

修士論文

学習理論に基づく免疫系 T 細胞集団動態の理解

37-196514 堀口修平

東京大学大学院工学系研究科電気系工学専攻

指導教員 小林徹也 准教授

2021 年 1 月

## 概要

免疫系は体内環境を監視し状況に応じて適切な応答をすることで、感染病原体やがん細胞の排除、生体組織の再構築などの重要な役割を担っている。特に、T細胞集団の多様性が免疫応答の決定に寄与していること、この多様性は末梢リンパ節でのT細胞の増殖と死滅、および胸腺からの新規T細胞の流入といった過程によって制御されていることが分かっている。しかしこれらのT細胞集団動態によって、どのように状況に応じた適切な免疫応答が実現されているのかは十分に理解されていない。

本研究では、数理的な基盤が確立されつつあり工学的にも成功を収めている学習理論を援用してこの問題に取り組んだ。具体的には、T細胞の増殖・死滅過程を強化学習アルゴリズム、胸腺からの流入過程を転移学習アルゴリズムの実装として捉える、T細胞集団動態モデルを提案した。強化学習では現在の個体による試行錯誤的な経験をもとに、転移学習では進化的な時間スケールでの先祖個体の経験をもとにして適切な免疫応答を惹起するT細胞多様性が調節される。

この考え方のもと、リーマン多様体上での勾配系の考え方をを用いることで具体的な数理モデルを構築し、実際に適切な免疫応答が学習されることを数値シミュレーションによって確認した。構築した数理モデルは実際の免疫系の分子や細胞との対応付けが可能であり、実験的に観測されている結果を再現することもできた。特に、胸腺の退縮という現象と定性的に類似した現象が学習の性能が高いモデルで確認できること、その原因はバイアスとバリエーションのトレードオフという機械学習の概念によって理解できることを示した。この結果から、胸腺の退縮の機能的意義という免疫学上の未解決問題に対する新たな仮説を提案した。

# 目次

<b>第 1 章</b>	<b>序論</b>	<b>5</b>
1.1	研究背景 . . . . .	5
1.2	研究目的 . . . . .	7
1.3	論文の構成 . . . . .	8
<b>第 2 章</b>	<b>免疫学の基礎</b>	<b>9</b>
2.1	病原体感染における免疫応答の全体像 . . . . .	9
2.2	獲得免疫系の多様性による特異的な免疫応答 . . . . .	11
2.3	T 細胞多様性の発生および制御機構 . . . . .	13
<b>第 3 章</b>	<b>T 細胞集団動態の数理モデル</b>	<b>14</b>
3.1	細胞生物学的な T 細胞集団動態モデル . . . . .	14
3.2	現象論的な T 細胞集団動態モデル . . . . .	17
3.3	規範的な T 細胞集団動態モデル . . . . .	20
<b>第 4 章</b>	<b>勾配系による T 細胞集団動態のモデル化</b>	<b>24</b>
4.1	リーマン多様体上の勾配系 . . . . .	24
4.2	機械学習・情報分野における勾配系の応用例 . . . . .	27
4.3	T 細胞集団動態のリーマン多様体構造 . . . . .	29
<b>第 5 章</b>	<b>マルコフ決定過程による免疫系のモデル化</b>	<b>33</b>
5.1	マルコフ決定過程 . . . . .	33
5.2	強化学習アルゴリズム . . . . .	35
5.3	免疫系における状態・観測・行動 . . . . .	38
5.4	強化学習としての T 細胞の増殖死滅過程 . . . . .	41

目次	2
5.5 強化学習の数値シミュレーション	46
<b>第 6 章 胸腺からの T 細胞の流入とその意義</b>	<b>50</b>
6.1 転移学習と胸腺の機能	50
6.2 転移学習としての流入動態の導出	52
6.3 流入分布の影響と胸腺選択	56
6.4 流入率の影響と胸腺の退縮	63
<b>第 7 章 結論</b>	<b>76</b>
7.1 研究結果のまとめ	76
7.2 計算論的生物学	77
7.3 工学・医学	78
謝辞	80
参考文献	81
<b>付録 A 力学系の安定性</b>	<b>101</b>
A.1 平衡点の安定性	101
A.2 平衡点集合の安定性	102
<b>付録 B 距離・ダイバージェンス</b>	<b>106</b>
B.1 距離とダイバージェンスの定義	106
B.2 ダイバージェンスの例	107

# 目次

5.1	免疫系における状態観測から行動決定の実装 . . . . .	41
5.2	免疫系細胞の相互作用による学習の実装 . . . . .	46
5.3	免疫系強化学習モデルの数値シミュレーション結果 . . . . .	49
6.1	免疫系における転移学習の数値シミュレーション結果 . . . . .	54
6.2	免疫系における転移学習の数値シミュレーション結果 2 . . . . .	55
6.3	胸腺からの流入分布 $m$ と正規化した報酬予測誤差の関係 (1) . . . . .	60
6.4	胸腺からの流入分布 $m$ と正規化した報酬予測誤差の関係 (2) . . . . .	60
6.5	胸腺からの流入細胞数の加齢による変化の概念図 . . . . .	64
6.6	相対頻度 $\mu$ と報酬予測誤差の関係 (実験 1) . . . . .	68
6.7	相対頻度 $\mu$ と胸腺からの流入速度 $\gamma$ の組に対する正規化した平均報酬予測誤差の関係 (実験 1) . . . . .	68
6.8	T 細胞種類数 $K$ と報酬予測誤差の関係 (実験 2) . . . . .	69
6.9	T 細胞種類数 $K$ と胸腺からの流入速度 $\gamma$ の組に対する正規化した平均報酬予測誤差の関係 (実験 2) . . . . .	69
6.10	学習率 $\alpha$ と報酬予測誤差の関係 (実験 3) . . . . .	70
6.11	学習率 $\alpha$ と胸腺からの流入速度 $\gamma$ の組に対する正規化した平均報酬予測誤差の関係 (実験 3) . . . . .	70
6.12	バイアスとバリエーションの流入率 $\gamma$ 依存性 . . . . .	73

# 表目次

2.1	自然免疫系のエフェクター機構 . . . . .	10
5.1	部分観測マルコフ決定過程としての免疫系 . . . . .	39
5.2	強化学習アルゴリズムとしての免疫系 [スカラー報酬信号モデルの場合 . . . . .	43
5.3	強化学習アルゴリズムとしての免疫系 [ベクトル報酬信号モデル . . . . .	46
5.4	強化学習の数値シミュレーションで用いたパラメータ . . . . .	47
6.1	流入分布 $m$ と報酬予測誤差の関係 . . . . .	61
6.2	最適な流入率 $\gamma$ を調べる数値シミュレーションで用いたパラメータ . . . . .	66

# 第 1 章

## 序論

### 1.1 研究背景

本研究では 2 つの異なる分野が扱われる。1 つ目が免疫系についての生物学である免疫学, もう一つが学習について学習理論である。

#### 免疫学

2019 年コロナウイルス感染症のパンデミックが発生し, ウイルスや免疫系についての関心が高まっている。免疫学は 18 世紀の Jenner による天然痘ワクチン開発から始まったと言われており, それから 200 年以上の研究によって多くのことが分かってきた。たとえば, 免疫系は感染症に対する生体防御だけでなく, アレルギーやがん [1], 精神疾患 [2] に至るまで多くの病気に関与していることが明らかになった。特に近年は生物学における計測技術の発展が目覚ましく, 免疫学にも大きな影響を与えている。

免疫系についての膨大な知識が蓄積されていく一方で, 免疫系をどう理解したら良いのかを明らかにするのが理論研究である。免疫学の歴史において理論研究は重要な役割を果たしてきた。Burnet によるクローン選択説 [3] は, 免疫系がダーウィン進化的機構を内包していると提唱して, 抗原に特異的に反応する抗体が出来る理由を説明し, 後に実験的に証明されている。Jerne と Hoffmann による免疫ネットワーク理論 [4], [5] は, 免疫細胞と分子の特異的な反応で作られるネットワークを考え, 複数の定常状態の間を遷移することで免疫記憶が実現できることを示した。Burnet と Jerne はこれらの功績でノーベル生理学・医学賞を受賞している。しかしこれらの研究はその後の実験的研究で明らかになった多くの重要な要素を欠いており, 免疫系の一側面を捉えているに過ぎない。以降, 数理モデルではなく概念的なモデルとして免疫系をどう見るかといった研究 [6]–[12] や, 免疫系

の一部の現象に注目し分子・細胞生物学的知見から構築した数理モデルを用いて実験では捉えられない現象や測定が困難な性質を予測・推定するといった研究が多くなされてきた [13]. しかし、これらの研究では**免疫系全体がなぜ上手く機能しているのか**という問いに対して定量的で明快な説明を与えられていない。

生命現象の数理モデルを構築する方法として、分子・細胞生物学的なミクロな機構から数理モデルを構築する方法 (分子・細胞生物学的モデル化)、マクロな現象からなるべく簡潔な数理モデルを構築する方法 (現象論的モデル化)、果たすべき機能からあるべき姿をモデル化する方法 (規範的モデル化)、などがある。細胞生物学的モデル化と現象論的モデル化では、構築されたモデルが生物学的な機能を再現することによって免疫系がなぜ上手く機能しているのかということの説明する。規範的モデル化では、構築したモデルと実際の現象との対応を検証することによって説明する。規範的なモデル化においては、免疫系の機能をまず同定する必要がある。免疫系の機能としてまず挙げられるのが未知の病原体を学習・記憶することである。

## 学習理論

学習という言葉は日常的にも用いられ、旺文社国語辞典によれば「(学校などで) 学問や技術の基礎的知識を学び習うこと」である。通常学習する主体は人間である。一方、専門用語としては分野によって意味が異なり、まず学習する主体が異なる。心理学や生理学、動物行動学では、ヒトなどの動物において脳神経系によって実現される現象が学習と呼ばれることが多い。一方機械学習の分野では、コンピュータープログラムによって実現される現象が学習と呼ばれる。

心理学の分野において学習という言葉は、「類似する経験を繰り返すことにより生じる、比較的永続的な行動変容の過程」と定義されている [14]. 免疫系においては免疫応答が「行動」に対応すると考えれば、それは過去の免疫学的な「経験」によって変化する。例えば、T 細胞や B 細胞による獲得免疫系は記憶を持つ。一度感染した病原体に再び感染した際の免疫応答は一度目よりも速くなることはよく知られており、ワクチン接種が感染症対策に有効である根拠の 1 つである。つまり、心理学の学習の定義のもとで免疫系は学習するといえる。

一方機械学習の分野においては、「性能指標  $P$  で測定されるタスク  $T$  の性能が経験  $E$  により改善される場合、そのタスク  $T$  のクラスおよび性能指標  $P$  に関して経験  $E$  から学習する」という定義がある [15]. この定義においては、ただ行動が変わるだけでなく具体的なタスクと性能指標を想定して良くなることが求められる。免疫系が機械学習の定義のもとで学習すると言えるかどうかは自明ではないが、免疫系を学習システムとしてモデル



化する試みは人工知能研究の歴史と並行して実際に行われてきている。

遺伝的アルゴリズムを開発した Holland[16] などによるルールベース機械学習の流れの中で、1980年代に免疫系を学習システムとしてモデル化する人工免疫系という分野が興った [17], [18] もの、その後人工知能研究の冬の時代を経て人工免疫系研究は下火になった。異なる人工知能研究の流れとして、**強化学習**や**深層学習**といった概念の発達がある。心理学において Thorndike や Skinner による試行錯誤学習やオペラント条件付けといった概念が強化学習の一つの源流である。一方、工学や数学の分野では Bellman が 1953 年に確立した動的計画法 [19] という強力な基盤があった。これらは、1980年代の Sutton による TD 学習や Watkins による Q 学習 [20] として合流した [21]。さらに強化学習と深層学習技術 [15] が組み合わさることで囲碁やロボット制御など工学応用として顕著な成果を挙げたほか、脳神経科学では計算論的神経科学 [22] という分野の拡大にもつながっている。このような流れの中で、2019年には Kato らによって強化学習によって免疫系を捉え直す先駆的な試みが発表された [23]。

## 1.2 研究目的

本研究では学習理論を活用して規範的にモデルを構築し、現象論的モデルや実際の免疫系におけるミクロおよびマクロな現象との一致を確かめることで、免疫系全体がなぜ上手く動いているのかの理解を試みる。これによって、免疫系がある種の学習を行っていると考えられることは、免疫系についての理解を深め、免疫学的に意義のあるアプローチであると示すことが本研究の目的である。

本研究において学習という言葉は機械学習分野の定義で用いられる。すなわち、具体的なタスクや性能指標を提示した上で経験から性能が向上することをもって学習と呼ぶ。ただし、学習の主体はコンピュータプログラムとは限らない。また、学習の主体が生物個体やその免疫系でない場合にも学習という言葉を用いる。たとえば、第6章で転移学習という機械学習における概念を導入して免疫系の胸腺という臓器の役割を調べるが、転移学習する主体は T 細胞でも胸腺でも個体でもない。あえて言えば、胸腺のふるまいを決定する遺伝子が主体であり、多数の世代を経て進化することを指している。

その前提にあるのは、Darwin と Wallace による自然選択説である。生物の形や振る舞いが何らかの機能を果たすために合目的的に設計されている、あるいは最適化されているように見える、ということは古くから指摘されていた。これはダーウィンとウォレスの自然淘汰の原理によって説明される。現代的な進化の総合説によれば、選択だけでなく中立的な遺伝的浮動など様々な要因が存在することがわかっている。あらゆる生物のすべての

形や振る舞いが最適化されているとは限らないが、個体の生存や繁殖の成功に大きく寄与するようなものはある程度の最適化がされていると考えられる。すなわち、この自然選択が機械学習アルゴリズムを実装している、という意味でも学習という言葉を用いる [24]。

この進化により実装される学習と対をなすのが、動物個体の行動として実装される学習である。このような個体の経験に基づく学習については、強化学習という枠組みによって非常に一般的な形で整理されており、本研究でもこの枠組みを用いる。既に述べたように、強化学習という枠組みは生物を理解する上でも工学的にもものをつくる上でも有用であることが示されている。強化学習の枠組みが十分一般的であり、生物個体が経験をもとに学習するように自然選択の力が働いているとすれば、その生物のふるまいを強化学習の枠組みで理解するという考え方は自然である。

このように学習理論を活用することで免疫系の理解を試みる上で、既存の学習理論の枠組みでは不足している部分がある。それが学習のアルゴリズムをどのように生物に実装するかという点である。この点について、本研究ではリーマン多様体上での勾配系 (第4章) という考え方が有用であることを示す。免疫系だけでなく生命現象をモデル化する上でも一般的に役立つと思われる。何らかの現象を勾配系としてのモデル化することは、その現象が最適化という機能を持つことを示すことができ、安定性という概念とも深く関わっている。工学分野でも、与えられたハードウェアにおいてどのように最適化アルゴリズムを実装するかを考える際に有用だと思われる。

### 1.3 論文の構成

本論文ではまず第2で本研究の主題となる T 細胞を含めた免疫学の基礎を紹介し、第3では T 細胞の集団動態をモデル化する研究を紹介する。続いて第4と5では勾配流と強化学習の考え方を導入して、これらを用いた免疫系のモデル化手法を示す。第6では胸腺にまつわる免疫系の現象が、このモデルによって理解できることを示す。最後に第7章でこれまでの結果を整理し、今後の展望を述べる。

## 第2章

# 免疫学の基礎

本章では、主にヒトやマウスなど哺乳類の免疫系について、実験によって明らかにされた事実をまとめる。特に、免疫系の中でも司令塔的な役割を果たす T 細胞に注目して概説する。

### 2.1 病原体感染における免疫応答の全体像

免疫系の重要な役割の一つが、病原体から個体を防御することである。病原体 (pathogen) は病気の原因となり生体組織に損傷や病変をもたらす微生物のことであり、ウイルス、細菌、真菌、寄生虫 (原虫・蠕虫) がある。各病原体の感染経路、生息環境、増殖機序などは異なるため、それぞれに適した免疫系の応答 (エフェクター機構) が必要になる [25]。

病原体からの防御は3つの段階的障壁によって実現されている [25]:

1. 上皮による障壁 (皮膚, 口腔粘膜, 気道上皮, 小腸)
2. 自然免疫系 (補体系, マクロファージ, 顆粒球, ナチュラルキラー細胞など)
3. 獲得免疫系 (B 細胞・抗体, T 細胞)

上皮は、外界と生体内を分けるバリアであり、物理的・化学的・微生物学的な機構によって病原体が生体 (宿主) 内へ侵入することを防いでいる。まず上皮細胞が互いに密着結合することで内外を遮断し、上皮の外側を流れる気体や液体によって上皮への病原体の接着を防ぐという物理的な機構がある。また細菌の細胞壁を破壊する抗菌タンパク質による化学的な機構に加え、病原性のない常在細菌の存在も上皮での病原微生物の増殖を防ぐ重要なバリア機構である。

環境に多種存在する通常の微生物は上皮障壁によって侵入を防げるが、病原体とされる微生物はこれを突破して体内に侵入する。この侵入にまず対処するのが自然免疫系である。自然免疫系の応答はさらに二段階に分けられる。

第一段階では、上皮直下において補体系や貪食細胞による病原体の認識と排除が試みられる。補体系は補体と呼ばれる30種類以上もの可溶性タンパク質からなり、通常は不活性型で血中を循環している。一部の補体は微生物表面の糖鎖などを認識して活性化することができ、別の補体分子を切断(タンパク質分解)することによって活性化させる。このような活性化経路を完了すると、補体系は特定の病原体の細胞膜に穴を開けたり、貪食細胞を誘引したりすることができる。マクロファージなどの貪食細胞は、補体系によって付けられた目印によって侵入した微生物を速やかに発見、貪食して消化する。補体系による目印の他にも、貪食細胞は**パターン認識受容体 (PRR)** とよばれる多種の受容体を表面にもっているため、微生物を認識して特異的に貪食することができる。このパターン認識受容体は、病原体に特徴的で生体内に存在しない分子構造のパターン(病原体関連分子パターン, PAMP)や、宿主の細胞が傷害やストレス・がん化することによって誘導される分子のパターン(傷害関連分子パターン, DAMP)を認識する。

第二段階は、パターン認識受容体を介して活性化した貪食細胞や補体系が炎症を誘導することで始まる。貪食細胞、とくにマクロファージはサイトカインやケモカインと呼ばれる分子を周辺に分泌することで、周辺組織が炎症状態となり他の自然免疫系の細胞を動員する。自然免疫系の細胞はそれぞれ異なるサイトカインの受容体を持っており、サイトカインを認識することで活性化し、さらにサイトカインを産生する。この連鎖によって、病原体に適したエフェクター機構が誘導される。情報伝達分子サイトカインは60種類以上あり、誘導されるエフェクター機構も多くの種類の細胞が関わる複雑なものであるが、その機能や対処する病原体の種類によって表2.1のように3つに分類されて理解されることが多い [26], [27].

名称	自然免疫細胞	活性化するサイトカイン	病原体の種類
1型免疫	NK細胞, ILC1	IL-12, IL-18	細胞内寄生細菌・ウイルス
2型免疫	ILC2	IL-25, IL-33, TSLP	細胞外寄生体(原虫・蠕虫)
3型免疫	ILC3, LTi細胞	IL-23, IL-1 $\beta$	細胞外寄生細菌・真菌

表2.1 自然免疫系のエフェクター機構。略称：ILC1(1型自然リンパ球), ILC2(2型自然リンパ球), ILC3(3型自然リンパ球), NK細胞(ナチュラルキラー細胞), LTi細胞(リンパ組織誘導細胞)。

大抵の病原体はこの自然免疫系による障壁も突破することができる。たとえば、パターン認識受容体によって感知されるような分子を表面からなくしたような細菌は生体内で増殖することができる。このような場合には、より精密な標的設定が可能な獲得免疫系が対応することになる。ただし、獲得免疫系も基本的に自然免疫系と同じエフェクター機構を誘導する。つまり、標的となる病原体の種類や量・場所に応じてエフェクター機構を効果的に誘導するのが獲得免疫系の役割である。

ただし、実際には病原体の感染以外にも免疫系は幅広い役割を担っている。たとえば、免疫系は正常であった自己組織に腫瘍が発生するとすぐに見つけて排除する仕組みがあり、通常はがんを防いでいる。その他、皮膚の創傷治癒 [28], [29] や血管の新生・再構築 [30], 内分泌組織の異常細胞の除去 [31], オタマジャクシがカエルに変態する際の尻尾の退縮 [32] など幅広く生体組織の維持と再構築に携わっている。したがって、病原体だけでなく生体内の異常を総合的に監視しており、その異常の種類や場所、タイミングを見極め、適切な応答をしているといえる。

## 2.2 獲得免疫系の多様性による特異的な免疫応答

T 細胞と B 細胞からなる獲得免疫系は、非常に多様な抗原を多様な抗原認識受容体によって識別する点が自然免疫系の病原体認識と明確に異なる。抗原とは、これらの受容体に結合しうる分子である。誘導するエフェクター機構は自然免疫系と同じだが、抗原認識受容体と抗原の特異的な結合によって、狙いを絞った効率的なエフェクター機構を誘導できる。以下では獲得免疫系の T 細胞・B 細胞の多様性について述べ、感染応答においてその多様性がどのように機能するかを述べる。

B 細胞では免疫グロブリン (Ig) と呼ばれるタンパク質が抗原認識受容体であり、細胞膜に B 細胞受容体 (BCR) としてもつほか、抗体として細胞外に分泌することもできる。免疫グロブリンの一部は可変部といって細胞ごとに異なる立体構造をもつが、同じ細胞であれば同じ構造をもつ。つまり、同じ個体の細胞にもかかわらず B 細胞の免疫グロブリン遺伝子の配列は細胞ごとに異なる。この受容体可変部の多様性が形成される仕組みについては T 細胞受容体と類似しており後述する。抗体の機能としては病原体に結合することによって毒素やウイルスが細胞に侵入するのを防ぐこと (中和) や、補体と同じように貪食細胞による捕食を推進すること (オプソニン化)、補体系を活性化させることなどがある。抗体には IgM, IgG, IgE, IgA といったようにいくつかの種類 (クラス) があり、それぞれ機能や空間的分布が異なる。そして免疫グロブリンの立体構造は細胞ごとに異なる。異なるクラスの抗体は異なるクラスの B 細胞によって分泌されるが、B 細胞は抗原認識とともに受

け取るサイトカインによってクラスを切り換えることが知られている [25].

一方 T 細胞では, T 細胞受容体 (TCR) として膜において抗原認識するのみである. T 細胞受容体も B 細胞と同様に可変部を持ち, 細胞ごとに異なる立体構造を持っている. 1 細胞で 2 種類の T 細胞受容体をもつような T 細胞が存在することも分かっているが, 多くの T 細胞は 1 種類の T 細胞受容体を発現する [33]. 同じ T 細胞受容体を持つ T 細胞集団をクローンとよぶ. T 細胞受容体も免疫グロブリンと同様の機構で多様性が形成される. T 細胞は通常 CD4 か CD8 のどちらかのタンパク質を膜上にもっているため, これによって CD4 陽性 T 細胞と CD8 陽性 T 細胞に分類でき, 機能も大きく異なる. CD8 陽性 T 細胞は, 抗原を認識すると細胞傷害性 T 細胞となる. ウイルス等に感染した細胞を排除することが主な機能であるためこの細胞自身が自然免疫系とは別系統のエフェクター機構といえる. 一方で CD4 陽性 T 細胞は, 免疫系のエフェクター機構全体を制御している. たとえば, B 細胞の抗体産生を開始するには CD4 陽性 T 細胞からの刺激が必要であり, CD8 陽性 T 細胞が細胞傷害性 T 細胞となるためにも多くの場合 CD4 陽性 T 細胞の関与が必要である. また, 自然免疫系のエフェクター機構も CD4 陽性 T 細胞から放出されるサイトカインによって制御される. CD4 陽性 T 細胞にはさらにいくつかのサブタイプに分類されており, ヘルパー T 細胞と制御性 T 細胞に分けられる. 一般に, ヘルパー T 細胞は免疫応答を促進するのに対し, 制御性 T 細胞は免疫応答を抑制するという違いがある. ヘルパー T 細胞の中でも, 抗原に出会ったことのないナイーブ T 細胞や, 抗原に出会ってから分化してそれぞれ異なる機能をもつエフェクター T 細胞 (Th1, Th2, Th17 など), 免疫応答収束後も長く残り再度の感染に備えるメモリー T 細胞などがある [25].

免疫グロブリンは抗原そのものに結合することができるが, T 細胞受容体は抗原提示という仕組みを通してのみ抗原認識する. 様々な細胞が表面に持つ主要組織適合遺伝子複合体 (MHC) と, そこに結合している抗原の断片のペプチドの複合体 (ペプチド-MHC 複合体) が T 細胞受容体と結合することで, 抗原が認識される. MHC を持つ細胞は, 細胞内で分解処理されたペプチドを MHC 分子の溝に結合させて細胞膜に提示する. MHC にはクラス I とクラス II という 2 つのクラスがある. クラス I はほぼ全ての有核細胞がもち, その細胞に感染したウイルスなどの病原体に由来するペプチドを細胞傷害性 T 細胞に提示する. もし細胞傷害性 T 細胞がウイルス由来のペプチドと MHC に結合した場合, T 細胞はそのペプチドを提示した細胞を破壊する. 一方, クラス II の MHC は樹状細胞や B 細胞, マクロファージなど抗原提示細胞とよばれる細胞にのみ発現する. 樹状細胞に提示された抗原を認識したナイーブ T 細胞は活性化・増殖してエフェクター T 細胞に分化する. B 細胞に提示された抗原を認識したヘルパー T 細胞はサイトカインを放出してその B 細胞が産生すべき抗体のクラスを指示する [25].

## 2.3 T細胞多様性の発生および制御機構

T細胞に限らず生体を構成するすべての細胞は、受精卵から分裂増殖した子孫細胞である。特殊化していない母細胞から、機能や形態を変化して特殊化した娘細胞が生まれることを細胞分化と呼ぶ。分裂によって自身と同じ細胞を複製することも、別の種類の細胞に分化することもでき、際限なく増殖できる細胞は幹細胞とよばれる。T細胞やB細胞を含めた免疫系の細胞と赤血球・血小板は、骨髄に存在する造血幹細胞に由来する。分化途上にあるT細胞(胸腺細胞)は骨髄から血流にのってまずは胸腺に入る。胸腺の中では胸腺皮質から胸腺髄質に移動しながら以下の3つの過程を経る。

まず、皮質で胸腺細胞はT細胞受容体遺伝子の再構成が行われてT細胞受容体を細胞膜上に発現するようになる。T細胞受容体遺伝子は、 $\alpha$ 鎖と $\beta$ 鎖、V(D)J領域に分けられ、それぞれ多数の候補の中からランダムに選ばれる。その組み合わせと、それぞれの結合部へのランダムなDNAの挿入や削除によって、T細胞受容体の膨大な多様性が生まれる。

次に、皮質に存在する胸腺皮質上皮細胞(cTEC)と相互作用する。cTECのMHCクラスI分子に抗原提示された自己由来の抗原を認識したT細胞はCD8陽性T細胞となり、MHCクラスII分子の抗原を認識したT細胞はCD4陽性T細胞となる。この自己ペプチド-自己MHC複合体と結合できないT細胞受容体をもつT細胞は無視によって死滅する。この過程は正の選択とよばれる[34]。

最後に、皮質および髄質において負の選択と呼ばれる過程を経たT細胞が胸腺から出て全身を循環するようになる。負の選択においては胸腺髄質上皮細胞(mTEC)に加え、樹状細胞・B細胞などの抗原提示細胞がT細胞と相互作用する。樹状細胞は胸腺以外の末梢リンパ節に偏在する自己抗原を提示し、mTECは全身の組織に特異的に発現する自己抗原を提示する。これらの自己抗原を認識したT細胞は死滅する。正の選択と負の選択を比較すると、同じ自己抗原への反応によって生存と死滅という一見逆の結果がもたらされているように見える。胸腺選択の親和性説によれば、正の選択の閾値は負の選択の閾値より低く、弱すぎず強すぎない反応性をもつT細胞が選ばれとされる。また、負の選択の閾値よりもわずかに弱い刺激を受けたT細胞は制御性T細胞に分化するとされる[34]。

胸腺から流出したナイーブT細胞は、血管と二次リンパ組織(リンパ節・脾臓・腸管のパイエル板など)を循環する。二次リンパ組織において、T細胞は樹状細胞によって抗原提示された抗原を認識すると活性化して増殖し、エフェクターT細胞に分化する。

## 第 3 章

# T 細胞集団動態の数理モデル

前章で示したように、ヘルパー T 細胞の集団には T 細胞受容体の遺伝子型の多様性と表現型の多様性の両方が存在し、この多様性が獲得免疫系によって惹起される免疫応答の要となっている。この多様性がどのように調整されているのかを理解し、定量的に予測することは感染症やがん、免疫系の関与する疾患において非常に重要であり、本研究の主題である。ただし、本研究では表現型の多様性については区別せず T 細胞受容体の遺伝子型の多様性に注目する。表現型の多様性も含めた扱いは今後の課題である。また、以下では主に CD4 陽性のヘルパー T 細胞のことを単に T 細胞と呼ぶことにする。

本章では、T 細胞集団の多様性の動態を記述するいくつかの数理モデルを紹介する。モデルを細胞生物学的・現象論的・規範的の 3 つに分類することで、モデル化の考え方の違いを明らかにする。現象論的モデル以外は先行研究によるモデルである。

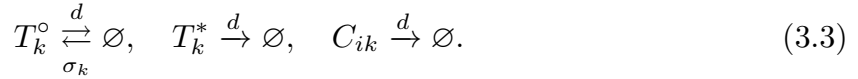
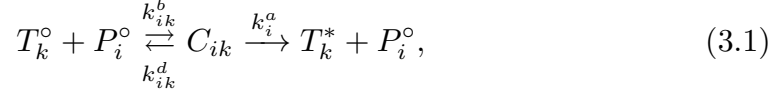
### 3.1 細胞生物学的な T 細胞集団動態モデル

T 細胞集団の多様性の時間変化を記述する数理モデルを構築する手法として最も基本的なのは、分子や細胞による素過程から導出する手法である。1994 年の De Boer と Perelson による T 細胞集団動態のモデル [35] およびその発展 [36] が決定論的なモデルとしては古典的であり、以下では [37] の解析と共にこれを紹介する。

$N$  種類の抗原提示されたペプチドと  $K$  種類のクローンの T 細胞による競争的な相互作用を考える。 $i$  番目のペプチドのうち T 細胞と結合していない自由なものを  $P_i^o$  と書き、ペプチドと結合していない  $k$  番目の T 細胞を  $T_k^o$  と書く ( $i = 1, \dots, N, k = 1, \dots, K$ )。これらは反応速度定数  $k_{ik}^b$  で結合して複合体  $C_{ik}$  をつくり、この複合体は反応速度定数  $k_{ik}^d$  でそのまま解離するか、 $k_{ik}^a$  で活性化した T 細胞  $T_k^*$  として解離する。活性化した T 細胞



はある速度定数  $k^p$  で2つに分裂する. また, T細胞は  $\sigma_k$  の速度で胸腺から流入し,  $d$  の速度定数で死亡する. 以上を化学反応式のようにまとめて書くと以下ようになる:



質量作用の法則を仮定して, 細胞や分子の名前と同じ文字でその個数 (細胞数, 分子数) を表すことにすれば,  $T_k^\circ$ ,  $C_{ik}$  および  $T_k^*$  の時間変化は以下ようになる:

$$\frac{dT_k^\circ}{dt} = \sigma_k - \sum_{i=1}^N k_{ik}^b T_k^\circ P_i^\circ + \sum_{i=1}^N k_{ik}^d C_{ik} + 2k^p T_k^* - dT_k^\circ, \quad (3.4)$$

$$\frac{dC_{ik}}{dt} = k_{ik}^b T_k^\circ P_i^\circ - (k_{ik}^d + k_{ik}^a + d) C_{ik}, \quad (3.5)$$

$$\frac{dT_k^*}{dt} = -(k^p + d)T_k^* + \sum_{i=1}^N k_{ik}^a C_{ik}. \quad (3.6)$$

これらをもとに,  $k$  番目の全 T細胞数  $T_k := T_k^\circ + T_k^* + \sum_i C_{ik}$  の時間変化

$$\frac{dT_k}{dt} = \frac{dT_k^\circ}{dt} + \frac{dT_k^*}{dt} + \sum_{i=1}^N \frac{dC_{ik}}{dt} \quad (3.7)$$

を調べる. 今注目する T細胞数の時間変化にくらべて, 受容体とペプチドの結合解離反応や細胞分裂が十分速いと仮定すると, 複合体  $C_{ik}$  および活性化した T細胞数  $T_k^*$  は概定常状態にあると近似できる. まず  $dT_k^*/dt \approx 0$  から

$$T_k^* = \frac{1}{k^p + d} \sum_{i=1}^N k_{ik}^a C_{ik} \quad (3.8)$$

が得られるので,  $T_k$  の時間変化は

$$\frac{dT_k}{dt} = \sigma_k + \kappa \sum_{i=1}^N k_{ik}^a C_{ik} - dT_k \quad (3.9)$$

のように,  $C_{ik}$  で表される. ただし  $\kappa = k^p/(k^p + d)$  とおいた.  $dC_{ik}/dt \approx 0$  より

$$C_{ik} = K_{ik} T_k^\circ P_i^\circ = K_{ik} \left( P_i - \sum_{l=1}^K C_{il} \right) \left( T_k - \sum_{i=1}^N C_{ik} \right). \quad (3.10)$$

ただし,  $K_{ik} := k_{ik}^b / (k_{ik}^d + k_{ik}^a + d)$  とおき, 各ペプチドの総量  $P_i = P_i^o + \sum_k C_{ik}$  は時間によらない定数 (保存量) になることを用いた. 式3.10は多変数の非線形代数方程式になっており解析的には解くのは難しいが,  $\sum_i C_{ik} \ll T_k$  のとき近似解が解析的に求められる [37]. これは T 細胞のうちペプチドと結合している T 細胞の割合が少ない場合に対応する. 式3.10より

$$C_{ik} \approx K_{ik} T_k \left( P_i - \sum_{l=1}^K C_{il} \right) \quad (3.11)$$

であり, 両辺の  $i$  について和を取ると  $\sum_i C_{ik}$  について解ける. 最終的に

$$C_{ik} = P_i \frac{K_{ik} T_k}{1 + \sum_{l=1}^K K_{il} T_l} \quad (3.12)$$

となる. よって T 細胞数の時間発展は

$$\frac{dT_k}{dt} = \sigma_k + T_k \left( \sum_{i=1}^N \frac{\kappa k_{ik}^a K_{ik} P_i}{1 + \sum_{l=1}^K K_{il} T_l} - d \right) \quad (3.13)$$

となる.

このモデルの長時間的な振る舞いを調べるために, 全ての  $k$  で  $dT_k/dt = 0$  を満たす T 細胞数の分布  $\{T_k^{st}\}_{k=1}^K$  を考える. ここでは簡単のため, 流入はない  $\sigma_k = 0 (k = 1, \dots, K)$  とする. すると, 全ての  $k$  において  $T_k^{st} = 0$  あるいは

$$\sum_{i=1}^N \frac{\kappa k_{ik}^a K_{ik} P_i}{1 + \sum_{l=1}^K K_{il} T_l^{st}} - d = 0 \Leftrightarrow \sum_{i=1}^N \kappa_{ik} \alpha_i = d \quad (3.14)$$

が成立する. ここで, 未知の細胞数分布を含む項を  $\alpha_i := 1 / (1 + \sum_{k=1}^K K_{ik} T_k^{st})$ , その係数を  $\kappa_{ik} := \kappa k_{ik}^a K_{ik} P_i$  としてまとめた. もし全ての  $T_k$  で式3.14が成立すると仮定すると, 未知数  $N$  個, 式が  $K$  本の線形連立方程式となる.  $\{\kappa_{ik}\}$  の値の組み合わせによっては線形従属性によって実質的に式の数が減るが, ランダムな値を取る場合はそのようなことが起こる確率は低い. そのようなことが起こらないとすれば,  $K$  本のうち最大でも  $N$  本の式しか同時に成立しない. 式3.14が成立しないような  $k$  では  $T_k = 0$  となるので, 結果としてペプチドの種類数  $N$  と等しいかそれより小さい種類しか T 細胞が最終的に生存できないということが分かる. また一般に, 各ペプチドに対して最も結合できる種類の T 細胞のみが生き残ることも分かる. すなわち, T 細胞クローンは抗原提示されたペプチドという「資源」を巡って競争しており, いずれかの「資源」の競争において一番にならない限り絶滅してしまう. 類似の現象は生態学において競争排他原理 (competitive exclusion principle) [38] としてよく知られている.

一方実際のヒト健常者の T 細胞クローンの種類数は、見積もりが難しいが、少なめの見積もりでも  $10^6$  種類以上とされており [39], この多様性の大きさと競争排他原理は一見矛盾している. いくつか理由が考えられるが、まず明確にしておくべき点は、T 細胞は病原体由来のペプチドのみを認識するのではなく、正常な組織や細胞に由来するペプチドも認識するという点である. これによって、競争排他原理が成立するとしても正常な組織由来のペプチドの多様性と同じ程度には T 細胞の多様性が大きくても矛盾しない. また、そもそも競争排他原理は成立しないと考えられる. その要因として上述の競争排他則の導出では、胸腺からの流入を無視している点、T 細胞数分布の定常状態への到達を仮定している点、空間的不均一性を無視している点が挙げられる [35].

