



Műszaki Kémiai Napok '04



Veszprém, 2004. április 20-22.

SPEKTRÁLIS ADATTÖMBÖK KOMPONENS-PROFILJAINAK KINYERÉSE
EXPLOITING COMPONENT-PROFILES FROM SPECTRAL DATA ARRAY

RAJKÓ RÓBERT

SZTE Szegei Élelmiszeripari Főiskolai Kar
 Élelmiszeripari Műveletek és Környezettechnika Tanszék
 6725 Szeged, Moszkvai krt. 5-7.
 6701 Szeged, Pf. 433.
 rajko@sol.cc.u-szeged.hu
 www.kemometria.hu/Rajko

A különböző analitikai mérőkészülékeket, ill. módszereket aszerint csoportosíthatjuk, hogy milyen típusú adatokat szolgáltatnak. Nulladrendű készülék lesz az egyszerű számadatot nyújtó berendezés, mivel az egyszerű skalár nulladrendű *tenzor* (a tenzor fogalmát a fizikában megszokott módon, a rendezett adatstruktúrák általánosításaként értelmezzük). Abban az esetben, ha a kalibrációt egy nulladrendű készülék (pl. ionszelektív elektród, vagy egyszerű fotométer) segítségével végezzük, úgy nulladrendű (1 utas) kalibrációt kell végrehajtanunk.

A gyakorlatban zavaróhatásokkal (interferencia) is számolnunk kell, melyek két forrásból származhatnak. Egyrészt az adott alkotó mellett előforduló többi alkotó is szolgáltathat jelet (a szelektivitás nem kielégítő volta miatt), ezt hívjuk háttérnek (background). Másrészt az oldószer és a készülék is okozhat zavart, amelyet eltolódásnak (offset) nevezünk. Az említett két ok bár kémiaiilag különbözőképpen eredeztethető, matematikailag együttesen le tudjuk írni őket. Az eltolódás és a háttér közötti fő különbség az, hogy a háttér mintáról mintára változhat, míg az eltolódás gyakran állandónak tekinthető, így vakmintával (amely csak az oldószert tartalmazza) elvégezve a mérést, a készüléket nullázhatjuk.

Ha a zavaróhatás sok változnak a kalibrációs sorozatban (ez főleg a háttérre jellemző), a valódi egyenestől jelentős mértékben eltérő kalibrációs görbét kapunk. Egy másik probléma jelentkezik, amikor a kalibrációs sorozatban ugyan nincs zavaróhatás, a meghatározandó mintában viszont van. Az így jelentkező torzítás mértéke gyakran katasztrofális is lehet.

Belátható tehát, hogy a nulladrendű kalibrációval, pusztán a kalibrációs adatokra támaszkodva, a zavaróhatásoknak még a kimutatása sem lehetséges.

Az elsőrendű készülék nulladrendű érzékelők sorából áll (pl. diódasoros spektrométerek). Az így nyert adatok egy vektort alkotnak, a vektor pedig elsőrendű tenzor (egy dimenzióval rendelkező adatstruktúra).

Lineáris függvénykapcsolat esetén (pl. a Lambert-Beer-Bouguer-törvény érvényességi tartományán belül) a következő mátrix összefüggést nyerjük:

$$\mathbf{R} = \mathbf{B} \cdot \mathbf{C}^T + \mathbf{E}.$$

Ez a mátrix egyenlet áttekinthetőbb, ha komponensei szerint kifejtjük:

