Reading Notes of "A Sweet Introduction to the Mathematical Analysis of TimeResolved Spectra and Complex Kinetic Mechanisms: The Chameleon Reaction Revisited"

Ref: J. Chem. Educ. 2022, 99, 2327−2337

Independent Author: Jingdan Chen 陈 敬丹(2020302031063)

Here, I pick the Glucose dataset analyzed by *TRSpec_Bilinear_FitData.m* (with *kineticsKmat_simu.m* and *IRFconvol_SP.m* being called in it) as the example to provide a detailed explanation of the algorithm developed in this paper. In this example workflow, an ideal system of first order reaction containing three components described by ordinary differential equations(ODEs) are proposed to analyze the chameleon reactions.(eq 1.1 and 1.2) Scree plot and singular value decomposition (SVD) are used to judge the constitute of the system and non-Linear solver are used to get the solution.

Arrived from Lambert-Beer Law, we usually view the strength of absorption spectrum linearly dependent with concentration and could be described as the linear combination of concentration and absorption coefficient.(eq 1.3) The time-resolved UV-Vis spectral data of chameleon reaction has been recorded, and the aim of the algorithm is to give the **S**(λ)(species-associated spectra, SAS) and **C**(t)(concentration evolution) of the three proposed species in the solution. Based on SAS and **C**(t), fitted time-resolved spectrum and residual plot (difference between the fitted spectrum and the true spectrum) could be given to clearly show the effectiveness of the fitting method. In this way, we could get a deeper insight into the reaction system knowing the concentration evolution and spectrum components.

$$\mathrm{A} \xrightarrow{k_1} \mathrm{B} \xrightarrow{k_2} \mathrm{X} \tag{1.1}$$

$$\begin{cases} \dfrac{\mathrm{d}[\mathrm{A}]_t}{\mathrm{d}t} = -k_1[\mathrm{A}]_t \\ \dfrac{\mathrm{d}[\mathrm{B}]_t}{\mathrm{d}t} = +k_1[\mathrm{A}]_t - k_2[\mathrm{B}]_t \\ \dfrac{\mathrm{d}[\mathrm{X}]_t}{\mathrm{d}t} = +k_2[\mathrm{B}]_t \end{cases} \tag{1.2}$$

$$\mathbf{D}(t,\lambda) = \mathbf{C}(t) \cdot \mathbf{S}(\lambda) \tag{1.3}$$

I am showing details in this script (Input Data, Solver, Important Functions Defined and Called) and I conclude the script into a workflow figure. Besides, I write a script to test, analyze and refine the SVD used in this workflow (see more in Supporting Information).

Input Data:

1) A float 1127*468 matrix, with one row denoting "time scale" and one column denoting wavelength, 1126*467 absorption data. I don't really know why there could be negative

values in the absorption;

2) A vector of initial concentration;(optional)
3) Lower and upper limit of rate constant for the sake of optimization;(optional)
4) Initial Guess for rate constants and IRF parameters (FWHM and t0);(optional)

Solver:

lsqnonlin(nonlinear least-squares solver)

Important Functions Defined and Called:

1. Ct = kineticsKmat_simu(t,K,C0,MakePlot,NameArray)

   MakePlot: bool, determine if we should plot; NameArray: defines the name of species(optional)

   t,K,C0: parameters giving needed parameters to solve ordinary differential equation;

   Ct: concentration of t (concentration evolution);

2. [res,Dfit,CConc,Sfit] = FitFunc(p,t,Kmodel,C0,Dexp)

   res: Loss Function value, given by $\|D_{\exp}-D_{fit}\|_2^2+GAMMA*NNR^2$ (mean squared error of D and penalty for negative values of Sfit), GAMMA is the parameter to adjust the non-negative penalty;

   Dfit: fitting result of D (fitted with Dexp);

   CConc: Concentration evolution of each species given by fitted model;

   Sfit: Fitted absorption for each species;

   p: initial guess (rate constants and IRF parameters);

   t: time series of Dexp;

