Institución Universitaria
Reacreditada en Alta Calidad

Optimización

Técnicas para NLPs

Docente: Cristian Guarnizo Lemus

Contenido

1. Métodos de Penalización y de Barrera.
2. Serie de problemas mas simples con Restricciones (LCL, SQP)

Métodos de Penalización y Barrera

- Idea: reemplazar el problema de restricciones por una secuencia de problemas de optimización sin restricciones. Como remover las restricciones?

- Método de penalización cuadrático (QPM): reemplazar las restricciones adicionando una penalización cuadrático en la función objetivo.
  - Aproximación desde los puntos infactibles.

- Método del Langrangiano Aumentado (ALM): Mejora de QPM que evita los mal-condicionamientos por medio de la estimación de los parámetros de Lagrange.

- Método de Log-Barrier (LBM): usar la barrera del logaritmo para forzar una estricta satisfacción de las desigualdades.
  - Aproximación desde los puntos factibles.

Método de Barrera logarítmica (LBM): desigualdades

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

$$\text{s.t.} \quad c_i(\boldsymbol{x}) \le 0,\ i \in I$$

- Reemplazar la restricciones por una barrera de logaritmo en la función objetivo:

$$P(\boldsymbol{x};\mu) = f(\boldsymbol{x}) - \mu \sum_{i \in I} \log[-c_i(\boldsymbol{x})]$$

  con parámetro de barrera $\mu > 0$.
- La barrera impone iteraciones estrictamente factibles
  - $P(\boldsymbol{x};u) \to \infty$ para $0 > c_i(\boldsymbol{x}) \to 0$. Entonces, para $\mu > 0$ se impone $\boldsymbol{c}(\boldsymbol{x}) < \boldsymbol{0}$.

- Similar a QPM, solucionar una secuencia de problemas sin restricción
  - Solucionar en una iteración el inicio de la siguiente iteración.
  - Mientras $\mu \to 0$, las aproximaciones de $\boldsymbol{x}$ se vuelven mejores.

