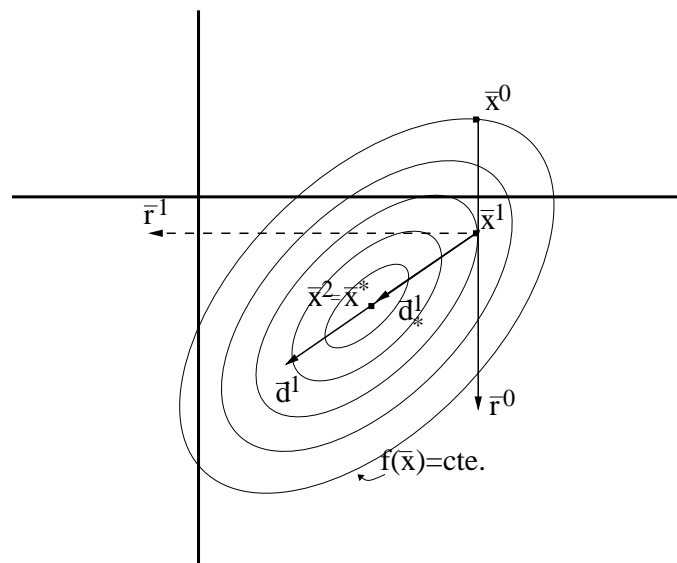


UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE MINAS

DEPARTAMENTO DE MATEMÁTICA APLICADA Y MÉTODOS INFORMÁTICOS



Resolución de sistemas lineales de ecuaciones:

Método del gradiente conjugado

Ultano Kindelán

Índice

1. Introducción	6
2. Notación	7
3. Funciones cuadráticas	7
4. El método del gradiente	11
5. El método del gradiente conjugado	16
5.1. Introducción	16
5.2. Determinación de la dirección de descenso	16
5.3. Propiedades del método del gradiente conjugado	20
5.4. Algoritmo del método del gradiente conjugado	21
A. Demostración de la conjugación de las direcciones de descenso	25
B. Condicionamiento de un sistema	27

Índice de Figuras

1.	<i>Representación geométrica de un sistema lineal de dos ecuaciones. La solución es el punto de intersección de ambas rectas.</i>	8
2.	<i>Gráfica de la forma cuadrática (5). El punto en donde esta función alcanza su valor mínimo es la solución de $A\bar{x} = \bar{b}$.</i>	9
3.	<i>Curvas de nivel de la forma cuadrática (5). A cada elipse le corresponde un valor constante de f</i>	9
4.	<i>Gradiente de la forma cuadrática (5). Para cada \bar{x} el gradiente apunta en la dirección de máximo crecimiento de f y es ortogonal a la curva de nivel que pasa por \bar{x}</i>	10
5.	<i>(a) Forma cuadrática con matriz definida positiva; (b) con matriz definida negativa; (c) con matriz singular (y positiva), en este caso una línea atraviesa el fondo del valle (no hay un único mínimo); (d) con una matriz indefinida; como en este caso el punto crítico es un punto de silla, los métodos de descenso no funcionarán correctamente. En dimensión superior a dos, las matrices singulares también pueden tener puntos de silla.</i>	11
6.	<i>El método del gradiente: (a) Se comienza en el punto $(3, -1)^t$ y se realiza la primera iteración siguiendo la dirección de máximo descenso. (b) Hay que encontrar el punto de \mathbb{R}^2 en el que la curva (parábola) intersección de las dos superficies alcanza su mínimo absoluto. (c) La parábola intersección de las dos superficies. (d) El gradiente en el punto donde se alcanza el mínimo es ortogonal al gradiente en el punto obtenido en la iteración anterior.</i>	13
7.	<i>Dibujo de las dos primeras iteraciones del método del gradiente sobre las curvas de nivel de la función $f(\bar{x}) = \frac{1}{2}\bar{x}^t A\bar{x} - \bar{b}^t \bar{x}$.</i>	17
8.	<i>Dibujo de las dos primeras iteraciones del método del gradiente conjugado sobre las curvas de nivel de la función $f(\bar{x}) = \frac{1}{2}\bar{x}^t A\bar{x} - \bar{b}^t \bar{x}$. En dos iteraciones se alcanza la solución exacta del sistema $A\bar{x} = \bar{b}$.</i>	18

Índice de Algoritmos

1.	Algoritmo que resuelve el sistema $A\bar{x} = \bar{b}$ mediante el método del gradiente. Versión preliminar.	14
2.	Algoritmo que resuelve el sistema $A\bar{x} = \bar{b}$ mediante el método del gradiente. Versión más eficiente con “memoria”.	15
3.	Algoritmo que resuelve el sistema $A\bar{x} = \bar{b}$ mediante el método del gradiente. Versión más eficiente sin “memoria”.	16
4.	Estructura general de los algoritmos de los métodos de descenso.	16
5.	Algoritmo que resuelve el sistema $A\bar{x} = \bar{b}$ mediante el método del gradiente gradiente conjugado. Versión provisional.	22
6.	Algoritmo que resuelve el sistema $A\bar{x} = \bar{b}$ mediante el método del gradiente gradiente conjugado. Versión final.	23

1. Introducción

El método del gradiente conjugado es el método iterativo más utilizado para resolver grandes sistemas lineales de ecuaciones*. El método del gradiente conjugado es efectivo para sistemas de la forma

$$A\bar{x} = \bar{b}, \quad (1)$$

en donde \bar{x} es un vector desconocido, \bar{b} es un vector conocido y A es una matriz cuadrada simétrica y definida positiva[†] también conocida. Estos sistemas de ecuaciones aparecen en la aplicación de técnicas numéricas de gran importancia como por ejemplo los métodos de diferencias finitas o elementos finitos para resolver ecuaciones diferenciales en derivadas parciales. También surgen en la resolución de problemas de ingeniería importantes como el análisis estructural o la teoría de circuitos.

El método del gradiente conjugado es un caso particular de método de descenso. Los métodos de descenso están especialmente indicados para la resolución de sistemas huecos (sistemas lineales cuya matriz de coeficientes tiene un número importante de ceros). Si A es una matriz llena suele ser más aconsejable factorizar la matriz A y después resolver el sistema mediante la técnica de remonte (método directo). El tiempo consumido factorizando la matriz A es más o menos el mismo que el necesario para resolver el sistema por un método iterativo[‡]; y una vez que la matriz de coeficientes está factorizada, el sistema se puede resolver rápidamente para múltiples valores de \bar{b} . Por el contrario si la matriz A es una matriz hueca, en su factorización triangular aparecerán muchos elementos nulos, con el consiguiente desperdicio de operaciones y memoria. Los métodos de descenso, y en general los métodos iterativos, consiguen ahorrar memoria y operaciones operando únicamente sobre los elementos no nulos.

El orden que se ha seguido para la elaboración de estos apuntes es el siguiente: en el apartado 2 se hace un breve repaso de la notación que se utilizará en el resto de las secciones. En el apartado 3 se repasan algunas ideas básicas de optimización de funciones de varias variables, con el objetivo de mostrar la equivalencia entre resolver el sistema $A\bar{x} = \bar{b}$ y hallar el mínimo de una función cuadrática cuando A es una matriz simétrica y definida positiva. En el apartado 4 se explica el método del gradiente. El interés de estudiar el método del gradiente radica en su utilidad pedagógica para introducir el método del gradiente conjugado y no en su eficacia para resolver sistemas lineales puesto que no puede competir en velocidad de convergencia con este último. Finalmente en el apartado 5 se explica el método del gradiente conjugado demostrando las propiedades más importantes del mismo, exceptuando la demostración de la conjugación de las direcciones de descenso que por su longitud se incluye separadamente en un apéndice.

Estos apuntes están escritos con la idea de utilizarlos en la asignatura de Programación y Métodos Numéricos del segundo curso de la titulación de Ingeniero de Minas de la Universidad Politécnica de Madrid. Presuponen, por tanto, unos conocimientos básicos de Álgebra lineal y Análisis en varias variables.

*Cuando se dice que el método del gradiente conjugado es el método más utilizado para resolver grandes sistemas, se entiende que nos estamos refiriendo a él directamente o a algunas de sus variantes como, por ejemplo, el método GMRES.

[†]Existen variantes del método para resolver sistemas cuya matriz no sea simétrica ni definida positiva.

[‡]En algunos casos, consultar [6], dependiendo de cómo sea la matriz del sistema, el método del gradiente conjugado y sobre todo alguna de sus variantes, pueden ser más rápidos que los métodos directos incluso cuando la matriz del sistema es una matriz llena.

2. Notación

En este capítulo se utilizarán letras latinas mayúsculas para designar matrices. Los vectores se denotarán con una raya encima y generalmente se utilizarán letras latinas minúsculas. Los escalares (números reales) se representarán mediante letras griegas minúsculas, salvo en el caso de las componentes de un vector (letra del vector con un subíndice) o los coeficientes de una matriz (letra de la matriz en minúscula con dos subíndices que indican su posición en la matriz). De este modo la ecuación (1) desarrollada sería:

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

En ocasiones se utilizarán superíndices para diferenciar unos vectores de otros:

$$(\bar{x}^1)^t = (x_1^1, \dots, x_n^1), \quad (\bar{x}^2)^t = (x_1^2, \dots, x_n^2) \quad \dots \quad (\bar{x}^p)^t = (x_1^p, \dots, x_n^p).$$

El producto escalar de dos vectores de \mathbb{R}^n se representará indistintamente como $\bar{x}^t \bar{y}$ o $\bar{x} \cdot \bar{y}$. Por lo tanto:

$$\bar{x}^t \bar{y} = \bar{y}^t \bar{x} = \bar{x} \cdot \bar{y} = \bar{y} \cdot \bar{x} = \sum_{i=1}^n x_i y_i.$$

Por último se recuerda que una matriz de dimensión n que verifique

$$\bar{x}^t A \bar{x} > 0, \quad \forall \bar{x} \in \mathbb{R}^n (\bar{x} \neq \bar{0}). \quad (2)$$

se dice que es una matriz **definida positiva**. Si además es simétrica se dirá que es una matriz **simétrica definida positiva**.

