Fundamentos de Optimización
Propuestas de proyecto final

Mayo 2023

Propuesta 1: Factorización de matriz simétrica

Sea A una matriz simétrica definida positiva $n \times n$, y d un natural $d < n$. El objetivo es encontrar una solución al problema:

$$\min_{X \in \mathbb{R}^{n \times d}} \|A - XX^T\|_F^2$$

donde $\|\cdot\|_F$ es la norma Frobenius. Este problema es no convexo, pero admite una solución en términos de la descomposición espectral de A, inlcuida en el notebook que acompaña esta propuesta.

1. Verifique que la solución no es única. Sugerencia: dada una solución X, busque soluciones de la forma XC, con C matriz de tamaño y características a definir.
2. Compruebe numéricamente que la función objetivo no es convexa. Sugerencia, tome dos soluciones, basándose en el item anterior.
3. Calcule el gradiente de la función objetivo[1], y verifique numéricamente el resultado.
4. Implemente un método de descenso por gradiente para encontrar un punto crítico de la función. Compare la solución y su valor funcional con la solución dada por la descomposición espectral de A.

Propuesta 2: Proyección a conjuntos convexos

El objetivo de esta propuesta es estudiar un método para hallar la proyección a un conjunto convexo, escrito como intersección de otros dos.

1. Estudiar el algoritmo de proyección alternada del documento *Alternating Projections*, de Boyd y Dattorro. Explicarlo brevemente.
2. Analizar la convergencia del método.
3. Implementar tanto el método del primer item, como la generalización de Dykstra[2]. Elegir la intersección de dos conjuntos convexos para compararlos.
4. Sea $DS = \{M \in \mathbb{R}^{n \times n} : M\mathbf{1} = \mathbf{1}, M^T\mathbf{1} = \mathbf{1}, M_{ij} \geq 0\}$, el conjunto de matrices doblemente estocásticas[3]. Escribir un método que calcule la proyección a DS[4].

[1] Una opción es escribirla como la traza de $(A - XX^T)$ por su traspuesta, y luego usar derivadas de trazas conocidas que hay en el matrix cookbook (`https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf`). Por ejemplo, la (123), poniendo B=C=Identidad, y las (109) y (113).

[2] Hay una descripción en `https://www.di.ens.fr/~aspremon/PDF/MVA/FirstOrderMethodsPartTwo.pdf`, página 48/52.

[3] La definición dice: las filas y columnas de la matriz suman uno ($M\mathbf{1} = \mathbf{1}$, $M^T\mathbf{1} = \mathbf{1}$), y las entradas son positivas.

[4] Sugerencia: escribirlo como intersección del conjunto de matrices de entradas positivas y del conjunto de matrices cuyas filas y columnas suman uno. Se puede utilizar el resultado del Ccorolario 2.2 de `http://www.people.vcu.edu/~rbreams/research/doubstoch1stmom.pdf`.

Propuesta 3: Factorización de matriz con BCD

Sea A una matriz simétrica definida positiva $n \times n$ con ceros en la diagonal, y d un natural $d < n$. El objetivo es encontrar una solución al problema:

$$\min_{X \in \mathbb{R}^{n \times d}} \|M \circ (A - XX^T)\|_F^2$$

donde $\| \cdot \|_F$ es la norma Frobenius, M es una matriz que tiene unos en todas las entradas y ceros en la diagonal, y $\circ$ es el producto de Hadarmard o producto entrada a entrada. Este problema está relacionado con el de la propuesta 1, pero ahora la solución en términos de la descomposición espectral de A (que resuelve el problema de esa propuesta) ya no es una solución, al incluir la matriz M de máscara.

1. Calcule el gradiente de $\|A - XX^T\|_F^2$ y verifique numéricamente que el gradiente de la función objetivo con la máscara corresponde a sustituir $(A - XX^T)$ por $M \circ (A - XX^T)$ en la expresión resultante.

   Consideremos ahora un método de Block Coordinate Descent, donde los bloques serán las filas de X. Así, actualizaremos cada fila de X de forma consecutiva. Llamemos x_i a la fila i-ésima.

2. Corroborar numéricamente que el gradiente de la función objetivo respecto a la fila i es $\nabla_i f(X) = -4A_i X + 4x_i(X^\top X - x_i^\top x_i)$, donde A_i es la fila i de la matriz A. Observar que x_i es un vector fila, por lo que $x_i^\top x_i$ es una matriz $d \times d$.

3. Argumentar que la matriz $R = X^\top X - x_i^\top x_i$ no depende de x_i

4. Implementar el método de Block Coordinate Descent para encontrar un punto crítico del problema planteado[5]. Testearlo con el ejemplo del notebook, y comparar tiempos de ejecución con la solución aproximada calculada con la descomposición espectral de A.

Propuesta 4: Puntos de Fekete

En esta propuesta buscaremos puntos en la esfera que estén "bien distribuidos", de acuerdo a la denominada energía logartímica.

Dados N puntos x_i en la esfera $S^2 = \{x \in \mathbb{R}^3 : \|x\|_2 = 1\}$, definimos la energía logarítmica del conjunto de puntos como

$$E(x_1, x_2, \ldots, x_N) = -\sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} log(\|x_i - x_j\|_2)$$

El objetivo es buscar una configuración de los N puntos que sea un mínimo local de E.

1. Hallar el gradiente de E respecto a un punto x_i. Implementarlo, e implementar también el gradiente respecto a toda la configuración de puntos $x = (x_1, x_2, \ldots, x_N) \in \mathbb{R}^{3N}$.

2. Implementar un método para hallar un mínimo local de E, de la siguiente forma: para cada uno de los N puntos (en forma consecutiva), actualizar la posición de x_i según un método de gradiente proyectado, con un paso fijo, usando como condición de parada que la norma de la diferencia con el iterado anterior sea menor que cierta tolerancia. Repetir el ciclo hasta convergencia (tomar una condición de parada similar, pero a nivel de toda la configuración de puntos)

3. Implementar un método de gradiente proyectado, tomando como variable la configuración de puntos completa $x \in \mathbb{R}^{3N}$, con condición de parada similar que el item anterior.

4. Comparar los métodos en tiempos de ejecución, y valores de la función objetivo alcanzados. Grafcar la configuración de puntos resultante en $\mathbb{R}^3$. Puede superponerlos a la esfera.

[5] Puede ser útil la función `np.outer` para calcular el producto externo $x_i^\top x_i$.