



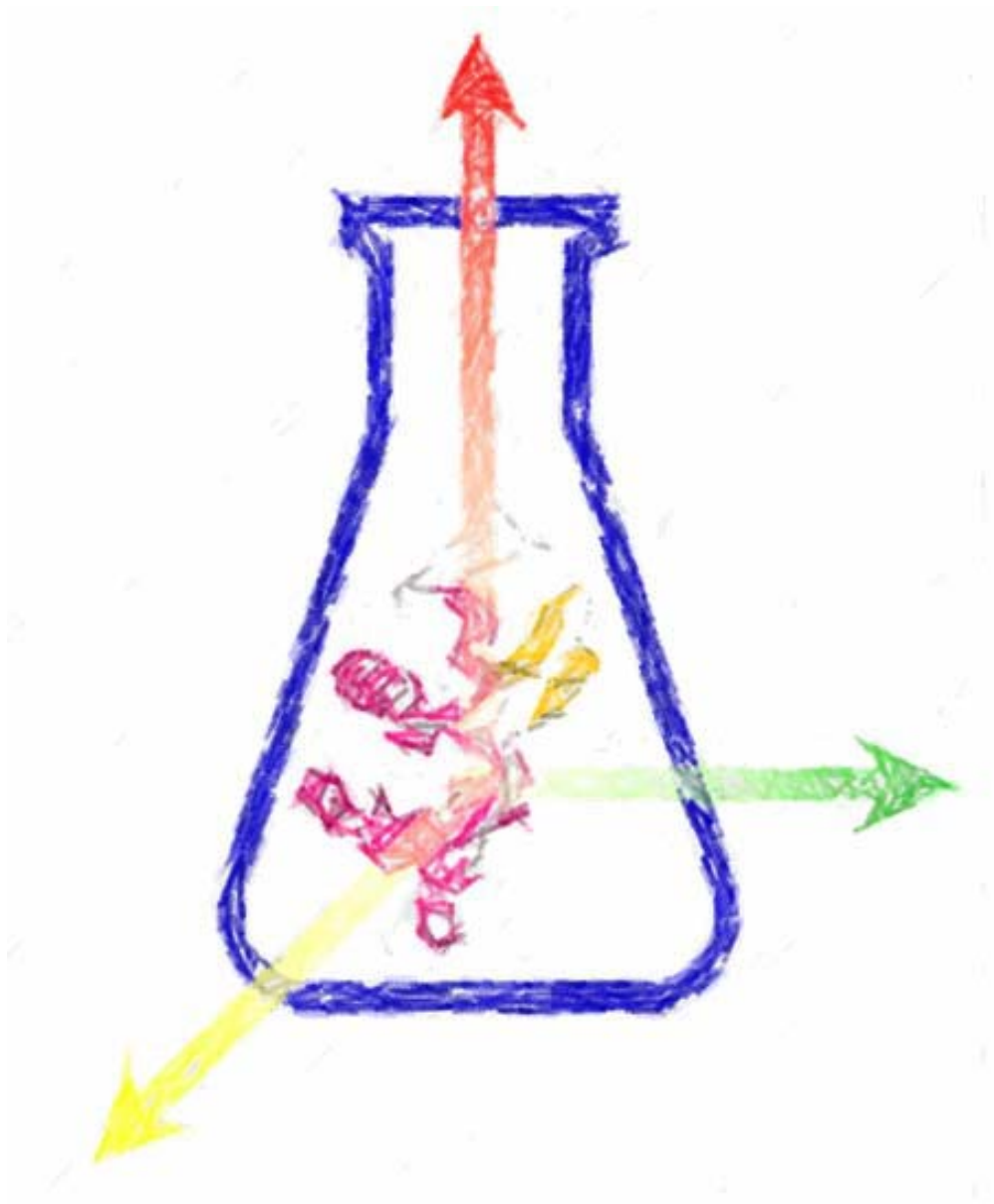
UNIVERSITAT DE BARCELONA



**Nous desenvolupaments,
aplicacions bioanalítiques i
validació de mètodes de
resolució multivariant**

Joaquim Jaumot Soler

**Tesi Doctoral
20 de juny de 2006**

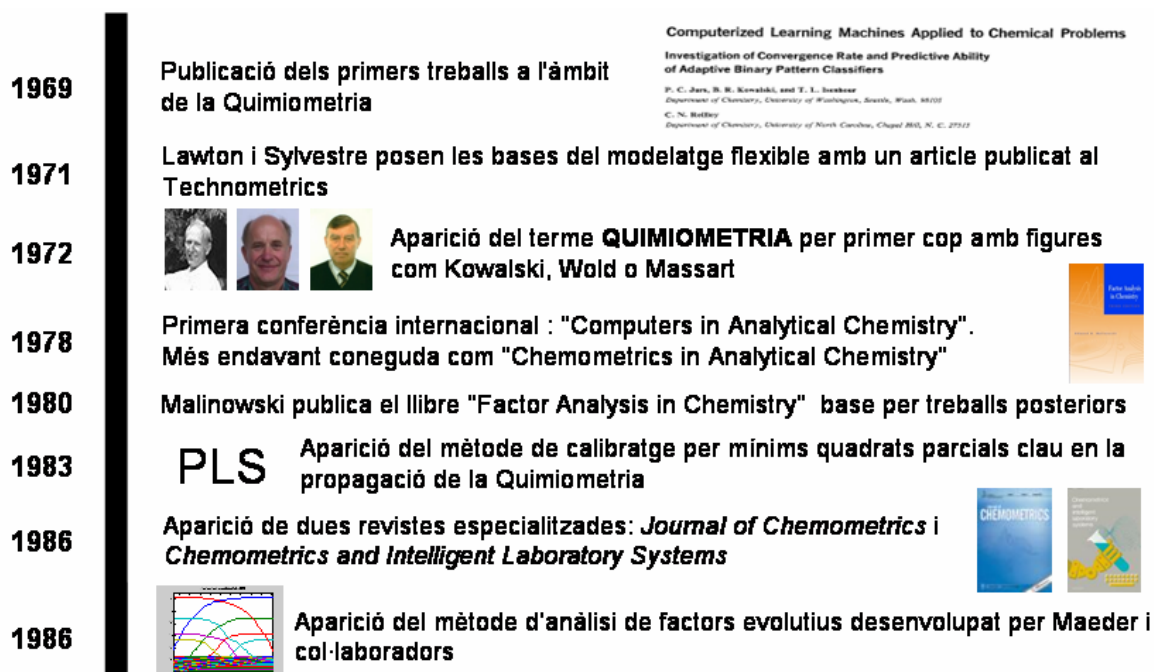


Capítol 1

Mètodes quimiomètrics

Les arrels de la Quimiometria s'han de buscar a l'any 1969 quan Kowalski, Jurs i Isenhour van publicar una sèrie de treballs sobre l'aplicació de màquines d'aprenentatge lineal per classificar espectres de masses de baixa resolució [1-7]. Seguint aquestes primeres iniciatives, altres grups van començar a treballar en problemes similars de classificació, liderats per científics com Wold amb algorismes de classificació com el SIMCA [8-11], o Massart en el camp de l'optimització de mètodes cromatogràfics [12, 13]. No va ser fins l'any 1972 quan va aparèixer per primer cop el nom de Quimiometria en un treball de Svante Wold en un treball publicat en la revista sueca *Kemisk Tidskrift* [14]. Alhora, Kowalski va fer servir el nom de "*Laboratory of Chemometrics*" pel seu grup de recerca. A partir d'aquí, la irrupció de la informàtica en els laboratoris químics va permetre l'extensió i la consolidació arreu de la Quimiometria.

A la Figura 1.1. es mostra un esquema d'aquest procés evolutiu de la Quimiometria des de l'any 1969 fins l'actualitat amb els fets més destacats en relació a aquesta Tesi Doctoral.



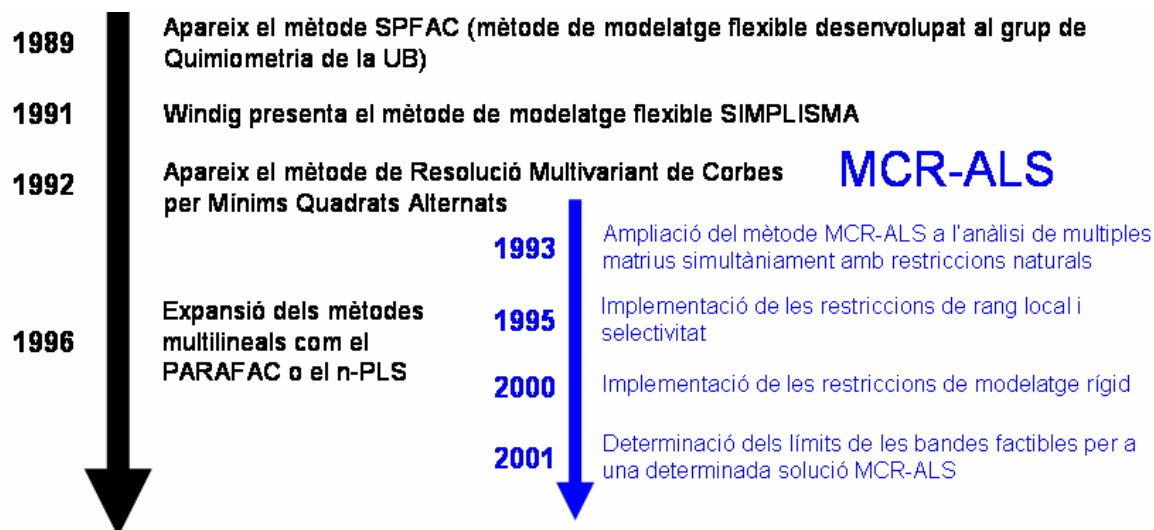


Figura 1.1. Breu història de la Quimiometria amb especial atenció als mètodes de modelització flexible. Per a més informació veure les referències següents: [14-27].

La *International Chemometrics Society* defineix la Quimiometria com:

“La Quimiometria és la ciència que relaciona les mesures fetes en un sistema o procés químic amb el seu estat mitjançant l’aplicació de mètodes matemàtics i estadístics”.

Donat el gran nombre de camps on s’aplica la Quimiometria s’han fet altres definicions més concretes del concepte Quimiometria com la que es pot trobar en el *Handbook of Chemometrics and Qualimetrics* [16]:

“La Quimiometria és la disciplina de la química que utilitza les matemàtiques, l’estadística i la lògica formal per a:

- el disseny i la selecció de condicions experimentals òptimes,
- l’extracció del màxim d’informació química rellevant a partir de l’anàlisi de dades d’origen químic, i
- l’obtenció de coneixement de processos químics.”

Aquesta darrera definició permet explicar la aplicació de la Quimiometria en camps tan diversos com:

- el disseny i l’optimització de paràmetres experimentals,
- el calibratge analític univariant o multivariant,

- el tractament de senyals,
- l'obtenció d'informació a partir de dades químiques utilitzant procediments estadístics,
- l'anàlisi i resolució de compostos i espècies químiques i dels seus senyals a partir de les seves mesclures,
- la classificació de mostres per a reconeixement dels patrons de les dades,
- la predicció de l'activitat o les propietats de compostos mitjançant l'estudi de les relacions d'estructura-activitat (QSAR, *Quantitative Structure-Activity Relationships*) o les relacions d'estructura-propietats (QSPR, *Quantitative Structure-Properties Relationships*)

Per tant, es pot dir que la Quimiometria és clarament una ciència que permet molts àmbits d'aplicació, tal i com es mostra a la Figura 1.2.

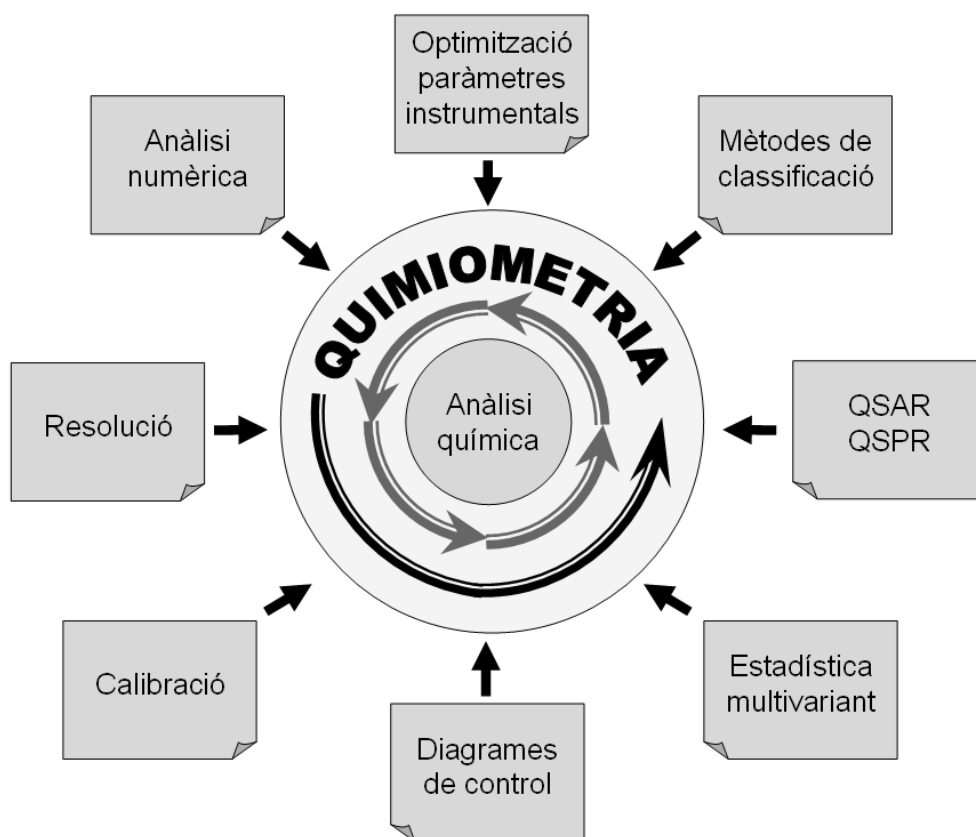


Figura 1.2. Representació esquemàtica de diferents aplicacions de la Quimiometria relacionades amb l'anàlisi química.

1.1. Mètodes multivariants d'anàlisi de dades

Les tècniques instrumentals actuals permeten l'obtenció de dades multivariants. Per exemple, el conjunt de mesures adquirides per un espectrofotòmetre en enregistrar l'espectre d'una mostra es pot ordenar en forma de vector \mathbf{s} amb n elements corresponents al nombre de longituds d'ona mesurades. Si s'agrupen tots els espectres (vectors) mesurats durant un experiment s'obté una taula o matriu de dades $\mathbf{D}(m,n)$, on m és el nombre d'espectres i n el nombre de longituds d'ona de cada espectre [28]. Aquesta matriu \mathbf{D} contindrà informació relativa al procés seguit, tal i com es representa a la Figura 1.3.

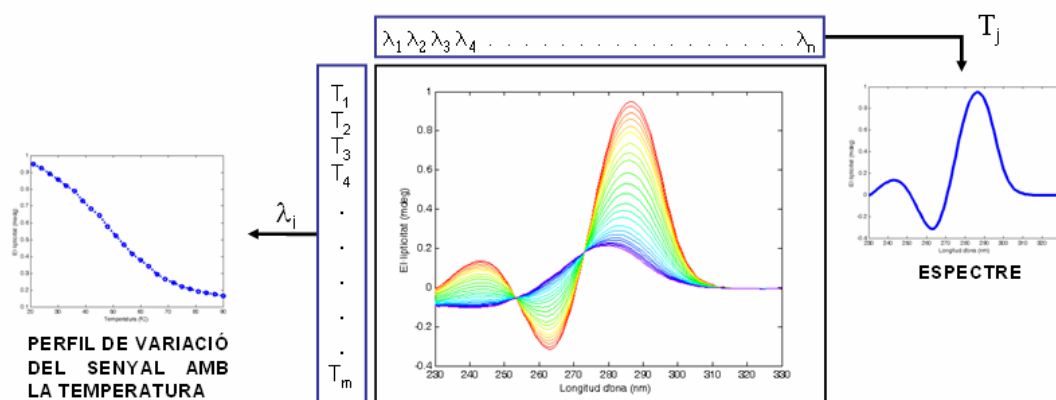


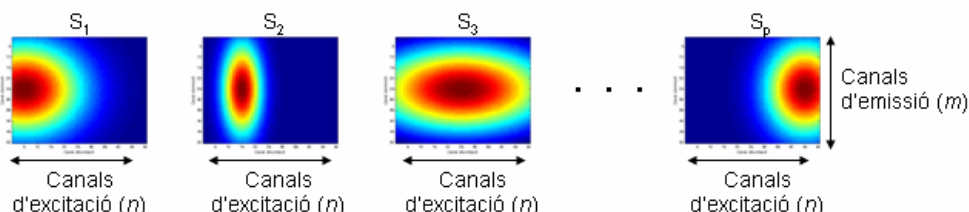
Figura 1.3. Ordenació de les matrius de dades on es mostra a la figura central una representació dels espectres enregistrats a les m temperatures de treball. A l'esquerra es mostra el perfil de variació del senyal amb la temperatura a una determinada longitud d'ona λ_i i, a la dreta, l'espectre enregistrat a una determinada temperatura T_j .

En aquesta matriu de dades \mathbf{D} , les columnes descriuen la variació del senyal amb la longitud d'ona i les files descriuen els espectres a cadascun dels valors de la variable externa (T , en aquest cas).

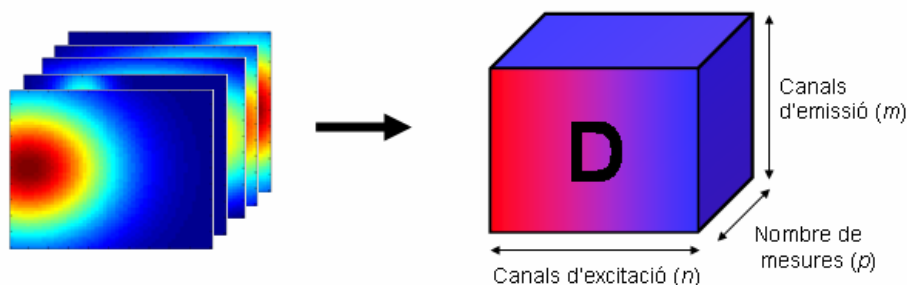
La incorporació, en els últims anys, de la informàtica als laboratoris químics ha permès l'obtenció i l'anàlisi de conjunts de dades encara més complexos que el que s'acaba de descriure [29-31]. Així, algunes tècniques instrumentals permeten l'adquisició a cada punt discret de matrius de dades \mathbf{S}_i ($i = 1, 2, \dots$, fins a p) de dimensions m (nombre de canals en una dimensió) per n (nombre de canals en una altra dimensió). Al tenir p matrius de dades es formarà una estructura en forma de cub \mathbf{S} de dimensions $m \times n \times p$ (Figura 1.4.). Exemples d'aquestes tècniques instrumentals són

la fluorescència d'excitació-emissió [32] o les tècniques multidimensionals de RMN [33].

Adquisició d'espectres excitació-emissió de fluorescència



Formació del cub de dades



Desplegament del cub de dades per donar una matriu augmentada

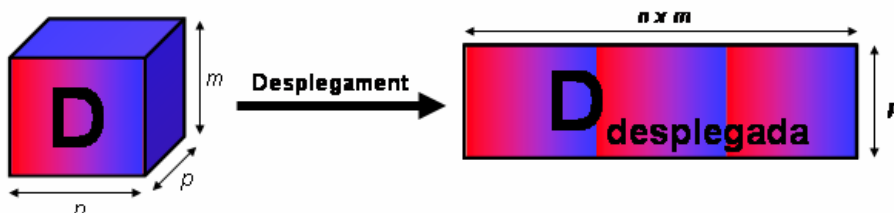


Figura 1.4. Adquisició de dades multidimensionals i formació d'estructures d'ordre superior.

Per tal d'analitzar aquests cubs de dades s'han proposat diverses estratègies i, en alguns casos com el que es mostra a la Figura 1.4., es proposa el desplegament del cub de dades per formar una matriu de dades augmentada similar a **D** [34].

