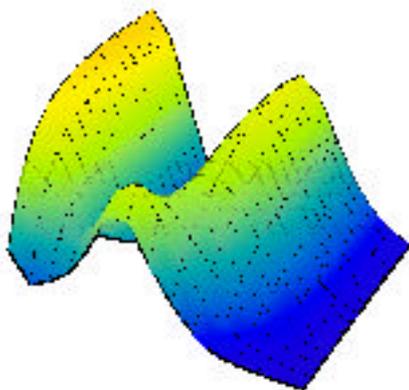




Universitat
Autònoma
de Barcelona

**APLICACIÓN DE TÉCNICAS QUIMIOMÉTRICAS
PARA EL DESARROLLO DE NUEVOS MÉTODOS
CINÉTICO-ESPECTROFOTOMÉTRICOS
DE ANÁLISIS**



MARTA PORCEL



**Universitat
Autònoma
de Barcelona**

**APLICACIÓN DE TÉCNICAS QUIMIOMÉTRICAS
PARA EL DESARROLLO DE NUEVOS MÉTODOS
CINÉTICO-ESPECTROFOTOMÉTRICOS
DE ANÁLISIS**

Memoria presentada por **Marta Porcel García**
para optar al grado de Doctora en Ciencias
Químicas

Bellaterra, Julio 2001

Departament de Química
Unitat de Química Analítica
Edifici Cn
08193 Bellaterra (Barcelona), Spain



**Universitat
Autònoma
de Barcelona**

Dra. Hortensia Iturriaga Martínez, Catedrática de Química Analítica de la Facultad de Ciencias de la Universidad Autónoma de Barcelona, y

Dr. Jordi Coello Bonilla, Profesor Titular de Química Analítica de la Facultad de Ciencias de la Universidad Autónoma de Barcelona,

CERTIFICAN: Que el presente trabajo de investigación titulado “**Aplicación de técnicas quimiométricas para el desarrollo de nuevos métodos cinético-espectrofotométricos de análisis**”, que constituye la Memoria presentada por Marta Porcel García para aspirar al grado de doctor en Ciencias Químicas, ha sido realizado bajo nuestra dirección y reuniendo, según nuestro parecer, las condiciones exigidas para este tipo de trabajo.

Y para que quede constancia, expiden y firman el presente certificado en Bellaterra, a 2 de julio de 2001.

Dra. H. Iturriaga

Dr. J. Coello

Todo trabajo realizado con placer no es completo si no se ofrece en dedicatoria. Y las dedicatorias siempre van dirigidas hacia aquellos que han hecho posible en mayor o en menor medida dicho trabajo. Por tanto, esta memoria va especialmente dedicada a todos los que han colaborado y me han apoyado, porque sin su intervenci3n no hubiese logrado esta meta.

ÍNDICE

OBJETIVOS	1
INTRODUCCIÓN	3
1. CALIBRACIÓN MULTIVARIABLE	5
1.1 INTRODUCCIÓN	5
1.1.1 Clasificación de los métodos de calibración	6
1.1.2 Construcción de modelos de calibración multivariable	11
1.1.3 Pretratamiento de los datos registrados	12
1.1.4 Evaluación de la capacidad predictiva del modelo	13
1.2 REGRESIÓN LINEAL MÚLTIPLE	15
1.2.1 Regresión lineal múltiple clásica	15
1.2.2 Regresión lineal múltiple inversa	17
1.3 MÉTODOS LINEALES PARA EL TRATAMIENTO DE DATOS DE PRIMER ORDEN BASADOS EN LA REDUCCIÓN DE VARIABLES	18
1.3.1 Consideraciones previas	18
1.3.2 Regresión en componentes principales	20
1.3.2.1 Etapa de análisis en componentes principales	20
1.3.2.2 Etapa de regresión en componentes principales	22

1.3.3	Regresión parcial por mínimos cuadrados	24
1.3.4	Regresión continua	25
1.3.5	Elección del número óptimo de componentes principales o factores	26
1.4	MÉTODOS LINEALES PARA EL TRATAMIENTO DE DATOS DE SEGUNDO ORDEN BASADOS EN LA REDUCCIÓN DE VARIABLES	28
1.4.1	Introducción	28
1.4.2	<i>Parallel Factor Analysis</i> , PARAFAC	28
1.4.2.1	Introducción	28
1.4.2.2	El modelo PARAFAC	30
1.4.2.3	Algoritmo: <i>Alternating least squares</i>	31
1.4.2.4	Criterio de parada	33
1.4.2.5	Forzando la solución	33
1.4.2.6	Preprocesado de los datos	33
1.4.2.7	Determinación del número de componentes o factores	34
1.4.2.8	Soluciones degeneradas	35
1.4.3	PLS Multi-lineal (Multi-linear PLS, N-PLS)	36
1.4.3.1	Introducción	36
1.4.3.2	El modelo N-PLS	37
1.4.3.3	Algoritmo	38
1.5	MÉTODOS NO LINEALES: REDES NEURONALES ARTIFICIALES (ANN)	39
1.5.1	Introducción	39
1.5.2	Redes Multi-capas Perceptron (MLP, Multi-Layer Perceptron)	41
1.5.2.1	Propagación de una MLP	42
1.5.2.2	Aprendizaje por retropropagación	44
1.5.2.3	Consideraciones prácticas para la construcción de una red	47
1.5.2.3.1	Muestras	47

1.5.2.3.2	Topología de la red: diseño y optimización	49
1.5.3	Entrenamiento y evaluación de una red neuronal	51
2.	MÉTODOS CINÉTICOS DE ANÁLISIS	55
2.1	INTRODUCCIÓN	55
2.2	CONCEPTOS BÁSICOS	57
2.2.1	Cálculo de las constantes de velocidad	59
2.2.2	Cálculo de la energía de activación, E_a	59
2.2.3	Cálculo de los órdenes parciales de reacción	60
2.3	INSTRUMENTACIÓN	60
2.3.1	Introducción	60
2.3.2	Sistemas de mezcla	61
2.3.2.1	Mezcla manual-agitación magnética	62
2.3.2.2	Sistema de mezcla de flujo interrumpido (<i>stopped-flow mixing</i>)	62
2.3.3	Seguimiento de la reacción	65
2.3.3.1	Absorción de luz y dicroísmo circular (DC)	66
2.4	CLASIFICACIÓN DE LOS MÉTODOS CINÉTICOS	69
2.4.1	Introducción	69
2.4.2	Métodos cinéticos diferenciales clásicos	70
2.4.3	Métodos cinéticos diferenciales modernos	72
2.5	TENDENCIA DE LOS MÉTODOS CINÉTICOS	73
3.	RESOLUCIÓN DE MEZCLAS MEDIANTE PROCEDIMIENTOS CINÉTICOS Y CALIBRACIÓN MULTIVARIABLE	77
3.1	INTRODUCCIÓN	77
3.2	VISIÓN GLOBAL	78
3.3	TRANSPORTE Y MEZCLA DE REACTIVOS	82

3.4	SISTEMA DE DETECCIÓN	83
3.5	TRATAMIENTO NUMÉRICO DE LOS DATOS	84
3.5.1	Tratamiento multivariable de los métodos diferenciales clásicos	84
3.5.2	Calibración con datos de primer orden	85
3.5.3	Calibración con datos de segundo orden	88
4.	REFERENCIAS	95
	RESULTADOS Y DISCUSIÓN	101
1.	INTRODUCCIÓN	103
2.	DETERMINACIÓN SIMULTÁNEA DE METANOL Y ETANOL	105
2.1	Introducción	105
2.2	Sistema químico	105
2.3	Cuantificación simultánea de metanol y etanol	107
3.	DETERMINACIÓN SIMULTÁNEA DE LOS ENANTIÓMEROS DE LA 1-FENILETILAMINA	109
3.1	Introducción	109
3.2	Sistema químico	110
3.3	Resolución de mezclas de enantiómeros	111
4.	EVALUACIÓN DE LOS MÉTODOS DE CALIBRACIÓN MULTIVARIABLE, BI Y TRIDIMENSIONALES, EN ANÁLISIS CINÉTICO DIFERENCIAL	113
4.1	Introducción	113
4.2	Simulaciones	114
4.3	Sistema químico	116
4.4	Cuantificación de mezclas de Co-Ni y de Co-Ni-Ga	116

5. SELECCIÓN DE LOS INTERVALOS DE LONGITUDES DE ONDA Y TIEMPOS PARA EL CALIBRADO DE UN SISTEMA CINÉTICO DE DOS COMPONENTES	118
5.1 Introducción	118
5.2 Simulaciones	118
5.3 Sistema químico	120
5.4 Cuantificación simultánea de difilina y proxifilina	122
6. DETERMINACIÓN SIMULTÁNEA DE METILXANTINAS EN UN FÁRMACO	123
6.1 Introducción	123
6.2 Sistema químico	123
6.3 Cuantificación simultánea de teofilina, difilina y proxifilina	124
CONCLUSIONES	127
ANEXOS	130
LISTA DE ABREVIATURAS	

OBJETIVOS

El objetivo principal de la memoria es el desarrollo de métodos de análisis cinético diferencial y su aplicación a la determinación de mezclas de compuestos con una naturaleza química muy similar. Pueden ser compuestos con los mismos grupos funcionales o incluso enantiómeros.

Existen otros objetivos de segundo nivel, pero necesarios y complementarios al principal. Por una parte, la profundización en el análisis e interpretación de la información que proporcionan los modernos sistemas de calibración aplicados a sistemas dinámicos; para ello se utilizan procedimientos de calibración multivariable, incluyendo redes neuronales y métodos tridimensionales. Por otra, el diseño y optimización de los procedimientos analíticos cinéticos basados en la variación con el tiempo de una señal multivariable.

La determinación de compuestos de estructura química semejante se abordará, tanto a partir del uso de enzimas de grupo, no específicas, como por la reacción con un agente derivatizante adecuado, buscando las condiciones experimentales para la diferenciación cinética-espectral óptima de los componentes de las mezclas. La resolución de mezclas de enantiómeros se abordará a partir de su diferente velocidad de reacción con reactivos quirales.

Entre las técnicas de calibración multivariable, se estudiarán: la regresión lineal múltiple (MLR), regresión en componentes principales (PCR), regresión parcial por mínimos cuadrados (PLS), regresión continua (CR), redes neuronales artificiales (ANN), regresión parcial por mínimos cuadrados multidimensional (N-PLS) y PARAFAC. Se aplicarán a los sistemas químicos reales estudiados o a las condiciones experimentales simuladas por ordenador para extraer conclusiones sobre tendencias generales de los métodos.

La consecución de estos objetivos se ve reflejada en los trabajos que se mencionan a continuación y que aparecen en los anexos:

1. *“Simultaneous enzymatic spectrophotometric determination of ethanol and methanol by use of artificial neural networks for calibration.”*
M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, M. Porcel
Analytica Chimica Acta, **398**, 83, **1999**.
2. *“Use of circular dichroism and artificial neural networks for the kinetic-spectrophotometric resolution of enantiomers.”*
M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, M. Porcel
Analytica Chimica Acta, **431**, 115, **2001**.
3. *“Evaluation of classical and three-way multivariate calibration procedures in kinetic-spectrophotometric analysis.”*
S. R. Crouch, J. Coello, S. Maspoch, M. Porcel
Analytica Chimica Acta, **424**, 115, **2000**.
4. *“Multi component kinetic-spectrophotometric analysis. Selection of wavelength and time ranges.”*
H. Iturriaga, J. Coello, S. Maspoch, M. Porcel
The Analyst, **126**, 1135, **2001**
5. *“Kinetic-spectrophotometric determination of theophylline, dyphylline and proxyphylline in a pharmaceutical preparation by use of partial least squares regression.”*

INTRODUCCIÓN

1. CALIBRACIÓN MULTIVARIABLE.
2. MÉTODOS CINÉTICOS DE ANÁLISIS.
3. RESOLUCIÓN DE MEZCLAS MEDIANTE PROCEDIMIENTOS CINÉTICOS Y CALIBRACIÓN MULTIVARIABLE.
4. REFERENCIAS.

1. CALIBRACIÓN MULTIVARIABLE

1.1 INTRODUCCIÓN.

El análisis de muestras que contienen mezclas de compuestos químicos constituye uno de los retos más importantes de la química analítica. La dificultad aumenta cuando se pretende determinar simultáneamente estos compuestos a bajas concentraciones o en presencia de otras sustancias que tienen estructuras y propiedades similares, las cuales pueden interferir en su determinación. El análisis cuantitativo de muestras complejas requiere utilizar un conjunto de muestras de referencia que tenga las mismas características que las muestras a analizar y de algún método instrumental que mida alguna propiedad característica de los analitos. En la utilización de un sistema de detección para las especies a analizar nos podemos encontrar con problemas debido a interferencias, a efectos matriz, ruido de fondo, etc., que precisan una especial atención.

El método tradicionalmente utilizado para resolver la falta de selectividad en muestras complejas, consiste en la separación previa de los compuestos químicos puros antes de analizarlos. Si la separación ha sido completa (ej. mediante cromatografía) tan solo es necesaria una medida univariable para tener la información cuantitativa deseada. No obstante, ésta presenta un coste elevado, ya que se necesita disponer de una instrumentación analítica sofisticada e invertir bastante tiempo para optimizar las condiciones de separación. Como alternativa, se han desarrollado los métodos quimiométricos, los cuales permiten extraer la información requerida sobre los componentes de interés de las muestras a partir del tratamiento de datos, sin necesidad de su separación previa.

La quimiometría es una disciplina ampliamente utilizada en el área de la química, nacida

aproximadamente al final de los años 70, y que, actualmente, está siendo ampliamente utilizada en diversos campos, especialmente en química analítica. El desarrollo rápido de la quimiometría en química analítica se debe fundamentalmente a la utilización masiva de los ordenadores acoplados a la moderna instrumentación. Se ha pasado de hacer medidas puntuales a una longitud de onda (univariables), a medidas espectrales (multivariables) con gran contenido de información y, por tanto, los métodos tradicionales de análisis y procesado de datos resultan totalmente insuficientes. Por tanto, uno de los campos de trabajo fundamental de la quimiometría es el desarrollo y aplicación de los nuevos métodos de calibración multivariable, que permiten extraer la máxima información química del análisis de muestras a partir de señales o respuestas instrumentales con muy poca selectividad.

En este capítulo se van a describir algunos métodos de calibración multivariable, haciendo especial énfasis en aquellos que han sido utilizados en esta memoria.

1.1.1 Clasificación de los métodos de calibración.

Los métodos de calibración pueden tener una base totalmente empírica o bien estar soportados por una base teórica que explica el fenómeno, físico o químico, responsable de la señal analítica (p.ej. la ley de Lambert-Beer en espectrofotometría de absorción). Se han propuesto diversas clasificaciones de los métodos de calibración. Una de ellas es la presentada por Booksh y Kowalski [Booksh, 1994] en su teoría de química analítica; los autores clasifican los métodos de calibración desde un punto de vista matemático, en función de la dimensión de los datos disponibles por muestra, los cuales han sido generados por instrumentos de diferente dimensión. Así, los datos pueden ser clasificados como:

(a) Datos univariables

- *Datos escalares. Instrumentos de orden cero.*

El tipo más sencillo de datos en química es de naturaleza univariable, es decir, medidas en las que únicamente se obtiene un valor numérico, un escalar. Esta situación se da en muchos casos, por ejemplo, en las medidas potenciométricas de pH o en las medidas espectrofotométricas realizadas a una sola longitud de onda. El método matemático utilizado en la extracción de la información analítica es la estadística univariable y la regresión lineal univariable. La aplicación de estos métodos supone que la respuesta instrumental es

consecuencia de la presencia de un solo componente químico, es decir, que no hay interferencias ni contribuciones desconocidas al ruido de fondo en la señal instrumental. La relación de la señal con la concentración del analito se reduce, generalmente, a una línea recta, siendo necesario realizar una separación previa de los componentes interferentes y eliminar el ruido de fondo antes realizar la calibración. Además, no es posible el análisis de multicomponentes o la determinación simultánea de los diferentes analitos que contiene la muestra.

(b) Datos multivariados.

- *Datos vectoriales. Instrumentos de primer orden.*

El instrumento proporciona un vector de datos (ej. espectro) al analizar cada muestra. Los métodos matemáticos utilizados son los derivados del análisis multivariable de los datos y de la estadística multivariable. En particular, cabe mencionar el desarrollado por los métodos llamados de calibración multivariable [Martens, 1989; Sekulink, 1993]. En este caso no es necesario modelar explícitamente las interferencias ni el ruido de fondo antes de extraer la información química relevante, aunque es necesario que los patrones sean de la misma naturaleza que las muestras estudiadas. Cuando el problema contiene interferencias no presentes en los patrones, los métodos de calibración multivariable tampoco darán resultados satisfactorios, pero sí que permitirán la detección de estas interferencias. Con este tipo de instrumentos no se puede resolver todavía de forma completa el problema de las interferencias y la señal del ruido de fondo, aunque con tratamientos previos de la información analítica éstos pueden ser eliminados casi por completo. El centrado y la utilización de las derivadas pueden corregir los efectos de la línea base, o simplemente las calibraciones pueden modelarlos como un interferente más. Las no linealidades (p.ej. efectos de la matriz) pueden ser corregidas escogiendo el método de calibración no lineal más adecuado a los datos disponibles.

- *Matriz de datos. Instrumentos de segundo orden.*

Es este caso, para cada muestra analizada se obtiene una matriz de datos. En esta matriz se consideran dos direcciones, las filas y las columnas, que corresponden, generalmente, a dos tipos diferentes de medida. Este es el caso que se da cuando se combinan las medidas de emisión fluorescente a diferentes longitudes de onda de excitación; o cuando se efectúa un

seguimiento de respuesta de tipo multivariable con el tiempo, pH o concentración. Los instrumentos de segundo orden permiten el análisis en presencia de algún componente en la muestra que no esté incluido en la calibración, por tanto, la calibración será posible en presencia de interferencias. Esta característica particular se llama “ventaja de segundo orden”, la cuantificación puede ser llevada a cabo con tan solo una muestra de calibración. En algunas ocasiones, se pueden determinar incluso las respuestas puras de cada componente en cada uno de los órdenes de medida.

La construcción de un modelo de calibración con datos de segundo orden se puede llevar a cabo de manera similar al de primer orden si se efectúa un desdoblamiento de los datos (*unfolding*), de tal manera que para cada muestra se tenga un tensor de primer orden (figura 1). En estas situaciones se mezcla la información de ambos órdenes y tan solo se pueden aprovechar las ventajas de la calibración de primer orden. Cuando se mantiene la estructura tridimensional de los datos (p. ej. tiempo-longitud de onda-muestra) la calibración es de segundo orden y se mantienen las ventajas anteriormente mencionadas. En este caso se pueden utilizar los llamados “métodos de análisis de tres vías” (*three-way data análisis*) [Kruskal, 1989].

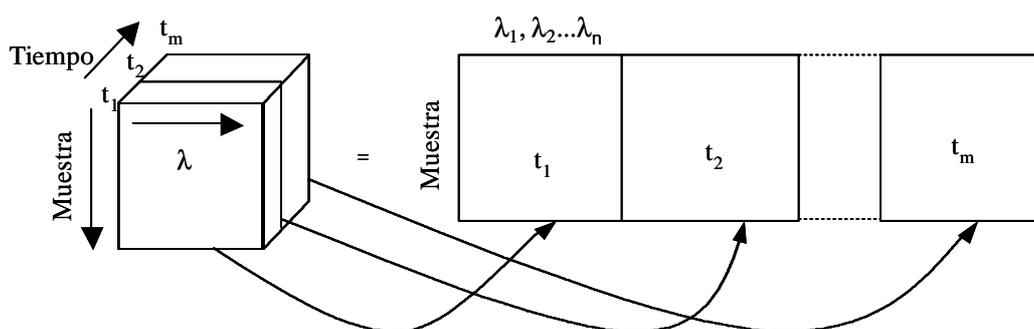


Figura 1. Desdoblamiento de una estructura tridimensional de datos.

- *Datos de orden superior. Instrumentos de orden superior.*

La extensión de los conceptos anteriores nos lleva hacia los instrumentos de orden superior y los datos obtenidos con ellos. No hay límite en el máximo orden de los datos que pueden ser obtenidos. Los instrumentos modernos de espectroscopia de fluorescencia pueden

proporcionar la variación de los espectros excitación-emisión con el tiempo, produciendo una estructura tridimensional de datos por muestra. Una ventaja añadida al utilizar instrumentos de orden superior es el aumento de la sensibilidad. También en este caso se pueden aplicar los métodos de análisis de “N vías”.

