書　評

タンパク質構造とトポロジー
—パーシステントホモロジー群入門—

平岡裕章　著，共立出版，2013 年

北海道大学大学院理学研究院
荒井　迅

本書は，パーシステントホモロジー理論について初めて日本語で書かれた書籍である．パーシステントホモロジーは，生命科学や物質科学，情報処理などの幅広い分野のデータを解析するためのツールとして近年急速に広まっており，その基礎理論をコンパクトにまとめた本書の意義は大きい．パーシステントホモロジーは代数トポロジーの技法を用いて構築されるが，本書は代数トポロジーの予備知識がなくても読めるように配慮されている．なお，タイトルにあるようにタンパク質の構造解析への応用がゴールとして設定されているが，タンパク質はあくまでも応用の一例であり，タンパク質への興味や知識がなくても面白く読める．ただし，「何がしたいのか」という動機があったほうがストーリーがわかりやすいので，タンパク質，もしくは各自のお気に入りの幾何的な構造を念頭に読み進めると良いと思う．

著者の平岡裕章氏は日本におけるパーシステントホモロジーの第一人者である．また，過去には符号理論や力学系など様々な分野を渡り歩きながら活躍してきた開拓者でもある．応用数学の将来を期待された研究者のひとりであるといえよう．ちなみに彼の名前の読みは「やすあき」であり「ひろあき」ではない．英語で話すときは “Yasu” と呼ばれているので，そちらで覚えると忘れにくい．

本書で印象的なのは，代数的な計算を具体的かつ丁寧に追いかけているという点である．このあたりは，様々なバックグラウンドを持つ研究者にパーシステントホモロジーを伝道してきた著者の経験が生きているものと思われる．もちろん，最終的に計算機で動くアルゴリズムとして実装するという目的のためには，行列演算を具体的に書き下さなくてはならないという側面もある．

本書は 3 つの章からなり，第 1 章が「単体複体」，第 2 章が「ホモロジー群」，第 3 章が「パーシステントホモロジー群」である．最初の 2 つの章だけであれば，通常のホモロジー群の入門書として読むこともできる．

第 1 章で扱う単体複体は多面体の高次元化であり，トポロジー全般において基礎的な概念であるが，計算トポロジーにおいてはより重要な役割を果たす．それは単体複体が完全に離

散な組合せ的データ構造であり，特異チェイン複体などの無限次元を経由するデータよりも計算機と相性が良いことから来ている．また本章では，空間内の点列として与えらえた入力データ（例えばタンパク質を構成する原子の座標）をどのように単体複体として表現するかも議論している．そのために導入されるのがヴィートリス・リップス複体とアルファ複体である．パーシステントホモロジーのデータ解析における強みのひとつに，トポロジー的な情報だけでなく幾何学的な情報も表現できる点があるが，それはこれらの単体複体（とそこに入るフィルトレーションの構造）が幾何学的な情報を取り込んでいるから可能になるのである．

第 2 章はホモロジー群への入門である．アーベル群の定義から始まり，ホモロジー群やその上に誘導される準同型の定義までを解説する．標準的なホモロジーの教科書とほぼ同様の内容を辿るものになっているが，特徴的なのはアルゴリズム的な視点が強調されていることである．例えば有限生成 $\mathbb{Z}$ 加群の構造定理においても，単因子を求めるためのスミス標準形はその計算方法がフローチャートで与えられている（図 2.2）．

