

新版への序

AI が実用化されるにつれ、情報の分野でもこれに対応した理論の基礎が求められるようになってきた。情報幾何はその一つの方法を提供するもので、ここに至って AI、信号処理、画像工学、脳科学、さらに物理学と、多くの分野で注目を浴びるようになっていく。本書の発刊および少し遅れた Springer からの英文の出版に留まらず、多くの情報幾何の解説書が見受けられる。また Springer 社から、統計数理研究所の江口真透教授を編集長とする国際学術誌「Information Geometry」が昨年より刊行され、研究者が世界的に増加している。

本書も刊行以来 5 年を経過し、その後の発展を少しでも取り入れたいと思った。旧版での誤りを訂正するとともに、内容の一部を削除し、その代わり新しく深層学習と Wasserstein 距離にかかわる 2 章を書き下した。

深層学習はコンピュータの性能の向上と大規模データベースに助けられ、怒涛の如く発展している。まさに驚異である。しかしそれは 1950 年代に始まる Rosenblatt のパーセプトロンの構想の実現であり、その理論的な基礎は確率勾配降下学習法と普遍関数近似性能に頼っていた。もちろん、その他のいろいろな工夫、発明はあり、驚くほどの性能を有するが、しかしどちらかと言えば腕力に頼ったものであった。

しかるに近年、変化が現れ始めた。一つは、かつて統計的機械学習の花形であった訓練誤差と汎化誤差の関係である。サンプル数 N に対してデータを説明するのに使うモデルのパラメータ数 P を増やせば、訓練誤差は単調に減少する。しかしあるところを過ぎると、訓練誤差は減るものの、汎化誤差はかえって増えてしまい、 P の大きいモデルは使いものにならない。また、大規模の非線形モデルでは、極小解が多数あって、確率勾配降下法はそのどれか一つに捉われてしまうから、うまく行かない。これが従来の常識であった。

深層学習はこの予想を覆して、やってみればうまく行くという成果を示した。それならばそれを示す理論的な根拠が必要である。近年、 P が N 以上に十分大きいときの挙動に関する理論が出始めた。このとき極小解問題は生ぜず、 P が十分に大きければ、訓練誤差は 0 に減ること、また、 N を超えて P を大きくすれば、汎化誤差が再び減少に転じることなどである。

ランダムに結合した回路網を初期値として学習を開始したとき、 P が十分に大きければ最適解はそのごく近くにあるという、驚くべき発見があった。これは neural tangent kernel という構想で発表され、世界に広まっている。ランダム回路は確率分布にも関係し、統計神経力学として私が研究してきたもので、情報幾何とも関係が深い。さらに、深層回路網の多様体の Fisher 情報計量も計算できる。自然勾配学習法がうまく実用化される。こんな思いで、書籍にするのはまだ早いこの話題をあえて取り上げて第 15 章で紹介した。これからの発展を見守りたい。

Wasserstein 輸送問題は、フランスの数学者 Monge に始まる 250 年もの歴史を持つ。二つの確率分布間の距離を、一つの分布を他の分布に輸送するコストで定義する。これは線形計画問題になり、ロシアの Kantorovich が整備して、ノーベル経済学賞を受賞した。ある空間上での分布の輸送には距離が関係するから、これは大変自然な分布間の距離である。

しかし、これは情報幾何の意味で不変ではない。分布のもととなる空間にどのような変換を施しても、分布間の距離は不変であることを情報幾何は要請するからである。しかし、画像などは2次元空間上の分布とみなせるから、画像間の距離は元の空間の距離構造に深く関係している。つまり、不変性の要請はここでは有害である。

Villani らの活躍があって、Wasserstein 距離が数学の世界で深く考究される一方、画像処理、AI など、多くの応用の分野で大きな話題になっている。そこで問われるのが、情報幾何との関係である。

もちろん両者は別物で、それぞれに特徴がある。しかしその関係を探求することは急務であろう。本改訂版の第 16 章では、Wasserstein 距離の問題を、情報幾何の手法を用いて明らかにする試み考えた。これも建設途上の理論ではあるが、基礎、応用を含めてこの分野に新しい方向を示すものと自負している。

