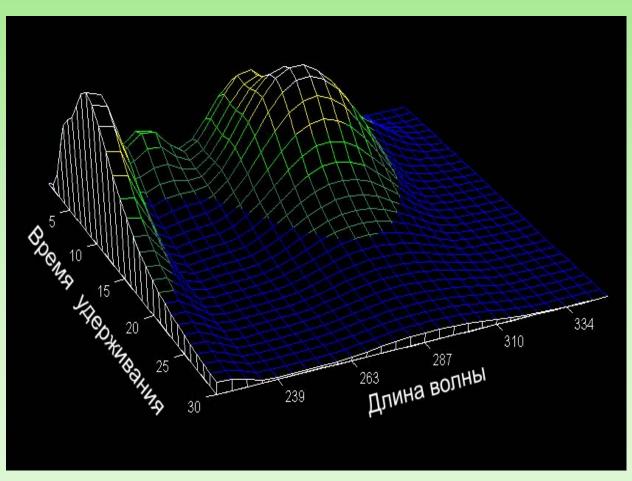
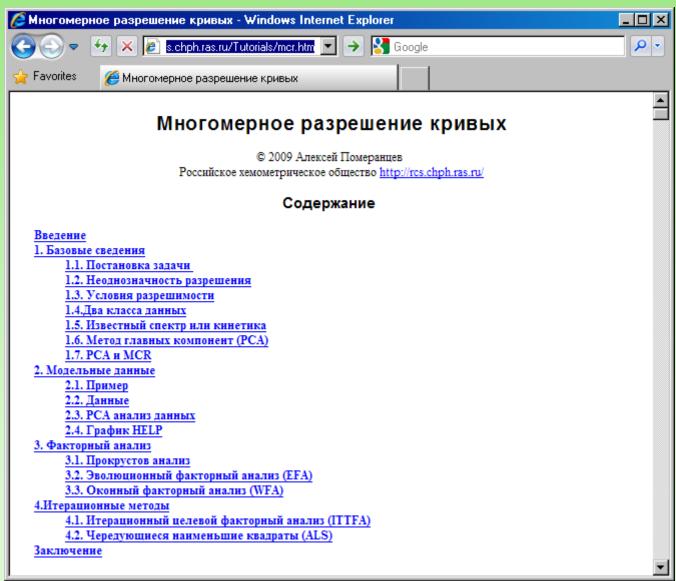
Разрешение многомерных кривых (MCR)

Институт химической физики РАН Алексей Померанцев

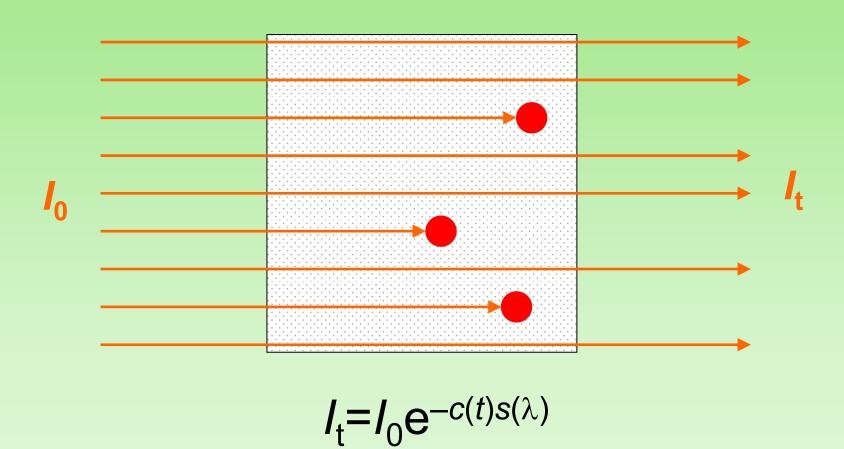


Российское хемометрическое общество

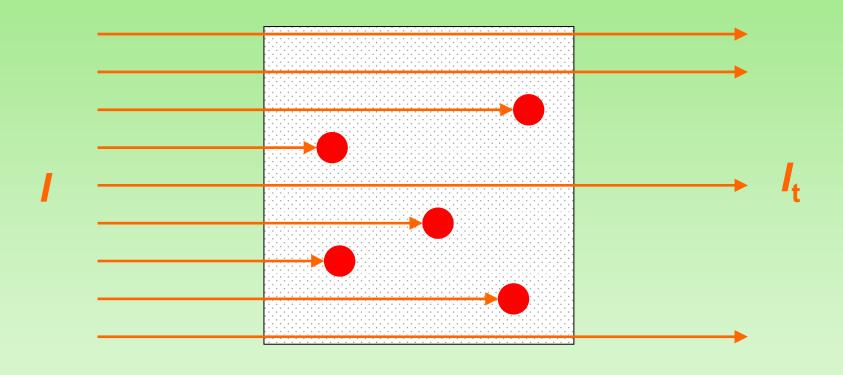
http://rcs.chph.ras.ru/Tutorials/mcr.htm



Закон Бугера — Ламберта — Бера

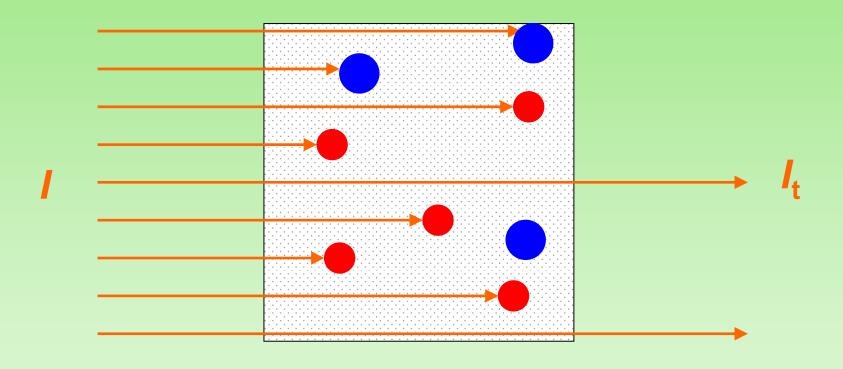


Закон Бугера — Ламберта — Бера



$$x(t,\lambda) = -\ln(I_t/I_0) = c(t)s(\lambda)$$

Закон Бугера — Ламберта — Бера



$$x(t,\lambda) = c_A(t)s_A(\lambda) + c_B(t)s_B(\lambda)$$

Устройство данных S_{true} 1.2 ☐ C_{true} 0.8 1.0 0.8 0.6 0.4 0.2 0.2 20 265 310 355 0.0 wavenumber 10 20 0 1.2 -0.9 -0.3 220 03.10.11 6 Лекция в Уфе

