Soluciones - Obligatorio 1 (Propuesta 1)

Propuesta 1: Un problema cuadrático con restricciones

1. a) Sea $b \in \mathbb{R}^n$ un vector no nulo. Probar que la solución del problema de minimizar $b^t x$ sujeto a $||x|| = 1$ viene dado por $x^* = -b/||b||$ (sugerencia: desigualdad de Cauchy-Schwarz).

Por Cauchy-Schwarz $|\langle b, x\rangle| \leq ||b||$

en particular $\langle b, x\rangle = b^t x \geq -||b||$ (I)

La igualdad en Cauchy-Schwarz se da sii los vectores son colineales, luego la igualdad en (I) solo es posible si $x = \lambda b$ con $\lambda \in \mathbb{R}$. Además $||x|| = |\lambda| . ||b|| = 1 \Rightarrow |\lambda| = \frac{1}{||b||}$,

o sea $\lambda = \frac{\delta}{||b||}$ con $\delta \in \{1, -1\}$ y $x = \lambda b = \frac{\delta . b}{||b||}$

$\Rightarrow b^t x = \delta \frac{b^t b}{||b||} = \delta . ||b||$ y el mínimo se obtiene con $\delta = -1$

$\Rightarrow x^* = \frac{-b}{||b||}$ como queríamos probar.

b) Sea $A \in \mathbb{R}^{n\times n}$ una matriz simétrica y sea $\lambda_0 := \min\{\lambda \in \mathbb{R} : \lambda \text{ es valor propio de } A\}$. Discutir según λ_0 si el problema $\min_{x\in\mathbb{R}^n} x^t Ax$ tiene o no solución, en el caso que tenga solución determine el valor óptimo.

- tenemos $A \in \mathbb{R}^{n\times n}$ simétrica con valor propio mínimo $\lambda_0 \in \mathbb{R}$.
- Si $\lambda_0 \geq 0 \Rightarrow A$ es semidef positiva y $\min_{x\in\mathbb{R}^n} x^t A x = 0$.
- Si $\lambda_0 < 0$, $A v_0 = \lambda_0 . v_0$ con $||v_0|| = 1$, tomando $x = t v_0$ con $t \in \mathbb{R}$ tenemos: $x^t A x = t^2 v_0^t A v_0 = t^2 . \lambda_0 \to -\infty$ si $t \to +\infty$

Luego el mínimo no existe pues $\inf_{x\in\mathbb{R}^n} x^t A x = -\infty$.

2. Sea $A \in \mathbb{R}^{n\times n}$ una matriz simétrica indefinida (como en el notebook) y $B = \{x \in \mathbb{R}^n : ||x||^2 \leq 1\}$.

   a) Implementar el método del gradiente condicional y gradiente proyectado para el problema de optimización con restricciones[1]
   $$\min_{x\in B} x^t A x.$$

   b) Compare ambos métodos en términos de la cantidad de iteraciones necesarias y los tiempos de ejecución. Para el valor óptimo x^* obtenido en la parte anterior, graficar en ambos casos la función de error $e(k) = ||x_k - x^*||$.

Observaciones importantes:

Obs 1. El método de gradiente condicional (Franke-Wolf) implica resolver un subproblema de optimización lineal, que en el ejemplo particular de lo propuesto puede resolverse de forma explícita. Más concretamente, se considera una iteración $x_{k+1} = x_k + \alpha_k d_k$ donde $\alpha_k \in \mathbb{R}$ y $d_k = \bar{x}_k - x_k$ es un vector tal que $\bar{x}_k \in B$ se elige de forma de minimizar $\frac{\partial f}{\partial d_k}(x_k) = \nabla f(x_k)^t \cdot (\bar{x}_k - x_k)$, siendo B el conjunto restricción del problema, en nuestro caso $B = \{x \in \mathbb{R}^n : ||x|| \leq 1\}$.

En nuestro caso específico debemos resolver el subproblema $\min_{\bar{x}\in B} \nabla f(x_k)^t \cdot (\bar{x} - x_k)$, o equivalentemente

$$\min_{\bar{x}\in B} \nabla f(x_k)^t \bar{x} = \min_{\bar{x}\in B} (\underbrace{2Ax_k}_{b})^t \cdot \bar{x}$$

Por la parte 1a, el menor valor es $\bar{x}_k = -\frac{b}{||b||} = -\frac{Ax_k}{||Ax_k||}$

quedando $d_k = -\left(\frac{Ax_k}{||Ax_k||} + x_k\right)$.

Obs2. El método de gradiente condicional construye una secuencia vía recurrencia $x_{k+1} = x_k + \alpha_k d_k$

Pero esta vez la dirección $d_k = P_B(x_k - s\nabla f(x_k)) - x_k$

(no confundir el parámetro $s>0$ que afecta la dirección con el paso fijo α_k)

(Se puede considerar $s=1$ como en el curso y determinar α_k óptimo con un estudio de parámetros)

Era necesario comentar y justificar las elecciones de los parámetros (salvo tomar $s=1$ que fue el caso visto en clase)

3. a) Sea $I \in \mathbb{R}^{n\times n}$ la matriz identidad. Encontrar explícitamente[2] un $R > 0$ (en función de las entradas de la matriz A) tal que $A + \lambda I$ sea semidefinida positiva para todo $\lambda \geq R$.

