Métodos Numéricos

Fernando Roldán Contreras

Universidad de Concepción
Facultad de Ciencias Físicas y Matemáticas
Departamento de Ingeniería Matemática

Optimización No Lineal
2025

- Bauschke, H.H., Combettes, P.L.: Convex analysis and monotone operator theory in Hilbert spaces, second edn.
  CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, Cham (2017).
  DOI 10.1007/978-3-319-48311-5
- Nesterov, Y.: Lectures on convex optimization, *Springer Optimization and Its Applications*, vol. 137, second edn.
  Springer, Cham (2018).
  DOI 10.1007/978-3-319-91578-4
- Nocedal, J., Wright, S.J.: Numerical optimization, second edn.
  Springer Series in Operations Research and Financial Engineering.
  Springer, New York (2006)

Problema de Optimización

$$\min_{x \in S} f(x)$$

- $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ es una función.
- $S \subset \mathbb{R}^n$.

Estudiaremos métodos iterativos de la forma

Método iterativo

Dado $x_0 \in S$, iteramos,

$$(\forall n \in \mathbb{N}) \quad \lfloor \; x_{n+1} = T(x_n),$$

donde T es un operador para el cual nos gustaría que $x_n \to x$ donde x es un minimizador.

Método iterativo

Dado $x_0 \in S$, iteramos,

$$(\forall n \in \mathbb{N}) \quad \lfloor \; x_{n+1} = T(x_n),$$

En la practica, se calcula x_n para un número finito de iteraciones hasta cumplir algún criterio de parada:

- Número máximo de iteraciones.
- Tolerancia criterio de optimalidad: $\|\nabla f(x_n)\| \leq \varepsilon$, ε pequeño, por ejemplo, $\varepsilon = 10^{-6}$.
- Tolerancia error relativo: $\frac{\|x_{n+1}-x_n\|}{\|x_n\|} \leq \varepsilon$.

Método iterativo: Direcciones de descenso

Dado $x_0 \in S$, iteramos,

$$(\forall n \in \mathbb{N}) \qquad \lfloor\; x_{n+1} = x_n + \alpha_n d_n,$$

- $T(x_n) = x_n + \alpha_n d_n$
- $(\alpha_n)_{n\in\mathbb{N}}$ es una sucesión positiva.
- $(d_n)_{n\in\mathbb{N}}$ es una sucesión de descenso, es decir $f(x_{n+1}) = f(x_n + \alpha_n d_n) \leq f(x_n)$.

Supongamos que f es diferenciable.

- Llamaremos $d \in \mathbb{R}^n$ una dirección de descenso si

$$\nabla f(x)^\top d < 0.$$

- Revisaremos diferentes formas de elegir d_n.

Direcciones de descenso

Direcciones de descenso

Método iterativo: Direcciones de descenso

Dado $x_0 \in S$, iteramos,

$$(\forall n \in \mathbb{N}) \qquad \lfloor\; x_{n+1} = x_n + \alpha_n d_n,$$

- $T(x_n) = x_n + \alpha_n d_n$
- $(\alpha_n)_{n\in\mathbb{N}}$ es una secesión positiva.
- $(d_n)_{n\in\mathbb{N}}$ debe generar un descenso, es decir,

$$f(x_{n+1}) = f(x_n + \alpha_n d_n) \leq f(x_n).$$

Direcciones de descenso

Método iterativo: Direcciones de descenso

Dado $x_0 \in S$, iteramos,

$$(\forall n \in \mathbb{N}) \quad \lfloor \; x_{n+1} = x_n + \alpha_n d_n,$$

Supongamos que f es diferenciable.

- Llamaremos $d \in \mathbb{R}^n$ una dirección de descenso si

$$\nabla f(x)^\top d < 0.$$

- Si d_n es una dirección de descenso, para α_n pequeño, aproximando por un desarrollo de taylor de primer orden, se tiene

$$f(x_{n+1}) \approx f(x_n) + \alpha_n \nabla f(x_n)^\top d_n < f(x_n).$$

- Revisaremos diferentes formas de elegir d_n.

Ejemplo: $d = -\nabla f(x_n)$

Método del gradiente

Dado $x_0 \in S$, iteramos,

$$(\forall n \in \mathbb{N}) \quad \lfloor \; x_{n+1} = x_n - \alpha_n \nabla f(x_n),$$

- $d = -\nabla f(x_n)$ es dirección de descenso:

$$\nabla f(x_n)^\top d = -\|\nabla f(x_n)\|^2 < 0.$$

- El valor de α_n influye en la convergencia del método.
- Ver ejemplo practico para $f(x) = x^2$.

Elegir α_n: Búsqueda Lineal Exacta

Dada una dirección de descenso d_n, revisemos formas de escoger α_n.

Búsqueda Lineal exacta

Elegimos α_n como la solución del problema:

$$\min_{\alpha>0} f(x_n + \alpha d_n).$$

1. Esta elección da el mejor α_n.
2. Resolver este problema puede difícil.
3. Para problemas cuadráticos estrictamente convexo se puede encontrar una fórmula explícita.