Los casos estudiados en esta memoria son todos de naturaleza multivariable, por lo que en esta introducción sólo se describen los métodos de calibración multivariable más relevantes. La clasificación descrita anteriormente no es la única que se utiliza. Martens propone otras clasificaciones de los métodos calibración basadas en conceptos menos matemáticos [Martens, 1989]:

(a) Según la relación entre las variables dependiente e independiente:

- *Calibración lineal* descrita por un modelo lineal,

$$y = b_0 + \sum_{k=1}^K b_k x_k \quad (1)$$

siendo b_0 , y b_k los parámetros a determinar y, x_k las variables independientes. Es decir, son modelos lineales en los parámetros.

- *Calibración no-lineal*: basada en modelos no lineales en los parámetros:

$$y = a x^b \quad (2)$$

donde a y b son los parámetros a determinar.

(b) Según la forma de encontrar la relación entre las variables:

- *Métodos directos*: los parámetros de la calibración se calculan directamente a partir de la señal de cada uno de los analitos de forma individual.

- *Métodos indirectos*: los parámetros de la calibración se calculan a partir de las señales analíticas de mezclas de los componentes.

(c) *Rígidos - Flexibles*: en los primeros es necesario tener información de todas las especies presentes que pueden contribuir a la señal, mientras que en los flexibles únicamente es necesario tener información de los analitos que se desea cuantificar.

(d) Según cuál sea la variable dependiente y cuál la independiente:

- *Calibración clásica*: sigue un criterio directamente relacionado con la ley de Beer-Lambert. La señal analítica actúa como variable dependiente de la concentración, que es la variable independiente.

-*Calibración inversa*: se utiliza la concentración como variable dependiente y la señal analítica como independiente.

Teniendo en cuenta esta última clasificación, cabe destacar como calibración clásica la regresión lineal múltiple clásica (*CLS, Classical Least Squares*). Dentro de la calibración inversa la regresión lineal múltiple inversa (*ILS, Inverse Least Squares*) y los métodos de regresión sesgados, basados en la reducción de variables. Estos últimos, teniendo en cuenta la ordenación de los datos cinéticos-espectrofotométricos utilizados en esta memoria, se pueden dividir en dos grandes bloques:

- Métodos “*unfolded*”; entre los que destaca la regresión en componentes principales (PCR), la regresión parcial por mínimos cuadrados (PLS) y la regresión continua (CR).
- Métodos “*N-way*”; entre los que destaca el *parallel factor analysis* (PARAFAC) y el PLS multidimensional (N-PLS).

Los métodos basados en la reducción de variables se caracterizan por que la información contenida en la señal analítica registrada puede ser concentrada o reducida en un menor número de variables, sin pérdida de la información relevante. La regresión de las respuestas no se realiza en los datos originales, sino en estas nuevas variables, simplificándose los modelos y la interpretación de los resultados. Presentan las ventajas de la calibración inversa, permitiendo realizar el análisis de únicamente alguno de los componentes que contribuyen a la absorción sin necesidad de conocer nada de las demás especies presentes. Por tanto, se les incluye también dentro de los métodos flexibles de calibración.

Entre los métodos de calibración no lineal cabe mencionar las redes neuronales artificiales (ANN) y variantes del PLS (*nonlinear PLS*). Una de las técnicas más aplicadas para datos cinético-espectrofotométricos ha sido el filtro de Kalman. Éste requiere de las variables medidas y de unos valores iniciales de los parámetros. Su variante, el filtro de Kalman extendido, puede ser aplicada también a modelos no lineales. Las limitaciones de esta técnica, aplicable a sistemas sencillos con conocimiento previo del modelo, hacen que sea substituida por métodos lineales como el PLS y, en sistemas muy complejos, por las ANN.

1.1.2 Construcción de modelos de calibración multivariable.

El objetivo de los métodos de calibración multivariable es establecer modelos de calibración que sean capaces de predecir el valor de la propiedad estudiada para nuevas muestras, a las que se ha registrado la señal analítica. Para que estas predicciones sean fiables, los modelos obtenidos han de ser robustos y, para ello, se deben seguir las etapas que se describen a continuación:

- 1) *Definición de un conjunto de calibración o conjunto de entrenamiento (training set)*. De él se conoce la propiedad a determinar, es representativo del sistema en estudio y debe permitir realizar predicciones futuras sobre un conjunto nuevo de muestras. El conjunto de calibración para realizar el análisis cuantitativo de una especie debe ser representativo de todo el intervalo de concentraciones, así como, de las posibles interferencias y otros componentes presentes en las muestras, aunque no vayan a ser determinados (siempre que no se aproveche la “ventaja de segundo orden” de las calibraciones de segundo orden). Se deben considerar tanto las fuentes de variación químicas como las físicas (origen de las muestras, tratamientos recibidos, etc), las cuales pueden afectar al espectro registrado.
- 2) *Registro de la información analítica*. En los trabajos presentados en esta memoria, esta información es espectrofotométrica, obtenida a partir de los espectros de absorción, dentro de la región del UV-Vis, y por la espectroscopia de dicroísmo circular (CD).
- 3) *Pretratamiento de los datos*. En esta etapa, se minimizan las contribuciones no deseadas presentes en la señal analítica que disminuyen la reproducibilidad y pueden provocar que el sistema presente comportamientos que darían lugar a estimaciones erróneas de los parámetros deseados. Los pretratamientos utilizados en los trabajos de la memoria se detallan en el apartado siguiente.
- 4) *Selección del modelo*. Establece la relación más simple posible entre la propiedad a determinar y la señal analítica. Esto, en muchas situaciones, implica un estudio complejo de las matriz de datos analíticos obtenidos, ya que no siempre es necesaria la utilización

de toda la información registrada.

- 5) *Validación del modelo.* Implica su aplicación a un número de muestras de las que se conoce la propiedad a determinar y que no han sido utilizadas en la etapa de construcción del modelo. De esta forma se verifica que el modelo construido constituye una correcta descripción del sistema en estudio.
- 6) *Predicción de nuevas muestras.* Con el modelo construido y validado, se calcula la propiedad estudiada en nuevas muestras a las que, previamente, se ha registrado la señal analítica.

1.1.3 Pretratamiento de los datos registrados.

Junto con la contribución del analito a la señal, se encuentran componentes o efectos no deseados, los cuales, de una manera genérica, son denominados ruido. Este ruido puede ser un ruido no estructurado, aleatorio, o bien un ruido estructurado fruto de interferencias químicas y/o físicas. Además, no siempre el instrumento responde linealmente a los cambios de concentración del analito por lo que la calidad de la señal puede ser deficiente. Existen bastantes tratamientos matemáticos que se pueden aplicar antes de intentar hallar la relación señal-concentración para mejorar la calidad de los datos. En esta memoria se han utilizado:

a) *Promediado de espectros.* El ruido instrumental es aleatorio, por lo que promediando n señales analíticas obtenidas a partir de una misma muestra, la relación señal/ruido aumenta en un factor $n^{1/2}$. El promediado de los espectros es una operación que casi siempre acompaña al registro del espectro de una muestra, siendo el espectro final obtenido el resultado del promedio de un número definido por el usuario del espectrofotómetro.

b) *Suavizado espectral.* Promediar espectros puede no ser suficiente para disminuir el ruido de alta frecuencia en aquellos casos en que la relación señal/ruido es pequeña, por lo que es necesario aplicar algoritmos sobre los espectros que minimicen o

eliminen este efecto. Para suavizar los espectros se pueden encontrar diferentes algoritmos, entre los que podemos destacar el de Savitzky-Golay [Savitzky, 1964] o el filtrado utilizando la transformada de Fourier [Horlick, 1972].

c) *Derivadas*. La derivada es uno de los pretratamientos más utilizados en espectroscopia por su capacidad de disminuir variaciones de línea base, como las debidas a la turbidez de las muestras o a la presencia de burbujas de aire. La utilización de la primera derivada elimina los términos constantes a todas las longitudes de onda, es decir, desplazamientos de la línea base. La segunda derivada elimina los términos que varían linealmente con la longitud de onda. No es común el uso de derivadas de orden superior. Uno de los métodos más utilizados para el cálculo de las derivadas es el propuesto por Savitzky-Golay [Savitzky, 1964]. Debido a que el uso de este pretratamiento espectral disminuye siempre la relación señal/ruido se tiene que ser cauteloso en su utilización.

1.1.4 Evaluación de la capacidad predictiva del modelo.

El objetivo de la calibración es obtener unos parámetros de regresión que permitan calcular la concentración en futuras muestras de forma que, para cada muestra i y analito j , el residual de la concentración, f_{ij} , sea el menor posible,

$$f_{ij} = \hat{y}_{ij} - y_{ij} \quad (3)$$

donde \hat{y}_{ij} es la concentración calculada.

Se desea minimizar algún tipo de error de predicción medio para la población a la que se aplicará la calibración. Para evaluar esta capacidad predictiva, se suele usar el sumatorio del cuadrado de los residuales, $(\sum (y_{ij} - \hat{y}_{ij})^2)$, denominado habitualmente PRESS (*Predicted Residual Error Sum of Squares*) o su valor medio, obtenido dividiendo el PRESS por el número de muestras, MSE (*Mean Squared Error*).

El cálculo de la concentración utilizando el modelo construido se denomina “predicción”. Así, se puede calcular el MSE de la predicción (MSEP) como

$$\text{MSEP} = \frac{\sum_{i=1}^{m_p} (\hat{y}_{ij} - y_{ij})^2}{m_p} \quad (4)$$

donde m_p es el número de muestras de predicción. También se suele utilizar la raíz cuadrada de este valor (*RMSEP*, *Root Mean Squared Error of Prediction*), debido a que tiene las mismas unidades en que se mide la concentración. De una manera similar, cuando se calcula con el modelo la concentración de las muestras del calibrado, se puede definir el MSEC y el RMSEC utilizando el PRESS y el número de muestras de calibración. Cuando se pretende comparar errores de calibración y/o predicción entre varios analitos, es útil el error cuadrático relativo (*RSE*, *Relative Standard Error*) definido para el analito j como

$$\text{RSE}_j = \sqrt{\frac{\sum_{i=1}^m (\hat{y}_{ij} - y_{ij})^2}{\sum_{i=1}^m (y_{ij})^2}} \times 100 \quad (5)$$

donde el sumatorio se extiende a las m muestras. Los símbolos utilizados para referirse a las muestras de calibración o de predicción son RSEC y el RSEP, respectivamente. Este parámetro es independiente de los rangos de concentración individuales, por tanto se podrán comparar errores entre analitos a diferentes concentraciones.

Para poder asegurar la bondad de un modelo en la predicción de muestras distintas a las utilizadas en calibración, es necesario un proceso de validación. Esta validación puede ser externa o interna.

Para la validación externa de los modelos, se suelen utilizar dos grupos de muestras, uno llamado propiamente de calibración y otro llamado conjunto de prueba o *test set*. El número de muestras del *test set* ha de ser elevado, y deben ser muestras independientes de las del conjunto de calibración pero representativas del mismo y de las futuras muestras a analizar. La concentración de las muestras del *test set* es conocida y, por tanto, es posible comprobar cómo se comporta el modelo frente a muestras distintas a las utilizadas en la construcción del mismo.

Cuando el número de muestras disponible es relativamente pequeño, se utiliza un método de validación interna. La metodología a seguir es la denominada “validación cruzada” (*cross validation*) [Geladi, 1985b; Wold, 1978], la cual, utiliza para comprobar el modelo muestras

del propio conjunto de calibración. Mediante este método, el conjunto de muestras de calibración se divide en varios bloques o segmentos. El modelo se construye reservando uno de los segmentos como conjunto de datos para comprobar los resultados y el resto para construir el modelo. El proceso se repite tantas veces como número de segmentos se ha elegido, de forma que cada vez se deja un segmento fuera del calibrado y el resto se utiliza para construir el modelo. Al final todos los segmentos han actuado para la construcción y para la validación del modelo. Esto se realiza para cada factor o componente principal, calculando el MSE para cada segmento y acumulándolo, de forma que se consigue una buena estimación del poder predictivo de las muestras de calibración. El MSECv (*Mean Squared Error of prediction by Cross Validation*) para cada factor y cada analito es

$$\text{MSECv} = \frac{\sum_{i=1}^{m_c} (\hat{y}_{ij} - y_{ij})^2}{m_c} \quad (6)$$

donde m_c es el número de muestras de calibración en el segmento de *cross validation*.

Tanto en modelos construidos por *test set* (construidos con validación externa) como en los obtenidos por *cross validation* (con validación interna), las muestras que definen el conjunto de validación han sido utilizadas indirectamente para construir el modelo, por lo que algunas veces se utiliza un tercer conjunto de muestras, al que llamaremos conjunto de predicción externo (*external prediction set*), para comprobar realmente el comportamiento del modelo con nuevas muestras. Por tanto, para este tercer conjunto se pueden definir los parámetros anteriormente mencionados con los siguientes acrónimos: RMSEEP y RSEEP.

1.2 REGRESIÓN LINEAL MÚLTIPLE.

La regresión lineal múltiple (MLR), introducida en el año 1960 por Sternberg y col. [Sternberg, 1960], es el tratamiento multivariable más sencillo. Calcula la relación lineal entre la señal analítica y la concentración aplicando el método de mínimos cuadrados [Draper, 1981], y se utiliza tanto en la calibración directa como inversa (ILS).

1.2.1 Regresión lineal múltiple clásica.

Asume el cumplimiento de la ley de Beer-Lambert para cada uno de los componentes de la

mezcla en todo el intervalo de trabajo y la aditividad de las absorbancias en las mezclas. El error en el modelo se considera que es debido a los datos espectrales. Para una mezcla de p componentes, si se realizan medidas a k longitudes de onda ($k \geq p$), se obtiene un sistema de k ecuaciones, que puede expresarse en forma matricial como:

$$\mathbf{A} = \mathbf{K} \cdot \mathbf{C} + \mathbf{E} \quad (7)$$

donde \mathbf{A} ($k \times 1$) es una matriz columna de absorbancias de una muestra a k longitudes de onda, \mathbf{K} es la matriz de coeficientes de dimensiones $k \times p$, \mathbf{C} es la matriz de concentraciones de dimensiones $p \times 1$ y \mathbf{E} la matriz de los residuales de las absorbancias que no ajustan correctamente el modelo, de dimensiones $k \times 1$.

Dependiendo del método con que se calcule la matriz \mathbf{K} estaremos ante un modelo de calibración directa o indirecta:

a) En el primer caso, los valores se pueden calcular a partir de los espectros de los componentes puros presentes en la mezcla, de manera que, si el camino óptico es constante, cada columna de la matriz \mathbf{K} es el espectro del componente puro patrón por su concentración. Se le suele denominar CLS por las iniciales de “*Classical Least-Squares*” [Haaland, 1980; Thomas, 1990]. Para la predicción, la solución por mínimos cuadrados para la matriz concentraciones no conocidas de una muestra es:

$$\hat{\mathbf{C}} = (\mathbf{K}^T \mathbf{K})^{-1} \mathbf{K}^T \mathbf{A} \quad (8)$$

donde \mathbf{A} es el espectro de la muestra, y donde el símbolo $\hat{}$ indica que es un valor estimado.

b) En la calibración indirecta, ésta se efectúa utilizando mezclas de todos los componentes. Si se utilizan m muestras de calibración (con más muestras que analitos), se tiene una matriz de absorbancias \mathbf{A} , de dimensiones $k \times m$; y una matriz de concentraciones \mathbf{C} de dimensiones $p \times m$. La matriz de residuales \mathbf{E} , en este caso, tiene dimensiones $k \times m$. Los valores de la matriz \mathbf{K} son estimados por mínimos cuadrados:

$$\hat{\mathbf{K}} = \mathbf{A} \mathbf{C}^T (\mathbf{C}^T \mathbf{C})^{-1} \quad (9)$$

donde $\hat{\mathbf{K}}$ es el estimado de \mathbf{K} .

La predicción de una muestra de concentración desconocida, se efectúa igual que en el caso anterior mediante la ecuación de $\hat{\mathbf{C}}$ estimada.

Puesto que estos métodos son de espectro completo, pueden proporcionar mejoras en la precisión frente a otros métodos en los que se trabaja con un número limitado de variables, y se pueden aplicar a la determinación simultánea de analitos si se conocen todos los componentes que contribuyen a la señal analítica, incluidas las interferencias.

La precisión de las concentraciones calculadas viene determinada por la precisión de las medidas de absorbancia y por la validez del modelo elegido. Los espectros de los analitos han de ser razonablemente distintos, ya que si están muy solapados provocan una gran colinealidad de las columnas de la matriz \mathbf{K} (es decir, relaciones casi lineales entre absorbancias de diferentes analitos) y consecuentemente, dificultades para el cálculo de la inversa de $\mathbf{K}^T\mathbf{K}$.

1.2.2 Regresión Lineal Múltiple Inversa.

El método de Regresión Lineal Múltiple Inversa (ILS, *Inverse Least Squares*) [Thomas, 1990] asume que la concentración es función de la absorbancia, siguiendo este modelo, la ecuación inversa de la ley de Beer-Lambert. De esta manera, se obtiene una ecuación donde la concentración actúa como variable dependiente, mientras que los valores espectrales registrados a distintas longitudes de onda son las variables independientes. Si se desea conocer la concentración de mezclas que contienen p analitos, se registran los espectros de m muestras para la calibración y se obtiene una ecuación a resolver del tipo

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E} \quad (10)$$

donde la matriz \mathbf{Y} contiene las concentraciones, con dimensiones $m \times p$, y la matriz \mathbf{X} los datos espectroscópicos, con dimensiones $m \times k$. La matriz \mathbf{E} de dimensiones $m \times p$ es la matriz de residuales aleatorios de las concentraciones y la matriz \mathbf{B} , de dimensiones $k \times p$, es la matriz de regresores desconocida que se calcula en la calibración. Se asume que el error está en las concentraciones, por lo que el método minimiza el cuadrado de los errores en las concentraciones.

Utilizar este método de calibración tiene la ventaja de que no es necesario conocer la contribución a la absorbancia total de todas las especies absorbentes de la mezcla, sino sólo la de los analitos en los que estemos interesados. Sin embargo, los componentes no incluidos en la cuantificación deben estar presentes en todas las muestras y son modelados implícitamente.

La solución por mínimos cuadrados de la ecuación de calibración es

$$\hat{\mathbf{B}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (11)$$

La desventaja de este método es que el número de muestras debe ser mayor que el número de variables, por lo que el modelo se restringe únicamente a unas pocas longitudes de onda. Seleccionar cuántas y cuáles serán las variables a utilizar es una cuestión difícil, especialmente en problemas complejos [Honigs, 1983]. Además, cuando se utiliza un número elevado de longitudes de onda se pueden tener los mismos problemas de colinealidad (diferentes variables independientes aportan la misma información espectral ya que están altamente correlacionadas) que mediante la calibración clásica, lo que disminuye la precisión de los resultados.

1.3 MÉTODOS LINEALES PARA EL TRATAMIENTO DE DATOS DE PRIMER ORDEN BASADOS EN LA REDUCCIÓN DE VARIABLES.

1.3.1 Consideraciones previas.

Los métodos de reducción de variables no se suelen aplicar directamente a los datos originales, sino que estos son centrados y/o autoescalados. Con estos tratamientos se intenta transformar matemáticamente los datos a la forma más adecuada para la obtención del modelo de calibración.

Consideremos una matriz \mathbf{X} de datos, donde cada fila se corresponde con el espectro de una muestra y cada columna es la absorbancia medida a una determinada longitud de onda (variable). Si se denomina x_{ik} al elemento de la matriz que está en la fila i y en la columna k , es habitual referir todos los valores de las variables a un origen común mediante el:

- *centrado de los datos*. Se calcula el valor medio de cada variable de la matriz y se resta a cada punto de la columna. El valor medio corresponde al centro del modelo, y los valores de todas las variables están ahora referidos a dicho centro.

$$x_{ik} - \bar{x}_k \quad (12)$$

Este tratamiento permite mantener las unidades originales y equivale a tratar la matriz de varianza-covarianza de las variables originales.

Si las variables originales están expresadas en unidades distintas, o cuando sus varianzas son muy diferentes, se aplica el:

- *autoescalado de los datos*. Después de centrar cada columna, se divide el resultado por la desviación estándar de la misma, s_k , de forma que la varianza de cada variable vale la unidad.

$$\frac{x_{ik} - \bar{x}_k}{s_k} \quad (13)$$

Se han cambiado las unidades originales a unidades de desviación estándar. Equivale a tratar la matriz de correlación de las variables originales. Geométricamente equivale a cambiar la longitud de los ejes de coordenadas. De esta forma, todos los ejes tienen la misma longitud y cada variable tiene la misma influencia en el cálculo.

