

Problema de optimización: $\min f(x)$ con $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $c(x): \mathbb{R}^n \rightarrow \mathbb{R}^m$

$$\text{s.a. } c_i(x) \geq 0$$

f convexa, c linear o bien continuamente diferenciable

Condiciones de optimalidad (KKT):

$x^* \in \mathbb{R}^n$ es sol. del problema anterior $\Leftrightarrow \exists \lambda^* \in \mathbb{R}^m$ tal que:

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0$$

$$c_i(x^*) \geq 0 \quad \forall i=1 \dots m$$

$$\lambda_i^* \geq 0 \quad \forall i=1 \dots m$$

$$\lambda_i^* c_i(x^*) = 0 \quad \forall i=1 \dots m \quad (\text{condición de complementariedad})$$

donde $\mathcal{L}(x, \lambda) = f(x) - \lambda^T c(x)$ se llama función lagrangiana asociada al problema \textcircled{A}

La última condición implica que $\lambda_i^* = 0$ o $c_i(x^*) = 0$. Por lo tanto:

$$0 = \nabla_x \mathcal{L}(x^*, \lambda^*) = \nabla f(x^*) - \sum_{i \in A(x^*)} \lambda_i^* \nabla c_i(x^*)$$

donde $A(x^*)$ es el conjunto de restricciones activas en x^*

Dualidad

Para el problema anterior se define la función dual $g: \mathbb{R}^m \rightarrow \mathbb{R}$ como:

$$g(\lambda) = \inf_x \mathcal{L}(x, \lambda)$$

$$D = \{ \lambda \in \mathbb{R}^m \mid g(\lambda) > -\infty \}, \text{ es decir, el conjunto de valores } \lambda \in \mathbb{R}^m \text{ para los que } g(\lambda) \text{ es finita}$$

Obs. \Rightarrow El cálculo del "íntimo "inf" requiere encontrar el mínimo global (recordar que tenemos un problema convexo) de \mathcal{L} para una λ dada

El problema dual de \textcircled{A} se define como:

$$\max_{\lambda \in \mathbb{R}^m} g(\lambda) \quad \text{s.a. } \lambda \geq 0$$

$$\text{Ej. } \min_{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}} \frac{1}{2}(x_1^2 + x_2^2) \text{ s.a. } x_1 - 1 \geq 0 \text{ entonces}$$

$\mathcal{L}(x_1, x_2, \lambda) = \frac{1}{2}(x_1^2 + x_2^2) - \lambda_1(x_1 - 1)$ y para x_1 fija \mathcal{L} es convexa como función de $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$. El mínimo respecto a $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ se obtiene con las paradas respectivas a x_1, x_2 e igualando a cero:

$$0 = \nabla_{\mathcal{L}_X}(x_1, x_2, \lambda) = \begin{pmatrix} x_1 - \lambda_1 \\ x_2 \end{pmatrix} \text{ con } x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$\therefore q(\lambda) = \frac{1}{2}(\lambda_1^2 + 0) - \lambda_1(\lambda_1 - 1) = -\frac{1}{2}\lambda_1^2 + \lambda_1 \quad (\text{sustituyendo valores de } x_1, x_2)$$

$$\text{y el problema dual es: } \max_{\lambda_1 \geq 0} -\frac{1}{2}\lambda_1^2 + \lambda_1 \text{ con solución } \lambda_1 = 1$$

Obs.) Una forma alternativa del problema dual es:

$$\max_{x, \lambda} \mathcal{L}(x, \lambda) \text{ s.a. } \nabla_x \mathcal{L}(x, \lambda) = 0, \lambda \geq 0$$

•) Ver capítulo 12 del libro Nocedal & Wright para problemas de optimización que consideran restricciones de igualdad y desigualdad con sus correspondientes condiciones de optimidad

PL

$$\text{En Programación Lineal: } \min_{\substack{x \in \mathbb{R}^n \\ \text{s.a.}}} c^T x \quad \text{con } A \in \mathbb{R}^{m \times n} \text{ de rango completo} \\ Ax = b \\ x \geq 0 \quad m \leq n$$

$$\text{Función Lagrangiana: } \mathcal{L}(x, \lambda, s) = c^T x - \lambda^T (Ax - b) - s^T x$$

(condiciones KKT: $\nabla_{\mathcal{L}_X}(x, \lambda, s) = c - A^T \lambda - s = 0$ donde $\lambda \in \mathbb{R}^m$, $s \in \mathbb{R}^n$

$$\nabla_{\lambda}(x, \lambda, s) = \begin{aligned} Ax - b &= 0 \\ x &\geq 0 \\ s &\geq 0 \\ x^T s &= 0 \quad (\text{complementariedad}) \end{aligned}$$