Elegir α_n: Regla de Armijo

Debido a la dificultad anterior, existen reglas de búsqueda lineal inexactas.

Regla de Armijo

Dado $\omega \in]0,1[$, elegimos α_n tal que:

$$f(x_n + \alpha_n d_n) \leq f(x_n) + \omega \alpha_n \nabla f(x_n)^\top d_n.$$

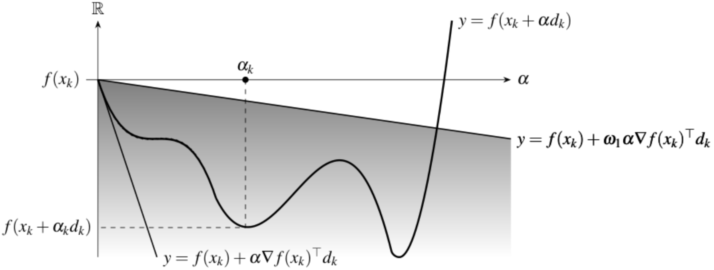
- El decrecimiento es proporcional a ω.
- Para encontrar α_n se utiliza la siguiente rutina:
  1. Elegir $\alpha \in \ ]0,+\infty[$ y $\tau \in]0,1/2[$.
  2. Si α satisface la regla de Armijo detenerse con $\alpha_n = \alpha$.
  3. Si no, probar un nuevo α en $[\tau\alpha,(1-\tau)\alpha]$.
- Ejemplo Numérico.

Elegir α_n: Regla de Armijo

Regla de Armijo

Dado $\omega \in]0,1[$, elegimos α_n tal que:

$$f(x_n + \alpha_n d_n) \leq f(x_n) + \omega \alpha_n \nabla f(x_n)^\top d_n.$$



Regla de Armijo

Imagen tomada del apunte “Optimización no Lineal”, de Luis Briceño y Cristopher Hermosilla

Ejercicios

Resuelva los siguientes problemas utilizando el método del gradiente y los pasos indicados. En cada caso grafique la sucesión obtenido y función objetivo con los valores de la sucesión en la función objetivo.

1. $min_{x \in \mathbb{R}} x^2$
   1. α_n constante, pruebe pasos $\alpha \in]0,1[$ y $\alpha \in [1,+\infty[$. ¿Cual es la diferencia para estos intervalos de pasos?
   2. Con la regla de Armijo. Pruebe distintos valores de ω y τ y vea en que cambiar en la convergencia y la búsqueda lineal.
2. $\min_{x \in \mathbb{R}} 5 \left(\frac{x^4}{2} - \frac{7x^3}{9} + \frac{x^2}{3} \right)$
   1. α_n constante y dado por la regla de Armijo con diferentes valores de ω y τ.

Elegir α_n: Regla de Goldstein

Hecho: puede ocurrir que α_n converja a 0 demasiado rápido en la regla de Armijo y el algoritmo no converja a un mínimo local. Para agregar una cota inferior a α_n existe la regla de Goldstein.

Regla de Goldstein

Dado $\omega \in]0,1[$, elegimos α_n tal que:

$$f(x_n + \alpha_n d_n) \leq f(x_n) + \omega \alpha_n \nabla f(x_n)^\top d_n,$$

$$f(x_n + \alpha_n d_n) \geq f(x_n) + (1-\omega) \alpha_n \nabla f(x_n)^\top d_n.$$

- Esta segunda condición ayuda a que α_n no sea tan pequeño.
- Resultado: Si $f : \mathbb{R}^n \to \mathbb{R}$ es diferenciable, entonces para todo $\omega \in]0, 1/2[$, dado una dirección de descenso d_n existe, $\alpha_n > 0$ que cumple la regla de Goldstein (y entonces la de Armijo).

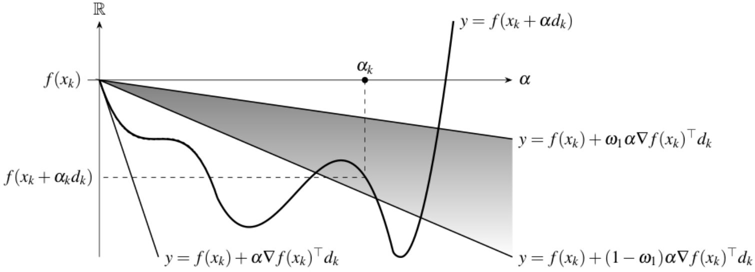
Elegir α_n: Regla de Goldstein

Regla de Goldstein

Dado $\omega \in]0,1[$, elegimos α_n tal que:

$$f(x_n + \alpha_n d_n) \leq f(x_n) + \omega \alpha_n \nabla f(x_n)^\top d_n,$$

$$f(x_n + \alpha_n d_n) \geq f(x_n) + (1-\omega) \alpha_n \nabla f(x_n)^\top d_n.$$



Regla de Goldstein.

