

# Análisis aplicado

Proyecto: método de máximo descenso

Guillermo Santiago Novoa Pérez

000125089

2 de febrero, 2015.

## Objetivo

El proyecto tiene como finalidad estudiar el comportamiento teórico y práctico del método de máximo descenso con búsqueda lineal exacta. En este punto es importante recordar que todos los métodos estudiados en el curso están guiados por las propiedades de la dirección opuesta al gradiente. Por lo tanto, resulta fundamental comprender el funcionamiento de dicho método en las funciones no lineales más elementales: las cuadráticas estrictamente convexas.

1. Las cuadráticas que vamos a utilizar en el proyecto son de la forma

$$f(x) = \frac{1}{2}x^T Hx + g^T x$$

en donde  $H$  es una matriz simétrica positiva definida de  $n \times n$  con número de condición conocido; el vector  $g$  es generado en forma aleatoria con la función `randn`.

2. El resultado sobre la convergencia global se formula como sigue:

**Teorema:** El método de máximo descenso aplicado a cuadráticas estrictamente convexas satisface la siguiente propiedad

$$\|x_{k+1} - x^*\|_H^2 \leq \left[ \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}} \right]^2 \|x_k - x^*\|_H^2, \quad (1)$$

en donde  $\|\cdot\|_H$  representa a la norma vectorial inducida por  $H$ ;  $\lambda_{\min}, \lambda_{\max}$  son los valores propios extremos de la matriz  $H$ .

El método parte de cualquier aproximación inicial  $x_0$ ; la iteración general tiene la forma

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k)$$

en donde  $\alpha_k$  es el minimizador global de

$$\phi(\alpha) = f(x_k - \alpha \nabla f(x_k)),$$

es decir

$$\alpha_k = \frac{\|\nabla f(x_k)\|^2}{\nabla f(x_k)^T H \nabla f(x_k)}$$

La prueba del resultado anterior es relativamente elaborada y requiere cierto cuidado para entender todos los pasos<sup>1</sup>

## Actividades del proyecto

1. Genera resultados con el programa desarrollado en la sesión anterior de laboratorio. Utiliza  $n = 10$  con puntos de inicio aleatorios. Utiliza números de condición  $10^1, 10^2, 10^3$ .
2. Utiliza la desigualdad (2) para predecir el número de pasos que le toma al método alcanzar la condición

$$\|x_{k+1} - x^*\|_H^2 \leq 10^{-5}$$

3. Construye los lados izquierdo y derecho de la desigualdad (2). ¿Es frecuente que ocurra la igualdad?
4. Reconstruye la prueba de Luenberger y justifica cada paso.
5. Escribe un reporte en L<sup>A</sup>T<sub>E</sub>X con los resultados del proyecto.

## Proyecto

### Condiciones

1. Resumen de datos respuesta del método de máximo descenso.  
Para el proyecto se utilizaron las siguientes condiciones de diseño:

- $n = 10$ , es decir, la dimensión en la que se trabaja es  $\mathbb{R}^n$ .
- El número máximo de iteraciones es 10,000, si el método no llegase a convergir antes de las 10,000 iteraciones, un mensaje aparecería comentándolo.
- La *tolerancia* es  $10^{-9}$ , lo que quiere decir que si la norma del gradiente en algún momento es menor a  $10^{-9}$ , el programa para y da por encontrado al  $x^*$  que minimiza  $f(x)$ .

---

<sup>1</sup>David G. Luenberger. Linear and Nonlinear Programming, second edition, Addison-Wesley Publishing Company. 1984.

- El método se ejecutó cinco veces en total. Las primeras tres veces se generaron matrices aleatoriamente, con números de condición dado por  $10^1$ ,  $10^2$ ,  $10^3$  respectivamente, un vector  $g$  generado de forma aleatoria con la función `randn` y un vector inicial  $x_0$  generado de la misma forma. Para poder comparar y comprobar los resultados a las preguntas (2) y (3), se realizaron dos iteraciones más, generando de nuevo matrices con números de condición  $10^1$  y  $10^3$  de manera aleatoria. Para generar el gradiente (o vector  $g$ ), se estableció una  $x^*$  óptima de inicio y se multiplicó por (menos) la matriz aleatoria generada anteriormente para obtener  $g$ ; de esta manera se pudieron comprobar resultados que incluyeran el óptimo de la función  $f(x)$ .
- Existe un criterio adicional para la impresión de los datos. Lo que esto ocasiona es que sólo se impriman las iteraciones en las cuales el cambio entre el valor de  $f$  impreso por último y el valor de  $f$  en la iteración actual sea significativo (ver código para ver a qué se refiere significativo).
- Todas las tablas vienen al final del trabajo junto con las conclusiones para dar mayor facilidad de lectura e impresión.
- Las tablas se separan por número de ejecución del programa. Las especificaciones de las ejecuciones vienen en cada tabla.

