

2

DESCRIPCIÓN MATEMÁTICA DEL PROBLEMA INVERSO.

CAPÍTULO 2

DESCRIPCIÓN MATEMÁTICA DEL PROBLEMA INVERSO.

2.1 CRITERIOS DE IDENTIFICACIÓN.

La formulación del problema inverso en los términos planteados anteriormente, requiere, en primer lugar, la adopción de un criterio de identificación, que permitirá definir una función a maximizar o minimizar, que denominaremos función objetivo.

Escogido un modelo cualquiera que relaciona unos parámetros p con unas variables medibles x : $x = M(p)$, se supone que se dispone de "m" medidas de dichas variables $x_1^*, x_2^*, x_3^*, \dots, x_m^*$, y se desea encontrar los parámetros del modelo, $p_1, p_2, p_3, \dots, p_n$, que mejor se ajustan a dichas medidas. Existen muchas formas de interpretar este "ajuste", y su definición en términos objetivos se concreta al adoptar un criterio de identificación, de manera que la elección de un criterio define el problema matemático que debe resolverse. Una vez escogido un criterio de identificación, hay diferentes maneras de minimizarlo, si bien el resultado final es independiente del método.

De todas las posibilidades, se ha escogido en este trabajo una visión fundamentalmente estadística del problema inverso, debido a su mayor generalidad y a que permite representar la propia aleatoriedad que acompaña a todo proceso de medida.

El primer paso para poder abordar matemáticamente el problema inverso es, como ya se ha indicado, el establecer un "criterio de identificación". De los posibles criterios utilizables, hay cinco que son sin duda los más empleados (Eykhoff, 1974), y que se describen a continuación. En todos ellos se considera que el modelo que caracteriza el sistema en estudio es determinista.

2.1.1 Minimización De Distancias.

El criterio más simple consiste en minimizar la "distancia" entre las variables medidas y las predichas con el modelo (error de predicción). No sólo puede usarse la distancia Euclídea como medida del error, sino que puede generalizarse el concepto al de norma de orden "j" :

$$\begin{aligned}
 \text{- Norma 1} \quad & \| \mathbf{x}^* - \mathbf{x} \|_1 = \left[\sum_i |x_i^* - x_i| \right] & x_i = M_i(p) \\
 \text{- Norma 2} \quad & \| \mathbf{x}^* - \mathbf{x} \|_2 = \left[\sum_i |x_i^* - x_i|^2 \right]^{1/2} \\
 \dots & \\
 \text{- Norma j} \quad & \| \mathbf{x}^* - \mathbf{x} \|_j = \left[\sum_i |x_i^* - x_i|^j \right]^{1/j}
 \end{aligned}$$

A medida que aumenta el orden de la norma, se hace más importante el término de mayor error. En el límite:

$$\| \mathbf{x}^* - \mathbf{x} \|_\infty = \max_i |x_i^* - x_i|$$

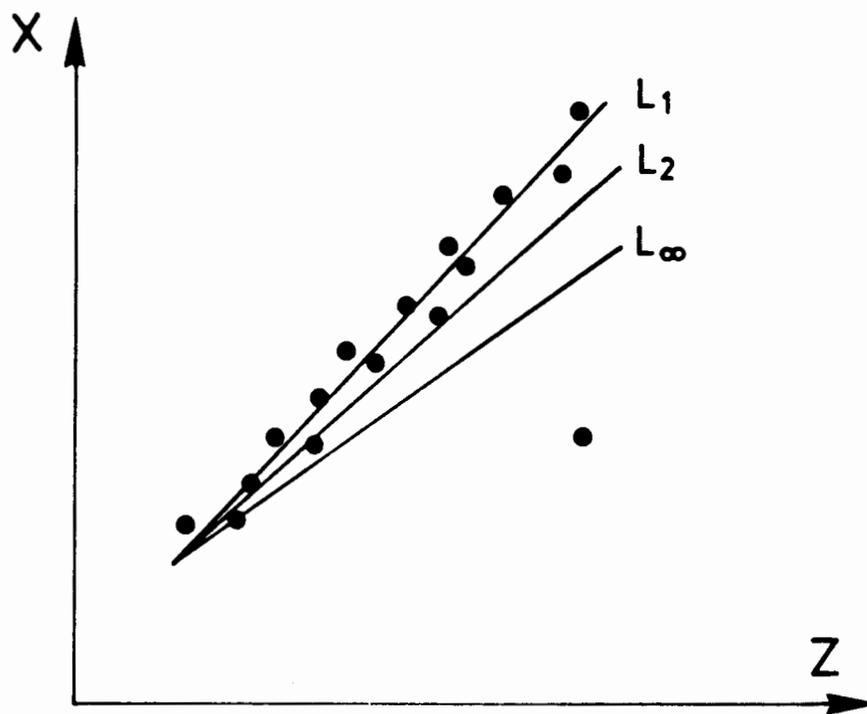


Figura 2.1. Ejemplo ilustrativo de un modelo, representado por una recta, ajustado a las observaciones utilizando diferentes criterios de minimización de distancias. Estos criterios de identificación corresponden matemáticamente a la minimización de las normas euclídeas de orden 1,2,...etc.

En la figura 2.1 se representa esquemáticamente una recta tomada como modelo que minimiza las normas de diferente orden, y puede comprobarse que si se utiliza la minimización de una norma de orden muy alto como criterio de identificación, se está dando mucha importancia a los términos de mayor error.

Hay varios motivos, además del anterior, que han convertido la minimización de orden uno o dos en criterios tradicionales de identificación. Algunos de estos motivos son de origen estadístico y los veremos posteriormente.

En concreto, la minimización de la norma de orden 2, que da lugar al conocido método de mínimos cuadrados, es el criterio más tradicional y frecuente en la resolución de problemas inversos. La función objetivo a minimizar es:

$$J = \sum_{i=1}^m \left[x_i^* - M_i(p) \right]^2$$

o en notación matricial:

$$J = \left[\mathbf{x}^* - \mathbf{M}(p^*) \right]^T \left[\mathbf{x}^* - \mathbf{M}(p^*) \right] \quad [2.1]$$

donde \mathbf{x}^* es el vector de variables medidas, y p^* es el vector de parámetros a determinar, que minimizan [2.1].

2.1.2 Markov.

Es una generalización natural del criterio anterior. La función objetivo es:

$$J = \left[\mathbf{x}^* - \mathbf{M}(p^*) \right]^T \mathbf{W} \left[\mathbf{x}^* - \mathbf{M}(p^*) \right] \quad [2.2]$$

donde \mathbf{W} es una matriz diagonal, que puede interpretarse como una

matriz de pesos que se asigna a priori en función de la calidad de cada una de las observaciones. El error cometido en el proceso de medida se refleja en esta matriz, dando más peso a las medidas más fiables. La utilización de este criterio requiere disponer de información inicial sobre dicho proceso de medida, para poder definir correctamente la matriz de pesos.

2.1.3 Máxima Verosimilitud.

La diferencia entre medida y predicción se atribuye a un error (cuyo sentido es específicamente aleatorio), y se le asigna una descripción probabilista. Debido a esta característica aleatoria, es conveniente utilizar un criterio estadístico para la definición de la función objetivo. La adopción de un criterio de este tipo requiere un mayor conocimiento del problema: considerando los errores en las observaciones como variables aleatorias, se necesita conocer su función de probabilidad, y sus posibles correlaciones. Como contrapartida, se puede definir un criterio mucho más general.

En este caso la solución al problema de identificación se supone que es aquella que maximiza la probabilidad de realizar unas observaciones, supuestas unas hipótesis sobre el modelo y la estructura de error de las medidas (representada por unas correlaciones conocidas). Matemáticamente, este criterio se puede expresar como:

$$\max_{\mathbf{p}} f(\mathbf{x}^*/\mathbf{p})$$

donde $f(\mathbf{x}^*/\mathbf{p})$ es la función densidad de probabilidad de las medidas, dados los parámetros.

La verosimilitud (Edwards, 1972) se define como proporcional a esta probabilidad, y los parámetros estimados que la maximizan son la solución al problema planteado. En concreto, si se supone que los errores de las medidas se distribuyen según una función normal (gausiana), con una matriz de covarianzas C_x , la verosimilitud para un conjunto de parámetros viene dada por:

$$L(\mathbf{x}) = \frac{K |C_x|^{-1/2}}{(2\pi)^{m/2}} \exp \left[-\frac{1}{2} (\Delta \mathbf{x})^T C_x^{-1} (\Delta \mathbf{x}) \right] \quad [2.3]$$

donde $\Delta \mathbf{x}$ es la diferencia entre variables medidas y calculadas con el modelo, y K es una constante de proporcionalidad arbitraria. La expresión [2.3] es proporcional a la probabilidad de la realización simultánea de todos los "errores" de las medidas.

Por comodidad de cálculo, se suele trabajar con el logaritmo natural de la función de verosimilitud [2.3], denominado función de soporte.

Maximizar [2.3] es equivalente a minimizar

$$S = -2 \ln L$$

gracias al crecimiento monotónico de la función logaritmo. Resulta más cómodo operar con la función S , especialmente si se trabaja con funciones de distribución normal. Su expresión desarrollada será:

$$S = (\Delta \mathbf{x})^T C_x^{-1} (\Delta \mathbf{x}) + m \ln 2\pi + \ln |C_x| - 2 \ln K$$

Obsérvese que una vez fijada C_x los tres últimos términos son constantes, por lo que a efectos de minimización, la función objetivo definitiva es de la forma:

$$J = (\Delta \mathbf{x})^T \mathbf{C}_x^{-1} (\Delta \mathbf{x}) \quad [2.4]$$

que tiene una estructura similar a la de los criterios anteriores. Hay que distinguir, sin embargo, que en este caso la matriz de pesos tiene un claro significado estadístico: es la inversa de la matriz de covarianzas de los errores de las medidas. Si dicha matriz es diagonal, aquel resulta ser el criterio de Markov. En el caso particular de que además sea la matriz identidad, se reproduce el criterio de mínimos cuadrados. El criterio de mínimos cuadrados puede contemplarse ahora desde un punto de vista estadístico, en el sentido de que es el criterio de máxima verosimilitud cuando se supone que los errores de las observaciones se representan mediante una función de distribución normal, son independientes entre sí, y tienen la misma varianza.

Se ha visto que el criterio de identificación de máxima verosimilitud es capaz de reproducir los criterios anteriores como casos particulares, y esta circunstancia le otorga un grado importante de generalidad.

2.1.4 Máxima Verosimilitud Con Información Previa.

Aumentando en el nivel de generalización, puede considerarse que para el proceso de identificación se dispone de cierta información previa de los parámetros, que puede expresarse en forma de una estimación "a priori" de los mismos. La función objetivo deberá incluir esta información, además de las medidas ya realizadas. En general, siguiendo con la representación estadística del problema, habrá que introducir la función densidad

de probabilidad de los errores en una estimación previa de los parámetros. Es decir, deberá considerarse un término que incluya los errores en los parámetros.

La expresión [2.3] puede generalizarse con este concepto, definiendo la verosimilitud como proporcional a la probabilidad conjunta de errores de medidas y de parámetros, de forma que si estos errores son independientes, resulta:

$$L(\mathbf{p}) = KP(\mathbf{x}, \mathbf{p}) = KP(\mathbf{x})P(\mathbf{p}) \quad [2.5]$$

Estas funciones de probabilidad, si se consideran normales o gaussianas, tendrán la expresión:

$$P(\mathbf{x}) = \frac{|\mathbf{C}_x|^{-1/2}}{(2\pi)^{m/2}} \exp \left[-\frac{1}{2} (\Delta \mathbf{x})^T \mathbf{C}_x^{-1} (\Delta \mathbf{x}) \right] \quad [2.6]$$

para los errores en las medidas, y para los errores en los parámetros:

$$P(\mathbf{p}) = \frac{|\mathbf{C}_p^0|^{-1/2}}{(2\pi)^{p/2}} \exp \left[-\frac{1}{2} (\Delta \mathbf{p}^0)^T \mathbf{C}_p^{0-1} (\Delta \mathbf{p}^0) \right] \quad [2.7]$$

donde "m" y "p" son el número de medidas y de parámetros respectivamente. \mathbf{C}_p^0 es la matriz de covarianzas de los parámetros estimados "a priori", y $\Delta \mathbf{p}^0$ es el vector de diferencias entre los parámetros que se identifican y los iniciales definidos en base a información previa del problema. De la misma forma que en el apartado anterior, se define:

$$S = -2 \ln L(\mathbf{p}) \quad [2.8]$$

y maximizar la verosimilitud equivale a minimizar [2.8]

Desarrollando esta expresión utilizando [2.6] y [2.7], resulta:

$$S = (\Delta \mathbf{x})^T \mathbf{C}_x^{-1} (\Delta \mathbf{x}) + (\Delta \mathbf{p}^o)^T \mathbf{C}_p^{0-1} (\Delta \mathbf{p}^o) + \ln |\mathbf{C}_x| + \ln |\mathbf{C}_p^o| + \\ + m \ln(2\pi) + p \ln(2\pi) - 2 \ln(K)$$

[2.9]

Para un problema dado, los tres últimos términos de esta expresión son constantes, siendo la función objetivo a minimizar:

$$S = (\Delta \mathbf{x})^T \mathbf{C}_x^{-1} (\Delta \mathbf{x}) + (\Delta \mathbf{p}^o)^T \mathbf{C}_p^{0-1} (\Delta \mathbf{p}^o) + \ln |\mathbf{C}_x| + \ln |\mathbf{C}_p^o|$$

[2.10]

La función objetivo depende por tanto de las matrices de covarianza de errores de las medidas de \mathbf{x} y de \mathbf{p} , del vector de diferencias entre variables medidas y calculadas y del vector de diferencias entre los parámetros y su estimación previa.

