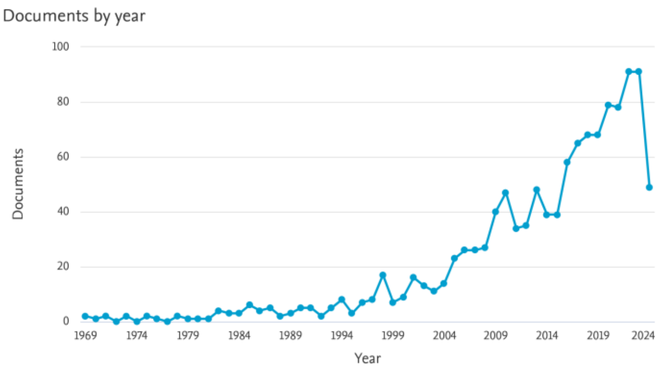
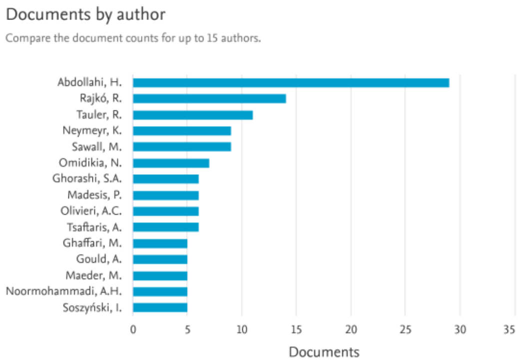
Válasz Dr. Abonyi János egyetemi tanár, az MTA doktora bírálatára

Köszönöm, hogy bírálóm végül elkészítette értékelését értekezésemről.

> A probléma tehát jelentős (amit az is igazol, hogy a feladat megoldására alkalmazott "Multivariate Curve Resolution" módszer széles körben alkalmazott), ugyanakkor SMCR módszer fejlesztésével kapcsolatban az érdeklődés nem csupán mérsékelt, hanem csökkenő tendenciát is mutat a Scopus adatbázis szerint (1. ábra).

Én is elvégeztem a Scopus adatbázisban a keresést, de az általánosabb ’unique AND curve AND resolution’ kulcsszavakkal:





A szakirodalomban ’self-modeling curve resolution, SMCR’ elnevezés főképp az analitikai kémiában, azon belül is a kemometria szakterületen terjedt el. Természetesen mátrixalapú mérési adatfeldolgozás a mátrixok dekompozicóját alkalmazva más tudományterületeken is elkezdődött és fejlődött más-más elnevezéssel: pl. ’blind source/signal separation, BSS’, ’independent component analysis, ICA’, ’dependent component analysis, DCA’, ’hyperspectral endmember extraction, HEE’, ’non-negative matrix factorization, NMF’, ‘positive matrix factorization, PMF’, ‘N-dimensional edge detection, NDEE’, ‘source apportionment by factors with explicit restrictions, SAFER’, ’polar qualification system, PQS’, stb.

> Ugyanakkor szeretném felhívni a figyelmet arra a más SMCR módszer kapcsán említett tényre, hogy a fejlesztett módszerek csak elenyésző mértékben vizsgáltak és alkalmazottak. Például a Google Scholar 79, Scopus 20 olyan publikációt regisztrált amelyben a dolgozat vizsgálatának alapját adó Borgen grafikon (Borgen plot) említésre kerül.

Önmagában a kevésnek tartott tanulmányszám nem jelenti a téma tudományos jelentőségének hiányát, én most a Boole algebrát említeném, amit a kortársak akkor teljesen felesleges és csak önmagának való matematikai fejtegetésnek tartottak, de szerencsére modern korunk már használni tudta pl. az információelmélet és a Neumann-elvű számítógépek megalkotása kapcsán. Bolyai Jánosnak is volt idevágó tapasztalata.

Irodalmi áttekintés

A harmadik fejezet mélyrehatóan foglalkozik a görbe nélküli komponensprofil-kinyerés irodalmával. Ez a legjobban megírt fejezet magába foglalja az elméleti alapok részletes tárgyalását, a használt matematikai háttér bemutatását. Ezt a fejezetet nagyon jól átgondoltnak így nagyon értékesnek tartom.

Köszönöm bírálom pozitív véleményét és a későbbiekben visszautalok majd erre.

Az ábrák sajnos nem konzisztensek a jelölsekkel (pl. 3 ábra bal-jobb oldala inkonzisztens).

A bal oldali ábra csak illusztráció, ahogy a 3. ábra aláírása is jelzi, tehát nem konkrét adatok felhasználásával készült, hanem „szabadkézi rajz”, viszont könnyen áttekinthető és értelmezhető, míg a jobb oldali ábra már konkrét számítási példa megjelenítése.

Rettentően nehéz olvashatóságot a jelölések bevezetésének és az egyenletek értelmezésnek hiánya eredményezi. Pl. a 4.3 fejezet amely a P3 cikk eredményeit ismerteti egyáltalán nem ismerteti, hogy mi a változók jelentése a 20 és 21 egyenletekben. A hiányos tárgyalásnak köszönhetően sajnos egyáltalán nem derül ki, hogy az SCF mérték miként függ a T rotációs mátrixtól (persze alapos megfontolásokkal minden visszafejthető, az eredeti közleményekben fellelhető, ugyanakkor véleményem szerint ez a munka egyáltalán nem várható el az olvasótól). A 2. táblázatot csupán a dolgozat alapján nem tudtam értelmezni.

