Optimización multidimensional

Lección 07

Dr. Pablo Alvarado Moya

CE3102 Análisis Numérico para Ingeniería
Área de Ingeniería en Computadores
Tecnológico de Costa Rica

II Semestre 2017

Contenido

Introducción

- Optimización multidimensional sin restricciones es central en ingeniería
- Existen versiones:
  - uniobjetivo y multiobjetivo
  - que usan gradiente y que no lo usan
  - heurísticos
- Aplicación uniobjetivo usual: se busca un conjunto de parámetros que optimiza un criterio dado (p. ej. potencia, velocidad, presión, etc.)

Métodos sin gradiente

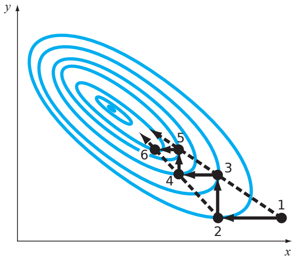
- Ventaja: no se requiere calcular el gradiente
- Desventaja: tienen convergencia lenta
- Algunos métodos:
  1. Búsqueda aleatoria
  2. Búsquedas univariables y búsquedas patrón
  3. Simplex en bajada

Búsqueda aleatoria

- Genere números aleatorios con alguna distribución probabilística en el rango de definición y evalúe la función en esos puntos.
- Seleccione valor máximo
- Ventajas:
  - funciona con discontinuidades y funciones sin derivadas
  - no se atasca en extermos locales
- Desventajas:
  - Ineficiente
  - Excesivo número de evaluaciones de la función
- Principios aleatorios se utilizan en métodos heurísticos más sofisticados

Búsqueda univariable y búsquedas patrón

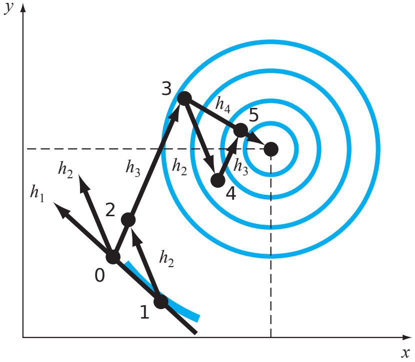
- Realice búsquedas lineales de máximos alternando los ejes
- Búsqueda lineal emplea los métodos vistos en clase anterior (Método de Brent)



- Direcciones patrón apuntan en dirección del máximo

Método de Powell

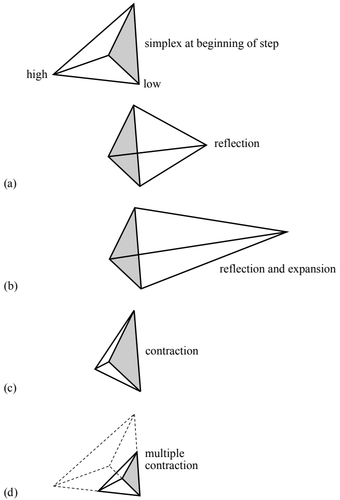
- Método de Powell explota direcciones patrón para encontrar máximo
- Solo se esboza idea general:



Simplex en bajada
Downhill simplex

- También conocido como “descenso de ameba” o algoritmo de Nelder-Mead
- Un **simplex** es la generalización de un triángulo. En espacio d-dimensional tiene $d+1$ vértices.
- Se inicia con cualquier simplex no degenerado, es decir, un simplex con volumen no nulo.
  - $\Rightarrow$ los vectores formados entre un vértice cualquiera del simplex y el resto engendran todo el espacio.
- El proceso paso a paso mueve un vértice del simplex a la vez para acercarse al extremo

Modificaciones del simplex



- Se toma siempre el vértice del simplex con *peor* valor asociado

  En imagen a la izquierda se busca minimizar
- Dicho vértice se *refleja*, pasándolo al otro lado del simplex pero manteniendo el volumen
- Si reflexión lleva a un mejor valor, hay expansión, y sino, a compresión del simplex
- Ejemplo 1
- Ejemplos 2