3. Funciones cuadráticas

En lo sucesivo se denominará **función cuadrática** a toda función escalar definida sobre un espacio vectorial de dimensión n de la forma:

$$f(\bar{x}) = \frac{1}{2} \bar{x}^t A \bar{x} - \bar{b}^t \bar{x} + c, \quad (3)$$

en donde A es una matriz cuadrada, \bar{x} y \bar{b} son vectores y c es una constante escalar. Se demostrará posteriormente que si A es una matriz simétrica definida positiva, f alcanza un mínimo en la solución de $A\bar{x} = \bar{b}$.

A lo largo de estos apuntes se ilustrarán algunos de los conceptos que vayan surgiendo con el siguiente ejemplo:

$$A = \begin{pmatrix} 3 & -2 \\ -2 & 4 \end{pmatrix}, \quad \bar{b} = \begin{pmatrix} 4 \\ 8 \end{pmatrix}, \quad c = 0. \quad (4)$$

La representación geométrica del sistema $A\bar{x} = \bar{b}$ se muestra en la figura 1. En general la solución \bar{x} se encuentra en el punto de intersección de n hiperplanos, cada uno de ellos de

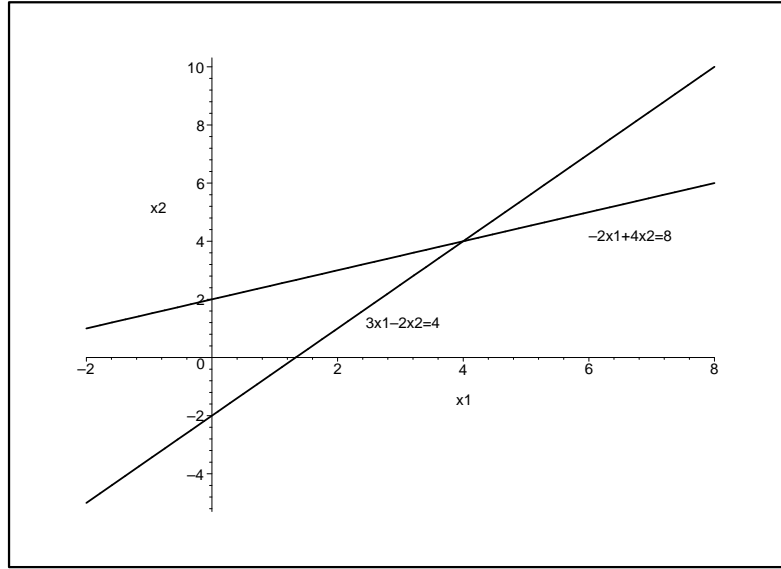


Figura 1: Representación geométrica de un sistema lineal de dos ecuaciones. La solución es el punto de intersección de ambas rectas.

dimensión $n - 1$. Para este problema, la solución es $\bar{x} = (4, 4)^t$. La correspondiente forma cuadrática es

$$f(\bar{x}) = \frac{3}{2}(x_1)^2 + 2(x_2)^2 - 2x_1x_2 - 4x_1 - 8x_2 \quad (5)$$

y aparece representada en la figura 2. Un dibujo con las curvas de nivel de f aparece en la figura 3. Debido a que la matriz A es definida positiva, la superficie definida por f en \mathbb{R}^2 tiene la forma de un paraboloide.

El gradiente de una función de n variables se define como:

$$\bar{f}'(\bar{x}) = \nabla f(\bar{x}) = \begin{pmatrix} \frac{\partial}{\partial x_1} f(\bar{x}) \\ \frac{\partial}{\partial x_2} f(\bar{x}) \\ \vdots \\ \frac{\partial}{\partial x_n} f(\bar{x}) \end{pmatrix}. \quad (6)$$

El gradiente es un campo vectorial que tiene la importante propiedad de que en un punto dado, \bar{x} , el vector de campo correspondiente apunta en la dirección del máximo crecimiento de f (obsérvese la figura 4). Derivando parcialmente en (3) se obtiene:

$$\nabla f(\bar{x}) = \frac{1}{2}A^t\bar{x} + \frac{1}{2}A\bar{x} - \bar{b}. \quad (7)$$

Si A es simétrica, la ecuación (7) se simplifica:

$$\nabla f(\bar{x}) = A\bar{x} - \bar{b}. \quad (8)$$

Igualando el gradiente a cero se recupera el sistema (1). De este modo la solución del sistema $A\bar{x} = \bar{b}$, que en lo sucesivo se denominará \bar{x}^* , es un punto crítico de $f(\bar{x})$. Si la matriz A , además de ser simétrica, es definida positiva, entonces **la solución del sistema $A\bar{x} = \bar{b}$ es un mínimo de f** ; esto se puede demostrar fácilmente estudiando el valor de f en un punto distinto de \bar{x}^* , p.e. $\bar{x}^* + \bar{e}$: si se supone que A es simétrica (definida positiva

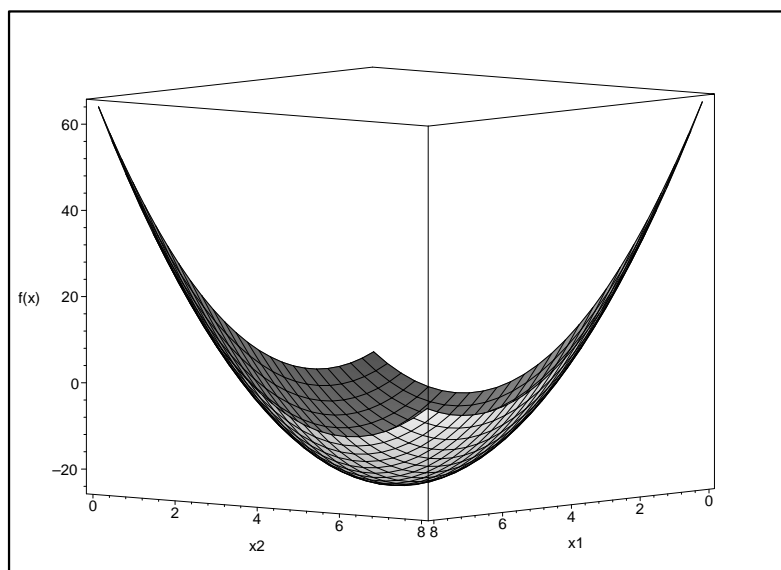


Figura 2: Gráfica de la forma cuadrática (5). El punto en donde esta función alcanza su valor mínimo es la solución de $A\bar{x} = \bar{b}$.

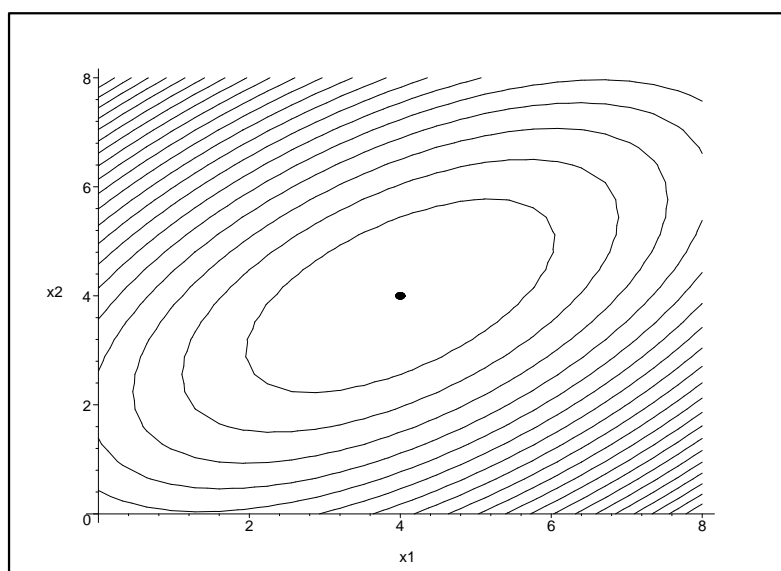


Figura 3: Curvas de nivel de la forma cuadrática (5). A cada elipse le corresponde un valor constante de f

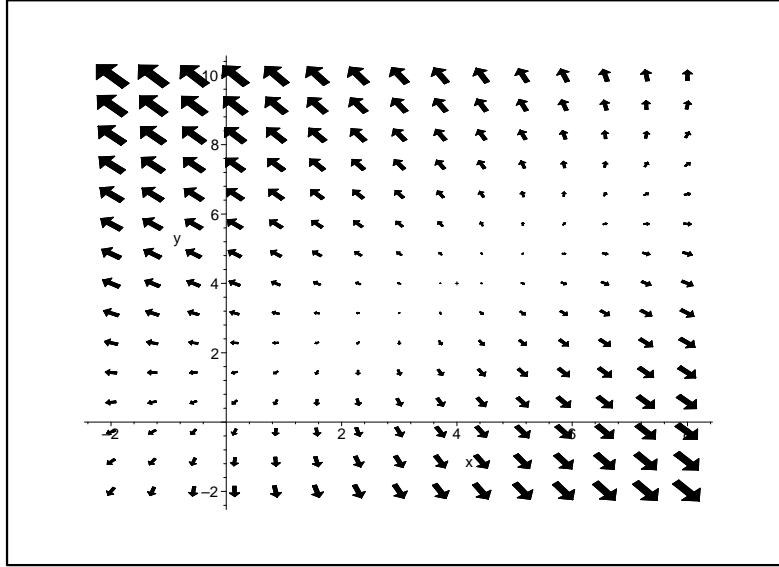


Figura 4: *Gradiente de la forma cuadrática (5). Para cada \bar{x} el gradiente apunta en la dirección de máximo crecimiento de f y es ortogonal a la curva de nivel que pasa por \bar{x}*

o no) se verifica, partiendo de (3), que $f(\bar{x}^* + \bar{e}) = f(\bar{x}^*) + (1/2)\bar{e}^t A \bar{e}$. Si además A es definida positiva se deduce que el segundo sumando del segundo miembro de la ecuación anterior es siempre positivo, alcanzándose, por tanto, un mínimo global de f [§] en \bar{x}^* .