Itt kell visszautalnom az ’Irodalmi áttekintés’-re vonatkozó dicséretre, mivel a szükséges egyenlet ott szerepel:

A PCA jelöléseket megtartva, megfelelő transzformáció után fizikailag és/vagy kémiailag értelmezhető megoldást kaphatunk az absztrakt helyett:

$$\underset{I\times J}{\mathbf{A}} = \left(\underset{I\times N}{\mathbf{X}^*}\underset{N\times N}{\mathbf{T}}\right)\left(\underset{N\times N}{\mathbf{T}^{-1}}\underset{N\times J}{\mathbf{V}^{*\mathrm{T}}}\right) + \underset{I\times J}{\mathbf{R}} = \underset{I\times N}{\mathbf{C}}\underset{N\times J}{\mathbf{S}^{\mathrm{T}}} + \underset{I\times J}{\mathbf{R}} \tag{5}$$

ahol $\mathbf{T}$ a transzformációs mátrix (négyzetes mátrix, amelynek létezik az inverze ($\mathbf{T}^{-1}$), ami a reciprokképzés általánosítása mátrixokra), $\mathbf{C}$ a koncentráció mátrix, $\mathbf{S}$ az egységnyi koncentrációhoz tartozó tiszta komponensek spektrumait tartalmazó mátrix.

A 2. táblázatra vonatkozóan: ’A rotációs bizonytalanság számszerűsítésére bevezettem a komponensek koncentráció és spektrum szerinti Borgen grafikonbéli megengedett megoldás-tartományainak területe és a külső poligon területe között számolt arányt, valamint a Gemperline-Tauler módszerrel kapott maximum és minimum SCF értékek különbségét, melyeket a 2. táblázatban szerepeltetek.’ … ’A 2. táblázatból rögtön szembe tűnhet, hogy a Gemperline-Tauler módszer komponensenként, a koncentráció és spektrum hatás megkülönböztetése nélkül nyújt információt (utolsó sor), míg a terület-arány külön a koncentrációra és külön a spektrumokra is meghatározható (1. és 2. sorok).’

Sajnos a 4.5 fejezet megértése is kihívásokat okozott számomra. Az „Elméleti fejtegetések .. „ cím igazán volt alkalmas, hogy felhívja a figyelmem arra, hogy mi is az az érdemi mondanivaló amelyre érdemes fókuszálni az olvasás közben.

Segítségül a tézisfüzetet kell megnézni:

'*5. TÉZISPONT*

Kétkomponensű rendszerek megengedett megoldás-tartomány határait értelmeztem és analitikus számítási módszert adtam meg [P5]

Korábban Abdollahi és mtsai egy kétkomponensű rendszer SMCR-éhez a transzformációs mátrix elemeinek kiszámításához és sávhatárok meghatározásához a Tauler-féle *mcrbands* Matlab szkriptet és a Maeder-féle rácsmódszert használták. Egyik módszer sem analitikus, hanem iteratív, meglehetősen nagy számítási igénnyel. Régóta köztudott, hogy Lawton és Sylvestre megközelítése analitikusan, tehát nem iteratívan is megadhatja a megengedett megoldás sávhatárait. Elsőként a szakirodalomban, egyértelmű kapcsolatot mutattam be Lawton és Sylvestre, Tauler, valamint Maeder megközelítései között.' stb.

A 9. ábra „A levezetésekhez és értelmezéshez szükséges illusztrációk" címe sem adott információt arra vonatkozóan, hogy mi az érdemi mondanivalója a fejezetnek. Az ábrát sajnos azért sem tudtam értelmezni, mert a bal és jobb oldali ábrán (feltehetőleg hibásan) ugyanazok a változók szerepelnek, teljesen más tartományban.

Nem hibás a 9. ábra bal és jobb oldala, egyszerűen a jobb oldalon kinagyítottam a bal oldali ábrát (bal oldalon 1 az x-tengelyen felvett maximum, jobb oldalon 0,15), hogy a szögek és a megfelelő pontok, változók jobban látszanak, segítsék a megértést.

A szűkös terjedelmet tovább csökkenti a fejezetben a teljesen felesleges MAtlab script amely semmilyen többlet információt nem tartalmaz.

A Matlab szkriptet a programozást jobban kedvelők számára közöltem, hogy a bemutatott képleteket megfelelően értelmezzék és ha használni szeretnék azt, kipróbálhatják, valóban adekvát az általam levezett elméleti fejtegetés más adatsorokra is. Egy programnyelvi kód megadása a tudományos munkában általában is jó gyakorlat, különösen alkalmas a közölt eredmények reprodukálására, segíti a lehetséges hibák felderítését és a hiba forrásának azonosítását, ill. nagyszerű segítség az esetleges általánosítás megsejtésére és kidolgozására is.

A 4.6 fejezet a P6 publikációt kivonatolja. Ebben az esetben három oldalból kellene, hogy kiderüljön a tudományos eredmény újdonságtartalma és elve. A fejezet első bejkezdése a SPSO optimalizációs algoritmus alkalmazásban rejlő lehetőségekre hívja fel a figyelmet. Sajnos egyáltalán nem derül ki a dolgozatból, hogy mi ennek a megoldásnak a jelentősége, célja, eredménye. Ma már számtalan populáció alapú optimalizációs lejárás érhető el. Miért ezt választotta a szerző? A 11. ábra egyáltalán nem értelmezhető számomra.

A természet inspirálta globális optimum kereső eljárások újból és újból reneszánszukat élik. A legismertebbek természetesen pl. a genetikus algoritmus, a szimulált hőkezelés

(simulated annealing) és a részecske raj optimalizálás (Particle Swarm Optimization, PSO) [X-S. Yang: Nature-Inspired Optimization Algorithms. Elsevier, London, 2014.]. Prof.Dr. Abdollahival közösen vezetett doktoranduszunk, Samira Beyramysoltan választotta a PSO-t a Matlabban lévő rácskereső függvény, az *fminsearch* használata helyett. Ő volt az aki utánanézett az irodalomban és az SPSO-t választotta, mint multimodális optimalizáló eljárás, mivel nem csak egyetlen globális optimumot kellett találunk. Akkori legjobb tudásunk szerint ez volt az első olyan módszer, amely képes volt az összes összetartozó megoldást egyszerre megadni. A klaszterek (komponensek megengedett megoldás-tartományai) egy-egy jellegzetes pontját meg tudjuk adni a külső politópok megfelelő pontjaival. Itt a probléma természetesen az, hogy a többszörös globális optimumok diszjunkt, de egyenként összefüggő területek. Az eredmények azt mutatták, hogy a módszer alkalmas az adatok elemzésére, azonban a módszer meglehetősen időigényes. A kifejlesztett módszerrel az egyenlőség kényszerfeltételt (ismert koncentráció vagy spektrális profil figyelembe vétele: ismert-profil) is alkalmaztuk és várakozásunknak megfelelően a forgatási bizonytalanság és a megengedett megoldás-tartományok drasztikusan csökkentek. Itt mutattuk meg részletekbe menően, hogy az ismert-profil kényszerfeltétel 3 komponensű rendszerekben a komplementer komponensekre vonatkozóan egy lineáris egyenes megfelelő szakaszaira egyszerűsödő megengedett-tartományokat eredményez. Ezek a megállapítások fontosabbak voltak, mint az SPSO alkalmazásának részletes elemzése.
A 11. ábra a megjelent cikk grafikus összefoglalója, sajnálom, hogy értelmezési problémát okozott.