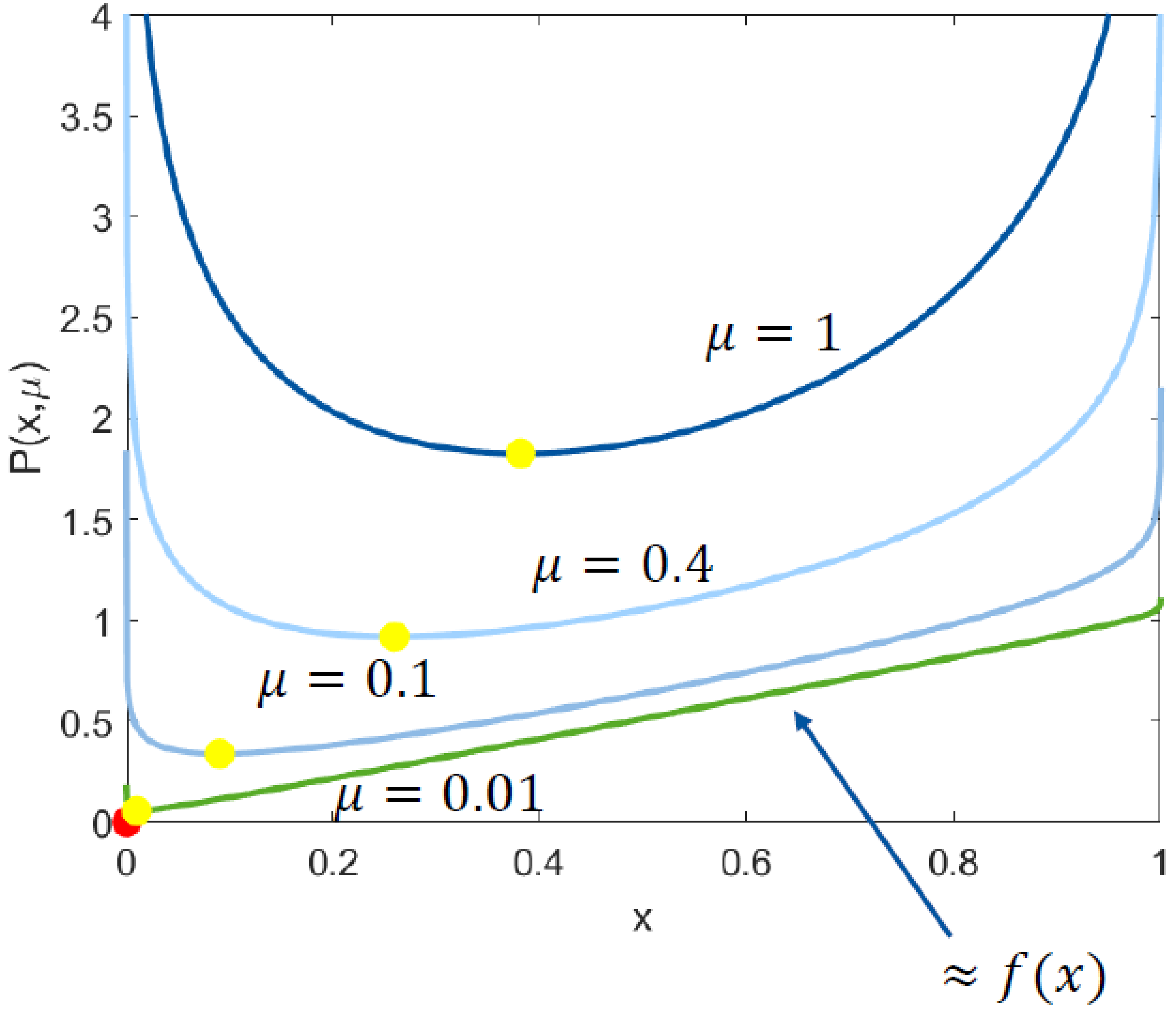
La barrera logarítmica LogBar

$$\min_{x \in R} \quad x$$

$$\text{s.t.} \quad x \geq 0$$

$$x \leq 1$$

$$P(x;\mu) = x - \mu(\log[x] + \log[1-x])$$



Barrera logarítmica (LBM): Igualdades y Desigualdades

Problema NLP general: $\min_{x \in R^n} \; f(\boldsymbol{x})$

s.t. $c_i(\boldsymbol{x}) = 0,\ i \in E$

$c_i(\boldsymbol{x}) \leq 0,\ i \in I$

- Reemplazar las restricciones de desigualdad por una barrera logarítmica
  Reemplazar las restricciones de igualdad por una penalización cuadrática

$$B(\boldsymbol{x};\mu) = f(\boldsymbol{x}) + \frac{1}{2\mu}\sum_{i \in E}[c_i(\boldsymbol{x})]^2 - \mu \sum_{i \in I} \log[-c_i(\boldsymbol{x})]$$

- Similar al caso de las desigualdades, solucionamos con una secuencia de problemas sin restricciones.
- Método de punto interior con respecto a: $c_i\left(x^{(k)}\right) < 0, i \in I$
- Violación de las restricciones de igualdad: $c_i\left(x^{(k)}\right) \neq 0, i \in E$
  - Las igualdades no tienen interior.

Chequeo

Cual es la idea principal en los métodos de barrera?

Cual es la diferencia entre los métodos de penalización cuadrático y el método de barrera?

Contenido

1. Métodos de Penalización y de Barrera.
2. Serie de problemas mas simples con Restricciones (LCL, SQP)

Problema de optimización no-lineal (NLP)

Formulación general:

$$\min_{\boldsymbol{x} \in D} f(\boldsymbol{x})$$

$$\text{s.t.} \quad c_i(\boldsymbol{x}) = 0, i \in E$$

$$c_i(\boldsymbol{x}) \leq 0, i \in I$$

$\boldsymbol{x} = [x_1, x_2, \dots, x_n]^\top \in D$ un vector (punto n-dimensional)

D conjunto anfitrión

$f: D \to R$ función objetivo

$c_i: D \to R$ funciones de restricción $\forall i \in E \cup I$

E el conjunto índice de las restricciones de igualdad

I el conjunto índice de las restricciones de desigualdad

- Tres estrategias de solución:
  - Eliminación de variables (convertir a un problema sin restricciones).
  - Aproximación como una serie de problemas sin restricción.
  - **Aproximación como una serie de problemas mas simples con restricciones**.

