

多関節制御器の協調運動の構造抽出と制御戦略の提案

Structural extraction of coordinated motion of articulated controllers and proposal of control strategies

田澤 龍之介[†], 鳥居 拓馬[†], 日高 昇平[†]
 Ryunosuke Tazawa, Takuma Torii, Shohei Hidaka

[†] 北陸先端科学技術大学院
 Japan Advanced Institute of Science and Technology University
 s2120023@jaist.ac.jp

概要

多自由度・高次元系の運動制御に強化学習手法が適用されている。強化学習は、報酬を最大化する制御則を学習する枠組みである。報酬をコスト関数とすれば、最適制御の枠組みと近い。制御器が取り得る軌道は無数に存在する。無数の軌道から1つを選択する、つまり不良設定問題は最適制御における最大化問題である。本研究は、不良設定問題を解決する身体運動の制御機構を基に、多関節制御器のための新たなカリキュラム学習の枠組みを提案する。

キーワード：不良設定性, 自由度, 協調, カリキュラム学習

1. 背景と目的

作業を人間からロボットへ移行するロボタイゼーション技術の急速な進歩により、我々の日常生活の多くの場面において、ロボットの活躍が期待されている。日常の様々な場面で適切に動作するロボットの実現には、未知な環境に対応できるシステムの付与が求められる。動的に変化し続ける実環境において、環境との相互作用を通じ、何らかの形でシステム内部に目的動作を生成するための環境の記述が必要であり、学習過程が欠かせない。ロボットの学習手法として、種々のアプローチが提案されているが、目標状態のみを示すだけで、そこに至る入出力関係を構築できるという実装の簡便さと反響的かつ適応的な行動を獲得できることから、強化学習手法がロボットの学習手法として、注目を浴びている [1]。

これまで、マニピュレーター [2], 車輪型 [3], 二足 [4] 及び多足型 [5] などの移動ロボット, 他にも、蛇型ロボット [6] の制御などが実現されてきた。しかし、ロボットの運動制御に強化学習を従来の枠組みのまま適用するだけでは、多くの場合で、動作が困難となる。要因として以下が挙げられる。

- 制御自由度の増加に伴い、状態・行動空間が連続

かつ高次元となり、次元の呪いが発生する。

- 報酬を用いたゴール状態の定義について、明示的な指針がない。タスク達成時のみに報酬を与える場合、その状態に至ることする難しい場合がある。

これらの要因は、運動制御における不良設定性によるものである。運動制御における不良設定性とは、目標を達成可能な身体部位の軌道が無数に存在し、一意に定まらないことである。本研究の目的は、この運動制御における不良設定性の解消に向けた新たな制御手法の提案である。次章では、強化学習の基本的な枠組みと強化学習における問題について述べる。

2. 強化学習手法を用いた運動学習

2.1 強化学習の枠組み

強化学習の枠組みは、エージェント、環境及びそれらの間の相互作用からなる。エージェントは、行動決定の主体であり、環境は、エージェントが相互作用を行う対象である。相互作用とは、情報の受け取りと引き渡しを行うことを意味する。

相互作用を記述する数理モデルとして代表的なのがマルコフ決定過程 (Markov Decision Process: MDP) である [7]。マルコフ決定過程に従う場合の強化学習の枠組みを図1に示す。エージェントと環境は、1時間ステップごとに、状態、行動、報酬の3つの情報を受

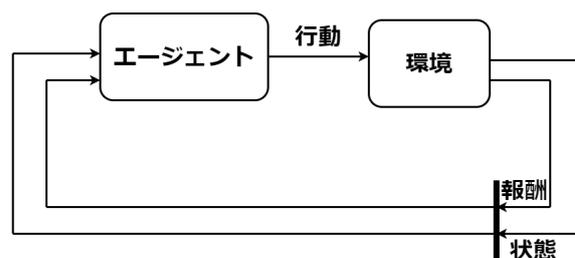


図1 強化学習の基本的な枠組み

け取ったり引き渡したりする。この図が示すように、情報のやり取りは、ループ構造となる。太い線は、その始点を表している。状態とはエージェントが置かれている状況を表し、倒立振り子を制御する場合、典型的には振り子の角度と角速度で状態を記述する。行動とはエージェントが環境に対して行う働きかけの種類であり、振り子の場合、トルクの大きさに相当する。最後に、報酬はその行動の即時的な評価を表す。エージェントは環境から現在の状態と報酬を受け取り、行動集合の中から行動を決定し、これを環境に引き渡す

