Sistemas dinámicos y algoritmos en optimización convexa

Juan PEYPOUQUET
Universidad Técnica Federico Santa María

Universidad de La Habana

La Habana, 16 al 20 de noviembre de 2015

CLASE I

Aspectos generales

Funciones en los reales extendidos

Consideraremos funciones definidas en un conjunto X con valores en los reales extendidos $\mathbb{R} \cup \{+\infty\}$ y escribiremos

$$\operatorname{dom}(f) = \{x \in \mathbb{R}^N : f(x) < +\infty\}.$$

Ejemplo

La función indicatriz de un conjunto C es

$$\delta_C(x) = \begin{cases} 0 & si \quad x \in C \\ +\infty & si \quad x \notin C \end{cases}$$

Problemas de optimización

Estudiaremos el problema de existencia, caracterización y aproximación numérica de los puntos donde una función $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ alcanza su menor valor:

El valor del problema es

$$\min(f) = \min\{f(x) : x \in \mathbb{R}^N\}.$$

El conjunto de soluciones es

$$\operatorname{argmin}(f) = \{x \in \mathbb{R}^N : f(x) = \min(f)\}.$$

Problemas de optimización

Estudiaremos el problema de existencia, caracterización y aproximación numérica de los puntos donde una función $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ alcanza su menor valor:

El valor del problema es

$$\min(f) = \min\{f(x) : x \in \mathbb{R}^N\}.$$

El conjunto de soluciones es

$$\operatorname{argmin}(f) = \{x \in \mathbb{R}^N : f(x) = \min(f)\}.$$

Problemas de optimización

Estudiaremos el problema de existencia, caracterización y aproximación numérica de los puntos donde una función $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ alcanza su menor valor:

El valor del problema es

$$\min(f) = \min\{f(x) : x \in \mathbb{R}^N\}.$$

El conjunto de soluciones es

$$\operatorname{argmin}(f) = \{x \in \mathbb{R}^N : f(x) = \min(f)\}.$$

Semicontinuidad inferior

Una función $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ es inferiormente semicontinua si para cada $x \in \mathrm{dom}(f)$ y cada $\varepsilon > 0$ existe $\delta > 0$ tal que $\|x - y\| < \delta \Rightarrow f(y) > f(x) - \varepsilon$.

Ejemplo

Toda función continua es inferiormente semicontinua.
Si C es cerrado, δ_C es inferiormente semicontinua.

Semicontinuidad inferior

Una función $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ es inferiormente semicontinua si para cada $x \in \mathrm{dom}(f)$ y cada $\varepsilon > 0$ existe $\delta > 0$ tal que $\|x - y\| < \delta \Rightarrow f(y) > f(x) - \varepsilon$.

Ejemplo

Toda función continua es inferiormente semicontinua.
Si C es cerrado, δ_C es inferiormente semicontinua.

Coercividad

Diremos que $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ es coerciva si

$$\lim_{\|x\|\to+\infty} f(x) = +\infty.$$

Es decir, si para cada $\gamma \in \mathbb{R}$ el conjunto de subnivel γ de f:

$$[f \leq \gamma] = \{x \in \mathbb{R}^N : f(x) \leq \gamma\}$$

es acotado.

Ejemplo

Si C es acotado, δ_C es coerciva.

Coercividad

Diremos que $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ es coerciva si

$$\lim_{\|x\| \to +\infty} f(x) = +\infty.$$

Es decir, si para cada $\gamma \in \mathbb{R}$ el conjunto de subnivel γ de f:

$$[f \leq \gamma] = \{x \in \mathbb{R}^N : f(x) \leq \gamma\}$$

es acotado.

Ejemplo

Si C es acotado, δ_C es coerciva.

Coercividad

Diremos que $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ es coerciva si

$$\lim_{\|x\| \to +\infty} f(x) = +\infty.$$

Es decir, si para cada $\gamma \in \mathbb{R}$ el conjunto de subnivel γ de f:

$$[f \leq \gamma] = \{x \in \mathbb{R}^N : f(x) \leq \gamma\}$$

es acotado.

Ejemplo

Si C es acotado, δ_C es coerciva.