## 3.2 現象論的な T 細胞集団動態モデル

分子や細胞による素過程から導いたモデルは実際の多様性をうまく説明できなかった. 特に近年は、次世代シーケンサー等計測技術の発展により T 細胞受容体の多様性を実験的に観測できるようになったため、その多様性を説明する数理モデルの構築が重要になってきている. とくに、文章中の単語の出現頻度と同じようなべき乗則 (ジップ則) が T 細胞や B 細胞受容体の頻度分布でも観測されるということが見いだされ [40]–[43], その数理モデルによる再現が試みられている [23], [44]–[46]. これらのモデルは、各 T 細胞クローンの増殖率がばらついていたり、時間的に変動したりする様子を適当な確率分布や確率過程によって確率論的にモデル化したものである. 一方、ニュートンの運動方程式のような 2 階の常微分方程式による決定論的な T 細胞集団動態モデルでも多様性を説明できることが示されている [47]. どのモデルも分子や細胞の素過程からではなく、現象論的 [48] になるべく簡潔な数理モデルを構築するという観点からのモデル化であると言える. 本節では、最低限の仮定から T 細胞集団動態を表すような決定論的モデルを考える.

### 仮定 1

第一に T 細胞は分裂増殖・細胞死・胸腺からの流入という 3 つの素過程によってその細胞数分布が変化すると仮定する.  $K$  種類あるうちの  $k$  番目の T 細胞数を  $n_k$  で表すと、その時間変化は以下のようなになる:

$$\frac{dn_k}{dt} = n_k \lambda_k + \sigma_k. \quad (3.15)$$

ここで、 $\lambda_k$  は増殖率で増殖・死滅の効果を表し、 $\sigma_k$  は胸腺からの時間あたりの流入細胞数で、どちらも時間や T 細胞数分布  $\{n_k\}$  に依存する. 細胞数は非負なので  $n_k$  は常に非

負であるべきだが, この力学系は  $\mathbb{R}_{\geq 0}^K := \{\mathbf{n} \in \mathbb{R}^K \mid n_k \geq 0, k = 1, \dots, K\}$  上の点  $\mathbf{n}_0$  を初期値とすると,  $\mathbb{R}_{\geq 0}^K$  内にとどまり続けることが示せる.

### 仮定2

2つ目の仮定として, 総細胞数  $\|\mathbf{n}\| = \sum_k n_k$  が  $n_{tot}$  より大きい, つまり  $\|\mathbf{n}\| > n_{tot}$  ならば全ての  $k$  で  $\lambda_k < 0$  となるような定数  $n_{tot} > 0$  が存在するとする. 総細胞数が  $n_{tot}$  より多くなると増殖より死滅が支配的になるので, これによって細胞数が無限に発散するような非現実的な解が排除される. 総細胞数が  $n_{tot}$  以下になるような領域を  $\mathcal{N} = \{\mathbf{n} \mid \|\mathbf{n}\| \leq n_{tot}\}$  とかくと, 定常状態の集合  $\mathcal{I}$  は  $\mathcal{N}$  に含まれ,  $\mathcal{I} \subset \mathcal{N}$  である.

### 仮定3

また, 全てのクローンが絶滅せずに生き残るような定常解が存在することを仮定する. すなわち,  $N < K$  でも定常状態において競争排除則は成り立たないとする. 全てのクローンが生存する解を  $\mathbf{n}^\dagger$  として表すことにすると,  $\mathbf{n}^\dagger$  は唯一とは限らずその集合を  $\mathcal{I}^\dagger$  と書く. 仮定2と合わせると

$$\mathcal{I}^\dagger \subset \mathcal{I} \subset \mathcal{N} \subset \mathbb{R}_{\geq 0}^K \quad (3.16)$$

で,  $\mathcal{I}^\dagger \subset \mathbb{R}_{> 0}^K$  である.

### 仮定4

最後に, この力学系は散逸的で, 長期的にはいずれかの平衡点へ収束すること (安定性) を仮定する. これは, 病原体への感染応答で一時的に T 細胞数が変化しても, 病原体が排除されて十分時間が経過すると健康な状態へ戻るというような状況を想定している. 実際, ヒトを対象に免疫細胞数の「状態」を調べると, 個体間のばらつきに比べて同じ個体の時間的なばらつきはとても小さく, 各個体の免疫系は感染やワクチン接種といったイベントのあとにもそれ以前と同様の状態へ戻る人が多いことが分かっている [49].

以上の仮定1から4を元に, 特に増殖率  $\lambda_k$  の関数形を考える. 増殖率を細胞数  $\{n_k\}$  の関数にすることで, T 細胞の競争的な様子を表現できるがここでは最も簡単な一次関数

$$\lambda_k = b_k + \sum_{l=1}^K a_{kl} n_k \quad (3.17)$$

の場合を考える. ベクトル・行列表示すれば

$$\frac{d\mathbf{n}}{dt} = \text{diag}(\mathbf{n})(\mathbf{b} + A\mathbf{n}) + \boldsymbol{\sigma} \quad (3.18)$$

となる. ここで  $\text{diag}(\mathbf{n})$  は  $\{n_k\}$  を並べた対角行列で,  $A$  は  $i, j$  番目の要素が  $a_{ij}$  であるような  $K \times K$  行列である. この増殖率をもつ流入がない場合の集団動態は一般化 Lotka-Volterra モデルとして知られ, 異なるクローン間での競争 ( $a_{kl} < 0$ ) や協力 ( $a_{kl} > 0$ ) の存在を自然と表現できている. 1次関数より複雑なものも考えられるが, 種類数が5より大きい ( $K \geq 5$ ) 場合は1次関数でも, 平衡点やリミットサイクルだけでなくカオスなどあらゆる長時間的振る舞いが現れることが示されている [50], [51].

仮定2および3と整合するための  $A$  や  $\mathbf{b}$  の条件を流入がない場合で考えてみる.  $\mathbf{n} \in I$  は  $n_k \lambda_k = 0 (k = 1, \dots, K)$  を満たすので, 全ての  $k$  で  $n_k = 0$  あるいは  $\lambda_k = 0$  である. 絶滅するクローンの集合を  $\mathcal{K}_0 = \{k \mid 1 \leq k \leq K, n_k = 0\}$ , 生き残るクローンの集合を  $\mathcal{K}_1 = \{k \mid 1 \leq k \leq K, n_k > 0\}$  と表すことにすれば,  $\mathcal{K}_0$  の要素が0で,  $\mathcal{K}_1$  の要素が非ゼロとなる定常解  $\mathbf{n}$  は, 線形連立方程式

$$\mathbf{b}_1 + A_1 \mathbf{n}_1 = 0 \quad (3.19)$$

の解である. ここで,  $\mathbf{b}_1, A_1, \mathbf{n}_1$  はそれぞれ  $\mathbf{b}, A, \mathbf{n}$  の  $\mathcal{K}_1$  に対応する部分を抜き出したものである. 仮定2より, この線形連立方程式の解が有界でなければならない. [52] より, 任意の  $\mathbf{b}$  に対して,  $\mathbf{n} \in \mathbb{R}_{\geq 0}^K$  の範囲で解が有界になることは,  $A_1$  の零空間  $\ker A_1$  と  $\mathbb{R}_{\geq 0}^{|\mathcal{K}_1|}$  の共通部分が原点しかないことと同値である. また, 仮定3より正象限  $\mathbb{R}_{\geq 0}^K$  に要素を持たなければならない. 少なくとも非負象限に解があることは Farkas の補題 [53] によって,  $A^\top \mathbf{v} \geq 0$  かつ  $\mathbf{b}^\top \mathbf{v} < 0$  を満たすような  $\mathbf{v} \in \mathbb{R}^K$  が存在しないことと同値である. また, 疑似逆行列  $A^+$  を使った最小ノルム解  $A^+ \mathbf{b}$  が存在して正象限にあることは十分条件を与える.

次に, 細胞生物学的な T細胞のモデルと近づけるため, 増殖率が  $N$  個の資源をめぐる競争 (あるいは協力) によって決まっているような状況を考える. これは1次関数の係数を  $\mathbf{b} = G\mathbf{r}, A = -GF$  と置くことによって表現できて,

$$\frac{d\mathbf{n}}{dt} = \text{diag}(\mathbf{n})G(\mathbf{r} - F\mathbf{n}) + \boldsymbol{\sigma}, \quad (3.20)$$

あるいは

$$\frac{dn_k}{dt} = n_k \sum_{i=1}^N G_{ki} \left( r_i - \sum_{l=1}^K F_{il} n_l \right) + \sigma_k \quad (3.21)$$

となる. ただし,  $G$  は  $K \times N$  行列,  $F$  は  $N \times K$  行列で  $\mathbf{r}$  は  $N$  次元ベクトルである. これは細胞生物学的なモデルの式3.13において  $\sum_k K_{ik} T_k \ll 1$  の極限における近似式

$$\frac{dT_k}{dt} = T_k \left( \sum_{i=1}^N \kappa k_{ik}^a K_{ik} P_i \left( 1 - \sum_{l=1}^K K_{il} T_l \right) - d \right) + \sigma_k \quad (3.22)$$

と同様の式になっている。つまり、 $N$  種類の抗原やペプチドとの相互作用をめぐって  $K$  種類の T 細胞クローンが競争するようなモデルとして解釈できる。

最後に、このモデルが仮定4の安定性を満たすための条件を考える。平衡点の漸近安定性やリアプノフ安定性の十分条件として、リアプノフ関数の存在がある。安定性の定義やリアプノフ関数を用いる手法については付録Aにまとめた。ここでは、安定性の必要条件は考えず、リアプノフ関数が構成できる例として以下のモデルを紹介する。全ての  $k = 1, \dots, K, i = 1, \dots, N$  で  $G_{ki} = q_i F_{ik}$  つまり  $G = F^\top \text{diag}(\mathbf{q})$  を満たすような  $\mathbf{q} = \{q_i\} \geq 0$  が存在する場合、すなわち

$$\frac{d\mathbf{n}}{dt} = \text{diag}(\mathbf{n})F^\top \text{diag}(\mathbf{q})(\mathbf{r} - F\mathbf{n}) + \boldsymbol{\sigma}, \quad (3.23)$$

あるいは

$$\frac{dn_k}{dt} = n_k \sum_{i=1}^N q_i F_{ik} \left( r_i - \sum_{l=1}^K F_{il} n_l \right) + \sigma_k \quad (3.24)$$

のときを考える。非負定値な関数  $V: \mathbb{R}_{\geq 0}^K \rightarrow \mathbb{R}$  を

$$V(\mathbf{n}) = \frac{1}{2} \sum_{i=1}^N q_i \left( r_i - \sum_{k=1}^K F_{ik} n_k \right)^2 \quad (3.25)$$

とすると、流入がないとき ( $\sigma_k = 0$ ) は

$$\frac{dV(\mathbf{n})}{dt} = \sum_{k=1}^K \frac{\partial V}{\partial n_k} \frac{dn_k}{dt} = - \sum_{k=1}^K n_k \left( \sum_{i=1}^N q_i F_{ik} \left( r_i - \sum_{l=1}^K F_{il} n_l \right) \right)^2 \leq 0 \quad (3.26)$$

を満たすので  $V$  はリアプノフ関数である。集合  $\{\mathbf{n} \in \mathbb{R}_{\geq 0}^K \mid \frac{dV(\mathbf{n})}{dt} = 0\}$  を考えると、これは平衡点の集合  $\mathcal{I}$  と一致する。また、仮定2より全ての解は有界である。したがって、付録Aの定理A.2.7より、任意の点  $\mathbf{n}_0 \in \mathbb{R}_{\geq 0}^K$  を初期値とする解軌道は  $t \rightarrow \infty$  で  $\mathcal{I}$  に近づく。ただし、 $\mathcal{I}$  は孤立した有限個の平衡点ではなく、無限個の連結した平衡点となりうるため、これらの平衡点は安定ではない。なぜなら、ある平衡点  $\mathbf{n}^*$  の安定性を示そうとすると近傍の任意の点  $\mathbf{n}$  で  $V(\mathbf{n}) > V(\mathbf{n}^*)$  を示す必要があるが、近傍には別の平衡点  $\mathbf{n}^{**}$  があ  $V(\mathbf{n}^*) = V(\mathbf{n}^{**})$  となる。しかし、平衡点の集合としては大域的に漸近安定である。

### 3.3 規範的な T 細胞集団動態モデル

本節では、T 細胞や B 細胞の集団動態をある種の最適化の過程として規範的にモデル化している先行研究を紹介する。ここ数年で T 細胞や B 細胞の振る舞いを最適化のように

モデル化する研究は増えている。T細胞とB細胞の双方に適用できるようなやや抽象的なモデル [54], [55] もあるが、より最近の研究では、抗体 [56] や記憶B細胞 [57], ヘルパーT細胞 [23] や制御性T細胞 [58] など個別の現象に注目した研究が多い。

ここでは、2015年のMayerらによる先行研究 [54] を紹介する。まず異なる受容体  $k = 1, \dots, K$  を持った獲得免疫系のT細胞あるいはB細胞の集団（細胞数分布  $\mathbf{n} = \{n_k\}$ ）と、病原体由来の異なる種類  $i = 1, \dots, N$  の抗原の集団（頻度分布  $\mathbf{q} = \{q_i\}$ ）を考える。各抗原はある確率で受容体をもつ細胞と出会い、そのペアの結合親和性によって決まるある確率で細胞は抗原を認識できる。抗原をうまく認識できないと生体に害をなすと考えられるのでその大きさをコスト  $C$  として表現する。細胞集団と抗原集団による多対多の反応なので、各ペアの和によって  $C$  を記述してもよいが、ここでは細胞集団の頻度分布  $p_k = n_k / \sum_l n_l$  を、抗原被覆率  $P_i := \sum_k f_{ik} p_k$  に変換して考える。ここで、 $f_{ik}$  は  $k$  番目の受容体と  $i$  番目の抗原の結合親和性を表すような値である。抗原被覆率を用いることによって、抗原の種類に関する和として以下のようにコスト関数を定義する。

$$C(\mathbf{n}) = \sum_{i=1}^N q_i c(P_i). \quad (3.27)$$

ここで、 $c: \mathbb{R} \rightarrow \mathbb{R}$  は各抗原の抗原被覆率に依存して決まる害の大きさを、様々な関数形が考えられる。 $c(P) = P^{-\alpha}$  のときをべき乗型コスト、 $c(P) = -\log(P)$  のときを対数型コストと呼ぶことにする。病原体の指数的な増殖などを仮定するとべき乗型コストが自然であることが導ける [54]。

最適な免疫細胞の細胞数分布は、このコスト関数  $C$  を最小化するものとして定義される。細胞数分布ではなく、抗原被覆率は解析的に最適なものを求められる場合があるのでまずはこれを示す。コスト関数を抗原被覆率の関数として  $C(\mathbf{P})$  と書くと、指数  $\alpha$  のべき乗型コストのとき

$$C(\mathbf{P}) = \sum_{i=1}^N q_i P_i^{-\alpha} \quad (3.28)$$

$$= Z^{1+\alpha} \sum_{i=1}^N \frac{(q_i^{1/(1+\alpha)} / Z)^{1+\alpha}}{P_i^\alpha} \quad (3.29)$$

$$= Z^\beta \exp \left\{ (\beta - 1) D_\beta \left( \frac{\mathbf{q}^{1/\beta}}{Z} \parallel P \right) \right\}. \quad (3.30)$$

ただし,  $Z = \sum_i q_i^{1/(1+\alpha)}$ ,  $\beta = 1 + \alpha$  で,  $D_\beta$  は Renyi ダイバージェンス

$$D_\beta(\mathbf{p}||\mathbf{q}) = \frac{1}{\beta-1} \log \left( \sum_i \frac{p_i^\beta}{q_i^{\beta-1}} \right) \quad (3.31)$$

である. Renyi ダイバージェンスは  $\mathbf{p} = \mathbf{q}$  のときのみ  $D(\mathbf{p}||\mathbf{q}) = 0$  でそれ以外のとき  $D(\mathbf{p}||\mathbf{q}) > 0$  となる性質を持つので, 最適な  $\mathbf{P}$  は  $P_i = q_i^\beta / Z$  で与えられる. 対数型コストのときは  $C$  が交差エントロピーとなって, 同様にして最適な  $\mathbf{P}$  は  $P_i = q_i$  と求まる.

次に, このコスト関数を最小化するような免疫細胞の集団動態を考える. 細胞生物学的なモデルや現象論的なモデルと同様に,  $N$  種類の抗原と反応できた細胞だけが増殖でき, 抗原をめぐる細胞間で競争が生じているというようなモデルを用いる. 細胞生物学的なモデル (式3.13) では  $1/(1 + \sum_k K_{ik} T_k)$ , 現象論的なモデル (式3.24) では  $r_i - \sum_k F_{ik} n_k$  という項が, 競争によって反応できる抗原数が実質的に減るような効果を表していた. ここではその効果を抗原被覆量  $N_i := \sum_k f_{ki} n_k$  の減少関数  $A: \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$  として表す. すると以下のような細胞数の時間発展となる

$$\frac{dn_k}{dt} = n_k \left( \sum_{i=1}^N q_i f_{ik} A \left( \sum_{l=1}^K f_{il} n_l \right) - d \right). \quad (3.32)$$

定常状態での総細胞数を  $n_{st}$  としたときに,  $A$  と前述のコスト関数を与える  $c$  との間に以下の関係式

$$A(N_i) = -\gamma c' \left( \frac{N_i}{n_{st}} \right) \quad (3.33)$$

が成立して, かつ後述する  $A$  のスケール不変性が満たされる場合, この細胞数動態の平衡点はコスト関数を最小化することが示せる. 実際, コスト関数  $C(\mathbf{n})$  はリアプノフ関数になっていることが以下のように示せる.

まず増殖率に相当する項を  $\lambda_k(\mathbf{n})$  とかくと式3.32は

$$\frac{dn_k}{dt} = n_k (\lambda_k(\mathbf{n}) - d) \quad (3.34)$$

となる. 細胞数  $n_k$  を総細胞数  $n_{tot} = \sum_l n_l$  で割った割合  $p_k = n_k / n_{tot}$  に変数変換する.

$$\frac{dp_k}{dt} = \frac{1}{n_{tot}} \frac{dn_k}{dt} - \frac{n_k}{n_{tot}^2} \frac{dn_{tot}}{dt} \quad (3.35)$$

$$= p_k \left( \lambda_k(\mathbf{n}) - \sum_{l=1}^K p_l \lambda_l(\mathbf{n}) \right). \quad (3.36)$$



$A$  がスケール不変性を満たす, つまり関数  $\nu$  を用いて任意のスカラー  $a$  で  $A(ax) = \nu(a)A(x)$  とできると仮定する. べき乗型コスト  $c(P) = P^{-\alpha}$  のときは満たされる. すると, 増殖率が  $\lambda_k(\mathbf{n}) = \nu(n_{tot})\lambda_k(\mathbf{p})$  のように分解できる. よって細胞数割合の時間発展は

$$\frac{dp_k}{dt} = \nu(n_{tot})p_k (\lambda_k(\mathbf{p}) - \bar{\lambda}(\mathbf{p})) \quad (3.37)$$

となる. ここで,  $\bar{\lambda}(\mathbf{p})$  は  $\lambda_k$  の平均値  $\sum_k p_k \lambda_k(\mathbf{p})$  である.

コスト関数の時間発展は

$$\frac{dC}{dt} = \sum_{k=1}^K \frac{\partial C}{\partial p_k} \frac{dp_k}{dt} \quad (3.38)$$

$$= \nu(n_{tot}) \sum_{k=1}^K p_k \left( \sum_{i=1}^N q_i f_{ik} c' \left( \sum_{l=1}^L f_{il} p_l \right) \right) (\lambda_k(\mathbf{p}) - \bar{\lambda}(\mathbf{p})) \quad (3.39)$$

$$= -\frac{\nu(n_{tot})}{\gamma} \sum_{k=1}^K p_k \left( \sum_{i=1}^N q_i f_{ik} A \left( n_{st} \sum_{l=1}^K f_{il} p_l \right) \right) (\lambda_k(\mathbf{p}) - \bar{\lambda}(\mathbf{p})) \quad (3.40)$$

$$= -\frac{\nu(n_{tot})\nu(n_{st})}{\gamma} \sum_{k=1}^K p_k \left( \sum_{i=1}^N q_i f_{ik} A \left( \sum_{l=1}^K f_{il} p_l \right) \right) (\lambda_k(\mathbf{p}) - \bar{\lambda}(\mathbf{p})) \quad (3.41)$$

$$= -\frac{\nu(n_{tot})\nu(n_{st})}{\gamma} \sum_{k=1}^K p_k \lambda_k(\mathbf{p}) (\lambda_k(\mathbf{p}) - \bar{\lambda}(\mathbf{p})) \quad (3.42)$$

$$= -\frac{\nu(n_{tot})\nu(n_{st})}{\gamma} \sum_{k=1}^K p_k (\lambda_k(\mathbf{p}) - \bar{\lambda}(\mathbf{p}))^2 \leq 0. \quad (3.43)$$

したがって, コスト関数  $C$  はリアプノフ関数になっている.

式3.33は, 指数  $\alpha$  のべき乗型コスト  $c(P) = P^{-\alpha}$  のとき

$$A(N_i) = \gamma\alpha \left( \frac{n_{st}}{N_i} \right)^{1-\alpha} \quad (3.44)$$

となる.  $\alpha \rightarrow 0$  で対数型コスト近づいて,

$$A(N_i) = \gamma\alpha \frac{n_{st}}{N_i} \quad (3.45)$$

となり, 細胞生物学的モデル3.13の  $1 \ll \sum_l K_{il} T_l$  のときに相当する.

## 第4章

# 勾配系による T 細胞集団動態のモデル化

第3章では、様々な異なる観点からの T 細胞集団動態のモデル化について紹介した。現象論的なモデルでも安定性の要求からリアプノフ関数が登場し、規範的なモデルでもリアプノフ関数を通してコスト関数が最小化されることを議論していた。リアプノフ関数が存在するような力学系の中でも、勾配系という特別な力学系がある。本章ではリーマン多様体における勾配系の考え方を導入する。勾配系は最適性という観点からみてより直接的で自然な解釈を与えるため、規範的なモデル化に適している。また、異なるリーマン多様体上での勾配系を考えることで、様々な力学系を表現できるという点で汎用性が高い。この枠組みを T 細胞の増殖や流入に応用することで対応するリーマン多様体を同定する。この結果は、以降の第5-6章で用いられる。

### 4.1 リーマン多様体上の勾配系

本章では、有限次元の距離空間  $\mathcal{X}$ 、例えば  $K$  次元ユークリッド空間  $\mathcal{X} = \mathbb{R}^K$  における勾配系を考える。勾配系とは、リアプノフ関数に対応する  $\mathcal{F} : \mathcal{X} \rightarrow \mathbb{R}$  の勾配に従って時間発展するような力学系  $\{x(t) \in \mathcal{X} \mid t \in [0, \infty)\}$  で、微分方程式でかけば

$$\frac{dx}{dt} = -\text{grad } \mathcal{F}(x) \quad (4.1)$$

のようになるもののことである。以下では  $\mathcal{F}$  は滑らかで、(広義の) 凸関数だとする。すなわち十分大きな  $n$  について  $n$  階連続的微分可能で、任意の  $x_1, x_2 \in \mathcal{X}$  と任意の  $s \in [0, 1]$  で  $\mathcal{F}(sx_1 + (1-s)x_2) \leq s\mathcal{F}(x_1) + (1-s)\mathcal{F}(x_2)$  が成り立つとする。  $\mathcal{X}$  を  $d$  次元ユーク

リッド空間だと考えると,  $\text{grad}$  は通常の偏微分を表すので,

$$\frac{dx_i}{dt} = -\frac{\partial \mathcal{F}}{\partial x_i}(x) \quad (i = 1, \dots, d) \quad (4.2)$$

となる. 式4.1は  $\mathcal{X}$  が滑らかなリーマン多様体である場合にも拡張できる. リーマン多様体は任意の  $x \in \mathcal{X}$  における内積  $\langle \cdot, \cdot \rangle_x : T_x \mathcal{X} \times T_x \mathcal{X} \rightarrow \mathbb{R}$  によって同定される. ここで,  $T_x \mathcal{X}$  は  $x \in \mathcal{X}$  における接ベクトル空間である. 内積は計量テンソル  $G(x) = \{g_{ij}(x)\}$  を用いて  $\langle v_1, v_2 \rangle_x = \sum_{ij} g_{ij}(x)v_{1,i}v_{2,j}$  と書ける.

勾配  $\text{grad } \mathcal{F}(x)$  は任意の接ベクトル  $\xi \in T_x \mathcal{X}$  に対して

$$\langle \text{grad } \mathcal{F}(x), \xi \rangle_x = D_x \mathcal{F}(\xi) \quad (4.3)$$

を満たす接ベクトルとして定義される. ここで,  $D_x \mathcal{F}(\xi)$  は  $x$  における  $\xi$  に沿った方向微分である. 正定値行列である計量テンソル  $G(x)$  を用いれば,

$$\text{grad } \mathcal{F}(x) = G^{-1}(x) \frac{\partial \mathcal{F}(x)}{\partial x}, \quad (4.4)$$

$$[\text{grad } \mathcal{F}(x)]_i = \sum_j g_{ij}^{-1}(x) \frac{\partial \mathcal{F}(x)}{\partial x_j} \quad (4.5)$$

のようになる. ただし,  $G^{-1}(x)$  は  $G(x)$  の逆行列で,  $g_{ij}^{-1}(x)$  はその  $i, j$  番目の要素を表す.  $\{x(t)\}_t$  が式4.1に従う曲線だとすると,  $x(t)$  に沿った  $\mathcal{F}$  の時間変化は

$$\frac{d\mathcal{F}(x(t))}{dt} = -\|\text{grad } \mathcal{F}(x(t))\|^2 \leq 0 \quad (4.6)$$

となるため, 勾配系 (式4.1) において関数  $\mathcal{F}$  はリアプノフ関数である.

一方, 距離空間においては別の方法で勾配系を特徴づけることができる. 時刻  $t \in [0, \infty)$  を時間幅  $h$  で分割し,  $t = 0$  での  $x_0$  をもとに時刻  $t = h\tau$  ( $\tau = 1, 2, \dots$ ) における  $x(t) = x^{(\tau)}$  を以下のように定める:

$$x^{(\tau+1)} = \arg \min_{x \in \mathcal{X}} \left\{ \mathcal{F}(x) + \frac{1}{2h} d(x^{(\tau)}, x)^2 \right\}. \quad (4.7)$$

ここで,  $d : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  は距離関数である. この手続によって与えられる離散的な時刻の間の時刻  $t \in (h\tau, h(\tau+1))$  においても適切に補間し時間幅  $h \rightarrow 0$  の極限をとることも勾配系の時間発展を記述できる. これは移動最小化スキーム (Minimizing Movement Scheme) と呼ばれる [59]. 実際,  $F$  の性質がよければ, 時間幅無限小の極限は収束して

式4.1と一致する [60]. その際, リーマン多様体上では, 距離  $d$  は測地線の長さとして自然に定義される

$$d(x_1, x_2)^2 = \inf_{x \in \mathcal{X}} \left\{ \int_0^1 \left\| \frac{dx_t}{dt} \right\|_{x_t}^2 dt \mid x_1 \overset{x}{\rightsquigarrow} x_2 \right\}. \quad (4.8)$$

ここで,  $x_1 \overset{x}{\rightsquigarrow} x_2$  は  $x$  が  $x_1$  から  $x_2$  をつなぐ滑らかな曲線であることを意味する. ノルム  $\|\cdot\|_x$  は内積から誘導されるノルムである.

移動最小化スキームの式4.7は, 以下のような制約付き最適化として考えてもほぼ等価である:

$$\begin{aligned} x^{(\tau+1)} = \arg \min_{x \in \mathcal{X}} \mathcal{F}(x) \\ \text{subject to } d(x^{(\tau)}, x)^2 \leq \rho. \end{aligned} \quad (4.9)$$

実際, ラグランジュ未定定数  $\lambda \geq 0$  を用いれば最適性の必要条件の一つは

$$\frac{d}{dx} \left\{ \mathcal{F}(x) + \lambda d(x^{(\tau)}, x)^2 \right\} = 0 \quad (4.10)$$

であり,  $\mathcal{F}(x) + \lambda d(x^{(\tau)}, x)^2$  が凸関数なので, 以下のような制約のない最適化問題として書き直せる:

$$x^{(\tau+1)} = \arg \min_{x \in \mathcal{X}} \left\{ \mathcal{F}(x) + \lambda d(x^{(\tau)}, x)^2 \right\}. \quad (4.11)$$

ここで,  $\lambda$  は最適性の他の条件とともに決まる  $\rho$  の関数であり,  $\lambda > 0$  のとき, その逆数を  $\varepsilon = \lambda(\rho)^{-1}$  と書くことにする.  $\varepsilon$  は式4.7では時間幅  $h$  の2倍に対応していた.  $\mathcal{X} = \mathbb{R}^d$  で  $\rho$  や  $\varepsilon$  が十分小さいときを考えると,  $x^{(\tau+1)} = x^{(\tau)} + \varepsilon \Delta x + o(\varepsilon)$  となって,

$$x^{(\tau+1)} = x^{(\tau)} - \varepsilon G^{-1}(x^{(\tau)}) \frac{\partial \mathcal{F}}{\partial x}(x^{(\tau)}) + o(\varepsilon). \quad (4.12)$$

ただし,  $G^{-1}(x^{(\tau)})$  は距離の二乗を  $d(x^{(\tau)}, x^{(\tau)} + \varepsilon \Delta x)^2 = \frac{\varepsilon^2}{2} \Delta x^\top H(x^{(\tau)}) \Delta x + o(\varepsilon^2)$  のように展開したヘッセ行列  $H(x^{(\tau)})$  で,  $x^{(\tau)}$  における計量テンソルの逆行列になる.  $o(\varepsilon)$  の項を無視すれば, これは元の定義の勾配系 (式4.1, 式4.4) のオイラー法による前進差分化になっている.  $\mathcal{X}$  がユークリッド空間でなく, その部分空間である場合には式4.12による更新で  $x^{(\tau+1)}$  が  $\mathcal{X}$  からはみ出る可能性がある. その場合は指数写像やレトラクションによって射影する [61].

制約付き最適化としての定式化から, 勾配系が目的関数最小化という機能の観点で最も優れていることが示唆される. 同じリアプノフ関数をもつ力学系は無限に多く考える事ができ, それらも目的関数としてのリアプノフ関数を最小化するという機能を持っていると

いえる。一方勾配系は式4.9のように、現在の点  $x^{(\tau)}$  の近傍の中で目的関数  $\mathcal{F}$  を最も小さくする点に移動する、という意味で局所的には考える最も効率の良い選択を繰り返している。

## 4.2 機械学習・情報分野における勾配系の応用例

勾配系は目的関数の最小化という機能をもつことを示したが、実際に機械学習や数理最適化の分野では勾配系の時間離散化に相当する勾配降下法 (gradient descent method) がよく用いられる。ただし、距離の二乗  $d^2$  のかわりにダイバージェンス  $D$  を用いることが多い。ダイバージェンスは距離のように2つの与えられた点の非類似度を測る関数で、非負性と非退化性を満たすが対称性と三角不等式を満たすとは限らない [62]。ダイバージェンスについては付録Bにまとめた。また、勾配降下法の枠組みでは、ラグランジュ定数の逆数に対応する量  $\varepsilon$  はステップサイズあるいは学習率と呼ばれる。勾配系の離散化という観点からは学習率は一定の値で、小さい方が連続時間の勾配系に近づくためその挙動が理解しやすい一方で、学習率を変化させ、時には大きくすることで最適解への収束を加速できることが知られている [63]。

### 4.2.1 鏡像降下法

以下では勾配降下法の例をいくつか紹介する。近接勾配法 (proximal gradient method) や鏡像降下法 (mirror descent method)[64], [65] は移動最小化スキーム (式4.7) の類いの式を用いて反復計算を行うアルゴリズムである。実際には式4.7の目的関数  $\mathcal{F}(x)$  を線形近似したものが用いられる。鏡像降下法は  $d^2$  の代わりに Bregman ダイバージェンス  $D_\psi$  (詳細は付録B) を用いて

$$x^{(\tau+1)} = \arg \min_{x \in \mathcal{X}} \left\{ (x - x^{(\tau)})^\top \frac{\partial \mathcal{F}}{\partial x}(x^{(\tau)}) + \frac{1}{2h} D_\psi(x^{(\tau)} \| x) \right\} \quad (4.13)$$

とする。たとえば、離散確率分布の空間である確率単体  $\mathcal{X} = \{x \in \mathbb{R}^K \mid 0 \leq x_k \leq 1, \sum_k x_k = 1\}$  でダイバージェンスとして Kullback-Leibler(KL) ダイバージェンス (付録B) を用いると指数勾配降下法 (exponentiated gradient descent method)[66] というアルゴリズムになり、以下のような更新式となる:

$$x_k^{(\tau)} = \frac{w_k^{(\tau)} x_k^{(\tau)}}{\sum_l w_l^{(\tau)} x_l^{(\tau)}}. \quad (4.14)$$

ここで、 $w_k^{(\tau)} = \exp(-\varepsilon \partial \mathcal{F} / \partial x_k)$  である。

### 4.2.2 自然勾配降下法と情報幾何

鏡像降下法とは異なり, 自然勾配降下法 (natural gradient descent method)[67] は計量テンソルによる式4.12を用いた反復計算を行うアルゴリズムである. このアルゴリズムは情報幾何学という確率分布の空間を微分幾何学的に考える分野と深く関係している. 情報幾何学では  $\theta$  をパラメータにもつ確率変数  $x$  の確率モデル  $p(x; \theta)$  の空間を, Fisher 情報行列

$$I_{ij}(\theta) = \sum_x p(x; \theta) \frac{\partial \log p(x; \theta)}{\partial \theta_i} \frac{\partial \log p(x; \theta)}{\partial \theta_j} \quad (4.15)$$

を計量テンソル (Fisher 計量) として与えたりーマン多様体として考える. この計量を用いて式4.12を繰り返すアルゴリズムが自然勾配降下法である.

情報幾何学では, 測地線の長さとして定義される距離 (式4.8) よりも, 距離の公理を満たさない Bregman ダイバージェンスや  $f$ -ダイバージェンスが本質的であるとしてよく用いられる. その2つのダイバージェンス族に属する唯一のダイバージェンスが,  $\alpha$ -ダイバージェンスである [68]. 詳細は付録Bにまとめたが, ここでは  $f$ -ダイバージェンスのみ紹介する. また, 確率分布の空間を考える場合は確率の和が1に正規化されているが, ここでは正規化されているとは限らない非負測度の空間を考えることにする.

Csiszár の  $f$ -ダイバージェンス [69] は

$$D_f(x||y) = \sum_k x_k f\left(\frac{y_k}{x_k}\right) \quad (4.16)$$

として定義される. ここで,  $f$  は  $f(1) = 0, f'(1) = 0$  を満たす凸関数で, 標準的な  $f$ -ダイバージェンスでは微分係数が  $f''(1) = 1$  を満たすような  $f$  を用いる.

