Entregable 2: Lagrange

29 de septiembre de 2020

Ejercicio 1 - (Bertsekas 3.1.1)

Utilizando las condiciones de optimalidad de Lagrange para problemas con restricciones de igualdad, resolver:

a) $\text{mín} \|x\|^2$ sujeto a $\sum_i x_i = 1$.

b) $\text{mín} \sum_i x_i$ sujeto a $\|x\|^2 = 1$.

c) $\text{mín} \|x\|^2$ sujeto a $x^T Q x = 1$, donde Q es definida positiva y simétrica. (sugerencia: considere el problema en términos de vectores y valores propios de Q).

Ejercicio 2 - (Bertsekas 3.1.6) Desigualdad geométrica–aritmética

Sean $\alpha_1, \dots, \alpha_n$ escalares positivos con $\sum_i \alpha_i = 1$. Utilice el teorema de Lagrange para resolver el problema

$$\begin{aligned} &\text{mín} \sum_i \alpha_i x_i \\ \text{sujeto a} \quad &\prod_i x_i^{\alpha_i} = 1, \\ &x_i > 0, i = \mathit{0}, \dots, n. \end{aligned}$$

Se sugiere utilizar el cambio de variables $y_i = \ln x_i$ (eso evita la restricción de desigualdad estricta, entre otras cosas). A partir de lo anterior, deduzca que para cualquier conjunto de n números positivos x_i y α_i tales que $\sum_i \alpha_i = 1$ se cumple,

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

Ejercicio 3 - (Bertsekas 3.1.7)

Considere un conjunto de m vectores $a_1, \dots, a_m \in \mathbb{R}^n$; sea $\hat{a}$ su centroide, $\hat{a} = (1/m) \sum_i a_i$. Considere ahora el problema

$$\begin{aligned} &\text{mín} \sum_{i=1}^{m} \|x - a_j\|^2 \\ \text{sujeto a} \quad &\|x\|^2 = 1. \end{aligned}$$

Utilice las condiciones de primer y segundo orden de Lagrange para demostrar que si $\hat{a} \neq 0$, el problema tiene un máximo y un mínimo únicos. ¿Qué pasa si $\hat{a} = 0$?

Ejercicio 4 - Métodos de consenso

Introducción Los métodos de consenso son aquellos que involucran restricciones de igualdad *entre* variables a minimizar. Exigir que dos variables sean iguales en un problema es conceptualmente redundante, pero tiene enormes ventajas prácticas en muchos escenarios, por ejemplo cuando lo que se busca es partir un problema grande en muchos pequeños. Esto es fundamental por ejemplo en técnicas de computación distribuida masivamente. El problema que vamos a resolver en este caso es el siguiente:

$$\min_{x,y,z\in\mathbb{R}^n} \frac{1}{2}\|Ax-\alpha\|_2^2 + \frac{1}{2}\|By-\beta\|_2^2 + \frac{1}{2}\|Cz-\gamma\|_2^2 \quad \text{sujeto a} \quad x=y=z. \tag{1}$$

Tal problema puede surgir por ejemplo en una situación donde α,β,γ son medidas físicas de distinto tipo vinculadas a una misma variable física x $(=y=z)$ que se desea estimar indirectamente en base a las primeras, asumiendo que hay una relación lineal entre ellas y la variable objetivo dada por las matrices A,B y C respectivamente.

Solución exacta El problema (1) se puede resolver de manera analíticamente sin mucha dificultad. (Esto es cierto en muchos casos en los problemas de consenso reales, sólo que los datos involucrados no están accesibles a todos los nodos de procesamiento). En nuestro caso, el cálculo de la solución nos permitirá tener una referencia confiable para evaluar los distintos métodos a implementar.

a) Resuelva el problema (1) analíticamente, calcule y muestre su solución. Esta solución x^* será usada como referencia para evaluar el desempeño de los algoritmos a probar luego.

b) Se define ahora una nueva variable $w=[x,y,z]$, la concatenación de las tres variables a consensuar. El problema (1) puede reescribirse en función de w como

$$\min_{w\in\mathbb{R}^{3n}} \frac{1}{2}\|Dw-\delta\|_2^2 \quad \text{sujeto a} \quad Hw=0. \tag{2}$$

con $\delta=[\alpha,\beta,\gamma]$, la concatenación de los términos independientes. Derive D y H para que el problema (2) sea equivalente a (1).

c) A partir de las condiciones de optimalidad de Lagrange, derive en forma analítica el multiplicador de Lagrange óptimo λ^* para el problema (2).

d) Plantee el Lagrangeano aumentado con $L_\tau(w,\lambda)$ para (2), donde τ es la penalización del término de penalización cuadrático. Calcule analíticamente $\arg\min_w L_\tau(w,\lambda)$ para un λ arbitrario.