Existencia de minimizadores

Teorema

Si $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ es inferiormente semicontinua y coerciva, entonces

$$argmin(f) \neq \emptyset.$$

Corolario

Si $f : C \subset \mathbb{R}^N \to \mathbb{R}$ es semicontinua y C es compacto, entonces

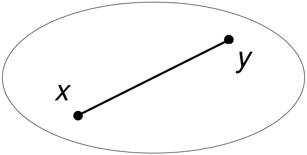
$$argmin(f) \neq \emptyset.$$

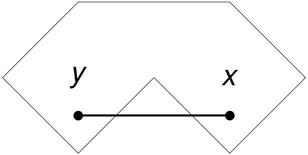
Conjuntos convexos

Un subconjunto C de $\mathbb{R}^N$ es convexo si para cada $x, y \in C$, el segmento que une x con y:

$$\mathrm{seg}[x, y] = \{\lambda x + (1-\lambda)y : \lambda \in [0, 1]\}$$

está contenido en C.





Condición necesaria de optimalidad

Teorema

Sea $f : \mathbb{R}^N \to \mathbb{R}$ *y sea* $C \subset \mathbb{R}^N$ *un conjunto convexo. Si* $f(x^*) \leq f(x)$ *para todo* $x \in C$ *y* f *es diferenciable en* x^**, entonces*

$$\nabla f(x^*) \cdot (x - x^*) \geq 0$$

para todo $x \in C$*. Si, además,* $x^* \in int(C)$*, entonces* $\nabla f(x^*) = 0$*.*

Funciones convexas

Una función $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ es convexa si para cada $x, y \in \mathbb{R}^N$ y cada $\lambda \in (0, 1)$ se cumple

$$f\big(\lambda x + (1-\lambda)y\big) \leq \lambda f(x) + (1-\lambda) f(y)$$

Esto es equivalente a que el epígrafo de f:

$$\mathrm{epi}(f) = \{(x, z) \in X \times \mathbb{R} : f(x) \leq z\}$$

sea convexo.

Lo anterior implica que el conjunto $[f \leq \gamma]$ es convexo para cada γ (pero no es equivalente!).

Funciones convexas

Una función $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ es convexa si para cada $x, y \in \mathbb{R}^N$ y cada $\lambda \in (0, 1)$ se cumple

$$f\big(\lambda x + (1-\lambda)y\big) \leq \lambda f(x) + (1-\lambda)f(y)$$

Esto es equivalente a que el epígrafo de f:

$$\mathrm{epi}(f) = \{(x, z) \in X \times \mathbb{R} : f(x) \leq z\}$$

sea convexo.

Lo anterior implica que el conjunto $[f \leq \gamma]$ es convexo para cada γ (pero no es equivalente!).

Funciones convexas

Una función $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ es convexa si para cada $x, y \in \mathbb{R}^N$ y cada $\lambda \in (0, 1)$ se cumple

$$f\big(\lambda x + (1-\lambda)y\big) \leq \lambda f(x) + (1-\lambda)f(y)$$

Esto es equivalente a que el epígrafo de f:

$$\mathrm{epi}(f) = \{(x, z) \in X \times \mathbb{R} : f(x) \leq z\}$$

sea convexo.
Lo anterior implica que el conjunto $[f \leq \gamma]$ es convexo para cada γ (pero no es equivalente!).

Propiedades del gradiente

Sea $f : \mathbb{R}^N \to \mathbb{R}$ convexa y diferenciable. Entonces

1. Se cumple la desigualdad del gradiente:

$$f(y) \geq f(x) + \nabla f(x) \cdot (y - x).$$

2. El gradiente es monótono:

$$\big(\nabla f(x) - \nabla f(y)\big) \cdot (x - y) \geq 0.$$

Propiedades del gradiente

Sea $f : \mathbb{R}^N \to \mathbb{R}$ convexa y diferenciable. Son equivalentes:

1. ∇f es Lipschitz-continuo con constante L:

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|.$$

2. Se cumple la desigualdad de descenso:

$$f(y) \leq f(x) + \nabla f(x) \cdot (y - x) + \frac{L}{2}\|x - y\|^2.$$

3. ∇f es cocoercivo con constante $1/L$:

