UNIVERSIDAD SIMÓN BOLÍVAR
CO5412: Optimización No Lineal I
Enero-Marzo 2011
Enrique Areyán

Tarea 1
Sartenejas, 26 de Enero de 2011

1. *a)* Cálculo del gradiente:

Para hallar el gradiente podemos consider la función $c(t) = f(x+td)$ y usar el siguiente hecho:

$$\left.\frac{dc(t)}{dt}\right|_{t=0} = \nabla f(x+td)^t d\Big|_{t=0} = \nabla f(x)^t d$$

Con $t \in \mathbb{R}$ y $d \in \mathbb{R}^n$. Luego,

$c(t) = f(x+td) = \frac{1}{2}[(x+td)^t A(x+td)] + b^t(x+td) + c =$

$\frac{1}{2}[(x^t + td^t)(Ax + tAd)] + b^t x + tb^t d + c =$

$\frac{1}{2}(x^t Ax + tx^t Ad + td^t Ax + t^2 d^t Ad) + b^t x + tb^t d + c.$

$c'(t) = \frac{1}{2}(x^t Ad + d^t Ax + 2td^t Ad) + b^t d.$

$c'(0) = \frac{1}{2}(x^t Ad + d^t Ax) + b^t d = (\frac{1}{2}((A^t x)^t d + (Ax)^t d)) + b^t d = [\frac{1}{2}((A^t x) + (Ax)) + b)^t]d$

$\Rightarrow \nabla f(x) = \frac{1}{2}(A^t x + Ax) + b$

Cálculo del Hessiano:

Para hallar la matriz hessiana también usamos la función $c(t)$ definida anteriormente y el siguiente hecho:

$$\left.\frac{d^2 c(t)}{dt}\right|_{t=0} = d^t \nabla^2 f(x) d$$

Como $c'(t) = \frac{1}{2}(x^t Ad + d^t Ax + 2td^t Ad) + b^t d$, entonces $c''(t) = d^t Ad \Rightarrow \nabla^2 f(x) = A$

b) Cálculo del gradiente:

Sea $\bar{x} \in \mathbb{R}^n$. Para hallar el gradiente podemos consider la función $c(t) = f(x+td)$ y usar el siguiente hecho:

$$\left.\frac{dc(t)}{dt}\right|_{t=0} = \left.\nabla f(x+td)^t d\right|_{t=0} = \nabla f(x)^t d$$

Con $t \in \mathbb{R}$ y $d \in \mathbb{R}^n$. Luego,

$$c(t) = f(x+td) = g(h(x+td))$$

$$c'(t) = \left.h'(x+td)g'(h(x+td))\right|_{t=0} = \nabla f(x)^t d$$

Cálculo del Hessiano:

Para hallar la matriz hessiana también usamos la función $c(t)$ definida anteriormente y el siguiente hecho:

$$\left.\frac{d^2c(t)}{dt}\right|_{t=0} = d^t \nabla^2 f(x) d$$

Como $c'(t) = h'(x+td)g'(h(x+td)$, entonces:

$$c''(t) = h''(x+td)g'(h(x+td)) + h'(x+td)g''(h(x+td))$$

$$\Rightarrow \nabla^2 f(x) d = \left.h''(x+td)g'(h(x+td)) + h'(x+td)g''(h(x+td))\right|_{t=0}$$

2. *a*) El problema puede ser interpretado como conseguir el mínimo valor tal que el vector x operado con una transformación lineal, representada por A, este lo más cercano posible, o se ajuste lo mejor al vector b. En otras palabras se busca minimizar la distancia entre el vector Ax y el vector b.

   b) La condición para la optimalidad la podemos conseguir calculando el gradiente y el hessiano, ya que si x es un mínimo entonces $\nabla f(x) = 0$ y $\nabla^2 f(x)$ será una matriz semi-definida postivia.

