Multidimensional Signal Analysis Part III – Lecture 14

February 5, 2004

Contents

References

[Gemperline, 1999] Gemperline, P. (1999). Computation of the range of feasible solutions in self-modeling curve resolution algorithms. *Anal. Chem.*, 71:5398–5404.

[Gruninger and Sundberg, 1997] Gruninger, J. and Sundberg, R. L. (1997). Bounds on component spectra of multispectral images. In *SPIE Technologies for Synthetic Environments: Hardware-in-the-Loop Testing II Proceedings*, volume 3084. SPIE.

[Jiang and Ozaki, 2002] Jiang, J.-H. and Ozaki, Y. (2002). Self-modeling curve resolution (smcr): principles, techniques, and applications. *Applied Spectroscopy Reviews*, 37(3):321–345.

1 Local linear embedding

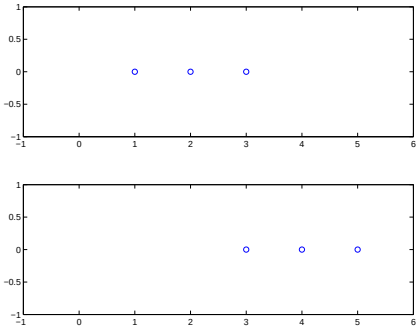
1. find neighbors of each datum
2. explain each datum as linear weighted combination of its neighbors
3. find embedding such that each datum is as close as possible to linear weighted combination of its neighbors, using weights from before

1.1 Step 2

$$\varepsilon(W) \quad = \quad \mathrm{tr}(X - WX)(X - WX)^T$$

minimize error function subject to two conditions:
i) weights are non-zero for nearest neighbors only
ii) weights sum to one

The latter condition is necessary to ensure translation invariance. Example:



Middle point can be reconstructed in both cases using weights $[1/2, 0, 1/2]$ but not using $[0, 0, 2/3]$.

Finding the optimum weights is a constrained least squares problem:

for a particular datum x $(1 \times p)$ with k nearest neighbors Ξ $(k \times p)$ and w $(1 \times k)$ the row vector of W $(N \times N)$ corresponding to x,

$$\varepsilon = (x - w\Xi)(x - w\Xi)^T$$

with the weights summing to one, $wl = 1$, $l = [1, \ldots, 1]^T$ we have

$$\begin{aligned} \varepsilon &= (wlx - w\Xi)(wlx - w\Xi)^T \\ &= w(lx - \Xi)\left(w(lx - \Xi)\right)^T \\ &= w(lx - \Xi)(lx - \Xi)^T w^T \\ &= wCw^T \end{aligned}$$

with C the local covariance matrix.

So

$$\min wCw^T \text{ s. t. } w^T l = 1$$

Using Lagrange multipliers to enforce weight normalization, we get

$$\min wCw^T + \lambda(w^T l - 1)$$

with derivatives

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial w} &= 2w^T C + \lambda l^T = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda} &= w^T l - 1 = 0 \end{aligned}$$

so that

$$\begin{aligned} w^T &= -\frac{\lambda}{2} l^T C^{-1} \\ w^T l &= -\frac{\lambda}{2} l^T C^{-1} l = 1 \\ \lambda &= -\frac{2}{l^T C^{-1} l} \\ w^T &= \frac{l^T C^{-1}}{l^T C^{-1} l} \end{aligned}$$

In practice, inversion of C is not necessary because the solution of $Cw^T = l$, divided by wl, is a solution to the above.

If the C is (nearly) singular, e.g. when $k > p$, add a small constant to the diagonal (as in ridge regression).

1.2 Step 3

Now that weights W are found, minimize the following cost function:

$$\begin{aligned} \Phi(Y) &= \operatorname{tr}\left((Y - WY)(Y - WY)^T\right) \\ &= \operatorname{tr}\left((Y - WY)^T(Y - WY)\right) \\ &= \operatorname{tr}\left(Y^T Y - Y^T W Y - Y^T W^T Y + Y^T W^T W Y\right) \\ &= \operatorname{tr}\left(Y^T Y - Y^T W Y - Y^T W^T Y + Y^T W^T W Y\right) \\ &= \operatorname{tr}\left(Y^T \left(\mathbb{I} - W - W^T + W^T W\right) Y\right) \end{aligned}$$

Remove translatory DOF by fixing center of mass at the origin $l^T Y = 0$ and choose one out of a family of "similar" $(\operatorname{tr} U^T A U = \operatorname{tr} A U U^T = \operatorname{tr} A)$ point clouds by imposing $Y^T Y = N\mathbb{I}$.

