HANDLING RANK DEFICIENCY IN *UV/VIS* SPECTRA FROM KINETIC STUDIES BY VARIOUS MCR TECHNIQUES

A.N. Skvortsov

Biophysics Department, St-Petersburg State Polytechnical University.
Laboratory of Molecular Biology of Stem Cells, Institute of Cytology. St-Petersburg, Russia
E-mail: colbug@mail.ru

Background

Monitoring UV/vis absorption is one of the oldest instruments for studying chemical reactions. It combines versatility, high sensitivity, low response time etc. *Multivariate curve resolution* (MCR) is usually employed to resolve highly overlapping absorption spectra of the components:

Given $\mathbf{X}$ is the data $M \times N$ matrix

$$\mathbf{X} = \mathbf{C}_0\mathbf{S}_0^{\mathrm{t}} + \mathbf{E},$$

$\mathbf{C}_0$ is the concentrations ($M \times A$) matrix, $\mathbf{S}_0$ is the spectra ($N \times A$) matrix, $\mathbf{E}$ is the noise matrix. Matrices $\mathbf{C}_0$, $\mathbf{S}_0$ are subject to *chemical constrains*.

MCR aims to find A (number of substances), and $\mathbf{C}_X$ and $\mathbf{S}_X$ close to $\mathbf{C}_0$, $\mathbf{S}_0$ in some sense. Usu. a least-square (LS) sense is used:

$$\| \mathbf{X} = \mathbf{C}_X\mathbf{S}_X^{\mathrm{t}} \|_2 = \min$$

Problem

Many important reactions in coordination chemistry have a following generic mechanism:

$$[MX_2] + 2Y \leftrightarrow [MXY] + Y + X \leftrightarrow [MY_2] + 2X$$

M here is coordination center, X and Y are some ligands.

Most intensive UV/vis bands usually originate from M-X and M-Y bonds. So, we quite often observe:

$$\mathbf{s}_{|MXY|} \approx \tfrac{1}{2}(\mathbf{s}_{|MX2|} + \mathbf{s}_{|MY2|})$$

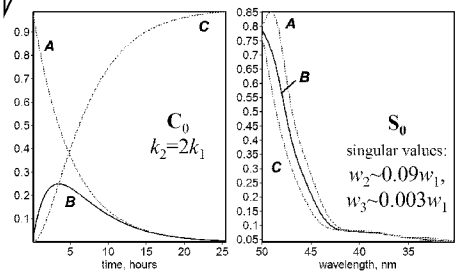
Hence, the efficient rank of matrices $\mathbf{S}_0$ and $\mathbf{X}$ is less, than the number of independent concentrations => we observe additional ***rank deficiency***.

This type rank deficiency is not obvious *a priori*, but its is not as practically rare as it may seem.

AIM: Crash-test various MCR approaches against "ill-conditioned" $\mathbf{S}_X$ matrices.

Model Data

Model reaction: $A \rightarrow B$, k_1; $B \rightarrow C$, k_2



k_1 and $\mathbf{S}_0$ were taken to directly emulate 2-step hydrolysis reaction of anticancer drug cisplatin. M 175, N 247.

SOFT only

Direct ***Alternating least squares*** approach (MCR-ALS)

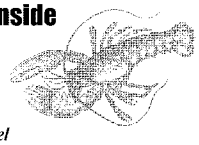
1. LS-estimate spectra $\hat{\mathbf{S}}=\mathbf{X}^{\mathrm{t}}\mathbf{C}(\mathbf{C}^{\mathrm{t}}\mathbf{C})^{-1}$
2. Correct $\hat{\mathbf{S}}$ for nonnegativity, $\hat{\mathbf{S}} \rightarrow \mathbf{S}$
3. LS-estimate concentrations $\hat{\mathbf{C}}=\mathbf{X}\mathbf{S}(\mathbf{S}^{\mathrm{t}}\mathbf{S})^{-1}$
4. Correct $\hat{\mathbf{C}}$ for closure, windows and nonneg., $\hat{\mathbf{C}} \rightarrow \mathbf{C}$
5. Loop (1) until convergence on $\|\mathbf{X}-\mathbf{C}\mathbf{S}^{\mathrm{t}}\|_2$

SOFT-outside & HARD-inside

Alternating least squares constrained ***with kinetic model***

1. LS-estimate spectra $\hat{\mathbf{S}}=\mathbf{X}^{\mathrm{t}}\mathbf{C}(\mathbf{C}^{\mathrm{t}}\mathbf{C})^{-1}$
2. Correct $\hat{\mathbf{S}}$ for nonnegativity, $\hat{\mathbf{S}} \rightarrow \mathbf{S}$
3. LS-estimate concentrations $\hat{\mathbf{C}}=\mathbf{X}\mathbf{S}(\mathbf{S}^{\mathrm{t}}\mathbf{S})^{-1}$
   4a. Generate $\mathbf{C}$ from ***kinetic model***
   4b. NLSF: loop (4a) until convergence on $\|\hat{\mathbf{C}}-\mathbf{C}\|_2$
5. Loop (1) until convergence on $\|\mathbf{X}-\mathbf{C}\mathbf{S}\mathrm{t}\|_2$

