Optimización continua no restringida

2 de septiembre de 2003

1. Optimización en una variable

- Problema general:

$$\text{minimizar } \varphi(x), \quad \varphi : \mathbb{R} \to \mathbb{R}$$

- Caso especial muy útil:

$$\text{minimizar } \varphi(x) = f(x + \lambda d)$$

con: $f : \mathbb{R} \to \mathbb{R}$, $x, d \in \mathbb{R}$fijos, $\lambda \in \Lambda$

- Posibles dominios:
  - $\Lambda = \mathbb{R}$
  - $\Lambda = [0, \infty]$
  - $\Lambda = [0, \lambda_{max}]$

1.1. Criterios de terminación

- Generalmente, se construye una sucesión $\{\lambda_k\}$ tal que

$$\lim_{k \to \infty} \lambda_k = \lambda^*,$$

donde λ^* es un minimizador global o local.

- $\{\lambda_k\}$ es infinita, de manera que el proceso debe detenerse para un valor de k finito. Algunos criterios de terminación son:
  - $|\lambda_{k+1} - \lambda_k| \leq \epsilon_\lambda$
  - $\left|\frac{\lambda_{k+1} - \lambda_k}{\lambda_k}\right| \leq \epsilon'_\lambda$, u opcionalmente $\left|\frac{\lambda_{k+1} - \lambda_k}{\lambda_k + 1}\right| \leq \epsilon'_\lambda$
  - $|\varphi(\lambda_{k+1}) - \varphi(\lambda_k)| \leq \epsilon_\varphi$

- $|\varphi'(\lambda_k)| \leq \epsilon_{\varphi'}$ (si se busca un punto crítico)
- $b_k - a_k \leq \epsilon$ (si se construyen intervalos de certidumbre $[a_k, b_k]$)

- Cómo escoger ϵ?
  - Debe ser dependiente en la precisión, la cual depende a su vez de la plataforma (compilador y máquina).
  - $\epsilon_{\text{maq}} = \min\{\epsilon : 1 + \epsilon' \neq' 1\}$ ($' \neq'$ de acuerdo a la plataforma)
  - $\epsilon = 1$
$\epsilon = 1 + \epsilon$
mientras $s \neq 1$
  $\epsilon = \epsilon/2$
  $s = 1 + \epsilon$
fin_mientras
$\epsilon'_{\text{maq}} = 2\epsilon$

1.2. Método de Newton

- Es un método para resolver $g(\lambda) = 0$
- $\lambda_{k+1} = \lambda_k - \frac{g(\lambda_k)}{g'(\lambda_k)}$
- Se puede usar como método de minimización (maximización) resolviendo $\varphi'(\lambda) = 0$

$$\lambda_{k+1} = \lambda_k - \frac{\varphi'(\lambda_k)}{\varphi"(\lambda_k)}$$

- Desventajas:
  - puede no converger,
  - puede converger a un maximizador,
  - se requiere $\varphi"(\lambda_k) \neq 0$,
  - se necesitan primeras y segundas derivadas de φ.

1.3. Método de la secante

- La idea es reemplazar en el método de Newton la segunda derivada ($\varphi"$) por una aproximación:

$$\varphi"(\lambda_k) = \frac{\varphi'(\lambda_k) - \varphi'(\lambda_{k+1})}{\lambda_k - \lambda_{k+1}}$$

- $\lambda_{k+1} = \lambda_k - \frac{\lambda_{k-1}\varphi'(\lambda_k) - \lambda_k\varphi'(\lambda_{k+1})}{\varphi'(\lambda_k) - \varphi'(\lambda_{k+1})}$
- No es necesario evaluar la segunda derivada.

- Desventajas:
  - puede no converger,
  - puede converger a un maximizador,
  - se requiere $\varphi'(\lambda_k) - \varphi'(\lambda_{k-1}) \neq 0$,
  - se necesita calcular primeras \ derivadas de φ.
- En general, es posible reemplazar las primeras derivadas en el método de Newton por aproximaciones numéricas.

1.4. Métodos de encajonamiento

- La idea es refinar iterativamente un intervalo $[a_k, b_k]$ (llamado intervalo de certidumbre).
- Cada sucesivo intervalo debe estar contenido en el anterior:

$$a_k \leq a_{k+1} \text{ y } b_k \geq b_{k+1}$$

- La longitud de los intervalos debe tender a 0:

$$\lim_{k \to \infty} b_k - a_k = 0$$

- Converge al valor óptimo si:
  - La función es cuasiconvexa:

$$\forall x, y \in D, \forall \sigma \in [0, 1] \; f((1 - \sigma)x + \sigma y) \leq \max\{f(x), f(y)\}$$

  - El valor óptimo $\lambda^* \in [a_1, b_1]$

1.4.1. Búsqueda secuencial

- La idea es evaluar la función en m posiciones y usar estos valores para refinar el intervalo.
- Entradas: a, b, ϵ, m
  $l = b - a$
  mientras $l \geq \epsilon$
  $\delta = l/(m + 1)$
  $\mu_j = a + \delta j$, $j = 1, \ldots, m$
  $\mu^* = \arg\min\{\varphi(\mu_1), \ldots, \varphi(\mu_m)\}$
  $a = \mu^* - \delta$
  $b = \mu^* + \delta$
  $l = b - a$
  fin-mientras
  $\lambda^* = \min\{a, \mu^*, b\}$