Si los datos son espectros de muestras puede ser más interesante no autoescalarlos, ya que al autoescalar se dará igual importancia a las variables con baja absorbancia, las que principalmente tienen ruido, que a los máximos de absorción. Si los datos provienen del registro de la absorbancia a una única longitud de onda con el tiempo, el escalado puede ser una buena solución, ya que se dará igual importancia a todos los tiempos. Esto es especialmente interesante cuando las reacciones son rápidas y en los primeros tiempos de registro los analitos presentan velocidades muy diferenciadas. Si los datos son espectros registrados a diferentes tiempos posiblemente se tendrán que probar ambos tratamientos. Si los analitos que contribuyen a la señal tienen un comportamiento cinético-espectral similar, las pequeñas diferencias existentes entre ambos serán acentuadas con el autoescalado pero esto puede suponer que también se vea afectado el ruido y empeore la calidad de los resultados. Por otra parte, escalar los datos puede favorecer la determinación de compuestos minoritarios en presencia de absorbentes mayoritarios. Para cada problema en concreto hay que adoptar la solución más conveniente.

Los efectos de estos pretratamientos han sido discutidos en la bibliografía [Geladi, 1985a; Blanco, 1994].

1.3.2 Regresión en componentes principales (PCR).

Este método de reducción de variables consta de dos etapas. En la primera se efectúa una reducción de variables mediante un análisis en componentes principales (PCA) y en la segunda se realiza la regresión de estos componentes principales sobre la matriz de parámetros a estimar que en este trabajo siempre han sido concentraciones.

1.3.2.1 Etapa de análisis en componentes principales.

El análisis en componentes principales (PCA) se aplica en química con diferentes finalidades [Jackson, 1991; Wold, 1987]. Entre ellas cabe destacar la clasificación de muestras y la reducción en la dimensionalidad de los datos, siendo éste el paso previo de muchos métodos multivariados de análisis de datos.

En la técnica multivariable PCA, un número elevado de variables, más o menos correlacionadas, son transformadas en un número inferior de variables ortogonales (no correlacionadas) y que contienen la información principal del sistema, eliminando la información redundante o reduciendo la contribución del ruido.

Se parte de una matriz de datos \mathbf{X} , donde cada fila es una muestra u objeto, y cada columna una variable, que puede ser la absorbancia a una longitud de onda y tiempo determinados. La dimensionalidad de esta matriz, si se han registrado m muestras y k variables, será $m \times k$.

Geométricamente, el PCA es un cambio en la dirección y dimensionalidad de los ejes (figura 2). Busca las direcciones ortogonales que explican la máxima variabilidad de las muestras y las utiliza como nuevos ejes de coordenadas llamados componentes principales (PCs) para representarlas. El primer componente principal es la dirección que explica la máxima variabilidad; el segundo se escoge de tal forma que sea perpendicular al primero y que explique la máxima variabilidad una vez extraída la explicada por el primer componente principal, y así sucesivamente. Para poder definir estos nuevos ejes se utilizan sus *loadings* que son los cosenos de los ángulos que forman con los ejes antiguos. Los *scores* son las coordenadas de las muestras en estos nuevos ejes.

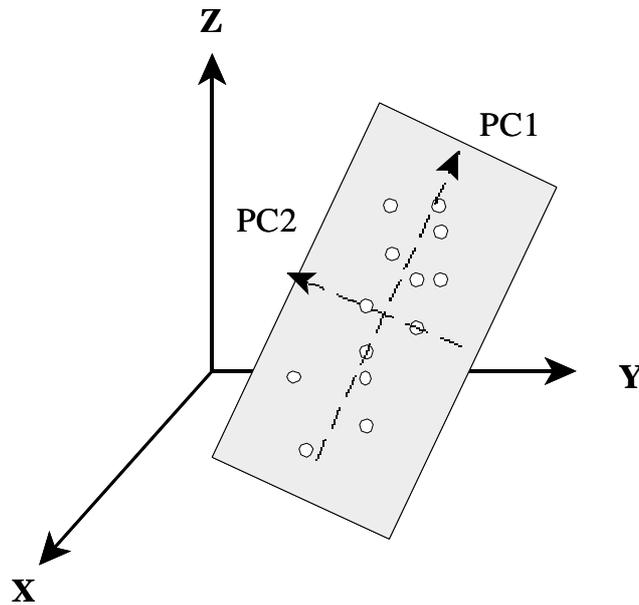


Figura 2. Interpretación geométrica de un PCA.

Matemáticamente, la matriz de datos \mathbf{X} (datos cinético-espectrales) se descompone en el producto de dos matrices, \mathbf{T} (matriz de *scores*) y \mathbf{P} (matriz de *loadings*), más una matriz \mathbf{E} de residuales de \mathbf{X}

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (14)$$

Los diferentes componentes principales no contienen la misma información. Los primeros describen la fuente de variación más importante de los datos, que se puede asociar a la información más relevante. Para tener una reproducción perfecta de la matriz \mathbf{X} se necesitan k vectores de *loadings* y *scores*, sin embargo, es posible representar la matriz original con un número $a < k$ de vectores

$$\mathbf{X} = \mathbf{t}_1\mathbf{p}_1^T + \mathbf{t}_2\mathbf{p}_2^T + \dots + \mathbf{t}_a\mathbf{p}_a^T + \mathbf{E} \quad (15)$$

puesto que la información relevante está contenida en los a primeros componentes, mientras que los últimos sólo describen variaciones debidas al ruido. Esto permite una reducción importante del número de variables. Resumiendo, se puede decir que el conjunto de datos \mathbf{X} , que estaba descrito por variables correlacionadas, en este momento está definido por un nuevo conjunto de variables, llamadas componentes principales, que son variables no correlacionadas entre sí, en un nuevo sistema de ejes ortogonales.

Existen diferentes algoritmos de cálculo para obtener las matrices \mathbf{T} y \mathbf{P} . Uno de ellos es el NIPALS (*Nonlinear Iterative Partial Least Squares*) [Wold, 1966], y es el algoritmo que utiliza el software *The Unscrambler* [CAMO AS, 1996] utilizado para realizar alguno de los cálculos de esta memoria.

Otra forma de efectuar una reducción de variables se basa en la descomposición en valores singulares de la matriz de datos, \mathbf{X} (*singular value decomposition, SVD*).

Cualquier matriz de dimensiones $m \times k$ puede ser escrita por el producto de tres términos, \mathbf{U} , \mathbf{S} y \mathbf{V} .

$$\mathbf{X} = \mathbf{USV}^T \quad (16)$$

donde \mathbf{U} y \mathbf{V} son una matrices ortonormales de dimensiones $m \times r$ y $k \times r$, respectivamente. \mathbf{S} es una matriz diagonal de dimensiones $r \times r$ cuyos elementos de la diagonal son los valores singulares y estan colocados en orden descendiente, es decir, $s_{11} \geq s_{22} \geq s_{33} \dots \geq s_{rr}$. Los elementos fuera de la diagonal son cero. Los *loadings* \mathbf{P} son idénticos a \mathbf{V} y los *scores* $\mathbf{T} = \mathbf{US}$. Por tanto, el modelo PCA puede ser escrito como:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} = \mathbf{USV}^T + \mathbf{E} \quad (17)$$

El resultado es el mismo que el obtenido mediante el algoritmo NIPALS.

1.3.2.2 Etapa de regresión en componentes principales (PCR).

La Regresión en Componentes Principales (PCR, *Principal Component Regression*) utiliza los componentes principales encontrados mediante el PCA y realiza una regresión múltiple inversa (ILS) de la propiedad a determinar sobre los *scores*, en lugar de realizarla sobre los datos originales, con lo que se elimina cualquier posible problema en la inversión de la matriz, $(\mathbf{X}^T\mathbf{X})^{-1}$.

Si tenemos una muestra con un conjunto de p especies absorbentes, tendremos p variables, $y_1, y_2, y_3, \dots, y_p$, representado la concentración de cada componente, que pueden ser escritas en forma de un vector \mathbf{y} . El espectro de la misma, registrado a k longitudes de onda constituye un conjunto de k variables independientes $x_1, x_2, x_3, \dots, x_k$ que pueden ser escritas en forma de vector \mathbf{x} . Si se construye un conjunto de calibración con m objetos, se pueden agrupar los vectores que describen cada uno de ellos en dos matrices: la matriz \mathbf{Y} , que contiene las concentraciones o las propiedades a determinar de cada muestra, de dimensiones $(m \times p)$ y la

matriz \mathbf{X} , que contiene los espectros de cada muestra, de dimensiones ($m \times k$). De este modo, dentro de las matrices la información espectral o las propiedades a determinar de cada muestra están escritas en una fila, mientras que cada columna contiene la información de una variable concreta para todas las muestras presentes.

El primer paso es realizar una descomposición de la matriz \mathbf{X} en sus componentes principales como se ha indicado en el apartado anterior (ecuación 15).

Una vez elegido el número a de componentes principales óptimo para describir la matriz \mathbf{X} , ésta se puede representar por su matriz de scores \mathbf{T}

$$\mathbf{T} = \mathbf{X}\mathbf{P} \quad (18)$$

Hasta este punto se ha realizado un PCA, obteniendo a partir de la matriz de datos \mathbf{X} la matriz de scores \mathbf{T} y la de loadings \mathbf{P} .

Ahora se puede calcular la matriz \mathbf{Y} según:

$$\mathbf{Y} = \mathbf{T}\hat{\mathbf{B}} + \mathbf{E} \quad (19)$$

donde $\hat{\mathbf{B}}$ es la matriz de regresores que se halla por mínimos cuadrados conociendo los valores de \mathbf{Y} del conjunto de calibración:

$$\hat{\mathbf{B}} = (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{Y} \quad (20)$$

donde el símbolo $\hat{}$ indica valores calculados.

Una vez establecido el modelo de calibración, se pueden realizar los cálculos para predecir un conjunto de nuevas muestras. En primer lugar, la matriz de datos cinético-espectroscópicos desdoblada del conjunto de muestras de predicción, \mathbf{X}^* , se centra o autoescala utilizando los valores calculados a partir de la matriz de datos \mathbf{X} empleada en la calibración. A partir de la matriz de loadings calculada en la calibración, para el número a de componentes principales óptimo, se calculan los scores de las muestras de predicción, \mathbf{T}^* .

$$\mathbf{T}^* = \mathbf{X}^*\mathbf{P} \quad (21)$$

y, por último, se utiliza la matriz de regresores calculada en la calibración, junto con los scores de estas muestras, para el cálculo de la propiedad a determinar en las muestras desconocidas

$$\mathbf{Y}^* = \mathbf{T}^*\hat{\mathbf{B}} \quad (22)$$

Uno de los principales problemas con PCR es que los componentes principales escogidos que

mejor representan la matriz de los datos, \mathbf{X} , pueden no ser los óptimos para la predicción de las propiedades de los analitos que queremos determinar [Joliffe, 1982; Sutter, 1992]. Por este motivo se han desarrollado otras técnicas de calibración que intenta concentrar el máximo poder predictivo en los primeros componentes principales como es el caso de la regresión parcial por mínimos cuadrados.

1.3.3 Regresión parcial por mínimos cuadrados.

El método de Regresión Parcial por Mínimos Cuadrados (PLS, *Partial Least-Squares Regression*) fue desarrollado por H. Wold en 1975 [Wold, H.; 1975]. La diferencia con el PCR es que se intenta que los primeros componentes contengan la mayor información para la predicción de la matriz \mathbf{Y} . Para ello, durante la etapa de calibración, el algoritmo PLS utiliza tanto la información contenida en la matriz de datos (matriz \mathbf{X} , p. ej. Datos cinético-espectrofotométricos) como la información contenida en la matriz de la propiedad a determinar (matriz \mathbf{Y} , p. ej. concentraciones), obteniéndose unas variables auxiliares llamadas variables latentes, factores o componentes que tienen gran parecido a los componentes principales que se hallan a partir de un PCR.

Antes de realizar la descomposición en factores, las matrices \mathbf{X} e \mathbf{Y} se centran o autoescalan como en el caso del PCA. Cada una de las matrices se descompone simultáneamente en una suma de A factores de forma que:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} = \sum_{i=1}^a \mathbf{t}_i \mathbf{p}_i^T + \mathbf{E} \quad (23)$$

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F} = \sum_{i=1}^a \mathbf{u}_i \mathbf{q}_i^T + \mathbf{F} \quad (24)$$

donde \mathbf{T} es la matriz de *scores*, \mathbf{P} la de *loadings* y \mathbf{E} la matriz de residuales para la matriz de datos (matriz \mathbf{X}); \mathbf{U} es la matriz de *scores*, \mathbf{Q} la matriz de *loadings* y \mathbf{F} la matriz de residuales para la matriz de la propiedad a determinar (matriz \mathbf{Y}). Si tenemos m muestras, a factores, k variables y p analitos, la dimensionalidad de las matrices es la siguiente: \mathbf{T} y \mathbf{U} ($m \times a$), \mathbf{P}^T ($a \times k$) y \mathbf{Q}^T ($a \times p$). En este caso, los *loadings* no coinciden exactamente con la dirección de máxima variabilidad de las muestras como en el caso de PCA, ya que están corregidos para obtener la máxima capacidad predictiva para la matriz \mathbf{Y} .

La descomposición de ambas matrices no es independiente, sino que se realiza de forma simultánea, estableciéndose una relación interna entre los *scores* de los bloques \mathbf{X} e \mathbf{Y} de forma que para cada componente a , se cumpla que

$$\hat{\mathbf{u}}_a = \mathbf{b}_a \mathbf{t}_a \quad (25)$$

donde el símbolo $\hat{}$ indica que es un valor calculado y b_a es el coeficiente de regresión para cada uno de los componentes.

A partir de aquí se calcula el valor de \mathbf{Y} utilizando la relación interna \hat{u}_a

$$\mathbf{Y} = \mathbf{T}\mathbf{B}\mathbf{Q}^T + \mathbf{F} \quad (26)$$

donde \mathbf{B} es la matriz de los regresores b_a , de dimensiones $(a \times a)$, y \mathbf{F} la matriz de los residuales de \mathbf{Y} .

En el caso de calcular una sola propiedad de la matriz \mathbf{Y} el algoritmo recibe el nombre de PLS1 y si se determinan simultáneamente varias propiedades recibe el nombre de PLS2.

Una vez establecido el modelo de calibración se puede realizar la predicción de la propiedad modelada en un nuevo conjunto de muestras según la siguiente expresión

$$\mathbf{y}_i^T = \hat{\mathbf{b}}_0^T + \mathbf{x}_i^T \hat{\mathbf{B}} \quad (27)$$

siendo \mathbf{x}_i el vector que define la señal analítica de la muestra, \mathbf{y}_i el vector de concentraciones o propiedades a determinar, y $\hat{\mathbf{b}}_0^T$ un vector que permite realizar la predicción de una muestra sin necesidad de descomponerla en matrices de *scores* y *loadings*.

1.3.4 Regresión continua (CR).

Las técnicas de regresión MLR, PCR y PLS pueden ser unificadas bajo la llamada Regresión Continua (CR) [Lorber, 1987; Stone, 1990 y 1994]. La idea central de la CR tiene en cuenta que el PLS captura la covarianza entre las variables de entrada o datos \mathbf{X} , y las de salida \mathbf{Y} , así como la reducción en la dimensionalidad de los datos \mathbf{X} . En los extremos de este método están: el PCR el cual empieza con un modelo que únicamente describe la varianza en las \mathbf{X} y la correlaciona con la matriz \mathbf{Y} ; y el MLR que busca tan solo correlacionar la matriz \mathbf{X} e \mathbf{Y} sin tener en cuenta la estructura de la matriz \mathbf{X} . El método convencional PLS intenta hacer las dos aproximaciones y, por tanto, ocupa una posición intermedia entre PCR y MLR.

Inicialmente, el algoritmo lleva a cabo una descomposición en valores singulares SVD de la matriz \mathbf{X} (ecuación 16). A partir de aquí, la matriz de datos es modificada a:

$$\mathbf{X}(\gamma) = \mathbf{U}\mathbf{S}^{(\gamma)}\mathbf{V}^T \quad (28)$$

los valores singulares son elevados a una cierta potencia γ y se construye una nueva matriz de datos que es función de esta potencia. Se aplica la regresión entre $\mathbf{X}(\gamma)$ e \mathbf{Y} . Cambiando la potencia desde 0 a ∞ el modelo cambia desde MLR ($\gamma = 0$) vía PLS ($\gamma = 1$) hasta PCR ($\gamma = \infty$). El programa de cálculo en código MatLab utilizado para esta técnica permite que los valores de γ pueden ser prefijados o que sean determinados simultáneamente junto con el número de factores mediante *cross validation*, de tal manera que se obtenga el mínimo PRESS.

1.3.5 Elección del número óptimo de componentes principales o factores.

La elección del número de componentes principales (PCs) o factores que configura el modelo óptimo es el punto clave en la utilización de cualquier técnica de calibración que realice una reducción de variables. Se han propuesto diferentes técnicas para deducir el número de PCs que se basan, en general, en el análisis del error de predicción al utilizar distinto número de éstos. Algunas de ellas han sido descritas en la bibliografía [Malinowsky, 1991].

Puesto que el método más habitual para la construcción del modelo es el de la validación cruzada (*cross validation*), una forma muy popular de seleccionar el número de factores óptimo es la sugerida por Wold [Wold, 1978], que consiste en representar el valor de MSECv frente al número de PCs y buscar el mínimo. Se parte de la idea de que el error disminuye al aumentar el número de PCs, ya que se modela cada vez mejor el sistema, hasta que llega un momento en que los nuevos PCs introducidos únicamente explican ruido y el MSECv vuelve a aumentar debido al sobreajuste del modelo. Aunque es una idea razonable, el hecho de utilizar únicamente un número limitado de muestras (como máximo todas las presentes en el conjunto de calibración) hace que el método esté sujeto a un cierto error y tenga el peligro de generar un cierto sobreajuste de los datos [Osten, 1988]. Otros autores prefieren utilizar el primer mínimo local que aparece en la representación del MSECv frente al número de PCs [Martens, 1989], aunque puede ser que en este caso se produzca un subajuste de los datos. Cuando la construcción del modelo se realiza por *test set*, el número de PCs se selecciona en función de la evolución del error para este conjunto de muestras.

Otro método es el descrito por Haaland y Thomas [Haaland, 1988] que escoge el número de componentes cuyo MSECv no es significativamente mayor que el mínimo MSECv del modelo, evitando de esta manera el sobreajuste. El mínimo valor del MSECv vendrá dado por un número de componentes que denominamos a^* . Cada valor de MSECv obtenido con un número de componentes menor al a^* se compara con el valor de $MSECv(a^*)$ mediante una prueba de Fisher. Con esto se pretende encontrar el modelo más pequeño cuyo MSECv no sea significativamente más grande que aquel modelo que proporciona el mínimo, evitando así el sobreajuste. Para cada componente $a = 1, 2, \dots, a^*$ se calcula

$$F(a) = \frac{MSECv(a)}{MSECv(a^*)} \quad (29)$$

Como número de componentes óptimo se escoge el menor a tal que $F(a) < F_{\alpha, m, m}$ donde $F_{\alpha, m, m}$ es el valor tabulado para una prueba F unilateral con un porcentaje de nivel de significación de $(1-\alpha)$ y m grados de libertad. Basándose en un criterio meramente empírico, el valor recomendado por Haaland y Thomas es 0.25. Los grados de libertad del numerador y denominador son los mismos (m en el caso de PLS1 y mp en el caso de PLS2, siendo m el número de muestras y p el número de analitos).

El programa *The Unscrambler* [CAMO, AS; 1996] utilizado en esta memoria propone como número óptimo de componentes el calculado por: $\min(MSE_{val 0} \times 0.01 \times a + MSE_{val a})$, donde MSE_{val} es el error de validación (que puede ser MSECv o MSEP), a indica el número de componentes principales y el subíndice 0 y a indican, respectivamente, el valor 0 y a componentes. Es decir, se añade a veces un 1 % de varianza inicial de las Y a la varianza correspondiente al componente a , de forma que se evita escoger un componente que no mejore de forma apreciable la predicción para que no exista un sobreajuste.

1.4 MÉTODOS LINEALES PARA EL TRATAMIENTO DE DATOS DE SEGUNDO ORDEN BASADOS EN LA REDUCCIÓN DE VARIABLES.

1.4.1 Introducción.

Todos los métodos anteriormente mencionados están diseñados para tratar datos de primer orden. Si se dispone de datos de segundo orden o superior (multidimensionales) deben ser desdoblados para poder utilizar dichos métodos de calibración y esto resulta en una pérdida de la estructura tridimensional y por tanto, de su información implícita. Recientemente, se han desarrollado técnicas capaces de tratar este tipo de datos directamente sin la necesidad de un desdoblamiento previo, es el caso del PARAFAC y el nPLS.