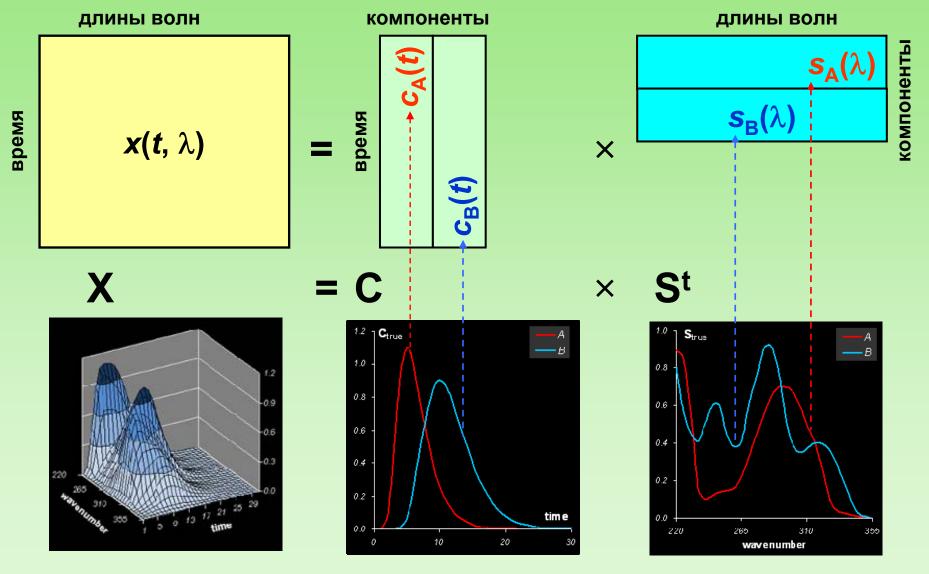
Постановка задачи

Данные, получаемые в эксперименте, можно представить в матричном виде —

$$X = CS^t + E$$

Задача разрешения кривых (МСR) состоит в том, чтобы по заданной матрице данных \mathbf{X} определить число химических компонентов (A) и найти матрицы концентраций \mathbf{C} и чистых спектров \mathbf{S} .

Структура «хроматографических» данных



Неоднозначность решения

$$X = CS^{t} + E$$

$$X = \widetilde{C}\widetilde{S}^{t} + E$$

Вращательная неопределенность

Пусть \mathbf{R} — это ортогональная матрица вращения размерностью $A \times A$, т.е такая матрица, что $\mathbf{R}^{t} = \mathbf{R}^{-1}$. Тогда

$$CS^{t} = CRR^{t}S^{t} = (CR)(SR)^{t} = \widetilde{C}\widetilde{S}^{t}$$

Масштабная неопределенность

Пусть \mathbf{R} — это диагональная матрица размерностью $A \times A$. Тогда

$$\mathbf{C}\mathbf{S}^{t} = \mathbf{C}\mathbf{R}\mathbf{R}^{-1}\mathbf{S}^{t} = (\mathbf{C}\mathbf{R})(\mathbf{S}\mathbf{R}^{-1})^{t} = \widetilde{\mathbf{C}}\widetilde{\mathbf{S}}^{t}$$

Глобальная неопределенность

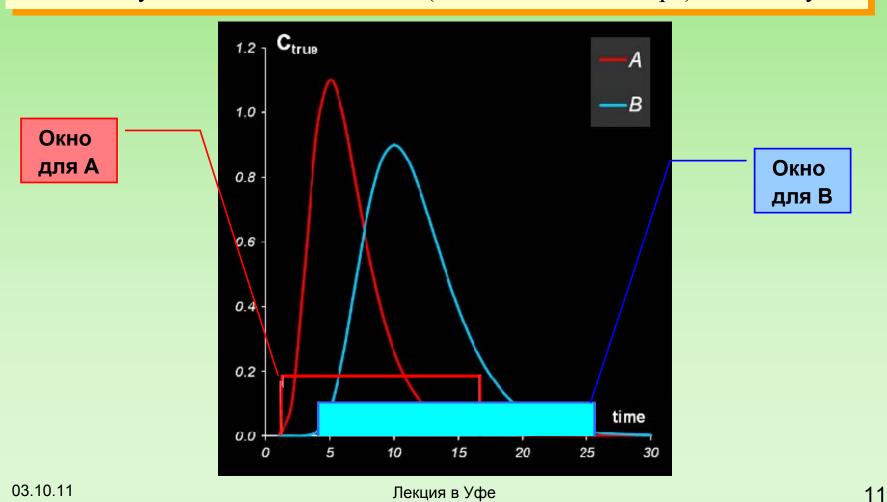
Избавиться от масштабной неопределенности нельзя. Любое найденное решение будет определено только с точностью до умножения на константу,

$$x_{ij} = \sum_{a=1}^{A} c_{ia} s_{aj} = \sum_{a=1}^{A} \left(\frac{c_{ia}}{r_a} \right) (s_{aj} r_a)$$
.

Иными словами каждый профиль концентраций c_a может быть в r_a разменьше, если соответствующий чистый спектр s_a увеличен в r_a раз. Однако форма кинетики и спектра при этом сохраняется.

Разрешимость: окно

Концентрационное (спектральное) окно для химического компонента a – это область переменных (соответственно t или λ), для которой величина соответствующего чистого сигнала (кинетики или спектра) больше нуля.

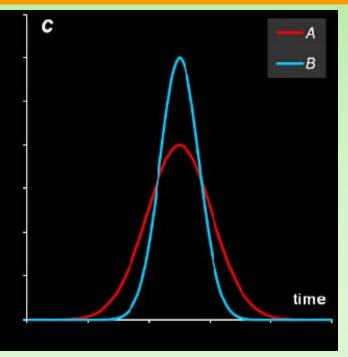


Условия разрешимости

- 1. Концентрационный профиль компонента может быть восстановлен, если все остальные концентрационные окна, пересекающиеся с окном этого компонента, имеют участки вне этого окна.
- 2. Спектральный профиль компонента может быть восстановлен, если его концентрационное окно не находится целиком внутри концентрационного окна какого-нибудь другого компонента.

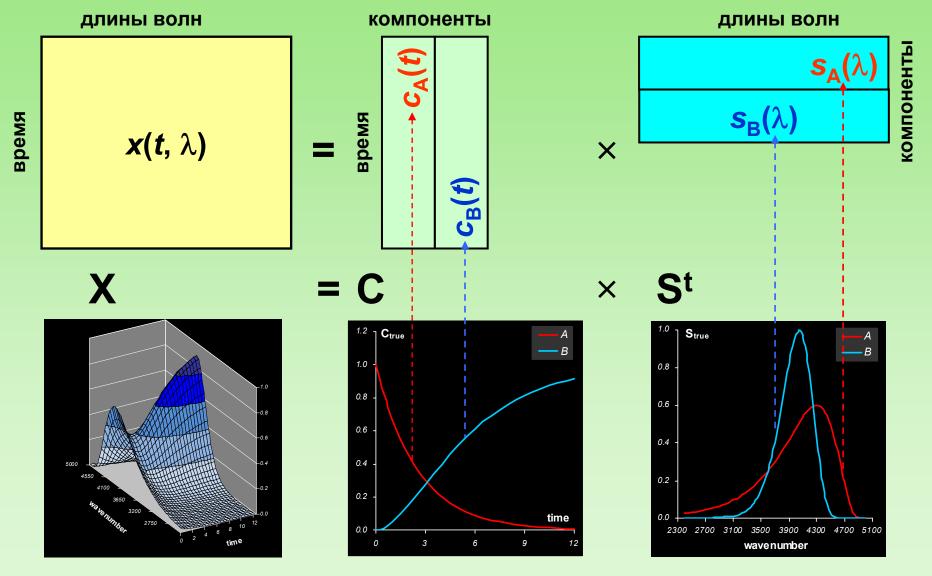


Roll Manne, 1978



Можно восстановить концентрационный профиль В, но нельзя найти его чистый спектр. С другой стороны, для А можно найти чистый спектр, но нельзя найти концентрационный профиль.