La incorporación de la información previa al criterio de máxima verosimilitud permite generalizarlo mediante un enfoque "bayesiano" del problema, manteniendo el modelo como determinista.

2.1.5 Mínimo Riesgo.

También denominado criterio de mínimo coste, se basa en definir una función de riesgo o de coste que penalice el hecho de que se identifiquen unos parámetros incorrectos.

La función de pérdida o daño es una función real no negativa que refleja la pérdida de tomar como parámetros \mathbf{p}^* en vez de \mathbf{p} . La función de riesgo se define como la esperanza matemática de la función de daño. Con estos supuestos, el estimador de mínimo

riesgo viene definido por la condición (Bury, 1975):

$$\min_{p^*} \int_p C(p^*, p) f(p/x^*) dp \quad \rightarrow \quad \frac{\partial}{\partial p^*} \int_p C(p^*, p) f(p/x^*) dp \Big|_{p=p^*} = 0$$

[2.11]

donde $C(p^*, p)$ es la función de daño, que tiene un mínimo para $p^* = p$, y $f(p/x^*)$ es la función densidad de probabilidad de los parámetros, dadas las medidas.

Lógicamente este es el criterio que puede plantearse en los términos más generales, pero también es el que requiere mayor información sobre el problema: Hay que definir una función de riesgo que exige conocer la función densidad de probabilidad de las observaciones y de los parámetros, además de tener muy bien definido para qué se va a usar el modelo y, sobre todo, los "costes" de los errores.

2.1.6 Elección De Criterio.

Los cinco criterios indicados anteriormente están en orden creciente de información necesitada para plantear el problema. Es decir, el criterio de mínimo riesgo es el más general, pero es el que requiere mayor información inicial. De manera que la elección de un criterio u otro para resolver el problema inverso, vendrá condicionada, entre otros factores, por la información de que se disponga.

Como consecuencia de ello, no se puede hablar de un criterio de identificación idóneo y general para todos los casos, sino que para cada tipo de problemas puede ser más adecuado uno u otro. En el tipo de problemas que aquí se estudian parece razonable adoptar la hipótesis de funciones de distribución normales, tanto para los errores en las variables de medida como en los parámetros. La definición de una función de penalización o de coste sólo parece indicada si se escoge adecuadamente dicha función para cada caso en concreto.

Por este motivo, a efectos de desarrollar una metodología de estimación de parámetros suficientemente general, el criterio de identificación escogido ha sido el de maximizar la función de verosimilitud. Por otra parte, tanto el criterio de Markov, como el de mínimos cuadrados, son un caso particular de aquél. En algunas deducciones se ha considerado la existencia de información previa de los parámetros, utilizando así una formulación más general.

La definición de las matrices de covarianza que se usan en estos criterios es uno de los problemas que presenta su aplicación, y a ello se dedica buena parte del capítulo cuarto. Definidas éstas, el planteamiento del problema es directo y queda definido por [2.4], con las hipótesis de normalidad para los errores de las medidas.

La mayoría de trabajos realizados hasta la fecha sobre el problema inverso en Ingeniería Geotécnica han utilizado, por simplicidad, los criterios de Markov y de mínimos cuadrados (Maier et al 1981; Gioda et al, 1980, 1981, 1985). El enfoque bayesiano

del criterio de máxima verosimilitud (caso de información previa) ha sido menos utilizado en ingeniería civil e identificación de parámetros (Cividini et al, 1983). Por otra parte, el método de máxima verosimilitud ha sido estudiado extensamente en la literatura matemática desde un punto de vista exclusivamente estadístico (Edwards, 1972; Bury, 1975; Graupe, 1972), y también se ha usado con éxito en algunos casos para la estimación de parámetros (Carrera, 1984). En lo que sigue se desarrollará la metodología de identificación utilizando este criterio, se generalizará en algunos casos incluyendo información previa y se particularizará para los otros criterios en algún caso más simple.

Escoger un criterio de identificación supone, como ya se ha visto, definir una función objetivo cuyo mínimo proporcionará la solución al problema inverso. Desde una óptica matemática, el problema inverso es ahora un problema de extremos.

2.2 ESTIMADORES

Antes de pasar revista a los diferentes métodos disponibles para calcular mínimos de funciones, es conveniente destacar una visión exclusivamente estadística del problema inverso.

Entre los siglos XVIII y XIX, muchos de los trabajos de investigación de los matemáticos hacían referencia también a la identificación de parámetros de los propios modelos estadísticos. Se deseaba establecer la metodología de cálculo de los parámetros de las funciones de probabilidad propuestas.

Supóngase una variable aleatoria que se considera definida por un modelo normal. Este modelo queda determinado por dos parámetros: la media y la desviación típica. Si se dispone de una serie de observaciones de esa variable, hay varias formas de obtener estos parámetros. Una de ellas puede ser, por ejemplo, utilizar el criterio de mínimos cuadrados.

La estadística ha definido para este tipo de problemas varios métodos de cálculo, todos ellos encaminados a la obtención de unos parámetros de modelos a partir de observaciones de variables aleatorias. El problema es parecido al problema de identificación planteado en el primer capítulo, si bien es fundamental distinguir entre los modelos estadísticos y los modelos que en cada caso se usan en ciencias aplicadas. A pesar de esta diferencia de aspecto, el fundamento matemático es similar. En estadística, una fórmula de estimación de parámetros de un modelo a partir de observaciones de las variables aleatorias recibe el nombre de "estimador". De esta forma muchos estimadores tienen su correspondencia con los criterios de identificación presentados en el apartado anterior. Así, puede hablarse del estimador de mínimos cuadrados, del de máxima verosimilitud, etc.

Un desarrollo completo de la teoría de la estimación desde un punto de vista estadístico puede verse en Bury (1975) y en Mann et al (1974). Como algunas propiedades estadísticas de estos estimadores pueden ser útiles para los problemas de identificación, se realiza un breve repaso de éstas a continuación.

Denominando p a los parámetros ciertos y p^* a los estimados, puede definirse en primer lugar, considerando p^* como variable aleatoria, algunos conceptos básicos:

- Valor esperado $E[p^*]$
- Sesgo $E[p^*] - p^*$
- Covarianza $cov[p^*] = E \left[\{p^* - E[p^*]\} \{p^* - E[p^*]\}^T \right]$

Se definen también una serie de propiedades que son deseables para los estimadores p^* de los parámetros p :

- Estimador no sesgado: cumple, para cualquier valor de "m" (número de medidas),

$$E[p^*] = p$$

- Estimador consistente: cumple

$$\lim_{m \rightarrow \infty} P \left[|p^* - p| > \epsilon \right] = 0$$

con ϵ arbitrariamente pequeño. Es decir, p^* converge en términos de probabilidad, al valor cierto p .

- Estimador eficiente: En términos generales, un estimador es eficiente si extrae toda la información que los datos contienen respecto a los parámetros que se trata de estimar. Estadísticamente, un estimación p^* es la más eficiente si para cualquier otro estimador no sesgado p^+ ,

$$cov[p^*] = E \left[(p^* - p)^T (p^* - p) \right] \leq E \left[(p^+ - p)^T (p^+ - p) \right] = cov[p^+]$$

Una visión probabilista del problema, que no se utiliza aquí, permite suponer que tanto las medidas como los parámetros son variables aleatorias. En este caso no tiene sentido hablar de parámetros ciertos, y un estimador no sesgado se define entonces

como aquel que cumple, para cualquier valor de "m":

$$E[p^*] = E[p]$$

En general, será conveniente que el estimador que se escoja cumpla el mayor número posible de estas propiedades.

Siguiendo un paralelismo con el apartado anterior, a partir de la expresión de la función de verosimilitud [2.5], se define el estimador de máxima verosimilitud con información previa mediante:

$$\left. \frac{\partial}{\partial p} L(x, p) \right|_{\substack{x=x^* \\ p=p^*}} = 0 \quad [2.12]$$

que equivale a imponer la condición de extremo a [2.5]. Este estimador ha sido ampliamente utilizado en la literatura estadística, debido a sus interesantes propiedades:

- Normalidad asintótica. La función de probabilidad de los parámetros tiende a una distribución normal cuando $m \rightarrow \infty$.
- Estimador no sesgado asintóticamente. Para $m \rightarrow \infty$

$$E[p^*] = p$$

- Estimador eficiente cuando $m \rightarrow \infty$.
- Estimador consistente.
- Invarianza. Si p^* es una estimación de máxima verosimilitud de p , entonces una función $g(p^*)$ es también una estimación de máxima verosimilitud de $g(p)$.

Obsérvese que algunas de estas propiedades se cumplen de forma asintótica, es decir, cuando el número de medidas es muy alto.

Restringiendo el problema, si se supone que tanto el error de las medidas como el de la estimación "a priori" de los parámetros tiene una función de distribución normal, la expresión [2.12] equivale a encontrar el mínimo de la función S ([2.10]).

Si no se considera la información previa, la minimización de la función objetivo [2.4] conduce a:

$$\frac{\partial}{\partial \mathbf{p}} \left[(\Delta \mathbf{x})^T \mathbf{C}_x^{-1} (\Delta \mathbf{x}) \right] = 0 \quad \Delta \mathbf{x} = \mathbf{x}^* - \mathbf{M}(\mathbf{p}) \quad [2.13]$$

lo que constituye la versión simplificada del estimador de máxima verosimilitud. En el caso de que el modelo $\mathbf{M}(\mathbf{p})$ sea lineal respecto a los parámetros, la expresión [2.13] puede desarrollarse:

$$\begin{aligned} \frac{\partial}{\partial \mathbf{p}} \left[(\mathbf{x}^* - \mathbf{M}\mathbf{p})^T \mathbf{C}_x^{-1} (\mathbf{x}^* - \mathbf{M}\mathbf{p}) \right] \Big|_{\mathbf{p}=\mathbf{p}^*} = 0 \rightarrow \\ \rightarrow \mathbf{M}^T \mathbf{C}_x^{-1} \mathbf{M} \mathbf{p}^* - \mathbf{M}^T \mathbf{C}_x^{-1} \mathbf{x}^* = 0 \end{aligned}$$

y si $\mathbf{M}^T \mathbf{C}_x^{-1} \mathbf{M}$ tiene inversa, el estimador anterior toma la expresión:

$$\mathbf{p}^* = \left[\mathbf{M}^T \mathbf{C}_x^{-1} \mathbf{M} \right]^{-1} \mathbf{M}^T \mathbf{C}_x^{-1} \mathbf{x}^* \quad [2.14]$$

Esta ecuación [2.14] proporciona una forma directa de estimar los parámetros \mathbf{p}^* con las condiciones indicadas. \mathbf{C}_x es la matriz de covarianzas de los errores en las medidas, indicada en el apartado anterior.

En el caso de que las medidas sean independientes entre sí, la matriz de covarianzas correspondiente es diagonal, y la fórmula [2.14] representa en este caso el estimador de Markov o de mínimos cuadrados generalizados, donde C_x^{-1} representa una matriz de pesos.

Finalmente, si esta matriz es del tipo $C_x = \sigma^2 I$, lo que implica además de las condiciones restrictivas indicadas hasta ahora, que las medidas tienen la misma varianza y que son independientes entre sí, se obtiene el estimador de mínimos cuadrados:

$$p^* = \left[M^T M \right]^{-1} M^T x^* \quad [2.15]$$

Se ha visto cómo desde un punto de vista exclusivamente estadístico pueden definirse estimadores de parámetros que guardan una relación directa con los criterios definidos en el apartado anterior. Existe una diferencia, ya apuntada anteriormente, entre ambas formas de identificar parámetros: el modelo. Esta diferencia es suficientemente importante como para cambiar la metodología de cálculo en la mayoría de los casos.

En general, el problema de identificación planteado en ciencias aplicadas, y en concreto en Geotecnia, no podrá utilizar directamente las expresiones [2.14] o [2.15], ya que el modelo, salvo contadas ocasiones, no será una función lineal de los parámetros. En muchos casos, ni siquiera será expresable analíticamente. De manera que, en principio, habrá que partir de las expresiones más generales [2.12] y [2.13], minimizando las

funciones objetivo definidas por el criterio escogido.

En el apartado anterior se han indicado los motivos por los que se escoge el criterio de máxima verosimilitud para la resolución del problema inverso. El problema matemático subyacente es un problema de extremos, y por este motivo es necesario estudiar los diferentes métodos disponibles para intentar resolverlo desde un punto de vista numérico.

2.3 METODOS NUMÉRICOS DE MINIMIZACIÓN DE FUNCIONES.

2.3.1 Generalidades. Condiciones De Mínimo.

Existen varios trabajos clásicos sobre minimización de funciones, ya sea desde un punto de vista analítico o numérico. La mayoría de ellos se han desarrollado paralelamente a los trabajos sobre optimización y programación matemática. Algunos desarrollos de tipo exclusivamente matemático pueden verse en Pchenitchny y Daniline (1977), Lawson y Hanson (1974) y Lanczos (1961). Métodos de minimización refiriéndose a problemas de identificación de sistemas pueden encontrarse en Eykhoff (1974), Graupe (1972) y en Maier y Gioda (1981). Finalmente, algoritmos de minimización aplicados a problemas de optimización pueden encontrarse en Fletcher (1981), Morris (1982) y Scales (1986).