Imagen tomada del apunte “Optimización no Lineal”, de Luis Briceño y Cristopher Hermosilla

Elegir α_n: Regla de Wolfe

Regla de Wolfe

Dado $\omega \in]0,1[$ y $\omega' \in]\omega, 1[$, elegimos α_n tal que:

$$f(x_n + \alpha_n d_n) \leq f(x_n) + \omega \alpha_n \nabla f(x_n)^\top d_n,$$

$$\nabla f(x_n + \alpha_n d_n)^\top d_n \geq \omega' \nabla f(x_n)^\top d_n.$$

- Esta condición también ayuda a evitar que α_n sea demasiado pequeño.
- Resultado: Si $f : \mathbb{R}^n \to \mathbb{R}$ es diferenciable, entonces para todo $\omega \in]0,1[$ y $\omega' \in]\omega, 1[$, dado una dirección de descenso d_n existe, $\alpha_n > 0$ que cumple la regla de Wolfe .

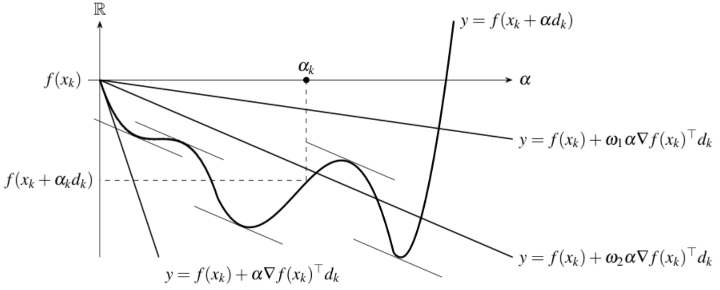
Elegir α_n: Regla de Wolfe

Regla de Wolfe

Dado $\omega \in]0,1[$ y $\omega' \in]\omega, 1[$, elegimos α_n tal que:

$$f(x_n + \alpha_n d_n) \le f(x_n) + \omega \alpha_n \nabla f(x_n)^\top d_n,$$

$$\nabla f(x_n + \alpha_n d_n)^\top d_n \ge \omega' \nabla f(x_n)^\top d_n.$$



Regla de Wolfe

Imagen tomada del apunte "Optimización no Lineal", de Luis Briceño y Cristopher Hermosilla

Convergencia del método de direcciones de descenso

Teorema (Convergencia)

Sea $(x_n)_{n\in\mathbb{N}}$ generada por el algoritmo de direcciones de descenso, con $x_0 \in \mathbb{R}^n$, $(d_n)_{n\in N}$ direcciones de descenso y $(\alpha_n)_{n\in\mathbb{N}}$ pasos generados por la regla de Wolfe. Sea $\mathcal{N}$ un abierto que contiene a $\Gamma_{f(x_0)}(f)$ donde f es diferenciable y ∇f es Lipschitz continuo. Entonces

$$\sum_{n\in N} \cos^2(\theta_n)\|\nabla f(x_n)\|^2 < +\infty, \tag{1}$$

para $\cos(\theta_n) = -\dfrac{\nabla f(x_n) d_n}{\|\nabla f(x_n)\|\|d_n\|}$, el coseno del ángulo entre $\nabla f(x_n)$ y d_n.

- En el caso que $\cos(\theta_n)$ es acotado por abajo por un número positivo (por ejemplo cuando $d_n = -\nabla f(x_n)$), se tiene que $\nabla f(x_n) \to 0$.
- Resultados similares se pueden obtener para la regla de Armijo y la de Goldstein.

Método de Newton y Quasi-Newton

Método iterativo: Direcciones de descenso

Dado $x_0 \in S$, iteramos,

$$(\forall n \in \mathbb{N}) \quad \lfloor\ x_{n+1} = x_n + \alpha_n d_n,$$

Método de Newton: $\alpha_n d_n = -[\nabla^2 f(x_n)]^{-1} \nabla f(x_n)$

Si f es dos veces diferenciable, dado $x_0 \in S$, iteramos,

$$(\forall n \in \mathbb{N}) \quad \lfloor\ x_{n+1} = x_n - [\nabla^2 f(x_n)]^{-1} \nabla f(x_n),$$

Método Quasi Newton: $d_n = -H_n \nabla f(x_n)$

H_n es una matriz simétrica y definida positiva que aproxima $[\nabla^2 f(x_n)]^{-1}$.

$$(\forall n \in \mathbb{N}) \quad \lfloor\ x_{n+1} = x_n - \alpha_n H_n \nabla f(x_n),$$

Convergencia del método de Newton

Teorema

Sea $f\colon \mathbb{R}^n \to R \cup \{+\infty\}$ una función propia y dos veces diferenciable en $\operatorname{dom}(f)$ el cual es abierto. Supongamos que existe $\overline{x} \in \mathbb{R}^n$ tal que $\nabla f(\overline{x}) = 0$ y $\nabla^2 f(\overline{x})$ es simétrica definida positiva y $\nabla^2 f$ es localmente Lipschitz continua en torno a $\overline{x}$. Entonces, existe $\rho > 0$ tal que si $x_0 \in B(\overline{x}, \rho)$, la sucesión $(x_n)_{n\in\mathbb{N}}$, generada por el metodo de Newton:

$$(\forall n \in \mathbb{N}) \qquad \lfloor\ x_{n+1} = x_n - [\nabla^2 f(x_n)]^{-1} \nabla f(x_n),$$

converge a $\overline{x}$. Adicionalmente

$$\limsup \frac{\|x_{n+1} - \overline{x}\|}{\|x_n - \overline{x}\|^2} < +\infty. \tag{2}$$

- La propiedad (2) dice que el método del gradiente converge *de manera cuadrática* a $\overline{x}$.