Gradiente

- El gradiente $\nabla f(\underline{\mathbf{x}})$ de una función indica la dirección de mayor cambio de la función

$$\nabla f(\underline{\mathbf{x}}) = \begin{bmatrix} \frac{\partial f(\underline{\mathbf{x}})}{\partial x_1} & \frac{\partial f(\underline{\mathbf{x}})}{\partial x_2} & \cdots & \frac{\partial f(\underline{\mathbf{x}})}{\partial x_n} \end{bmatrix}^T$$

  con $\underline{\mathbf{x}} = [x_1, x_2, \ldots, x_n]^T$
- Los extremos ocurren en puntos donde no hay cambio ($\nabla f(\underline{\mathbf{x}}) = \underline{\mathbf{0}}$)
- El cálculo del gradiente en problemas reales se realiza
  - Analíticamente
  - Por diferenciación numérica
  - Por diferenciación automática

Matriz Hessiana (1)

- Con la segunda derivada en el caso unidimensional se determina si un extremo es máximo o mínimo.
- Por ejemplo, si $f'(t_0) = \left.\frac{df(t)}{dt}\right|_{t=t_0} = 0$ entonces
  - $f(t_0)$ es máximo si $f''(t_0) < 0$
  - $f(t_0)$ es mínimo si $f''(t_0) > 0$
- El equivalente de la segunda derivada multidimensional es la matriz Hessiana (o hessiano):

$$\mathbf{H}(\underline{\mathbf{x}}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Matriz Hessiana (2)

- Se sabe para los extremos de $f(\underline{\mathbf{x}})$ que se encuentran donde $\nabla f(\underline{\mathbf{x}}) = 0$ y
  - Si $|\mathbf{H}| > 0$ y $\partial^2 f / \partial x^2 > 0$, entonces $f(x, y)$ tiene un mínimo local
  - Si $|\mathbf{H}| > 0$ y $\partial^2 f / \partial x^2 < 0$, entonces $f(x, y)$ tiene un máximo local
  - Si $|\mathbf{H}| < 0$, $f(x, y)$ tiene un punto de silla

Máxima inclinación (1)

Steepest descent/ascent

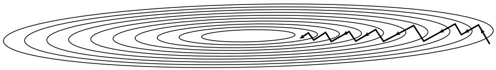
- Estrategia: seguir dirección del gradiente para maximizar (o dirección opuesta para minimizar)

$$\underline{\mathbf{x}}_{i+1} = \underline{\mathbf{x}}_i \pm \nabla f(\underline{\mathbf{x}}_i) \arg \min_s f(\underline{\mathbf{x}}_i \pm s \nabla f(\underline{\mathbf{x}}_i))$$

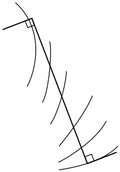
- Problema: si se minimiza linealmente en dirección del gradiente, nueva dirección del gradiente en mínimo siempre será ortogonal a la última dirección, lo que fuerza un efecto zig-zag:

Máxima inclinación (2)

Steepest descent/ascent



(a)



(b)

- Conforme se aproxima el extremo, los desplazamientos lineales disminuyen, lo que hace la convergencia cada vez más lenta

Regla Δ

- Una técnica utilizada para evitar el zig-zag es la llamada regla Δ

$$\underline{\mathbf{x}}_{i+1} = \underline{\mathbf{x}}_i \pm \underbrace{\lambda \nabla f(\underline{\mathbf{x}}_i)}_{\Delta} = \underline{\mathbf{x}}_i \pm \Delta$$

  donde la aplicación restringe el valor adecuado de λ
- De nuevo, el signo $\pm$ se elige como "-" para minimizar y "+" para maximizar.
- Una mala elección de λ conduce a secuencias largas antes de la convergencia (si se elige λ muy pequeño), o a oscilaciones indefinidas al acercarse al extremo (si se elige λ muy grande)
- Metodo se usa por facilidad de implementación.