### 4.2.3 確率的勾配降下法

目的関数  $\mathcal{F} : \mathcal{X} \rightarrow \mathbb{R}$  が期待値の形で

$$\mathcal{F}(x) = \mathbb{E}[L(z, x)] \quad (4.17)$$

のように与えられることは多々ある. ここで,  $L$  はスカラー関数  $L : \mathcal{Z} \times \mathcal{X} \rightarrow \mathbb{R}$  で, 期待値は  $\mathcal{Z}$  に値をとる確率変数  $z$  についてとっている. 簡単のため  $\mathcal{Z}$  は有限集合とする. ユ

ークリッド距離を用いた通常の勾配降下法では

$$x^{(\tau+1)} = x^{(\tau)} - \alpha_\tau \frac{\partial}{\partial x} \mathbb{E} \left[ L(z, x^{(\tau)}) \right] \quad (4.18)$$

$$= x^{(\tau)} - \alpha_\tau \frac{\partial}{\partial x} \sum_{z \in \mathcal{Z}} P(z) L(z, x^{(\tau)}) \quad (4.19)$$

$$= x^{(\tau)} - \alpha_\tau \sum_{z \in \mathcal{Z}} P(z) \frac{\partial}{\partial x} L(z, x^{(\tau)}) \quad (4.20)$$

の反復計算を行う。ここで  $\alpha$  は学習率で、 $P(z)$  は  $z$  の確率分布である。しかし実際には  $z$  の確率分布  $P(z)$  は与えられず、各時刻  $\tau$  で  $P(z)$  からのサンプル  $z_\tau$  のみ与えられるような場合は勾配の期待値を計算できない。そこで、期待値を実現値  $z_\tau$  で置き換えて

$$x^{(\tau+1)} = x^{(\tau)} - \alpha_\tau \frac{\partial}{\partial x} L(z_\tau, x^{(\tau)}) \quad (4.21)$$

による反復計算を行う手法がある。このように確率的な勾配を用いる方法は確率的勾配降下法 (stochastic gradient descent) と呼ばれる。確率的な勾配は真の勾配の不偏推定量になっている。真の勾配ではなく確率的な勾配を用いることにより、勾配系の性質は一般に満たさなくなるため、解析には注意を要する。よく知られた一般的な結果としては、通常のユークリッド空間の確率的勾配降下法について、学習率  $\alpha \geq 0$  を時間とともに小さくしていくと確率 1 で最適解へ収束することが示されている [70], [71]。その際、 $\{\alpha_t\}$  は

$$\alpha_t > 0, \quad \alpha_t \rightarrow 0, \quad \sum_{t=1}^{\infty} \alpha_t = \infty, \quad (4.22)$$

および

$$\sum_{t=1}^{\infty} \alpha_t^2 < \infty \quad (4.23)$$

を満たす必要がある。リーマン多様体上の確率的勾配降下法でも同様の結果が示されている [72]。

### 4.3 T 細胞集団動態のリーマン多様体構造

現象論的なモデルや規範的なモデルにおいて述べたように、T 細胞集団動態が正常な状態や最適な状態への緩和という役割を持っているとすると、その緩和の速さは個体の生存率に影響を与えると考えられる。よって、自然選択の原理が適用できるとすれば、T 細胞集団動態が勾配系としてモデル化できるというのは自然な仮定である。実際の T 細胞の集

団動態が勾配系であるとは限らないが、この仮定によって見通しが良くなり、理論的に多くの性質が導ける。最終的には導かれた性質を実際の免疫系と照らし合わせることによって、モデルの妥当性を検証することを目指す。

目的関数の具体的な形については次章で検討することにして、本節では T 細胞の増殖・死滅と流入という素過程に対応してリーマン多様体の構造を同定する。さらに、同定されたリーマン多様体のダイバージェンスがその素過程に付随する生物学的コストとして自然に解釈できることを示す。

### 増殖・死滅過程のリーマン多様体構造

まず、細胞数分布  $\mathbf{n}$  は非負の  $K$  次元ベクトル  $\mathbf{n} \in \mathbb{R}_{\geq 0}^K$  である。増殖率が何らかの目的関数  $\mathcal{F}_p$  の偏微分  $\partial \mathcal{F}_p / \partial \mathbf{n}$  で表されるとすれば、増殖・死滅は

$$\frac{d\mathbf{n}}{dt} = -\text{diag}(\mathbf{n}) \frac{\partial \mathcal{F}_p}{\partial \mathbf{n}}(\mathbf{n}) \quad (4.24)$$

として表される。つまり、 $\text{grad } \mathcal{F}_p(\mathbf{n}) = \text{diag}(\mathbf{n}) \frac{\partial \mathcal{F}_p}{\partial \mathbf{n}}(\mathbf{n})$  となるようなリーマン多様体を見つけたい。式4.4から、計量テンソルとして  $G_p(\mathbf{n}) = \text{diag}(\mathbf{n})^{-1}$  とすればよい。これは、Fisher 情報行列と等価であるため、情報幾何学と同様のリーマン多様体を考えることになる。ただし、和が1に正規化されている確率分布ではなく非負測度の空間としてのリーマン多様体である。

Csiszár の  $f$ -ダイバージェンスは Fisher 計量を導くことが分かる。なぜなら、

$$D_f(x \| x + \varepsilon \Delta x) = \sum_{k=1}^K x_k f\left(\frac{x_k + \varepsilon \Delta x_k}{x_k}\right) \quad (4.25)$$

$$= \sum_{k=1}^K x_k \left( f(1) + \varepsilon \Delta x_k f'(1) + \frac{1}{2} (\varepsilon \Delta x_k)^2 f''(1) + o(\varepsilon^2) \right) \quad (4.26)$$

$$= \sum_{k=1}^K x_k \left( \frac{1}{2} (\varepsilon \Delta x_k)^2 + o(\varepsilon^2) \right) \quad (4.27)$$

$$= \frac{\varepsilon^2}{2} \Delta x^\top \text{diag}(\mathbf{n}) \Delta x_k + o(\varepsilon^2) \quad (4.28)$$

となるからである。すなわち、式4.24は、以下のような  $f$ -ダイバージェンスを用いた移動最小化スキーム (式4.7) の連続時間極限としてみなせる：

$$\mathbf{n}^{(\tau+1)} = \arg \min_{\mathbf{n} \in \mathbb{R}_{\geq 0}^K} \left\{ \mathcal{F}_p(\mathbf{n}) + \frac{1}{\varepsilon} \sum_{k=1}^K n_k^{(\tau)} f\left(\frac{n_k}{n_k^{(\tau)}}\right) \right\}. \quad (4.29)$$



この式は生物学的な解釈が以下のように可能である。この  $\tau$  から  $\tau + 1$  への時間変化において、本来の目標は目的関数  $\mathcal{F}_p$  を小さくすることである。しかし、最小化されているものは目的関数と  $f$ -ダイバージェンスの重み付き和であるので、 $f$ -ダイバージェンスの項は最小化にあたっての生物学的なコストを表していると考えたと自然である。例えば、 $f$ -ダイバージェンスは細胞を増殖したり死滅させるにあたって必要となるエネルギーや物質の量と対応していると考えられる。実際、 $f$ -ダイバージェンスは、 $f(n_k/n_k^{(\tau)})$  の値を現在の細胞数  $n_k^{(\tau)}$  で重み付けして和をとった形になっている。 $f$  は正の値をとるので、これを  $k$  番目のクローンの細胞一つあたりの必要なエネルギーや物質の量 (コスト) と考えると自然である。 $f(n_k/n_k^{(\tau)})$  は、現在と比べて細胞数が何倍に増殖するかという細胞数比率  $\eta = n_k/n_k^{(\tau)}$  によって  $f$  で評価されている。 $f$  の関数形より、細胞が増殖も死滅もしない場合が  $\eta = 1$  で最も小さく、増殖が起こる  $\eta > 1$  や死滅が起こる  $\eta < 1$  では大きな値になる。この点でも、 $f(n_k/n_k^{(\tau)})$  を 1 細胞あたりのコストと考えて自然である。

$f$ -ダイバージェンスの中でも、KL ダイバージェンスを用いる場合は特徴的である。 $\varepsilon$  の 2 次以降も考慮した更新は、式 4.12 ではなく、指数勾配降下法の式 4.14 のようになる。これは、細胞生物学的なモデルの式 3.13 の  $1 \ll \sum_l K_{il} T_l$  極限の式と類似している。

### 流入過程のリーマン多様体構造

次に、流入に対応するリーマン多様体を同定する。流入率も何らかの目的関数  $\mathcal{F}_i$  の偏微分で表されるとすれば、流入は

$$\frac{d\mathbf{n}}{dt} = -\frac{\partial \mathcal{F}_i}{\partial \mathbf{n}}(\mathbf{n}) \quad (4.30)$$

として表される。これは計量テンソルが単位行列  $G_i(\mathbf{n}) = I$  となるようなリーマン多様体での勾配系になっている。明らかにユークリッド空間はそのようなリーマン多様体である。ここで、ユークリッド距離の二乗を少し拡張した以下のようなダイバージェンス (ここでは  $g$ -ダイバージェンスと呼ぶことにする) を考える。

$$D_g(x||y) = \sum_{k=1}^K g(y_k - x_k). \quad (4.31)$$

ただし、 $g: \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$  は凸関数で、 $g(0) = 0, g'(0) = 0, g''(0) = 1$  を満たすとする。すると  $g(u) = \frac{1}{2}u^2$  のときに  $g$ -ダイバージェンスはユークリッド距離の二乗と定数倍を除いて一致する。 $g$ -ダイバージェンスはユークリッド空間と同じ単位行列を計量テンソルとして

導く:

$$D(x||x + \varepsilon\Delta x) = \sum_k g(x_k + \varepsilon\Delta x_k - x_k) \quad (4.32)$$

$$= \sum_k g(\varepsilon\Delta x_k) \quad (4.33)$$

$$= \sum_k \left( g(0) + \varepsilon\Delta x_k g'(0) + \frac{1}{2}(\varepsilon\Delta x_k)^2 g''(0) \right) \quad (4.34)$$

$$= \sum_k \left( \frac{1}{2}(\varepsilon\Delta x_k)^2 \right). \quad (4.35)$$

したがって、流入という素過程に対応する移動最小化スキーム (式4.7) は以下のようになる:

$$\mathbf{n}^{(\tau+1)} = \arg \max_{\mathbf{n} \in \mathbb{R}_{\geq 0}^K} \left\{ \mathcal{F}_i(\mathbf{n}) + \frac{1}{\varepsilon} \sum_{k=1}^K g(n_k - n_k^{(\tau)}) \right\}. \quad (4.36)$$

増殖・死滅の場合と同様に考えると、 $g$ -ダイバージェンスは生物学的なコストに対応する。実際、 $g$ -ダイバージェンスは現在の細胞数  $n_k^{(\tau)}$  の絶対値ではなく、その増分  $n_k - n_k^{(\tau)}$  によりのみ依存して関数  $g$  で評価され、クローン全体の和を取る形になっている。胸腺からの新たな T 細胞の流入にあたって必要なコストは、新しい T 細胞をつくって成熟させるコストだと考えられ、既に存在する細胞の数には依存しないと考えるのが自然であり、 $g$ -ダイバージェンスと整合が取れている。

## 第5章

# マルコフ決定過程による免疫系のモデル化

第4章では、最適化の過程を表す自然なモデルとして勾配系の考え方を紹介し、T細胞の集団動態を導くようなリーマン多様体構造を同定した。それではT細胞集団はどのような目的関数を最適化しているのだろうか。すなわち、T細胞の機能とは何か、どのような問題を解いているのかというのが本章の主題である。それにあたってT細胞集団だけでなく免疫系全体に視野を広げ、逐次的な意思決定の問題を解いているという考え方を紹介する。そしてその具体的なモデルとしてKatoらによるマルコフ決定過程と強化学習によるモデル [23] とその拡張を示す。

### 5.1 マルコフ決定過程

動物の学習行動をモデル化するために作られた枠組みが**マルコフ決定過程 (Markov decision processes)** および**強化学習 (reinforcement learning)** である。これらの枠組みは、現在では機械学習としてロボットの制御などに用いられている。マルコフ決定過程は、不確実性のある意思決定を逐次的に繰り返すという問題設定を、「環境」と「エージェント」の相互作用としてモデル化したものである。ここではマルコフ決定過程より一般的な部分観測マルコフ決定過程 (partially observable Markov decision processes) を紹介する。離散的な時刻  $t = 0, 1, \dots$  ごとに状態  $z_t \in \mathcal{Z}$ , 観測  $s_t \in \mathcal{S}$ , 行動  $a_t \in \mathcal{A}$ , 報酬  $r_t \in \mathbb{R}$  という変数がある。状態観測行動空間の大きさは有限  $|\mathcal{Z}|, |\mathcal{S}|, |\mathcal{A}| < \infty$  とする。まず時刻  $t = 0$  において環境は初期状態分布  $P_0(z_0)$  から初期状態  $z_0$  をサンプリングする。時刻  $t = 0, 1, \dots$  においてエージェントは環境の状態  $z_t$  を直接得ることはできず、状態で

条件付けられた観測確率  $\phi(s_t|z_t)$  によってサンプリングされた観測  $s_t$  を受け取る. この観測に基づいて, エージェントは方策 ( $\mathcal{S}$  で条件付けられた  $\mathcal{A}$  の確率)  $\pi(a|s)$  に従って行動  $a_t$  をサンプリングして決定する. 環境は現在の状態に対して行動がどれほど良いかを評価して報酬  $r_t$  としてエージェントに返し, 状態遷移確率  $P(z'|z, a)$  に従って, 時刻  $t+1$  の状態  $z_{t+1}$  へ確率的に遷移する. ここで, 報酬は状態と行動に条件付けられた確率分布  $P(r|z, a)$  に従ってサンプルされる確率変数で, 有限な平均  $R(z, a)$  と有限な分散をもつとする. エージェントは与えられる報酬あるいはその将来にわたる累積和の期待値

$$\mathbb{E} \left[ \sum_{t=0}^{\infty} \gamma^t r_t \right] = \lim_{T \rightarrow \infty} \sum_{\substack{z_0:T, s_0:T, \\ a_0:T, r_0:T}} \mathcal{P}(z_0:T, s_0:T, a_0:T, r_0:T) \left[ \sum_{t=0}^T \gamma^t r_t \right] \quad (5.1)$$

$$= \lim_{T \rightarrow \infty} \sum_{\substack{z_0:T, s_0:T, \\ a_0:T, r_0:T}} \mathcal{P}(z_0:T, s_0:T, a_0:T) \left[ \sum_{t=0}^T \gamma^t R(z_t, a_t) \right] \quad (5.2)$$

を最大化することを目標とする. ここで,  $\gamma \in [0, 1)$  は割引率と呼ばれる値で,  $x_{s:t}$  は時刻  $s$  から  $t$  までの変数  $x$  の集合  $\{x_\tau | \tau = s, s+1, \dots, t\}$  を表す. また,  $\mathcal{P}$  は経路確率で

$$\mathcal{P}(z_0:T, s_0:T, a_0:T, r_0:T) = P_0(z_0) \prod_{t=1}^T P(z_t|z_{t-1}, a_t) \prod_{t=0}^T \phi(s_t|z_t) \pi(a_t|s_t) P(r_t|z_t, a_t), \quad (5.3)$$

$$\mathcal{P}(z_0:T, s_0:T, a_0:T) = P_0(z_0) \prod_{t=1}^T P(z_t|z_{t-1}, a_t) \prod_{t=0}^T \phi(s_t|z_t) \pi(a_t|s_t) \quad (5.4)$$

である. エージェントは各時刻  $t$  で現在までの観測・行動・報酬の情報  $\{(s_\tau, a_\tau, r_\tau) | \tau = 0, \dots, t\}$  を元に方策  $\pi$  を更新することができる.

観測過程  $\phi$  が決定論的で  $\mathcal{Z}$  から  $\mathcal{S}$  への単射な写像  $\psi$  となっている場合, すなわち  $\phi(s|z) = \delta_{\psi(z)}(s)$  となっている場合はエージェントが観測から状態を完全に同定できるため部分観測ではない通常のマルコフ決定過程となる. ただし,  $\delta_A(B)$  は  $B = A$  のときのみ 1 でそれ以外のときは 0 をとる関数である.  $\psi$  の像を  $\bar{\mathcal{S}}$  とおくと,  $\mathcal{Z}$  の要素と  $\bar{\mathcal{S}}$  の要素は  $\psi$  によって一対一に対応する. すると観測  $s$  を状態のようにみなすことができ, この新しい状態  $s \in \bar{\mathcal{S}}$  について期待報酬関数  $R$  や状態遷移確率  $P$  を以下のように定義できる.

$$R(s, a) := R(\psi^{-1}(s), a), \quad (5.5)$$

$$P(s'|s, a) := P(\psi^{-1}(s') | \psi^{-1}(s), a), \quad (5.6)$$

$$P_0(s) := P_0(\psi^{-1}(s)). \quad (5.7)$$

以下ではこの部分観測マルコフ決定過程だが実質的には部分観測でないマルコフ決定過程を考える。

さらに、本研究では報酬の累積和ではなく各時刻で得られる報酬そのものを最大化する問題を考える。これは割引率  $\gamma = 0$  に相当し、最大化すべき式5.1は

$$\mathbb{E}[r_0] = \sum_{s,a} P_0(s)\pi(a|s)R(s,a) \quad (5.8)$$

となる。また、状態遷移が行動に依存しない場合  $P(s'|s,a) = P(s'|s)$  にも同様の問題に帰着する。実際、状態は単純なマルコフ連鎖としてふるまうため、エルゴード性を仮定すれば唯一の定常分布  $P_\infty(s)$  に収束する。この定常分布のもとでは

$$\mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r_t\right] = \lim_{T \rightarrow \infty} \sum_{s_0:T, a_0:T} \prod_{t'=0}^T P_\infty(s_{t'})\pi(a_{t'}|s_{t'}) \left[\sum_{t=0}^T \gamma^t R(s_t, a_t)\right] \quad (5.9)$$

$$= \sum_{t=0}^{\infty} \gamma^t \sum_{s_t, a_t} P_\infty(s_t)\pi(a_t|s_t)R(s_t, a_t) \quad (5.10)$$

$$= \frac{1}{1-\gamma} \sum_{s,a} P_\infty(s)\pi(a|s)R(s,a) \quad (5.11)$$

となって、式5.8の初期状態分布  $P_0$  を定常分布  $P_\infty$  で置き換えたものと定数倍を除いて一致する。このような各時刻の報酬を最大化する問題は連想強化学習 (associative reinforcement learning) の問題などと呼ばれる [73]。さらに、状態がマルコフ連鎖でなく独立同分布からのサンプリングである場合  $P(s'|s) = P(s')$  は連想バンディット問題 (associative bandit problem) あるいは文脈付きバンディット問題 (contextual bandit problem, bandit problem with side information) などと呼ばれる。

## 5.2 強化学習アルゴリズム

マルコフ決定過程の問題を解くためのアルゴリズムが強化学習である。強化学習は様々な手法があるが、多くの手法は (1) 未知である報酬関数を推測する段階と、(2) 報酬の予測値に基づいて方策を決定する段階の2段階を繰り返すようなアルゴリズムである [21]。アルゴリズム1に全体の流れをまとめた。以下ではまず (2) の方策決定の段階における難しさとその代表的な解決方法を示し、その後 (1) の推定問題の解法を紹介する。

**アルゴリズム 1** 強化学習アルゴリズム

---

```

パラメータ  $\theta$  の初期化  $\theta^{(0)} \leftarrow \theta_0$ 
for  $t = 0$  to  $T$  do
  環境から状態  $z_t$  と観測  $s_t = \psi(z_t)$  を得る
  パラメータ  $\theta^{(t)}$  に依存する方策  $\pi(a_t|s_t)$  に従って行動  $a_t$  を決める
  環境から報酬  $r_t$  を得る
   $(s_t, a_t, r_t)$  を元にパラメータを更新する (式5.15)
end for

```

---

**方策決定**

方策を決める際の難しさは、エージェントは真の期待報酬関数  $R$  ではなくその推定値  $\hat{R}$  の情報しか得られないという点にある。もし期待報酬関数  $R$  が既知の場合、得られる報酬を最大化する方策は

$$\pi(a|s) = \delta_{a^\dagger(s)}(a), \quad a^\dagger(s) = \arg \max_{a \in \mathcal{A}} R(s, a). \quad (5.12)$$

ただし、 $a^\dagger(s)$  は状態  $s$  で得られる報酬が最も高い行動である。本研究では扱わない報酬の累積和を最大化するような問題設定においても期待報酬関数  $R$  や状態遷移確率  $P$  が既知であれば、動的計画法 [19] によって完全に解くことができる。しかし一般の場合にはこのような解は得られない。その難しさを表現したのが**探索と活用のジレンマ**と呼ばれる問題である。たとえば、真の報酬関数  $R$  の代わりに推定値  $\hat{R}$  を決定論的な方策の式5.12に代入した場合、すなわちエージェントが現在までの経験の「活用」のみする場合、実際に選択したことの無い行動でより高い報酬が得られる可能性を無視しているため平均的には得られる報酬は低くなる。一方で、今まで選んだことの無い行動を取り続け（「探索」）、状態行動空間のすべての点における報酬の値を確定させてから式5.12の方策をとる場合も、大きな状態行動空間の探索が終了するまでに膨大な時間がかかってしまうため効率が悪い。よって、探索と活用のバランスをとることが良いアルゴリズムの条件である。

典型的には、 $\varepsilon$ -貪欲方策やソフトマックス方策と呼ばれる確率的な方策が用いられる。ソフトマックス方策 (Boltzmann 探索方策とも呼ばれる) は、期待報酬関数の推定  $\hat{R}$  を用いて

$$\pi(a|s) = \frac{\exp(\beta \hat{R}(s, a))}{\sum_{a'} \exp(\beta \hat{R}(s, a'))} \quad (5.13)$$

と表される。ここで、 $\beta \geq 0$  は探索と活用のバランスを調整するパラメータになっている。実際、 $\beta = 0$  とすると  $\pi$  は一様分布となり探索のみをするような方策になり、 $\beta \rightarrow \infty$  とす

ると、 $\pi$  は式5.12と同様の決定論的方策で、 $R$  の代わりに  $\hat{R}$  を代入したもの、すなわち活用のみをする方策となる。学習の初期では小さな  $\beta$  から始めて、反復回数が増えるにつれて、徐々に大きくしていくというのが一般的な方針であり、事前に決められたスケジュールで反復回数だけに依存させる方法や、状態の訪問回数に依存させるような手法 [74] がよく使われる。

### 報酬推定

次に、報酬の推定について考える。割引率  $\gamma > 0$  の場合は TD 誤差法、その中でも SARSA [75] や Q 学習法 [20] といった手法がよく用いられる。割引率  $\gamma = 0$  のときはこれらのアルゴリズムは一致し、ある種の勾配降下法とみなすことができる。以下ではこの手法を紹介する。

まず報酬の推定値を表現する方法を選ぶ。状態と行動の数が少ない場合は各状態・行動対  $(s, a)$  に対する期待報酬の推定値  $\hat{R}(s, a)$  を表として保持できる。状態と行動の数が多しあるいは無限個の場合には、パラメータ  $\theta$  を持つなんらかの関数近似器  $\hat{R}_\theta(s, a)$  によって推定値を表現する。関数近似器としては、ニューラルネットワークのほか基底関数の線形結合などが用いられる。基底関数を用いる手法では、事前に用意された  $K$  個の関数  $g_k : \bar{\mathcal{S}} \times \mathcal{A} \rightarrow \mathbb{R}$  ( $k = 1, \dots, K$ ) に対してパラメータ  $\theta \in \mathbb{R}^K$  を用いて

$$\hat{R}_\theta(s, a) = \sum_{k=1}^K \theta_k g_k(s, a) \quad (5.14)$$

のように表現する。状態行動対の個数と同じ数のデルタ関数を基底関数に用いる ( $K = |\bar{\mathcal{S}}| |\mathcal{A}|$ ) ことで、表としての推定値の表現も表すことができるため、以下この線形結合による表現のみ扱う。

この推定値は新たな情報を得るとともに更新していくことができる。まず時刻 0 で推定値のパラメータ  $\theta^{(0)}$  を適当に初期化しておく (例えば、 $\theta^{(0)} = 0$ )。時刻  $t$  で状態  $s_t$  で行動  $a_t$  をとったあとに、報酬信号  $r_t$  を得たとすると、パラメータ  $\theta^{(t)}$  のすべての要素  $k = 1, \dots, K$  について以下のように更新する:

$$\theta_k^{(t+1)} = \theta_k^{(t)} + \alpha_t g_k(s_t, a_t) (r_t - \hat{R}_{\theta^{(t)}}(s_t, a_t)). \quad (5.15)$$

ここで  $\alpha_t \in (0, 1]$  は学習率である。これは、時刻  $t$  ごとに異なる (確率変数としての) 目的関数

$$\mathcal{F}_t(\theta) = \frac{1}{2} \left( r_t - \hat{R}_\theta(s_t, a_t) \right)^2 \quad (5.16)$$

の勾配降下法になっている。あるいは、時刻  $t$  に依存しない

$$\mathcal{F}(\theta) = \mathbb{E} \left[ \frac{1}{2} (r - \hat{R}_\theta(s, a))^2 \right] = \mathbb{E} \left[ \frac{1}{2} (R(s, a) - \hat{R}_\theta(s, a))^2 \right] \quad (5.17)$$

の期待値をサンプルで置き換えたものとも考えられる。実際、期待値と微分が交換できるとすれば

$$\frac{\partial \mathcal{F}}{\partial \theta_k}(\theta) = \mathbb{E} \left[ -g_k(s, a)(r - \hat{R}_\theta(s, a)) \right] \quad (5.18)$$

となるので式5.15の更新部分の期待値は  $\mathcal{F}(\theta)$  の勾配と一致する。この点で第4章で紹介した確率的勾配降下法と類似している。ただし、データ  $\{(s_t, a_t, r_t)\}_t$  が方策に依存しており独立同分布からのサンプルとは限らない点で一般化された確率的勾配降下法になっており、通常の収束証明は直接適用できない。表により推定値を表現する場合は割引率  $\gamma \geq 0$  の一般的な設定で最適解へ確率1で収束することが示されている [20], [76]。ただし、ソフトマックス方策でいえば  $\beta$  を十分ゆっくり無限大に大きくしていくことや、学習率を式4.22に従ってゼロに近づけることが必要である。線形結合による表現を用いる場合は、基底関数を行列として並べたときにフルランクとなることを仮定すれば唯一の最適解への確率1での収束が示せる [77]。典型的な収束証明においては、 $t$  が十分大きいところで  $\theta^{(t)}$  の挙動が決定論的な常微分方程式 (勾配系) で記述できることを利用して、勾配系の安定性から収束を示す [78]。

### 5.3 免疫系における状態・観測・行動

Kato ら [23] はヘルパー T 細胞を中心とした免疫系を (部分観測) マルコフ決定過程としてモデル化した。まず環境の状態  $z$  に相当するのが生体内に存在する正常な組織や病原体など免疫系の標的となりうるものである。  $P$  個の状態からなる状態空間を  $\mathcal{Z}$  とする。免疫系はどの病原体に感染しているかなどの情報を直接得ることはできない。第2章で紹介したように、免疫系は自然免疫系の細胞のパターン認識受容体と獲得免疫系の B 細胞および T 細胞受容体を介して生体内を監視している。いまは獲得免疫系の T 細胞に注目することにすれば、観測に対応するものは樹状細胞から抗原提示されるペプチドである。簡単のため抗原とペプチドが一対一対応しており  $N$  種類の抗原 (ペプチド) が存在しうるとする。1つの病原体  $z$  から同時に複数の抗原が生じることを考慮して、 $N$  種類の抗原それぞれの存在の有無を二値のベクトルで表したものを観測  $s \in \mathcal{S} = \{0, 1\}^N$  とする。免疫系は観測を元にして行動、すなわち免疫応答を惹起する。第2章の表2.1では免疫応答は3つのエフェクター機構に分類されていたが、ここでは一般的に  $M$  種類のエフェ



クター細胞が存在してそれらの活性化の有無が行動に対応すると考える. よって行動は  $a \in \mathcal{A} = \{0, 1\}^M$  とする. 以上を表5.3にまとめた.

また, 本章では簡単のため観測過程  $\phi(s|z)$  は決定論的で全単射な写像  $\psi: \mathcal{Z} \rightarrow \bar{\mathcal{S}} \subset \mathcal{S}$  によって与えられるとする. つまり, 全ての標的  $z \in \mathcal{Z}$  は互いに異なる抗原パターン  $s = \phi(z) \in \mathcal{S}$  をもたらす. この仮定のもとでは観測  $s$  から状態  $z$  を常に同定できるため部分観測ではない通常マルコフ決定過程となる. そのため以下では状態  $z$  と観測  $s$  をいづれも状態と呼ぶことにする.

変数名	変数空間	免疫系
状態 $z$	$\mathcal{Z} = \{1, \dots, P\}$	正常な組織や病原体などの標的
観測 $s$	$\mathcal{S} = \{0, 1\}^N$	$N$ 種類の抗原 (ペプチド) の存在の有無
行動 $a$	$\mathcal{A} = \{0, 1\}^M$	$M$ 種類のエフェクター細胞の活性化の有無

表5.1 部分観測マルコフ決定過程としての免疫系

### 方策 $\pi(a|s)$

この観測  $s$  から行動  $a$  への過程 (方策  $\pi(a|s)$ ) を,  $K$  種類の T 細胞クローンによる相互作用として以下のように具体的にモデル化する [23]. まず,  $k$  番目の T 細胞クローンの細胞数を  $n_k$  とおき, そのうち  $\tilde{n}_k$  個の細胞が  $N$  個の抗原の刺激によって活性化するとする. 細胞生物学的なモデル (3.1節) では T 細胞の活性化を単純な一次反応とみなしていた. その場合はペプチドと T 細胞受容体の結合定数に対応するような量  $K_{ik}$  が大きいことによって対応する T 細胞の活性化割合も高くなっていた. しかし近年の研究では, T 細胞の活性化には複雑な機構が存在しており, 必ずしも結合の強さと T 細胞の活性化は正の相関を持たない [79], [80] など, 未だ不明点も多い. よって, ここではペプチドと受容体の結合などの詳細な反応やその動力学を全て無視し, T 細胞の活性化割合が以下で与えられるとする:

$$\frac{\tilde{n}_k}{n_k} = \sigma \left( \sum_{i=1}^N s_i w_{ik} \right) =: h_k(s). \tag{5.19}$$

ここで,  $w_{ik}$  は  $i = 1, \dots, N$  番目の抗原存在することによる  $k = 1, \dots, K$  番目の T 細胞を活性化させる強さで,  $\sigma$  はシグモイド関数  $\sigma(x) = 1/(1 + \exp(-x))$  である. T 細胞が認識できる抗原が少ないと活性化割合は 0 になり, 多いと活性化割合は 1 となる. 活性化した (ナイーブ) T 細胞は Th1 細胞, Th17 細胞などに分化し, サイトカインを放出することによってエフェクター細胞を活性化させる. この過程も多種のサイトカインや細胞内の

複雑な機構によっており [81], 全体として定量的な理解には程遠い. よって, T 細胞の表現型の違いを区別せず, 各 T 細胞クローンがエフェクター細胞に作用して活性化させるような, 単純なモデルを用いる. 具体的には T 細胞の活性化と同様にシグモイド関数を用いて,  $m = 1, \dots, M$  番目のエフェクター細胞の活性化確率を

$$P(a_m = 1 | \tilde{\mathbf{n}}) = \sigma \left( \beta u_{0m} + \beta \sum_{k=1}^K \tilde{n}_k u_{km} \right) \quad (5.20)$$

とする. ここで,  $u_{km}$  は各 T 細胞クローン  $k$  のうち活性化した細胞が  $m$  番目のエフェクター細胞に作用する重みで,  $u_{0m}$  は T 細胞以外の細胞からのサイトカインを受け取ることやエフェクター細胞自身の自発的な活性化の影響を表す.  $\beta$  はエフェクター細胞の全体的な感受性を決める値である. これらをまとめると, 抗原の観測  $s$  で条件付けられた行動  $a$  の確率は

$$\pi(a|s) = \prod_{m=1}^M P(a_m | n_k h_k(s)) \quad (5.21)$$

$$= \prod_{m=1}^M \left[ a_m P(a_m = 1 | n_k h_k(s)) + (1 - a_m) \{1 - P(a_m = 1 | n_k h_k(s))\} \right] \quad (5.22)$$

$$= \prod_{m=1}^M \frac{a_m \exp(\beta u_{0m} + \beta \sum_k n_k h_k(s) u_{km}) + (1 - a_m)}{1 + \exp(\beta u_{0m} + \beta \sum_k n_k h_k(s) u_{km})} \quad (5.23)$$

$$= \frac{\exp(\beta Q_{\mathbf{n}}(s, a))}{\sum_{a' \in \mathcal{A}} \exp(\beta Q_{\mathbf{n}}(s, a'))} \quad (5.24)$$

となり, これが方策に対応する. ただし,  $Q_{\mathbf{n}}$  は

$$Q_{\mathbf{n}}(s, a) := Q_0(s, a) + \sum_k n_k f_k(s, a), \quad (5.25)$$

ただし,

$$f_k(s, a) := h_k(s) \sum_{m=1}^M a_m u_{km}, \quad (5.26)$$

$$Q_0(s, a) := \sum_{m=1}^M a_m u_{0m} \quad (5.27)$$

である. したがって, 免疫系は T 細胞数分布  $\mathbf{n} = \{n_k\}$  をパラメータとして持っており, これを変えることで方策を変えることができると考える. その他の重み  $\{w_{ik}\}$  や  $\{u_{km}\}$  などは固定された値である.

以上をまとめたものを図5.1に示す.

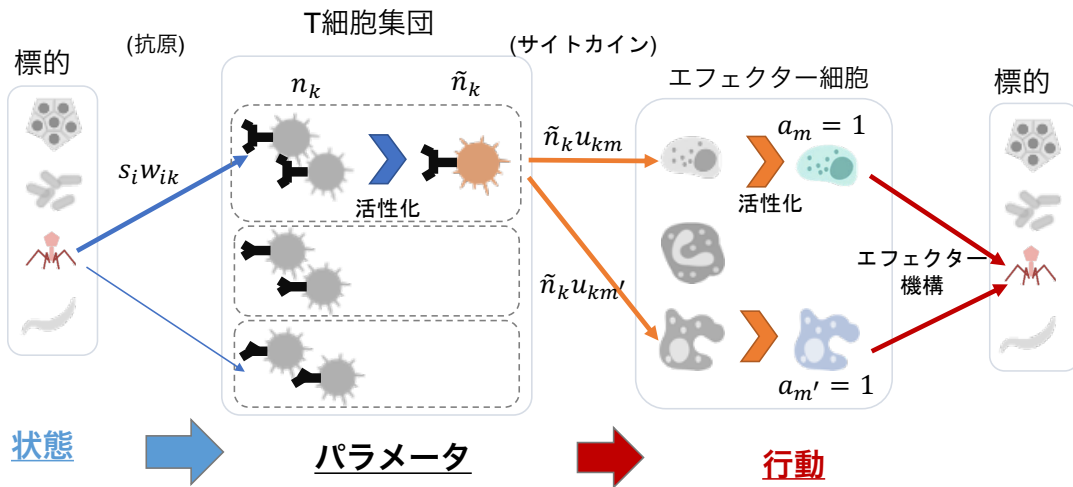


図5.1 免疫系における状態から行動への流れ. 標的・細胞から細胞への矢印は抗原やサイトカインによる情報伝達を示す. T細胞やエフェクター細胞は受け取った信号の大きさによって活性化する.

## 5.4 強化学習としての T 細胞の増殖死滅過程

### 期待報酬関数 $R(s, a)$ の存在

それぞれの標的  $z$  (あるいは対応する抗原分布  $s = \psi(z)$ ) には対応する適切な免疫応答  $a$  が存在することは既に述べた. たとえば, 正常な組織に対しては炎症を起こさず, 切創等により損傷した組織に対しては修復を促し, 侵入した病原体に対してはその種類に応じて適切なエフェクター機構を誘導する必要がある. よって免疫応答の適切さを評価する指標が存在すると考えるのは自然である. これが期待報酬関数  $R(s, a)$  に相当すると考えられる. しかし免疫系において, どのように報酬の情報を得ているのか自明ではない. 本来のマルコフ決定過程であれば期待値が  $R(s, a)$  と一致するような報酬信号  $r_t$  を直接用いる. 免疫応答が適切であればあるほど値が大きくなるような免疫系の変数 (分子の濃度や細胞数など) はあるだろうか. 病原体に対する初期の免疫応答については, 多くの分子や細胞を動員する大きな炎症を引き起こすことが迅速な病原体排除につながると考えられるが, 一方で大きすぎる炎症は正常な組織を傷つけてしまう [8]. よって炎症の大きさが報酬の大きさに対応するとは考えにくい.

### ソフトマックス方策

ここで、報酬の実態をひとまず置いておいて、免疫系において5.2節で紹介したような強化学習アルゴリズムが実装されていると考えてみる。すると、免疫系における方策の式5.24は、一般の強化学習アルゴリズムで用いられるソフトマックス方策の式5.13と同じ形をしていることに気付く。すなわち、エフェクター細胞全体の感受性を決める  $\beta$  は探索と活用のバランスを定めており、関数  $Q_n$  は期待報酬関数の推定  $\hat{R}$  に対応する。また関数  $Q$  の式5.25と関数  $\hat{R}$  の式5.14を比べれば、基底関数  $\{g_k\}$  が免疫系では  $\{f_k\}$  であり、その重みであるパラメータ  $\theta$  が T 細胞数分布  $n$  に対応することがわかる。

### 報酬推定

このパラメータを変化させて期待報酬関数をうまく予測できるできるようになれば、強化学習アルゴリズムとして成立していることになる。パラメータが T 細胞数分布に対応することから、T 細胞の集団動態がパラメータの更新に対応することになる。ここでは、T 細胞の増殖や死滅がこの機能を担っていると考え、胸腺からの流入については次の6章で議論する。よって確率的勾配降下法によるパラメータ更新の式5.15を T 細胞数分布  $n$  について適用すれば、T 細胞がどのように増殖・死滅すれば強化学習に対応するかという条件を与えることになる。ただし、ここで前章のリーマン多様体上の勾配降下法の考え方を組み合わせる。4.3節で同定されたように、T 細胞の増殖・死滅過程は  $f$ -ダイバージェンスを備えたリーマン多様体の上で考えるのが妥当である。よって、時刻  $t$  でクローン  $k$  の細胞数変化は

$$n_k^{(t+1)} = n_k^{(t)} + \alpha_t n_k^{(t)} f_k(s_t, a_t)(r_t - Q_{n^{(t)}}(s_t, a_t)) \quad (5.28)$$

となる。この式は1細胞あたりの正味の増殖率

$$\lambda_k^{(t)} = f_k(s_t, a_t)(r_t - Q_{n^{(t)}}(s_t, a_t)) \quad (5.29)$$

による増殖 ( $\lambda_k^{(t)} > 0$  のとき) あるいは死滅 ( $\lambda_k^{(t)} < 0$  のとき) を表している。この増殖死滅動態を用いた免疫系の強化学習モデルを、**スカラー報酬信号モデル**と呼ぶことにする。このモデルについて表5.4にまとめた。この増殖死滅動態は第3.2節で現象論的なモデルとして導入したものと期待値をとった連続時間極限で一致する。ただし以下で議論するように、このモデルの学習アルゴリズムが実際の免疫系で実装されていると考えるのは難しい。そこで、以降では報酬信号の実態について議論することで免疫学的により妥当なモデルを提案する。

強化学習	強化学習での意味	免疫系	免疫系での解釈
$R(s, a) \in \mathbb{R}$	期待報酬	$R(s, a) \in \mathbb{R}$	免疫応答の適切さの指標
$r \in \mathbb{R}$	(スカラー) 報酬信号	$r \in \mathbb{R}$	-
$\hat{R}_\theta(s, a) \in \mathbb{R}$	期待報酬の推定値	$Q_n(s, a) \in \mathbb{R}$	-
$g_k(s, a) \in \mathbb{R}$	基底関数	$f_k(s, a) \in \mathbb{R}$	クローン $k$ の寄与
$\theta \in \mathbb{R}^K$	パラメータ	$\mathbf{n} \in \mathbb{R}_{\geq 0}^K$	T 細胞数分布
式5.15	勾配降下法による報酬推定	式5.28	T 細胞の増殖死滅動態

表5.2 ]

強化学習アルゴリズムとしての免疫系. [スカラー報酬信号モデルの場合.]