Método Lagrangiano con restricciones lineales

El método del Lagrangiano con restricciones lineales (LCL) es una modificación del método del Lagrangiano Aumentado.

- En cada paso, se linealizan las restricciones.
- Para el problema $\min_{x \in R^n} \; f(\boldsymbol{x})$

$$\text{s.t.} \quad c_i(\boldsymbol{x}) = 0, \quad i \in E$$

en la iteración k resolver $\min_{x \in R^n} \; F^{(k)}(\boldsymbol{x})$

$$\text{s.t.} \; \nabla c_i\left(\boldsymbol{x}^{(k)}\right)^T \left(\boldsymbol{x} - \boldsymbol{x}^{(k)}\right) + c_i\left(\boldsymbol{x}^{(k)}\right) = 0, \quad i \in E$$

- Para $F^{(k)}$, a menudo se selecciona el Lagrangiano Aumentado.

$$F^{(k)}(\boldsymbol{x}) = f(\boldsymbol{x}) + \sum_{i \in E} \lambda_i^{(k)} \bar{c}_i^{(k)}(\boldsymbol{x}) + \frac{1}{2\mu} \sum_{i \in E} \left[\bar{c}_i^{(k)}(\boldsymbol{x})\right]^2$$

$$\bar{c}_i^{(k)}(\boldsymbol{x}) = c_i(\boldsymbol{x}) - c_i\left(\boldsymbol{x}^{(k)}\right) - \nabla c_i\left(\boldsymbol{x}^{(k)}\right)^T \left(\boldsymbol{x} - \boldsymbol{x}^{(k)}\right)$$

Programación Secuencial Cuadrática (SQP)

- SQP provee la base para algunos buenos códigos de optimización.
- Se considera el siguiente problema de optimización $\min_{\boldsymbol{x}\in R^n} f(\boldsymbol{x})$

$$\text{s.t.} \quad c_i(\boldsymbol{x}) = 0, \quad i \in E$$

- Idea principal: solucionar una secuencia $\{k\}$ de QPs, donde se aproxima el NLP en la iteración $\boldsymbol{x}^{(k)}$ por una QP.
- La forma mas fácil: Aproximación por series de Taylor

$$\min_{\boldsymbol{p}\in R^n} \frac{1}{2}\boldsymbol{p}^T \nabla^2 f(\boldsymbol{x}^{(k)})\boldsymbol{p} + \left(\nabla f(\boldsymbol{x}^{(k)})\right)^T \boldsymbol{p}$$

$$\text{s.t.} \quad \left(\nabla \boldsymbol{c}(\boldsymbol{x}^{(k)})\right)^T \boldsymbol{p} + \boldsymbol{c}(\boldsymbol{x}^{(k)}) = \boldsymbol{0}$$

- Se puede interpretar como el método de Newton para solucionar las condiciones KKT.

Algoritmo Básico SQP

Puntos importantes no discutidos:

- Aproximación de la matriz Hessiana (p.e., actualización BFGS, en métodos quasi Newtons).
- Solución de las ecuaciones Newton-Lagrange en cada paso del QP.
- Criterio de parada.
- Inclusión de desigualdades.

Algoritmo Básico:

- Seleccionar $\boldsymbol{x}^{(0)}$.
- Para $k = 1,2,\ldots$

  - Calcular $f^{(k)} = f\left(\boldsymbol{x}^{(k-1)}\right), \nabla f^{(k)} = \nabla f\left(\boldsymbol{x}^{(k-1)}\right), \boldsymbol{c}^{(k)} = \boldsymbol{c}\left(\boldsymbol{x}^{(k-1)}\right), \boldsymbol{A}^{(k)} = \nabla \boldsymbol{c}\left(\boldsymbol{x}^{(k-1)}\right)$, actualizar $\boxed{\boldsymbol{B}^{(k)}}$.
  - Solucionar el QP para $\boldsymbol{p}$.
  - $\boldsymbol{x}^{(k)} = \boldsymbol{x}^{(k-1)} + \boldsymbol{p}$.
  - Si las condiciones de optimalidad se satisfacen, PARAR.

