Universidad de la República
Facultad de Ingeniería

Teoría y Algoritmia de Optimización

Año 2023

Entregable 2

JUAN VARELA

Autor:

· Juan Manuel Varela

Docentes:
Ignacio Ramírez
Matías Valdes

2 de octubre de 2023

i) He leído y estoy de acuerdo con las Instrucciones especificadas en la carátula obligatorio. ii) He resuelto por mi propia cuenta los ejercicios, sin recurrir a informes de otros compañeros, o soluciones existentes. iii) Soy el único autor de este trabajo. El informe y todo programa implementado como parte de la resolución del obligatorio son de mi autoría y no incluyen partes ni fragmentos tomados de otros informes u otras fuentes, salvo las excepciones mencionadas.

Ejercicio 1

a) Se tiene $f(\mathbf{x}) = ||\mathbf{x}||^2$ y $h(\mathbf{x}) = \sum_i x_i - 1$:

Entonces $L(\mathbf{x}, \lambda) = ||\mathbf{x}||^2 + \lambda(\sum_i x_i - 1)$

Se plantea la condición de optimalidad $\nabla_x L(\mathbf{x}^*, \lambda^*) = \nabla f(\mathbf{x}^*) + \lambda^* \nabla h(\mathbf{x}^*) = 0$

Calculando los gradientes:

$$\nabla f(\mathbf{x}^*) = (2x_1^*, ..., 2x_n^*)$$
$$\nabla h(\mathbf{x}^*) = (1, ..., 1)$$

$$\Rightarrow (2x_1^*, ..., 2x_n^*) + \lambda^*(1, ..., 1) = 0$$

Esto es equivalente a:

$$2x_i^* + \lambda^* = 0$$

Por lo tanto:

$$x_i^* = -\frac{\lambda^*}{2}$$

Ahora se toma la condición $\nabla_\lambda L(\mathbf{x}^*, \lambda^*) = h(\mathbf{x}^*) = 0$:

$$\Rightarrow \sum_i x_i^* - 1 = n(-\lambda^*/2) - 1 = 0 \Rightarrow \lambda^* = -2/n$$

Entonces $\mathbf{x}^* = (1/n, ..., 1/n)$

Se calcula la Hessiana del lagrangeano en $(\mathbf{x}^*, \lambda^*)$:

$$\nabla^2_{xx} L(\mathbf{x}^*, \lambda^*) = \begin{bmatrix} 2 & 0 & \cdots & 0 \\ 0 & 2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 2 \end{bmatrix}$$

Se puede ver que esta matriz es definida positiva por lo que $d^T \nabla^2_{xx} L(\mathbf{x}^*, \lambda^*) d > 0 \ \forall \ d$, en particular para las direcciones factibles. Por lo que el punto hallado cumple las condiciones suficientes, entonces se trata de un mínimo. Al ser el único mínimo, es el mínimo global.

b) Se tiene $f(\mathbf{x}) = \sum_i x_i$ y $h(\mathbf{x}) = ||\mathbf{x}||^2 - 1$:

Entonces $L(\mathbf{x}, \lambda) = \sum_i x_i + \lambda(||\mathbf{x}||^2 - 1)$

Se plantea la condición de optimalidad $\nabla_x L(\mathbf{x}^*, \lambda^*) = \nabla f(\mathbf{x}^*) + \lambda^* \nabla h(\mathbf{x}^*) = 0$

Calculando los gradientes:

$$\nabla f(\mathbf{x}^*) = (1, ..., 1)$$
$$\nabla h(\mathbf{x}^*) = (2x_1^*, ..., 2x_n^*)$$

$$\Rightarrow (1, ..., 1) + \lambda^*(2x_1^*, ..., 2x_n^*) = 0$$

Esto es equivalente a:

$$1 + \lambda^* 2x_i^* = 0$$

Por lo tanto:

$$x_i^* = -\frac{1}{2\lambda^*}$$

Ahora se toma la condición $\nabla_\lambda L(\mathbf{x}^*, \lambda^*) = h(\mathbf{x}^*) = 0$:

$$\Rightarrow ||\mathbf{x}^*||^2 - 1 = n\frac{1}{4(\lambda^*)^2} - 1 = 0 \Rightarrow \lambda^* = \pm\frac{\sqrt{n}}{2}$$

Entonces los candidatos son $x^* = (-\frac{1}{\sqrt{n}}, ..., -\frac{1}{\sqrt{n}})$ y $x^{*'} = (\frac{1}{\sqrt{n}}, ..., \frac{1}{\sqrt{n}})$
Con $\lambda^* = \frac{\sqrt{n}}{2}$ y $\lambda^{*'} = -\frac{\sqrt{n}}{2}$ respectivamente.

Se calcula la Hessiana del lagrangeano en $(\mathbf{x}^*, \lambda^*)$:

$$\nabla^2_{xx} L(\mathbf{x}^*, \lambda^*) = \begin{bmatrix} \sqrt{n} & 0 & \cdots & 0 \\ 0 & \sqrt{n} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{n} \end{bmatrix}$$

Se puede ver que esta matriz es definida positiva por lo que $d^T \nabla^2_{xx} L(\mathbf{x}^*, \lambda^*) d > 0 \ \forall \ d$, en particular para las direcciones factibles. Por lo que el punto $\mathbf{x}^*$ cumple las condiciones suficientes, entonces se trata de un mínimo.