$$\begin{bmatrix} r_{11} & \Lambda & r_{1n} & \Lambda & r_{1N} \\ r_{21} & \Lambda & r_{2n} & \Lambda & r_{2N} \\ \mathbf{M} & & \mathbf{M} & & \mathbf{M} \\ r_{s1} & \Lambda & r_{sn} & \Lambda & r_{sN} \\ \mathbf{M} & & \mathbf{M} & & \mathbf{M} \\ r_{S1} & \Lambda & r_{Sn} & \Lambda & r_{SN} \end{bmatrix} = \begin{bmatrix} b_{10} & b_{11} & \Lambda & b_{1k} & \Lambda & b_{1K} \\ b_{20} & b_{21} & \Lambda & b_{2k} & \Lambda & b_{2K} \\ \mathbf{M} & \mathbf{M} & & \mathbf{M} & & \mathbf{M} \\ b_{s0} & b_{s1} & \Lambda & b_{sk} & \Lambda & b_{sK} \\ \mathbf{M} & \mathbf{M} & & \mathbf{M} & & \mathbf{M} \\ b_{S0} & b_{S1} & \Lambda & b_{Sk} & \Lambda & b_{SK} \end{bmatrix} \cdot \begin{bmatrix} 1 & \Lambda & 1 & \Lambda & 1 \\ c_{11} & \Lambda & c_{1n} & \Lambda & c_{1N} \\ \mathbf{M} & & \mathbf{M} & & \mathbf{M} \\ c_{k1} & \Lambda & c_{kn} & \Lambda & c_{kN} \\ \mathbf{M} & & \mathbf{M} & & \mathbf{M} \\ c_{K1} & \Lambda & c_{Kn} & \Lambda & c_{KN} \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} & \Lambda & \varepsilon_{1n} & \Lambda & \varepsilon_{1N} \\ \varepsilon_{21} & \Lambda & \varepsilon_{2n} & \Lambda & \varepsilon_{2N} \\ \mathbf{M} & & \mathbf{M} & & \mathbf{M} \\ \varepsilon_{s1} & \Lambda & \varepsilon_{sn} & \Lambda & \varepsilon_{sN} \\ \mathbf{M} & & \mathbf{M} & & \mathbf{M} \\ \varepsilon_{S1} & \Lambda & \varepsilon_{Sn} & \Lambda & \varepsilon_{SN} \end{bmatrix}$$

r_{sn} az s -edik szenzor által szolgáltatott jel az n -edik mintára vonatkozóan, b_{sk} ha $k \neq 0$ az s -edik szenzor k -edik komponensre vonatkozó érzékenységet (az s -edik szenzornak megfelelő hullámhosszhoz tartozó k -edik komponens moláris abszorpciós koefficiensét a fényúttal szorozva, a példánál maradván), ha $k=0$ akkor az s -

edik szenzorhoz tartozó q értéket, azaz a háttérret (eltolást, alapvonalat), c_{kn} a k -edik komponens n -edik mintabeli koncentrációját jelenti, ε_{sn} az n -edik mintában az s -edik szenzorhoz tartozó mérési hiba.

bilineáris modellnek hívjuk, mivel \mathbf{R} minden egyes sora ugyanazon lineáris függvénye \mathbf{B} sorainak, ill. \mathbf{R} -nek minden egyes oszlopa ugyanazon lineáris függvénye \mathbf{C}^T oszlopainak:

$$\mathbf{R}[s, \cdot] = \mathbf{B}[s, \cdot] \cdot \mathbf{C}^T \quad s = 1, \dots, S$$

$$\mathbf{R}[\cdot, n] = \mathbf{B} \cdot \mathbf{C}^T[\cdot, n] \quad n = 1, \dots, N$$

Ha az \mathbf{R} mátrixon kívül a \mathbf{B} vagy \mathbf{C} mátrixok valamelyikét ismerjük, akkor a másik mátrix műveletekkel kifejezhető. Ha csak az \mathbf{R} mátrix ismert, akkor fel kell bontanunk két mátrix szorzatára remélve azt, hogy fizikailag azonosítható spektrális és koncentráció profilokat nyerünk. A felbontás elvégezhető a főkomponens elemzés (principal component analysis, PCA) vagy a szinguláris érték felbontás (singular value decomposition, SVD) segítségével:

$$\mathbf{R} = \mathbf{U} \cdot \mathbf{V}^T + \mathbf{E}, \quad \mathbf{R} = \mathbf{B} \cdot \mathbf{C}^T + \mathbf{E}, \quad \mathbf{R} = \mathbf{U}\mathbf{T} \cdot \mathbf{T}^{-1}\mathbf{V}^T + \mathbf{E} = \mathbf{B} \cdot \mathbf{C}^T + \mathbf{E}$$

$$\mathbf{B} = \mathbf{U} \cdot \mathbf{T} \quad \mathbf{C}^T = \mathbf{T}^{-1} \cdot \mathbf{V}^T$$

ekkor azonban absztrakt faktorokat kapunk, amelyekből a \mathbf{T} forgatómátrix segítségével kaphatjuk a fizikailag értelmezhető profilokat. A feladat a \mathbf{T} mátrix meghatározása. Az angolszász irodalomban ehhez a feladathoz kapcsolódó módszerek összefoglaló elnevezése *self-modeling curve resolution, SMCR* (önmodellező görbefelbontás). A szó szerinti fordítás nem igazán adja vissza, mi is a módszerek szerepe, helyesebb lenne a *görbeillesztés nélküli profilkeresés* elnevezés, hiszen a spektrum vagy a koncentráció profilt leíró valódi függvény ismerete nélkül (görbeillesztési feladat elhagyásával) keressük az adatokat legjobban leíró profilokat. A gyakran használt *modellmentes* (modell-free) elnevezés sem helyes, hiszen van modellünk, a Lambert-Beer-Bouguer-törvény által megfogalmazott bilineáris modell.

