Trabajo final

Teoría y Algoritmia de Optimización

18 de noviembre de 2021

Instrucciones

Objetivo El objetivo del examen es mostrar que el estudiante es capaz de resolver una serie de problemas teóricos e implementar y analizar una serie de problemas prácticos. En el primer caso, es fundamental justificar cualquier paso no trivial de la resolución. En el caso de problemas prácticos, es fundamental analizar y comentar todo resultado que se obtenga.

Contenido El entregable debe contener: resolución detallada de problemas teóricos, resultados de los problemas prácticos, análisis y discusión de ellos. No es necesario (ni aconsejable) incluir: letra de ejercicios, código de los ejercicios prácticos. El código de los ejercicios prácticos debe incluirse en un archivo aparte para posible referencia por parte de los docentes.

Autoría Esta es un examen *individual*. Sus ejercicios deben ser resueltos por el estudiante cuyo nombre, cédula y firma se deben incluir en la carátula del informe. No es admisible la realización colectiva de ninguno de los ejercicios ni sus partes. Tampoco es admisible la búsqueda y/o reutilización, total o parcial, de material en Internet u otros medios, así como entregas disponibles de años anteriores.

Sí es admisible y aconsejable consultar, cotejar, e intercambiar ideas y sugerencias con otros estudiantes.

También es admisible utilizar material de referencia tales como: documentación sobre lenguajes de programación, resultados, definiciones y propiedades matemáticas, incluyendo todo el material expuesto en el teórico de este curso, tanto teórico como práctico.

También es admisible la reproducción e inclusión de recetas y código relacionado con aspectos auxiliares, tales como el graficado de funciones, etc., que no hacen al objetivo de los ejercicios. Se sugiere citar la fuente para evitar posibles confusiones.

Sanciones Cualquier violación a las anteriores reglas constituye una *falta disciplinaria*. En primera instancia, dicha falta implica la pérdida de los puntos del obligatorio en su totalidad. En caso de reincidencia, se desvinculará al estudiante del curso y quedará registrado como reprobado.

Conformidad El examen debe incluir una carátula identificando claramente a) fecha y b) nombre, cédula, y firma (preferentemente digital) del autor y c) el siguiente texto, copiado textualmente:

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

Puntaje El puntaje total de todos los ejercicios del examen es de **150 puntos**. Con **100 puntos** se alcanza la nota máxima de la prueba (que cuenta como un **40 %** del total del curso).

Puntos obtenidos por encima de 100 se tendrán en cuenta y permitirán sumar hasta un **10 % adicional** del total del curso.

Ejercicio 1 - Relajación de problemas binarios (50 puntos)

Considere el siguiente problema:

$$(PB) \quad \min_{x\in\mathbb{R}^n} f(\mathbf{x}) \qquad s.a. \quad x_i^2 = 1, \; i = 1\dots,n$$

donde

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{q}^T\mathbf{x} + r$$

con $\mathbf{Q}$ definida positiva. Sea $\mathbf{x}^*$ la solucion exacta de PB.

a) Escriba el problema dual (DB) correspondiente a (PB) como un problema SDP. Sea ρ^* su valor máximo. Para esta parte se puede usar sin demostrar el resultado en las notas del curso para problemas cuadráticos no convexos.

Considérese la siguiente solución aproximada $\hat{\mathbf{x}}$ del problema (PB):

1. Se obtiene $\bar{\mathbf{x}}$, la solución exacta de la siguiente *relajación convexa* de (PB):

$$(PC) \quad \bar{\mathbf{x}} = \arg\min_{x\in\mathbb{R}^n} f(\mathbf{x}) \qquad s.a. \quad x_i^2 \leq 1, \; i = 1\dots,n$$

2. Hallado $\bar{\mathbf{x}}$, la solución aproximada $\hat{\mathbf{x}}$ se obtiene mediante: $\hat{x}_i \leftarrow \text{sign}(\bar{x}_i), i = 1,\dots,n$.

b) Ordene $f(\mathbf{x}^*)$, ρ^* y $f(\hat{\mathbf{x}})$. Nótese que de este orden puede obtenerse un intervalo en el cual debe caer el valor funcional óptimo de (PB), $f(\mathbf{x}^*)$, y una cota superior para el *gap* resultante de utilizar la solución aproximada $\hat{\mathbf{x}}$ en lugar de la solución exacta $\mathbf{x}^*$.

De aquí en más se pasará a resolver los problemas numéricamente usando la matriz $\mathbf{Q} \in \mathbb{R}^{10\times 10}$, vector $\mathbf{q} \in \mathbb{R}^{10\times 1}$ y escalar $r \in \mathbb{R}$ que se encuentran en los archivos adjuntos `binP`, `binq.txt` y `binr.txt`. Para la resolución de estos problemas puede utilizarse Matlab/CVX, Julia/Convex, o Python/cvxpy.

c) Obtenga la solución aproximada $\hat{\mathbf{x}}$ de (PB) mediante el procedimiento descrito anteriormente.

d) Resuelva (DB), verificando que se respeta el orden obtenido en la parte c).

e) Resuelva (PB) de manera exhaustiva probando todas las $1024 = 2^{10}$ combinaciones posibles de valores binarios y verifique que respeta el orden de la parte c).