1.4.2 *Parallel Factor Analysis*, PARAFAC.

1.4.2.1 Introducción.

Para descomponer datos multidimensionales se puede, simplemente, desdoblar el cubo de datos en una matriz y aplicar un método estándar tal como PCA (capítulo 1.3.2.1). Si se desea mantener la estructura multidimensional, se puede utilizar el método PARAFAC o el método Tucker3. El método Tucker3 suele llamarse también *N-mode principal component analysis* y no se describirá aquí [Geladi, 1989].

PARAFAC es un método multidimensional que tiene su origen en la psicometría [Harshman, 1970; Carroll, 1970]. Este método ha ganado interés en quimiometría por diferentes razones: por su simplicidad y posibilidades, por el incremento de la complejidad de los datos obtenidos con instrumentos cada vez más sofisticados y por el incremento del poder computacional de los ordenadores de hoy en día que permiten procesar grandes cantidades de información [Geladi, 1989; Smilde, 1992]. PARAFAC, Tucker3 y PCA son todos métodos que descomponen los datos en conjuntos de *scores* y *loadings* para describirlos de forma más condensada. Todos estos métodos presentan ventajas e inconvenientes, por tanto, puede ser necesario probar varios de ellos para encontrar el más adecuado a cada problema en concreto. Sin entrar en detalles sobre PCA (ya descrito anteriormente) y el método de Tucker, es importante considerar ciertos aspectos sobre la jerarquía de estos métodos. Kiers [Kiers, 1991a] muestra que PARAFAC puede ser considerado como una versión restrictiva del método Tucker3, y que el método Tucker3 es una versión restrictiva de PCA. Cualquier conjunto de datos multidimensionales puede ser modelado adecuadamente tanto utilizando PARAFAC como Tucker3 o PCA, pero PARAFAC utiliza menos grados de libertad. Un

modelo PCA siempre ajusta los datos mejor que un modelo Tucker3 -entendiendo la palabra ajuste como el proceso de modelado- el cual, a su vez, ajusta mejor los datos que un modelo PARAFAC, todo esto excepto para casos extremos donde los modelos pueden ajustar igualmente bien. Si un modelo PARAFAC es adecuado, los modelos Tucker3 y PCA tenderán a utilizar un exceso de grados de libertad para modelar ruido o variaciones sistemáticas de manera redundante. Además, normalmente se prefiere utilizar el modelo más simple posible. Como el PCA utiliza más grados de libertad puede ser considerado el modelo más complejo y flexible, mientras que el PARAFAC es el modelo más simple y restrictivo.

La experiencia práctica existente en la utilización del PCA hace pensar que éste sea más simple que los métodos multidimensionales, sin embargo, si consideramos la estructura multidimensional de los datos esto no es cierto, ya que al aplicar un PCA los datos han de ser previamente desdoblados en una matriz bidimensional. Las variables aparecen mezcladas y, por tanto, el efecto de una variable no está asociado con un único elemento del vector *loading*, sino con varios de ellos. Cuando no se tiene en cuenta la estructura tridimensional, el modelo utiliza por igual todos los datos, usando los grados de libertad necesarios para obtener el mejor ajuste. Así, PCA puede dar lugar a modelos más complejos y difíciles de interpretar, debido a que el número de parámetros que se calculan es superior que en el caso de PARAFAC.

La razón para utilizar los métodos multidimensionales no es obtener el mejor modelado de los datos, sino el modelo más adecuado, robusto e interpretable. Esto es similar a la comparación entre MLR y PLS para calibración multivariable. Se sabe que la MLR, que es un método no sesgado, proporciona el mejor modelado para la variable dependiente del conjunto de calibración, pero en la mayoría de casos PLS posee mejor poder predictivo. Los métodos multidimensionales son menos sensibles al ruido que los bilineales y proporcionan *loadings* que pueden ser directamente relacionados con los diferentes modos o direcciones de la estructura multidimensional de los datos. En este apartado de la memoria tan solo se tendrán en cuenta los datos tridimensionales por su simplicidad y por la estructura de los datos utilizados, pero la mayoría de resultados son válidos para datos y modelos de orden superior.

1.4.2.2 El modelo PARAFAC.

El PARAFAC [Bro, 1997] es un método de descomposición que, conceptualmente, puede ser comparado a un PCA bilineal. El modelo fue propuesto independientemente por Harshman [Harshman, 1970] y por Carroll y Chang [Carroll, 1970] los cuales lo llamaron CANDECOMP (descomposición canónica). La descomposición de los datos se realiza en componentes trilineales, pero en vez de un vector de *scores* y uno de *loadings* como en el caso del PCA bilineal, cada componente consiste de un vector de *scores* y dos vectores de *loadings* (figura 3).

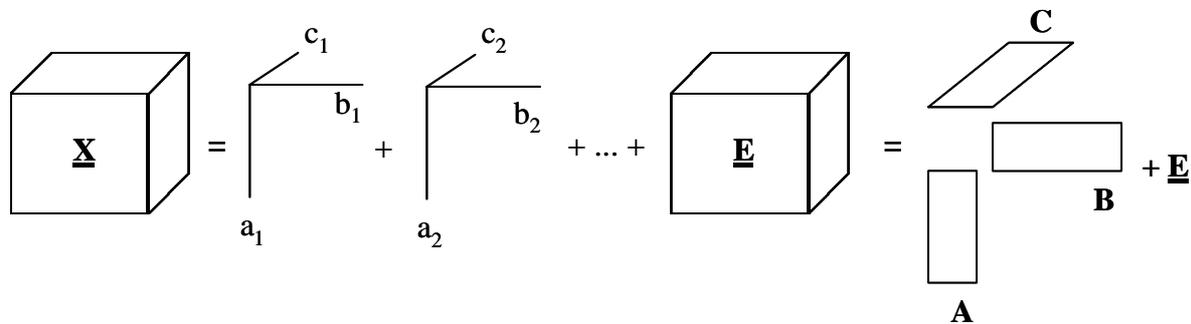


Figura 3. Descomposición PARAFAC de datos con estructura tridimensional.

En la práctica no se distingue entre *scores* y *loadings*, ya que son tratados numéricamente por igual. Por tanto, un modelo PARAFAC de una estructura tridimensional viene definido por tres matrices de loadings, \mathbf{A} ($I \times F$), \mathbf{B} ($J \times F$) y \mathbf{C} ($K \times F$), cuyos elementos son a_{if} , b_{jf} y c_{kf} , respectivamente. F representa el número de factores; I , J y K pueden ser considerados como el número de muestras, longitudes de onda y tiempos, respectivamente. Se busca el modelo trilineal que minimice la suma de cuadrados de los residuales, e_{ijk} en el modelo:

$$x_{ijk} = \sum_{f=1}^F a_{if} b_{jf} c_{kf} + e_{ijk} \quad (30)$$

El modelo puede ser también escrito de la siguiente manera:

$$\underline{\mathbf{X}} = \sum_{f=1}^F \mathbf{a}_f \mathbf{b}_f \mathbf{c}_f \quad (31)$$

donde $\underline{\mathbf{X}}$ es la matriz de datos de estructura tridimensional cuyas dimensiones son $I \times J \times K$, \mathbf{a}_f , \mathbf{b}_f y \mathbf{c}_f son las columnas f de las matrices de *loadings* \mathbf{A} , \mathbf{B} y \mathbf{C} , respectivamente [Burdick, 1995].

Características del modelo PARAFAC:

- La solución es única. En los métodos bilineales es bien conocido el problema que existe de libertad rotacional. En PCA se pueden rotar los *scores* y *loadings* manteniendo el ajuste del modelo, por tanto, no es posible estimar, por ejemplo, los espectros puros, porque hay infinitas soluciones (*scores* y *loadings*) que proporcionan el mismo ajuste. Este no es el caso del PARAFAC; si los datos son trilineales (es decir, cada especie química debe ser definida en las diferentes matrices de datos por el mismo perfil espectral y de concentración), se podrá obtener el espectro verdadero utilizando el número correcto de factores, siempre y cuando la relación señal/ruido sea apropiada.
- A pesar de la interpretabilidad del PARAFAC, hay que tener en cuenta dos aspectos: el orden y la escala de los factores. El hecho de que no hay un orden intrínseco significa que el componente debe ser identificado después de la construcción del modelo. El que la escala no esté definida significa que tan solo se puede determinar espectros, concentraciones etc. realizando un escalado. No se puede medir la absorbancia de una especie y determinar directamente la concentración. Se necesita una referencia que proporcione la escala. Por tanto, si se conoce la concentración en una muestra, se podrá escalar los resultados para obtener la concentración deseada y, en consecuencia, el del resto de muestras.

1.4.2.3 Algoritmo: *Alternating least squares*.

La solución al modelo PARAFAC puede ser encontrada por *alternating least squares* (ALS) utilizando iteraciones sucesivas, en las que se asume que los *loadings* en dos de los modos son conocidos y, entonces, se estima el conjunto desconocido de parámetros del último modo.

El algoritmo general PARAFAC-ALS puede ser escrito como:

- (0) decisión en el número de componentes, F
- (1) inicialización de \mathbf{B} y \mathbf{C} .
- (2) Estimación de \mathbf{A} a partir de $\underline{\mathbf{X}}$, \mathbf{B} y \mathbf{C} por mínimos cuadrados.
- (3) Estimación de \mathbf{B} de la misma manera.
- (4) Estimación de \mathbf{C} de la misma manera.

- (5) Se prosigue con el paso (2) hasta convergencia (cambio pequeño en el ajuste o *loadings*.)

Las matrices **A**, **B** y **C** están definidas igual que anteriormente. En el paso (2) $\underline{\mathbf{X}}$ está desdoblada en una matriz **X** ($I \times JK$) y la columna f en la matriz **Z** ($F \times JK$) está definida como:

$$\mathbf{z}_f = (\mathbf{b}_f \ddot{\mathbf{A}} \mathbf{c}_f) \quad (32)$$

Entonces la estimación de **A** se lleva a cabo resolviendo por mínimos cuadrados:

$$\begin{aligned} \mathbf{X} &= \mathbf{AZ} \\ \mathbf{A} &= \mathbf{XZ}^T (\mathbf{ZZ}^T)^{-1} \end{aligned} \quad (33)$$

Para estimar **B**, $\underline{\mathbf{X}}$ es desdoblada en una matriz $J \times IK$ y ahora **Z** es la matriz $F \times IK$ calculada a partir de **A** y **C**. De una manera similar se procede con **C**.

El algoritmo ALS mejorará, en cada iteración, el ajuste del modelo. Si el algoritmo converge a un mínimo global, que es lo que sucede en los casos ideales, se encuentra la solución por mínimos cuadrados del modelo. Un punto negativo es el tiempo requerido para estimar los modelos, especialmente cuando el número de variables es elevado. A veces, son necesarias cientos o miles de iteraciones antes de conseguir la convergencia. Este punto es problemático cuando se necesita recalculan los modelos, por ejemplo, durante la detección de muestras anómalas (*outliers*). Para que el PARAFAC sea más manejable es importante desarrollar algoritmos más rápidos. La utilización de ordenadores más potentes podría solucionar el problema, pero siempre existe una tendencia a obtener conjuntos enormes de datos que no son compensados con el poder computacional. Existen algunos métodos de aceleración descritos en la bibliografía, como la compresión de los datos y la extrapolación, aunque en esta memoria no se ha utilizado ninguno de ellos [Bro, 1997; Harshman, 1970; Harshman, 1994; Kiers, 1991b]. Unos buenos valores iniciales para el algoritmo ALS podrían, en principio, acelerar el algoritmo y asegurar que se ha encontrado un mínimo global. Han sido propuestos varios tipos de iniciación, como por ejemplo, empezar desde valores iniciales aleatorios [Harshman, 1984a] o iniciar del algoritmo desde diferentes puntos de partida. En otros trabajos se propone usar valores iniciales basados en la descomposición en valores singulares [Sands, 1980; Burdick, 1990; Sánchez, 1990; Li, 1993].

1.4.2.4 Criterio de parada.

A veces sucede que pequeños cambios en el ajuste pueden estar asociados con grandes diferencias en los *loadings*, debido a que la superficie de respuesta de la función error es muy plana [Harshman, 1970]. Esto es cierto si algunas variables están altamente correlacionadas, por lo que es importante utilizar un criterio de parada conveniente. Como medida preventiva contra ello se puede poner en marcha el algoritmo por duplicado; si el algoritmo converge, las dos soluciones serán prácticamente idénticas. Un criterio común que se puede utilizar es detener las iteraciones cuando el cambio relativo en el ajuste entre iteraciones es inferior a un cierto valor (p. ej. 10^{-6}). En algunas situaciones, se utiliza como criterio de parada que el cambio en los *loadings* para dos iteraciones sucesivas sea muy pequeño [Tu, 1992].

1.4.2.5 Forzando la solución.

El forzar la solución del PARAFAC puede ser de ayuda en términos de interpretabilidad o estabilidad de la solución. En la etapa de modelado, el ajuste de este modelo será siempre menor que el ajuste de modelos no forzados, pero la obtención de un modelo más interpretable y real puede justificar su utilización y la disminución en el ajuste.

Una de las restricciones utilizadas en quimiometría es, por ejemplo, forzar la no negatividad de los *loadings* en un conjunto de datos espectrales. Esta restricción es escogida a partir del conocimiento previo de los datos, ya que los espectros de absorbancia deben ser positivos si se ha utilizado un blanco correcto. Pueden ser aplicadas otro tipo de restricciones, como es el caso de la ortogonalidad [Harshman, 1994].

1.4.2.6 Preprocesado de los datos.

El procesado de los datos tridimensionales es más complicado que en el caso de los bidimensionales, debido a la mayor dimensionalidad de los mismos. El centrado de un modo puede ser efectuado desdoblado primero los datos de calibración en una matriz $I \times JK$ y centrando esta matriz como en el PCA ordinario:

$$X_{ijk}^{\text{cent}} = X_{ijk} - \bar{X}_{jk} \quad (34)$$

donde

$$\bar{x}_{jk} = \frac{\sum_{i=1}^I x_{ijk}}{I} \quad (35)$$

Esto se refiere a un centrado simple y se denomina “centrado a través del primero modo” [Ten Berge, 1989]. El centrado puede ser aplicado a cualquiera de los modos, dependiendo del problema. Si se tiene que efectuar el centrado a través de más de un modo se debe efectuar primero el centrado en un modo y, posteriormente, centrar nuevamente el resultado obtenido. En la bibliografía [Ten Berge, 1989; Kruskal, 1983; Harshman, 1984b] se describe el efecto de escalar y centrar en datos trilineales. El escalado también tiene que ser efectuado teniendo en cuenta la estructura tridimensional de los datos. Si la variable j del segundo modo tiene que ser escalada (comparada con el resto de variables en el segundo modo), es necesario escalar todas las columnas donde interviene la variable j . Esto significa que se deben escalar matrices enteras en vez de columnas. Matemáticamente el escalado puede ser descrito como,

$$x_{ijk}^{scal} = \frac{x_{ijk}}{s_i} \quad (36)$$

donde s_i puede ser definido como

$$s_i = \sqrt{\left(\sum_{j=1}^J \sum_{k=1}^K x_{ijk}^2 \right)} \quad (37)$$

Cuando se desea realizar el escalado dentro de diferentes modos, la situación se complica, porque el escalado de un modo afecta la escala de los otros. Si se desea escalar a varianza unidad dentro de varios modos, debe ser efectuado simultáneamente hasta convergencia.

En el software utilizado, se dispone de un programa en código MatLab para realizar el escalado iterativo y el centrado.

1.4.2.7 Determinación del número de componentes o factores.

Decidir el mejor rango de un modelo PARAFAC, es mucho más fácil que para problemas bidimensionales, debido a la propiedad de solución única que posee el algoritmo. En general, se dispone de los mismos métodos que se disponen para análisis bidimensional: observando los histogramas de residuales, valorando la interpretabilidad de los *loadings*, comparando la variación del residual con el nivel de ruido intrínseco y utilizando técnicas de reordenación de

muestras, de las que es un ejemplo el método de *cross validation* [Louwerse, 1999]. Si el modelo PARAFAC tiene que ser utilizado para calibración, como sucede en esta memoria, se puede utilizar la *cross validation* en la predicción de la variable dependiente para obtener el modelo óptimo. A menudo, se conocen aspectos en los datos que están relacionados con ciertos fenómenos, como es el caso del conocimiento de los espectros de los productos de reacción, y esto puede ser de gran importancia cuando se comparan diferentes modelos y se intenta buscar el número apropiado de componentes. En [Ross, 1995; Durell, 1990] se dan algunos ejemplos de cómo utilizar los residuales y el conocimiento externo del sistema para escoger el número de componentes óptimo.

1.4.2.8 Soluciones degeneradas.

A veces se pueden obtener soluciones degeneradas, las cuales son difíciles de interpretar. Esto significa que los modelos calculados son, a menudo, inestables y poco fiables. Un signo típico de soluciones degeneradas, es que los vectores de *loadings* del mismo modo están altamente correlacionados. A menudo, una solución degenerada se caracteriza por tener dos componentes PARAFAC que muestran igual forma de los vectores *loading* en todos los modos. Esto sucede cuando, por ejemplo, dos analitos presentan espectros de absorción y velocidades de reacción muy similares. La degeneración puede ser observada calculando el ángulo entre factores. Cuando el coseno del ángulo tiene un valor cercano a -1 significa la presencia de una degeneración temporal mientras se está calculando el modelo [Kruskal, 1989; Mitchell, 1993 y 1994].

Las soluciones degeneradas pueden ser un artificio del cálculo iterativo en la construcción del modelo (pseudo-degeneraciones) y pueden ser eliminadas realizando nuevos cálculos disminuyendo el criterio de convergencia. Si de esta manera no se eliminan las degeneraciones, éstas se denominan reales y pueden ser debidas a que [Harshman, 1984b; Kruskal, 1984]:

- (a) Se ha escogido un número elevado de factores y, por tanto, disminuyéndolo se obtendrán soluciones no degeneradas.
- (b) Se ha aplicado un pobre preprocesado.
- (c) El modelo es inapropiado: los datos no son trilineales como exige el modelo. Algunas

de estas situaciones son denominadas “degeneraciones de dos factores” [Kruskal, 1989] y cuando aparecen es preferible emplear el algoritmo de Tucker o modelos bilineales desdoblados para tratar los datos.

En presencia de degeneraciones, la bibliografía indica que puede ser preferible hacer varias veces el cálculo con pocas iteraciones y escoger aquellos cálculos que no estén sujetos a degeneraciones [Mitchell, 1993 y 1994]. Otra manera de solucionar el problema es aplicar al modelo restricciones de ortogonalidad [Kruskal, 1989] y no-negatividad [Krijnen, 1992]. También existen modelos alternativos, como el PARAFAC2, propuesto por Harshman, todavía no muy utilizado [Harshman, 1972; Kiers, 1993].

Software.

Bro ofrecen un conjunto de rutinas en MatLab para calcular modelos PARAFAC que pueden ser utilizados teniendo en cuenta restricciones de unimodalidad, no negatividad, ortogonalidad etc., y también proporcionan procedimientos para el preprocesado de los datos [Bro, 1998].

1.4.3 PLS Multi-lineal (Multi-linear PLS, N-PLS).

1.4.3.1 Introducción.

En este apartado se hace una extensión de la regresión PLS a los datos multidimensionales, haciendo énfasis en los datos tridimensionales característicos de esta memoria. El significado de descomposición trilineal en PLS es que la estructura cúbica $\underline{\mathbf{X}}$ se descompone en un conjunto de cubos de manera óptima y que, a diferencia del PARAFAC, el rango puede ser superior que el número de especies absorbentes. Las ventajas del N-PLS son varias:

- La principal característica del algoritmo es que produce vectores de *scores* que, en sentido trilineal, tienen máxima covarianza con la parte inexplicada de la variable dependiente.
- La solución es fácil de interpretar, comparada con los métodos *unfolding*, y esto es especialmente importante cuando se dispone de un número muy grande de variables.
- El algoritmo es rápido debido a la poca cantidad de parámetros a calcular y al hecho

de que el problema se reduce a una descomposición en valores propios.

- Otra ventaja de este método de calibración frente a otros métodos trilineales como el PARAFAC, es la posibilidad de estabilización de la solución en caso de baja contribución neta del analito, debido a la incorporación de la variable dependiente en la descomposición.

Pero, evidentemente, como en todos los métodos de calibración, hay ciertas desventajas. Se pierde ajuste en el modelo trilineal comparándolo con uno bilineal ya que hay restricciones más severas.