### 報酬の実態と学習の実装

まず、報酬関数が満たすと期待される関数形について考えてみる。免疫系のエフェクター機構はモジュール化されており [25], 独立して動作するのでその評価も独立していると考え、報酬関数はエフェクター細胞の種類ごとの和によって以下のように表せる:

$$R(s, a) = \sum_{m=1}^M [a_m R_m^1(s) + (1 - a_m) R_m^0(s)] \quad (5.30)$$

$$= \sum_{m=1}^M [a_m \Delta R_m(s) + R_m^0(s)]. \quad (5.31)$$

ただし,  $R_m^1: \bar{S} \rightarrow \mathbb{R}$  は状態  $s$  のときに  $m$  番目のエフェクター細胞が活性化していること ( $a_m = 1$ ) に対する評価,  $R_m^0: \bar{S} \rightarrow \mathbb{R}$  は活性化していないこと ( $a_m = 0$ ) に対する評価で,  $\Delta R_m(s) := R_m^1(s) - R_m^0(s)$  である。さらにまとめると

$$R(s, a) = \Delta R(s, a) + R^0(s), \quad (5.32)$$

$$R^0(s) := \sum_{m=1}^M R_m^0(s), \quad (5.33)$$

$$\Delta R(s, a) := \sum_{m=1}^M a_m \Delta R_m(s) \quad (5.34)$$

と表せる。 $R^0(s)$  が行動  $a$  によらない成分を表し,  $\Delta R(s, a)$  が  $a$  への依存性を表している。

次に、シグモイド関数、ひいてはソフトマックス方策は並進不変性をもつことに注意する。つまり、 $s$  には依存するかもしれないが  $a$  によらない任意の関数  $C: \bar{S} \rightarrow \mathbb{R}$  について

$$\frac{\exp(\beta R(s, a))}{\sum_{a'} \exp(\beta R(s, a'))} = \frac{\exp(\beta R(s, a) + C(s))}{\sum_{a'} \exp(\beta R(s, a') + C(s))} \quad (5.35)$$

が成り立つ。従って、分解された報酬関数の式5.32より

$$\frac{\exp(\beta R(s, a))}{\sum_{a'} \exp(\beta R(s, a'))} = \frac{\exp(\beta \Delta R(s, a))}{\sum_{a'} \exp(\beta \Delta R(s, a'))} \quad (5.36)$$

となる。これが  $Q_n$  を用いた方策  $\pi(a|s) \propto \exp(\beta Q_n(s, a))$  と一致すれば良いので、 $Q_n$  は期待報酬関数  $R$  の推定値ではなく、ベースライン  $R^0$  を引いた  $\Delta R$  の推定値となっても良い。よって  $\Delta R(s, a)$  を正規化された報酬関数と呼べば、報酬関数を推定する問題は、正規化された報酬関数を推定する問題と等価である。さらに、正規化された報酬  $\Delta R(s, a)$  が式5.34のようにエフェクター細胞の種類ごとに分解されたことを踏まえると、その各成分  $\Delta R_m(s)$  を推定する問題に帰着できる。

そこで、和としての報酬信号  $R(s, a)$  に対応するスカラー信号ではなく、各成分  $\Delta R_m(s)$  に対応するベクトル信号として報酬が与えられるという問題設定を考える。具体的には、ベクトル報酬信号として  $\mathbf{r} = \{r_m\}_{m=1}^M \in \mathbb{R}^M$  を考え、それぞれの  $r_m$  は平均が  $\Delta R_m(s)$  で分散が有限な確率分布からサンプリングされるとする。以下では表記をわかりやすくするため行動  $a_t$  と報酬  $r_t$  の時刻の添字を右上につけて  $a^{(t)}, r^{(t)}$  のように書くことにすれば、式5.29の増殖率  $\lambda_k^{(t)}$  は以下のように書ける：

$$\lambda_k^{(t)} = h_k(s_t) \sum_{l=1}^M a_l^{(t)} u_{kl} \sum_{m=1}^M a_m^{(t)} (r_m^{(t)} - U_m^{(t)}). \quad (5.37)$$

ただし、 $U_m^{(t)}$  は時刻  $t$  で  $m$  番目のエフェクター細胞へ入力される信号の総量

$$U_m^{(t)} := u_{0m} + \sum_{k=1}^K n_k^{(t)} h_k(s_t) u_{km} \quad (5.38)$$

である。

ここで、異なる種類のエフェクター細胞の活性化という事象の間には相関がなく  $\mathbb{E}[a_l^{(t)} a_m^{(t)}] = 0 (m \neq l)$ 、さらに  $r_m^{(t)} - U_m^{(t)}$  とは無相関だと仮定する。すると、増殖率の式5.37において  $(a_m^{(t)})^2 = 1$ 、 $a_l^{(t)} a_m^{(t)} = 0 (l \neq m)$  としても期待値としては一致して

$$\lambda_k^{(t)} \approx h_k(s_t) \sum_{m=1}^M a_m^{(t)} u_{km} (r_m^{(t)} - U_m^{(t)}) \quad (5.39)$$

となる．この近似された増殖率を用いると T 細胞数の時間変化を以下のように記述できる：

$$n_k^{(t+1)} = n_k^{(t)} + \alpha_t \tilde{n}_k^{(t)} \sum_{m=1}^M a_m^{(t)} u_{km} (r_m^{(t)} - U_m^{(t)}). \quad (5.40)$$

方策の図5.1に示したような順方向の情報の流れ (標的 → T 細胞 → エフェクター細胞 → 標的) とは反対に, T 細胞の増殖死滅動態による学習は逆方向の情報の流れ (標的 → エフェクター細胞 → T 細胞) を示しているといえる．実際, 式5.40は活性化した T 細胞  $\tilde{n}_k$  が活性化したエフェクター細胞  $a_m = 1$  から信号を受け取ることで実現できる．エフェクター細胞  $m$  と T 細胞  $k$  の間のやり取りには順方向と同じ重み  $u_{km}$  が存在し, エフェクター細胞は  $r_m - U_m$  という信号を送っている．ここで,  $U_m$  は順方向伝播におけるエフェクター細胞  $m$  への入力 of 総量であったから, 細胞は既にその情報をもっている．

最後に報酬  $r_m$  は, エフェクター細胞  $m$  の活性化が標的  $z$  にどのような効果をもたらしたかを表現しているもので, センサー細胞が標的  $z$  について得た情報をエフェクター細胞に情報伝達した結果の信号であると考えるのが自然である．ここで, センサー細胞は, 第2章で紹介したパターン認識受容体をもつ貪食細胞のことである．パターン認識受容体は, 侵入してきた病原体や傷害された自己組織の存在を感知して, 獲得免疫系を含む免疫系全体の炎症を開始する役割をもつ．パターン認識受容体には多くの種類があり, たとえば Toll 様受容体 (TLR) と呼ばれる受容体はヒトでは 10 種類同定されており, それぞれが異なる種類の病原体由来物質を認識する [82]．そこで, センサー細胞は標的の数だけ存在し, それぞれの標的  $z_t = 1, \dots, P$  の存在に対応してセンサー細胞の活性化状態  $y_z^{(t)} \in \{0, 1\}$  が決まるとする．ここでは簡単のため現在の標的  $z_t = z$  に対応するセンサー細胞が確実に活性化し  $y_z^{(t)} = 1$ , それ以外は必ず活性化していない  $y_z^{(t)} = 0$  とする．活性化したセンサー細胞はなんらかのサイトカイン分子を放出してエフェクター細胞に情報を伝える．このときの  $z$  番目のセンサー細胞から  $m$  番目のエフェクター細胞への信号の重みを  $b_{zm}$  とすると, 時刻  $t$  の報酬  $r_m^{(t)}$  はそれらの総和として

$$r_m^{(t)} = \sum_{z=1}^P y_z^{(t)} b_{zm} = b_{z_t, m} \quad (5.41)$$

と表せる．よってこの重み  $\{b_{zm}\}$  が期待報酬関数  $R$  の情報を符号化したものだと考えることができる．エフェクター細胞の種類ごとの分解 (式5.30) を仮定した場合には, 明らかに任意の期待報酬関数  $R(z, a)$  をなんらかの重み  $\{b_{zm}\}$  で表現できる．

以上をまとめると, 表5.4および図5.2のようになる．

強化学習	強化学習での意味	免疫系	免疫系での解釈
$\Delta R(s, a) \in \mathbb{R}$	(正規化された) 期待報酬	$\Delta R(s, a)$	免疫応答の適切さの指標
$r \in \mathbb{R}^M$	(ベクトル) 報酬信号	$\{r_m\} \in \mathbb{R}^M$	センサー細胞の信号
$\hat{R}_\theta(s, a) \in \mathbb{R}^M$	期待報酬の推定値	$\{U_m\} \in \mathbb{R}^M$	エフェクター細胞への入力
$g_k(s, a) \in \mathbb{R}$	基底関数	$f_k(s, a) \in \mathbb{R}$	クローン $k$ の寄与
$\theta \in \mathbb{R}^K$	パラメータ	$n \in \mathbb{R}_{\geq 0}^K$	T細胞数分布
式5.15	勾配降下法による報酬推定	式5.40	T細胞の増殖死滅動態

表5.3 ]

強化学習アルゴリズムとしての免疫系. [ベクトル報酬信号モデルの場合.]

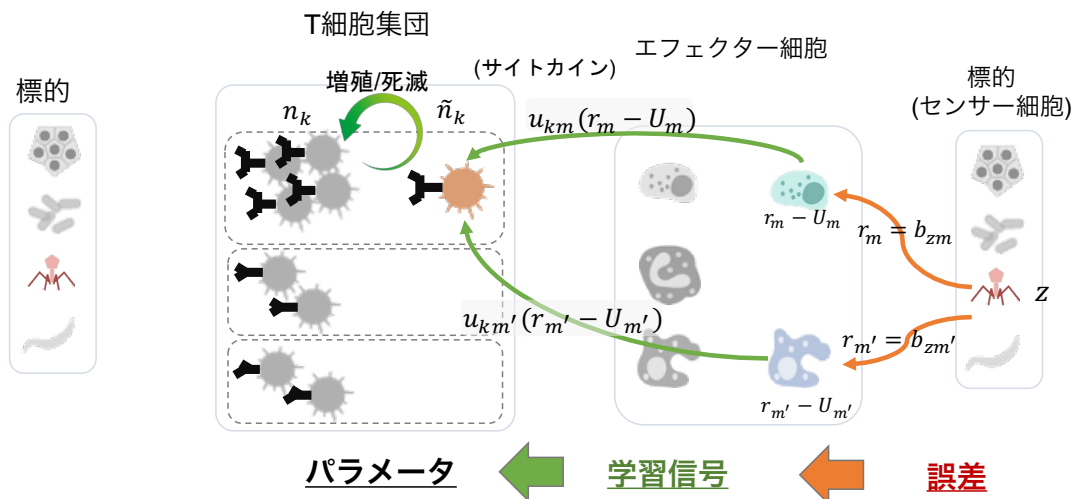


図5.2 免疫系細胞の相互作用による学習の実装. 細胞から細胞への矢印はサイトカインによる情報伝達を示す.

### 5.5 強化学習の数値シミュレーション

前節では免疫系を強化学習アルゴリズムとらえることで、スカラー報酬信号モデルとベクトル報酬信号モデルの2つのモデルを紹介した. スカラー報酬信号モデルは [23] で提案されたものと同じであり, Kato らが数値シミュレーションで確認しているようにうまく学習することが可能である. 証明はできていないが, 確率的勾配降下法の観点からもその



収束は妥当だと考えられる。一方でベクトル報酬信号モデルは、免疫学的な実態を明らかにするために近似を用いて導出したものであり、勾配降下法としての構造が明らかではない。よって、ベクトル報酬信号モデルが実際に正しく学習できるアルゴリズムになっていることを数値シミュレーションによって確かめる。

パラメータは [23] にならって表5.5のように設定した。  $i = 1, \dots, N$ ,  $k = 1, \dots, K$ ,  $m = 1, \dots, M$  について、重み  $w_{ik}$  は平均 0 分散  $2/N$  の正規分布から、重み  $u_{km}$  は平均 0 分散  $2/K$  の正規分布からサンプリングし、簡単のため  $u_{0m} = 0$  とした。

パラメータ名	値
T細胞の種類数 $K$	5,000
標的の種類数 $P$	20
抗原の種類数 $N$	100
エフェクター細胞の種類数 $M$	20
学習率 $\alpha_t$	0.1 で一定
探索活用バランス $\beta_t$	1.0 から 10.0 へ $t$ の線形増加

表5.4 強化学習の数値シミュレーションで用いたパラメータ

報酬  $r_t$  は  $s_t$  と  $a_t$  のもとで決定論的で期待報酬関数  $R$  の値がそのまま得られるとし、期待報酬関数は以下のように定めた。各状態  $s \in \bar{\mathcal{S}}$  に対する最適な行動  $a^\dagger(s)$  を  $\mathcal{A} = \{0, 1\}^M$  上の一様分布に従ってランダムに決定し、期待報酬関数は実際にとった行動  $a$  と  $a^\dagger(s)$  のハミング距離  $\text{ham}$  を用いて

$$R(s, a) = M - \text{ham}(a, a^\dagger(s)) \tag{5.42}$$

$$= \sum_{m=1}^M \left[ a_m \delta_{a_m^\dagger(s)}(1) + (1 - a_m) \delta_{a_m^\dagger(s)}(0) \right] \tag{5.43}$$

と定めた。これは実際の活性化  $a_m$  と最適な活性化  $a_m^\dagger(s)$  が一致しているようなエフェクター細胞の種類数であり、全く一致していなければ最小値の 0、完全に一致していれば最大値の  $M$  をとる。正規化された報酬関数  $\Delta R(s, a)$  の  $m$  番目の成分は

$$\Delta R_m(s) = \delta_{a_m^\dagger(s)}(1) - \delta_{a_m^\dagger(s)}(0) \tag{5.44}$$

$$= \begin{cases} -1 & (a_m^\dagger(s) = 0) \\ 1 & (a_m^\dagger(s) = 1) \end{cases} \tag{5.45}$$

である。

また、各状態  $z = 1, \dots, P$  について対応する観測  $s = \psi(z)$  を  $\mathcal{S} = \{0, 1\}^N$  上の一様分布に従ってランダムに決め、状態遷移確率はハミング距離  $\text{ham}$  を用いて以下のように定めた:

$$P(z'|z, a) = \frac{\text{ham}(a^\dagger(\psi(z')), a)}{\sum_{z''=1}^P \text{ham}(a^\dagger(\psi(z'')), a)} \propto \text{ham}(a^\dagger(\psi(z')), a). \quad (5.46)$$

$z, z' \in \{1, \dots, P\}$  でなく  $s, s' \in \bar{\mathcal{S}}$  を用いれば以下のようなのである:

$$P(s'|s, a) = \frac{\text{ham}(a^\dagger(s'), a)}{\sum_{s'' \in \bar{\mathcal{S}}} \text{ham}(a^\dagger(s''), a)} \propto \text{ham}(a^\dagger(s'), a). \quad (5.47)$$

これは負の報酬  $M - R(s', a)$  に比例する形になっている。すなわち、現在とった行動  $a$  によって対処可能な (同じ行動をとっても報酬が高いような) 状態は次の時刻では出現確率が低く、対処が難しい状態に遷移する確率が高い。

細胞数  $n_k$  の初期値はすべての  $k$  で 1 として一様分布からランダムに決定した状態  $z_0$  からシミュレーションを開始した。  $T = 10^5$  回の反復を 1 試行として、異なる乱数シードで 10 回試行を行った。スカラー報酬信号モデルとベクトル報酬信号モデルの両方の結果を図5.3に示す。各時刻で得られるの報酬と二乗誤差をプロットしてある。

どちらのモデルも  $t$  が大きくなるにつれて得られる報酬が高くなり、誤差はほぼゼロに収束している。よって、これらのモデルは正しく強化学習するアルゴリズムになっているといえる。また、同じパラメータ設定にも関わらず、ベクトル報酬信号モデルのほうがやや収束が速いことがわかる。これは、スカラー報酬信号モデルに比べて各時刻で得られる情報が増えているため、問題として易しくなっているからだと考えられる。

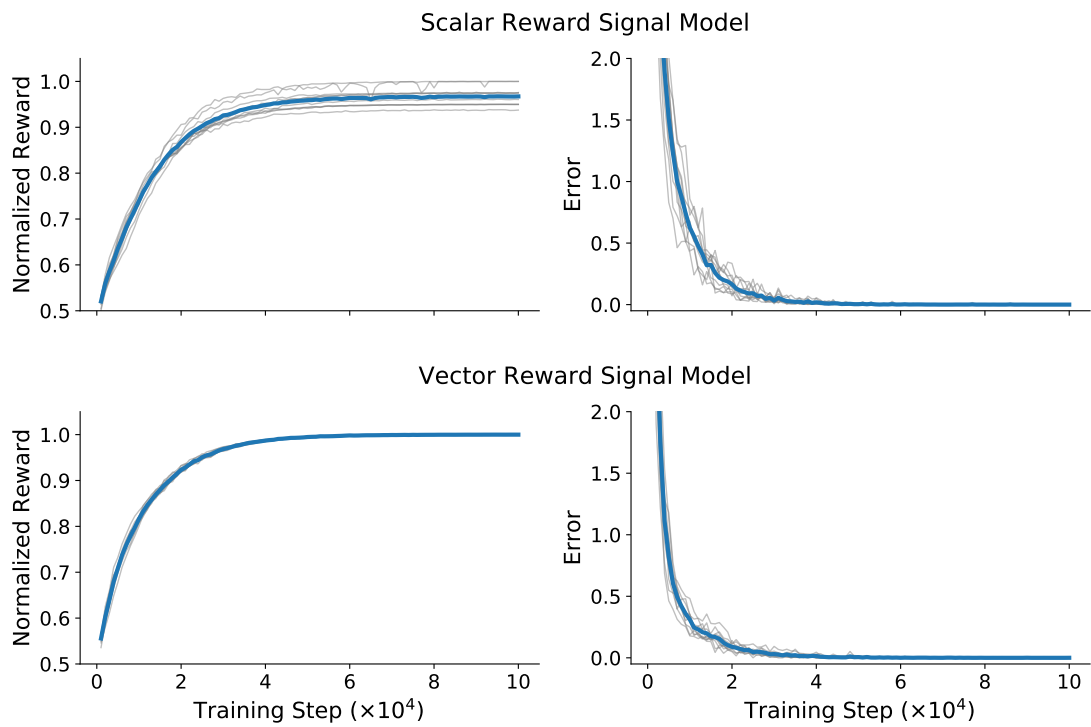


図5.3 免疫系強化学習モデルの数値シミュレーション結果. 上段はスカラー報酬信号モデル, 下段はベクトル報酬信号モデルの結果である. 左列は各時刻で得られた報酬  $r_t$  の値を最大値  $M$  で割った値の時系列を示している. 右列は各時刻の二乗誤差  $(r_t - Q_{\mathbf{n}^{(t)}}(s_t, a_t))^2$  の時系列である. 細線は各試行の時系列で青の太線は平均値の時系列を表す.

## 第6章

# 胸腺からの T 細胞の流入とその意義

前章においては、末梢 (胸腺以外の全身循環系) における T 細胞の増殖死滅過程が強化学習アルゴリズムにおける期待期待関数の推定という機能をもつという考え方を示した。本章では T 細胞が胸腺から末梢へ流入する過程に注目する。流入過程のリーマン多様体構造は第4章で既に明らかにしているため、その機能あるいは目的関数をまずは同定する。これによって流入の T 細胞集団動態が導かれる。本章の後半 2 節ではこのモデルが現実の免疫系と整合する振る舞いをすることを示す。特に胸腺の退縮という現象にまつわる免疫学の未解決問題について学習理論の観点からの 1 つの解答を与える。

### 6.1 転移学習と胸腺の機能

時刻  $t$  における胸腺からの T 細胞の流入は、単位時間あたりの  $k$  番目のクローンの流入細胞数  $\sigma_k^{(t)}$  として表現できる。 $\sigma_k^{(t)}$  については免疫学的に分かっていることは少ないが、 $k$  によらない総数  $\sigma^{(t)} = \sum_{k=1}^K \sigma_k^{(t)}$  については個体の年齢やストレスによって変動することが知られている。この現象は胸腺の退縮とよばれ、本章の最後で再び議論する。一方で各  $k$  の割合の分布  $\sigma_k^{(t)}/\sigma^{(t)}$  については胸腺選択という仕組みによってバイアスされていることが知られている。しかしそのバイアスされた分布が年齢や個体の状況によって変化するという報告は、我々が調べた範囲ではなかった。従って、胸腺からの流入という現象が前章5で示したような強化学習による報酬の推定という機能を持つとは考えにくい。なぜなら強化学習はエージェントが試行錯誤する経験をもとに行動を変えていくための枠組みである一方で、胸腺からの流入  $\sigma_k^{(t)}$  は (少なくともその割合は) 個体の経験には依存し

ないからである。

しかし、胸腺からどのような細胞数分布で流入させるかは個体の生存に重要な役割を果たす。実際、胸腺選択を担う胸腺の上皮細胞で発現している AIRE と呼ばれる遺伝子をノックアウトすると、流入する分布が変化して、致命的な自己免疫疾患を引き起こすことが知られている [83]。よって、流入分布は進化的な時間スケールにおいて自然選択によって変化してきたと考えられる。

強化学習でいえば、エージェントは過去のエージェントが経験して学習がうまく行ったかどうかという情報を間接的に受け取っていることに相当する。これは、強化学習に限らず機械学習において一般的に利用される **マルチタスク学習** や **転移学習** と呼ばれる手法である。解くべき問題 (タスク) が複数存在  $\{T_m\}_{m=0}^M$  しているときに、全てのタスクでの性能を同時に向上させることを目指すのがマルチタスク学習で、特定のタスク  $T_0$  での性能を向上させるために他のタスク  $\{T_m\}_{m=1}^M$  の情報を事前知識として用いるのが転移学習である。各タスクで最適化対象となるパラメータを  $\theta^{[m]} (m = 0, \dots, M)$  とすると、異なるタスク間でパラメータの情報を共有することが転移学習あるいはマルチタスク学習における典型的な手法の一つである [84]。具体的には、ハードなパラメータ共有とソフトなパラメータ共有という 2 つの手法がある。ハードなパラメータ共有では、全てのタスクで同じパラメータを用いる。ソフトなパラメータ共有では、タスク間でパラメータが類似するように正則化を加える [85]。たとえば、0 番目のタスクの本来の学習の目的関数を  $\mathcal{F}(\theta^{[0]})$  とすると、それに正則化項として異なるタスクのパラメータ間の L2 距離の和を加えて

$$\min_{\theta^{[0]}} \mathcal{F}(\theta^{[0]}) + \gamma \sum_{m=1}^M \|\theta^{[0]} - \theta^{[m]}\|^2 \quad (6.1)$$

という最適化問題に変形させる。ここで、 $\gamma$  は正則化の強さを決めるようなハイパーパラメータである。

転移学習やマルチタスク学習の手法は多くの応用分野で高い性能を示している。上手くいく理由として、Ruder らは暗黙的にデータ拡張になっていること (注目するタスクから本来得られるデータに加えて、異なるタスクのデータを利用できる)、複数のタスクで上手くいくよう学習する正則化に対応すること、などを挙げている [85]。後者については、多くのタスクで性能が高い学習器は新規のタスクについても高い確率で性能が高くなることが理論的に示されている [86]。

## 6.2 転移学習としての流入動態の導出

増殖死滅が報酬学習という本来の目的関数  $\mathcal{F}(\theta^{[0]})$  の最小化に対応すると考えれば、流入が正則化項  $\sum_m \|\theta^{[0]} - \theta^{[m]}\|^2$  の最小化に対応すると考えられる。以降ではこの目的関数を用いて流入のリーマン多様体上で勾配降下法を適用することで流入動態を導出する。ここで、簡単のためタスクは2つだけ存在して、1つ目のタスクが現在の個体と環境のもとでの報酬最大化、2つ目のタスクが仮想的な祖先の個体と環境のもとでの報酬最大化とする。現在の個体の最適化したいパラメータはT細胞分布  $\mathbf{n} = \{n_k\}_{k=1}^K \in \mathbb{R}_{\geq 0}^K$  であり、祖先のパラメータを  $\mathbf{m} = \{m_k\}_{k=1}^K \in \mathbb{R}_{\geq 0}^K$  とする。祖先のパラメータがどのようなべきかは次節で議論するが、進化を経て個体の遺伝子に埋め込まれていると考える。よって、目的関数  $\mathcal{F}_i(\mathbf{n})$  は

$$\mathcal{F}_i(\mathbf{n}) = \frac{1}{2} \|\mathbf{n} - \mathbf{m}\|^2 = \frac{1}{2} \sum_{k=1}^K (n_k - m_k)^2 \quad (6.2)$$

である。第4.3節で議論したように、流入動態は  $g$ -ダイバージェンスを用いることで勾配降下法として以下のように求まる:

$$n_k^{(t+1)} = (1 - \gamma_t)n_k^{(t)} + \gamma_t m_k. \quad (6.3)$$

ここで、 $\gamma_t \geq 0$  は学習率である。この式は、1細胞単位時間あたり  $\gamma_t$  の速度で死亡し、各クローン  $k$  は単位時間あたり  $\gamma_t m_k$  細胞ずつ流入することを示している。すなわち、流入分布  $\sigma_k = \gamma_t m_k$  であり、 $k$  依存的な流入割合は実質的に祖先のパラメータ  $\{m_k\}$  と一致している。増殖死滅と流入を同時に実装すると、式5.28と式6.3を組み合わせると

$$n_k^{(t+1)} = (1 + \alpha_t \lambda_k^{(t)} - \gamma_t)n_k^{(t)} + \gamma_t m_k \quad (6.4)$$

となる。

次に、このモデルが報酬学習と転移学習の両方を実現していることを数値シミュレーションによって確認する。ここでは、理論的に扱いやすいスカラー報酬信号モデルを採用して増殖率を式5.29のように設定した。シミュレーションするにあたって用いた主なパラメータは増殖のみのときと同様(表5.5)である。簡単のため、状態遷移は  $P = 20$  個の状態の一様分布からの独立なサンプリングとする。流入させる  $\mathbf{m}$  としては、まず同じ状況設定で流入せずにシミュレーションを行い、終了時刻での細胞数分布  $\mathbf{n}^{(T)}$  を用いた。ソフトマックス方策の探索と活用のバランスを決める  $\beta$  を表5.5のように時間とともに増加(アニ

ーリング) させるか 10.0 で一定にするかの 2 通りと, 流入の学習率  $\gamma_t$  を 0 とするか  $10^{-4}$  で一定とするかの 2 通りで, 計 4 通りの実験条件で  $T = 10^5$  の試行をそれぞれ異なる乱数シードで 10 回行った. 流入分布  $m$  を決める事前学習においては  $\beta$  はアニーリングし,  $\gamma_t = 0$  である. シミュレーション結果を図 6.1 に示す.  $\beta$  をアニーリングして流入しない場合が事前学習に相当するが,  $\beta$  をアニーリングして事前学習結果を流入させるほうが素早く報酬が大きく, 誤差が小さくなっている. さらに,  $\beta$  が一定の場合を見てみると, 転移学習しない場合には探索が十分にできず報酬があまり大きくならない一方で, 転移学習する場合には  $t = 0.2T$  で既に報酬を最大化し, 誤差がほぼゼロに収束している. 転移学習する場合に,  $\beta$  をアニーリングするより一定にした方が速く収束しているのは, 転移学習によって  $Q$  は十分真の報酬関数を近似しているにもかかわらず,  $\beta$  をアニーリングする場合は学習初期での  $Q$  を信用せずに探索してしまうからだと考えられる. このシミュレーションでは, 解きたいタスクと転移させるタスクが同じため, 当然ではあるが実際に転移学習がうまく働いていることが示された.

さらに, 解きたいタスク (現在の環境) と転移させるタスク (先祖の環境) が異なる場合にも流入による転移学習が有効であることを確かめる. 先程のシミュレーションでは事前学習でも評価するときも  $P = 20$  個の状態  $S_{\text{pre}}$  を等確率で遷移していた. ここでは, 事前学習は同じ  $S_{\text{pre}}$  を遷移する環境だが, 評価する際の環境として新規に 10 個の状態  $S_{\text{new}}$  を加えた  $P = 30$  個の状態集合  $S_{\text{pre}} \cup S_{\text{new}}$  を等確率で遷移するとした. 結果を図 6.2 に示す. 青色 (same) は図 6.1 に示したものと同一である. 橙色 (general) は事前学習していない状態を含む環境での結果で, 各条件で same よりも報酬は低く, 誤差は大きくなっている. しかし, general で流入のある場合とない場合を比べると,  $\beta$  を一定にした場合は流入したほうが明らかに高い報酬を得られている. しかし誤差については  $t = 0.2T$  までは流入の有無に関わらず同じように下がるが,  $t = 0.2T$  以降流入がある場合には誤差が下がる. また  $\beta$  をアニーリングする場合も報酬の上昇は速いが, 最終的な値は低くなっている.

これらの結果は, 各状態  $s$  における報酬の大きさは  $Q(s, a)$  が大きいような  $a$  において  $R(s, a)$  によって決まると考えると理解できる.  $\beta$  一定のときは, 探索をしないので  $Q(s, a)$  が大きいような  $a$  は学習初期に決まっている. 流入がある場合には少なくとも事前学習した状態  $S_{\text{pre}}$  において  $R(s, \cdot)$  と  $Q(s, \cdot)$  は類似しているため,  $S_{\text{pre}}$  では高い報酬を得続けられる. 流入がない場合には全ての状態  $s$  でたまたま  $Q(s, a)$  が大きかった行動  $a$  しか取り続けられないため, そのような  $(s, a)$  で学習が進んで  $Q(s, a)$  と  $R(s, a)$  が近づいても報酬はあまり大きくならない.  $\beta$  をアニーリングする場合は探索をして多くの  $(s, a)$  について  $Q(s, a)$  が  $R(s, a)$  に近づくため, 学習後期には各状態  $s$  で  $Q(s, a)$  が大き

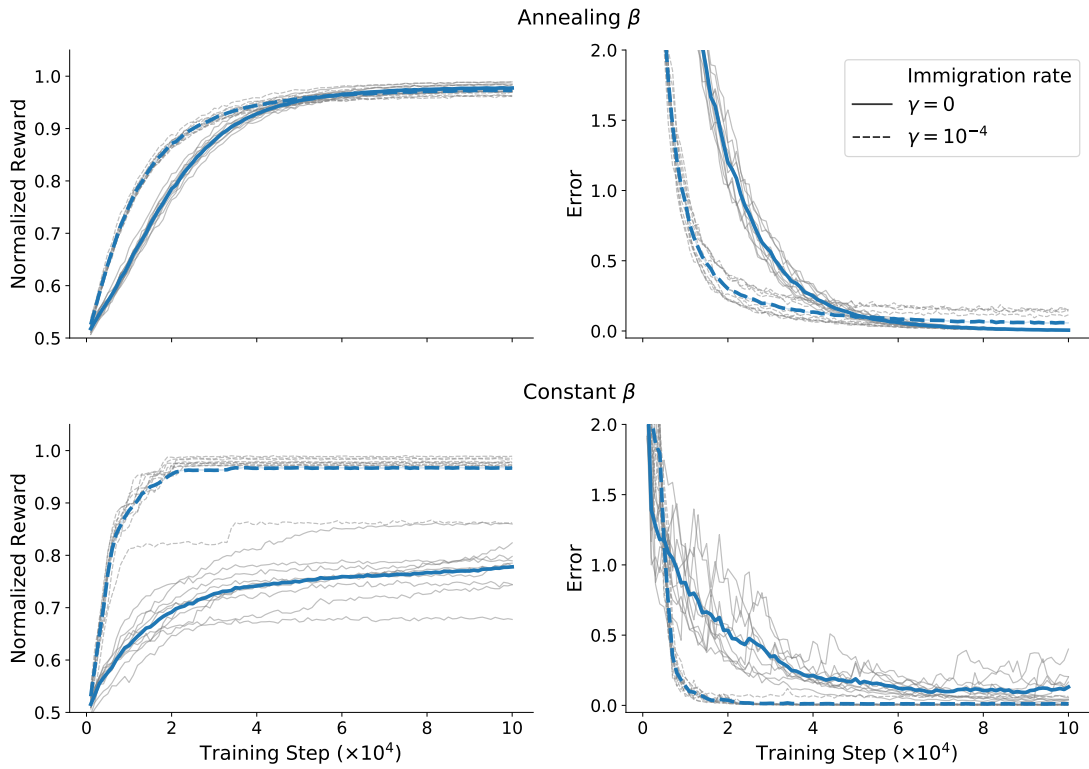


図6.1 免疫系強化学習モデルの数値シミュレーション結果. 上段は  $\beta$  をアニーリングした場合, 下段は  $\beta$  を一定にした場合の結果である. 左右の列は図5.3と同様で正規化した報酬と誤差の時系列を表す. 実線は  $\gamma_t = 0$  で転移学習しない場合, 破線は  $\gamma_t = 10^{-4}$  で転移学習する場合である. 細線は各試行の時系列で青の太線は平均値の時系列を表す.

いような  $a$  では  $R(s, a)$  も大きい. これが流入がないときの場合に相当する.

では  $\beta$  をアニーリングして流入する場合になぜ報酬が低いのだろうか. これは,  $Q$  を  $R$  と近づけようとする強化学習と,  $n$  を  $m$  と近づけようとする転移学習との間で, ある種の競合が生じているからだと考えると理解できる. 実際,  $\beta$  によらず流入がない場合は誤差がゼロ付近に収束しているのに対し, 流入があると誤差は大きい. たとえば,  $\beta$  を一定にした場合は学習中期以降は同じような  $(s, a)$  を遷移する定常状態に至るが, 流入があると誤差はそれ以降下がらない.  $\beta$  をアニーリングする場合には学習中期以降も異なる  $(s, a)$  に遷移するため誤差は下がり続けるが, 流入に依る転移学習と増殖による強化学習の競合によって誤差は大きい. したがって,  $Q(s, a)$  と  $R(s, a)$  の誤差によって, この場合には報酬を大きくする事ができない.



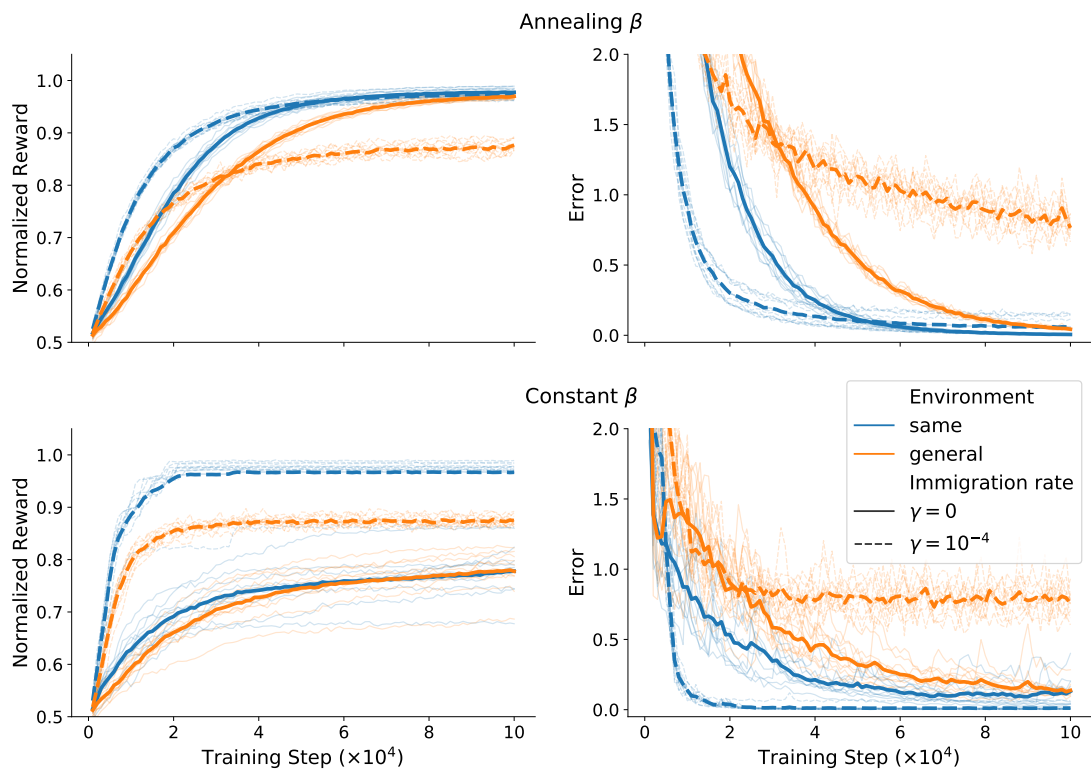


図6.2 免疫系強化学習モデルの数値シミュレーション結果. 図の読み方は図6.1と同様で, 色は環境の違いを表す. 青色 (same) は事前の学習と全く同じ  $P = 20$  個の状態を等確率で遷移する環境で, 橙色 (general) は事前の学習と同じ 20 個の状態に加えて別の 10 個の状態等確率で遷移する環境である.