Todo lo anterior implica que **la solución de $A\bar{x} = \bar{b}$ se puede calcular buscando el punto \bar{x}^* que minimice $f(\bar{x})$.**

El hecho de que en \mathbb{R}^2 f sea un paraboloide permite interpretar geoméricamente el que A sea una matriz definida positiva: la función f alcanza un mínimo absoluto y a partir del punto en el que alcanza dicho mínimo la función crece en todas las direcciones. Si A no es definida positiva, se pueden dar los siguientes casos:

- A puede ser definida negativa, en este caso la función f sería un paraboloide abierto hacia abajo, alcanzaría un máximo absoluto y a partir del punto en el que se alcanzase el máximo la función decrecería en todas las direcciones.
- A puede ser una matriz singular en cuyo caso la solución de (1), de existir, no sería única: existirían infinitos puntos críticos formando una recta o un plano dependiendo del rango de A .
- En los casos en los que A no es definida ni singular, será una matriz indefinida. En estos casos el punto \bar{x}^* será un “punto silla” y tanto el método del gradiente como el método del gradiente conjugado fallarán.

La figura 5 muestra los casos que se acaban de mencionar. Los valores de \bar{b} y c determinan el lugar en que se encuentra el mínimo de f pero no afectan a la forma de la superficie definida por f .

[§]También se puede demostrar comprobando que la matriz Hessiana de f es constante e igual a A ; para ello basta con derivar parcialmente las n componentes del gradiente respecto a las n variables x_1, \dots, x_n .

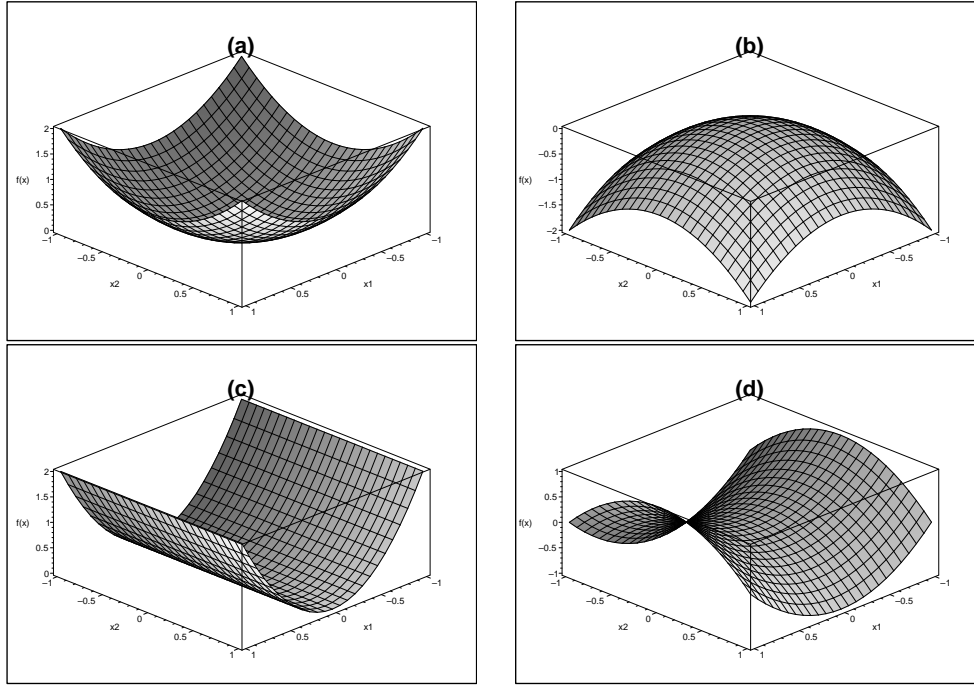


Figura 5: (a) Forma cuadrática con matriz definida positiva; (b) con matriz definida negativa; (c) con matriz singular (y positiva), en este caso una línea atraviesa el fondo del valle (no hay un único mínimo); (d) con una matriz indefinida; como en este caso el punto crítico es un punto de silla, los métodos de descenso no funcionarán correctamente. En dimensión superior a dos, las matrices singulares también pueden tener puntos de silla.

4. El método del gradiente

El método del gradiente es un método de descenso en el que se comienza a iterar en un punto arbitrario \bar{x}^0 y se continúa siguiendo la línea de máximo descenso del paraboloide obteniéndose una sucesión de puntos $\bar{x}^1, \bar{x}^2, \dots$ hasta que se obtiene un punto lo suficientemente cercano a la solución \bar{x}^* .

En cada iteración se elige la dirección para la que f decrece más rápidamente, que es la dirección contraria a $\nabla f(\bar{x}^i)$. De acuerdo con la ecuación (8), esta dirección es $-\nabla f(\bar{x}^i) = \bar{b} - A\bar{x}^i$.

A continuación se van a dar dos definiciones que serán de utilidad en el resto de los apartados de estos apuntes. El **error** $\bar{e}^i = \bar{x}^i - \bar{x}^*$ es un vector que indica lo alejado que se está de la solución. El **residuo** $\bar{r}^i = \bar{b} - A\bar{x}^i$ es otro vector que indica cuan lejos está el vector $A\bar{x}$ del vector \bar{b} . Es fácil comprobar que $\bar{r}^i = -A\bar{e}^i$, por lo que se debe pensar en el residuo como el error transformado por $-A$. Más importante, $\bar{r}^i = -\nabla f(\bar{x}^i)$, y por lo tanto también se debe pensar en el residuo como la dirección de máximo descenso.

Supóngase que en el ejemplo dado en (4) se comienza a iterar en $\bar{x}^0 = (3, -1)^t$. El primer paso, a lo largo de la dirección de máximo descenso, llegará hasta algún punto de la línea continua de la figura 6(a). Dicho de otra manera, se escogerá el punto

$$\bar{x}^1 = \bar{x}^0 + \rho_0 \bar{r}^0. \quad (9)$$

La pregunta que hay que hacer para calcular ρ_0 es: ¿de qué tamaño se dará el paso en la iteración 1? Para contestar a esta pregunta, supondremos que en cada iteración se realizará una búsqueda lineal a lo largo de la dirección de descenso correspondiente (en el caso del método del gradiente en la dirección contraria al gradiente):

$$f(\bar{x}^0 + \rho_0 \bar{r}^0) = \min_{\rho \in \mathbb{R}} f(\bar{x}^0 + \rho \bar{r}^0). \quad (10)$$

Una búsqueda lineal es un procedimiento que busca un ρ_0 que minimice ρ a lo largo de una línea. En la figura 6 se muestra el procedimiento: nos restringimos a buscar un punto en la curva (la parábola que se muestra en la figura 6(c)) intersección del plano vertical con el paraboloide, por lo que se trata de resolver un problema de optimización unidimensional: hallar ρ tal que

$$\frac{d}{d\rho} f(\bar{x}^0 + \rho \bar{r}^0) = 0. \quad (11)$$

Al derivar se obtiene:

$$\begin{aligned} \frac{d}{d\rho} f(\bar{x}^0 + \rho \bar{r}^0) &= \frac{d}{d\rho} \left(\frac{1}{2} (\bar{x}^0 + \rho \bar{r}^0)^t A (\bar{x}^0 + \rho \bar{r}^0) - \bar{b}^t (\bar{x}^0 + \rho \bar{r}^0) + c \right) = \\ &\rho (\bar{r}^0)^t A \bar{r}^0 + (\bar{r}^0)^t A \bar{x}^0 - \bar{b}^t \bar{r}^0 = 0. \end{aligned} \quad (12)$$

De donde se deduce:

$$\rho_0 = \frac{(\bar{b} - A\bar{x}^0)^t \bar{r}^0}{(\bar{r}^0)^t A \bar{r}^0} = \frac{(\bar{r}^0)^t \bar{r}^0}{(\bar{r}^0)^t A \bar{r}^0}. \quad (13)$$

Una vez conocidos la dirección de descenso (\bar{r}^0) y ρ_0 , se calcula \bar{x}^1 mediante (9). Después de obtener \bar{x}^1 se calcula \bar{x}^2 de idéntica manera:

$$\bar{x}^2 = \bar{x}^1 + \rho_1 \bar{r}^1,$$

en donde ρ_1 se calcula igual que ρ_0 . El proceso seguiría de forma iterativa calculando los sucesivos valores aproximados de la solución de $A\bar{x} = \bar{b}$:

$$\begin{aligned} \bar{x}^1 &= \bar{x}^0 + \rho_0 \bar{r}^0 \\ \bar{x}^2 &= \bar{x}^1 + \rho_1 \bar{r}^1 \\ \bar{x}^3 &= \bar{x}^2 + \rho_2 \bar{r}^2 \\ &\vdots \\ \bar{x}^i &= \bar{x}^{i-1} + \rho_{i-1} \bar{r}^{i-1} \\ \bar{x}^{i+1} &= \bar{x}^i + \rho_i \bar{r}^i, \\ &\vdots \end{aligned}$$

o, en forma abreviada:

$$\bar{x}^{i+1} = \bar{x}^i + \rho_i \bar{r}^i, \quad i = 0, \dots \quad (14)$$

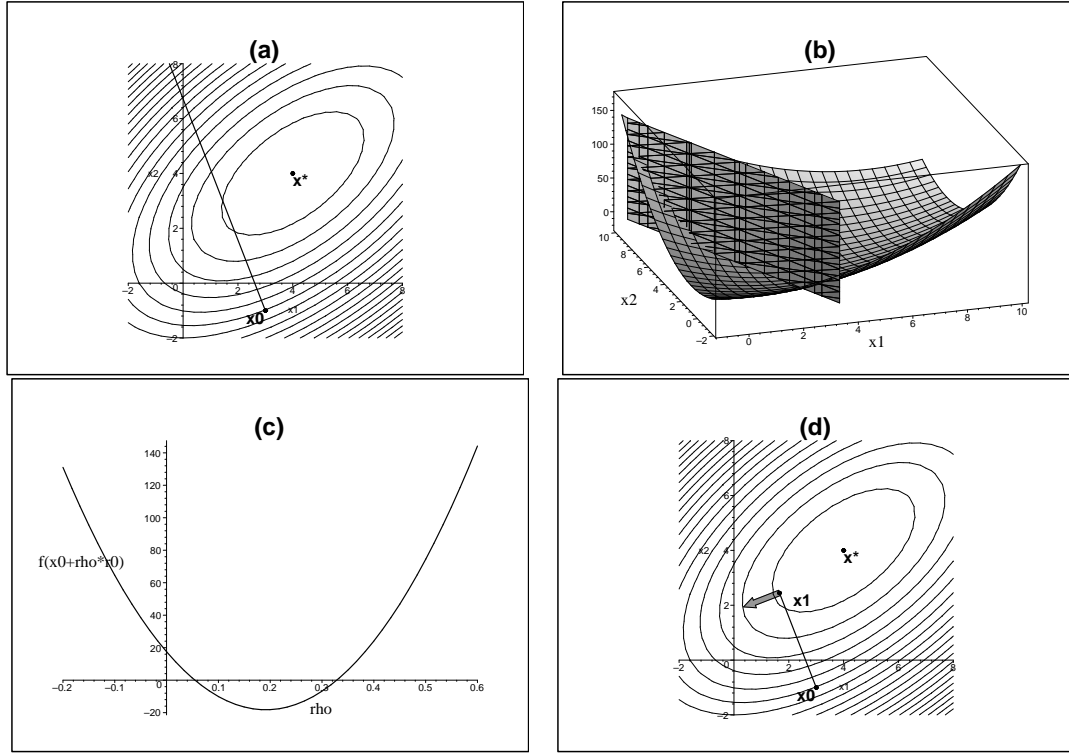


Figura 6: El método del gradiente: (a) Se comienza en el punto $(3, -1)^t$ y se realiza la primera iteración siguiendo la dirección de máximo descenso. (b) Hay que encontrar el punto de \mathbb{R}^2 en el que la curva (parábola) intersección de las dos superficies alcanza su mínimo absoluto. (c) La parábola intersección de las dos superficies. (d) El gradiente en el punto donde se alcanza el mínimo es ortogonal al gradiente en el punto obtenido en la iteración anterior.