(a) 1<sub>a</sub> ejecución :  $N_{cond} = 10^1$

(b) 2<sub>a</sub> ejecución :  $N_{cond} = 10^2$

(c) 3<sub>a</sub> ejecución :  $N_{cond} = 10^3$

————CASOS  $X^*$  CONOCIDA (  $x^* = e$  ) —————

(d) 4<sub>a</sub> ejecución :  $N_{cond} = 10^1$

(e) 5<sub>a</sub> ejecución :  $N_{cond} = 10^3$

2. Predecir número de pasos que le toma al método.

$$\begin{aligned}
 \|x_{k+1} - x^*\|_H^2 &\leq 10^{-5} \\
 \|x_{k+1} - x^*\|_H^2 &\leq \left[ \frac{\lambda_{min} - \lambda_{max}}{\lambda_{min} + \lambda_{max}} \right]^2 \|x_k - x^*\|_H^2 \\
 &= \left[ \frac{cond(H) - 1}{cond(H) + 1} \right]^2 \|x_k - x^*\|_H^2
 \end{aligned}$$

Sea  $q := \left\lceil \frac{\text{cond}(H)-1}{\text{cond}(H)+1} \right\rceil$

$$\therefore \|x_{k+1} - x^*\|_H^2 \leq (q^2) \|x_k - x^*\|_H^2$$

$$\begin{aligned} \text{Pero } \|x_k - x^*\|_H^2 &\leq (q^2) \|x_{k-1} - x^*\|_H^2 \\ &\dots \\ \|x_1 - x^*\|_H^2 &\leq (q^2) \|x_0 - x^*\|_H^2 \end{aligned}$$

$$\begin{aligned} \therefore \|x_{k+1} - x^*\|_H^2 &\leq (q^2)^{k+1} \|x_0 - x^*\|_H^2 \\ \iff \|x_{k+1} - x^*\|_H^2 &\leq (q)^{2(k+1)} \|x_0 - x^*\|_H^2 \end{aligned}$$

Queremos que  $\|x_{k+1} - x^*\|_H^2 \leq 10^{-5}$ , por lo que si  $(q)^{2(k+1)} \|x_0 - x^*\|_H^2 \leq 10^{-5}$ , ya está

$$\begin{aligned} (q)^{2(k+1)} \|x_0 - x^*\|_H^2 &\leq 10^{-5} \\ \iff (q)^{2(k+1)} &\leq \frac{10^{-5}}{\|x_0 - x^*\|_H^2} \\ \iff (2k+2) \log q &\leq \log \left[ \frac{10^{-5}}{\|x_0 - x^*\|_H^2} \right] \\ \iff 2(k+1) &\geq \frac{-5 \log 10 - \log [\|x_0 - x^*\|_H^2]}{\log q} \\ \iff k &\geq (-1) \frac{5 \log 10 + \log [\|x_0 - x^*\|_H^2]}{2 \log q} - 1 \\ &= \text{ceil} * (-1) \frac{5 \log 10 + \log [\|x_0 - x^*\|_H^2]}{2 \log (\text{cond}(H) - 1) - 2 \log (\text{cond}(H) + 1)} - 1 \end{aligned}$$

Para poder encontrar el valor exacto de k se necesitaría conocer el valor de  $\|x_0 - x^*\|_H^2$ , lo cuál en la práctica (y para los primeros 3 ejemplos que se ejecutaron) no es posible y sólo se puede aproximar. Para poder evaluar qué tan buena es esta predicción se usarán los datos de la cuarta y de la quinta ejecución, conociendo  $x^*$  de antemano para poder comparar.