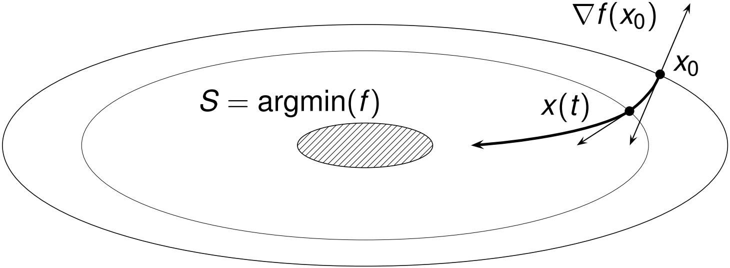
$$\big(\nabla f(x) - \nabla f(y)\big) \cdot (x - y) \geq \frac{1}{L}\|\nabla f(x) - \nabla f(y)\|^2.$$

Clase II

Métodos básicos de descenso

Dinámica del máximo descenso

Ecuación diferencial $\dot{x}(t) = -\nabla f(x(t))$, $x(0) = x_0$



Dinámica del máximo descenso

La función $t \mapsto f(x(t))$ es decreciente:

$$\frac{d}{dt} f(x(t)) = \nabla f(x(t)) \cdot \dot{x}(t) = -\|\nabla f(x(t))\|^2 = -\|\dot{x}(t)\|^2 \leq 0.$$

Anclaje al conjunto de soluciones: Si $x^* \in S$, entonces

$$\frac{d}{dt} \|x(t) - x^*\|^2 \leq 0.$$

Dinámica del máximo descenso

La función $t \mapsto f(x(t))$ es decreciente:

$$\frac{d}{dt} f(x(t)) = \nabla f(x(t)) \cdot \dot{x}(t) = -\|\nabla f(x(t))\|^2 = -\|\dot{x}(t)\|^2 \leq 0.$$

Anclaje al conjunto de soluciones: Si $x^* \in S$, entonces

$$\frac{d}{dt} \|x(t) - x^*\|^2 \leq 0.$$

Dinámica del máximo descenso

Minimización: Si $x^* \in S$, entonces

$$f(x(t)) - \min(f) \leq \frac{\mathrm{dist}(x_0, S)^2}{2t}.$$

Aproximación de las soluciones: Si $S \neq \emptyset$, entonces

$$\lim_{t \to +\infty} x(t)$$

existe y pertenece a S.

Dinámica del máximo descenso

Minimización: Si $x^* \in S$, entonces

$$f(x(t)) - \min(f) \leq \frac{\operatorname{dist}(x_0, S)^2}{2t}.$$

Aproximación de las soluciones: Si $S \neq \emptyset$, entonces

$$\lim_{t \to +\infty} x(t)$$

existe y pertenece a S.

Método del gradiente

Para aproximar las soluciones de la ecuación

$$\dot{x}(t) = -\nabla f(x(t)),$$

hacemos una discretización explícita en diferencias finitas:

$$\frac{x_{k+1} - x_k}{\lambda} = -\nabla f(x_k).$$

En otras palabras,

$$x_{k+1} = x_k - \lambda \nabla f(x_k).$$

Método del gradiente

Para aproximar las soluciones de la ecuación

$$\dot{x}(t) = -\nabla f(x(t)),$$

hacemos una discretización explícita en diferencias finitas:

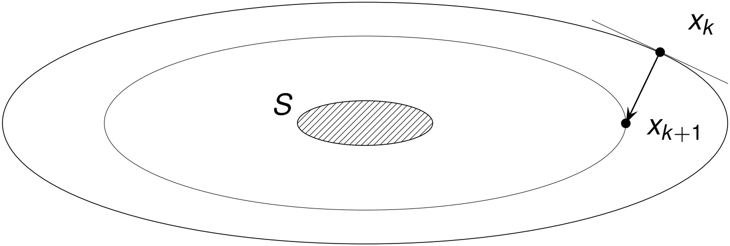
$$\frac{x_{k+1} - x_k}{\lambda} = -\nabla f(x_k).$$

