

1 はじめに

化学プロセスでは、多くの場合でラボスケールから工 場スケールに至るまで、そのプロセスでは混合物を扱わ ざるを得ない. 医薬品等を含む化合物の合成から大規模 な製鉄や石油精製など基礎から実用展開に至るまで混合 物の取り扱いは必須である.やむを得ず得られた混合物 から必要な化学成分の定性及び定量のために、一般には クロマトグラフィーや分留など様々な分離プロセスを経 たのち, 種々の分析法を用いた測定及びその解析が行わ れる.一方で、分離が困難な場合、やむを得ず混合物の まま分析する必要がある場合も多い. 混合物の分光分析 法によって分析する場合、スペクトルの解析ではリファ レンススペクトルによって定性したり、各成分のピーク を微分スペクトルの極大・極小波長から求め、あらかじ め既知量の試料をもちいて作製した検量線による各成分 の定量が基本である.しかし、成分数が多い場合など、 分析対象となるスペクトルは多成分のスペクトルが重な る複雑な混合物スペクトルとなり、ピークの選別が困難 となるケースは多い. そのため、混合物からなるスペク トルは純成分スペクトルとは異なり、各成分の定性や定 量には限界がある.実際に未知の中間体を含むような反 応系などでは、成分の分離及び特定自体が困難である. そこで、得られた混合物のスペクトルから純スペクトル と濃度情報を直接得ることができれば、時として煩雑で 困難な分離プロセスを経ずに、試料の定性・定量が可能 となり、分析プロセスの簡略化が期待できる.

このように実験で得られる混合物のスペクトルなど 種々の分析データから有益な情報を抽出するために,多 変量解析を分析化学に用いるケモメトリックスが発展し てきた¹⁾.分析化学で扱うデータ,特にスペクトルデー タは, n番目のサンプルにおいて,各波長λにおける吸 光度 *d*_(n,λ) を記録した多変量データであることから,ス ペクトル解析はケモメトリックスの主要な応用例であ る.多変量解析では,数値データを縦横に並べた行列と して表現し,線形代数を用いた様々な数値解析を行う. 例えば,n個の混合物サンプルについて測定したn個の 混合物スペクトルの行列表現を考える.ある一つの混合 物サンプルにおいて,各波長λにおける吸収強度*d*_(λ)を とすると,ランベルトベールの法則から*d*_(λ)は成分*k*の 純粋なスペクトル強度*s*_(4,λ)とその濃度*c*_(k)の線形和とし て(1)のように表現できる.

 $d_{(\lambda)} = \sum_{j=1}^{K} c_{(k)} s_{(k,\lambda)}$ (1)

ただし*K*は純成分の総数を表す. 各波長におけるスペ クトル強度 $d_{(\lambda)}$ を横方向(行方向)に並べることで,混 合物サンプル n のスペクトルは行ベクトル $d_{(n)}$ として (2)のように表すことができる.

この行ベクトルを縦方向(列方向)に並べることで, n サンプルの混合物スペクトルは行列 d_(n,λ)として表現で きる. 同様に,ある純成分 kのスペクトル強度 s_(k,λ)を行 ベクトルとし,それを縦方向に K 個並べることで,行 列として表現できる.さらに,それぞれのサンプルにお ける各成分の濃度を横方向に配列した行ベクトルを, n 個縦に並べることで,各サンプルにおけるそれぞれの成 分濃度も行列として式(3)のように表せる.



式 (3) を

と書く. ここで, Dが混合物スペクトル, Cが濃度分 布, Sが純スペクトルデータを表し, Tは行列の転置を 表す記号である.以上より, 混合物スペクトルから純ス

Qualitative and Quantitative Analysis of Mixture Spectral Data by Multivariate Curve Resolution.

ペクトルとその組成を推定することは、一つの行列を二 つの行列に分解する行列分解に対応する.

主成分分析は、行列分解に用いられる計算手法の中で 最も代表的な方法である.主成分分析は n 個のサンプル について、波長 λ_m などの m 種の変数を配列させた行列 データ X について、データ全体をより少数の変数で表 現できるような変換を行う手法である²⁾.数学的には、 n 個のサンプルの m 次元スペクトルデータ X を m より もずっと少ない k 次元の超平面上(図1 中 P) への射影 し、超平面を張る単位ベクトル上での大きさを求める.