*k* predicción *vs* *k* real

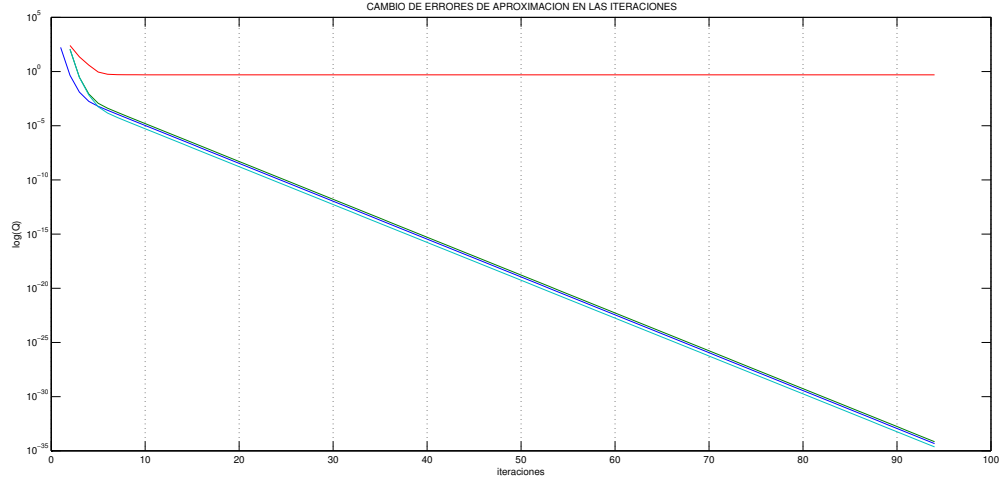
$k_{predic}$	$k_{real}$	$  x_{k_{predic}}  $
41	11	1.24E-08
3824	16	1.79E-10

Los resultados indican que aunque se tenía previsto alcanzar la cota del error de aproximación hasta las  $41_{ava}$  y  $3824_{ava}$  iteraciones, los resultados deseados (o la cercanía de las aproximaciones) se alcanzaron mucho antes. Para mayor análisis de esto, ver las conclusiones del final.

3. Comparación lado izquierdo y lado derecho de la desigualdad.

Se intenta ver qué tan laxa es la cota superior que se demostrará en el apartado 4 de este trabajo. Dado que se está tratando con números que pueden tener errores o variaciones muy pequeñas que, aunque a simple vista muchas veces pueden parecer casi el mismo, al compararse en la máquina arrojan resultados de falsedad al preguntar si son el mismo, se intentará, mejor, comparar o calcular esa diferencia para intentar ver si en algún momento esa cota se alcanza o cada cuánto parece esto ocurrir.

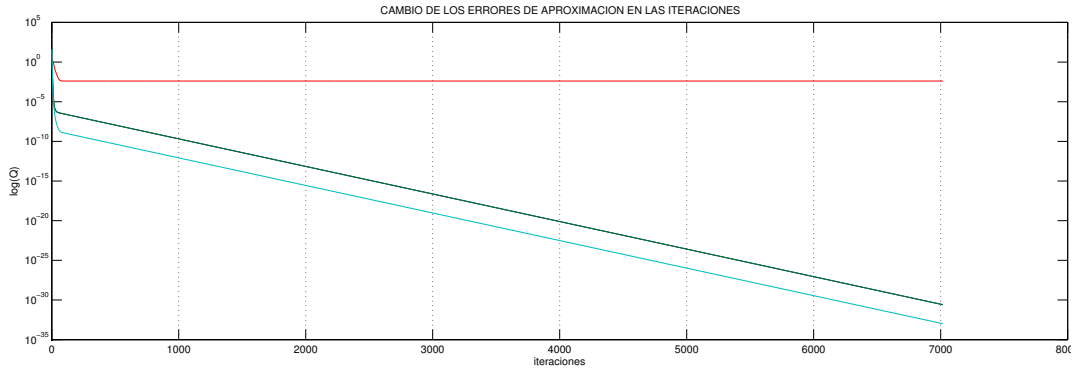
Después de haber almacenado los valores de  $x_k$  en cada iteración (de nuevo en la  $4_a$  y  $5_a$  iteración para poder comparar con el valor "óptimo" de  $x$ ,  $x^*$ ,) y de haber operado estos valores de tal manera que los datos que quedaran fueran el lado izquierdo y el lado derecho se realizaron dos diferentes comparaciones. Antes de seguir y para poder entender mejor los resultados se mostrará una gráfica a continuación con los datos ya modificados de la  $4_a$  ejecución del programa ( $N_{con} = 10^1$ ).