De forma análoga se ve que la Hessiana del lagrangeano en $(\mathbf{x}^{*'}, \lambda^{*'})$ es definida negativa, por lo que se trata de un máximo.

Por lo tanto el mínimo global se da en $\mathbf{x}^* = (-\frac{1}{\sqrt{n}}, ..., -\frac{1}{\sqrt{n}})$.

c) Se tiene $f(\mathbf{x}) = ||\mathbf{x}||^2$ y $h(\mathbf{x}) = \mathbf{x^T Q x} - 1$:

Entonces $L(\mathbf{x}, \lambda) = ||\mathbf{x}||^2 + \lambda(\mathbf{x^T Q x} - 1)$

Se plantea la condición de optimalidad $\nabla_x L(\mathbf{x}^*, \lambda^*) = \nabla f(\mathbf{x}^*) + \lambda^* \nabla h(\mathbf{x}^*) = 0$

Calculando los gradientes:

$$\nabla f(\mathbf{x}^*) = (2x_1^*, ..., 2x_n^*)$$
$$\nabla h(\mathbf{x}^*) = 2\mathbf{Q}\mathbf{x}^*$$

$$\Rightarrow 2\mathbf{x}^* + \lambda^* 2\mathbf{Q}\mathbf{x}^* = 0$$

Esto es equivalente a:

$$\mathbf{Q}\mathbf{x}^* = -\frac{1}{\lambda^*}\mathbf{x}^*$$

Por lo tanto los candidatos serán los vectores propios de Q.

Si α es el valor propio de $\mathbf{Q}$ asociado al vector propio $\mathbf{x}_\alpha^*$, entonces $\lambda_\alpha^* = -\frac{1}{\alpha}$.
Por lo tanto, tomando la condición $\nabla_\lambda L(\mathbf{x}_\alpha^*, \lambda_\alpha^*) = h(\mathbf{x}_\alpha^*) = 0$:

$$\Rightarrow \mathbf{x}_\alpha^{*\mathbf{T}} \mathbf{Q} \mathbf{x}_\alpha^* - 1 = \alpha \mathbf{x}_\alpha^{*\mathbf{T}} \mathbf{x}_\alpha^* = \alpha ||\mathbf{x}_\alpha^*||^2 - 1 = 0 \Rightarrow ||\mathbf{x}_\alpha^*||^2 = \frac{1}{\alpha}$$

Como todos los valores propios de $\mathbf{Q}$ son positivos, $||\mathbf{x}_\alpha^*||^2$ será la menor posible cuando α sea el valor propio más grande de $\mathbf{Q}$

Entonces se puede concluir que el mínimo global se dará en el punto $\mathbf{x}_\alpha^*$, que es el vector propio asociado al mayor valor propio de $\mathbf{Q}$ (α), escalado de forma tal que $||\mathbf{x}_\alpha^*|| = \frac{1}{\sqrt{\alpha}}$

Ejercicio 2 - Desigualdad geométrica-aritmética

- Primero se hace el cambio de variables sugerido: $y_i = \ln x_i$. Por lo tanto el problema pasa a ser:

$$min_{\mathbf{y}} \sum_i \alpha_i e^{y_i}$$

$$s.a \sum_i \alpha_i y_i = 0$$

Se tiene $f(\mathbf{y}) = \sum_i \alpha_i e^{y_i}$ y $h(\mathbf{y}) = \sum_i \alpha_i y_i$.

Entonces $L(\mathbf{y}, \lambda) = \sum_i \alpha_i e^{y_i} + \lambda(\sum_i \alpha_i y_i)$

Se plantea la condición de optimalidad $\nabla_y L(\mathbf{y}^*, \lambda^*) = \nabla f(\mathbf{y}^*) + \lambda^* \nabla h(\mathbf{y}^*) = 0$

Calculando los gradientes:

$$\nabla f(\mathbf{y}^*) = (\alpha_1 e^{y_1^*}, ..., \alpha_n e^{y_n^*})$$
$$\nabla h(\mathbf{y}^*) = (\alpha_1, ..., \alpha_n)$$

$$\Rightarrow (\alpha_1 e^{y_1^*}, ..., \alpha_n e^{y_n^*}) + \lambda^*(\alpha_1, ..., \alpha_n) = 0$$

Esto es equivalente a:

$$\alpha_i e^{y_i^*} + \lambda^* \alpha_i = 0$$

Por lo tanto:

$$y_i^* = \ln(-\lambda^*)$$

Ahora se toma la condición $\nabla_\lambda L(\mathbf{y}^*, \lambda^*) = h(\mathbf{y}^*) = 0$:

$$\Rightarrow \sum_i \alpha_i y_i^* = \ln(-\lambda^*) \sum_i \alpha_i = \ln(-\lambda^*) = 0 \Rightarrow \lambda^* = -1$$

Entonces $\mathbf{y}^* = (0, ..., 0)$

Se calcula la Hessiana del lagrangeano en $(\mathbf{y}^*, \lambda^*)$:

$$\nabla^2_{yy} L(\mathbf{y}^*, \lambda^*) = \begin{bmatrix} \alpha_1 & 0 & \cdots & 0 \\ 0 & \alpha_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_n \end{bmatrix}$$

Como todos los α_i son positivos, esta matriz es definida positiva por lo que $d^T \nabla^2_{yy} L(\mathbf{y}^*, \lambda^*) d > 0 \ \forall \ d$, en particular para las direcciones factibles. Por lo que el punto hallado cumple las condiciones suficientes, entonces se trata de un mínimo. Al ser el único mínimo, es el mínimo global.