2.2 強化学習における問題

エージェントの行動は方策により決定される。強化学習問題を解くということは、できるだけ多くの報酬を受け取れるように、この方策を設計することである。しかし、制御器の最適制御において、環境の挙動は未知であり、あらかじめ方策（制御則）を設計することは不可能である。従って、エージェントの行動結果（状態・報酬）を繰り返し観測することで最適な制御則を獲得するアルゴリズムの設計が具体的な課題となる。しかし、制御器の自由度の増加に伴い、探索空間は巨大となる。先にも述べたように、多自由度系には、システムの最終出力である運動軌道は無数に存在するため、探索空間から最適な制御側の発見は不良設定問題となる。これにより、多くの試行回数を要し、学習時間の増加及びサンプリング効率の低下を招く。この問題を解消するために、人間が容易にこの不良設定性を解消することに着目し研究を進めた。

3. 身体運動機構

3.1 不良設定性を有する運動系

人間の身体は多自由度系であり、各身体部位が複雑に相互作用する。しかし、実際、人間は滑らかに運動を生成し、目標の運動を容易に達成する。

脊椎動物は筋肉の内側に骨格を持つことで、柔軟に動く身体を獲得した。骨格が筋肉に覆われた筋骨格構造を人間は有し、筋張力により関節を駆動させたり、関節の剛性を変化することができる。この時、筋は収縮方向の力しか発生できないため、1自由度の関節を駆動させるには2本以上の筋が必要である。従って、身体運動制御は多数の筋骨格系を制御して始めて成立し、多数の関節と筋肉の冗長性の問題を解決しなければならない。

3.2 協調運動と制御自由度の変化

では、実際のところ人間はどのようにこの冗長問題を解決しているのかについて、これまで多くの議論がされてきた。この身体の冗長問題はベルンシュタイン問題 [8] と呼ばれ、ベルンシュタインは運動の自由度に着目し、運動制御における協調構造の存在を示唆した [8]。自由度とは、制御のために決定しなくてはならない値であり、各部位が動くことができる方向の数とも言える。例えば、単純な到達運動の場合、関節は7自由度を持つ（肩関節が3、肘関節は1、尺骨は1、手首は2方向に動ける）。体肢全体を考えれば、100以上の自由度がある。それゆえに、こうした動作の要素（関節や筋）を1つ1つ個別に制御することは非現実的である。この運動要素を制御可能なシステムへと転換しているのが、協調構造である [8]。協調とは、複数の独立な自由度が互いを機能的に拘束し合い、1つの機能的な単位として、結合することを言う。この状態にある時、それぞれの自由度はバラバラでなく、相互補完的に変動する。つまり、それまで、莫大であった自由度が制御可能な範囲になるまで、減少する。このように、身体運動には莫大な自由度が潜んでおり、協調により、各身体部位の自由度が増加・減少し、自在な運動が実現される。まず、身体運動の制御自由度の変化の調査を行った。具体的には、身体を力学系とみなし、投球運動課題を与え、その運動中の制御自由度を分析した。身体運動において、多くの座標変換が行われる。特に、複数部位が協調する時、動作間の空間座標系の変換の組み合わせ数が多くなり、逆モデルの計算コストが膨大となる。そのため、本研究では、身体運動の時間発展を記述する体系として、日高ら [9] の力学的不変量仮説を採用した。これは、末梢・中枢神経系・筋骨格系と作業空間との間の相互作用をそれらの系の状態とその状態遷移の系列とし、身体運動を力学系と捉えることである。ある運動の多くの軌道は、相空間上の特定の部分空間のみを通過し、この部分空間はその力学系のアトラクタに対応する。アトラクタの不変量の1つであるフラクタル次元は滑らかな座標変換に対する不変性を持つため、表層的な運動ではなく、それに潜在する運動制御、すなわち制御自由度を抽出できると考えられる。また、投球運動課題の主な選定理由は、課題成績が評価しやすく、投球目標の距離により容易に課題の難易度を操作できるため、身体の制御自由度と課題難度の関係を調べるのに適した課題であるからである。

