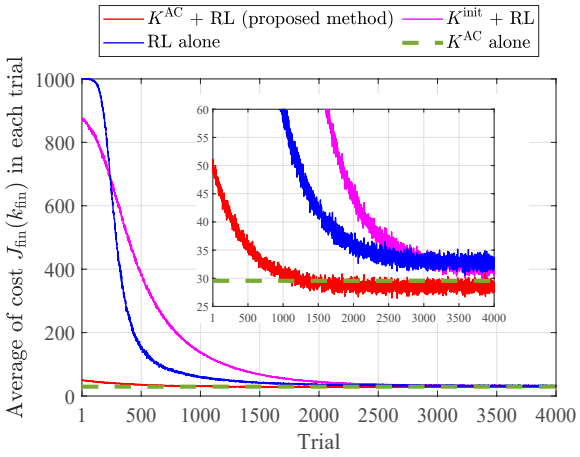

Fig. 2: Cost $J_{\mathrm{fin}}(k_{\mathrm{fin}})$ in each trial in Step 2. Inset is plotted with in a different scale of the vertical access.

3. 数値例による検証

数値シミュレーションを用いて，提案法 (K^{AC} + RL)，K^{init} と強化学習を組み合わせた方法（K^{init} + RL），強化学習単独 (RL alone)，補助制御器単独 (K^{AC} alone) の 4 つの場合について，倒立振子の制御器設計を行った．Step 2 における各試行ごとの二次コストを Fig. 2 に示す．Fig. 2 より，提案法は他の方法と比べて，学習初期のコストが小さく，過渡学習性能が良いことが示された．

4. 結論と今後の展望

本稿では，数理モデルが未知な非線形制御対象に対する最適制御則の再設計において，過渡的な学習性能を向上させた学習法である二段階強化学習を提案した．具体的には，まずオフラインの学習則で一定の性能を実現する線形制御則を設計し，次に強化学習器と線形制御則を並列に用いて強化学習を行うことで，線形制御則だけでは実現できない性能を実現する非線形制御則を設計した．今後は Step 2 における学習中の安全性を加味した方法を構築する予定である．

参考文献

[1] R. S. Sutton *et al.*, 2nd ed. *MIT Press*, 2018.

[2] B. Kiumarsi *et al.*, *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 6, pp. 2042–2062, 2018.