- Cuál es el número óptimo de evaluaciones por iteración (m)?
  - El objetivo es minimizar el número total de evaluaciones: mk (k corresponde al número de iteraciones)
  - Asumiendo que el criterio de terminación es $b_k - a_k \leq \epsilon$, es posible demostrar que:

$$mk \approx \frac{m}{\log \frac{m+1}{2}} L_0 = c(m)L_0$$

donde $L_0 = \log \frac{b_0 - a_0}{\epsilon}$

m	$c(m)$
2	4.93
3	4.33
4	4.37
5	4.55
6	4.79
7	5.05

1.5. Otros métodos

- Sección dorada (o áurea)
  Método de encajonamiento que tiene un mejor desempeño que la búsqueda secuencial. Solo se realiza una evaluación por iteración y los puntos no están igualmente espaciados.

- Minimización por interpolación cuadrática
  La idea es encontrar tres valores t_1, t_2 y $t_3$que permitan obtener el minimizador t_p^*de la parábola que pasa por los puntos $(t_i, \varphi(t_i))$ $(i = 1, 2, 3)$. Este punto sirve como nuevo valor para la siguiente iteración.

2. Optimización en varias variables

- Problema general:

$$\text{minimizar } f(x), \quad f : \mathbb{R}^n \to \mathbb{R}$$

- Estrategia general (aplica a varios métodos):
- Dado un punto x^k se construye una dirección $d^k \neq 0$ y se obtiene λ_k^* el cual es un minimizador de $\varphi(\lambda) = f(x^k + \lambda d^k)$ cuando λ varía en $\mathbb{R}$:

$$\begin{aligned} \lambda_k^* &= \arg\min f(x^k + \lambda d^k),\ \lambda \in \mathbb{R} \\ x^{k+1} &= x^k + \lambda_k^* d^k \end{aligned}$$

2.1. Método de Newton

- Es una generalización del método para una variable:

$$x_{k+1} = x_k - \frac{g(x_k)}{g'(x_k)}$$

- La anterior fórmula puede ser reescrita como:

$$x_{k+1} = x_k - (g'(x_k))^{-1} g(x_k)$$

- Si $g : \mathbb{R}^n \to \mathbb{R}^n$, la fórmula se puede generalizar a:

$$x^{k+1} = x^k - (g'(x^k))^{-1} g(x^k)$$

donde $g'(x_k)$ corresponde al jacobiano de la función g.

- El método de Newton se expresa en dos partes:

$$\begin{aligned} \text{resolver} \quad f"(x^k) d^k &= -f'(x^k) \\ x^{k+1} &= x^k + d^k \end{aligned}$$

donde $f"$ es el hessiano y f' es el gradiente.

2.2. Método del gradiente o del descenso más pendiente

- La idea es escoger la dirección d^k como la dirección que produce mayor decenso. Esta dirección corresponde al gradiente multiplicado por -1.
- El algoritmo es el siguiente:

Entrada: x^0, ϵ_g, MAXIT
para $k = 0, \ldots,$MAXIT
  si $\left\| f'(x^k) \right\| \leq \epsilon_g$ **entonces parar**
  $d^k = -f'(x^k)$
  $\lambda_k^* = \arg\min f(x^k + \lambda d^k),\ \lambda \geq 0$
  $x^{k+1} = x^k + \lambda_k^* d^k$
fin-para

2.3. Método del gradiente conjugado

- El algoritmo es el siguiente:

Entrada: x^1, ϵ_g, MAXIT
para $K = 1, \ldots,$MAXIT
  para $k = 1, \ldots, n$
    si $\left\| f'(x^k) \right\| \leq \epsilon_g$ **entonces parar**
    si $k = 1$ **entonces** $d^k = -f'(x^k)$
    sino
      $\alpha_k = \frac{\left\| f''(x^k) \right\|_2^2}{\left\| f''(x^{k-1}) \right\|_2^2}$
      $d^k = -f'(x^k) + \alpha_k d^{k-1}$
    fin-sino
    $\lambda_k^* = \arg\min f(x^k + \lambda d^k),\ \lambda \geq 0$
    $x^{k+1} = x^k + \lambda_k^* d^k$
  fin-para
  $x^1 = x^{n+1}$
fin-para

2.4. Método cíclico coordenado

- El algoritmo es el siguiente:

Entrada: x^0, ϵ_x, MAXIT
para $k = 0, \ldots,$MAXIT
  $y^1 = x^k$
  para $j = 1, \ldots, n$
    $\lambda_j^* = \arg\min f(y^j + \lambda e^j)$
    $y^{j+1} = y^j + \lambda_j^* e^j$
  fin-para
  $x^{k+1} = y^{n+1}$
  si $\left\| x^{k+1} - x^k \right\| < \epsilon_x$ **entonces parar**
fin-para

Donde e^jcorresponde al vector $(e_1, ..., e_n)$ con $e_j = 1$ y $e_i = 0$ para todo $i \neq j$.