Структура «кинетических» данных



Частные решения задачи

Нужно решить задачу

Если известен спектр S, тогда

$$\mathbf{C}_{\text{hat}} = \mathbf{X}\mathbf{S}(\mathbf{S}^{\mathsf{t}}\mathbf{S})^{-1}$$

Если известна кинетика С, тогда

$$S_{hat} = X^t C(C^t C)^{-1}$$

Методы разрешения кривых

Факторные

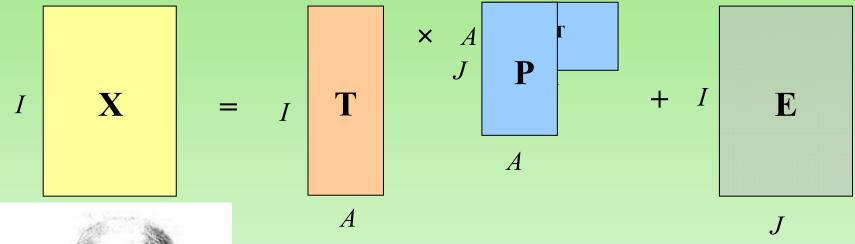
- Прокрустово вращение
- Эволюционный факторный анализ (EFA)
- Оконный факторный анализ (WFA)

Итерационные

- Итерационный целевой факторный анализ (ITTFA)
- Чередующиеся наименьшие квадраты (ALS)

Метод главных компонент (РСА)

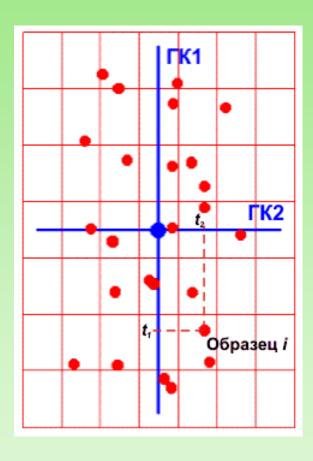
Исходные данные Матрица счетов Матрица нагрузок Матрица ошибок

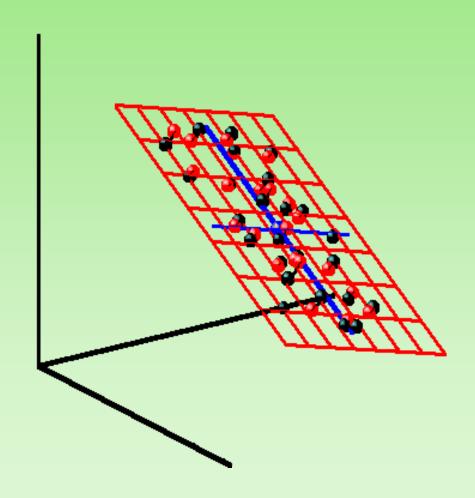


$$X=TP^T+E$$

Лекция в Уфе

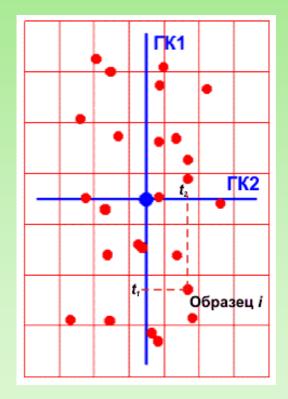
Представление данных в подпространстве

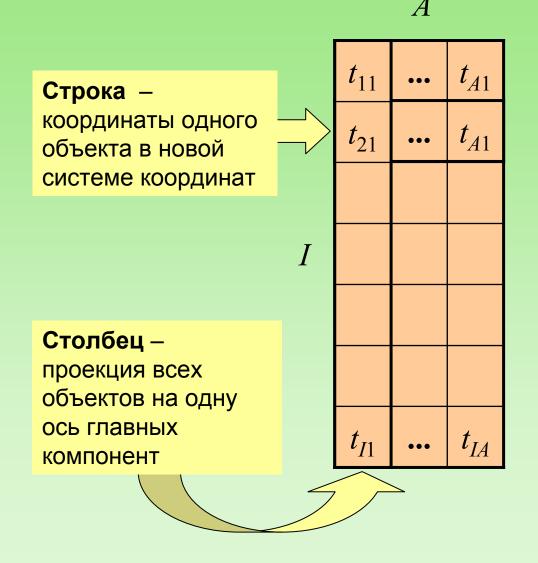




Матрица счетов T (scores)





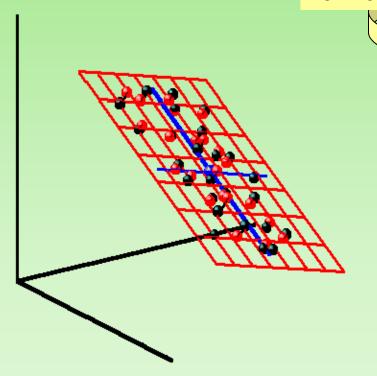


Матрица нагрузок P (loadings)



проекция всех переменных на одну ось главных компонент





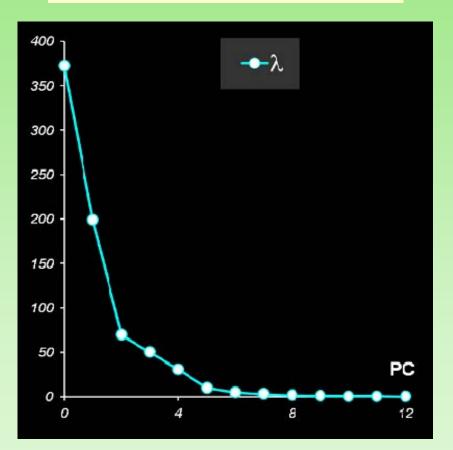
p_{11}	p_{12}		•••	p_{1J}
•••	•••		•••	•••
p_{A1}	p_{A2}		•••	p_{AJ}

Столбец –

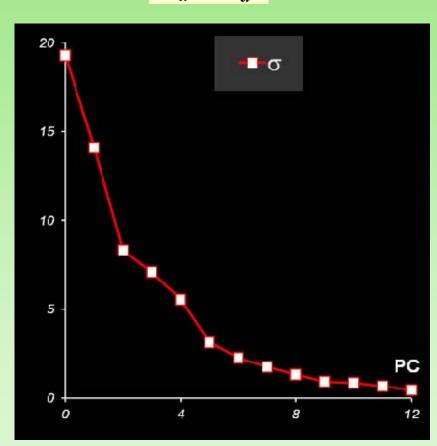
проекция одной переменной на новую систему координат

Собственные и сингулярные значения

$$\mathbf{T}^{\mathsf{t}}\mathbf{T} = \mathbf{\Lambda} = \operatorname{diag}\{\lambda_1, \dots, \lambda_A\}$$



$$\sigma_a = \sqrt{\lambda_a}$$



Chemometrics Add-in for Excel

Позволяет пользователю использовать все возможности Excel: расчеты, стандартные функции, графики, и т.п.

Программирование в VBA помогает упростить работу

Все вычисления выполняются очень быстро – «на лету».