新たに付け加えた二つの章は、私が引退してから、自分の趣味として続けている研究の一端を紹介したものである。見苦しい点はご容赦願いたい。

2019 年 9 月

甘利 俊一

序

情報幾何はその名のとおりに、情報の豊かな構造を幾何学により表現する方法である。情報科学では、論理、代数、解析、確率など、多彩な方法が使われているが、幾何学は情報要素の相互の関係を明らかにするのに有用である。情報幾何は、確率分布を情報要素とみなし、統計的な推論の仕組みを明らかにする理論として登場したが、いまは機械学習、信号処理、神経回路網など、広い分野で注目され、大きな話題になってきた。

情報幾何は現代の微分幾何を用いて構築されたため、取り付きにくいきらいがあった。しかし、その内容は直感的に理解できるものである。そこで、微分幾何に無用に深入りすることなく、その本質を直接的に紹介するとともに、広い応用分野を眺望する解説が必要であると痛感するに至った。このため「数理科学」誌に 30 回にわたる連載を試み、それを整理して大幅に改訂したものが本書である。

情報幾何を数学としてみた場合、それは従来のリーマン空間に 3 次の対称テンソルを付け加え、双対接続という新しい概念を導入したものである。これは数学としても発展しつつあるし、関数空間での情報幾何などいまだに未解決の部分もある。しかし、本書では数学的な厳密性には一切こだわらず、直感的な理解で話を進めようと試みた。

本書は 4 部構成である。第 I 部は情報の空間（多様体）に 2 点間の分離の度合いを表すダイバージェンスを導入し、そこから得られる豊かな幾何構造を調べた。ダイバージェンスとは、2 点間の非対称な関数で、ここから双対構造が生み出される。第 II 部では、これを数学的に定式化するために、微分幾何が必要になる場合に備えて、微分幾何入門を試みた。これまでの微分幾何の教科書は数学者が数学者のために書いたものである。ここでは、使う人のために厳密性にこだわらず直感的な入門を旨としたが、読者はこの部分を読み流して構わない。

第 III 部は、統計的な推論にあたって情報幾何が基本的な役割を果たすことを明らかにし、多くの統計的な手法を導入した。第 IV 部は、種々の分野への応用である。ここでは、機械学習、画像処理、信号処理、神経情報、最適化など、いま“熱い”話題を扱っている。本書は、雑誌で連載したという性格もあって、各部分が比較的独立に読めるようになっている。わからない部分や興味の少ない部分は読み飛ばして先に進んで一向に差支えない。

情報幾何の建設は国内外の多くの研究者と共同で進めたものである。名前こそ挙げないが、この方々に深く感謝したい。また、「数理科学」誌連載と本書をまとめる過程で、大溝良平、伊崎修通両氏には多くの助言と協力をいただいた。忘れてならないのは、私の汚い手書き原稿を整理し、忍耐強く $\text{\LaTeX}$ 原稿に仕立ててくれた浪岡恵美さんである。彼女の献身がなければ、本書は完成しなかったであろう。

2014 年 6 月

甘利 俊一

目 次

第 I 部	多様体とダイバージェンス関数	1
第 1 章	多様体とダイバージェンス関数	2
1.1	工学に現れる空間：多様体	2
1.2	多様体上のダイバージェンス関数	6
第 2 章	凸関数の導くダイバージェンスと双対平坦構造	9
2.1	凸関数とダイバージェンス	9
2.2	Legendre 変換と双対性	11
2.3	ダイバージェンスとリーマン幾何：接空間	12
2.4	凸関数と双対平坦多様体	15
2.5	拡張ピタゴラスの定理	16
2.6	拡張射影定理	18
第 3 章	指数型分布族の双対平坦構造	20
3.1	指数型分布族	20
3.2	指数型分布族の凸関数と双対平坦構造	22
3.3	e -平坦と m -平坦	23
3.4	指数型分布族, 凸関数, Bregman ダイバージェンス	24
3.5	指数型分布族の拡大	24
3.6	射影定理とピタゴラスの定理の応用	25
第 4 章	確率分布族における不変なダイバージェンス	29
4.1	不変性	29
4.2	粗視化と単調性	30
4.3	標準 f 関数と Fisher 情報行列	32
4.4	いろいろな f ダイバージェンス	33
4.5	正測度空間における f ダイバージェンス	34
4.6	平坦で不変なダイバージェンス：KL ダイバージェンスと α ダイバージェンス	36
4.7	正測度空間で双対平坦性を与える f ダイバージェンス	37
4.8	KL ダイバージェンスの性質	39
第 5 章	確率分布族, 正測度族, 正定値行列空間に導入する非不変な双対平坦構造	43
5.1	正測度空間における (u, v) ダイバージェンス	43
5.2	正定値行列空間の不変なダイバージェンス	45