Ejercicios

Aplique el métodos del gradiente con paso constante y de Newton para las siguientes funciones. En cada caso grafique la sucesión obtenida y función objetivo con los valores de la sucesión en la función objetivo.

3. $f(x) = x^2$. ¿En cuantas iteraciones converge el método Newton?
4. $f(x) = x \arctan(x) - \dfrac{1}{2}\log(1+x^2)$. Pruebe diferentes pasos iniciales, por ejemplo $x_0 = 0.1$ y $x_0 = 5$.
5. $f(x) = x^\top Q x - b^\top x$, donde Q es una matriz simétrica y definida positiva de orden n y $b \in \mathbb{R}^n$. Pruebe diferentes valores de n.
6. $f(x, y) = (2-x)^2 + 5(y - x^2)^2$. Pruebe diferentes puntos iniciales x_0 de manera aleatoria. Dependiendo de x_0, el Hessiano puede ser no invertible.

Métodos Quasi-Newton

Método Quasi Newton: $d_n = -H_n \nabla f(x_n)$

$$(\forall n \in \mathbb{N}) \quad \lfloor \; x_{n+1} = x_n - \alpha_n H_n \nabla f(x_n),$$

Algunas elecciones clásicas: $\gamma_n = \nabla f(x_{n+1}) - \nabla f(x_n)$, $\delta_n = x_{n+1} - x_n$

- Corrección de rango uno: $H_{n+1} = H_n + \dfrac{(\delta_n - H_n\gamma_n)(\delta_n - H_n\gamma_n)^\top}{\gamma_n^\top(\delta_n - H_n\gamma_n)}.$
- Formula DFP (Davidon- Fletcher-Powell):
$$H_{n+1} = H_n + \frac{\delta_n\delta_n^\top}{\gamma_n^\top\delta_n} - \frac{H_n\gamma_n\gamma_n^\top H_n}{\gamma_n^\top H_n\gamma_n}.$$
- Formula BFGS (Broyden-Fletcher-Goldfarb-Shanno):
$$H_{n+1} = H_n + \left(1 + \frac{\gamma_n^\top H_n\gamma_n}{\gamma_n^\top\delta_n}\right)\frac{\delta_n\delta_n^\top}{\gamma_n^\top\delta_n} - \frac{H_n\gamma_n\delta_n^\top + \delta_n\gamma_n^\top H_n}{\gamma_n^\top\delta_n}.$$

Método Quasi-Newton

- La regla BFGS es una de las más estables computacionalmente.
- Existen teoremas que garantizan la convergencia a un mínimo (local) del método quasi Newton con formula BFGS y paso dado por la regla de Wolfe.
- En general, el peor escenario de convergencia global de los métodos quasi Newton no son mejores que los métodos tipo gradiente. Más aún, estos métodos requieren guardar en cada iteraciones matrices de orden $n \times n$ lo que eleva considerablemente el costo computacional.

Métodos de Gradiente Conjugado

Métodos de Gradiente Conjugado: $d_n = -\nabla f(x_n) + \beta_n d_{n-1}$

Dado $x_0 \in \mathbb{R}^n$, definir $d_{n-1} = 0$.

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} d_n = -\nabla f(x_n) + \beta_n d_{n-1} \\ x_{n+1} = x_n + \alpha_n d_n, \end{array} \right.$$

Algunas elecciones clásicas:

- Dai–Yuan: $\beta_n = \dfrac{\|\nabla f(x_n)\|^2}{(\nabla f(x_n) - \nabla f(x_{n-1}))^\top d_{n-1}}$.
- Fletcher–Reeves: $\beta_n = \dfrac{\|\nabla f(x_n)\|^2}{\|\nabla f(x_{n-1})\|^2}$.
- Polak–Ribbiere: $\beta_n = \dfrac{(\nabla f(x_n) - \nabla f(x_{n-1}))^\top \nabla f(x_n)}{\|\nabla f(x_n)\|^2}$.

Método Gradiente Conjugado

- Los métodos tipo gradiente conjugado son motivados a partir de problemas cuadráticos.
  - En este tipo de problemas, el algoritmo genera una sucesión tal que los gradientes de la función objetivo son ortogonales.
  - En este caso, el algoritmo termina en a lo más n iteraciones (la dimensión del problema).
  - Además, todas las formas de elegir β_n coinciden.
- Las iteraciones son menos costosas que en los métodos de Newton y Quasi Newton.