En general, la función objetivo J a minimizar dependerá de las observaciones y de los parámetros del modelo escogido. Es decir,

$$J = J(\mathbf{x}^*, \mathbf{p}) \quad [2.16]$$

Habitualmente esta dependencia no es lineal. Se trata de encontrar un conjunto de parámetros \mathbf{p}^* que hagan mínimo [2.16], imponiendo si se desea alguna condición restrictiva a las variables del problema, como por ejemplo:

$$p_i^1 \leq p_i \leq p_i^2 \quad i = 1, \dots, n$$

Este tipo de restricciones son habituales en los problemas de optimización, en los que las variables solución del problema deben cumplir unas condiciones determinadas. En los problemas de identificación, sin embargo, no es imprescindible imponerlas, puesto que si la estimación que se obtiene proporciona parámetros sin sentido físico, se deduce que las medidas no son buenas o el modelo es inadecuado. En general no es necesario, por tanto, limitar los rangos de los parámetros. En caso de que se limitara, si las medidas o el modelo no fueran adecuados, la solución se encontraría en un punto del contorno. Con frecuencia, los problemas de optimización utilizan funciones objetivo mas simples que los problemas de identificación de parámetros, pero requieren un tratamiento más complejo de los contornos, ya que es en ellos donde suele estar el punto óptimo.

Consideremos ahora la función objetivo desde un punto de vista matemático, suponiéndola continua y derivable al menos hasta la segunda derivada.

Para unas medidas determinadas, interesa hallar \mathbf{p}^* que minimice J . Recordemos que el gradiente de la función objetivo es:

$$\mathbf{g}(\mathbf{p}) = \begin{bmatrix} \partial J / \partial p_1 \\ \cdot \\ \cdot \\ \cdot \\ \partial J / \partial p_n \end{bmatrix} \quad [2.17]$$

El elemento "i" de $\mathbf{g}(\mathbf{p})$ es la pendiente de J en la dirección p_i . La derivada de la función objetivo a lo largo de una dirección ($\|\mathbf{u}\| = 1$) será:

$$\mathbf{u}^T \mathbf{g}(\mathbf{p})$$

Por otra parte, el hesiano de J es la matriz simétrica:

$$\mathbf{G}(\mathbf{p}) = \begin{bmatrix} \partial^2 J / \partial p_1^2 & \dots & \partial^2 J / \partial p_1 p_n \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \partial^2 J / \partial p_n p_1 & \dots & \partial^2 J / \partial p_n^2 \end{bmatrix} \quad [2.18]$$

Y de nuevo, la derivada segunda de la función objetivo a lo largo de una dirección \mathbf{u} será:

$$\mathbf{u}^T \mathbf{G}(\mathbf{p}) \mathbf{u}$$

Con esta notación, el desarrollo en serie de Taylor de J puede escribirse así:

$$J(\mathbf{p} + \Delta \mathbf{p}) = J(\mathbf{p}) + \Delta \mathbf{p}^T \mathbf{g}(\mathbf{p}) + \frac{1}{2} \Delta \mathbf{p}^T \mathbf{G}(\mathbf{p}) \Delta \mathbf{p} + \dots \quad [2.19]$$

En general, una función puede tener varios mínimos relativos o locales, si bien nosotros estaremos interesados en encontrar el mínimo global. En los problemas típicos de identificación, no es frecuente obtener muchos mínimos locales, pero sí que aparecen a menudo mínimos débiles (por ejemplo, en tres dimensiones, funciones que presentan un mínimo a lo largo de una línea).

El problema de la no unicidad de solución es de los más importantes en estimación de parámetros. En general, aunque la solución sea única, desde un punto de vista numérico pueden encontrarse diversas estimaciones de parámetros correspondientes a un mínimo débil y con un valor de la función objetivo muy bajo. Es decir, pueden existir varias combinaciones de parámetros que proporcionen resultados parecidos. En algunos casos puede ser el sentido físico del problema el que permitirá distinguir la combinación adecuada.

Matemáticamente se considera que un punto p^* es un mínimo de J si y sólo si

$$\forall p^{**} | p^{**} \neq p^* \rightarrow J(p^*) < J(p^{**})$$

Es importante definir a priori las condiciones que debe cumplir un mínimo, para facilitar en lo posible su búsqueda. Para la existencia de un mínimo es necesario y suficiente que se satisfagan las dos condiciones siguientes:

a) **Condición de primer orden:** Utilizando hasta el término de primer orden del desarrollo [2.19], resulta, en el mínimo:

$$J(p^* + \Delta p) = J(p^*) + \Delta p^T g(p^*)$$

[2.20]

donde se supone que $\|\Delta p\|$ es suficientemente pequeño. Si se cumple $\Delta p^T g(p^*) < 0$, entonces $J(p^* + \Delta p) < J(p^*)$, lo cual no puede ser si $J(p^*)$ es mínimo. Por otra parte, si $\Delta p^T g(p^*) > 0$ entonces $J(p^* - \Delta p) < J(p^*)$, lo que de nuevo no puede ser en el mínimo. Se deduce, por tanto que

$$\Delta p^T g(p^*) = 0 \quad \forall \Delta p$$

es decir, en el mínimo $g(p^*) = 0$.

b) **Condición de segundo orden:** Escogiendo hasta el término de segundo orden de la expresión [2.19], e imponiendo la condición anterior, se llega a $\Delta p^T G(p^*) \Delta p \geq 0$, es decir, el hesiano de la función en el mínimo es semidefinido positivo. Si el mínimo es estricto, el hesiano debe ser definido positivo.

2.3.2 Métodos Numéricos De Minimización.

La complejidad de la función objetivo "J", o el hecho de que ésta no tenga una expresión analítica explícita aconsejan estudiar los diferentes métodos de cálculo numérico disponibles para minimizar funciones.

La mayoría de estos métodos numéricos se basan en suponer que la función objetivo, en las proximidades del mínimo, se comporta como una función cuadrática. Una función de este tipo con un punto estacionario en el origen y que vale cero en él puede expresarse en la forma:

$$J(p) = \frac{1}{2} p^T G p$$

[2.21]

donde G es el hesiano de la función. Por otra parte, G admite una descomposición en la forma:

$$G = V \Lambda V^T \quad [2.22]$$

donde Λ es la matriz diagonal de valores propios del hesiano y V es la matriz ortogonal de vectores propios. Entonces la derivada segunda de J en una dirección cualquiera u será:

$$\frac{u^T G u}{\|u\|^2} \quad [2.23]$$

donde el denominador representa el cuadrado módulo del vector de dirección, en el caso de que éste no sea unitario. Utilizando la expresión [2.16], y la descomposición $u = Vc$ basada en el hecho de que los vectores propios forman una base del espacio R^n , se tiene:

$$\frac{u^T G u}{\|u\|^2} = \frac{c^T V^T (V \Lambda V^T) V c}{c^T V^T V c} = \frac{c^T \Lambda c}{\|c\|^2} = \frac{\sum_{i=1}^n \lambda_{ii} c_i^2}{\sum_{i=1}^n c_i^2} \quad [2.24]$$

luego

$$\lambda_{mín} \leq \frac{u^T G u}{\|u\|^2} \leq \lambda_{máx} \quad Y = \frac{\lambda_{máx}}{\lambda_{mín}} \quad [2.25]$$

Como la derivada segunda de J es una indicación de la magnitud y signo de su curvatura en la dirección de u , se deduce de [2.19] que los vectores propios correspondientes a los valores propios máximo y mínimo son las direcciones de máxima y mínima curvatura (Figura 2.2). El cociente entre el valor propio máximo y mínimo se denomina número de condición.

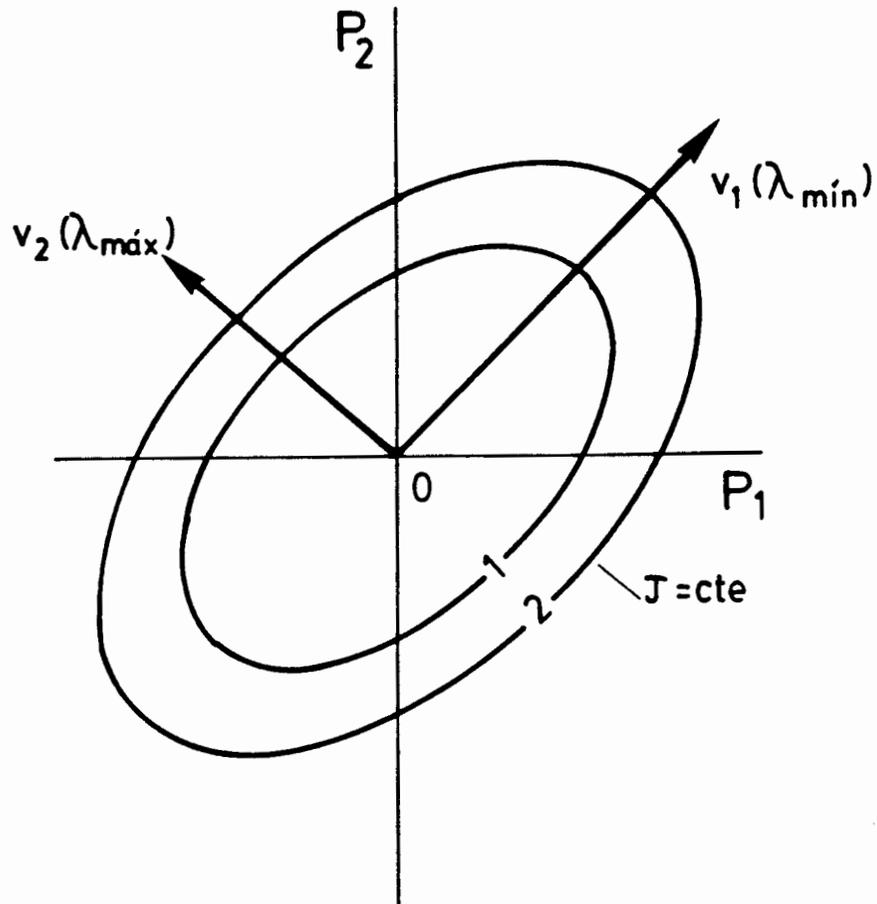


Figura 2.2. Esquema simplificado para el caso de dos parámetros, indicando las direcciones de máxima y mínima curvatura del paraboloides que aproxima a la función objetivo en el mínimo. Estas direcciones están asociadas a los valores y vectores propios del hessiano de dicha función.

Para el caso de dos parámetros, representado en la figura 2.2, el número de condición representa la excentricidad de las curvas de nivel (elípticas) en la vecindad del mínimo. Un número de condición bajo, es decir, curvas de nivel casi circulares, permite asegurar una convergencia más rápida.

2.3.3 Algoritmos De Minimización.

En general, los algoritmos para minimización de funciones se basan en escoger un punto inicial, y avanzar hacia el mínimo con un esquema de la forma:

$$P_{k+1} = P_k + \alpha_k u_k \quad J_{k+1} < J_k \quad [2.26]$$

donde u_k es un vector unitario que representa una dirección de búsqueda.

Existen muchos esquemas de minimización adaptados a menudo a tipos de funciones concretas. La literatura sobre el tema es muy dispersa, y es difícil realizar una clasificación de los diferentes métodos existentes. Aquí se presenta un resumen de los más utilizados, y se justifica la elección de uno de los métodos para la resolución del problema inverso en Geotecnia. Una relación más exhaustiva de estos esquemas puede verse en Eykhoff (1974), Fletcher (1981) y Scales (1986).

2.3.3.1 Métodos De Búsqueda Directa. -

Puede realizarse una primera división de los algoritmos de minimización en dos grupos generales: Métodos de búsqueda directa y métodos de gradiente. Los métodos de búsqueda directa no requieren una evaluación explícita del gradiente de la función objetivo, mientras que el segundo grupo se basa en el supuesto de que dicho gradiente puede ser evaluado.

Los métodos directos de búsqueda se reducen, en general, a calcular la función objetivo en una serie de puntos del espacio de parámetros, localizando de forma directa el mínimo. Existen variantes que permiten seguir una cierta dirección de descenso: por ejemplo, el método "simplex" se basa en calcular la función objetivo en "n+1" puntos equiespaciados en el espacio de n - dimensiones de los parámetros, y posteriormente decidir una dirección de búsqueda a partir del valor de la función en todos ellos, avanzando en sentido opuesto a los puntos de mayor valor. Estos métodos directos de búsqueda pueden dar buenos resultados en el caso de que la función objetivo tenga discontinuidades, o el gradiente no sea fácilmente evaluable. Sin embargo, no son en general muy eficientes, en particular si el número de parámetros o el número de condición es alto.

2.3.3.2 Métodos De Gradiente: Cálculo Del Gradiente. -

Los métodos del gradiente forman un gran grupo de técnicas de minimización, y permiten, en general, una búsqueda mucho más sistemática del mínimo. Su requisito inicial es que la función

objetivo sea continua en todo el dominio de definición.

En todos estos métodos se expresa de alguna forma el valor de α_k y u_k en la expresión [2.26].

Hay una condición general que debe satisfacer cualquier dirección de descenso: a partir de [2.26],

$$J_{k+1} = J(p_k + \alpha_k u_k) \approx J_k + \alpha_k g_k^T u_k$$

$$\alpha_k > 0 \rightarrow g_k^T u_k < 0$$

[2.27]

Luego nuestra dirección de búsqueda debe cumplir $g_k^T u_k < 0$. Por otra parte, puede hacerse una evaluación de la magnitud del vector de descenso: se trata de calcular α_k , una vez fijados p_k y u_k tal que minimice $J(p_k + \alpha_k u_k)$. Para ello puede hacerse:

$$\frac{\partial}{\partial \alpha_k} \left[J(p_k + \alpha_k u_k) \right] \Big|_{\alpha_k = \alpha_k^*} = 0$$

[2.28]

que equivale, derivando, a:

$$g^T(p_k + \alpha_k u_k) \frac{\partial}{\partial \alpha_k} (p_k + \alpha_k u_k) \Big|_{\alpha_k = \alpha_k^*} = 0 \rightarrow$$

$$\rightarrow g_{k+1}^T u_k = 0$$

[2.29]

Luego, en principio, el valor óptimo de α_k es aquel que cumple que el gradiente en el punto $k+1$ es perpendicular al vector de avance u_k . Sin embargo, no siempre es fácilmente calculable este valor, especialmente debido a limitaciones de precisión de ordenador.