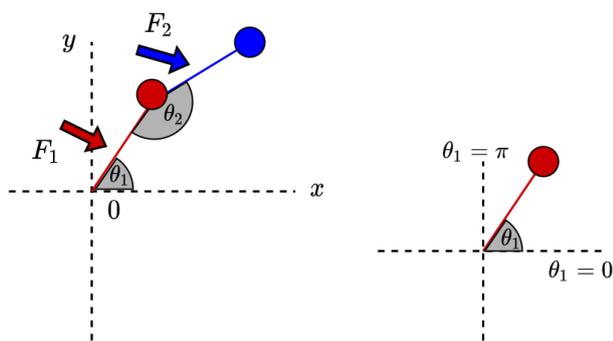


図2 腕の振り子モデルと角度軸

4. 二重振り子による投球運動課題

二重振り子の運動制御を力学系の構成と捉え、人間の腕を単純化したモデルとして、目標点に向けてボールを投げる投球運動課題に対して運動学習を行った。

4.1 人間の腕の振り子モデル

図2は設計した制御器を表している。上下関節はそれぞれ人間の肘・肩関節に対応し、両関節の質点は $M_1, M_2 = 1[\text{kg}]$ 、長さは $L_1, L_2 = 1[\text{m}]$ である。観測される状態は両関節の角度と角速度 $(\theta_1, \dot{\theta}_1), (\theta_2, \dot{\theta}_2)$ の4変数である。角度は、上関節は下関節となす角度を、下関節は x 軸となす角度とした。それぞれ、時計周り方向に角度の値が減少していくもとする。また、各関節への入力トルク F_1, F_2 は時計周り方向にのみ与え、両関節とも $0 \leq (F_1, F_2) \leq 50[\text{N}]$ と制限した。

二重振り子の運動方程式はラグランジュ関数 [10] を用いて導出され、以下の式で表される (重力加速度 $g = 9.8[\text{m/s}^2]$)。

$$\begin{pmatrix} \ddot{\theta}_1 \\ \ddot{\theta}_2 \end{pmatrix} = \begin{bmatrix} (M_1+M_2)L_1 & M_2L_2 \cos(\theta_1-\theta_2) \\ M_2L_1 \cos(\theta_1-\theta_2) & M_2L_2 \end{bmatrix}^{-1} \begin{bmatrix} (M_1+M_2)g \sin \theta_1 & M_2L_2 \dot{\theta}_2^2 \sin(\theta_1-\theta_2) \\ M_2L_1 \dot{\theta}_1^2 \sin(\theta_1-\theta_2) & M_2g \sin \theta_2 \end{bmatrix} \begin{pmatrix} -F_1 \\ -F_2 \end{pmatrix} \quad (1)$$

オイラー法 [11] を用いて、この運動方程式を数値的に解いた。

4.2 投球課題と課題パフォーマンス

投球運動課題において、振り子の端点から、初期条件を $(\theta_2, \dot{\theta}_2)$ として、ボールが発射される。空気抵抗

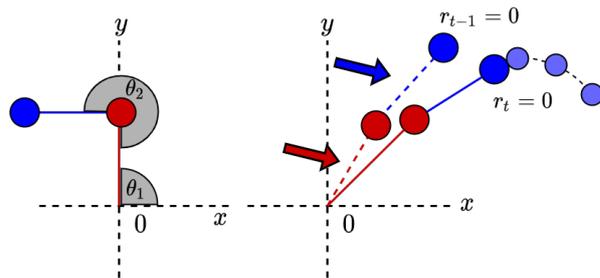


図3 初期値と報酬

(抗力)の影響は受けず、ボールは水平投射運動に従う。ボールの初速度は上関節の角速度の水平方向成分のみで構成される。ボールの位置と速度をデカルト座標系に変換し、 $(x_0, y_0, \dot{x}_0, \dot{y}_0) = h(\theta_1, \dot{\theta}_1, \theta_2, \dot{\theta}_2)$ の形で表現すると、位置座標は $(x_0, y_0) = (x_p + L_1 \cos \theta_1 + L_2 \cos \theta_2, y_p - L_1 \sin \theta_1 - L_2 \sin \theta_2)$ 、ボールの速度は $(\dot{x}_0, \dot{y}_0) = (\dot{\theta}_2 \cos \theta_2, 0)$ となる。 $(x_p, y_p) = (0, 0)$ は振り子の旋回軸である。デカルト座標系におけるボールの初期条件 $(x_0, y_0, \dot{x}_0, \dot{y}_0)$ から、地面に衝突したボールの到達点 (x_*, y_*) は次の式となる。また、ボールが最初に地面に到達した点から枢軸 (x_p, y_p) までの距離を課題距離としたため、 $y_* = 0$ である。

$$x_*(x_0, y_0, \dot{x}_0, \dot{y}_0) = x_0 + \dot{x}_0 \sqrt{(2/g)y_0} \quad (2)$$

$x_0 > 0, y_0 > 0$ である。よって、課題距離 d とボールの到達点との誤差 s は以下の式となる。

$$s(\theta_1, \dot{\theta}_1, \theta_2, \dot{\theta}_2 | d) = \sqrt{(d - x_*(h(\theta_1, \dot{\theta}_1, \theta_2, \dot{\theta}_2)))^2} \quad (3)$$

つまり、この誤差 s が0に近づくほど、課題パフォーマンスが良いとする。課題距離は $d = 5[\text{m}]$ とした。

4.3 学習手法

制御関数 F_1, F_2 の学習には、Soft Actor-Critic と呼ばれる深層強化学習手法を用いた [12]。Soft Actor-Critic (SAC) は連続行動空間を対象とする state of the art の強化学習手法であり、そのサンプル効率の高さと学習の頑健さからロボット制御などに用いられている。最大の特徴は目的関数 (期待収益の最大化) にエントロピー最大項を加えたものである。