   La función f la vamos a trabajar así:

   $$f(x) = ||Ax-b||_2^2 = (Ax-b)^t(Ax-b) = (Ax-b)^t(Ax) - (Ax-b)^t(b) =$$
   $$(Ax)^t Ax - b^t Ax - (Ax)^t b + b^t b = x^t(A^tA)x - 2(Ax)^t b + b^t b$$

   Cálculo del gradiente:

   Para hallar el gradiente podemos consider la función $c(t) = f(x+td)$ usar el siguiente hecho:

   $$\left.\frac{dc(t)}{dt}\right|_{t=0} = \left.\nabla f(x+td)^t d\right|_{t=0} = \nabla f(x)^t d$$

Con $t \in \mathbb{R}$ y $d \in \mathbb{R}^n$. Luego,

$c(t) = f(x+td) = (x+td)^t(A^tA)(x+td) - 2(A(x+td))^tb + b^tb =$

$(x+td)^t(((A^tA)x) + ((A^tA)td)) - 2(Ax + Atd)^tb + b^tb =$

$x^t(A^tA)x + x^t(A^tA)td + td^tA^tAx + t^2d^tA^tAd - 2(Ax)^tb - 2(Atd)^tb + b^tb =$

$c'(t) = x^t(A^tA)d + d^tA^tAx + 2td^tA^tAd - 2(Ad)^tb$

$c'(0) = x^t(A^tA)d + d^tA^tAx - 2(Ad)^tb =$

$x^t(A^tA)d + x^t(A^tA)^td - 2b^t(Ad); A = A^t \Rightarrow c'(0) = 2[(A^tA)x - A^tb]d$

$\Rightarrow \nabla f(x) = 2[(A^tA)x - A^tb]$

La condición de optimalidad dice que $\vec{0} = \nabla f(x) = 2[(A^tA)x - A^tb] \iff A^tAx = A^tb$

<u>Cálculo del Hessiano:</u>

Para hallar la matriz hessiana también usamos la función $c(t)$ definida anteriormente y el siguiente hecho:

$$\left.\frac{d^2c(t)}{dt}\right|_{t=0} = d^t\nabla^2 f(x)d$$

Como $c'(t) = x^t(A^tA)d + d^tA^tAx + 2td^tA^tAd - 2(Ad)^tb$, entonces $c''(t) = 2d^tA^tAd \Rightarrow \nabla^2 f(x) = 2A^tA$

Por lo tanto, un punto $x \in \mathbb{R}^n$ será un minímo del problema si $\nabla f(x) = 2[(A^tA)x - A^tb] = 0$ y A^tA es una matriz S.D.P

$c)$ El problema tendrá solución única si la matriz A tiene rango n, es decir, todos los vectores que componen la matriz son L.I.

$d)$ $A^tAx = A^tb$, suponiendo que $rango(A) = n$

$e)$

$$\begin{pmatrix} 1 & -1 & 0 \\ 0 & 2 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}^t \begin{pmatrix} 1 & -1 & 0 \\ 0 & 2 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & -1 & 1 \\ -1 & 6 & -2 \\ 1 & -2 & 2 \end{pmatrix}$$

$$\begin{pmatrix} 1 & -1 & 0 \\ 0 & 2 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}^t \begin{pmatrix} 2 \\ 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}$$

Debemos resolver:

$$\begin{pmatrix} 2 & -1 & 1 \\ -1 & 6 & -2 \\ 1 & -2 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}$$

Este sistema tiene solución única ya que la matriz es invertible. La solución es $(-\frac{5}{3}, 0, -\frac{4}{3})^t$. Este candidato a mínimo cumple que $\nabla f(-\frac{5}{3}, 0, -\frac{4}{3}) = \vec{0}$. Esta función es estrictamente convexa ya que si inspeccionamos el hessiando:

$$\nabla^2 f(x) = 2 \times \begin{pmatrix} 2 & -1 & 1 \\ -1 & 6 & -2 \\ 1 & -2 & 2 \end{pmatrix}$$

Nos damos cuenta que todos los subdeterminantes de la matriz son mayores que cero. Por lo tanto la matriz es positiva definida y la función es estrictamente convexa. Luego el mínimo hallado es único.