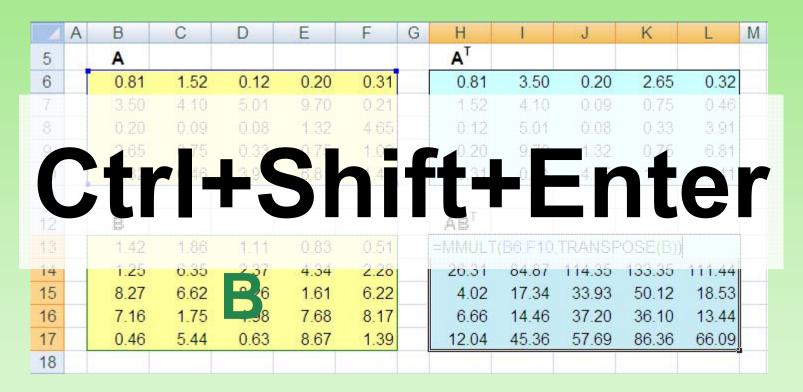
Input Excel User Interface Functions Calculations

刀

sults

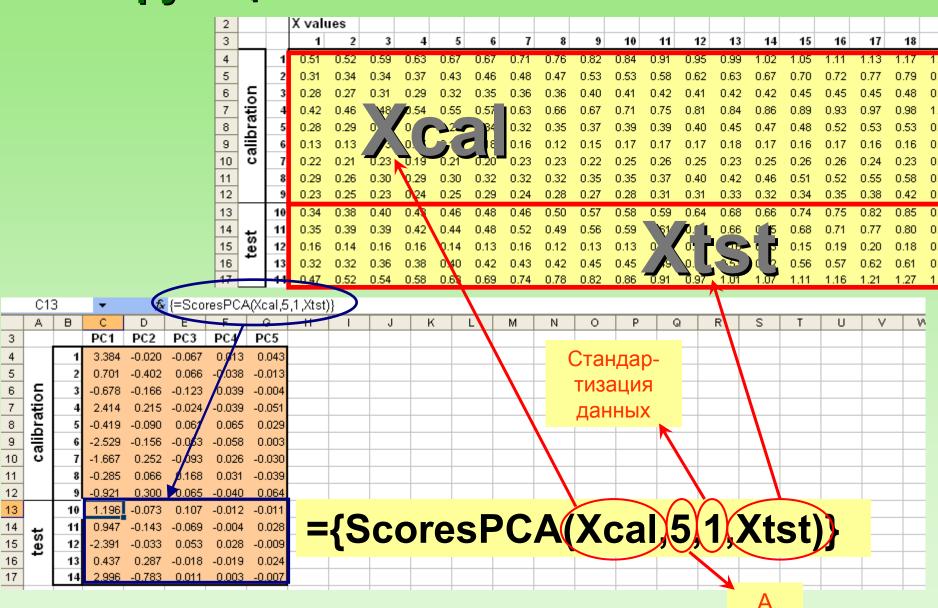
Матричные вычисления в Excel

={TRANSPOSE(B6:F10)}



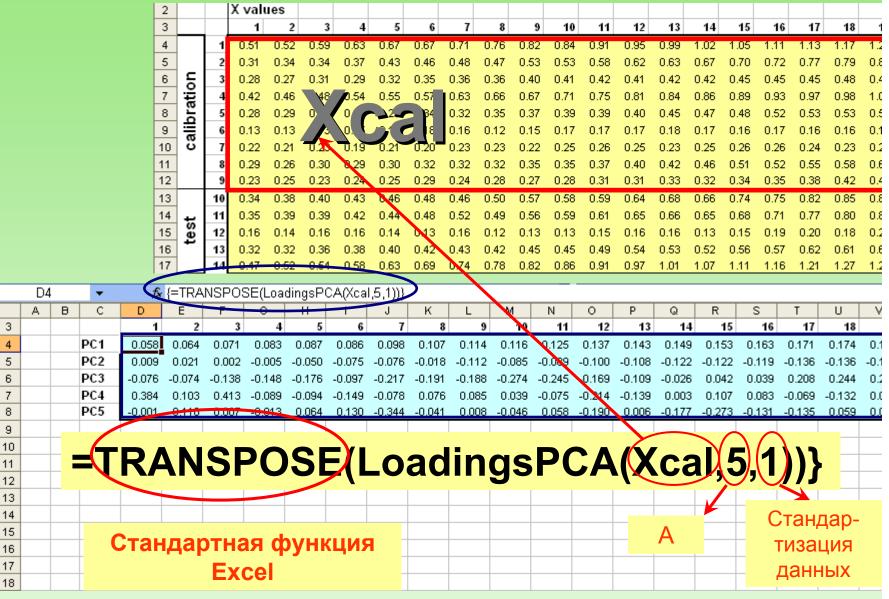
={MMULT(B6:F10,TRANSPOSE(B))}

МГК: функция ScoresPCA

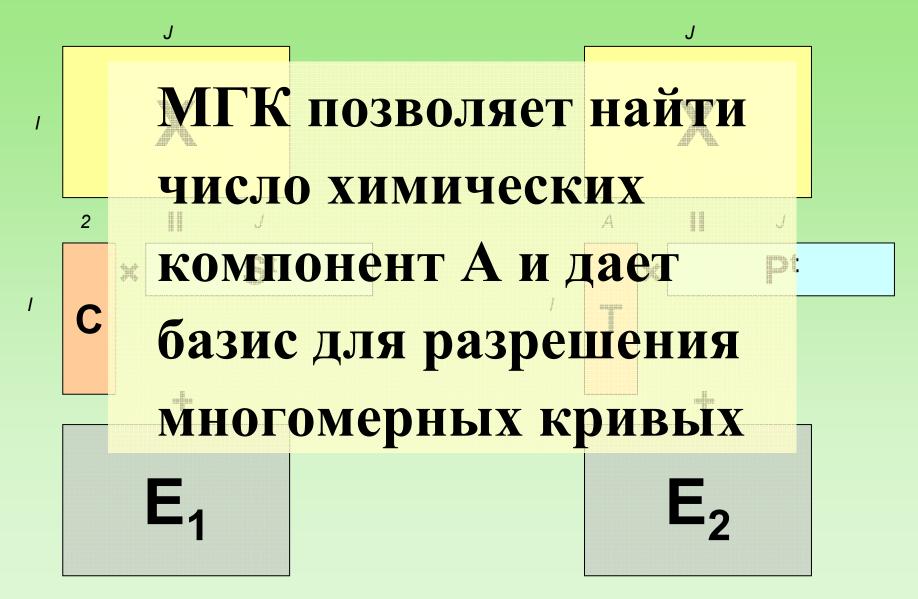


03.10.11

МГК: функция LoadingsPCA



Факторный анализ и анализ ГК



03.10.11

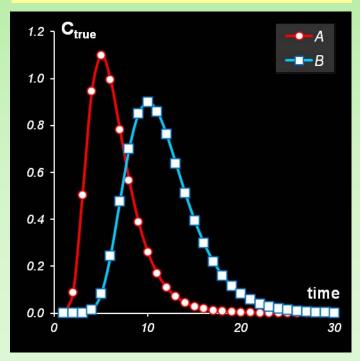
Модельные данные

Прообразом для модельного примера служит задача разделения пиков в высокоэффективной жидкостной хроматографии с детектированием на диодной матрице (ВЭЖХ-ДДМ). В системе имеются два вещества A и B, с профилями элюции $c_{\rm A}(t)$ и $c_{\rm B}(t)$. Тогда спектр смеси

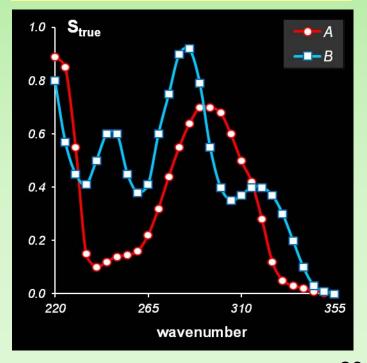
$$x = c_{\mathbf{A}} \cdot s_{\mathbf{A}} + c_{\mathbf{B}} \cdot s_{\mathbf{B}},$$

где s_{A} и s_{B} — спектры чистых веществ.