Deshaciendo el cambio de variables se tiene que $x_i^* = e^{y_i^*} \Rightarrow x_i^* = 1$.

Por lo tanto $\mathbf{x}^* = (1, ..., 1)$ es el mínimo global del problema original.

Sustituyendo este $\mathbf{x}^*$ hallado en la función objetivo del problema original, se tiene que $f(\mathbf{x}^*) = \sum_i \alpha_i = 1$ es igual a $h(\mathbf{x}^*) = \prod_i 1^{\alpha_i} = 1$

Ahora, se procede a resolver el problema original pero cambiando la restricción por $\prod_i x_i^{\alpha_i} = a$ con $a > 0$ real.

Se resuelve forma análoga, haciendo el cambio de variables $y_i = \ln x_i$, por lo tanto el problema pasa a ser:

$$min_{\mathbf{y}} \sum_i \alpha_i e^{y_i}$$

$$s.a \sum_i \alpha_i y_i = \ln a$$

Se tiene $f(\mathbf{y}) = \sum_i \alpha_i e^{y_i}$ y $h(\mathbf{y}) = \sum_i \alpha_i y_i - \ln a$
Entonces $L(\mathbf{y}, \lambda) = \sum_i \alpha_i e^{y_i} + \lambda(\sum_i \alpha_i y_i - \ln a)$
Se plantea la condición de optimalidad $\nabla_y L(\mathbf{y}^*, \lambda^*) = \nabla f(\mathbf{y}^*) + \lambda^* \nabla h(\mathbf{y}^*) = 0$

Calculando los gradientes:

$$\nabla f(\mathbf{y}^*) = (\alpha_1 e^{y_1^*}, ..., \alpha_n e^{y_n^*})$$
$$\nabla h(\mathbf{y}^*) = (\alpha_1, ..., \alpha_n)$$

$$\Rightarrow (\alpha_1 e^{y_1^*}, ..., \alpha_n e^{y_n^*}) + \lambda^*(\alpha_1, ..., \alpha_n) = 0$$

Esto es equivalente a:

$$\alpha_i e^{y_i^*} + \lambda^* \alpha_i = 0$$

Por lo tanto:

$$y_i^* = \ln(-\lambda^*)$$

Ahora se toma la condición $\nabla_\lambda L(\mathbf{y}^*, \lambda^*) = h(\mathbf{y}^*) = 0$:

$$\Rightarrow \sum_i \alpha_i y_i^* - \ln a = \ln(-\lambda^*) \sum_i \alpha_i - \ln a = \ln(-\lambda^*) - \ln a = 0 \Rightarrow \lambda^* = -a$$

Entonces $\mathbf{y}^* = (\ln a, ..., \ln a)$
Se calcula la Hessiana del lagrangeano en $(\mathbf{y}^*, \lambda^*)$:

$$\nabla^2_{yy} L(\mathbf{y}^*, \lambda^*) = \begin{bmatrix} \alpha_1 a & 0 & \cdots & 0 \\ 0 & \alpha_2 a & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_n a \end{bmatrix}$$

Como todos los α_i son positivos y a es positivo, esta matriz es definida positiva, por lo que el punto hallado es un mínimo. Al ser el único mínimo es el mínimo global.

Deshaciendo el cambio de variables se tiene que $x_i^* = e^{y_i^*} \Rightarrow x_i^* = a$.

Por lo tanto $\mathbf{x}^* = (a, ..., a)$ es el mínimo global del problema original generalizado.

Sustituyendo este $\mathbf{x}^*$ hallado en la función objetivo del problema original, se puede ver que $f(\mathbf{x}^*) = a \sum_i \alpha_i = a$ es igual a $\prod_i a^{\alpha_i} = a$

En conclusión se tiene que para cualquier valor que tome $\prod_i x_i^{\alpha_i}$, el valor funcional mínimo de $\sum_i \alpha_i x_i$ será igual. Por lo tanto se puede deducir que para cualquier conjunto n de números positivos x_i y α_i tal que $\sum_i \alpha_i = 1$ se cumple que:

$$\prod_i x_i^{\alpha_i} \leq \sum_i \alpha_i x_i$$

La igualdad se cumple cuando todos los x_i son iguales, en cualquier otro caso es una desigualdad estricta

Ejercicio 3

- Se tiene $f(\mathbf{x}) = \sum_{i=1}^m ||\mathbf{x} - \mathbf{a_i}||^2$ y $h(\mathbf{x}) = ||\mathbf{x}||^2 - 1$:

  Entonces $L(\mathbf{x}, \lambda) = \sum_{i=1}^m ||\mathbf{x} - \mathbf{a_i}||^2 + \lambda(||\mathbf{x}||^2 - 1)$

  Se plantea la condición de optimalidad $\nabla_x L(\mathbf{x}^*, \lambda^*) = \nabla f(\mathbf{x}^*) + \lambda^* \nabla h(\mathbf{x}^*) = 0$

