Análisis convexo, cálculo diferencial y aplicaciones

Juan PEYPOUQUET
Universidad Técnica Federico Santa María

V-Escuela 2016

Valparaíso, 11 al 21 de octubre

BÚSQUEDA DE MINIMIZADORES

Sucesiones minimizantes convergentes

Sea $f : X \to \mathbb{R} \cup \{+\infty\}$ una función propia e inferiormente semicontinua, y sea (x_n) una sucesión minimizante para f. Es decir, tal que

$$\lim_{n\to\infty} f(x_n) = \inf(f).$$

Tenemos lo siguiente:

- Si x_n converge a $\bar{x}$, entonces $\bar{x} \in S$.
- Si f es convexa y x_n converge débilmente a $\bar{x}$, entonces $\bar{x} \in S$.

Sucesiones minimizantes convergentes

Sea $f : X \to \mathbb{R} \cup \{+\infty\}$ una función propia e inferiormente semicontinua, y sea (x_n) una sucesión minimizante para f. Es decir, tal que

$$\lim_{n\to\infty} f(x_n) = \inf(f).$$

Tenemos lo siguiente:

- Si x_n converge a $\bar{x}$, entonces $\bar{x} \in S$.
- Si f es convexa y x_n converge débilmente a $\bar{x}$, entonces $\bar{x} \in S$.

Sucesiones minimizantes convergentes

Sea $f : X \to \mathbb{R} \cup \{+\infty\}$ una función propia e inferiormente semicontinua, y sea (x_n) una sucesión minimizante para f. Es decir, tal que

$$\lim_{n\to\infty} f(x_n) = \inf(f).$$

Tenemos lo siguiente:

- Si x_n converge a $\bar{x}$, entonces $\bar{x} \in S$.
- Si f es convexa y x_n converge débilmente a $\bar{x}$, entonces $\bar{x} \in S$.

Sucesiones minimizantes convergentes

Sea $f : X \to \mathbb{R} \cup \{+\infty\}$ una función propia e inferiormente semicontinua, y sea (x_n) una sucesión minimizante para f. Es decir, tal que

$$\lim_{n\to\infty} f(x_n) = \inf(f).$$

Tenemos lo siguiente:

- Si x_n converge a $\bar{x}$, entonces $\bar{x} \in S$.
- Si f es convexa y x_n converge débilmente a $\bar{x}$, entonces $\bar{x} \in S$.

Una condición suficiente para la convergencia

Teorema

Sea X un espacio de Hilbert y sea $f : X \to \mathbb{R} \cup \{+\infty\}$ una función propia, convexa e inferiormente semicontinua con $S \neq \emptyset$. Supongamos que (x_n) es una sucesión minimizante para f tal que para cada $s \in S$ existe

$$\lim_{n\to\infty} \|x_n - s\|.$$

Entonces, x_n converge débilmente a algún $\bar{x} \in S$.

Estrategia para buscar minimizadores

Para aproximar un minimizador de f basta encontrar una sucesión (x_n) tal que

1. $\lim_{n\to\infty} f(x_n) = \inf(f)$; y
2. $\lim_{n\to\infty} \|x_n - s\|$ existe para cada $s \in S$.

Bajo estas condiciones, x_n converge débilmente a algún $\bar{x} \in S$.

Estrategia para buscar minimizadores

Para aproximar un minimizador de f basta encontrar una sucesión (x_n) tal que

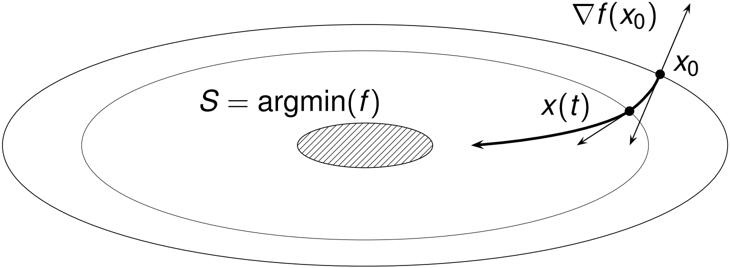
1. $\lim_{n\to\infty} f(x_n) = \inf(f)$; y
2. $\lim_{n\to\infty} \|x_n - s\|$ existe para cada $s \in S$.

Bajo estas condiciones, x_n converge débilmente a algún $\bar{x} \in S$.