Профили элюции

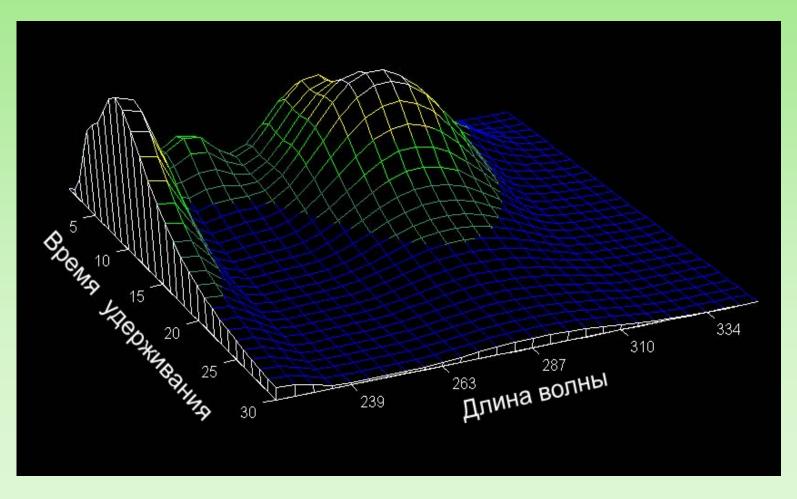


Спектры чистых компонент



Хроматограмма смеси

$$X=CS^t+E$$



Метод главных компонент

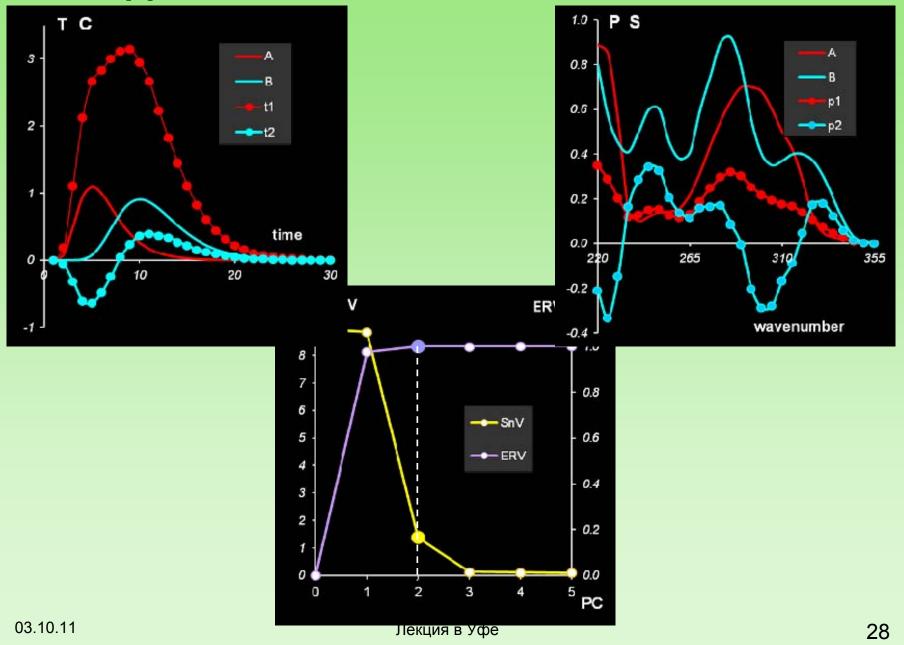
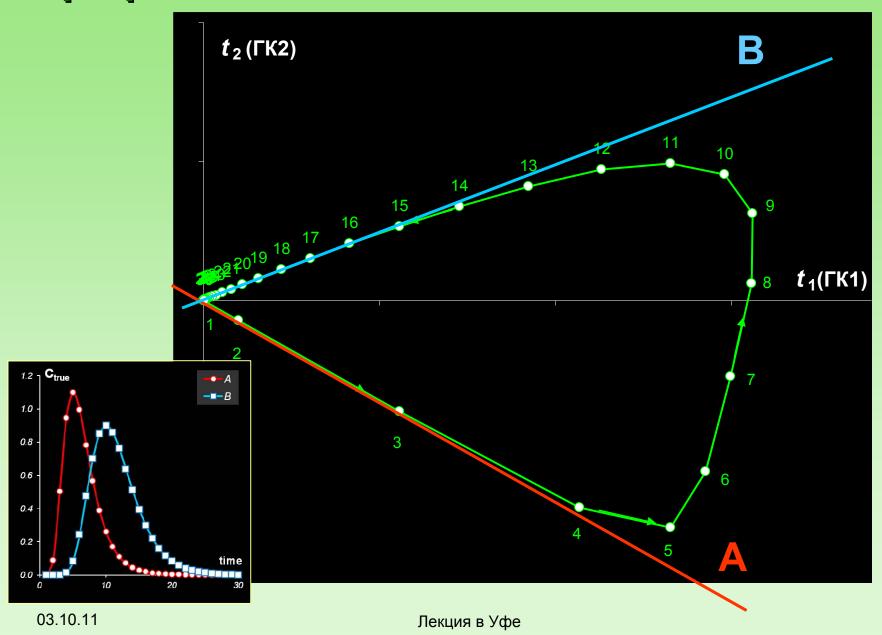


График счетов



29

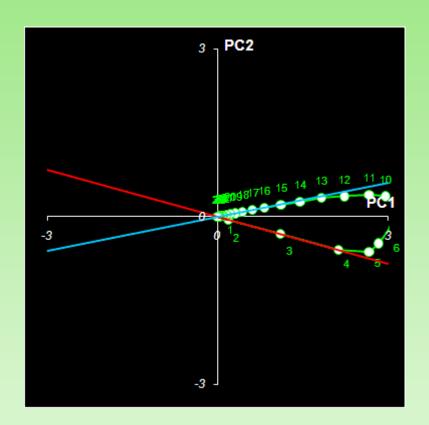
Выводы из графика счетов

- 1. линейные участки = чистые компоненты
- 2. кривые участки = коэлюция
- 3. ближе к началу = меньше интенсивность
- 4. число поворотов = число чистых компонент

Прокрустово преобразование

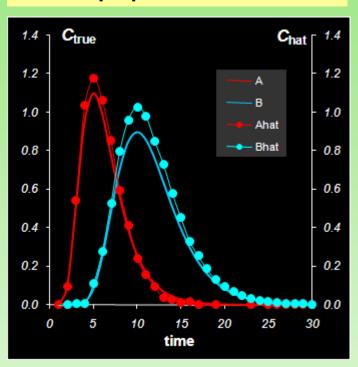


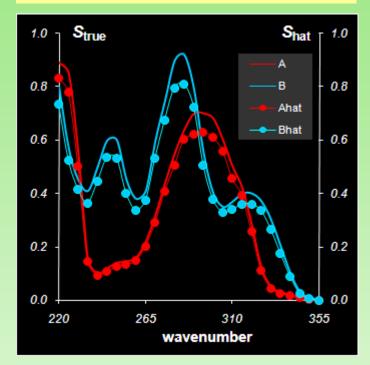
Προκρούστης, 13 в. до н.э.



