ICS 3153 Optimización Avanzada, Sem 2021-1

Tomás Valenzuela

Tarea 2

Pregunta 1
(a) Sabemos que el sistema $Ax \geq b + \Delta b \quad x \geq 0$ es infactible si y solo si $Ax - v = b + \Delta b \quad x \geq 0 \quad v \geq 0$ es infactible. Escribiendo la condición anterior de forma matricial, obtenemos:

$$\begin{bmatrix} A & -\mathbb{I} \end{bmatrix} \begin{bmatrix} x \\ v \end{bmatrix} = b + \Delta b \qquad \begin{bmatrix} x \\ v \end{bmatrix} \geq 0$$

Luego, por lema de Farkas existe w tal que:

$$\begin{bmatrix} A^t \\ -\mathbb{I} \end{bmatrix} w \geq 0 \qquad \begin{bmatrix} b^t + \Delta b^t \end{bmatrix} w < 0 \qquad \text{es factible}$$

$$\Leftrightarrow \qquad A^t w \geq 0 \qquad \left(b^t + \Delta b^t\right) w < 0 \qquad -w \geq 0$$

$$\Leftrightarrow \qquad \exists u \quad \text{tal que} \quad A^t u \leq 0 \qquad \left(b^t + \Delta b^t\right) w > 0 \qquad u \geq 0$$

Pregunta 2
(a): Notemos que se trata de un problema cuadrático. Por lo tanto, puede ser expresado de la forma:

$$f(x) = \frac{1}{2} x^t Q x + c^t x + d$$

En este caso:

$$Q = \begin{bmatrix} 10 & -1 \\ -1 & 10 \end{bmatrix} \qquad b = \begin{bmatrix} -11 \\ 11 \end{bmatrix} \qquad d = 11$$

La condición necesaria de primer orden es $\nabla f(x) = 0$. En este caso, necesitamos que:

$$Qx + c = 0 \Leftrightarrow \begin{bmatrix} 10x_1 - x_2 - 11 \\ -x_1 + 10x_2 + 11 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
$$\Leftrightarrow \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

(b): Mostraremos que la función es estrictamente convexa. En efecto, notemos que $\nabla^2 f = Q$. Luego podemos obtener los determinantes de sus submatrices principales verificándose que:

$$\left|10\right| > 0 \quad y \quad \begin{vmatrix} 10 & -1 \\ -1 & 10 \end{vmatrix} = 99 > 0$$

Lo anterior comprueba que la matriz Hessiana es definida positiva $\forall x \in \mathbb{R}^2$ lo cual implica convexidad estricta. Así, el punto hallado que satisface la condición necesaria de primer orden es mínimo global del problema.

(c): Al tratarse de un problema cuadrático con Q definida positiva, sabemos que el orden de convergencia es lineal satisfaciendo la relación:

$$f(x^k) - f(x^*) \leq \left[\frac{\kappa(Q) - 1}{\kappa(Q) + 1}\right]^2 (f(x^{k-1}) - f(x^*))$$

Para obtener la tasa, calculamos los valores propios de Q:

$$\begin{aligned} |Q - \lambda \mathbb{I}| &= \begin{vmatrix} 10-\lambda & -1 \\ -1 & 10-\lambda \end{vmatrix} \\ &= 100 - 20\lambda + \lambda^2 - 1 \\ &= \lambda^2 - 20\lambda + 99 \\ &= 0 \Leftrightarrow \lambda_1 = 0 \ \ y \ \ \lambda_2 = 11 \end{aligned}$$

(d): En primer lugar consideremos que $\kappa(Q) = \frac{11}{9} \Rightarrow \delta = \left[\frac{\frac{11}{9}-1}{\frac{11}{9}+1}\right]^2 = 0{,}01$.
Por otra parte, Sabemos que:

$$\begin{aligned} e^k &\leq \delta e^{k-1} \\ &= \delta^2 e^{k-2} \\ &\ \ \vdots \\ &= \delta^k e^0 \end{aligned}$$

De esta forma, si queremos $e^k < \epsilon$, basta pedir que $\lambda^k e^0 < \epsilon$. De esta última expresión podemos despejar k, obteniendo:

$$k \geq \frac{ln(e^0/\epsilon)}{ln(1/\delta)}$$

En este caso tenemos:

$$e^0 = 11 \qquad \epsilon = 10^{-11} \qquad \delta = 0{,}01$$

Reemplazando estos valores en la expresión obtenida para k, se obtiene que el número de iteraciones necesarias es k=6.