3. 3

4. Sea $f(x, y, z) = 2x^2 + xy + y^2 + yz + z^2 - 6x - 7y - 8z + 9$. Utilizamos las condiciones de optimalidad para hallar el mínimo:

   Cálculo del gradiente:

   Sea $\bar{x} \in \mathbb{R}^3$. Si $\bar{x}$ es un mímino local de f entonces $\nabla f(\bar{x}) = \vec{0}$. Resolvemos:

$$\vec{0} = \nabla f(\bar{x}) = \begin{pmatrix} \frac{\partial f(\bar{x})}{\partial x} \\ \frac{\partial f(\bar{x})}{\partial y} \\ \frac{\partial f(\bar{x})}{\partial z} \end{pmatrix} = \begin{pmatrix} 4x + y - 6 \\ x + 2y + z - 7 \\ y + 2z - 8 \end{pmatrix} \iff \begin{array}{rrr} 4x+ & y & = 6 \\ x+ & 2y+ & z = 7 \\ & y+ & 2z = 8 \end{array}$$

   Debemos resolver el sistema

$$\begin{pmatrix} 4 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 6 \\ 7 \\ 8 \end{pmatrix}$$

   Este sistema tiene solución única, ya que como su determinante es 10 (distinto de cero) la matriz es invertible. Se puede verificar fácilmente que la solución del sistema es:

$$x = y = \frac{6}{5}, z = \frac{17}{5}$$

   Para verificar que el punto es un mínimo global verifiquemos primero si la función es convexa. Para ello examinemos al hessiano.

   Cálculo del Hessiano:

   Note que $\nabla^2 f(\bar{x}) \in M^{3\times 3}$

$$\nabla^2 f(\bar{x}) = \frac{\partial}{\partial x_i}(\frac{\partial g(\bar{x})}{\partial x_j}) = \begin{pmatrix} 4 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}$$

Para verificar si esta matriz es P.S.D o P.D, inspeccionemos todos sus determinantes. Es facil ver que tanto el determinante de la matriz $M^{1\times1} = 4$ como el de la matriz $M^{2\times2} = 7$ son ambos positivos mayores que cero. Verifiquemos el determinante que falta:

$$|\nabla^2 f(\bar{x})| = \begin{vmatrix} 4 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{vmatrix} = 4 \begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix} - \begin{vmatrix} 1 & 1 \\ 0 & 2 \end{vmatrix} = 4(3) - 2 = 12 - 2 = 10$$

Como todos los subdeterminantes de la matriz son mayores que cero, la función f es estrictamente convexa. Luego, el punto $(\frac{6}{5}, \frac{6}{5}, \frac{17}{5})$ es mínimo global. También podemos concluir que el mínimo global en este caso es único.

5. Por defínción de convexidad sabemos que se cumple lo siguiente para f:

   $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y); \lambda \in [0,1]$

   Como g es no decreciente se cumple que:

   $g(f(\lambda x + (1-\lambda)y)) \leq g(\lambda f(x) + (1-\lambda)f(y))$ (el argumento de g se puede ver como un par de puntos)

   A su vez como g es convexa:

   $g(f(\lambda x + (1-\lambda)y)) \leq g(\lambda f(x) + (1-\lambda)f(y)) \leq \lambda g(f(x)) + (1-\lambda)g(f(y)) \iff$

   $g(f(\lambda x + (1-\lambda)y)) \leq \lambda g(f(x)) + (1-\lambda)g(f(y))$, por lo tanto $h = g(f(x))$ es convexa

6. ($\Rightarrow$) Por definición sea $\bar{x} \in \mathbb{R}$ un minimizador global de $f(x)$. Entonces:

   $$f(\bar{x}) \leq f(x), \forall x \in \mathbb{R}$$

   Como g es creciente se cumple que:

   $$x_1 \leq x_2 \Rightarrow g(x_1) \leq g(x_2)$$

   En particular si consideramos $x_1 = f(\bar{x})$ y $x_2 = f(x)$, entonces $g(f(\bar{x})) \leq g(f(x))$ $\forall x \in \mathbb{R} \iff \bar{x}$ es mínimo de $g(f(x))$