con:

$$\begin{aligned}\bar{x}^0 &= \text{vector inicial arbitrario} \\ \bar{r}^i &= \bar{b} - A\bar{x}^i \\ \rho_i &= \frac{(\bar{r}^i)^t \bar{r}^i}{(\bar{r}^i)^t A \bar{r}^i}\end{aligned}$$

Para definir de forma completa el esquema iterativo anterior hace falta determinar cuándo se para de iterar (hasta que valor de i se llega). Para ello hay que estimar el error cometido en cada iteración. Se citan a continuación dos formas de estimar el error en la iteración i :

1. La norma del residuo, $\|\bar{r}^i\|$. Como se ha explicado más arriba $\bar{r}^i = -A\bar{e}^i$, lo cual implica que si $\bar{r}^i = 0$ entonces $\bar{e}^i = 0$ (siempre y cuando A sea una matriz no singular). En consecuencia cuanto más pequeño sea $\|\bar{r}^i\|$ más cerca se estará de la solución.
2. La norma del vector diferencia entre los resultados en la iteración $i + 1$ y la iteración i , $\|\bar{x}^{i+1} - \bar{x}^i\|$. Si la norma de este vector diferencia se hace cada vez más pequeña entonces la sucesión de vectores $\{\bar{x}^i\}$ tiene un límite. En el caso de que se demuestre que el esquema iterativo converge hacia la solución del sistema (lo cual sucede en el

método del gradiente si la matriz A es definida positiva), entonces el límite anterior coincide con la solución del sistema. Por tanto, al igual que en el caso anterior, cuanto más pequeño sea $\|\bar{x}^{i+1} - \bar{x}^i\|$ más cerca se estará de la solución.

A estas dos formas de estimar el error les denominaremos *criterios de parada*. En la práctica, al aplicar estos dos criterios de parada, no se exige llegar a un valor de i que haga cero alguna de las dos normas anteriores (valor de i con el que se obtendría la solución exacta) pues, en la mayoría de los casos, es imposible en un número finito de operaciones. Lo que se hace es exigir que se llegue hasta un valor de i que asegure que la norma correspondiente esté por debajo de un cierto valor impuesto por el usuario del algoritmo y que en lo sucesivo se designará por ϵ .

A partir del esquema iterativo anterior se puede construir un algoritmo para resolver un sistema lineal de ecuaciones por el método del gradiente:

Algoritmo 1 Algoritmo que resuelve el sistema $A\bar{x} = \bar{b}$ mediante el método del gradiente. Versión preliminar.

```

Leer  $A, \bar{b}, \bar{x}^0, maxiter, \epsilon$ 
 $i \leftarrow 0, tol \leftarrow 2\epsilon$ 
mientras  $i < maxiter$  y  $tol > \epsilon$  hacer
     $\bar{r}^i \leftarrow \bar{b} - A\bar{x}^i$ 
     $\rho_i \leftarrow \frac{(\bar{r}^i)^t \bar{r}^i}{(\bar{r}^i)^t A \bar{r}^i}$ 
     $\bar{x}^{i+1} \leftarrow \bar{x}^i + \rho_i \bar{r}^i$ 
     $tol \leftarrow \|\bar{r}^i\|$ 
     $i \leftarrow i + 1$ 
fin mientras

```

Observación 4.1. En el algoritmo anterior y en todos los restantes se va a utilizar como estimador del error la norma del residuo. Tiene la ventaja respecto a $\|\bar{x}^{i+1} - \bar{x}^i\|$ que no hay que efectuar ninguna resta puesto que \bar{r}^i ya está calculado y además no obliga a almacenar más vectores de la sucesión que el último calculado. El valor de $\|\bar{r}^i\|$ se almacena en tol , obsérvese que inicialmente se le asigna a tol un valor artificialmente grande (mayor que ϵ) para que el algoritmo realice al menos una iteración.

Observación 4.2. El dato $maxiter$ representa el número máximo de iteraciones que el algoritmo puede realizar independientemente de que haya alcanzado la solución o no.

Si analizamos el coste en operaciones del algoritmo 1, llegamos a lo siguiente:

Operación vectorial	Nº de operaciones elementales	Operación más costosa
$\bar{r}^i \leftarrow \bar{b} - A\bar{x}^i$	$2n^2$	$A\bar{x}^i$
$\rho_i \leftarrow \frac{(\bar{r}^i)^t \bar{r}^i}{(\bar{r}^i)^t A \bar{r}^i}$	$2n^2 + 3n - 1$	$A\bar{r}^i$
$\bar{x}^{i+1} \leftarrow \bar{x}^i + \rho_i \bar{r}^i$	$2n$	$\rho_i \bar{r}^i$
$tol \leftarrow \ \bar{r}^i\ $	$2n$	$\ \bar{r}^i\ $

En la tabla anterior se ha especificado el número de operaciones elementales* que hay que efectuar para realizar cada una de las cuatro operaciones vectoriales que componen

*Se entiende por operación elemental una suma(resta) o una multiplicación (división) entre escalares.

el algoritmo del gradiente. Observando la tabla se comprueba que las dos operaciones que determinan el “costo” de cada una de las iteraciones del método del gradiente son $A\bar{x}^i$ y $A\bar{r}^i$. La primera se puede evitar si tenemos en cuenta que:

$$\bar{r}^i = \bar{b} - A\bar{x}^i = \bar{b} - A(\bar{x}^{i-1} + \rho_{i-1}\bar{r}^{i-1}) = \bar{b} - A\bar{x}^{i-1} - \rho_{i-1}A\bar{r}^{i-1} = \bar{r}^{i-1} - \rho_{i-1}A\bar{r}^{i-1}. \quad (15)$$

Por lo tanto, salvo en la iteración inicial, no es necesario calcular el producto $A\bar{x}^i$; basta con calcular $A\bar{r}^i$ y utilizarlo en la siguiente iteración.

Teniendo en cuenta lo anterior se puede introducir una ligera modificación en el algoritmo 1 para conseguir un algoritmo del gradiente más eficiente:

Algoritmo 2 Algoritmo que resuelve el sistema $A\bar{x} = \bar{b}$ mediante el método del gradiente. Versión más eficiente con “memoria”.

```

Leer  $A, \bar{b}, \bar{x}^0, maxiter, \epsilon$ 
 $i \leftarrow 0, \bar{r}^0 \leftarrow \bar{b} - A\bar{x}^0, tol \leftarrow \|\bar{r}^0\|$ 
mientras  $i < maxiter$  y  $tol > \epsilon$  hacer
     $\bar{b} \leftarrow A\bar{r}^i$ 
     $\rho_i \leftarrow \frac{(\bar{r}^i)^t \bar{r}^i}{(\bar{r}^i)^t \bar{b}}$ 
     $\bar{x}^{i+1} \leftarrow \bar{x}^i + \rho_i \bar{r}^i$ 
     $\bar{r}^{i+1} \leftarrow \bar{r}^i - \rho_i \bar{b}$ 
     $tol \leftarrow \|\bar{r}^{i+1}\|$ 
     $i \leftarrow i + 1$ 
fin mientras

```

Se ha conseguido que en cada iteración se realicen $2n^2 + 9n - 1$ operaciones elementales en lugar de las $4n^2 + 7n - 1$ que necesitaba el algoritmo preliminar. Por lo tanto el método del gradiente realiza aproximadamente, si n es grande, $2n^2$ operaciones elementales en cada iteración y será competitivo con los métodos directos si se necesitan menos de $n/6^\dagger$ iteraciones para alcanzar la solución. Al final del siguiente apartado se analizará cuántas iteraciones son necesarias para alcanzar la convergencia (velocidad de convergencia).

En el algoritmo anterior se han almacenado todos los vectores de la sucesión. En la práctica es suficiente con almacenar el último (algoritmo 3); que coincidirá con la solución del sistema buscada.

Salvo que por algún motivo se quiera tener almacenadas las evoluciones del residuo y de la sucesión de vectores $\{\bar{x}^i\}$, es siempre el algoritmo 3, en la práctica, el que se implementa, debido al ahorro de memoria que supone respecto al anterior.

[†]El método de Cholesky necesita $\frac{n^3}{3}$ iteraciones en cada iteración.