En otras palabras,

$$x_{k+1} = x_k - \lambda \nabla f(x_k).$$

Método del gradiente

$$x_{k+1} = x_k - \lambda \nabla f(x_k)$$



Puede salir bien o mal

Ejemplo

Sean $f(x) = x^2$ *y* $x_0 = 1$*. Entonces*

$$x_k = (1 - 2\lambda)^k.$$

En particular,

1. *Si* $\lambda > 1$, $\lim_{k\to+\infty} |x_k| = +\infty$;
2. *Si* $\lambda = 1$, $x_k = (-1)^k$, $|x_k| \equiv 1$ *pero* $\nexists \lim_{k\to+\infty} x_k$; *y*
3. *Si* $\lambda < 1$, $\lim_{k\to+\infty} x_k = 0$.

Puede salir bien o mal

Ejemplo

Sean $f(x) = x^2$ *y* $x_0 = 1$. *Entonces*

$$x_k = (1 - 2\lambda)^k.$$

En particular,

1. *Si* $\lambda > 1$, $\lim_{k \to +\infty} |x_k| = +\infty$;
2. *Si* $\lambda = 1$, $x_k = (-1)^k$, $|x_k| \equiv 1$ *pero* $\nexists \lim_{k \to +\infty} x_k$; *y*
3. *Si* $\lambda < 1$, $\lim_{k \to +\infty} x_k = 0$.

Puede salir bien o mal

Ejemplo

Sean $f(x) = x^2$ *y* $x_0 = 1$. *Entonces*

$$x_k = (1 - 2\lambda)^k.$$

En particular,

1. *Si* $\lambda > 1$, $\lim_{k\to+\infty} |x_k| = +\infty$;
2. *Si* $\lambda = 1$, $x_k = (-1)^k$, $|x_k| \equiv 1$ *pero* $\nexists \lim_{k\to+\infty} x_k$; *y*
3. *Si* $\lambda < 1$, $\lim_{k\to+\infty} x_k = 0$.

Puede salir bien o mal

Ejemplo

Sean $f(x) = x^2$ *y* $x_0 = 1$. *Entonces*

$$x_k = (1 - 2\lambda)^k.$$

En particular,

1. *Si* $\lambda > 1$, $\lim_{k \to +\infty} |x_k| = +\infty$;
2. *Si* $\lambda = 1$, $x_k = (-1)^k$, $|x_k| \equiv 1$ *pero* $\nexists \lim_{k \to +\infty} x_k$; *y*
3. *Si* $\lambda < 1$, $\lim_{k \to +\infty} x_k = 0$.

Convergencia

Supongamos que ∇f es Lipschitz-continuo con constante L y

$$\lambda L < 2.$$

Recordemos que se tiene la desigualdad del descenso:

$$f(y) \leq f(x) + \nabla f(x) \cdot (y - x) + \frac{L}{2}\|x - y\|^2$$

y la cocoercividad:

$$(\nabla f(x) - \nabla f(y)) \cdot (x - y) \geq \frac{1}{L}\|\nabla f(x) - \nabla f(y)\|^2.$$

Convergencia

Supongamos que ∇f es Lipschitz-continuo con constante L y

$$\lambda L < 2.$$

Recordemos que se tiene la desigualdad del descenso:

$$f(y) \leq f(x) + \nabla f(x) \cdot (y - x) + \frac{L}{2}\|x - y\|^2$$

y la cocoercividad:

$$\big(\nabla f(x) - \nabla f(y)\big) \cdot (x - y) \geq \frac{1}{L}\|\nabla f(x) - \nabla f(y)\|^2.$$

Convergencia

La sucesión $f(x_k)$ es decreciente:

$$f(x_{k+1}) \leq f(x_k).$$

Anclaje al conjunto de soluciones: Si $x^* \in S$, entonces

$$\|x_{k+1} - x^*\|^2 + \delta\|x_{k+1} - x_k\|^2 \leq \|x_k - x^*\|^2.$$

Convergencia

La sucesión $f(x_k)$ es decreciente:

$$f(x_{k+1}) \leq f(x_k).$$

Anclaje al conjunto de soluciones: Si $x^* \in S$, entonces

$$\|x_{k+1} - x^*\|^2 + \delta\|x_{k+1} - x_k\|^2 \leq \|x_k - x^*\|^2.$$

Convergencia

Minimización: Si $S \neq \emptyset$, entonces

$$f(x_k) - \min(f) \leq \frac{\mathrm{dist}(x_0, S)^2}{\lambda(2 - \lambda L)k}.$$

Convergencia: Si $S \neq \emptyset$, entonces

$$\lim_{k \to +\infty} x_k$$

existe y pertenece a S.