La gráfica muestra los cambios en  $(d)$  logaritmos de valores diferentes a través de las iteraciones. Se utilizaron logaritmos de los valores para poder entender mejor la gráfica dado que los cambios entre los errores de aproximación (casi desde el principio de las iteraciones) son mínimo de  $O(1/k^2)$

- Las curvas morado y verde representan al error absoluto de la aproximación, al tiempo  $k$  y al tiempo  $k+1$ , al cuadrado:  $\|x_k - x^*\|_H^2$ .
- La curva azul claro representa la diferencia absoluta entre el lado derecho de la desigualdad y el lado izquierdo, es decir,  $LD - LI$ .
- Parece ser que la curva más interesante es la roja, ésta representa la diferencia relativa entre el lado derecho y el lado izquierdo, es decir,  $(LD - LI)/LD$ .

Antes de analizar lo que esto puede significar, se expondrá la misma gráfica con los valores de los datos para la 5ª ejecución del programa ( $N_{con} = 10^3$ ).



- Las curvas de colores verde fuerte representan, de nuevo, los valores de el error absoluto de la aproximación al tiempo  $k$ .

- La curva de color azul claro es el resultado de restar el lado derecho al lado izquierdo (diferencia absoluta).
- La curva roja es la diferencia relativa entre los dos lados de la desigualdad.

Si, bien, el número de iteraciones en  $(e)$  crece (por el peor condicionamiento de la matriz), se puede notar el mismo comportamiento en los datos. Tanto los errores de aproximación como la diferencia absoluta (dos temas muy diferentes) decrecen de manera lineal (en base al logaritmo de los datos). Aún así existe una diferencia mayor entre el lado izquierdo y el lado derecho para la segunda gráfica que para la primera, es decir, la distancia entre la línea verde y la azul es mayor que entre la línea morada y la azul. Esto podría explicarse porque la manera de acotar es diferente en los dos casos, depende de  $\text{cond}(H)$ . Sin embargo, lo que parece más interesante de las dos gráficas es como la diferencia relativa entre lado derecho y lado izquierdo parece estancarse en vez de seguir decreciendo. Al analizar los números que estas gráficas arrojan, se ve la fuerza del condicionamiento de la matriz en ambos casos. Tanto el valor al que converge la diferencia relativa como la rapidez con la que lo hace es diferente en los dos casos. Para la matriz mejor condicionada, esta convergencia aparece casi de inmediato, además de que el valor al que lo hace es menor que en la matriz peor condicionada. Entonces, respondiendo la pregunta de si el lado izquierdo y el derecho son iguales, la respuesta sería que no, pero que, dependiendo de los valores propios de la matriz, se podría hacer un argumento en cuanto a su convergencia.

#### 4. Reconstrucción de la prueba de Leunberger justificando.

**Teorema:** El método de máximo descenso aplicado a cuadráticas estrictamente convexas satisface la siguiente propiedad

$$\|x_{k+1} - x^*\|_H^2 \leq \left[ \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}} \right]^2 \|x_k - x^*\|_H^2, \quad (2)$$

#### Prueba:

Recordando, las cuadráticas que vamos a utilizar en el proyecto son de la forma

$$f(x) = \frac{1}{2}x^T Hx + g^T x$$

en donde  $H$  es una matriz simétrica positiva definida de  $n \times n$  con número de condición conocido; el vector  $g$  es generado en forma aleatoria con la función `randn`.

Y la iteración general del método del gradiente tiene la forma

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k)$$

con

$$\alpha_k = \frac{||\nabla f(x_k)||^2}{\nabla f(x_k)^T H \nabla f(x_k)}$$

Definamos

- $E(x) = \frac{1}{2} (x - x^*)^T H (x - x^*) = \frac{1}{2} x^T H x - x^T H x^* + \frac{1}{2} x^{*T} H x^* = \frac{1}{2} x^T H x + g + \frac{1}{2} x^{*T} H x^* = f(x) + \frac{1}{2} x^{*T} H x^*$
- $g_k = \nabla f(x_k) = H x_k + g = \nabla E(x_k)$

y entonces

$$\alpha_k = \frac{g_k^T g_k}{g_k^T H g_k}$$

Lema 1: El proceso iterativo anterior satisface

$$\frac{E(x_k) - E(x_{k+1})}{E(x_k)} = \left[ 1 - \frac{(g_k^T g_k)^2}{(g_k^T H g_k)(g_k^T H^{-1} g_k)} \right] E(x_k)$$

