

Válasz Dr. Szederkényi Gábor az MTA doktora bírálatára

Nagyon köszönöm a bíráló alapos, körültekintő munkáját, kérdéseit és pozitív véleményét.

A bírálat első részében a dolgozat rövid, összefoglaló ismertetése található, amit nagy örömmel olvastam, mivel újszerű nézőpontból találkozhattam a már általam jól ismert munkámmal. Köszönöm, hogy megállapította miszerint a formátum és a szerkesztési munka igényes, az ábrák informatívak, elgépelés szinte egyáltalán nem található, a matematikai formulák jól értelmezhetők, a mátrixok és vektorok dimenzióinak következetes jelzése is segíti a megértést.

A bíráló kérdései és válaszaim:

1. Kérem, a számítási komplexitások összehasonlításával fejtse ki, miért hatékonyabb a 4.1 szakaszban leírt számítógépes geometriai megoldás mint a mások által alkalmazott lineáris programozáson alapuló megközelítés. Hogyan skálázódik ez a számítási előny az adathalmaz dimenziójának növekedésével?

A 4.1 fejezetben két számítógépes geometriai eljárás bemutatása és használata történt meg: konvex burok (convex hull) és dupla deskripció (double description, DD). A 32-es és 33-as számokon hivatkozott lineáris programozást (LP) használó tanulmányokban a belső határok (inner boundary) meghatározását sokkal nehezebb feladatként látták, mint a külső határok (outer boundary) meghatározását, hiszen a külső határokat a nem-negativitás kikényszerítése adta meg, ami a lineáris programozás során kapott megoldáshalmaz aktív feltételek által meghatározott burka (most nem kell optimális megoldás(oka)t kiválasztanunk ezek segítségével!). A belső határok által körülzárt terület csökkenti a megengedett megoldások halmazát, amit egy másik LP feladattal oldottak meg és így a konvex burkot hipersíkok metszeteiként kapták meg. Mi a Matlabban implementált *convhull* függvényt használtuk, ami csúcspontokkal adta meg ugyanezt a belső konvex politópot, ami lényeges számítási sebességnövekedést jelentett. A külső és a belső politópok meghatározása azonban még nem jelenti a feladat végét és végül kiderült, hogy Kim és Henry [32,33] csak a konvex „engedélyezett területeket” (permitted areas) tudták meghatározni, amelyeken belül helyezkednek el az általában nem-konvex megengedett megoldás-tartományok (feasible regions).

A DD módszer a Fourier-Motzkin eliminációs módszer megvalósítása, azaz a redundáns egyenlőtlenségek észlelése és kihagyása. Amennyiben olyan LP implementációt alkalmazunk, ami nem kezeli a redundáns egyenlőtlenségeket, akkor többlet számítások fordulhatnak elő, ill. a kerekítési hibák akkumulációival az eredményeink is torzulhatnak. Minél nagyobb skálájú az LP feladat, annál kritikusabb lesz ez a hatás. Megítélésem szerint az adathalmaz dimenziójának növekedésével exponenciális mértékben romolhat az LP teljesítménye a DD-hez képest.

2. 4.6 szakasz: Az adatok ismeretében tudunk előzetes becslést adni a fizibilis megoldások számára? Van rá (elméleti) garancia, hogy az alkalmazott SPSO algoritmus valóban megtalálja valamennyi fizibilis megoldást?

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 találnunk. 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, mivel pl. nincs elméleti garancia, hogy az alkalmazott SPSO algoritmus valóban megtalálja valamennyi fizibilis megoldást. Természetesen elegendő lenne csak a határfelületek összes pontját meghatározni, mert a belső pontok ezek segítségével számolhatók lesznek; azonban ezek a módszerek inkább a klaszterek középpontjait igyekeznek meghatározni. Ezután már nem kíséreltünk tovább globális optimumkereső eljárások alkalmazásával.

3. 5.2 szakasz: Hogyan befolyásolja a becslés minőségét az eredeti (módosítatlan) MCR-ALS algoritmusnál az engedélyezett iterációk száma? (Pl. több iterációt engedve javulhat-e a megoldás fizibilitása?)