En molt casos, la resposta instrumental de les tècniques espectroscòpiques segueix un model bilineal anàleg al de la llei de Lambert-Beer generalitzada per a múltiples longituds d'ona en l'anàlisi simultània dels espectres corresponents a un procés, reacció o sistema químic [28]:

$$d_{ij} = \sum_{k=1}^n (c_{ik} s_{kj} + e_{ij})$$

Equació 1.1.

on d_{ij} és el senyal individual mesurat experimentalment a la longitud d'ona j en l'espectre i . Quan el model lineal es compleix, d_{ij} és la suma del senyal de les diferents espècies k presents en l'espectre i a la longitud d'ona j .

c_{ik} representa la concentració de l'espècie k a l'espectre i ,

s_{kj} representa el senyal unitari de l'espècie k a la longitud d'ona j ,

e_{ij} és l'error residual

En forma matricial aquest model bilineal s'escriu com:

$$\mathbf{D}_{(m,n)} = \mathbf{C}_{(m,nc)} \mathbf{S}_{(nc,n)}^T + \mathbf{E}_{(m,n)} \quad \text{Equació 1.2.}$$

on \mathbf{D} és la matriu de dades experimentals, que conté els espectres de la mostra o procés enregistrats a valors diferents d'una variable experimental (temperatura, pH, ...).

\mathbf{C} és la matriu de concentracions de les diferents espècies presents en el sistema químic. Cada fila de la matriu \mathbf{C} conté les concentracions de totes les espècies durant l'experiment o un conjunt de mostres. Per tant, cada columna de \mathbf{C} conté l'evolució de la concentració d'una determinada espècie. És a dir, el perfil de concentracions de cada espècie a partir del qual es pot deduir el mecanisme de la reacció o els paràmetres termodinàmics o cinètics associats.

\mathbf{S}^T és la matriu d'espectres purs corresponent a cadascuna de les espècies presents en el sistema químic. A partir d'aquests espectres purs es podran determinar les característiques pròpies d'aquestes espècies.

\mathbf{E} és la matriu dels residuals que inclou la variació no explicada pel model bilineal de l'Equació 1.2. Si el sistema està ben descrit per a aquest model, \mathbf{E} inclourà únicament l'error experimental.

m és el nombre d'espectres analitzats,

n és el nombre de longituds d'ona de cada espectre.

nc és el nombre d'espècies o components que participen en el sistema químic i que tenen resposta espectroscòpica.

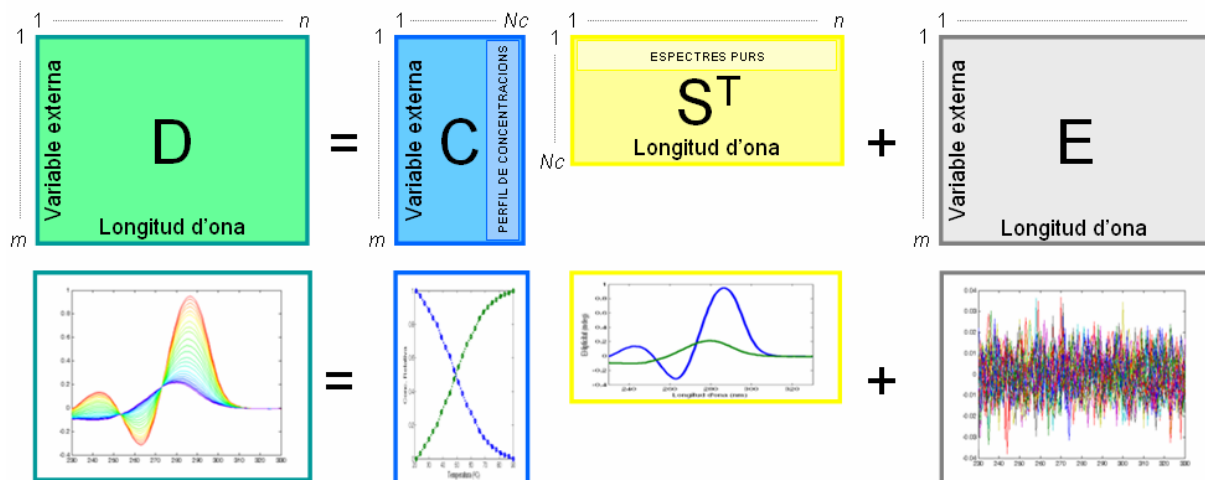


Figura 1.5. Representació en forma matricial de les dades experimentals.

Gràficament, l'Equació 1.2 es mostra a la Figura 1.5., on la matriu **E** s'ha representat enfront de la longitud d'ona.

A partir de les mesures experimentals contingudes a la matriu **D**, es vol extreure informació amb significat químic. El principal objectiu dels mètodes de resolució multivariant és el càlcul de les matrius **C** i **S^T** a partir de l'anàlisi de la matriu experimental **D** [16]. És a dir, es volen resoldre les contribucions de cada un dels *nc* components que participen en el sistema químic estudiat, de manera que s'obtinguin els seus perfils de concentració (**C**) i els seus espectres purs (**S^T**).

Hi ha dos grups de mètodes de resolució multivariant que es poden fer servir per resoldre l'Equació 1.2. En primer lloc, hi ha els mètodes que no fan servir cap model químic explícit en el càlcul i que es coneixen com a mètodes de modelització flexible o *soft-modelling* [15, 17, 35]. En segon lloc, hi ha els mètodes de modelització rígida o *hard-modelling* que són aquells mètodes que fan servir un model químic, definit per una llei quimicofísica (per exemple, la llei d'acció de masses o un mecanisme cinètic) i una estequiometria que relaciona les concentracions de les espècies en el procés [36, 37]. Recentment, han aparegut els mètodes coneguts com de modelització híbrida rígida-flexible o *hard-soft modelling* que conjunten els avantatges d'aquestes dues aproximacions [38-41].

1.2. Modelització flexible

Els mètodes de modelització flexible descomponen la matriu de dades \mathbf{D} en les contribucions en els perfils de concentració \mathbf{C} i espectres \mathbf{S}^T , basant-se únicament en l'aplicació de restriccions naturals com, per exemple, la no-negativitat de concentracions i espectres de les espècies del sistema.

1.2.1. Consideracions bàsiques: nombre de components i estimacions inicials

El concepte de modelització flexible va aparèixer per primera vegada a la literatura en el treball pioner de Lawton i Sylvestre [15]. En aquest treball s'estudiava una matriu de dades espectrals que s'havia obtingut per mescla de dos components. A partir d'aquest treball inicial i de la seva extensió a tres o més components, han aparegut diversos mètodes de resolució [42-44]. Alguns d'aquests mètodes són el *Iterative Target Transformation Factor Analysis* (ITTFA) [45, 46] amb *Needle search* [47] o el mètode *Heuristic Evolving Latent Projections* (HELP) [48, 49]. Aquests mètodes no necessiten la postulació d'un model químic determinat però poden utilitzar informació prèvia del sistema incloent restriccions en l'algoritme d'optimització les quals poden guiar aquest procés cap a la solució òptima.

En aquesta Tesi Doctoral s'ha fet servir el mètode de Resolució Multivariant de Corbes per Mínims Quadrats Alternats (*Multivariate Curve Resolution by Alternating Least Squares* [24, 25, 50-52], MCR-ALS).

La descomposició de la matriu \mathbf{D} mitjançant el mètode MCR-ALS es produeix en tres etapes:

A) Determinació del nombre d'espècies químiques presents en el sistema. S'utilitza l'anàlisi per components principals (*Principal Component Analysis* [16, 53-55], PCA), l'anàlisi de factors evolutius (*Evolving Factor Analysis* [19, 56, 57], EFA) i els mètodes de detecció de variables pures (com el *SIMPLe-to-use Interactive Self-modelling Mixture Analysis* [21, 58], SIMPLISMA).

B) Creació d'estimacions inicials de:

- perfils de concentració (\mathbf{C}_{inic}) obtinguts a partir de EFA;
- espectres ($\mathbf{S}_{\text{inic}}^T$), mitjançant un mètode de selecció de les variables o objectes més purs de la matriu \mathbf{D} (com el que es proposa en el mètode SIMPLISMA).

C) Càlcul de les matrius \mathbf{C} i \mathbf{S}^T mitjançant l'aplicació d'un procés d'optimització iteratiu de mínims quadrats alternats sota restriccions (*Alternating Least Squares* [24, 25, 50], ALS), que s'inicia amb les estimacions inicials obtingudes en el pas anterior.

A continuació, es fa una descripció detallada d'aquestes etapes.

A) Estimació del nombre de components

Quan es planteja la resolució d'un determinat sistema químic podem trobar-nos amb dues situacions: tenir un coneixement previ sobre el nombre possible de components o no tenir aquest coneixement previ. En aquest últim cas, s'ha de recórrer a mètodes que ens indiquin, a partir de la informació experimental disponible, quin pot ésser aquest nombre de components.

L'anàlisi per components principals (PCA) és una tècnica d'anàlisi multivariant basada en un model bilineal de les dades experimentals \mathbf{D} que permet explicar la variació experimental observada amb un nombre mínim de components (factors) ortogonals [17, 54, 55].

L'anàlisi per components principals descompon matemàticament la matriu de dades experimentals \mathbf{D} en el producte de dues matrius \mathbf{U} i \mathbf{V}^T .

$$\mathbf{D} = \mathbf{U} \mathbf{V}^T + \mathbf{E} \quad \text{Equació 1.3.}$$

on \mathbf{U} i \mathbf{V}^T s'anomenen, respectivament, la matriu de mapa de mostres o *scores* i la matriu de variables o *loadings* [17], i \mathbf{E} és la matriu de residuals no explicats pel model. Aquesta descomposició matemàtica es realitza sota restriccions de màxima variància explicada i ortogonalitat. A la pràctica, aquesta descomposició coincideix amb el càlcul dels vectors i valors propis de la matriu \mathbf{D} . En l'anàlisi per components principals, els

factors s'ordenen en funció de la variància explicada en ordre descendent. Per tant, el primer component està associat amb el valor propi més gran i explica el percentatge més gran de variància de la matriu **D**. El segon component és ortogonal al primer i explica el percentatge més gran de variància no descrita pel primer component i, així successivament. El nombre màxim de components calculats és *s*, on *s* és el nombre de files o de columnes de la matriu **D** (el que sigui de dimensió més petita), els quals expliquen tota la variància de les dades, incloent-hi també la deguda a l'error experimental.

De forma anàloga es pot realitzar l'anomenada descomposició en valors singulars (*Singular Value Decomposition* [53], SVD). Amb aquest procediment, la matriu de dades es descompon en:

$$\mathbf{D}_{(m,n)} = \mathbf{U}_{(m,nc)} \mathbf{S}_{(nc,nc)} \mathbf{V}_{(nc,n)}^T + \mathbf{E}_{(m,n)} \quad \text{Equació 1.4.}$$

On **S** és una matriu diagonal en la qual els elements de la diagonal tenen els valors singulars que són l'arrel quadrada dels valors propis. Gràficament aquesta descomposició es mostra a la Figura 1.6.

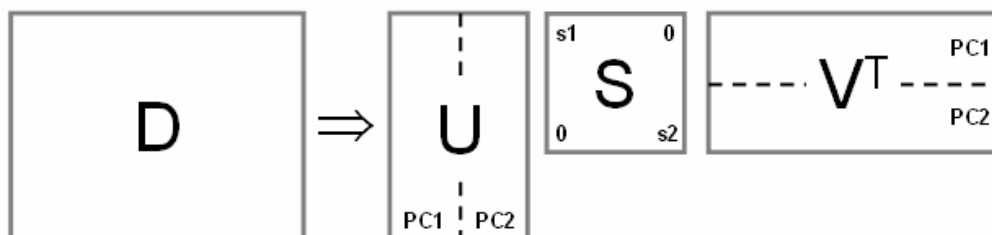


Figura 1.6. Esquema gràfic de la descomposició en valors singulars d'una matriu.

En aquest cas es consideren 2 valors singulars.

La forma més fàcil i intuïtiva de determinar el nombre de components necessaris per explicar la variació química (no el soroll experimental) de la matriu **D** és la representació gràfica dels valors singulars com en el cas de la Figura 1.7. Aquest nombre de components és el que es coneix com rang químic, que es pot definir com el rang matemàtic en absència d'error experimental. La magnitud dels valors singulars relacionats amb les causes químiques de la variació solen ser clarament superiors a la dels valors que només expliquen soroll experimental. Cal tenir en compte que si es

selecciona un nombre de components més gran que el real es corre el perill de sobreajustar les dades, és a dir, introduir soroll experimental en els perfils de concentració i espectrals resolts.

A la pràctica, la determinació del nombre de components d'un sistema a partir de la representació gràfica dels valors singulars pot no ser evident. En un sistema on no es coneixin, a priori, el nombre de components, es poden donar diverses possibilitats potencialment acceptables. Això és degut a la impossibilitat d'afirmar si un valor

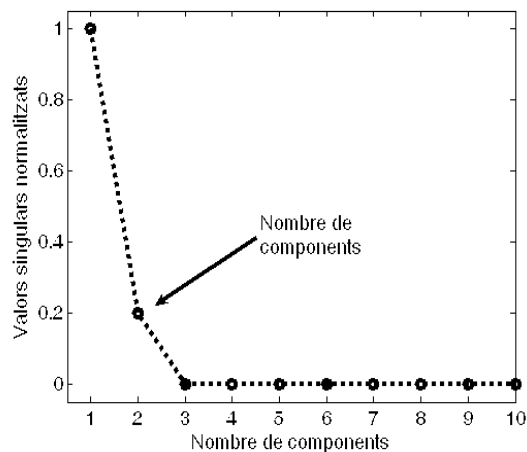


Figura 1.7. Representació gràfica de la determinació del nombre de components.

singular explica variacions produïdes per factors químics o si explica variacions associades a l'error experimental. En aquests casos, és recomanable efectuar el procés de resolució per a diferent nombre de components, i adoptar com a solució final aquella que proporcioni un millor ajust de les dades i, al mateix temps, proporcioni perfils de concentració i espectres acceptables i amb significat químic. Altres mètodes proposats per a la determinació del nombre de components [17], no aconsegueixen en general assegurar quin és el nombre de components adequat de forma unívoca i és necessari repetir tota l'anàlisi per a diferent nombre de components.

B) Estimacions inicials de perfils de concentració o espectrals

Un cop es coneix quin és el nombre més probable de factors que es poden relacionar amb espècies químiques o conformacions (*nc*) presents en el sistema, es passa a buscar una estimació inicial dels perfils de concentració o dels espectres purs d'aquestes espècies.

En alguns casos, el càlcul d'aquestes estimacions inicials es pot fer a partir de perfils de concentració o d'espectres coneguts prèviament. Si no disposem d'aquesta informació, existeixen mètodes quimiomètrics que permeten obtenir estimacions dels perfils de concentració a partir de l'anàlisi de factors evolutius i, també, a partir de mètodes d'estimació dels espectres més representatius de la matriu **D** a partir de la detecció de les variables o objectes més pures.

En aquest treball s'han fet servir, principalment, aquests dos mètodes:

1) L'anàlisi de factors evolutius (EFA)

L'anàlisi de factors evolutius (*Evolving Factor Analysis*, EFA) utilitza informació pròpia dels processos químics evolutius. Va ser desenvolupat el 1985 per Maeder i col·laboradors [19, 57], i posteriorment millorat i aplicat a molts tipus de processos químics en evolució [56, 59-64] .

Donat que en aquesta Tesi s'han estudiat generalment processos evolutius (per exemple, la fusió d'un àcid nucleic, una valoració àcid-base, ...), els espectres enregistrats estaran ordenats segons una determinada variable (temperatura, pH, ...). Aquesta ordenació dels espectres proporciona informació addicional, que no s'explota a l'anàlisi de components principals. En canvi, l'anàlisi de factors evolutius (EFA) es basa en seguir el rang químic de la matriu de dades (nombre de components principals relacionats amb la variació sistemàtica, és a dir, rang matemàtic en absència de soroll i error experimental) a mesura que avança el procés. L'aparició de cada nou component químic estarà associada a un increment en el rang químic de la matriu. A més de la informació de rang local en cada valor de la variable externa, a partir de les línies que uneixen els components més significatius, es pot obtenir una estimació inicial de l'evolució dels perfils de les concentracions de les diferents espècies químiques del sistema durant el procés [19, 57].

Així doncs, el mètode d'anàlisi de factors evolutius calcula els valors propis de les submatrius de dades, que provenen de la matriu de dades original, la mida de les quals va augmentant progressivament (fila a fila) a mesura que el procés avança. A cada submatriu generada, se li aplica de nou l'anàlisi de components principals, fins arribar a completar les dimensions totals de la matriu **D**. Així, per una banda s'obté l'EFA creixent (*Forward EFA*) on es comença per analitzar una submatriu formada per les dues primeres files (espectres) de la matriu de dades experimentals. A partir d'aquí es van analitzant les submatrius formades per l'addició de la fila següent a la submatriu anterior, fins arribar a completar les dimensions de la matriu experimental. D'altra banda, l'EFA decreixent o *Backward EFA* és el resultat de l'anàlisi inicial de la submatriu formada per les dues últimes files de la matriu

experimental i d'anar augmentant, aquesta submatriu, cada vegada amb la fila anterior, fins completar les dimensions de la matriu experimental.

Com s'observa a la Figura 1.8., en el diagrama EFA, s'uneixen els valors singulars corresponents al mateix component trobats en l'anàlisi per components principals de cadascuna de les submatrius. El punt on una d'aquestes línies sobresurt de la zona dels valors singulars associats amb el soroll experimental indica la fila de la matriu (variable) on aquesta espècie és detectada a l'EFA creixent o a on desapareix a l'EFA decreixent. La combinació de les anàlisis EFA creixents i decreixents proporciona informació sobre el nombre total d'espècies i sobre les finestres d'existència d'aquestes, a partir de les quals es poden obtenir estimacions inicials dels perfils de concentració C_{inic} .

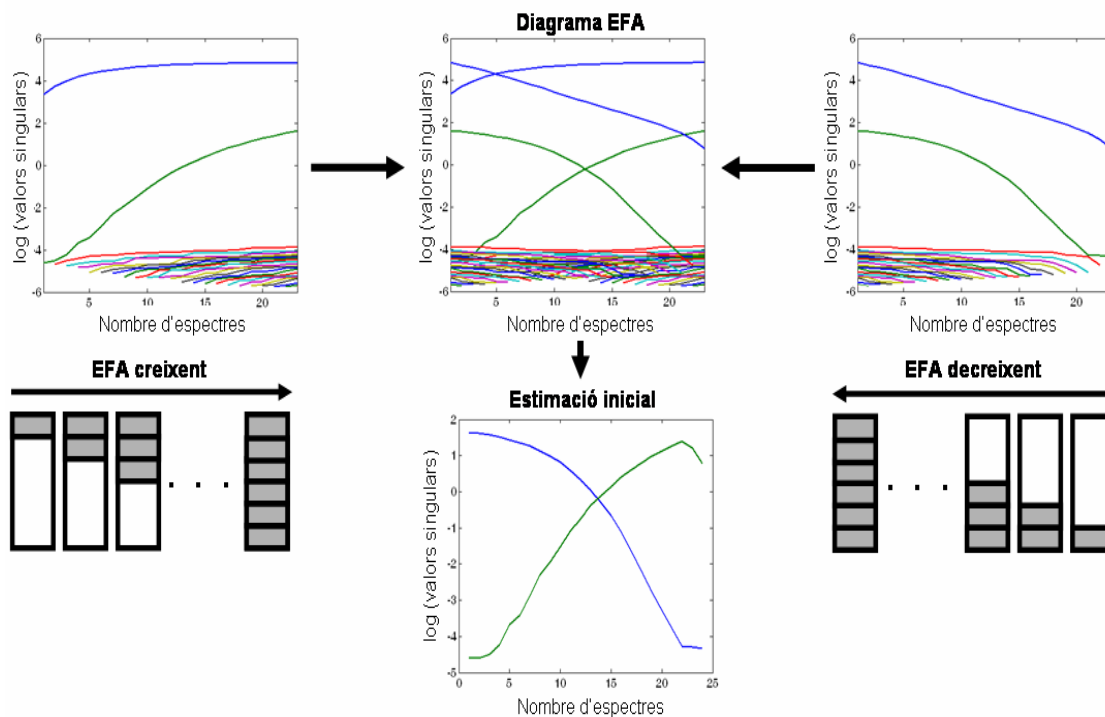


Figura 1.8. Representació esquemàtica del mètode EFA. Diagrama EFA format a partir del EFA creixent i de l'EFA decreixent i s'obtenen les estimacions inicials dels perfils de concentració.

Recentment, ha aparegut una modificació de l'algoritme EFA que permet la utilització d'aquest mètode en el cas de matrius deficientes de rang utilitzant una estratègia de matrius augmentades [62].

2) Detecció de variables pures

Cal diferenciar entre els conceptes de variable pura i d'objecte pur. Una variable pura es defineix com aquella variable o longitud d'ona on absorbeix majoritàriament un determinat component i ens serveix per obtenir estimacions inicials dels perfils de concentració dels components del sistema. Un objecte pur es defineix com aquell objecte o espectre on majoritàriament només absorbeix un determinat component i ens serveix per obtenir estimacions inicials dels espectres dels components del sistema.

