

まえがき

我々は人生においてベストを尽くしたい。最高の性能を発揮する装置を作りたい。環境の汚染を最小に留めたい。社会的な問題を解決するためにいちばんよい答を選びたい。このように我々は、何かを最大あるいは最小にしたいという要求をいつも持っている。さらに、自然が今あるように観測されるのは、その状態（あるいは過程）が最も出現頻度の高い状態（あるいは過程）であるからかもしれない。例えば、熱平衡の条件下で最も高い確率で現れるのは、自由エネルギー最小の状態である。古典力学の運動は作用を最小にする過程として実現する。つまり、自然を理解する基礎研究においても、何かの条件を最もよく満たすものを見つけることは、避けて通ることのできない大事な問題である。

こうしたベストなものを見つける問題が直観的に解けるときはよいが（日常の多くの場面では、我々はそうして生活しているが）、条件が複雑になるとそういうわけにはいかない。そのときは問題を数量化して、計算によって最適な解を探ることが必要となるであろう。すなわち、システムがとる N 個の変数を x_1, x_2, \dots, x_N 、あるいはそれらをまとめて N 次元のベクトル \mathbf{x} と書くとする、ある評価関数、例えば $F(\mathbf{x})$ の最適解、つまり F を最大あるいは最小にする \mathbf{x} を見つける問題を解けばよい。例えば、 $F(\mathbf{x})$ が利潤や適応度、エントロピーなら最大に、 $F(\mathbf{x})$ がコストやエネルギー、自由エネルギーなら最小にする \mathbf{x} を見つける問題を解くことになるが、最大か最小かは評価関数の符号の付け方の違いにすぎず、計算の方法としては区別する必要はない。本書はこうした最適解を見つける計算法について、基礎から最前線に至るまで、系統的なわかりやすい解説を目標として書かれた本である。最適化の計算法は個々の問題の性質を超えた一般的な方法であるため、本講座では、多くの分野を横断す

る教科書として本書を位置づけ、広い分野の方々に読んでいただくことを企図している。

関数 $F(\mathbf{x})$ の性質が複雑になると、最適解を見つけることは簡単ではなくなる。例えば、解が満たさなければならない条件が複数あるときを考えよう。それぞれの条件の達成度を \mathbf{x} の関数として表現し、評価関数 F を達成度を表す全ての関数の和と定義する。ある解ではある条件をよく満たすが別の条件を満たさない、別の解ではその他の条件をよく満たすが、やはり全ての条件を満たすわけではない、つまり、あちら立てればこちら立たずの状況は、我々のよく出会う問題である。こうしたことが起こると、 \mathbf{x} を変化させるとともに F の値は上下を繰り返し、 F はたいへんでこぼこの激しい関数になる。物理学ではこうした問題のことを、フラストレーションのある問題（フラストレートした問題）と呼ぶことが多い。こうした、でこぼこの評価関数の問題、あるいはフラストレートした問題は $N \approx 3, 4$ のうちは試行錯誤で容易に解くことができるが、 N が大きくなるにつれて急速に難しくなる。多くの場合、評価関数のでこぼこ（つまり局所的な極小、極大）の数は N の指数関数、あるいはそれ以上の速さで増大するので、その中から最適解を見つけるために網羅的な探索をすると、探索に必要な計算ステップ数も指数関数的に増大して、現実的な時間では実行不可能になるからである。このような問題は NP 困難な問題^{†1}と呼ばれ、より効率的に最適解を計算する方法を求めて離散数学の分野で長く研究されてきた。しかし実際には、厳密な最適解を求めなくても、充分最適に近い解であれば、たいへん有用であることが多い。そこで、こうした実際的な解を数値的に探す方法が重要となる。本書は大きな N の系、すなわち、超多自由度系の最適解を探す計算科学をテーマとしている。

自然の現象では、自然は厳密解を数学的に導く計算を実行しているわけではなく、多くの変数の間の相互作用の結果、実際的な解が試行錯誤の末に選ばれる過程が実現している。したがって、自然に学ぶことによって、良い解を効率よく計算する新しい方法を発見することができるのではないだろうか？ 本書では、でこぼこの評価関数の問題、あるいはフラストレートした問題を解くた