Método Gradiente Conjugado

- En el caso no lineal general, se pierde la interpretación de gradientes ortogonales.
- Para el caso general, luego de n iteraciones, se puede hacer un *reinicio*, es decir tomar β_n, para garantizar convergencia (por la convergencia del gradiente clásico).
- Para **Fletcher–Reeves**, si f tiene gradiente Lipschitz y su curva de nivel en el mínimo local es acotada, utilizando una busqueda lineal tipo Wolfe, se puede probar que $\liminf \|\nabla f(x_n)\| = 0$.
- El resultado anterior también se puede probar para **Polak-Ribbiere** si adicionalmente f es fuertemente convexa y se utiliza una búsqueda lineal exacta.
- Si f no es convexa, se puede construir un ejemplo en $\mathbb{R}^3$ donde el método de **Polak-Ribbiere** no converge a un punto donde $\nabla f(x) = 0$.

Ejercicio

7. Resuelta el problema

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} x^\top Q x - b^\top x,$$

donde Q es una matriz simétrica y definida positiva de orden n y $b \in \mathbb{R}^n$ mediante:

- Gradiente paso constante
- Newton
- Quasi-Newton BFGS y la regla de Wolfe (ver [3, Paginas 61 y 62]).
- Gradiente Conjugado Fletcher-Reeves

Problemas con restricciones

Problemas con restricciones

$$\begin{aligned} \min \quad & f(x) \\ & g_i(x) \leq 0, i = 1, \ldots, p, \\ & h_i(x) = 0, i = 1, \ldots, q. \end{aligned}$$

Métodos para resolver este problema de una manera aproximada.

- Lagrangiano y Lagrangiano aumentado.
- Barrera
- Penalización

Método Lagrangiano y Lagrangiano aumentado

$$\begin{aligned} \min \quad & f(x) \\ & h_i(x) = 0, i = 1, \ldots, q. \end{aligned}$$

- Restricciones de desigualdad: se puede incorporar restricciones $g_i(x) + y_i^2 = 0$ donde $y_i \in \mathbb{R}$ es una variable auxiliar llamada holgura.
- Se considera el Lagrangiano (aumentado $r > 0$)

$$L(x, \lambda, r) = f(x) + \sum_{j=1}^{q} \lambda_j h_j(x) + \frac{r}{2} \sum_{k=1}^{q} h_j^2(x).$$

- Esquema: Para $\lambda_0 \in \mathbb{R}$ y $r_0 > 0$
  1. $x_{n+1} = \arg\min_{x \in \mathbb{R}^n} L(x, \lambda_n, r_n)$.
  2. Si $h_j(x_{n+1}) \approx 0$, detenerse.
  3. Si no, repetir 1 con $\lambda_{n+1} = \lambda_n + r_{n+1} h_j(x_{n+1})$.
- La convergencia de este método es bajo hipótesis fuertes sobre los multiplicadores, los gradientes (independencia lineal) y condiciones de segundo.

Método de barrera

$$\begin{aligned} \min \quad & f(x) \\ & g_i(x) \leq 0, i = 1, \dots, p. \end{aligned}$$

- $F : \mathbb{R}^n \to \mathbb{R}^m$ es una función barrera para $S \subset \mathbb{R}^n$ cerrado y de interior no vacío, si F es continua, no negativa y $F(x) \to +\infty$ si x se acerca a un valor en la frontera de S (se asume la condición de Slater).
- Esquema: $(\lambda_n)_{n\in\mathbb{N}}$ positiva tal que $\lambda_n \to +\infty$ y $x_0 \in \mathbb{R}^n$ que satisface Slater, iteramos
  1. $x_{n+1} = \arg\min_{x\in\mathbb{R}^n} f(x) + \frac{1}{\lambda_n} F(x)$.
- La idea es resolver 1 con un método iterativo comenzando en x_n.
- Hecho: Si la función de barrera F es acodada por abajo en el interior del conjunto restricción C, entonces $min_x f(x) + \dfrac{1}{\lambda_n} F(x) \to \min_C f$.
  - Barrera tipo potencia $F(x) = \sum_{j=1}^{p} \frac{1}{-g_j(x)^p}, \quad p \geq 1.$
  - Barrera logarítmica $F(x) = -\sum_{j=1}^{p} \ln(-g_j(x))$.
  - Barrera exponencial $F(x) = -\sum_{j=1}^{p} \exp\left(\frac{1}{-g_j(x)}\right)$.

Ejercicio

8. Resuelva el problema

$$\min_{(x,y)\in\mathbb{R}^2} x^2 + y^2$$

$$x + y \geq 4$$

$$2x + y \geq 5$$

mediante el método de barrera logarítmica. Compare la solución numérica con la analítica.