   ($\Leftarrow$) Por definición sea $\bar{x} \in \mathbb{R}$ un minimizador global de $g(f(x))$. Entonces:

   $$g(f(\bar{x})) \leq g(f(x)), \forall x \in \mathbb{R}$$

   Como g es creciente entonces tiene inversa g^{-1}. Aplicamos la inversa a ambos lados de la desigualdad anterior y queda

   $$g^{-1}(g(f(\bar{x}))) \leq g^{-1}(g(f(x))), \forall x \in \mathbb{R} \Rightarrow f(\bar{x}) \leq f(x)$$

   Entonces $\bar{x}$ también es un minimizador de $f(x)$. Queda demostrada la equivalencia.

7. Sea $f(x_1, x_2) = 2x_1^3 - 3x_1^2 - 6x_1x_2(x_1 - x_2 - 1) = 2x_1^3 - 3x_1^2 - 6x_1^2x_2 + 6x_1x_2^2 + 6x_1x_2$.

Cálculo del gradiente:

$$\vec{0} = \nabla f(\bar{x}) = \left(\begin{array}{c} \frac{\partial f(\bar{x})}{\partial x_1} \\ \frac{\partial f(\bar{x})}{\partial x_2} \end{array} \right) = \left(\begin{array}{c} 6x_1^2 - 6x_1 - 12x_1x_2 + 6x_2^2 + 6x_2 \\ -6x_1^2 + 12x_1x_2 + 6x_1 \end{array} \right) \iff \begin{array}{l} 6(x_1^2 - x_1 - 2x_1x_2 + x_2^2 + x_2) = 0 \\ -6(x_1^2 - 2x_1x_2 - x_1) = 0 \end{array}$$

Por lo tanto debemos resolver estas dos ecuaciones:

$$\begin{array}{l} x_1^2 - x_1 - 2x_1x_2 + x_2^2 + x_2 = 0 \\ x_1^2 - 2x_1x_2 - x_1 = 0 \qquad \iff x_1(x_1 - 2x_2 - 1) = 0 \iff x_1 = 0; x_1 - 2x_2 - 1 = 0 \end{array}$$

Despejamos x_2 de la última ecuación $x_2 = \frac{x_1-1}{2}$ (ec. *) y la reemplazamos en la primera:

$$\begin{array}{l} x_1^2 - x_1 - 2x_1(\frac{x_1-1}{2}) + (\frac{x_1-1}{2})^2 + (\frac{x_1-1}{2}) = \\ x_1^2 - x_1 - x_1^2 + x_1 + \frac{x_1^2-2x_1+1}{4} + \frac{x_1-1}{2} = \\ \frac{x_1^2-2x_1+1+2x_1-2}{4} = \\ \frac{x_1^2-1}{4} = 0 \Rightarrow x_1^2 - 1 = 0 \iff x_1 = 1 \vee x_1 = -1 \end{array}$$

Reemplazamos estas soluciones en (ec. *) y obtenemos los siguientes puntos $m_1 = (1,0)$ y $m_2 = (-1,-1)$. Ambos puntos hacen que el gradiente sea cero y por tanto son puntos estacionarios. También al inspeccionar las ecuaciones nos damos cuenta que el punto $(0,0)$ es un punto estacionario ya que se satisface $\nabla f((0,0)) = \vec{0}$. En conclusión tenemos los siguientes tres puntos estacionarios:

$$m_1 = (1,0); m_2 = (-1,-1); m_3 = (0,0)$$

Para determinar si estos puntos son mínimos o máximos locales nos vamos a la condición de segundo orden.