Solución numérica En lo que sigue las matrices A,B,C y los vectores α,β,γ serán cargadas de los archivos **A.txt**, **B.txt**, **C.txt**, **alpha.txt**, **beta.txt** y **gamma.txt** respectivamente. En todos los casos se pide implementar los métodos mencionados. La notación w^k indica el valor de w en la k-ésima iteración. Se fijará un máximo de iteraciones $k_{\text{máx}}=100$. Los algoritmos deben detenerse si se alcanza $k=k_{\text{máx}}$ o bien la diferencia relativa entre iterandos,

$$\epsilon^k = \frac{\|w^k-w^{k-1}\|}{\|w^k\|},$$

cae bajo un umbral de $\epsilon_{\text{máx}}=1e^{-8}$. Se define también un valor multiplicador *base* $\tau_0=1/\|D\|_2$ (donde $\|D\|_2$ es la norma ℓ_2 matricial de D; no es la Frobenius). En todos los casos, se calcularán las variables originales x,y,z del problema como el promedio de las tres partes correspondientes de w^k,

$$x^k=y^k=z^k=\frac{w^k_{1:10}+w^k_{11:20}+w^k_{21:23}}{3},$$

donde $w_{i:j}$ es el subvector de w que va entre los índices i y j (ojo, en la mayoría de los lenguajes de programación el primer elemento tiene índice 0 y no incluyen el último índice. Por ejemplo, en Python, el primer subvector sería $w[0:10]$).

e) Resuelva numéricamente el problema mediante el método de penalización cuadrática utilizando $\lambda = \lambda^*$ y $\tau^k = \tau_0 2^k$. (Este no es un caso común en la práctica pero sirve para ver el efecto de tener una buena –en este caso perfecta– estimación de λ).

f) Repita el método anterior utilizando $\lambda = 0$ y $\tau = \tau_0 2^k$.

g) Resolver el problema mediante el *método de los multiplicadores* con $\lambda^0 = 0$ para $\tau^k = 10\tau_0$ y $\tau^k = 1000\tau_0$ respectivamente. Recordar que en este método se actualiza en cada iteración $\lambda^k = \lambda^{k-1} + \tau H w$.

h) Resolver el problema mediante el método combinado de penalización cuadrática/ método de los multiplicadores con $\lambda^0 = 0$, $\lambda^k = \lambda^{k-1} + \tau^k H w$. y $\tau^k = \tau_0 2^k$.

Entregable

El informe debe contener:

1. Todos los desarrollos analíticos de las soluciones exactas
2. La evaluación numérica de la solución exacta para los datos provistos
3. El código fuente de la última parte (λ y τ variables); los otros deberían ser prácticamente idénticos.
4. Comparación gráfica de la convergencia de todas las variantes implementadas. Concretamente, se pide graficar $\|x^k - x^*\|_2 / \|x^*\|_2$ para cada algoritmo y elección de parámetros fijos pedida en la letra (serían 5 curvas en total). Lo ideal es combinar todas las curvas en una misma gráfica y utilizar etiqueta (por ej. `matplotlib.pyplot.label` en Python). De lo contrario, deberá verificarse que *los ejes de todas las gráficas tengan la misma escala y rango.* Se sugiere graficar con eje vertical logarítmico (ej `semilogy` en Python, Matlab) y agregar grillas.
5. **Comentario detallado sobre lo que se observa en las gráficas**. Esto es *obligatorio y excluyente; el no comentar los resultados suficientemente implicará una reducción importante en la nota más allá de los resultados.* Debe comentarse sobre el desempeño de cada método de por sí y comparativamente.
6. Conclusiones. Qué se puede concluir en este caso respecto a cáda método.