Результат

Профили элюции



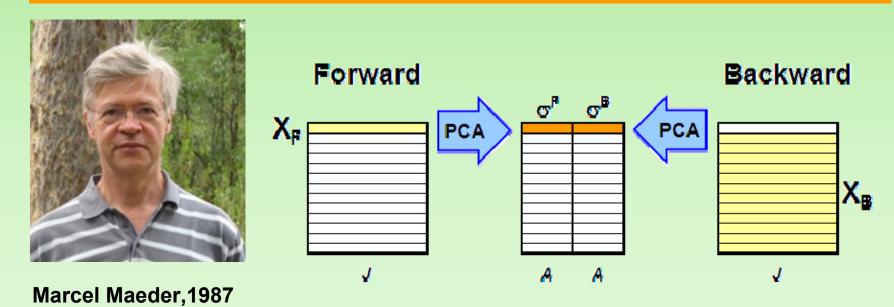


$$\mathbf{C}_{\text{hat}} = \mathbf{T}\mathbf{R}_1\mathbf{R}_2 = \mathbf{T}\mathbf{R}$$

$$\mathbf{S}_{\text{hat}} = \mathbf{P}(\mathbf{R}^{\text{t}})^{-1}$$

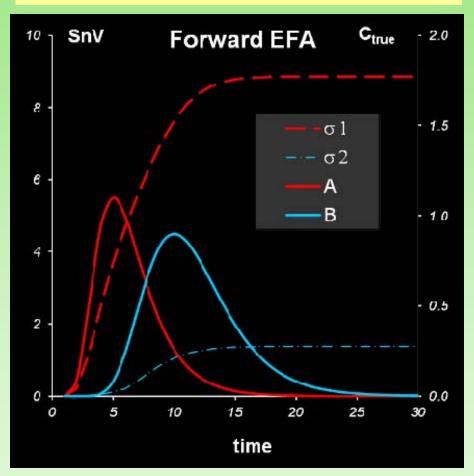
Эволюционный факторный анализ (EFA)

Главная идея эволюционного факторного анализа (EFA, Evolving factor analysis) состоит в том, что метод главных компонент применяется не ко всей матрице X, а к последовательности ее подматриц. В методе EFA используют две последовательности: расширяющаяся матрица (*прямой проход*) и сокращающаяся матрица (*обратный проход*). К каждой подматрице из последовательности применяется PCA и определяются соответствующие сингулярные значения (c.з.) $\sigma_1,...,\sigma_4$.

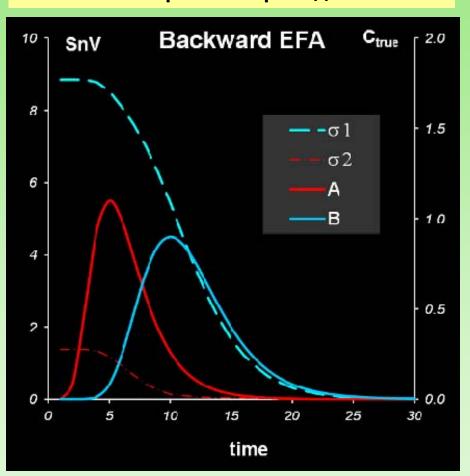


Изменение сингулярных значений





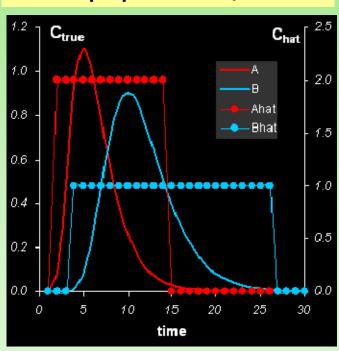
Обратный проход



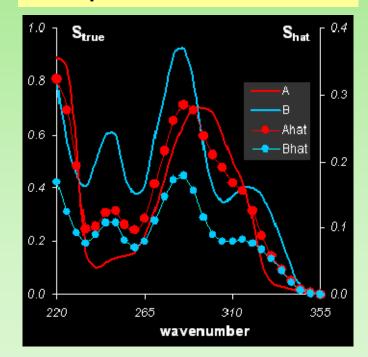
Определение концентрационных окон

$$W_a(t) = \begin{cases} 0, & \min(\sigma_a^{F}, \sigma_{A-a+1}^{B}) < \varepsilon \\ H_a, & \min(\sigma_a^{F}, \sigma_{A-a+1}^{B}) \ge \varepsilon \end{cases}$$

Профили элюции



Спектры чистых компонент



$$S_{hat} = X^t C(C^t C)^{-1}$$

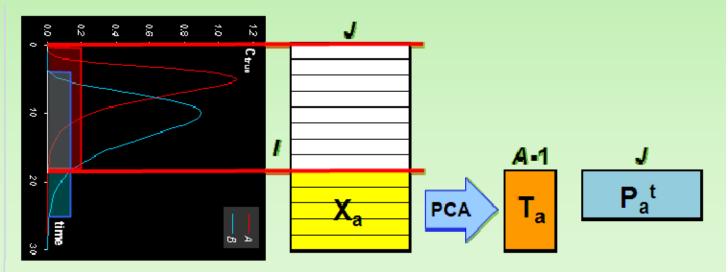
Оконный факторный анализ (WFA)

Идея *оконного факторного анализа* (Windows Factor Analysis, WFA) состоит в удалении из исходных данных ${\bf X}$ тех строчек, в которых концентрация каждого химического компонента отлична от нуля. Оставшуюся после удаления матрицу ${\bf X}_a$ разлагают с помощью PCA и определяют матрицу нагрузок ${\bf P}_a$.

Матрица \mathbf{P}_a несет в себе информацию о спектрах всех химических компонентов, кроме исключенного. Поэтому для того, чтобы определить этот исключенный компонент, надо найти вектор ортогональный к \mathbf{P}_a .



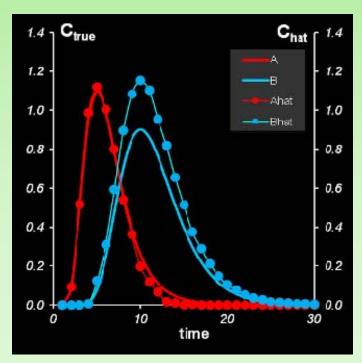
Edmund R. Malinowski, 1992

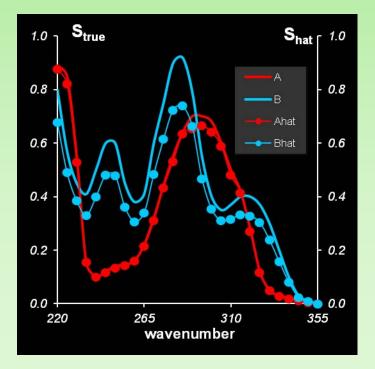


Алгоритм WFA

Для каждого химического компонента a=1,...,A:

- 1. Определяется матрица \mathbf{X}_a
- 2. Выполняется МГК: $\mathbf{X}_a = \mathbf{T}_a (\mathbf{P}_a)^{\mathrm{t}} + \mathbf{E}_a$, и находится матрица \mathbf{P}_a
- 3. Вычисляется матрица $\mathbf{Z}_a = \mathbf{X} \mathbf{X} \mathbf{P}_a (\mathbf{P}_a)^{\mathsf{t}}$
- 4. Опять выполняется МГК: $\mathbf{Z}_a = \mathbf{t}_{\mathbf{Z}}(\mathbf{p}_{\mathbf{Z}})^t + \mathbf{E}_{\mathbf{Z}}$, и находится вектор счетов $\mathbf{t}_{\mathbf{Z}}$
- 5. Все отрицательные элементы в \mathbf{t}_{Z} заменяются нулями. Это и есть \mathbf{c}_a