Convergencia

Minimización: Si $S \neq \emptyset$, entonces

$$f(x_k) - \min(f) \leq \frac{\mathrm{dist}(x_0, S)^2}{\lambda(2 - \lambda L)k}.$$

Convergencia: Si $S \neq \emptyset$, entonces

$$\lim_{k \to +\infty} x_k$$

existe y pertenece a S.

CLASE III

Optimización no diferenciable

Optimización no diferenciable

El algoritmo proximal se define de la siguiente forma: dados $z_k \in \mathbb{R}^N$ y $\lambda > 0$, escogemos z_{k+1} como la única solución del problema

$$\operatorname{argmin} \left\{ f(z) + \frac{1}{2\lambda} \|z - z_k\|^2 \right\}.$$

Observación

- *Está bien definido.*
- *No es necesario que f sea diferenciable.*

Optimización no diferenciable

El algoritmo proximal se define de la siguiente forma: dados $z_k \in \mathbb{R}^N$ y $\lambda > 0$, escogemos z_{k+1} como la única solución del problema

$$\operatorname{argmin}\left\{ f(z) + \frac{1}{2\lambda}\|z - z_k\|^2 \right\}.$$

Observación

- *Está bien definido.*
- *No es necesario que f sea diferenciable.*

Relación con los métodos anteriores

Dinámica del máximo descenso:

$$\dot{x}(t) = -\nabla f(x(t))$$

Método del gradiente:

$$\frac{x_{k+1} - x_k}{\lambda} = -\nabla f(x_k) \qquad \Leftrightarrow \qquad x_{k+1} = x_k - \lambda \nabla f(x_k).$$

Algoritmo proximal (si f es diferenciable):

$$\frac{z_{k+1} - z_k}{\lambda} = -\nabla f(z_{k+1}) \qquad \Leftrightarrow \qquad z_{k+1} + \lambda \nabla f(z_{k+1}) = z_k.$$

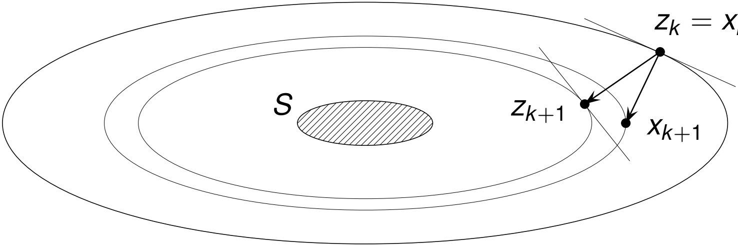
Comparación geométrica

Gradiente

$x_{k+1} = x_k - \lambda \nabla f(x_k)$

Proximal

$z_{k+1} + \lambda \nabla f(z_{k+1}) = z_k$



Lo bueno y lo malo

Método del gradiente

+ Fácil implementación (fórmula explícita).
– Se necesita f diferenciable con ∇f Lipschitz-continuo para tener convergencia. Hay que escoger bien el paso λ.

Algoritmo proximal

+ Garantía de convergencia aunque f no sea diferenciable, y sin importar el paso.
– La implementación puede ser más difícil (la fórmula es implícita).

Lo bueno y lo malo

Método del gradiente

+ Fácil implementación (fórmula explícita).
– Se necesita f diferenciable con ∇f Lipschitz-continuo para tener convergencia. Hay que escoger bien el paso λ.

Algoritmo proximal

+ Garantía de convergencia aunque f no sea diferenciable, y sin importar el paso.
– La implementación puede ser más difícil (la fórmula es implícita).

Ejemplos

Ejemplo

En $\mathbb{R}^N$, $f(x) = \delta_C(x)$:

$$z_k = proy_C(z_0)$$

para todo $k \geq 1$.