**Prueba**

$$\begin{aligned} \frac{E(x_k) - E(x_{k+1})}{E(x_k)} &= \frac{\frac{1}{2} (x_k - x^*)^T H (x_k - x^*) - \left[ \frac{1}{2} (x_{k+1} - x^*)^T H (x_{k+1} - x^*) \right]}{E(x_k)} \\ \iff \frac{E(x_k) - E(x_{k+1})}{E(x_k)} &= \frac{\frac{1}{2} [x_k^T H x_k - 2x_k^T H x^* + x^{*T} H x^* - x_{k+1}^T H x_{k+1} + 2x_{k+1}^T H x^* - x^{*T} H x^*]}{E(x_k)} \\ \iff \frac{E(x_k) - E(x_{k+1})}{E(x_k)} &= \frac{1}{2E(x_k)} \left[ x_k^T H x_k - 2x_k^T H x^* - (x_k - \alpha_k g_k)^T H (x_k - \alpha_k g_k) + 2(x_k - \alpha_k g_k)^T H x^* \right] \\ \iff \frac{E(x_k) - E(x_{k+1})}{E(x_k)} &= \frac{[x_k^T H x_k - 2x_k^T H x^* - x_k^T H x_k + 2\alpha_k x_k^T H g_k - \alpha_k^2 g_k^T H g_k + 2x_k^T H x^* - 2\alpha_k g_k^T H x^*]}{2E(x_k)} \\ \iff \frac{E(x_k) - E(x_{k+1})}{E(x_k)} &= \frac{[2\alpha_k x_k^T H g_k - \alpha_k^2 g_k^T H g_k - 2\alpha_k g_k^T H x^*]}{2E(x_k)} \end{aligned}$$



$$\begin{aligned}
&\Longleftrightarrow \frac{E(x_k) - E(x_{k+1})}{E(x_k)} = \frac{[2\alpha_k g_k^T H(x_k - x^*) - \alpha_k^2 g_k^T H g_k]}{2E(x_k)} \\
&\Longleftrightarrow \frac{E(x_k) - E(x_{k+1})}{E(x_k)} = \frac{2\alpha_k g_k^T H(x_k - x^*) - \alpha_k^2 g_k^T H g_k}{(x_k - x^*)^T H(x_k - x^*)} \\
&\Longleftrightarrow \frac{E(x_k) - E(x_{k+1})}{E(x_k)} = \frac{2\alpha_k g_k^T g_k - \alpha_k^2 g_k^T H g_k}{g_k^T H^{-1} g_k} \\
&\Longleftrightarrow \frac{E(x_k) - E(x_{k+1})}{E(x_k)} = \frac{2 \left( \frac{g_k^T g_k}{g_k^T H g_k} \right) g_k^T g_k - \left( \frac{g_k^T g_k}{g_k^T H g_k} \right)^2 g_k^T H g_k}{g_k^T H^{-1} g_k} \\
&\Longleftrightarrow \frac{E(x_k) - E(x_{k+1})}{E(x_k)} = \frac{2 \frac{(g_k^T g_k)^2}{g_k^T H g_k} - \frac{(g_k^T g_k)^2}{g_k^T H g_k}}{g_k^T H^{-1} g_k} \\
&\Longleftrightarrow \frac{E(x_k) - E(x_{k+1})}{E(x_k)} = \frac{(g_k^T g_k)^2}{(g_k^T H g_k)(g_k^T H^{-1} g_k)} \quad \square
\end{aligned}$$

La Desigualdad de Kantorovich:

$$\frac{(x^T x)^2}{(x^T H x)(x^T H^{-1} x)} \geq \frac{4\lambda_{\min}\lambda_{\max}}{(\lambda_{\min} + \lambda_{\max})^2}$$

para toda H matriz simétrica definida positiva y para cualquier vector x

### Prueba

Como A es simétrica, por el Teorema Espectral, existe una matriz ortogonal P y una transformación correspondiente  $y = P^T x$  tal que  $y^T H y = \frac{1}{2} y^T P^T D P^T y = x^T D x = \sum_{i=1}^n \lambda_i x_i^2$  donde  $\lambda_i$  son los vectores propios de H