Чередующиеся наименьшие квадраты (ALS)

Метод чередующихся наименьших квадратов (Alternating Least-Squares, ALS) — это итерационный метод решения зада МСR. Вместо МГК он использует принцип наименьших квадратов, последовательно применяя на каждом шагу формулы

По известному спектру S, находим концентрационный профиль C



Romá Tauler, 1993

$$\mathbf{C}_{\text{hat}} = \mathbf{X}\mathbf{S}(\mathbf{S}^{\mathsf{t}}\mathbf{S})^{-1}$$

По известному концентрационному профилю C, находим спектр S

$$\mathbf{S}_{\text{hat}} = \mathbf{X}^{\text{t}}\mathbf{C}(\mathbf{C}^{\text{t}}\mathbf{C})^{-1}$$

Алгоритм ALS

- 1. Методом PCA определяется число компонент, а методом EFA концентрационные окна.
- 2. Задается начальное приближение для матрицы концентраций \mathbf{C}_{in}
- 3. Матрица \mathbf{C}_{in} подправляется т.ч. \mathbf{C}_{in} ≥0. Получается матрица $\mathbf{C}_{\mathsf{hat}}$
- 4. Находится оценка матрицы спектров \mathbf{S}_{in}

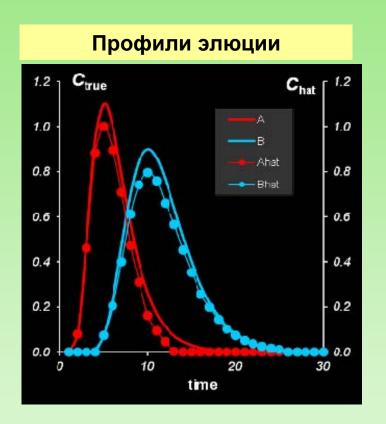
$$\mathbf{S}_{\text{in}} = \mathbf{X}^{\text{t}} \mathbf{C}_{\text{hat}} (\mathbf{C}_{\text{hat}}^{\text{t}} \mathbf{C}_{\text{hat}})^{-1}$$

- 5. Матрица \mathbf{S}_{in} подправляется т.ч. $\mathbf{S}_{\text{in}} {\ge} 0$. Получается матрица \mathbf{S}_{hat}
- 6. Определяется оценка матрицы концентраций $\mathbf{C}_{\mathrm{out}}$

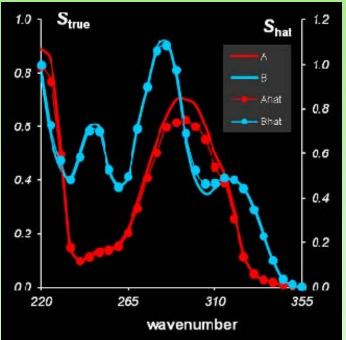
$$\mathbf{C}_{\text{out}} = \mathbf{X} \mathbf{S}_{\text{hat}} (\mathbf{S}_{\text{hat}}^{\text{t}} \mathbf{S}_{\text{hat}})^{-1}$$

- $m{m{U}}$. Матрица $m{C}_{ ext{in}}$ заменяется матрицей $m{C}_{ ext{out}}$
- 8. Шаги 3-7 повторяются до сходимости

Результат ALS



Спектры чистых компонент S_{true} 1.0 7



Дуальность MCR

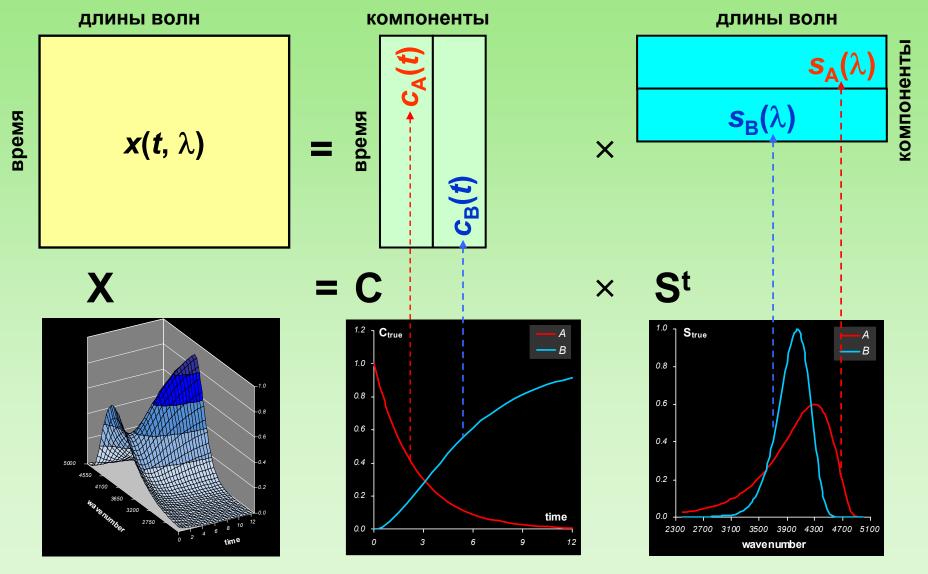
«Хроматографическая» и «кинетическая» задачи переходят друг в друга при транспонировании –

$$X = CS^t + E$$

$$X^t = SC^t + E^t$$

Однако, помимо формальной стороны дела есть и содержательная — априорное знание. Чем полнее мы его учитываем, тем точнее решение.

Кинетическая задача



Кинетическая схема

Предположим, что нам известна кинетика изменения концентраций в задаче MCR, например – это последовательная реакция первого порядка

$$A \xrightarrow{k_1} B \xrightarrow{k_2} C$$

Такая кинетическая схема описывается системой дифференциальных уравнений

$$\frac{dA}{dt} = -k_1 A; \qquad A(0) = A_0$$

$$\frac{dB}{dt} = k_1 A - k_2 B; \qquad B(0) = 0$$

$$\frac{dC}{dt} = k_2 B; \qquad C(0) = 0$$

Постараемся учесть это априорное знание при решении задачи разрешения спектральных данных, снятых по ходу этого процесса.

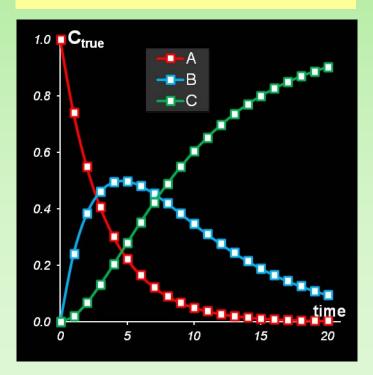
Модельные данные

Рассчитаем концентрации для нашей схемы

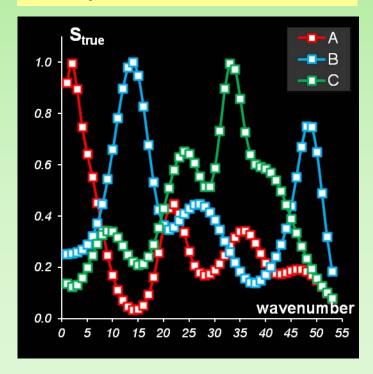
$$A \xrightarrow{k_1} B \xrightarrow{k_2} C$$

при A_0 =1, k_1 =0.30, k_2 =0.15.