主成分分析の定義は行列を用いると以下のように表現 できる.

$$\begin{bmatrix} t_{(1,1)} & t_{(1,2)} & \cdots & t_{(1,k)} \\ t_{(2,1)} & t_{(2,2)} & \cdots & t_{(2,k)} \\ \vdots & \vdots & \ddots & \vdots \\ t_{(n,1)} & t_{(n,2)} & \cdots & t_{(n,k)} \end{bmatrix} = \begin{bmatrix} x_{(1,\lambda_1)} & x_{(1,\lambda_2)} & \cdots & x_{(1,\lambda_n)} \\ x_{(2,\lambda_1)} & x_{(2,\lambda_2)} & \cdots & x_{(2,\lambda_n)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{(n,\lambda_1)} & x_{(n,\lambda_2)} & \cdots & x_{(n,\lambda_n)} \end{bmatrix} \begin{bmatrix} p_{(\lambda_1,1)} & p_{(\lambda_2,2)} & \cdots & p_{(\lambda_n,k)} \\ p_{(\lambda_{n-1})} & p_{(\lambda_{n-2})} & \cdots & p_{(\lambda_n,k)} \\ \vdots & \vdots & \ddots & \vdots \\ p_{(\lambda_{n-1})} & p_{(\lambda_{n-2})} & \cdots & p_{(\lambda_n,k)} \end{bmatrix} \cdots$$
(5)

ここで、Tはn次元の列ベクトル $t_{(k)} = [t_{(1,k)}, t_{(2,k)}, \cdots, t_{(n,k)}]^T$ を k 列配列したスコア行列, p は m 次元の列ベクトル $p_{(k)} = [p_{(\lambda_1,k)}, p_{(\lambda_2,k)}, \dots, p_{(\lambda_n,k)}]^T$ を k 列配列させたローディン グであり、**p**(k)をローディングベクトルと呼ぶ.式(3) と式(5)は形式上類似しているが、式(5)のローディ ングベクトルは互いに直交するが、式(3)の純スペク トルは直交しているとは限らないという点でこれらの2 式には違いがある. 式 (5) の解釈としては、スペクト ルデータX(m次元)にローディング行列Pを演算し て,スコア行列T(k次元)に次元圧縮することとなる. 次元圧縮の際に m-k 次元変換の情報ロス(図1中の点 線)が最小となるように P が求められ, データ X を X の次元 (m) よりも低次元 (k) で効率よく表現するこ とが可能となる.よって、XはTとPの積で近似でき、 主成分分析は行列分解としても近似的に利用できる. す なわち, 式 (5) は

と表現できる.式(4)及び(6)を対応させると,混 合物スペクトルに主成分分析を適用することで得たロー ディング Pとスコア Tをそれぞれ純スペクトルと濃度 分布として解釈することができる.これは得られた k 個 の主成分が混合物中の「純粋な」化学成分に対応する場 合は正しい.しかし,主成分はデータ全体を最も効率的 に表現できるように,分散が最大となるように選択され る数学的な仮想上の成分である.そのため,主成分は必 ずしも化学成分に対応せず,Pや Tが物理的化学的に直 接の解釈が可能な純スペクトルや濃度行列とならないこ とも多い.例えば,ローディングは純スペクトルとして は不自然な負の値をとったり,混合物スペクトルを純成 分と扱ったりするため、解釈が困難となる場合がある.