Cost function is minimized by composing Y of eigenvectors pertaining to lowest eigenvectors of the sparse Hermitian matrix $\mathbb{I} - W - W^T + W^T W$ (Rayleigh-Ritz theorem).

1.3 Rayleight-Ritz theorem

For a Hermitian matrix $A = A^\dagger$ with $\lambda_1 \leq \ldots \leq \lambda_n$,

$$\lambda_1 x^\dagger x \leq x^\dagger A x \leq \lambda_n x^\dagger x \qquad \forall x \in \mathbb{C}^n$$
$$\lambda_n = \max_{x \neq 0} \frac{x^\dagger A x}{x^\dagger x} = \max_{x^\dagger x = 1} x^\dagger A x$$
$$\lambda_1 = \min_{x \neq 0} \frac{x^\dagger A x}{x^\dagger x} = \min_{x^\dagger x = 1} x^\dagger A x$$

Proof: since $A = A^\dagger, \exists UU^\dagger = \mathbb{I}$ such that $A = U\Lambda U^\dagger$ with $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$.

$$x^\dagger A x = x^\dagger U \Lambda U^\dagger x = (U^\dagger x)^\dagger \Lambda (U^\dagger x) = \textstyle\sum_{i=1}^n \lambda_i |[U^\dagger x]_i|^2$$
$$\lambda_1 \textstyle\sum_{i=1}^n |[U^\dagger x]_i|^2 \leq \sum_{i=1}^n \lambda_i |[U^\dagger x]_i|^2 \leq \lambda_n \sum_{i=1}^n |[U^\dagger x]_i|^2$$
$$\textstyle\sum_{i=1}^n |[U^\dagger x]_i|^2 = (U^\dagger x)^\dagger (U^\dagger x) = x^\dagger U U^\dagger x = x^\dagger x$$
$$\Rightarrow \lambda_1 x^\dagger x \leq x^\dagger A x \leq \lambda_n x^\dagger x$$

The inequalities are tight (take eigenvectors corresponding to λ_1 or λ_n as x).

For $x \neq 0$, $x^\dagger A x / x^\dagger x \geq \lambda_1$, thus $\min_{x \neq 0} x^\dagger A x / x^\dagger x = \lambda_1$.

Moreover, $\min_{y^\dagger y = 1} y^\dagger A y = \lambda_1$ for $y = x/\sqrt{x^\dagger x}$.

2 Bilinear decomposition

is a set of problems which can be formulated as

$$X = CS^T(+\varepsilon)$$

with observations X given and C, S sought.

2.1 Factor analysis

considers

S as a collection of **latent variables** (typically modelled as Gaussian RV)

C as a collection of **factor loadings** which are used to interpret the factors.

Fitting of the model is by maximum likelihood and is unique only up to rotations.

Example

Outcome of (oblique Promax) rotation with a factor analysis of 24 measures of mental ability (from R.L. Gorsuch *Factor Analysis* (1983) Hillsdale, NJ: Erlbaum).

	Verbal	Simple repetitive tasks	Visual	Recog-nition
General information	.80	.10	-.01	-.06
Paragraph comprehension	.81	-.10	.02	.09
Sentence completion	.87	.04	.01	-.10
Word classification	.55	.12	.23	-.08
Word meaning	.87	-.11	-.01	.07
Add	.08	.86	-.30	.05
Code	.03	.52	-.09	.29
Counting groups of dots	-.16	.79	.14	-.09
Straight & curved capitals	-.01	.54	.41	-.16
Woody-McCall mixed	.24	.43	.00	.18
Visual perception	-.08	.03	.77	-.04
Cubes	-.07	-.02	.59	-.08
Paper form board	-.02	-.19	.68	-.02
Flags	.07	-.06	.66	-.12
Deduction	.25	-.11	.40	.20
Numerical puzzles	-.03	.35	.37	.06
Problem reasoning	.24	-.07	.36	.21
Series completion	.21	.05	.49	.06
Word recognition	.09	-.08	-.13	.66
Number recognition	-.04	-.09	-.02	.64
Figure recognition	-.16	-.13	.43	.47
Object-number	.00	.09	-.13	.69
Number-figure	-.22	.23	.25	.42
Figure-word	.00	.05	.15	.37

```
[R] varimax and promax rotation
Prof Brian D Ripley ripley at stats.ox.ac.uk
Thu Nov 26 07:30:24 MET 1998

    * Previous message: [R] varimax and promax rotation
    * Next message: [R] Lindsey libraries for Win
    * Messages sorted by: [ date ] [ thread ] [ subject ] [ author ]

On Wed, 25 Nov 1998, Jorge M. S. Magalhaes wrote:

> How i can make Varimax  and Promax Rotation  in R 0.63

As I understand those terms, they apply to the output of a factor analysis,
and I am not aware of that being implemented in R.

You are here in a rather specialised (and not at all widely accepted) area
of statistics, and would be better off using a specialist package. I made a
deliberate decision not to implement factor analysis in R.

Rotation is sometimes applied to the output of a principal components
analysis, but that is even more controversial.

--
Brian D. Ripley,                  ripley at stats.ox.ac.uk
Professor of Applied Statistics,  http://www.stats.ox.ac.uk/~ripley/
University of Oxford,             Tel:  +44 1865 272861 (self)
1 South Parks Road,                     +44 1865 272860 (secr)
Oxford OX1 3TG, UK                Fax:  +44 1865 272595

-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-
r-help mailing list -- Read http://www.ci.tuwien.ac.at/~hornik/R/R-FAQ.html
Send "info", "help", or "[un]subscribe"
(in the "body", not the subject !)  To: r-help-request at stat.math.ethz.ch
-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-
```