En todos los métodos de gradiente se requiere estar en condiciones de calcular al menos la primera derivada de la función objetivo, lo que puede entrañar dificultades. El capítulo tercero se dedica explícitamente a la formulación conjunta de la estimación de parámetros y del método de los elementos finitos, describiendo un procedimiento para evaluar estas derivadas.

Alternativamente a los métodos analíticos para evaluar el gradiente, puede utilizarse la técnica de diferencias finitas, aunque suele conducir a mayores errores: Utilizando el desarrollo en serie de Taylor de la función $J(\mathbf{p})$, se tiene

$$\begin{aligned} J(\mathbf{p} + d_j \mathbf{e}_j) &\approx J(\mathbf{p}) + d_j \mathbf{e}_j^T \mathbf{g}(\mathbf{p}) + \frac{1}{2} d_j^2 \mathbf{e}_j^T \mathbf{G}(\mathbf{p}) \mathbf{e}_j = \\ &= J(\mathbf{p}) + d_j g_j(\mathbf{p}) + \frac{1}{2} d_j^2 G_{jj}(\mathbf{p}) \end{aligned} \quad [2.30]$$

donde d_j es un escalar suficientemente pequeño, \mathbf{e}_j es el vector unitario en la dirección coordenada "j", g_j es la componente "j" del vector gradiente y G_{jj} es la componente "jj" del hesiano. Puede aproximarse entonces el gradiente en la forma:

$$g_j(\mathbf{p}) \approx \frac{J(\mathbf{p} + d_j \mathbf{e}_j) - J(\mathbf{p})}{d_j} - \frac{1}{2} d_j G_{jj} \quad [2.31]$$

Despreciando el término del hesiano se obtiene la fórmula clásica de las diferencias finitas de primer orden. Esta expresión es exacta para funciones lineales. Usando el mismo criterio evaluando $J(\mathbf{p} - d_j \mathbf{e}_j)$ puede obtenerse la fórmula de las diferencias finitas centrales:

$$g_j(\mathbf{p}) \approx \frac{J(\mathbf{p} + d_j \mathbf{e}_j) - J(\mathbf{p} - d_j \mathbf{e}_j)}{2d_j} \quad [2.32]$$

que es exacta para funciones cuadráticas. Desarrollando en serie de Taylor la función gradiente, puede obtenerse una expresión similar para aproximar el hesiano:

$$G_j \approx \frac{g(\mathbf{p} + d_j \mathbf{e}_j) - g(\mathbf{p})}{d_j} \quad [2.33]$$

que es también exacta para funciones cuadráticas. El valor de d_j debe escogerse en función de la precisión del ordenador (Gill et al. 1981), aconsejándose un valor de $\sqrt{\varepsilon}$ donde ε es la precisión relativa de la máquina.

Existen otros métodos alternativos para calcular el gradiente. En algunos problemas determinados se ha utilizado el método del estado adjunto (Neuman et al, 1979, Carrera, 1984 en hidrología subterránea) que consiste en aumentar la función objetivo con un término adicional de forma que al minimizar condicionado a las ecuaciones de estado del problema, los mínimos coinciden. El gradiente de la nueva función objetivo puede ser más sencillo de calcular.

Hay finalmente métodos de cálculo del gradiente que son particulares de un determinado tipo de funciones. Los criterios de identificación escogidos dan lugar a un tipo de función objetivo de una estructura muy concreta, definida siempre por [2.4]. Esto supone algunas ventajas, y en el apartado siguiente se indica el método de cálculo del gradiente y de minimización para este tipo

particular de funciones.

2.3.3.3 Métodos De Gradiente: Tipos. -

A continuación se indican brevemente los fundamentos de los métodos de gradiente más utilizados.

a) Método de máximo descenso.

Utilizando la expresión [2.27], puede escribirse:

$$J_{k+1} = J(p_k + \alpha_k u_k) \approx J_k + \alpha_k g_k^T u_k \quad [2.34]$$

es decir,

$$J_{k+1} - J_k \approx \alpha_k \|g_k\| \|u_k\| \cos \theta \quad [2.35]$$

donde θ puede interpretarse como el ángulo que forman los dos vectores g_k , u_k . El primer término de esta expresión tomará el valor más negativo posible para $\theta = \pi$, fijados α_k , g_k , u_k . Luego, supuesto α_k pequeño, la mayor reducción en la función objetivo se obtiene avanzando en la dirección del gradiente y en sentido contrario: $u_k = -g_k$. Si además se avanza con α_k definido por la condición [2.30], puede demostrarse que este método conduce a un punto estacionario de la función objetivo, aunque puede no ser fácil evaluar α_k . Sin embargo, también se comprueba (Scales, 1986) que su convergencia es lenta, por lo que, en general, es poco utilizado. En la figura 2.3 puede observarse el proceso iterativo del método en una función cuadrática.

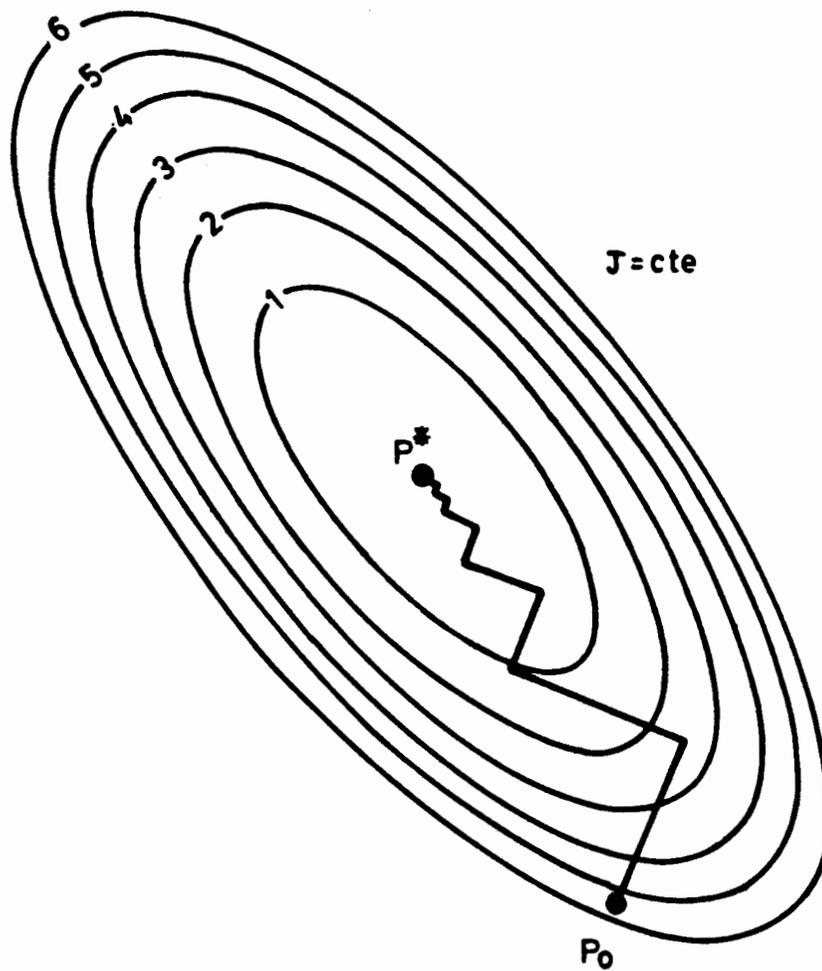


Figura 2.3. Esquema del proceso iterativo seguido por el método de máximo descenso en la minimización de una función cuadrática. Obsérvese que en cada punto se avanza según la dirección del gradiente hasta un nuevo punto en el que el gradiente es perpendicular al anterior.

b) Método de Newton.

Se basa en el hecho de que cerca del mínimo puede aproximarse $J(\mathbf{p})$ por una cuádrica. Utilizando el desarrollo en serie de Taylor del gradiente en un punto $\mathbf{p}_{k+1} = \mathbf{p}_k + \mathbf{v}_k$, se tiene:

$$\mathbf{g}_{k+1} = \mathbf{g}(\mathbf{p}_k + \mathbf{v}_k) = \mathbf{g}_k + \mathbf{G} \mathbf{v}_k \quad [2.36]$$

ya que para una función cuadrática los términos de orden superior son nulos. Como en el mínimo el gradiente debe ser nulo, $\mathbf{g}_{k+1} = 0$, se debe cumplir:

$$\mathbf{v}_k = -\mathbf{G}^{-1} \mathbf{g}_k \quad [2.37]$$

y por tanto \mathbf{v}_k es la dirección de avance. Si la función objetivo es cuadrática, esta expresión conduce al mínimo en un solo salto. En caso contrario puede procederse de forma iterativa. Puede comprobarse que \mathbf{v}_k es una dirección de descenso, y que por tanto cumple la condición general [2.27]:

$$\mathbf{g}_k^T \mathbf{v}_k < 0 \rightarrow -\mathbf{g}_k^T \mathbf{v}_k = \mathbf{g}_k^T \mathbf{G}^{-1} \mathbf{g}_k > 0 \quad [2.38]$$

ya que la matriz hessiano es definida positiva cerca del mínimo.

Puede hacerse otra interpretación de la expresión [2.37]: Minimizar una función cuadrática con matriz simétrica y definida positiva es equivalente a resolver un sistema lineal de ecuaciones. Esto ha contribuido a que exista cierto paralelismo entre algunos métodos de resolución de sistemas de ecuaciones y métodos de minimización.

El método de Newton es en general relativamente rápido, pero requiere el cálculo del hessiano de la función, lo que en muchos problemas no es factible, y en cualquier caso aumenta el tiempo de

cálculo. De hecho, se han desarrollado modificaciones al método de Newton que no requieren el cálculo directo del hesiano.

c) Métodos de gradientes conjugados.

Es una alternativa al método de Newton cuando no se puede o es demasiado costoso calcular el hesiano. Se basa en la definición de vectores conjugados: dos vectores u , v son conjugados respecto a G si $u^T G v = 0$. Los vectores propios de G cumplen esta condición. Uno de los primeros trabajos sobre el método corresponde a Hestenes y Stiefel (1952). Posteriormente se han desarrollado muchas variantes sobre el mismo.

Puede demostrarse (Fletcher 1981, Scales 1986) que dada una función cuadrática con su hesiano definido positivo, y un algoritmo de minimización del tipo $p_{k+1} = p_k + \alpha_k u_k$, donde α_k se calcula según la condición [2.29], y los vectores de avance sucesivos u_k son conjugados respecto a G , se llega a un mínimo, como mucho, en "n" iteraciones, donde "n" es el orden de la matriz hesiano.

El proceso, que no requiere el cálculo de G se basa en lo siguiente:

$$u_j^T G u_k = 0 \quad j \neq k \quad \rightarrow \quad \alpha_j u_j^T G u_k = \Delta p_j^T G u_k = \Delta g_j^T u_k = 0 \quad j \neq k$$

[2.39]

Por tanto, basta con que u_0 sea arbitrario, u_1 cualquier vector perpendicular a Δg_0 , etc. Un esquema bastante utilizado para definir los vectores de búsqueda es el siguiente:

$$u_k = -g_k + \beta_k u_{k-1} \quad k = 1, 2, \dots, n-1 \quad [2.40]$$

donde β_k es un escalar positivo y $u_0 = -g_0$. Hay varias opciones para evaluar β_k . Fletcher y Reeves (1964) proporcionaron el valor (Fletcher, 1981):

$$\beta_k = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}} \quad [2.41]$$

Otras opciones para β_k pueden verse en Fletcher (1981) y Scales (1986).

d) Métodos cuasi-Newton.

Constituye este un grupo de algoritmos que suelen converger mejor, si bien requieren un poco más de capacidad de memoria en ordenador. En muchos casos dan lugar a algoritmos del tipo de direcciones conjugadas.

Se basan en encontrar una aproximación al hesiano de la función objetivo. La expresión [2.37] proporciona la dirección de avance en el método de Newton:

$$v_k = G^{-1} g_k$$

Se aproxima el inverso del hesiano G^{-1} por una matriz H que se evalúa de forma recurrente con una expresión del tipo:

$$H_{k+1} = H_k + Q_k^H \quad [2.42]$$

La aproximación H debe cumplir algunas condiciones del hessiano: de la propiedad $G\Delta p_k = \Delta g_k$ se obtiene la condición $H_{k+1}\Delta g_k = \Delta p_k$. Se han desarrollado muchas fórmulas del tipo [2.42]. Una relación de ellas puede verse en Fletcher (1981) y Scales (1986). Como ejemplo se indica la fórmula de Davidon - Fletcher - Powell (1963):

$$H_{k+1} = H_k + \frac{\Delta p_k \Delta p_k^T}{\Delta p_k^T \Delta g_k} - \frac{H_k \Delta g_k \Delta g_k^T H_k}{\Delta g_k^T H_k \Delta g_k} \quad [2.43]$$

e) Métodos para tipos de funciones particulares.