Método de penalización

$$\begin{aligned} \min \quad & f(x) \\ & g_i(x) \leq 0, i = 1, \ldots, p, \\ & h_i(x) = 0, i = 1, \ldots, q. \end{aligned}$$

- Penalización: $F \colon x \mapsto \sum_{i=1}^{p} \phi(g_i(x)) + \sum_{j=1}^{q} \psi(h_j(x))$, donde

$$\phi(t) \in \begin{cases} \{0\} \text{ si } t \leq 0 \\ ]0, +\infty[\text{ si } t > 0 \end{cases} \qquad \psi(t) \in \begin{cases} \{0\} \text{ si } t = 0 \\ ]0, +\infty[\text{ si } t \neq 0. \end{cases}$$

- Esquema: $(\lambda_n)_{n \in \mathbb{N}}$ positiva tal que $\lambda_n \to +\infty$ y $x_0 \in \mathbb{R}^n$, iteramos
  1. $x_{n+1} = \arg\min_{x \in \mathbb{R}^n} f(x) + \lambda_n F(x)$.
- La idea es resolver 1 con un método iterativo comenzando en x_n.
- Hecho: Si $(x_n)_{n \in \mathbb{N}}$ es acotada, todo punto de acumulación es solución.
- Se puede tomar
  - $\phi(t) = (\max\{0, t\})^m$, $m \in \mathbb{N}$.
  - $\psi(t) = |t|^m$.

Ejercicio

9. Dado $\alpha \in \mathbb{R}$. Resuelva el problema

$$\begin{aligned} \min_{(x,y,w,z)\in\ \mathbb{R}^4} \quad & x^2 + y^2 + w^2 + z^2 \\ \text{s.a.} \quad & x + y + w + z = 1 \\ & z \leq \alpha. \end{aligned}$$

mediante el método de penalización cuadrática. Compare la solución numérica con la analítica.

Problemas Convexos

Problema de Optimización Convexa

$$\min_{x \in \mathbb{R}^n} f(x)$$

- $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ es una función convexa.

Resultados Previos

Lema de Opial

Sea $(x_n)_{n\in\mathbb{N}}$ una sucesión en $\mathbb{R}^n$ y sea C subconjunto de $\mathbb{R}^n$ no vacío. Suponga que

1. Para todo $x \in C$, $(\|x_n - x\|)_{n\in\mathbb{N}}$ es una sucesión convergente; y
2. Todo punto de acumulación de $(x_n)_{n\in\mathbb{N}}$ pertenece a C.

Entonces, $(x_n)_{n\in\mathbb{N}}$ converge a algún $x \in C$.

Lema *de* sucesiones

Sean $(\alpha_n)_{n\in\mathbb{N}}$ y $(\beta_n)_{n\in\mathbb{N}}$ sucesiones en $[0, +\infty[$ tales que

$$(\forall n \in \mathbb{N}) \quad \alpha_{n+1} \leq \alpha_n - \beta_n. \tag{3}$$

Entonces $(\alpha_n)_{n\in\mathbb{N}}$ es convergente y $\sum_{n\in\mathbb{N}} \beta_n < +\infty$.

Resultados Previos

Lema de descenso

Sea $f\colon \mathbb{R}^n \to \mathbb{R}$ una función Gâteux diferenciable tal que ∇f es β-Lipschitz para $\beta \in]0, +\infty[$. Entonces

$$\forall (x, y) \in \mathcal{H}^2 \quad f(y) \leq f(x) + \langle \nabla f(x) \mid y - x \rangle + \frac{\beta}{2}\|x - y\|^2. \tag{4}$$

Teorema de Baillon-Haddad

Sea $f\colon \mathbb{R}^n \to \mathbb{R}$ una función convexa y diferenciable (Fréchet) y sea $\beta \in]0, +\infty[$. Entonces, ∇f es β^{-1}-cocoercivo si y solamente si ∇f es β-Lipschitz.

Sea $D \subset \mathcal{H}$ no vacío, $T\colon D \to \mathbb{R}^n$ y $\beta \in]0, +\infty[$. T es β^{-1}-cocoercivo si

$$(\forall (x, y) \in D^2) \quad \langle x - y \mid Tx - Ty \rangle \geq \frac{1}{\beta}\|Tx - Ty\|^2.$$

Convergencia Método del Gradiente

Teorema

Sea $f\colon \mathbb{R}^n \to \mathbb{R}$ una función continuamente diferenciable tal que ∇f es β-Lipschitz para $\beta \in \,]0, +\infty[$. Suponga además que f es acotada inferiormente y $\arg\min f \neq \varnothing$. Sea $\tau \in]0, 2/\beta[$, sea $x_0 \in \mathbb{R}^n$, y considere la siguiente sucesión

$$x_{n+1} = x_n - \tau \nabla f(x_n). \tag{5}$$

Entonces, las siguientes afirmaciones son verdaderas

1. $(f(x_n))_{n\in\mathbb{N}}$ es decreciente.
2. $\sum_{n\in\mathbb{N}} \|\nabla f(x_n)\|^2 < +\infty$.
3. $\lim_{n\to\infty} \nabla f(x_n) = 0$.
4. Si f es convexa, entonces todo punto de acumulación de $(x_n)_{n\in\mathbb{N}}$ minimiza f.
5. Si f es convexa, $(x_n)_{n\in\mathbb{N}}$ converge a algún punto en $\arg\min f$.

Demostración

1. Del Lema des descenso se sigue que

$$f(x_{n+1}) - f(x_n) \leq \langle \nabla f(x_n) \mid x_{n+1} - x_n \rangle + \frac{\beta}{2} \|x_{n+1} - x_n\|^2$$
$$= \frac{\tau(\tau\beta - 2)}{2} \|\nabla f(x_n)\|^2.$$

   Ya que $\tau < 2/\beta$, se tiene, para todo $n \in \mathbb{N}$, $f(x_{n+1}) < f(x_n)$.