> A 4.7 fejezet „Újabb fejlesztések … „ címet viseli, amely cím sajnos ismét nem informatív. Az eredetileg 14 oldalas publikáció eredményeinek pár oldalas képeskönyvként történő ismertetése véleményem szerint szokatlan egy MTA doktora értekezés stílusától.

Most is a tézisfüzetet kell megnézni:
'*7. TÉZISPONT*

Az ismert-profil és unimodalitás kényszerfeltételek alkalmazását dolgoztam ki dualitás alapú görbeillesztés nélküli komponensprofil-kinyeréshez [P7]

A görbeillesztés nélküli komponensprofil-kinyerés (SMCR) módszerei a mátrix alapú adatkészletekhez a teljes megengedett megoldás-tartományokat adják meg csak nemnegatív kényszerfeltétel alkalmazásával. Úgy tűnik, hogy konvex geometriai szemlélet is szükséges a görbefelbontó módszerek megértéséhez és fejlesztéséhez, mivel a gyakran komplikált (lineáris) algebrai koncepciók kizárólagos alkalmazása megakasztotta a Borgen módszer alaposabb megértését 20 éven keresztül. Az analitikai SMCR módszereket vizsgáltuk felül és írtuk le a dualitás alapelveihez kapcsolódó egyszerű fogalmak segítségével. Ezenkívül, először a szakirodalomban, az egyenlőség (ismert-profil) és unimodalitás kényszerfeltételeket sikeresen implementáltuk a Lawton–Sylvestre módszerhez. A Borgen grafikonok megrajzolásához is egy korszerűbb eljárást fejlesztettünk az ismert-profil kényszerfeltétel alkalmazásához. Két- és háromkomponensű HPLC-DAD adatsorokat szimuláltunk és elemeztük őket az újabb kényszerfeltételek felhasználásával és azok nélkül. Az *LSandBP* Matlab szkriptünk a dualitás elvén alapuló egyszerűsítéseket tartalmazza így a Borgen grafikonok néhány másodperc alatt elkészültek. A dualitás elvén alapuló egyszerűsítés pl., hogy az egyik absztrakt térben (mondjuk az oszlop térben) konvex burok algoritmussal meghatározott belső poligon a duális absztrakt térben (mondjuk a sor térben) a külső poligont határozza meg, ami igaz fordítva is, azaz a másik absztrakt térben (mondjuk a sor térben) meghatározott belső poligon az első absztrakt tér (mondjuk az oszlop tér) külső poligonját rögzíti. Így nem kell a számításigényes dupla-deskripció (DD) (*cddmex*) számítógépes geometriai algoritmust igénybe vennünk. A másik egyszerűsítés, hogy elegendő csak az egyik absztrakt térben meghatározni a Borgen szimplexeket (Borgen háromszögek három komponens esetén), mert a dualitás elvének alkalmazásával a másik absztrakt térben iteráció nélkül megkapjuk azokat.'

A 4.10 fejezet önmagában szintén érthetetlen volt számomra. A változók jelentése egyáltalán nincs ismertetve.

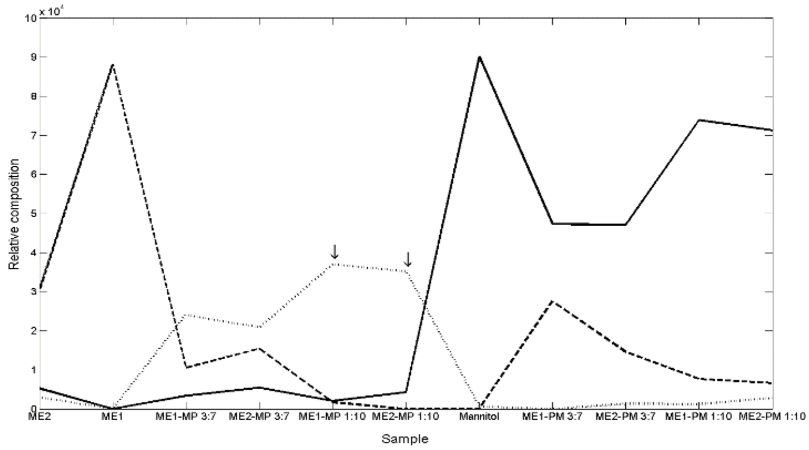
Az értekezésben a következőképpen definiáltam a változókat (*F* nyilvánvalóan a komponensek számát jelkenti): 'Kruskal által bevezetett egyenlőtlenség alkalmazható. Ten Berge és Sidiropoulos [47] megmutatta, hogy három-utas CP modellt feltételezve $F = 2$ és $F = 3$ esetén a ($k_E + k_S + k_C \geq 2F + 2$) Kruskal egyenlőtlenség teljesülése szükséges és elégséges feltétele az egyértelmű felbontásnak, viszont $F > 3$ esetén már nem lesz szükséges. Itt k_E, k_S, és k_C az egyes **E**, **S**, **C** lódingú (loading) mátrixok ún. *k*-rangja, azaz egy mátrix minden lehetséges módon kiválasztott *k* oszlopa lineárisan független, de van legalább egy *k*+1 db oszlopból álló részmátrix amely oszlopai már nem lesznek lineárisan függetlenek. A következő jelölést vezetjük be: $^k\{F_1,F_2,F_3\}$ jelentse, hogy $k_E=F_1$, $k_S=F_2$, és $k_C=F_3$.' stb.
Az **E**, **S**, **C** lóding mátrixok az ábra feliratokból következtethetően mást és mást jelenthetnek az a), b), c) alattiaknak megfelelően:
'22. ábra Kétkomponensű szimulált másodrendű kalibráció spektro–pH-metriás titrálás-sal. a) abszorpciós spektrumok, b) pH-függő összetétel profilok, c) relatív kezdeti koncentrációk a különböző mintákban'
'23. ábra Háromkomponensű szimulált HPLC-DAD adattömb. a) abszorpciós spektrumok, b) elúciós profilok, c) minták szerinti koncentráció profilok'
'24. ábra Háromkomponensű szimulált másodrendű kalibráció spektro–pH-metriás titrálással. a) abszorpciós spektrumok, b) pH-függő összetétel profilok, c) relatív kezdeti koncentrációk a különböző mintákban'
'25. ábra Háromkomponensű szimulált gerjesztett emissziós spektrum (EEM). a) ger-jesztő spektrumok, b) emissziós spektrumok, c) minták szerinti koncentráció profilok'
'26. ábra Háromkomponensű szimulált gerjesztett emissziós spektrum (EEM). a) ger-jesztő spektrumok, b) emissziós spektrumok, c) minták szerinti koncentráció profilok'