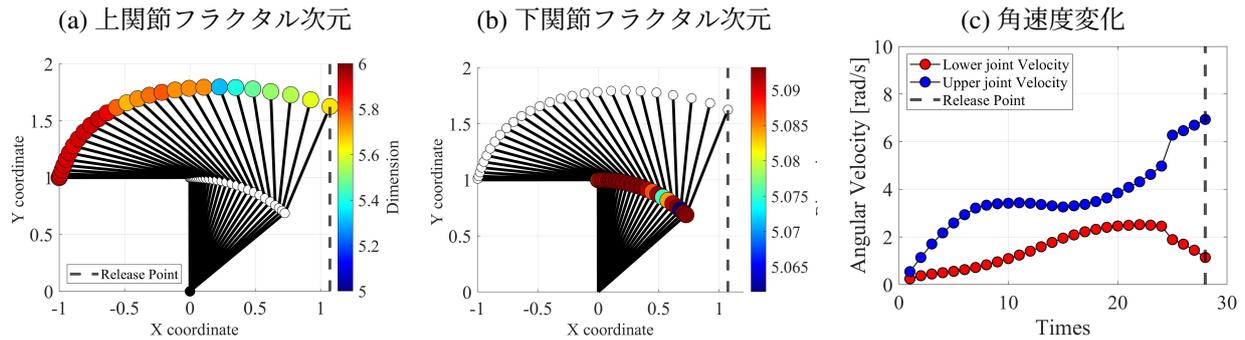


図4 ベストパフォーマンス軌道のフラクタル次元と角速度変化 $d = 5[m]$.

4.4 リリース条件と報酬の設定

投球運動課題において、図3で示すように、上限関節の学習初期値は $(\theta_1, \dot{\theta}_1) = (\frac{\pi}{2}, 0), (\theta_2, \dot{\theta}_2) = (\frac{3}{2}\pi, 0)$ とした。3.2節で述べたように、ボールの速度 \dot{x} は上限関節の角速度の水平方向成分である。ある時刻 t の状態 s_t におけるボールの速度 \dot{x}_t が前の状態の角速度の水平方向成分 \dot{x}_{t-1} より、減速した時、ボールが振り子の端点からリリースされる。報酬関数 r_t は課題距離とボールの到達点との誤差 s (課題パフォーマンス) とし、リリース時のみに与えられる。

$$r(\theta_t, \dot{\theta}_t) = \begin{cases} s(\theta_1, \dot{\theta}_1, \theta_2, \dot{\theta}_2) & \text{if } \dot{x}_t < \dot{x}_{t-1} \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

5. 投球運動課題結果

課題距離を $d = 5[m]$ とし、計 20000 回の試行を行った。以下では、ベストパフォーマンス (誤差 s が 0 に近い) 軌道の結果とその考察について述べる。

5.1 制御自由度変化

本課題では、力学系の制御自由度の分析手法として、力学的不変量の 1 つであるフラクタル次元推定を行った。まず、実際に観測される身体の状態は限られ、外乱の影響を受けるため、観測データから相空間を再構成する必要がある。そのため、相空間再構成法として、縮退した軌道のある次元の空間へ写像することで、特異点の無い滑らかな軌道に変換する時間遅れ座標への埋め込みを行った [14]。この埋め込み操作により、再構成されたアトラクタは距離などの幾何学的性質を持たず、アトラクタの位相的構造のみにより、特徴づけられる。アトラクタ軌道の各点の自由度は次元によって特徴づけられる。日高ら [15] によって

提案された次元推定法を用いて、力学系とみなした運動データの特徴次元の抽出を行った。図4の(a,b)は、学習で得られたベストパフォーマンス軌道における両関節のフラクタル次元の変化である。ヒートマップはフラクタル次元の大きさを、点線はリリース状態を示している。下関節のフラクタル次元変化が微量であるのに対して、上関節は端点の最上点からリリース付近で次元が減少している。

5.2 角速度変化

図4の(c)はベストパフォーマンス軌道の角速度の変化を表したものであり、両関節とも滑らかに変化している。上関節の角速度 (図4(c)の青線) は運動開始直後に、加速、その後、一時的に減速し、リリース状態付近で再び加速する。最終的に、リリース状態時にピークとなる。一方で、下関節の角速度 (図4(c)の赤線) は、緩やかに加速するが、リリース状態付近で、減速する。