Cálculo del Hessiano:

Note que $\nabla^2 f(\bar{x}) \in M^{2\times 2}$

$$\nabla^2 f(\bar{x}) = \frac{\partial}{\partial x_i}(\frac{\partial g(\bar{x})}{\partial x_j}) = \left(\begin{array}{cc} 12x_1 - 6 - 12x_2 & -12x_1 + 12x_2 + 6 \\ -12x_1 + 12x_2 + 6 & 12x_1 \end{array} \right)$$

Evaluamos los puntos estacionarios m_1, m_2 y m_3 en el hessiano:

Para m_1:

$$\nabla^2 f(1,0) = \left(\begin{array}{cc} 12 & -6 \\ -6 & 12 \end{array} \right)$$

Como la matriz del hessiano evaluada en $m_1 = (1,0)$ es positiva definida, es decir se cumple que $d^t(\nabla^2 f(1,0))d > 0, \forall d \in \mathbb{R}^2$ (ya que todos los subdeterminantes son mayores que cero), entonces este punto es un mínimo local.

Para m_2:

$$\nabla^2 f(-1,-1) = \left(\begin{array}{cc} -6 & 6 \\ 6 & -12 \end{array} \right)$$

La matriz anterior es negativa definida ya que sus autovalores son todos negativos ($-9+3\sqrt{5}$ y $-9-3\sqrt{5}$). Por lo tanto el punto $(-1,-1)$ cumple la condición suficiente para ser un máximo local.

Para m_3:

$$\nabla^2 f(0,0) = \begin{pmatrix} -6 & 6 \\ 6 & 0 \end{pmatrix}$$

Como los autovalores de la matriz son $-3-3\sqrt{5} < 0$ y $-3+3\sqrt{5} > 0$, no podemos decir nada sobre este punto.

8. <u>Nota:</u> Enunciado ambiguo

9. Hacemos el siguiente cambio de variables:

$$y_i = ln(x_i), (e^{y_i} = x_i)$$

Nuestro problema original queda así:

$$e^{(\frac{y_1+y_2+...+y_n}{n})} \leq \frac{e^{y_1+y_2+...+y_n}}{n}$$

Más aún, si consideramos que $y_1+y_2+...+y_n = m$, entonces la inecuación anterior la podemos replantear como:

$$ne^{\frac{m}{n}} \leq e^{y_1} + e^{y_2} + ... + e^{y_n}$$

Ahora consideremos la función $f(y_1, y_2, ..., y_n) = e^{y_1}+e^{y_2}+...+e^{y_n}$. A ésta le podemos eliminar la variable y_n despejando de la ecuación: $y_1+y_2+...+y_n = m \iff y_n = m-y_1-y_2-...-y_{n-1}$.

Finalmente tenemos una nueva función a minimizar $g : \mathbb{R}^{n-1} \mapsto \mathbb{R}$ tal que:

$$g(y_1, y_2, ..., y_{n-1}) = e^{y_1} + e^{y_2} + ... + e^{y_n-1} + e^{m-y_1-y_2-...-y_{n-1}}$$

Si logramos demostrar que el mínimo global de la función g es $ne^{\frac{m}{n}}$, entonces habremos demostrado el problema original (la desigualdad de la media aritmética-geométrica).

Comenzamos calculando las condiciones de optimalidad.

<u>Cálculo del gradiente:</u>

Sea $\bar{x} \in \mathbb{R}^{n-1}$. Si $\bar{x}$ es un mímino local de g entonces $\nabla g(\bar{x}) = \vec{0}$. Resolvemos:

$$\vec{0} = \nabla g(\bar{x}) = \begin{pmatrix} \frac{\partial g(\bar{x})}{\partial y_1} \\ \frac{\partial g(\bar{x})}{\partial y_2} \\ ... \\ \frac{\partial g(\bar{x})}{\partial y_{n-1}} \end{pmatrix} = \begin{pmatrix} e^{y_1} - e^{m-y_1-y_2-...-y_{n-1}} \\ e^{y_2} - e^{m-y_1-y_2-...-y_{n-1}} \\ ... \\ e^{y_{n-1}} - e^{m-y_1-y_2-...-y_{n-1}} \end{pmatrix} \iff \begin{array}{l} e^{y_1} = e^{m-y_1-y_2-...-y_{n-1}} \\ e^{y_2} = e^{m-y_1-y_2-...-y_{n-1}} \\ ... \\ e^{y_{n-1}} = e^{m-y_1-y_2-...-y_{n-1}} \end{array}$$