Algoritmo 3 Algoritmo que resuelve el sistema $A\bar{x} = \bar{b}$ mediante el método del gradiente. Versión más eficiente sin “memoria”.

```

Leer  $A, \bar{b}, \bar{x}, maxiter, \epsilon$ 
 $i \leftarrow 0, \bar{r} \leftarrow \bar{b} - A\bar{x}, tol \leftarrow \|\bar{r}\|$ 
mientras  $i < maxiter$  y  $tol > \epsilon$  hacer
     $\bar{b} \leftarrow A\bar{r}$ 
     $\rho \leftarrow \frac{(\bar{r})^t \bar{r}}{(\bar{r})^t \bar{b}}$ 
     $\bar{x} \leftarrow \bar{x} + \rho \bar{r}$ 
     $\bar{r} \leftarrow \bar{r} - \rho \bar{b}$ 
     $tol \leftarrow \|\bar{r}\|$ 
     $i \leftarrow i + 1$ 
fin mientras

```

5. El método del gradiente conjugado

5.1. Introducción

En general los métodos de descenso responden al siguiente algoritmo (generalización del método del gradiente):

Algoritmo 4 Estructura general de los algoritmos de los métodos de descenso.

```

Leer  $A, \bar{b}, \bar{x}^0, maxiter, \epsilon$ 
 $i \leftarrow 0, tol \leftarrow 2\epsilon$ 
mientras  $i < maxiter$  y  $tol > \epsilon$  hacer
    Elegir la dirección de descenso a seguir  $\rightarrow \bar{d}^i$ 
    Determinar el parámetro de descenso  $\rightarrow \rho_i$ 
     $\bar{x}^{i+1} \leftarrow \bar{x}^i + \rho_i \bar{d}^i$ 
     $tol \leftarrow \sim \|\bar{e}^{i+1}\|$ 
     $i \leftarrow i + 1$ 
fin mientras

```

Ya se ha explicado cómo se calcula el parámetro de descenso (ver la ecuación (13)). Se debe tener en cuenta que la dirección de descenso ya no es necesariamente igual a \bar{r}^i , por lo que en el caso general:

$$\rho_i = \frac{(\bar{r}^i)^t \bar{d}^i}{(\bar{d}^i)^t A \bar{d}^i}. \quad (16)$$

Para estimar el error se seguirá utilizando la norma del residuo: $tol \leftarrow \|\bar{r}^i\|$. Por consiguiente para definir un nuevo método de descenso será necesario determinar cuál es la dirección de descenso (\bar{d}^i) en cada iteración.

5.2. Determinación de la dirección de descenso

Supóngase que se utiliza el método del gradiente para resolver un sistema de dos ecuaciones con dos incógnitas: $A\bar{x} = \bar{b}$ ($A \in M_2(\mathbb{R})$, simétrica y definida positiva). La representación

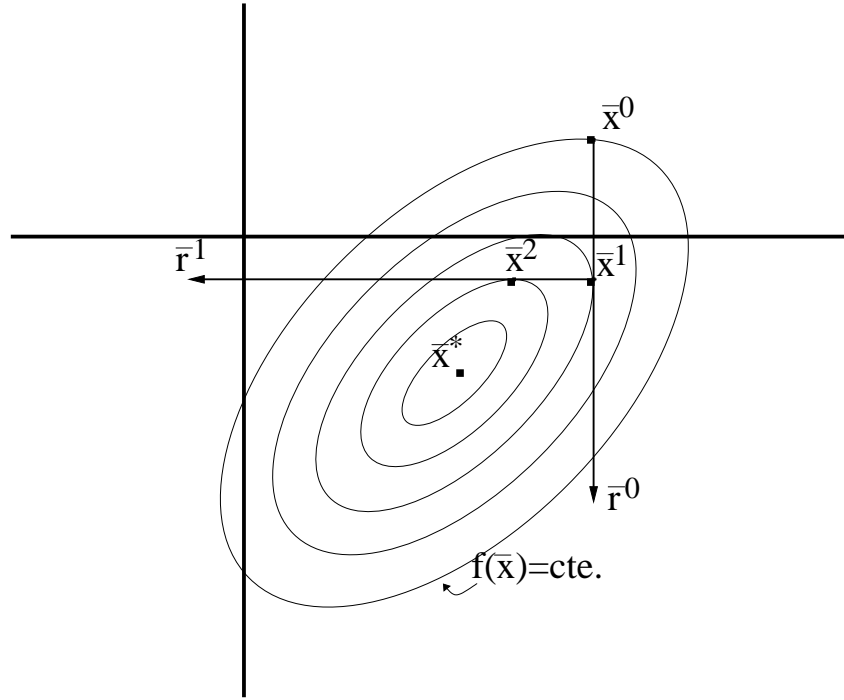


Figura 7: Dibujo de las dos primeras iteraciones del método del gradiente sobre las curvas de nivel de la función $f(\bar{x}) = \frac{1}{2}\bar{x}^t A \bar{x} - \bar{b}^t \bar{x}$.

geométrica de las dos primeras iteraciones del método del gradiente aparece en la figura 7.

Nos podríamos plantear modificar el método del gradiente eligiendo una dirección de descenso distinta al gradiente en la segunda iteración: la dirección que une \bar{x}^1 con la solución exacta \bar{x}^* . En este caso tendríamos un método que obtiene (si se trabaja con aritmética exacta) la solución exacta en dos iteraciones (ver la figura 8). El problema es cómo determinar esta nueva dirección sin conocer la solución exacta del sistema. Para resolver el problema llamaremos \bar{d}_*^1 al vector que une \bar{x}^1 con \bar{x}^* :

$$\bar{d}_*^1 = \bar{x}^* - \bar{x}^1 = \bar{x}^* - \bar{x}^0 - \rho_0 \bar{d}^0.$$

En donde se ha denominado \bar{d}^0 a la primera dirección de descenso, $\bar{d}^0 = \bar{r}^0$.

Proposición 5.1. Los vectores \bar{d}^0 y \bar{d}_*^1 son conjugados respecto a la matriz A .

Demostración. Basta con comprobar que $(\bar{d}^0)^t A \bar{d}_*^1 = 0$:

$$\begin{aligned} (\bar{d}^0)^t A \bar{d}_*^1 &= (\bar{d}^0)^t A (\bar{x}^* - \bar{x}^0 - \rho_0 \bar{d}^0) = (\bar{d}^0)^t \left(\underbrace{A \bar{x}^*}_{\bar{b}} - A \bar{x}^0 - \rho_0 A \bar{d}^0 \right) = \bar{d}^0 \cdot \bar{d}^0 - \rho_0 (\bar{d}^0)^t A \bar{d}^0 = \\ &= \bar{d}^0 \cdot \bar{d}^0 - \frac{\bar{r}^0 \cdot \bar{d}^0}{(\bar{d}^0)^t A \bar{d}^0} (\bar{d}^0)^t A \bar{d}^0 = \bar{d}^0 \cdot \bar{d}^0 - \bar{d}^0 \cdot \bar{d}^0 = 0. \quad \square \end{aligned}$$

Esquema iterativo para resolver un sistema de dos ecuaciones con dos incógnitas ($A\bar{x} = \bar{b}$) mediante el método del gradiente conjugado

$$\begin{aligned}
\bar{d}^0 &= \bar{r}^0 = \bar{b} - A\bar{x}^0 \\
\rho_0 &= \frac{\bar{r}^0 \cdot \bar{d}^0}{(\bar{d}^0)^t A \bar{d}^0} \\
\bar{x}^1 &= \bar{x}^0 + \rho_0 \bar{d}^0 \\
\bar{r}^1 &= \bar{b} - A\bar{x}^1 \\
\alpha_0 &= -\frac{(\bar{d}^0)^t A \bar{r}^1}{(\bar{d}^0)^t A \bar{d}^0} \\
\bar{d}^1 &= \bar{r}^1 + \alpha_0 \bar{r}^0 \\
\rho_1 &= \frac{\bar{r}^1 \cdot \bar{d}^1}{(\bar{d}^1)^t A \bar{d}^1} \\
\text{Solución exacta} &\rightarrow \boxed{\bar{x}^2} = \bar{x}^1 + \rho_1 \bar{d}^1
\end{aligned}$$

A continuación vamos a generalizar el método anterior a dimensión n . En este caso la dirección de descenso en la iteración i será una combinación lineal del residuo en la iteración i y de la dirección de descenso en la iteración $i - 1$:

$$\bar{d}^i = \bar{r}^i + \alpha_{i-1} \bar{d}^{i-1}. \quad (19)$$

con

$$\alpha_{i-1} = -\frac{(\bar{d}^{i-1})^t A \bar{r}^i}{(\bar{d}^{i-1})^t A \bar{d}^{i-1}}$$

Por lo tanto el esquema iterativo será:

Esquema iterativo para resolver un sistema de n ecuaciones con n incógnitas ($A\bar{x} = \bar{b}$) mediante el método del gradiente conjugado

$$\begin{aligned}
\bar{d}^0 &= \bar{r}^0 = \bar{b} - A\bar{x}^0 \\
\rho_0 &= \frac{\bar{r}^0 \cdot \bar{d}^0}{(\bar{d}^0)^t A \bar{d}^0} \\
\bar{x}^1 &= \bar{x}^0 + \rho_0 \bar{d}^0 \\
\bar{r}^1 &= \bar{b} - A\bar{x}^1 \\
\alpha_0 &= -\frac{(\bar{d}^0)^t A \bar{r}^1}{(\bar{d}^0)^t A \bar{d}^0} \\
\bar{d}^1 &= \bar{r}^1 + \alpha_0 \bar{r}^0 \\
&\vdots
\end{aligned}$$

$$\begin{aligned}
\rho_i &= \frac{\bar{r}^i \cdot \bar{d}^i}{(\bar{d}^i)^t A \bar{d}^i} \\
\bar{x}^{i+1} &= \bar{x}^i + \rho_i \bar{d}^i \\
\bar{r}^{i+1} &= \bar{b} - A \bar{x}^{i+1} \\
\alpha_i &= -\frac{(\bar{d}^i)^t A \bar{r}^{i+1}}{(\bar{d}^i)^t A \bar{d}^i} \\
\bar{d}^{i+1} &= \bar{r}^{i+1} + \alpha_i \bar{d}^i \\
&\vdots
\end{aligned}$$

En principio nada asegura que el método anterior converja hacia la solución del sistema $A\bar{x} = \bar{b}$. En la siguiente sección se demostrará no solamente que el esquema anterior converge hacia la solución de $A\bar{x} = \bar{b}$, sino que obtiene la solución exacta en menos de n iteraciones si la dimensión de A es n y se utiliza aritmética exacta en los cálculos.

5.3. Propiedades del método del gradiente conjugado

En esta sección se supondrá que se está resolviendo mediante el método del gradiente conjugado el sistema $A\bar{x} = \bar{b}$ de n ecuaciones con n incógnitas con matriz de coeficientes simétrica y definida positiva.