6. 考察

6.1 運動の分節化の可能性

図4(a)の上関節のフラクタル次元の変化から、運動開始時は高次元で制御され、端点が最上点に至る時に、次元が低下する。さらに、リリース状態付近で、フラクタル次元がわずかに増加する。このことから、手先が通るべき経由点、つまり、サブゴールが存在し、サブゴールを経由することで、運動全体の目標が達成されるという、階層構造の存在が予想される。従って、適切な分節点をフラクタル次元により推定することができれば、獲得する運動の要所や経由点を定める等の運動計画の大枠が定められ、各サブゴールにおいて、低自由度系の運動学習が可能となる。

6.2 ムチ運動

下関節（肩）→上関節（肘）の順に角速度のピークを迎える。各部位の速度が順々にピークとなる運動をムチ動作という。スポーツにおける投球や打撃では、身体末端の速度をいかに高めるかが重要となる。ムチ動作は、多関節より構成される人間の身体運動において、身体をムチのように動かすことによって、末端部の速度を高めようとする動作メカニズムである。運動連鎖[16]とも呼ばれ、協調運動によって、生成される。投球運動に伴う腕関節部位の速度は、肩、肘、ボールの順に速度がピークとなる。両関節の角速度の変化から、このムチ動作が生成されたことが言える。

複雑な運動を分節することができれば、局所な運動の学習となり、探索空間を削減ができる。そのため、どこで運動が分けられるかを明らかにする必要がある。適切なタイミングにおいて、両関節の互いに独立した自由度が、凍結・解放、つまり、協調することでムチ動作が生成される。次節では、協調運動の構造の抽出について述べる。協調運動の関連する両関節のダイナミクスを構成する3つの項に着目した。

7. 協調構造の抽出

7.1 二重振り子の3つ項

まず、二重振り子の運動方程式に着目すると、3つの項で構成されていると考えることができる。式(1)に身体パラメータ(質量 $(M_1, M_2) = 1[kg]$ 、長さ $(L_1, L_2) = 1[m]$)を代入、また、 $A = \cos(\theta_1 - \theta_2)$ 、 $B = \sin(\theta_1 - \theta_2)$ とする。

$$\ddot{\theta}_1 = \frac{-B\dot{\theta}_2^2 - AB\dot{\theta}_1^2 - 2g \sin \theta_1 + gA \sin \theta_2 + F_1 - AF_2}{2 - A^2} \quad (5)$$

$$\ddot{\theta}_2 = \frac{-AB\dot{\theta}_2^2 - 2B\dot{\theta}_1^2 - 2gA \sin \theta_1 + 2g \sin \theta_2 + AF_1 - 2F_2}{2 - A^2} \quad (6)$$

$(\dot{\theta}_1, \dot{\theta}_2)$ は、前の状態 s_{t-1} に依存する速度依存項、重力依存項及び方策 π の影響を直接受けるトルク項 (F_1, F_2) の3つの項から構成される。トルク F_1, F_2 はそれぞれの関節に入力されるが、両関節の角加速度 $\ddot{\theta}_1, \ddot{\theta}_2$ と依存関係となる。その後、角速度、角度の順で、値が更新される。この3つの項のうち、トルク項と速度依存項の持つ効果とムチ運動生成との関係について調査した。

7.2 下関節定トルク及び定速度制御

下関節定トルク制御の場合の両関節の運動方程式は式(7)(8)と同じであるが、下関節の入力トルク F_1 を一定値 $5, 10, 20[N]$ とした。従って、制御器は F_2 のみを SAC で学習する。

次に、下関節定速度制御における運動方程式は以下で表される。

$$\dot{\theta}_1 = 0 \quad (7)$$

$$\ddot{\theta}_2 = \frac{-AB\dot{\theta}_2^2 - 2B\dot{\theta}_1^2 - 2gA \sin \theta_1 + 2g \sin \theta_2 - 2F_2}{A^2 - 2} \quad (8)$$

上関節の制御トルク F_2 のみを学習する課題であり、式(11)から、下関節のトルク F_1 項が除かれる。下関節の速度は $1, 2, 3[rad/s]$ とした。

7.3 制御自由度変化の比較

図5.6は、下関節定トルク及び定速度制御におけるベストパフォーマンス軌道のフラクタル次元推定結果である。振り子の端点が最上点に達する付近で、次元が低下する傾向が見られ、特に、 $F_1 = 5, 20[N]$ の時、運動開始時に、次元が下がる。投球運動の達成には、上関節の振り上げが必須である。そのため、トルクが小さいまたは大きい時 $F_1 = 5, 20[N]$ 、振り上げに向けて、低自由度制御が行われたと考えられる。