Los métodos anteriores son aplicables, en principio, a cualquier función, y sirven de base teórica para desarrollos posteriores. Sin embargo, a menudo es posible plantear algoritmos más eficientes para algunos tipos determinados de funciones. En el caso de la estimación de parámetros, las funciones objetivo habituales suelen ser normas de vectores, o suma de funciones al cuadrado, tal como se indicaba en la expresiones [2.1] y [2.2]. En este caso particular, la función objetivo siempre es positiva en todos sus puntos, y puede representarse en la forma:

$$J(\mathbf{p}) = \sum_{i=1}^m f_i^2(\mathbf{p}) \rightarrow J(\mathbf{p}) = \mathbf{f}^T(\mathbf{p})\mathbf{f}(\mathbf{p}) \quad [2.44]$$

El gradiente es, en este caso:

$$g_j = 2 \sum_{i=1}^m f_i \frac{\partial f_i}{\partial p_j} \quad [2.45]$$

y definiendo la matriz jacobiano como :

$$\mathbf{J} = \begin{bmatrix} \frac{\partial f_1}{\partial p_1} & \dots & \frac{\partial f_1}{\partial p_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial p_1} & \dots & \frac{\partial f_m}{\partial p_n} \end{bmatrix}$$

[2.46]

resulta

$$\mathbf{g} = 2\mathbf{J}^T \mathbf{f}$$

[2.47]

Diferenciando [2.45] respecto a p_k se obtiene la componente "kj" del hesiano:

$$G_{kj} = 2 \sum_{i=1}^m \left[\frac{\partial f_i}{\partial p_k} \frac{\partial f_i}{\partial p_j} + f_i \frac{\partial^2 f_i}{\partial p_k \partial p_j} \right]$$

[2.48]

Llamando \mathbf{T}_i a la matriz hesiano de f_i :

$$\mathbf{T}_i(\mathbf{p}) = \nabla^2 f_i(\mathbf{p})$$

el hesiano de la función objetivo es:

$$\mathbf{G}(\mathbf{p}) = 2\mathbf{J}^T(\mathbf{p})\mathbf{J}(\mathbf{p}) + 2 \sum_{i=1}^m f_i(\mathbf{p})\mathbf{T}_i(\mathbf{p})$$

[2.49]

resultan finalmente dos términos:

$$\mathbf{G} = 2\mathbf{J}^T \mathbf{J} + 2\mathbf{Q}$$

[2.50]

Si se aplica el método de Newton en este caso, se obtiene, a partir de [2.37]:

$$\mathbf{G}_k \mathbf{u}_k = -\mathbf{g}_k \rightarrow (\mathbf{J}_k^T \mathbf{J}_k + \mathbf{Q}_k) \mathbf{u}_k = -\mathbf{J}_k^T \mathbf{f}_k$$

[2.51]

y se avanza con el esquema $p_{k+1} = p_k + u_k$. El principal problema es calcular la componente Q_k del hesiano, por lo que suele no considerarse, y trabajar con la expresión aproximada:

$$J_k^T J_k u_k = -J_k^T f_k \quad [2.52]$$

Este método, denominado de Gauss - Newton, coincide en la estructura con el estimador de mínimos cuadrados [2.15]. Cerca del mínimo, tiende al método de Newton, ya que el término Q se hace pequeño. De hecho, es de esperar que este término sea pequeño en los problemas de identificación, ya que en ellos las funciones f_i dependen de las diferencias entre variables medidas y calculadas.

El método puede plantear problemas de resolución en el caso de que la matriz $J^T J$ sea singular. En estos casos es especialmente útil la modificación al método debida a Levenberg - Marquardt (Marquardt, 1963) :

$$(J_k^T J_k + \mu_k I) u_k = -J_k^T f_k \quad [2.53]$$

donde μ_k es un escalar positivo variable en cada iteración, y que en el mínimo se hace nulo. El término $\mu_k I$ de [2.53] facilita, en general, la convergencia.

Como los criterios de identificación escogidos conducen a funciones objetivo del tipo [2.44], es de esperar que este método proporcione buenos resultados en los problemas de estimación de parámetros. Así se ha comprobado por varios autores: El método de Gauss-Newton y Marquardt ha sido utilizado con éxito en la

resolución del problema inverso en problemas muy variados. Por este motivo, en el apartado siguiente se estudian con más detalle estos métodos.

2.4 FORMULACIÓN Y MÉTODOS DE RESOLUCIÓN ADOPTADOS.

Como ya se ha indicado, de los algoritmos de minimización explicados en el apartado anterior, el método de Gauss-Newton y la variante debida a Marquardt son los más apropiados para el problema que nos ocupa. Tienen la característica de que al ser la función objetivo del tipo

$$J(\mathbf{p}) = \sum_{i=1}^m f_i^2(\mathbf{p}) \rightarrow J(\mathbf{p}) = \mathbf{f}^T(\mathbf{p})\mathbf{f}(\mathbf{p}) \quad [2.54]$$

una buena aproximación al hesiano cerca del mínimo es:

$$\mathbf{G} \approx 2 \mathbf{J}^T \mathbf{J}$$

donde \mathbf{J} es la matriz jacobiana definida en [2.46]. De esta forma puede emplearse el método de Newton de minimización. Como la función objetivo no será, en general cuadrática, habrá que proceder por iteraciones sucesivas.

Se desarrolla a continuación el cálculo de la función objetivo y se formula su minimización en el problema de identificación planteado, utilizando estos algoritmos. Para ello se define el problema en primer lugar con el criterio más simple de mínimos cuadrados, para pasar posteriormente al criterio de máxima verosimilitud. En cada caso se indica la forma de combinar estos criterios con los algoritmos de minimización escogidos.

2.4.1 Utilización Del Criterio De Mínimos Cuadrados.

Supongamos que el modelo geotécnico planteado puede formularse en forma explícita como

$$\mathbf{x} = \mathbf{M}(\mathbf{p}) \quad [2.55]$$

donde, con la notación habitual, \mathbf{x} son las variables medibles y \mathbf{p} son los parámetros desconocidos. La relación expresada por [2.55] será, en general, no lineal.

Por otra parte, el criterio de identificación de mínimos cuadrados propone que los parámetros solución al problema inverso son aquellos que minimizan la función objetivo [2.1], cuya expresión es:

$$J = \left[\mathbf{x}^* - \mathbf{M}(\mathbf{p}^*) \right]^T \left[\mathbf{x}^* - \mathbf{M}(\mathbf{p}^*) \right] \quad [2.56]$$

donde \mathbf{x}^* son las variables que se han medido, y \mathbf{p}^* son los parámetros solución al problema. En principio, $\mathbf{x}^* \neq \mathbf{M}(\mathbf{p}^*)$ puesto que las medidas contendrán un cierto error, y el modelo no será totalmente cierto. Por este motivo se realiza una estimación de \mathbf{p} utilizando alguno de los criterios de identificación.

Para facilitar el trabajo puede linealizarse la expresión del modelo [2.55]. Esto obligará a trabajar de forma iterativa, pero simplificará el cálculo en cada paso. Desarrollando en serie de Taylor [2.55] alrededor de un punto \mathbf{p}_0 de parámetros iniciales,

$$\mathbf{x} = \mathbf{M}(\mathbf{p}_0) + \left. \frac{\partial \mathbf{M}(\mathbf{p})}{\partial \mathbf{p}} \right|_{\mathbf{p}_0} \Delta \mathbf{p} + \dots \quad [2.57]$$

que puede escribirse en la forma:

$$\Delta \mathbf{x} = \mathbf{x} - \mathbf{M}(\mathbf{p}_0) \approx \left. \frac{\partial \mathbf{M}(\mathbf{p})}{\partial \mathbf{p}} \right|_{\mathbf{p}_0} \Delta \mathbf{p} = \mathbf{A} \Delta \mathbf{p} \quad [2.58]$$

Si se introducen las variables medidas \mathbf{x}^* en [2.58], se obtiene la expresión:

$$\Delta \mathbf{x}^* = \mathbf{x}^* - \mathbf{M}(\mathbf{p}_0) \approx \mathbf{A} \Delta \mathbf{p}^* \quad [2.59]$$

donde

$$\Delta \mathbf{p}^* = \mathbf{p}^* - \mathbf{p}_0 \quad \mathbf{A} = \left. \frac{\partial \mathbf{x}}{\partial \mathbf{p}} \right|_{\mathbf{p}_0}$$

Conviene observar que el número de parámetros "n" es, habitualmente, menor que el de variables medidas "m", por lo que la matriz \mathbf{A} es de dimensiones $m \times n$.

Un sistema de ecuaciones del tipo

$$\Delta \mathbf{x} = \mathbf{A} \Delta \mathbf{p} \quad [2.60]$$

con \mathbf{A} no cuadrada, no tiene en general solución. Sin embargo, puede plantearse el problema como uno clásico de mínimos cuadrados en la forma siguiente: Dada una matriz real \mathbf{A} "m x n" de rango $k < \min(m, n)$, y dado un vector real $\Delta \mathbf{x}$ de dimensión "m", encontrar un vector $\Delta \mathbf{p}$ de dimensión "n" que minimice la norma euclídea del vector $\mathbf{A} \Delta \mathbf{p} - \Delta \mathbf{x}$. El desarrollo matemático puede verse en Lawson y Hanson (1974).

La solución al problema así planteado se obtiene utilizando el operador o estimador de mínimos cuadrados, ya definido en [2.15]. En este caso se usa una expresión incremental, debido a la no

linealidad del modelo:

$$\Delta x = A \Delta p$$

$$A^T \Delta x = A^T A \Delta p \rightarrow (A^T A)^{-1} A^T \Delta x = \Delta p$$

[2.61]

donde se ha supuesto que la matriz $(A^T A)$ "n x n", tiene inversa. La matriz $H = (A^T A)^{-1} A^T$ se suele denominar inversa generalizada, pseudoinversa, o inversa de Lanczos (Lanczos, 1961).

La expresión [2.61] proporciona el vector corrección Δp que hay que aplicar en cada iteración:

$$p_{k+1} = p_k + \Delta p_k \quad k = 0, 1, \dots$$

[2.62]

donde

$$\Delta p_k = (A_k^T A_k)^{-1} A_k^T \Delta x_k^* \quad \Delta x_k^* = x^* - M(p_k)$$

[2.63]

y en cada iteración, usando el operador de mínimos cuadrados, se minimiza

$$\|A_k \Delta p_k - \Delta x_k^*\|$$

[2.64]

El proceso iterativo converge hacia unos parámetros p^* al reducirse progresivamente el vector de incrementos Δp , de manera que la expresión [2.64] tiende hacia $\|\Delta x_k^*\|$. Por tanto, la solución del algoritmo iterativo descrito por [2.62], [2.63] y [2.64] minimiza la función objetivo $J = (\Delta x^*)^T (\Delta x^*)$, que era nuestro propósito inicial.

En definitiva este algoritmo es una utilización incremental del método de mínimos cuadrados. Si se compara con las expresiones generales desarrolladas en [2.51] y [2.52], es fácil ver la equivalencia. En concreto, al ser

$$J = f^T f \qquad f = (\Delta x^*) = x^* - x \qquad [2.65]$$

la matriz jacobiano definida según [2.46] es :

$$J = \frac{\partial f}{\partial p} = - \frac{\partial x}{\partial p} = - A \qquad [2.66]$$

y por tanto, las expresiones que definen la dirección de búsqueda del mínimo en el método de Gauss-Newton [2.52] y Marquardt [2.53], son:

$$\begin{aligned} A_k^T A_k u_k &= A_k^T f_k \\ (A_k^T A_k + \mu_k I) u_k &= A_k^T f_k \end{aligned} \qquad [2.67]$$

donde ahora $f_k = \Delta x_k^*$ y la dirección de búsqueda es $u_k = \Delta p_k$.

Se comprueba, por tanto, que el estimador de mínimos cuadrados [2.61] y la fórmula del algoritmo de Gauss-Newton [2.66] coinciden en este caso. La modificación de [2.66] del algoritmo de Marquardt sirve en cualquier caso para mejorar la convergencia. Aunque las dos fórmulas se han obtenido por caminos diferentes, en ambas se ha partido de la forma especial de la función objetivo [2.54] y [2.56].

2.4.2 Utilización Del Criterio De Máxima Verosimilitud.

La metodología de cálculo indicada en el apartado anterior puede generalizarse fácilmente cuando se utiliza el criterio de identificación de máxima verosimilitud. La expresión [2.14] representa una generalización del operador de mínimos cuadrados, y en este caso, trabajando incrementalmente, se escribe en la forma :

$$\Delta p = \left[A^T C_x^{-1} A \right]^{-1} A^T C_x^{-1} \Delta x^* \quad [2.68]$$

Esto no es más que una generalización de [2.61] utilizando una interpretación probabilista del problema inverso. La función objetivo es en este caso:

$$J = (\Delta x)^T C_x^{-1} (\Delta x) \quad [2.69]$$

donde C_x representa la matriz de covarianzas de las medidas (m x m), que por propia definición es simétrica y definida positiva.

Los desarrollos anteriores son igualmente aplicables a una función objetivo de este tipo. Para comprobar que este problema puede formularse en términos similares que el de mínimos cuadrados, considérese la descomposición de la matriz de covarianzas en la forma:

$$C_x = F \Lambda_d F^T \quad [2.70]$$

donde F es la matriz ortogonal de vectores propios y Λ_d la matriz diagonal de valores propios. Haciendo el cambio de variables

$$\Delta \mathbf{x}' = \Lambda_d^{-\frac{1}{2}} \mathbf{F}^T \Delta \mathbf{x} \qquad \mathbf{A}' = \Lambda_d^{-\frac{1}{2}} \mathbf{F}^T \mathbf{A}$$

[2.71]

la ecuación [2.69] se transforma en

$$J = (\Delta \mathbf{x}')^T (\Delta \mathbf{x}')$$

[2.72]

y de la misma forma, la ecuación $\Delta \mathbf{x} = \mathbf{A} \Delta \mathbf{p}$ se transforma en

$$\Lambda_d^{-\frac{1}{2}} \mathbf{F}^T \Delta \mathbf{x} = \Lambda_d^{-\frac{1}{2}} \mathbf{F}^T \mathbf{A} \Delta \mathbf{p}$$

$$\Delta \mathbf{x}' = \mathbf{A}' \Delta \mathbf{p}$$

[2.73]

De manera que la formulación indicada en la sección anterior puede aplicarse para minimizar [2.69] utilizando un criterio de máxima verosimilitud. La expresión a utilizar en este caso será la [2.68].