^{†1} Non-deterministic Polynomial (NP) time の難しさを持つ問題。

めに、自然に学ぶことによって発展した二つの方法、進化的計算の方法と拡張アンサンブルの方法について解説する。

第1章では進化的計算について解説がなされている。1.1節では遺伝的アルゴリズム (genetic algorithm: GA) の基礎が解説されている。進化的計算は生物の進化過程を工学的に模倣した最適化のアルゴリズムである。進化的計算の最大の特徴は、評価関数 $F(\mathbf{x})$ がでこぼこの多い NP 困難な問題に対して、現実的な計算時間内に実際的な解を探し出そうとするところにある。進化的計算の基本はランダム探索であり、探索の効率化を図る工夫に応じて、GA の他にも進化的プログラミング、進化戦略、遺伝的プログラミング、蟻コロニー最適化、粒子群最適化などが提唱されている。この中で GA は進化的計算の代表と言ってよいであろう。GA は地球上の生物が自然選択と DNA レベルでの交叉と突然変異により進化したとする進化論に基づいている。今日の地球上に見られる生物の多様性を作り出している進化のメカニズムをアルゴリズム化することで、効率的な解探索を実現できると期待されている。

GA の基本的演算は初期世代の個体群を生成した後に、選択・淘汰と増殖・交叉・突然変異により後継世代を生成するものである。交叉演算は雌雄交配における遺伝子の交叉の模倣であり、突然変異演算は遺伝子複製時の突然変異の模倣である。いずれの演算も確率に基づいて適用される。評価値の高い個体を選択されて次世代にコピーを残す確率が高められている。GA は、多数の個体によるランダムな多点探索である。単なるランダム探索と異なるのは、探索点における情報を交叉により交換して探索の効率化が図られている点にある。でこぼこの多い評価関数空間に多くの個体がランダムにばらまかれ、その中で比較的評価値の良い個体を選択・淘汰で生き残る。生き残った個体どうしが交叉・突然変異により子孫を自らの近傍、もしくは全く離れた新しい場所に生成することで、実際的な解を効率よく探索する。

1.2節では進化型多目的最適化が解説されている。現実世界の多くの問題では、複数の目的を同時に考慮しなければならないことが多い。しかしこのとき、あちらを立てればこちらが立たずというようなトレードオフの関係が存在して、複数の目的を同時に達成することが難しいことがよくある。こうした状況では、複数の目的への達成度を足し算した一つの評価関数を扱うのではな

く、複数の評価関数をあらわに扱う計算法がユーザにとって有用であろう。すなわち、 \mathbf{x} の評価関数を複数扱う多目的最適化問題として、さまざまな妥協解を呈示することで、ユーザの意思決定を支援することが効果的である。進化的計算は多様な妥協解生成に有効である。1.2 節では、進化型多目的最適化問題における重要な概念である非劣解とパレート最適解の厳密な定義が示されている。すなわち、「多目的最適化問題の全ての実行可能解の中で他の解に優越されない非劣解をパレート最適解と呼び、任意の解集合の中で他の解に優越されない解をその解集合における非劣解と呼ぶ」として、両者が明確に区別されている。その上で、パレート最適解集合の効率的な探索が重要課題と位置づけられ、パレート最適解集合への個体群の収束の速さと多様性の維持に関する簡潔明快な計算例が紹介されている。さらには、目的関数を多数持つ「進化型多目的最適化」の新しい研究動向とそこで語られているホットな話題が紹介されている。なお、本節では進化型多目的最適化アルゴリズムの発展の経緯も詳説されているので、読者はこの分野の全容を知ることができる。

1.3 節では並列進化的計算について解説されている。近年の PC クラスタなどの並列計算機の普及とグリッドコンピューティング、クラウドなどの並列計算環境の発展に伴い、多点探索の並列実行による進化的計算の超高速化が現実のものとなりつつある。本節で紹介されている実験例では、徳島大、東工大、産総研など複数のサイトにある PC クラスタをインターネットで相互接続し、1,196CPU からなるグリッド計算環境において、1 台の CPU では約 200 日かかる NMR 立体構造決定問題を約 2 時間 40 分で終了させることに成功している。

本節は、まず、進化的計算の並列化において、分散計算モデルと、並列実行(実装)モデルが明確に区別されなければならないとの指摘から始まっている。計算としての並列化と、ハードウェアに実装する際の並列化を混同してはならないとした上で、分散計算モデルの代表としてマスタースレーブモデル、分割母集団モデル、セルラーモデル、P2P モデルの概要が述べられている。