A becslés minőségét (a becslés egyre közelebb lesz egy-egy fizibilis megoldáshoz) természetesen javítja az iterációk számának növelése. Az MCR-ALS-nél elég egyetlen becslést kapunk, nem igény annak megvizsgálása, hogy egyértelmű megoldás-e, esetleg van több azonos optimum tulajdonsággal bíró más megoldás is. Ez az egyik legveszélyesebb tulajdonsága az ilyen típusú algoritmusok alkalmazásának: a tájékozatlan alkalmazó hajlamos elhinni, hogy ez az egyetlen és optimális megoldás. Persze a kémiai/fizikai alapokon nyugvó kényszerfeltételekkel (nem-negativitás, unimodalitás, ismert profilok, kinetikai egyenletek stb.) szűkíthetjük a megengedett-tartományokat. Azt azonban, hogy egyetlen megoldást kapjunk, sohasem tehetjük

matematikai kényszerfeltételek bevezetésével: így járunk el pl. a PCA (SVD) kapcsán is (a gyakorlatban a legritkább esetben ortogonálisok a spektrális profilok!).

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.

4. A 7. fejezetben említi, hogy kinetikai példákat mutatott helytelen kémiai rendszer alkalmazására. Kérem, hogy ezt fejtsse ki kissé részletesebben.

Az azóta megjelent tanulmányomban [R. Rajkó: On problematic practice of using normalization in self-modeling/multivariate curve resolution (S/MCR). *Chemometrics and Intelligent Laboratory Systems* 244: 105033, 2024. <https://doi.org/10.1016/j.chemolab.2023.105033>] többek között egy korábban megjelent rangvesztéses adatmátrixokra vonatkozó algoritmus [<https://doi.org/10.1016/j.chemolab.2022.104577>] bemutatására szolgáló kinetikai rendszereket is újra elemeztem. Összefoglalva és most csak az $S + K \xrightleftharpoons[k_{-1}]{k_1} [SK] \xrightarrow{k_2} P + K$ Michaelis-Menten

enzimkinetikát bemutatva, Matlabban a Simbiology toolbox-ot használtam az alapbeállításokkal, ami az ode15s integrátor használatát is jelentette, a következőket kaptam a kinetikai koncentráció profilok SVD felbontásával:

$\text{svd}(C) = [6.650693175896303; 4.230055090441462; 0.425321106032236; 0.000000862499289]$.

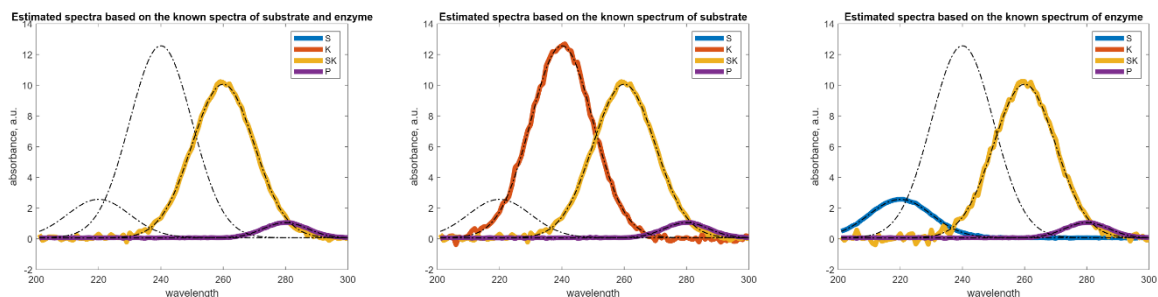
A 4. szinguláris érték bár kicsi, de szignifikáns (zajmentes adatokkal dolgoztam), így a mátrix rangja 4 és nincs rangvesztés. Ugyanakkor, figyelembe véve a szokásos kezdeti koncentráció feltételeket és az anyagmegmaradás elvén nyugvó két egyenletet: $c_{SK}(0) = 0$, $c_P(0) = 0$; $c_K(t) + c_{SK}(t) = c_K(0) = 0.1$ és $c_S(t) + c_{SK}(t) + c_P(t) = c_S(0) = 1$, egy affin és egy lineáris kombináció írható fel: $c_S(t) + c_K(t) + 2c_{SK}(t) + c_P(t) = c_S(0) + c_K(0) = 1.1$, és $c_S(t)/c_S(0) - c_K(t)/c_K(0) + (1/c_S(0) - 1/c_K(0))c_{SK}(t) + c_P(t)/c_S(0) = c_S(t) - 10c_K(t) - 9c_{SK}(t) + c_P(t) = 0$. Ez utóbbi lineáris kombináció rangvesztést okoz! Alaposabban elemeztem az ode15s megoldást és azt tapasztaltam, hogy egyik kombinációs egyenlet sem teljesült. Tehát ennél az enzimkinetikánál az alapbeállításként használt ode15s integrátor az alapparaméterekkel nem alkalmas megfelelő kinetikai koncentráció profil meghatározására. Az ode89 nagypontosságú integrátor már megfelelő eredményt szolgáltatott:

$\text{svd}(C) = [10.487255413923357; 7.918856903253778; 0.656617933338373; 0.000000000000000]$

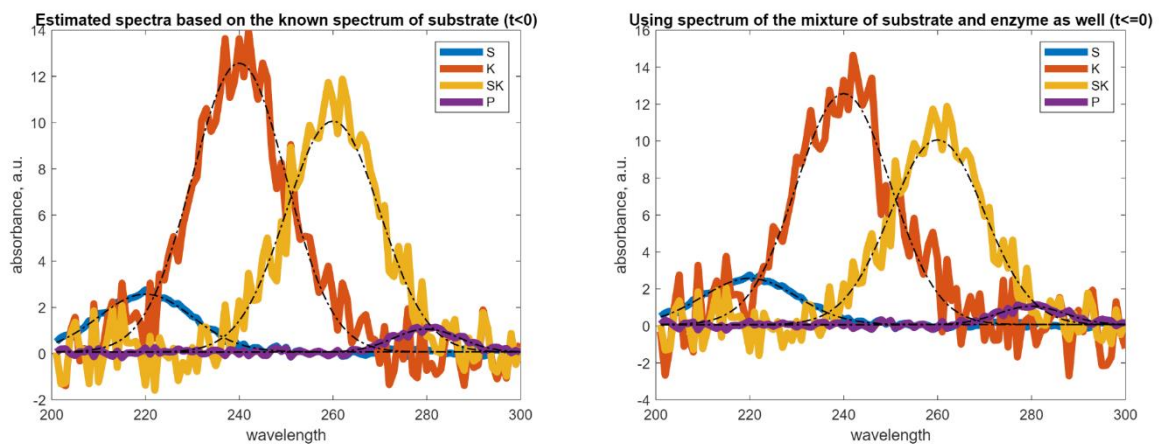
Bár az ode89 nem elérhető a Simbiology toolboxban, az ode45 viszont igen és ezzel is megkaphatjuk a kívánt rangvesztéses eredményt. Tehát a Michaelis-Menten enzimkinetika valóban rangvesztéses kinetikai koncentráció profilokat eredményez, ha megfelelő integrátort

használunk és/vagy megfelelően paraméterezzük őket. Igen ám, de ez a példa nem éppen a legjobb a rangvesztéses koncentrációmátrixok felhasználásával a spektrális profilok sávmegoldásának bemutatására, hiszen analitikai/fizikai-kémia ismeretek felhasználásával megszüntethetjük a rangvesztést, amit a cikkem mellékleteként biztosított Matlab Live Editorral készített Notebookban részletesen be is mutattam (zajos adatokra is):

- az enzim kis részletét szétosztva folyamatosan adagoljuk a rendszerhez,
- az enzim egy kis részletét egyszerre adagoljuk (spike) egy alkalmas időpontban a rendszerhez,
- a szubsztrát és az enzim spektruma ismert, vagy az enzimreakció előtt mérhető,



- ha az enzim spektruma nem ismert, a reakció előtti szubsztrát spektruma ($t < 0$) egyértelműen ismert (mert mérhető), utána pedig a szubsztrát és az enzim kezdeti keverékének ($t = 0$) spektrumából eliminációval az enzim spektruma becsülhető.



Végül nagy tisztelettel megköszönöm, hogy a bíráló elfogadta az értekezésemben megfogalmazottakon alapuló téziseimet új tudományos eredményeimnek, az értekezésben összefoglalt eredményeimet elegendőnek tartotta az MTA doktora cím megszerzéséhez, és javasolta a nyilvános vita kitűzését és részemre az MTA doktora cím odaítélését.

Szeged, 2024. június 26.

R. Rajkó

Prof.Dr. Rajkó Róbert