2.4.3 Consideración De La Información Previa.

Generalizando el criterio de identificación, puede obtenerse una expresión para el vector incremento de parámetros en el caso de que se disponga de información previa.

Esta información previa sobre los valores de los parámetros se suele incluir como un término adicional en la función objetivo a minimizar. La razón de esta adición se ha visto en el apartado 2.1.4, al considerar los parámetros como variables aleatorias y tener en cuenta su función de distribución en la expresión de la verosimilitud.

La función objetivo resultante tiene la expresión:

$$J = (\Delta \mathbf{x})^T \mathbf{C}_x^{-1} (\Delta \mathbf{x}) + (\Delta \mathbf{p}^o)^T \mathbf{C}_p^{o-1} (\Delta \mathbf{p}^o) \quad [2.74]$$

donde $\Delta \mathbf{p}^o = \mathbf{p} - \mathbf{p}^o$ y \mathbf{p}^o es un vector inicial de parámetros conocidos "a priori". \mathbf{C}_p^o representa una matriz de covarianzas de los parámetros conocidos a priori.

En este caso, el vector incremento de parámetros que minimiza [2.74] es (Jackson, 1972 y Menke, 1984):

$$\Delta \mathbf{x} = \mathbf{A} \Delta \mathbf{p}$$

$$\Delta \mathbf{p} = \Delta \mathbf{p}^o + \left[\mathbf{A}^T \mathbf{C}_x^{-1} \mathbf{A} + \mathbf{C}_p^{o-1} \right]^{-1} \mathbf{A}^T \mathbf{C}_x^{-1} (\Delta \mathbf{x} - \mathbf{A} \Delta \mathbf{p}^o) \quad [2.75]$$

Esta es una de las expresiones más generales para estimar parámetros utilizando información previa. Existen fórmulas alternativas, en función del tipo de información previa que se considere, pero la formulación anterior es la más común, ya que se deduce de la generalización del concepto de verosimilitud incluyendo medidas y parámetros.

2.4.4 Descomposición En Valores Singulares.

Históricamente el método de mínimos cuadrados ha sido el más utilizado en múltiples aplicaciones. Es posible encontrar en la literatura una gran variedad de interpretaciones físicas de este método y de sus propiedades. Así, por ejemplo, en Geofísica su uso ha sido muy extendido. En este sentido son clásicos los trabajos de Backus (1970), Jackson (1972) y Wiggins (1972). Algunas de las propiedades del operador estudiado matemáticamente por Lanczos

(1961) admiten en estos trabajos un significado físico interesante. Por otra parte, estas propiedades se pueden extender si se utiliza el criterio de máxima verosimilitud. Por este motivo, se ha considerado oportuno estudiar en detalle las características de este operador.

La matriz de derivadas de las variables medibles respecto a los parámetros, también denominada matriz de sensibilidad, A , definida en la expresión [2.58], tiene dimensiones $m \times n$, y admite una descomposición única del tipo (Lanczos, 1961):

$$A = U \Lambda V^T \quad [2.76]$$

donde:

U ($m \times k$) es una matriz ortogonal cuyas columnas son los vectores propios de la matriz $(A A^T)$ correspondientes a sus valores propios no nulos,

V ($n \times k$) es una matriz ortogonal cuyas columnas son los vectores propios de la matriz $(A^T A)$, correspondientes a sus valores propios no nulos,

Λ ($k \times k$) es una matriz con sólo "k" elementos de su diagonal principal diferentes de cero y positivos. "k" es el rango de la matriz A , con $k < \min(m, n)$. Los elementos de Λ se denominan valores singulares de A .

Esta descomposición es muy útil para ilustrar el proceso de minimización.

Puede demostrarse (Lanczos 1961) que existen "k" valores propios distintos de cero comunes a las matrices $(A^T A)$ y $(A A^T)$ $\Gamma_1, \dots, \Gamma_k$. Se comprueba que los valores singulares de A son las raíces cuadradas positivas de los valores propios $\Gamma_1, \dots, \Gamma_k$. Es decir:

$$\Lambda = \begin{bmatrix} \lambda_1 & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & \lambda_k \end{bmatrix} \quad \begin{aligned} A^T A v_j &= \Gamma_j v_j \\ A A^T u_i &= \Gamma_i u_i \\ \lambda_i &= +\sqrt{\Gamma_i} \end{aligned} \quad [2.77]$$

En realidad el hecho de que A esté asociado a dos espacios, uno de dimensión "n" y otro de dimensión "m", permite plantear el problema de valores propios en forma cruzada:

$$A v_j = \lambda_j u_j \quad A^T u_i = \lambda_i v_i \quad [2.78]$$

Conviene observar que la descomposición [2.78] permite reconstruir la matriz A sólo con los valores propios diferentes de cero, y se puede calcular la matriz pseudoinversa H aun cuando no exista la inversa de $(A^T A)$, a partir de la expresión [2.78] :

$$H = V \Lambda^{-1} U^T \quad [2.79]$$

H es, por tanto, una inversa en el sentido de mínimos cuadrados. Es además la única matriz que cumple las condiciones de Penrose (Lawson y Hanson 1974):

$$AHA = A$$

$$HAH = H$$

$$(AH)^T = AH$$

$$(HA)^T = HA$$

[2.80]

La matriz $R = HA$ ($n \times n$) se denomina matriz de resolución. Sus filas se denominan núcleos de resolución. Esta matriz da idea de la unicidad de la solución. Si el sistema es completo y $(A^T A)^{-1}$ existe, R coincide con la identidad, ya que, de [2.61],

$$H = V \Lambda^{-1} U = (A^T A)^{-1} A^T$$

y se tendrá:

$$R = HA = (A^T A)^{-1} A^T A = I$$

y en este caso los parámetros están determinados de forma única.

Por otra parte, la matriz $S = AH$ ($m \times m$) recibe el nombre de matriz densidad de información, y representa una medida de la independencia de los datos y de cómo están relacionados entre sí a través del modelo. En el caso de que el número de medidas coincida con el de parámetros, y el sistema sea determinado, la matriz es la identidad. El estudio de esta matriz revela también la estructura del modelo para cada valor de los parámetros, y la importancia relativa de unas u otras medidas.

2.4.5 Análisis De Errores.

Una de las ventajas de un tratamiento estadístico del problema inverso es que se puede obtener información acerca de la calidad de la identificación realizada, medida en términos de varianzas de los parámetros.

En efecto, a partir de una relación entre medidas y parámetros del tipo [2.62] (criterio de máxima verosimilitud), y conocida la matriz de covarianzas de las medidas, puede evaluarse la matriz de covarianzas de los parámetros estimados (Bury, 1975). De la relación lineal $\Delta p = H \Delta x$ se obtiene, utilizando las propiedades de la covarianza:

$$\text{cov}[\Delta p] = H \text{cov}[\Delta x] H^T \quad [2.81]$$

Cerca del mínimo de la función objetivo, las covarianzas incrementales coinciden con las covarianzas de las medidas y de los parámetros, y sustituyendo por el valor de H definido por [2.68], resulta

$$\text{cov}[p] = \left[(A^T C_x^{-1} A)^{-1} A^T C_x^{-1} \right] C_x \left[(A^T C_x^{-1} A)^{-1} A^T C_x^{-1} \right]^T \quad [2.82]$$

obteniéndose la expresión:

$$C_p = \text{cov}[p] = (A^T C_x^{-1} A)^{-1} = (A^T A)^{-1} \quad [2.83]$$

En realidad este valor de la covarianza de los parámetros es sólo una cota inferior, debido a la linealidad supuesta en [2.81], pero permite disponer de una estimación del error con que identificamos nuestros parámetros. Esto es especialmente útil en problemas reales, y su importancia se constatará en los capítulos posteriores.

Si se utiliza el operador de mínimos cuadrados, implícitamente se está suponiendo que $C_x = \sigma_x^2 I$, y la matriz de covarianzas de los parámetros es:

$$C_p = \sigma_x^2 (A^T A)^{-1} \quad [2.84]$$

Las varianzas de los parámetros estimados serán los elementos de la diagonal principal de C_p . Si se utiliza la descomposición singular definida en [2.76], puede hacerse un cambio de variables que permite conocer mejor la estructura del problema (Wiggins, 1972). Haciendo

$$\begin{aligned} \Delta p^+ &= V^T \Delta p \\ \Delta x^+ &= U^T \Delta x \end{aligned} \quad [2.85]$$

la ecuación $\Delta p = V \Lambda^{-1} U^T \Delta x$ se transforma en:

$$\Delta p^+ = \Lambda^{-1} \Delta x^+ \quad \rightarrow \quad \Delta p_i^+ = \left[1/\lambda_i \right] \Delta x_i^+ \quad [2.86]$$

Esta ecuación indica cómo una combinación lineal de las medidas (definida por los vectores propios U), está directa y unívocamente relacionada con una combinación lineal de los parámetros (definida por los vectores propios V), a través del inverso del correspondiente valor propio.

La observación directa de los vectores propios calculados en el mínimo de la función objetivo, puede ayudar a conocer mejor la estructura del modelo. Por ejemplo, puede ocurrir que un parámetro esté muy asociado a un valor propio concreto, lo cual es bueno, puesto que se reducen las correlaciones entre parámetros (correlaciones "a posteriori").

La expresión [2.86] permite deducir también la fórmula de la varianza de los parámetros:

$$\begin{aligned} \text{var}(\Delta p_i^+) &= \left(1/\lambda_i\right) \text{var}(\Delta x_i^+) \\ \text{var}(\Delta p_i) &= \sum_{j=1}^n (V_{ij}^2/\lambda_j^2) \text{var}(\Delta x_j^+) \end{aligned} \quad [2.87]$$

Se observa directamente que los valores propios pequeños originan varianzas grandes, que con frecuencia están asociadas a algunos parámetros en concreto.

2.4.6 Aplicación Del Algoritmo De Gauss-Newton.

Se ha visto la manera de resolver iterativamente el problema inverso utilizando los diferentes criterios de identificación considerados. El vector de avance en el proceso iterativo tiene, para el algoritmo de Gauss-Newton y con el criterio de mínimos cuadrados la siguiente expresión:

$$u_k = \left(A_k^T A_k\right)^{-1} A_k^T f_k \quad f_k = \Delta x_k \quad [2.88a]$$

y con el criterio de máxima verosimilitud la siguiente:

$$u_k = \left(A_k^T C_x^{-1} A_k\right)^{-1} A_k^T C_x^{-1} f_k \quad [2.88b]$$

En todas estas expresiones aparece la matriz de sensibilidad como elemento básico para llevar a cabo el proceso de identificación. Esta matriz de derivadas de los desplazamientos respecto a los parámetros no siempre es sencilla de calcular. En el apartado 2.3.3.2 se indicaron algunos métodos de cálculo del gradiente de la función objetivo, relacionado directamente con la matriz de sensibilidad a través de la expresión [2.65] y [2.66].

Para el tipo de problemas objeto de estudio, es aconsejable plantear una metodología de cálculo de esta matriz que esté directamente relacionada con el cálculo del problema directo. Como éste se realiza en muchos casos mediante el método de los elementos finitos, en el capítulo tercero se desarrolla un procedimiento para utilizar el mismo método en el cálculo de la matriz A . En el caso de que el modelo esté definido en forma analítica, el cálculo de esta matriz es directo.

Una vez evaluada la matriz de sensibilidad para unos parámetros iniciales, p_0 , se aplica el proceso iterativo en la forma indicada por [2.26]:

$$p_{k+1} = p_k + u_k$$

donde ahora u_k es el vector de avance definido por las expresiones anteriores [2.88a] y [2.88b] correspondiente al algoritmo de Gauss-Newton.

2.4.7 Aplicación Del Algoritmo De Marquardt.

La modificación del algoritmo de Gauss-Newton debida primero a los trabajos de Levenberg (1944) y posteriormente a Marquardt (1963) permite facilitar la convergencia del proceso en problemas conflictivos desde este punto de vista.

La expresión del vector de avance es, para este algoritmo, y con el criterio de mínimos cuadrados:

$$u_k = \left(A_k^T A_k + \mu_k I \right)^{-1} A_k^T f_k \quad [2.89a]$$

y utilizando el criterio de máxima verosimilitud:

$$u_k = \left(A_k^T C_x^{-1} A_k + \mu_k I \right)^{-1} A_k^T C_x^{-1} f_k \quad [2.89b]$$

El algoritmo de Gauss-Newton se convierte en un caso particular del de Marquardt, al hacer $\mu_k = 0$. De manera que, de forma general, puede escribirse:

$$\begin{aligned} \Delta p_k &= (A_k^T C_x^{-1} A_k + \mu_k I)^{-1} A_k^T C_x^{-1} \Delta x^* \\ p_{k+1} &= p_k + \Delta p_k \end{aligned} \quad [2.90]$$

donde $\mu_k \rightarrow 0$ cerca del mínimo.

Desde un punto de vista numérico el algoritmo de Gauss-Newton se ha mostrado muy eficaz en la mayoría de los problemas que se plantean en este trabajo. Sin embargo, en algunos casos con matrices de covarianza complejas ha sido necesario recurrir a la generalización de Marquardt para garantizar la convergencia del método. El uso de uno u otro depende, en definitiva, de la complejidad (desde un punto de vista de extremos) de la función objetivo.