En la práctica, durante el proceso de calibración no se mira tan solo la falta de ajuste, sino también, si el modelo es el apropiado para el problema en concreto. Por tanto, se deben probar diferentes métodos para ver cuál es el más apropiado. Si diferentes métodos describen más o menos por igual los datos, se debe escoger siempre el más simple e interpretable.

El preprocesado de los datos tridimensionales para el N-PLS merece especial atención y se tienen en cuenta las mismas consideraciones que en el PARAFAC.

1.4.3.2 El modelo N-PLS.

En la versión trilineal del PLS los datos con estructura tridimensional se descomponen en un modelo similar al PARAFAC, pero en este caso el modelo se ajusta siguiendo la filosofía del PLS, describiendo la covarianza de las variables dependientes e independientes. Si asumimos que tanto el bloque de variables dependientes como el de independientes tienen estructura tridimensional, $\underline{\mathbf{X}}$ es el conjunto de variables independientes de dimensiones $I \times J \times K$ y $\underline{\mathbf{X}}$ la matriz desdoblada $I \times JK$. $\underline{\mathbf{Y}}$ es el conjunto de variables dependientes de dimensiones $I \times L \times M$ e $\underline{\mathbf{Y}}$ la matriz desdoblada de dimensiones $I \times LM$. De forma general el modelo N-PLS puede escribirse:

$$\underline{\mathbf{X}} = \mathbf{T}(\mathbf{W}^K | \ddot{\mathbf{A}} | \mathbf{W}^J)^T + \mathbf{E}_x \quad (38)$$

$$\underline{\mathbf{Y}} = \mathbf{U}(\mathbf{Q}^M | \ddot{\mathbf{A}} | \mathbf{Q}^L)^T + \mathbf{E}_y \quad (39)$$

Donde \mathbf{T} y \mathbf{U} son las matrices de *scores*, \mathbf{W} y \mathbf{Q} las de *loadings* (el superíndice indica el orden considerado) y \mathbf{E}_x y \mathbf{E}_y los errores para las variables independientes y dependientes, respectivamente. El modelo de regresión se construye asumiendo la relación lineal interna de las matrices de scores \mathbf{T} y \mathbf{U} ,

$$\mathbf{U} = \mathbf{TB} + \mathbf{E}_u \quad (40)$$

donde \mathbf{B} es una matriz diagonal.

1.4.3.3 Algoritmo.

El software utilizado en esta memoria, escrito en código MatLab, está basado en el algoritmo descrito por Bro [Bro, 1996 y 1998]. El algoritmo descompone $\underline{\mathbf{X}}$ (se asume que está centrada en la dirección de I) en conjuntos de tríadas. Una tríada consiste en un vector \mathbf{t} de *scores* y dos vectores de pesos, uno en el segundo orden llamado \mathbf{w}^J ($J \times 1$) y otro en el tercero llamado \mathbf{w}^K ($K \times 1$). El modelo de $\underline{\mathbf{X}}$ puede ser escrito como:

$$x_{ijk} = t_i w_j^J w_k^K \quad (41)$$

y el problema reside en encontrar los vectores \mathbf{w}^J y \mathbf{w}^K que satisfagan

$$\max \left[\text{cov}(\mathbf{t}, \mathbf{y}) / \min \left(\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (x_{ijk} - t_i w_j^J w_k^K)^2 \right) \right] \quad (42)$$

proporcionando un vector de scores \mathbf{t} que tenga máxima covarianza con \mathbf{y} . Cuando los pesos son conocidos y su módulo es 1, la solución por mínimos cuadrados al problema para encontrar \mathbf{t} se resuelve proyectando \mathbf{X}_i ($1 \times JK$) sobre los pesos

$$\max \left[\text{cov}(\mathbf{t}, \mathbf{y}) / \left(t_i = \sum_{j=1}^J \sum_{k=1}^K (x_{ijk} w_j^J w_k^K)^2 \right) \right] \quad (43)$$

La covarianza entre \mathbf{t} e \mathbf{y} puede ser expresada por el sumatorio

$$\max \left[\sum_{i=1}^I t_i y_i / t_i = \sum_{j=1}^J \sum_{k=1}^K (x_{ijk} w_j^J w_k^K)^2 \right] = \max \left[\sum_{j=1}^J \sum_{k=1}^K (z_{jk} w_j^J w_k^K)^2 \right] \quad (44)$$

Como $\underline{\mathbf{X}}$ e \mathbf{y} son conocidos, el sumatorio sobre i puede ser efectuado y se obtiene la matriz $\mathbf{Z} = \mathbf{X}^T \mathbf{y}$ de dimensiones $J \times K$ cuya SVD proporcionará los vectores \mathbf{w}^J y \mathbf{w}^K .

La ecuación 43 puede ser escrita utilizando el producto de Kronecker como

$$\max \left[\mathbf{t}^T \mathbf{y} / \mathbf{t} = \mathbf{X} (\mathbf{w}^K \otimes \mathbf{w}^J) \right] \quad (45)$$

A continuación se representa un esquema del algoritmo:

Centrado de $\underline{\mathbf{X}}$ e \mathbf{y}

$f=1$

- (a) Cálculo de la matriz \mathbf{Z} .
- (b) Determinación de \mathbf{w}^J y \mathbf{w}^K por SVD
- (c) Cálculo de \mathbf{t}
- (d) $\mathbf{b} = (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{y}$
- (e) Cada muestra \mathbf{X}_i es substituida por $\mathbf{X}_i - \mathbf{t}_i \mathbf{w}^J (\mathbf{w}^K)^T$ e $\mathbf{y} = \mathbf{y} - \mathbf{Tb}$
- (f) $f = f+1$. Se prosigue con el paso (a) hasta la descripción completa de \mathbf{y} .

Como los vectores *scores* de los diferentes factores no son ortogonales, el cálculo de los coeficientes de regresión en el paso (d) se realiza teniendo en cuenta todos los vectores *scores* calculados hasta el momento. Por tanto, \mathbf{T} tiene dimensiones $I \times F$. Para este método de calibración existen los mismos métodos para determinar el número de factores óptimos que para el caso del PLS bidimensional.

Se pueden hallar descritas en la bibliografía más propiedades del N-PLS [Smilde, 1997; De Jong, 1998; Bro, 1998].

1.5 MÉTODOS NO LINEALES: REDES NEURONALES ARTIFICIALES (ANN).

1.5.1 Introducción.

Las redes neuronales artificiales (ANN, *Artificial Neural Networks*) se pueden definir como un sistema iterativo de cálculo que intenta reproducir, de forma simple y sencilla, el sistema de conexiones que existe entre las neuronas del cerebro humano.

El tejido nervioso de los seres vivos se compone de unas unidades de transmisión denominadas neuronas. Éstas reciben una señal de entrada por medio de unos canales denominados dendritas que, en el momento en que sobrepasan un cierto valor, provocan que la neurona transfiera otra señal a través del axón a otras neuronas. La unión por la cual se transfiere la señal de una neurona a otras se denomina sinapsis, y la magnitud o características de esta señal es la fuerza sináptica.

En las redes neuronales artificiales, la neurona artificial (denominada simplemente neurona o nodo) intenta simular la neurona biológica [Wythoff, 1993]. La sinapsis se representa por una conexión entre dos neuronas y la fuerza sináptica por un peso asociado a esta conexión (un

número real). Las señales de entrada pasan a las neuronas, donde se realiza su suma ponderada. A continuación, se transforman pasando a través de una función de transferencia hacia la salida. La propagación de la señal viene determinada por las conexiones entre las neuronas y por sus pesos asociados (figura 4).

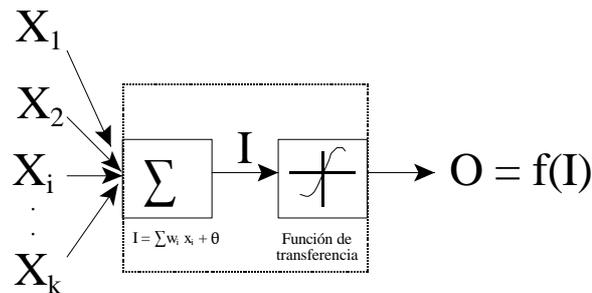


Figura 4. Representación de una neurona artificial.

La manera como están conectadas entre sí las neuronas individuales es lo que define las redes, la estructura de las cuales se acostumbra a plantear en términos de capas o niveles. De esta manera, se puede definir una “capa de entrada” como aquella en la que los datos se presentan a la red y una “capa de salida”, a partir de la cual se obtienen las respuestas, denominando al resto de capas como “capas ocultas”. El número de nodos que hay en cada capa define la “arquitectura” de la red neuronal, representada habitualmente por (i, h_i, o) , donde i es el número de neuronas de la capa de entrada, que normalmente es igual al número de variables; h_i el número de neuronas de la capa oculta i y o el número de neuronas de la capa de salida, que en esta memoria corresponde a las concentraciones de los analitos a determinar.

Un proceso llevado a cabo con ANN consta de dos etapas: la de aprendizaje (*learning*) y la de respuesta. Durante el aprendizaje, la red neuronal ‘aprende’ a partir de los ejemplos que se le presentan, adaptando los pesos de las conexiones en respuesta a las señales que le llegan de la capa de entrada y, opcionalmente, de la respuesta deseada. El aprendizaje puede ser supervisado, si existe una respuesta correcta a la que el método deba aproximarse; no supervisado, si no se conoce la respuesta que el método deba proporcionar; o por refuerzo si el aprendizaje no se basa en una respuesta correcta, sino únicamente en una indicación sobre si es buena o no, en función de lo cual, se produce un refuerzo o inhibición de las conexiones.

Para cada clase de aprendizaje existe una regla de aprendizaje que indica cómo deben adaptarse los pesos en respuesta a los ejemplos de los que está aprendiendo. En esta memoria, se ha utilizado la regla delta, basada en reducir el error entre la salida obtenida por la red y la salida deseada. Es, por tanto, un esquema de aprendizaje supervisado.

La etapa de respuesta se refiere a cómo la red procesa globalmente las señales que llegan a su capa de entrada y proporciona la respuesta en la capa de salida. La respuesta se integra continuamente en el proceso de aprendizaje comparándose con la respuesta deseada, creando así una función de error.

Existen diferentes tipos de redes neuronales y se utilizan unas u otras dependiendo del tipo de problema que se desee resolver. Todas ellas se definen por los tres elementos ya citados: los elementos de cálculo (neuronas), la arquitectura de la red (distribución y conexión entre las neuronas) y la regla de entrenamiento utilizada. En la bibliografía [Zupan, 1993; Zupan, 1991] se pueden encontrar diferentes tipos de redes, al igual que algunas aplicaciones de éstas en el ámbito de la química.

1.5.2 Redes Multi-Capas Perceptrón (MLP, Multi-Layer Perceptron).

La mayor parte de las aplicaciones en análisis químico de las redes neuronales se basan en los sistemas de multi-capas denominadas *multi-layer perceptron (MLP)*, también llamadas *multi-layer feed-forward network (MLF)*. Estas redes usan como algoritmo de aprendizaje el algoritmo de propagación de los errores hacia atrás (*back-propagation algorithm*). El tipo de aprendizaje de estas redes es supervisado, con conexiones totales entre neuronas de capas consecutivas y siempre hacia delante, es decir, los elementos de cada nivel se conectan con todos los elementos del siguiente nivel (figura 5).

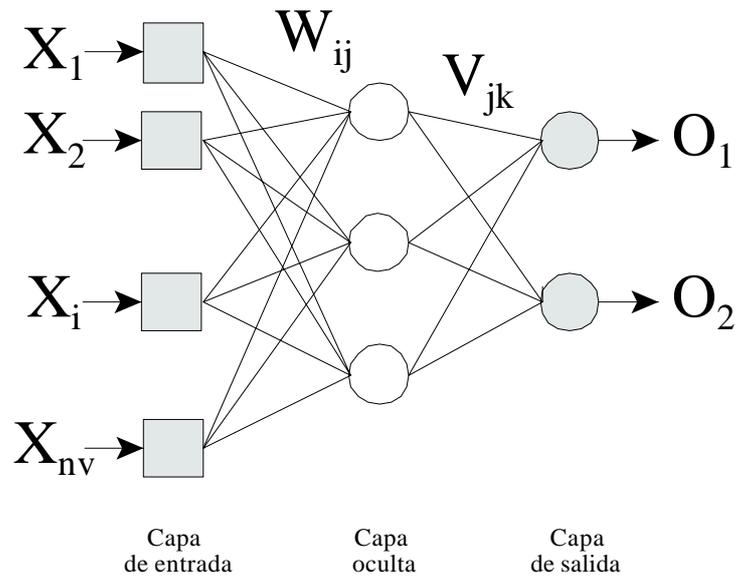


Figura 5. Estructura de una red MLP.

Estas redes se caracterizan por su capacidad de modelar correctamente sistemas en que la relación entre datos experimentales y la propiedad a determinar es no lineal [Long, 1990; Despagne, 1998], aunque también han sido aplicadas con éxito a problemas de clasificación [McCarrick, 1991].

La estructura de una red MLP viene definida inicialmente por una capa de entrada, con un número de neuronas igual al número de variables que definen a cada uno de los objetos. Esta capa es la que recibe la información y la traspasa a la siguiente capa, la capa oculta, que será la responsable de procesar y transmitir la información a la siguiente capa o a la capa de salida. El número de capas ocultas, así como las neuronas que las configuran, deben ser optimizadas en cada caso particular. La capa de salida siempre estará constituida por tantas neuronas como variables de salida deseemos obtener.

1.5.2.1 Propagación de una MLP.

Los datos de entrada \mathbf{X} son introducidos en la red a través de la capa de entrada y se relacionarán con los datos obtenidos en la capa de salida, *no*, a través de una o más capas ocultas de neuronas (*nh*: número de neuronas o nodos en el nivel oculto). Para resolver no linealidades es necesaria, al menos, una capa oculta. El número de nodos o neuronas está

directamente relacionado con la complejidad del sistema que se esté modelando.

El sistema estudiado lo podemos codificar con las siguientes matrices:

\mathbf{X} ($np \times nv$), matriz de datos x_{pi} , donde np es el número de muestras y nv el número de variables de entrada.

\mathbf{O} ($np \times no$), matriz que contiene las salidas o_{pk} obtenidas por la red.

\mathbf{W} ($nv \times nh$), matriz que contiene los pesos w_{ij} de las conexiones entre la entrada y el nivel oculto.

\mathbf{V} ($nh \times no$), matriz que almacena los pesos v_{jk} de las conexiones entre el nivel oculto y el de salida.

siendo $p = 1, \dots, np$, muestras u objetos, $i = 1, \dots, nv$, entradas o variables, $j = 1, \dots, nh$, nodos ocultos y $k = 1, \dots, no$, salidas de la red.

El objetivo es obtener, para una matriz de señales de entrada \mathbf{X} , una matriz \mathbf{Y} ($np \times no$) que contendrá los valores de salida conocidos, (y_{pk}) .

Las etapas que tienen lugar en la propagación son las siguientes:

(a) Los nodos de la capa de entrada ponderan la información recibida:

$$I_{pj} = \sum_{i=0}^{nv} x_{pi} w_{ij} \quad (46)$$

(b) Para cada valor de p y j sea $x_{p0} = 1$, w_{0j} es el sesgo asociado a la unidad j oculta. Dada una muestra x_p , el nodo oculto j , \mathbf{H}_j , recibe como entrada \mathbf{I}_{pj} y produce como salida,

$$H_{pj} = f(I_{pj}) = f\left[\sum_{i=0}^{nv} x_{pi} w_{ij}\right] \quad (47)$$

valor que será la entrada para el nivel de la salida de la red y donde f es la función de transferencia correspondiente a un nodo de la capa oculta (generalmente f será una función lineal o sigmoideal)

(c) Si ahora se define $\mathbf{H}_{p0} = 1$ para cada p , y se denota por v_{0k} el sesgo asociado a la neurona de salida k , ésta recibe como entrada,

$$I_{pk} = \sum_{j=0}^{nh} H_{pj} v_{jk} \quad (48)$$

y tiene por salida,

$$O_{pk} = g(I_{pk}) \quad (49)$$

siendo, $k = 1, \dots, no$, número de variables de salida y g la función de transferencia de un nodo de la capa de salida.

Por tanto, la ecuación general que describe la salida de una red MLP con una capa oculta, para cada o_{pk} y cada muestra x_{pi} , es:

$$O_{pk} = g \left[v_{0k} + \sum_{j=1}^{nh} v_{jk} f \left[w_{0j} + \sum_{i=1}^{nv} x_{pi} w_{ij} \right] \right] \quad (50)$$

Esta última expresión es indicativa de un modelo paramétrico no lineal, que nos sirve para relacionar \mathbf{Y} con todas las variables \mathbf{X} . Los pesos v_{jk} y w_{ij} normalmente se calculan mediante retropropagación.

1.5.2.2 Aprendizaje por retropropagación.

El algoritmo de retropropagación [Zupan, 1993], no es más que un aprendizaje supervisado en el que los pesos de las conexiones son corregidos en función de los valores de referencia de las salidas obtenidas. Este proceso de corrección de pesos puede hacerse de dos maneras: después de cada nueva entrada individual (“corrección inmediata”), o después de que todas las entradas hayan sido comprobadas (“corrección pospuesta”). En el primer caso, la corrección se efectúa inmediatamente después de que el error haya sido detectado; en el segundo, el error individual para todos los pares de datos se acumula y el error acumulado de todo el conjunto de aprendizaje se utiliza para la corrección. La más utilizada es la corrección inmediata. En la bibliografía puede hallarse más información sobre las ventajas e inconvenientes de los modos de corrección mencionados [Finnoff, 1994; López-Fandiño, 1997].

Durante el aprendizaje, los datos \mathbf{X} son presentados a la red neuronal y la matriz de salida, \mathbf{O} es inmediatamente comparada con la matriz de referencia \mathbf{Y} , que es el valor correcto de salida para los datos \mathbf{X} . Una vez se conoce el error producido por la red, **la regla delta generalizada** lo aprovecha para corregir los pesos a lo largo de la misma en la dirección opuesta a la entrada de datos, de ahí viene el nombre de retropropagación.

La regla delta generalizada propone que la corrección de pesos ΔW en una capa es proporcional a \mathbf{d} (función que depende del error) y a la entrada \mathbf{X} que, en este caso, es la salida de la capa anterior \mathbf{O}^{ant} . Una vez corregido el vector de pesos, la señal de salida

debería aproximarse más al valor de salida conocido. La regla delta se puede expresar como:

$$\Delta W = \eta \delta X \quad (51)$$

donde h es la constante de proporcionalidad llamada “velocidad de aprendizaje” (*learning rate*), que nos determina la velocidad a la que serán cambiados los pesos y d es la constante de corrección (función del error) que se pretende buscar.

El algoritmo de corrección sigue los siguientes pasos:

(a) Dada la entrada de un objeto o muestra $\mathbf{x}_p(x_1, x_2, \dots, x_i, \dots, x_{nv})$, en primer lugar se etiquetan los componentes x_i del objeto de entrada \mathbf{x}_p como o_i^0 y se añade un componente de valor 1 como sesgo; así, el vector de entrada se convierte en $\mathbf{o}^0(o_1^0, o_2^0, \dots, o_{nv}^0, 1)$.

(b) Seguidamente, se propaga \mathbf{o}^0 a través de las capas de la red, evaluándose el vector de salida \mathbf{o}^l siendo l la capa correspondiente. Para ello se utilizan los pesos w_{ij}^l de la l -capa y la salida o_i^{l-1} de la capa anterior (que actúa de entrada de la capa l):

$$o_j^l = f \left(\sum_{i=1}^m w_{ij}^l o_i^{l-1} \right) \quad (52)$$

donde f es la función de transferencia y m el número de neuronas de la capa $l-1$.

Cada uno de los valores o_j serán ahora las entradas a las neuronas de la siguiente capa, volviéndose a calcular otros o_j para esta capa, y así sucesivamente hasta la última capa. En ella se obtiene un valor o_j^{ult} , que es el valor de respuesta obtenido por la red.

(c) A continuación, se calcula el factor de corrección para todos los pesos en la capa de salida d_j^{ult} utilizando el vector \mathbf{o}_j^{ult} y el valor de \mathbf{y} conocido, es decir el de referencia:

$$\delta_j^{ult} = (y_j - o_j^{ult}) \quad (53)$$

Este error debe multiplicarse por la derivada de la función de transferencia utilizada. Por ejemplo, si la función de transferencia es la función lineal la expresión no queda modificada. Sin embargo, si la función es sigmoidea el factor de corrección sería:

$$\delta_j^{ult} = (y_j - o_j^{ult}) o_j^{ult} (1 - o_j^{ult}) \quad (54)$$

(d) Calculado el error para todas las neuronas de la última capa, se corrigen entonces sus pesos, w_{ij}^{ult} :

$$\Delta w_{ij}^{ult} = \eta \delta_j^{ult} o_i^{ult-1} + \mu \Delta w_{ij}^{ult(\text{previa})} \quad (55)$$

Esta ecuación corresponde a la fórmula de la regla delta a la que se ha añadido un segundo término en el que aparece la constante m (*momentum*), multiplicada por la corrección de pesos realizada en la iteración previa. Esta constante es utilizada para prevenir cambios bruscos en la dirección en la que están siendo efectuadas las correcciones, evitando así que el modelo se pare en un mínimo local.