2.2 Independent component analysis

In **blind source separation**, observations are assumed to result from underlying signals which have been mixed by a linear transformation ("cocktail party problem").

Independent component analysis (ICA) assumes the sources are non-Gaussian and independent (not merely uncorrelated!).

Individual sources are estimated such that their product matches the joint distribution as closely as possible. This is equivalent to maximizing departure from normality of individual sources.

The combination of non-Gaussianity and the independence assumption lifts the rotational ambiguity.

Applications in the deconvolution of biomedical signals: EEG, MEG, fMRI.

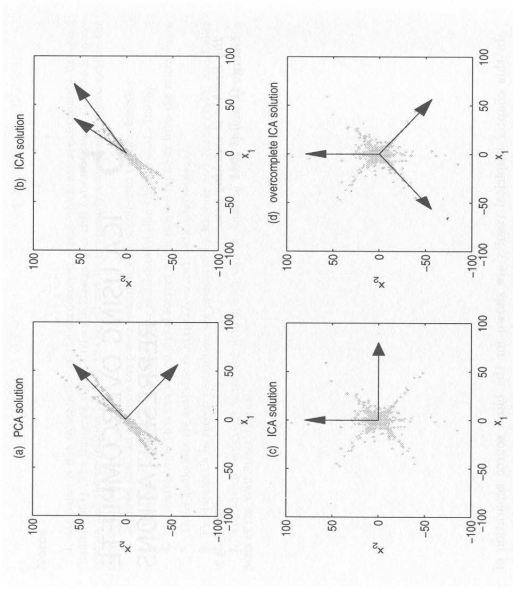


Figure 5.1. Illustration of basis vectors in a two-dimensional data space with two 2 sparse sources (top) or three sparse sources (bottom). (a) PCA finds orthogonal basis vectors and (b) ICA representation finds independent basis vectors. (c) ICA cannot model the data distribution adequately with three sources but (d) the overcomplete ICA representation finds 3 basis vectors that match the underlying data distribution (see Lewicki and Sejnowski).

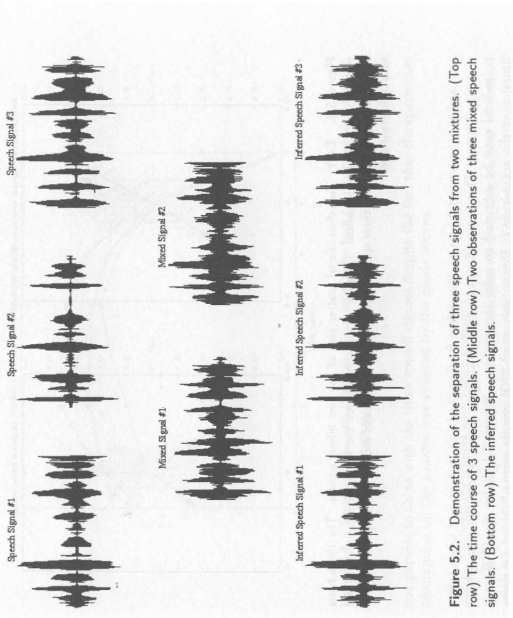
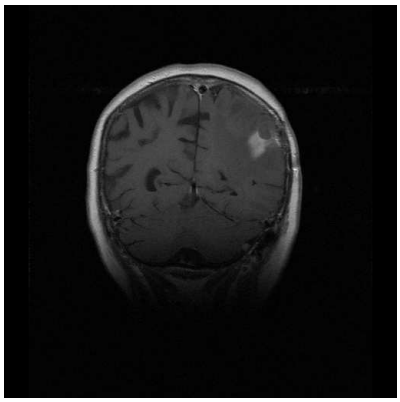


Figure 5.2. Demonstration of the separation of three speech signals from two mixtures. (Top row) The time course of 3 speech signals. (Middle row) Two observations of three mixed speech signals. (Bottom row) The inferred speech signals.

