

# Zastosowanie analizy faktorowej do równoczesnego oznaczania aspiryny, salicylamidu i kofeiny w środkach uśmierzających ból

## 1. Wprowadzenie

Równoczesne oznaczenie ilości poszczególnych składników obecnych w próbce analitycznej stanowi częsty problem w praktyce laboratoryjnej rozwiązywany zazwyczaj przez zastosowanie rozmaitych technik rozdzielania składników takich jak na przykład chromatografia gazowa lub cieczą, spektroskopia masowa, ekstrakcja etc. To samo zadanie wypełniają z powodzeniem techniki chemometryczne objęte ogólnym mianem *wielowymiarowych technik kalibracyjnych*. Jedną z takich technik, konceptualnie dopracowanych i skutecznych w praktyce, jest namiarowa analiza faktorowa (*Target Factor Analysis - TFA*). Namiarowa analiza faktorowa jest jedną z wielu metod analizy faktorowej (*Factor Analysis - FA*) zwanej również zastępczo analizą głównych składników (*Principal Component Analysis - PCA*). Jej zasady i sposób praktycznego zastosowania do ilościowego oznaczenia składu próbki analitycznej będącej mieszaniną trzech znanych składników stanowią jeden z tematów niniejszego ćwiczenia.

Podstawą wszystkich metod analizy faktorowej są macierze danych obrazujące zmiany w analizowanym układzie wywołane zazwyczaj przez ewolucję składu będącą następstwem procesów chemicznych, zmian pH, zmian temperaturowych itp. Na ogół przyjmuje się, iż każdy wiersz takiej macierzy odpowiada pojedynczemu obiektowi (próbce), a każda kolumna związana jest z pojedynczą zmienną. Przykładem tego typu macierzy danych może być macierz, której wiersze odpowiadają poszczególnym widmom a kolumny reprezentują zmiany intensywności dla określonej długości fali np. macierz widm absorpcyjnych lub fluorescencyjnych. Jest rzeczą oczywistą, iż w przypadku jednej próbki wieloskładnikowej można zarejestrować tylko jedno widmo, a zatem zastosowanie analizy faktorowej nie jest tutaj możliwe. Problem ten można jednak rozwiązać stosując na przykład *uogólnioną metodę dodatku wzorca*, co zostanie zademonstrowane poniżej w dalszej części tego opracowania.

*Notacja.* W analizie faktorowej dużą tłustą czcionką oznacza się macierze, natomiast małą tłustą czcionką wektory. Macierz transponowaną w stosunku do macierzy  $\mathbf{X}$  oznacza się jako  $\mathbf{X}^T$ , macierz będącą oszacowaniem macierzy  $\mathbf{X}$  jako  $\hat{\mathbf{X}}$  i wreszcie tak zwaną macierz obciętą jako  $\bar{\mathbf{X}}$ . Ta ostatnia macierz jest macierzą reprodukującą macierz  $\mathbf{X}$  w oparciu o określoną liczbę istotnych faktorów, o czym mówimy poniżej.

Ogólnie, w analizie faktorowej przyjmuje się, że każdą macierz danych widmowych  $\mathbf{X}$ , o wymiarach  $m \times n$ , można rozłożyć w taki sposób, że

$$\mathbf{X} = \bar{\mathbf{X}} + \mathbf{E} = \bar{\mathbf{C}}\bar{\mathbf{S}}^T + \mathbf{E} \quad (1)$$

gdzie macierz  $\bar{\mathbf{C}}$ , o wymiarach  $m \times f$ , jest macierzą abstrakcyjnych profili stężeniowych czyli abstrakcyjnych zmian w składzie, macierz  $\bar{\mathbf{S}}$ , o wymiarach  $n \times f$ , jest macierzą abstrakcyjnych profili widmowych czyli abstrakcyjnych widm, natomiast macierz  $\mathbf{E}$ , o wymiarach  $n \times m$ , jest macierzą szumów (błędów) doświadczalnych. Wskaźnik  $f$  oznacza liczbę istotnych faktorów i jest równoważny rzędowi macierzy  $\mathbf{X}$ . Po wyznaczeniu właściwej liczby istotnych faktorów  $f$  macierz  $\bar{\mathbf{X}}$  powinna wyjaśniać prawie całą wariancję macierzy

danych  $\mathbf{X}$ , to znaczy powinna reprodukować macierz  $\mathbf{X}$  z dokładnością do szumów, które ujęto w macierzy  $\mathbf{E}$ .

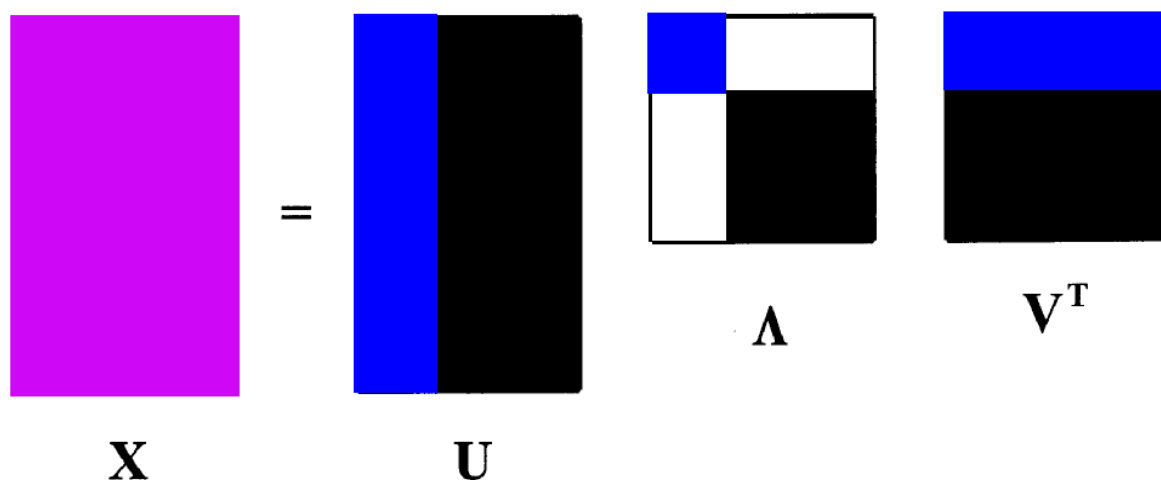
Istnieją liczne algorytmy, które prowadzą do rozkładu macierzy  $\mathbf{X}$  zgodnie z zapisem równania (1). Do najpowszechniej stosowanych zalicza się *rozkład według wartości szczególnych* (*Singular Value Decomposition - SVD*). Macierz danych  $\mathbf{X}$  o wymiarach  $m \times n$  rozkłada się na iloczyn nie dwóch, lecz trzech macierzy, przy czym macierz  $\mathbf{U}$  jest również macierzą o wymiarach  $m \times n$ , macierz  $\mathbf{\Lambda}$  jest macierzą diagonalną o wymiarach  $n \times n$ , a macierz  $\mathbf{V}$  posiada identyczne wymiary jak macierz  $\mathbf{\Lambda}$ . Mamy zatem

$$\mathbf{X} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T \quad (2)$$

przy czym macierz  $\mathbf{U}$  jest kolumnowo ortogonalna, tzn.  $\mathbf{U}^T \mathbf{U} = \mathbf{1}$ , macierz  $\mathbf{V}$  jest kolumnowo ortogonalna, ale również wierszowo ortogonalna, czyli  $\mathbf{V}^T \mathbf{V} = \mathbf{1}$  oraz  $\mathbf{V} \mathbf{V}^T = \mathbf{1}$ , natomiast macierz  $\mathbf{\Lambda}$  zawiera na głównej diagonali wartości szczególne uporządkowane od wartości największej do najmniejszej. Jeśli znana jest liczba czynników  $f$  zawartych w macierzy danych  $\mathbf{X}$ , to równanie (2) możemy formalnie zapisać w następujący sposób

$$\mathbf{X} = \sum_{l=1}^f \mathbf{u}_l \lambda_l \mathbf{v}_l^T + \sum_{k=l+1}^n \mathbf{u}_k \lambda_k \mathbf{v}_k^T \quad (3)$$

Pierwsza suma w równaniu (3) definiuje macierz obcięta  $\bar{\mathbf{X}}$  w równaniu (1), natomiast druga suma w równaniu (3) odpowiada macierzy szumów doświadczalnych  $\mathbf{E}$  w równaniu (1). Rozkład według wartości szczególnych zobrazowano graficznie na Rys. 1. Macierz obcięta  $\bar{\mathbf{X}}$  powstaje w wyniku przemnożenia „niebieskich” podmacierzy „wyciętych” z macierzy  $\mathbf{U}$ ,  $\mathbf{\Lambda}$  i  $\mathbf{V}^T$ , zwanych  $\bar{\mathbf{U}}$ ,  $\bar{\mathbf{\Lambda}}$  i  $\bar{\mathbf{V}}^T$ , natomiast macierz  $\mathbf{E}$  to efekt przemnożenia „czarnych” podmacierzy macierzy  $\mathbf{U}$ ,  $\mathbf{\Lambda}$  i  $\mathbf{V}^T$ , które nie posiadają swoich specyficznych nazw. „Niebieskie” i „czarne” podmacierze wycięte z macierzy  $\mathbf{\Lambda}$  to oczywiście kwadratowe macierze diagonalne (patrz Rys. 1). W nawiązaniu do równania (1) przyjmuje się na ogół, że  $\bar{\mathbf{C}} = \bar{\mathbf{U}}$ , natomiast  $\bar{\mathbf{S}}^T = \bar{\mathbf{\Lambda}} \bar{\mathbf{V}}^T$ .