2 Multivariate curve resolution (MCR) 法

主成分分析に対して、MCR 法では解釈が可能な解を 得ることができる. その誕生は 50 年前にさかのぼる³⁾. Lawton と Sylvestre は、2 成分の重なった混合物スペク トルについて、スペクトル分離の手法を提唱した⁴⁾.混 合物スペクトル強度について濃度と純スペクトルの線形 結合で表し、スペクトル概形について仮定を用いない点 など今日の MCR 法に通じる重要な概念が示されてい る. 開発当初のアルゴリズムは, 濃度行列 C ないし, 純スペクトル行列Sを推定し,他方を逆行列計算で算 出することで、解を得るものだった.しかし、反応プロ セスデータなどに存在する濃度分布の連続性を仮定して 解析する evolving factor analysis (EFA)⁵⁾など,計算上に 何らかの仮定や制限をおくが必要があった. こうした仮 定や制限は行列分解に普遍的に存在する Rotational ambiguity⁶⁾に対処するために用いられる. Rotational ambiguity は行列分解の解の自由度のために、解が一意 に決まらないことを意味する. 行列分解の自由度は, 例 えばxy=36を満たすx, yは無数に存在する, すなわち 自由度が高く解が一意に決まらないが、x+y=15、x>y などの制約条件を加えることで、x=12、y=3と一意に 決まるのと本質的には同一である. EFA ではプロセス データなどに存在する各成分の濃度分布の連続性を仮定 し解の自由度を低下させることが可能だが、プロセス データ以外の様々な系へと MCR 法を利用していくため には制約条件を緩く設定する必要がある.しかし、制約 条件を弱めると解析的に行列分解ができないため、数値 的に解を得る必要がある. そこで交互最小二乗法(Alternating Least Square) を導入した MCR-ALS 法が提唱さ れた7). MCR-ALS は初期推定の濃度行列ないし純ス ペクトル行列を交互に最適化し、行列Dとの誤差が最 小となるまで計算を反復する最小二乗計算である. MCR-ALS 法の汎用的なソフトウェアとして、MCR-ALS GUI 2.0⁸⁾があり、プログラムやチュートリアル等 が web ページ⁹⁾で公開されている. 図2に計算プロセス





この最適化の過程で制約条件を加えることで解の一意 性を高められる。制約条件には、スペクトルや濃度分布 はどちらも正でなくてはならないという非負条件、純ス ペクトル形状または濃度分布の形状をピークが一つの単 調な形状に制限する Unimodality 制約、各成分での物質 収支を考慮して、サンプルごとの各成分濃度をその総和 で規格化する Closure 制約などがある⁶⁾.

MCR-ALS 法の初期推定法について述べる. MCR-ALS 法は初期値をもとにした最小二乗計算であるため、 最終的な解は初期値に大きく依存する. そのため、より 良い解を得るために様々な初期値推定法が提案・利用さ れており, 初期推定法の手段として, 主成分分析や Pure variable detection (PVD) 法など¹⁰⁾¹¹⁾がある. PVD 法は、混合物スペクトルを各波長におけるスペクトル強 度がサンプルごとに変動するベクトルとみなして、これ らの強度変動ベクトルの中で最も独立した変動を示す波 長を,純粋な物質のみに起因する pure variable として選 択する手法である. 選択された pure variable における強 度変動ベクトルに対し,他の波長での強度変動ベクトル の独立性を行列式から求め、この行列式が最小となるよ うな pure variable を逐次的に選別することで、互いに独 立した変動を示す波長 (pure variable) の組み合わせを 決定できる. 相異なる化学成分は互いに独立した変動を 示すので, 選択されたすべての pure variable における強 度変動ベクトルは相異なる純成分同士の濃度分布とみな すことができる.よって、これらの濃度分布を初期推定 値として ALS 計算を適用することで、純スペクトルと 濃度行列を得る. ALS を用いたことで MCR 法の適用 範囲はプロセスデータ以外にも拡張されてきた. 顕微分 光では、測定位置ごとに得られたスペクトルデータの集 合データに MCR-ALS が適用されており、ハイパース ペクトルイメージングデータや環境中での微粒子モニタ リングデータなどで用いられている¹²⁾¹³⁾.

MCR 法は、様々なスペクトル解析において適用され

てきた優れた手法であるが、実測した混合物データに対 して既存の初期値推定法を用いて MCR 法で計算する と、最小二乗計算が初期値近傍の極小値で収束してしま い、物理・化学的に有意でない誤った解が得られること が多いことが問題としてあげられる(図3上).そこで 最初からより理にかなった初期値を推定することで、計 算を最小値で収束させ、正しい解を導くことが考えられ た(図3下).



本稿では、著者らのグループで行われた、コサイン類 似度を用いて正解に近い初期値を導出する初期値推定法 と MCR 法を組み合わせた手法である cos-s MCR 法を紹 介する¹⁴⁾. さらに、よりロバストな解析が可能となる ように改良した cos-s map MCR 法¹⁵⁾を紹介し、様々な実 スペクトルデータ解析に適用した結果を示す.