S'ha fet servir el mètode de detecció de variables pures basat en l'algoritme SIMPLISMA (*Simple-To-Use-Interactive Self-Modelling Mixture Analysis*) proposat per Windig [21, 58] per determinar els espectres més purs de la matriu de dades experimentals [65, 66]. Aquest mètode es basa en la selecció inicial com a espectre més pur, de l'espectre que és més diferent de l'espectre mitjana. Aquest espectre s'obté a partir de l'avaluació de les desviacions estàndards relatives entre els espectres que formen la matriu de dades experimentals **D**.

Cada valor experimental de la matriu de dades, **D**, estarà format per contribucions d'un o més components. Es calcula la puresa (p_i) de cada objecte o espectre segons l'Equació 1.5.:

$$p_i = \frac{w_i \cdot \sigma_i}{(\mu_i + \zeta)} \quad \text{per a } i = 1, \dots, m \quad \text{Equació 1.5.}$$

on m és el nombre d'espectres.

w_i és un vector d'escalat (*weight vector*) que depèn de l'espectre (objecte) que s'ha seleccionat i de la normalització aplicada a la matriu de dades. El seu valor per a l'espectre més pur és igual a 1 i va disminuint en augmentar el nombre d'espectres seleccionats.

μ_i és el component i del vector mitjana dels espectres de la matriu de dades experimental **D**. Es troba a partir de l'Equació 1.6.:

$$\mu_j = \sum_{i=1}^n \frac{d_{ij}}{n} \quad \text{per a } j = 1, \dots, n \text{ longituds d'ona} \quad \text{Equació 1.6.}$$

σ_i és el component i del vector desviació estàndard obtingut a partir de:

$$\sigma_i = \frac{\sum_{j=1}^n \frac{d_{ij} - \bar{d}_i}{n}}{n-1} \quad \text{per a } j = 1, \dots, n \text{ longituds d'ona} \quad \text{Equació 1.7.}$$

ζ és un factor de correcció per evitar que siguin escollits com a espectres purs aquells que representin gairebé només soroll amb mitjana molt petita. Així, la puresa dels espectres que representen només soroll disminuirà a causa del factor afegit al denominador. S'escull un valor d'acord amb el nivell d'error experimental (per exemple, un 1% de la intensitat del senyal experimental).

El següent pas és el càlcul del vector d'escalat. Aquest vector es troba a partir de:

$$w_i = (\sin \alpha_i)^2 \quad \text{per a } i = 1, \dots, m \quad \text{Equació 1.8.}$$

on α_i és l'angle entre cada espectre experimental i l'espectre mitjana de tots els espectres experimentals.

Els valors de w_i seran petits quan l'espectre experimental i l'espectre mitjana siguin semblants.

Es calcula una nova matriu de dades experimentals \mathbf{D}^s a partir de l'eliminació de l'espectre (objecte) seleccionat i es torna a començar el procés per obtenir el següent espectre pur. Es repeteix el procés fins que els valors que obtenim del vector d'escalat són tots aproximadament igual a 0, la qual cosa indicarà que la variància no deguda a l'error experimental està explicada.

El mètode de detecció de variables pures pot servir també per determinar el nombre de components del sistema a partir del nombre d'espectres purs que es necessiten per explicar les dades experimentals.

1.2.2. Optimització per mínims quadrats alternats

El tercer pas en el procés de resolució multivariant de corbes és optimitzar els perfils (C_{inic} i S_{inic}^T) mitjançant mínims quadrats alternats (*Alternating Least Squares* [53, 67], ALS). Una vegada es tenen les estimacions inicials dels perfils de concentració o espectrals s'optimitzen aquestes aplicant restriccions que aportin significat químic a les solucions purament matemàtiques. Quan la matriu reproduïda (D^*) és suficientment semblant a la matriu de dades experimentals (D) es finalitza el procés iteratiu obtenint C i S^T . Aquest procés d'optimització es pot resumir en Figura 1.9.:

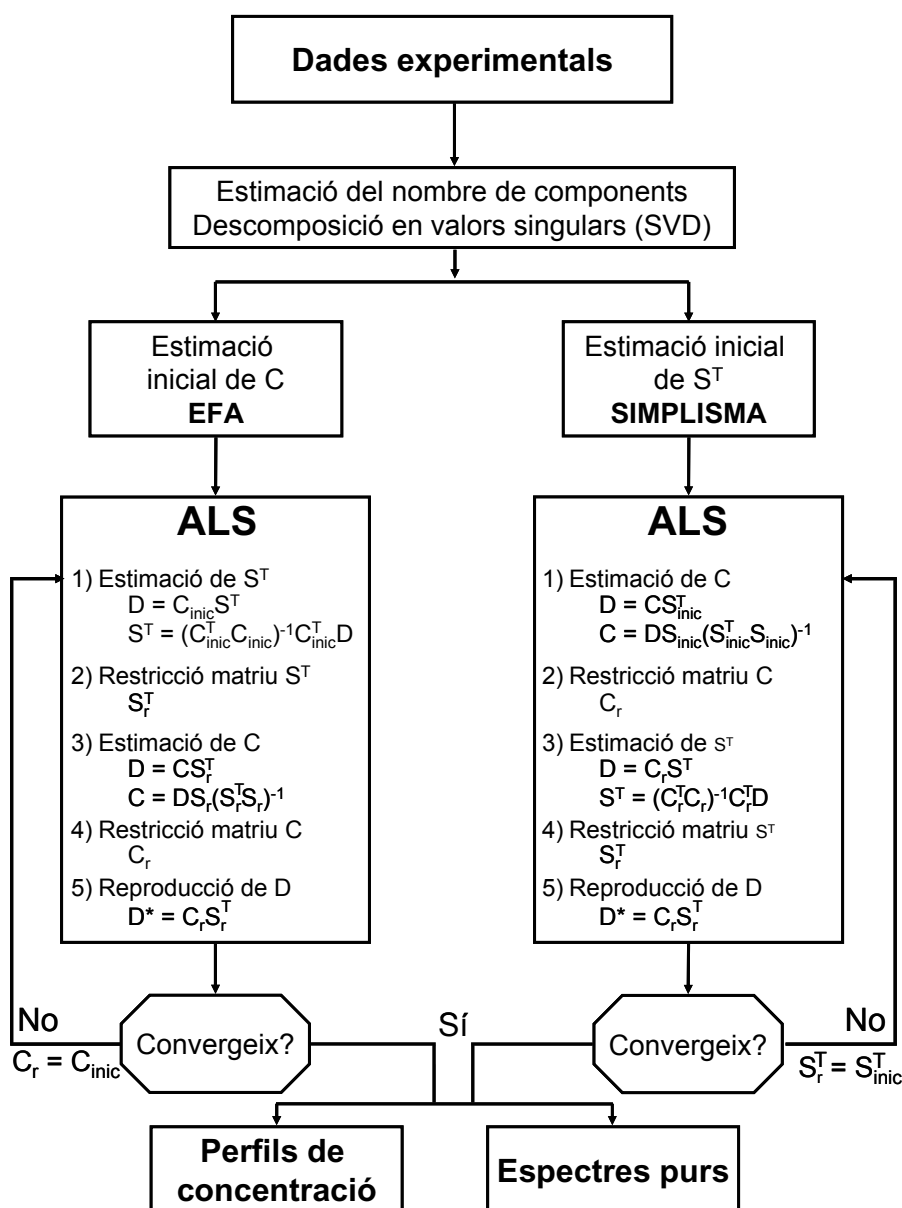


Figura 1.9. Esquema del procés de resolució del mètode MCR-ALS.

El procés iteratiu finalitza quan es compleix un criteri de convergència que s'ha establert a l'inici de l'optimització i que es basa en que la diferència relativa de les desviacions estàndards dels residuals entre els valors experimentals i els calculats en l'optimització ALS siguin menors que un valor preestablert (habitualment 0.1%).

Es pot avaluar la qualitat de l'optimització ALS a partir del percentatge de manca d'ajust, el percentatge de variància explicada i la desviació estàndard dels residuals.

L'ajust de les dades al model, o manca d'ajust (*lack of fit*) [68], es pot mesurar a partir de l'Equació 1.9:

$$\text{manca d'ajust (\%)} = 100 \cdot \sqrt{\frac{\sum_{ij} e_{ij}^2}{\sum_{ij} d_{ij}^2}} \quad \text{Equació 1.9}$$

on d_{ij} correspon a un valor de la matriu de dades original \mathbf{D} i e_{ij} correspon als residuals obtinguts segons l'Equació 1.10. a partir dels valors de la matriu original (d_{ij}) i dels valors de la matriu calculada ($d_{calc_{ij}}$) per ALS.

$$e_{ij} = d_{ij} - d_{calc_{ij}} \quad \text{Equació 1.10.}$$

Idealment, aquesta manca d'ajust hauria de tenir un valor similar al del soroll de fons, l'error experimental o variància no química present a les dades experimentals.

El percentatge de variància explicada i la desviació estàndard dels residuals respecte les dades experimentals es calculen d'acord amb les equacions següents:

$$\text{Variància explicada} = R^2 = \frac{\sum_{ij} d_{ij}^2 - \sum_{ij} e_{ij}^2}{\sum_{ij} d_{ij}^2} \quad \text{Equació 1.11.}$$

$$\text{Desviació dels residuals} = \sigma = \sqrt{\frac{\sum_{ij} e_{ij}^2}{n_{files} \cdot n_{columnes}}} \quad \text{Equació 1.12.}$$

Habitualment, es fa necessària l'aplicació de diverses restriccions per tal que les solucions matemàtiques obtingudes en el procés d'optimització adquireixin sentit químic. Una restricció es pot definir com una propietat, d'origen matemàtic o químic, que ha de ser complerta pels perfils de concentració i/o espectrals de les espècies pures per tal que la solució final tingui sentit químic [28]. Durant l'aplicació de la restricció, els elements del perfil (de concentració o espectral) que no compleixin la restricció es canvien per altres que sí ho facin. Hi ha diverses maneres d'aplicar restriccions que poden donar lloc a diversos algoritmes com es descriu a continuació.

Una breu descripció de les possibles restriccions que es poden aplicar durant l'optimització ALS conjuntament amb una representació gràfica d'elles (Figura 1.10.) és la següent:

- ♦ No-negativitat. S'aplica sempre als perfils de concentració (donat que una concentració química només pot ser positiva o igual a zero) [35]. Aquesta restricció també es pot aplicar a les respostes instrumentals on la senyal de la mostra no té valors negatius (per exemple, l'absorció molecular a l'ultraviolat o la fluorescència). En canvi, no es pot aplicar als espectres de dicroïsmo circular ja que aquests poden tenir valors negatius. Hi ha diferents formes d'aplicar la restricció de no-negativitat com la substitució directa de valors negatius per 0 o els mètodes més rigorosos basats en l'optimització per mínims quadrats no negatius (*non-negative least squares*, nnls, i la seva variant *fast non-negative least squares* [69], fnnls). En aquest treball s'ha fet servir la basada en el mètode de mínims quadrats no negatius ràpids (fnnls).
- ♦ Unimodalitat. Aquesta restricció imposa l'existència d'un únic màxim en cada perfil de resposta [70, 71]. Molts perfils de concentració associats a reaccions químiques són unimodals ja que únicament tenen un màxim o pic. Existeixen diverses formes d'imposar unimodalitat. Una d'aquestes consisteix en eliminar la part no unimodal d'un perfil (els màxims menors) substituint aquests valors per altres més petits que compleixin la restricció. Altres formes es basen en tallar els màxims secundaris vertical o horitzontalment, o en la suavització adequada dels perfils resolts.

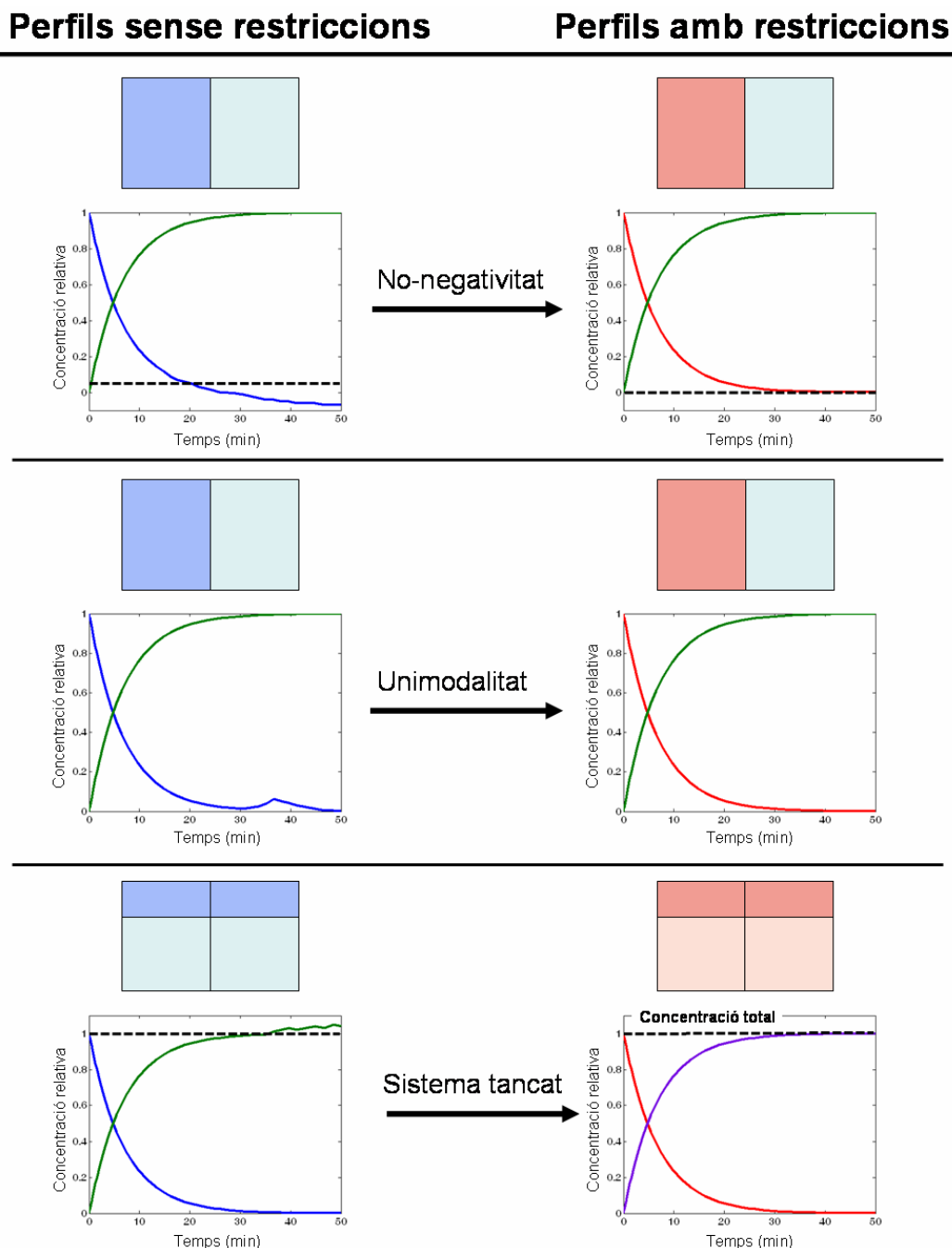


Figura 1.10. Aplicació de restriccions durant el procés d'optimització. a) no-negativitat, b) unimodalitat, c) sistema tancat.

- ◆ Sistema tancat. Aquesta condició s'aplica a alguns sistemes en els quals es sap que la suma de les concentracions de totes o algunes de les espècies detectables en un moment donat de l'experiment, és un valor determinat conegut degut al compliment d'un balanç de matèria, o bé es coneix la variació que pot tenir aquesta concentració analítica (per exemple, en els casos d'experiments on es produeix

dilució) [34, 51, 72, 73]. Aquesta restricció pot ser aplicada de dues formes diferents [74]. Una possibilitat és forçar la suma de les concentracions a ser igual a un determinat valor. Una segona possibilitat és permetre que la suma dels perfils sigui igual o inferior a aquest valor predeterminat. Aquesta segona opció pot ser molt útil en el cas de tenir petites variacions difícils de quantificar. Si en el procés d'optimització no s'aplica cap restricció de sistema tancat sobre els perfils de concentració és recomanable aplicar una normalització d'alçades o de pic als perfils espectrals per tal d'evitar indeterminacions d'escala durant la resolució [25].

- ♦ Selectivitat. Aquesta restricció es pot aplicar tant als perfils de concentració com als d'espectres [25] i permet fixar determinats valors (en el cas que siguin coneguts) durant el procés d'optimització. Les zones o regions selectives són aquelles regions, en qualsevol de les dues direccions de la matriu de dades, on només és present una espècie química. És a dir, les regions de la matriu de dades on el rang químic és igual a la unitat (per exemple, aquelles longituds d'ona on només una espècie química absorbeix) i, per tant, en aquestes regions no hi ha problemes de resolució. Si existeix selectivitat per a una espècie química determinada en les dues direccions de la matriu de dades, aquesta espècie es troba caracteritzada sense cap tipus d'ambigüitat. En canvi, si únicament existeix selectivitat a una de les dues direccions, l'ambigüitat es resol només per al perfil d'aquesta espècie en la direcció oposada. Per exemple, si a una longitud d'ona únicament absorbeix un component llavors el seu perfil de concentració es pot trobar totalment determinat sense ambigüitats, però el seu espectre encara podrà no estar totalment caracteritzat. Es pot definir la informació selectiva a partir d'igualar a zero o a valors molt petits, determinats valors dels perfils de concentració o dels espectres dels components resolts [74]. Cal destacar que aquesta restricció es pot aplicar també de forma relaxada, ja que es pot permetre que els valors a restringir siguin iguals o menors que un valor determinat.
- ♦ Aplicació de models químics. Aquesta restricció que està basada en la imposició d'un determinat model fisicoquímic durant el procés d'optimització ALS, dona lloc als mètodes de modelització rígida-flexible [26, 41, 74]. La part rígida del model explicarà la variància de processos fisico químics coneguts com poden ser, per exemple, reaccions d'equilibri àcid-base o cinètiques [26, 40, 41, 75]. La part

flexible del model explicarà la resta de la variància a la qual no se li pot aplicar el model.

Com es pot veure a l'esquema de la Figura 1.11., el funcionament del mètode MCR-ALS híbrid és pràcticament idèntic al que s'ha estudiat en el cas de la modelització flexible de les dades [26, 41, 74]. D'aquesta forma l'única modificació important de l'algoritme es produeix quan es calcula la matriu de concentracions **C** sobre la que s'imposa el compliment del model químic postulat.

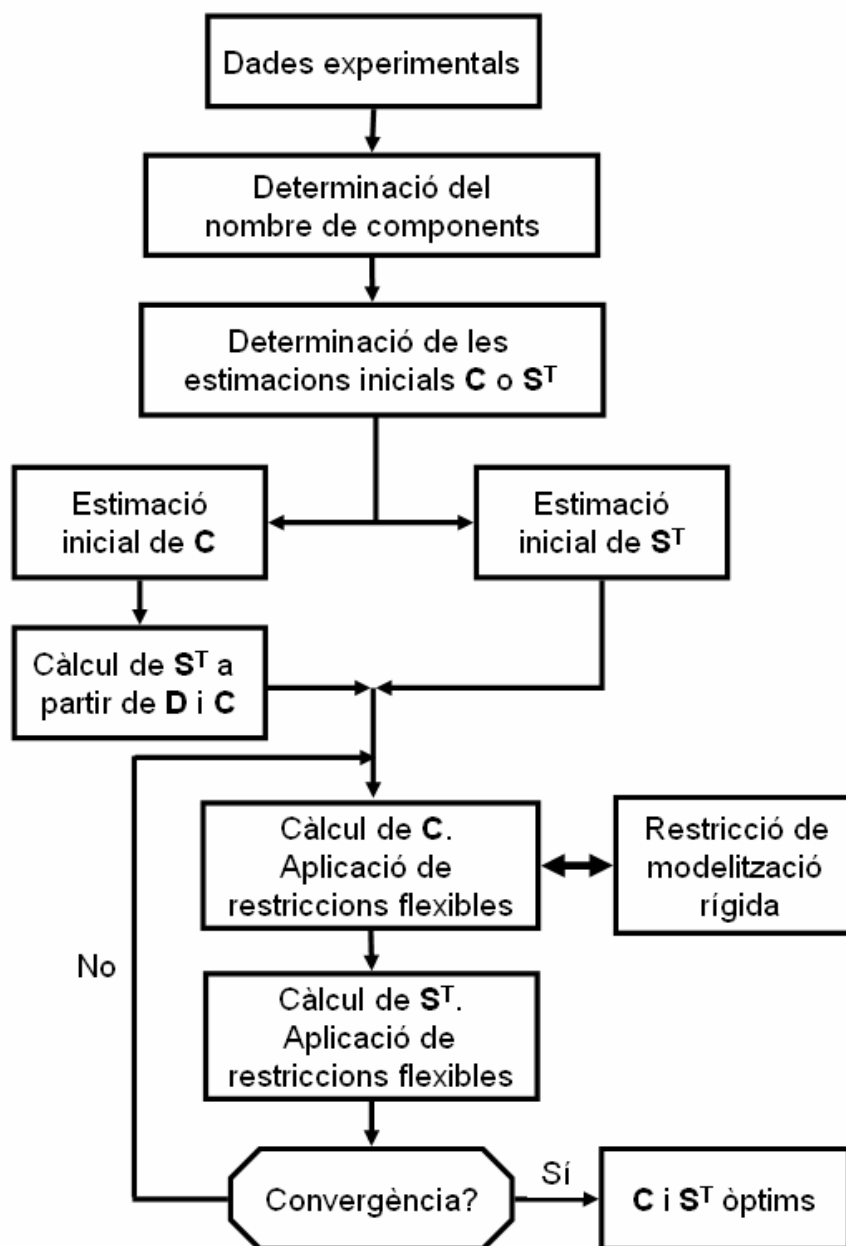


Figura 1.11. Esquema del procediment MCR-ALS amb restriccions de modelització rígida.