   Kmodel: kinetic matrix that defines the model;

   C0: initial concentration vector;

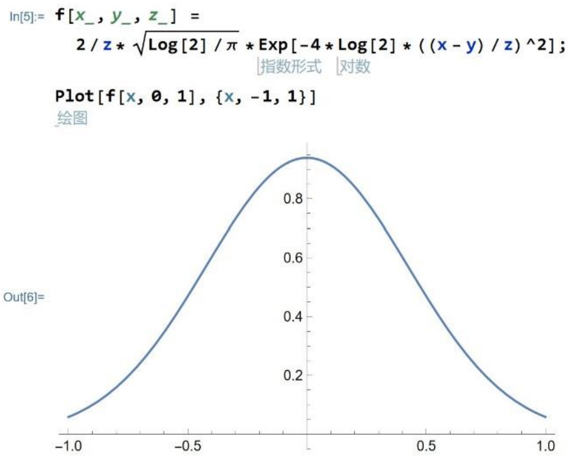
   Dexp: experimental spectrum;

3. [CConc,tIRF,simIRF] = IRFconvol_SP(IRFfnc,t,t0,w,ConcFnc)

   Calculate ode derived concentration and use IRF function to convolve the specified part(initial part of time series during which the bias caused by mixing exists) to get revised

   Hint: IRF is usually Gaussian-like(Figure 1)

Figure 1: IRF defined in the paper (with t_0=0, w=1)



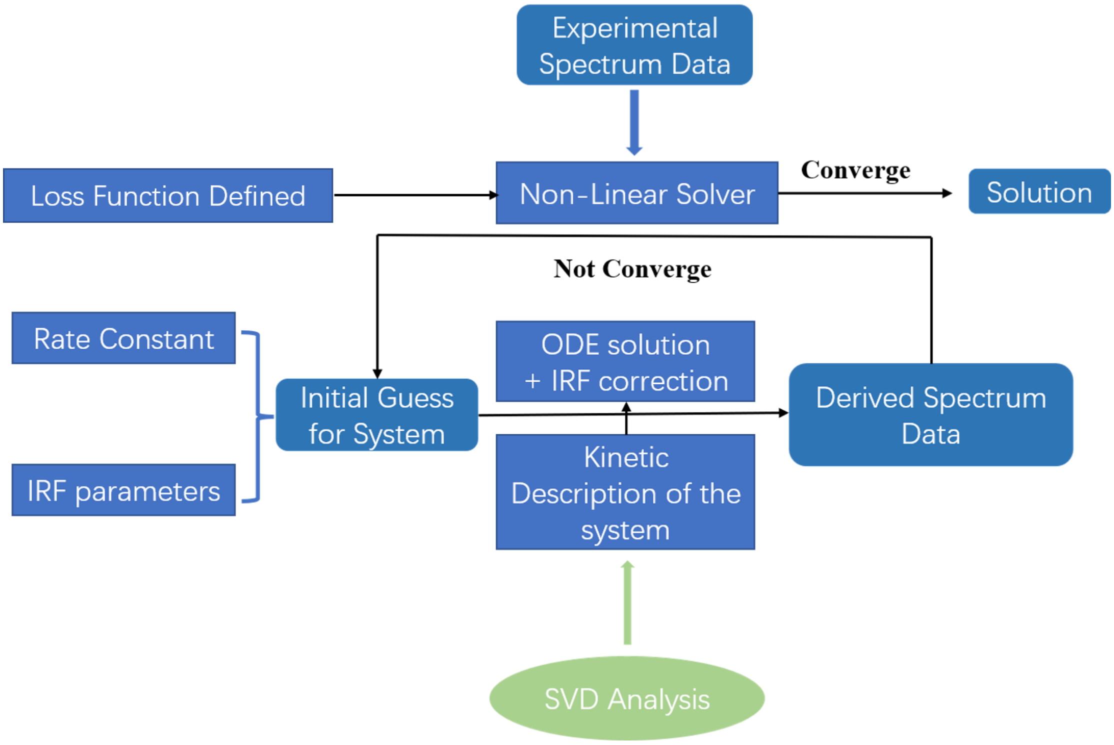
```
In[5]:= f[x_, y_, z_] =
          2 / z * √(Log[2] / π) * Exp[-4 * Log[2] * ((x - y) / z)^2];

        Plot[f[x, 0, 1], {x, -1, 1}]
```