## 6.3 流入分布の影響と胸腺選択

### 背景

前節のシミュレーションのように、現在の経験する環境と先祖の経験した環境が同じとは限らない。例えば、紀元前から人類を苦しめてきた天然痘ウイルスは、18 世紀の Edward Jenner による世界初のワクチンの発明もあって、1980 年には自然界から消滅したと撲滅宣言がなされた [87]。一方で近年パンデミックを起こしている SARS-CoV-2 は 2019 年に始めて確認されたコロナウイルスである。よって、祖先の経験した環境に適した T 細胞数分布をそのまま現在の環境に転移させることは望ましいとは限らず、どのような分布が良いのかは自明でない。

一方、実際の免疫系の流入分布は第 2 章でも述べたように、以下の 2 つの機構によってつくられる。

- (1) T 細胞受容体遺伝子の再構成
- (2) 胸腺上皮細胞などの相互作用による正と負の選択

これらによって形成される流入分布が免疫応答の成否に重要な役割を果たすとすれば、これらは自然選択によって最適化されている可能性が高い。(1) では、ランダムな再構成によって生成可能な T 細胞受容体遺伝子のアミノ酸配列が全て等確率で出現するのではなく、配列の頻度分布には偏りがあることが分かっている [88]。(2) では、親和性選択説によれば生体内に存在する自己抗原への結合親和性が強すぎず弱すぎない T 細胞受容体をもつ T 細胞のみが生き残り、残りは死滅する [34], [89]。さらに、個体の中ではなく進化的な時間スケールで

- (0) 再構成前の V(D)J 遺伝子の配列

も最適化されているはずである。この配列は (2) の正の選択において自身の MHC と結合できる T 細胞受容体が多く生成されるように自然選択されていると考えられている [90]。

理論上のあらゆる T 細胞受容体が等確率で出現する一様分布から、(0)~(2) の過程を経て T 細胞受容体の頻度分布が変形されるとみなし、これらの変形はどのような環境に対する事前知識を反映しているかを考える。ヒトがしばしば感染する病原体に結合できる T 細胞が多く流入するよう進化しているのではないかとする仮説 [91] もあったが、代表的なウイルスに特異的な T 細胞受容体は (1) における生成確率および (2) における選択確率のいずれにおいてもその他の T 細胞受容体と差がないことが示されている [88], [92]。

一方、(0) と (2), 少なくとも (2) は自己の生体組織という環境への事前知識を反映していると解釈できるため、結果としての流入分布は自己環境の事前知識を反映しているといえる。実際、2017 年の Asaf Madi らの研究 [93] では、T 細胞受容体のアミノ酸配列の類似度を利用したネットワーク解析により、どのような配列が頻度が高くネットワークの中でクラスターをつくるのかが調べられた。その結果、病原体など非自己に対する T 細胞受容体に対するクラスターよりも自己や自己に類似した抗原 (自己免疫疾患や腫瘍免疫・同種移植) に対する T 細胞受容体が多くクラスターを作っており、多くの T 細胞受容体はそのクラスターに属していることがヒトとマウスの双方で確認された。これらのクラスターには種の中であるいは種を超えて共通して出現する受容体が多く含まれた。また、そのクラスターは胸腺選択がない場合にはなくなることが示された [93]。

このように、胸腺からの流入分布は自己に関連した環境への事前知識を反映していると考えられるが、なぜそのような流入分布になっているのかは明確にわかっていない。負の選択で自己に強く反応する T 細胞が除去されることは自己免疫疾患を予防するという観点から理解できる。一方で、正の選択ではある程度自己に反応するものが選ばれるため、一見矛盾している。正の選択ではペプチドではなく MHC と T 細胞受容体の結合がテストされていると以前は考えられていたため、正の選択の役割はペプチド MHC 複合体と相互作用ができて細胞内のシグナル伝達が正常である T 細胞を選定することだと考えられていた。しかし、正の選択においてもペプチドに対する特異的結合が寄与していること、正の選択において提示される自己抗原のペプチドが特別な機構によって生成されていること、などが明らかになっているため、正の選択の役割はそれ以上のものがあるとされている [94]。そこで正の選択の役割として異なる仮説が提案されている。1 つ目は、末梢の T 細胞が活性化はしない程度に自己抗原を認識することが、T 細胞の恒常性維持や外来抗原への鋭敏な応答に役立つとする説である [95]–[97]。2 つ目の仮説は、自己抗原への応答の強さと外来抗原への応答の強さは正の相関があることに注目する説である。つまり、自己に対して選択することで病原体に対しての応答を良くすることができると想定している [94], [98]。

### 目的・手法

胸腺の正の選択を含む流入分布の形成機構が自己に対する事前知識を積極的に活用している点について、既存の説とは矛盾しないが別の観点からの説を提案する。そのため、過去と現在で環境が異なることを考慮に入れたシミュレーションを行い、どのような流入分布  $\{m_k\}$  を用いるのが良いのかを調べる。流入分布が自然選択によって最適化されていると仮定すると、学習モデルの元では最も学習がうまくいくような流入分布が、現存する免疫系が採用する流入分布である。

まずシミュレーション結果の評価指標を示す. これまでは学習の成功度合いを, 報酬  $r_t$  の大小によって主に評価してきた. これは前章で議論したように免疫応答の適切さを直接測るものである. 一方以降では報酬の推定誤差の大きさ

$$L_t = (r_t - Q_{\mathbf{n}^{(t)}}(s_t, a_t))^2 \quad (6.5)$$

を用いる. 誤差がゼロであれば報酬は最大化され, 誤差が大きいとでたらめな行動をしているのと同じなので報酬は小さくなる. 実際, 前節のシミュレーション結果 (図6.2) をみても,  $\beta$  が小さければ誤差が小さければ報酬は大きくなることが確認できる. ただし, 誤差が有意に小さくなくても報酬が有意に大きくなるが多々ある. これは経験した状態行動対での誤差が小さくなくても適切に探索するような方策を用いなければ報酬は大きくならないからである. 報酬の大きさではなく誤差の小ささによって評価することは, 期待報酬関数の推定という機能のみに注目することに相当し, その機能についての性能の差がよく見えるようになる. 誤差を小さくするアルゴリズムと適切に探索する方策を組み合わせれば, 報酬も大きくなることが期待される.

次に行ったシミュレーションの設定を示す. 上記の誤差  $L_t$  は勾配系の目的関数 (式5.17)

$$\mathcal{F}_p(\mathbf{n}) = \mathbb{E} \left[ \frac{1}{2} (r_t - Q_{\mathbf{n}}(s, a))^2 \right] \quad (6.6)$$

の期待値の中身と同じ形をしている. ここで, 環境の状態遷移を独立同分布  $P(s)$  からのサンプリング, 方策のパラメータ  $\beta$  は0つまり  $\pi(a|s) = 1/|\mathcal{A}|$  とする. すると, 全ての  $t$  で状態行動対  $(s_t, a_t)$  は独立同分布  $P(s, a) = P(s)/|\mathcal{A}|$  からのサンプリングとなるため, 誤差  $L_t$  の期待値は目的関数  $\mathcal{F}_p(\mathbf{n}^{(t)})$  の2倍と一致し, 勾配系としての性質が明確になる. 状態遷移確率に相当する  $P(s)$  あるいは  $P(z)$  については, 状態集合全体  $\bar{\mathcal{S}}$  を高頻度に遭遇する状態集合  $\mathcal{S}_0$  と低頻度に遭遇する  $E$  個の状態集合  $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_E$ , 遭遇しない状態集合  $\mathcal{S}_{E+1}$  へ分割して定義する. 簡単のため全ての状態集合  $\mathcal{S}_i$  は等しく  $P_0$  個の状態から構成されるとする. 確率は

$$P(s) = \begin{cases} \mu p & (s \in \mathcal{S}_0) \\ p & (s \in \cup_{i=1}^E \mathcal{S}_i) \\ 0 & (s \in \mathcal{S}_{E+1}) \end{cases} \quad (6.7)$$

と定義する.  $\mu$  は  $\mathcal{S}_0$  に属する状態の相対的な遭遇頻度のようなもので,  $\mu = 1$  のとき  $\mathcal{S}_{E+1}$  を除く全ての状態は等確率で出現する. 今回の実験では  $\mu \in \{10, 100, 1000\}$ ,  $P_0 = 10$ ,  $E = 4$ , 流入率は  $\gamma = 10^{-6}$  で一定とした.

流入分布としては, あらゆるものを試すことは難しいため代表的なものを持ちいた. まず転移学習ではなくでたらめな流入分布として, 何も流入しない  $\text{none}(m_k = 0(k =$

$1, \dots, K$ ) と一様分布  $\text{uniform}(m_k = 1(k = 1, \dots, K))$  を用いた. また, 既に示したように必ず良い性能を発揮する流入分布として, 誤差を評価する試行と同一の状態遷移確率のもとで事前に学習させた後のパラメータ `same` を用いた. それ以外に,  $i = 0, 1, \dots, 5$  について  $S_i$  のみを遷移する環境で学習させたものを用いる 0 から 5 や,  $S_0 \cup S_1$  を等確率で遷移する環境で学習した 0+1 などを用いた. 事前の学習も評価する際の学習も  $T = 10^5$  回の反復を 1 試行として, 各条件において異なる乱数シードで 5 回行った.

## 結果

シミュレーション結果の誤差二乗和の時間平均  $\hat{\mathcal{L}} = \sum_t L_t$  について表6.3に示す. 大抵の場合で想定通りに `none` と `uniform` はその他の条件よりも誤差が大きく, `same` は誤差が最も小さいことが確認できる. 相対頻度  $\mu$  が高くなるほど, どの流入分布でも誤差は小さくなるが流入分布によってその挙動は異なる. 特に, 最も頻繁に遭遇する  $S_0$  を事前に学習した 0 は,  $\mu$  の増加にともなう誤差の減少が著しい.

この点を明示するため, 特徴的な流入分布のみ抜き出して正規化した誤差を計算して図6.3を作成した. ここで, `uniform` の誤差が最大値 1.0 で `same` の誤差が最小値 0.0 となるように線形変換により正規化を行った. 図の中の流入分布の名前は `frequent` が 0, `rare` が 1, `not appear` が 5 と同じである. 図より, `frequent` の正規化した誤差は  $\mu$  を大きくすると減少し,  $\mu = 1000$  においてはほぼゼロ, すなわち真の状態遷移確率にて事前学習したものと同等の性能を発揮している.

図6.3ではいずれかの状態集合  $S_i$  のみで事前学習したものについて比較した. 次に  $S_0$  を含む複数の状態集合で事前学習した場合の結果を図6.4に示す. 図より,  $\mu = 10$  の場合は実際に経験する状態をより多く事前学習したほうが誤差が小さくなることが分かる. 特にこのとき 0+1+2+3+4 は `same` よりも誤差が小さい. 一方  $\mu$  が 100 以上の場合,  $S_0$  以外の状態集合を事前学習するほうが誤差は大きくなる.

その他, 実際には遭遇しない状態集合  $S_5$  を事前学習させることは性能の悪化をもたらす. 実際, 図6.3でも `not appear` すなわち 5 は  $\mu = 10$  においては最も誤差が大きくなっている. また表6.3より実際に経験するすべての状態を事前学習する 0+1+2+3+4 に  $S_5$  を加えた 0+1+2+3+4+5 は誤差が大きくなることも分かる.

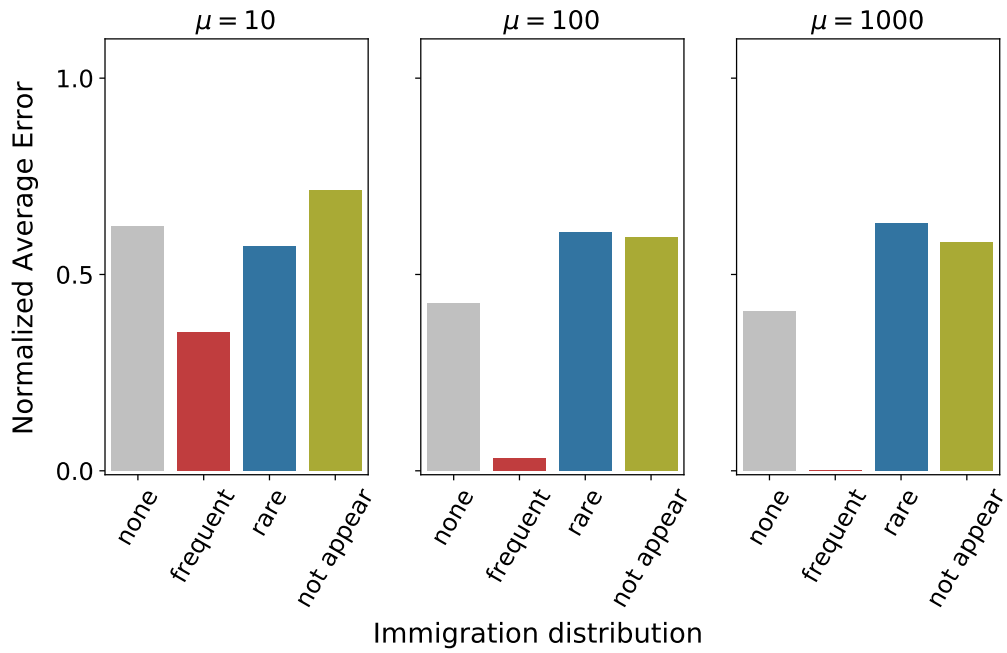


図6.3 胸腺からの流入分布  $m$  と正規化した報酬予測誤差の関係 (1). **frequent** は最も頻繁に遭遇する  $S_0$  で事前学習したもの (0 と同じ), **rare** は低頻度で遭遇する  $S_1$  で事前学習したもの (1 と同じ), **not appear** は実際には遭遇しない  $S_5$  で事前学習したもの (5 と同じ) である.

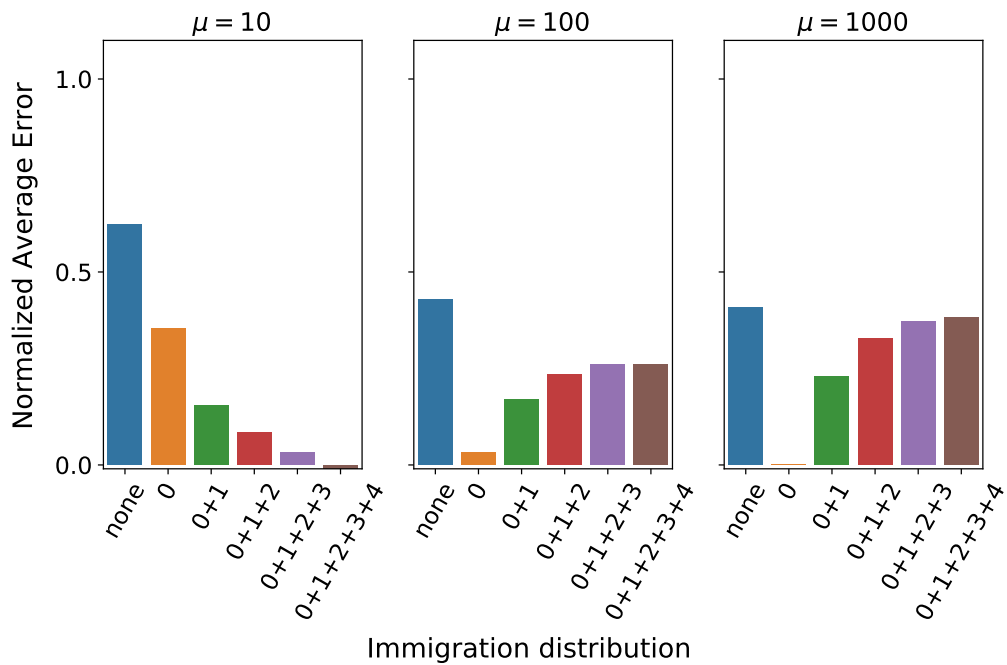


図6.4 胸腺からの流入分布  $m$  と正規化した報酬予測誤差の関係 (2).

流入分布 $m$	$\mu = 10$	$\mu = 100$	$\mu = 1000$
none	7.4052 ± 1.1393	6.3127 ± 1.1283	6.0929 ± 1.1535
uniform	7.4283 ± 1.1338	6.3380 ± 1.1214	6.1181 ± 1.1466
same	7.3670 ± 1.1439	<b>6.2938</b> ± 1.1263	<b>6.0755</b> ± 1.1509
0 (frequent)	7.3886 ± 1.1361	6.2952 ± 1.1257	6.0756 ± 1.1509
1 (rare)	7.4020 ± 1.1400	6.3206 ± 1.1253	6.1024 ± 1.1501
2	7.4032 ± 1.1398	6.3197 ± 1.1259	6.1017 ± 1.1505
3	7.4018 ± 1.1396	6.3181 ± 1.1260	6.0999 ± 1.1507
4	7.3976 ± 1.1413	6.3170 ± 1.1263	6.0995 ± 1.1508
5 (not appear)	7.4107 ± 1.1372	6.3201 ± 1.1253	6.1003 ± 1.1505
0+1	7.3765 ± 1.1417	6.3013 ± 1.1252	6.0854 ± 1.1494
0+1+2	7.3721 ± 1.1442	6.3042 ± 1.1256	6.0895 ± 1.1493
0+1+2+3	7.3690 ± 1.1461	6.3054 ± 1.1263	6.0913 ± 1.1498
0+1+2+3+4	<b>7.3650</b> ± 1.1485	6.3053 ± 1.1272	6.0918 ± 1.1505
0+1+2+3+4+5	7.3729 ± 1.1461	6.3070 ± 1.1270	6.0924 ± 1.1506
0+5	7.3923 ± 1.1364	6.3029 ± 1.1246	6.0842 ± 1.1496

表6.1 異なる相対頻度  $\mu$  と異なる流入分布  $m$  を用いたときの報酬予測誤差の平均と標準偏差. 各  $\mu$  について, 最も小さい平均値を太字で示した. 図6.3で用いる流入分布の名前も括弧に示してある.

## 考察

結果から、流入分布  $\mathbf{m}$  を学習した環境と、誤差を評価する現在の環境が近いほど転移学習としての真価が発揮されて誤差が小さくなる傾向が見られる。このことは、以下のよう理論的にも説明できる。確率的勾配降下法の収束証明でよく用いられるように [78],  $(s_t, a_t)$  について期待値をとって連続時間極限をとった以下のような決定論的な力学系について考える。  $\alpha \rightarrow 0$  で、  $\gamma$  は  $\gamma_0 := \gamma/\alpha$  が定数になるように  $\alpha$  と同じように 0 へ近づけると、

$$\frac{dn_k}{dt} = n_k \sum_{x \in \mathcal{S} \times \mathcal{A}} P(x) f_k(x) (R(x) - Q_{\mathbf{n}}(x)) + \gamma_0 (m_k - n_k) \quad (6.8)$$

$$= n_k [B - A\mathbf{n}]_k + \gamma_0 (m_k - n_k) \quad (6.9)$$

となる。ここで、簡単のため状態行動対を  $x \in \bar{\mathcal{S}} \times \mathcal{A}$  と表し、  $x$  の定常分布を  $P(x)$  とした上で、  $K$  次元ベクトル  $B$  と  $K \times K$  行列  $A$  を以下のような要素をもつものとして定義した。

$$[B]_k = \sum_{x \in \bar{\mathcal{S}} \times \mathcal{A}} P(x) f_k(x) (R(x) - Q_0(x)), \quad (6.10)$$

$$[A]_{kl} = \sum_{x \in \bar{\mathcal{S}} \times \mathcal{A}} P(x) f_k(x) f_l(x). \quad (6.11)$$

これは現象論的モデルの式3.24とほぼ等価である。定常状態  $\frac{dn_k}{dt} = 0 (k = 1, \dots, K)$  において、

$$n_k^2 [B - A\mathbf{n}]_k^2 = \gamma_0 (n_k - m_k)^2 \quad (6.12)$$

が成立する。左辺は報酬予測誤差の大きさと関係しており、右辺は流入分布と実際の分布の誤差である。  $\gamma_0 = 0$  であれば報酬予測誤差に関係する  $B - A\mathbf{n}$  をゼロにできるが、  $\gamma_0 > 0$  の場合は  $\mathbf{m}$  が  $B - A\mathbf{m} = 0$  を満たさない限りは  $B - A\mathbf{n}$  はゼロにならない。これは強化学習と転移学習の間で競合が発生しているようにみえるという前節の結果を説明している。また、流入分布  $\mathbf{m}$  が  $B - A\mathbf{m}$  を小さくする場合は定常分布  $\mathbf{n}$  でも  $B - A\mathbf{n}$  が小さくなる点で、今回の実験結果と整合している。

このシミュレーションにおいて、出現頻度の高い  $S_0$  は自己抗原に相当すると考えられる。なぜなら、病原体やがんなど異常が発生する頻度は稀であり、通常時は自己組織由来の正常な抗原を常に免疫細胞は認識しているからである。T細胞が自己抗原を認識していることは実験で確かめられており、自己抗原を認識することがT細胞が生存し続けるために必要な信号であるとされている [96]。



シミュレーション結果から、高頻度に出現する自己抗原について事前に学習したものを流入させることが現在の環境においても高い性能を発揮することが示された。現在の環境で自己抗原の相対出現頻度  $\mu$  が高い場合は、自己環境のみを事前に学習する方がよくなり、現在の環境と同一の環境で事前に学習する理想的な場合とほぼ同等の性能を発揮する。また、自己環境以外も同時に事前学習することは、 $\mu$  が高いときには悪影響しかない。 $\mu$  が低いときにも現在は遭遇しない環境を事前に学習してしまうことによる悪影響がある。これらのシミュレーション結果は、胸腺選択が自己抗原についての偏りをもたらしており、ヒトがよく経験する病原体に対する偏りが無いという実験事実と整合している。すなわち、胸腺選択が自己抗原についての選択である理由の 1 つとして、病原体などの他の抗原と比べて、自己由来の抗原を非常に頻繁に経験するという点があることが示唆される。

本研究では  $S_0$  に含まれる抗原間の違いについては特に考慮しなかった。これらの抗原のシミュレーションでの実態は、ランダムに割り当てられたバイナリ列だが、ランダムに初期化された重み  $w$  や  $u$  などの実現値によっては特定の抗原のバイナリ列が学習されやすく、別のバイナリ列は学習されにくいということが考えられる。実際、胸腺選択で提示される自己抗原はあらゆる自己抗原ではなくその一部である。提示できる自己抗原の数に制限があるとすれば、どのように提示するものを選ぶかが重要になる。これは、有限の訓練データから訓練されていない入力へ予測を汎化させることに相当する。2020 年の Wortel らの数理モデルによる研究 [99] では特定の自己抗原を胸腺で提示することで提示されなかった自己抗原と外来抗原の識別性能が上がる現象が確認されている。我々のモデルにおいても同様の現象が確認できると期待される。

## 6.4 流入率の影響と胸腺の退縮

### 背景

次にクローン型  $k$  によらない流入全体の多さを決める  $\gamma_t$  について考える。免疫学的には  $\gamma_t$  は年齢  $t$  とともに減少することが知られており、**胸腺の退縮** (thymic involution, thymic atrophy) と呼ばれる。感染や妊娠・ストレスによって一時的に退縮する減少も知られているが、ここでは加齢に伴う退縮に注目する。この現象は免疫老化 (immunosenescence) の一例であり、一見すると本来果たすべき胸腺の機能が損なわれる過程に思える。しかし以下で述べるように退縮自体がなんらかの機能をもっていて、退縮することが生存に有利であるとも考えられる。

まず胸腺からの流入細胞数の変化は図6.5のように主に3つの段階で説明できる。

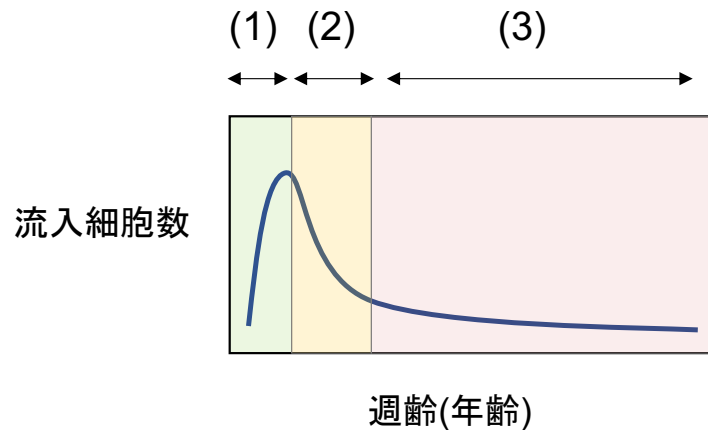


図6.5 胸腺からの流入細胞数の加齢による変化の概念図

- (1) 生後直後, 幼児期の急激な増加
- (2) 小児期の初期に減少に転じる
- (3) 青年～老齢期では減少の傾きは緩やかになる

Hale らは胸腺での T 細胞受容体の構成時に発現する遺伝子 RAG2 と同時に緑色蛍光タンパク質 (GFP) が発現するようなマウスを作製し, 末梢の免疫系の臓器である脾臓から T 細胞を抽出して GFP の蛍光強度が強い T 細胞数を数えた [100]. これによって胸腺から 2,3 週間以内に末梢に流入した T 細胞が推定できる. 図6.5は Hale らによる結果 [100] の図 1D を参考に作成したものである. 段階 (1) の急速な流入の後, およそ生後 6 週で段階 (2) に移行した. 段階 (3) で流入細胞数の減少は緩やかになり, 生後 100 週でも流入が確認できた. 胸腺に存在する T 細胞数を数えた場合にも同様の傾向が確認されている [101]. ヒトでは直接的に細胞数を数える事はできないが, 突然死した遺体の胸腺を取り出し, T 細胞を含む領域の体積を調べると, 10 歳より前に退縮がはじまること, 青年期以降では退縮が緩やかになることが観察されている [102].

生体内の他の臓器と比べて年齢依存的な変化が大きいため, 古くから注目されていた胸腺の退縮だが, なぜ退縮するかについては至近的な要因と究極的な要因の双方で分からないことが多い. 至近的な要因, つまり分子や細胞のどのような機構によって退縮するのかについては, 以前は性ホルモンと結びつけて考えられていたが否定的な証拠が多く, 近年は脂肪細胞の蓄積が直接的な要因だと考えられている [103].

究極的な要因とは, 胸腺の退縮がどのような意味で生存に有利であるかということである. 上述の (1~3) の段階ごとに考えると, 生誕時には末梢には T 細胞が存在しないため,

最低限必要な数を供給するという意味で(1)生後直後の高い流入数は理解できる[104]。また、(3)の後期で高齢になると流入数がほぼゼロに近づいてしまうことはいわゆる個体の老化によって正常な機能を保てなくなっているからだと考えられる。実際高齢者にはワクチンの効果が低く、感染症に罹りやすい[105]ことから、この段階が進化的に説明できるとは考えにくい。一方で(2)の急激に減少する過程は、正常な機能が果たせなくなっていく過程ではなく何らかの積極的な理由によって生存に有利なため減少する過程だと考えられている。なぜなら、胸腺をもつ多くの脊椎動物で同様の胸腺の退縮が観察されていること[106]や、胸腺の退縮に多くの複雑な分子的機構が備わっていること[103]が挙げられる。

では胸腺の退縮にどのような進化的意義があるのだろうか。1つ目の仮説は、生体の有限なエネルギーを効率的に配分した結果だとする説である。胸腺選択の過程でT細胞は90%以上が死滅するため、胸腺は非常に多くのエネルギーを必要とするといえる。胸腺の役目は十分なT細胞数を供給した時点である程度完了しており、小児期に免疫系以外の生体組織を正常に発達させることにエネルギーを割り当てるために胸腺は退縮するとされる[104], [107]。この説ではT細胞の多様性ではなく総数に注目しているといえる。

2つ目の仮説ではT細胞の総数を固定して考えたときに、認識できる抗原の数を増やすことと、特定の抗原により強力に応答することのバランスによって流入率が決まっていると考える[108]。胸腺からの流入は前者を促進し、末梢での増殖は後者を促進する。よってどちらを優先するかによって、どの程度胸腺から流入させるかが決まるとされる。生後直後に十分な多様性を供給した後は、特定の抗原に対する応答を準備することで適切なバランスになるため、増殖の効果を優先させるために流入は減らされる[108]。この説では流入の効果と増殖の効果は互いに打ち消し合うことが暗に仮定されている。

どちらの仮説も流入が減少することは説明できているが、図6.5に示したような減少の動態を完全には説明できない。まず(2)小児期での急激な流入の減少について考える。1つ目の仮説では有限のエネルギー配分を考えていたため、免疫系以外の組織の発達を促進するためには胸腺のエネルギー消費を減らさざるを得ないので、急激な減少と矛盾しない。一方2つ目の仮説では、暗に仮定されてように流入と増殖の間で有限の資源を取り合っているとは言い切れない。そのため、特定の抗原に対する応答を準備するために増殖を促進することは、流入を急激に減らすという結論を導かない。

また、いずれの仮説でも(3)青年期以降にも流入が続いていることを説明できていない。特に1つ目の仮説に基づけば、胸腺の役目は小児期で完了しているため完全に機能を停止させるほうが自然である。実際、青年期以降の流入は生存に有利である可能性が高い。ヒトでは先天的な心臓の疾患で手術時に胸腺を取り除くことがあり、その術後の状況を調べた研究がある。胸腺の摘出は十分な数が供給された(1)より後に行われたと考えられてい

る [109]. この手術が施されるようになってから長い時間が経っていないので, 青年期や老齢期での観察は十分できていないが, 22 歳程度の若い青年を対象にワクチン接種前後の免疫系のいくつかの指標が調べられた. その結果によると, 明らかな免疫系の異常は見られないが, ダニ媒介性脳炎のワクチンに対する応答が同年齢の健常者と比べて明らかに小さく [110], 高齢者と同様の傾向が見られた. 一方マウスでは青年期に胸腺摘出手術を行い, いくつかの病原体への感染実験を行った研究がある [111]–[113]. 胸腺の摘出されなかったマウスと比べて, 胸腺を摘出したマウスでは病原体 (リステリア・モノサイトゲネス (*Listeria monocytogenes*) およびサルモネラ菌 (*Salmonella enterica*)) への免疫応答が弱く, 生存率が有意に下がることが示されている.

### 目的・手法

以上のように胸腺の退縮の進化的な意義を完全に説明できる理論は存在しない. そこで, 本研究ではまず免疫系の学習モデルの数値シミュレーションによってどのような流入率  $\gamma_t$  が良いのかを調べる. その結果が学習理論によって明瞭に解釈できることを示し, 実際の免疫系の退縮と比較を行う.

報酬関数や各種パラメータは前節と同様である. 状態遷移確率はパラメータ  $\mu$  を持つ式6.7の独立同分布を用い,  $\mathcal{S}_0$  内を等確率で遷移する環境で  $T = 10^5$  まで学習させた後の細胞数分布を  $m$  として流入させた. 流入率  $\gamma_t$  は本来時刻  $t$  に依存する量だが, ここでは時間に依存しない  $\gamma$  を用いた. 事前に学習した環境の相対的遭遇頻度  $\mu$ , T細胞の種類数  $K$ , 学習率  $\alpha$  をそれぞれ変えたときに誤差  $L_t$  の経験的平均値が最も小さくなるような  $\gamma$  の値を探る. そのために以下の表6.4のような設定で3つのシミュレーション実験を行った. いずれの実験でも流入率  $\gamma$  は0および  $10^{-7}$  から  $10^{-5}$  の間を対数スケールで20等分した値のそれぞれの場合で調べた.  $T = 10^5$  までの試行を各条件の異なる乱数シードで5回行った.

	相対頻度 $\mu$	T細胞種類数 $K$	学習率 $\alpha$
実験1	{0.05, 0.10, ..., 5.0}	5,000	0.10
実験2	5.0	{1000, 2000, ..., 10000}	0.10
実験3	5.0	5,000	{0.01, 0.02, ..., 0.10}

表6.2 最適な流入率  $\gamma$  を調べる数値シミュレーションで用いたパラメータ

## 結果

実験 1 の結果を図6.6と6.7に、実験 2 の結果を図6.8と6.9に、実験 3 の結果を図6.10と6.11に示す。各実験の 1 つ目の図 (図6.6, 6.8, 6.10) は流入率について平均をとって注目するパラメータ ( $\mu$ ,  $K$  or  $\alpha$ ) 依存的な報酬予測誤差の平均値をプロットしたものである。黒の破線 (Total) は  $t = 0$  から  $T$  まで全区間での平均値で、(1)~(5) は平均を取る区間を一部に制限したものである。例えば (2) は  $t = 0.2T$  から  $0.4T$  の区間で平均をとっている。各実験の 2 つ目の図 (図6.7, 6.9, 6.11) は、横軸流入率  $\gamma$  と縦軸注目するパラメータ ( $\mu$ ,  $K$  or  $\alpha$ ) として色で報酬予測誤差の平均値をプロットしたものである。ただし、見やすいように流入率について平均をとった値 (1 つ目の図) を引いた値をプロットした。また緑のひし形は各パラメータ ( $\mu$ ,  $K$  or  $\alpha$ ) の値において平均報酬予測誤差が最も小さくなった流入率  $\gamma$  の点を示している。6 つある図のうち、左上の図 (Total) は全区間の平均値、それ以外 ((1)~(6)) は部分的な区間の平均値である。

各パラメータ ( $\mu$ ,  $K$  or  $\alpha$ ) の大きさが報酬予測誤差の意味で平均的に性能にどのような影響を与えるのかは各実験の 1 つ目のグラフの黒い破線を見るとわかる。図6.6より、 $\mu = 1$  付近で性能が悪化し  $\mu = 0$  や  $\mu = 10$  では性能が向上している。これは状態の確率分布の乱雑さ (エントロピー) が  $\mu = 1$  で最大になり、 $\mu$  が 0 や大きくなると小さくなることに対応していると考えられる。つまり経験する状態が乱雑であればあるほど、学習が困難になる。図6.9より T 細胞は種類が多ければ多いほど性能が向上し、図6.11より学習率には最適な値がある、というような傾向がみられる。また、平均をとる時間区間が後期になるほど学習が進んで誤差が小さくなっていく様子もわかる。

次に図6.7, 6.9, 6.11で全区間の平均値 (Total) についての最適な流入率についてみると、流入がまったくない  $\gamma = 0$  や実験した中で最大の  $\gamma = 10^{-5}$  が最適になる場合もあるが、中間の値が最適になる場合も多いことがわかる。また、多くの場合で初期 (1) や (2) では高い流入率  $\gamma = 10^{-5}$  が最適になるが後期 (5) には流入しない  $\gamma = 0$  が最適になっている。この傾向は図6.7では  $\mu = 0$  以外の全ての場合でみられ、図6.9では  $K$  が 3000 以上の場合、図6.11では  $\alpha$  が 0.05 以上の場合でみられる。

## 考察

まず全区間の平均報酬予測誤差が最小になっている  $\gamma$  の結果から、流入がない  $\gamma = 0$  が最適でないことが多々あることは注目に値する。T 細胞数分布の初期値として全クローンが存在する  $n_k^{(0)} = 1 (k = 1, \dots, K)$  ため、図6.5で (1) として示した最低限の T 細胞を供給する段階は終わった (2) の段階からシミュレーションが開始しているとみなせる。ただ

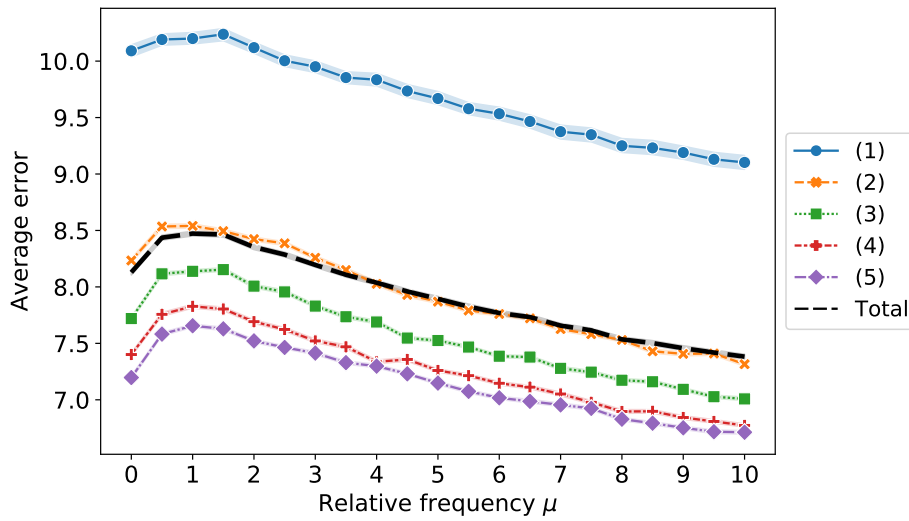


図6.6 相対頻度  $\mu$  と平均報酬予測誤差の関係 (実験 1). 異なる流入率と異なる乱数シードの結果についての平均値をプロットし, 95% 信頼区間が網掛けで示されている. 黒の破線 (Total) は全時間区間での平均値で, (1)~(5) は平均を取る区間を一部に制限したもの.

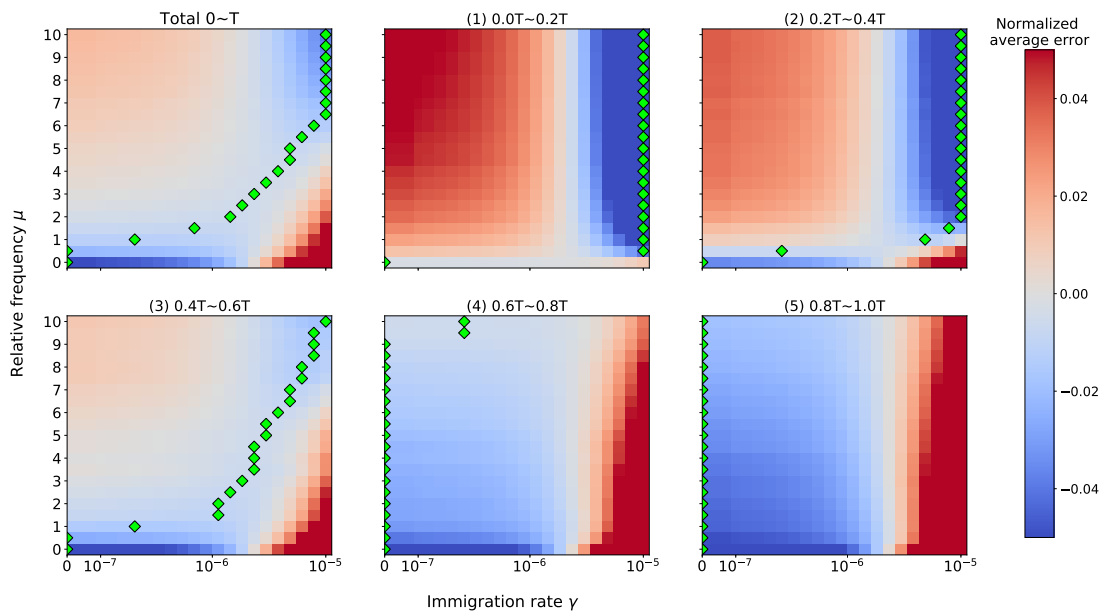


図6.7 相対頻度  $\mu$  と胸腺からの流入速度  $\gamma$  の組に対する正規化した平均報酬予測誤差の関係 (実験 1). 異なる乱数シードの結果の平均値が色で示されている. 緑のひし形は各  $\mu$  の値において平均報酬予測誤差が最も小さくなった流入率  $\gamma$  の点を示している. 6 つある図は図6.6と同様に平均をとる区間が異なる.