Pregunta 3
Notemos que si $h(y) = f(A^{-1}y)$ entonces se obtienen las siguientes igualdades:

$$\nabla h(y) = A^{-1^t} \nabla f(A^{-1}y)$$

$$\nabla^2 h(y) = A^{-1^t} \nabla^2 f(A^{-1}y^k) A^{-1}$$

Mostraremos por inducción que $x^k = A^{-1}y^k, \ \ \forall k \in \mathbb{N}$. En efecto, por enunciado tenemos el caso base $x^0 = A^{-1}y^0$. Supongamos que esto se satisface para $k > 0$, es decir, $x^k = A^{-1}y^k$. Notemos que:

$$\begin{aligned}
A^{-1}y^{k+1} &= A^{-1}\left[y^k - \lambda \left[\nabla^2 h(y)\right]^{-1} \nabla h(y^k)\right] \\
&= A^{-1}\left[y^k - \lambda \left[A^{-1^t} \nabla^2 f(A^{-1}y^k) A^{-1}\right]^{-1} A^{-1^t} \nabla f(A^{-1}y^k)\right] \\
&= A^{-1}y^k - \lambda A^{-1} A \left[\nabla^2 f(A^{-1}y^k)\right]^{-1} A^t A^{-1^t} \nabla f(A^{-1}y^k) \\
&= A^{-1}y^k - \lambda \left[\nabla^2 f(A^{-1}y^k)\right]^{-1} \nabla f(A^{-1}y^k) \\
&= x^k - \lambda \left[\nabla^2 f(x^k)\right]^{-1} \nabla f(x^k) \\
&= x^{k+1}
\end{aligned}$$

Es decir, $x^k = A^{-1}y^k \quad \forall k \in \mathbb{N}$. Esto implica que el método de Newton es invariante bajo transformaciones lineales.

Podemos observar que en general esto no ocurre para el método del gradiente. En efecto, si $x^0 = A^{-1}y^0$ entonces:

$$x^1 = x^0 - \lambda \nabla f(x^0)$$

Mientras que:

$$\begin{aligned}
A^{-1}y^1 &= A^{-1}\left[y^0 - \lambda A^{-1^t} \nabla f(A^{-1}y^0)\right] \\
&= A^{-1}y^0 - \lambda A^{-1} A^{-1^t} \nabla f(A^{-1}y^0) \\
&= x^0 - \lambda A^{-1} A^{-1^t} \nabla f(x^0) \neq x^1
\end{aligned}$$

Pregunta 4
Notemos que el Lagrangeano está dado por:

$$L(x, \lambda) = \sum_j^n \frac{c_j}{x_j} + \lambda \left[\sum_{j=1}^n a_j x_j - b\right]$$

Las condiciones KKT para este problema son:

$$\begin{array}{llcll}
(1) & \frac{\partial L}{\partial x_j} \geq 0 & \Leftrightarrow & -\frac{c_j}{x_j^2} + \lambda a_j \geq 0, & j = 1, ..., m \\
(2) & \frac{\partial L}{\partial x_j} x_j = 0 & \Leftrightarrow & \left[-\frac{c_j}{x_j^2} + \lambda a_j\right] x_j \geq 0, & j = 1, ..., m \\
(3) & \frac{\partial L}{\partial \lambda} = 0 & \Leftrightarrow & \sum_{j=1}^n a_j x_j - b = 0 & \\
(4) & x_j \geq 0 & & & j = 1, ..., m
\end{array}$$

Notamos que no puede existir $j = 1, ..., m$ tal que $x_j = 0$, pues se indefinirían restricciones. Luego, por (2) si consideramos la solución $+x_j$ tenemos:

$$\begin{aligned} -\frac{c_j}{x_j^2} + \lambda a_j = 0 &\Leftrightarrow \frac{c_j}{x_j^2} = \lambda a_j \\ &\Leftrightarrow x_j^2 = \frac{c_j}{\lambda a_j} \\ &\Leftrightarrow x_j = \pm\frac{1}{\sqrt{\lambda}}\sqrt{\frac{c_j}{a_j}} \end{aligned}$$