本節の特徴の一つは並列実行モデルの効果に関する基礎研究の成果の紹介にある。マスタースレーブモデルによる並列遺伝的アルゴリズム(マスタースレーブ GA)の並列実行においては、GA の遺伝的演算(選択・淘汰、交叉・突

然変異)の並列実行は効果がないとしている。効果が大きいのは、GAの個体の評価に時間がかかるような複雑・大規模な問題である。このとき、スレーブごとに逐次GAを割り振るマスタースレーブGAが効果を発揮する。スレーブの数が増えるにつれて、1台当たりの評価計算時間の短縮と、通信頻度の増加にトレードオフの関係が生じる。本節ではGAの実行時間を最小化するスレーブ数が導出されている。

本節のもう一つの特徴は、グリッド計算環境における並列実行モデルであるグリッド向けMGG (minimum generation gap) システムとその実験結果の紹介である。グリッド環境を活かすGAの課題が明らかにされ、解決策の一つとして世代交代モデル(MGG)を用いたシステムによる並列実行の効果が示されている。

本節の最後では進化的多目的最適化のための並列実行モデルが解説されている。多目的GAの並列実行モデルの研究はごく最近始められたばかりであり、この節の著者らによる提案手法が先駆的である。近傍交叉と計算資源の性能に応じた増殖個体数の適応的な変化手法の効果が数値実験により示されている。

第2章では、扱うシステムがカノニカル分布に従うときに使用できる、非常に強力な最適化計算法が解説される。すなわち、システムが状態 \mathbf{x} にあるときのコストを $E(\mathbf{x})$ と書き、状態 \mathbf{x} の出現確率を $P(\mathbf{x})$ ^{†2}としたとき、 $P(\mathbf{x}) \propto \exp(-\beta E(\mathbf{x}))$ と書ける場合に使う方法である。ここで β は適当な定数であるが、物理的な問題のときは温度エネルギーの逆数 $\beta = 1/k_B T$ の意味を持ち、 $E(\mathbf{x})$ はエネルギーである。もちろん、このカノニカル分布は、温度やエネルギーが定義されるようなあらゆるシステムで成り立っている分布であるから、その適用範囲は極めて広い。この方法は、カノニカル分布で表される自然の熱揺らぎをうまくコントロールしたシミュレーションを計算機の中で行い、最適解を含む広い解を探索する方法である。

さて、システムがカノニカル分布に従うとき、システムを記述する変数を、ゆっくり変化する変数 \mathbf{r} と速く変化する変数 \mathbf{s} に分けることができ、 $E(\mathbf{x}) = E(\mathbf{r}, \mathbf{s})$ と書けると考えよう。例えば蛋白質の問題なら、蛋白質の周

^{†2} \mathbf{x} が連続変数の場合、状態が \mathbf{x} と $\mathbf{x} + d\mathbf{x}$ の間にあるときの確率密度。

りの溶媒分子の座標は 10^{-12} 秒以下で動く速い変数 \mathbf{s} 、蛋白質の構造を表す座標は 10^{-9} 秒以上で動く遅い変数 \mathbf{r} と考えてよい。たいていの場合、大事な変数は遅い変数 \mathbf{r} であり、速い変数は遅い変数に影響を与える環境の変数であるため、我々は遅い変数に注目したい。このとき、 \mathbf{r} の確率密度 $P(\mathbf{r})$ は、 $P(\mathbf{r}) \propto \int d\mathbf{s} \exp(-\beta E(\mathbf{r}, \mathbf{s}))$ と、先に \mathbf{s} を積分して求めることができるであろう。ただし、指数関数は変化が激しくて扱いづらいので、対数をとって自由エネルギー $F(\mathbf{r})$ を $F(\mathbf{r}) = -(1/\beta) \log\{\int d\mathbf{s} \exp(-\beta E(\mathbf{r}, \mathbf{s}))\}$ と定義すると、最も高い確率で出現する \mathbf{r} を求める問題は、自由エネルギー $F(\mathbf{r})$ を最小にする \mathbf{r} を求める問題となる。このように、カノニカル分布の問題で最適解を求める問題は極めて重要な問題である。ただしここで、システムの変数を \mathbf{r} と \mathbf{s} に分けたのは、概念的な説明のために導入した技巧であり、第2章で紹介する実際のシミュレーションでは使われていないことに注意されたい。すなわち、変数を \mathbf{r} と \mathbf{s} を分けるという余分な仮定なしで、モンテカルロ法あるいは分子動力学法によって \mathbf{r} と \mathbf{s} を区別しないシミュレーションを行い、その結果の分析の際に、 \mathbf{r} に注目して図を作ったり、平均量を計算したりしているわけである。