Tomando logaritmo por ambos lados de las ecuaciones queda que:

$$
\begin{aligned}
&y_1 = m - y_1 - y_2 - ... - y_{n-1} \\
&y_2 = m - y_1 - y_2 - ... - y_{n-1} \\
&... \\
&y_{n-1} = m - y_1 - y_2 - ... - y_{n-1}
\end{aligned}
$$

Sumando todas las ecuaciones:

$$y_1 + y_2 + ... + y_{n-1} = (n-1)m - (n-1)y_1 - (n-1)y_2 - ... - (n-1)y_{n-1} \iff$$

$$y_1 + y_2 + ... + y_{n-1} = (n-1)m - ny_1 - y_1 - ny_2 - y_2 - ... - ny_{n-1} - y_{n-1} \iff$$

Cancelando los y_i,

$$0 = nm - m - ny_1 - ny_2 - ... - ny_{n-1} \iff$$

$$0 = n(m - y_1 - y_2 - ... - y_{n-1}) - m \iff$$

$$\frac{m}{n} = m - y_1 - y_2 - ... - y_{n-1}$$

Por lo tanto, como sabemos que $m - y_1 - y_2 - ... - y_{n-1} = y_i$, con $i = 1, 2, ..., n-1$, concluimos que $y_i = \frac{m}{n}$ y el mínimo candidato es: $(\frac{m}{n}, \frac{m}{n}, ...\frac{m}{n}) \in \mathbb{R}^{n-1}$ Para verificar que efectivamente este candidato es el mínimo global, vamos a calcular el Hessiano (segunda derivada), correspondiente a la condición necesaria de segundo orden.

<u>Cálculo del Hessiano:</u>

Note que $\nabla^2 g(\bar{x}) \in M^{n-1 \times n-1}$

$$\nabla^2 g(\bar{x}) = \frac{\partial}{\partial y_i}(\frac{\partial g(\bar{x})}{\partial y_j}) = \begin{pmatrix} e^{y_1} + e^{m-y_1-y_2-...-y_{n-1}} & -e^{m-y_1-y_2-...-y_{n-1}} & ... & -e^{m-y_1-y_2-...-y_{n-1}} \\ -e^{m-y_1-y_2-...-y_{n-1}} & e^{y_2} + e^{m-y_1-y_2-...-y_{n-1}} & ... & -e^{m-y_1-y_2-...-y_{n-1}} \\ ... & ... & ... & ... \\ -e^{m-y_1-y_2-...-y_{n-1}} & -e^{m-y_1-y_2-...-y_{n-1}} & ... & e^{y_{n-1}} + e^{m-y_1-y_2-...-y_{n-1}} \end{pmatrix}$$

Como todos los subdeterminantes de la matriz hessiana son positivos entonces la matriz es P.S.D, lo que implica que la función es convexa y por lo tanto el candidato a mínimo es un mínimo global.

¿Cuál es el valor mínimo de la función? Para hallarlo simplemente evaluamos el punto que mínimiza la función:

$$g(\frac{m}{n}, \frac{m}{n}, ..., \frac{m}{n}) = e^{\frac{m}{n}} + e^{\frac{m}{n}} + ... + e^{m - \frac{m}{n} - \frac{m}{n} - ... - \frac{m}{n}} = e^{\frac{m}{n}} + e^{\frac{m}{n}} + ... + e^{m - ((n-1)\frac{m}{n})} = e^{\frac{m}{n}} + e^{\frac{m}{n}} + ... + e^{\frac{m}{n}} = ne^{\frac{m}{n}}$$

Efectivamente el mínimo de la función g es el que estábamos buscando. Por lo tanto concluimos que se cumple la desigualdad de la media aritmética-geométrica.