Problema dual: $\max_{\lambda} b^T \lambda$ s. a. $C - A^T \lambda - s = 0$
 $\lambda \in \mathbb{R}^m, s \in \mathbb{R}^n$ $s \geq 0$

PC
 En programación cuadrática: $\min_{x \in \mathbb{R}^n} \frac{1}{2} x^T H x + c^T x$ con $H \in \mathbb{R}^{n \times n}$ simétrica
 s.a. $Ax - b = 0$ positiva definida (s.p.d)
 $x \geq 0$ $A \in \mathbb{R}^{m \times n}$, rango completo

Función Lagrangiana: $\mathcal{L}(x, \lambda, s) = \frac{1}{2} x^T H x + c^T x - \lambda^T (Ax - b) - s^T x$

Condiciones KKT: $\nabla_x \mathcal{L}(x, \lambda, s) = c + Hx - A^T \lambda - s = 0$
 $\nabla_\lambda \mathcal{L}(x, \lambda, s) = Ax - b = 0$ con $\lambda \in \mathbb{R}^m, s \in \mathbb{R}^n$
 $x \geq 0$
 $s \geq 0$
 $x + s = 0$ (complementariedad)

- PC
- ↳ optimizar una función objetivo cuadrática sujeta a restricciones lineales
 - ↳ Si en un programa cuadrático la Hessiana de la función objetivo es positivo definido o semidefinida el programa cuadrático se llama convexo
 - ↳ Si el programa cuadrático es convexo entonces
 - El mínimo local es un mínimo global
 - Las condiciones de optimidad (KKT) son necesarias y suficientes

Método de Newton

Problema: Dada $f: \mathbb{R} \rightarrow \mathbb{R}$ no lineal se busca $x^* \in \mathbb{R}$ s.t. $f(x^*) = 0$ A

- El método de Newton es un método iterativo que resuelve el problema anterior para funciones que cumplen con condiciones particulares

• El costo computacional de los métodos iterativos se mide por el # de iteraciones y por el # de operaciones en /iteración.

- Se define la tasa a la que convergen los métodos para realizar comparaciones entre ellos:

Un método se dice q' converge con tasa r si:

$$\lim_{k \rightarrow \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^r} = C \quad \text{para alguna } C \neq 0$$

Con $\{x_k\}_{k=0}^n$ sucesión convergente a x^* tal qe $x_k \neq x^* \forall k$

Entonces r sea más alta el método converge más rápido (más iteraciones)

Algunas tasas de convergencia importantes:

- $r = 1, C < 1$ lineal: # de dígitos de precisión por cada iteración
- $r = 2$ cuadrática duplica el # de dígitos de precisión por iteración
- El método de Newton tiene convergencia cuadrática
- Algunos criterios de paro utilizados en los métodos iterativos para resolver el problema X:

Dado un umbral de tolerancia $Tol > 0$, $\{x_k\}_{k=0}^n$ sucesión

1) $|f(x_k)| > Tol$ 2) $|x_{k+1} - x_k| > Tol$ 3) $\frac{|x_{k+1} - x_k|}{|x_k|} > Tol$

4) $|x - x_k| > Tol$ 5) $\frac{|x - x_k|}{|x_k|} > Tol, x \neq 0$ $|x_{k+1}|$
or $x_{k+1} \neq 0$

Método de Newton

Sea $f \in C^2[a, b]$, $x_0, x^* \in [a, b]$. Entonces por Teorema de Taylor:

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + f''(\zeta_x) \frac{(x - x_0)^2}{2} \text{ con } \zeta_x \text{ entre } x, x_0$$

Si x^* es tal que $f(x^*) = 0$ entonces:

$$0 = f(x_0) + f'(x_0)(x^* - x_0) + f''(\zeta_x) \frac{(x^* - x_0)^2}{2}$$

• Suponemos $|x^* - x_0| < \varepsilon$ con $0 < \varepsilon < 1$ entonces

$$(x^* - x_0)^2 < |x^* - x_0|$$

$$\therefore 0 \approx f(x_0) + f'(x_0)(x^* - x_0)$$

• Suponemos $f'(x_0) \neq 0$ entonces: $x^* \approx x_0 - \frac{f(x_0)}{f'(x_0)}$ DIF

Método:

Sea x_0 una aprox. inicial y x^* "suficientemente cercano"

Sea sucesión $\{x_k\}_{k=0}^{\infty}$ dada por la siguiente relación de recurrencia:

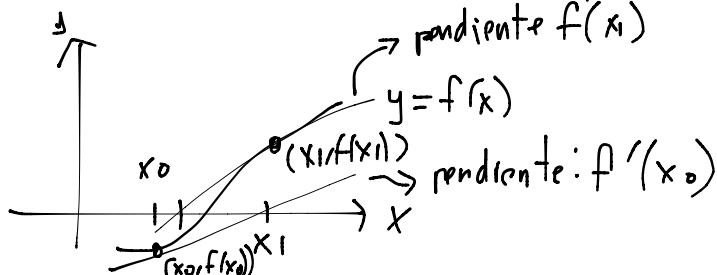
$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} \quad k=0, 1, 2, \dots$$

Obs

• El método se detiene si $f'(x_k) = 0$ para alguna k (monotono de $|f'(x_k)|$)

• Si $f'(x_n)$ es cercano a cero $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$ puede estar lejos de x^*

Gráficamente:



- Tiene convergencia cuadrática si $f'(x^*) \neq 0$ y $|x_0 - x^*|$ pequeño aunque hay ejemplos en los que no es necesario esto último
- El método se extiende a un sistema de ecuaciones no lineales:

$f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ se busca $x^* \in \mathbb{R}^n$ s.t. $f(x^*) = 0$, $f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_n(x) \end{pmatrix}$

La actualización:

$$x_{k+1} = x_k - J(x_k)^{-1} f(x_k)$$

donde $(J(x))_{ij} = \frac{\partial f_i(x)}{\partial x_j}$ $J(x)$ es la jacobiana: $\begin{pmatrix} \nabla f_1(x)^\top \\ \vdots \\ \nabla f_n(x)^\top \end{pmatrix}$

en la práctica no se calcula explícitamente $J(x_k)^{-1}$ por problemas de redondeo, por ello, se resuelve el siguiente sistema de ecuaciones lineales:

para el vector s_k : $J(x_k) s_k = -f(x_k)$

y se actualiza: $x_{k+1} = x_k + s_k$

Se monitorea el número de condición de $J(x_k)$

La convergencia es cuadrática si x_0 está cerca de la solución x^* y $J(x^*)$ es no singular

Apliaciones

- Resolver problemas de optimización: Si $f: \mathbb{R}^n \rightarrow \mathbb{R}$ entonces $\min_x f(x)$ implica buscar puntos estacionarios de f , es decir: buscar $x \in \mathbb{R}^n$ tal que $g(x) = \nabla f(x) = 0$ con $g: \mathbb{R}^n \rightarrow \mathbb{R}^n$ entonces el método se aplica como sigue:

$$x_{k+1} = x_k - J(x_k)^{-1} g(x_k) = x_k - \nabla^2 f(x_k)^{-1} \nabla f(x_k) \text{ donde}$$

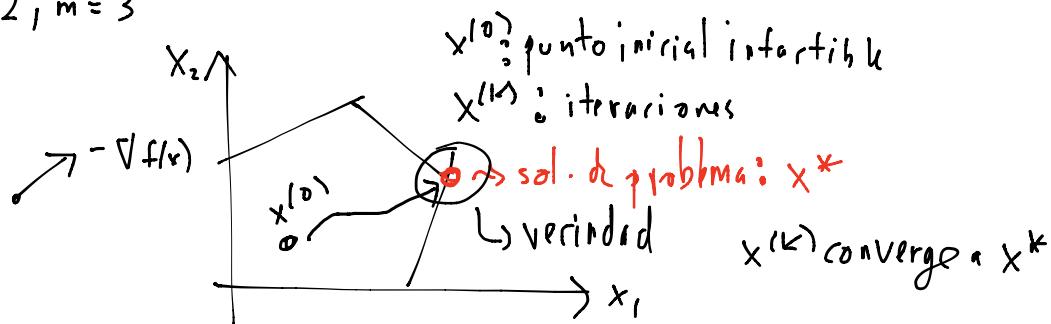
$\nabla^2 f(x_k)$ es la Hessiana de f

Puntos interiores (PI):

- >) Desarrollados en los 80's para resolver problemas lineales tienen la característica de mantener para todas las iteraciones las restricciones de igualdad de manera estricta
-) Entre los métodos por PI los llamados primales-dobles son los que han recibido mayor atención por su eficiencia en tiempo de ejecución y otras características