Кинетика

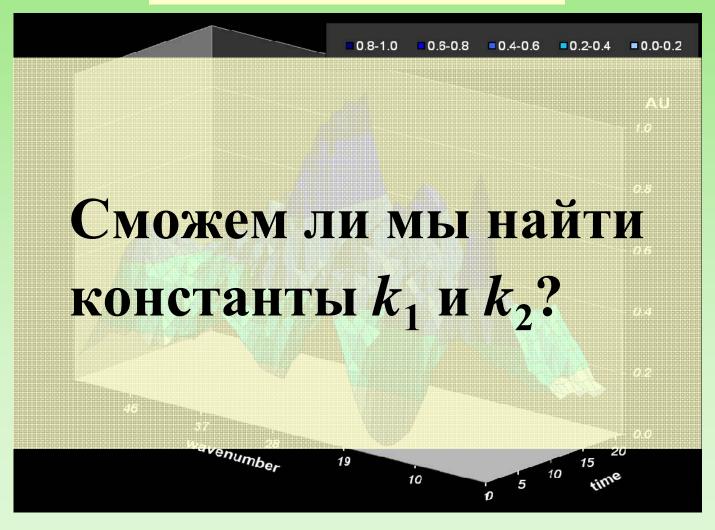


Спектры чистых компонент



Спектральные данные

$$X=CS^t+E$$

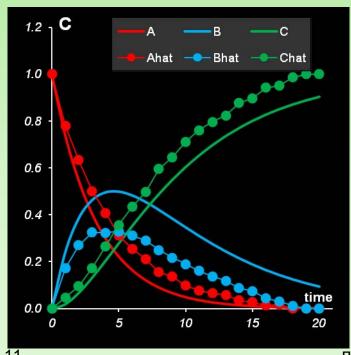


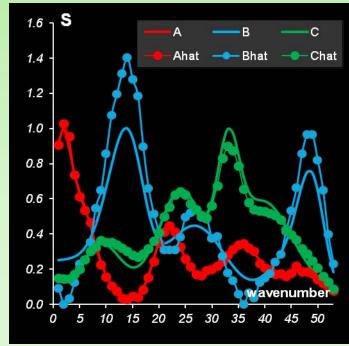
Формальное моделирование (Soft ALS)

Для анализа данных **X** можно применить метод ALS с учетом:

- 1. Число компонент известно их три: А, В, С;
- 2. Спектры и концентрации неотрицательны причем A(0)=1, B(0)=C(0)=0
- 3. Система замкнута, т.е. в любой момент t:

$$A(t)+B(t)+C(t)=\text{Const}=1$$
.



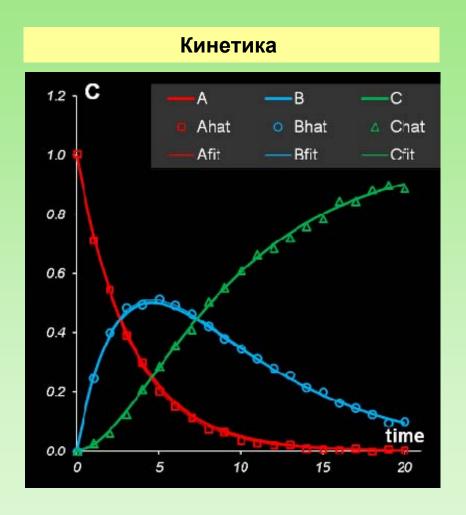


03.10.11 Лекция в Уфе 46

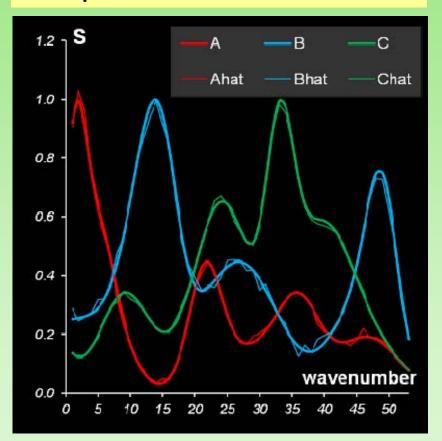
Содержательное моделирование (Hard ALS)

- 1. Задается начальное приближение для матрицы концентраций \mathbf{C}_{in}
- 2. Задаются начальные значения кинетических параметров $\mathbf{k} = (k_1, \dots k_p)$
- ightharpoonup. Матрица \mathbf{C}_{in} подправляется т.ч $\mathbf{C}_{\text{in}} \ge 0$. Получается матрица \mathbf{C}_{hat}
- 4. По кинетическим уравнениям вычисляется матрица $\mathbf{C}_{\mathrm{fit}}(\mathbf{k})$
- 5. Подбираются значения $k_1, ..., k_p$ для которых min ($\|\mathbf{C}_{hat} \mathbf{C}_{fit}(\mathbf{k})\|^2$)
- 6. Находится оценка матрицы спектров: $\mathbf{S}_{in} = \mathbf{X}^t \mathbf{C}_{fit} (\mathbf{C}_{fit}^t \mathbf{C}_{fit})^{-1}$
- 7. Матрица \mathbf{S}_{in} подправляется т.ч. $\mathbf{S}_{\text{in}} {\ge} 0$. Получается матрица \mathbf{S}_{hat}
- 8. Находится оценка матрицы концентраций: $\mathbf{C}_{\text{out}} = \mathbf{X} \mathbf{S}_{\text{hat}} (\mathbf{S}_{\text{hat}}^{\text{t}} \mathbf{S}_{\text{hat}})^{-1}$
 - Матрица \mathbf{C}_{in} заменяется матрицей $\mathbf{C}_{\mathsf{out}}$
- 10. Шаги 3-9 повторяются до сходимости.

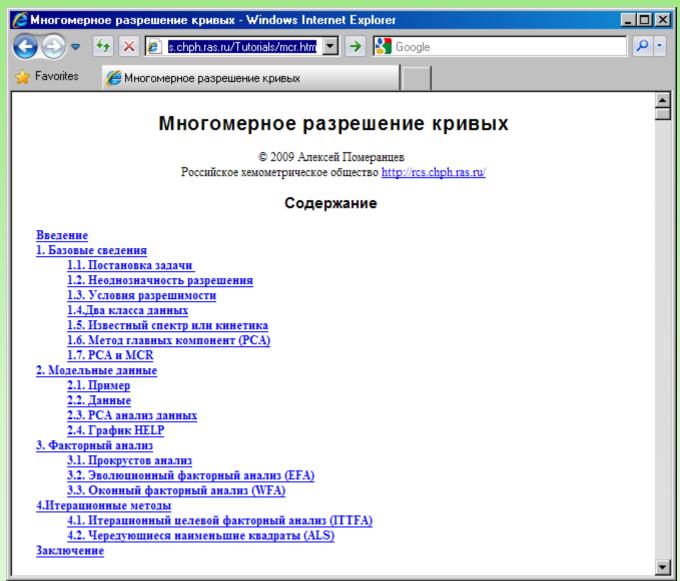
Результат Hard ALS



Спектры чистых компонент



http://rcs.chph.ras.ru/Tutorials/mcr.htm



8th Winter Symposium on Chemometrics



Школа: 25.02-26.02.12, МГУ

Симпозиум: 27.02-02.03.12, Московская обл.