Ejemplo

En $\mathbb{R}$, $f(x) = |x|$:

$$z_{k+1} = \begin{cases} z_k - \lambda & si \quad z_k > \lambda \\ z_k + \lambda & si \quad z_k < -\lambda \\ 0 & si \quad z_k \in [-\lambda, \lambda]. \end{cases}$$

Ejemplos

Ejemplo

En $\mathbb{R}^N$, $f(x) = \delta_C(x)$*:*

$$z_k = proy_C(z_0)$$

para todo $k \geq 1$.

Ejemplo

En $\mathbb{R}$, $f(x) = |x|$*:*

$$z_{k+1} = \begin{cases} z_k - \lambda & si \quad z_k > \lambda \\ z_k + \lambda & si \quad z_k < -\lambda \\ 0 & si \quad z_k \in [-\lambda, \lambda]. \end{cases}$$

Convergencia

La sucesión $f(z_k)$ es decreciente:

$$f(z_{k+1}) \leq f(z_k).$$

Anclaje al conjunto de soluciones: Si $x^* \in S$, entonces

$$\|z_{k+1} - x^*\| \leq \|z_k - x^*\|.$$

Convergencia

La sucesión $f(z_k)$ es decreciente:

$$f(z_{k+1}) \leq f(z_k).$$

Anclaje al conjunto de soluciones: Si $x^* \in S$, entonces

$$\|z_{k+1} - x^*\| \leq \|z_k - x^*\|.$$

Convergencia

Minimización: Si $S \neq \emptyset$, entonces

$$f(z_k) - \min(f) \leq \frac{\mathrm{dist}(z_0, S)^2}{2\lambda k}.$$

Convergencia: Si $S \neq \emptyset$, entonces

$$\lim_{k \to +\infty} z_k$$

existe y pertenece a S.

Convergencia

Minimización: Si $S \neq \emptyset$, entonces

$$f(z_k) - \min(f) \leq \frac{\mathrm{dist}(z_0, S)^2}{2\lambda k}.$$

Convergencia: Si $S \neq \emptyset$, entonces

$$\lim_{k \to +\infty} z_k$$

existe y pertenece a S.

El subdiferencial

Sabemos que si $f : \mathbb{R}^N \to \mathbb{R}$ es convexa y diferenciable, entonces

$$f(y) \geq f(x) + \nabla f(x) \cdot (y - x).$$

Diremos que $v \in \mathbb{R}^N$ es un subgradiente de f en x si

$$f(y) \geq f(x) + v \cdot (y - x)$$

para todo $y \in \mathbb{R}^N$. El subdiferencial de f en x es el conjunto de todos los subgradientes:

$$\partial f(x) = \{v \in \mathbb{R}^N : f(y) \geq f(x) + v \cdot (y - x) \text{ para todo } y \in \mathbb{R}^N\}$$

Algunas propiedades

Caracterización de las soluciones:

$$x^* \in \operatorname{argmin}(f) \qquad \Longleftrightarrow \qquad 0 \in \partial f(x^*)$$

Monotonía: Si $\hat{x} \in \partial f(x)$ y $\hat{y} \in \partial f(y)$, entonces

$$(\hat{x} - \hat{y}) \cdot (x - y) \geq 0$$

Algunas propiedades

Caracterización de las soluciones:

$$x^* \in \text{argmin}(f) \qquad \Longleftrightarrow \qquad 0 \in \partial f(x^*)$$

Monotonía: Si $\hat{x} \in \partial f(x)$ y $\hat{y} \in \partial f(y)$, entonces

$$(\hat{x} - \hat{y}) \cdot (x - y) \geq 0$$

Ejemplos

Ejemplo

Si f es diferenciable, entonces $\partial f(x) = \{\nabla f(x)\}$.

Ejemplo

En $\mathbb{R}$, si $f(x) = |x|$:

$$\partial f(x) = \begin{cases} \{1\} & si \quad x > 0 \\ \{-1\} & si \quad x < 0 \\ [-1, 1] & si \quad x = 0. \end{cases}$$

Ejemplos

Ejemplo

Si f es diferenciable, entonces $\partial f(x) = \{\nabla f(x)\}$.