Proposición 5.2. *Cada dirección de descenso del método del gradiente conjugado es conjugada, respecto a la matriz A , con todas las direcciones de descenso calculadas anteriormente:*

$$(\bar{d}^i)^t A \bar{d}^j = 0, \quad 0 \leq j < i. \quad (20)$$

Demostración. Consultar el apéndice A.

Proposición 5.3. *Los gradientes de f ($f(\bar{x}) = \frac{1}{2}\bar{x}^t A \bar{x} - \bar{b}^t \bar{x}$) en los puntos generados en el método del gradiente conjugado son ortogonales a las direcciones de descenso de las iteraciones anteriores:*

$$\bar{r}^i \cdot \bar{d}^j = 0, \quad 0 \leq j < i.$$

Demostración. Para realizar la demostración se estudiarán por separado los casos $j = i - 1$ y $j < i - 1$:

1. $\boxed{j = i - 1}$

$$\begin{aligned}
\bar{d}^{i-1} \cdot \bar{r}^i &= \bar{d}^{i-1} \cdot (\bar{r}^{i-1} - \rho_{i-1} A \bar{d}^{i-1}) = \bar{d}^{i-1} \cdot \bar{r}^{i-1} - \frac{\bar{r}^{i-1} \cdot \bar{d}^{i-1}}{(\bar{d}^{i-1})^t A \bar{d}^{i-1}} (\bar{d}^{i-1})^t A \bar{d}^{i-1} = \\
&\bar{d}^{i-1} \cdot \bar{r}^{i-1} - \bar{d}^{i-1} \cdot \bar{r}^{i-1} = 0.
\end{aligned}$$

En donde se ha utilizado la propiedad $\bar{r}^i = \bar{r}^{i-1} - \rho_{i-1} A \bar{d}^{i-1}$, propiedad que es fácil de demostrar sin más que repetir la “cuenta” realizada en (15) utilizando como dirección de descenso \bar{d}^{i-1} en vez de \bar{r}^{i-1} .

$$2. \quad \boxed{j < i - 1}$$

$$\bar{d}^j \cdot \bar{r}^i = \bar{d}^j \cdot (\bar{r}^{i-1} - \rho_{i-1} \bar{A} \bar{d}^{i-1}) \stackrel{\text{prop. 5.2}}{=} \bar{d}^j \cdot \bar{r}^{i-1} = \dots = \bar{d}^j \cdot \bar{r}^{j+1} \stackrel{\text{parte 1.}}{=} 0. \quad \square$$

Proposición 5.4. Si se utiliza aritmética exacta al realizar las operaciones, el método del gradiente conjugado alcanza la solución exacta del sistema $A\bar{x} = \bar{b}$ en no más de n iteraciones.*

$$\exists k \leq n/\bar{r}^k = \bar{0}.$$

Demostración. Supóngase que se han realizado $n - 1$ iteraciones sin alcanzar la solución de $A\bar{x} = \bar{b}$, o, lo que es lo mismo, sin hallar el mínimo de la función $f(\bar{x}) = \frac{1}{2}\bar{x}^t A \bar{x} - \bar{b}^t \bar{x}$. Si a continuación se realiza la iteración n se obtiene:

$$\bar{x}^n = \bar{x}^{n-1} + \rho_n \bar{d}^{n-1}.$$

Por la proposición 5.3 se tiene que \bar{r}^n es ortogonal a $V_n = \langle \bar{d}^0, \dots, \bar{d}^{n-1} \rangle$. Por otro lado al ser los vectores $\{\bar{d}^0, \dots, \bar{d}^{n-1}\}$ linealmente independientes (por la proposición 5.2 son vectores conjugados dos a dos respecto a A) la dimensión de V_n es n y por lo tanto $V_n = \mathbb{R}^n$. Dado que el único vector de \mathbb{R}^n que es ortogonal a todos los de \mathbb{R}^n es el vector nulo, se deduce que $\bar{r}^n = \bar{0}$. \square

Proposición 5.5.

$$\bar{r}^i \cdot \bar{d}^i = \bar{r}^i \cdot \bar{r}^i.$$

Demostración.

$$\bar{r}^i \cdot \bar{d}^i = \bar{r}^i \cdot (\bar{r}^i + \alpha_{i-1} \bar{d}^{i-1}) \stackrel{\text{prop. 5.3}}{=} \bar{r}^i \cdot \bar{r}^i. \quad \square$$

Proposición 5.6.

$$\bar{r}^i \cdot \bar{r}^j = 0 \quad 0 \leq j < i.$$

Demostración. En la iteración j del gradiente conjugado se verifica $\bar{d}^j = \bar{r}^j + \alpha_{j-1} \bar{d}^{j-1}$, si se despeja \bar{r}^j de la expresión anterior se obtiene:

$$\bar{r}^j = \bar{d}^j - \alpha_{j-1} \bar{d}^{j-1}$$

y por lo tanto $\bar{r}^j \in V_{j+1} = \langle \bar{d}^0, \dots, \bar{d}^j \rangle$. Como por otro lado (proposición 5.3) \bar{r}^i es ortogonal a V_{j+1} entonces $\bar{r}^i \cdot \bar{r}^j = 0$. \square

5.4. Algoritmo del método del gradiente conjugado

A partir del esquema del método del gradiente conjugado para el caso de dimensión n es fácil diseñar un algoritmo que resuelva el sistema $A\bar{x} = \bar{b}$ en donde A es una matriz cuadrada de dimensión n simétrica y definida positiva:

*Esta proposición implica que el método del gradiente conjugado, en teoría, es un método directo. En la práctica no es así debido a los errores de redondeo y a que el método se parará en una iteración muy inferior a la iteración n .

Algoritmo 5 Algoritmo que resuelve el sistema $A\bar{x} = \bar{b}$ mediante el método del gradiente conjugado. Versión provisional.

Leer $A, \bar{b}, \bar{x}^0, maxiter, \epsilon$
 $i \leftarrow 0, \bar{r}^0 \leftarrow \bar{b} - A\bar{x}^0, tol \leftarrow \|\bar{r}^0\|$
mientras $i < maxiter$ y $tol > \epsilon$ **hacer**
 $\rho_i \leftarrow \frac{(\bar{r}^i)^t \bar{d}^i}{(\bar{d}^i)^t A \bar{d}^i}$
 $\bar{x}^{i+1} \leftarrow \bar{x}^i + \rho_i \bar{d}^i$
 $\bar{r}^{i+1} \leftarrow \bar{b} - A\bar{x}^{i+1}$
 $\alpha_i \leftarrow -\frac{(\bar{r}^{i+1})^t A \bar{d}^i}{(\bar{d}^i)^t A \bar{d}^i}$
 $\bar{d}^{i+1} \leftarrow \bar{r}^{i+1} + \alpha_i \bar{d}^i$
 $tol \leftarrow \|\bar{r}^{i+1}\|$
 $i \leftarrow i + 1$
fin mientras

El número de operaciones en cada iteración del algoritmo anterior es: $8n^2 + 11n - 1$. El problema, al igual que en el caso del gradiente, son los productos matriz por vector. Utilizando las propiedades que se han demostrado en el apartado anterior se va a modificar el algoritmo para conseguir que en cada iteración solamente haya que realizar un producto matriz vector.

1. Modificación del término

$$\alpha_i = -\frac{(\bar{r}^{i+1})^t A \bar{d}^i}{(\bar{d}^i)^t A \bar{d}^i}. \quad (21)$$

Despejando $A \bar{d}^i$ de $\bar{r}^{i+1} = \bar{r}^i - \rho_i A \bar{d}^i$ se obtiene:

$$A \bar{d}^i = \frac{1}{\rho_i} (\bar{r}^i - \bar{r}^{i+1}),$$

e introduciendo $A \bar{d}^i$ en (21)

$$\alpha_i = -\frac{\frac{1}{\rho_i} (\bar{r}^{i+1})^t (\bar{r}^i - \bar{r}^{i+1})}{\frac{1}{\rho_i} (\bar{d}^i)^t (\bar{r}^i - \bar{r}^{i+1})} = -\frac{\bar{r}^{i+1} \cdot \bar{r}^i - \bar{r}^{i+1} \cdot \bar{r}^{i+1}}{\bar{d}^i \cdot \bar{r}^i - \bar{d}^i \cdot \bar{r}^{i+1}} = \boxed{\frac{\bar{r}^{i+1} \cdot \bar{r}^{i+1}}{\bar{r}^i \cdot \bar{r}^i}}.$$

En donde $\bar{r}^{i+1} \cdot \bar{r}^i = 0$ por la proposición 5.6, $\bar{d}^i \cdot \bar{r}^{i+1} = 0$ por la proposición 5.3 y $\bar{d}^i \cdot \bar{r}^i = \bar{r}^i \cdot \bar{r}^i$ por la proposición 5.5.

2. Modificación del término

$$\bar{r}^{i+1} = \bar{b} - A\bar{x}^{i+1}. \quad (22)$$

De la igualdad (15) y tomando como dirección de descenso \bar{d}^i en vez de \bar{r}^i se obtiene:

$$\boxed{\bar{r}^{i+1} = \bar{r}^i - \rho_i A \bar{d}^i}. \quad (23)$$

3. Modificación del término

$$\rho_i = \frac{(\bar{r}^i)^t \bar{d}^i}{(\bar{d}^i)^t A \bar{d}^i}. \quad (24)$$

De nuevo utilizando la proposición 5.6:

$$\rho_i = \frac{(\bar{r}^i)^t \bar{d}^i}{(\bar{d}^i)^t A \bar{d}^i} = \boxed{\frac{(\bar{r}^i)^t \bar{r}^i}{(\bar{d}^i)^t A \bar{d}^i}}. \quad (25)$$

Por lo tanto para calcular α_i ya no es necesario realizar ningún producto matriz vector y para calcular \bar{r}^{i+1} y ρ_i basta con un solo producto matriz vector ($A \bar{d}^i$) que es el mismo en ambos casos. Estas modificaciones permiten construir un nuevo algoritmo del gradiente conjugado (algoritmo 6).