Rys. 1 Rozkład „ekonomiczny” macierzy danych  $\mathbf{X}$  w oparciu o metodę SVD.

Wektory  $\mathbf{u}_l$  i  $\mathbf{v}_l$  łączą następujące relacje

$$\mathbf{X}\mathbf{v}_l = \lambda_l \mathbf{u}_l \quad (4a)$$

$$\mathbf{X}^T \mathbf{u}_l = \lambda_l \mathbf{v}_l \quad (4b)$$

co prowadzi do następujących macierzowych zagadnień własnych

$$\mathbf{X}^T \mathbf{X} \mathbf{v}_l = \gamma_l \mathbf{v}_l \quad (5a)$$

$$\mathbf{X} \mathbf{X}^T \mathbf{u}_l = \gamma_l \mathbf{u}_l \quad (5b)$$

Rozwiązanie równań (5a) i (5b) stanowi treść najstarszego algorytmu PCA, zwanego algorytmem Jacobi'ego, przy czym jeśli  $m > n$  to macierz  $\mathbf{X}^T \mathbf{X}$  (o wymiarach  $n \times n$ ) nosi miano rdzeniowej albo małej macierzy kowariancji, natomiast macierz  $\mathbf{X} \mathbf{X}^T$  (o wymiarach  $m \times m$ ) nazywa się dużą albo pełną macierzą kowariancji. Wektory  $\mathbf{v}_l$  i  $\mathbf{u}_l$  to wektory własne, odpowiednich zagadnień własnych (5a) i (5b);  $\gamma_l$  to stowarzyszone z nimi wartości własne. Z porównania równań (4) i (5) wynika, że pomiędzy wartościami własnymi  $\gamma_l$  i wartościami szczególnymi  $\lambda_l$  zachodzi związek

$$\Lambda = \Gamma^{\frac{1}{2}} \quad (6)$$

Łatwo pokazać, że oba algorytmy analizy faktorowej, algorytm Jacobi'ego i metoda SVD, prowadzą do tych samych rozwiązań. Weźmy dla przykładu macierz  $\mathbf{X}^T \mathbf{X}$ . Formalnie możemy napisać

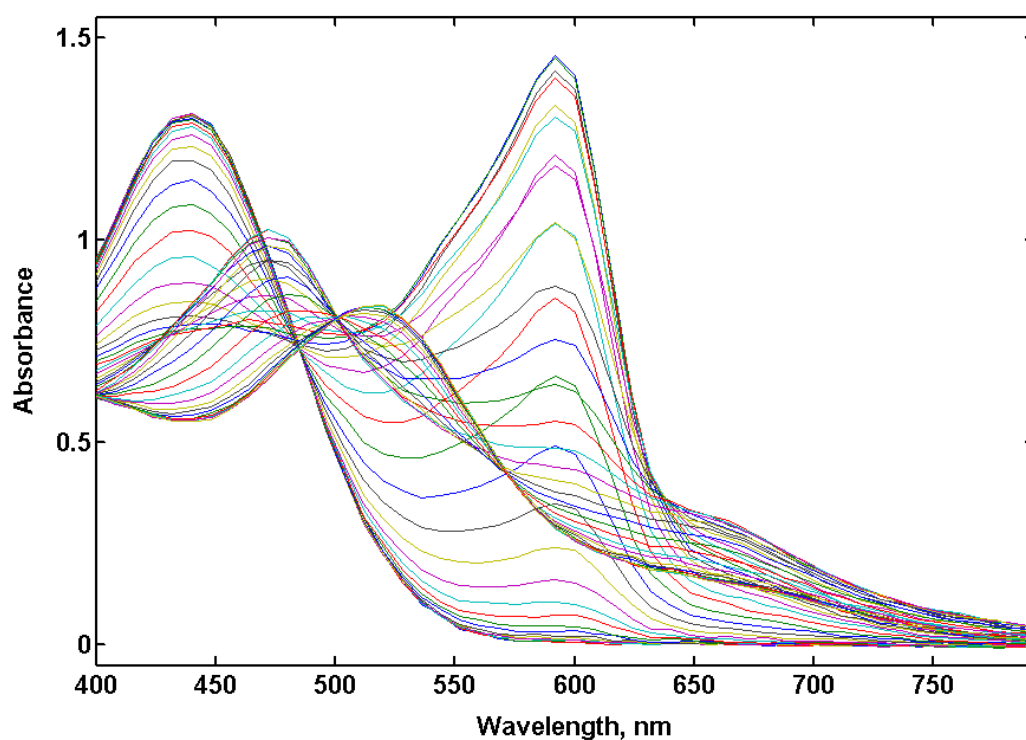
$$\begin{aligned} \mathbf{X}^T \mathbf{X} &= (\mathbf{U} \Lambda \mathbf{V}^T)^T \mathbf{U} \Lambda \mathbf{V}^T \\ &= (\mathbf{V}^T)^T \Lambda^T \mathbf{U}^T \mathbf{U} \Lambda \mathbf{V}^T \\ &= \mathbf{V} \Lambda^T \Lambda \mathbf{V}^T \\ &= \mathbf{V} \Lambda^2 \mathbf{V}^T \end{aligned} \quad (7)$$

Mnożąc obie strony równania (7) prawostronnie przez  $\mathbf{V}$  otrzymamy

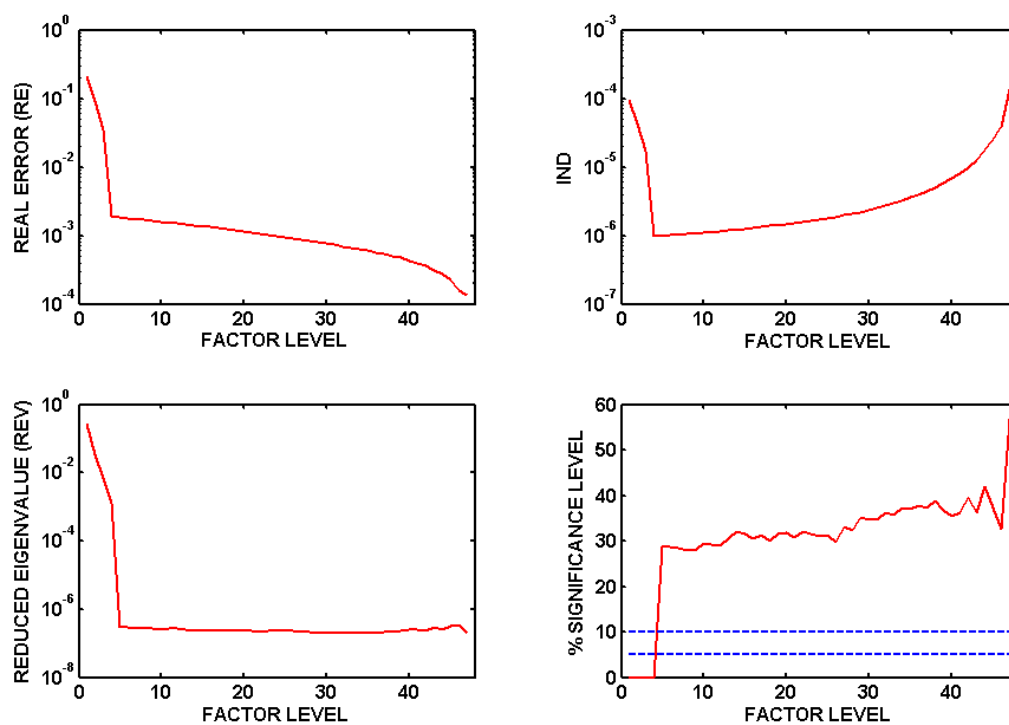
$$\begin{aligned} \mathbf{X}^T \mathbf{X} \mathbf{V} &= \mathbf{V} \Lambda^2 \mathbf{V}^T \mathbf{V} \\ &= \mathbf{V} \Lambda^2 \\ &= \mathbf{V} \Gamma \end{aligned} \quad (8)$$

Równanie (8) jest zatem uogólnionym zapisem szczególnego zagadnienia własnego przedstawionego w równaniu (5a). W analogiczny sposób, wychodząc z macierzy  $\mathbf{XX}^T$ , można znaleźć uogólnienie równania (5b).