Se puede observar que, para la corrección de pesos, se tiene en cuenta el valor de salida de esta capa, o_j^{ult} (que está incluido en d_j^{ult}), y el valor de salida de la capa anterior o_i^{ult-1} (que es el valor de entrada de esta capa).

(e) Consecutivamente, se calculan los errores d_j^l para cada una de las capas ocultas, desde $l=(\text{última}-1)$ hasta $l=1$:

$$\delta_j^l = \left(\sum_{k=1}^r \delta_k^{l+1} w_{jk}^{l+1} \right) o_j^{ult} (1 - o_j^{ult}) \quad (56)$$

donde r indica el número de neuronas de la capa $l+1$.

(f) Finalmente, se corrigen entonces todos los pesos w_{ij}^l en la capa l :

$$\Delta w_{ij}^l = \eta \delta_j^l o_i^{l-1} + \mu \Delta w_{ij}^{l(\text{previa})} \quad (57)$$

(g) Todo este proceso se repite con una nueva pareja (\mathbf{X}, \mathbf{Y}) .

La constante de velocidad de aprendizaje es un parámetro que debe tenerse en cuenta, ya que determina la rapidez con la que van a cambiar los pesos. Si éstos se modifican rápidamente, el procedimiento puede finalizar en un mínimo local donde el error de predicción puede oscilar o crecer. Si la velocidad de cambio es pequeña, el tiempo de aprendizaje de la red puede ser demasiado elevado. Desafortunadamente, no hay un valor óptimo de este parámetro y dependerá, en particular, de la red definida, del número de muestras, etc. Tampoco puede darse un valor óptimo para la constante m . Su valor será función de la velocidad de aprendizaje [Barnard, 1992].

1.5.2.3 Consideraciones prácticas para la construcción de una red.

Vamos a tratar ahora diferentes aspectos prácticos que deben tenerse en cuenta cuando se

desarrollan métodos de calibración utilizando redes MLP. Podemos distinguir dos tipos de cuestiones: las relacionadas con las muestras y aquellas que tienen que ver con la búsqueda de la topología óptima de la red.

1.5.2.3.1 Muestras.

En principio, las muestras que se utilizan en el aprendizaje deben definir de forma correcta el espacio de aplicación de la red, ya que la ecuación construida no es una descripción del sistema completo sino que está restringida por los objetos utilizados en el proceso de aprendizaje.

a) Conjuntos de muestras.

Cuando se trabaja con redes neuronales son necesarios, como mínimo, tres conjuntos de muestras. En la etapa de construcción del modelo se utilizan dos de los conjuntos: el de aprendizaje (*training set*) y el de validación (*test set*). Este segundo conjunto de objetos es independiente del de aprendizaje, pero debe ser representativo del mismo. Se utiliza para optimizar el modelo y evitar un posible sobreajuste. Finalmente, para asegurar la capacidad predictiva real de la red construida debe utilizarse un tercer conjunto de muestras, conjunto de predicción externo (*external prediction set*).

La selección de estos conjuntos está fuertemente influenciada por el tipo de red a utilizar pero siempre ha de mantenerse la representatividad del sistema en estudio en los tres conjuntos. Es importante evitar introducir la dependencia temporal o de orden de asignación de los objetos seleccionándolos de forma aleatoria o según un diseño experimental previamente definido.

b) Número de muestras.

Durante el proceso de aprendizaje se optimizan los pesos de la red a partir de los ejemplos que se utilizan en el conjunto de entrenamiento. Por tanto, será necesario conseguir el número de objetos suficientes para poder modelar la relación entre \mathbf{X} e \mathbf{Y} . El número de pesos a optimizar va a aumentar, lógicamente, al aumentar el número de neuronas de las capas, y ello determina el número de objetos a considerar. Por ejemplo, con una sola capa oculta, el número de pesos corresponde a:

$$n^{\circ} \text{ de pesos} = (k+1) \times h + (h+1) \times p \quad (58)$$

donde k es el número de neuronas de la capa de entrada, h el número de neuronas de la capa oculta y p el número de neuronas de la capa de salida.

Se observa claramente cómo el número de pesos (es decir, el número de parámetros ajustables) crece rápidamente al aumentar el número de neuronas, con lo que necesitaremos un gran número de objetos en el conjunto de aprendizaje para evitar tener más parámetros que datos. En general, se recomienda [Zupan, 1993] que el número mínimo de muestras en el conjunto de aprendizaje sea el doble que el número de pesos totales. Otros autores aconsejan que el número de muestras sea 10 veces el número de pesos a ajustar [Bos, 1993]. Por otro lado, también van a ser necesarias más muestras para la validación del modelo. La gran cantidad de muestras necesarias es una de las mayores limitaciones prácticas en el uso de redes neuronales.

c) Detección de objetos anómalos.

Este es un paso difícil en redes neuronales. Si previamente a la utilización de las redes neuronales se ha aplicado alguna otra técnica de calibración multivariable, como es el caso de un PLS o PCR, puede suceder que algunas muestras hayan sido consideradas *outliers* y, sin embargo, sean modeladas correctamente por las redes neuronales. La manera más habitual y sencilla de hallar objetos anómalos es por medio de representaciones gráficas de los datos \mathbf{X} e \mathbf{Y} . La representación puede simplificarse si se realiza un análisis en componentes principales, PCA.

d) Pretratamiento de datos.

Cada tipo de variable presentada al modelo neuronal tiene un rango de operación individual. Algunas variables tendrán un valor medio más alto y/o una varianza más elevada que el resto. Si estas variables son aplicadas directamente al modelo neuronal, es de suponer que tendrán unos pesos asociados más grandes. Como consecuencia, estas variables de entrada tendrán una gran influencia en la respuesta del modelo.

Al existir los términos de sesgo en las capas de entrada y las ocultas, no es necesario centrar las variables. Tampoco es necesario escalarlas a varianza unidad ya que el método no se basa

en la maximización de la varianza-covarianza. Normalmente las variables se escalan en un cierto intervalo (*range scaling*) de forma que queden distribuidas adecuadamente según la función de transferencia escogida. Análogamente se recomienda este escalado para las variables \mathbf{Y} .

e) Número de variables de entrada.

Ya hemos dicho que la relación muestras/pesos debe ser lo mayor posible. En general, cuando se tiene un número elevado de objetos y se ha controlado correctamente el proceso de toma de datos, la red neuronal puede ser construida utilizando datos originales. Sin embargo, esta situación no es la más frecuente. Normalmente, necesitaremos reducir significativamente el número de variables de entrada para que el número de pesos a utilizar esté en concordancia con la ecuación 58. Ello puede realizarse a partir de métodos que permitan disminuir la dimensionalidad inicial de nuestro sistema como puede ser, por ejemplo, la utilización de los *scores* obtenidos en un análisis en componentes principales (PCA) como variables de entrada a la red. Es una buena estrategia y proporciona ciertas ventajas adicionales, como es una mayor velocidad de aprendizaje, la disminución del riesgo de sobreajuste y el aumento de la robustez de la red cuando el número de objetos que son utilizados en el proceso de aprendizaje es pequeño.

De todos modos, la utilización de técnicas de reducción de variables debe realizarse siempre en el conjunto de entrenamiento y, a partir de los valores hallados, calcular los *scores* de los objetos que configuran los otros dos conjuntos de muestras.

1.5.2.3.2 Topología de la red: diseño y optimización.

Una vez se dispone de los datos y objetos representativos del problema, y se han pretratado convenientemente, el siguiente paso es el diseño de la red.

a) Arquitectura de la red.

La selección de la arquitectura adecuada, es decir, la elección del número de capas y el de

neuronas en cada una de ellas, es un punto de gran importancia para conseguir modelar la relación existente entre unos datos de entrada y los de salida.

El número de neuronas de la capa de entrada y de salida viene determinado por la naturaleza del problema. Tendremos tantas neuronas de entrada como variables (originales o procedentes de una reducción previa) y tantas neuronas en la capa de salida como parámetros a estimar (ej. concentraciones). En general, con una sola capa oculta es suficiente. El número de neuronas que conforman esta capa oculta debe optimizarse. La utilización de un número de neuronas inferior al adecuado provoca que la red no funcione correctamente, mientras que el uso de un número demasiado grande puede crear problemas de sobreajuste. Por tanto, es conveniente ensayar diferente número de neuronas en la capa oculta y escoger aquella red que proporcione un mínimo error sin llegar a sobreajustar el sistema.

b) La función de transferencia.

En la capa de entrada, la función de transferencia es una función de paso, la cual sólo introduce los datos originales en la red sin hacerles ninguna transformación.

Las funciones que se utilizan en las neuronas de la capa oculta deben cumplir una serie de condiciones para poder aplicar la regla de aprendizaje:

- S** Ser monótona creciente en el intervalo de aplicación (implica que la función es continua en el intervalo y está acotada).
- S** Tener una derivada continua en dicho intervalo.

Generalmente, en este tipo de redes, la función de transferencia de las neuronas de la capa oculta es del tipo sigmoideal. Una de las más habituales es la denominada función logística:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (59)$$

También se utiliza la función tangente hiperbólica:

$$f(x) = \tanh(x) \quad (60)$$

Las funciones de transferencia utilizadas en las neuronas de la capa de salida pueden ser lineales o no lineales.

c) Otros parámetros optimizables.

Existen otros parámetros a optimizar que pueden influir de manera decisiva en la velocidad de aprendizaje y en el comportamiento de la red, como pueden ser los valores iniciales de los pesos de las conexiones y su distribución (uniforme o gaussiana) que, en general, deben ser valores pequeños y aleatorios; el método de actualización de los pesos, es decir, corrección inmediata o corrección pospuesta; los valores de los parámetros velocidad de aprendizaje, η , y momento μ , ya definidos anteriormente etc.

1.5.3 Entrenamiento y evaluación de una red neuronal.

Cuando se está desarrollando un modelo neuronal el objetivo es encontrar la solución global en el dominio de un problema, el cual puede contener una gran cantidad de soluciones subóptimas o locales.

Una solución global se obtiene con aquel modelo que produce el mínimo error posible. La red ideal es aquella que proporciona un error cero. Generalmente, encontrar este modelo no es posible, aunque debemos aproximarnos a él. Para ello, necesitamos controlar la variación del error cometido por las distintas redes o, lo que es lo mismo, por las combinaciones de pesos que vamos obteniendo durante el aprendizaje.

La mayoría de las redes neuronales aprenden para obtener salidas que minimicen el error entre los valores obtenidos y los de referencia. Sin embargo, y como se observa en la figura 6, cualquier problema complejo tiene varias soluciones subóptimas. Para evitar que el proceso de aprendizaje se pare en una de estas soluciones es importante definir correctamente la configuración de la red y también realizar una adecuada división de todos los objetos o muestras disponibles en los diferentes conjuntos con los que se trabaja: entrenamiento y validación.

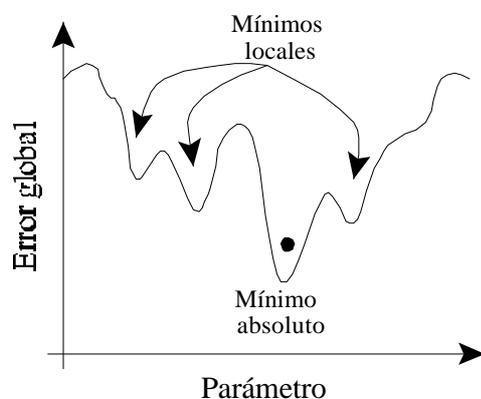


Figura 6. Posibles soluciones de una ANN.

El comportamiento de una ANN supervisada durante el proceso de aprendizaje se suele seguir utilizando el RMS (*root mean square error*),

$$\text{RMS} = \sqrt{\frac{1}{np} \sum_{i=1}^n \sum_{j=1}^p (y_{ij} - \hat{y}_{ij})^2} \quad (61)$$

donde n es el número de muestras en el conjunto considerado (entrenamiento, validación o externo), p el número de neuronas de la capa de salida, y el valor de referencia de la respuesta e \hat{y} la respuesta calculada por la red. Puesto que es una medida cuadrática del error, la utilización del RMS tiende a amplificar las grandes diferencias y a amortiguar las pequeñas. El software utilizado en el desarrollo de este trabajo calcula este error individualmente para cada salida. Por tanto, si lo que se está cuantificando es la concentración de varios analitos se podrá observar el ajuste de la red para cada especie, independientemente, mediante la evolución del RMS_j para cada iteración (figura 7).

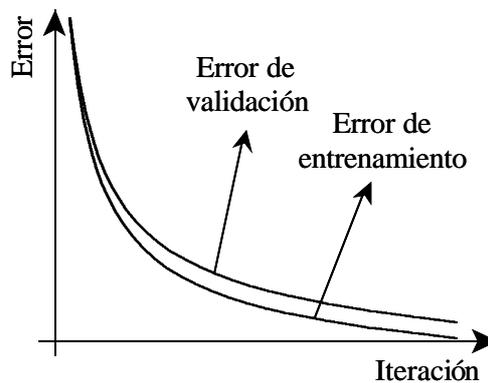


Figura 7. Variación del error en el entrenamiento de una ANN.

La evolución del RMS_j a lo largo de todo el proceso de entrenamiento proporciona información importante. Podemos encontrarnos ante situaciones muy variadas:

- a) Tanto el RMS_j del conjunto de entrenamiento como el de validación tienden hacia un mínimo (situación óptima).
- b) El RMS_j converge con dificultad ya que se producen grandes oscilaciones en torno a valores mínimos. Se puede solventar tomando un valor alto al principio de velocidad de aprendizaje, η , e ir disminuyéndolo a medida que transcurre el entrenamiento.
- c) El entrenamiento puede detenerse o evolucionar muy lentamente debido a que la red ha caído en un mínimo local y no puede salir de él. Para solucionarlo deben tomarse pesos iniciales aleatorios más pequeños y/o volver a escalar las variables de entrada de otra forma. Es decir, se reinicia el cálculo cambiando la configuración de la red.
- d) El fenómeno de sobreentrenamiento (*overtraining*) se puede detectar claramente porque la curva de validación no sigue la misma tendencia que la de entrenamiento. A partir de un momento determinado se separan y la curva de validación comienza a subir para después estabilizarse. La situación puede ser debida a dos motivos: el número de neuronas de la capa oculta es demasiado alto o se han efectuado demasiadas iteraciones de entrenamiento. En general, la disminución del número de neuronas en la capa oculta suele acabar con el problema.
- e) La última situación sucede cuando hay una falta de concordancia entre el conjunto de entrenamiento y el de validación. La curva de validación, aunque sigue la misma tendencia que la de calibración, se separa de ésta después del inicio y no presenta mínimo. No se han

repartido correctamente las muestras disponibles entre los dos conjuntos, por tanto, es necesario redistribuirlas.

2. MÉTODOS CINÉTICOS DE ANÁLISIS

2.1 INTRODUCCIÓN.

Los métodos cinéticos se basan en medidas de los cambios que tienen lugar en el transcurso de una reacción química o proceso físico, a diferencia de los métodos basados en medidas en el equilibrio o estado estacionario. En este apartado, se abordará la cinética de reacciones químicas y su interés en química analítica, pero hay que tener en cuenta que los métodos cinéticos son más generales y se pueden aplicar a muchos fenómenos físicos dependientes del tiempo, como los fenómenos de difusión, migración, etc.

Si consideramos el caso de una reacción química, de ecuación general



donde A es el analito, R es un reactivo que se considera en gran exceso, y P el producto de reacción, la figura 8 muestra los cambios de concentración con el tiempo

En los métodos cinéticos, las medidas son efectuadas en la zona que se denomina región cinética (figura 8). La velocidad de formación del producto o de desaparición de A es la velocidad de la reacción, v , y puede ser usada como una medida de la concentración inicial de A. En los métodos basados en el equilibrio, las medidas son efectuadas en la región donde las concentraciones ya no cambian. Tales métodos, como determinaciones espectrofotométricas, varios tipos de valoraciones, determinaciones gravimétricas y muchas otras, están basadas en medidas obtenidas en condiciones estáticas después del equilibrio o cuando se ha alcanzado el estado estacionario. La cantidad total de P producida está relacionada con la concentración inicial de A.

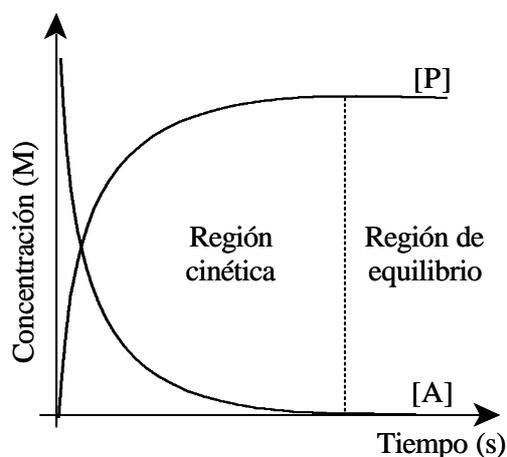


Figura 8. Región cinética y de equilibrio durante una reacción química.

Los métodos cinéticos presentan ventajas respecto los métodos basados en el equilibrio:

- a) Pueden ser más rápidos. Si las medidas son efectuadas en la región cinética, no es necesario esperar a que se llegue al equilibrio.
- b) Son medidas relativas y libres de interferencias de fondo. Los métodos cinéticos son inherentemente relativos y no dependen de las concentraciones absolutas, sino de los cambios de concentración.
- c) Pueden ser más selectivos. Los métodos cinéticos pueden ser más selectivos si las especies interferentes reaccionan con el reactivo R a una velocidad más lenta o más rápida que el analito de interés.
- d) Estas diferencias de velocidad pueden ser aprovechadas para desarrollar determinaciones cinéticas de multicomponentes, llamadas en muchas ocasiones métodos cinéticos diferenciales.
- e) Debido a que las medidas son, a menudo, más rápidas con los métodos cinéticos, muchas reacciones muy lentas son analíticamente útiles. Esto amplía el rango de reacciones útiles, desde reacciones que requieren horas para llegar al equilibrio, hasta las que lo alcanzan en milisegundos.
- f) Finalmente, los métodos cinéticos permiten la determinación de la concentración de catalizadores, los cuales, por supuesto, no pueden ser determinados por los métodos basados en el equilibrio.

Sin embargo, los métodos cinéticos presentan ciertas desventajas. Como principales limitaciones nos encontramos con:

- a) Tienen menos precisión que los métodos basados en el equilibrio. Esto es debido a que las medidas se basan en pequeños cambios en la señal (absorbancia, fluorescencia etc.) a diferencia al cambio total disponible cuando se alcanza el equilibrio.
- b) Son a menudo bastante dependientes de las condiciones experimentales (T, pH, fuerza iónica, concentración de reactivos, etc.)
- c) La medida de pequeños cambios de la concentración requiere sistemas de medida muy sensibles.
- d) Las velocidades de reacción deben de estar como máximo dentro del rango del sistema de mezcla ($t_{1/2}$ desde ms al minuto).

En el pasado, la utilización de los métodos cinéticos estaba desmotivada por estas limitaciones, en particular por la dependencia de las condiciones de reacción. Actualmente, los métodos automatizados pueden reproducir las condiciones experimentales y mezclar cantidades exactas de muestra y reactivos en el sistema de mezcla de forma muy rápida y reproducible. Igualmente, se están desarrollando métodos de cálculo muy potentes que permiten procesar toda la información obtenida de forma más precisa.

2.2 CONCEPTOS BÁSICOS.

Las reacciones se pueden clasificar en homogéneas y heterogéneas. Una reacción homogénea ocurre enteramente en una fase y una reacción heterogénea implica la presencia de especies en dos o más fases. En esta memoria, las reacciones utilizadas se caracterizan por ser homogéneas en disolución, por tanto, tan solo haremos mención de los aspectos que conciernen a este tipo de reacciones.