7.4 角速度変化の比較

図7, 8は各制御条件における上下関節の変化である。両制御条件とも、上関節はリリース時に速度がピークとなる。下関節定トルク制御において、ムチ動作が生成され、角速度は滑らかに変化する。入力される上関節トルク F_2 の増減は両関節の角加速度及び角速度に影響を与える。下関節のトルクが小さいほど1時間ステップの角速度変化が小さくなり、角速度変化が滑らかでなくなる。従って、トルクは1時間ステップの角速度変化に対して瞬間的効果を有する。

一方で、下関節定速度制御の結果から、下関節の角速度が小さいほど、上関節角速度の変化がより滑らかとなる。角速度の大きさは1時間ステップ当たりの角度更新量を意味する $\dot{\theta}_1 = 1[rad/s]$ 時に上関節角速度の滑らかな変化から、速度依存項は累積的效果を持つと考えられる。

次章では、これまでの知見から、学習時間、サンプル効率の向上を図る学習手法を提案し、結果は、本学会にて述べる。

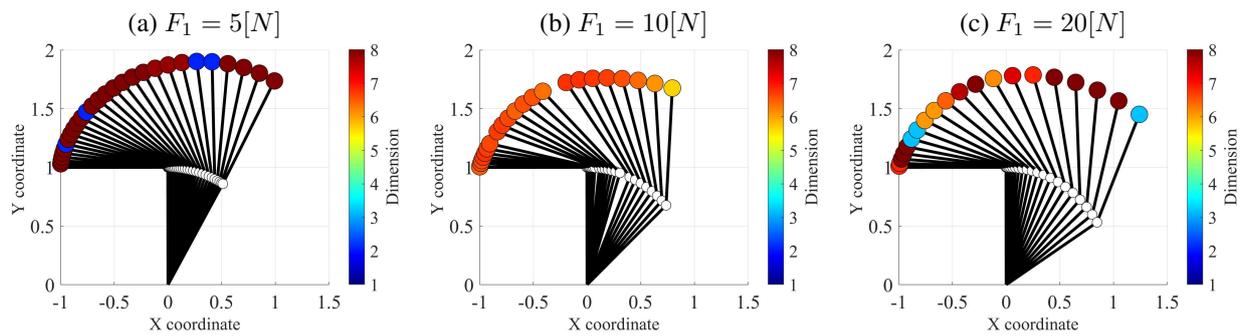


図5 下関節定トルク制御フラクタル次元

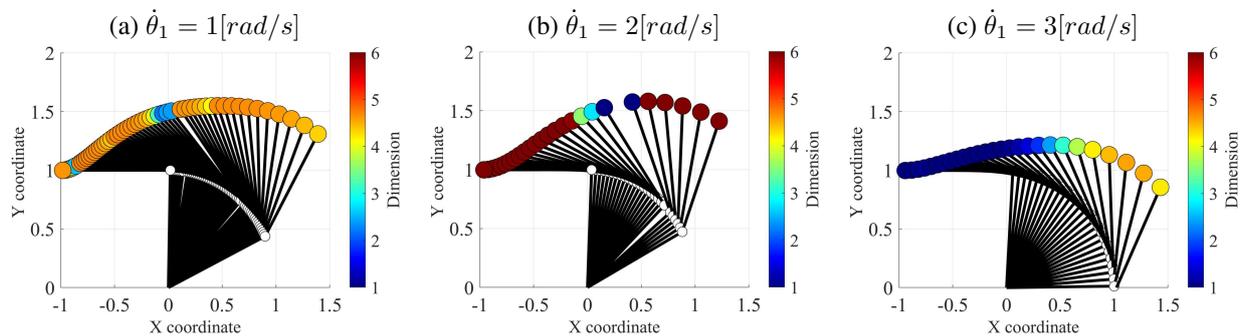


図6 下関節定速度制御フラクタル次元

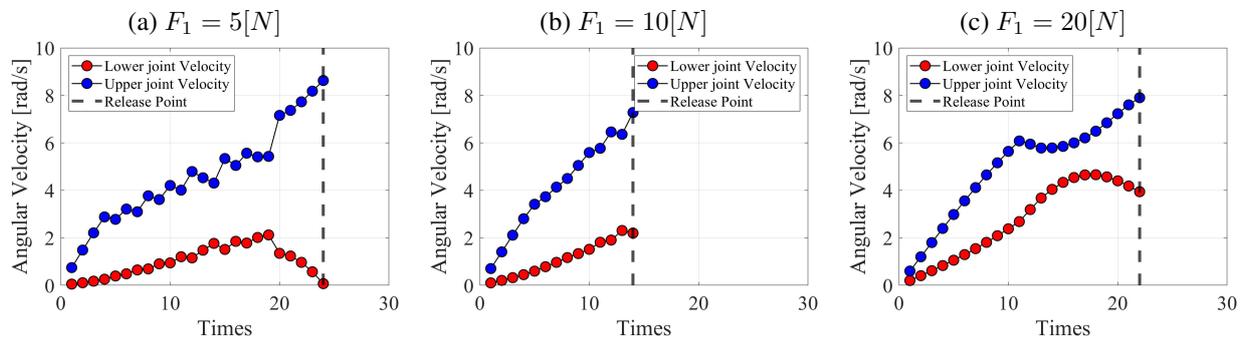


図7 下関節定トルク制御角速度変化

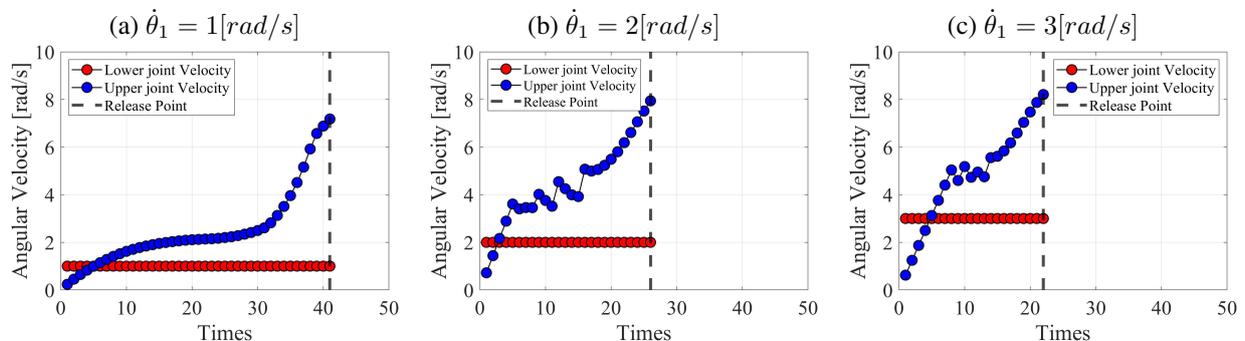


図8 下関節定速度制御角速度変化