El método de Marquardt admite una interpretación geométrica simple: Si $\mu_k = 0$ se avanza según el método de Gauss-Newton hacia el mínimo de un paraboloides tangente a la función objetivo en el punto p_k . Si μ_k es muy grande, entonces

$$\Delta p_k \rightarrow (\mu_k I)^{-1} A_k^T C_x^{-1} \Delta x^* = \mu_k^{-1} A_k^T \Delta x^* = - \mu_k^{-1} J_k^T \Delta x^* \quad [2.91]$$

donde J_k es el jacobiano en la iteración k , definido según [2.46] y [2.66]. Es decir, si μ_k es muy grande, se avanza un incremento

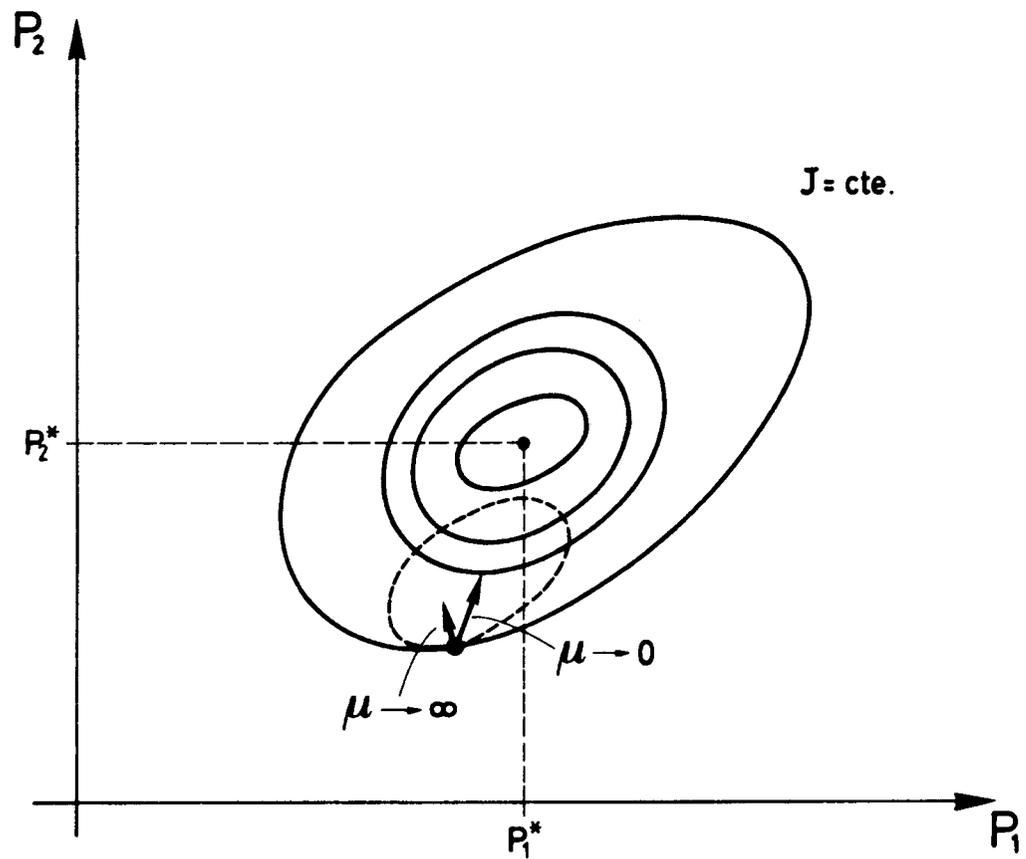


Figura 2.4. Esquema indicando las direcciones de avance en el método de Marquardt en un ejemplo de dos parámetros. Para $\mu = 0$ se avanza hacia el centro del paraboloide tangente al punto considerado. Para μ muy grande, se avanza según el gradiente.

muy pequeño en la dirección del gradiente. En la figura 2.4 se representa esquemáticamente esta propiedad del algoritmo de Marquardt.

Existen también otras interpretaciones de este algoritmo. En principio puede considerarse que el término adicional que se añade respecto al método de Gauss-Newton hace el término entre paréntesis en [2.89b] "más definido positivo", por analogía con las expresiones [2.50] y [2.51]. Sin embargo, puede darse un significado concreto a este término adicional. Puede demostrarse (Menke, 1984) que las expresiones [2.89] son la solución al problema inverso en el caso de que se minimice la función objetivo:

$$J = \Delta x^T C_x^{-1} \Delta x + \mu_k \Delta p^T \Delta p \quad [2.92]$$

que incluye el término habitual de las medidas y un término de información previa de los parámetros. El valor de μ_k controla la importancia de este término, que en el mínimo es nula al ser $\mu_k \rightarrow 0$.

Marquardt (1963) propuso un esquema de variación de μ_k para optimizar el número de iteraciones:

- Se fija, por ejemplo, $\mu_0 = 10^{-2}$ y un valor inicial $\rho > 1$, por ejemplo $\rho = 10$.
- En la iteración "k" se calcula la función objetivo J_k y se compara con el valor de la iteración anterior. Si $J_{k+1} < J_k$, entonces se ha avanzado correctamente y se hace $\mu_{k+1} = \mu_k / \rho$, de manera que si el algoritmo converge hacia el mínimo, progresivamente $\mu_k \rightarrow 0$.
- Si por el contrario, $J_{k+1} > J_k$, se sigue en el punto p_k y

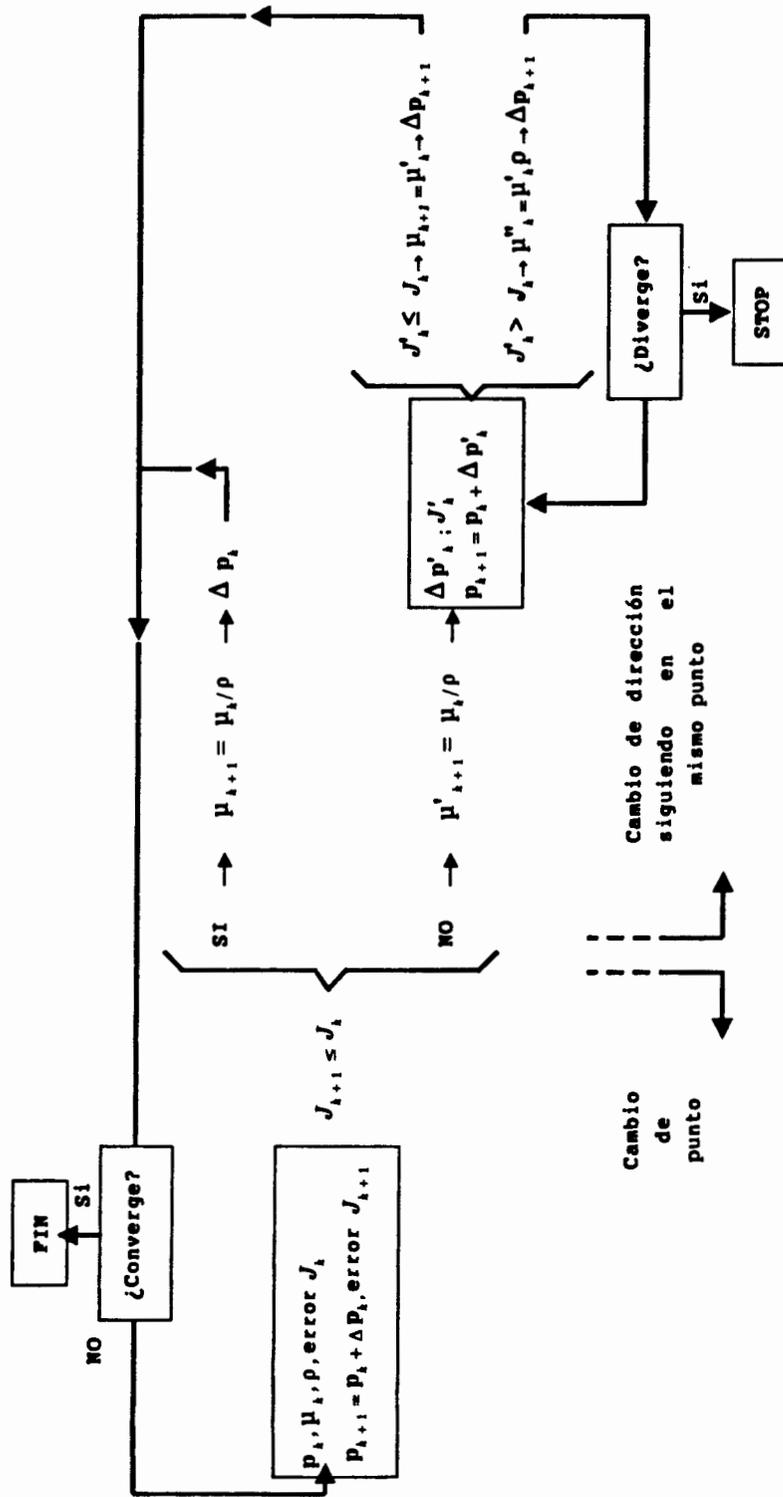


Figura 2.5. Esquema de la metodología del algoritmo de Levenberg - Marquardt.

se intenta avanzar en una nueva dirección que se calcula con

$$\mu_{k+1} = \mu_k \rho \quad .$$

En la figura 2.5 se representa esquemáticamente la estrategia de variación de μ_k y ρ en el algoritmo de Marquardt, y que ha sido utilizada con éxito en los problemas de identificación de parámetros que se resuelven en este trabajo.

2.5 EJEMPLO DE APLICACIÓN.

Con el fin de ilustrar los conceptos indicados en los apartados anteriores, se presenta a continuación un ejemplo simple que permite mostrar la metodología escogida, de entre todas las presentadas, para resolver los problemas de identificación de parámetros en sistemas geotécnicos.

Se trata de identificar los parámetros elásticos, E , módulo de elasticidad, y ν , coeficiente de Poisson, de un espacio elástico, homogéneo indefinido e isótropo, sobre el que se excava un túnel de sección circular. El problema se plantea en dos dimensiones y en el supuesto de deformación plana. Idealizando, en la medida de lo posible, la situación real en la excavación de un túnel, se supone que existen unas tensiones iniciales de valor "q" actuando en la clave del túnel, y de valor $K_0 q$ actuando en la zona de hastiales (Figura 2.6), donde K_0 tiene el sentido de coeficiente de empuje al reposo (aunque se trabaja en tensiones totales), con la notación habitual en Ingeniería Geotécnica. En estas condiciones el campo de desplazamientos tiene expresión analítica (Rodríguez y Serrano, 1980) y en el contorno del túnel ($r=a$) el desplazamiento radial

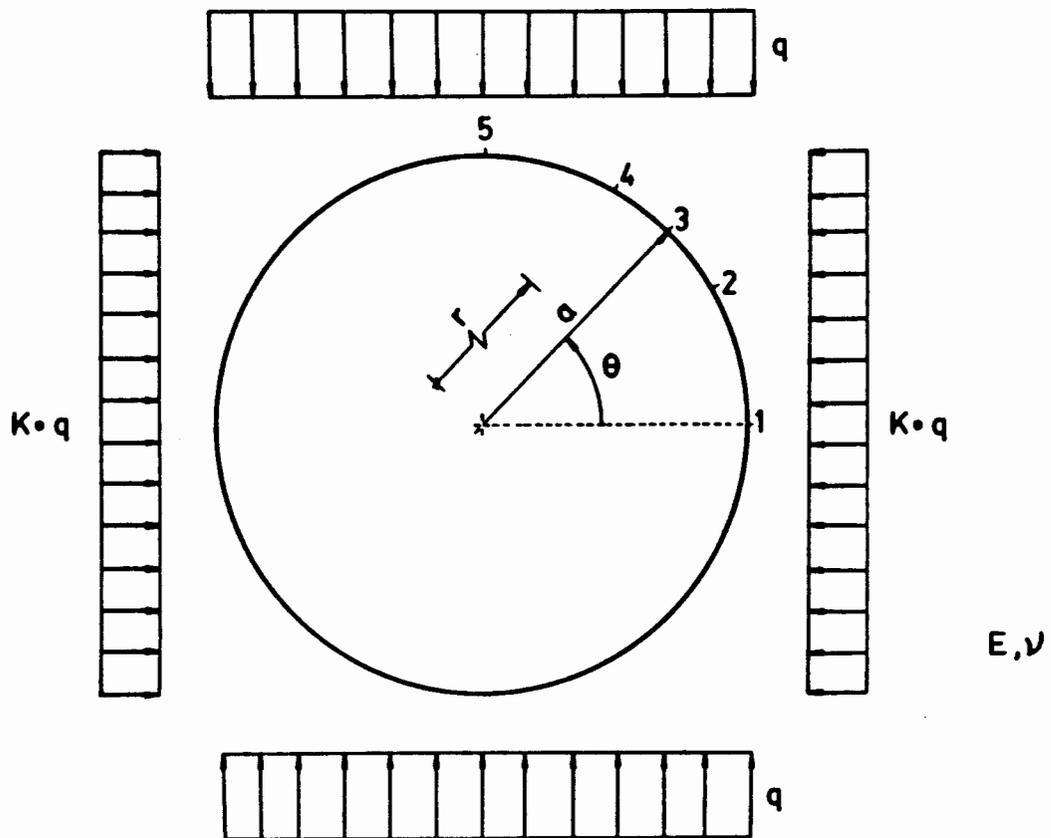


Figura 2.6. Geometría utilizada para el ejemplo teórico de túnel circular en un espacio indefinido, elástico lineal, homogéneo e isótropo. Se indican los cinco puntos considerados como puntos de medida.

vale:

$$u_r(r=a) = \frac{qa}{E} (1 + \nu) \left[\left(\frac{1 + K_o}{2} \right) - 3 \left(\frac{1 - K_o}{2} \right) (1 - \nu) \cos 2\theta \right] \quad [2.93]$$

donde el significado de los símbolos se indica en la figura 2.6.

Es habitual en la excavación de túneles realizar medidas de los desplazamientos en la sección, con el fin de estudiar su evolución. La expresión [2.93] representa precisamente el modelo que relaciona unos parámetros con unas variables medibles, $\mathbf{x} = \mathbf{M}(\mathbf{p})$.