3 Cosine similarity MCR 法

コサイン類似度を用いた初期推定法は、混合物のスペ クトルデータに対し、同一の化学種に由来するピークは サンプル方向に類似した強度変動を示し、異なる成分に 由来するピークは異なる強度変動を示すことを念頭に置 き、サンプル方向の強度変動の違いを化学種の違いとみ なし、それを初期推定に反映する手法である. 本初期 推定法は PVD 法と同様にサンプル方向の強度変動に着 目するが、PVD 法が強度変動の独立性に基づき、純成 分に由来する波長を選択するのに対し、コサイン類似度 を用いた初期推定法は強度変動の類似性に基づき純成分 スペクトルの波形を推定する点で PVD 法と異なる. 強 度変動の類似性をコサイン類似度という指標を用いるこ とで定量的に解釈し、コサイン類似度をいくつかの基準 値(閾値)をもとに判定を行う.詳細には,(1)混合 物スペクトルの個々のピークにおいて、純成分に由来す るピークと複数成分が重なるピークの判別、(2) ピー ク同士において,同一化学成分によるピークの識別, (3) 複数成分が重なるピークから純成分のピークの分 配を行うことで、構成成分数の決定とそれぞれの成分に 化学的に有意にスペクトル強度を分配できる(図4).







図 5 (a) 混合物の¹HNMR スペクトル, (b) コサイン類似度を用いて導出した初期 推定スペクトル, (c) cos-s MCR 法による計算結果, (d) 純スペクトル

このようにして得られた初期値を MCR 法で最適化する cos-s MCR 法によって純スペクトルとその組成を高精度 に求めることができる.

cos-s MCR 法を 4 種類の化学物質(THF, 酢酸エチ ル, アセトン,シクロペンタノン)を任意の混合比で含 む 16 個の混合物サンプルの核磁気共鳴(¹HNMR)スペクトルに適用した¹⁴⁾.図 5 (a)に測定した¹HNMR スペクトルを示す.また,コサイン類似度を用いた初期 推定法を用いて強度変動の類似度を評価し,ピーク領域 を各成分に割り当てることで得た初期推定スペクトルを 図 5 (b)に示す.得られた初期推定スペクトルをもと に,MCR 法で最適化することで得た純スペクトルを図 5 (c)に示す.得られた結果はシグナル分裂及び積分比 について図 5 (d)に示した純スペクトルと一致し,そ の一致度は 85 % 程度であった.また,濃度についてその誤差は 8 % 未満であった.

4 Cosine similarity map MCR 法

cos-s MCR 法は、リファレンススペクトルや各成分の 濃度プロファイル等のサンプル系に関する事前情報がな くても、種々の分光法で測定された混合物スペクトルに 適用可能である.しかし実際には、本手法の適用が困難 な系も存在する.例えば、バックグラウンドが大きい、 スペクトル重なりが大きい、SN 比が低いスペクトル データである.cos-s MCR 法の初期推定においてはリ ファレンススペクトルを利用せず、成分が未知であって も決定できるが、計算プロセスには複数の閾値とそれに 基づく判定が含まれる.これらの判定に基づき各ピーク 領域での重なりの有無を判定し、その分離が行われるた め、誤判定により計算精度が低下する. とくに含有成分 の濃度や純スペクトルの強度と関連しないノイズやバッ クグラウンドが存在すると誤判定が起こりやすくなる. もちろん個々の応用例で閾値パラメータを調整すること で分類精度を改善することができるが、これらは分析手 法ごとに異なり、解析結果が恣意的になる可能性があ る. そこで、パラメータの調整を用いないノンパラメト リックな手法が必要である.