Zazwyczaj liczba składników posiadających wkłady widmowe w analizowanym zakresie spektralnym odpowiada liczbie istotnych czynników, ale może się też zdarzyć, że zmiany widmowe wywołane zostaną przez czynniki natury fizycznej takie jak systematyczne dryfowanie tła w czasie pomiarów lub zmiany temperaturowe. Liczbę istotnych czynników  $f$  wyznacza się tradycyjnie w oparciu o kilka kryteriów empirycznych i statystycznych, do których należą między innymi: reszkowe odchylenie standardowe czyli tzw. błąd rzeczywisty (RE) i związana z nim czynnikowa funkcja wskaźnikowa Malinowskiego (IND), zredukowane wartości własne (REV), test Fishera sprawdzający negatywną hipotezę przynależności poszczególnych wartości własnych do grupy „szumowych” wartości własnych (SL), i inne. Na Rys. 2 przedstawiono dla celów ilustracyjnych 48 widm absorpcyjnych barwnika fioletu pyrokatechinowego, zestawionych w macierz danych widmowych  $\mathbf{X}$  o wymiarach  $50 \times 48$  i zarejestrowanych w zakresie widmowym 400 – 792 nm z krokiem 8 nm. Widma te nie są wystarczająco gładkie, gdyż zastosowany krok widmowy był zbyt duży. Przedstawione widma pochodzą z pakietu programów do Analizy Faktorialnej *TARGET96M* opracowanego przez Malinowskiego. Z powtórzonych pomiarów dla tej samej serii roztworów wyznaczono błąd pomiarowy absorbancji równy w przybliżeniu  $\pm 0.002$  umownej jednostki absorbancji. Fiolet pyrokatechinowy jest barwnikiem, który może wymieniać cztery jony wodorowe. Pięć różnych indywiduów chemicznych wywodzących się z fioletu pyrokatechinowego można przedstawić ogólnie jako  $\text{H}_4\text{I}$ ,  $\text{H}_3\text{I}^-$ ,  $\text{H}_2\text{I}^{2-}$ ,  $\text{H}_3\text{I}^{3-}$  i  $\text{I}^{4-}$ . 48 różnych roztworów fioletu pyrokatechinowego odpowiada 48 różnym wartościom pH zmieniającym się od 4.6 do 14.0 z krokiem 0.2. Na Rys. 3 przedstawiono wyniki kilku kryteriów określających liczbę czynników odpowiedzialnych za macierz  $\mathbf{X}$  widm absorpcji fioletu pyrokatechinowego. Wszystkie kryteria wskazują zgodnie na obecność 4 czynników. Są to jonowe formy chemiczne fioletu pyrokatechinowego; forma niezdisocjowana  $\text{H}_4\text{I}$  nie wnosi swojego wkładu widmowego do macierzy  $\mathbf{X}$ , gdyż pojawia się dopiero poniżej pH 4.0.



Rys. 2 Widma absorpcji różnych form jonowych fioletu pyrokatechinowego w funkcji pH.



Rys. 3 Cztery kryteria wyznaczające liczbę czynników dla macierzy X fioletu pyrokatechinowego: u góry RE i IND, na dole REV i SL. Wyznaczona

jednoznacznie przez wszystkie kryteria liczba czynników  $f$  wynosi 4. Wniosek ten jest konsekwencją następujących zachowań: błąd rzeczywisty RE jest duży i obniża się gwałtownie dla pierwszych 4 czynników, a później obniża się o wiele wolniej, funkcja wskaźnikowa IND wykazuje minimum dla  $f = 4$ , zredukowane wartości własne REV osiągają plateau począwszy od  $f = 5$ , co wskazuje na to, że wszystkie wartości własne powyżej  $f = 4$  należą do grupy wartości „szumowych”, poziom istotności wartości własnych SL podnosi się wyraźnie ponad dopuszczalne granice począwszy od  $f = 5$ , co lokuje wszystkie wartości własne powyżej  $f = 4$  wśród grupy „szumowych” wartości własnych.

Macierze abstrakcyjnych profili  $\bar{\mathbf{C}}$  i  $\bar{\mathbf{S}}$  stanowią kombinacje liniowe odpowiednio rzeczywistych profili stężeniowych  $\mathbf{C}$  i rzeczywistych profili widmowych  $\mathbf{S}$  pojedynczych składników. Po ustaleniu właściwej liczby czynników  $f$  i odrzuceniu macierzy błędów  $\mathbf{E}$ , otrzymuje się w przybliżeniu prawdziwą relację

$$\mathbf{X} \cong \bar{\mathbf{C}}\bar{\mathbf{S}}^T = \bar{\mathbf{X}} \quad (9)$$

Głównym zadaniem analizy czynnikowej jest teraz znalezienie takiej macierzy transformacji  $\mathbf{R}$  (macierzy rotacji), która przekształci abstrakcyjne profile stężeniowe  $\bar{\mathbf{C}}$  i widmowe  $\bar{\mathbf{S}}$  w przewidywane rzeczywiste profile stężeniowe  $\hat{\mathbf{C}}$  i widmowe  $\hat{\mathbf{S}}$  poszczególnych składników.

$$\mathbf{X} \cong \bar{\mathbf{C}}\mathbf{R}\mathbf{R}^{-1}\bar{\mathbf{S}}^T = (\bar{\mathbf{C}}\mathbf{R})(\mathbf{R}^{-1}\bar{\mathbf{S}}^T) = \hat{\mathbf{C}}\hat{\mathbf{S}}^T = \hat{\mathbf{X}} \quad (10)$$

Problem wyznaczenia wszystkich elementów macierzy transformacji  $\mathbf{R}$ , o wymiarach  $f \times f$ , nie jest problemem trywialnym i jego rozwiązanie klasycznymi metodami analizy czynnikowej jest możliwe tylko wtedy, gdy w analizowanej mieszaninie spektralnej zachodzi specyficzna ewolucja składu.

### 1.1 Klasyczna metoda najmniejszych kwadratów (Classical Least Squares - CLS)

Wyznaczenie wszystkich elementów macierzy rotacji  $\mathbf{R}$  staje się automatycznie bardzo proste, jeśli macierz danych widmowych  $\mathbf{X}$  odpowiada roztworom kalibracyjnym  $\mathbf{C}$ , o znanym udziale każdego składnika w każdym roztworze. Każda kolumna macierzy  $\mathbf{C}$  zawiera wówczas zmiany stężeniowe tylko jednego składnika. Możemy w takim przypadku napisać

$$\mathbf{C} = \hat{\mathbf{C}} \cong \bar{\mathbf{C}}\mathbf{R} \quad (11)$$

Rozwiązując to równanie macierzowe przy użyciu metody najmniejszych kwadratów otrzymuje się zależność

$$\mathbf{R} = (\bar{\mathbf{C}}^T\bar{\mathbf{C}})^{-1}(\bar{\mathbf{C}}^T\hat{\mathbf{C}}) \quad (12)$$

która pozwala wyznaczyć macierz rzeczywistych profili widmowych wszystkich składników

$$\hat{\mathbf{S}}^T = \mathbf{R}^{-1}\bar{\mathbf{S}}^T \quad (13)$$

Otrzymane rzeczywiste profile widmowe składników (a ściślej przewidywane rzeczywiste widma) można teraz wykorzystać do wyznaczenia ilości tych składników w innych próbkach zawierających te same składniki. Możliwe są tutaj dwa warianty: *bezpośredniego oznaczenia ilości składników* lub pośredniego z zastosowaniem *uogólnionej metody dodatku wzorca*. W obu przypadkach obecność innych substancji wnoszących swój wkład do analizowanej macierzy danych nie pozwala na poprawne oznaczenie ilości analizowanych składników.

A. *Metoda bezpośredniego oznaczenia ilości składników*. Posiadając rzeczywiste widma wszystkich składników otrzymane poprzez analizę faktorową przeprowadzoną dla roztworów kalibracyjnych, można użyć ich dalej do oznaczenia ilości tych składników w próbkach zawierających te same składniki o nieznanym składzie i generujących macierz danych widmowych  $\mathbf{Y}$ , o wymiarach  $l \times n$ . Ponieważ

$$\mathbf{Y} \equiv \hat{\mathbf{C}}_Y \hat{\mathbf{S}}^T \quad (14)$$

a zatem

$$\hat{\mathbf{C}}_Y = \mathbf{Y} \hat{\mathbf{S}} (\hat{\mathbf{S}}^T \hat{\mathbf{S}})^{-1} \quad (15)$$

Należy tutaj zaznaczyć, iż bezpośrednie wyznaczenie  $\hat{\mathbf{C}}_Y$  z relacji  $\hat{\mathbf{C}}_Y = \mathbf{Y} (\hat{\mathbf{S}}^T)^{-1}$  nie jest możliwe, gdyż macierz  $\hat{\mathbf{S}}$  jest macierzą prostokątną i jej odwrotność nie istnieje, stąd konieczność odwołania się do metody najmniejszych kwadratów i w konsekwencji wprowadzenia macierzy  $\hat{\mathbf{S}} (\hat{\mathbf{S}}^T \hat{\mathbf{S}})^{-1}$  zwanej „prawą pseudoodwrotnością” macierzy  $\hat{\mathbf{S}}^T$ . W celu otrzymania  $l$  roztworów wieloskładnikowych, gdzie  $l > f$ , pobiera się z roztworu podstawowego różne alikwoty objętościowe tego roztworu i uzupełnia rozpuszczalnikiem.