2. Sumando en $\mathbb{N}$ la desigualdad anterior, se sigue que

$$\frac{\tau(2 - \beta\tau)}{2} \sum_{n \in \mathbb{N}} \|\nabla f(x_n)\|^2 \leq f(x_0) - \inf(f) < +\infty.$$

3. Directo de 2.

4. Sea $(x_{n_k})_{k\in\mathbb{N}}$ una subsucesión de $(x_n)_{n\in\mathbb{N}}$ tal que $x_{n_k} \to x \in \mathbb{R}^n$. Ahora, por la desigualdad del subdiferencial

$$(\forall y \in \mathcal{H})(\forall k \in \mathbb{N}) \quad \langle \nabla f(x_{n_k}) \mid y - x_{n_k} \rangle + f(x_{n_k}) \leq f(y). \tag{6}$$

Debido a que $\nabla f(x_{n_k}) \to 0$ y $(x_{n_k})_{k\in\mathbb{N}}$ es acotada, se tiene que

$$\langle \nabla f(x_{n_k}) \mid y - x_{n_k} \rangle \to 0.$$

Por otro lado, la continuidad de f nos permite concluir

$$(\forall y \in \mathbb{R}^n) \quad f(x) \leq f(y). \tag{7}$$

5. Para $x \in \arg\min f$, $\nabla f(x) = 0$. Entonces, por el Teorema de Baillon-Haddad, se sigue que:

$$\begin{aligned}
&\|x_{n+1} - x\|^2 \\
&= \|x_{n+1} - x_n\|^2 + 2\langle x_{n+1} - x_n \mid x_n - x\rangle + \|x_n - x\|^2 \\
&= \tau^2\|\nabla f(x_n)\|^2 + 2\tau\langle \nabla f(x_n) \mid x - x_n\rangle + \|x_n - x\|^2 \\
&= \tau^2\|\nabla f(x_n) - \nabla f(x)\|^2 + 2\tau\langle \nabla f(x_n) - \nabla f(x) \mid x - x_n\rangle + \|x_n - x\|^2 \\
&\leq -\frac{\tau}{\beta}(2 - \tau\beta)\|\nabla f(x_n) - \nabla f(x)\|^2 + \|x_n - x\|^2.
\end{aligned}$$

Por lema de las sucesiones, concluimos que $(\|x_n - x\|)_{n\in\mathbb{N}}$ es convergente. El resultado se sigue utilizando Lema de Opial.

Convergencia Método de Newton

Teorema

Sea $f\colon \mathbb{R}^n \to \mathbb{R}$, $x \in \arg\min f$ y suponga que f es dos veces Gâteux diferenciable en una vecindad de x, su Hessiano $\nabla^2 f$ es continuo en x y $\nabla^2 f(x)$ es invertible. Sea $(x_n)_{n\in\mathbb{N}}$ genera por el método de Newton. Entonces:

1. Existe una vecindad $U \subset \mathbb{R}^n$ de x tal que para $x_0 \in U$ y todo $\varepsilon \in]0,1[$ existe $R > 0$ tal que si $\|x_N - x\| < R$ para algún $N \in \mathbb{N}$, entonces
$$(\forall m \in \mathbb{N}) \quad \|x_{N+m} - x\| \leq R\varepsilon^m. \tag{8}$$
2. Si $\nabla^2 f$ es β-Lipschitz en U para $\beta \in]0,+\infty[$ y $\|x_N - x\| \leq \frac{2\varepsilon}{\beta C}$ donde $C \in]0,+\infty[$ es tal que $\sup_{x\in U} \|[\nabla^2 f(z)]^{-1}\| \leq C$, entonces
$$(\forall m \in \mathbb{N}) \quad \|x_{N+m} - x\| \leq \frac{2}{\beta C}\varepsilon^{2m}. \tag{9}$$

Demostración: Ya que $\nabla^2 f$ es continuo en x y $\nabla^2 f(x)$ es invertible, existe una vecindad U de x donde f es estrictamente convexa, $\nabla^2 f(y)$ es invertible para todo $y \in U$, y existe $C > 0$ que $\sup_{x\in U} \|[\nabla^2 f(z)]^{-1}\| \leq C$.