HARD-outside & SOFT-inside

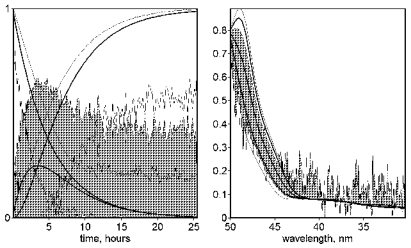


Full-matrix non-linear least squares fit

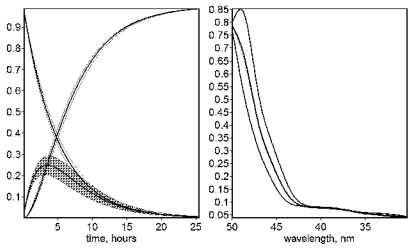
1. Generate $\mathbf{C}$ from ***kinetic model***
   2. LS-estimate spectra $\hat{\mathbf{S}}=\mathbf{X}^{\mathrm{t}}\mathbf{C}(\mathbf{C}^{\mathrm{t}}\mathbf{C})^{-1}$
   3. Correct $\hat{\mathbf{S}}$ for nonnegativity, $\hat{\mathbf{S}} \rightarrow \mathbf{S}$
4. NLSF loop (1) until convergence on $\|\mathbf{X}-\mathbf{C}\mathbf{S}^{\mathrm{t}}\|2$

Minimum/Maximum Bands of the Solution, $k_2 = 2k_1$

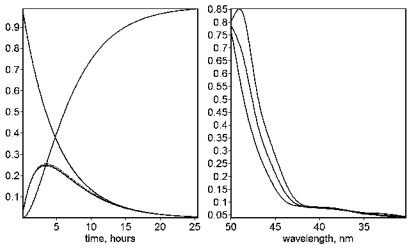
100 replicas, noise 0.3%. Band for ***B*** is painted gray



98% converged to ε 10^{-4}, <L.S.Eval>=**530**



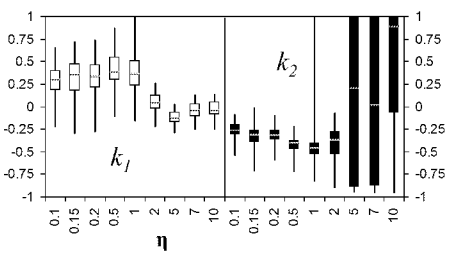
100% convgd. to ε 10^{-4}, <L.S.Eval>=**157,** <Kin.Eval>=**613**

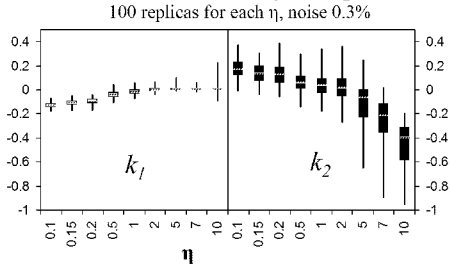


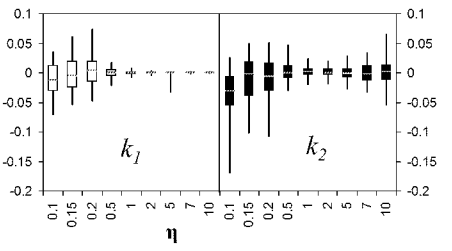
100% convgd. to ε 10^{-4}, <L.S.Eval>=**18,** <Kin.Eval>=**19**

Quartile Plots of Relative Errors for k_1 and k_2 Prediction, $k_2 = (1+\eta)k_1$

100 replicas for each η, noise 0.3%

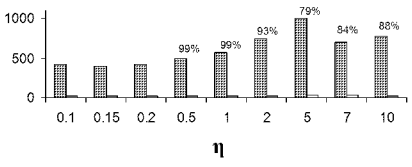


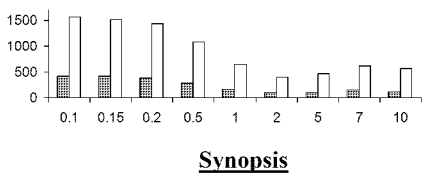


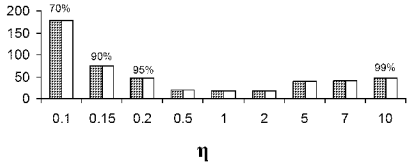


Convergence Expenses, Gray Bars – <Full-matrix LS-fits>, White Bars – <Kinetic Model Evaluations>

% of converged cases is indicated if less than 100%







Synopsis

WRECKED

BARELY PASSABLE

TOLERABLE