2.3 Self-modeling curve resolution

Aim: given a number of spectra from mixtures, estimate pure spectra of constituent components.



Example: in magnetic resonance spectroscopy, find spectra of white matter, gray matter, liquor.

$$X = CS^T$$

X – observed spectra of mixtures
C – mixing coefficients
S^T – one pure spectrum in each row

with non-negativity constraints: $[C]_{ij} \geq 0, [S]_{ij} \geq 0$

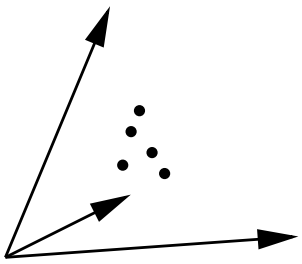
How many components to fit? $\rightarrow$ estimate rank by singular value decomposition (SVD)

How to find admissible C, S? $\rightarrow$ alternating least squares:

$$\begin{aligned}
C^{(i+1)} &= XS^{(i)} \left({S^{(i)}}^T S^{(i)} \right)^{-1} \\
\left[C^{(i+1)} \right]_{ij} &= I_{\left[C^{(i+1)} \right]_{ij} \geq 0} \left[C^{(i+1)} \right]_{ij} \\
{S^{(i+2)}}^T &= \left({C^{(i+1)}}^T C^{(i+1)} \right)^{-1} {C^{(i+1)}}^T X \\
\left[S^{(i+2)} \right]_{ij} &= I_{\left[S^{(i+2)} \right]_{ij} \geq 0} \left[S^{(i+2)} \right]_{ij}
\end{aligned}$$

etc.

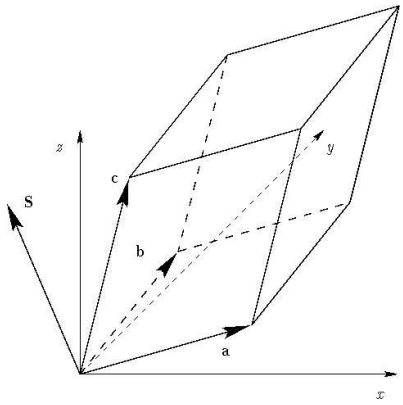
Interpretation



- vectors represent (estimated) pure spectra; these must lie in the positive orthant to guarantee non-negativity
- mixed spectra are represented by dots; these must reside in the parallelepiped spanned by the vectors to ensure positive mixture coefficients
- parallepiped can usually be rotated and widened / narrowed without violating the constraints $\Rightarrow$ "rotational ambiguity"
- problem becomes unique iff set of observations contains pure spectra; the corners of this simplex and the origin span a cone comprising all observations; simplex corners correspond to pure spectral directions

2.3.1 How to reduce rotational ambiguity?

Range of possibilities can be narrowed down by minimizing or maximizing the volume of the cone [Gruninger and Sundberg, 1997]. The determinant of [a, b, c] gives the volume of the parallelepiped spanned by these vectors:



These extreme decompositions convey information on the extent of the ambiguity.

Alternatively, prior information on "zero-concentration regions" can be exploited [Jiang and Ozaki, 2002].

Gemperline proposes to minimize / maximize the contribution of each normalized component in turn to the overall observations, while respecting non-negativity for all other components.

$$\begin{aligned} X &= U\Sigma V^T \\ &= U\Sigma R R^{-1} V^T \\ &= CS^T \end{aligned}$$

$$\min / \max \; l(U\Sigma R)\mathrm{diag}(0,\ldots,0,1,0,\ldots,0)(R^{-1}V^T)l, \quad l = [1,\ldots,1]^T$$

These extreme solutions for individual components can be plotted and thus give an idea of the extent of ambiguity.

(If it was not for the positivity constraints, the minimized / maximized component would point in the direction of the last and first right singular vector / principal component).