  $f(\mathbf{x})$ se puede reescribir como $f(\mathbf{x}) = \sum_{i=1}^m \sum_{j=1}^n (x_j - a_{i,j})^2$. Siendo x_j el j-ésimo elemento del vector $\mathbf{x}$ y $a_{i,j}$ el j-ésimo elemento del vector $\mathbf{a_i}$

  Calculando los gradientes:

$$\nabla f(\mathbf{x}^*) = (\sum_{i=1}^m 2(x_1^* - a_{i,1}), ..., \sum_{i=1}^m 2(x_n^* - a_{i,n}))$$

$$\nabla h(\mathbf{x}^*) = (2x_1^*, ..., 2x_n^*)$$

$$\Rightarrow (\sum_{i=1}^m 2(x_1^* - a_{i,1}), ..., \sum_{i=1}^m 2(x_n^* - a_{i,n})) + \lambda^*(2x_1^*, ..., 2x_n^*) = 0$$

  Esto es equivalente a:

$$2\sum_{i=1}^m (x_j^* - a_{i,j}) + 2\lambda^* x_j^* = 0$$

Por lo tanto:

$$x_j^* = \frac{\sum_{i=1}^m a_{i,j}}{m + \lambda^*}$$

$$\Rightarrow \mathbf{x}^* = \frac{\sum_{i=1}^m \mathbf{a_i}}{m + \lambda^*} = \frac{m\hat{\mathbf{a}}}{m + \lambda^*}$$

Ahora se toma la condición $\nabla_\lambda L(\mathbf{x}^*, \lambda^*) = h(\mathbf{x}^*) = 0$:

$$\Rightarrow ||\frac{m\hat{\mathbf{a}}}{m + \lambda^*}||^2 - 1 = (\frac{m}{m + \lambda^*})^2 ||\hat{\mathbf{a}}||^2 - 1 = 0$$

$$\Rightarrow \lambda^{*2} + 2m\lambda^* + m^2(1 - ||\hat{\mathbf{a}}||^2) = 0$$

Esto es una ecuación cuadrática que tiene dos soluciones para λ^*:

$$\lambda_1^* = -m(1 + ||\hat{\mathbf{a}}||^2)$$

$$\lambda_2^* = -m(1 - ||\hat{\mathbf{a}}||^2)$$

Se puede ver que cuando $\hat{\mathbf{a}} = 0$, $\lambda_1^* = \lambda_2^* = -m$. Por lo que sustituyendo este valor en la ecuación anterior para $\mathbf{x}^*$, el denominador se anula. Esto indica que no existe ningún punto que minimice la función objetivo y cumpla con la restricción en este caso, por lo que el problema no tiene solución.

Ahora, considerando el caso que $\hat{\mathbf{a}} \neq 0$, se obtienen dos puntos estacionarios, uno para cada valor de λ hallado.

Se calcula la Hessiana para $\mathbf{x_1^*} = -\frac{\hat{\mathbf{a}}}{||\hat{\mathbf{a}}||}$ y $\lambda_1^* = -m(1 + ||\hat{\mathbf{a}}||^2)$.

$$\nabla_{xx}^2 L(\mathbf{x_1^*}, \lambda_1^*) = \begin{bmatrix} -2||\hat{\mathbf{a}}|| & 0 & \cdots & 0 \\ 0 & -2||\hat{\mathbf{a}}|| & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -2||\hat{\mathbf{a}}|| \end{bmatrix}$$

Como $||\hat{\mathbf{a}}|| > 0$, esta matriz es definida negativa, por lo que $\mathbf{x_1^*}$ se trata de un máximo.

De igual forma, se calcula la Hessiana para $\mathbf{x_2^*} = \frac{\hat{\mathbf{a}}}{||\hat{\mathbf{a}}||}$ y $\lambda_2^* = -m(1 - ||\hat{\mathbf{a}}||^2)$.

$$\nabla_{xx}^2 L(\mathbf{x_1^*}, \lambda_1^*) = \begin{bmatrix} 2||\hat{\mathbf{a}}|| & 0 & \cdots & 0 \\ 0 & 2||\hat{\mathbf{a}}|| & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 2||\hat{\mathbf{a}}|| \end{bmatrix}$$

Esta matriz es definida positiva, por lo que $\mathbf{x_2^*}$ se trata de un mínimo.

De esta manera queda probado que si $\hat{\mathbf{a}} \neq 0$ el problema tiene un máximo y un mínimo únicos. Mientras que si $\hat{\mathbf{a}} = 0$ el problema no tiene solución.

La explicación de que no tenga solución es que cuando $\hat{\mathbf{a}} = 0$ todos los puntos que cumplen con la restricción tienen el mismo valor funcional, por lo tanto se podría interpretar que todos son mínimos y máximos al mismo tiempo.

