実験スペクトル自動分解のための線形独立非負行列因子分解法の提案 Linear independent non-negative matrix factorization for automatic decomposition of experimental spectra

吉川信明 鈴村彰敏 田島伸 山川俊輔 旭良司 Nobuaki Kikkawa Akitoshi Suzumura Shin Tajima Shunsuke Yamakawa Ryoji Asahi

> 株式会社豊田中央研究所 Toyota Central R&D Labs., Inc.

We propose a new algorithm to decompose a set of arbitrary complex spectra into independent basis spectra. Based on the nonnegative matrix factorization (NMF) algorithm we introduce a regulation term to a cost function in order to ensure the linear independency between basis spectra. We apply it to decompose a set of complex X-ray diffraction (XRD) spectra, which is typically obtained from high-throughput materials libraries. Each of the decomposed spectra nicely matches with one of the single phase spectra in an available XRD database. Alternatively, the method facilitates to find new phases when the decomposed spectra are not found in the database. These features are particularly useful for materials informatics where big spectrum data obtained from experiment are to be handled.

1. はじめに

近年の情報科学の発展を受け材料科学の分野においても情報技術の活用が急務となっている。このような材料科学と情報科学の境界領域はmaterials informatics (MI)と呼ばれ、米国Materials Genome Initiative[1]をはじめ、各国のプロジェクトにより研究が進められている。例えば、コンビナトリアル合成と高速(high-throughput, HT)測定を組み合わせることで大量の実験データを取得し、そのデータを情報科学の知見に基づき利用することで材料開発を効率化しようという試みが進められている[2]。我々は、薄膜合成技術と放射光を用いたHT-XRD(X-raydiffraction)測定技術を組み合わせることで組成の異なる多数のサンプルを一度に合成し、各サンプルのXRDスペクトルを得ることに成功しており、これらのデータを材料開発に活かす技術の構築が望まれている。

XRD スペクトルはサンプル内の結晶相ごとに特有のピークパ ターン(単相スペクトル)を持ち、その強度がサンプル中の結晶 相の割合に対応する。単相スペクトル(から同定された結晶相) とその相の割合は物理的に良い記述子となることから、これを求 めることが XRD 解析の目的となる。特に HT-XRD 測定では、

- 各サンプルから得られたスペクトルは、サンプル数に比べ 少数の単相スペクトルの線形結合で近似できる。
- ② 測定結果、単相スペクトル、線形結合係数のいずれも非負 である。

という性質を持つ場合が多く、非負行列因子分解(Nonnegative Matrix Factorization, NMF)、あるいは多変量スペクト ル分解(Multivariate Curve Resolution, MCR)と呼ばれる方法 が有効となる[3-8]。これらの手法は HT 測定で得られる XRD スペクトルが数種類の非負基底の非負線形結合で表されると仮 定して、極小値探索により非負基底、および線形結合係数を求 める手法であり、理想的には非負基底が単相スペクトル、線形 結合係数が各結晶相の割合に対応する。

NMF 法による XRD スペクトル分解の問題は、非負基底および非負線形結合係数に対する制約が少なく解が一意でないため、非負基底が単相スペクトルに対応する保証がない点である。

この問題に対処するには極小値探索時に正則化項を加えるこ とが有効であり、L1 正則化[9, 10] や L2 正則化[8]をはじめ、直 交性[11]、滑らかさ[12] 等を正則化項として加味することが提案 されている。しかしながらいずれの方法も満足のいく分解がで きているとは言い難い。

複数の結晶相が混合された混合相の XRD スペクトルでは通 常、一部のピークを除き各結晶相の単相スペクトルは重ならな い。これは数学的には単相スペクトル間の線形独立性が高いこ とに対応する。したがって線形独立性を向上させる正則化項は、 結晶相の XRD スペクトルの分解において良い正則化項になる ことが予想される。そこで本研究では線形独立性の定量化、お よび線形独立性を考慮した NMF の開発を目的として研究を行 った。

2. 方法

正則化項を含む NMF は $Y \in \mathbb{R}^{N \times M}$ 、 $C \in \mathbb{R}^{N \times L}$ 、 $S \in \mathbb{R}^{L \times M}$ として

minimize $f(Y, CS) \equiv d(Y, CS) + g(C, S)$

subject to $C_{nl} \ge 0$, $S_{lm} \ge 0$

で表される。HT-XRD スペクトルの文脈では、Y は HT-XRD 測定で得られたスペクトルのセット、C は線形結合の係数、S は単相スペクトルに対応する。また、d(Y, CS) は入力値と分解値の 乖離度、g(C, S) は正則化項である。なお、本研究では乖離度 として、

$$d(Y, CS) = \frac{1}{2} \|Y - CS\|_F^2 = \frac{1}{2} \sum_{nm} \left(Y_{nm} - \sum_{l} C_{nl} S_{lm} \right)^2$$

を使用している。

本研究で提案する線形独立性を考慮した正則化項は、

$$g^{\text{indep}}(S) = -\frac{k_1^{\text{indep}}}{2}\log[\det(SS^{\text{T}})] + \frac{k_2^{\text{indep}}}{2}\operatorname{tr}(SS^{\text{T}})$$

である。ここで $k_1^{\text{indep}}, k_2^{\text{indep}}$ はハイパーパラメータであり正値を取る。本正則化項はSの各成分が線形従属となるときに ∞ 、直交する場合に0となる関数であり、特に線形従属性が高い場合に急激に大きくなる。このような特性は NMF の分解結果が線形従属となることを強く抑制する効果を持つ。

連絡先:吉川信明,豊田中央研究所,

^{〒480-1192} 愛知県長久手市横道 41-1,0561-71-8044, e1703@mosk.tytlabs.co.jp



図1 HT-XRD 測定により測定される XRD スペクトルのセット.