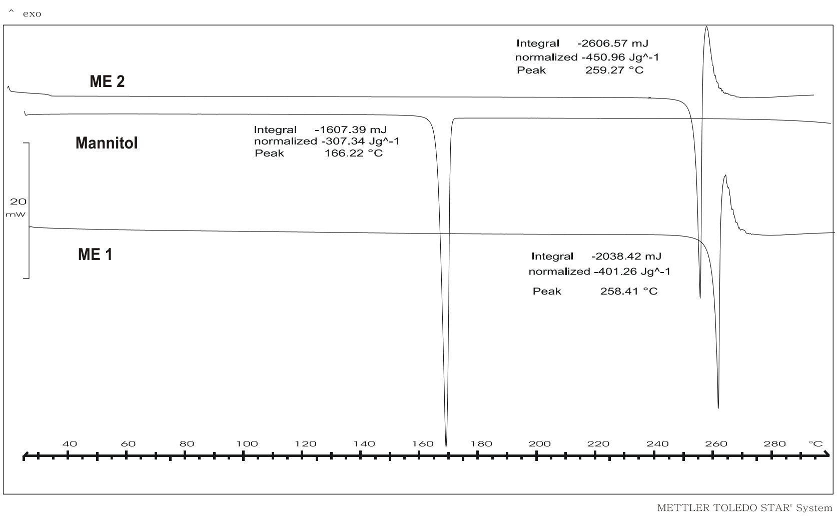
Az 5.2 alfejezetnél már tényleg szembesültem azzal a ténnyel, hogy a dolgozat teljes mértékben olvashatatlan és érthetetlen számomra. A P14 cikket a szerző három bekezdésben ismereti, két számomra teljes mértékben áttekinthetetlen ábrával.

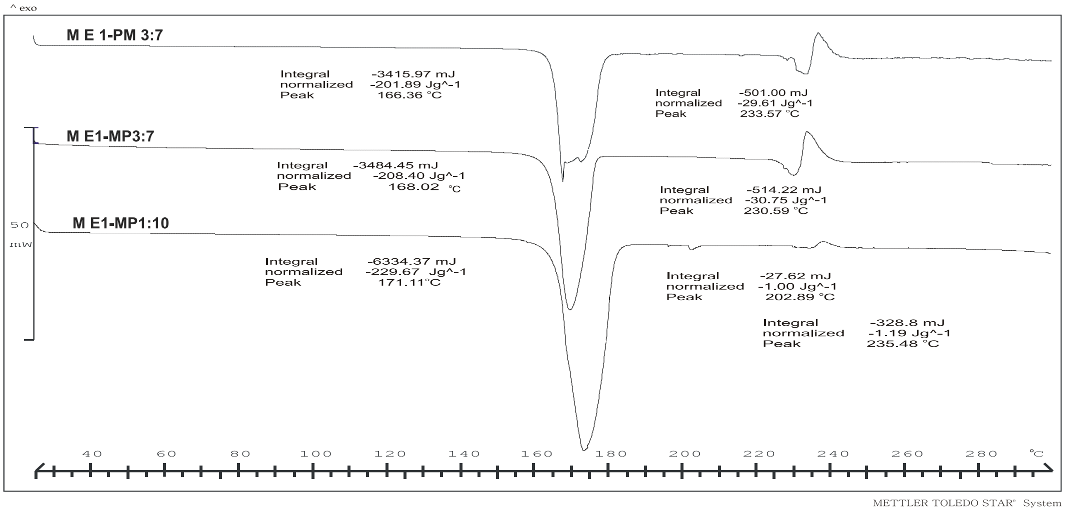
Az ábrák természetszerűleg az adatmátrix sortereit mutatják, a nullterek ezekre ortogonálisan helyezkednek el, és ezek vetítései lesznek az MCR-ALS iterációs trajektóriái, amik a külső poligonon kívülre (így negatív elemeket is tartalmazni fognak) vagy a megengedett tartományokon kívülre kerülnek. A zaj (mérési hiba) a mátrix nullterét és sorterét összezavarja, a zajmentes mátrix nulltér részlete átkerülhet a sortérbe és a zajmentes mátrix sortere átmosódhat a nulltérbe. Ezáltal az MCR-ALS algoritmus stabilizálódhat és gyorsabban konvergálhat, hiszen ha nincs összehasonlító standard profilunk, csak a kényszerfeltételek teljesülését tudjuk ellenőrizni, amik torz becslésekre is teljesülhetnek. A gyakorlati szakemberek nem tartják jelentősnek az MCR-ALS nulltér konvergenciáját, hiszen ritkán tudunk zajmentes adatmátrixokat műszeresen produkálni, azonban össze kell vetnünk ezt a kérdést a hipotézisvizsgálattal: ha pl. a generikus és az eredeti gyógyszer hasonlóságát kell kimutatnunk, akkor sikeresebbek lehetünk, ha nagy szórással mért adatokat használunk, mivel így a nullhipotézist nagyobb eséllyel fogadjuk el, de csakis a megbízhatatlan mérések miatt.