Luego, por (3) tenemos que:

$$\begin{aligned} \sum_{j=1}^{n} a_j \left[\frac{1}{\sqrt{\lambda}}\sqrt{\frac{c_j}{a_j}}\right] = b &\Leftrightarrow \frac{1}{\sqrt{\lambda}}\sum_{j}^{n}\sqrt{a_j c_j} = b \\ &\Leftrightarrow \sqrt{\lambda} = \frac{\sum_j^n \sqrt{a_j c_j}}{b} \\ &\Leftrightarrow \lambda = \left[\sum_{j=1}^{n}\left(\frac{a_j c_j}{b^2}\right)^{1/2}\right]^2 \end{aligned}$$

A partir del último resultado, podemos notar que la solución $-x_j$ no es factible, pues en este caso se tendría $x_j < 0$ lo que se contradice con la condición (4).
Por otra parte, observemos que:

$$\frac{\partial^2 L}{\partial x_j \partial x_i} = \begin{cases} \frac{c_j}{x_j^3} & si \quad j = i \\ \\ 0 & si \quad j = i \end{cases}$$

Así, la matrix Hessiana es una matriz diagonal de pivotes positivos y, por lo tanto, es definida positiva en el punto encontrado. Por otra parte, las restricciones también son convexas con lo cual la solución encontrada es óptima.

Pregunta 5
Por enunciado tenemos:

$$A\bar{x} \leq b, \qquad A_\beta \bar{x}, \qquad A_\eta \bar{x} < b_\eta$$

Adicionalmente, P es matriz de proyección sobre el $Ker(A_\beta)$, esto implica que $A_\beta P x = 0$

(a): En primer lugar mostraremos la factibilidad de $\bar{x} + \lambda\bar{d}$. En efecto, asumiendo sin pérdida de generalidad que A queda particionada superiormente por las filas β e inferiormente por las correspondientes a η, podremos entonces notar que:

$$
\begin{aligned}
A(\bar{x}+\lambda\bar{d}) &= A\bar{x}+\lambda A\bar{d}\\
&= A\bar{x}-\lambda AP\nabla f(\bar{x})\\
&= \begin{bmatrix} A_\beta\bar{x}\\ A_\eta\bar{x}\end{bmatrix} - \lambda \begin{bmatrix} A_\beta\\ A_\eta\end{bmatrix}\left(\mathbb{I}-A_\beta^t\left[A_\beta A_\beta^t\right]^{-1}A_\beta\right)\nabla f(\bar{x})\\
&= \begin{bmatrix} A_\beta\bar{x}\\ A_\eta\bar{x}\end{bmatrix} - \lambda \begin{bmatrix} A_\beta - A_\beta A_\beta^t\left[A_\beta A_\beta^t\right]^{-1}A_\beta\\ A_\eta - A_\eta A_\beta^t\left[A_\beta A_\beta^t\right]^{-1}A_\beta\end{bmatrix}\nabla f(\bar{x})\\
&= \begin{bmatrix} A_\beta\bar{x}\\ A_\eta\bar{x}\end{bmatrix} - \begin{bmatrix} 0\\ \lambda A_\eta\nabla f(\bar{x})\end{bmatrix}\\
&= \begin{bmatrix} b_\beta\\ A_\eta\bar{x}-\lambda A_\eta\nabla f(\bar{x})\end{bmatrix}
\end{aligned}
$$

Como $A_\eta < b_\eta$, entonces existe $\lambda > 0$ tal que $A_\eta\bar{x}-\lambda A_\eta\nabla f(\bar{x}) = A_\eta(\bar{x}-\lambda\nabla f(\bar{x})) \leq b_\eta$ mostrando así la factibilidad del punto sugerido.

Pregunta 6

En primer lugar resolveremos el problema de forma analítica. Notemos que:

$$
\begin{aligned}
\frac{\partial f}{\partial x_1} &= 400x_1^3 - 400x_1x_2 + 2x_1 - 2\\
\frac{\partial f}{\partial x_2} &= 200(x2 - x_1^2)
\end{aligned}
$$

Luego,

$$
\begin{aligned}
\nabla f(x) = 0 \Leftrightarrow \begin{bmatrix} 400x_1^3 - 400x_1x_2 + 2x_1 - 2\\ 200(x_2 - x_1^2)\end{bmatrix} &= \begin{bmatrix}0\\0\end{bmatrix}\\
\Leftrightarrow (x_1, x_2) &= (1,1)
\end{aligned}
$$