Se tomaron E y ν como parámetros independientes a identificar, y como datos: $a = 1$ m, $q = 1$ MPa, y $K_o = 0.5$. Por otra parte, se supuso que los puntos de medida, señalados del 1 al 5 en la figura habían proporcionado los siguientes valores:

$$\begin{aligned} u_1 &= 0.18 \cdot 10^{-2} \text{ m} & u_2 &= 0.54 \cdot 10^{-2} \text{ m} & u_3 &= 0.9 \cdot 10^{-2} \text{ m} \\ u_4 &= 1.26 \cdot 10^{-2} \text{ m} & u_5 &= 1.62 \cdot 10^{-2} \text{ m} \end{aligned} \quad [2.94]$$

valores que en realidad corresponden a la expresión [2.93] particularizada para $E = 100$ MPa y $\nu = 0.2$. Se trata, por tanto, de "medidas" teóricas, es decir, sin errores en el proceso de medición (sólo están afectadas por el error de redondeo derivado del propio cálculo).

El problema queda planteado en los siguientes términos: con las medidas [2.94] y el modelo definido por [2.93], hay que identificar los parámetros $E = 100$ MPa y $\nu = 0.2$.

La matriz de sensibilidad $A = \frac{\partial \mathbf{x}}{\partial \mathbf{p}}$ puede calcularse analíticamente, derivando directamente [2.93]. Como se dispone de 5 medidas y se identifican 2 parámetros, A es una matriz 5 x 2 :

$$A = \begin{bmatrix} \frac{\partial u_1}{\partial E} & \frac{\partial u_1}{\partial v} \\ \frac{\partial u_2}{\partial E} & \frac{\partial u_2}{\partial v} \\ \frac{\partial u_3}{\partial E} & \frac{\partial u_3}{\partial v} \\ \frac{\partial u_4}{\partial E} & \frac{\partial u_4}{\partial v} \\ \frac{\partial u_5}{\partial E} & \frac{\partial u_5}{\partial v} \end{bmatrix} = \begin{bmatrix} -\frac{750}{E}(1+v)v & 750(1+2v) \\ -\frac{375}{E}(1+v^2) & 750(1+v) \\ -\frac{500}{E}(1+v) & 500 \\ -\frac{375}{E}(1+v)(3-v) & 750(1-v) \\ -\frac{750}{E}(1+v)(2-v) & 750(1-2v) \end{bmatrix} \quad [2.95]$$

Un ejemplo sintético de este tipo, además de su valor como elemento de aprendizaje, ha servido para comprobar el buen funcionamiento de los programas de cálculo. El hecho de que sólo dependa de dos parámetros permite el dibujo de la función objetivo. En la figura 2.7 se presenta esta función calculada a partir de las diferencias entre los desplazamientos medidos [2.94] y los calculados con [2.93]:

$$J = (\Delta \mathbf{x})^T (\Delta \mathbf{x}) = (\mathbf{x}^* - \mathbf{x})^T (\mathbf{x}^* - \mathbf{x})$$

$$\mathbf{x}^* = \begin{bmatrix} u_1 \\ \cdot \\ \cdot \\ \cdot \\ u_5 \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} u_{r_1}(\theta = 0) \\ u_{r_2}(\theta = \pi/6) \\ u_{r_3}(\theta = \pi/4) \\ u_{r_4}(\theta = \pi/3) \\ u_{r_5}(\theta = \pi/2) \end{bmatrix} \quad \begin{matrix} E = E_i \\ v = v_i \end{matrix} \quad [2.96]$$

Cada punto de la figura corresponde a un valor de la función

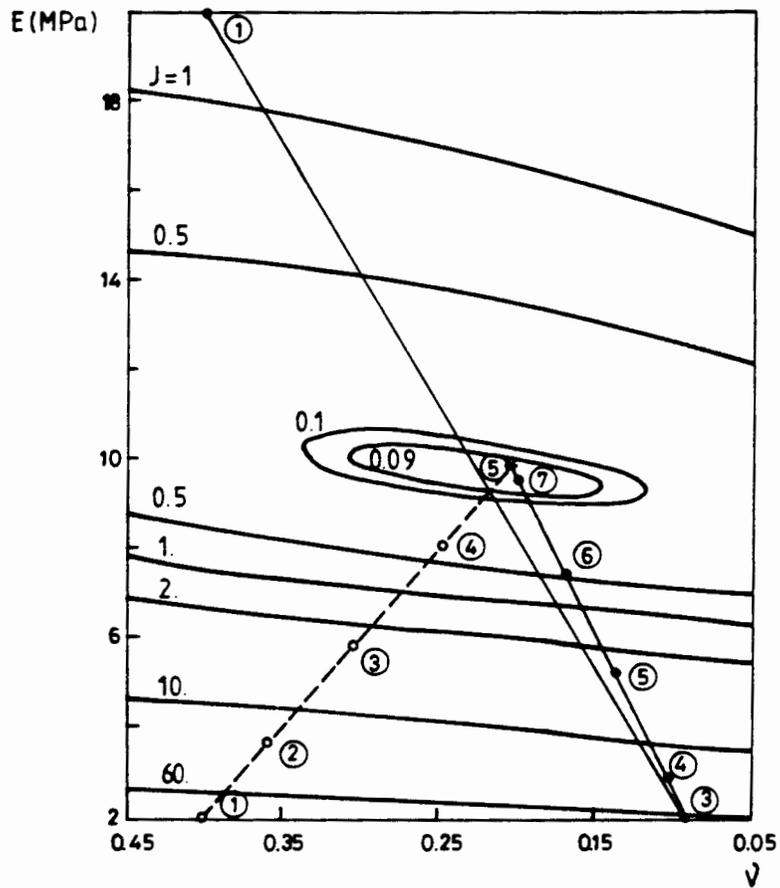


Figura 2.7. Representación de la función objetivo (J, m^2) para el ejemplo teórico de túnel circular en un espacio indefinido, elástico lineal, homogéneo e isótropo. Se indican dos caminos seguidos por el proceso iterativo en la obtención del mínimo.

objetivo, y por tanto, ha requerido el cálculo del problema directo con la ecuación [2.93]. En este caso, la propia forma de las curvas de nivel de dicha función indican que la minimización no va a ser difícil, debido al relativo parecido con un paraboloide. Por otra parte, la excentricidad de las curvas de nivel en la zona del mínimo no es muy alta. Esta excentricidad está relacionada directamente con el número de condición, definido en el apartado 2.3, que representa una medida de la buena o mala convergencia de los procedimientos de minimización anteriormente descritos. Un valor bajo asegura una buena convergencia.

Como en todo proceso iterativo, hay que escoger unos valores iniciales, lo más cercanos posible a la solución final. En un primer tanteo se tomó:

$$E_0 = 200 \text{ MPa} \quad \nu = 0.4$$

valores relativamente apartados de la solución verdadera. El algoritmo de resolución quedaba así:

$$\mathbf{p}_{k+1} = \mathbf{p}_k + \Delta \mathbf{p}_k$$

$$\Delta \mathbf{p}_k = (\mathbf{A}_k^T \mathbf{A}_k)^{-1} \mathbf{A}_k^T \Delta \mathbf{x}_k \quad \Delta \mathbf{x}_k^* = \mathbf{x}^* - \mathbf{M}(\mathbf{p}_k) \quad [2.97]$$

siguiendo el esquema definido por el operador de mínimos cuadrados en la ecuación [2.63].

El criterio de convergencia adoptado consistió en comparar la norma del término $\Delta \mathbf{p}_k$ en cada iteración. Esta debe tender a cero de forma natural. El proceso se detiene o bien cuando la norma aumenta mucho en una iteración con respecto a la anterior, en cuyo caso se produce divergencia, o bien cuando la diferencia entre las normas de dos iteraciones consecutivas es inferior a un cierto

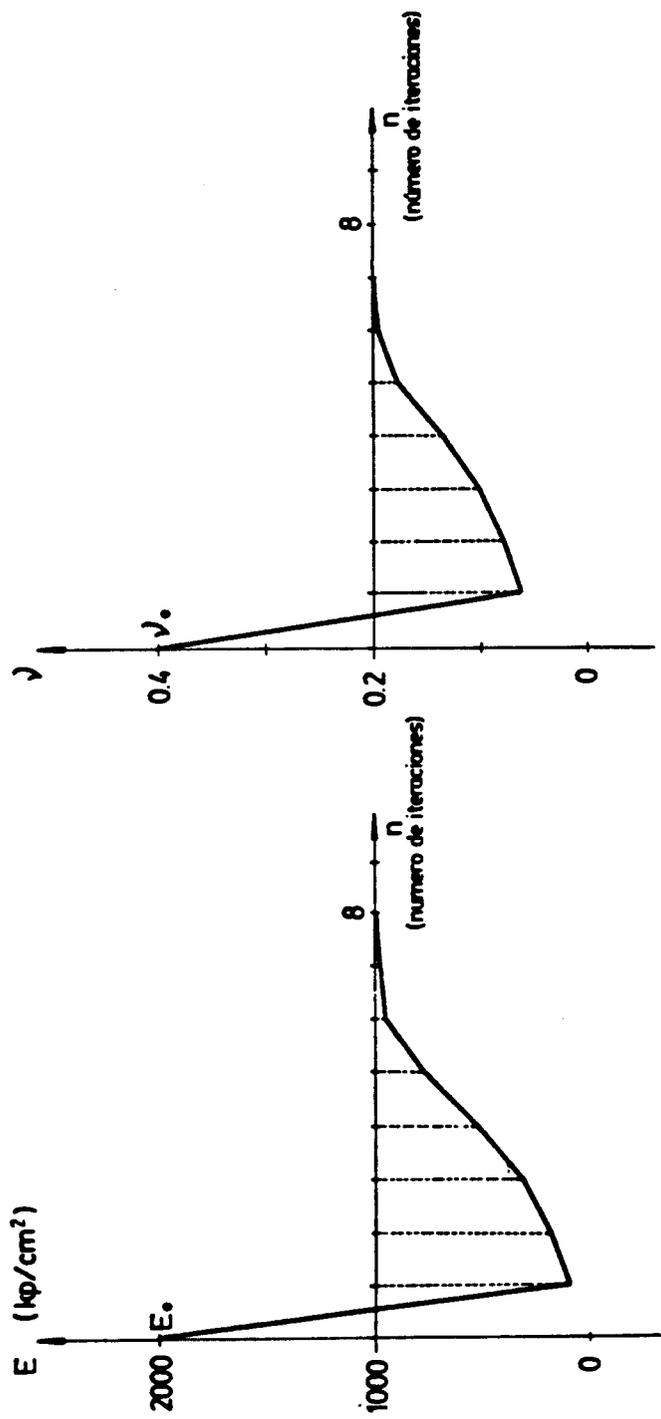


Figura 2.8. Valores de los parámetros en el proceso iterativo.

número prefijado. En la figura 2.8 se indica la evolución de los parámetros con el proceso iterativo. Puede comprobarse que con 8 iteraciones se consigue llegar a los valores identificados previstos. En este caso, como las medidas eran "exactas", el mínimo de la función objetivo corresponde, teóricamente, a cero. Debido a errores de redondeo del ordenador, se llega un mínimo del orden de 10^{-9} .

La interpretación del operador de mínimos cuadrados como matriz inversa generalizada permite definir las matrices de resolución y de densidad de información (apartado 2.4.4). Como la matriz de sensibilidad tiene rango 2 (su rango, si está bien formada, coincide con el número de parámetros), y el sistema es completo, es decir existe $(A^T A)^{-1}$, la matriz de resolución es la identidad. La matriz densidad de información obtenida en la última iteración, correspondiente al mínimo, es:

$$S = A H = \begin{bmatrix} 0.588 & 0.405 & 0.202 & 0.062 & 0.0 \\ 0.405 & 0.306 & 0.198 & 0.094 & 0.0 \\ 0.202 & 0.198 & 0.199 & 0.199 & 0.199 \\ 0.617 & 0.094 & 0.199 & 0.298 & 0.401 \\ 0.0 & 0.0 & 0.199 & 0.401 & 0.608 \end{bmatrix}$$

La estructura de esta matriz indica que las medidas son bastante interdependientes entre sí, ya que se aleja de la matriz identidad. En realidad la matriz S refleja la estructura del modelo que hemos utilizado, interrelacionado con el hecho de que se haya medido en esos puntos determinados.

En la figura 2.7 se ha representado sobre la función objetivo "J" el proceso iterativo seguido, comenzando con dos conjuntos diferentes de parámetros iniciales. Se observa que en pocas iteraciones se llega al mínimo, lo cual es consecuencia, en buena medida, de la sencillez de la función objetivo. Por este motivo no sería necesario, en este caso, utilizar el método de Marquardt en la resolución:

$$\Delta p_k = (A_k^T A_k + \mu_k I)^{-1} A_k^T \Delta x^o$$

Podría ser incluso contraproducente, puesto que se requerirían más iteraciones para resolver el problema, ya que en la medida en que μ_k es grande, se avanza según el gradiente en cada punto, y ya se ha indicado que no es el camino óptimo de avance. Sólo se trata de un camino seguro de descenso, en el caso de que las otras direcciones conduzcan a puntos de mayor valor de J. Por este motivo conviene tomar para el proceso iterativo valores iniciales de μ_k bajos.

Este ejemplo se trata, sin duda, de un modelo muy simple, pero ha permitido ilustrar la metodología a seguir en el proceso de identificación. La mayoría de los problemas que se plantean en Geotecnia requieren una resolución utilizando modelos más complejos. Por otra parte, la versatilidad del método de los elementos finitos aconseja tomarlo como punto de partida en el planteamiento del problema inverso en Geotecnia. Este es el motivo por el cual el siguiente capítulo se dedica al estudio de una formulación acoplada Método de los Elementos Finitos - Identificación de Parámetros.