> Az 5.5 fejezet már csak egy oldalban vázolja a P17 publikáció tartalmát, amely közlemény véleményem szerint semmilyen új tudományos eredményt nem tartalmaz a módszertan fejlesztése szempontjából.

Másik két bírálóm elfogadta az MCR-ALS alkalmazását tudományos eredményként, én nem is akartam többet demonstrálni ennél az értekezésemben, kemometriai módszertani fejlesztésről nem is volt szó. Azonban Parya Reisi Nassab a tanulmány első szerzője PhD disszertációjában megvédte a tanulmányban állítottakat, mivel a DSC görbék összehasonlításából egyértelműen kirajzolódik a blend megjelenése.
A lenti 1. komponens-profil ábrán az ME1-MP1:10 és ME2-MP1:10 esetében a meloxikám (szaggatott görbe) és a mannit (folytonos vonal) befolyása szinte elhanyagolható, ahogy a lenti 2. DSC vizsgálatnál az ME1-MP1:10 görbéje az elvándorolt és kiszélesedő csúcs után kisimul, teljesen eltérve a lenti 1. DSC vizsgálatnál tapasztaltaktól. Nyilván lehetne részletesebben is elemezni a DSC görbéket, de ez a kvalitatív kép is elegendő igazolás kell legyen, hogy új tudományos eredményt értünk el az MCR-ALS megfelelő alkalmazásával.







> A hetedik fejezet az eredmények további hasznosítási lehetőségeit tárgyalja. Sajnos ez a fejezet is egy ömlesztett kivonata a szerző dolgozat érdemi részéhez nem kapcsolódó szakmai tevékenységének. A sok bonyolultnak tűnő ábra, megfogalmazás egyáltalán nem a tudományos alaposságot és igényességet tükrözi. Az 50. ábrának már cím sem jutott. Ezeket a munkákat nem is értem, hogy miért szerepelteti a szerző (pl. TDK munka ismertetése és a kapcsolódó triviális ábra a többváltozós adatelemzés jelentőségéről egyáltalán nem a dolgozathoz illő). Az itt felsorolt információk többsége nem kapcsolódik a vizsgált tématerülethez. Ez a fejezet a címével ellentétben tényleges kutatási kérdést nem vet fel, kutatási irányt nem jelöl ki.

Az 50. ábra hiányzó aláírása: 'Bal oldalon a PLS Toolbox PLSDA funkciójával kapott előzetes eredmény a teljes adatkészleten – főkomponens-együttható ábra (score-plot, 3D absztrakt tér), jobb oldalon a permutációs teszt eredménye annak ellenőrzésére, hogy a PLSDA modell nem generálható véletlenszerűen'.

Egy a nyugdíjkorhatárt közelítő egyetemi oktató-kutatótól elég különös elvárás, hogy perspektivikusan több évtizedre fogalmazzon meg előre kutatási irányokat, főképpen úgy, hogy tudományos kutatási pályázatait rendszerint elutasítják.

Persze folytatom tudományos tevékenységemet és igen, fontosnak tartom a múltbeli tudományos sikereket, így OTDK eredményeket is, amit a tanítványaim elértek.

Van még megoldatlan kemometriai probléma és éppen mostanában jelentkezik doktori képzésre Zanjanban (IASBS) egy újabb iráni hallgató hölgy, akit Prof.Dr. Hamid Abdollahival közös témavezetéssel segítünk.

> Az első tézis nem fogalmaz meg egy egyértelmű állítást, csupán a kutatáshoz kapcsolódó tevékenységet és az érintett cikk tartalmát ismerteti. Ennek köszönhetően formai szempontból sajnos nem tudom elfogadni a tézist, holott a kapcsolódó publikációban szereplő eredményeket jelentősnek tartom, azok véleményem szerint tézisértékűek.

A tézisfüzetben szereplő néhány soros tézispontot kell figyelembe venni és a bővebb összefoglaló leírást:

1. TÉZISPONT

Számítógépes geometriai módszerekkel a megengedett megoldás-tartományokat analitikusan határoztam meg háromkomponensű rendszerekben

'Ún. számítógépes geometriai módszereket alkalmaztam a kemometriai irodalomban elsőként, így egy igen gyors Matlab implementációt sikerült készítenem. Az általam Borgen grafikonnak (Borgen plot) elnevezett ábra elkészítéséhez először az adatmátrixunk normalizálását végeztem el az egyes sorok, ill. oszlopok összegeit felhasználva. A normalizálás okozta dimenzió-csökkenés miatt az absztrakt 2-dimenziós síkon lévő adatpontok konvex burkának meghatározását a *convhull* Matlab függvénnyel végeztem (első számítógépes geometriai módszer). Megkaptuk a belső poligont (inner polygon), amit egy megengedett megoldás szimplexe - vagyis három komponens esetén háromszög -, egyetlen oldala sem metszhet, legfeljebb érintheti. A Fourier-Motzkin eliminációs módszer (FMEM) egy lineáris egyenlőtlenség-rendszerből a redundáns egyenlőtlenségeket szűri ki. A redundáns egyenlőtlenségek észlelése a csúcsok és szélső irányok listájának kezelésével a dupla-deskripció (double description (DD)) módszerrel történt, amit a Matlabban a *cddmex* harmadik fél által fejlesztett C/C++ alacsony szintű programnyelven készített és előre lefordított külső függvénnyel valósítottam meg (második számítógépes geometriai módszer). A nem redundáns félsíkok a külső poligont (outer polygon) határozzák meg. A megengedett megoldás-tartományok a belső és a külső poligonok között keresendők, tehát a nemnegatív profilok absztrakt pontjai a külső poligonon belül helyezkednek el, és egy háromszög csúcsaként a belső poligont körbeveszik, azaz lineáris konvex kompozícióként az adatpontok előállíthatók. Ebben a tanulmányban vezettük be az Average Orthogonal Distance from the Linear Subspace (AODLS) elnevezésű mértéket, ami egy ponthalmaz átlagos ortogonális távolságát adja meg egy lineáris altértől. Az AODSL segítségével kimutattuk, hogy a Lawton és Sylvestre által használt adathalmaz akár négy komponenst is tartalmazhat kettő helyett.'