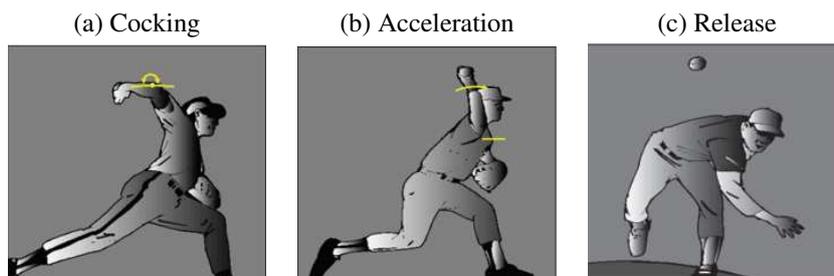


図9 投球運動における腕関節の運動フェーズ [22]

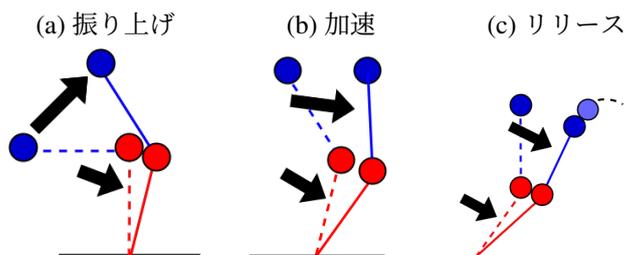


図10 二重振り子の投球運動フェーズ

8. カリキュラム学習の提案に向けて

カリキュラム学習とは、シンプルな課題から学習を開始し、徐々に課題の難易度を上げて学習を行う手法である [17]。強化学習において、主要な学習信号は、学習中に受け取る報酬である。しかし、より複雑あるいは困難な課題において、報酬が疎になり、ほとんど獲得されない場合が生じる。シンプルな課題から開始することで、エージェントが容易に課題を学習でき、得られた経験を足場に次の課題へと移行することができる。さらに、カリキュラムを作成することで探索空間の削減でき、不良設定性を解消すると期待される。

8.1 運動学習へのカリキュラム学習の適用

これまで、人間や動物の試行錯誤を通じた運動獲得のメカニズムを、強化学習のアルゴリズムをモデルとして理解する試みが多くなされてきた [18]。中でも、カリキュラム学習は、文法学習 [19]、ロボティクス [20] や分類問題 [21] など、様々な分野の学習で用いられ、学習収束速度の向上やサンプル効率の改善を実現できることが知られている。しかし、全ての課題に対して、カリキュラムが作成できるという保証はない。従って、運動学習へのカリキュラム学習の適用性を調べる。4章で得られたベストパフォーマンス軌道 (図4) を Q-Learning 手法を用いて、1時間ステップ状態ずつ学習する。