Ejercicio 4 - Métodos de consenso

a) Como se tiene la restricción $\mathbf{x} = \mathbf{y} = \mathbf{z}$, se escribe la función objetivo únicamente en función de $\mathbf{x}$:

$$f(\mathbf{x}) = \frac{1}{2}\|\mathbf{Ax} - \boldsymbol{\alpha}\|^2 + \frac{1}{2}\|\mathbf{Bx} - \boldsymbol{\beta}\|^2 + \frac{1}{2}\|\mathbf{Cx} - \boldsymbol{\gamma}\|^2$$

Ahora, para hallar el mínimo $\mathbf{x}^*$, se deriva respecto a $\mathbf{x}$ y se iguala a 0:

$$\mathbf{A^T}(\mathbf{Ax}^* - \boldsymbol{\alpha}) + \mathbf{B^T}(\mathbf{Bx}^* - \boldsymbol{\beta}) + \mathbf{C^T}(\mathbf{Cx}^* - \boldsymbol{\gamma}) = 0$$

Se desarrollan las multiplicaciones:

$$\mathbf{A^T}\mathbf{Ax}^* - \mathbf{A^T}\boldsymbol{\alpha} + \mathbf{B^T}\mathbf{Bx}^* - \mathbf{B^T}\boldsymbol{\beta}) + \mathbf{C^T}\mathbf{Cx}^* - \mathbf{C^T}\boldsymbol{\gamma} = 0$$

Se agrupan los términos:

$$(\mathbf{A^T}\mathbf{A} + \mathbf{B^T}\mathbf{B} + \mathbf{C^T}\mathbf{C})\mathbf{x}^* = \mathbf{A^T}\boldsymbol{\alpha} + \mathbf{B^T}\boldsymbol{\beta} + \mathbf{C^T}\boldsymbol{\gamma}$$

Se despeja $\mathbf{x}^*$:

$$\mathbf{x}^* = (\mathbf{A^T}\mathbf{A} + \mathbf{B^T}\mathbf{B} + \mathbf{C^T}\mathbf{C})^{-1}(\mathbf{A^T}\boldsymbol{\alpha} + \mathbf{B^T}\boldsymbol{\beta} + \mathbf{C^T}\boldsymbol{\gamma})$$

Al tratarse de una función cuadrática y positiva, es claro que $\mathbf{x}^*$ es el mínimo.

Por lo tanto el mínimo del problema original es $(\mathbf{x}, \mathbf{y}, \mathbf{z}) = (\mathbf{x}^*, \mathbf{x}^*, \mathbf{x}^*)$

Sustituyendo las matrices $\mathbf{A}$, $\mathbf{B}$ y $\mathbf{C}$ proporcionadas en la expresión obtenida se llega a:

$$\mathbf{x}^* = \begin{bmatrix} -0{,}45209767 \\ -0{,}11645923 \\ -0{,}00416928 \\ -0{,}06615063 \\ 0{,}15325308 \\ 0{,}12561526 \\ -0{,}0518167 \\ -0{,}00062787 \\ -0{,}04588603 \\ 0{,}00636417 \end{bmatrix}$$

Con este valor se compararán los valores obtenidos con los diferentes métodos numéricos.

b) Se tiene $\mathbf{w}$, que es la concatenación de $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$. Este es un vector de dimensión $3n$.

Se quiere reescribir el problema original como:

$$min_{\mathbf{w}} \ \frac{1}{2}||\mathbf{D}\mathbf{w} - \boldsymbol{\delta}||^2 \ s.a \ \mathbf{H}\mathbf{w}$$

Siendo $\boldsymbol{\delta}$ la concatenación de $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$. Este también es un vector de dimensión $3n$.

A partir de esto, se puede ver que $\mathbf{D}$ debe ser una matriz de $3n \times 3n$ que multiplicada por $\mathbf{w}$ resulte en:

$$\begin{bmatrix} \mathbf{Ax} \\ \mathbf{By} \\ \mathbf{Cz} \end{bmatrix}$$

Por lo tanto:

$$\mathbf{D} = \begin{bmatrix} \mathbf{A} & 0 & 0 \\ 0 & \mathbf{B} & 0 \\ 0 & 0 & \mathbf{C} \end{bmatrix}$$

Para las restricciones se necesita una matriz $\mathbf{H}$ tal que $\mathbf{Hw} = \mathbf{0}$ las represente.

$\mathbf{x} = \mathbf{y}$ es lo mismo que $\mathbf{x} - \mathbf{y} = 0$. De igual forma $\mathbf{y} = \mathbf{z}$ es lo mismo que $\mathbf{y} - \mathbf{z} = 0$.

A partir de esto, se puede ver que $\mathbf{H}$ debe ser una matriz de $2n \times 3n$ que multiplicada por $\mathbf{w}$ resulte en:

$$\begin{bmatrix} \mathbf{x} - \mathbf{y} \\ \mathbf{y} - \mathbf{z} \end{bmatrix}$$

Por lo tanto:

$$\mathbf{H} = \begin{bmatrix} \mathbf{I} & -\mathbf{I} & 0 \\ 0 & \mathbf{I} & -\mathbf{I} \end{bmatrix}$$

Siendo $\mathbf{I}$ la matriz identidad de $n \times n$

c) El lagrangeano del problema es el siguiente:

$$L(\mathbf{w}, \boldsymbol{\lambda}) = \frac{1}{2}||\mathbf{Dw} - \boldsymbol{\delta}||^2 + \boldsymbol{\lambda}^{\boldsymbol{T}}\mathbf{Hw}$$