3. 結果と議論

図 1 は斜めスパッタ法により製膜した 81 サンプルの Ga₂O₃-GeO₂-Cr₂O₃ 系コンビナトリアル傾斜組成薄膜の HT-XRD スペクトルであり、図 2 はその分解結果と分解結果を既存 の XRD スペクトル DB と照合した結果である。分解されたスペ クトルと DB 照合結果は良い一致を示しており、提案手法により 複数の相を含んだ HT-XRD スペクトルから良好な精度で単相 スペクトルを抽出できていることが分かる。提案手法は通常の NMF に線形独立性の正則化項を加えるだけという非常に単純 な方法であり、本手法がうまくいくということは、線形独立性が単 相スペクトルの特徴を良く捉えていることを意味する。また、本 手法ではスペクトルの分解に DB を利用しないため、DB に存 在しない新しい結晶相の抽出が可能である。実際、本手法で 見つかった Ga₄GeO₈ 相は(DB には登録されていたが、)専門 家の手動解析では見逃されていた相であった。

4. まとめ

本研究では線形独立性という単相スペクトルの特徴をうまく捉 えた正則化項を用いることで、HT-XRD スペクトルから単相スペ クトルを抽出する NMF 手法を開発した。本手法は DB を参照 することなく単相スペクトルを得ることが可能であるため、未知の 相を抽出する能力を持つ。また、今回は特に言及していないが 単相スペクトルとともにその強度も同時に取得されるため、この 強度に注目することで相図の自動作成等も可能と考えられる。

本研究では提案した線形独立性の正則化項を加える以外に 特殊な手法は用いていない。したがって既存の正則化項と組 み合わせたり、NMF 手法を改良することにより、より高精度な分 解が達成できる可能性を持つ。なお、XRD スペクトルは条件 によってピークのブロード化やシフトが起こるが、これらついても 今後本手法を改良することで対応する予定である。

5. 謝辞

本研究を進めるに当たり、弊社西川和孝氏、松原賢東氏から実験データを提供頂いた。また岸田佳大氏にはデータ測定 に協力いただいた。

参考文献

 A. Jain *et al.*, "Commentary: The Materials Project: A materials genome approach to accelerating materials innovation," *Apl Materials*, vol. 1, no. 1, p. 011002, 2013.



図 2 HT-XRD スペクトルの分解結果(青)と DB 照 合の結果(黄). (b)~(d) については細かいピーク特 徴まで捉えられている. (a) については抽出精度は高 くないが、実験条件を考慮すれば同定は可能であ る.

[2] R. Potyrailo, K. Rajan, K. Stoewe, I. Takeuchi, B.

Chisholm, and H. Lam, "Combinatorial and highthroughput screening of materials libraries: review of state of the art," *ACS combinatorial science*, vol. 13, no. 6, pp. 579-633, 2011.

- [3] D. D. Lee and H. S. Seung, "Algorithms for nonnegative matrix factorization," in Advances in neural information processing systems, 2001, pp. 556-562.
- [4] C. Long, D. Bunker, X. Li, V. Karen, and I. Takeuchi, "Rapid identification of structural phases in combinatorial thin-film libraries using x-ray diffraction and non-negative matrix factorization," *Review of Scientific Instruments*, vol. 80, no. 10, p. 103902, 2009.
- [5] S. Ermon *et al.*, "Pattern Decomposition with Complex Combinatorial Constraints: Application to Materials Discovery," in *AAAI*, 2015, pp. 636-643.
- [6] J. Gregoire, D. Van Campen, C. Miller, R. Jones, S. Suram, and A. Mehta, "High-throughput synchrotron X-ray diffraction for combinatorial phase mapping," *Journal of synchrotron radiation*, vol. 21, no. 6, pp. 1262-1268, 2014.
- [7] R. Tauler, "Multivariate curve resolution applied to second order data," *Chemometrics and intelligent laboratory systems*, vol. 30, no. 1, pp. 133-146, 1995.
- [8] V. P. Pauca, J. Piper, and R. J. Plemmons, "Nonnegative matrix factorization for spectral data analysis," *Linear algebra and its applications*, vol. 416, no. 1, pp. 29-47, 2006.
- [9] V. Pomareda, D. Calvo, A. Pardo, and S. Marco, "Hard modeling multivariate curve resolution using LASSO: application to ion mobility spectra," *Chemometrics and Intelligent Laboratory Systems*, vol. 104, no. 2, pp. 318-332, 2010.
- P. O. Hoyer, "Non-negative sparse coding," in *Neural* Networks for Signal Processing, 2002. Proceedings of the 2002 12th IEEE Workshop on, 2002, pp. 557-565: IEEE.
- [11] C. Ding, T. Li, W. Peng, and H. Park, "Orthogonal nonnegative matrix t-factorizations for clustering," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2006, pp. 126-135: ACM.

[12] S. Essid and C. Févotte, "Smooth nonnegative matrix factorization for unsupervised audiovisual document structuring," *IEEE Transactions on Multimedia*, vol. 15, no. 2, pp. 415-425, 2013.