B. *Uogólniona metoda dodatku wzorca*. Do każdego roztworu kalibracyjnego o znanym składzie dodaje się ustaloną objętość pobraną z roztworu podstawowego badanej próbki wieloskładnikowej. Dla  $m$  tak zmodyfikowanych roztworów kalibracyjnych zestawia się macierz danych widmowych  $\mathbf{Z}$ . Poszukuje się rozwiązania w postaci

$$\mathbf{Z} \equiv \hat{\mathbf{C}}_Z \hat{\mathbf{S}}^T = \hat{\mathbf{C}}_Y \hat{\mathbf{S}}^T + \hat{\mathbf{C}} \hat{\mathbf{S}}^T \quad (16)$$

gdzie  $\hat{\mathbf{C}} \hat{\mathbf{S}}^T = \hat{\mathbf{X}}$ , jest macierzą danych widmowych otrzymaną dla roztworów kalibracyjnych przed dodaniem ustalonej objętości badanej próbki. Po wyznaczeniu widm czystych składników zgodnie z równaniem (13) otrzymujemy ostatecznie

$$\hat{\mathbf{C}}_Y = \mathbf{Z} \hat{\mathbf{S}} (\hat{\mathbf{S}}^T \hat{\mathbf{S}})^{-1} - \hat{\mathbf{C}} \quad (17)$$

## 1.2 Namiarowa analiza faktorowa (Target Factor Analysis – TFA)

Widma pojedynczych czystych składników są zazwyczaj bardzo dobrze znane i opisane szczegółowo w odpowiednich atlasach widm. Dotyczy to w szczególności widm absorpcyjnych. Powstaje zatem pytanie czy podejrzewając obecność danego składnika w mieszaninie danych widmowych można skorzystać ze zmierzonego wcześniej widma takiego

składnika i stwierdzić jego obecność w badanej mieszaninie ? Odpowiedź na to pytanie jest pozytywna, a zadanie to wypełnia z powodzeniem *namiarowa analiza faktorowa*. Nazwa tej odmiany analizy faktorowej wywodzi się właśnie z procesu *namierzania* faktorów czyli składników chemicznych w analizowanej macierzy danych widmowych. Aby zrozumieć ideę namierzania faktorów w takiej wersji tzn. z użyciem istniejących widm czystych składników, przekształcimy horyzontalnie zorientowaną macierz danych widmowych  $\mathbf{X}$  przedstawioną w równaniu (10) w macierz ustawioną wertykalnie

$$\mathbf{X}^T \equiv (\bar{\mathbf{C}}\mathbf{R}\mathbf{R}^{-1}\bar{\mathbf{S}}^T)^T = (\bar{\mathbf{S}}^T)^T(\mathbf{R}^{-1})^T\mathbf{R}^T\bar{\mathbf{C}}^T = \bar{\mathbf{S}}\mathbf{R}\mathbf{R}^{-1}\bar{\mathbf{C}}^T = \hat{\mathbf{X}}^T \quad (18)$$

Ostateczny wynik tego przekształcenia jest konsekwencją ortogonalności macierzy rotacji tzn. istnienia relacji  $\mathbf{R}^{-1} = \mathbf{R}^T$ . Odtworzona w równaniu (18) macierz danych widmowych  $\hat{\mathbf{X}}$  nie jest identyczna z zarejestrowaną macierzą danych widmowych  $\mathbf{X}$ , stanowi jednak jej najlepsze możliwe przybliżenie. Wniosek wynikający z równania (18) można inaczej zapisać następująco

$$\hat{\mathbf{X}}^T = (\bar{\mathbf{S}}\mathbf{R})(\mathbf{R}^{-1}\bar{\mathbf{C}}^T) = \hat{\mathbf{S}}\hat{\mathbf{C}}^T \quad (19)$$

Równanie (19) mówi nam, że przewidywaną macierz widm  $\hat{\mathbf{X}}^T$  otrzymuje się z przemnożenia przewidywanej macierzy widm czystych składników  $\hat{\mathbf{S}}$  przez transponowaną przewidywaną macierz zmian stężeniowych czystych składników  $\hat{\mathbf{C}}^T$ . Jedynymi znanymi macierzami w równaniu (19) są macierze obcięte  $\bar{\mathbf{S}}$  i  $\bar{\mathbf{C}}$  otrzymane w wyniku zastosowania metody SVD lub innego algorytmu PCA do rozkładu macierzy widm  $\mathbf{X}^T$ . Z równania (19) wynika, że  $l$ -tą kolumnę macierzy  $\hat{\mathbf{S}}$  otrzymuje się z przemnożenia macierzy  $\bar{\mathbf{S}}$  przez  $l$ -tą kolumnę macierzy  $\mathbf{R}$ , tzn.

$$\hat{\mathbf{s}}_l = \bar{\mathbf{S}}\mathbf{r}_l \quad (20)$$

Wektor  $\hat{\mathbf{s}}_l$  nazywa się *wektorem przewidywanym* (czyli w naszym przypadku przewidywanym widmem), natomiast wektor  $\mathbf{r}_l$  jest stowarzyszonym z nim *wektorem transformacji*. Aby wygenerować wektor  $\mathbf{r}_l$ , który pozwoli otrzymać przewidywany wektor  $\hat{\mathbf{s}}_l$ , w największym stopniu zbliżony do wektora próbnego  $\mathbf{s}_l$  zwanego „namiarem” (czyli do posiadanego widma czystego składnika), należy dalej skorzystać z metody najmniejszych kwadratów, która zminimalizuje odchylenia pomiędzy wektorem namierzanym a wektorem przewidywanym i w ten sposób wytworzy najlepszy z możliwych wektorów transformacji dla rozważanego testu namierzania indywidualnego faktora.

Wektor transformacji  $\mathbf{r}_l$  posiada składowe  $r_{1l}, r_{2l}, \dots, r_{fl}$ . Każdy wiersz macierzy  $\bar{\mathbf{S}}$  można traktować jako wektor wierszowy, na przykład  $i$ -ty wektor macierzy  $\bar{\mathbf{S}}$  to wektor mający składowe  $\bar{s}_{i1}, \bar{s}_{i2}, \dots, \bar{s}_{if}$ . Można go oznaczyć umownie jako wektor  $\bar{\mathbf{s}}_i$ , dla odróżnienia od  $i$ -tej kolumny macierzy  $\bar{\mathbf{S}}$ , która jest wektorem oznaczanym odpowiednio jako  $\bar{\mathbf{s}}_i$ . Iloczyn skalarny wektora  $\bar{\mathbf{s}}_i$  i wektora  $\mathbf{r}_l$  określa  $i$ -ty element wektora  $\hat{\mathbf{s}}_l$ , czyli

$$\hat{s}_{il} = \bar{\mathbf{s}}_i \cdot \mathbf{r}_l = \bar{s}_{i1}r_{1l} + \bar{s}_{i2}r_{2l} + \dots + \bar{s}_{if}r_{fl} \quad (21)$$



Sumowanie we wzorze (21) przebiega po wszystkich  $f$  głównych faktorach. Mnożąc każdy wiersz macierzy  $\bar{\mathbf{S}}$  przez wektor  $\mathbf{r}_l$  otrzymujemy wszystkie elementy wektora przewidywanego  $\hat{\mathbf{s}}_l$ , czyli  $\hat{s}_{1l}, \hat{s}_{2l}, \dots, \hat{s}_{nl}$ . Każdy element tego wektora należy następnie porównać z odpowiednim elementem wektora próbnego  $\mathbf{s}_l$ , posiadającego składowe  $s_{1l}, s_{2l}, \dots, s_{nl}$ . Różnicę pomiędzy wartością  $\hat{s}_{il}$  a wartością  $s_{il}$  zapisujemy jako  $\Delta s_{il}$ .

$$\Delta s_{il} = \hat{s}_{il} - s_{il} = \bar{s}_{i1}r_{1l} + \bar{s}_{i2}r_{2l} + \dots + \bar{s}_{if}r_{fl} - s_{il} \quad (22)$$

Aby uzyskać najlepszy wektor  $\mathbf{r}_l$  minimalizuje się odchylenie wektora próbnego (namierzanego) od wektora przewidywanego przyrównując do zera sumę pochodnych kwadratów różnic względem odpowiedniego elementu wektora transformacji. Na przykład, pochodna kwadratu różnicy względem elementu  $r_{lj}$  określona jest następująco

$$\frac{d(\Delta s_{il})^2}{dr_{jl}} = 2\bar{s}_{i1}^2 r_{il} + 2\bar{s}_{i1}\bar{s}_{i2} r_{2l} + \dots + 2\bar{s}_{i1}\bar{s}_{if} r_{fl} - 2\bar{s}_{i1} s_{il} \quad (23)$$