Se plantea la condición de optimalidad $\nabla_w L(\mathbf{w}^*, \boldsymbol{\lambda}^*) = \mathbf{0}$

$$\Rightarrow \mathbf{D^T}(\mathbf{Dw}^* - \boldsymbol{\delta}) + \mathbf{H^T}\boldsymbol{\lambda^*} = \mathbf{0}$$

Desarrollando:

$$\mathbf{D^T}\mathbf{Dw}^* - \mathbf{D^T}\boldsymbol{\delta} + \mathbf{H^T}\boldsymbol{\lambda^*} = \mathbf{0}$$

$$\mathbf{D^T}\mathbf{Dw}^* = \mathbf{D^T}\boldsymbol{\delta} - \mathbf{H^T}\boldsymbol{\lambda^*}$$

Despejando $\mathbf{w}^*$ se tiene que:

$$\mathbf{w}^* = (\mathbf{D^T}\mathbf{D})^{-1}(\mathbf{D}^T\boldsymbol{\delta} - \mathbf{H}^T\boldsymbol{\lambda^*})$$

Sustituyendo este valor en la condición $\nabla_\lambda L(\mathbf{w}^*, \boldsymbol{\lambda^*}) = \mathbf{Hw}^* = \mathbf{0}$:

$$\mathbf{H}(\mathbf{D^T}\mathbf{D})^{-1}(\mathbf{D}^T\boldsymbol{\delta} - \mathbf{H}^T\boldsymbol{\lambda^*}) = \mathbf{0}$$

Desarrollando:

$$\mathbf{H}(\mathbf{D^T}\mathbf{D})^{-1}\mathbf{D}^T\boldsymbol{\delta} - \mathbf{H}(\mathbf{D^T}\mathbf{D})^{-1}\mathbf{H}^T\boldsymbol{\lambda^*} = \mathbf{0}$$

$$\mathbf{H}(\mathbf{D^T}\mathbf{D})^{-1}\mathbf{H}^T\boldsymbol{\lambda^*} = \mathbf{H}(\mathbf{D^T}\mathbf{D})^{-1}\mathbf{D}^T\boldsymbol{\delta}$$

Despejando $\boldsymbol{\lambda^*}$ se llega a:

$$\boldsymbol{\lambda^*} = (\mathbf{H}(\mathbf{D^T}\mathbf{D})^{-1}\mathbf{H}^T)^{-1}\mathbf{H}(\mathbf{D^T}\mathbf{D})^{-1}\mathbf{D}^T\boldsymbol{\delta}$$

Sustituyendo en la expresion anterior las matrices formadas con los datos proporcionados se obtiene:

$$\lambda^* = \begin{bmatrix} -35{,}5039017 \\ -58{,}13500458 \\ -49{,}5687729 \\ -6{,}00575292 \\ -20{,}74913018 \\ -25{,}67972695 \\ -39{,}77207172 \\ -42{,}25391356 \\ -37{,}92853762 \\ -69{,}8592178 \\ 43{,}96561715 \\ -16{,}9286349 \\ -36{,}54879524 \\ -20{,}9741974 \\ -24{,}17324522 \\ -90{,}33607098 \\ -151{,}8607936 \\ -25{,}27311201 \\ -111{,}95744081 \\ -24{,}69133404 \end{bmatrix}$$

d) El lagrangeano aumentado es:

$$L_\tau(\mathbf{w}, \boldsymbol{\lambda}) = \frac{1}{2}||\mathbf{Dw} - \boldsymbol{\delta}||^2 + \boldsymbol{\lambda}^{\boldsymbol{T}}\mathbf{Hw} + \frac{\tau}{2}||\mathbf{Hw}||^2$$

Para calcular el $\mathbf{w}$ mínimo, se deriva la expresión anterior y se iguala a $\mathbf{0}$:

$$\mathbf{D^T}(\mathbf{Dw}^* - \boldsymbol{\delta}) + \mathbf{H^T}\boldsymbol{\lambda} + \tau\mathbf{H^T}\mathbf{Hw}^* = \mathbf{0}$$

Desarrollando:

$$\mathbf{D^T}\mathbf{Dw}^* - \mathbf{D^T}\boldsymbol{\delta} + \mathbf{H^T}\boldsymbol{\lambda} + \tau\mathbf{H^T}\mathbf{Hw}^* = \mathbf{0}$$
$$(\mathbf{D^T}\mathbf{D} + \tau\mathbf{H^T}\mathbf{H})\mathbf{w}^* - \mathbf{D^T}\boldsymbol{\delta} + \mathbf{H^T}\boldsymbol{\lambda} = \mathbf{0}$$
$$(\mathbf{D^T}\mathbf{D} + \tau\mathbf{H^T}\mathbf{H})\mathbf{w}^* = \mathbf{D^T}\boldsymbol{\delta} - \mathbf{H^T}\boldsymbol{\lambda}$$

Despejando $\mathbf{w}^*$ se obtiene:

$$\mathbf{w}^* = (\mathbf{D^T}\mathbf{D} + \tau\mathbf{H^T}\mathbf{H})^{-1}(\mathbf{D^T}\boldsymbol{\delta} - \mathbf{H^T}\boldsymbol{\lambda})$$

e) Se resuelve el problema utilizando el método de penalización cuadrática con $\boldsymbol{\lambda} = \boldsymbol{\lambda}^{\boldsymbol{*}}$ y $\tau^k = \tau_0 2^k$