Así si consideramos la reacción homogénea:



donde a, b, ..., e, f, ... son los coeficientes estequiométricos y A, B, ..., E, F, ... las especies químicas. La velocidad a la cual cualquier reactivo se consume es proporcional a su

coeficiente estequiométrico; por tanto

$$\text{velocidad} = v = -\frac{1}{a} \frac{d[A]}{dt} = -\frac{1}{b} \frac{d[B]}{dt} = \dots = \frac{1}{e} \frac{d[E]}{dt} = \frac{1}{f} \frac{d[F]}{dt} = \dots \quad (62)$$

Para muchas reacciones (pero no todas), la forma de v hallada experimentalmente es

$$v = k[A]^\alpha [B]^\beta \dots \quad (63)$$

donde los exponentes α , β ,... son, en general, enteros o semienteros. La constante de proporcionalidad k , denominada constante cinética (de velocidad) o coeficiente de velocidad, depende de la temperatura y, en menor grado, de la presión. Se dice que la reacción tiene orden α respecto a A, orden β respecto a B, etc. Los exponentes α , β ,... se denominan también órdenes parciales y la suma de ellos es el orden total, n , de la reacción. La expresión de v en función de las concentraciones, a una temperatura dada, se denomina ecuación cinética.

La determinación de una especie mediante un método cinético se basa en la medida directa o indirecta de la velocidad de reacción. Para ello, deben medirse los cambios de concentración de los reaccionantes o productos en función del tiempo. Estas concentraciones pueden ser substituidas por alguna magnitud de medida que sea directamente proporcional a las mismas, normalmente, alguna propiedad física de la disolución: absorbancia, fluorescencia, potencial, temperatura, conductividad, etc. La curva de la señal obtenida en función del tiempo es de tipo exponencial, positiva o negativa, según se monitorice la formación del producto o la desaparición de algún reactivo. Así pues, si suponemos la reacción genérica



La velocidad en un tiempo t viene dada por:

$$v = \frac{d[P]}{dt} = -\frac{d[A]}{dt} = -\frac{d[B]}{dt} \quad (64)$$

o bien,

$$v = \frac{d[P]}{dt} = k[A][B] \quad (65)$$

donde k es la constante de velocidad. Para el análisis cinético de la reacción, el tramo más interesante es el primero, cuando la reacción apenas se ha desarrollado en un 1-3%, ya que éste es lineal, siendo su pendiente proporcional a la concentración de la especie determinada.

Este aspecto es muy importante cuando se está poniendo a punto un método cinético diferencial, ya que un estudio preliminar de las especies puras a determinar, puede indicar si existe suficiente diferencia en las constantes de velocidad o, también, si se ven afectadas de forma distinta por los cambios de temperatura.

2.2.1 Cálculo de las constantes de velocidad.

Cuando en la ecuación anterior se mantiene constante y en exceso la concentración de uno de los reactivos (B, por ejemplo), podemos suponer que su concentración a lo largo de la reacción no se modifica, incluyéndose el valor de [B] en la constante k [Pérez-Bendito, 1984]. Se dice entonces que la reacción es de pseudo-primer orden con respecto a [A].

$$v = k_{ap} [A] \quad (66)$$

Si S_o es la señal al comienzo de la reacción y S_t la señal en un tiempo t, tendremos que

$$\log(S_t - S_o) = \log(S_\infty - S_o) - \frac{kt}{2.303} \quad (67)$$

y la representación gráfica de $\log(S_t - S_o)$ en función de t es una línea recta cuya pendiente negativa vale $k/2.303$, lo que nos permite conocer el valor de k. Para las reacciones de orden superior a 1 se llega a otras expresiones.

2.2.2 Cálculo de la energía de activación, E_a .

Los métodos cinéticos exigen el perfecto control y optimización de las variables experimentales. Éstas se refieren, fundamentalmente, a la temperatura, concentración de reactivos, pH, fuerza iónica y constante dieléctrica, así como la existencia de reacciones controladas. La constante de velocidad se ve modificada por la temperatura de la forma expresada por la ecuación de Arrhenius [Pérez-Bendito, 1984]:

$$k = Ae^{-\frac{E_a}{RT}} \quad (68)$$

siendo T la temperatura absoluta y R la constante de los gases; E_a es la energía de activación y A una constante denominada factor de frecuencia o de Arrhenius. Si se conocen los valores que adopta k para distintas temperaturas y transformamos la ecuación de Arrhenius en su forma logarítmica:

$$\log k = \log A - \frac{E_a}{2.303RT} \quad (69)$$

la representación gráfica de $\log k$ en función de $1/T$ es la ecuación de una recta de pendiente $-E_a/2.303 R$, que permite calcular la energía de activación de la reacción.

2.2.3 Cálculo de órdenes parciales de reacción.

Para proponer una ecuación de velocidad de una reacción química es necesario el cálculo previo de los órdenes parciales de reacción respecto a las distintas especies que intervienen en el equilibrio. Este cálculo se realiza variando convenientemente la concentración de la especie en cuestión, manteniendo constante y en exceso el resto de las concentraciones [Pérez-Bendito, 1984]. De esta manera, la velocidad de reacción solamente se verá afectada por la especie cuya concentración está en defecto.

Si, a partir de las curvas $S(t)$ obtenidas para cada concentración de la especie cuyo orden queremos obtener C , calculamos sus velocidades iniciales (tangente del perfil cinético a tiempos iniciales) tendremos

$$\text{velocidad inicial} = v_o = k_{ap} [C]^n \quad (70)$$

donde n es el orden en cuestión y k_{ap} la constante de velocidad aparente, que engloba las concentraciones de las demás especies.

$$\log v_o = \log k_{ap} + n \log C \quad (71)$$

Se define así una recta en función del $\log C$ cuya pendiente es n . Lo mismo se haría con los demás reactivos, hasta establecer la ecuación cinética.

2.3 INSTRUMENTACIÓN.

2.3.1 Introducción

Las medidas de la señal analítica en un sistema cinético requieren varias unidades modulares para llevar a cabo el seguimiento de la reacción. Conceptualmente estas unidades se utilizan para:

- a) mezcla rápida y a temperatura constante de los reactivos,
- b) monitorización del progreso de la reacción (sistema de detección) a temperatura

constante,

- c) sistemas electrónicos auxiliares para la adecuación de la señal y su modificación, y
- d) procesado y tratamiento de la señal.

El proceso que acompaña al seguimiento de la reacción se puede dividir en tres pasos esenciales susceptibles de ser automatizados:

- a) preparación, transporte y mezcla de las especies involucradas en la reacción,
- b) medida y transducción de la variación de la señal analítica con el tiempo a temperatura constante, y
- c) procesado de gran cantidad de datos mediante ordenadores.

El seguimiento de los cambios químicos se lleva a cabo, principalmente, mediante métodos espectroscópicos, pero también con técnicas electroanalíticas. La señal obtenida es convertida, mediante convertidores analógico-digitales, a una señal numérica que, posteriormente, será procesada para obtener la información deseada.

A pesar de que en cada uno de los niveles de automatización anteriormente mencionados hay muchos aspectos a tratar, en este apartado tan solo haremos referencias a algunos que tienen especial interés en esta memoria.

2.3.2 Sistemas de mezcla.

El sistema de mezcla de reactivos depende de la velocidad de cambio del proceso que está siendo monitorizado. Va desde la mezcla manual y agitación magnética hasta sistemas más complicados, como el de flujo continuo (*continuous flow analysers*), flujo continuo segmentado (*air-segmented continuous-flow analysers*) o el de flujo interrumpido (*stopped-flow analysers* y *stopped-flow-FIA analysers*). Para reacciones con un tiempo de vida media de aproximadamente 10s o menos, el sistema de *stopped-flow* es una buena alternativa.

Los métodos cinéticos de análisis basados en reacciones lentas no son automatizados fácilmente. Normalmente, se utilizan métodos simples y convencionales de mezcla, como la manual, la agitación magnética o ambos y los analizadores de flujo continuo en especial los segmentados con aire (*air-segmented continuous flow analysers*). En estos últimos el problema de la dispersión que aparece en los analizadores de flujo convencionales se minimiza y el sistema está muy bien desarrollado.

El análisis por inyección de flujo (*flow injection analysis, FIA*) también ha sido propuesto para medidas cinéticas. Sin embargo, éste no es ideal, debido a la dispersión física que tiene lugar en el sistema. Para solucionar algunos inconvenientes de esta técnica, se han utilizado los *stopped-flow-FIA analysers*.

Cabe mencionar también, los sistemas de flujo reverso (*flow reversal analysers*) y los de recirculación (*flow re-cycle analysers*), situados entre los de flujo convencional, que tan solo proporcionan como datos un par de puntos, y los sistemas de flujo interrumpido, donde se puede obtener más información realizando el seguimiento completo de la reacción. Éstos pueden ser, o no, segmentados con aire. Estos sistemas también minimizan el problema de la dispersión y se utilizan en reacciones lentas.

2.3.2.1 Mezcla manual-agitación magnética.

Este sistema ha sido el más utilizado en la realización de esta memoria (anexos 1, 2, 4 y 5). De las disoluciones a mezclar, se introduce en la cubeta de medida, con la ayuda de micropipetas, la o las que van a ser utilizadas para la corrección de la línea de base (blanco) y se mantiene la cubeta termostatizada y agitada durante 2 ó 3 minutos. Posteriormente, se realiza la corrección del blanco y, manteniendo la agitación y termostatación, se añade el resto de disoluciones. Es conveniente que las disoluciones añadidas posteriormente estén a la misma temperatura que aquélla a la que se desea realizar la reacción. Siempre se añade en última posición la disolución que inicia la reacción. Es importante que durante todo el tiempo de registro, la agitación y la temperatura permanezcan constantes. La agitación evita la formación de precipitados y mantiene homogéneo el sistema, aspecto que hay que tener muy en cuenta cuando se mezclan disoluciones con diferentes propiedades.

2.3.2.2 Sistema de mezcla de flujo interrumpido (*stopped-flow mixing*).

Muchos métodos cinéticos aprovechan la ventaja de reacciones con tiempos de vida medios de menos de 10 segundos y utilizan este sistema de mezcla. Este es el caso presentado en el anexo 3 y que sucede en la mayoría de reacciones de complejación de metales. En un sistema de flujo interrumpido convencional, las soluciones de muestra y reactivo se hacen circular a gran velocidad y presión por la acción combinada de dos jeringas propulsoras, las cuales

2. Métodos cinéticos de análisis

aseguran la mezcla rápida. El flujo resultante es, seguidamente, interrumpido de forma brusca, en la celda de medida con la ayuda de una tercera jeringa, para poder seguir el desarrollo de la reacción (figura 9).

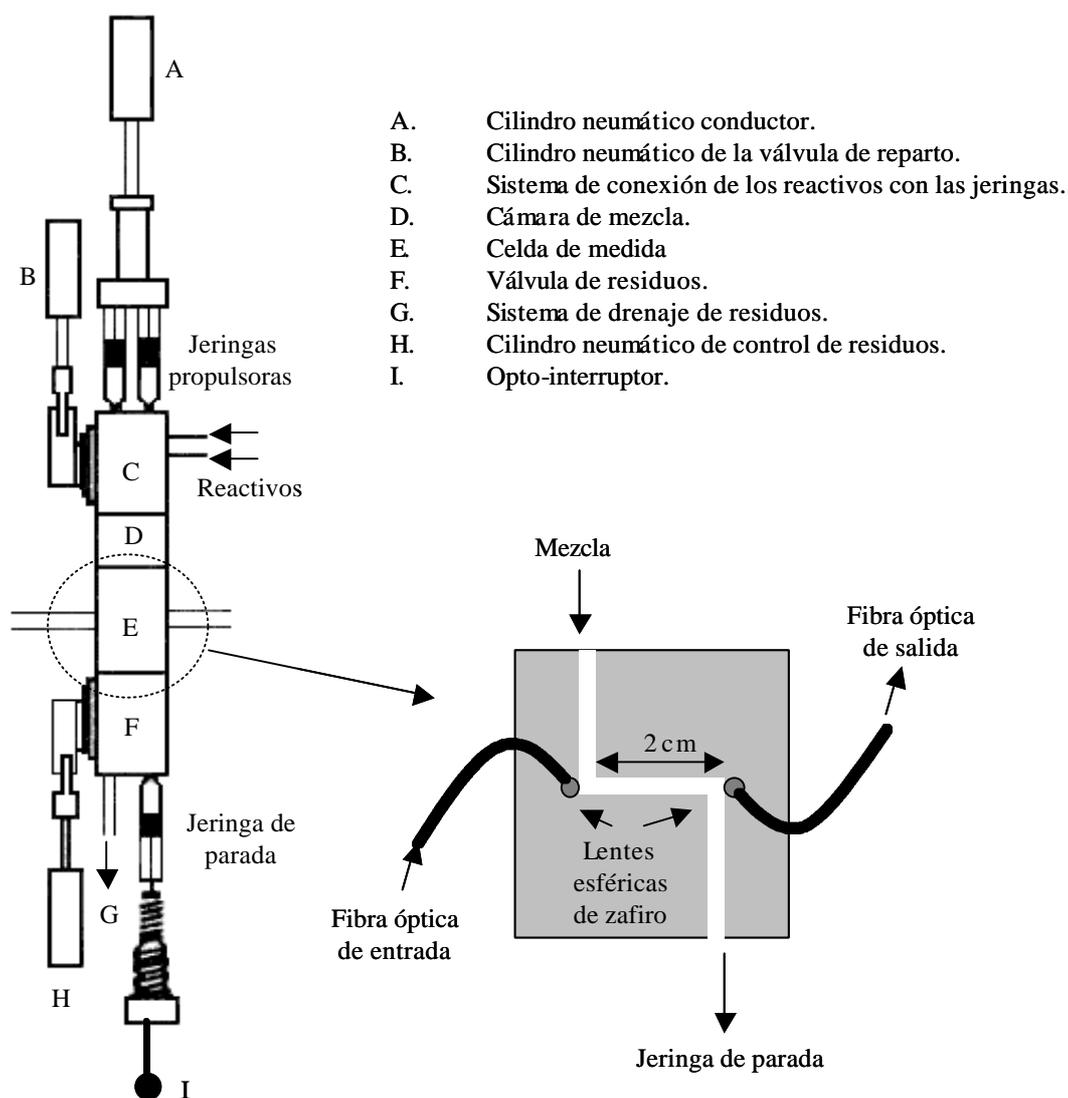


Figura 9. Esquema de una sistema *stopped-flow* automatizado.

En principio, este sistema es inapropiado para la implementación de los métodos cinéticos basados en reacciones lentas, ya que después de que el flujo sea interrumpido y la reacción haya empezado, los productos de reacción se difunden a las jeringas propulsoras, donde se encuentran las disoluciones de partida. En algunas situaciones, los sistemas *stopped-flow*

pueden acelerar algunas reacciones relativamente lentas, ya que la mezcla tiene lugar a presiones más elevadas que en los métodos cinéticos convencionales. Por tanto, la técnica de *stopped-flow* puede ser aplicada a este tipo de reacciones. Existen varias modificaciones sobre la técnica básica, así como diferentes híbridos, pero el fundamento es siempre el mismo [Mottola, 1988; Silva, 1993].

El efecto de la temperatura ha sido estudiado y es un parámetro crítico en estos sistemas de mezcla, como consecuencia de la fricción generada cuando se fuerza a las soluciones, a gran velocidad y presión, a través de las jeringas.

Uno de los problemas que presentan la mayoría de los sistemas de *stopped-flow* es que están limitados a mezclar tan solo los dos reactivos contenidos en las jeringas. La mayoría de las aplicaciones analíticas en cinética requieren la mezcla de varios reactivos antes de la medida.

En esta memoria, se ha utilizado un sistema de *stopped-flow* automatizado [Beckwith, 1972] que ha sido posteriormente rediseñado [Cullen, 1999]. Este sistema utiliza un conjunto de válvulas neumáticas para dirigir los líquidos y para eliminar los residuos, y una jeringa neumática para la mezcla rápida. El sistema de flujo es vertical para minimizar problemas con las burbujas de aire. El ciclo completo de operación puede ser movido por pulsos de un sistema secuenciador digital o mediante interruptores. El ciclo puede repetirse de 1 a 10 veces sin necesidad de rellenar las jeringas, con el propósito de obtener replicados.

El conjunto óptico del sistema consta de una fibra óptica que transporta la luz desde una lámpara de tungsteno hasta la celda de medida (figura 9). La luz pasa a través la celda e interacciona con la muestra; posteriormente, la luz emergente es recogida por una segunda fibra óptica la cual la transporta al detector de diodos en línea (*diode array detector*). Como se observa en la figura, las fibras ópticas están insertadas en el bloque que contiene la celda de medida y terminan en lentes esféricas de zafiro. Estas lentes tienen una función colimadora y de sellado de la celda. El sistema de adquisición de datos para este tipo de celda ha sido caracterizado [Cullen, 1999] y aplicado en análisis cinético diferencial [Cullen, 2000].

Las características principales del sistema son:

- a) 75 ms de tiempo de espera (*delay time*), definido como el tiempo transcurrido entre la mezcla y la adquisición del primer dato.

- b) La velocidad máxima de adquisición de un espectro es función del tiempo de integración deseado. El mínimo tiempo de integración práctico es 13 ms y a este tiempo los espectros pueden ser registrados a 10.5 Hz.
- c) La relación señal/ruido (S/N) es función de la radiación transmitida en la celda, la ganancia del intensificador y del tiempo de integración. El ruido negro es la principal fuente y es minimizado enfriando el detector mediante un *peltier* a aprox. -5.9°C y realizando el blanco. Para minimizar la relación también se intenta que los valores de absorbancia estén situados entre 0.2 y 1.
- d) El rango lineal ha sido determinado de dos maneras diferentes. Utilizando filtros llega hasta una unidad de absorbancia. Con una disolución de dicromato, la linealidad es observada incluso hasta 1.5 unidades de absorbancia.

2.3.3 Seguimiento de la reacción.

Tan solo consideraremos los sistemas que tengan en cuenta los fenómenos relacionados con la luz. Un porcentaje muy elevado de los métodos cinéticos de análisis se basa en las medidas de absorbancia en la región UV-visible. También se utilizan detectores en los procesos que se caracterizan por la emisión de fotones, tales como la luminiscencia (fluorescencia, quimioluminiscencia y bioluminiscencia) y de los cuales también se encuentran aplicaciones.

Los detectores de diodos en línea (*diode array detectors*) en absorbancia y fluorescencia son de uso muy común en los métodos cinéticos. Permiten el seguimiento de la reacción a múltiples longitudes de onda y, por tanto, el potencial más importante de este tipo de detectores es el análisis de multicomponentes.

Las medidas de absorbancia realizadas en esta memoria (anexos 1, 4 y 5) se han llevado a cabo con dos espectrofotómetros de diodos en línea: Hewlett Packard HP8452A y Hewlett Packard HP8453, gobernados por ordenadores a través de una interfase HP-IB. El software del equipo, HP89530 MS-DOS UV/Vis y ChemStation, respectivamente, incorpora módulos para el control del instrumento y para el tratamiento de los espectros.

Son menos utilizados los detectores *CCDs* (*Charge-coupled devices*) y los *CIDs* (*charge-injection devices*) pero éstos son más prometedores para espectroscopia óptica, particularmente cuando puede ser usada su naturaleza bi-dimensional. Por tanto, son

especialmente útiles en espectrómetros de fluorescencia bi-dimensionales, que muestran simultáneamente espectros de emisión a varias longitudes de onda de excitación. Estos detectores son comparables, en términos de sensibilidad y rango lineal, a los detectores de diodos en línea.

En este apartado, vamos a considerar las técnicas quirópticas y, en concreto, la llamada dicroísmo circular (DC). El fundamento de estas técnicas se basa en la interacción entre moléculas con un centro asimétrico y la radiación electromagnética polarizada. Normalmente, las aplicaciones son de elucidación de estructuras moleculares, particularmente de productos naturales para los que es crítico disponer de una técnica capaz de confirmar o determinar la estequiometría absoluta. La creciente utilización de especies quirales en campos muy diversos, ha llevado a la utilización de estas técnicas en química analítica y esta memoria va un poco más allá, utilizando el dicroísmo circular en un método cinético para la resolución de enantiómeros (anexo 2). Por este motivo, se realizará una breve descripción de esta técnica.

2.3.3.1 Absorción de luz y dicroísmo circular (DC).

Una propiedad física que distingue entre sustancias quirales (p. ej. enantiómeros) y no quirales, es la capacidad de las primeras de rotar el plano de la luz polarizada plana. Un rayo de luz consta de dos planos oscilantes perpendiculares, correspondientes a los campos eléctrico y magnético. Estos planos, a su vez, son perpendiculares a la dirección de propagación del rayo de luz. En un rayo de luz ordinaria, se observa cómo las oscilaciones de los campos eléctrico y magnético ocurren en todos los planos posibles, mientras que la luz polarizada plana oscila en un plano. Por tanto, la luz polarizada plana es la suma de dos componentes polarizados circularmente, uno a la izquierda y otro a la derecha, de la misma amplitud (figura 10).