上記の背景のもと、よりロバストな手法として cos-s MCR 法の初期値推定法の改良を図るために、Cosine similarity mapping (cos-s map) を導入し、これと MCR 法 を組み合わせた cos-s map MCR 法が開発された¹⁵⁾. cos-s mapは、混合物スペクトルある波長 A の強度変化傾向 と別の波長 λ_B の強度変動のコサイン類似度を求め、 λ_A とλ_Bのすべての組み合わせで求めたコサイン類似度を ヒートマップとして可視化したものである. ここで、同 一の化学種に由来する成分は各波長で同一の変動を示し て類似度が高くなるため、2次元 NMR と同様に cos-s map 上の相関が高い領域同士を同一成分に由来すると解 釈する. これにより各成分スペクトルの概形を予測でき る. この cos-s map に対し, データの分類を行うソフト クラスタリングの一種である FCM (fuzzy c-means) 法 を適用することで、純スペクトルの初期推定値を得る方 法を開発した. FCM 法では各波長における構成成分の 重みづけを得ることができるので、測定したスペクトル

の平均を基準として、その重みづけを使ってスペクトル 強度を分配することで初期推定スペクトルを得ることが できる.このことにより、ユーザーが決める閾値が不要 になるため、任意性の高い方法になっている.この初期 推定値をもとに MCR 法による最適化を行うことで、純 スペクトルと濃度情報を算出する.

cos-s map MCR 法の実例としてはじめに,複数成分が 重なるブロードな混合物スペクトル系である pH が異な る八つのブロモチモールブルー(BTB) 溶液の紫外可 視吸収スペクトルが用いられた¹⁵⁾.以下の図 6 に調製 された BTB 溶液の UV-Vis スペクトルを示す.

BTB 溶液の UV-Vis スペクトルから, 図 7 (a) に示す cos-s map が得られ, これに FCM 法を適用して得た初期



図 6 pH が異なる BTB 溶液の UV-Vis スペクトル



図 7 (a) pH の異なる BTB 溶液の吸収スペクトルから得た cos-s map, (b) cos-s map から算出 した初期推定スペクトル, (d) 純スペクトル, (d) cos-s map MCR 法による分離結果

推定スペクトルが図7(b)のように得られた.また, 強酸または強塩基条件下で測定された純スペクトルを図 7(c)に示す.これらを比較すると,初期推定スペクト ルと純スペクトルのピーク位置は一致するが,形状は異 なるため最適化が必要である.この初期値をもとに MCR法による最適化を行って得られた純スペクトルを 図7(d)に示す.図7より,MCR法により初期推定ス ペクトルの形状が最適化され,得られたスペクトルは, 純スペクトルと形状が99%以上一致した.また,濃度 についてもBTBの酸塩基平衡式(7)を用いた最小二 乗法によるフィッティングを行うことで,pK_a(K_aは酸 解離定数)を算出したところpK_aは文献値¹⁶⁾と一致し 7.07であった.以上より,スペクトルの重なりの大き い吸収スペクトルから,純成分の定性と定量が示され た.

$$[HA] = \frac{[H^+]}{[H^+] + K_a}, [A^-] = \frac{K_a}{[H^+] + K_a} \dots \dots \dots (7)$$

さらに,孤立した多数のピーク群を持つ XRD パター ンに cos-s map MCR 法を適用した.その実例として酸化 鉄(III),酸化鉄(II,III),酸化アルミニウム(III),二 酸化ケイ素の組成を任意に変化させながら混合した 12 個の粉末サンプルの XRD パターンが用いられた¹⁵⁾.図 8 に測定した XRD パターンを示す.

これらの XRD パターンついて,同様に求めた cos-s

map を図9(a)に、初期推定スペクトルを図9(b)に 示す. さらに、純物質のスペクトルを図9(c)に、得 られた初期推定をもとにして MCR 法を適用した結果を 図9(d)に示す. 図9より、MCR 法により各成分のス ペクトル形状と強度が最適化されたことで推定結果と純 スペクトルは83%以上一致した.よって、物質同定が 可能なレベルで純スペクトルを推定できた.また、各サ ンプルにおける各成分の濃度比を求めたところ、実際の 調製量との絶対誤差は6%未満であり、高い精度で各 成分の濃度が推定されている.



図8 4 成分混合物粉末試料の XRD パターン



図 9 (a) 混合物の XRD パターンから得た cos-s map, (b) cos-s map から算出した初期推定ス ペクトル XRD パターン, (c) 純スペクトル, (d) cos-s map MCR 法による分離結果

以上より,形状の大きく異なる吸収スペクトルと XRDの混合物データの双方で,純スペクトルを高い精 度で計算できたことから, cos-s map MCR 法は様々な混 合物スペクトルデータの解析に適用可能な一般化された 計算手法といえる.