Para esta parte y las siguientes $\tau_0 = 1/||\mathbf{D}||_2 = 0{,}018053248521226337$, $\epsilon_{max} = 1e^{-8}$ y $k_{max} = 100$

El resultado después de 2 iteraciones es:

$$\mathbf{x}^* = \begin{bmatrix} -0{,}45209767 \\ -0{,}11645923 \\ -0{,}00416928 \\ -0{,}06615063 \\ 0{,}15325308 \\ 0{,}12561526 \\ -0{,}0518167 \\ -0{,}00062787 \\ -0{,}04588603 \\ 0{,}00636417 \end{bmatrix}$$

f) Se resuelve el problema utilizando el método de penalización cuadrática con $\boldsymbol{\lambda} = \mathbf{0}$ y $\tau^k = \tau_0 2^k$

El resultado después de 42 iteraciones es:

$$\mathbf{x}^* = \begin{bmatrix} -0{,}45209767 \\ -0{,}11645923 \\ -0{,}00416928 \\ -0{,}06615063 \\ 0{,}15325307 \\ 0{,}12561526 \\ -0{,}0518167 \\ -0{,}00062787 \\ -0{,}04588603 \\ 0{,}00636417 \end{bmatrix}$$

g) Primero, se resuelve el problema utilizando el método de los multiplicadores con $\boldsymbol{\lambda^0} = \mathbf{0}$, $\boldsymbol{\lambda^k} = \boldsymbol{\lambda^{k-1}} + \tau^k \mathbf{Hw}$ y $\tau^k = 10\tau_0$

El resultado después de 100 iteraciones es:

$$\mathbf{x}^* = \begin{bmatrix} 0{,}49428732 \\ -0{,}359564 \\ -0{,}86401806 \\ -0{,}27124434 \\ -0{,}03539417 \\ 0{,}17024413 \\ -0{,}51730138 \\ 0{,}54697099 \\ 0{,}06391845 \\ -0{,}08557211 \end{bmatrix}$$

Luego, se resuelve el problema utilizando el método de los multiplicadores con $\boldsymbol{\lambda^0} = \mathbf{0}$, $\boldsymbol{\lambda^k} = \boldsymbol{\lambda^{k-1}} + \tau^k \mathbf{Hw}$ y $\tau^k = 1000\tau_0$

El resultado después de 100 iteraciones es:

$$\mathbf{x}^* = \begin{bmatrix} -0{,}45180941 \\ -0{,}11718768 \\ -0{,}00445706 \\ -0{,}06611046 \\ 0{,}15273706 \\ 0{,}1260817 \\ -0{,}05162811 \\ -0{,}0011988 \\ -0{,}04811751 \\ 0{,}00630212 \end{bmatrix}$$

h) Por último, se resuelve el problema utilizando el método combinado de penalización cuadrática/ método de los multiplicadores con $\boldsymbol{\lambda^0} = \mathbf{0}$, $\boldsymbol{\lambda^k} = \boldsymbol{\lambda^{k-1}} + \tau^k \mathbf{Hw}$ y $\tau^k = \tau_0 2^k$

Se muestra el código utilizado:

```

    # Multiplicador base
    tau_0 = 1/ np.linalg.norm(D, 2)

    # Maximo de iteraciones
    k_max = 100

    # Condicion de parada (diferencia relativa de entre iterandos)
    epsilon_max = 1e-8

    # Inicializo lambda, w, epsilon y k
    lambda_k = np.zeros(2*n)
    w_prev = np.zeros(3*n)
    epsilon_k = np.inf
```

```
k=0

# Lista para almacenar x_k
x_k_lista5 = []

while k < k_max and epsilon_k > epsilon_max:
    # Calculo tau para esta iteracion
    tau_k = tau_0 * (2**k)

    # Calculo w para esta iteracion
    w_k = np.linalg.inv(D.T @ D + tau_k * H.T @ H) @ (D.T @ delta - H.T @ lambda_k)

    # Calculo la diferencia relativa entre iterandos
    epsilon_k = np.linalg.norm(w_k - w_prev) / np.linalg.norm(w_k)

    # Calculo de x_k como el promedio de las entradas de w_k
    x_k = (w_k[0:10] + w_k[10:20] + w_k[20:30]) / 3
    x_k_lista5.append(x_k)

    # Actualizo valores
    lambda_k += tau_k * H @ w_k
    w_prev = w_k
    k += 1

x_k
```

El resultado después de 23 iteraciones es:

$$\mathbf{x}^* = \begin{bmatrix} -0{,}45209767 \\ -0{,}11645923 \\ -0{,}00416928 \\ -0{,}06615063 \\ 0{,}15325308 \\ 0{,}12561526 \\ -0{,}0518167 \\ -0{,}00062787 \\ -0{,}04588603 \\ 0{,}00636417 \end{bmatrix}$$

En la Figura 1 se presenta una gráfica donde se muestra la convergencia de cada uno de los métodos utilizados.