El fenómeno de actividad óptica viene determinado por la diferencia en los índices de refracción para la luz polarizada circularmente a la derecha y la luz polarizada circularmente a la izquierda del medio considerado. Por tanto, ambos componentes de la luz se desfasan, es decir, viajan en el medio a diferente velocidad, dando lugar a un vector resultante igual al de incidencia, pero con el plano de polarización rotado α grados. Cada uno de los enantiómeros hace girar la luz polarizada plana en igual magnitud pero en direcciones opuestas

(dextrorrotatorio y levorrotatorio).

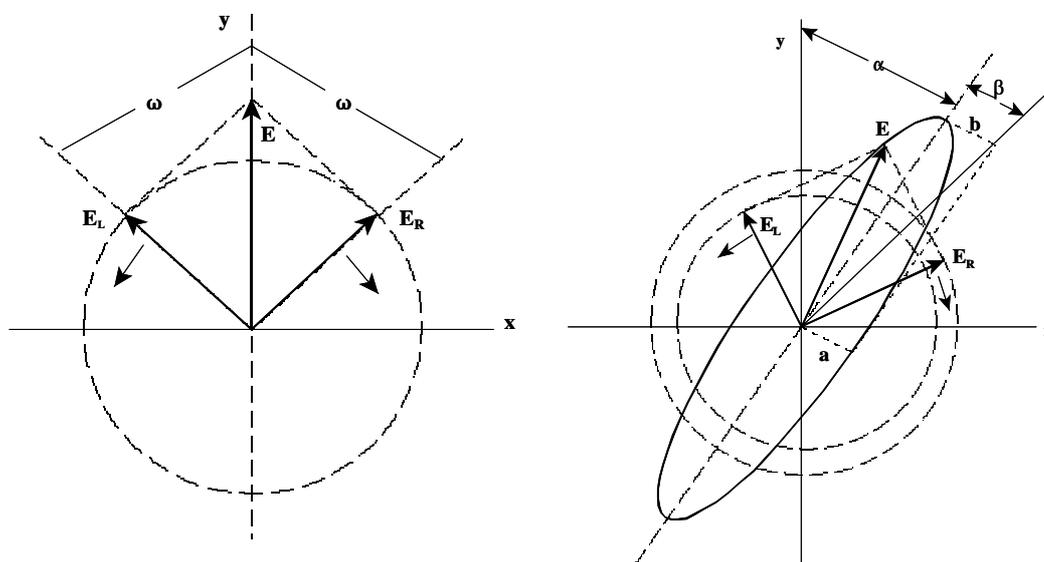


Figura 10. Representación de los vectores de los campos eléctricos antes y después de pasar por una substancia ópticamente activa y absorbente.

Además de la rotación del plano de polarización, se puede producir una absorción de los componentes circulares en diferente grado ($\epsilon_L \neq \epsilon_R$), de forma que el rayo incidente de luz polarizada plana emerge como un rayo polarizado elípticamente, cuyo eje mayor está rotado α grados. Por convenio, la diferencia $\Delta\epsilon = \epsilon_L - \epsilon_R$ es llamada dicroísmo circular. La magnitud del dicroísmo circular se mide mediante la relación entre el eje mayor y el menor de la elipse (Ψ , elipticidad) o bien mediante la diferencia de absorbancia medida para ambos rayos (ΔA). La elipticidad (en radianes) se puede expresar como:

$$\Psi = \frac{2.303Cb(\epsilon_L - \epsilon_R)}{4} \quad (72)$$

siendo b el camino óptico, en centímetros, y C la concentración de la especie quiral absorbente, en mol/l. La misma expresión en grados queda:

$$\theta = 32.98(\epsilon_L - \epsilon_R)Cb \quad (73)$$

relación con semejanza a la Ley de Beer-Lambert. La elipticidad molar $[\theta]$ es una magnitud intrínseca que da una medida del poder dicroico del compuesto y se calcula según:

$$[\theta] = \frac{\theta M}{b \cdot C' \cdot 100} \quad (74)$$

donde M es el peso molecular de la sustancia y C' la concentración del soluto en g/l. El hecho que la relación sea entre Ψ y ΔA (o $[\theta]$ y $\Delta \epsilon$) y no entre Ψ y A (o $[\theta]$ y ϵ) implica que las señales de DC pueden ser intensas, incluso cuando las absorbancias son débiles, si ΔA es alta.

Un enantiómero (con grupo cromóforo) podrá ser analizado y cuantificado por dicroísmo circular. Pero, cuando se presente un sistema con mezcla de enantiómeros, el dicroísmo circular tan solo permitirá observar proporciones relativas y hará falta una técnica complementaria, como la UV-visible o fluorescencia, para poder resolver el sistema. Una alternativa a la utilización de una segunda técnica sería la utilización de reactivos quirales, capaces de formar diastereoisómeros, los cuales pueden presentar espectros de dicroísmo circular diferente y, si además, se puede hacer el registro de la reacción, la información cinética y espectral obtenida puede ser aprovechada para la cuantificación de mezclas.

Instrumentación básica de dicroísmo circular.

El dicroísmo circular puede ser medido por dos métodos que son realmente equivalentes. En el primer caso, se mide separadamente la absorción de la muestra para cada uno de los dos componentes, polarizados circularmente, y se determina la diferencia entre ambos. En el segundo caso, un rayo de luz polarizada linealmente incide en la muestra ópticamente activa y la diferente absorción de las dos componentes circulares da lugar a un rayo de luz emergente, polarizado elípticamente, cuya elipticidad está directamente relacionada con $\Delta \epsilon$. Un esquema básico de un instrumento se muestra en la figura 11. La luz, en el intervalo de longitudes de onda seleccionado, se obtiene a partir de una fuente de luz policromática y una combinación de monocromadores. La luz monocromática es entonces polarizada circularmente mediante un polarizador circular, que no es más que un polarizador lineal combinado con un modulador fotoelástico o una célula de Pockels. La detección de la luz se realiza habitualmente con un tubo fotomultiplicador, puesto que la absorción de fondo hace que sólo una porción de la luz incidente abandone la muestra.

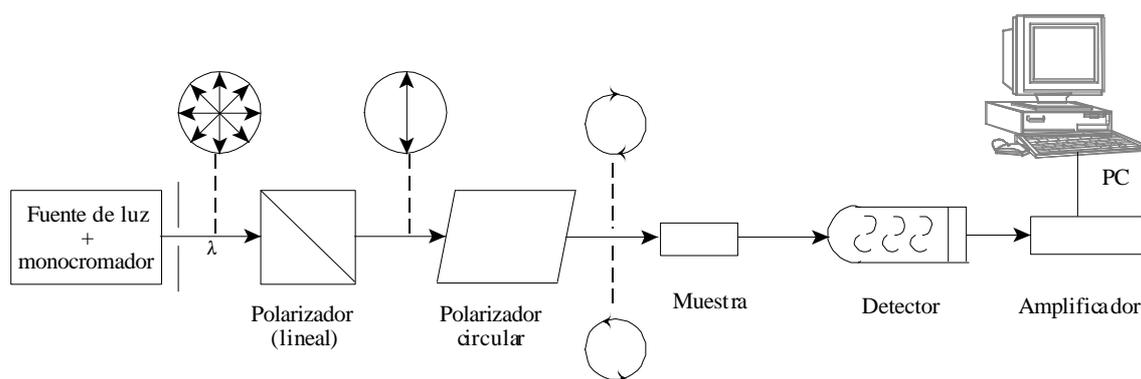


Figura 11. Esquema de los elementos básicos de un espectrómetro de DC.

El espectropolarímetro utilizado en uno de los trabajos (anexo 2) ha sido un JASCO (Japan Spectroscopy Company, Tokio, Japón) modelo J-715, equipado con una lámpara de xenón de 150 W, y purga de nitrógeno. Se ha utilizado para las medidas una cubeta rectangular de cuarzo de 1 cm de camino óptico. La temperatura durante todo el registro en el compartimento de medida ha sido fijada a 25.0 ± 0.2 °C mediante un PTC-351S 6-Position Peltier Effect Cell Changer. La señal de DC es comunicada a través de una interfase RS-232C a un PC que permite el tratamiento de datos.

Al ser un espectropolarímetro de barrido, son muchos los factores a tener en cuenta en el desarrollo del método cinético. Entre ellos, hay que considerar la velocidad de barrido, la anchura de banda, la constante de tiempo (o respuesta), la resolución y el intervalo de longitudes de onda y tiempo. Todos estos parámetros tendrán que ser optimizados y adecuados al tipo de reacción estudiada. Evidentemente, las reacciones muy rápidas no podrán ser seguidas cuando el registro se quiere efectuar a múltiples longitudes de onda.

2.4 CLASIFICACIÓN DE LOS MÉTODOS CINÉTICOS.

2.4.1 Introducción.

Se han desarrollado muchos métodos cinéticos a lo largo de los años. Pardue [Pardue, 1989] describe una amplia clasificación para ellos, basada en la medida y las opciones disponibles para el procesado de datos. En la clasificación propuesta por Pardue, los métodos cinéticos son divididos en dos grandes bloques: métodos de componente único y métodos de multicomponentes. Tanto unos como otros pueden ser o no de compensación de errores (p. ej.

cambios de temperatura, pH, fuerza iónica etc.).

Otros esquemas están centrados en el tipo de reacción utilizada: enzimática, oxidación-reducción y complejación; o en el tipo de determinación: catalítica (enzimática y no enzimática) y no catalítica (determinación de un único componente o de multicomponentes).

La utilización de una clasificación u otra es indiferente, siempre y cuando se tenga en consideración el tipo de reacción utilizada y la finalidad del método, es decir, si se pretende cuantificar uno o más analitos. En esta memoria, se presentan varios trabajos destinados siempre a la cuantificación de varias especies, pero siempre utilizando diferentes tipos de reacciones. A pesar de ello, es importante considerar diferentes aspectos, como es el caso de los métodos cinéticos diferenciales, para clarificar el contenido de esta memoria.

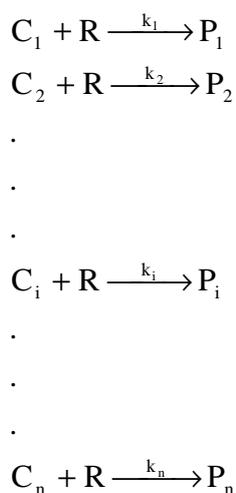
2.4.2 Métodos cinéticos diferenciales clásicos.

Los métodos cinéticos diferenciales se basan en la resolución de mezclas de especies afines según su diferente velocidad de reacción con un reactivo común, normalmente, y a la diferencia en la señal analítica obtenida en el transcurso de la reacción. Comparados con los métodos de determinación directa; por ejemplo, de un catalizador que posibilita la cuantificación a nivel trazas e incluso inferior, los métodos cinéticos diferenciales, no son, en general tan sensibles, pero permiten el análisis simultáneo de dos o más especies sin separación previa.

Por lo general, las reacciones que sirven de base a estos métodos son no catalizadas, aunque se han aplicado también a la determinación de mezclas de especies que catalizan diferencialmente una misma reacción (anexo 1).

Las reacciones consideradas pueden ser de naturaleza muy diversa y son aplicadas tanto a análisis orgánico como inorgánico.

Si consideramos varios analitos C_i que reaccionan con un reactivo común R para dar el mismo o diferentes productos de reacción P_i :



k_i es la constante de velocidad para la especie reaccionante C_i . La ecuación de velocidad para el componente i puede ser escrita como

$$\frac{d[P_i]}{dt} = -\frac{d[C_i]}{dt} = k_i [C_i][R] \quad (75)$$

Mark y Rechnitz [Mark, 1968] propusieron una clasificación de los métodos cinéticos diferenciales de análisis basada en el orden de reacción. Sin embargo, esta clasificación no tiene en cuenta los avances en quimiometría y el procesado de gran cantidad de información mediante calibración multivariable, por tanto, se les puede considerar como métodos cinéticos clásicos.

Tabla 1. Clasificación de los métodos cinéticos diferenciales clásicos en función del orden de reacción.

Orden de reacción	Concentración inicial de las especies	Método (numérico o gráfico)
Pseudo-primer orden respecto los analitos	$[R]_o \gg \Sigma[C_i]$	Extrapolación logarítmica
		Punto único
		Tangentes
		Ecuaciones proporcionales
Segundo orden	$[R]_o > \Sigma[C_i]$	Extrapolación logarítmica
	$[R]_o = \Sigma[C_i]$	Punto único
	$[R]_o < \Sigma[C_i]$	Extrapolación lineal
Pseudo-primer orden respecto el reactivo	$[R]_o \ll \Sigma[C_i]$	Punto único (Roberts y Regan)
		Ecuaciones proporcionales

En la tabla 1 se muestra un esquema de estos métodos teniendo en cuenta el sistema de multicomponentes.

Se puede observar que, dependiendo del orden reacción, se pueden utilizar unos métodos u otros. Los más utilizados son los métodos gráficos de extrapolación logarítmica y lineal y, dentro de los métodos numéricos, el de las ecuaciones proporcionales. En la bibliografía [Mottola, 1988] se describen cada uno de estos métodos con detalle y se citan las diferentes aplicaciones.

Estos métodos asumen que las velocidades de los analitos son independientes unas de otras, es decir, la presencia de un analito no modifica el comportamiento de los otros. Cuando esto no es cierto, se dice que estamos en presencia de un efecto sinérgico. Este efecto no es muy común, pero puede ser una fuente de error muy importante. En estos casos sólo los métodos que no necesitan el conocimiento previo de las constantes de velocidad, como es la extrapolación logarítmica de segundo orden, o bien otros métodos específicos, pueden ser aplicados.

La aplicabilidad de estos métodos está condicionada también por los valores relativos de las constantes de velocidad, así como por la relación de las concentraciones de los componentes de la mezcla y la complejidad de la reacción.

Un análisis de los efectos de los errores en los parámetros medidos, de la relación de las constantes de velocidad, y de la composición de la mezcla en los métodos clásicos más utilizados, se da en el libro de Mark y Rehnitz [Mark, 1968].

2.4.3 Métodos cinéticos diferenciales modernos.

El uso de los ordenadores juega un papel cada vez más importante, tanto en la automatización de la recogida de datos cinéticos como en el tratamiento de éstos para el análisis cinético diferencial. Los métodos numéricos y gráficos expuestos anteriormente, hacen uso por lo general, de un número mínimo de datos puntuales. Sin embargo, el análisis de mezclas de multicomponentes (y por supuesto el análisis de un solo compuesto) puede mejorar si se considera la respuesta total obtenida en función del tiempo. Los ordenadores llevan a cabo otro trabajo en los métodos cinéticos: la simulación de reacciones químicas. Este aspecto es muy importante para estudiar reacciones muy complejas y para observar cómo los métodos

modernos de cálculo (calibración multivariable) se ven afectados por los diferentes parámetros a tener en cuenta en un método cinético (ruido instrumental, fluctuaciones en las constantes de velocidad, etc.).

2.5 TENDENCIA DE LOS MÉTODOS CINÉTICOS.

Crouch [Crouch, 1990] ya menciona que los métodos cinéticos se prestan a la automatización inteligente. Los métodos cinéticos precisan de un control riguroso del tiempo y de las condiciones experimentales, tales como la temperatura, pH y concentración de los reactivos. Precisamente este aspecto es uno de los puntos negativos por los cuales muchos químicos analíticos son reticentes a utilizar los métodos cinéticos.

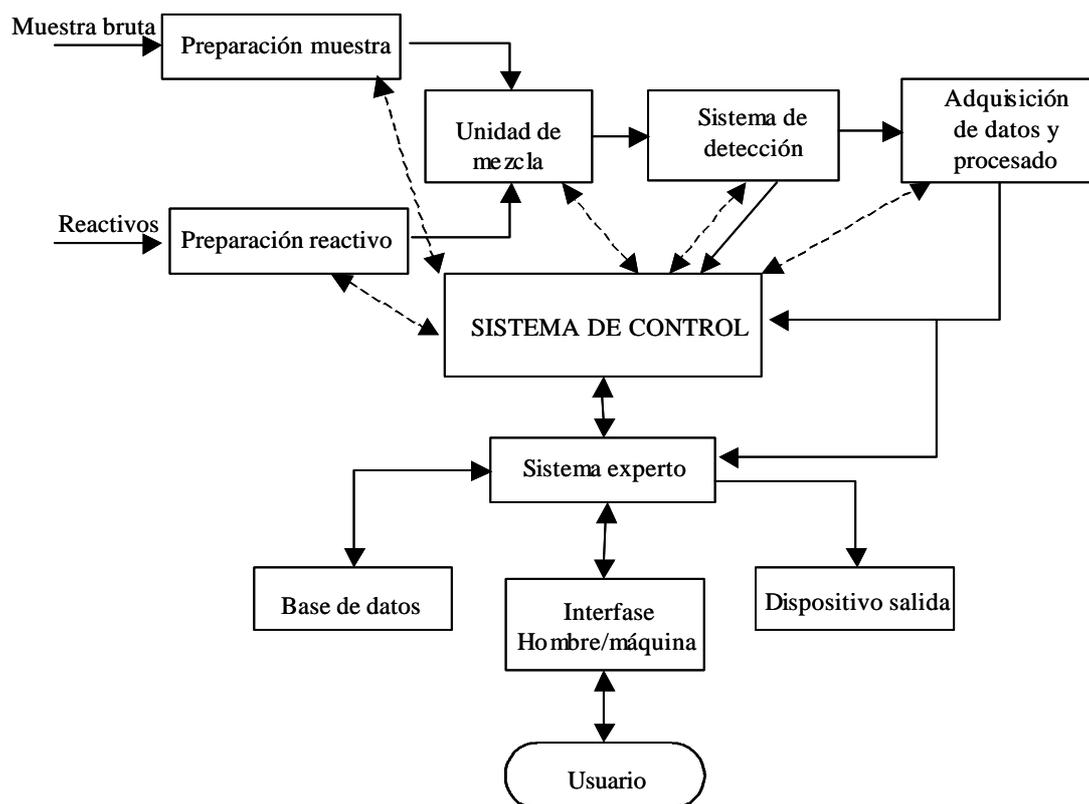


Figura 12. Analizador cinético inteligente.

Los sistemas automáticos son ideales para proporcionar estas características.

Muchos de los avances en los métodos cinéticos aprovechan la automatización y los

tratamientos de datos quimiométricos para proporcionar analizadores cinéticos, reducción y almacenamiento de grandes conjuntos multidimensionales de datos, determinación de multicomponentes y compensación de errores debidos a las variaciones en las condiciones. Basado en estos conceptos y los avances más recientes, Crouch propone y describe las características más importantes que debería contemplar el analizador cinético inteligente (figura 12). En él, el usuario interactúa con un sistema experto vía una interfase hombre-máquina. Además de su propio conocimiento, el sistema experto tiene acceso a una base de datos y se comunica con el sistema de control el cual es otro programa u ordenador que controla el instrumento. En este analizador, tanto la manipulación de la muestra como la de reactivos y su mezcla está bajo completo control. Existen muchos bucles *feedback* y la intervención humana puede estar presente en cualquiera de las etapas. El analizador cinético automático tiene que ser también un sistema adaptable proporcionando una adaptación a tiempo real tanto del instrumento como del experimento en estudio. El sistema tendría que ser capaz de detectar errores y ser capaz de corregirlos (p. ej. problemas instrumentales, de temperatura, contaminación de reactivos y su estabilidad, concentraciones fuera de rango, etc.).

Silva [Silva, 1993], posteriormente, describe y evalúa los avances de la automatización en los métodos cinéticos de análisis y lo hace según las tres etapas mencionadas en el apartado 2.3.1. Los ordenadores dominan el tercer nivel de automatización en el laboratorio analítico, haciendo posible la comunicación con muchas unidades de los otros dos niveles.

Al mismo tiempo, Crouch [Crouch, 1993] identifica y describe las tendencias en los métodos cinéticos de análisis. Éstas incluyen de nuevo:

- El uso de la automatización
- La instrumentación multidimensional (espectrofotómetros de *diode-array*, detectores *CCD* y sistemas de detectores múltiples).
- También considera como tendencias el continuo desarrollo de sofisticados métodos numéricos de procesamiento de datos y de compensación de errores.
- Las mejoras en la determinación de uni- y multicomponentes.
- La expansión del enfoque cinético a sistemas dinámicos que no son función de una reacción química.

2. Métodos cinéticos de análisis

Estas tendencias, mencionadas hace casi una década, siguen siendo las que destacan hoy en día y en un futuro próximo. La consecución de éstas hace que los métodos cinéticos sean más competitivos con los basados en el equilibrio y sean cada vez más tenidos en cuenta por los químicos analíticos.