Actualment hi ha dos tipus de sistemes químics que es poden estudiar mitjançant el mètode de resolució MCR-ALS amb restriccions de modelització rígida i que són els sistemes d'equilibri químic i els de reaccions cinètiques. El principal avantatge que té l'aplicació de restriccions de modelització rígida durant la resolució del mètode MCR-ALS és que permet l'obtenció dels valors de les constants d'equilibri o velocitat dels processos estudiats juntament amb els seus errors associats. A més, la restricció de modelització rígida redueix els efectes de l'ambigüitat rotacional associada als perfils resolts permetent, en moltes ocasions, obtenir una solució única sense ambigüitats [74]. Així, es fa servir la part de modelització rígida per caracteritzar el procés químic conegut mentre que s'utilitza la part de modelització flexible per modelitzar variacions desconegudes, com poden ser els interferents que presenten senyal espectral o les desviacions importants de la línia base durant l'experiment. D'aquesta forma durant l'optimització ALS es podran tenir uns components descrits d'acord amb un model químic mentre que la resta de components es modelitzaran a partir només de restriccions flexibles com les que s'han descrit prèviament pel mètode MCR-ALS.

Evidentment, es poden aplicar o no les mateixes restriccions a \mathbf{C} i \mathbf{S}^T , a més, es poden aplicar aquestes restriccions a alguns o a tots els perfils espectrals o de concentració resolts.

Malgrat que l'aplicació de les restriccions guia el procés d'optimització cap a l'òptim global, en els mètodes basats en l'anàlisi de factors es poden presentar diferents problemes d'identificació de les solucions correctes. Així, tant els perfils de concentració com els espectres resolts en l'anàlisi individual MCR-ALS d'un únic experiment poden no ser els correctes donada l'existència de dos tipus d'ambigüitats en el procés de resolució [24, 25, 28]. Aquestes ambigüitats són de dos tipus:

- ♦ l'ambigüitat d'intensitat o d'escala, ja que tant els perfils de concentració com els espectres es poden trobar escalats per un factor desconegut [24]. Aquesta ambigüitat es pot minimitzar aplicant la restricció de sistema tancat. Quan no s'aplica la restricció de sistema tancat és recomanable normalitzar els perfils en un dels dos modes (concentració o espectres) per evitar que surtin fora d'escala.

- ♦ l'ambigüitat rotacional, ja que els espectres o els perfils de concentració resolts poden ésser combinació lineal dels espectres o perfils de concentració vertaders desconeguts amb un ajust similar de les dades experimentals. Aquesta ambigüitat és present sempre que al sistema estudiat li falti selectivitat, ja sigui perquè dues o més espècies tinguin un rang d'existència similar o perquè absorbeixin a la mateixa zona [24, 25, 27].

A l'apartat 1.2.3., es descriuen detalladament aquests dos tipus d'ambigüitats.

A més, es pot presentar un altre tipus de problema en l'anàlisi de les dades experimentals que és el problema de la deficiència de rang [76-78]. Aquest problema es presenta quan en un sistema només es detecta un nombre de components linealment independents inferior al nombre vertader de components químics existents en el sistema, és a dir, es detecten menys espècies de les que realment estan presents. Aquest problema és conseqüència de la presència de dependències lineals en els perfils de les espècies en un dels modes, generalment en el perfils de concentració.

Un exemple de deficiència de rang és la desnaturalització d'un ADN de cadena doble:



on és impossible resoldre mitjançant un únic experiment el diagrama de distribució de les espècies A i B ja que els perfils de concentració corresponents als dos polinucleòtids de cadena senzilla són idèntics. Així, el rang químic d'aquest sistema només podrà ser dos encara que el nombre d'espècies químiques diferents sigui tres.

Una possible solució per tal de reduir i, eventualment, resoldre totalment aquests problemes d'ambigüitats rotacionals i de deficiència de rang, és el tractament simultani MCR-ALS de diversos experiments dissenyats apropiadament [50]. D'aquesta manera es redueix l'ambigüitat rotacional de forma considerable i s'aconsegueix resoldre més fàcilment sense ambigüitat els espectres purs i els perfils de concentració vertaders dels diferents components dels sistemes estudiats.

Anàlisi simultània de dades procedents de múltiples experiments

Una de les característiques principals del mètode MCR-ALS és la possibilitat d'analitzar simultàniament dades de diferents experiments [51], estratègia que s'ha revelat molt útil ja que pot servir per superar les ambigüitats rotacionals intrínseques de l'anàlisi de factors i els problemes de deficiència de rang.

Aquesta anàlisi simultània s'aconsegueix a partir de la construcció d'una matriu **D** augmentada on diverses matrius (**D_i**) s'encadenen una al costat de l'altra per formar la matriu augmentada [25, 79]. Es pot tenir matrius augmentades de diferents formes:

- ♦ Matrius de dades augmentades en la direcció de les files (*row-wise augmented data matrix*). Aquest esquema permet l'anàlisi simultània de diverses matrius de dades, **D_i**, corresponents a un mateix sistema químic estudiat mitjançant més d'una tècnica espectroscòpica diferent. La resolució d'aquesta matriu augmentada donarà lloc a una única matriu de perfil de concentracions i a dues o més submatrius de perfils espectrals (Figura 1.12.b).

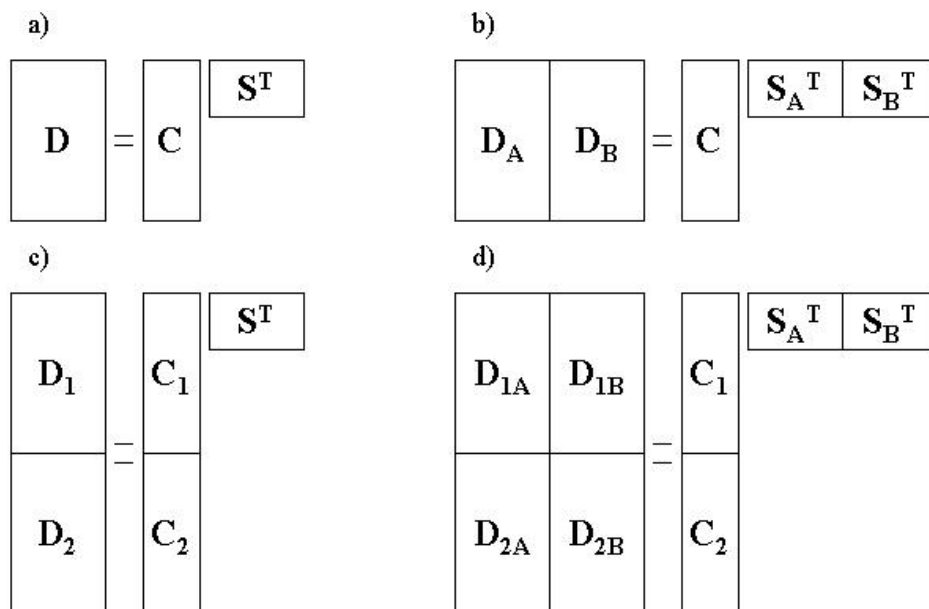


Figura 1.12. Tipus de matrius de dades. a) Matriu de dades sense augmentar; b) Matriu de dades augmentada en la direcció de les files (mateix experiment seguit mitjançant diferents tècniques); c) Matriu de dades augmentada en la direcció de les columnes (diferents experiments seguits mitjançant la mateixa tècnica); d) Matriu de dades augmentada en la direcció de les files i les columnes (diferents experiments seguits mitjançant diferents tècniques).

- ♦ Matrius de dades augmentades en la direcció de les columnes (*column-wise augmented data matrix*). Aquest esquema permet l'anàlisi simultània de diverses matrius de dades corresponents a sistemes químics amb espècies comunes estudiats amb una única tècnica experimental. En aquest cas es poden estudiar conjuntament matrius de dades d'experiments diversos de desnaturalització, valoracions, ... En l'estudi d'un sistema químic amb espècies comunes és necessari que tots els experiments tinguin el mateix nombre de columnes (longituds d'ona). D'aquesta forma, la resolució d'aquesta matriu augmentada donarà lloc a una única matriu de perfils espectrals però dues o més submatrius de perfils de concentració (Figura 1.12.c).

- ♦ Matrius de dades augmentada en la direcció de les files i en la de les columnes (*row- and column-wise augmented data matrix*). Aquest esquema permet l'estudi simultani de diversos experiments amb espècies comunes i emprant diferents tècniques instrumentals. Aquesta és l'aproximació més potent i la que permet obtenir resultats més fiables. En aquest cas es resoldran dos o més submatrius de perfils de concentració i dues o més submatrius espectrals, com es mostra a la Figura 1.12.d.

Des d'un punt de vista metodològic i algorítmic, la complexitat del procés de resolució mitjançant el mètode MCR-ALS no varia significativament, malgrat que en augmentar el nombre d'experiments analitzats pugui semblar que s'augmenta la complexitat del sistema. Així, en la resolució de matrius augmentades es generen perfils de concentració i/o espectres purs augmentats, sobre els quals es poden aplicar restriccions similars a les que s'apliquen en l'anàlisi MCR-ALS d'una única matriu de dades. A més, hi ha dues restriccions més que es poden aplicar només en el cas de treballar amb matrius de dades augmentades:

- Identificació i correspondència de les espècies entre les diferents matrius. Amb aquesta restricció es pot fixar quines espècies són presents o absents en cada submatriu augmentada [24].

- Trilinearitat. Si les dades tenen l'estructura d'un cub o d'un paral·lelepípede (per exemple, matrius de dades que tenen el mateix nombre de files i de

columnes) es poden descriure mitjançant un model trilineal. Així, els perfils resolts poden ser forçats a obeir característiques pròpies dels models trilineals [34, 80]. Aquesta restricció permet forçar els perfils resolts del mateix component en les diferents matrius (\mathbf{C}_i) a tenir la mateixa forma.

Com és lògic aquestes restriccions i les prèviament descrites (no-negativitat, unimodalitat, sistema tancat, ...) es poden aplicar a totes les submatrius de concentracions (\mathbf{C}_i) o espectrals (\mathbf{S}_i^T) o a només a algunes d'elles, és a dir, només a aquelles en les quals l'aplicació d'aquestes restriccions tingui sentit.

1.2.3. Avaluació de la qualitat dels resultats

Les principals fonts d'incertesa associada a les solucions obtingudes en la resolució són les ambigüitats rotacionals i l'error experimental present en les dades. Per tant, l'estudi de l'efecte de les ambigüitats i del soroll és necessari quan es volen comparar els mètodes de resolució multivariant amb altres procediments quimiomètrics i avaluar la qualitat dels resultats obtinguts.

A) Ambigüitats d'intensitat i rotacionals

Com s'ha esmentat anteriorment, el mètode MCR-ALS es basa en la descomposició bilineal seguint l'equació:

$$\mathbf{D} = \mathbf{C} \mathbf{S}^T + \mathbf{E} \quad \text{Equació 1.13.}$$

La solució d'aquesta equació és ambigua degut a la presència de dos tipus d'ambigüitats associades a aquest tipus de descomposició: les ambigüitats d'intensitat i les ambigüitats rotacionals [15, 43, 81-83].

L'ambigüitat d'intensitat és present sempre en aquest tipus de descomposicions ja que per a qualsevol escalar m , es complirà que,

$$\mathbf{D} = \mathbf{C} \mathbf{S}^T = (\mathbf{C} \cdot m) \cdot \left(\frac{1}{m} \cdot \mathbf{S}^T \right) = \mathbf{C}' \mathbf{S}'^T, \quad \text{Equació 1.14.}$$

Això vol dir que les concentracions i els espectres estimats es trobaran escalats per un factor desconegut m , la qual cosa pot incidir negativament en intentar realitzar determinacions quantitatives [24].

L'ambigüitat rotacional apareixerà quan hi ha dos o més components que es superposen [24]. Com a conseqüència d'aquesta ambigüitat rotacional, els perfils estimats per a qualsevol d'aquests components pot ésser una combinació lineal desconeguda dels perfils reals. Es pot representar aquesta ambigüitat amb l'equació:

$$\mathbf{D} = \mathbf{C}_{\text{vell}} \mathbf{S}_{\text{vell}}^T = (\mathbf{C}_{\text{vell}} \mathbf{T}) (\mathbf{T}^{-1} \mathbf{S}_{\text{vell}}^T) = \mathbf{C}_{\text{nou}} \mathbf{S}_{\text{nou}}^T \quad \text{Equació 1.15.}$$

A partir d'aquesta equació es pot observar que pot haver-hi un nombre infinit de solucions possibles per a qualsevol matriu \mathbf{T} no singular i invertible. Cal destacar que aquesta expressió és més general que la que s'ha tractat anteriorment referent a les ambigüitats d'intensitat i es transforma en aquesta en el cas de que \mathbf{T} sigui un escalar [27]. S'ha de tenir en compte que la matriu \mathbf{T} serà de dimensions molt més reduïdes que les dimensions de les matrius \mathbf{C} , \mathbf{S}^T i \mathbf{D} . Quan aquesta ambigüitat existeix, els perfils d'un component s'haurien de representar per una banda de solucions factibles i no per un únic perfil .

En general, es pot disminuir l'efecte de les ambigüitats d'intensitat i rotacionals mitjançant l'aplicació de restriccions durant el procés de resolució [25, 74]. Així, l'aplicació de restriccions basades en el coneixement químic com, per exemple, la de només permetre valors positius o zero per a les concentracions, disminueixen el nombre de solucions possibles, restringint els valors que pot adoptar la matriu \mathbf{T} . La restricció de sistema tancat o la de dur a terme la normalització dels perfils espectrals, ja sigui respecte a la intensitat o l'àrea dels espectres, serveix per minimitzar les ambigüitats d'intensitat. Malgrat això, la restricció que té un efecte més important per tal d'eliminar les ambigüitats és la selectivitat, ja que permet, en molts casos, determinar la solució vertadera sense cap tipus d'ambigüitat. Una altra forma de disminuir les ambigüitats associades al procés de resolució és l'estratègia d'augmentar les matrius. D'aquesta forma, a partir d'un disseny apropiat d'experiments es poden aconseguir millors condicions de selectivitat i rang local. A més, la sobredeterminació del sistema d'equacions proporcionada per l'anàlisi de diverses matrius alhora també

contribueix a disminuir els valors que pugui prendre la matriu de rotació **T** de forma que les solucions obtingudes per l'anàlisi simultània de diferents experiments sempre ens proporcionarà resultats més fiables.

El càlcul de les bandes factibles de solucions no és senzill malgrat que des de fa temps existeixen diferents aproximacions per a sistemes de 2 o 3 components. En els últims anys, han aparegut mètodes per tal de determinar l'efecte de les ambigüitats a la solució obtinguda per a sistemes sense límits de components sota unes determinades restriccions [27, 84]. Així, tenim que, per a una determinada solució sota unes determinades restriccions, es pot definir una banda de solucions factibles i aquesta banda pot ser caracteritzada pels seus límits màxim i mínim de les solucions factibles segons una determinada funció d'optimització com la que es mostra a l'Equació 1.16.

$$f_i(T) = \frac{\sum_i \mathbf{c}_{i,k} \mathbf{s}_{k,i}^T}{\sum_k \sum_i \mathbf{c}_{i,k} \mathbf{s}_{k,i}^T} \quad \text{Equació 1.16.}$$

Un dels possibles mètodes de determinació d'aquests límits es basa en la minimització (f_{min}) i maximització (f_{max}) sota restriccions de funcions objectives definides a partir de la relació entre la contribució al senyal d'un component i el senyal global de tots els components del sistema.

L'optimització d'aquesta funció ens dóna una estimació dels límits de les solucions factibles en forma de valors òptims de les matrius de rotació **T** per a cada espècie del sistema (**T**_{min} i **T**_{max}) [27]. D'aquesta forma, considerant els *k* components dels sistema, les solucions (**C**_{inic, k} i **S**^T_{inic, k}) per a un conjunt de dades sota un seguit de restriccions, i els límits de les solucions factibles per les concentracions (**C**_{max, k} i **C**_{min, k}) i els espectres (**S**^T_{max, k} i **S**^T_{min, k}) compleixen,

$$\begin{aligned} \mathbf{D} &= \mathbf{C}_{inic} \mathbf{S}_{inic}^T = \mathbf{C}_{inic} \mathbf{T}_{min} \mathbf{T}_{min}^{-1} \mathbf{S}_{inic}^T = \mathbf{C}_{min,k} \mathbf{S}_{min,k}^T = \\ &= \mathbf{C}_{inic} \mathbf{T}_{max} \mathbf{T}_{max}^{-1} \mathbf{S}_{inic}^T = \mathbf{C}_{max,k} \mathbf{S}_{max,k}^T \end{aligned} \quad \text{Equació 1.17.}$$

B) Efecte del soroll experimental en la resolució

Un cop s'han tingut en compte les incerteses presents en una determinada solució MCR-ALS degudes a les ambigüitats rotacionals o en absència d'aquestes, es pot avaluar com el soroll present a les dades experimentals es pot propagar en aquestes solucions. Aquesta font d'incertesa afecta sempre als resultats obtinguts mitjançant qualsevol mètode quimiomètric. Així, en els mètodes de calibratge lineal univariant i multivariant lineals, la determinació de l'efecte del soroll es pot estimar a partir de fórmules matemàtiques analítiques [36]. Aquest no és el cas dels mètodes com el MCR-ALS, que es pot considerar que és un mètode d'optimització no lineal (ja que es duu a terme aquesta optimització sota restriccions no lineals on s'han de determinar un gran nombre de paràmetres), on la determinació d'aquestes fórmules analítiques és un problema complex. Per superar aquesta dificultat, i proporcionar una estimació de la propagació de l'error experimental als resultats obtinguts pel mètode MCR-ALS, es proposa fer servir altres estratègies [85, 86]. En primer lloc, es pot considerar l'estratègia de dur a terme una sèrie de replicats experimentals i analitzar estadísticament els resultats obtinguts mitjançant el mètode MCR-ALS. Aquesta estratègia presenta l'avantatge de treballar únicament amb dades experimentals, la qual cosa proporciona certesa experimental a les estimacions dels errors obtinguts [16, 87]. Malgrat això, presenta inconvenients com el cost econòmic o de temps que fan falta per dur a terme un nombre gran de replicats del mateix experiment. A causa d'això han aparegut els mètodes de remostratge o *resampling* [88] que es basen en estimar la dispersió dels resultats obtinguts en l'anàlisi de múltiples replicats simulats amb un determinat nivell de soroll afegit. Per tal de simular aquests replicats hi ha diferents estratègies que proporcionen una gran quantitat de resultats independents que permeten l'estimació de la incertesa deguda al soroll als perfils resolts. Finalment, cal esmentar els mètodes basats en l'estimació dels paràmetres a partir de criteris de probabilitat màxima (*maximum likelihood*) enlloc d'emprar criteris de mínims quadrats [89, 90]. En aquests mètodes, les dades es sospesen adequadament per tal de tenir unes estimacions més robustes dels paràmetres i de l'error associat a la determinació d'aquests paràmetres.