Ejercicio 2 - Problemas inversos en estadística (50 puntos)

A menudo es necesario optimizar funciones en donde el argumento es una distribución de probabilidades. Toda distribución válida $\mathbf{p}$ debe cumplir $p_i \geq 0$ y $\sum_i p_i = 1$. El conjunto de distribuciones válidas, $S = \{\mathbf{p} : \sum_i p_i = 1, p_i \geq 0\}$ es conocido como *simplex de probabilidades*. En problemas como los mencionados, es importante tener una forma eficiente de calcular proyecciones sobre S. Concretamente, buscamos resolver:

$$\mathbf{p}^+ = \text{proj}_S(\mathbf{p}) = \arg\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{u} - \mathbf{p}\|_2^2 \quad \text{s.a.} \quad u_i \geq 0, \sum_{i=1}^{\alpha} u_i = 1. \tag{1}$$

a) Plantee las condiciones de KKT para el problema (1) (se sugiere expresar las condiciones de forma canónica, es decir, $-u_i \leq 0$ en lugar de $u_i \geq 0$).

b) Muestre que el orden relativo de los elementos se mantiene luego de la proyección, es decir que $p_i^+ \leq p_j^+$ sii $p_i \leq p_j$ (sugerencia: suponer por absurdo que no se cumple lo anterior y evaluar la función objetivo)

c) Supondremos ahora que los elementos de $\mathbf{p}$ están ordenados en forma *decreciente*. Muestre que

$$\lambda = \frac{1}{k}\left(\sum_{i=1}^{k} p_i - 1\right)$$

para algún $k = 1, \ldots, \alpha$ en donde λ es el multiplicador de Lagrange asociado a la condición de igualdad. Notar que sólo hay α valores posibles de λ, por lo que se puede obtener una solución exacta simplemente probando con cáda valor.

d) A partir del resultado anterior y las condiciones KKT, desarrolle un algoritmo exacto (no iterativo) que calcule la proyección de un vector $\mathbf{p}$ sobre el simplex S.

Sean X e Y dos variables aleatorias que toman valores en el conjunto $A = \{1, 2, \ldots, \alpha\}$. Sean $\mathbf{p}$ y $\mathbf{q}$ las distribuciones de X e Y respectivamente, de modo que $p_i = P(X = i)$ y $q_j = P(Y = j)$. Finalmente, definimos la matriz de transición de probabilidades $\Pi \in [0, 1]^{\alpha\times\alpha}$ tal que $\Pi_{ij} = P(Y = j|X = i)$. Según la fórmula de la probabilidad, se tiene que:

$$q_j = P(Y = j) = \sum_{i=1}^{\alpha} P(Y = j|X = i)P(X = i) = \sum_{i=1}^{\alpha} \Pi_{ij} p_i.$$

Lo anterior puede escribirse de manera compacta como $\mathbf{q} = \Pi^T \mathbf{p}$.

Nuestro objetivo es obtener una estimación $\hat{\mathbf{p}}$ de la dist. de X, a partir de una estimación $\hat{\mathbf{q}}$ de la dist. de Y. Desafortunadamente, en muchos casos de interés, resolver el problema directamente mediante $\hat{\mathbf{p}} = \Pi^{-T}\hat{\mathbf{q}}$ no suele arrojar buenos resultados. Por un lado, la matriz Π puede estar mal condicionada. En segundo lugar, incluso si Π está bien condicionada, los errores de aproximación en $\hat{\mathbf{q}}$ pueden hacer que $\Pi^{-T}\hat{\mathbf{q}}$ no sea una distribución válida. Se plantea entonces calcular $\hat{\mathbf{p}}$ como la solución de siguiente problema:

$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p}\in S} \frac{1}{2}\|\Pi^T\mathbf{p} - \hat{\mathbf{q}}\|_2^2. \tag{2}$$

e) Resuelva numéricamente el problema (2) utilizando algún método visto en el curso (por ejemplo, gradiente proyectado). Los datos para este problema se encuentran en los archivos `statsPi.txt` y `statsq.txt` respectivamente.´

Ejercicio 3 - Regresión lineal con función de Huber (50 puntos)

La función de Huber es un híbrido entre la norma ℓ_1 y la ℓ_2 que combina la diferenciabilidad de la segunda con la robustez frente a *outliers* de la primera. Se define de la siguiente forma:

$$h(x) = \begin{cases} \delta(-x - \frac{\delta}{2}), & x < -\delta \\ \frac{1}{2}x^2, & |x| \leq \delta \\ \delta(x - \frac{\delta}{2}), & x > \delta \end{cases}$$

Aplicado a un vector, la función $h(\mathbf{x})$ es la suma de $h(\cdot)$ escalar aplicada elemento a elemento, es decir: $h(\mathbf{x}) = \sum_i h(x_i)$.

a) Qué puede decir de la función $h(x)$? Es continua? Es diferenciable? Es doblemente diferenciable?

b) Calcule el operador proximal de $h(x)$ (se sugiere trabajar en cada una de las tres regiones de x).

El objetivo ahora es resolver numéricamente el siguiente problema de regresión no lineal:

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} h(\mathbf{A}\mathbf{x} - \mathbf{b}).$$

La matriz $\mathbf{A}$ y el vector $\mathbf{b}$ están almacenados respectivamente en los archivos `huberA.txt` y `huberb.txt`.

c) Resuelva el problema para $\delta = 1$ utilizando descenso por gradiente con el paso seleccionado utilizando el método de Armijo.

d) Calcule la Hessiana H de la función de costo en un punto arbitrario.

e) Resuelva el problema utilizando el siguiente método cuasi-Newton con dirección de descenso $\mathbf{d}^k = \left(\nabla^2 h(\mathbf{x}^k) + 0{,}1\mathbf{I}\right)^{-1} \nabla h(\mathbf{x}^k)$ y paso $s_k = 0{,}1/k$.