Por lo tanto,

$$\frac{(x^T x)^2}{(x^T H x)(x^T H^{-1} x)} = \frac{(\sum_{i=1}^n x_i^2)^2}{(\sum_{i=1}^n \lambda_i x_i^2) \left( \sum_{i=1}^n \frac{x_i^2}{\lambda_i} \right)} = \frac{\psi(\xi)}{\phi(\xi)}$$

con  $\xi_i = \frac{x_i^2}{\sum_{i=1}^n x_i^2}$

Notemos que  $0 \leq \xi_i \leq 1$ , ya que  $x_i \geq 0$  y  $\sum_{i=1}^n x_i^2 \geq x_i^2$ .

Además,

$$\sum_{i=1}^n \xi_i = \sum_{i=1}^n \frac{x_i^2}{\sum_{i=1}^n x_i^2} = 1.$$

$\therefore \sum_{i=1}^n \xi_i \lambda_i$  es una combinación convexa de las  $\lambda$ 's y  $\sum_{i=1}^n \frac{\xi_i}{\lambda_i}$  es una combinación convexa de las  $\frac{1}{\lambda}$ 's. (i)

Más aún, sabemos que

$$\lambda_{min} \leq \lambda_i \leq \lambda_{max} \quad (ii)$$

para  $i = 1 \dots n$ , y por lo mismo,

$$\frac{1}{\lambda_{max}} \leq \frac{1}{\lambda_i} \leq \frac{1}{\lambda_{min}}$$

para  $i = 1 \dots n$ .

Por (i) y (ii),  $\lambda_{min} \leq \sum_{i=1}^n \xi_i \lambda_i \leq \lambda_{max}$  y

$$\frac{1}{\lambda_{max}} \leq \sum_{i=1}^n \frac{\xi_i}{\lambda_i} \leq \frac{1}{\lambda_{min}} \quad (iii)$$

Esto quiere decir que para todo valor  $\xi$  definido como lo definimos (combinación convexa),  $\psi(\xi)$  va a pertenecer a un punto de la recta  $\xi_1 \lambda_{min} + \xi_n \lambda_{max}$ ; de la misma manera,  $\phi(\xi)$  será un punto de  $\frac{\xi_1}{\lambda_{min}} + \frac{\xi_n}{\lambda_{max}}$ .

Por continuidad, existe  $\lambda^*$  combinación convexa de  $\lambda_{min} = \lambda_1$  y  $\lambda_{max} = \lambda_n$  tal que

$$\frac{\psi(\xi^*)}{\phi(\xi^*)} \geq \frac{f(\lambda^*)}{g(\lambda^*)}$$

Como  $\xi_1^*, \xi_n^*$  hacen una combinación convexa de  $\lambda_1, \lambda_n$ ;  $\xi_1^* + \xi_n^* = 1$

$$\begin{aligned} \frac{\xi_1^*}{\lambda_1} + \frac{\xi_n^*}{\lambda_n} &= \frac{\lambda_n \xi_1^* + \lambda_1 \xi_n^*}{\lambda_1 \lambda_n} \\ &= \frac{\lambda_n (1 - \xi_n^*) + \lambda_1 (1 - \xi_1^*)}{\lambda_1 \lambda_n} \\ &= \frac{\lambda_n + \lambda_1 - \lambda^*}{\lambda_1 \lambda_n} \\ \therefore \frac{\psi(\xi)}{\phi(\xi)} &\geq \frac{\frac{1}{\lambda^*}}{(\lambda_n + \lambda_1 - \lambda^*)(\lambda_1 \lambda_n)} \end{aligned}$$

Para encontrar el mínimo:

$$\begin{aligned} \frac{d}{d\lambda} \left( \frac{\lambda_1 \lambda_n}{\lambda (\lambda_1 + \lambda_n - \lambda)} \right) &= 0 \\ \iff -\lambda_1 \lambda_n ((\lambda_1 + \lambda_n - \lambda) \lambda)^{-2} (\lambda_1 + \lambda_n - 2\lambda) &= 0 \\ \iff \lambda &= \frac{\lambda_1 + \lambda_n}{2} \end{aligned}$$