El punto recién encontrado satisface la condición necesaria de primer orden, sin embargo, debemos verificar que cumpla las de segundo orden. Para esto notamos que:

$$
H(x) = \begin{bmatrix} 1200x_1^2 - 400x_2 + 2 & -400x_1\\ -400x_1 & 200\end{bmatrix}
$$

Si evaluamos el Hessiano en el punto encontrado verificamos que:

$$
H\left(\begin{bmatrix}1\\1\end{bmatrix}\right) = \begin{bmatrix} 802 & -400\\ -400 & 200\end{bmatrix}
$$

Que es una matriz definida positiva, pues sus submatrices poseen determinantes positivos. Así, el punto propuesto es solución analítica al problema.

Ahora bien, el problema anterior fue resuelto en forma computacional mediante el método del gradiente ocupando linesearch clásico y el linesearch bajo condiciones de Wolfe-Armijo con $\alpha = 0{,}9$. Adicionalmente se utilizó el método de Newton. Las convergencias del error para estos casos pueden observarse en los siguientes gráficos:

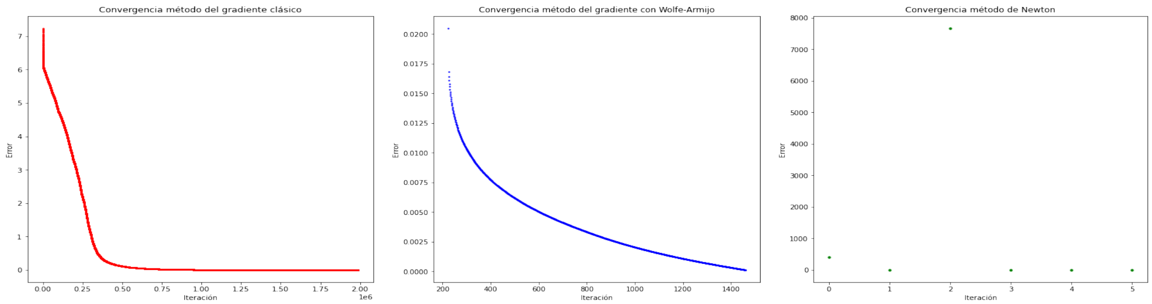

Figura 1: Gráficos error convergencia para los distintos algoritmos

Por otra parte, se presenta a continuación una tabla comparativa en el comportamiento de estos tres algoritmos aplicado al problema:

Modelo	**Gradiente clásico**	**Gradiente Wolfe-Armijo**	**Newton**
Tiempo Convergencia (seg)	3228	0.523	0.031
Número Iteraciones	1985112	1460	5
Valor función Objetivo	$1{,}2 \cdot 10^{-4}$	$1{,}2 \cdot 10^{-4}$	$1{,}809 \cdot 10^{-19}$
Norma del gradiente	$9{,}9 \cdot 10^{-3}$	$9{,}9 \cdot 10^{-3}$	$8{,}507 \cdot 10^{-10}$

Se puede observar que el método del gradiente con linesearch clásico es muy ineficiente para este problema, demorándose mucho más tiempo e iteraciones que los otros dos. Por otra parte, el método del gradiente usando condiciones de Wolfe Armijo tuvo una muy buena convergencia respecto al gradiente clásico, esto puede deberse a que es menos greedy en la búsqueda de la solución, característica positiva en este caso, pues el alto condicionamiento del Hessiano de la función objetivo afecta notablemente la convergencia utilizando un linesearch clásico, pues los pasos ortogonales del algoritmo lo llevan a una alta ineficiencia.
Por otra parte, el método de Newton funcionó más rápido aún en cuanto a tiempo y el número de iteraciones requeridas fue muy bajo. Esto se debe a que este método trabaja con información adicional (segundas derivadas) que permite dar pasos no ortogonales y más directos hacia el óptimo, aunque el costo computacional por iteración es más elevado.

El código correspondiente a los algoritmos fue desarrollado en el entorno Jupyter de Python. Puede ser encontrado en el link Código algoritmos de descenso que lo va a redirigir al repositorio Github donde se encuentra la tarea, aprovechando las bondades de este y de jupyter notebook, que les permitirá visualizar no solo el código sino que la ejecución de este, evitando así adjuntar imágenes de la ejecución para mantener el orden del informe.