A continuació es descriuen algunes d'aquestes estratègies:

- Simulacions de MonteCarlo

Els mètodes de MonteCarlo són força emprats com a mètodes de simulació numèrica [16]. Les simulacions de MonteCarlo simulen la repetició d'un conjunt de dades en unes condicions determinades. Les dades es generen mitjançant un model conegut i s'afegeix soroll artificial a aquestes dades per així simular el seu comportament experimental. D'aquesta manera es generen numèricament una gran quantitat de rèpliques en un temps molt petit, de forma que s'obtenen un gran nombre d'estimacions replicades, a partir de les quals es pot fer una estimació acurada de la seva incertesa.

De forma esquemàtica aquest procés es mostra a la Figura 1.13.:

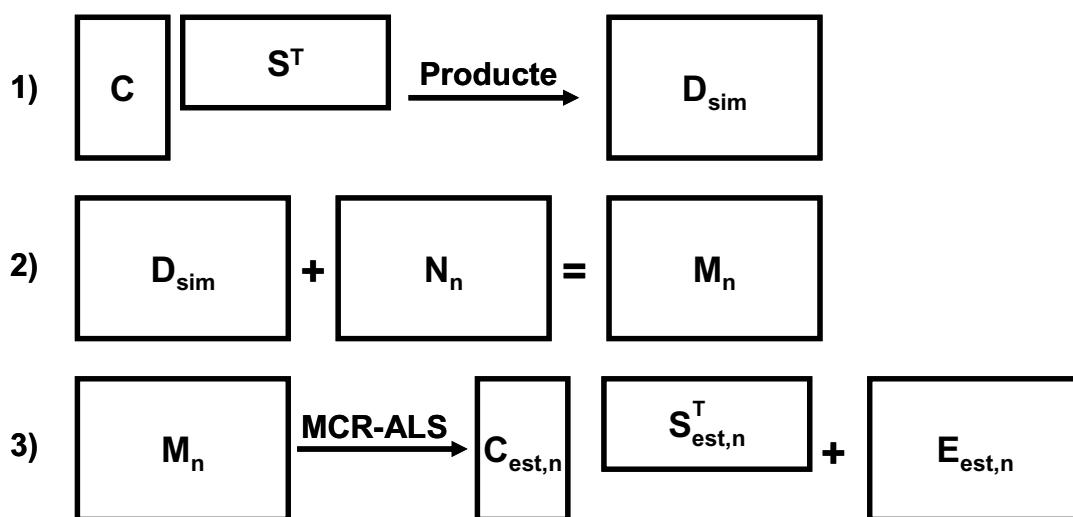


Figura 1.13. Esquema del mètode de les simulacions de MonteCarlo. 1) Simulació de les dades teòriques, 2) Addició de soroll conegut a les dades simulades, 3) Resolució amb el mètode MCR-ALS de les dades simulades amb soroll afegit.

En primer lloc, les matrius de concentració (**C**) i d'espectres (**S^T**) teòriques es multipliquen seguint el model bilineal de la llei de Lambert-Beer, descrit anteriorment per obtenir la matriu simulada de dades sense soroll (**D_{sim}**). Aleshores, s'afegeix una matriu de soroll conegut per tal d'obtenir una matriu de dades simulada que conté una quantitat de soroll coneguda **M_n**. El conjunt de matrius simulades **M_n**, són aleshores analitzades individualment per MCR-ALS, obtenint-se un conjunt de solucions MCR-ALS de les matrius **C_{est, n}** i **S^T_{est, n}** a

partir de les quals és possible estimar les seves incerteses i aplicar tests estadístics sobre l'avaluació de la propagació del soroll.

- Addició de soroll

Les simulacions mitjançant l'addició de soroll consisteixen en la repetida addició de soroll a una matriu de dades experimentals **D** (a diferència de les simulacions de MonteCarlo on aquesta matriu era una matriu teòrica simulada a partir d'un model) [16, 86]. Idealment, cada addició de soroll es pot considerar com una nova realització de tot el procés experimental sota condicions experimentals controlades i semblants. L'assumpció bàsica d'aquest mètode de remostratge és que el soroll afegit té un efecte en els resultats finals similar al soroll experimental. Esquemàticament es pot veure a la Figura 1.14:

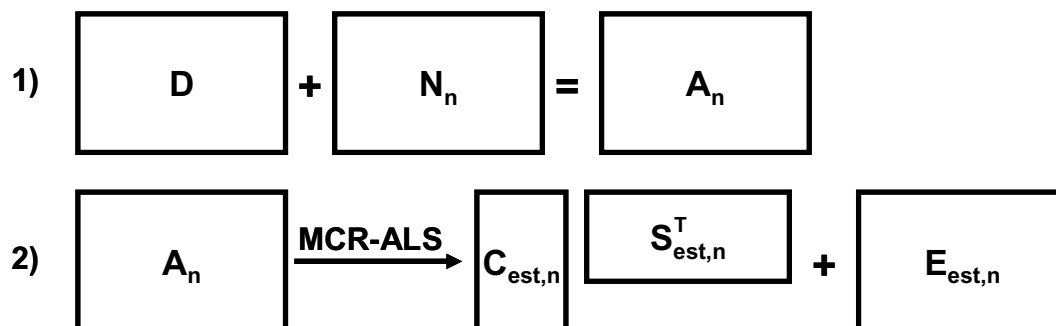


Figura 1.14. Esquema del mètode de l'addició de soroll. 1) Addició de soroll conegut a les dades experimentals. 2) Resolució amb el mètode MCR-ALS de les dades amb soroll afegit.

D'aquesta forma, aquestes matrius amb soroll afegit **A_n** tenen dos tipus diferents de soroll: una quantitat desconeguda de soroll experimental intrínsec associat a les mesures experimentals i una quantitat coneguda de soroll corresponent al soroll afegit. En aquest cas, el conjunt de matrius simulades **A_n** poden analitzar-se individualment pel mètode MCR-ALS de forma que s'obté un conjunt de solucions MCR-ALS de les matrius **C_{est, n}** i **S_{est, n}^T**. A partir d'aquestes solucions es troben les estimacions de les seves incerteses i l'avaluació de la propagació del soroll a les solucions MCR-ALS.

- *Jackknife*

Aquest mètode s'ha emprat extensivament en el món de l'estadística com a mètode de remostratge per tal de determinar la incertesa associada a models no lineals [91-93]. En l'àmbit de la Quimiometria aquest mètode s'ha fet servir per tal de determinar la incertesa associada a mètodes de regressió multivariant com la regressió per mínims quadrats parcials (PLS) [94] o l'anàlisi de factors paral·lels (PARAFAC) [85, 95, 96]. A continuació, es fa una breu descripció de com s'ha aplicat aquest mètode pel cas del MCR-ALS.

Com en el cas anterior dels mètodes d'addició de soroll, s'afegeix una quantitat coneguda de soroll a la matriu de dades experimental obtenint una nova matriu \mathbf{A}_n de dimensions m files per n columnes. A diferència dels mètodes anteriors, en els mètodes de *Jackknife* s'eliminen successivament d'aquesta matriu \mathbf{A}_n un nombre determinant (p) de files. Si repetim aquest procés de forma controlada s'obté una sèrie de matrius de dades \mathbf{J}_{ri} de dimensions ($m-p$ files i n columnes). Un esquema d'aquest procés es pot observar a la Figura 1.15.:

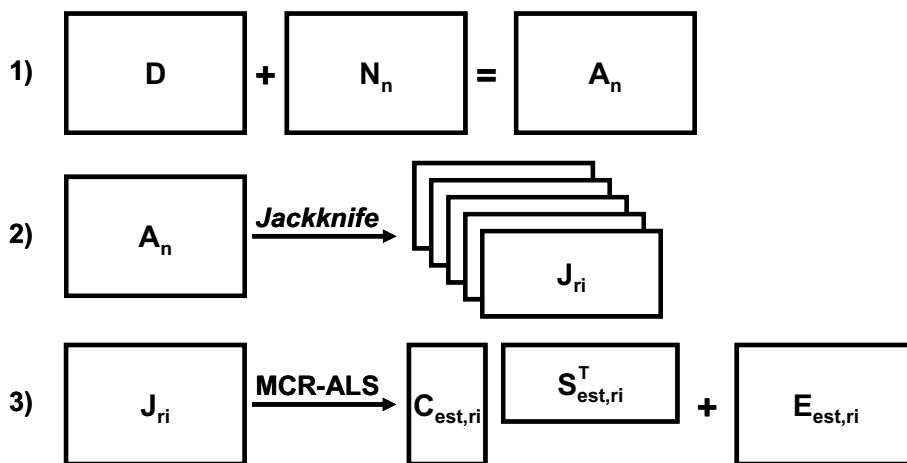


Figura 1.15. Esquema del mètode de *Jackknife*. 1) Addició de soroll conegut a les dades experimentals. 2) Aplicació del mètode de *Jackknife* per obtenir un nombre determinat de submatrius \mathbf{J}_{ri} a partir de la matriu original \mathbf{A}_n . 3) Resolució amb el mètode MCR-ALS de cada matriu obtinguda pel mètode de *Jackknife*.

El conjunt de les \mathbf{J}_{ri} matrius obtingudes mitjançant el mètode de *Jackknife* de la matriu experimental \mathbf{D} serà aleshores analitzat individualment mitjançant el mètode MCR-ALS. A partir dels perfils de concentració i espectrals resolts per

a cadascuna de les matrius obtingudes, es podrà determinar la incertesa i avaluar la propagació del soroll experimental en les solucions obtingudes pel mètode MCR-ALS. D'aquesta forma, a partir de la mateixa mesura experimental que en la cas anterior pel mètode d'addició de soroll, es poden estimar les incerteses associades als paràmetres (\mathbf{C} i \mathbf{S}^T) a partir d'un nombre molt elevat de mesures replicades amb avantatges, com la major precisió que permet obtenir una estimació del mínim de soroll que pot ser causat per motius instrumentals [85], però també amb inconvenients, com els majors requeriments computacionals, que comporta.

Malgrat que les ambigüitats inherents al procediment MCR-ALS i que la propagació del soroll experimental són dues fonts diferents d'incertesa, el seus efectes no es poden considerar de forma independent. Així, les incerteses de les solucions obtingudes es veuran distorsionades per la quantitat de soroll afegit al sistema, però al mateix temps això pot provocar l'augment de l'efecte de les ambigüitats rotacionals sobre les solucions obtingudes en la resolució per MCR-ALS del sistema considerat.

1.3. Modelització rígida

Com ja s'ha dit anteriorment, els mètodes de modelització rígida es basen també en el model bilineal de descomposició de les dades experimentals ($\mathbf{D} = \mathbf{C} \mathbf{S}^T$). En aquest cas, però, es tractarà d'obtenir els millors valors de les constants d'equilibri o velocitat (k) que determinen la matriu de concentracions \mathbf{C} a partir d'un model fisicoquímic (lleis d'acció de masses o de velocitat) i que alhora ens permeten obtenir la matriu \mathbf{S}^T amb les contribucions espectrals i ajustar les dades experimentals de la matriu \mathbf{D} [36]. L'ajust d'aquestes dades a un determinat model químic pot ser una feina complexa sobretot si no hi ha un coneixement previ suficient d'aquest model i si es pretén ajustar un nombre gran de paràmetres.

1.3.1. Consideracions bàsiques

En els mètodes de modelització rígida, el primer pas serà, determinar quin és el model químic que ajusta millor les dades experimentals. El segon objectiu serà l'optimització dels valors dels paràmetres pel model escollit. Per tal d'aconseguir aquesta

optimització, es busquen aquells paràmetres que fan que els residuals siguin mínims [97, 98]. Els residuals, e_i , es defineixen com en el cas de la modelització flexible, com la diferència entre la mesura experimental d_i i el seu valor calculat d_{calc_i} com s'ha indicat a l'Equació 1.10. Per determinar la qualitat de l'ajust obtingut es fa servir com a funció objectiva a minimitzar la suma de quadrats dels residuals [53, 98-100]. Aquesta suma de quadrats dels residuals (ssq) es defineix a l'Equació 1.18.

$$ssq = \sum_i e_i^2 = \sum_i (d_i - d_{calc_i})^2 \quad \text{Equació 1.18.}$$

Eliminació de paràmetres lineals

El treball amb dades multivariants comporta un augment considerable del nombre de paràmetres a ajustar per a un model determinat. Per exemple, si tenim el model bilineal ($\mathbf{D} = \mathbf{C} \mathbf{S}^T$) i s'estudia un sistema químic amb 2 valors de pK_a i s'ha seguit l'evolució d'aquest sistema adquirint 50 espectres a 80 longituds d'ona cadascun, el nombre total de paràmetres que s'hauran de determinar és de 242. Aquest nombre de paràmetres és igual al nombre de pK_a a determinar ($n_k = 2$) més el nombre d'absortivitats molars a partir dels quals es determinen els espectres, \mathbf{S}^T , i que són 240 ja que hi ha 3 espècies i 80 longituds d'ona. No obstant aquest elevat nombre de paràmetres, el problema es pot simplificar de forma considerable quan es considera que aquests paràmetres es poden dividir en dos grups [53, 99, 100]. Així, hi ha el primer conjunt petit de constants (de velocitat o d'equilibri) que són paràmetres no lineals i, en segon lloc, hi ha el conjunt gran d'absortivitats molars que són paràmetres lineals. Aquests dos tipus de paràmetres influeixen de forma independent sobre els termes de l'equació bilineal i, per tant, aquests dos termes es poden separar linealment. A partir d'aquí, l'objectiu de l'optimització serà aconseguir ajustar els valors dels paràmetres no lineals que, en ser un nombre petit, no faran necessaris grans recursos computacionals [97].

Com s'ha dit anteriorment, a partir de les constants que defineixen el sistema químic (conjuntament amb els valors de les concentracions inicials i el model a aplicar), es pot estimar la matriu \mathbf{C} de perfils de concentració. Així, un cop estimada la matriu de concentracions \mathbf{C} es pot avaluar la matriu d'absortivitats molars \mathbf{S}^T , mitjançant la seva estimació per mínims quadrats a partir del model bilineal:

$$\mathbf{S}^T = \mathbf{C}^+ \mathbf{D} \quad \text{Equació 1.19.}$$

on la matriu \mathbf{C}^+ correspon a la matriu pseudoinversa de \mathbf{C} [97]. La matriu pseudoinversa \mathbf{C}^+ es defineix com la matriu que multiplicada per \mathbf{C} dona la matriu identitat. Així, no cal que la matriu \mathbf{C} sigui quadrada en el cas de treballar amb matrius de rang complet. En aquest darrer cas la pseudoinversa d'una matriu \mathbf{M} equival a la inversa generalitzada que es calcula segons l'equació:

$$\mathbf{M}^+ = \mathbf{M}^T [\mathbf{M}^T \mathbf{M}]^{-1} \quad \text{Equació 1.20.}$$

La resolució de l'Equació 1.19. permet obtenir la matriu de residuals \mathbf{E} i, en conseqüència, calcular el valor de la suma de quadrats dels residuals (Equació 1.18.):

$$\mathbf{E} = \mathbf{D} - \mathbf{C}\mathbf{S}^T = \mathbf{D} - \mathbf{C}\mathbf{C}^+\mathbf{D} \quad \text{Equació 1.21.}$$

Com es pot observar en l'Equació 1.21., el càlcul dels residuals (i, per tant, la suma de quadrats dels residuals) depèn únicament de la matriu de concentracions i, per tant, del valor dels paràmetres del model químic [97].

Optimització

L'optimització dels paràmetres consisteix en la determinació dels valors dels paràmetres d'un determinat model que fan que la suma de quadrats dels residuals (*ssq*) sigui mínima [36].

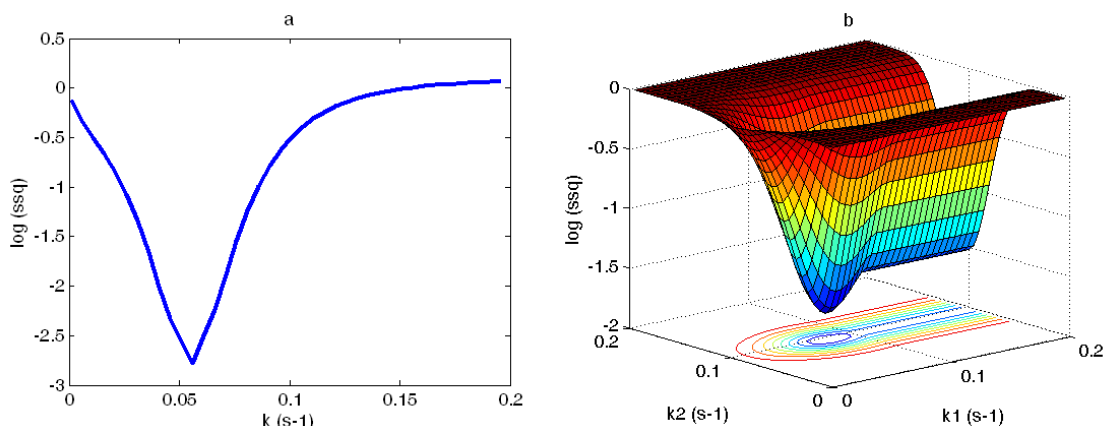


Figura 1.16. Exemples de superfícies de resposta obtingudes en l'optimització, a) d'un paràmetre i b) de dos paràmetres.

Per exemple, quan es pretén optimitzar un sol paràmetre, s'obté una representació del logaritme de la suma de quadrats dels residuals respecte aquest paràmetre que presenta un únic valor mínim (Figura 1.16.a). Si es volen optimitzar dos paràmetres s'obté una superfície de resposta bidimensional (Figura 1.16.b). Si es volen optimitzar tres o més paràmetres tindrem representacions multidimensionals que no es poden representar gràficament sobre el pla.

Hi ha diferents mètodes d'optimització de paràmetres no lineals. Aquests mètodes es poden classificar en dos grups [36, 101]. En el primer grup es troben els mètodes iteratius directes, que es basen en minimitzar la suma de quadrats dels residuals trobant el mínim de les figures anteriors sense utilitzar el càlcul de les derivades. Exemples d'aquests mètodes són el mètode Simplex [102] o els algoritmes genètics [103, 104]. En el segon grup es troben els basats en el mètode de Newton-Gauss en els quals els residuals guien el procés iteratiu cap al mínim [36]. A continuació es presenta una explicació detallada d'aquest mètodes que és el que s'ha fet servir en aquesta Tesi Doctoral.

- Mètode de Newton-Gauss-Levenberg/Marquardt (NGL/M)

S'inicia el procés d'optimització amb unes estimacions inicials dels paràmetres \mathbf{k}_0 . A continuació, es milloren aquests paràmetres a cada iteració mitjançant el càlcul del vector de canvis dels paràmetres, $\Delta\mathbf{k}$, el qual es determina a partir de les derivades dels residuals respecte dels paràmetres [97].

La matriu dels residuals, després d'aplicar el vector de canvi de paràmetres ($\Delta\mathbf{k}$), es pot aproximar mitjançant una sèrie de Taylor de primer ordre:

$$\mathbf{R}(\mathbf{k}_0 + \Delta\mathbf{k}) \cong \mathbf{R}(\mathbf{k}_0) + \Delta\mathbf{k} \frac{d\mathbf{R}(\mathbf{k}_0)}{d\mathbf{k}} \quad \text{Equació 1.22.}$$

La derivada $\frac{d\mathbf{R}(\mathbf{k}_0)}{d\mathbf{k}}$ s'anomena Jacobià \mathbf{J} . Per trobar, el vector de canvis dels paràmetres, $\Delta\mathbf{k}$, es minimitza el residual $\mathbf{R}(\mathbf{k}_0 + \Delta\mathbf{k})$ igualant la seva aproximació de Taylor de primer ordre a zero.

$$\mathbf{R}(\mathbf{k}_o + \Delta \mathbf{k}) \cong \mathbf{R}(\mathbf{k}_o) + \mathbf{J} \Delta \mathbf{k} \cong \mathbf{0} \quad \text{Equació 1.23.}$$

A partir d'aquesta expressió, es pot calcular:

$$\Delta \mathbf{k} = -\mathbf{J}^+ \mathbf{R}(\mathbf{k}_o) \quad \text{Equació 1.24.}$$

L'estructura iterativa de l'algoritme Newton-Gauss es pot veure a la següent figura:

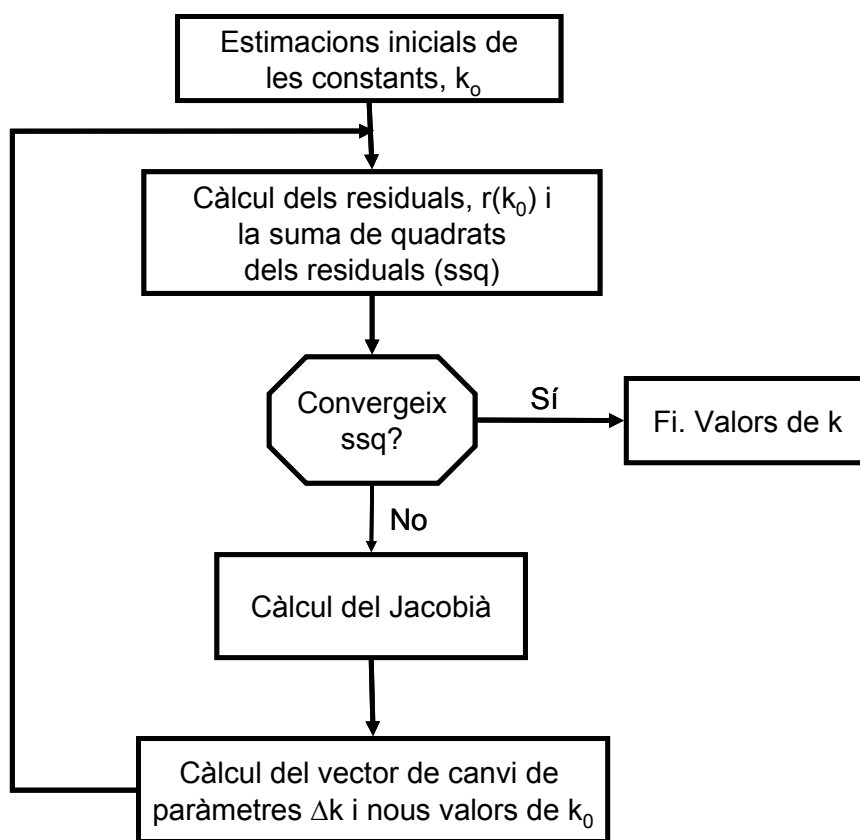


Figura 1.17. Esquema del procediment de l'algoritme Newton-Gauss.