Kérem, gondolja át bírálom és fogadja el az 1. tézispontot is.

> A 2-12 téziseket a 9-es tézis kivételével elfogadom, jelezve, hogy ezek megfogalmazása sok esetben nem ideális, hasznosabb lett volna az eredményeket strukturálni, téziscsoportokba foglalni.

Köszönöm, hogy bírálom elfogadta a 2-8. és a 10-12. tézispontokat.

> A 9 tézis egy többszerzős irodalmi áttekintéshez kapcsolódik. A módszerekre vonatkozó megállapítások korábbi publikációkban már ismertetésre kerültek, nem találtam olyan megállapítást, mely a szerzőhöz egyértelműen köthető lett volna és tézisértékű új tudományos eredményként lehetne elismerni.

Másik bírálóm, Dr. Tóth Gergely megfogalmazásánál nem tudok jobbat, így őt idézem (4.9 alfejezet – P9 cikk (2016) – 9. tézispont - BP): „*Kutatói berkekben nagyon ritka az ilyen összefogás egy területen, főleg, ha mint egymással konkurens módszerek fejlesztőiként nézzük őket. Tekinthetjük összefoglaló cikknek az MCR és az SMCR területéről, de fontosabb a cikk összehasonlító jellege. Három módszert vizsgálnak és minősítenek. Rajkó Róbert korábbi munkája az etalonnak tekinthető Borgen-ábrák részleteinek felderítésével esszenciális volt ennek a munkának a létrejöttéhez.*”

Kérem, gondolja át bírálom és fogadja el a 9. tézispontot is.

A 15 tézis egy „Comments on “Near-Infrared Hyperspectral Unmixing Based on a Minimum Volume Criterion for Fast and Accurate Chemometric Characterization of Counterfeit Tablets” publikációhoz kapcsolódik. A szerzővel ellentétben úgy gondolom, hogy egy MTA doktora dolgozat esetében jóval jelentősebb tudományos eredményeket kell hogy tézisekbe foglaljunk mint egy publikációval kapcsolatos kiegészítés, amely felhívja a figyelmet az alkalmazott MCR-ALS módszer ismert hátrányaira.

Azért tartottam fontosnak ezt a tézispontot megfogalmaznom, mert rávilágít arra az esetre, amikor egy már elismert, de az analitikai kémia területétől távol eső kutatónak (José M. Bioucas Dias) a szerkesztők még akkor is igazat adnak, amikor egyértelműen hibáztak.

’Elsőként azt jegyeztem meg, hogy a tiszta pixel nélküli hiperspektrumok esetére Craig [61] földtudományok területén megjelent munkáját hivatkozták, mivel ő az absztrakt térben az adatpontokat tartalmazó minimális térfogatú szimplex csúcspontjait javasolta becslésként. Azonban Perczel et al. [62,63] már Craig előtt javasolták ugyanazon elven alapuló CCA (convex constraint analysis) módszerüket kémiai szaklapban megjelent cikkeikben.’ Erre azt a választ adták, hogy ők találtak egy korábbi Craig előadást, de egy még korábbi Perczel cikket is: Perczel, A., Hollósi, M., Tusnady, G., and Fasman, D. *Croatica Chim. Acta* 1989, 62, 189–200. https://hrcak.srce.hr/en/175392

Nos, van egy még korábbi, konferencia kiadványban megjelent publikáció a témában ugyanattól a csoporttól: G. Tusnády, A. Perczel, M. Hollósi: Mixture decomposition on convex constrains for CD curves of proteins. pp. 340-346 in *Second International Conference on Circular Dichroism*, August 15-18, 1987, Budapest, Conference proceedings ed. by M. Kajtár; publ. by the Institute of Organic Chemistry L. Eötvös University.

’Másodsorban arra hívtam fel a figyelmet, hogy a forgatási bizonytalanság nem az MCR-ALS vagy más algoritmus tulajdonsága, hanem a bilineáris adatmátrix belső sajátja!’ Amit bírálóm ma már triviálisnak tarthat, de akkoriban, 15 évvel ezelőtt főképp a villamos-mérnöki/informatikai (IEEE) szakirodalomban egyáltalán nem volt az, de még előfordult, hogy a kemometria területén sem. Az NMF (nonnegative matrix factorization) területén csak mostanában került előtérbe és az értekezésben 100-as sorszámmal hivatkozott cikkben vezettük be a ’partial identifiability’ fogalmát, mert addig csak a teljes identifikációt (a teljesen egyértelmű felbontást szolgáltató NMF megoldásokat) tanulmányozták.

’Lopes et al. a nemnegativitást alkalmazták kényszerfeltételként, valamint egy szükséges feltételt: „a szimplex minden (p-1) dimenziós oldala, azaz lapja (facet) tartalmazzon p-1 spektrális adatvektort”.’ … ’Rávilágítottam, hogy ez olyan erős megkötés, ami csak ritkán fordul elő a gyakorlatban.’ Ezzel arra céloztam, hogy az analitikai kémiában nem az erőltetett kényszerfeltételeket kell alkalmazni pusztán azért, hogy egyértelmű megoldást kapjunk. A téveszme, hogy akár absztrakt matematikai kényszerfeltételek által is, de egyértelmű megoldást kapjunk a kemometriában manapság is hódít.

’Továbbiakban illusztráltam, hogy a koncepció, amit használtak nem minden esetben ad elfogadható eredményt: nemnegatív kényszerfeltétel alkalmazása mellett a minimális térfogatú szimplex (három komponensű rendszerek esetén minimális területű háromszög) a duális térben negatív értékeket is tartalmazó profilok absztrakt pontjait eredményezheti.’