Para comprobar que verdaderamente estamos en un mínimo, chocamos la segunda derivada:

$$\frac{d^2}{d\lambda^2} \left( \frac{\lambda_1 \lambda_n}{\lambda (\lambda_1 + \lambda_n - \lambda)} \right) = \lambda_1 \lambda_n \left( \frac{\lambda_1 + \lambda_n}{2} \left( \frac{\lambda_1 + \lambda_n}{2} \right) \right)^{-2} (2) > 0$$

y

$$\therefore \frac{\psi(\xi)}{\phi(\xi)} \geq \frac{2(\lambda_1 \lambda_n)}{(\lambda_1 + \lambda_n) (\lambda_1 + \lambda_n - \frac{\lambda_1 + \lambda_n}{2})} = \frac{4\lambda_1 \lambda_n}{(\lambda_1 + \lambda_n)^2}$$

y de aquí obtenemos:

$$\begin{aligned} \frac{(x^T x)^2}{(x^T H x)(x^T H^{-1} x)} &\geq \frac{4\lambda_{\min} \lambda_{\max}}{(\lambda_{\min} + \lambda_{\max})^2} \\ \iff -\frac{(x^T x)^2}{(x^T H x)(x^T H^{-1} x)} &\leq -\frac{4\lambda_{\min} \lambda_{\max}}{(\lambda_{\min} + \lambda_{\max})^2} \\ \iff \left[ 1 - \frac{(x^T x)^2}{(x^T H x)(x^T H^{-1} x)} \right] &\leq \left[ 1 - \frac{4\lambda_{\min} \lambda_{\max}}{(\lambda_{\min} + \lambda_{\max})^2} \right] \\ \iff \left[ 1 - \frac{(x^T x)^2}{(x^T H x)(x^T H^{-1} x)} \right] E(x_k) &\leq \left[ 1 - \frac{4\lambda_{\min} \lambda_{\max}}{(\lambda_{\min} + \lambda_{\max})^2} \right] E(x_k) \\ \therefore E(x_{k+1}) &\leq \left[ 1 - \frac{4\lambda_{\min} \lambda_{\max}}{(\lambda_{\min} + \lambda_{\max})^2} \right] E(x_k) \\ \iff E(x_{k+1}) &\leq \left[ \frac{\lambda_{\min} - \lambda_{\max}}{\lambda_{\min} + \lambda_{\max}} \right]^2 E(x_k) \\ \therefore \|x_{k+1} - x^*\|_H^2 &\leq \left[ \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}} \right]^2 \|x_k - x^*\|_H^2 \quad \square \end{aligned}$$

## Tablas

Table 1:  $N_o$ . condición =  $10^1$

(a)

$k$	$f_k$	$\ \nabla f_k\ $	$\alpha_k$
1	-7.1720741	1.21E+00	1.53E+00
2	-8.6990872	9.78E-01	2.08E+00
4	-9.9450389	6.24E-01	2.09E+00
30	-10.873992	3.02E-03	2.09E+00

resultados 1<sub>a</sub> ejecución

$k_{final}$	$f_{final}$	$  \nabla f_{final}  $
103	-10.874013	8.40E-10

Table 2:  $N_o$ . condición =  $10^2$

(b)

$k$	$f_k$	$  \nabla f_k  $	$\alpha_k$
1	-10.085693	1.94E+00	2.06E+00
2	-13.546051	1.62E+00	1.84E+00
3	-16.059326	1.47E+00	1.92E+00
4	-18.111677	1.34E+00	1.90E+00
6	-21.447377	1.20E+00	1.93E+00
8	-24.200737	1.12E+00	1.95E+00
10	-26.610402	1.05E+00	1.95E+00
13	-29.792113	9.95E-01	2.00E+00
17	-33.434125	9.15E-01	2.00E+00
22	-37.224506	8.19E-01	1.96E+00
28	-40.874672	7.26E-01	1.96E+00
37	-44.942445	6.13E-01	2.00E+00
51	-48.967301	4.63E-01	2.00E+00
86	-53.006247	2.28E-01	1.96E+00
604	-54.327843	7.19E-06	1.96E+00

resultados 2<sub>a</sub> ejecución

$k_{final}$	$f_{final}$	$  \nabla f_{final}  $
1048	-54.327843	1.00E-09

Table 3:  $N_o$  de condición =  $10^3$ 

(c)