5 ま と め

本稿では、混合物のスペクトル解析法である行列分解 法の基本から、今日での主要な解析法である MCR 法, さらにこれらを発展させた cos-s MCR や cos-s map MCR 法について解説してきた. MCR 法により混合物スペク トルから純成分スペクトルと濃度比率を求めることは, 複雑で多数の情報から解釈可能な少数の重要な情報を得 る特徴抽出であると解釈できる.近年では分析機器の性 能向上により様々な分析データがビックデータ化してき ており、 高次元の多数のデータの中から特徴抽出を行う ために、行列分解を利用することで解釈可能な解析を行 うことが期待できる. cos-s MCR とそれをさらに改良し た cos-s map MCR 法を用いることで,異なる分析方法か ら得られた形状の差異が大きいスペクトルデータについ て、純成分スペクトルと各成分の濃度分布を高い精度で 得ることができる.今後は混合物中の各成分同士の相互 作用により、スペクトルのピークが漸次的にシフトする 系や微量成分の定量や定性など cos-s map MCR 法のさら なる改良・改善が進み、適用範囲がさらに広がることが 期待される.

文 献

- 1) S. Wold : Chemom. Intell. Lab. Syst., 30, 109 (1995).
- 加納 学:主成分分析,京都大学大学院工学研究科化学 工学専攻,http://manabukano.brilliant-future.net/document/ text-PCA.pdf (2021年10月29日,最終確認).
- 3) A. de Juan, R. Tauler : Anal. Chim. Acta, 1145, 59 (2020).
- 4) W. H. Lawton, E. A. Sylvestre : *Technometrics*, **13**, 617 (1971).
- 5) H. Gampp, M. Maeder, C. J. Meyer, A. D. Zuberbuehler :

Comments Inorg. Chem., 6, 41 (1987).

- A. de Juan, R. Tauler: Crit. Rev. Anal. Chem., 36, 163 (2006).
- R. Tauler: Multivariate resolution in chemistry Lecture 2, https://digital.csic.es/bitstream/10261/37051/1/Tauler_Roma. pdf (2021 年 10 月 29 日, 最終閲覧).
- J. Jaumot, A.de Juan, R. Tauler : Chemom. Intell. Lab. Syst., 140, 1 (2015).
- R. Tauler, A. de Juan, J. Jaumot : Multivariate Curve Resolution Homepage, https://mcrals.wordpress.com/ (2021 年 11 月 26 日,最終確認)
- W. Windig, A. Bogomolov, S. Kucheryavskiy: "Comprehensive Chemometrics (2nd edition)", Edited by S. D. Brown, R. Tauler, B. Walczak, p.107 (2020), (Elsevier, Amsterdam).
- W. Windig, J.L. Lippert, M.J. Robbins, K.R. Kresinske, J.P. Twist, A.P. Snyder : *Chemom. Intell. Lab. Syst.*; 9, 7 (1990).
- 12) J. P. Smith, F. C. Smith, K. S. Booksh : Analyst, 142, 3140 (2017).
- M. Terrado, D. Barceló, R. Tauler : *Environ. Sci. Technol.*, 43, 5321 (2009).
- 14) Y. Nagai, W. Y. Sohn, K. Katayama : Analyst, 144, 5986 (2019).
- 15) Y. Nagai, K. Katayama : Analyst, 146, 5045 (2021).
- 16) T. Shimada, T. Hasegawa : Spectrochim. Acta. A. Mol. Biomol. Spectrosc., 185, 104 (2017).



永井優也 (Yuya NAGAI) 中央大学 (〒112-8551 東京都文京区春 日1-13-27). 中央大学理工学研究科応 用化学専攻博士前期課程. 修士 (工学). 《現在の研究テーマ》分析データからの特 徴抽出とそれを基にした帰納的な材料開 発.《趣味》ガジェット収集, 散歩.



片山建二(Kenji KATAYAMA) 中央大学理工学部(〒112-8551 東京都 文京区春日1-13-27).東京大学工学系 研究科応用化学専攻.博士(工学).《現 在の研究テーマ》時間分解顕微分光法と データ科学的手法を組み合わせた光触媒, 太陽電池,光応答性液晶の機能探索.《趣 味》犬の散歩.

E-mail : kkata@kc.chuo-u.ac.jp