Gradientes conjugados (1)

- Asúmase que la superficie es cuadrática y por tanto

$$f(\underline{\mathbf{x}}) \approx f(\underline{\mathbf{x}}_0) + (\underline{\mathbf{x}} - \underline{\mathbf{x}}_0)^T \nabla f(\underline{\mathbf{x}}_0) + \frac{1}{2}(\underline{\mathbf{x}} - \underline{\mathbf{x}}_0)^T \mathbf{H}(\underline{\mathbf{x}}_0)(\underline{\mathbf{x}} - \underline{\mathbf{x}}_0)$$

- Esta aproximación de Taylor es más exacta mientras más pequeña sea la vecindad alrededor de $\underline{\mathbf{x}}_0$
- Bajo esta suposición, la idea es

Direcciones conjugadas

Encontrar direcciones de minimización lineal tales que, la componente del gradiente que el paso anterior ya hizo cero no sea alterada en el nuevo paso

Gradientes conjugados (2)

- Supóngase que ya optimizamos en el paso anterior en la dirección $\underline{\mathbf{d}}_i$, iniciando en $\underline{\mathbf{x}}_i$, para llegar a un nuevo punto $\underline{\mathbf{x}}_{i+1}$. Entonces, por estar en un mínimo se debe cumplir

$$\underline{\mathbf{d}}_i^T \nabla f(\underline{\mathbf{x}}_{i+1}) = 0$$

- Necesitamos que a lo largo de la siguiente dirección $\underline{\mathbf{d}}_{i+1}$ la componente del gradiente paralela a la dirección anterior se mantenga cero:

$$\underline{\mathbf{d}}_i^T \nabla f(\underline{\mathbf{x}}_{i+1} + \lambda \underline{\mathbf{d}}_{i+1}) = 0$$

- Expandiendo en serie de Taylor con respecto a λ se puede demostrar que esto equivale a

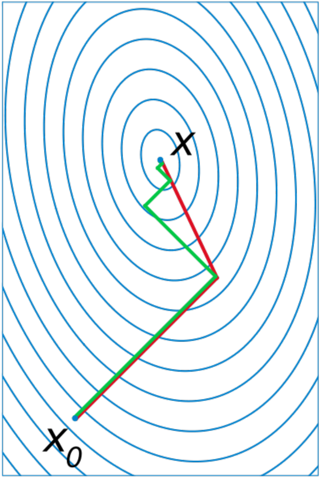
$$\underline{\mathbf{d}}_{i+1}^T \mathbf{H}(\underline{\mathbf{x}}_{i+1}) \underline{\mathbf{d}}_i = 0$$

Gradientes conjugados (3)

- Direcciones que cumplen esta propiedad se denominan **sin interferencia** o **conjugadas**.
- Polak y Ribiere demostraron que es posible obtener las direcciones conjugadas sin necesidad de calcular la matriz hessiana con:

$$\begin{aligned} \underline{\mathbf{d}}_{i+1} &= \pm\nabla f(\underline{\mathbf{x}}_{i+1}) + \beta_i \underline{\mathbf{d}}_i \\ \beta_i &= \frac{(\nabla f(\underline{\mathbf{x}}_{i+1}) - \nabla f(\underline{\mathbf{x}}_i))^T \nabla f(\underline{\mathbf{x}}_{i+1})}{\|\nabla f(\underline{\mathbf{x}}_i)\|^2} \end{aligned}$$

Gradientes conjugados (4)



Otros métodos con gradiente

- Método de Newton

$$\underline{\mathbf{x}}_{i+1} = \underline{\mathbf{x}}_i - \mathbf{H}^{-1}(\underline{\mathbf{x}}_i)\nabla f(\underline{\mathbf{x}}_i)$$

  - Converge de forma cuadrática cerca del óptimo
  - Requiere cálculo del hessiano e inversión de matriz
- Método de Levenberg-Marquardt
  - Similar al método de Newton, reemplazando el hessiano con

$$\tilde{\mathbf{H}}_i = \mathbf{H}_i + \alpha_i \mathbf{I}$$

    con $\alpha_i \mathbf{I}$ un término de regularización que se adapta durante el algoritmo
  - Al inicio se comporta como algoritmo de máxima inclinación y cerca del óptimo se comporta como algoritmo de Newton

Resumen

1. Introducción
2. Métodos sin gradiente
   - Búsqueda aleatoria
   - Búsqueda univariable y búsquedas patrón
   - Simplex en bajada
3. Métodos con gradiente
   - Máxima inclinación
   - Gradientes conjugados
   - Otros métodos

Este documento ha sido elaborado con software libre incluyendo LATEX, Beamer, GNUPlot, GNU/Octave, XFig, Inkscape, LTI-Lib-2, GNU-Make y Subversion en GNU/Linux