Análisis Convexo: Subdiferenciales

Proyecto Final

Roberto Vásquez Martínez

Universidad de Guanajuato
Departamento de Matemáticas

Profesor: Dr. Joaquín Peña Acevedo

10 de junio de 2022

Enero-Junio 2022

Contenidos

Contenidos

Motivación

- Curso de Optimización: Métodos en los cuales se utiliza la derivada y segundas derivadas.
- $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ es la función objetivo y no es diferenciable no podemos utilizar ninguno de estos métodos.

Convexidad

Un caso importante es considerar f una función convexa, donde vimos condiciones necesarias y suficientes para el punto óptimo de f cuando esta función es diferenciable.

Relajando esta condición, analizaremos el caso cuando f es convexa pero no diferenciable.

Convexidad

Un caso importante es considerar f una función convexa, donde vimos condiciones necesarias y suficientes para el punto óptimo de f cuando esta función es diferenciable.
Relajando esta condición, analizaremos el caso cuando f es convexa pero no diferenciable.

Esquema

1. Obtener condiciones de optimalidad para funciones convexas no diferenciables.
2. Generalizar la noción de aproximación lineal y derivada direccional.
3. Construir generalizaciones del método de descenso y ejemplificar el uso de subdiferenciales en la optimización convexa.

Esquema

1. Obtener condiciones de optimalidad para funciones convexas no diferenciables.
2. Generalizar la noción de aproximación lineal y derivada direccional.
3. Construir generalizaciones del método de descenso y ejemplificar el uso de subdiferenciales en la optimización convexa.

Esquema

1. Obtener condiciones de optimalidad para funciones convexas no diferenciables.
2. Generalizar la noción de aproximación lineal y derivada direccional.
3. Construir generalizaciones del método de descenso y ejemplificar el uso de subdiferenciales en la optimización convexa.

Contenidos

Enero-Junio 2022

Para obtener métodos de optimización como los de búsqueda en línea es fundamental la idea de *aproximación lineal*. El tipo de funciones que aproximaran a las funciones convexas serán **sublineales**, en ese sentido tendremos *aproximación sublineal*.

Para obtener métodos de optimización como los de búsqueda en línea es fundamental la idea de *aproximación lineal*. El tipo de funciones que aproximaran a las funciones convexas serán **sublineales**, en ese sentido tendremos *aproximación sublineal*.

Funciones sublineales #1

Definición (Sublinealidad)

Una función $\sigma : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ se dice **sublineal** si se cumple

$$\sigma(t_1 x_1 + t_2 x_2) \leq t_1 \sigma(x_1) + t_2 \sigma(x_2) \quad \forall (x_1, x_2) \in \mathbb{R}^n \times \mathbb{R}^n.$$

y $(t_1, t_2) \in \mathbb{R}^+ \times \mathbb{R}^+$

Funciones sublineales #2

Observamos que el conjunto de funciones sublineales es un subconjunto de la clase de funciones convexas.

Contenidos

Intuición

- Sabemos que forma tienen los funcionales lineales de espacio de Hilbert.
- Las aproximaciones lineales de funciones objetivo f diferenciables son de esta forma.
- El resultado teórico que nos caracteriza a estos funcionales lineales es el *Teorema de Representación de Riesz*.

Motivación

Quisieramos un teorema similar para el caso de funciones pues

1. Brinda intención geométrica.
2. Conocer la forma de estas funciones sublineales es útil en el esquema de optimización.

Funciones sublineales cerradas #1

Sin embargo, no podemos considerar la clase de todas las funciones sublineales. Nos restringimos a la clase de funciones **sublineales cerradas**.

Funciones sublineales cerradas #2

Definición (Epígrafe)

Sea $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ no idénticamente $+\infty$. La epígrafe de f la definimos como el conjunto

$$\operatorname{epi} f = \{(x, r) \in \mathbb{R}^n \times \mathbb{R} \ : \ r \geq f(x)\}.$$

Definición (Función cerrada)

Sea $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ no idénticamente $+\infty$. Decimos que f es cerrada si $\operatorname{epi} f$ es un conjunto cerrado en la topología producto.

Funciones sublineales cerradas #2

Definición (Epígrafe)

Sea $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ no idénticamente $+\infty$. La epígrafe de f la definimos como el conjunto

$$\operatorname{epi} f = \{(x, r) \in \mathbb{R}^n \times \mathbb{R} \;:\; r \geq f(x)\}.$$

Definición (Función cerrada)

Sea $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ no idénticamente $+\infty$. Decimos que f es cerrada si $\operatorname{epi} f$ es un conjunto cerrado en la topología producto.

Función Soporte

El tipo de representación que necesitamos se vale del siguiente concepto.

Definición (Función Soporte)

Sea $S \subset \mathbb{R}^n$ no vacío y la función $\sigma_S : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ definida como

$$\sigma_S(x) = \sup\{\langle s, x\rangle \ : \ s \in S\}.$$

Decimos que σ_S es la función soporte del conjunto S.