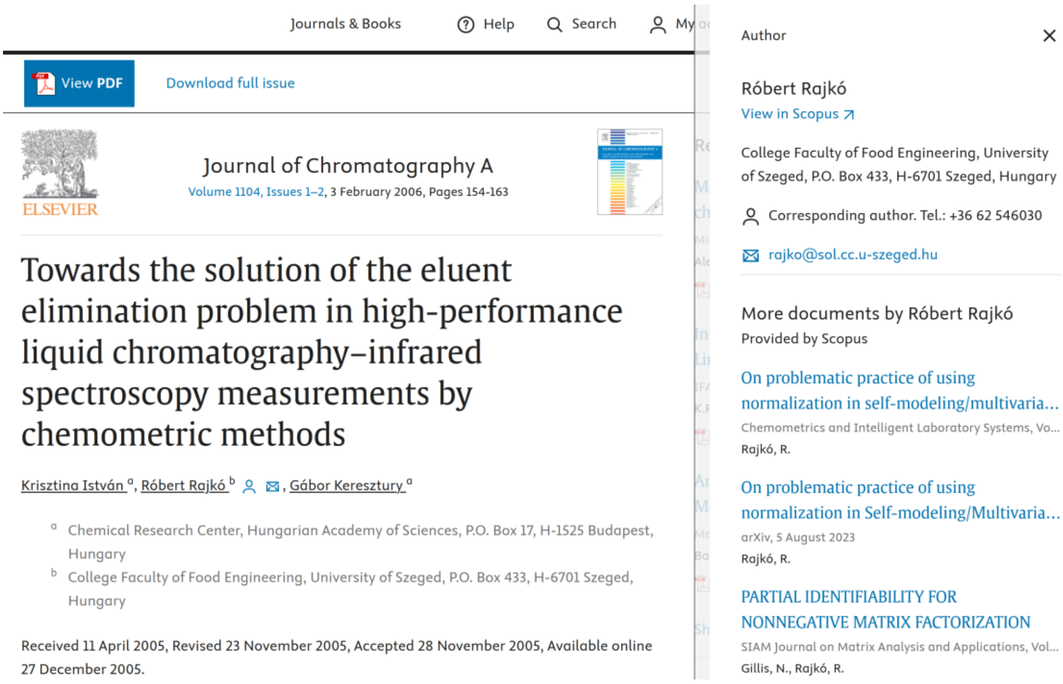
Bevezetnek egy a gyakorlatban ritkán teljesülő kényszerfeltételt, de azzal nem foglalkoznak, hogy a triviális nemnegativitás sérülhet.

'Egy további kritikus problémaként az alkalmazott normalizálást jártam körül.' Nem megismételve a részletes levezetést itt csak annyit jegyzek meg, hogy legújabb cikkemben ismét elővettem ezt a normalizálási problémát is: R. Rajkó: [On problematic practice of using normalization in self-modeling/multivariate curve resolution (S/MCR)](), *Chemometrics and Intelligent Laboratory Systems* 244: 105033, 2024.

Kérem, gondolja át bírálom és fogadja el a 15. tézispontot is.

> A 16-17 téziseket nem fogadom el. A kapcsolódó publikációkban és alfejezetekben nem találtam a módszertanhoz kapcsolódó érdemi eredményt, a közleményekben a szerző nem a meghatározó első vagy utolsó helyen szerepel.

A 16. tézispontnál valami fatális félreértés történt, hiszen annak alapját képző P17-es cikknek én vagyok a levelező szerzője https://doi.org/10.1016/j.chroma.2005.11.131:


Journals & Books Help Search My ac

View PDF Download full issue

ELSEVIER

Journal of Chromatography A
Volume 1104, Issues 1–2, 3 February 2006, Pages 154-163

Towards the solution of the eluent elimination problem in high-performance liquid chromatography–infrared spectroscopy measurements by chemometric methods

Krisztina István [a], Róbert Rajkó [b], Gábor Keresztury [a]

[a] Chemical Research Center, Hungarian Academy of Sciences, P.O. Box 17, H-1525 Budapest, Hungary
[b] College Faculty of Food Engineering, University of Szeged, P.O. Box 433, H-6701 Szeged, Hungary

Received 11 April 2005, Revised 23 November 2005, Accepted 28 November 2005, Available online 27 December 2005.

Author

Róbert Rajkó
View in Scopus

College Faculty of Food Engineering, University of Szeged, P.O. Box 433, H-6701 Szeged, Hungary

Corresponding author. Tel.: +36 62 546030

rajko@sol.cc.u-szeged.hu

More documents by Róbert Rajkó
Provided by Scopus

On problematic practice of using normalization in self-modeling/multivaria...
Chemometrics and Intelligent Laboratory Systems, Vo...
Rajkó, R.

On problematic practice of using normalization in Self-modeling/Multivaria...
arXiv, 5 August 2023
Rajkó, R.

PARTIAL IDENTIFIABILITY FOR NONNEGATIVE MATRIX FACTORIZATION
SIAM Journal on Matrix Analysis and Applications, Vol...
Gillis, N., Rajkó, R.

A HPLC/FT-IR mérések során (is) előfordul az a szerencsétlen helyzet, hogy az eluensnek is van jele, spektruma. Mivel tudtuk a tiszta eluens(ek) spektrumait, akkori tudásommal nem értettem, miért nem tudjuk kinyerni az eluens elúciós profilját. Ehhez kellett elmerülnöm a Borgen grafikonok előállításába, de előtte még kifejlesztettük az OSSS-IU-PARAFAC és OSSS-IU-PARAFAC2 módszereket. A probléma jelentőségét bizonyítja, hogy Julia Kuligowski és mtsai. a Scopus szerint 12 tanulmányban hivatkozták cikkünket, de nekik sem sikerült végleges megoldást találniuk:
https://www.scopus.com/results/citedbyresults.uri?sort=plf-f&cite=2-s2.0-30744459352&src=s&nlo=&nlr=&nls=&imp=t&sid=f205ac6a608390b4b890e4fa9a2277f3&sot=cite&sdt=cl&cluster=scoprefnameauid%2c%22Kuligowski%2c+J.%2323492474900%22%2ct%2c%22Lendl%2c+B.%2336150760000%22%2ct%2c%22Tauler%2c+R.%237005428187%22%2ct&sl=0&origin=resultslist&zone=leftSideBar&editSaveSearch=&txGid=6894ca297d0211d5d6875c4ea5ac9c7e

Ez az egyik legjelentősebb hozzájárulásom a többutas kemometriai módszerekkel kapcsolatban, sajnos a csoportunk együttműködését és a probléma általunk való megnyugtató lezárását az első szerző fiatalon bekövetkezett halála megakadályozta.
Kérem, gondolja át bírálom és fogadja el a 16. tézispontot is.