Ejemplo

En $\mathbb{R}$, *si* $f(x) = |x|$:

$$\partial f(x) = \begin{cases} \{1\} & si \quad x > 0 \\ \{-1\} & si \quad x < 0 \\ [-1, 1] & si \quad x = 0. \end{cases}$$

CLASE IV

Combinando funciones diferenciables y no diferenciables

Tenemos dos métodos básicos

Método del gradiente (para g diferenciable):

$$x_{k+1} = \mathrm{Grad}_{\lambda g}(x_k) = x_k - \lambda \nabla g(x_k)$$

Algoritmo proximal (para h cualquiera):

$$z_{k+1} = \mathrm{Prox}_{\lambda h}(z_k) = \mathrm{argmin} \left\{ h(z) + \frac{1}{2\lambda} \|z - z_k\|^2 \right\}$$

Estructura mixta

Queremos minimizar $f = g + h$, donde g es diferenciable pero h no.

Ejemplo

El problema con restricciones

$$\min\{g(x) : x \in C\}$$

es equivalente a

$$\min\{g(x) + \delta_C(x) : x \in \mathbb{R}^N\}.$$

Estructura mixta

Queremos minimizar $f = g + h$, donde g es diferenciable pero h no.

Ejemplo

El problema con restricciones

$$\min\{g(x) : x \in C\}$$

es equivalente a

$$\min\{g(x) + \delta_C(x) : x \in \mathbb{R}^N\}.$$

Estructura mixta

Queremos minimizar $f = g + h$, donde g es diferenciable pero h no.

Ejemplo

Sean $\mu > 0$, $b \in \mathbb{R}^M$ y A una matriz de tamaño $M \times N$. Considere las funciones $g, h : \mathbb{R}^N \to \mathbb{R}$ definidas por

$$g(x) = \mu \sum_{n=1}^{N} |x_n| \qquad y \qquad h(x) = \frac{1}{2}\|Ax - b\|^2$$

La función $f = g + h$ aparece en procesamiento de imágenes y compresión datos, entre otros.

Estrategia

Usar método del gradiente para la parte diferenciable y algoritmo proximal para la parte no diferenciable:

Método gradiente-proximal:

$$x_{k+1} = T_\lambda(x_k) = \mathrm{Prox}_{\lambda h}\big(\mathrm{Grad}_{\lambda g}(x_k)\big)$$

Caso $h = 0$: gradiente

Caso $h = \delta_C$: gradiente proyectado $x_{k+1} = \mathrm{Proy}_C\big(\mathrm{Grad}_{\lambda g}(x_k)\big)$

Caso $g = 0$: proximal

Una propiedad fundamental

Proposición

Si ∇g es Lipschitz-continuo con constante L y $\lambda L \leq 1$, entonces

$$f(T_\lambda(x)) + \frac{\|y - T_\lambda(x)\|^2}{2\lambda} \leq f(y) + \frac{\|y - x\|^2}{2\lambda}.$$

Principales consecuencias

Proposición

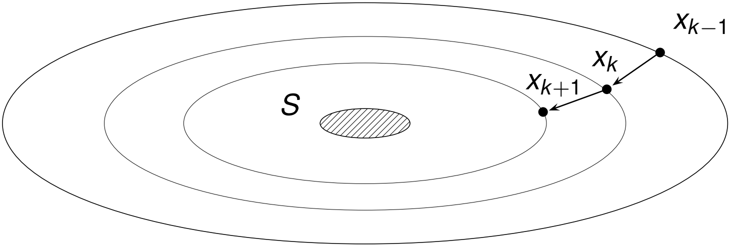
- *Decrecimiento:* $f(x_{k+1}) + \frac{1}{2\lambda}\|x_{k+1} - x_k\|^2 \leq f(x_k)$
- *Anclaje: Si* $x^* \in S$, $\|x_{k+1} - x^*\| \leq \|x_k - x^*\|$
- *Minimización: Si* $S \neq \emptyset$, $f(x_k) - \min(f) \leq \dfrac{dist(x_0, S)^2}{2\lambda k}$
- *Convergencia: Si* $S \neq \emptyset$, $\lim\limits_{k\to+\infty} x_k$ *existe y está en* S.