Podobne wyrażenia otrzymuje się dla pozostałych indeksów wierszowych. Po zsumowaniu tych wyrażen i zastosowaniu kryteriów metody najmniejszych kwadratów otrzymujemy ostatecznie następujące równanie

$$\sum_{i=1}^n \frac{d(\Delta s_{il})^2}{dr_l} = 0 = r_{1l} \sum_i \bar{s}_{i1}^2 + r_{2l} \sum_i \bar{s}_{i1} \bar{s}_{i2} + \dots + r_{fl} \sum_i \bar{s}_{i1} \bar{s}_{if} - \sum_i \bar{s}_{i1} \bar{s}_{il} \quad (24)$$

Powtarzając te rachunki otrzymuje się sumy kwadratów pochodnych względem pozostałych elementów wektora transformacji  $\mathbf{r}_l$  (tzn. względem  $r_{2l}, r_{3l}, \dots, r_{jl}$ ). Prowadzi to do następującego układu równań

$$\begin{aligned}
\sum \bar{s}_{i1}s_{il} &= r_{l1}\sum \bar{s}_{i1}^2 + r_{2l}\sum \bar{s}_{i1}\bar{s}_{i2} + \dots + r_{fl}\sum \bar{s}_{i1}\bar{s}_{if} \\
\sum \bar{s}_{i2}s_{il} &= r_{1l}\sum \bar{s}_{i1}\bar{s}_{i2} + r_{2l}\sum \bar{s}_{i2}^2 + \dots + r_{fl}\sum \bar{s}_{i2}\bar{s}_{if} \\
&\vdots \\
\sum \bar{s}_{if}s_{il} &= r_{1l}\sum \bar{s}_{i1}\bar{s}_{if} + r_{2l}\sum \bar{s}_{i2}\bar{s}_{if} + \dots + r_{fl}\sum \bar{s}_{if}^2
\end{aligned} \tag{25}$$

W notacji macierzowej równanie (25) przyjmuje postać

$$\mathbf{a}_i = \mathbf{Br}_i \quad (26)$$

gdzie

$$\mathbf{a}_l \equiv \begin{bmatrix} \sum \bar{s}_{i1} s_{il} \\ \sum \bar{s}_{i2} s_{il} \\ \vdots \\ \sum \bar{s}_{if} s_{il} \end{bmatrix} \quad (27)$$

$$\mathbf{r}_l \equiv \begin{bmatrix} r_{1l} \\ r_{2l} \\ \vdots \\ r_{fl} \end{bmatrix} \quad (28)$$

$$\mathbf{B} \equiv \begin{bmatrix} \sum \bar{s}_{i1}^2 & \sum \bar{s}_{i1} \bar{s}_{i2} & \cdots & \sum \bar{s}_{i1} \bar{s}_{if} \\ \sum \bar{s}_{i1} \bar{s}_{i2} & \sum \bar{s}_{i2}^2 & \cdots & \sum \bar{s}_{i2} \bar{s}_{if} \\ \vdots & \vdots & \ddots & \vdots \\ \sum \bar{s}_{i1} \bar{s}_{if} & \sum \bar{s}_{i2} \bar{s}_{if} & \cdots & \sum \bar{s}_{if}^2 \end{bmatrix} \quad (29)$$

Mnożąc obie strony równania (26) przez  $\mathbf{B}^{-1}$  otrzymujemy

$$\mathbf{r}_l = \mathbf{B}^{-1} \mathbf{a}_l \quad (30)$$

Dokładna analiza równania (29) ujawnia, że

$$\mathbf{B} = \bar{\mathbf{S}}^T \bar{\mathbf{S}} \quad (31)$$

natomiast równania (27), że

$$\mathbf{a}_l = \bar{\mathbf{S}}^T \mathbf{s}_l \quad (32)$$

gdzie  $\mathbf{s}_l$  jest wektorem próbnym (namierzonym). Równanie (30) można zatem zapisać w postaci

$$\mathbf{r}_l = (\bar{\mathbf{S}}^T \bar{\mathbf{S}})^{-1} \bar{\mathbf{S}}^T \mathbf{s}_l \quad (33)$$

Jest to najważniejsze równanie *namiarowej analizy faktorowej*. Przedstawia się je niekiedy w formie bardziej zwartej jako

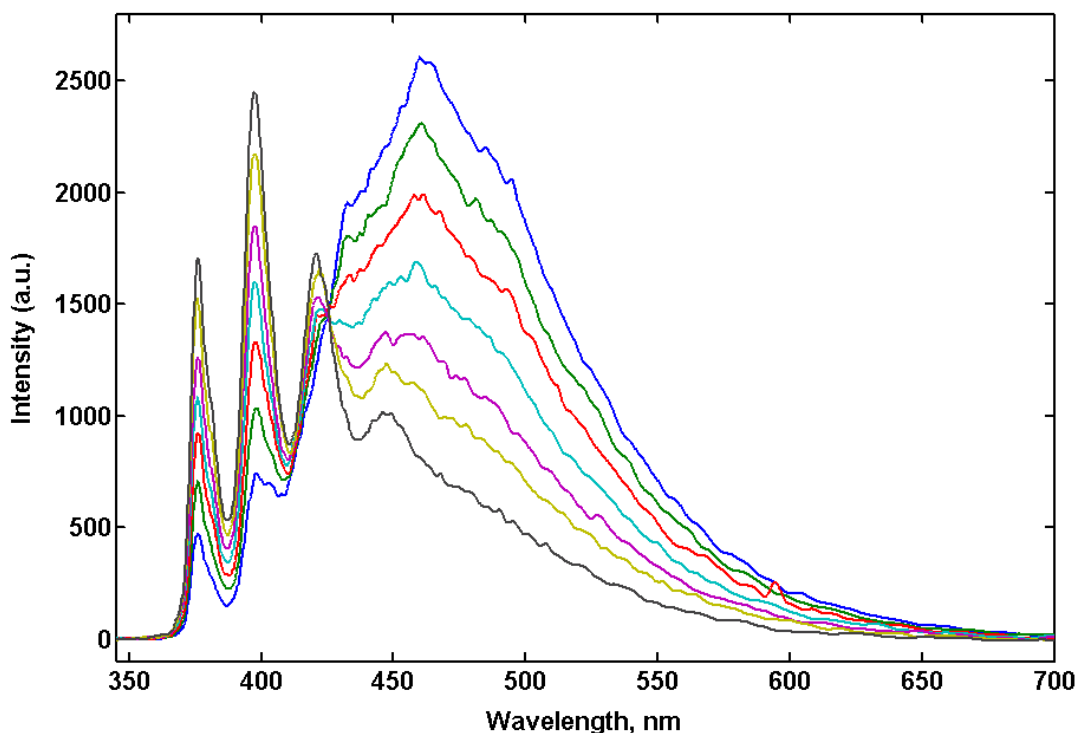
$$\mathbf{r}_l = \bar{\mathbf{S}}^+ \mathbf{s}_l \quad (34)$$

gdzie macierz  $\bar{\mathbf{S}}^+$  jest tzw. „lewą pseudoodwrotnością” macierzy  $\bar{\mathbf{S}}$ . Po wstawieniu równania (34) do równania (20) otrzymujemy ostatecznie

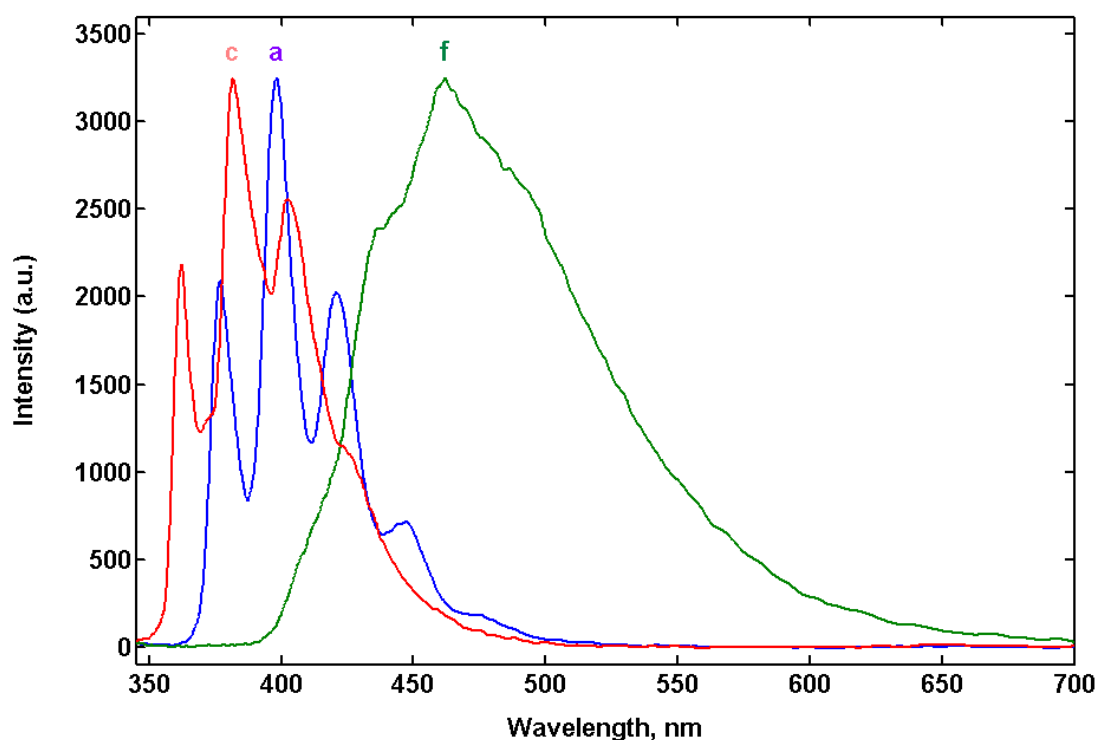
$$\hat{s}_l = \overline{SS}^+ s_l \quad (35)$$