Durant el procés d'optimització, els canvis del vector $\Delta \mathbf{k}$ juntament amb la suma dels quadrats dels residuals disminueixen constantment. Així, s'acostuma a utilitzar un criteri de convergència basat en el canvi relatiu de la suma dels quadrats dels residuals en dues iteracions successives.

Aquest mètode presenta l'inconvenient de que si les estimacions inicials són dolentes el vector $\Delta \mathbf{k}$ pot no dirigir-se cap al mínim de la funció i, per tant, es pot provocar una

optimització divergent. Per això, Marquardt [105], basant-se en les idees de Levenberg [106], va introduir una modificació d'aquest algoritme. Així, si s'afegeix un valor escalar mp , conegut com el paràmetre de Marquardt, als elements de la diagonal de la matriu quadrada $\mathbf{J}^T\mathbf{J}$ (necessària per calcular la pseudoinversa \mathbf{J}^*), es produirà un escurçament del vector $\Delta\mathbf{k}$ i la millora del procés d'optimització, ja que s'eviten optimitzacions divergents [99].

Així, l'esquema del mètode Newton-Gauss-Levenberg/Marquardt és el següent:

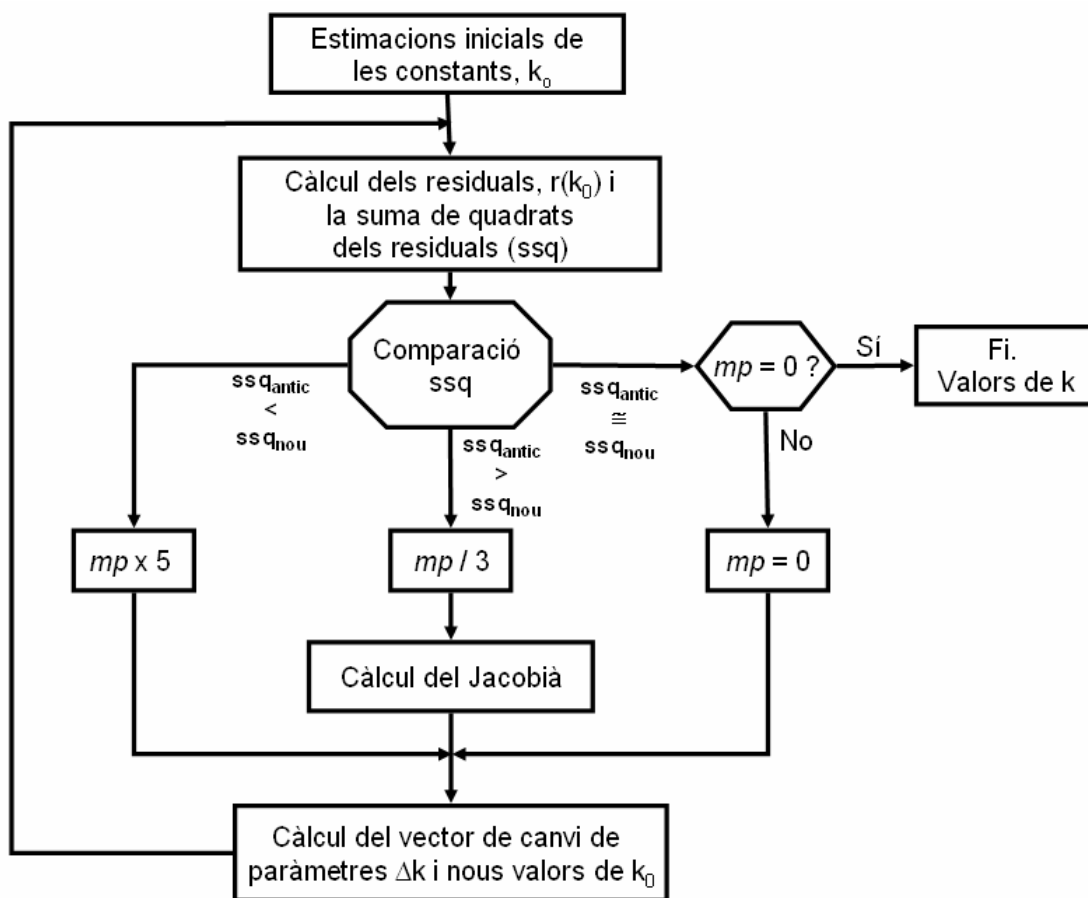


Figura 1.18. Esquema del procediment de l'algoritme Newton-Gauss-Levenberg/Marquardt

Els mètodes de Newton-Gauss presenten a més l'avantatge de permetre una estimació directa de l'error associat als paràmetres ajustats [87]. D'aquesta forma l'error estàndard σ_{k_i} dels k_i paràmetres ajustats es pot estimar a partir de l'expressió:

$$\sigma_{k_i} = \sigma_d \sqrt{h_{ii}}$$

Equació 1.25.

on h_{ii} és l'element i-éssim de la diagonal de la matriu hessiana invertida $(\mathbf{J}^T \mathbf{J})^{-1}$ (sense el paràmetre de Marquardt afegit) i σ_d representa la desviació estàndard de l'error mesurat en \mathbf{D} .

$$\sigma_d = \sqrt{\frac{\text{suma de quadrats dels residuals}}{\nu}} \quad \text{Equació 1.26.}$$

on ν representa el nombre de graus de llibertat del sistema que en aquest cas és igual al nombre de valors experimentals (nombre m d'espectres adquirits multiplicats pel nombre n de longituds d'ona mesurades) menys el nombre de paràmetres no lineals que s'optimitzen (que és el nombre nk de constants optimitzades) i menys el nombre de paràmetres lineals (equivalent a les dimensions de la matriu \mathbf{S}^T , és a dir, un nombre nc de components per un nombre n longituds d'ona mesurades). El càlcul del nombre de graus de llibertat serà doncs:

$$\nu = m \cdot n - nk - nc \cdot n \quad \text{Equació 1.27.}$$

Anàlisi simultània de diferents experiments

En moltes ocasions, la matriu \mathbf{C} té rang complet per la qual cosa els perfils de concentració són linealment independents i, per tant, es pot calcular sense problemes la matriu pseudoinversa \mathbf{C}^+ i es poden determinar els seus espectres associats. No obstant, en el cas d'estudis cinètics, és freqüent que els perfils de concentració siguin linealment dependents i que, per tant, la matriu \mathbf{C} no sigui de rang complet i que no es pugui calcular la seva inversa [78, 107, 108]. És el cas, per exemple, de la reacció $A + B \rightarrow P$ on el rang químic de la matriu només és dos i el sistema complet de tres components no es pot resoldre. Un altre problema es quan les espècies es troben a molt baixa concentració durant tot l'experiment, de forma que tant els paràmetres que les determinen com els seus espectres seran de baixa qualitat.

Per tal de superar aquest problemes, es pot dur a terme l'anàlisi simultània de diferents experiments sota condicions inicials diferents [107, 108]. Aquest tipus d'anàlisi de dades és similar al que s'ha descrit amb detall anteriorment per a l'anàlisi simultània de diverses matrius amb el mètode MCR-ALS [24].

Per fer aquesta anàlisi simultània es poden fer servir dues descomposicions diferents de la matriu de dades com es mostra a l'esquema següent:

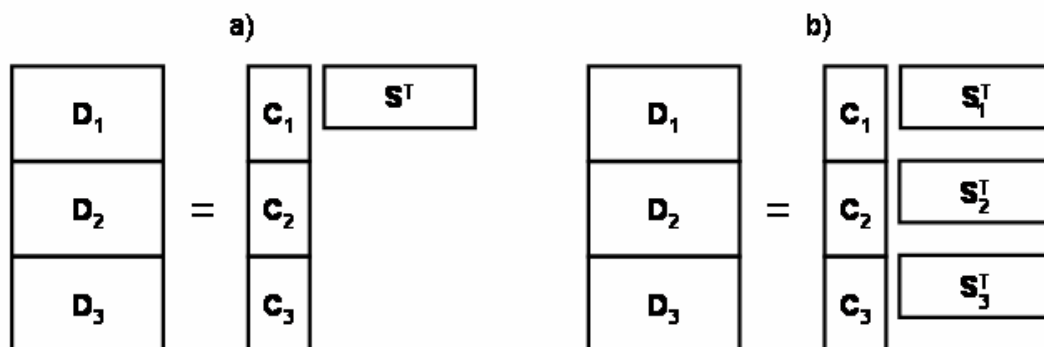


Figura 1.19. Representació esquemàtica de l'anàlisi global total (a) i de l'anàlisi global local (b).

En el primer cas es fa una anàlisi global total ja que, a partir de diferents experiments, s'obtenen diferents matrius de concentracions corresponents a cada experiment però una única matriu d'espectres. En el segon cas es tractaria d'una anàlisi global local, ja que per cada experiment s'obté una matriu de perfils de concentració i una matriu d'espectres [107]. A més, cal tenir en compte que, en ambdós casos, malgrat que es resolen perfils de concentració diferents per a cada experiment, s'obindrà un únic valor dels paràmetres a determinar, és a dir, que el valor dels paràmetres (per exemple, constants d'equilibri o cinètiques) serà global per a tot el conjunt de dades. La comparació entre aquests dos mètodes permet establir les situacions on la utilització d'un dels dos és preferible. Així, en el cas de tenir espècies linealment dependents és recomanable dur a terme una anàlisi global total mentre que si es tenen components mal definits només en alguns experiments, és recomanable dur a terme una anàlisi global local.

Una vegada s'han descrit els fonaments dels mètodes de resolució rígids emprats en aquesta Tesi es descriuran els dos tipus de processos químics als quals s'han aplicat.

1.3.2. Aplicació dels mètodes de modelització rígida a l'estudi de sistemes en equilibri químic

El fonament bàsic de la modelització de sistemes en equilibri químic és el compliment de la llei d'acció de masses. En els casos més simples es poden calcular els valors de

la matriu **C** (és a dir, els valors de les concentracions de cada espècie a cada punt de la valoració) a partir d'equacions explícites que relacionen la llei d'acció de masses i les concentracions totals de cada espècie. Quan es considera el cas general de l'equilibri:



la llei d'acció de masses relaciona les concentracions de la forma següent:

$$\beta_{abc} = \frac{[A]_a [B]_b [C]_c}{[A]^a [B]^b [C]^c} \quad \text{Equació 1.29.}$$

A partir de la llei d'acció de masses i de les concentracions analítiques de cada espècie es poden desenvolupar les equacions que hauran de complir les concentracions en cada punt experimental. Si es consideren mecanismes de reacció complexos la postulació d'aquestes equacions no és immediata i, per tant, s'utilitzen mètodes no explícits de càlcul numèric de les derivades parcials del Jacobià (Equació 1.22. i 1.23.) [109].

Així, una vegada es postulen les equacions que defineixen els paràmetres a modelitzar es tracta de calcular les concentracions de totes les espècies en l'equilibri i d'ajustar el valor òptim de les constants d'equilibri mitjançant un procés iteratiu. Aquesta optimització es duu a terme mitjançant l'algoritme Newton-Gauss-Levenberg/Marquandt i del compliment de la llei d'acció de masses, dels valors inicials de les constants d'equilibri i de les concentracions analítiques de les espècies presents [99]. Com es lògic, si es tenen estimacions inicials acurades dels valors de les constants d'equilibri, l'optimització serà més ràpida i fiable. El procediment d'optimització seguirà el model de la Figura 1.18. per l'algoritme Newton-Gauss-Levenberg/Marquandt.

Aquesta mateixa aproximació es pot fer servir per qualsevol model d'equilibri d'interacció lligand-metall o per valoracions àcid-base, malgrat que en augmentar el nombre de components també augmentarà la dificultat en la selecció correcta del model a ajustar.

Programari utilitzat

Existeixen diversos programes per tal de dur a terme la modelització rígida de reaccions en equilibri com, per exemple, el SQUAD [110, 111] o el SPECFIT [100]. En aquesta Tesi s'ha fet servir el programa EQUISPEC [107] desenvolupat per Maeder i col·laboradors. Aquest programa funciona en la línia de comanda per selecció de les diferents opcions dins de l'entorn de computació de MATLAB®.

L'optimització de les constants d'equilibri es realitza mitjançant l'algoritme de Newton-Gauss-Levenberg/Marquandt, descrit anteriorment, de forma que s'obté el valor de les constants i l'error associat a elles. A més, aquest programa permet la possibilitat de realitzar l'anàlisi simultània de diferents experiments tant de forma global total com de forma global local.

a

```

1 function [fname, drive, op, data, location_known] = m_znphen;
2
3 % M_ZNPHEN.M
4 %
5
6 %filename(s) for location of data
7 %drive in form 'c:\dirname\' - op in form 'xy'
8
9 fname = {'phen_67y'; 'phan_83y'; 'phzn_72y'; 'phan_87y'};
10 drive='c:\matlab7\quimio_tools\equispec\data\';
11 op='rd';
12
13 %
14 %      M      L      H logbeta loose?  col? known?
15 data=[ 1      0      0      0      0      0      0      0
16        0      1      0      0      0      0      1      0
17        0      1      1      4.93  1      1      1      0
18        1      1      0      6.58  1      1      1      0
19        1      2      0      11.91 1      1      1      0
20        1      3      0      17.12 1      1      1      0]; %2n/phen
21
22 % filenames corresponding to known spectra (single datafile method only)
23 location_known={' '};
    
```

Figura 1.20. Representació del programa EQUISPEC. a) Arxiu .m per dur a terme la selecció del model d'equilibri a aplicar.

b

```

Command Window
File Edit Debug Desktop Window Help
>> equispec

EQUISPEC - for Global Spectrophotometric Equilibrium Calculations

  _R_ead data from file
  _G_enerate data
  _C_alculate (A global)
  _2_nd calculation type (A local)

  _A_lter single data file (read/generate first)
  _W_rite results file
  _L_oad results file
  save _K_nown absorptivity spectra (plot A first)
  _P_lot

  _Q_uit

Enter letter of option> |
    
```

c

```

Command Window
File Edit Debug Desktop Window Help
0
0
4.8423
6.5500
11.8978
17.2161

Sum of squares = 0.0229083

normal termination
log beta(100) = 0
log beta(010) = 0
log beta(011) = 4.84235 +- 0.0130522
log beta(110) = 6.55005 +- 0.0206013
log beta(120) = 11.8978 +- 0.0592087
log beta(130) = 17.2161 +- 0.115097
sigY = 0.00552301

<enter>

EQUISPEC - for Global Spectrophotometric Equilibrium Calculations

  _R_ead data from file
  _G_enerate data
    
```

Figura 1.20. Representació del programa EQUISPEC. b) Línia de comanda on es poden seleccionar les opcions inicials del programa. c) Línia de comanda dels resultats obtinguts pel programa.

d

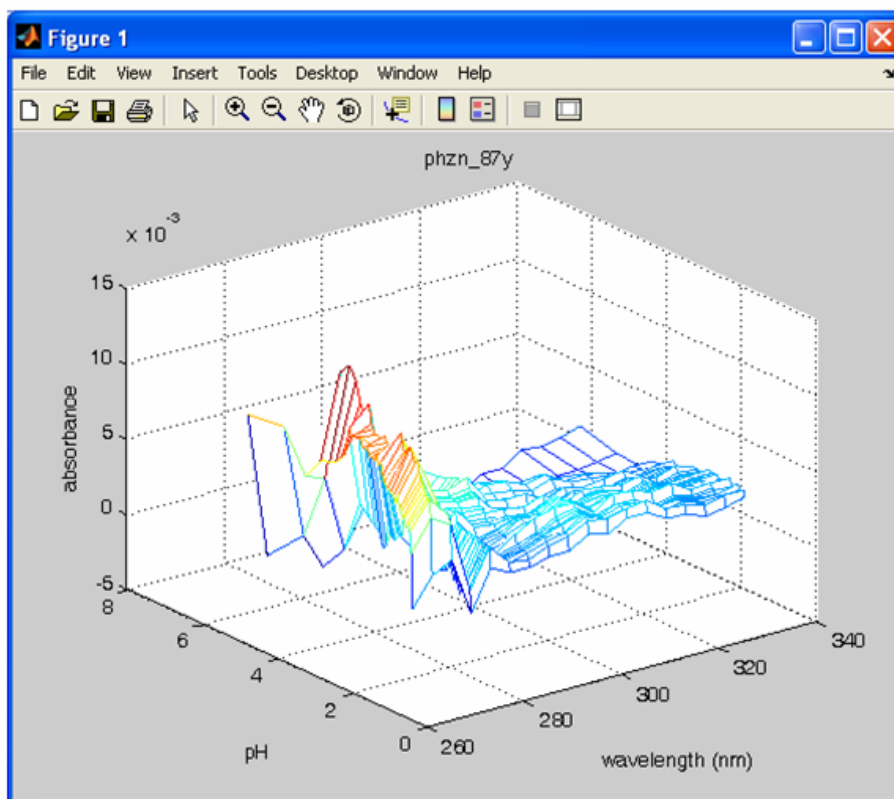


Figura 1.20. Representació del programa EQUISPEC. d) Representació gràfica dels resultats. En la figura es mostra un exemple dels residuals obtinguts després de l'optimització.

Per dur a terme aquesta optimització, en primer lloc s'ha de crear una descripció del model (arxiu de tipus .m) en el qual es dona la llista de totes les espècies del sistema i dels valors aproximats de les constants d'equilibri (Figura 1.20.a). El següent pas es cridar a la funció "equispec.m" amb la qual cosa s'obté el menú on es poden seleccionar les diferents opcions (Figura 1.20.b). Així, a partir d'aquest menú es poden importar les dades que es volen analitzar (*read*), dur a terme els càlculs (*calculate*) o, una vegada s'han obtingut els resultats, es poden obtenir diferents representacions dels perfils de concentració, espectrals o dels residuals.

Així, per exemple, els resultats obtinguts a l'ajustar el model (Figura 1.20.a.) corresponent a la complexació de la fenantrolina amb el zinc [107] proporciona els valors de constants d'equilibri que es poden observar a la Figura 1.20.c. A més, el programa permet l'obtenció de representacions com, per exemple, la dels residuals que es mostra a la Figura 1.20.d.

1.3.3. Aplicació dels mètodes de modelització rígida a l'estudi de sistemes químics cinètics

Freqüentment s'estudia l'evolució d'un sistema químic al llarg del temps a partir de la variació dels valors de concentració de les espècies del sistema en funció del temps.

El model químic que descriu el mecanisme de reacció ve definit per a un sistema d'equacions diferencials. En els casos més simples es poden calcular els valors de la matriu **C** (és a dir, els valors de les concentracions de cada espècie a cada temps donat) a partir d'equacions explícites que relacionen la concentració i el temps. Per exemple, en el cas de la reacció cinètica de primer ordre $A \rightarrow B$ es defineix una equació diferencial de primer ordre (ODE) de la forma $\frac{d[A]}{dt} = -\frac{d[B]}{dt} = -k[A]$. La integració d'aquesta equació ens proporciona els valors de concentració per a cada espècie en funció del temps: $[A] = [A]_0 e^{-kt}$ i $[B] = [A]_0 - [A]$ [112].