**Convolution is a basic technique processing matrix-like data and one of its main purpose is to build relationship between data point and data points nearby. It is widely used in image process and Gaussian-like kernel is often used as an image blurring tool. There are also many other kernels developed for other purposes. (like image sharpening and edge detection)
See: MIT 18.S191/6.S083/22.S092 (https://computationalthinking.mit.edu/Fall20/)

Workflow:

Figure 2: Workflow of *TRSpec_Bilinear_FitData.m*



Supporting Information: Preliminary Spectrum Analysis

(see preliminary_spectrum_analysis.html or preliminary_spectrum_analysis.ipynb, they are exactly the same)

This jupyter notebook aims to use SVD (Singular Value Decomposition, a method introduced in primary linear algebra and utilized in this paper), PCA (Principal Component Analysis, nearly one of the most classical unsupervised machine learning method), NMD (non-Negative Matrix Factorization, the algorithm I think fits the problem most which is not introduced in undergraduate mathematics nor this paper, developed based on PCA) to analyze the time-resolved spectrum preliminarily. I use preliminary spectrum analysis to judge how many components are there in this system, and hopefully we could roughly judge the shape of the UV-Vis spectrum of each component.

In the paper, the author used SVD combined with Scree plot to figure out the number of UV-Vis active species in the system. Since that SVD is classic and shows up in primary linear algebra context, it could be very pedagogical to connect the handful mathematical tools with chemical experiments by doing so. Here, I reproduce the result with Glucose data (Figure 1.1). The scale of the singular value represent the "importance" of corresponding "vector component" of the matrix. We could see that in Scree plot, first three points owns relatively high value while the points after the third own a similar and low magnitude, which shows that there are three UV-Vis active species in the system. While the author utilized a method which combines ode with least-square-error fitting with the correction of IRF curve to figure out the spectrum of each species in the paper, the column vector results of SVD can hopefully also give some spectrum shape information (Figure 1.2). We could see that the component 2 and component 3 are somehow like the spectrum of MnO_4^- and MnO_4^{2-}, but component 1 could not be attributed to any species(at least in the paper).

When it comes to the "decomposition" of digital signal, Fourier Transform and PCA could jump into my mind. In analytical chemistry, Fourier Transform could be another frequently mentioned concept in spectrum decomposition and SVD is exactly somehow the other format of Fourier's methodology. There is another commonly mentioned widely used tool for "decomposition" which is principal component Analysis (PCA). Here, I use PCA to analyze the spectrum of Glucose too, and I set the number of component to 50 to sketch the Scree plot to analyze the number of principal components. We could find that the PCA also did pretty well in finding the number of principal component (Figure 2.1), but when it comes to spectrum shape, only component 2 could be attributed to MnO_4^- while other 2 fails to fit any species well. The relatively poor performance of PCA and SVD could be originated from two issues: 1) IRF function is not used here to convolve the spectrum to get a corrected one; 2) SVD and PCA could not define the sign of "contribution" of each dimension of the component, namely "negative contribution" to a wavelength could occur in the component which is not physically meaningful. To solve the second question which is relatively easy to deal with, here I use the NMD which is developed based on PCA and could define each dimension of the "component vector" positive (the author introduced "a non-negativity of the SAS" to his model to solve this problem"). (Hint: Considering that the input matrix of NMD should be “all-positive”, I used absolute value of the origin matrix) From Figure 3 we could see that three components could be roughly well attributed to Mn(IV, aq), MnO_4^- and MnO_4^{2-}, reveling the success of NMF algorithm on this example. (Apologize for not providing further mathematics here, maybe google is enough here?)