Idea indywidualnego namierzania czynników staje się teraz oczywista. Można ją wyrazić obrazowo w następujący sposób: wektor próbny  $s_l$  zostaje „przefiltrowany” przez macierz  $\overline{SS}^+$ , wynikiem tej transformacji jest wektor przewidywany  $\hat{s}_l$ ; jeśli wektor  $s_l$  jest związany z rzeczywistym czynnikiem, to każdy element wektora  $\hat{s}_l$  będzie równy, w granicach dopuszczalnego błędu, odpowiedniemu elementowi wektora  $s_l$ ; jeśli przeciwnie, nie będzie związany z prawdziwym czynnikiem, to różnice pomiędzy odpowiednimi elementami będą większe niż należałoby oczekiwać. Kryteria sprawdzania tej odpowiedniości opisał Malinowski.<sup>1</sup> Stosując *namiarową analizę czynnikową* należy pamiętać o jednym istotnym szczególe: każda liniowa kombinacja widm odpowiadających rzeczywistym czynnikom, a w szczególności abstrakcyjne profile widmowe zawarte w macierzy  $\overline{S}$ , również z powodzeniem przechodzą test namierzania czynników. A zatem jeśli nie dysponuje się widmami czystych składników, lecz jedynie domniemanymi widmami takich składników otrzymanymi na przykład na drodze ich niekompletnego rozdzielania, to wynik namierzania czynników będzie poprawny, ale nie oznacza to, że namierzone widma odpowiadają widmom czystych składników.

Na Rys. 4 przedstawiono widma fluorescencji dwuskładnikowej mieszaniny antracenu-fluorantenu w *n*-heksanie. Nie ma wśród nich widm czystych składników. Rys. 5 przedstawia widma trzech czystych składników: antracenu, fluorantenu i chryzenu w tym samym rozpuszczalniku. Widm tych użyto jako wektorów próbnych do indywidualnego namierzania czynników. Rezultaty namierzania czynników pokazano na Rys. 6. Rys. 7 przedstawia natomiast przewidywane rzeczywiste profile stężeniowe (zmiany składu) obu namierzonych składników mieszaniny.



Rys. 4      Widma fluorescencji dwuskładnikowej mieszaniny antracen-fluoranten w n-heksanie (7 obiektów).

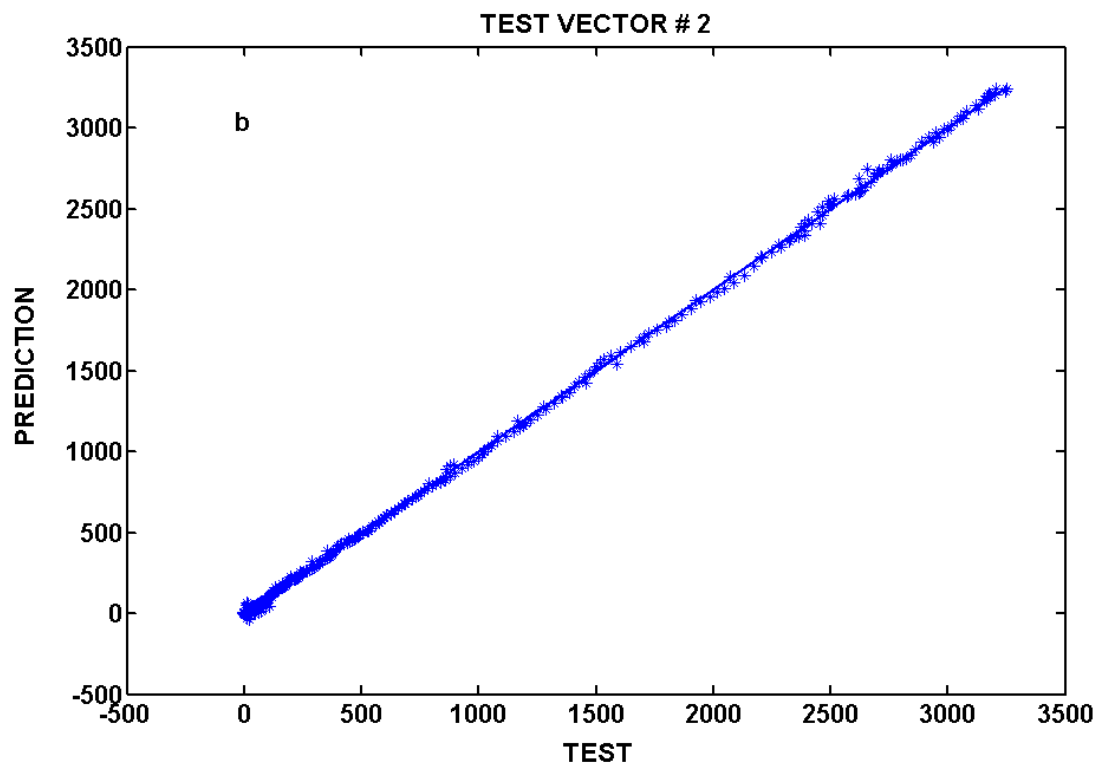
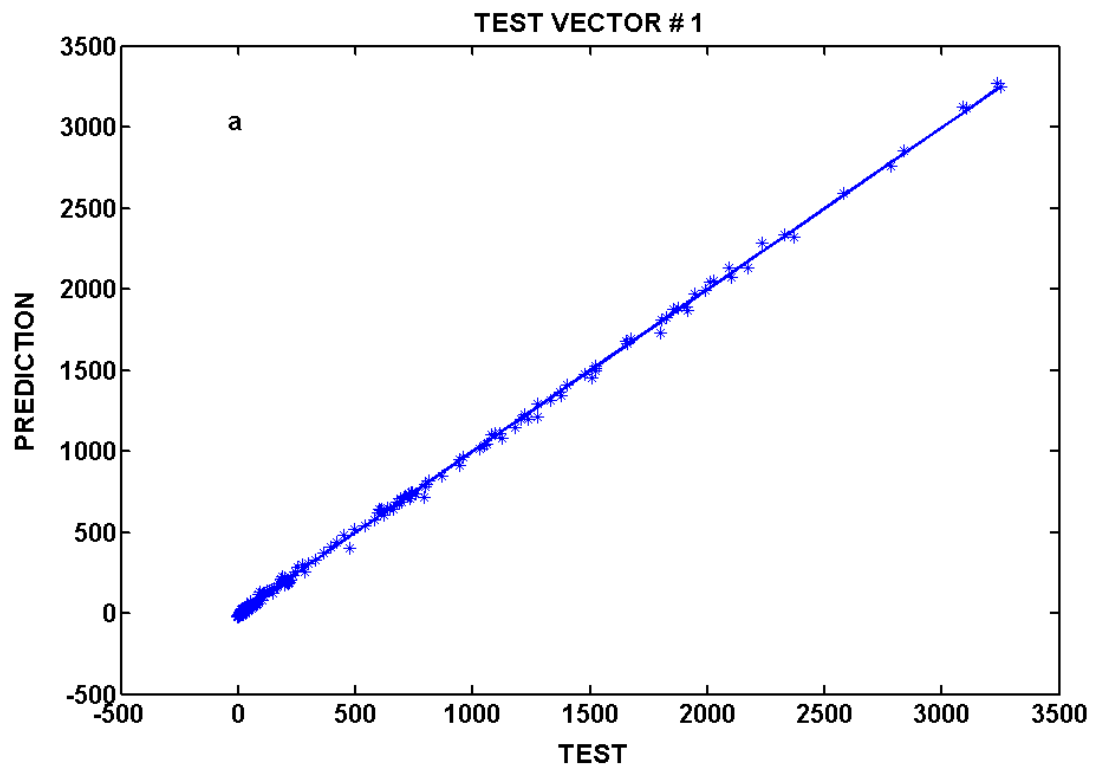