Malauradament, el nombre de mecanismes de reacció dels quals es pot tenir la seva solució analítica és molt limitat. Per tant, en general, s'utilitzen mètodes d'integració numèrica per determinar els valors de concentració a partir de les equacions diferencials ordinàries corresponents. Els principals mètodes que es fan servir per realitzar aquestes integracions numèriques són:

- El Mètode d'Euler

És el mètode més simple ja que es basa únicament en els valors inicials de les concentracions i en l'expressió de les derivades per calcular els valors al següent punt experimental [113]. Aquest procés es pot repetir indefinidament fins arribar al final de l'experiment com es veu a la Figura 1.21.

Matemàticament aquest mètode es pot explicar mitjançant la fórmula següent:

$$C_{i+1} = C_i + \Delta t_{i+1,i} \cdot f(t_i, C_i) \text{ per a } i=1, 2, 3, \dots \quad \text{Equació 1.30.}$$

on C_{i+1} és el nou valor de concentració, C_i és el valor de concentració original, $\Delta t_{i+1,i}$ és l'interval de temps entre els punts i $f(t_i, C_i)$ és la funció que relaciona els punts experimentals.

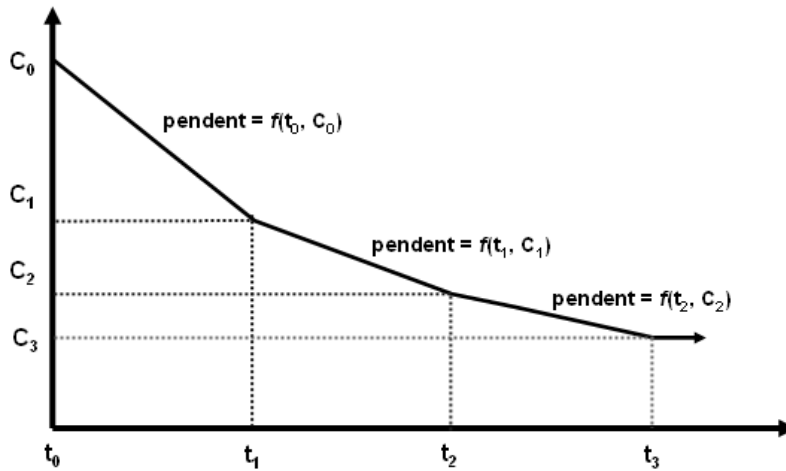


Figura 1.21. Representació esquemàtica de la integració mitjançant el mètode d'Euler.

Aquest mètode presenta limitacions en l'obtenció de perfils de concentració acurats ja que s'ha de treballar amb intervals de temps molt petits, la qual cosa fa augmentar considerablement el temps de càlcul.

- Mètode de Runge-Kutta de quart ordre

Aquest és el mètode de càlcul numèric d'integrals més habitualment emprat, ja que supera les limitacions del mètode d'Euler [113]. Es pot entendre aquest mètode com una ampliació del mètode d'Euler ja que la fórmula amb la qual es treballa és la següent:

$$C_{i+1} = C_i + \frac{1}{6}R + \frac{1}{3}R_1 + \frac{1}{3}R_2 + \frac{1}{6}R_3 \quad \text{Equació 1.31.}$$

on $R = \Delta t_{i+1,i} \cdot f(t_i, C_i)$, $R_1 = \Delta t_{i+1,i} \cdot f(t_i + \Delta t_{i+1,i} / 2, C_i + R/2)$,
 $R_2 = \Delta t_{i+1,i} \cdot f(t_i + \Delta t_{i+1,i} / 2, C_i + R_1 / 2)$, i
 $R_3 = \Delta t_{i+1,i} \cdot f(t_i + \Delta t_{i+1,i}, C_i + R_2)$.

En aquest cas particular no es considera el pendent de la funció al punt t_i sinó que es considera al punt mig entre t_i i t_{i+1} , és a dir, al punt $t_i + \Delta t_{i+1}/2$ de forma que es guanya precisió en el càlcul.

Una vegada es calculen els valors de les concentracions, es pot iniciar l'optimització iterativa dels paràmetres cinètics (constants de velocitat) mitjançant els algorismes descrits anteriorment com, per exemple, el de Newton-Gauss-Levenberg-Marquardt.

El càlcul dels valors de les concentracions presenta moltes similituds amb els càlculs que es realitzen en els experiments de reaccions en equilibri àcid-base o de complexació però amb una diferència important. En els càlculs d'equilibris, cada concentració en cada punt discret és tractada independentment de les altres mentre que en les anàlisis cinètiques, les concentracions a un temps qualsevol depenen de les concentracions calculades anteriorment.

Programari utilitzat

En aquesta Tesi s'ha utilitzat el programa GUIPRO [114] per tal de dur a terme el càlcul de les constants de velocitat de reaccions químiques mitjançant modelització rígida. Aquest programa, desenvolupat per Gemperline, consisteix en una interfície gràfica en l'entorn de treball de MATLAB® a partir de la qual es poden fer diferents tipus d'anàlisi de dades entre els quals es troba la modelització rígida de dades d'experiments cinètics.

Encara que GUIPRO pot fer servir diversos mètodes d'optimització, el que proporciona els millors resultats és el que s'ha descrit anteriorment basat en l'algoritme Newton-Gauss-Levenberg-Marquardt. Així, a partir de les dades experimentals, dels valors de concentracions inicials i de la postulació d'un mecanisme de reacció particular, s'obté el valor de les constants (amb els seus perfils de concentració i espectrals associats) i l'error associat a aquestes constants. En les versions d'aquest programa disponibles actualment, només es pot dur a terme l'anàlisi individual d'una matriu de dades de forma que no existeix la possibilitat de dur a terme l'anàlisi simultània (global) de diferents experiments cinètics a diferència dels programes EQUISPEC o MCR-ALS descrits prèviament.

Per tal de dur a terme l'ajust cinètic d'un determinat sistema, en primer lloc, s'importa la matriu de dades a analitzar dins l'entorn de treball del programa. En la finestra principal (Figura 1.22.a) es determina el nombre de components del sistema i s'inicia el procés d'optimització. Posteriorment, es defineixen els paràmetres característics del model a ajustar (Figura 1.22.b) de forma que s'escriu el mecanisme del model a ajustar, les concentracions inicials de cadascun dels reactius presents i uns valors inicials per a les constants de velocitat a determinar. A la part inferior d'aquesta finestra es mostra una representació dels perfils de concentració obtinguts a partir del model, els valors de les concentracions inicials i els valors de les constants proposades. Finalment, els resultats es mostren a la finestra que es mostra a la Figura 1.22.c. A més, a la finestra de comandes de MATLAB® es mostraran els valors obtinguts de les constants de velocitat, de les concentracions finals de cada espècie i la informació referent a la qualitat de l'ajust com, per exemple, la variància explicada o l'arrel de l'error quadràtic mitjà.

a

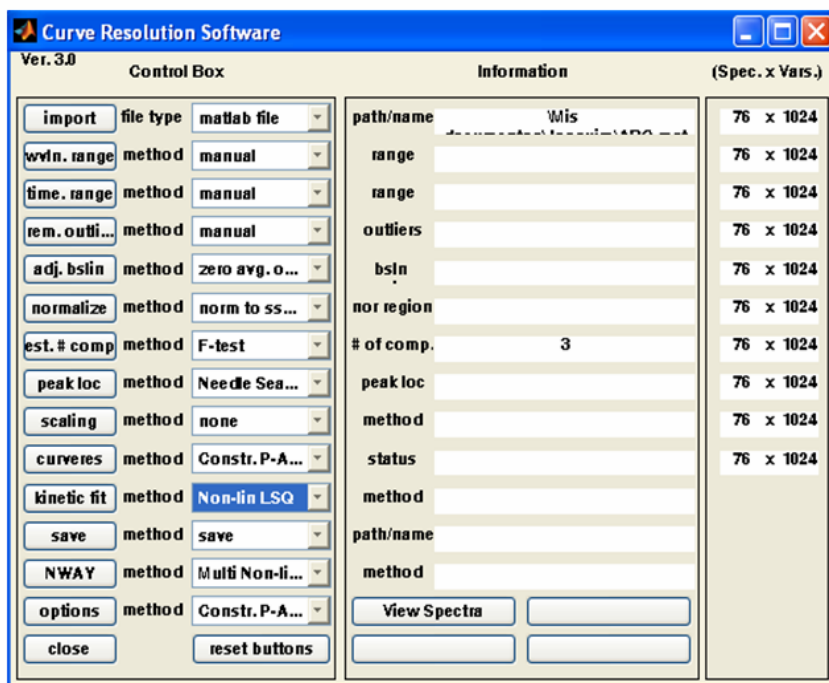
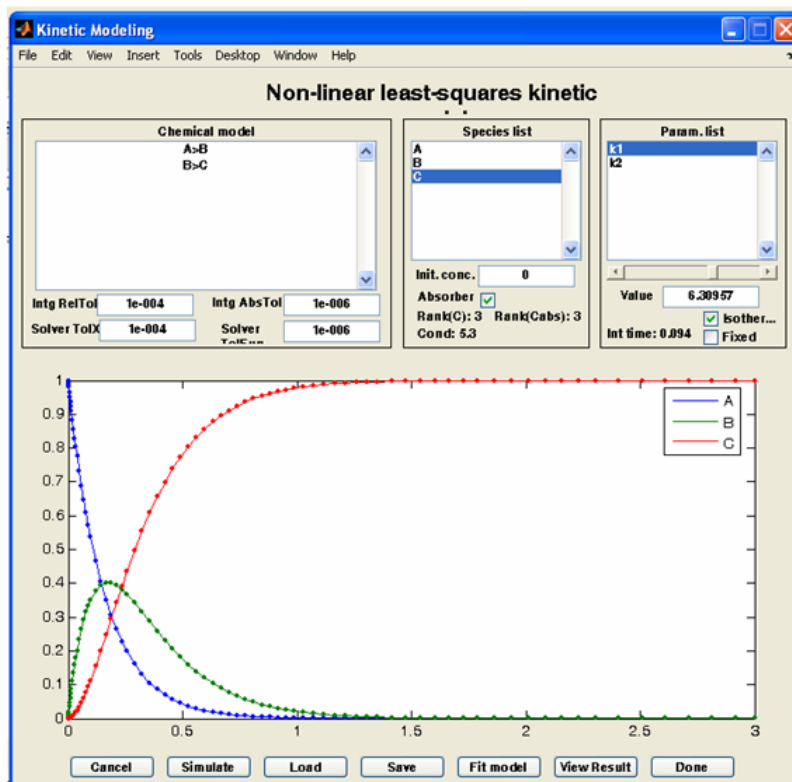


Figura 1.22. Finestres de la interfície gràfica en l'entorn MATLAB® GUIPRO. a) Finestra de selecció de la matriu de dades i del mètode d'optimització.

b



c

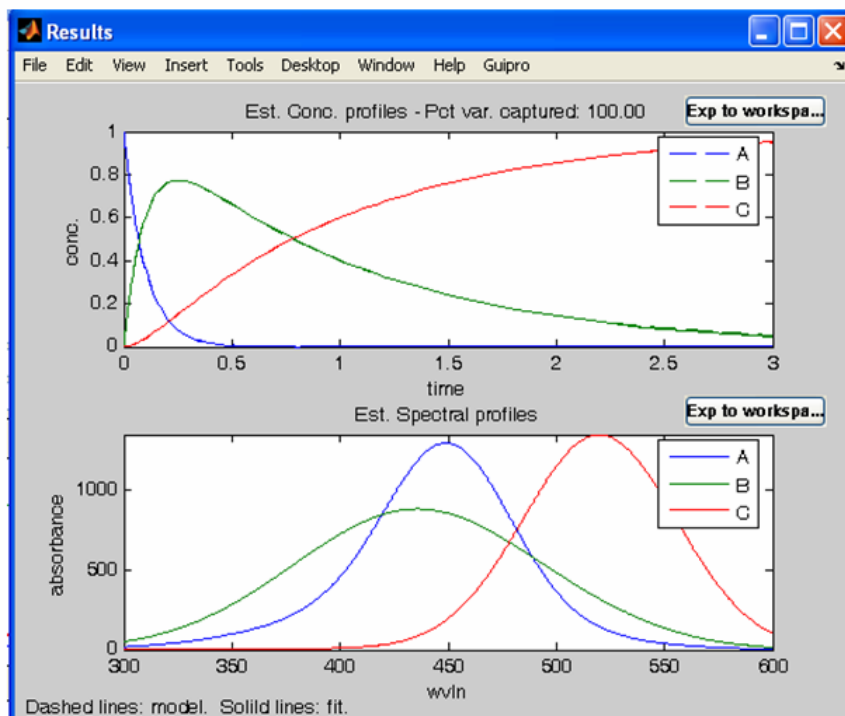


Figura 1.22. Finestres de la interfície gràfica en l'entorn MATLAB® GUIPRO. b) Finestra de definició del model, de les concentracions inicials i del valors inicial de les constants de velocitat, c) Finestra de resultats.

A més, GUIPRO també permet realitzar la resolució de sistemes mitjançant mètodes de modelització flexible com els algorismes de mínims quadrats alternats no negatius. Recentment, s'ha afegit també la possibilitat de dur a terme mínims quadrats alternats amb restriccions de no-negativitat, unimodalitat, sistema tancat i selectivitat. Cal tenir en compte que la imposició d'aquestes restriccions es pot dur a terme de forma rígida o flexible en el que es coneix com mínims quadrats alternats amb penalització a les restriccions (*Penalty constraints with Alternating Least Squares*, P-ALS). El programa també permet la determinació del nombre de components i de les estimacions inicials dels espectres o de les concentracions [114].

Aquest programa es troba disponible lliurement a l'adreça web http://personal.ecu.edu/gemperlinep/Research/gemper_software.html.

1.3.4. Limitacions dels mètodes de modelització rígida

De la mateixa manera que els mètodes de modelització flexible presenten el problema de les ambigüitats rotacionals, els mètodes de modelització rígida també presenten una sèrie de limitacions d'aquestes característiques que cal conèixer.

- Coneixement previ del model .

És necessari postular un model químic que descrigui de forma acurada el sistema estudiat. Si aquest coneixement no és disponible, aleshores s'hauran de dur a terme diversos assaigs per tal de determinar quin és el model més satisfactori mitjançant una estratègia prova-error. Per prendre aquesta decisió s'acostuma a tenir en compte diversos paràmetres relacionats amb la qualitat d'ajust de les dades experimentals com poden ser la variància experimental explicada o els errors associats a la determinació dels paràmetres. S'ha de vigilar no cometre errors de sobreajust ja que en intentar trobar el model amb el màxim ajust de les dades experimentals es pot incloure en aquest model variància deguda al soroll experimental.

- Estimacions inicials raonables

En els mètodes d'optimització, els valors de les estimacions inicials ens poden determinar si l'optimització tindrà èxit o no. Així, depenent de la qualitat d'aquestes estimacions inicials (per exemple, de les constants de velocitat en experiments cinètics) l'etapa d'optimització podrà convergir cap a l'òptim global ràpidament, divergir (amb la qual cosa s'haurà de repetir l'optimització amb noves estimacions inicials) o convergir cap a un mínim local amb la qual cosa s'agafarà com a solució correcta una solució no òptima.

- Presència d'ambigüitats quan s'estudien mecanismes complexos

En els casos on només es vol optimitzar un paràmetre, la superfície de resposta presenta un únic mínim [115, 116]. Aquest fet és, per exemple, vàlid únicament pels mecanismes més simples del tipus $A \xrightarrow{k_1} B$. En el cas de mecanismes més complexos es fa necessari optimitzar simultàniament més paràmetres i això pot provocar l'aparició de més d'un mínim global en la superfície de resposta corresponent, malgrat que únicament una solució tindrà sentit químic. Per exemple, en el cas de voler optimitzar dos constants de velocitat d'una reacció tipus $A \xrightarrow{k_1} B \xrightarrow{k_2} C$ hi haurà dos mínims a la superfície de resposta amb el mateix valor de suma de quadrats dels residuals.

- Distorsió dels models degut a la presència d'interferents

Una altra dificultat d'aquest tipus d'optimitzacions és la possible presència d'interferents desconeguts no inclosos en el model químic però que presenten senyal espectroscòpic. Aquests casos són habituals quan es volen analitzar processos químics en situacions no de laboratori, com és el cas de processos industrials, on la postulació d'un model químic incloent totes les espècies amb senyal espectroscòpic és impossible a la pràctica.

Per superar aquestes dificultats es poden fer servir els mètodes de modelització flexible i els mètodes de modelització híbrida, en els quals es conjuguen els avantatges dels mètodes de modelització flexible i els de modelització rígida [26, 40, 41]. La utilització d'aquests procediments és un tema d'investigació actual.

1.4. Bibliografia

- (1) Jurs, P. C.; Kowalski, B. R.; Isenhour, T. L. (1969) Computerized Learning Machines Applied to Chemical Problems - Molecular Formula Determination from Low Resolution Mass Spectrometry. *Analytical Chemistry*, **41**, 21-27.
- (2) Kowalski, B. R.; Jurs, P. C.; Isenhour, T. L.; Reilley, C. N. (1969) Computerized Learning Machines Applied to Chemical Problems - Interpretation of Infrared Spectrometry Data. *Analytical Chemistry*, **41**, 1945-1949.
- (3) Jurs, P. C.; Kowalski, B. R.; Isenhour, T. L.; Reilley, C. N. (1969) Computerized Learning Machines Applied to Chemical Problems. Molecular Structure Parameters from Low Resolution Mass Spectrometry. *Abstracts of Papers of the American Chemical Society*, 53-61.
- (4) Jurs, P. C.; Kowalski, B. R.; Isenhour, T. L.; Reilley, C. N. (1969) Computerized Learning Machines Applied to Chemical Problems - on Pattern Classification in Analytical Data Treatment. *Abstracts of Papers of the American Chemical Society*, 23-30.
- (5) Jurs, P. C.; Kowalski, B. R.; Isenhour, T. L.; Reilley, C. N. (1969) Computerized Learning Machines Applied to Chemical Problems - Investigation of Convergence Rate and Predictive Ability of Adaptive Binary Pattern Classifiers. *Analytical Chemistry*, **41**, 690-695.
- (6) Kowalski, B. R.; Jurs, P. C.; Isenhour, T. L.; Reilley, C. N. (1969) Computerized Learning Machines Applied to Chemical Problems - Multicategory Pattern Classification by Least Squares. *Analytical Chemistry*, **41**, 695-700.
- (7) Jurs, P. C.; Kowalski, B. R.; Isenhour, T. L.; Reilley, C. N. (1970) Computerized Learning Machines Applied to Chemical Problems - Molecular Structure Parameters from Low Resolution Mass Spectrometry. *Analytical Chemistry*, **42**, 1387-1394.
- (8) Wold, S. (1970) Computers and Studies of Kinetics. *Kemisk Tidskrift*, **82**, 16-17.
- (9) Wold, S.; Ahlberg, P. (1970) Evaluation of Activation Parameters for a First Order Reaction from One Kinetic Experiment - Theory, Numerical Methods and Computer Program. *Acta Chemica Scandinavica*, **24**, 618-632.
- (10) Wold, S. (1975) Analysis of Similarities and Dissimilarities between Chromatographic Liquid-Phases by Means of Pattern Cognition. *Journal of Chromatographic Science*, **13**, 525-532.
- (11) Wold, S. (1976) Pattern-Recognition by Means of Disjoint Principal Components Models. *Pattern Recognition*, **8**, 127-139.
- (12) Massart, D. L.; Declercq, H. (1974) Application of Numerical Taxonomy Techniques to Choice of Optimal Sets of Solvents in Thin-Layer Chromatography. *Analytical Chemistry*, **46**, 1988-1992.