蛋白質の場合、さまざまな引力相互作用や斥力相互作用が構成要素の間に働いているため、あちら立てればこちら立たず、すなわち、全ての相互作用エネルギーを小さくする構造 \mathbf{r} を見つけることができない。つまり、蛋白質はフラストレートした系であり、 $F(\mathbf{r})$ は非常にでこぼこした関数となる。 F を最小にする \mathbf{r} 、つまり自然の中で実現している \mathbf{r} を見つけることは生物学、医学、薬学などにとって非常に重要な問題であるため、優れた最適化計算法の登場が強く望まれてきた。

これまで、フラストレーションのない単純なシステムでカノニカル分布を再現できるシミュレーションは広く利用されてきたが、フラストレーションのある場合にこうした方法を用いても、計算は局所解に捕まってカノニカル分布を再現できず、最適解を発見することは難しい。第2章では、カノニカル分布の考え方を意図的に変更して局所解に捕まらない計算を実行し、計算の後に変更分を補正してもとのカノニカル分布に戻るマルチカノニカル法など、こうした困難を克服する拡張アンサンブル法について解説する。これらの方法を用いれば、最適解の \mathbf{r} を発見するに留まらずに、自然のシステムが最適解を外れて分

布する程度，すなわち，システムの揺らぎや熱的性質についても多くの情報を計算することができる。また，こうした方法は蛋白質の問題に留まらずに，温度が定義されているあらゆる問題に適用することができる。実際，拡張アンサンブルの源流となる方法は，スピン系などのシミュレーションのために物理学者が考案したものであるが，第2章の著者である岡本氏はその方法を世界で初めて生体分子に適用し，標準的な方法に成長させ，その上さらに重要な発展の方向を提案している研究者であり，本章の解説も基礎から最新の発展まで，創始者ならではの本質をついた解説となった。もちろん第2章で解説されている方法は，生体分子に限らず，カノニカル分布に従うあらゆる物理，化学システムに適用可能である。

さらに第2章について，興味のある読者のためにコメントを加えておきたい。熱平衡が定義できるシステムではカノニカル分布が定義できるが，カノニカル分布は熱平衡よりさらに広い概念である。 $P(\mathbf{x})$ の汎関数としてシャノンエントロピー， $S(\{P(\mathbf{x})\}) = -\int d\mathbf{x}P(\mathbf{x})\log P(\mathbf{x})$ を定義すると，このエントロピーは，システムの状態分布に如何に偏りが無いかという偏りの無さの程度を表す指標である。ここで， $E(\mathbf{x})$ の平均値 $\langle E \rangle$ が指定されている以外に状態の分布に偏りを生じさせる原因が無ければ，その条件下でシャノンエントロピーは最大になるべきである。 $P(\mathbf{x})$ の規格化条件と合わせて，二つの条件を二つのラグランジュ未定乗数 β と λ で表すと， $S' = S - \lambda(\int d\mathbf{x}P(\mathbf{x}) - 1) - \beta(\int d\mathbf{x}P(\mathbf{x})E(\mathbf{x}) - \langle E \rangle)$ を最大にする関数 $P(\mathbf{x})$ を求めればよい。その答がカノニカル分布である。熱平衡に関するシステムでは， β は温度エネルギーの逆数， $E(\mathbf{x})$ はエネルギーという意味づけができるが，熱平衡と関係ないシステムでは他の意味づけがあってもよい。そうしたより一般的なシステムでも，もし $P(\mathbf{x})$ を再現する計算を定義できるなら，第2章で紹介される拡張アンサンブル法は有力な方法であろう。もちろん，そうした一般化は今のところ編者の空想に過ぎないが，分野を超えた応用ができればおもしろい問題である。

こうして本書では，進化的計算と拡張アンサンブルという二つの独自に発展してきた方法を一つの本の中で紹介する。これは，さまざまな分野の学生や専門の方々が本書を手取ることで，新しい展開が始まればよいという期待のも

とに始まった企画である。それぞれのテーマについての体系的な教科書として利用していただくとともに、異なる発想が少しでもミックスされる機会となることを願って、本書をスタートさせたい。

2013年4月

編者 古橋 武
 笹井理生