CLASE V

Aceleración

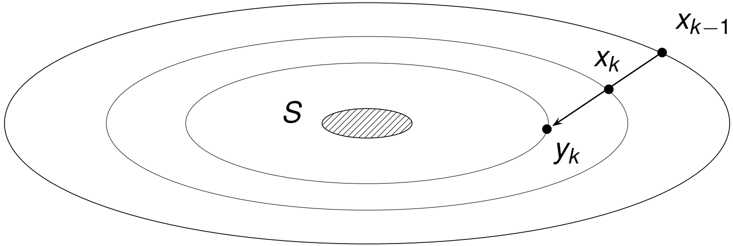
Aceleración

La idea es la siguiente: En lugar de hacer esto:



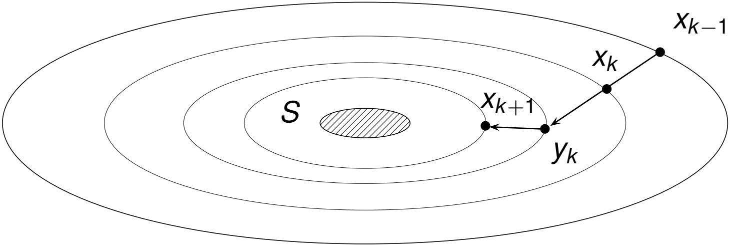
Aceleración

Probemos mejor lo siguiente:



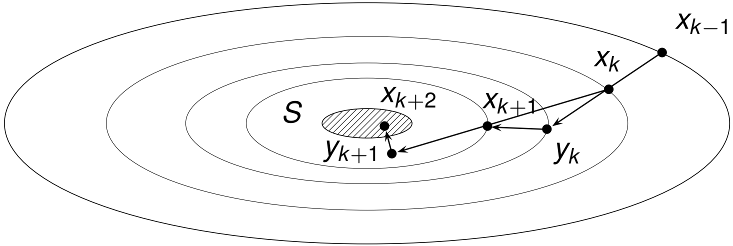
Aceleración

Probemos mejor lo siguiente:



Aceleración

Probemos mejor lo siguiente:



Aceleración

Algoritmo gradiente-proximal acelerado:

$$\begin{cases} y_k = x_k + \frac{k-2}{k+1}(x_k - x_{k-1}) \\ x_{k+1} = \mathrm{Prox}_{\lambda h}(\mathrm{Grad}_{\lambda g}(y_k)) \end{cases}$$

donde ∇g es Lipschitz-continuo con constante L, y $\lambda L \leq 1$.

Teorema

$$(g+h)(x_k) - \min(g+h) \leq \frac{dist(x_0, S)^2}{\lambda k^2}, \quad k \geq 1$$

Aceleración

Algoritmo gradiente-proximal acelerado:

$$\begin{cases} y_k = x_k + \frac{k-2}{k+1}(x_k - x_{k-1}) \\ x_{k+1} = \mathrm{Prox}_{\lambda h}\big(\mathrm{Grad}_{\lambda g}(y_k)\big) \end{cases}$$

donde ∇g es Lipschitz-continuo con constante L, y $\lambda L \leq 1$.

Teorema

$$(g+h)(x_k) - \min(g+h) \leq \frac{dist(x_0, S)^2}{\lambda k^2}, \quad k \geq 1$$

Aceleración

Usaremos la propiedad fundamental:

Proposición

Si ∇g es Lipschitz-continuo con constante L y $\lambda L \leq 1$, entonces

$$f(T_\lambda(x)) + \frac{\|y - T_\lambda(x)\|^2}{2\lambda} \leq f(y) + \frac{\|y - x\|^2}{2\lambda}.$$

Con:

$$\begin{aligned} x &= y_k &&= x_k + \tfrac{k-2}{k+1}(x_k - x_{k-1}) \\ T_\lambda(x) &= T_\lambda(y_k) &&= x_{k+1} \\ & \quad\;\; y &&= \tfrac{2}{k+1}x^* + \tfrac{k-1}{k+1}x_k \end{aligned}$$