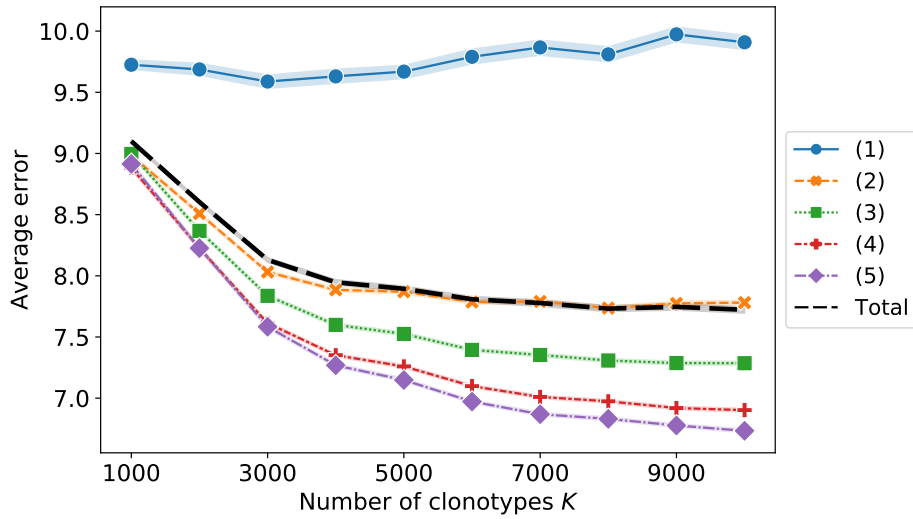


図6.8 T 細胞種類数  $K$  と平均報酬予測誤差の関係 (実験 2). 異なる流入率と異なる乱数シードの結果についての平均値をプロットし, 95% 信頼区間が網掛けで示されている. 黒の破線 (Total) は全時間区間での平均値で, (1)~(5) は平均を取る区間を一部に制限したもの.

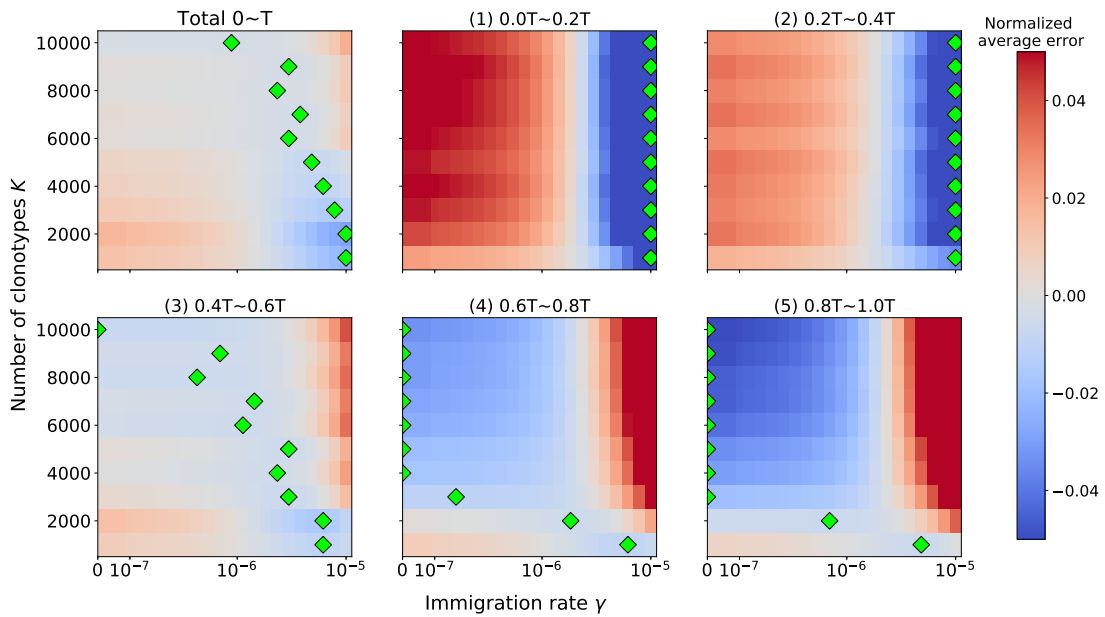


図6.9 T 細胞種類数  $K$  と胸腺からの流入速度  $\gamma$  の組に対する正規化した平均報酬予測誤差の関係 (実験 2). 異なる乱数シードの結果の平均値が色で示されている. 緑のひし形は各  $K$  の値において平均報酬予測誤差が最も小さくなった流入率  $\gamma$  の点を示している. 6 つある図は図6.8と同様に平均をとる区間が異なる.

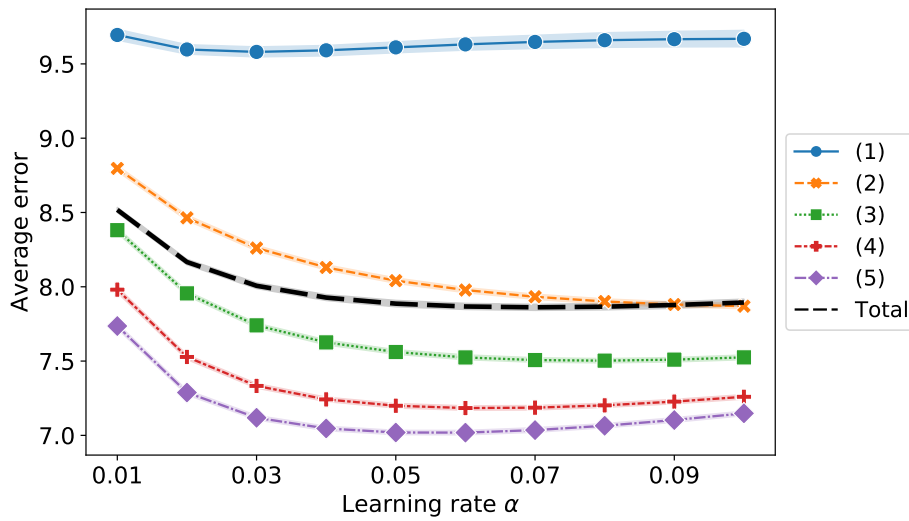


図6.10 学習率  $\alpha$  と平均報酬予測誤差の関係 (実験 3). 異なる流入率と異なる乱数シードの結果についての平均値をプロットし, 95% 信頼区間が網掛けで示されている. 黒の破線 (Total) は全時間区間での平均値で, (1)~(5) は平均を取る区間を一部に制限したものの.

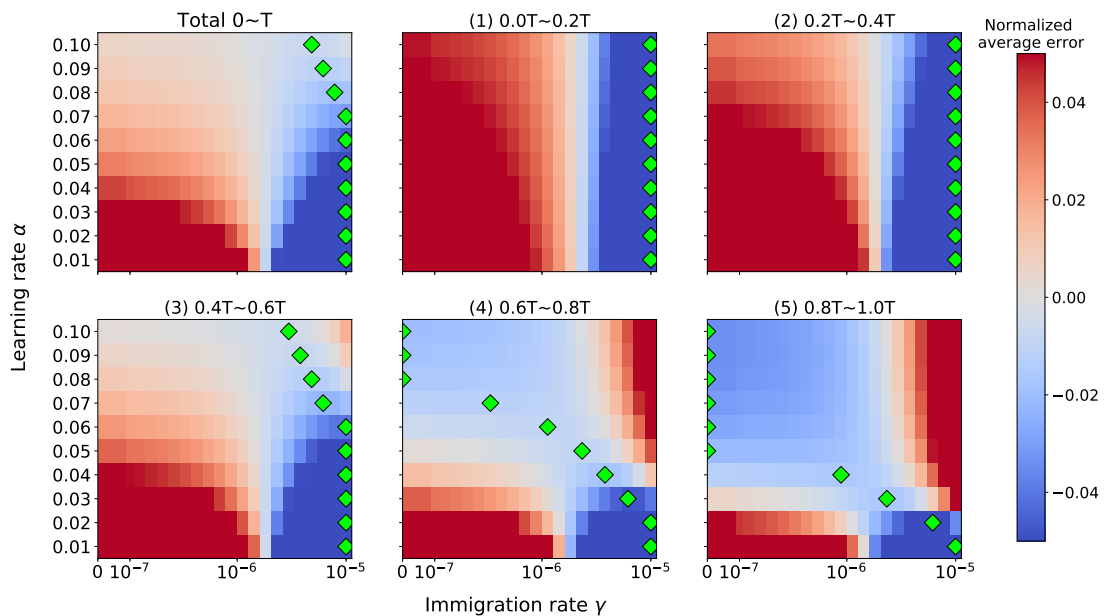


図6.11 学習率  $\alpha$  と胸腺からの流入速度  $\gamma$  の組に対する正規化した平均報酬予測誤差の関係 (実験 3). 異なる乱数シードの結果の平均値が色で示されている. 緑のひし形は各  $\alpha$  の値において平均報酬予測誤差が最も小さくなった流入率  $\gamma$  の点を示している. 6 つある図は図6.10と同様に平均をとる区間が異なる.



し、今回の実験では一定の流入率  $\gamma$  を用いているため、定常的な (3) の青年期以降に相当する流入率が最適値として求まるような状況設定になっていると考えられる。つまり先行研究の仮説では青年期以降の流入の意義を説明できていなかったが、今回の結果は流入を説明できる可能性がある。

最適な流入率がゼロでない正の値になることについては、機械学習で頻繁に議論されるバイアスとバリエーションのトレードオフという考え方によって理解できる。ある時刻  $t$  での報酬予測誤差の期待値

$$\mathcal{L} = \mathbb{E}_{(s,a),\mathbf{n}} \left[ (R(s,a) - Q_{\mathbf{n}}(s,a))^2 \right] \quad (6.13)$$

について考える。ここで、 $(s,a)$  は時刻  $t$  での状態行動対で確率変数である。また、 $\mathbf{n}$  もそれまでの経験  $\{(s_{\tau}, a_{\tau}, r_{\tau}) \mid \tau = 0, 1, \dots, t-1\}$  に依存した確率変数である。上式の期待値  $\mathbb{E}_{(s,a),\mathbf{n}}$  はこの2つの確率変数についてとったものである。以下では簡単のため、 $(s,a)$  と  $\mathbf{n}$  は独立であるとする。独立でなくても同様の分解は成り立つ。まず表記を簡単にするため状態行動対を  $x \in \bar{S} \times \mathcal{A}$  と表して  $x$  ごとに期待値を分解すると

$$\mathcal{L} = \mathbb{E}_x [\mathcal{L}(x)] \quad (6.14)$$

となる。ここで  $\mathcal{L}(x)$  は

$$\mathcal{L}(x) = \mathbb{E}_{\mathbf{n}} \left[ (R(x) - Q_{\mathbf{n}}(x))^2 \right] \quad (6.15)$$

である。これは更に以下のように式変形できる。

$$\mathcal{L}(x) = \mathbb{E}_{\mathbf{n}} \left[ (R(x) - \mathbb{E}_{\mathbf{n}}[Q_{\mathbf{n}}(x)] + \mathbb{E}_{\mathbf{n}}[Q_{\mathbf{n}}(x)] - Q_{\mathbf{n}}(x))^2 \right] \quad (6.16)$$

$$\begin{aligned} &= (R(x) - \mathbb{E}_{\mathbf{n}}[Q_{\mathbf{n}}(x)])^2 + \mathbb{E}_{\mathbf{n}} \left[ (Q_{\mathbf{n}}(x) - \mathbb{E}_{\mathbf{n}}[Q_{\mathbf{n}}(x)])^2 \right] \\ &\quad - 2(R(x) - \mathbb{E}_{\mathbf{n}}[Q_{\mathbf{n}}(x)]) \underbrace{\mathbb{E}_{\mathbf{n}} [Q_{\mathbf{n}}(x) - \mathbb{E}_{\mathbf{n}}[Q_{\mathbf{n}}(x)]]}_{=0} \end{aligned} \quad (6.17)$$

$$= (R(x) - \mathbb{E}_{\mathbf{n}}[Q_{\mathbf{n}}(x)])^2 + \mathbb{E}_{\mathbf{n}} \left[ (Q_{\mathbf{n}}(x) - \mathbb{E}_{\mathbf{n}}[Q_{\mathbf{n}}(x)])^2 \right]. \quad (6.18)$$

$x$  について期待値をとれば、

$$\mathcal{L} = \text{Bias} + \text{Var} \quad (6.19)$$

$$:= \mathbb{E}_x \left[ (R(x) - \mathbb{E}_{\mathbf{n}}[Q_{\mathbf{n}}(x)])^2 \right] + \mathbb{E}_{x,\mathbf{n}} \left[ (Q_{\mathbf{n}}(x) - \mathbb{E}_{\mathbf{n}}[Q_{\mathbf{n}}(x)])^2 \right] \quad (6.20)$$

となる。第一項をバイアス、第二項をバリエーションと呼ぶ。バイアスは予測の期待値  $\mathbb{E}_{\mathbf{n}}[Q_{\mathbf{n}}(x)]$  が真の値  $R(x)$  からどれほど離れているかを測ったもので、バリエーションは予測値  $Q_{\mathbf{n}}(x)$  の分散である。機械学習や統計的推測の分野一般において、式6.19は汎化誤差と

呼ばれる量に相当する。確率的なゆらぎのあるデータをもとにモデルのパラメータの値を推測するとき、得られたデータへのあてはまりの良さのみを考慮してパラメータを最適化するとバイアスは小さくなるが、バリエーションが大きくなってしまふことがよく知られている [15], [114]。その1つの解決方法として正則化がある。たとえば式6.1のように目的関数に正則化項を追加してその和を最適化する。胸腺からの流入は正則化項ではないが、実質的には正則化の役割を果たしていると考えられる。また、流入率  $\gamma$  は正則化の強さを調整するパラメータである。 $\gamma = 0$  のときは正則化がないのでバイアスは小さくなるがバリエーションが大きくなってしまふ。逆に  $\gamma$  が大きすぎるとバリエーションは小さくなるが本来の目的関数を最適化しないのでバイアスが大きくなる。よって正則化のパラメータ  $\gamma$  を調整することで、バイアスとバリエーションを制御して汎化誤差の意味で最適なパラメータ  $\mathbf{n}$  を得ることができる。ここで述べたような正則化パラメータとバイアスとバリエーションの関係は多くの確率モデルで調べられているが、胸腺からの流入についても同様の傾向が見られた。実際、図6.12に示すように、流入率を大きくするとバイアスに相当する量は大きくなるが、バリエーションに相当する量は小さくなる様子が確認できた。ここで、バイアスとバリエーションの推定値は  $Q_{\mathbf{n}}(x)$  の平均と分散の推定値を用いて計算する。 $x \in \bar{\mathcal{S}} \times \mathcal{A}$  について平均と分散の推定値は

$$\hat{\mu}(x) = \frac{1}{T} \sum_{t=0}^{T-1} \delta_x(x_t) Q_{\mathbf{n}^{(t)}}(x_t), \quad (6.21)$$

$$\hat{\sigma}^2(x) = \frac{1}{T-1} \sum_{t=0}^{T-1} \delta_x(x_t) (Q_{\mathbf{n}^{(t)}}(x_t) - \hat{\mu}(x_t))^2 \quad (6.22)$$

であり、Welford の方法 [115] によってオンラインで逐次的に計算した。これらを用いてバイアス Bias とバリエーション Var の推定値は

$$\widehat{\text{Bias}} = \frac{1}{T} \sum_{t=0}^{T-1} (R(x_t) - \hat{\mu}(x_t))^2, \quad (6.23)$$

$$\widehat{\text{Var}} = \frac{T-1}{T} \sum_{t=0}^{T-1} \hat{\sigma}^2(x_t) \quad (6.24)$$

として計算した。ただし、 $\mathbf{n}$  と  $(s, a)$  が独立でなく、双方の期待値を  $t$  に関する和に置き換えているため、以下のようにバイアスの推定値とバリエーションの推定値の和は誤差の時間平

均  $\widehat{\mathcal{L}}$  とは一致しない.

$$\widehat{\mathcal{L}} = \frac{1}{T} \sum_{t=0}^{T-1} L_t \tag{6.25}$$

$$= \frac{1}{T} \sum_{t=0}^{T-1} (R(x_t) - Q_{n^{(t)}}(x_t))^2 \tag{6.26}$$

$$= \widehat{\text{Bias}} + \widehat{\text{Var}} + -\frac{2}{T} \sum_{t=0}^{T-1} (R(x_t) - \hat{\mu}(x_t))(Q_{n^{(t)}}(x_t) - \hat{\mu}(x_t)). \tag{6.27}$$

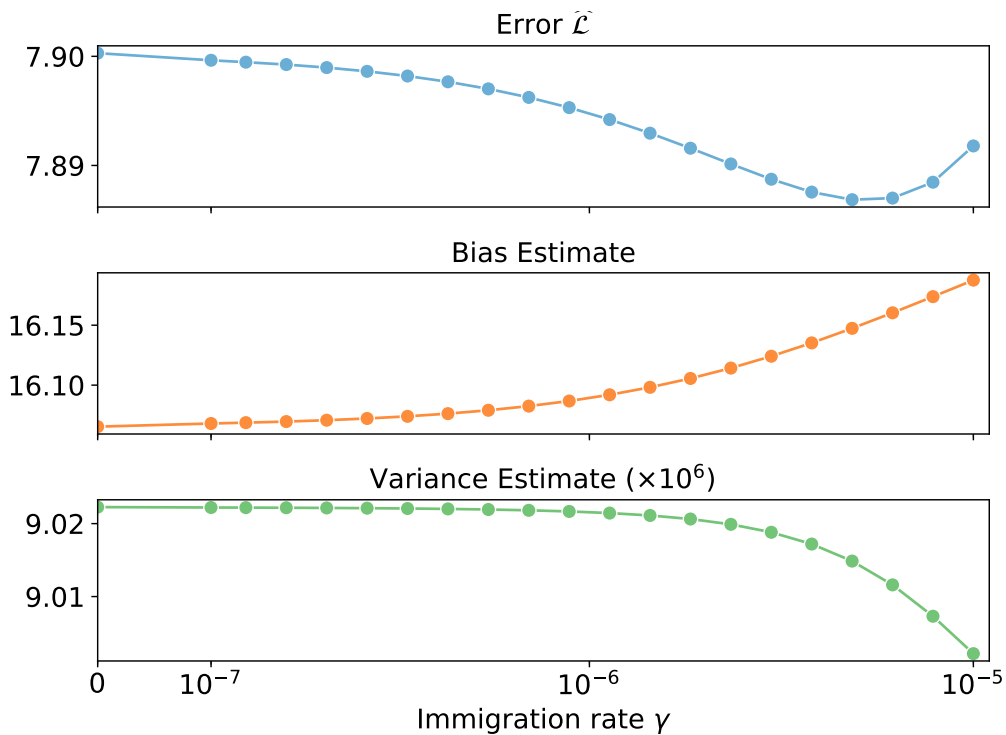


図6.12 バイアスとバリエアンスの流入率  $\gamma$  依存性. ここでは図6.7で  $\mu = 5.0$  に相当する場合の結果を示した

次に、時間の関数としての流入率  $\gamma_t$  について考える. 自然選択された免疫系で実装されていると思われるのは、全区間の平均報酬予測誤差を最小化するような流入率の時間変化  $\gamma_t$  である. しかし時間の関数について最適なものを今回のように直接列挙して同定することには膨大な計算時間がかかるため実施できていない. 今回は平均をとる時間区間を狭めることで各時刻での最適な  $\gamma_t$  を求めた. この2つの  $\gamma_t$  は異なるものだが、定性的な傾向は類似していると期待される. 今回のシミュレーションの全ての実験条件において最適な

流入率は時間について単調で非増加な関数になっており、多くの実験条件において初期は  $10^{-5}$  以上の流入率が最適だが中期に突然変化して、後期は  $10^{-7}$  以下の流入率が最適となった。これは図6.5における (2) の小児期の急激な退縮と類似している。

一方、 $K$  や  $\alpha$  が小さいときには最適な流入率はあまり変化せず、高い流入率を保つ方が良くなっている点は対照的である。このようにパラメータ依存的に最適な胸腺のあり方が変わることは、生物種によって異なる免疫系のあり方を理解するための示唆を与える。ここまでの説明では、ヒトとマウスで共通して胸腺の退縮が起こると述べてきたが、近年の研究ではヒトとマウスで大きく異なる点も見つかっている。2012年の Ineke den Braber らの研究 [116] では、いくつかの実験手法と数理モデルを組み合わせることで、胸腺からの流入と末梢での増殖が末梢の T 細胞集団の形成にどれほど寄与しているか、ヒトとマウスでそれぞれ見積もられた。青年期以降のヒトでは末梢の CD4 陽性ナイーブ T 細胞集団のうち 90% 程度が末梢での分裂増殖によって生まれた細胞であり、胸腺からの流入に相当する細胞は 10% 程度であることが分かった。一方、マウスでは週齢によらずナイーブ T 細胞は末梢でほぼ増殖せず、ほぼ全ての末梢 CD4 陽性ナイーブ T 細胞は胸腺由来であることが分かった。ただしメモリー T 細胞はある程度増殖によって維持されており二度目以降の迅速な免疫応答を担う [116]。つまり、ヒトと比較して、マウスでは増殖に対する流入の末梢 T 細胞集団への寄与が大きく、胸腺の退縮はあるが減少の効果は比較的穏やかであると言える。これは増殖率の全体的な大きさを決める  $\alpha$  が小さい場合に相当し、最適な流入率の減少は穏やかであるという今回のシミュレーション結果と整合している。また、マウスは全 T 細胞数が少なく種類数  $K$  も少ないと思われる [39], [117]。この場合もシミュレーション結果によれば流入率はあまり減少させないほうが良く、実験結果と矛盾しない。

マウス以外でも脊椎動物では胸腺の退縮が広く見られることは既に述べたが、ヒト [102]・マウス [100]・ヒツジ [118] などの哺乳類では明らかに年齢依存的な退縮が確認されている一方で、鳥類 [119] や魚類 [120] では年齢依存的な退縮が確認できた種もあるが、確認できなかった種もある。モデル生物として確立していない種が多いため退縮が確認できていないだけで実際には全ての脊椎動物で退縮している可能性もあるが、今回のシミュレーション結果は  $K$  や  $\alpha$  のように種ごとに異なる免疫系のパラメータのもとでは退縮しない胸腺も自然選択されうることを示唆する。 $K$  や  $\alpha$  といったパラメータが免疫系の誤差を小さくするという観点だけから最適なものが自然選択されるとすると、多くの種は  $K$  を大きくして  $\alpha$  を 0.05 にして、胸腺を劇的に退縮させるはずである。つまり遺伝的な効果(流入)よりも個体の経験に基づいた学習(増殖)を優先して変化する環境でも生存率を高める。しかし個体が利用できるエネルギーなどの資源のうち免疫系に割くことのできる資源は限られる、などの理由から  $K$  や  $\alpha$  を自由に変えられず、生涯を通して遺伝的な効果に

頼る (胸腺を退縮しない) ことがあり得る。

本研究では胸腺が一時的に退縮する現象については扱わなかった。哺乳類では感染や妊娠時, 精神的ストレスのある時に退縮する現象はよく知られている [103]。また爬虫類や冬眠する哺乳類では冬に退縮して春にもとに戻るような季節による変化も観察されている [121]。この一時的な退縮についても学習モデルを用いればその進化的な意義を明らかにできる可能性がある。

## 第7章

# 結論

### 7.1 研究結果のまとめ

本論文ではまず免疫学によって明らかにされた T 細胞を始めとする免疫系の知見 (第2章) と数理モデル (第3) を整理した. これらをもとに第5章では強化学習, 第6章では転移学習として理解できることを示した.

免疫系の強化学習としての見方を整理する. 免疫応答は自然免疫系のパターン認識受容体や, 獲得免疫系の抗原認識受容体への物質の結合という「観測」によって開始され, 様々なエフェクター機構の発動という「行動」をとることができる. 状況に応じて適切なエフェクター機構は異なるが, 観測と適切な行動の対応は必ずしも事前に与えられない. そこで免疫系は試行錯誤的な経験によって適切な行動を選ぶように自らを変容させる (強化学習する). 学習によって変容する生物学的実態は, ヘルパー T 細胞の多様性, すなわち T 細胞受容体の種類ごとの細胞数分布だと考えられる. 実際, 観測をもとにした行動決定においては, ヘルパー T 細胞の細胞数分布が大きく寄与し, 細胞数分布は T 細胞の増殖・死滅によってダイナミックに調節されている. T 細胞集団を中心に免疫系の観測と行動を捉え直すと, 観測は T 細胞受容体による抗原認識であり, 行動はエフェクター機構へのサイトカイン信号の発出である. 残る自然免疫系のパターン認識受容体由来の信号は, 免疫応答の適切性を報酬信号として T 細胞に伝達するものと考えられる.

次に, 免疫系の転移学習的な側面を整理する. T 細胞集団における増殖・死滅動態と比較して, 胸腺からの T 細胞の流入動態は免疫系の現在の状況に応じた機敏な変化は少ない. たとえば, 流入する全細胞数のうちの受容体の種類ごとの細胞数割合は, 個体の一生を通じて大きく変化しないと考えられる. この細胞数割合の分布は, 個体の経験ではなく進化的な経験, つまり何世代にもわたる自然選択によって最適化される. 増殖・死滅過程で

形成される末梢の T 細胞数分布は、流入過程と組み合わせられると流入分布へ近づくよう偏りが生じる。強化学習を補助して生存に有利となるような流入分布は、世代を超えて共通して頻繁に経験するような抗原、すなわち自己抗原に大して事前に学習したものであることが数値実験から分かった。これは機械学習でもちいられる転移学習という手法に相当する。

これらの学習理論的なモデル化において役立つのが第4章のリーマン多様体上での勾配系の考え方である。前述のような学習という機能を免疫系に想定した場合に、勾配系の考え方をを用いれば具体的な T 細胞集団動態を導出できた。また、勾配系として導出された集団動態の式を変形・近似することで生物学的により妥当な集団動態を得ることができた。

最後に、流入による転移学習と増殖死滅による強化学習の間にはある種のトレードオフが存在する。この2つの学習のバランスを調節するパラメータが時間あたりに流入する総 T 細胞数 (流入率) である。流入率を適切に選択することが平均的に最も良い性能を与える。哺乳類を含む多くの生物種の免疫系においては胸腺の退縮という現象が確認されており、小さい正の流入率を採用することが生存に有利であると言われている。本研究で構築した学習モデルの振る舞いは、胸腺の退縮と定性的に一致している。これによって胸腺の退縮の機能的意義という未解決問題に対して、背後にトレードオフの関係があるからであるという新たな仮説を提案した。このトレードオフは機械学習分野でバイアスとバリエーションのトレードオフとして知られる現象と関連しており、機械学習の概念によって生命現象を説明できることが示せた。

## 7.2 計算論的生物学

本研究では学習理論を用いて規範的に免疫系をモデル化することで生物学的な理解を得られることを示した。学習理論に限らずフィルタリング理論や最適制御理論など数理工学的な理論を用いて生命現象の理解を目指す学問分野を計算論的生物学と呼べば、この分野は今後も大きく発展する可能性が高い。

計算論的生物学としては計算論的神経科学が典型的であり多くの成果を挙げている。たとえば近年は深層学習と実際の脳神経系の対応が一層議論されている [122]。計算論的神経科学の研究手法は、Marr の三階層という概念によって整理される [22]。

- (1) 計算論 (どのような問題を解いているか, 目的関数)
- (2) アルゴリズム (どのような手続きで問題を解いているか)
- (3) 実装 (具体的にどのような分子や細胞の活動によって実現されているか)

脳神経系ではなく免疫系においても同様の階層的考え方が当てはまる。実際、本研究において免疫系は、(1) 観測をもとに行動を決定する逐次的な意思決定の問題を、(2) 強化学習や転移学習といったアルゴリズムで、(3) T細胞集団の増殖死滅・流入過程によって実現している、としてモデル化された。脳神経系や免疫系に限らず、胚発生や走化性など幅広い生命現象も計算論的生物学の対象となると期待される。

計算論とアルゴリズムの階層をつなげるのが数理工学だとすれば、計算論的生物学における理論の中心的な役割はアルゴリズムと実装の階層をつなげることである。たとえば、通常の工学応用においては学習のアルゴリズムは勾配降下法によって実装が導出される。しかし細胞や分子といった実装されるハードウェアの性質を踏まえると、リーマン多様体上の勾配降下法へと拡張するのが自然である。他にも、通常の学習理論の枠組みでは報酬や誤差といった学習信号を直接用いた勾配降下法の更新式が用いられる。しかし生命現象との対応を考える上では学習信号の実態が必ずしも明らかでなかったり、学習信号伝播の実態が理論の前提と異なったりする [123]。脳神経系においては、深層学習で用いられる逆誤差伝播法を変形することで、生物学的に妥当かつ学習が成功するようなモデルが盛んに研究されている [124], [125]。免疫系においては、本研究ではサイトカインによる情報伝達ネットワークが学習信号の伝播を担っていると提案した。サイトカインネットワークの実際の性質 [126] と矛盾しない学習モデルを探究することは今後の重要な課題である。

また、多くのアルゴリズムで普遍的に成り立つ法則があるとすれば、それを実装する生命現象にも対応する法則が見つかるはずである。たとえば本研究では、バイアスとバリエーションのトレードオフという統計推定における法則が登場した。単純に考えるとバリエーションの存在を無視してしまい、バイアスが小さい統計モデルを選択するのが良いという考えに陥りがちである。しかし実際にはバリエーションが存在することによって、バリエーションが小さければバイアスが大きい統計モデルを選択する方が良いという結論になりうる。この一見不合理だが実は合理的であるという事実は、本研究では胸腺の退縮を説明することに用いられた。他にもヒトの認知におけるバイアスを説明するためにこの統計推定の法則が用いられている [127]。脳神経系は情報処理に特化して高度に発達しているため、機械学習を実装したコンピュータと類似の法則が存在することは明らかだと思われる。一方で免疫系や単細胞生物など、原始的な情報処理機構にもアルゴリズム由来の隠れた法則が成立している可能性は高い。

### 7.3 工学・医学

本研究は工学や医学的にも有用な考え方や結果を示した。



まずリーマン多様体上での勾配降下法は学習アルゴリズムを実装した人工物をつくる際に有用だと思われる。例えば、強化学習するロボットをつくりたいとき、通常はマイクロコントローラなどノイマン型のコンピュータをロボットに搭載する。ノイマン型コンピュータでは浮動小数点演算によってユークリッド空間の通常の勾配降下法を実装可能である。しかしノイマン型コンピュータは汎用的であるがために、用途を絞った非ノイマン型コンピュータに凌駕され得る。深層学習に特化した専用チップ [128] や量子コンピュータ [129] は特定の問題を解くために設計される。いずれは細胞集団をプログラムすることができるかもしれない。与えられたハードウェアで勾配降下法を実装したいという状況においては、適切なリーマン多様体を同定することで具体的な更新式を導出できる。同定されたリーマン多様体構造には、ハードウェアの制約や動態が反映される。

また本研究の結果から、胸腺にまつわる医学的な示唆が得られる。高齢者はワクチンの効果が低く、自己免疫疾患やがんなどの発症リスクが高まる。その原因として、免疫老化とよばれる加齢に伴う免疫系の機能的な変容がある。免疫老化の顕著な例として古くから指摘されていたのが胸腺の退縮である。加齢と胸腺の退縮の相関関係から、胸腺の退縮を停止させたり逆に活性化させたりすることは免疫機能の向上をもたらすのではないかという考え方が一部では受け入れられており、胸腺を賦活化させる方法が多く検討されている [130]。しかし、胸腺の退縮が免疫機能の低下の要因であるという因果関係は直接示されていない。近年では胸腺の退縮以外にも T 細胞の疲弊や老化といった現象により免疫機能の低下が引き起こされることが分かっている [131]。また、胸腺の退縮速度では小児期が最も大きく青年期以降の退縮は緩やかであることから、老化に伴う免疫機能の低下とは関係がなく、むしろ積極的に退縮するよう自然選択されているということは以前から指摘されていた [103], [104], [106]–[108]。さらに本研究の結果から、胸腺が退縮すべきときに退縮しないことは無意味であるだけでなく却って悪影響があることが示唆された。したがって、年齢依存的な胸腺の退縮に対する医学的な介入には慎重になる必要がある。

一方で、本研究においても退縮が過度に進み胸腺からの T 細胞の供給が全く無くなることに対しては、免疫機能の低下が導かれることが示された。胸腺からの T 細胞の供給が極端に少なくなった高齢者に対しては胸腺の再賦活化が有効である可能性がある。他にも HIV ウイルス等への感染やがんの放射線/化学療法、骨髄移植に伴って T 細胞数が劇的に減ってしまった場合には胸腺への医学的な介入が重要になると思われる。近年では胸腺という臓器をまるごと人工的に再構成する研究 [132] もある。胸腺の役割をより正確に理解することで、どのようなタイミングでどのように胸腺へ介入することが免疫系の正常な働きを助けるのかが明らかになるだろう。

# 謝辞

本研究にあたり、指導教官の小林徹也先生には大変お世話になりました。先生のおかげで良い研究テーマに巡り会うことができました。研究として面白そうな話を沢山投げただけなので、研究として何が面白いかわかすまで学ばせて頂きましたし、私も負けずに面白そうな話を積極的に見つけて共有するようになりました。また本当に忙しい中、原稿のチェックを迅速に行っていただけて驚きました。大変感謝しています。

小林研究室の杉山友規さん、斎藤陽平さん、Dimitri Loutchko さん、上村淳さん、高倉勇氣さんには情報共有や活発な議論を通じて多くのことを学ばせて頂きました。とくにミーティング時の杉山さんの鋭い指摘や、斎藤さんとの雑談での空想的議論は刺激的でした。テクニカルスタッフの青木美帆さんには研究をすすめる上での多くの事務作業を手伝って頂きました。研究室の先輩および同期である中島蒼さん、中村絢斗さん、鳥取岳広さんには研究のことはもちろんのこと、学生生活を送る上で大事な申請書や提出書類のことなど多くのことを教えて頂きました。居心地が良く研究にも刺激的な環境をつくってくださった小林研究室の皆様、2年間ありがとうございました。

既に小林研究室を卒業された金子和正さん、加藤卓也さんも私の研究に大きく影響しています。金子さんは唯一の同じ専攻の先輩であり、大学院生活のアドバイス頂いたり、免疫系について色々な議論をさせて頂きました。加藤さんは私の研究の前身を構築してくださりました。研究の引き継ぎにあたってはお世話になりました。また、本田直樹先生と吉戸香奈さん、澤田哲先生、堀部和也さんと都筑拓さんや堀口和彦さんにも本研究に関して議論を通じて有益なアドバイスを頂きました。深く感謝しております。

東京大学の WINGS-ABC では異分野で活躍する同世代の学生との議論や情報交換ができました。本研究は複数の分野に関連するため、前提知識を共有しない異分野の方との率直な議論や査読練習は本論文を執筆する上でも大変有益でした。福島孝治先生を始めとするプログラム運営委員の先生方、プログラム生の皆様には深く感謝申し上げます。

最後に、私を支えてくれた家族に感謝の意を表して謝辞と致します。

## 参考文献

- [1] D. S. Chen and I. Mellman, “Oncology Meets Immunology: The Cancer-Immunity Cycle,” *Immunity*, vol. 39, no. 1, pp. 1–10, Jul. 2013, ISSN: 1074-7613. DOI: [10.1016/j.immuni.2013.07.012](https://doi.org/10.1016/j.immuni.2013.07.012). [Online]. Available: <https://doi.org/10.1016/j.immuni.2013.07.012>.
- [2] B. Sperner-Unterweger, “Immunological Aetiology of Major Psychiatric Disorders,” *Drugs*, vol. 65, no. 11, pp. 1493–1520, 2005, ISSN: 1179-1950. DOI: [10.2165/00003495-200565110-00004](https://doi.org/10.2165/00003495-200565110-00004). [Online]. Available: <https://doi.org/10.2165/00003495-200565110-00004>.
- [3] F. M. Burnet, “A Modification of Jerne’s Theory of Antibody Production using the Concept of Clonal Selection,” *CA: A Cancer Journal for Clinicians*, vol. 26, no. 2, pp. 119–121, 1976, ISSN: 0007-9235. DOI: [10.3322/canjclin.26.2.119](https://doi.org/10.3322/canjclin.26.2.119). [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.3322/canjclin.26.2.119>.
- [4] N. K. Jerne, “Towards a network theory of the immune system.,” eng, *Annales d’immunologie*, vol. 125C, no. 1-2, pp. 373–389, Jan. 1974, ISSN: 0300-4910 (Print).
- [5] G. W. Hoffmann, “A theory of regulation and self-nonsel self discrimination in an immune network.,” eng, *European journal of immunology*, vol. 5, no. 9, pp. 638–647, Sep. 1975, ISSN: 0014-2980 (Print). DOI: [10.1002/eji.1830050912](https://doi.org/10.1002/eji.1830050912).
- [6] P. Matzinger, “Tolerance, Danger, and the Extended Family,” *Annual Review of Immunology*, vol. 12, no. 1, pp. 991–1045, Apr. 1994, ISSN: 0732-0582. DOI: [10.1146/annurev.iy.12.040194.005015](https://doi.org/10.1146/annurev.iy.12.040194.005015). [Online]. Available: <https://doi.org/10.1146/annurev.iy.12.040194.005015>.