Ej. PL $\min f(x)$ con $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $c: \mathbb{R}^n \rightarrow \mathbb{R}^m$ lineales
 s.a $c(x) = 0$

Si $n = 2, m = 3$



Ventajas de PI:

- v) Complejidad polinomial en problemas lineales:
 → es un método que rompe con simplex (complejidad exponencial)
- v) Convergencia a partir de puntos iniciales infeasibles dentro/fuera de la región
- v) A medida que el tamaño del problema crece (en número de variables) el número de iteraciones crece lentamente

Desventajas de PI

- d) Cálculo de la siguiente iteración es costoso computacionalmente hablando (resolver un sistema de ecuaciones lineales)

PI en PL

Consideramos PL: $\min_{x \in \mathbb{R}^n} c^T x$ con $A \in \mathbb{R}^{m \times n}$ rango completo, $m \leq n$
 $\text{s.a. } Ax = b$
 $x \geq 0$

Condiciones de KKT: $\nabla_x L(x, \lambda, s) = c - A^T \lambda - s = 0$ donde $\lambda \in \mathbb{R}^m$, $s \in \mathbb{R}^n$
 $\nabla_{\lambda} L(x, \lambda, s) = Ax - b = 0$
 $x^T s = 0$ (complementariedad)
 $x \geq 0$
 $s \geq 0$

Definimos $F(w) = \begin{bmatrix} c - A^T \lambda - s \\ Ax - b \\ x^T s \end{bmatrix}$ con $w = \begin{pmatrix} x \\ \lambda \\ s \end{pmatrix}$, $X = \text{diag}(x_i)_{i=1}^n$

y buscamos resolver con el método de Newton: $F(w) = 0$

Para el método de Newton:

→ Sea w_0 un punto inicial

→ $w_{k+1} = w_k + \Delta_k$ donde Δ_k se le conoce como dirección de desplazamiento
afín y es solución de $J(w)\Delta_k = -F(w)$

con $J(w) = \begin{bmatrix} 0 & -A^T & -I \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix}$, $\Delta_k = \begin{pmatrix} \Delta x_k \\ \Delta \lambda_k \\ \Delta s_k \end{pmatrix}$, $S = \text{diag}(s_i)_{i=1}^n$

Para la implementación de PI:

+) No tal que $x_0, s_0 \geq 0$, definir $M_0 = \frac{x_0^T s_0}{n}$ (medida de doblidad)

+) Elegir $\alpha_k \in [0, 1]$ y resolver para $k = 0, 1, \dots$, hasta convergencia.

$$J(w_k)\Delta_k = -F(w_k) = - \begin{bmatrix} F_1(w_k) \\ F_2(w_k) \\ X_S - \alpha_k M_0 e \end{bmatrix} \quad \|F(w_k)\| \leq T_0$$

$$\text{con } e = (1)_{i=1}^n$$

σ_K se conoce como parámetro de centrado que junto con M_K controlan Δ_K hacia una solución que reduzca en cada iteración la medida de dualidad y puede elegirse de manera estática ($\sigma_K=0$) o dinámica (cambio en cada iteración)

- +) $W_{K+1} = W_K + \alpha_K \Delta_K$ donde $\alpha_K \in [0, 1]$ es un valor que se reduce como longitud de paso y corta Δ_K garantizando $x_{K+1}, s_K > 0$

Para lo anterior:

Si $x_K^{(i)}, \Delta x_K^{(i)}$ denota la i-ésima componente de $x_K, \Delta x_K$ respectivamente y $\Delta x_K^{(i)} > 0$ tal que: $x_{K+1}^{(i)} = x_K^{(i)} + \Delta x_K^{(i)} = 0$ (paso máximo a la frontera) entonces:

1º caso: si $\Delta x_K^{(i)} > 0$ se tiene $x_{K+1}^{(i)} > 0$ pues $x_K^{(i)} > 0$

2º caso: Si $\Delta x_K^{(i)} < 0$ se tiene $x_{K+1}^{(i)} > 0$ o $x_{K+1}^{(i)} < 0$

$$\therefore \text{calculamos } \alpha_x^m = \min_{i=1, \dots, n} \left\{ -\frac{x_K^{(i)}}{\Delta x_K^{(i)}} : \Delta x_K^{(i)} < 0 \right\}$$