Algoritmo 6 Algoritmo que resuelve el sistema $A\bar{x} = \bar{b}$ mediante el método del gradiente conjugado. Versión final.

```

Leer  $A, \bar{b}, \bar{x}, maxiter, \epsilon$ 
 $i \leftarrow 0, \bar{r} \leftarrow \bar{b} - A\bar{x}, \bar{d} \leftarrow \bar{r}, tol \leftarrow \|\bar{r}\|, \beta \leftarrow \bar{r}^t \bar{r}$ 
mientras  $i < maxiter$  y  $tol > \epsilon$  hacer
     $\bar{z} \leftarrow A\bar{d}$ 
     $\rho \leftarrow \frac{\beta}{(\bar{d}^i)^t \bar{z}}$ 
     $\bar{x} \leftarrow \bar{x} + \rho \bar{d}$ 
     $\bar{r} \leftarrow \bar{r} - \rho \bar{z}$ 
     $\gamma \leftarrow \beta$ 
     $\beta \leftarrow \bar{r}^t \bar{r}$ 
     $\alpha \leftarrow -\frac{\beta}{\gamma}$ 
     $\bar{d} \leftarrow \bar{r} + \alpha \bar{d}$ 
     $tol \leftarrow \|\bar{r}\|$ 
     $i \leftarrow i + 1$ 
fin mientras

```

En este algoritmo se realiza un solo producto matriz por vector, siendo el número de operaciones total $2n^2 + 11n$. Por lo tanto el número de operaciones por iteración es, al igual que en el método del gradiente, aproximadamente igual (para n grande) a $2n^2$. Sin embargo a la hora de escoger entre los dos métodos, siempre se deberá escoger el del gradiente conjugado puesto que el número de iteraciones necesarias para alcanzar una reducción del error dada (**velocidad de convergencia**) es mucho más pequeño en el método del gradiente conjugado que en el método del gradiente. La estimación de la velocidad de convergencia es una tarea complicada para ambos métodos, tarea que queda fuera de los objetivos de estos apuntes. Se aconseja consultar [1], [2], [7] y [8] para estudiar como se obtienen las expresiones de la velocidad de convergencia que se dan a continuación para ambos métodos:

1. **Velocidad de convergencia en el método del gradiente,**

$$k \sim \frac{\text{cond}(A)}{4} \ln \left(\frac{1}{\alpha} \right).$$

2. **Velocidad de convergencia en el método del gradiente conjugado,**

$$k \sim \frac{\sqrt{\text{cond}(A)}}{2} \ln \left(\frac{2}{\alpha} \right).$$

En donde α es el valor en que se reduce el error inicial, $\text{cond}(A)$ es el condicionamiento de la matriz A (ver el apéndice B) y k es el número de iteraciones necesarias para que el error se reduzca α veces respecto al error inicial.

Observación 5.1. Como ya se ha visto anteriormente, si la matriz no es hueca, el número de operaciones elementales a realizar en cada iteración es aproximadamente $2n^2$. En consecuencia el número total de operaciones necesarias para conseguir reducir el error inicial α veces en el método del gradiente conjugado es $n^2 \ln\left(\frac{2}{\alpha}\right) \sqrt{\text{cond}(A)}$ y el método podrá competir con el método de Cholesky cuando

$$\text{cond}(A) \leq \frac{n^2}{9 \left(\ln\left(\frac{2}{\alpha}\right)\right)^2}.$$

Observación 5.2. Si la matriz es hueca el número de operaciones a realizar en cada iteración del orden de $2nc$ en donde c es un coeficiente que indica el número medio de elementos no nulos en cada fila y que generalmente es mucho más pequeño que n . En este caso el número total de operaciones necesarias para conseguir reducir el error inicial α veces será $nc \ln\left(\frac{2}{\alpha}\right) \sqrt{\text{cond}(A)}$. Si, por ejemplo, $c = \sqrt{n}$ se tendría un número de operaciones igual a $n^{3/2} \ln\left(\frac{2}{\alpha}\right) \sqrt{\text{cond}(A)}$ y en este caso el gradiente conjugado podrá competir con Cholesky siempre que

$$\text{cond}(A) \leq \frac{n^3}{9 \left(\ln\left(\frac{2}{\alpha}\right)\right)^2}.$$

A. Demostración de la conjugación de las direcciones de descenso

Se recuerda la proposición 5.2:

Cada dirección de descenso del método del gradiente conjugado es conjugada, respecto a la matriz A , con todas las direcciones de descenso calculadas anteriormente:

$$(\bar{d}^i)^t A \bar{d}^j = 0, \quad 0 \leq j < i. \quad (26)$$

En donde A es la matriz de coeficientes (simétrica y definida positiva) del sistema $A\bar{x} = \bar{b}$ de n ecuaciones con n incógnitas.

Demostración. Se realizará la demostración por inducción:

1. La proposición es cierta para $i = 1$ (proposición 5.1).
2. Se supone que la proposición es cierta para $i - 1$:

$$(\bar{d}^{i-1})^t A \bar{d}^j = 0, \quad 0 \leq j < i - 1 \quad (27)$$

3. Se demuestra (26) utilizando (27):

- a) Se sabe que

$$\bar{d}^i = \bar{r}^i + \alpha_{i-1} \bar{d}^{i-1} \quad i = 1, 2, 3, \dots, \quad (28)$$

con $\bar{d}^0 = \bar{r}^0$.

- b) Para el caso $j = i - 1$ es fácil comprobar, utilizando (28) para expresar \bar{d}^i en función de \bar{d}^{i-1} y \bar{r}^i , que

$$(\bar{d}^i)^t A \bar{d}^{i-1} = (\bar{r}^i)^t A \bar{d}^{i-1} + \alpha_{i-1} (\bar{d}^{i-1})^t A \bar{d}^{i-1} = 0.$$

- c) Para el resto de los casos ($j < i - 1$) se deduce, volviendo a utilizar (28), que:

$$(\bar{d}^i)^t A \bar{d}^j = (\bar{r}^i)^t A \bar{d}^j + \alpha_{i-1} (\bar{d}^{i-1})^t A \bar{d}^j \quad 0 \leq j < i - 1. \quad (29)$$

El segundo sumando de la derecha de (29) es nulo por (27), queda, por tanto, demostrar que $(\bar{r}^i)^t A \bar{d}^j = 0$ para $j < i - 1$. Para esta última tarea será conveniente definir el siguiente subespacio:

$$V_{j+1} = \langle \bar{d}^0, \bar{d}^1, \dots, \bar{d}^j \rangle.$$

Por otro lado en la iteración j del gradiente conjugado se verifica $\bar{d}^j = \bar{r}^j + \alpha_{j-1} \bar{d}^{j-1}$, si se despeja \bar{r}^j de la expresión anterior se obtiene:

$$\bar{r}^j = \bar{d}^j - \alpha_{j-1} \bar{d}^{j-1}$$

y por lo tanto $\bar{r}^j \in V_{j+1} = \langle \bar{d}^0, \dots, \bar{d}^j \rangle$ por ser \bar{r}^j combinación lineal de dos vectores de V_{j+1} . Esto es cierto para todo $j < i - 1$, por lo tanto:

$$V_{j+1} = \langle \bar{d}^0, \bar{d}^1, \dots, \bar{d}^j \rangle = \langle \bar{r}^0, \bar{r}^1, \dots, \bar{r}^j \rangle.$$

Despejando de la igualdad (23) ($\bar{r}^{j+1} = \bar{r}^j - \rho_j A \bar{d}^j$) se obtiene:

$$A \bar{d}^j = \frac{\bar{r}^j - \bar{r}^{j+1}}{\rho_j}$$

En consecuencia:

$$A \bar{d}^j \in V_{j+2} = \langle \bar{r}^0, \bar{r}^1, \dots, \bar{r}^j, \bar{r}^{j+1} \rangle = \langle \bar{d}^0, \bar{d}^1, \dots, \bar{d}^j, \bar{d}^{j+1} \rangle,$$

y se habrá concluido la demostración si se demuestra que \bar{r}^i es ortogonal a V_{j+2} . (para $0 \leq j < i - 1$). Para ello bastará con demostrar que

$$\bar{r}^i \cdot \bar{d}^j = 0 \quad 0 \leq j < i. \quad (30)$$

También se demostrará por inducción:

1) La proposición es cierta para $i = 1$:

$$\bar{r}^1 \cdot \bar{d}^0 = \bar{r}^1 \cdot \bar{r}^0 = 0.$$

2) Se supone cierto que

$$\bar{r}^{i-1} \cdot \bar{d}^j = 0, \quad 0 \leq j < i - 1. \quad (31)$$

3) Se demostrará (30) utilizando (31):

$$\bar{r}^i \cdot \bar{d}^j = \bar{r}^{i-1} \cdot \bar{d}^j - \rho_{i-1} (\bar{d}^j)^t A \bar{d}^{i-1},$$

para demostrar que esta resta es cero distinguiremos dos casos:

■ $j = i - 1$

$$\bar{r}^{i-1} \cdot \bar{d}^{i-1} - \rho_{i-1} (\bar{d}^{i-1})^t A \bar{d}^{i-1} = \bar{r}^{i-1} \cdot \bar{d}^{i-1} - \frac{\bar{r}^{i-1} \cdot \bar{d}^{i-1}}{(\bar{d}^{i-1})^t A \bar{d}^{i-1}} (\bar{d}^{i-1})^t A \bar{d}^{i-1} = 0.$$

■ $j < i - 1$

$$\underbrace{\bar{r}^{i-1} \cdot \bar{d}^j}_0 - \rho_{i-1} \underbrace{(\bar{d}^j)^t A \bar{d}^{i-1}}_0 = 0,$$

en donde el primer sumando es nulo por (31) y el segundo por (27).

Por lo tanto ha quedado demostrado que \bar{r}^i es ortogonal a V_{j+2} y, en consecuencia, $(\bar{r}^i)^t A \bar{d}^j = 0$, para $1 \leq j < i - 1$. \square

*Obsérvese que esta es la proposición 5.3 que ya se ha demostrado en el apartado 5, pero entonces se hizo la demostración suponiendo que la proposición 5.2 era cierta, cosa que no se puede hacer ahora.