- (13) Massart, D. L.; Lenders, P.; Lauwerey, M. (1974) Selection of Preferred Liquid-Phases after Classification by Numerical Taxonomy Techniques. *Journal of Chromatographic Science*, **12**, 617-625.
- (14) Wold, S. (1972) Spline Functions, a New Tool in Data-Analysis. *Kemisk Tidsskrift*, **84**, 34-37.
- (15) Lawton, W. H.; Sylvestre, E. A. (1971) Self Modeling Curve Resolution. *Technometrics*, **13**, 617-633.
- (16) Massart, D. L.; Buydens, L. M. C.; Vandegiste, B. G. M. (1997) *Handbook of Chemometrics and Qualimetrics*. 1st edn. ed. Elsevier, Amsterdam, The Netherlands.
- (17) Malinowski, E. R.; Howery, D. G. (2002) *Factor Analysis in Chemistry*. 3rd edn. ed. Wiley, New York, NY, US.
- (18) Wold, S.; Martens, H.; Wold, H. (1983) The Multivariate Calibration-Problem in Chemistry Solved by the PLS Method. *Lecture Notes in Mathematics*, **973**, 286-293.
- (19) Gampp, H.; Maeder, M.; Meyer, C. J.; Zuberbuhler, A. D. (1985) Calculation of Equilibrium-Constants from Multiwavelength Spectroscopic Data.3. Model-Free Analysis of Spectrophotometric and Electron-Spin-Resonance Titrations. *Talanta*, **32**, 1133-1139.
- (20) Tauler, R.; Casassas, E. (1989) Application of Principal Component Analysis to the Study of Multiple Equilibria Systems - Study of Copper(II) Salicylate Monoethanolamine, Diethanolamine and Triethanolamine Systems. *Analytica Chimica Acta*, **223**, 257-268.
- (21) Windig, W.; Markel, S. (1993) Simple-to-Use Interactive Self-Modelling Mixture Analysis of FTIR Microscopy Data. *Journal of Molecular Structure*, **292**, 161-170.
- (22) Bro, R. (1996) Multiway calibration. Multilinear PLS. *Journal of Chemometrics*, **10**, 47-61.
- (23) Bro, R. (1997) PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems*, **38**, 149-171.
- (24) Tauler, R.; Kowalski, B.; Fleming, S. (1993) Multivariate Curve Resolution Applied to Spectral Data from Multiple Runs of an Industrial-Process. *Analytical Chemistry*, **65**, 2040-2047.
- (25) Tauler, R.; Smilde, A.; Kowalski, B. (1995) Selectivity, Local Rank, 3-Way Data-Analysis and Ambiguity in Multivariate Curve Resolution. *Journal of Chemometrics*, **9**, 31-58.
- (26) de Juan, A.; Maeder, M.; Martinez, M.; Tauler, R. (2000) Combining hard- and soft-modelling to solve kinetic problems. *Chemometrics and Intelligent Laboratory Systems*, **54**, 123-141.

- (27) Tauler, R. (2001) Calculation of maximum and minimum band boundaries of feasible solutions for species profiles obtained by multivariate curve resolution. *Journal of Chemometrics*, **15**, 627-646.
- (28) Tauler, R. (1997) *Anàlisi de mesclades mitjançant Resolució Multivariant de Corbes*. 1era. ed. ed. Institut d'Estudis Catalans, Barcelona.
- (29) Lavine, B. K. (2000) Chemometrics. *Analytical Chemistry*, **72**, 91R-97R.
- (30) Lavine, B. K.; Workman, J. (2002) Chemometrics. *Analytical Chemistry*, **74**, 2763-2769.
- (31) Lavine, B.; Workman, J. J. (2004) Chemometrics. *Analytical Chemistry*, **76**, 3365-3371.
- (32) Skoog, D. A.; Leary, J. J. (1994) *Análisis Instrumental*. 4ta Ed. ed. McGraw-Hill, Madrid.
- (33) Evans, J. N. S. (1995) *Biomolecular Nmr Spectroscopy*. 1era. ed. ed. Oxford University Press, Oxford, UK.
- (34) Tauler, R.; Marques, I.; Casassas, E. (1998) Multivariate curve resolution applied to three-way trilinear data: Study of a spectrofluorimetric acid-base titration of salicylic acid at three excitation wavelengths. *Journal of Chemometrics*, **12**, 55-75.
- (35) Lawson, C. L.; Hanson, R. J. (1974) *Solving Least-squares Problems*. 1st edn. ed. Prentice-Hall, Englewood Cliffs, NJ, US.
- (36) Gans, P. (1992) *Data fitting in the Chemical Sciences: By the Method of Least Squares*. 1st edn. ed. Wiley, New York, NY, US.
- (37) Motulsky, H.; Christopoulos, A. (2004) *Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting*. 1era. ed. ed. Oxford University Press, Oxford, UK.
- (38) Bezemer, E.; Rutan, S. C. (2001) Multivariate curve resolution with non-linear fitting of kinetic profiles. *Chemometrics and Intelligent Laboratory Systems*, **59**, 19-31.
- (39) Sanchez-Ponce, R.; Rutan, S. C. (2005) Steady state kinetic model constraint for Multivariate Curve Resolution-Alternating Least Squares analysis. *Chemometrics and Intelligent Laboratory Systems*, **77**, 50-58.
- (40) Diewok, J.; de Juan, A.; Maeder, M.; Tauler, R.; Lendl, B. (2003) Application of a combination of hard and soft modeling for equilibrium systems to the quantitative analysis of pH-modulated mixture samples. *Analytical Chemistry*, **75**, 641-647.
- (41) de Juan, A.; Maeder, M.; Martinez, M.; Tauler, R. (2001) Application of a novel resolution approach combining soft- and hard-modelling features to investigate temperature-dependent kinetic processes. *Analytica Chimica Acta*, **442**, 337-350.

- (42) Hamilton, J. C.; Gemperline, P. J. (1990) Mixture analysis using factor analysis. II: Self-modelling curve resolution. *Journal of Chemometrics*, **4**, 1-13.
- (43) Borgen, O. S.; Kowalski, B. R. (1985) An Extension of the Multivariate Component-Resolution Method to 3 Components. *Analytica Chimica Acta*, **174**, 1-26.
- (44) Henry, R. C.; Kim, B. M. (1990) Extension of Self-Modelling Curve Resolution to Mixtures of More Than 3 Components.1. Finding the Basic Feasible Region. *Chemometrics and Intelligent Laboratory Systems*, **8**, 205-216.
- (45) Vandeginste, B. G. M.; Derks, W.; Kateman, G. (1985) Multicomponent Self-Modelling Curve Resolution in High-Performance Liquid-Chromatography by Iterative Target Transformation Analysis. *Analytica Chimica Acta*, **173**, 253-264.
- (46) Roscoe, B. A.; Hopke, P. K. (1981) Comparison of Weighted and Unweighted Target Transformation Rotations in Factor-Analysis. *Computers & Chemistry*, **5**, 1-7.
- (47) Gemperline, P. J. (1986) Target Transformation Factor-Analysis with Linear Inequality Constraints Applied to Spectroscopic Chromatographic Data. *Analytical Chemistry*, **58**, 2656-2663.
- (48) Liang, Y. Z.; Kvalheim, O. M.; Keller, H. R.; Massart, D. L.; Kiechle, P.; Erni, F. (1992) Heuristic Evolving Latent Projections - Resolving 2-Way Multicomponent Data.2. Detection and Resolution of Minor Constituents. *Analytical Chemistry*, **64**, 946-953.
- (49) Kvalheim, O. M.; Liang, Y. Z. (1992) Heuristic Evolving Latent Projections - Resolving 2-Way Multicomponent Data.1. Selectivity, Latent-Projective Graph, Datascope, Local Rank, and Unique Resolution. *Analytical Chemistry*, **64**, 936-946.
- (50) Tauler, R. (1995) Multivariate curve resolution applied to second order data. *Chemometrics and Intelligent Laboratory Systems*, **30**, 133-146.
- (51) Tauler, R.; Izquierdo-Ridorsa, A.; Casassas, E. (1993) Simultaneous Analysis of Several Spectroscopic Titrations with Self-Modelling Curve Resolution. *Chemometrics and Intelligent Laboratory Systems*, **18**, 293-300.
- (52) Tauler, R.; Casassas, E.; Izquierdoridorsa, A. (1991) Self-Modelling Curve Resolution in Studies of Spectrometric Titrations of Multi-Equilibria Systems by Factor-Analysis. *Analytica Chimica Acta*, **248**, 447-458.
- (53) Golub, G. H.; Van Loan, C. F. (1996) *Matrix Computations*. 3rd. edn. ed. John Hopkins University Press, Baltimore, MD, US.
- (54) Carey, R. N.; Wold, S.; Westgard, J. O. (1975) Principal Component Analysis - Alternative to Referee Methods in Method Comparison Studies. *Analytical Chemistry*, **47**, 1824-1829.

- (55) Wold, S.; Sjöström, M. (1972) Statistical-Analysis of Hammett Equation.1. Methods and Model Calculations. *Chemica Scripta*, **2**, 49-55.
- (56) Maeder, M. (1987) Evolving Factor-Analysis for the Resolution of Overlapping Chromatographic Peaks. *Analytical Chemistry*, **59**, 527-530.
- (57) Gampp, H.; Maeder, M.; Meyer, C. J.; Zuberbühler, A. D. (1986) Calculation of Equilibrium-Constants from Multiwavelength Spectroscopic Data.4. Model-Free Least-Squares Refinement by Use of Evolving Factor-Analysis. *Talanta*, **33**, 943-951.
- (58) Windig, W.; Guilment, J. (1991) Interactive Self-Modelling Mixture Analysis. *Analytical Chemistry*, **63**, 1425-1432.
- (59) Mason, C.; Maeder, M.; Whitson, A. (2001) Resolving factor analysis. *Analytical Chemistry*, **73**, 1587-1594.
- (60) Malinowski, E. R. (1996) Automatic window factor analysis - A more efficient method for determining concentration profiles from evolutionary spectra. *Journal of Chemometrics*, **10**, 273-279.
- (61) Malinowski, E. R. (1992) Window Factor-Analysis - Theoretical Derivation and Application to Flow-Injection Analysis Data. *Journal of Chemometrics*, **6**, 29-40.
- (62) de Juan, A.; Navea, S.; Diewok, J.; Tauler, R. (2004) Local rank exploratory analysis of evolving rank-deficient systems. *Chemometrics and Intelligent Laboratory Systems*, **70**, 11-21.
- (63) de Juan, A.; Maeder, M.; Hanczewicz, T.; Tauler, R. (2005) Local rank analysis for exploratory spectroscopic image analysis. Fixed Size Image Window-Evolving Factor Analysis. *Chemometrics and Intelligent Laboratory Systems*, **77**, 64-74.
- (64) Whitson, A. C.; Maeder, M. (2001) Exhaustive evolving factor analysis (E-EFA). *Journal of Chemometrics*, **15**, 475-484.
- (65) Windig, W. (1994) The Use of 2Nd-Derivative Spectra for Pure-Variable Based Self-Modelling Mixture Analysis Techniques. *Chemometrics and Intelligent Laboratory Systems*, **23**, 71-86.
- (66) Windig, W.; Gallagher, N. B.; Shaver, J. M.; Wise, B. M. (2005) A new approach for interactive self-modelling mixture analysis. *Chemometrics and Intelligent Laboratory Systems*, **77**, 85-96.
- (67) Tauler, R.; Barcelo, D. (1993) Multivariate Curve Resolution Applied to Liquid-Chromatography Diode-Array Detection. *Trac-Trends in Analytical Chemistry*, **12**, 319-327.
- (68) Gargallo, R.; Tauler, R.; Cuesta Sanchez, F.; Massart, D. L. (1996) Validation of alternating least-squares multivariate curve resolution for chromatographic resolution and quantitation. *Trac-Trends in Analytical Chemistry*, **15**, 279-286.

- (69) Bro, R.; DeJong, S. (1997) A fast non-negativity-constrained least squares algorithm. *Journal of Chemometrics*, **11**, 393-401.
- (70) Bro, R.; Sidiropoulos, N. D. (1998) Least squares algorithms under unimodality and non-negativity constraints. *Journal of Chemometrics*, **12**, 223-247.
- (71) de Juan, A.; Van der Heyden, Y.; Tauler, R.; Massart, D. L. (1997) Assessment of new constraints applied to the alternating least squares method. *Analytica Chimica Acta*, **346**, 307-318.
- (72) Saurina, J.; Hernandez Cassou, S.; Tauler, R. (1995) Multivariate Curve Resolution Applied to Continuous-Flow Spectrophotometric Titrations - Reaction between Amino-Acids and 1,2-Naphthoquinone-4-Sulfonic Acid. *Analytical Chemistry*, **67**, 3722-3726.
- (73) Van Benthem, M. H.; Keenan, M. R.; Haaland, D. M. (2002) Application of equality constraints on variables during alternating least squares procedures. *Journal of Chemometrics*, **16**, 613-622.
- (74) de Juan, A.; Tauler, R. (2003) Chemometrics applied to unravel multicomponent processes and mixtures - Revisiting latest trends in multivariate resolution. *Analytica Chimica Acta*, **500**, 195-210.
- (75) Diaz-Cruz, M. S.; Diaz-Cruz, J. M.; Mendieta, J.; Tauler, R.; Esteban, M. (2000) Soft- and hard-modelling approaches for the determination of stability constants of metal-peptide systems by voltammetry. *Analytical Biochemistry*, **279**, 189-201.
- (76) Saurina, J.; Hernandez-Cassou, S.; Tauler, R.; Izquierdo-Ridorsa, A. (1998) Multivariate resolution of rank-deficient spectrophotometric data from first-order kinetic decomposition reactions. *Journal of Chemometrics*, **12**, 183-203.
- (77) Izquierdo-Ridorsa, A.; Saurina, J.; Hernandez Cassou, S.; Tauler, R. (1997) Second-order multivariate curve resolution applied to rank-deficient data obtained from acid-base spectrophotometric titrations of mixtures of nucleic bases. *Chemometrics and Intelligent Laboratory Systems*, **38**, 183-196.
- (78) Amrhein, M.; Srinivasan, B.; Bonvin, D.; Schumacher, M. M. (1996) On the rank deficiency and rank augmentation of the spectral measurement matrix. *Chemometrics and Intelligent Laboratory Systems*, **33**, 17-33.
- (79) Casassas, E.; Marques, I.; Tauler, R. (1995) Study of Acid-Base Properties of Fulvic-Acids Using Fluorescence Spectrometry and Multivariate Curve Resolution Methods. *Analytica Chimica Acta*, **310**, 473-484.
- (80) Saurina, J.; Hernandez Cassou, S.; Tauler, R. (1997) Multivariate curve resolution and trilinear decomposition methods in the analysis of stopped-flow kinetic data for binary amino acid mixtures. *Analytical Chemistry*, **69**, 2329-2336.
- (81) Manne, R. (1995) On the Resolution Problem in Hyphenated Chromatography. *Chemometrics and Intelligent Laboratory Systems*, **27**, 89-94.

- (82) Borgen, O. S.; Davidsen, N.; Zhu, M. Y.; Oyen, O. (1986) The Multivariate N-Component Resolution Problem with Minimum Assumptions. *Mikrochimica Acta*, **2**, 63-73.
- (83) Sylvestre E.A.; Lawton, W. H.; Maggio, M. S. (1974) Curve Resolution Using a Postulated Chemical-Reaction. *Technometrics*, **16**, 353-368.
- (84) Gemperline, P. J. (1999) Computation of the range of feasible solutions in self-modelling curve resolution algorithms. *Analytical Chemistry*, **71**, 5398-5404.
- (85) Bijlsma, S.; Smilde, A. K. (2000) Estimating reaction rate constants from a two-step reaction: a comparison between two-way and three-way methods. *Journal of Chemometrics*, **14**, 541-560.
- (86) Faber, N. M. (2001) Comment on a recently proposed resampling method. *Journal of Chemometrics*, **15**, 169-188.
- (87) Bevington, P. R.; Robinson, D. K. (2002) *Data Reduction and Error Analysis for the Physical Sciences*. ed. McGraw-Hill, New York, NY, USA.
- (88) Good, P. I. (2001) *Resampling Methods. A practical guide to data analysis*. ed. Birkhäuser Boston, New York, NY, US.
- (89) Leger, M. N.; Wentzell, P. D. (2004) Maximum likelihood principal components regression on wavelet-compressed data. *Applied Spectroscopy*, **58**, 855-862.
- (90) Vega-Montoto, L.; Wentzell, P. D. (2003) Maximum likelihood parallel factor analysis (MLPARAFAC). *Journal of Chemometrics*, **17**, 237-253.
- (91) Quenouille, M. H. (1949) Approximate Tests of Correlation in Time-Series-3. *Proceedings of the Cambridge Philosophical Society*, **45**, 483-484.
- (92) Quenouille, M. H. (1956) Notes on Bias in Estimation. *Biometrika*, **43**, 353-360.
- (93) Tukey, J. W. (1958) Bias and Confidence in Not-Quite Large Samples. *Annals of Mathematical Statistics*, **29**, 614-614.
- (94) Wold, S.; Sjostrom, M.; Eriksson, L. (2001) PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, **58**, 109-130.
- (95) Olivieri, A. C.; Faber, N. K. M. (2004) Standard error of prediction in parallel factor analysis of three-way data. *Chemometrics and Intelligent Laboratory Systems*, **70**, 75-82.
- (96) Riu, J.; Bro, R. (2003) Jack-knife technique for outlier detection and estimation of standard errors in PARAFAC models. *Chemometrics and Intelligent Laboratory Systems*, **65**, 35-49.
- (97) Maeder, M.; Zuberbuhler, A. D. (1990) Nonlinear Least-Squares Fitting of Multivariate Absorption Data. *Analytical Chemistry*, **62**, 2220-2224.
- (98) Gans, P. (1976) Numerical-Methods for Data-Fitting Problems. *Coordination Chemistry Reviews*, **19**, 99-124.

- (99) Gampp, H.; Maeder, M.; Meyer, C. J.; Zuberbuhler, A. D. (1985) Calculation of Equilibrium-Constants from Multiwavelength Spectroscopic Data.1. Mathematical Considerations. *Talanta*, **32**, 95-101.
- (100) Gampp, H.; Maeder, M.; Meyer, C. J.; Zuberbuhler, A. D. (1985) Calculation of Equilibrium-Constants from Multiwavelength Spectroscopic Data.2. Specfit - 2 User-Friendly Programs in Basic and Standard Fortran-77. *Talanta*, **32**, 257-264.
- (101) Saeber, G. A. F.; Wild, C. J. (1989) *Nonlinear Regression*. 1era. ed. ed. Wiley, New York, NY, USA.
- (102) Long, D. E. (1969) Simplex Optimization of Response from Chemical Systems. *Analytica Chimica Acta*, **46**, 193-206.
- (103) Maeder, M.; Neuhold, Y. M.; Puxty, G. (2004) Application of a genetic algorithm: near optimal estimation of the rate and equilibrium constants of complex reaction mechanisms. *Chemometrics and Intelligent Laboratory Systems*, **70**, 193-203.
- (104) Hibbert, D. B. (1993) Genetic Algorithms in Chemistry. *Chemometrics and Intelligent Laboratory Systems*, **19**, 277-293.
- (105) Marquardt, D. W. (1963) An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics*, **11**, 431-441.
- (106) Levenberg, K. (1944) A method for the solution of certain non-linear problems in least squares. *Quarterly Journal of Applied Mathematics*, **II(2)**, 164-168.
- (107) Dyson, R. M.; Kaderli, S.; Lawrance, G. A.; Maeder, M.; Zunderbuhler, A. D. (1997) Second order global analysis: the evaluation of series of spectrophotometric titrations for improved determination of equilibrium constants. *Analytica Chimica Acta*, **353**, 381-393.
- (108) Bugnon, P.; Chottard, J. C.; Jestin, J. L.; Jung, B.; Laurency, G.; Maeder, M.; Merbach, A. E.; Zuberbuhler, A. D. (1994) 2Nd-Order Globalization for the Determination of Activation Parameters in Kinetics. *Analytica Chimica Acta*, **298**, 193-201.
- (109) Leggett, D. J. (1977) Machine Computation of Equilibrium Concentrations - Some Practical Considerations. *Talanta*, **24**, 535-542.
- (110) Leggett, D. J. (1977) Numerical-Analysis of Multicomponent Spectra. *Analytical Chemistry*, **49**, 276-281.
- (111) Leggett, D. J.; McBryde, W. A. E. (1975) General Computer-Program for Computation of Stability-Constants from Absorbance Data. *Analytical Chemistry*, **47**, 1065-1070.
- (112) LEVINE, I. N. (2004) *Fisicoquímica*. 5a. ed. ed. McGraw-Hill/Interamericana de España, Madrid, Espanya.

- (113) Stoer, J.; Bulirsch, R. (1993.) *Introduction to numerical analysis*. 2a ed. ed. Springer, New York, NY, USA.
- (114) Gemperline, P. J.; Cash, E. (2003) Advantages of soft versus hard constraints in self-modelling curve resolution problems. Alternating least squares with penalty functions. *Analytical Chemistry*, **75**, 4236-4243.
- (115) Andraos, J. (1999) The problem of distinguishability of rate constants in the two-step consecutive sequence $A \rightarrow B \rightarrow C$. *Canadian Journal of Chemistry-Revue Canadienne De Chimie*, **77**, 565-576.
- (116) Vajda, S.; Rabitz, H. (1994) Identifiability and Distinguishability of General Reaction Systems. *Journal of Physical Chemistry*, **98**, 5265-5271.