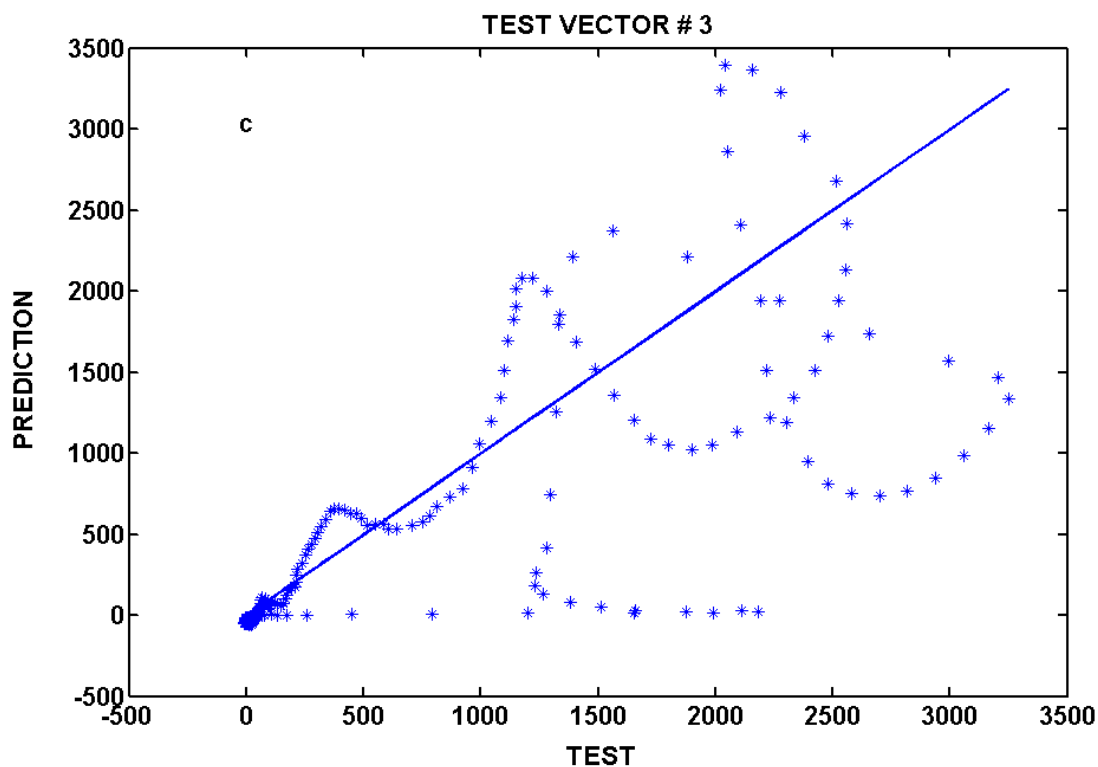
Rys. 5      Unormowane widma fluorescencji czystych substancji w *n*-heksanie: chryzenu (c – linia czerwona), antracenu (a – linia niebieska) i fluorantenu (f – linia zielona).

Tabela I.      Wyniki namierzania czynników w oparciu o program ttest.m z pakietu *TARGET96M* opracowanego przez Malinowskiego.

namiar #	AET	REP	RET	SPOIL	%SL
1	1.7216e+001	1.1292e+001	1.2995e+001	1.1508e+000	4.9602e+001
2	1.8082e+001	1.1212e+001	1.4186e+001	1.2653e+000	4.3106e+001
3	2.5195e+002	6.5744e+000	2.5187e+002	3.8310e+001	2.0854e-005

Znaczenie symboli w Tabeli I jest następujące: AET - pozorny błąd w wektorze próbnym, REP - rzeczywisty błąd wektora przewidywanego, RET - rzeczywisty błąd wektora próbnego, przy czym  $(AET)^2 = (REP)^2 + (RET)^2$ . Funkcja SPOIL zdefiniowana przez Malinowskiego kwalifikuje wektor próbny jako możliwy do przyjęcia jeśli przyjmuje wartości w przedziale 1.0 do 3.0. Podobną rolę odrywa poziom istotności %SL, który weryfikuje pozytywną hipotezę kwalifikacyjną wobec wektora próbnego. Jeśli jest większy niż 10, to wektor próbny jest możliwym do przyjęcia kandydatem na dobry namiar.

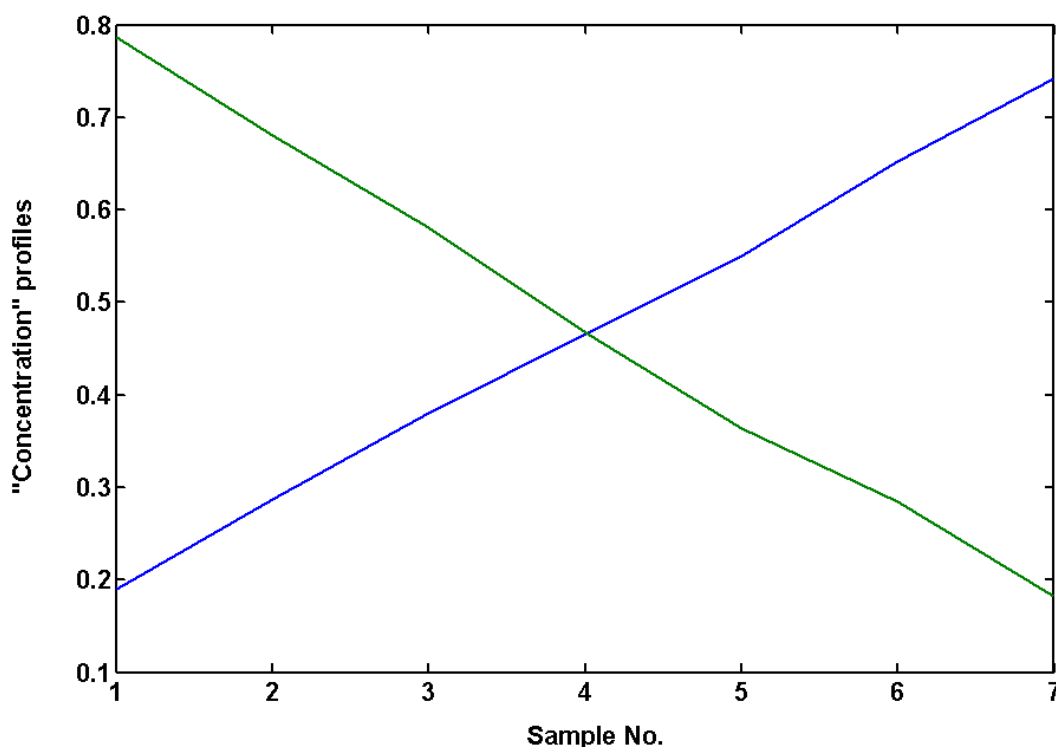




Rys. 6 Korelacja pomiędzy wektorem próbnym i wektorem przewidywanym: udana dla widm fluorescencji antracenu i fluorantenu (dwa pierwsze wykresy - a,b) i nieudana dla widma fluorescencji chryzenu (trzeci wykres - c).

Wyniki namierzania czynników przedstawione w Tabeli I i na Rys. 6 są jednoznaczne: chryzen jako składnik nie wchodzi w skład analizowanej mieszaniny widmowej.

Na Rys. 7 przedstawiono przewidywane profile „stężeniowe” (ściśle biorąc intensywnościowe) otrzymane przy pomocy programu tload.m z pakietu *TARGET96M*. Dla kolejnych 7 mieszanin spektralnych antracenu i fluorantenu względne stosunki „stężeniowe” powinny zmieniać się następująco: 2:8, 3:7, 4:6, 5:5, 6:4, 7:3, 8:2.



Rys. 7 Przewidywane rzeczywiste profile stężeniowe dla fluoryzującej mieszaniny antracen-fluoranten w *n*-heksanie: profil monotonicznie rosnący dla antracenu i profil monotonicznie opadający dla fluorantenu.

### Zagadnienia do opracowania

1. Podstawowe pojęcia algebry macierzowej: definicje wektorów i macierzy, działania na macierzach, rząd macierzy, równania macierzowe, rozwiązanie najmniejszych kwadratów, macierz odwrotna i pseudoodwrotna, wektory własne i wartości własne.
2. Pakiet Matlab 6.0 (lub wyższe wersje) – wczytywanie i zapamiętywanie danych, wykonywanie prostych działań na wektorach i macierzach, sporządzanie wykresów 2D i ich modyfikacja.
3. Analiza głównych składników (PCA): duża i mała macierz kowariancji, macierzowe zagadnienia własne, macierz rotacji, kryteria wyznaczania liczby istotnych czynników.
4. Rozkład macierzy według wartości osobliwych (SVD).
5. Własności transformacji namiarowej (*target transformation*), wektor próbny, wektor przewidywany, kryteria skuteczności namiaru.
6. Uogólniona metoda dodatku wzorca i jej wykorzystanie w spektroskopii absorpcyjnej do oznaczania ilości poszczególnych składników w próbce.

## Część doświadczalna

### Aparatura

3 kolby miarowe na 100 ml, 1 kolba miarowa na 50 ml, 15 kolbek miarowych na 10 ml, mikropipeta, naczynka wagowe, zlewki, kuwety kwarcowe do pomiarów widm absorpcyjnych, spektrofotometr Helios  $\beta$ , komputer z programem MATLAB wersja 6.0 lub wyższe, z oprogramowaniem do analizy faktorowej (pakiet *TARGET 96 M*).