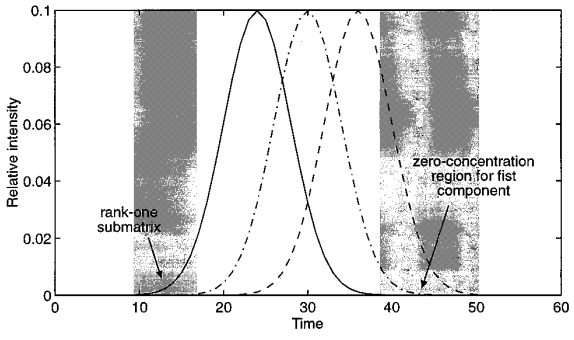


Figure 1. Simulated three-component chromatogram showing regions in time where a rank-one submatrix would be expected and a zero-component region for component one would be expected. Similar regions can be identified for other components of the chromatogram.

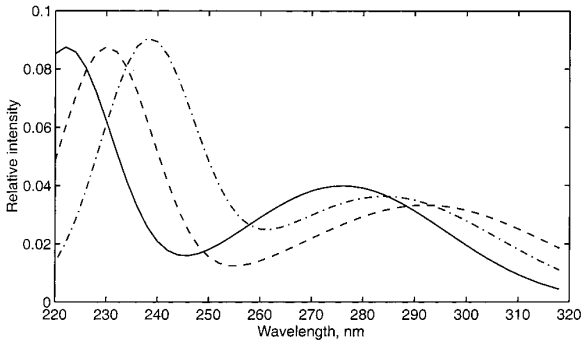


Figure 2. Simulated UV/visible spectra corresponding to the peaks shown in Figure 1.

After adding random normal deviations with a mean of zero

From [Gemperline, 1999]

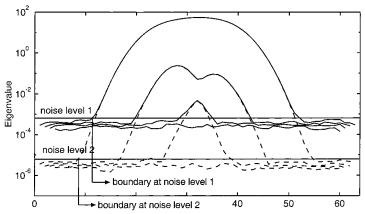


Figure 3. Evolving window factor analysis of a simulated spectrochromatographic data set at two noise levels produces different boundaries.

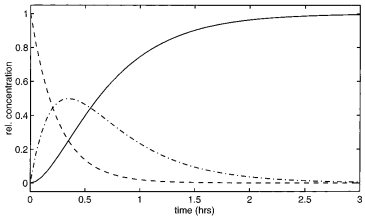


Figure 4. Simulated composition profiles for the sequential nonreversible reaction A + B ⇔ C ⇔ D + E. Reactant A (- - -), intermediate C (•••), and product D (—) are shown.

to the net analyte signal for the corresponding constituents. Our Matlab code attempts to compensate for this problem by finding an adaptive, robust noise estimate for each profile which is then used to "relax" the constraints slightly to allow small negative deviations. A detailed description of the algorithm is given later.

A simulated data set for a batch reaction was generated by numerical integration of the appropriate rate laws for the sequential reversible reaction A + B ⇔ C ⇔ D + E. Pure-component profiles for species A, C, and D are shown in Figure 4 with species B and E being nonabsorbing. The simulated spectra shown in Figure 5 were generated so that a nearly unique response was obtained for reactant A at 285 nm, intermediate C at 310 nm, and product D at 350 nm. These pure-component profiles and spectra were used to produce a data matrix of mixture spectra that was

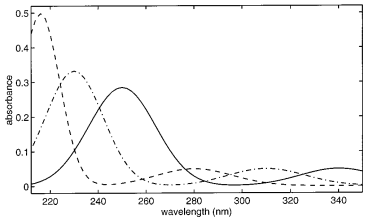


Figure 5. Simulated UV/Visible spectra for the sequential nonreversible reaction shown in Figure 4. Reactant A (- - -), intermediate C (•••), and product D (—) are shown.

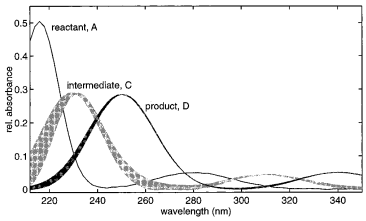


Figure 6. SMCR results and range of feasible profiles for simulated data described in Figures 4 and 5.

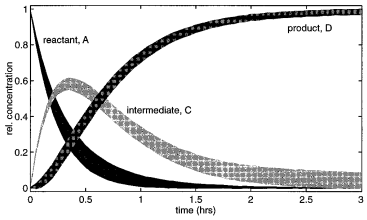


Figure 7. SMCR results and range of feasible pure-component spectra for the simulated data described in Figures 4 and 5.

From [Gemperline, 1999]