- [7] I. R. Cohen, “Real and artificial immune systems: computing the state of the body,” *Nature Reviews Immunology*, vol. 7, no. 7, pp. 569–574, 2007, ISSN: 1474-1741. DOI: [10.1038/nri2102](https://doi.org/10.1038/nri2102). [Online]. Available: <https://doi.org/10.1038/nri2102>.
- [8] R. Medzhitov, D. S. Schneider, and M. P. Soares, “Disease Tolerance as a Defense Strategy,” *Science*, vol. 335, no. 6071, pp. 936–941, 2012, ISSN: 0036-8075. DOI: [10.1126/science.1214935](https://doi.org/10.1126/science.1214935). [Online]. Available: <https://science.sciencemag.org/content/335/6071/936>.
- [9] T. Pradeu, S. Jaeger, and E. Vivier, “The speed of change: towards a discontinuity theory of immunity?” *Nature Reviews Immunology*, vol. 13, no. 10, pp. 764–769, 2013, ISSN: 1474-1741. DOI: [10.1038/nri3521](https://doi.org/10.1038/nri3521). [Online]. Available: <https://doi.org/10.1038/nri3521>.
- [10] S. Nataf, “Clonal selection versus clonal cooperation: the integrated perception of immune objects [version 1; peer review: 2 approved],” *F1000Research*, vol. 5, no. 2226, 2016. DOI: [10.12688/f1000research.9386.1](https://doi.org/10.12688/f1000research.9386.1).
- [11] H. Veiga-Fernandes and A. A. Freitas, “The S(c)ensory Immune System Theory,” *Trends in Immunology*, vol. 38, no. 10, pp. 777–788, Oct. 2017, ISSN: 1471-4906. DOI: [10.1016/j.it.2017.02.007](https://doi.org/10.1016/j.it.2017.02.007). [Online]. Available: <https://doi.org/10.1016/j.it.2017.02.007>.
- [12] I. R. Cohen and S. Efroni, “The Immune System Computes the State of the Body: Crowd Wisdom, Machine Learning, and Immune Cell Reference Repertoires Help Manage Inflammation,” vol. 10, p. 10, 2019. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fimmu.2019.00010>.
- [13] R. Eftimie, J. J. Gillard, and D. A. Cantrell, “Mathematical Models for Immunology: Current State of the Art and Future Research Directions,” eng, *Bulletin of mathematical biology*, vol. 78, no. 10, pp. 2091–2134, Oct. 2016, ISSN: 1522-9602. DOI: [10.1007/s11538-016-0214-9](https://doi.org/10.1007/s11538-016-0214-9). [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/27714570%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5069344/>.
- [14] 豊. 山本, 学習・教育. 新曜社, 2008, p. 134.
- [15] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
- [16] J. H. Holland, *Adaptation in Natural and Artificial Systems*. MIT Press, 1975, p. 232, ISBN: 9780262581110.

- [17] J. Farmer, N. H. Packard, and A. S. Perelson, “The immune system, adaptation, and machine learning,” *Physica D: Nonlinear Phenomena*, vol. 22, no. 1, pp. 187–204, 1986, ISSN: 0167-2789. DOI: [https://doi.org/10.1016/0167-2789\(86\)90240-X](https://doi.org/10.1016/0167-2789(86)90240-X). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/016727898690240X>.
- [18] D. Dasgupta and F. Nino, *Immunological Computation: Theory and Applications*, 1st ed. USA: Auerbach Publications, 2008, ISBN: 1420065459.
- [19] R. Bellman, “The theory of dynamic programming,” en, *Bull. Amer. Math. Soc.*, vol. 60, no. 6, pp. 503–515, 1954, ISSN: 0002-9904. [Online]. Available: <https://projecteuclid.org:443/euclid.bams/1183519147>.
- [20] C. J. C. H. Watkins and P. Dayan, “Q-learning,” *Machine Learning*, vol. 8, no. 3, pp. 279–292, 1992, ISSN: 1573-0565. DOI: [10.1007/BF00992698](https://doi.org/10.1007/BF00992698). [Online]. Available: <https://doi.org/10.1007/BF00992698>.
- [21] R. S. Sutton, Barto, and A. G., *Reinforcement Learning: An Introduction*, 2nd ed. MIT Press, 2018, ISBN: 9780262039246.
- [22] D. Marr, *Vision*. MIT Press, 1982.
- [23] T. Kato and T. J. Kobayashi, “Understanding Adaptive Immune System as Reinforcement Learning,” *ArXiv*, 2019. arXiv: [1904.05581 \[q-bio.PE\]](https://arxiv.org/abs/1904.05581). [Online]. Available: <https://arxiv.org/abs/1904.05581>.
- [24] L. Valiant, *Probably Approximately Correct: Nature’s Algorithms for Learning and Prospering in a Complex World*. Basic Books, 2014, p. 208, ISBN: 978-0465060726.
- [25] K. Murphy and C. Weaver, *Janeway’s Immunobiology*, 9th ed. Garland Science, 2016, p. 904, ISBN: 9781315533247.
- [26] H. Spits, D. Artis, M. Colonna, A. Diefenbach, J. P. Di Santo, G. Eberl, S. Koyasu, R. M. Locksley, A. N. J. McKenzie, R. E. Mebius, F. Powrie, and E. Vivier, “Innate lymphoid cells —a proposal for uniform nomenclature,” *Nature Reviews Immunology*, vol. 13, no. 2, pp. 145–149, 2013, ISSN: 1474-1741. DOI: [10.1038/nri3365](https://doi.org/10.1038/nri3365). [Online]. Available: <https://doi.org/10.1038/nri3365>.
- [27] M. L. Robinette and M. Colonna, “Immune modules shared by innate lymphoid cells and T cells,” *Journal of Allergy and Clinical Immunology*, vol. 138, no. 5, pp. 1243–1251, Nov. 2016, ISSN: 0091-6749. DOI: [10.1016/j.jaci.2016.09.006](https://doi.org/10.1016/j.jaci.2016.09.006). [Online]. Available: <https://doi.org/10.1016/j.jaci.2016.09.006>.

- [28] R. Gillitzer and M. Goebeler, “Chemokines in cutaneous wound healing,” *Journal of Leukocyte Biology*, vol. 69, no. 4, pp. 513–521, Apr. 2001, ISSN: 0741-5400. DOI: <https://doi.org/10.1189/jlb.69.4.513>. [Online]. Available: <https://doi.org/10.1189/jlb.69.4.513>.
- [29] M. Schäffer, M. Bongartz, W. Hoffmann, and R. Viebahn, “MHC-Class-II-Deficiency Impairs Wound Healing,” *Journal of Surgical Research*, vol. 138, no. 1, pp. 100–105, Mar. 2007, ISSN: 0022-4804. DOI: [10.1016/j.jss.2006.05.029](https://doi.org/10.1016/j.jss.2006.05.029). [Online]. Available: <https://doi.org/10.1016/j.jss.2006.05.029>.
- [30] C. Stockmann, D. Schadendorf, R. Klose, and I. Helfrich, *The Impact of the Immune System on Tumor: Angiogenesis and Vascular Remodeling*, 2014. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fonc.2014.00069>.
- [31] Y. Korem Kohanim, A. Tendler, A. Mayo, N. Friedman, and U. Alon, “Endocrine Autoimmune Disease as a Fragility of Immune Surveillance against Hypersecreting Mutants,” *Immunity*, vol. 52, no. 5, pp. 872–884.e5, 2020, ISSN: 1074-7613. DOI: <https://doi.org/10.1016/j.immuni.2020.04.022>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1074761320301801>.
- [32] K. Mukaigasa, A. Hanasaki, M. Maéno, H. Fujii, S.-i. Hayashida, M. Itoh, M. Kobayashi, S. Tochinai, M. Hatta, K. Iwabuchi, M. Taira, K. Onoé, and Y. Izutsu, “The keratin-related Ouroboros proteins function as immune antigens mediating tail regression in *Xenopus* metamorphosis,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 43, pp. 18309–18314, Oct. 2009. DOI: [10.1073/pnas.0708837106](https://doi.org/10.1073/pnas.0708837106). [Online]. Available: <http://www.pnas.org/content/106/43/18309.abstract>.
- [33] N. J. Schuldt and B. A. Binstadt, “Dual TCR T Cells: Identity Crisis or Multitaskers?” *The Journal of Immunology*, vol. 202, no. 3, pp. 637–644, Feb. 2019. DOI: [10.4049/jimmunol.1800904](https://doi.org/10.4049/jimmunol.1800904). [Online]. Available: <http://www.jimmunol.org/content/202/3/637.abstract>.
- [34] L. Klein, B. Kyewski, P. M. Allen, and K. A. Hogquist, “Positive and negative selection of the T cell repertoire: what thymocytes see (and don’t see),” *Nature*

- Reviews Immunology*, vol. 14, no. 6, pp. 377–391, 2014, ISSN: 1474-1741. DOI: [10.1038/nri3667](https://doi.org/10.1038/nri3667). [Online]. Available: <https://doi.org/10.1038/nri3667>.
- [35] R. J. De Boer and A. S. Perelson, “T cell repertoires and competitive exclusion,” eng, *Journal of theoretical biology*, vol. 169, no. 4, pp. 375–390, Aug. 1994, ISSN: 0022-5193 (Print). DOI: [10.1006/jtbi.1994.1160](https://doi.org/10.1006/jtbi.1994.1160).
- [36] ———, “Towards a general function describing T cell proliferation,” eng, *Journal of theoretical biology*, vol. 175, no. 4, pp. 567–576, Aug. 1995, ISSN: 0022-5193 (Print). DOI: [10.1006/jtbi.1995.0165](https://doi.org/10.1006/jtbi.1995.0165).
- [37] A. Mayer, Y. Zhang, A. S. Perelson, and N. S. Wingreen, “Regulation of T cell expansion by antigen presentation dynamics,” *Proceedings of the National Academy of Sciences*, vol. 116, no. 13, 5914 LP –5919, Mar. 2019. DOI: [10.1073/pnas.1812800116](https://doi.org/10.1073/pnas.1812800116). [Online]. Available: <http://www.pnas.org/content/116/13/5914.abstract>.
- [38] G. F. ( F. Gauze, *The struggle for existence*. Baltimore, The Williams & Wilkins company, 1934.
- [39] T. Mora and A. M. Walczak, “How many different clonotypes do immune repertoires contain?” *Current Opinion in Systems Biology*, vol. 18, pp. 104–110, 2019, ISSN: 2452-3100. DOI: <https://doi.org/10.1016/j.coisb.2019.10.001>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S2452310019300289>.
- [40] J. A. Weinstein, N. Jiang, R. A. 3. White, D. S. Fisher, and S. R. Quake, “High-throughput sequencing of the zebrafish antibody repertoire,” eng, *Science (New York, N.Y.)*, vol. 324, no. 5928, pp. 807–810, May 2009, ISSN: 1095-9203 (Electronic). DOI: [10.1126/science.1170020](https://doi.org/10.1126/science.1170020).
- [41] T. Mora, A. M. Walczak, W. Bialek, and C. G. Callan, “Maximum entropy models for antibody diversity,” *Proceedings of the National Academy of Sciences*, vol. 107, no. 12, 5405 LP –5410, Mar. 2010. DOI: [10.1073/pnas.1001705107](https://doi.org/10.1073/pnas.1001705107). [Online]. Available: <http://www.pnas.org/content/107/12/5405.abstract>.
- [42] J. M. Heather, M. Ismail, T. Oakes, and B. Chain, “High-throughput sequencing of the T-cell receptor repertoire: pitfalls and opportunities,” *Briefings in Bioinformatics*, vol. 19, no. 4, pp. 554–565, Jul. 2018, ISSN: 1477-4054. DOI:

- [10.1093/bib/bbw138](https://doi.org/10.1093/bib/bbw138). [Online]. Available: <https://doi.org/10.1093/bib/bbw138>.
- [43] P. C. de Greef, T. Oakes, B. Gerritsen, M. Ismail, J. M. Heather, R. Hermsen, B. Chain, and R. J. de Boer, “The naive T-cell receptor repertoire has an extremely broad distribution of clone sizes,” *eLife*, vol. 9, A. M. Walczak, S. Rath, and M. V. Pogorelyy, Eds., e49900, 2020, ISSN: 2050-084X. DOI: [10.7554/eLife.49900](https://doi.org/10.7554/eLife.49900). [Online]. Available: <https://doi.org/10.7554/eLife.49900>.
- [44] J. Desponds, T. Mora, and A. M. Walczak, “Fluctuating fitness shapes the clone-size distribution of immune repertoires,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 2, pp. 274–279, 2016, ISSN: 10916490. DOI: [10.1073/pnas.1512977112](https://doi.org/10.1073/pnas.1512977112). arXiv: [1507.00751](https://arxiv.org/abs/1507.00751).
- [45] R. Dessalles, M. D’Orsogna, and T. Chou, “How heterogeneous thymic output and homeostatic proliferation shape naive T cell receptor clone abundance distributions,” *bioRxiv*, 2019.
- [46] M. U. Gaimann, M. Nguyen, J. Desponds, and A. Mayer, “Early life imprints the hierarchy of T cell clone sizes,” *eLife*, vol. 9, A. Nourmohammad, Ed., e61639, 2020, ISSN: 2050-084X. DOI: [10.7554/eLife.61639](https://doi.org/10.7554/eLife.61639). [Online]. Available: <https://doi.org/10.7554/eLife.61639>.
- [47] C. F. Arias, M. A. Herrero, F. J. Acosta, and C. Fernandez-Arias, “Population mechanics: A mathematical framework to study T cell homeostasis,” *Scientific Reports*, vol. 7, no. 1, p. 9511, 2017, ISSN: 2045-2322. DOI: [10.1038/s41598-017-09949-w](https://doi.org/10.1038/s41598-017-09949-w). [Online]. Available: <https://doi.org/10.1038/s41598-017-09949-w>.
- [48] 克. 大野, 非線形な世界. 東京大学出版会, 2009, ISBN: 978-4-13-063352-9.
- [49] E. J. Carr, J. Dooley, J. E. Garcia-Perez, V. Lagou, J. C. Lee, C. Wouters, I. Meyts, A. Goris, G. Boeckxstaens, M. A. Linterman, and A. Liston, “The cellular composition of the human immune system is shaped by age and cohabitation,” *Nature Immunology*, vol. 17, no. 4, pp. 461–468, 2016, ISSN: 1529-2916. DOI: [10.1038/ni.3371](https://doi.org/10.1038/ni.3371). [Online]. Available: <https://doi.org/10.1038/ni.3371>.



- [50] S. Smale, “On the differential equations of species in competition,” *Journal of Mathematical Biology*, vol. 3, no. 1, pp. 5–7, 1976, ISSN: 1432-1416. DOI: [10.1007/BF00307854](https://doi.org/10.1007/BF00307854). [Online]. Available: <https://doi.org/10.1007/BF00307854>.
- [51] J. A. Vano, J. C. Wildenberg, M. B. Anderson, J. K. Noel, and J. C. Sprott, “Chaos in low-dimensional Lotka–Volterra models of competition,” *Nonlinearity*, vol. 19, no. 10, pp. 2391–2404, 2006, ISSN: 0951-7715. DOI: [10.1088/0951-7715/19/10/006](https://doi.org/10.1088/0951-7715/19/10/006). [Online]. Available: <http://dx.doi.org/10.1088/0951-7715/19/10/006>.
- [52] J. R. Jackson, “On the existence problem of linear programming,” *Journal of Mathematics*, vol. 4, no. 1, pp. 29–36, 1954. DOI: [10.2140/pjm.1954.4.29](https://doi.org/10.2140/pjm.1954.4.29).
- [53] J. Farkas, “Theorie der einfachen Ungleichungen.,” English, *Journal für die reine und angewandte Mathematik*, vol. 1902, no. 124, pp. 1–27, DOI: <https://doi.org/10.1515/crll.1902.124.1>. [Online]. Available: <https://www.degruyter.com/view/journals/crll/1902/124/article-p1.xml>.
- [54] A. Mayer, V. Balasubramanian, T. Mora, and A. M. Walczak, “How a well-adapted immune system is organized.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 19, pp. 5950–5, May 2015, ISSN: 1091-6490. DOI: [10.1073/pnas.1421827112](https://doi.org/10.1073/pnas.1421827112). [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/25918407><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4434741>.
- [55] A. Mayer, V. Balasubramanian, A. M. Walczak, and T. Mora, “How a well-adapting immune system remembers,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 116, no. 18, pp. 8815–8823, 2019, ISSN: 10916490. DOI: [10.1073/pnas.1812810116](https://doi.org/10.1073/pnas.1812810116). arXiv: [1806.05753](https://arxiv.org/abs/1806.05753).
- [56] A. Mozeika, F. Fraternali, D. Dunn-Walters, and A. Coolen, “Roles of repertoire diversity in robustness of humoral immune response,” *ArXiv*, Jan. 2019. DOI: [10.13140/RG.2.2.10811.57120](https://doi.org/10.13140/RG.2.2.10811.57120).
- [57] O. H. Schnaack and A. Nourmohammad, *Optimal evolutionary decision-making to store immune memory*, 2020. arXiv: [2007.01363](https://arxiv.org/abs/2007.01363) [q-bio.PE].
- [58] R. Marsland III, O. Howell, A. Mayer, and P. Mehta, *Tregs self-organize into a “computing ecosystem” and implement a sophisticated optimization algorithm for mediating immune response*, 2020. arXiv: [2012.00252](https://arxiv.org/abs/2012.00252) [q-bio.PE].

- [59] E. De Giorgi, “New problems on minimizing movements,” in *Boundary Value Problems for PDE and Applications*, 1993, pp. 81–98.
- [60] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Birkhäuser Basel, 2005, p. 334.
- [61] N. Boumal, “Optimization and estimation on manifolds,” Ph.D. dissertation, Université catholique de Louvain, Feb. 2014.
- [62] S.-i. Amari, “Divergence, Optimization and Geometry,” C. S. Leung, M. Lee, and J. H. Chan, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 185–193, ISBN: 978-3-642-10677-4.
- [63] Y. E. Nesterov, “A method for solving the convex programming problem with convergence rate  $O(1/k^2)$ ,” *Dokl. Akad. Nauk SSSR*, vol. 269, pp. 543–547, 1983. [Online]. Available: <https://ci.nii.ac.jp/naid/10029946121/>.
- [64] A. Nemirovsky and Y. D.B., *Problem complexity and method efficiency in optimization*, English, ser. Wiley-Interscience series in discrete mathematics. New York: Wiley, 1983, xv, 388 p. ISBN: 9780471103455 CN - QA402.5 .N45413 1983.
- [65] A. Beck and M. Teboulle, “Mirror descent and nonlinear projected subgradient methods for convex optimization,” *Operations Research Letters*, vol. 31, no. 3, pp. 167–175, 2003, ISSN: 0167-6377. DOI: [https://doi.org/10.1016/S0167-6377\(02\)00231-6](https://doi.org/10.1016/S0167-6377(02)00231-6). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0167637702002316>.
- [66] J. Kivinen and M. K. Warmuth, “Exponentiated Gradient versus Gradient Descent for Linear Predictors,” *Information and Computation*, vol. 132, no. 1, pp. 1–63, 1997, ISSN: 0890-5401. DOI: <https://doi.org/10.1006/inco.1996.2612>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0890540196926127>.
- [67] S.-I. Amari, “Natural Gradient Works Efficiently in Learning,” *Neural Comput.*, vol. 10, no. 2, pp. 251–276, Feb. 1998, ISSN: 0899-7667. DOI: [10.1162/089976698300017746](https://doi.org/10.1162/089976698300017746). [Online]. Available: <http://dx.doi.org/10.1162/089976698300017746>.
- [68] ———, “ $\alpha$ -Divergence is Unique, Belonging to Both f-Divergence and Bregman Divergence Classes,” *IEEE Trans. Inf. Theor.*, vol. 55, no. 11, pp. 4925–4931,

- Nov. 2009, ISSN: 0018-9448. DOI: [10.1109/TIT.2009.2030485](https://doi.org/10.1109/TIT.2009.2030485). [Online]. Available: <https://doi.org/10.1109/TIT.2009.2030485>.
- [69] I. Csiszár, “Why Least Squares and Maximum Entropy? An Axiomatic Approach to Inference for Linear Inverse Problems,” en, *Ann. Statist.*, vol. 19, no. 4, pp. 2032–2066, 1991, ISSN: 0090-5364. DOI: [10.1214/aos/1176348385](https://doi.org/10.1214/aos/1176348385). [Online]. Available: <https://projecteuclid.org:443/euclid.aos/1176348385>.
- [70] H. Robbins and S. Monro, “A Stochastic Approximation Method,” en, *Ann. Math. Statist.*, vol. 22, no. 3, pp. 400–407, 1951, ISSN: 0003-4851. DOI: [10.1214/aoms/1177729586](https://doi.org/10.1214/aoms/1177729586). [Online]. Available: <https://projecteuclid.org:443/euclid.aoms/1177729586>.
- [71] J. R. Blum, “Approximation Methods which Converge with Probability one,” en, *Ann. Math. Statist.*, vol. 25, no. 2, pp. 382–386, 1954, ISSN: 0003-4851. DOI: [10.1214/aoms/1177728794](https://doi.org/10.1214/aoms/1177728794). [Online]. Available: <https://projecteuclid.org:443/euclid.aoms/1177728794>.
- [72] S. Bonnabel, “Stochastic Gradient Descent on Riemannian Manifolds,” *IEEE Transactions on Automatic Control*, vol. 58, no. 9, pp. 2217–2229, 2013, ISSN: 1558-2523. DOI: [10.1109/TAC.2013.2254619](https://doi.org/10.1109/TAC.2013.2254619).
- [73] L. P. Kaelbling, M. L. Littman, and A. W. Moore, “Reinforcement Learning: A Survey,” *J. Artif. Int. Res.*, vol. 4, no. 1, pp. 237–285, May 1996, ISSN: 1076-9757.
- [74] B. O’Donoghue, “Variational Bayesian Reinforcement Learning with Regret Bounds,” *arXiv e-prints*, arXiv:1807.09647, Jul. 2018. arXiv: [1807.09647](https://arxiv.org/abs/1807.09647) [cs.LG].
- [75] G. A. Rummery and M. Niranjan, “On-Line {Q}-Learning Using Connectionist Systems,” Cambridge University Engineering Department, Cambridge, England, Tech. Rep. TR 166, 1994.
- [76] S. Singh, T. Jaakkola, M. L. Littman, and C. Szepesvári, “Convergence Results for Single-Step On-Policy Reinforcement-Learning Algorithms,” *Machine Learning*, vol. 38, no. 3, pp. 287–308, 2000, ISSN: 1573-0565. DOI: [10.1023/A:1007678930559](https://doi.org/10.1023/A:1007678930559). [Online]. Available: <https://doi.org/10.1023/A:1007678930559>.

- [77] F. S. Melo, S. P. Meyn, and M. I. Ribeiro, “An Analysis of Reinforcement Learning with Function Approximation,” in *Proceedings of the 25th International Conference on Machine Learning*, ser. ICML '08, New York, NY, USA: Association for Computing Machinery, 2008, pp. 664–671, ISBN: 9781605582054. DOI: [10 . 1145 / 1390156 . 1390240](https://doi.org/10.1145/1390156.1390240). [Online]. Available: <https://doi.org/10.1145/1390156.1390240>.
- [78] H. J. Kushner and G. G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, Second. Springer, New York, NY, 2003, ISBN: 978-0-387-21769-7. DOI: [10.1007/b97441](https://doi.org/10.1007/b97441).
- [79] M. Lever, H.-S. Lim, P. Kruger, J. Nguyen, N. Trendel, E. Abu-Shah, P. K. Maini, P. A. van der Merwe, and O. Dushek, “Architecture of a minimal signaling pathway explains the T-cell response to a 1 million-fold variation in antigen affinity and dose,” *Proceedings of the National Academy of Sciences*, vol. 113, no. 43, E6630 LP –E6638, Oct. 2016. DOI: [10.1073/pnas.1608820113](https://doi.org/10.1073/pnas.1608820113). [Online]. Available: <http://www.pnas.org/content/113/43/E6630.abstract>.
- [80] J. Gálvez, J. J. Gálvez, and P. García-Peñarrubia, *Is TCR/pMHC Affinity a Good Estimate of the T-cell Response? An Answer Based on Predictions From 12 Phenotypic Models*, 2019. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fimmu.2019.00349>.
- [81] J. Zhu, H. Yamane, and W. E. Paul, “Differentiation of Effector CD4 T Cell Populations,” *Annual Review of Immunology*, vol. 28, no. 1, pp. 445–489, 2010. DOI: [10.1146/annurev-immunol-030409-101212](https://doi.org/10.1146/annurev-immunol-030409-101212). [Online]. Available: <https://doi.org/10.1146/annurev-immunol-030409-101212>.
- [82] O. Takeuchi and S. Akira, “Pattern Recognition Receptors and Inflammation,” *Cell*, vol. 140, no. 6, pp. 805–820, Mar. 2010, ISSN: 0092-8674. DOI: [10.1016/j.cell.2010.01.022](https://doi.org/10.1016/j.cell.2010.01.022). [Online]. Available: <https://doi.org/10.1016/j.cell.2010.01.022>.
- [83] H. S. Scott, M. Heino, P. Peterson, L. Mittaz, M. D. Lalioti, C. Betterle, A. Cohen, M. Seri, M. Lerone, G. Romeo, P. Collin, M. Salo, R. Metcalfe, A. Weetman, M.-P. Papasavvas, C. Rossier, K. Nagamine, J. Kudoh, N. Shimizu, K. J. E. Krohn, and S. E. Antonarakis, “Common Mutations in Autoimmune Polyendocrinopathy-Candidiasis-Ectodermal Dystrophy Patients of Different Origins,” *Molecular Endocrinology*, vol. 12, no. 8, pp. 1112–1119, Aug. 1998,

- ISSN: 0888-8809. DOI: [10.1210/mend.12.8.0143](https://doi.org/10.1210/mend.12.8.0143). [Online]. Available: <https://doi.org/10.1210/mend.12.8.0143>.
- [84] Y. L. Zhang and Q. Yang, “A Survey on Multi-Task Learning,” *ArXiv*, 2017. arXiv: [1707.08114](https://arxiv.org/abs/1707.08114).
- [85] S. Ruder, “An Overview of Multi-Task Learning in Deep Neural Networks,” *ArXiv*, 2017. arXiv: [1706.05098](https://arxiv.org/abs/1706.05098).
- [86] J. Baxter, “A Model of Inductive Bias Learning,” *J. Artif. Int. Res.*, vol. 12, no. 1, pp. 149–198, Mar. 2000, ISSN: 1076-9757.
- [87] S. Riedel, “Edward Jenner and the history of smallpox and vaccination,” eng, *Proceedings (Baylor University. Medical Center)*, vol. 18, no. 1, pp. 21–25, Jan. 2005, ISSN: 0899-8280. DOI: [10.1080/08998280.2005.11928028](https://doi.org/10.1080/08998280.2005.11928028). [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/16200144/><https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1200696/>.
- [88] Z. Sethna, Y. Elhanati, C. G. Callan Jr, A. M. Walczak, and T. Mora, “OLGA: fast computation of generation probabilities of B- and T-cell receptor amino acid sequences and motifs,” *Bioinformatics*, vol. 35, no. 17, pp. 2974–2981, Sep. 2019, ISSN: 1367-4803. DOI: [10.1093/bioinformatics/btz035](https://doi.org/10.1093/bioinformatics/btz035). [Online]. Available: <https://doi.org/10.1093/bioinformatics/btz035>.
- [89] K. A. Hogquist, T. A. Baldwin, and S. C. Jameson, “Central tolerance: learning self-control in the thymus,” *Nature Reviews Immunology*, vol. 5, no. 10, pp. 772–782, 2005, ISSN: 1474-1741. DOI: [10.1038/nri1707](https://doi.org/10.1038/nri1707). [Online]. Available: <https://doi.org/10.1038/nri1707>.
- [90] K. Christopher Garcia, J. J. Adams, D. Feng, and L. K. Ely, “The molecular basis of TCR germline bias for MHC is surprisingly simple,” *Nature Immunology*, vol. 10, no. 2, pp. 143–147, 2009, ISSN: 1529-2916. DOI: [10.1038/ni.f.219](https://doi.org/10.1038/ni.f.219). [Online]. Available: <https://doi.org/10.1038/ni.f.219>.
- [91] P. G. Thomas and J. C. Crawford, “Selected before selection: A case for inherent antigen bias in the T-cell receptor repertoire,” *Current Opinion in Systems Biology*, vol. 18, pp. 36–43, 2019, ISSN: 2452-3100. DOI: <https://doi.org/10.1016/j.coisb.2019.10.007>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S245231001930054X>.
- [92] Z. Sethna, G. Isacchini, T. Dupic, T. Mora, A. M. Walczak, and Y. Elhanati, “Population variability in the generation and selection of T-cell repertoires,”

- PLOS Computational Biology*, vol. 16, no. 12, e1008394, Dec. 2020. [Online]. Available: <https://doi.org/10.1371/journal.pcbi.1008394>.
- [93] A. Madi, A. Poran, E. Shifrut, S. Reich-Zeliger, E. Greenstein, I. Zaretsky, T. Arnon, F. V. Laethem, A. Singer, J. Lu, P. D. Sun, I. R. Cohen, and N. Friedman, “T cell receptor repertoires of mice and humans are clustered in similarity networks around conserved public CDR3 sequences,” *eng, eLife*, vol. 6, e22057, Jul. 2017, ISSN: 2050-084X. DOI: [10.7554/eLife.22057](https://doi.org/10.7554/eLife.22057). [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/28731407%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5553937/>.
- [94] N. Vriskoop, J. P. Monteiro, J. N. Mandl, and R. N. Germain, “Revisiting Thymic Positive Selection and the Mature T Cell Repertoire for Antigen,” *Immunity*, vol. 41, no. 2, pp. 181–190, Aug. 2014, ISSN: 1074-7613. DOI: [10.1016/j.immuni.2014.07.007](https://doi.org/10.1016/j.immuni.2014.07.007). [Online]. Available: <https://doi.org/10.1016/j.immuni.2014.07.007>.
- [95] P. J. R. Ebert, S. Jiang, J. Xie, Q.-J. Li, and M. M. Davis, “An endogenous positively selecting peptide enhances mature T cell responses and becomes an autoantigen in the absence of microRNA miR-181a,” *Nature Immunology*, vol. 10, no. 11, pp. 1162–1169, 2009, ISSN: 1529-2916. DOI: [10.1038/ni.1797](https://doi.org/10.1038/ni.1797). [Online]. Available: <https://doi.org/10.1038/ni.1797>.
- [96] W.-L. Lo, N. J. Felix, J. J. Walters, H. Rohrs, M. L. Gross, and P. M. Allen, “An endogenous peptide positively selects and augments the activation and survival of peripheral CD4+ T cells,” *Nature Immunology*, vol. 10, no. 11, pp. 1155–1161, 2009, ISSN: 1529-2916. DOI: [10.1038/ni.1796](https://doi.org/10.1038/ni.1796). [Online]. Available: <https://doi.org/10.1038/ni.1796>.
- [97] M. Krogsgaard, Q.-j. Li, C. Sumen, J. B. Huppa, M. Huse, and M. M. Davis, “Agonist/endogenous peptide–MHC heterodimers drive T cell activation and sensitivity,” *Nature*, vol. 434, no. 7030, pp. 238–243, 2005, ISSN: 1476-4687. DOI: [10.1038/nature03391](https://doi.org/10.1038/nature03391). [Online]. Available: <https://doi.org/10.1038/nature03391>.
- [98] J. N. Mandl, J. P. Monteiro, N. Vriskoop, and R. N. Germain, “T Cell-Positive Selection Uses Self-Ligand Binding Strength to Optimize Repertoire Recognition of Foreign Antigens,” *Immunity*, vol. 38, no. 2, pp. 263–274, Feb. 2013,

- ISSN: 1074-7613. DOI: [10.1016/j.immuni.2012.09.011](https://doi.org/10.1016/j.immuni.2012.09.011). [Online]. Available: <https://doi.org/10.1016/j.immuni.2012.09.011>.
- [99] I. M. N. Wortel, C. Keşmir, R. J. de Boer, J. N. Mandl, and J. Textor, *Is T Cell Negative Selection a Learning Algorithm?* 2020. DOI: [10.3390/cells9030690](https://doi.org/10.3390/cells9030690).
- [100] J. S. Hale, T. E. Boursalian, G. L. Turk, and P. J. Fink, “Thymic output in aged mice,” *Proceedings of the National Academy of Sciences*, vol. 103, no. 22, 8447 LP–8452, May 2006. DOI: [10.1073/pnas.0601040103](https://doi.org/10.1073/pnas.0601040103). [Online]. Available: <http://www.pnas.org/content/103/22/8447.abstract>.
- [101] C. L. Ortman, K. A. Dittmar, P. L. Witte, and P. T. Le, “Molecular characterization of the mouse involuted thymus: aberrations in expression of transcription regulators in thymocyte and epithelial compartments,” *International Immunology*, vol. 14, no. 7, pp. 813–822, Jul. 2002, ISSN: 0953-8178. DOI: [10.1093/intimm/dxf042](https://doi.org/10.1093/intimm/dxf042). [Online]. Available: <https://doi.org/10.1093/intimm/dxf042>.
- [102] G. G. Steinmann, B. Klaus, and H.-K. Müller-Hermelink, “The Involution of the Ageing Human Thymic Epithelium is Independent of Puberty,” *Scandinavian Journal of Immunology*, vol. 22, no. 5, pp. 563–575, Nov. 1985, ISSN: 0300-9475. DOI: <https://doi.org/10.1111/j.1365-3083.1985.tb01916.x>. [Online]. Available: <https://doi.org/10.1111/j.1365-3083.1985.tb01916.x>.
- [103] J. Dooley and A. Liston, “Molecular control over thymic involution: From cytokines and microRNA to aging and adipose tissue,” *European Journal of Immunology*, vol. 42, no. 5, pp. 1073–1079, May 2012, ISSN: 0014-2980. DOI: <https://doi.org/10.1002/eji.201142305>. [Online]. Available: <https://doi.org/10.1002/eji.201142305>.
- [104] D. P. Shanley, D. Aw, N. R. Manley, and D. B. Palmer, “An evolutionary perspective on the mechanisms of immunosenescence,” *Trends in Immunology*, vol. 30, no. 7, pp. 374–381, 2009, ISSN: 1471-4906. DOI: <https://doi.org/10.1016/j.it.2009.05.001>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1471490609001057>.
- [105] R. D. Kovaïou, D. Herndler-Brandstetter, and B. Grubeck-Loebenstien, “Age-related changes in immunity: implications for vaccination in the elderly,” *Expert Reviews in Molecular Medicine*, vol. 9, no. 3, pp. 1–17, 2007. DOI: [DOI: 10.1017/S1462399407000221](https://doi.org/10.1017/S1462399407000221). [Online]. Available: <https://www.cambridge>.

- [org/core/article/agerelated-changes-in-immunity-implications-for-vaccination-in-the-elderly/7A47989867F9FA1E468C29DA7F3A3BF0](http://core/article/agerelated-changes-in-immunity-implications-for-vaccination-in-the-elderly/7A47989867F9FA1E468C29DA7F3A3BF0).
- [106] M. Torroba and A. G. Zapata, “Aging of the vertebrate immune system,” *Microscopy Research and Technique*, vol. 62, no. 6, pp. 477–481, Dec. 2003, ISSN: 1059-910X. DOI: <https://doi.org/10.1002/jemt.10409>. [Online]. Available: <https://doi.org/10.1002/jemt.10409>.
- [107] A. J. T. George and M. A. Ritter, “Thymic involution with ageing: obsolescence or good housekeeping?” *Immunology Today*, vol. 17, no. 6, pp. 267–272, 1996, ISSN: 0167-5699. DOI: [https://doi.org/10.1016/0167-5699\(96\)80543-3](https://doi.org/10.1016/0167-5699(96)80543-3). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0167569996805433>.
- [108] M. R. Dowling and P. D. Hodgkin, “Why does the thymus involute? A selection-based hypothesis,” *Trends in Immunology*, vol. 30, no. 7, pp. 295–300, Jul. 2009, ISSN: 1471-4906. DOI: [10.1016/j.it.2009.04.006](https://doi.org/10.1016/j.it.2009.04.006). [Online]. Available: <https://doi.org/10.1016/j.it.2009.04.006>.
- [109] V. Appay, D. Sauce, and M. Prelog, “The role of the thymus in immunosenescence: lessons from the study of thymectomized individuals,” *Aging*, vol. 2, no. 2, pp. 78–81, 2010, ISSN: 1945-4589. DOI: [10.18632/aging.100122](https://doi.org/10.18632/aging.100122). [Online]. Available: <https://doi.org/10.18632/aging.100122>.
- [110] M. Prelog, C. Wilk, M. Keller, T. Karall, D. Orth, R. Geiger, G. Walder, G. Laufer, M. Cottogni, B. Zimmerhackl Lothar, J. Stein, B. Grubeck-Loebenstein, and R. Wuerzner, “Diminished response to tick-borne encephalitis vaccination in thymectomized children,” *Vaccine*, vol. 26, no. 5, pp. 595–600, 2008, ISSN: 0264-410X. DOI: <https://doi.org/10.1016/j.vaccine.2007.11.074>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0264410X07013412>.
- [111] X.-Y. Zhang, K. Tanaka, Y. Koga, Y. Wang, and K. Nomoto, “The Effect of Adult Thymectomy on Immune Response to Infection with *Listeria monocytogenes* in Mice,” *Immunobiology*, vol. 188, no. 4, pp. 355–369, 1993, ISSN: 0171-2985. DOI: [https://doi.org/10.1016/S0171-2985\(11\)80219-6](https://doi.org/10.1016/S0171-2985(11)80219-6). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0171298511802196>.