Representación

Teorema de Representación

Sea $\sigma : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ función sublineal cerrada y

$$S_\sigma = \{s \in \mathbb{R}^n \ : \ \langle s, d \rangle \leq \sigma(d) \quad \forall d \in \mathbb{R}^n\}.$$

Entonces σ es la función soporte de S_σ.

Así, la representación de funciones sublineales cerradas se reduce al supremo de funcionales sinales que la minorizan.

Contenidos

Preliminares

Al igual que en el caso diferenciable, tenemos la noción de *derivada direccional*, que de manera similar nos permitirá generalizar la aproximación que tenemos en el sentido del Teorema de Taylor y así construir métodos de descenso para este caso.

En primer lugar, dado $x, d \in \mathbb{R}^n$

$$q(t) = \frac{f(x+td) - f(x)}{t} \quad \text{para } t > 0. \tag{1}$$

Definición (Derivada Direccional)

Sean $x, d \in \mathbb{R}^n$ y $f : \mathbb{R}^n \to \mathbb{R}$ función convexa. Entonces la derivada direccional de f en x en la dirección d es

$$f'(x,d) := \lim_{t \downarrow 0} q(t) = \inf\{q(t) \; : \; t > 0\}.$$

Se puede probar que la derivada direccional es una función sublineal cerrada.

Contenidos

Enero-Junio 2022

Subdiferencial I

Definición (Subdiferencial I)

El *subdiferencial* $\partial f(x)$ de f en x es el conjunto no vacío del cual $f'(x, \cdot)$ es función soporte, i.e.

$$\partial f(x) = \{s \in \mathbb{R}^n \ : \ \langle s, d \rangle \leq f'(x, d) \quad \forall d \in \mathbb{R}^n\}.$$

Un vector $s \in \partial f(x)$ decimos que es un *subgradiente* de f en x.

Subdiferencial II

Esta definición carece de facilidad de cómputo. Mostramos una equivalencia.

Definición (Subdiferencial II)

El subdiferencial $\partial f(x)$ es el conjunto de vectores s que satisfacen

$$f(y) \geq f(x) + \langle s, y - x \rangle \quad \forall y \in \mathbb{R}^n.$$

Contenidos

Condiciones de primer orden

A partir de las definiciones anteriores es fácil demostrar lo siguiente

Teorema (Condiciones de optimalidad)

Sea $f : \mathbb{R}^n \to \mathbb{R}$ convexa. Las siguientes condiciones son equivalentes

1. f es minimizada en x_* sobre $\mathbb{R}^n$.
2. $0 \in \partial f(x_*)$.
3. $f'(x_*, d) \geq 0$ para toda $d \in \mathbb{R}^n$.

Contenidos

Descenso Máximo

En el sentido del Teorema anterior podemos buscar la defininción de máximo descenso. La dirección de descenso máximo se conceptualiza de la siguiente forma

Definición (Dirección de Máximo Descenso)

Sea $||\cdot||$ una norma en $\mathbb{R}^n$. Una dirección de máximo descenso normalizada de f en x, con respecto a la norma $||\cdot||$, es una solución al problema de optimización

$$\text{mín}\{f'(x,d)\ :\ ||d|| = 1\}, \tag{2}$$

o de forma equivalente, podemos escribir en notación min-máx este problema de optimización ya que $f'(x,\cdot)$ es función soporte de $\partial f(x)$

$$\min_{||d||=1}\ \max_{s\in\partial f(x)} \langle s, d\rangle,$$

y ese máximo se alcanza gracias a que $\partial f(x)$ es un compacto.

Contenidos

Algoritmo de Descenso Máximo

Algoritmo de Descenso Máximo

Empezamos con $x_1 \in \mathbb{R}^n$. Ajustamos $k = 1$. Repetimos los siguientes pasos:

1. **(Criterio de Paro)** $0 \in \partial f(x_k)$
2. **(Dirección de Descenso)** Para la norma $||\cdot||$ en $\mathbb{R}^n$. Hallar d_k solución de (2)
3. **(Búsqueda en línea)** Encontrar $t_k > 0$ de forma que $x_{k+1} := x_k + t_k d_k$ es tal que

$$f(x_{k+1}) < f(x_k)$$

4. Hacer $k = k + 1$ y volver a (i).

Referencias

R. Tibshirani et al.
Statistical Learning with Sparsity.
Taylor & Francis Group, 2015.

J. Nocedal, S. Wright
Numerical Optimization.
Springer, 2006.

J. Hiriart-Urruty.
Convex Analysis and Minimization Algorithms I: Fundamentals.
Springer, 1993.

R. Tibshirani.
Regression Shrinkage and Selection via the LASSO.
Journal of the Royal Statistical Society. 58(1), 1996.