Método Quasi-Newton

Definiciones: $f^{(k)} = f(\boldsymbol{x}^{(k)})$, $\boldsymbol{g}^{(k)} := \nabla f(\boldsymbol{x}^{(k)})$ y $\boldsymbol{H}^{(k)} := \nabla^2 f(\boldsymbol{x}^{(k)})$

Del método de Newton:

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \boldsymbol{p}^{(k)}$$

$$[\nabla^2 f(\boldsymbol{x}^{(k)})]\boldsymbol{p}^{(k)} = -\nabla f(\boldsymbol{x}^{(k)}) \Rightarrow \boldsymbol{H}^{(k)}\boldsymbol{p}^{(k)} = -\boldsymbol{g}^{(k)}$$

Idea:

El sistema lineal $\boldsymbol{H}^{(k)}\boldsymbol{p}^{(k)} = -\boldsymbol{g}^{(k)}$ se puede solucionar aproximadamente por un método iterativo.

Observaciones:

Factorizaciones como LU o Cholesky – costo computacional alto.

Ocurren errores grandes para problemas mal condicionados.

No se requiere la solución exacta.

Método Quasi-Newton (2)

Idea: reducir la compejidad simplificando el calculo de $\boldsymbol{H}^{(k)}$ (Davidon):

- Reemplazar $\boldsymbol{H}^{(k)}$ por una aproximación $\boldsymbol{B}^{(k)}$.
- En vez de calular $\boldsymbol{B}^{(k)}$, buscamos una simple actualización usando la información de las ultimas iteraciones.

Metodo:

- Considerar la aproximación cuadrática de f en $\boldsymbol{x}^{(k)}$, $m^{(k)}(\boldsymbol{p}) = f^{(k)} + [\boldsymbol{g}^{(k)}]^T\boldsymbol{p} + \frac{1}{2}\boldsymbol{p}^T\boldsymbol{B}^{(k)}\boldsymbol{p}$.
- Condición de optimalidad de primer orden: $\boldsymbol{p}^{(k)} = -{\boldsymbol{B}^{(k)}}^{-1}\boldsymbol{g}^{(k)}$
- Por convexidad es necesaria y suficiente para la minimización de $m^{(k)}(\boldsymbol{p})$.
- Construir la aproximación cuadrática en $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k\boldsymbol{p}^{(k)}$

$$m^{(k+1)}(\boldsymbol{p}) = f^{(k+1)} + [\boldsymbol{g}^{(k+1)}]^T\boldsymbol{p} + \frac{1}{2}\boldsymbol{p}^T\boldsymbol{B}^{(k+1)}\boldsymbol{p}$$

- Que condiciones debe satisfacer $\boldsymbol{B}^{(k+1)}$?

Método Quasi-Newton (3)

Condiciones de $\boldsymbol{B}^{(k+1)}$:

1. Gradiente de $m^{(k+1)}$ en $\boldsymbol{x}^{(k+1)}$ debe ser igual al gradiente de f.

$\nabla m^{(k+1)}(\boldsymbol{p}) = \boldsymbol{g}^{(k+1)} + \boldsymbol{B}^{(k+1)}\boldsymbol{p}$	
At $\boldsymbol{x} = \boldsymbol{x}^{(k+1)}$, $\boldsymbol{p} = 0$ Queremos $\nabla m^{(k+1)}(0) = \boldsymbol{g}^{(k+1)}$	At $\boldsymbol{x} = \boldsymbol{x}^{(k)}$, $\boldsymbol{p} = -\alpha_k \boldsymbol{p}^{(k)}$ Queremos $\nabla m^{(k+1)}\left(-\alpha_k \boldsymbol{p}^{(k)}\right) = \boldsymbol{g}^{(k)}$
Satisface automáticamente	$\Rightarrow \boldsymbol{g}^{(k+1)} - \alpha_k \boldsymbol{B}^{(k+1)}\boldsymbol{p}^{(k)} = \boldsymbol{g}^{(k)}$ $\Rightarrow \boldsymbol{B}^{(k+1)} \underbrace{\alpha_k \boldsymbol{p}^{(k)}}_{=\boldsymbol{x}^{(k+1)}-\boldsymbol{x}^{(k)}} = \boldsymbol{g}^{(k+1)} - \boldsymbol{g}^{(k)}$ $\Rightarrow \boxed{\boldsymbol{B}^{(k+1)}\boldsymbol{s}^{(k)} = \boldsymbol{y}^{(k)}}$, donde $\boldsymbol{s}^{(k)} = \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}$ y $\boldsymbol{y}^{(k)} = \boldsymbol{g}^{(k+1)} - \boldsymbol{g}^{(k)}$