Como A es simétrica $\Rightarrow A+\lambda I$ también (y por lo tanto todos sus valores propios son reales por el teo. espectral), así que $A+\lambda I$ es semidefinida si y solo si sus valores propios son no negativos.

El teo. de Gershgorin nos asegura que los valores propios de $A+\lambda I$ están contenidos en la unión de los discos

$$D(a_{ii}+\lambda, R_i) \text{ con } R_i = \sum_{j\neq i} |a_{ij}| \quad i=1,\dots,n.$$

Queremos que $a_{ii}+\lambda - R_i \geq 0 \ \forall i \Leftrightarrow \lambda \geq R_i - a_{ii} \ \forall i$

$$\Leftrightarrow \lambda \geq \max\{R_i - a_{ii} : 1\leq i\leq n\} \overset{def}{=} R$$

(si ese máximo fuese negativo podemos tomar cualquier $R>0$, por ejemplo $R=1$)

Obs. $A+\lambda I$ podría ser semidefinida positiva para algunos valores de $\lambda < R$.

Solución alternativa: Usando polinomio característico en lugar de Gershgorin.

$A+\lambda I$ es semidef. positiva $\Leftrightarrow$ todas las raíces del poli. característico $p_{A+\lambda I}(x)$ son no negativas.

Sean $\lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{n-1}$ las raíces de $p_A(x)$ (i.e. los valores propios de A)

Tenemos: $p_{A+\lambda I}(x) = \det(A+\lambda I - xI) = \det(A-(x-\lambda)I) = p_A(x-\lambda)$

Luego $p_{A+\lambda I}(x)=0 \Leftrightarrow p_A(x-\lambda)=0 \Leftrightarrow x-\lambda=\lambda_i$ para algún i

$\Leftrightarrow x=\lambda_i+\lambda$ para algún i.

$\therefore \lambda_0+\lambda \leq \lambda_1+\lambda \leq \dots \leq \lambda_{n-1}+\lambda$ son los val. propios de $A+\lambda I$

Concluimos que $A+\lambda I$ es semidef pos. $\Leftrightarrow \lambda_0+\lambda \geq 0 \Leftrightarrow \lambda \geq -\lambda_0$ donde λ_0 es el menor valor propio de A

Obs. Cuando $\lambda_0<0$ como en nuestro problema entonces $R=-\lambda_0$ es óptimo.

b) Hallar el Lagrangiano del problema $L(x,\lambda)$ y calcule explicitamente[3] $g(\lambda) = \inf_{x\in\mathbb{R}^n} L(x,\lambda)$ para $\lambda \geq 0$. Con la expresión obtenida para $g(\lambda)$ determine numéricamente el óptimo del problema dual. Estime la brecha de dualidad y discuta si es razonable suponer dualidad fuerte.

Problema primal: $\min\limits_{f_1(x)\leq 0} f_0(x)$ con $f_0, f_1: D=\mathbb{R}^n \to \mathbb{R}$ / $f_0(x)=x^tAx$, $f_1(x)=x^tx-1$

El Lagrangiano $L(x,\lambda) = x^tAx + \lambda(x^tx-1) = x^t(A+\lambda I)x - \lambda$

Sea l_λ = menor valor propio de $A+\lambda I$, usando la parte 1b podemos calcular la función dual $g(\lambda) = \inf\limits_{x\in\mathbb{R}^n} x^t(A+\lambda I)x - \lambda = \begin{cases} -\lambda & \text{si } l_\lambda \geq 0 \\ -\infty & \text{si } l_\lambda < 0 \end{cases}$

Dos formas de seguir

(1) Computacionalmente: estimar $\max_{\lambda\geq 0} g(\lambda) = \max_{\lambda\in \mathcal{X}} g(\lambda)$
con $\mathcal{X} = \{\lambda \geq 0 \text{ t.q. } \ell_\lambda \geq 0\}$.

(2) Seguir de forma puramente teórica:
Observar que el polinomio característico de $A+\lambda I$ es
$p_{A+\lambda I}(x) = \det(A+\lambda I - xI) = \det(A-(x-\lambda)I) = p_A(x-\lambda)$
de donde $\ell_\lambda = \lambda_0 + \lambda$ y la función de dualidad resulta

$$g(\lambda) = \begin{cases} -\lambda & \text{si } \lambda \geq -\lambda_0 \\ -\infty & \text{si } \lambda < -\lambda_0 \end{cases}$$

Luego $d^* = \max_{\lambda\geq 0} g(\lambda) = \max_{\substack{\lambda\geq -\lambda_0 \\ \lambda \geq 0}} -\lambda = \begin{cases} 0 & \text{si } \lambda_0 \geq 0 \\ \lambda_0 & \text{si } \lambda_0 < 0 \end{cases}$

Si asumen $\lambda_0<0$ está bien

Finalmente, comparando con la parte 2 deberían concluir (tanto si lo hacen computacionalmente o de forma teórica) que hay dualidad fuerte