### Odczynniki

aspiryna, salicylamid, kofeina, metanol.

### Wykonanie ćwiczenia

1. Dokładnie odważone naważki zawierające ok. 20-25 mg aspiryny, salicylamidu oraz kofeiny rozpuszcza się w ok. 50 ml metanolu, a następnie rozcieńcza do kreski tym samym rozpuszczalnikiem w kolbach miarowych na 100 ml.
2. Z otrzymanych w pkt. 1 jednoskładnikowych roztworów podstawowych pobiera się mikropipetą po 0.50 ml tych roztworów i przenosi do trzech kolbek miarowych na 10 ml, a następnie uzupełnia do kreski metanolem. Dla rozcieńczonych w ten sposób roztworów jednoskładnikowych rejestruje się widma absorpcji w zakresie 220-370 nm stosując jako odnośnik metanol.
3. Z tych samych roztworów podstawowych sporządza się 6 roztworów trójskładnikowych. W tym celu pobiera się mikropipetą odpowiednie ilości roztworów podstawowych według załączonej przez asystenta tabeli i przenosi do kolbek miarowych na 10 ml. Kolbki uzupełnia się do kreski miarowej metanolem. Dla wszystkich 6 roztworów rejestruje się widma absorpcji (macierz **X**) w zakresie 220-370 nm stosując jako odnośnik metanol.

Tabela typowego składu roztworów kalibracyjnych.

Nr roztworu	Objętości roztworów podstawowych [ml]		
	Salicylamid	Aspiryna	Kofeina
1	0.32	0.40	0.32
2	0	0.72	0.40
3	0.24	1.20	0
4	0.40	0.48	0.08
5	0.48	0	0.24
6	0.16	0.24	0.48

4. Asystent lub wybrany student sporządza „nieznany” roztwór trójskładnikowy. W tym celu przenosi do kolby miarowej na 50 ml określone ilości mililitrów podstawowych roztworów jednoskładnikowych, np. 0.50 ml roztworu salicylamidu, 1.00 ml roztworu aspiryny i 0.50 ml roztworu kofeiny. Kolbę uzupełnia się do kreski miarowej rozpuszczalnikiem.



5. „Nieznany” roztwór trójskładnikowy przenosi się w ilościach po 5.00 ml do 6 kolbek miarowych na 10 ml. Kolbki te uzupełnia się do kreski miarowej 6 roztworami kalibracyjnymi opisanymi w pkt. 3 dodając do każdej kolbki jeden taki roztwór. Dla otrzymanych roztworów rejestruje się widma absorpcyjne (macierz  $\mathbf{Y}$ ) w zakresie 220-370 nm używając metanol jako odnośnika.

## Opracowanie wyników

1. Narysować wzory strukturalne analizowanych środków przeciwbólowych i odpowiadające im widma absorpcyjne.
2. Przedstawić na rysunku widma absorpcyjne roztworów kalibracyjnych (macierz  $\mathbf{X}$ ).
3. Stosując program **pcomp.m** (z pakietu *TARGET 96 M*) przeprowadzić analizę ilości istotnych czynników w macierzy  $\mathbf{X}$ . Wyniki zebrać w tabeli i zilustrować odpowiednim rysunkiem.
4. Dla wyznaczonej liczby istotnych czynników przeprowadzić rozkład macierzy kalibracyjnej  $\mathbf{X}$  na iloczyn macierzy abstrakcyjnych widm  $\bar{\mathbf{S}}$  i abstrakcyjnych profili stężeniowych  $\bar{\mathbf{C}}$ :

$$\mathbf{X} \approx \bar{\mathbf{X}} = \bar{\mathbf{S}}\bar{\mathbf{C}}^T$$

Do przeprowadzenia tego rozkładu wykorzystać procedurę matlabowską **svd.m** w jej wersji „ekonomicznej”. Wyniki zilustrować na odpowiednich rysunkach.

5. Na podstawie znajomości rzeczywistych profili stężeniowych roztworów kalibracyjnych  $\mathbf{C}$  odzyskać z macierzy  $\bar{\mathbf{X}}$  przewidywane rzeczywiste widma czystych składników mieszanin widmowych, posługując się następującym przekształceniem (wykonywalnym w programie Matlab):

$$\hat{\mathbf{S}} = \bar{\mathbf{X}}\mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}$$

Odzyskane w ten sposób widma absorpcyjne aspiryny, salicylamidu i kofeiny przedstawić na rysunku, a po unormowaniu do powierzchni jednostkowej porównać z podobnie unormowanymi zmierzonymi widmami czystych składników.

6. W stosunku do roztworów kalibracyjnych zastosować procedurę namiarowej analizy czynnikowej (program **ttest.m** z pakietu *TARGET 96 M*), używając jako namiarów widm czystych składników. Wyniki namiarowej analizy czynnikowej zestawzić w tabeli. Na rysunkach przedstawić korelacje między wektorem namierzonym a wektorem przewidywanym.
7. Punkty 2, 3 i 4 wykonać dla macierzy  $\mathbf{Y}$  (widma próbek uzyskanych przez dodanie określonej ilości mieszaniny trójskładnikowej do poszczególnych roztworów kalibracyjnych).
8. Korzystając z wyznaczonych w pkt. 5 przewidywanych widm rzeczywistych czystych składników  $\hat{\mathbf{S}}$  odzyskać z macierzy  $\bar{\mathbf{Y}}$  przewidywane rzeczywiste profile stężeniowe  $\hat{\mathbf{C}}$ , w oparciu o następujące przekształcenie (wykonywalne w programie Matlab):

$$\hat{\mathbf{C}} = \left[ (\hat{\mathbf{S}}^T \hat{\mathbf{S}})^{-1} \hat{\mathbf{S}}^T \bar{\mathbf{Y}} \right]^T$$

Przewidywane rzeczywiste profile stężeniowe uzyskane w uogólnionej metodzie dodatków wzorca zestawień w tabeli:

Nr próbki	Objętości roztworów podstawowych [ml]		
	Salicylamid	Aspiryna	Kofeina
1			
2			
3			
4			
5			
6			

Ponieważ roztwory próbek otrzymano mieszając ze sobą roztwory kalibracyjne i roztwór trójskładnikowy o „nieznanym” składzie zawsze w stosunku 1:1, tzn. 5 ml roztworu kalibracyjnego i 5 ml roztworu „nieznanej” próbki, dlatego stężenie salicylamidu, aspiryny i kofeiny w mieszaninie próbki trójskładnikowej o „nieznanym” składzie można obliczyć używając wzoru:  $c_x = 2 \cdot (c - 0.5 \cdot c_0) \cdot 5$ , ( $c_x$  - stężenie danego składnika w „nieznanej” próbce,  $c$  - stężenie danego składnika w roztworze powstałym z połączenia roztworu kalibracyjnego i „nieznanej” próbki,  $c_0$  - stężenie tegoż składnika w wyjściowym roztworze kalibracyjnym). Liczba 5 w powyższym wzorze wynika z przeliczenia objętości kolbek (50 ml/10 ml). Po uśrednieniu po wszystkich roztworach otrzymuje się stężenia składników w badanej próbce wyrażone w ml roztworów podstawowych, które można teraz przeliczyć na stężenia molowe.

9. Przeprowadzić tak jak w pkt. 6 namiarową analizę faktorową dla macierzy  $\mathbf{Y}$  wykorzystując:

- przewidywane rzeczywiste widma czystych składników  $\hat{\mathbf{S}}$  otrzymane w pkt. 5.
- zarejestrowane widma czystych składników  $\mathbf{S}$ .

Dla obu przypadków wyznaczyć przewidywane rzeczywiste profile stężeniowe w oparciu o program **tload.m** (z pakietu *TARGET 96 M*) i przedstawić w tabelach jak w pkt. 8. Wyznaczyć w analogiczny sposób jak w pkt. 8 średnie ilości składników wyrażone w ml roztworów podstawowych i przeliczyć na stężenia molowe.

## Literatura podstawowa

- [1] E. R. Malinowski, „*Factor Analysis in Chemistry*“, Wyd. 3, John Wiley & Sons, New York Chichester Brisbane Toronto Singapore, 2002 str. str. 32-59, 65-69.
- [2] S. Brandt, „*Analiza Danych*“, Wyd. 1, PWN, Warszawa, 1998, str. str. 506-513, 535-541, 553-559.
- [3] M. J. Charles, N. W. Martin, H. Z. Msimanga, *J. Chem. Education*, **74**, str. str. 1114-1117, (1997) – „*Simultaneous Determination of Aspirin, Salicylamide, and Caffeine in Pain Relievers by Target Factor Analysis*„

## Literatura uzupełniająca

- [4] M. Maeder, *Anal. Chem.*, **59**, 527 (1987).
- [5] J. Toft, *Chem. Intel. Lab. Syst.*, **29**, 189 (1995).
- [6] B. M. Wise, N. B. Gallagher, „ *PLS\_Toolbox 3.0 for use with MATLAB<sup>TM</sup>* “ , Eigenvector Research, Inc., 2003.