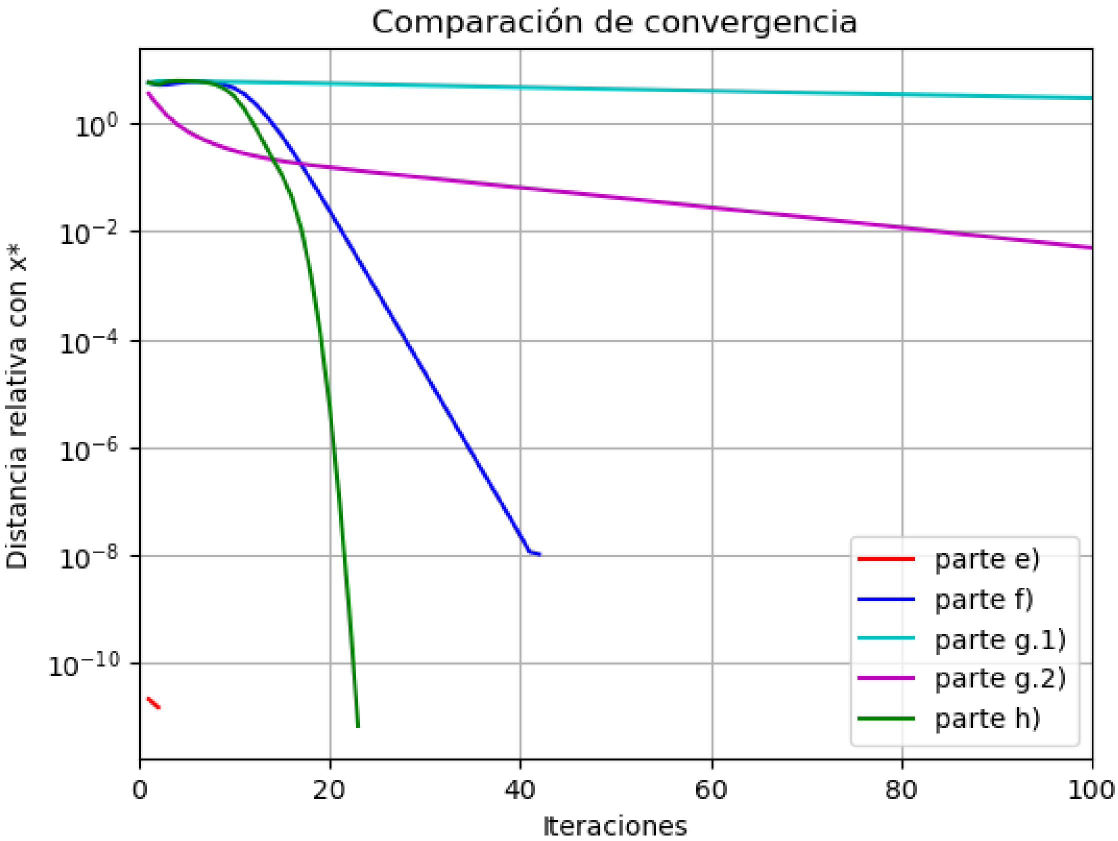

Figura 1: Distancia relativa respecto al óptimo para cada iteración, por método y elección de parámetros

Aquí se puede ver que la resolución de la parte e) es la que más rápido converge a un punto muy cercano al óptimo real, en sólo 2 iteraciones, esto se debe a la gran ventaja que tiene por comenzar utilizando una muy buena (perfecta) estimación de $\boldsymbol{\lambda}^*$.

La resolución de la parte f) converge a un punto muy cercano al óptimo real, pero al no tener una buena estimación de lambda lo hace en más iteraciones que en la parte anterior. Además, si bien es imperceptible al ver los valores, el punto hallado está algo más lejos del óptimo real que el punto hallado en la parte anterior.

Las resoluciónes de la parte g) no llegan a converger y se detienen debido al máximo de iteraciones fijado, por lo tanto el punto hallado es algo lejano al óptimo real en ambos casos. De todas formas, este punto es mucho más cercano al óptimo real cuando se utiliza un τ^k más grande. Esto es un indicio de que este método podría llegar a converger a un buen punto utilizando un τ^k suficientemente grande.

La resolución de la parte h) converge rápidamente a un punto múy cercano al óptimo real (el más cercano de todos). Esto se debe a que utiliza una combinación de ambos métodos (de penalización cuadrática y de los múltiplicadores) por lo tanto, el τ^k va creciendo en cada iteración y además se ajusta el $\boldsymbol{\lambda}^{\boldsymbol{k}}$ teniendo en cuenta qué tan bien se cumplen las restricciones. Esto significa que al converger no solo se llega a un punto muy cercano a $\mathbf{x}^*$ si no que tambien a una muy buena estimación $\boldsymbol{\lambda}^*$.

En conclusión, si por alguna razón se tuviera una buena estimación de $\boldsymbol{\lambda}^*$, el método de penalización cuadrática obtiene resultados excelentes sorprendentemente rápido. Mientras que en cualquier otro caso donde no se tiene esa información, el método que mejor garantiza la convergencia a un buen punto es el método combinado.

El método de penalización cuadrática sin una buena estimación de $\boldsymbol{\lambda}^*$ es válido, pero se debe estar dispuesto a realizar más iteraciones para llegar a la convergencia.

Por último, el método de los multiplicadores sería válido únicamente si se tiene el cuidado de utilizar un τ^k suficientemente grande, aunque es dificil saber a priori qué tan grande.