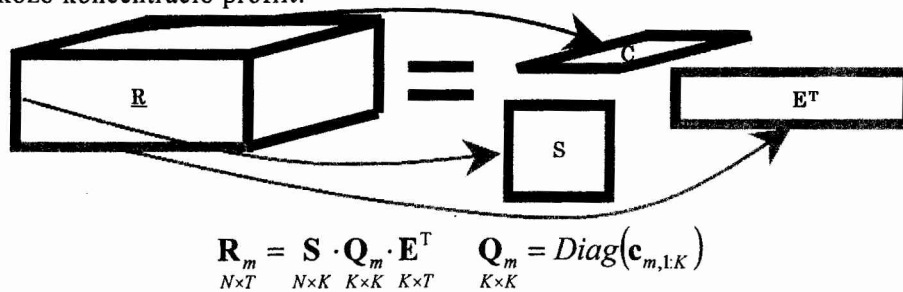
A megoldáshoz közelebb jutunk, ha feltételezzük azt a szinte triviális tényt, hogy mind a koncentráció, mind az abszorbancia értékek csak pozitívak lehetnek. De még e két feltétel teljesülése esetében sem tudunk mindig egyértelmű megoldáshoz jutni.

Az SMCR módszereket két csoportra oszthatjuk: azokra amelyek az egyértelmű megoldást tűzik ki célul, és azokra amelyek csak megoldástartományt szolgáltatnak.

Az *evolving factor analysis, EFA* (kifejlődő faktorelemzés), *window factor analysis, WFA* (ablakfüggvényes faktorelemzés), *heuristic evolving latent projection, HELP* (heurisztikus kifejlődő rejtett leképezés), *subwindow factor analysis, SWFA* (egy komponenses ablakfüggvényt alkalmazó faktorelemzés), *parallel vector analysis, PVA* (párhuzamos vektorelemzés) az első csoportba tartozik. Ezek a módszerek azokat az információkat használják, amelyeket lokális, helyi szinten nyerhetnek ki az adatokból, amikor szelektív vagy zérókoncentrációjú területeket keresnek.

A második csoport tagjai: *orthogonal projection analysis, OPA* (ortogonális kivetések elemzése), *alternating least squares, ALS* (alternáló /ide-oda végrehajtott/ legkisebb négyzetek), *simple-to-use interactive self-modeling mixture analysis, SIMPLISMA* (egyszerűen használható párbeszédes görbeillesztés nélküli profilkeresés), *iterative target transformation factor analysis, ITTFA* (ismétlődő célvektor transzformációs faktorelemzés), *positive matrix factorization, PMF* (csak pozitív profilt eredményező mátrixfelbontás), *simplex-based method, SIMPLEX* (szimplexeken alapuló módszerek), *dynamic Monte Carlo self-modeling curve resolution, DMC-SMCR* (görbeillesztés nélküli profilkeresés dinamikus Monte Carlo módszerrel), *alternating least squares with penalty functions, ALSwPF* (alternáló /ide-oda végrehajtott/ legkisebb négyzetek büntető függvényekkel), *source apportionment by factors with explicit restrictions, SAFER* (faktorok általi forrás felosztás explicit korlátozásokkal). Ezek a módszerek olyan tartományokat adnak eredményül, amelyekben nem sérülnek az *a priori* feltételezések, pl. a nemnegativitás, unimodalitás stb. A tartományok szélessége függ a tiszta profilok közötti korrelációtól, a kollinearitástól.

Amennyiben a spektrális adatsor valamilyen elválasztó művelettel kapott eredmény, akkor több különböző összetételű mintákra vonatkozó mérés esetén ezek összeilleszthetők és adatmátrix helyett 3 utas adattömb kiértékelése a feladat. Trilineáris adatsor esetén a *parallel factor analysis*, *PARAFAC* (párhuzamos faktorelemzés) enyhe feltételek teljesülése esetén szolgáltatja a normált spektrumot, elúciós profilt és a mintákra vonatkozó koncentráció profilt:



Az előadásban a felsorolt SMCR módszerek közül néhány és a PARAFAC által szolgáltatott eredmények bemutatása és elemzése hangzik el.