1. Sea $n \in \mathbb{N}$ fijo. Definamos $g\colon [0,1] \to \mathbb{R}^n$ dada por $g(t) = \nabla f(x + t(x_n - x))$. Note que

$$\nabla f(x_n) = g(1) - g(0) = \int_0^1 g'(t)\,dt = \int_0^1 \nabla^2 f(x + t(x_n - x))(x_n - x)\,dt.$$

Tomando $R > 0$ tal que si $\|y - x\| < R$ entonces $\|\nabla^2 f(y) - \nabla^2 f(x)\| \leq \varepsilon / C$, se tiene

$$\begin{aligned}
\|x_{n+1} - x\| &= \|x_n - x - [\nabla^2 f(x_n)]^{-1} \nabla f(x_n)\| \\
&= \|[\nabla^2 f(x_n)]^{-1} (\nabla^2 f(x_n)(x_n - x) - \nabla f(x_n)\| \\
&\leq C \left\| \int_0^1 [\nabla^2 f(x_n) - \nabla^2 f(x + t(x_n - x))](x_n - x)\,dt \right\| \\
&\leq C\|x_n - x\| \int_0^1 \|\nabla^2 f(x_n) - \nabla^2 f(x + t(x_n - x))\|\,dt \\
&\leq \varepsilon \|x_n - x\|.
\end{aligned}$$

2. Si $\nabla^2 f$ es β-Lipschitz en U, de la penultima desigualdad se tiene

$$\begin{aligned} \|x_{n+1} - x\| &\leq C\beta\|x_n - x\|^2 \int_0^1 (1-t)\,dt \\ &= \frac{C\beta}{2}\|x_n - x\|^2. \end{aligned}$$

Considerado que $\|x_N - x\| \leq \frac{2\varepsilon}{\beta C}$, el resultado se sigue inductivamente.

Problemas Convexos no diferenciables

Método Proximal

Teorema

Sea $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ convexa y s.c.i., tal que $\arg\min_f \neq \varnothing$. Sea $\gamma \in]0, +\infty[$, $x_0 \in \mathbb{R}^n$ y considere el algoritmo proximal dado por

$$x_{n+1} = \mathrm{prox}_{\gamma f}(x_n) := \arg\min_z \left(\gamma f(z) + \frac{1}{2}\|z - x_n\|^2 \right) \tag{10}$$

Entonces, $(x_n)_{n\in\mathbb{N}}$ converge a algún elemento en $\arg\min_f$.

- Note que directamente se tiene

$$f(x_{n+1}) + \frac{1}{2\gamma}\|x_{n+1} - x_n\|^2 \leq f(x_n).$$

Método Gradiente-Proximal

Problema que involucra dos funciones

$$\min_{x\in\mathbb{R}^n} f(x) + g(x). \tag{11}$$

Teorema

Sea $f\colon \mathbb{R}^n \to \mathbb{R}$ una función continuamente diferenciable tal que ∇f es β-Lipschitz para $\beta \in \,]0, +\infty[$. Sea $g\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ convexa y s.c.i. Suponga que tal que $\arg\min_{(f+g)} \neq \varnothing$. Sea $\gamma \in]0, 2/\beta[$, $x_0 \in \mathbb{R}^n$ y considere el algoritmo del gradiente proximal dado por

$$x_{n+1} = \mathrm{prox}_{\gamma g}(x_n - \gamma \nabla f(x_n)) \tag{12}$$

Entonces, $(x_n)_{n\in\mathbb{N}}$ converge a algún elemento en $\arg\min_{(f+g)}$.

Aplicación: Tratamiento de imágenes

Una imagen consiste en una matriz donde cada entrada corresponde a un pixel que da el color a esta. Supongamos que tomamos una fotografía $x \in \mathbb{R}^{n \times m}$ pero que está dañada por un ruido en el proceso de toma de la imágen.

$$b = x + \varepsilon$$

- b es la observación.
- x es la imagen sin ruido.
- ε es un ruido que distorciona la imagen.

Modelo de restauración

Problema de optimización convexo

$$\min_{x\in\mathbb{R}^{n\times m}} \frac{1}{2}\|x-b\|^2 + g(x). \tag{13}$$

- El término $\frac{1}{2}\|x-b\|^2$ se conoce como término de fidelidad con los datos.
- g debe ser una función que disminuya el ruido, conocida como regularización, por ejemplo

$$g(x) = \lambda h_\rho(Wx) \tag{14}$$

donde $\lambda \in]0,+\infty[$, h_ρ es la función de Huber y W es la transformada wavelets.

Modelo más sofisticado

$$b = Tx + \varepsilon$$

- T es un operador lineal que modela el proceso de obtención de la imágen.

$$\min_{x \in [0,1]^{n \times m}} \frac{1}{2} \| Tx - b \|^2 + \| \nabla x \|_1. \tag{15}$$

- Se añade la restricción en los pixeles: $x \in [0,1]^{n \times m}$
- ∇ es el gradiente discreto que incluye diferencias verticales y horizontales
- $\| \cdot \|_1$ es la norma 1 (suma de valores absolutos) no es diferenciable y $\mathrm{prox}_{\| \nabla \cdot \|_1}$ no tiene un expresión explicita

Bibliografía [1] [2] [3]

Bauschke, H.H., Combettes, P.L.: Convex analysis and monotone operator theory in Hilbert spaces, second edn.
CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, Cham (2017).
DOI 10.1007/978-3-319-48311-5

Nesterov, Y.: Lectures on convex optimization, *Springer Optimization and Its Applications*, vol. 137, second edn.
Springer, Cham (2018).
DOI 10.1007/978-3-319-91578-4

Nocedal, J., Wright, S.J.: Numerical optimization, second edn.
Springer Series in Operations Research and Financial Engineering. Springer, New York (2006)