$k$	$f_k$	$  \nabla f_k  $	$\alpha_k$	$k$	$f_k$	$  \nabla f_k  $	$\alpha_k$
1	-12.963893	2.56E+00	2.60E+00	150	-221.82464	8.37E-01	1.63E+00
2	-18.11055	1.85E+00	1.57E+00	168	-237.59771	8.06E-01	1.63E+00
3	-22.381152	2.16E+00	2.49E+00	188	-253.77346	7.73E-01	1.63E+00
4	-26.101566	1.63E+00	1.59E+00	210	-270.0583	7.38E-01	1.63E+00
5	-29.465744	1.96E+00	2.53E+00	234	-286.18088	7.01E-01	1.63E+00
6	-32.547435	1.50E+00	1.60E+00	261	-302.46954	8.36E-01	2.59E+00
8	-38.098002	1.41E+00	1.61E+00	290	-318.01565	6.23E-01	1.63E+00
10	-43.042571	1.34E+00	1.61E+00	323	-333.53893	7.34E-01	2.59E+00
12	-47.547999	1.29E+00	1.62E+00	359	-348.18049	6.80E-01	2.59E+00
15	-53.712654	1.55E+00	2.57E+00	400	-362.35829	4.94E-01	1.63E+00
18	-59.370449	1.18E+00	1.62E+00	446	-375.60083	4.48E-01	1.63E+00
22	-66.38432	1.14E+00	1.62E+00	497	-387.56884	5.08E-01	2.59E+00
26	-72.99046	1.11E+00	1.62E+00	555	-398.38835	4.50E-01	2.59E+00
31	-80.867188	1.38E+00	2.59E+00	622	-407.99631	3.09E-01	1.63E+00
36	-88.446741	1.07E+00	1.63E+00	700	-416.24236	2.62E-01	1.63E+00
42	-97.24925	1.05E+00	1.63E+00	791	-422.98734	2.73E-01	2.59E+00
49	-107.18941	1.31E+00	2.59E+00	901	-428.33676	2.17E-01	2.59E+00
56	-116.81781	1.02E+00	1.63E+00	1037	-432.29045	1.63E-01	2.59E+00
64	-127.46887	1.00E+00	1.63E+00	1212	-434.95299	8.90E-02	1.63E+00
73	-139.0248	1.24E+00	2.59E+00	1450	-436.49657	5.38E-02	1.63E+00
83	-151.35858	1.22E+00	2.59E+00	1802	-437.18647	2.56E-02	1.63E+00
94	-164.3372	9.42E-01	1.63E+00	2377	-437.37043	9.61E-03	2.59E+00
106	-177.82486	9.19E-01	1.63E+00	3397	-437.388	1.12E-03	2.59E+00
119	-191.68568	1.13E+00	2.59E+00	5261	-437.38824	2.18E-05	2.59E+00
134	-206.76172	8.66E-01	1.63E+00	8678	-437.38824	1.27E-08	1.63E+00

resultados 3<sub>a</sub> ejecución

$k_{final}$	$f_{final}$	$  \nabla f_{final}  $
9884	-437.38824	9.99E-10

Table 4:  $N_o$  de condición =  $10^1$  ,  $g = -H * e$ 

(d)

$k$	$f_k$	$  \nabla f_k  $	$\alpha_k$
1	-3.213243	6.53E-01	1.36E+00
4	-3.5340241	6.72E-02	1.64E+00

resultados 4<sub>a</sub> ejecución para  $x^* = e$

$k_{final}$	$f_{final}$	$  \nabla f_{final}  $
94	-3.5463799	8.90E-10

Table 5:  $N_o$  de condición =  $10^3$  ,  $g = -H * e$  (e)

$k$	$f_k$	$  \nabla f_k  $	$\alpha_k$
1	-1.3115673	8.35E-01	1.44E+00
2	-1.8557898	4.58E-01	1.56E+00
4	-2.096884	2.02E-01	1.71E+00
401	-2.1802517	5.58E-04	1.98E+00
3443	-2.1803304	1.27E-06	1.98E+00

resultados 5<sub>a</sub> ejecución para  $x^* = e$

$k_{final}$	$f_{final}$	$  \nabla f_{final}  $
7017	-2.1803304	9.98E-10