第 3 章では本書の主題であるパーシステントホモロジー群に入門する．問題設定は単体複体 K に部分複体 K^t によるフィルトレーション

$$\emptyset = K^0 \subset K^1 \subset \cdots \subset K^n = K$$

が入っているような状況である．タンパク質の原子の座標データの場合であれば，「どのくらいの距離にある原子を繋っているとみなすか」を決めるパラメータを t として，t を大きくするとどんどん繋がっていく原子たちのなすネットワーク構造の変化がフィルトレーションで表現されている．また，他にも時間とともに空間が増大する（繋がっていく）ようなデータの表現と思うこともできる．このとき，$C_k(K^t)$ の直和 $\bigoplus_{0\leq t} C_k(K^t)$ において K^t の元を K^{t+1} に埋め込む操作を x と書くと，$\bigoplus_{0\leq t} C_k(K^t)$ に $\mathbb{Z}_2[x]$ 加群構造が入る．これらを次数 k についてまとめた次数付き $\mathbb{Z}_2[x]$ 加群のホモロジーをとったものがパーシステントホモロジー群である．なぜ $\mathbb{Z}[x]$ でなく $\mathbb{Z}_2[x]$ なのか不思議に思われるかも知れないが，第 3 章の補足にあるように，係数が単項イデアル整域にならないとパーシステント区間などの定義がうまくいかず，有用な情報が引出せないのである．

本書では代数的簡明さから上のような多項式環を用いた定義を採用しているが，後で挙げる Edelsbrunner らの本では包含写像 $f^{i,j} : K_i \to K_j$ が $\mathbb{Z}_2$ 係数ホモロジー群

$$0 = H_p(K^0) \to H_p(K^1) \to \cdots \to H_p(K^n) = H_p(K)$$

に誘導する $f_p^{i,j} : H_p(K^i) \to H_p(K^j)$ の像のことをパーシステントホモロジー群と定義している．どちらの定義も本質的に同じものなので心配はいらないが，本書の定義で幾何学的な意味がつかみにくい読者は Edelsbrunner の本も参照するとよいだろう．

本書は全体でも 130 ページとコンパクトにまとまっているが，普通のホモロジー理論や PID 上の加群の構造定理に慣れている読者は，1.4 節でアルファ複体やヴィートリス・リップス複体などのデータ構造を学んだうえで，いきなり核心の第 3 章に読み進むことも可能である．その場合には 40 ページ程度でパーシステントホモロジー理論の基礎定理に辿りつける．

本書でパーシステントホモロジーやトポロジーの応用に興味を持たれた読者に関連してお勧めしたい文献をいくつか挙げておく．まずは，パーシステントホモロジー界の親玉である Herbert Edelsbrunner らの本 “Computational Topology” (AMS, 2010, ISBN: 978-0821849255) と “A Short Course in Computational Geometry and Topology” (Springer, 2014, ISBN: 978-3319059563) である．著者の思い入れが強すぎて少し読み難いところもあるが，ちゃんと考えながら読めば大変面白い．また，応用トポロジー界の大スターである Robert Ghrist による “Elementary Applied Topology” (Createspace, 2014, ISBN: 978-1502880857) も最近出版された．パーシステントホモロジーだけでなくトポロジーの応用に関連する話題を幅広く扱った本であり，Ghrist と平岡氏の共同研究による network coding の話にも触れられている．パーシステントホモロジーの話題からは外れるが，玉木大氏の「広がりゆくトポロジーの世界」(現代数学社，2012) もある．モデル圏のような高度な代数トポロジーの概念が科学全体に与えうるインパクトを議論したエキサイティングな本である．

パーシステントホモロジーは理論も応用も急速に発展しており，新しい展開がどんどん生まれている．そのひとつは，本書第 3 章の補足でも解説されているように，1 次元的に空間が並んだフィルトレーションではなく，より複雑な相互関係で結ばれた空間たちを扱おうというものである．複数のパラメータをもつデータを扱いたいという応用からの要請もあり研究が進んでいるが，表現論的な難しさがあり，直線的なフィルトレーションほど簡明にはいかない．また，統計学との関連も進展している．パーシステントホモロジーは入力データからこれまでになかった数値データを沢山取り出してくれるので，すぐにでも統計をとりたくなるが，パーシステントホモロジーの構造を活かすためには，平均や分散といった統計処理をどのような空間でとるかをよく考えないといけない．これらの話題については，平岡氏による解説が日本数学会編集の「数学」で予定されているとのことなので期待されたい．