8.2 投球フェーズ分割と報酬設定

図9は投球運動における腕関節の運動フェーズである [22]。一般的に、腕関節は3つの運動フェーズに分節化される。二重振り子の場合、この3運動フェーズは図10で示すものとなる。また、強化学習の場合、報酬の設定に関する明示的な定義はなく、学習結果に影響する。速度最大、エネルギー最大などというように、明確な報酬関数を与えられる運動フェーズで分節化する必要がある。適切な報酬関数の設定に向け、変分法 [23] などを用い、運動の分節化を図る。

9. 謝辞

本研究は、JST 次世代研究者挑戦的研究プログラム JPMJSP2102 の支援を受けたものである。

文献

- [1] J. H. Connell and S. Mahadevan. Rapid task learning for real robot. In J. H. Connell and S. Mahadevan, editors, *Robot Learning*, chapter 5. Kluwer Academic Publishers, 1993.
- [2] J. Kober and J. Peters: Policy Search for Motor Primitives in Robotics, *Machine Learning*, 84–(1-2), 171/203 (2011)
- [3] S. Mahadevan and J. Connell: Automatic Programming of Behavior-based Robots Using Reinforcement

- Learning, *Artificial Intelligence*, 55-(2-3), 311/365 (1991)
- [4] Y. Nakamura, T. Mori, M. Sato and S. Ishii: Reinforcement Learning for a Biped Robot Based on a CPG-actor-critic Method, *Neural Networks*, 20-6, 723/735 (2007)
- [5] M.S. Erden and K. Leblebicioaglu: Free Gait Generation with Reinforcement Learning for a Six-legged Robot, *Robotics and Autonomous Systems*, 56-3, 199/212 (2008)
- [6] 伊藤, 松野: GA により探索空間の動的生成を行う Q 学習による実多自由度ロボットの制御, *日本ロボット学会誌*, 21-5, 526/534 (2003)
- [7] Otterlo, Martijn van, and Marco Wiering. "Reinforcement learning and markov decision processes." *Reinforcement learning*. Springer, Berlin, Heidelberg, 2012. 3-42.
- [8] Bernstein, Nicholai A., Mark L. Latash, and Michael T. Turvey. *Dexterity and its development*. Psychology Press, 2014.
- [9] 日高昇平. "力学的不変量仮説: 運動制御の最適化理論の上位原理として." 第 15 回身体知研究会予稿集 (2013).
- [10] Shinbrot, Troy, et al. "Chaos in a double pendulum." *American Journal of Physics* 60.6 (1992): 491-499.
- [11] Korsch, H. J., and H-J. Jodl. "The Double Pendulum." *Chaos*. Springer, Berlin, Heidelberg, 1994. 89-114.
- [12] Haarnoja T, Zhou A, Abbeel P, Levine S (2018), Soft-actor-critic: off-policy maximum entropy reinforcement learning with a stochastic actor. *Proceedings of the 35th International Conference on Machine Learning*, 1861-1870.
- [13] Kingma, D. P. and Welling, M.: Auto-Encoding Variational Bayes, 2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings (Bengio, Y. and LeCun, Y., eds.) (online), available from (<http://arxiv.org/abs/1312.6114>) (2014).
- [14] Takens, F. (1981). Detecting strange attractors in turbulence. In *Dynamical systems and turbulence*, Warwick 1980 (pp. 366-381). Springer, Berlin, Heidelberg.
- [15] Hidaka, Shohei, and Neeraj Kashyap. "On the estimation of pointwise dimension.
- [16] Seroyer, Shane T., et al. "The kinetic chain in overhand pitching: its potential role for performance enhancement and injury prevention." *Sports health* 2.2 (2010): 135-146.
- [17] Bengio, Y., Louradour, J., Collobert, R., & Weston, J. (2009, June). Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning* (pp. 41-48).
- [18] Doya, Kenji. "What are the computations of the cerebellum, the basal ganglia and the cerebral cortex?" *Neural networks* 12.7-8 (1999): 961-974.
- [19] Jeffrey L Elman. *Learning and development in neural networks: The importance of starting small*. *Cognition*, 48(1):71-99, 1993.
- [20] Terence D Sanger. *Neural network learning control of robot manipulators using gradually increasing task difficulty*. *IEEE Transactions on Robotics and Automation*, 10(3):323-333, 1994.
- [21] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *International Conference on Machine Learning (ICML)*, pag
- [22] Calabrese, Gary J. "Pitching mechanics, revisited." *International journal of sports physical therapy* 8.5 (2013): 652.
- [23] Gelfand, I. M., & Silverman, R. A. (2000). *Calculus of variations*. Courier Corporation.