2. Debido que $\boldsymbol{B}^{(k+1)}$ es simétrica definida positiva: $\boldsymbol{s}^{(k)^T}\boldsymbol{B}^{(k+1)}\boldsymbol{s}^{(k)} > 0, \forall \boldsymbol{s}^{(k)} \neq \boldsymbol{0} \Rightarrow \boldsymbol{s}^{(k)^T}\boldsymbol{y}^{(k)} > 0$

Método Quasi-Newton (3)

Condiciones de $\boldsymbol{B}^{(k+1)}$:

$\boldsymbol{B}^{(k+1)}\boldsymbol{s}^{(k)} = \boldsymbol{y}^{(k)}$gives many solutions for $\boldsymbol{B}^{(k+1)}$

- Unique solution: $\boldsymbol{B}^{(k+1)}$ should be close to $\boldsymbol{B}^{(k)}$

$$\min_{\boldsymbol{B}} \left\|\boldsymbol{B} - \boldsymbol{B}^{(k)}\right\|_W \quad \leftarrow \text{weighted Frobenius-Norm}$$

$$\|A\|_W = \left\|W^{1/2} A W^{1/2}\right\|_F, \text{ for any } W \text{ s.t. } W y_k = s_k$$

$$\text{s.t.} \quad \boldsymbol{B}^T = \boldsymbol{B}$$

$$\|\cdot\|_F^2 : R^{n \times n} \to R_{\geq 0}, \|C\|_F^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}^2$$

$$\boldsymbol{B}\boldsymbol{s}^{(k)} = \boldsymbol{y}^{(k)}$$

$$\Rightarrow \boldsymbol{B}^{(k+1)} = \left(\boldsymbol{I} - \frac{1}{\boldsymbol{y}^{(k)^T}\boldsymbol{s}^{(k)}} \boldsymbol{y}^{(k)}\boldsymbol{s}^{(k)^T}\right) \boldsymbol{B}^{(k)} \left(\boldsymbol{I} - \frac{1}{\boldsymbol{y}^{(k)^T}\boldsymbol{s}^{(k)}} \boldsymbol{s}^{(k)}\boldsymbol{y}^{(k)^T}\right) + \frac{1}{\boldsymbol{y}^{(k)^T}\boldsymbol{s}^{(k)}} \boldsymbol{y}^{(k)}\boldsymbol{y}^{(k)^T} \quad \to \text{DFP formula}$$

$$\Rightarrow \boldsymbol{B}^{(k+1)^{-1}} = \left(\boldsymbol{I} - \frac{1}{\boldsymbol{y}^{(k)^T}\boldsymbol{s}^{(k)}} \boldsymbol{s}^{(k)}\boldsymbol{y}^{(k)^T}\right) \boldsymbol{B}^{(k)^{-1}} \left(\boldsymbol{I} - \frac{1}{\boldsymbol{y}^{(k)^T}\boldsymbol{s}^{(k)}} \boldsymbol{y}^{(k)}\boldsymbol{s}^{(k)^T}\right) + \frac{1}{\boldsymbol{y}^{(k)^T}\boldsymbol{s}^{(k)}} \boldsymbol{s}^{(k)}\boldsymbol{s}^{(k)^T} \quad \to \text{BFGS formula}$$

Institución Universitaria
Reacreditada en Alta Calidad

¡Gracias!

Somos Innovación Tecnológica con Sentido Humano

Alcaldía de Medellín