5.3	q 自由エネルギーが導く確率分布族空間の双対平坦構造	47
第 II 部 微分幾何学入門		51
第 6 章	アファイン接続, 共変微分, 測地線	52
6.1	接空間と計量	52
6.2	座標変換とテンソル	53
6.3	アファイン接続	55
6.4	共変微分	56
6.5	ベクトルの平行移動	57
6.6	測地線	58
6.7	リーマン接続	59
第 7 章	曲率と振率	62
7.1	リーマン曲率とベクトルの世界一周	62
7.2	振率と世界一周	65
7.3	部分空間の埋め込み曲率	66
第 8 章	双対接続の幾何	69
8.1	双対接続の導入	69
8.2	双対平坦なリーマン空間	71
8.3	α 接続と α 幾何学	73
8.4	双対平坦空間のポテンシャル関数と規範ダイバージェンス	73
8.5	ダイバージェンスと双対幾何	75
第 9 章	階層構造を持つ双対平坦空間	77
9.1	階層双対平坦空間	77
9.2	混合座標系と直交葉層化	79
9.3	神経スパイクの情報幾何	81
9.4	高次相関は情報を担うか	83
9.5	産業連関表の情報幾何	85
第 III 部 統計的推論の情報幾何		89
第 10 章	統計的推論と情報幾何：曲指数型分布族を用いて	90
10.1	統計推論の幾何学的な枠組み—例題を用いた直観的説明	90
10.2	Fisher 情報量と Crámer-Rao の定理	93
10.3	指数型分布族における統計的推論	94
10.4	曲指数型分布族における推論	95
10.5	統計的推定の漸近理論	99
10.6	指数型分布族における検定と分類	102
10.7	検定の漸近理論	103
第 11 章	Neyman-Scott 問題：局外母数とセミパラメトリック統計モデル	108
11.1	局外母数を含む統計モデル	108
11.2	Neyman-Scott 問題	110
11.3	セミパラメトリック統計モデルと推定関数	113

11.4	推定関数の情報幾何	115
11.5	Neyman–Scott 問題の解	118
11.6	線形関係の推定：その解	120
11.7	秤の問題の解	121
11.8	神経スパイクの解	122
第 12 章	隠れ変数のあるモデル：em と EM アルゴリズム，非忠実なモデル，Bayes 統計	125
12.1	隠れ変数のある統計モデル	125
12.2	em アルゴリズム	126
12.3	EM アルゴリズム	127
12.4	非忠実なモデルによる推論	130
12.5	変分 Bayes 法	131
12.6	Bayes 統計の情報幾何—deep learning に向けて	132
第 IV 部	情報幾何の様々な応用	135
第 13 章	機械学習の情報幾何	136
13.1	画像の分類，検索，認識：クラスタリングの情報幾何	136
13.2	サポートベクトル機械	144
13.3	確率推論，グラフィカルモデル，信念伝播	150
第 14 章	学習の力学と特異点：多層パーセプトロンと自然勾配学習法	161
14.1	学習機械：多層パーセプトロン—その栄枯盛衰	161
14.2	多層パーセプトロンの空間の特異構造	165
14.3	確率降下法と自然勾配学習法	171
14.4	多層パーセプトロンの学習力学	175
14.5	Bayes 推論と特異点：代数幾何による解析	178
第 15 章	深層学習の発展と統計神経力学	182
15.1	ランダム結合の深層回路	182
15.2	学習とカーネル	183
15.3	接神経核（NTK）理論	184
15.4	深層学習にかかわる驚異の発見	185
15.5	ランダム回路網の信号伝播	186
15.6	Fisher 情報量	189
15.7	Fisher 情報行列の固有値の分布	190
第 16 章	Wasserstein 距離の情報幾何	193
16.1	Wasserstein 距離	193
16.2	エントロピー制約を付けた Wasserstein 距離	194
16.3	最適輸送計画の多様体	195
16.4	エントロピー制約が与えるダイバージェンス	196
16.5	パターンの集まりの中心：形と位置の分離定理	197
第 17 章	信号処理と最適化の情報幾何	200
17.1	主成分分析	200
17.2	独立成分分析	204

17.3	非負行列分解	208
17.4	スパース解と圧縮測定 (compressed sensing)	210
17.5	凸計画法の情報幾何	216
あとがき		222
索引		223