- [112] J. A. Goggins II, “Exploring the role of thymic output and recent thymic emigrants during *Salmonella* infection,” English, Ph.D. dissertation, Ann Arbor, 2016, p. 165, ISBN: 978-1-369-68430-8. [Online]. Available: [https://search.proquest.com/dissertations-theses/exploring-role-thymic-output-recent-emigrants/docview/1884651688/se-2?accountid=14357%20http://vs2ga4mq9g.search.serialssolutions.com?ctx%7B%5C\\_%7Dver=Z39.88-2004%7B%5C%7Dctx%7B%5C\\_%7Denc=info:ofi/enc:UTF-8%7B%5C%7Drfr%7B%5C\\_%7Ddid=info:sid/ProQuest+Dissertations+%7B%5C%7D26+Theses+Global%7B%5C%7Drft%7B%5C\\_%7Dval%7B%5C\\_%7Dfmt=info:ofi/fmt:kev:mtx:dissertation%7B%5C%7Drft.genre=dissertations+%7B%5C%7D26+theses%7B%5C%7Drft.jtitle=%7B%5C%7Drft.atitle=%7B%5C%7Drft.au=Goggins%7B%5C%7D2C+James+Alan%7B%5C%7D2C+II%7B%5C%7Drft.aulast=Goggins%7B%5C%7Drft.aufirst=James%7B%5C%7Drft.date=2016-01-01%7B%5C%7Drft.volume=%7B%5C%7Drft.issue=%7B%5C%7Drft.spage=%7B%5C%7Drft.isbn=978-1-369-68430-8%7B%5C%7Drft.btitle=%7B%5C%7Drft.title=Exploring+the+role+of+thymic+output+and+recent+thymic+emigrants+during+Salmonella+infection%7B%5C%7Drft.issn=%7B%5C%7Drft%7B%5C\\_%7Ddid=info:doi/](https://search.proquest.com/dissertations-theses/exploring-role-thymic-output-recent-emigrants/docview/1884651688/se-2?accountid=14357%20http://vs2ga4mq9g.search.serialssolutions.com?ctx%7B%5C_%7Dver=Z39.88-2004%7B%5C%7Dctx%7B%5C_%7Denc=info:ofi/enc:UTF-8%7B%5C%7Drfr%7B%5C_%7Ddid=info:sid/ProQuest+Dissertations+%7B%5C%7D26+Theses+Global%7B%5C%7Drft%7B%5C_%7Dval%7B%5C_%7Dfmt=info:ofi/fmt:kev:mtx:dissertation%7B%5C%7Drft.genre=dissertations+%7B%5C%7D26+theses%7B%5C%7Drft.jtitle=%7B%5C%7Drft.atitle=%7B%5C%7Drft.au=Goggins%7B%5C%7D2C+James+Alan%7B%5C%7D2C+II%7B%5C%7Drft.aulast=Goggins%7B%5C%7Drft.aufirst=James%7B%5C%7Drft.date=2016-01-01%7B%5C%7Drft.volume=%7B%5C%7Drft.issue=%7B%5C%7Drft.spage=%7B%5C%7Drft.isbn=978-1-369-68430-8%7B%5C%7Drft.btitle=%7B%5C%7Drft.title=Exploring+the+role+of+thymic+output+and+recent+thymic+emigrants+during+Salmonella+infection%7B%5C%7Drft.issn=%7B%5C%7Drft%7B%5C_%7Ddid=info:doi/).
- [113] J. A. Goggins, J. R. Kurtz, and J. B. McLachlan, *Control of Persistent Salmonella Infection Relies on Constant Thymic Output Despite Increased Peripheral Antigen-Specific T Cell Immunity*, 2020. DOI: [10.3390/pathogens9080605](https://doi.org/10.3390/pathogens9080605).
- [114] G. H. Golub, M. Heath, and G. Wahba, “Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter,” *Technometrics*, vol. 21, no. 2, pp. 215–223, Jan. 1979, ISSN: 00401706. DOI: [10.2307/1268518](https://doi.org/10.2307/1268518). [Online]. Available: <http://www.jstor.org/stable/1268518>.
- [115] B. P. Welford, “Note on a Method for Calculating Corrected Sums of Squares and Products,” *Technometrics*, vol. 4, no. 3, pp. 419–420, Aug. 1962, ISSN: 0040-1706. DOI: [10.1080/00401706.1962.10490022](https://doi.org/10.1080/00401706.1962.10490022). [Online]. Available: <https://www.tandfonline.com/doi/abs/10.1080/00401706.1962.10490022>.
- [116] I. den Braber, T. Mugwagwa, N. Vrisekoop, L. Westera, R. Mögling, A. Bregje de Boer, N. Willems, E. H. Schrijver, G. Spierenburg, K. Gaiser, E. Mul, S. A. Otto, A. F. Ruiter, M. T. Ackermans, F. Miedema, J. A. Borghans, R. J.

- de Boer, and K. Tesselaar, "Maintenance of Peripheral Naive T Cells Is Sustained by Thymus Output in Mice but Not Humans," *Immunity*, vol. 36, no. 2, pp. 288–297, 2012, ISSN: 1074-7613. DOI: <https://doi.org/10.1016/j.immuni.2012.02.006>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1074761312000556>.
- [117] J. Nikolich-Žugich, M. K. Slifka, and I. Messaoudi, "The many important facets of T-cell repertoire diversity," *Nature Reviews Immunology*, vol. 4, no. 2, pp. 123–132, Feb. 2004, ISSN: 1474-1733. DOI: [10.1038/nri1292](https://doi.org/10.1038/nri1292). [Online]. Available: <http://www.nature.com/doifinder/10.1038/nri1292%20http://www.ncbi.nlm.nih.gov/pubmed/15040585>.
- [118] C. P. Cunningham, W. G. Kimpton, J. E. Holder, and R. N. P. Cahill, "Thymic export in aged sheep: a continuous role for the thymus throughout pre- and postnatal life," *European Journal of Immunology*, vol. 31, no. 3, pp. 802–811, Mar. 2001, ISSN: 0014-2980. DOI: [https://doi.org/10.1002/1521-4141\(200103\)31:3<802::AID-IMMU802>3.0.CO;2-P](https://doi.org/10.1002/1521-4141(200103)31:3<802::AID-IMMU802>3.0.CO;2-P). [Online]. Available: [https://doi.org/10.1002/1521-4141\(200103\)31:3%7B%5C%%7D3C802::AID-IMMU802%7B%5C%%7D3E3.0.CO%20http://2-p](https://doi.org/10.1002/1521-4141(200103)31:3%7B%5C%%7D3C802::AID-IMMU802%7B%5C%%7D3E3.0.CO%20http://2-p).
- [119] E. Ciriaco, P. P. Píñera, B. Díaz-Esnal, and R. Laurà, "Age-related changes in the avian primary lymphoid organs (thymus and bursa of Fabricius)," *Microscopy Research and Technique*, vol. 62, no. 6, pp. 482–487, Dec. 2003, ISSN: 1059-910X. DOI: <https://doi.org/10.1002/jemt.10416>. [Online]. Available: <https://doi.org/10.1002/jemt.10416>.
- [120] S. Chilmonczyk, "The thymus in fish: Development and possible function in the immune response," *Annual Review of Fish Diseases*, vol. 2, pp. 181–200, 1992, ISSN: 0959-8030. DOI: [https://doi.org/10.1016/0959-8030\(92\)90063-4](https://doi.org/10.1016/0959-8030(92)90063-4). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0959803092900634>.
- [121] L. M. Zimmerman, L. A. Vogel, and R. M. Bowden, "Understanding the vertebrate immune system: insights from the reptilian perspective," *The Journal of Experimental Biology*, vol. 213, no. 5, pp. 661 LP –671, Mar. 2010. DOI: [10.1242/jeb.038315](https://doi.org/10.1242/jeb.038315). [Online]. Available: <http://jeb.biologists.org/content/213/5/661.abstract>.

- [122] B. A. Richards, T. P. Lillicrap, P. Beaudoin, Y. Bengio, R. Bogacz, A. Christensen, C. Clopath, R. P. Costa, A. de Berker, S. Ganguli, C. J. Gillon, D. Hafner, A. Kepecs, N. Kriegeskorte, P. Latham, G. W. Lindsay, K. D. Miller, R. Naud, C. C. Pack, P. Poirazi, P. Roelfsema, J. Sacramento, A. Saxe, B. Scellier, A. C. Schapiro, W. Senn, G. Wayne, D. Yamins, F. Zenke, J. Zylberberg, D. Therien, and K. P. Kording, “A deep learning framework for neuroscience,” *Nature Neuroscience*, vol. 22, no. 11, pp. 1761–1770, 2019, ISSN: 1546-1726. DOI: [10.1038/s41593-019-0520-2](https://doi.org/10.1038/s41593-019-0520-2). [Online]. Available: <https://doi.org/10.1038/s41593-019-0520-2>.
- [123] E. Hunsberger, “Spiking Deep Neural Networks: Engineered and Biological Approaches to Object Recognition,” Ph.D. dissertation, 2018. [Online]. Available: <http://hdl.handle.net/10012/12819>.
- [124] T. P. Lillicrap, D. Cownden, D. B. Tweed, and C. J. Akerman, “Random synaptic feedback weights support error backpropagation for deep learning,” *Nature Communications*, vol. 7, no. 1, p. 13276, 2016, ISSN: 2041-1723. DOI: [10.1038/ncomms13276](https://doi.org/10.1038/ncomms13276). [Online]. Available: <https://doi.org/10.1038/ncomms13276>.
- [125] G. Bellec, F. Scherr, E. Hajek, D. Salaj, R. Legenstein, and W. Maass, “Biologically inspired alternatives to backpropagation through time for learning in recurrent neural nets,” *ArXiv*, 2019. arXiv: [arXiv:1901.09049](https://arxiv.org/abs/1901.09049).
- [126] Z. Frankenstein, U. Alon, and I. R. Cohen, “The immune-body cytokine network defines a social architecture of cell interactions.,” *eng, Biology direct*, vol. 1, p. 32, Oct. 2006, ISSN: 1745-6150 (Electronic). DOI: [10.1186/1745-6150-1-32](https://doi.org/10.1186/1745-6150-1-32).
- [127] G. Gigerenzer and H. Brighton, “Homo Heuristicus: Why Biased Minds Make Better Inferences,” *Topics in Cognitive Science*, vol. 1, no. 1, pp. 107–143, Jan. 2009, ISSN: 1756-8757. DOI: <https://doi.org/10.1111/j.1756-8765.2008.01006.x>. [Online]. Available: <https://doi.org/10.1111/j.1756-8765.2008.01006.x>.
- [128] N. P. Jouspi, C. Young, N. Patil, D. Patterson, G. Agrawal, R. Bajwa, S. Bates, S. Bhatia, N. Boden, A. Borchers, R. Boyle, P.-l. Cantin, C. Chao, C. Clark, J. Coriell, M. Daley, M. Dau, J. Dean, B. Gelb, T. V. Ghaemmghami, R. Gottipati, W. Gulland, R. Hagmann, C. R. Ho, D. Hogberg, J. Hu, R.

- Hundt, D. Hurt, J. Ibarz, A. Jaffey, A. Jaworski, A. Kaplan, H. Khaitan, D. Killebrew, A. Koch, N. Kumar, S. Lacy, J. Laudon, J. Law, D. Le, C. Leary, Z. Liu, K. Lucke, A. Lundin, G. MacKean, A. Maggiore, M. Mahony, K. Miller, R. Nagarajan, R. Narayanaswami, R. Ni, K. Nix, T. Norrie, M. Omernick, N. Penukonda, A. Phelps, J. Ross, M. Ross, A. Salek, E. Samadiani, C. Severn, G. Sizikov, M. Snelham, J. Souter, D. Steinberg, A. Swing, M. Tan, G. Thorson, B. Tian, H. Toma, E. Tuttle, V. Vasudevan, R. Walter, W. Wang, E. Wilcox, and D. H. Yoon, “In-Datacenter Performance Analysis of a Tensor Processing Unit,” in *Proceedings of the 44th Annual International Symposium on Computer Architecture*, ser. ISCA '17, New York, NY, USA: Association for Computing Machinery, 2017, pp. 1–12, ISBN: 9781450348928. DOI: [10.1145/3079856.3080246](https://doi.org/10.1145/3079856.3080246). [Online]. Available: <https://doi.org/10.1145/3079856.3080246>.
- [129] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J. R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Y. Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. Rieffel, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis, “Quantum supremacy using a programmable superconducting processor,” *Nature*, vol. 574, no. 7779, pp. 505–510, 2019, ISSN: 1476-4687. DOI: [10.1038/s41586-019-1666-5](https://doi.org/10.1038/s41586-019-1666-5). [Online]. Available: <https://doi.org/10.1038/s41586-019-1666-5>.
- [130] G. A. Holländer, W. Krenger, and B. R. Blazar, “Emerging strategies to boost thymic function,” *Current Opinion in Pharmacology*, vol. 10, no. 4, pp. 443–453, 2010, ISSN: 1471-4892. DOI: <https://doi.org/10.1016/j.coph.2010.04.008>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1471489210000445>.

- [131] A. N. Akbar and S. M. Henson, “Are senescence and exhaustion intertwined or unrelated processes that compromise immunity?” *Nature Reviews Immunology*, vol. 11, no. 4, pp. 289–295, 2011, ISSN: 1474-1741. DOI: [10.1038/nri2959](https://doi.org/10.1038/nri2959). [Online]. Available: <https://doi.org/10.1038/nri2959>.
- [132] S. Campinoti, A. Gjinovci, R. Ragazzini, L. Zanieri, L. Ariza-McNaughton, M. Catucci, S. Boeing, J.-E. Park, J. C. Hutchinson, M. Muñoz-Ruiz, P. G. Manti, G. Voza, C. E. Villa, D.-E. Phylactopoulos, C. Maurer, G. Testa, H. J. Stauss, S. A. Teichmann, N. J. Sebire, A. C. Hayday, D. Bonnet, and P. Bonfanti, “Reconstitution of a functional human thymus by postnatal stromal progenitor cells and natural whole-organ scaffolds,” *Nature Communications*, vol. 11, no. 1, p. 6372, 2020, ISSN: 2041-1723. DOI: [10.1038/s41467-020-20082-7](https://doi.org/10.1038/s41467-020-20082-7). [Online]. Available: <https://doi.org/10.1038/s41467-020-20082-7>.
- [133] H. K. Khalil, *Nonlinear Control*. Pearson Education, 2015, ISBN: 978-0-13-349926-1.
- [134] I. D. Čuešov and I. Chueshov, *Introduction to the Theory of Infinite-Dimensional Dissipative Systems*. “Acta” publishers, 2002, p. 418, ISBN: 966-7021-64-5.
- [135] J. A. Walker, *Dynamical Systems and Evolution Equations: Theory and Applications*, J. A. Walker, Ed. Boston, MA: Springer US, 1980, p. 236, ISBN: 978-1-4684-1038-9. DOI: [10.1007/978-1-4684-1036-5](https://doi.org/10.1007/978-1-4684-1036-5). [Online]. Available: <https://doi.org/10.1007/978-1-4684-1036-5>.
- [136] V. Răsvan, “Dynamical systems with several equilibria and natural Liapunov functions,” *Archivum Mathematicum*, vol. 34, no. 1, pp. 207–215, 1998. [Online]. Available: <http://hdl.handle.net/10338.dmlcz/107646>.
- [137] A. M. Lyapunov, “The general problem of the stability of motion,” *International Journal of Control*, vol. 55, no. 3, pp. 531–534, Mar. 1992, ISSN: 0020-7179. DOI: [10.1080/00207179208934253](https://doi.org/10.1080/00207179208934253). [Online]. Available: <https://doi.org/10.1080/00207179208934253>.
- [138] J. LaSalle, “Some Extensions of Liapunov’s Second Method,” *IRE Transactions on Circuit Theory*, vol. 7, no. 4, pp. 520–527, 1960, ISSN: 2331-3854 VO - 7. DOI: [10.1109/TCT.1960.1086720](https://doi.org/10.1109/TCT.1960.1086720).

- [139] G. A. Leonov, V. A. Yakubovich, and A. K. Gelig, *Stability Of Stationary Sets In Control Systems With Discontinuous Nonlinearities*, ser. Series On Stability, Vibration And Control Of Systems, Series A. World Scientific Publishing Company, 2004, ISBN: 9789814483483. [Online]. Available: <https://books.google.co.jp/books?id=hA7JCgAAQBAJ>.
- [140] C. F. Havrda Jan, “Quantification method of classification processes. Concept of structural entropy,” *Kybernetika*, vol. 03, no. 1, (30)–35, 1967. [Online]. Available: <http://eudml.org/doc/28681>.
- [141] A. Schied, “Sample path large deviations for super-Brownian motion,” *Probability Theory and Related Fields*, vol. 104, no. 3, pp. 319–347, 1996, ISSN: 1432-2064. DOI: [10.1007/BF01213684](https://doi.org/10.1007/BF01213684). [Online]. Available: <https://doi.org/10.1007/BF01213684>.
- [142] M. Liero, A. Mielke, and G. Savaré, “Optimal Transport in Competition with Reaction: The Hellinger–Kantorovich Distance and Geodesic Curves,” *SIAM Journal on Mathematical Analysis*, vol. 48, no. 4, pp. 2869–2911, Jan. 2016, ISSN: 0036-1410. DOI: [10.1137/15M1041420](https://doi.org/10.1137/15M1041420). [Online]. Available: <https://doi.org/10.1137/15M1041420>.

## 付録 A

# 力学系の安定性

以下は主に [133]–[136] による.

力学系とは完備距離空間  $X$  と,  $X$  から  $X$  への連続写像の族  $S_t$  からなる  $(X, S_t)$  である.  $S_t$  は非負実数  $t, \tau \in \mathbb{R}_{\geq 0}$  について

$$S_{t+\tau} = S_t \circ S_\tau, \quad S_0 = I \quad (\text{A.1})$$

を満たすとする. ただし  $I$  は恒等写像. ここでは  $X$  は有限の  $d$  次元のユークリッド空間の部分空間とする. 力学系  $(X, S_t)$  の軌道  $\gamma$  とは  $\{x(t) := S_t x_0 \mid t \geq 0\}$  と表せる集合で, 初期値  $x_0 \in X$  によって唯一に決まる. ここでは常微分方程式

$$\frac{dx_i}{dt} = f_i(x) \quad (i = 1, \dots, d) \quad (\text{A.2})$$

の形で表せるような自励的な力学系を扱う.  $x_e \in X$  が任意の  $t \geq 0$  で  $S_t x_e = x_e$  すなわち  $f_i(x_e) = 0 (i = 1, \dots, d)$  を満たすとき,  $x_e$  をこの力学系の**平衡点** (equilibrium point, fixed point, stationary point) とよぶ.

### A.1 平衡点の安定性

まずは平衡点の安定性について述べる.

**定義 A.1.1** (平衡点の安定性). 平衡点  $x_e$  が**(リアプノフ) 安定**であるとは, 任意の  $\varepsilon > 0$  について,  $\delta > 0$  が存在して, 任意の  $x \in X, t \geq 0$  で

$$d(x, x_e) < \delta \Rightarrow d(S_t x, x_e) < \varepsilon \quad (\text{A.3})$$

を満たすことである。また、平衡点  $x_e$  が漸近安定であるとは、それが安定でかつ  $\delta > 0$  が存在して

$$d(x, x_e) < \delta \Rightarrow \lim_{t \rightarrow \infty} d(S_t x, x_e) = 0 \quad (\text{A.4})$$

を満たすことである。任意に大きい  $\delta$  で成立するなら  $x_e$  は大域的に漸近安定であるという。

安定性について示すときに有用なのがリアプノフ関数である。力学系  $(X, S_t)$  と連続的微分可能な関数  $V : X \rightarrow \mathbb{R}$  について、 $\dot{V} : X \rightarrow \mathbb{R}$  を

$$\dot{V}(x) = \frac{dV}{dt} = \sum_{i=1}^d \frac{\partial V}{\partial x_i} \frac{dx_i}{dt} \quad (\text{A.5})$$

と定める。  $X$  の部分集合  $D$  のすべての点  $x \in D$  において  $\dot{V}(x) \leq 0$  を満たすとき、  $V$  を  $D$  におけるリアプノフ関数とよぶ。

**定理 A.1.2** (Lyapunov の定理 [137]). 平衡点  $x_e$  を含む開集合  $D$  におけるリアプノフ関数を  $V$  とする。任意の  $x \in D \setminus x_e$  において  $V(x) > V(x_e)$  を満たすとき、  $x_e$  は安定である。また、任意の  $x \in D \setminus x_e$  において  $\dot{V}(x) < 0$  なら  $x_e$  は漸近安定である。さらに、  $D = X$  でこれらが成立して、  $\|x\| \rightarrow \infty \Rightarrow V(x) \rightarrow \infty$  なら  $x_e$  は大域的に漸近安定である。

証明は [133] などを参照。

## A.2 平衡点集合の安定性

平衡点が複数ある場合などはリアプノフの定理で大域的な安定性を調べられない。以下では平衡点は無限個あってもよいが、平衡点の集合は有界だとする。力学系の長時間の挙動を調べたいときには不変集合・極限集合の概念が有用であるのでまずはこれを定義する。ただし点  $y$  と集合  $A$  の距離

$$\text{dist}(y, A) := \inf_{y \in A} d(z, y) \quad (\text{A.6})$$

を用いる。ここで、  $d(x, y)$  は距離空間  $X$  における  $x$  と  $y$  の距離である。

**定義 A.2.1** (不変集合).  $Y$  が正の不変集合であるとは、  $X$  の部分集合  $Y$  で、任意の  $x \in Y$  と  $t \geq 0$  において

$$S_t y \supseteq Y \quad (\text{A.7})$$



を満たすことである.

$X$  全体や空集合, 平衡点の集合, 周期軌道の閉曲線などは不変集合である.

**定義 A.2.2** (極限集合).  $x \in X$  について,  $x$  の  $\omega$ -極限集合あるいは正の極限集合とは

$$\omega(x) := \bigcap_{s \geq 0} \overline{\bigcup_{t \geq s} S_t x} \quad (\text{A.8})$$

と定義される. ただし,  $\bar{Y}$  は  $X$  における部分集合  $Y$  の閉包を表す. つまり, ある点  $y \in X$  が極限集合  $\omega(x)$  に属していることの必要十分条件は,  $y$  へ収束する部分列がとれること, つまり

$$\lim_{n \rightarrow \infty} d(S_{t_n} x, y) = 0 \quad (\text{A.9})$$

を満たすような  $\{t_n\} \subset \mathbb{R}_{\geq 0}, t_n \rightarrow \infty$  が存在することである. 定義より  $\omega(x)$  は閉集合.

**命題 A.2.3** (極限集合の性質). 軌道  $\{x(t) = S_t x \mid t \geq 0\}$  が有界のとき,  $x$  の極限集合  $\omega(x)$  は空でない連結した有界閉な正の不変集合である. さらに,

$$\text{dist}(S_t x, \omega(x)) \rightarrow 0 \quad (t \rightarrow \infty) \quad (\text{A.10})$$

を満たす.

**証明.**  $\omega(x)$  が連結であることは [135] を参照. 閉であることはすでに述べた. 軌道が有界なので  $\omega(x)$  も有界であり,  $\{x(t)\}$  に収束部分列をとれるので  $\omega(x)$  は空集合でない.  $\omega(x)$  が正の不変集合であることを示す. 任意の  $y \in \omega(x)$  について  $n \rightarrow \infty$  で  $t_n \rightarrow \infty$  と  $S_{t_n} x \rightarrow y$  を満たすような  $\{t_n\} \subset \mathbb{R}_{\geq 0}$  がとれる.  $\tau > 0$  について, 式A.1と  $S_\tau$  の連続性より,

$$S_{t_n + \tau} x = S_{t_n} S_\tau x \rightarrow S_\tau y. \quad (\text{A.11})$$

最左辺は  $\omega(x)$  に収束するから  $S_\tau y \in \omega(x)$ . つまり  $\omega(x)$  は正の不変集合.

最後に式A.10を背理法で示す. 式A.10が成り立たないと仮定すると,  $\varepsilon > 0$  と  $\{t_n\}, t_n \rightarrow \infty$  が存在して  $n = 1, 2, \dots$  で  $\text{dist}(S_{t_n} x, \omega(x)) > \varepsilon$  とできる. 一方, 軌道が有界だから収束部分列  $\{S_{t_m} x\}$  が存在する. つまり  $\{t_m\}$  が存在して  $m \rightarrow \infty$  で  $t_m \rightarrow \infty, \text{dist}(S_{t_m} x, \omega(x)) \rightarrow 0$  が成立する. その収束先  $y = \lim_{m \rightarrow \infty} S_{t_m} x$  は,  $y \in \omega(x)$  だが, 仮定より  $\text{dist}(y, \omega(x)) > \varepsilon$  となり矛盾.  $\square$

**定理 A.2.4** (LaSalle の定理 [138]).  $D$  を有界閉集合で,  $D$  を初期値にもつ任意の軌道は  $D$  内にとどまり,  $D$  におけるリアプノフ関数  $V$  が存在するとする. 集合  $N = \{y \in D \mid$

$\dot{V}(y) = 0$  の部分集合で最大の正の不変集合を  $M$  とすると,  $D$  内の任意の点  $x$  の  $\omega$ -極限集合は  $M$  に含まれる.

$$\omega(x) \subset M \subset N \subset D \quad (\text{A.12})$$

さらに,  $x \in D$  を初期値とすると  $\text{dist}(S_t x, M) \rightarrow 0$  ( $t \rightarrow \infty$ ) を満たす.

**証明.** 初期値  $x$  を任意に  $D$  の中にとって  $x \in D$ , その解  $x(t) = S_t x$  を考える.  $V$  は有界閉集合  $D$  で連続だから下に有界で,  $V(x(t))$  は  $t$  について非増加でなので極限值  $\beta = \inf_{t \geq 0} V(x(t))$  が存在する. 極限集合  $\omega(x)$  は  $D$  が閉集合なので  $\omega(x) \subset D$  で,  $V$  の連続性より任意の  $y \in \omega(x)$  において  $V(y) = \beta$ . 命題A.2.3より,  $\omega(x)$  は正の不変集合だが,  $\omega(x)$  の中では  $V$  は定数なので  $y \in \omega(x)$  で  $\dot{V}(y) = 0$  (つまり  $\omega(x) \subset N \subset D$ ).  $\omega(x)$  は不変集合だから,  $\omega(x) \subset M \subset N$ . 従って命題A.2.3より  $x \in D$  なら  $t \rightarrow \infty$  で  $\text{dist}(S_t x, M) \rightarrow 0$ .  $\square$

次に, 平衡点の集合 (stationary set[136], [139]) の安定性を定義して LaSalle の定理によって集合の安定性を示す. 力学系の性質としての安定性の定義と集合についての安定性の定義がある.

**定義 A.2.5.** 任意の有界な軌道  $x(t)$  が平衡点集合  $\mathcal{E}$  へ近づく, すなわち

$$\text{dist}(x(t), \mathcal{E}) \rightarrow 0 \quad (t \rightarrow \infty) \quad (\text{A.13})$$

を満たすとき, この力学系は**概単安定** (quasi-monostable, dichotomic) であるという. 有界とは限らない任意の軌道  $x(t)$  で式A.13が成立するなら, この力学系は**概勾配的** (quasi-gradient-like) であるという.

概勾配系的であれば振動し続けるような解は存在しない.

**定義 A.2.6** (平衡点集合の安定性). 平衡点集合  $\mathcal{E} \subset X$  が (リアプノフ) **安定** であるとは, 任意の  $\varepsilon > 0$  について  $\delta > 0$  が存在して, 任意の  $x \in X$ ,  $t \geq 0$  で

$$\text{dist}(x, \mathcal{E}) < \delta \Rightarrow \text{dist}(S_t x, \mathcal{E}) < \varepsilon \quad (\text{A.14})$$

であることとして定義する. 同様に, 集合  $\mathcal{E}$  が安定で,  $\delta > 0$  が存在して

$$\text{dist}(x, \mathcal{E}) < \delta \Rightarrow \lim_{t \rightarrow \infty} \text{dist}(S_t x, \mathcal{E}) = 0 \quad (\text{A.15})$$

のとき集合  $\mathcal{E}$  は**漸近安定** であるという. 任意の  $\delta > 0$  で漸近安定なら大域的に漸近安定である.

**定理 A.2.7** ([136], [139]).  $X$  全体におけるリアプノフ関数  $V$  が存在するとする. 有界な解軌道  $x(t)$  で,  $V(x(\tau)) = V(x(0))$  なる  $\tau > 0$  が存在するような軌道は平衡点しかない場合, この力学系は概単安定である. これに加えて,  $\|x\| \rightarrow \infty$  で  $V(x) \rightarrow \infty$  なら, 概勾配的である. すなわち平衡点全体の集合  $\mathcal{E}$  は大域的に漸近安定である.

**証明.** 平衡点集合  $\mathcal{E}$  を含む任意の有界閉集合  $D \subset X$  をとる. 仮定より,  $\dot{V}(y) = 0$  を満たすような  $y \in D$  は平衡点である. よって LaSalle の定理における集合  $N = \{y \in D \mid \dot{V}(y) = 0\}$  は, 平衡点で構成される ( $N \subset \mathcal{E}$ ). 平衡点の集合は不変集合なので,  $N$  の最大不変集合  $M$  は  $N$  と一致する ( $M = N$ ). LaSalle の定理より  $D$  に初期値をもつ軌道  $x(t)$  は  $\mathcal{E} \cap N = M$  に近づく, つまり  $\text{dist}(x(t), \mathcal{E}) \rightarrow 0 (t \rightarrow \infty)$ . よって概単安定性は示された.

追加の条件は,  $V(x)$  が有界なら  $x$  も有界であることを意味するから,  $\Omega_c = \{x \in X \mid V(x) \leq c\}$  は任意の  $c > 0$  について有界. 任意の初期値  $x_0 \in X$  について  $D = \Omega_{V(x_0)}$  とすれば前述の議論で平衡点集合へ近づく. よって概勾配系的であり, 平衡点集合は大域的に漸近安定である.  $\square$

漸近安定性は, 軌道が平衡点集合の周囲を周りながら近づいて収束先が存在しないような場合も含まれる. 任意の軌道が平衡点集合のいずれか一点に収束するときには**大域的に各点漸近安定** (pointwise globally stable) とよぶ. 大域的な各点漸近安定性を示すには任意の軌道に収束先が存在することを示す必要があり, 2変数のリアプノフ関数のような関数を用いて示せることが知られている ([139] の補題 1.7).

## 付録 B

# 距離・ダイバージェンス

### B.1 距離とダイバージェンスの定義

集合  $X$  上の関数  $d: X \times X \rightarrow \mathbb{R}$  が任意の  $x, y, z \in X$  について以下の条件を満たすとき,  $d$  は  $X$  における距離関数あるいは距離であるという.

- (非負性)  $d(x, y) \geq 0$
- (非退化性)  $x = y$  のときに限って  $d(x, y) = 0$
- (対称性)  $d(x, y) = d(y, x)$
- (三角不等式)  $d(x, y) + d(y, z) \geq d(x, z)$

座標が定義された多様体  $X$  上の関数  $D: X \times X \rightarrow \mathbb{R}$  が任意の  $x, y \in X$  について以下の条件を満たすとき,  $D$  は  $X$  におけるダイバージェンスであるという.  $x, y$  における値は  $D(x\|y)$  と書く.

- $D(x\|y) \geq 0$
- $x = y$  のときに限って  $D(x\|y) = 0$
- 小さい  $dx \in X$  について正定値二次形式でテーラー展開可能

$$D(x + dx\|x) \approx \frac{1}{2} \sum_{i,j} g_{ij} dx_i dx_j. \quad (\text{B.1})$$

対称性  $D(x\|y) = D(y\|x)$  は成り立つとは限らないが, 3つ目の条件により,  $O(dx)$  のみ考えれば

$$D(x + dx\|x) = \frac{1}{2} \sum_{i,j} g_{ij} dx_i dx_j = D(x\|x + dx) \quad (\text{B.2})$$

となるので, ダイバージェンスは局所的には対称性をみたく [62].

## B.2 ダイバージェンスの例

Bregman ダイバージェンスとは狭義凸で微分可能な関数  $\psi : X \rightarrow \mathbb{R}$  を用いて

$$D_\psi(x||y) := \psi(x) - \psi(y) - \nabla\psi(y) \cdot (x - y) \quad (\text{B.3})$$

と定義される。ここで、 $\nabla\psi(x)$  は  $x$  における偏微分  $\partial\psi/\partial x_i$  のベクトルで、 $\cdot$  は通常の内積である。

Bregman ダイバージェンスは凸関数のルジャンドル変換 (凸共役) と強く関係している。 $\psi$  のルジャンドル変換  $\psi^*$  を用いれば、鏡像降下法の反復式4.13の最適化を陽に解ける [65]。また、情報幾何学においては双対平坦性という幾何学的な構造を導く [62]。

以下では有限次元の非負測度の空間  $X$  を想定する。すなわち、 $x \in X$  は  $x_i \geq 0 (i = 1, \dots, n)$  を用いて  $x = (x_1, x_2, \dots, x_n)$  と表される。

$f$ -ダイバージェンス [69] とは  $f(1) = 0, f'(1) = 0$  を満たす微分可能な凸関数  $f : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$  を用いて

$$D_f(x||y) := \sum_i x_i f\left(\frac{y_i}{x_i}\right) \quad (\text{B.4})$$

と定義される。 $f(u)$  の  $u = 1$  における微分係数  $f''(1)$  の大きさはダイバージェンスの値の定数倍に相当するため、標準的には  $f''(1) = 1$  を満たすような  $f$  が用いられる。

$f$ -ダイバージェンスは情報単調性と呼ばれる性質をもち、座標  $i$  ごとの和に分解された形をしている (分解可能性)。逆にこの 2 つの性質をもつ任意のダイバージェンスは  $f$ -ダイバージェンスの形でかける [68]。

全ての  $f$ -ダイバージェンスは局所的には式B.1のように同じテーラー展開をもち

$$D_f(x||x + dx) = \frac{1}{2} \sum_i \frac{1}{x_i} (dx_i)^2 \quad (\text{B.5})$$

となる。つまり  $g_{ij} = x_i^{-1} \delta_{ij}$  である。 $X$  が確率分布の空間なら、 $g_{ij}$  は Fisher 情報行列 (式4.15) と一致する。

$f$ -ダイバージェンスの中でも、 $\alpha$ -ダイバージェンス [68], [140] は関数  $f$  としてスカラーパラメータ  $\alpha$  を用いた以下の形を用いるもののことである:

$$f_\alpha(u) = \begin{cases} \frac{4}{1-\alpha^2} \left(1 - u^{\frac{1+\alpha}{2}}\right) - \frac{2}{1-\alpha}(u-1) & (\alpha \neq \pm 1) \\ u \log u - (u-1) & (\alpha = +1) \\ -\log u + (u-1) & (\alpha = -1). \end{cases} \quad (\text{B.6})$$

すなわち,  $\alpha$ -ダイバージェンスは

$$D_\alpha(x||y) = \begin{cases} \frac{4}{1-\alpha^2} \sum_k \left( \frac{1-\alpha}{2} x_k + \frac{1+\alpha}{2} y_k - x_k^{\frac{1+\alpha}{2}} y_k^{\frac{1-\alpha}{2}} \right) & (\alpha \neq \pm 1) \\ \sum_k \left( x_k - y_k + y_k \log \frac{y_k}{x_k} \right) & (\alpha = +1) \\ \sum_k \left( y_k - x_k + x_k \log \frac{x_k}{y_k} \right) & (\alpha = -1) \end{cases} \quad (\text{B.7})$$

である.

$\alpha$ -ダイバージェンスは Bregman ダイバージェンスでもあることが示せるので,  $\alpha$ -ダイバージェンスは Bregman ダイバージェンスであり  $f$ -ダイバージェンスである. 逆に, Bregman ダイバージェンスであり  $f$ -ダイバージェンスであるようなダイバージェンスは  $\alpha$ -ダイバージェンスであることが知られている [68].

$\alpha$ -ダイバージェンスの例として, Kullback-Leibler(KL) ダイバージェンスや Hellinger 距離 (の二乗) がある. KL ダイバージェンスは

$$D_{KL}(x||y) = \sum_i x_i \log \left( \frac{x_i}{y_i} \right) \quad (\text{B.8})$$

と定義され,  $\alpha = +1$  の  $\alpha$ -ダイバージェンス  $D_{KL} = D_{+1}$  である. 引数の順番を逆にした  $D_{KL}(y||x)$  は  $\alpha = -1$  の  $D_{-1}(x||y)$  に相当する.

また,  $\alpha = 0$  は Hellinger 距離の二乗の定数倍

$$D_0(x||y) = 2 \sum_k (\sqrt{x_k} - \sqrt{y_k})^2 = 4 \text{Hel}(x, y)^2 \quad (\text{B.9})$$

を表す. Hellinger 距離 (の二乗) は,  $\alpha$ -ダイバージェンスの中で距離の公理を満たす唯一のダイバージェンスである点で特徴的である. さらに, リーマン多様体で自然に定義される距離 (式4.8) は Hellinger 距離と定数倍を除いて一致する [141], [142].