ALGORITMO GRADIENTE-PROXIMAL

Inspiración: Dinámica del máximo descenso

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



Principales propiedades

Proposición

1. $t \mapsto f(x(t))$ *es decreciente;*
2. $t \mapsto \|x(t) - s\|$ *es decreciente para cada* $s \in S$*;*
3. $f(x(t)) - \min(f) \leq \dfrac{dist(x_0, S)^2}{2t}$*; y*
4. *Cuando* $t \to +\infty$, $x(t)$ *converge débilmente a algún* $\bar{x} \in S$.

Principales propiedades

Proposición

1. $t \mapsto f(x(t))$ *es decreciente;*
2. $t \mapsto \|x(t) - s\|$ *es decreciente para cada* $s \in S$*;*
3. $f(x(t)) - \min(f) \leq \dfrac{dist(x_0, S)^2}{2t}$*; y*
4. *Cuando* $t \to +\infty$*,* $x(t)$ *converge débilmente a algún* $\bar{x} \in S$*.*

Principales propiedades

Proposición

1. $t \mapsto f(x(t))$ *es decreciente;*
2. $t \mapsto \|x(t) - s\|$ *es decreciente para cada* $s \in S$*;*
3. $f(x(t)) - \min(f) \leq \dfrac{dist(x_0, S)^2}{2t}$*; y*
4. *Cuando* $t \to +\infty$, $x(t)$ *converge débilmente a algún* $\bar{x} \in S$.

Principales propiedades

Proposición

1. $t \mapsto f(x(t))$ *es decreciente;*
2. $t \mapsto \|x(t) - s\|$ *es decreciente para cada* $s \in S$*;*
3. $f(x(t)) - \min(f) \leq \dfrac{dist(x_0, S)^2}{2t}$*; y*
4. *Cuando* $t \to +\infty$, $x(t)$ *converge débilmente a algún* $\bar{x} \in S$.

Método del gradiente (Cauchy, 1847)

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:

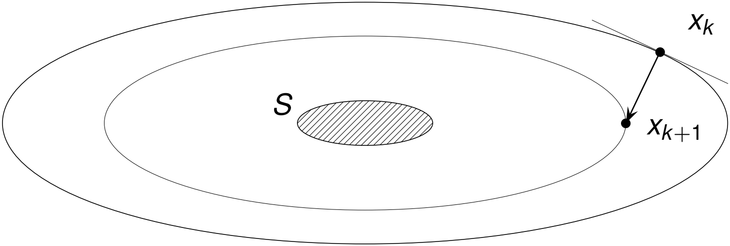
$$\frac{x_{k+1} - x_k}{\lambda} = -\nabla f(x_k).$$

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* $\lambda < 1$, $\lim_{k\to+\infty} x_k = 0$.

Iteraciones proximales (Martinet, 1970)

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.*

Ejemplos

Ejemplo

Si $f(x) = \delta_C(x)$*:*

$$z_k = proy_C(z_0)$$

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

Si $f(x) = \delta_C(x)$*:*

$$z_k = proy_C(z_0)$$

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

Si $f(x) = \delta_C(x)$*:*

$$z_k = proy_C(z_0)$$

para todo $k \geq 1$.

Ejemplo

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

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

Relación entre los 3

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:

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

Relación entre los 3

Dinámica del máximo descenso:

$$\dot{x}(t) \in -\partial f(x(t))$$

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:

$$\frac{z_{k+1} - z_k}{\lambda} \in -\partial f(z_{k+1}) \qquad \Leftrightarrow \qquad z_{k+1} + \lambda \partial f(z_{k+1}) \ni z_k.$$

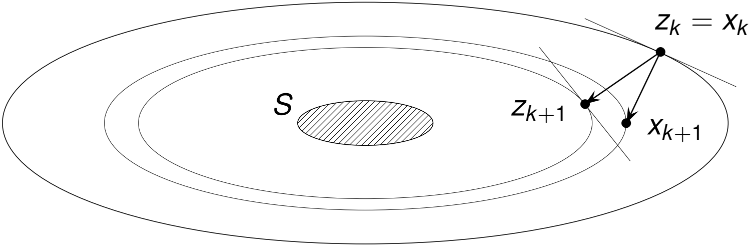
Comparación geométrica

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).

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):

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

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:

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

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

Comportamiento del método

Teorema

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