A kutatás az OTKA T037480 sz. pályázata, valamint a szerző Széchenyi István Kutatási Ösztöndíja támogatásával valósult meg.

Irodalom

Bro, R. (1998): *Multi-way Analysis in the Food Industry. Models, algorithms and applications.* Doctoral thesis, University of Amsterdam.

Horvai György (szerkesztő) (2001): *Sokváltozós adatelemzés (kemometria)*, Nemzeti Tankönyvkiadó, Budapest

Tartalom

I. Feltáró statisztika

Adatok értelmezése (*Szepesváry Pál*)

Csoportosítás (alakfelismerés) (*Borosy András*)

Faktor- és főkomponensanalízis és változataik (*Héberger Károly, Rajkó Róbert*)

II. Sokváltozós mennyiségi összefüggések alkotása, kezelése

Fizikai modellek jóságának és paramétereinek becslése és alkalmazhatóságának vizsgálata (*Rajkó Róbert*)

Modellépítés a regressziós számítások során, korrelált változók esetén (*Héberger Károly, Rajkó Róbert, Kolossváry István*)

Sztocasztikus folyamatok (*Paksy László*)

III. Matematikai összefoglaló

Lineáris algebrai áttekintés (*Szepesváry Pál*)

Valószínűségelméleti alapok (*Lengyel Attila*)

Gobális szélsőérték-kereső módszerek (*Borosy András*)

Mesterséges ideghálózatok (*Borosy András*)

Függelék

A, Tárgymutató magyar-angol szótárral (összes szerző)

B, Software tájékoztató (*Borosy András*)

Jiang, J-H., Liang, Y., Ozaki, Y. (2004): Principles and methodologies in self-modeling curve resolution. *Chemometrics and Intelligent Laboratory Systems*, 71, 1-12.

Jiang, J-H., Ozaki, Y. (2002): Self-modeling curve resolution (SMCR): Principles, techniques, and applications. *Applied Spectroscopy Reviews*, 37(3), 321-345.

- Malinowski, E.R. (2002):** *Factor Analysis in Chemistry* (3rd Ed.). Wiley, New York.
- Rajkó, R. (1994):** Treatment of model error in calibration by robust and fuzzy procedures. *Analytical Letters*, 27(1), 215–228
- Rajkó R. (1998):** *Regressziós módszerek alkalmazása és fejlesztése a kémiai és élelmiszeripari mérések kiértékelése során.* PhD Értekezés, BME, Budapest.
- Rajkó R. (1999):** Robusztus statisztikai eljárások a műszaki kémiában. *Műszaki Kémiai Napok'99* Veszprém, 1999. április 27-29. pp. 98-103.
- Rajkó R. (2001):** Kémiai kalibráció kemometriai jellegű problémáiról. *Műszaki Kémiai Napok'01* Veszprém, 2001. április 24-26. pp. 31-36.
- Rajkó R. (2001):** Kalibráció a kémiai méréseknél. Az analitikai információ minősége. *Magyar Kémiai Folyóirat*, 107(2), 45-59
- Rajkó, R. (2003):** Discussant lecture for M. Hubert: Robust multivariate calibration. *Gordon Research Conferences (GRC) Statistics in Chemistry & Chemical Engineering. Mount Holyoke College South Hadley, MA (USA), July 27-August 1, 2003.* [as invited lecturer]
- Rajkó, R., K. István, G. Keresztury (2003):** Towards the solution of the solvent/eluent problem in HPLC/IR (FTC) determination by chemometric methods. *Advances in Chromatography and Electrophoresis - Conferentia Chemometrica ACE-CC2003*, Budapest, Hungary, October 27-29, 2003.
- Rajkó R., István K., Keresztury G. (2004):** A HPLC/FT-IR csatolt rendszerrel kapott mérési adatok kemometriai kiértékeléséről. *MTA Kémiai Kutatóközpont Kémiai Intézet*, Budapest, 2004. február 26.
- Tauler, R. (1995):** Multivariate curve resolution applied to second order data. *Chemometrics and Intelligent Laboratory Systems*, 30, 133-146.
- Tauler, R., Smilde, A., Kowalski, B.R. (1995):** Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution. *Journal of Chemometrics*, 9, 31-58.
- Vandeginste, B.G.M., Massart, D.L., Buydens, L.M.C., de Jong, S., Lewi, P.J., Smeyers-Verbeke, J. (1998):** *Handbook of Chemometrics and Qualimetrics: Part A and Part B.* Elsevier, Amsterdam.