A 17. tézisponttal kapcsolatban megismétlem a fentebb leírtakat az ábrák nélkül:
Másik két bírálóm elfogadta az MCR-ALS alkalmazását tudományos eredményként, én nem is akartam többet demonstrálni ennél az értekezésemben, módszertani fejlesztésről nem is volt szó. Azonban Parya Reisi Nassab a tanulmány első szerzője PhD disszertációjában megvédte a tanulmányban állítottakat, mivel a DSC görbék összehasonlításából egyértelműen kirajzolódik a blend megjelenése. Az ME1-MP1:10 és ME2-MP1:10 esetében a meloxikám (szaggatott görbe) és a mannit (folytonos vonal) befolyása szinte elhanyagolható, ahogy a DSC vizsgálatnál az ME1-MP1:10 görbéje az elvándorolt és kiszélesedő csúcs után kisimul. Nyilván lehetne részletesebben is elemezni a DSC görbéket, de ez a kvalitatív kép is elegendő igazolás kell legyen, hogy új tudományos eredményt értünk el az MCR-ALS megfelelő alkalmazásával.
Kérem, gondolja át bírálom és fogadja el a 17. tézispontot is.

A 18-20 téziseket elfogadom.

Köszönöm, hogy bírálom elfogadta a 18-20. tézispontokat.

Kérdések

A hatodik tézisben ismertetett módszer kapcsán az SPSO optimalizációs algoritmust alkalmazta. Sajnos a cikkben sem találtam utalást arra, hogy mi alapján határozta meg a részecskék számát? Összevetést sem találtam más optimálási algoritmusokkal. Mi a módszer érdemi előnye? Ma milyen technikát választana?

A 12-es tézisben/alfejezetben/publikációban ritkás profilok meghatározására fokuszál. Kérem tekintse át a számítástudományban a ritkás mátrixok kezeléséhez kapcsolódó kutatásokat és vesse össze ezekkel a munkáját. Kérem taglalja, hogy mennyire is tekinthetők ritkásnak az Ön által vizsgált problémák, illetve mennyire is igényeltek speciális algoritmusokat és megfontolásokat.

1) Prof.Dr. Abdollahival közösen vezetett doktoranduszunk, Samira Beyramysoltan választotta a PSO-t a Matlabban lévő rácskereső függvény, az *fminsearch* használata helyett. Ő volt az aki utánanézett az irodalomban és az SPSO-t választotta, mint multimodális optimalizáló eljárás, mivel nem csak egyetlen globális optimumot kellett találnunk (ez volt az előnye). Itt a probléma természetesen az, hogy a többszörös globális optimumok diszjunkt, de egyenként összefüggő területek. Akkori legjobb tudásunk szerint a módszerünk volt az első, amely képes volt az összes összetartozó megoldást egyszerre megadni. Az eredmények azt mutatták, hogy a módszer alkalmas az adatok elemzésére, azonban a módszer meglehetősen időigényes.

A részecskék pozícióinak változása megadható a sebességükkel: $v_{new} = wv_{old} + c_1r_1(p_{best} - p) + c_2r_2(p_{gbest} - p)$, majd frissítjük a pozíciókat: $p_{new} = p + v_{new}$.
Pl. tegyük fel, hogy a 2. komponenst vizsgáljuk, így a rácskeresési módszer minden iterációjakor a **T** normalizált mátrix második komponensének koordinátája rögzítetten marad, az SPSO algoritmus részecskéinek az első és a harmadik komponens koordinátáinak kell lenniük, ezért a véletlenszám-generátor segítségével 150 mátrixot (2×2) állítottunk elő az absztrakt tér tartományában. A c_1, c_2 és r_s az SPSO algoritmusban 1,2-re, 0,8-ra és 0,08-ra volt beállítva, és a w lineárisan csökkent 0,9-ről 0,4-re az iteráció változásával együtt (a maximális iteráció=200).
A PSO-t azóta nem használtuk és nem is tervezem használni.

2) Kétkomponensű rendszerek esetén a legritkásabb megoldások a külső határolópontok által meghatározott profilok lesznek és ez triviálisan igaz. Azt hangsúlyoztuk ki, hogy pl. a LASSO csak egyváltozós esetben találja meg a legritkásabb megoldást az L_1 és L_2 normák kombinálásával kapott célfüggvény minimalizálásával, többváltozó esetében (a triviális nemnegatív kényszerfeltétel alkalmazásával) már egyes komponensekre éppen az ellentétes (nullát egyáltalán nem tartalmazó) megoldásokat szolgáltatja.
Itt egyébként az a félreértés érhető tetten, mint az SMCR negligálásánál, nevezetesen nem szolgáltat egyértelmű megoldást. A legritkásabb megoldás megkövetelése (kényszerfeltételként) a legritkább esetben természetes, a nullák nem maximálisan vagy minimálisan helyezkednek el egy koncentráció és/vagy spektrális profilban, hanem a természet által előre rögzítve. Csakhogy ezt a számot nem tudjuk előre megjósolni, így pl. a maximális ritkásság megkövetelése pusztán matematikai műtermék lesz.

Végül köszönöm bírálómnak, hogy a doktori művel kapcsolatos kritikai észrevételei ellenére az értekezés nyilvános vitára bocsátását és sikeres védés esetén számomra az MTA doktora cím odaítélését javasolta.

Szeged, 2024. június 26.

Prof.Dr. Rajkó Róbert