B. Condicionamiento de un sistema

Al hablar de condicionamiento de un sistema de ecuaciones nos referimos a las estabilidad que presenta la solución frente a pequeñas perturbaciones tanto en la matriz de coeficientes como en el vector de términos independientes. Veamos el siguiente ejemplo introductorio:

La solución del sistema:

$$\begin{pmatrix} 0,1 & 3,0 & 3,0 & 4,0 \\ 0,4 & 12,2 & 20,1 & 26,1 \\ 0,1 & 3,1 & 7,0 & 9,0 \\ 0,2 & 6,1 & 10,1 & 13,0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} = \begin{pmatrix} 10,1 \\ 58,8 \\ 19,2 \\ 29,4 \end{pmatrix}$$

es $x = y = z = t = 1$. Veamos lo que sucede si se perturba el vector de segundos miembros:

$$\begin{pmatrix} 0,1 & 3,0 & 3,0 & 4,0 \\ 0,4 & 12,2 & 20,1 & 26,1 \\ 0,1 & 3,1 & 7,0 & 9,0 \\ 0,2 & 6,1 & 10,1 & 13,0 \end{pmatrix} \begin{pmatrix} x' \\ y' \\ z' \\ t' \end{pmatrix} = \begin{pmatrix} 10,095 \\ 58,805 \\ 19,195 \\ 29,395 \end{pmatrix}$$

La solución es ahora $x' = 351,45$; $y' = -11$; $z' = 1,05$; $t' = 1,2$.

Si en vez de perturbar el vector de segundos miembros perturbamos la matriz de coeficientes:

$$\begin{pmatrix} 0,100 & 3,005 & 3,000 & 4,000 \\ 0,400 & 12,200 & 20,100 & 26,100 \\ 0,100 & 3,100 & 6,995 & 9,000 \\ 0,200 & 6,100 & 10,100 & 12,995 \end{pmatrix} \begin{pmatrix} x'' \\ y'' \\ z'' \\ t'' \end{pmatrix} = \begin{pmatrix} 10,1 \\ 58,8 \\ 19,2 \\ 29,4 \end{pmatrix}$$

llegaremos a la solución: $x'' = -3,425$; $y'' = 1,150$; $z'' = 1,053$; $t'' = 0,957$

Es evidente que sistemas de este tipo serán indeseables ya que, en general, los datos (matriz de coeficientes y vector de segundos miembros) procederán de cálculos previos realizados con ordenador y estarán sometidos a errores de redondeo, o procederán de medidas tomadas con el correspondiente error. Además, resolver un sistema como el precedente suele acarrear más iteraciones que resolver uno al que no le ocurra esto.

El llamado **número de condición** constituye una herramienta eficaz para predecir el comportamiento de un sistema de ecuaciones lineales frente a perturbaciones de la matriz de coeficientes y/o del vector de segundos miembros. Para definir este número, consideremos el sistema $A\bar{x} = \bar{b}$. Es obvio que si perturbamos el segundo miembro una cantidad $\delta\bar{b}$, la solución del nuevo sistema será $\bar{x} + \delta\bar{x}$, esto es:

$$\left. \begin{array}{l} A\bar{x} = \bar{b} \\ A(\bar{x} + \delta\bar{x}) = \bar{b} + \delta\bar{b} \end{array} \right\} \Rightarrow A\delta\bar{x} = \delta\bar{b} \Rightarrow \delta\bar{x} = A^{-1}\delta\bar{b}$$

Considerando una norma matricial subordinada a una norma vectorial cualquiera:

$$\left. \begin{aligned} \|\bar{b}\| &\leq \|A\| \|\bar{x}\| \\ \|\delta\bar{x}\| &= \|A^{-1}\| \|\delta\bar{b}\| \end{aligned} \right\} \Rightarrow \frac{\|\delta\bar{x}\|}{\|\bar{x}\|} \leq (\|A\| \|A^{-1}\|) \frac{\|\delta\bar{b}\|}{\|\bar{b}\|}$$

De forma análoga, si perturbamos en δA la matriz del sistema, se llega a la expresión:

$$\frac{\|\delta\bar{x}\|}{\|\bar{x} + \delta\bar{x}\|} \leq (\|A\| \|A^{-1}\|) \frac{\|\delta A\|}{\|A\|}$$

Es decir, tanto si perturbamos la matriz del sistema como si perturbamos el vector de segundos miembros, el error relativo de la solución está acotado en función de los errores en los coeficientes multiplicados por el número $\|A\| \|A^{-1}\|$.

Así pues, se denomina **condicionamiento** (o número de condición) del sistema $A\bar{x} = \bar{b}$, $\text{cond}(A)$, al número $\|A\| \|A^{-1}\|$, donde $\|\cdot\|$ es la norma elegida en el espacio de las matrices cuadradas de orden n e invertibles.

De la definición anterior se desprenden de forma inmediata las siguientes propiedades:

1. $\text{cond}(A) \geq 1$
2. $\text{cond}(A^{-1}) = \text{cond}(A)$
3. $\text{cond}(\lambda A) = \text{cond}(A), \quad \forall \lambda \neq 0$

Además, en el caso de usar $\|\cdot\|_2$, pueden demostrarse las dos propiedades siguientes:

1. Si la matriz A es **simétrica** y denotamos por λ_M a su valor propio de mayor módulo y por λ_m a su valor propio de menor módulo, se cumple:

$$\text{cond}_2(A) = \frac{|\lambda_M|}{|\lambda_m|}$$

2. Si la matriz A **no es simétrica** y denotamos por λ_M al valor propio de mayor módulo de $A^T A$ y por λ_m el valor propio de menor módulo de $A^T A$, se cumple:

$$\text{cond}_2(A) = \left(\frac{|\lambda_M|}{|\lambda_m|} \right)^{1/2}$$

Las técnicas de **precondicionamiento** permiten mejorar el número de condición de un sistema. La idea es considerar una matriz invertible M y resolver el sistema $M^{-1}A\bar{x} = M^{-1}\bar{b}$ en lugar de $A\bar{x} = \bar{b}$. Esta matriz M recibe el nombre de **matriz de precondicionamiento**. Naturalmente, para que este procedimiento sea efectivo se tiene que dar:

$$\text{cond}(M^{-1}A) < \text{cond}(A)$$

Como el condicionamiento de una matriz unidad es 1, cualquiera que sea la norma elegida, en principio, lo más conveniente sería tomar M de tal modo que $M^{-1}A = I$, es decir,

$M = A$, aunque esta elección tiene escaso interés práctico, pues equivale a resolver el sistema:

$$\begin{aligned} A\bar{x} &= \bar{b}; \quad M^{-1}A\bar{x} = M^{-1}\bar{b} \\ \text{Si } M &= A \Rightarrow A^{-1}A\bar{x} = A^{-1}\bar{b} \Rightarrow \bar{x} = A^{-1}\bar{b} \end{aligned}$$

No obstante, esta idea nos proporciona una pista de cómo actuar: buscar matrices de preconditionamiento “parecidas” a la matriz del sistema y fáciles de invertir. Un ejemplo clásico es el preconditionamiento diagonal que consiste en utilizar como matriz de preconditionamiento una matriz diagonal cuyos elementos no nulos coincidan con los elementos diagonales de A . Este preconditionamiento es muy económico desde el punto de vista computacional, pues:

$$\begin{aligned} A &= \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}, \quad M = \begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix}, \\ M^{-1} &= \begin{pmatrix} 1/a_{11} & 0 & \dots & 0 \\ 0 & 1/a_{22} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1/a_{nn} \end{pmatrix} \end{aligned}$$

y sólo añade $2n$ operaciones (divisiones) al proceso de cálculo.

Sin embargo, hay que tomar algunas precauciones a la hora de preconditionar un sistema. La más importante es tener en cuenta que se puede destruir la estructura de la matriz del sistema. Por ejemplo, si la matriz A es simétrica, la matriz $M^{-1}A$ no tiene por qué serlo, o si la matriz A es definida positiva, la matriz $M^{-1}A$ puede no serlo. En particular, si se trabaja con métodos de resolución de tipo gradiente se prefiere preconditionar de la siguiente manera:

$$A\bar{x} = \bar{b} \Rightarrow \underbrace{M^{-1}A}_{P} \underbrace{M^{-1}\bar{x}}_{\bar{y}} = \underbrace{M^{-1}\bar{b}}_{\bar{c}} \Rightarrow P\bar{y} = \bar{c}$$

resolviéndose a continuación el sistema: $P\bar{y} = \bar{c}$ y posteriormente el $M\bar{x} = \bar{y}$. Este método conserva las propiedades de la matriz A .

El estudio de otros preconditionadores más sofisticados cae fuera del objetivo de estas notas y, en este sentido, nos limitaremos a comentar que entre los más populares están los basados en la descomposición LU de la matriz A . En la idea de utilizar matrices de preconditionamiento “parecidas” a la matriz del sistema, este tipo de preconditionadores trabajan, en lugar de con la descomposición $A = LU$ exacta, con una descomposición aproximada, de tal forma que $A = L'U' + R$ y utilizan como matriz de preconditionamiento $M = L'U'$. Los preconditionadores de este tipo, **denominados LU incompletos (ILU)**, son particularmente efectivos en el caso, muy frecuente en la práctica, de que la matriz A tenga muchos elementos nulos, es decir, sea una matriz **hueca**. En este caso, se suele tomar una matriz L' con el mismo patrón de huecos que la parte triangular inferior de la matriz A .

Bibliografía

- [1] O. Axelsson. Iterative Solution Methods. Ed. Cambridge University Press, 1996.
- [2] O. Axelsson, V.A. Barker. Finite Element Solution of Boundary Value Problems. Theory and Computations. Ed. Academic Press, 1984.
- [3] R. L. Burden y J. D. Faires. Métodos Numéricos (3ª Edición). Ed. Thomson, 2004.
- [4] C. Conde y G. Winter. Métodos y algoritmos básicos del álgebra numérica. Ed. Reverté, 1989.
- [5] G. H. Golub y C. F. Van Loan. Matrix computations (III Ed.), 1996.
- [6] G. H. Golub y J. M. Ortega. Scientific computing: an introduction with parallel computing. Academic Press Inc., 1993.
- [7] P.Lascaux y R. Théodor. Analyse numérique appliquée a l'art de l'ingénieur (2 tomos). Ed. Dunod, 2000.
- [8] J.R. Shewchuk. An Introduction to the Conjugate Gradient Method Without the Agonizing Pain. School of Computer Science. Carnegie Mellon University, 1994.