↑ problema

El análisis anterior aplica para s_K y tenemos

$$\alpha_K = \underbrace{0,995 \min(\alpha_x^m, \alpha_s^m)}$$

para que las componentes de x_K y s_K no sean iguales a cero

+) $M_{K+1} = \frac{x_{K+1}^T s_{K+1}}{n}$

(Ob.) si σ_K se elige dinámicamente, entonces es pregunta si $\alpha_K \Delta_K$ reduce M_K de manera significativa

*) Ver capítulo 13 de Nocedal & Wright para corrección de Mehrotra (elección de σ_K de forma dinámica)

*) En cada iteración se tiene $x_K = x_K(M_K), s_K = s_K(M_K)$

- r) Al sistema $\mathcal{J}(w)\Delta = -F(w)$ se le conoce como sistema KKT
 o) A medida que W_K se approxima a la solución, cond($\mathcal{J}(w)$) crece

P.I en PC

Consideramos: $\min_{x \in \mathbb{R}^n} \frac{1}{2} x^T H x + c^T x$ con $H \in \mathbb{R}^{n \times n}$ simétrica
 s.a. $\begin{cases} Ax = b \\ x \geq 0 \end{cases}$ positiva definida (s.p.d), $A \in \mathbb{R}^{m \times n}$
 de rango completo

$$\text{Condicion KKT: } \nabla_x \mathcal{L}(x, \lambda, s) = c + Hx - A^T \lambda - s = 0$$

$$\begin{aligned} \nabla_\lambda \mathcal{L}(x, \lambda, s) &= Ax - b = 0 \quad \text{en } \lambda \in \mathbb{R}^m, s \in \mathbb{R}^n \\ &x \geq 0 \\ &s \geq 0 \\ &x + s = 0 \quad (\text{complementariedad}) \end{aligned}$$

Definimos $F(w) = \begin{bmatrix} c + H(x - A^T \lambda - s) \\ Ax - b \\ x + s - M \end{bmatrix}$ con $w = \begin{pmatrix} x \\ \lambda \\ s \end{pmatrix}$, $X = \text{diag}(x_i)_{i=1}^n$

$e = (1)_{i=1}^n$, M medida de dualidad, ϵ parámetro de cierre y buscamos
 resolver con el método de Newton: $F(w) = 0$

\Rightarrow Sea w_0 un punto inicial, con $x_0, s_0 > 0$

$\Rightarrow w_{k+1} = w_k + \alpha_k \Delta_k$ donde Δ_k es solución de $\mathcal{J}(w)\Delta_k = -F(w)$

con $\mathcal{J}(w) = \begin{bmatrix} H & -A^T & -I \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix}$, $\Delta_k = \begin{pmatrix} \Delta x_k \\ \Delta \lambda_k \\ \Delta s_k \end{pmatrix}$, $S = \text{diag}(s_i)_{i=1}^n$

$$\alpha_k \in (0, 1]$$

Puntos interiores con función de barrera logarítmica:

Son una clase de métodos que resuelven problemas lineales y cuadráticos
 perturbando las condiciones optimidad relacionadas con la
 complementariedad, a través de barreras logarítmicas
 y mantienen la división entre restricciones activas/inactivas
 el mayor número de iteraciones posibles hasta convergencia

$$\text{En PL: } \begin{array}{l} \min_{\substack{x \in \mathbb{R}^n \\ s.a. \quad Ax = b \\ x \geq 0}} c^T x \\ \text{con } A \in \mathbb{R}^{m \times n} \text{ de rango completo, } m \leq n \end{array}$$

con PI tenemos que resolver $J(w) \Delta = -F(w)$ y con las consideraciones en la implementación las condiciones de KKT son:

$$\left. \begin{array}{l} \text{condiciones} \\ \text{KKT} \\ \text{perturbadas} \end{array} \right\} \begin{array}{l} c - A^T \lambda - s = 0 \\ Ax - b = 0 \\ Xs = M \epsilon \end{array} \quad \text{con } \lambda \in \mathbb{R}^m, s \in \mathbb{R}^n, s \geq 0$$

Definimos el problema: $\min_{x \in \mathbb{R}^n} c^T x - \frac{1}{N} \sum_{i=1}^n \log(x_i)$ entonces:

$$s.a. \quad Ax = b$$

- 1) Este problema tiene las mismas condiciones KKT perturbadas
- 2) $\sum_{i=1}^n \log(x_i)$ se comporta como función de barrera logarítmica
- 3) M es el parámetro de barrera y en cada iteración decrece
- 4) El problema anterior es convexo, no tiene restricciones de desigualdad
- 5) Aplicamos la implementación de PI a este problema:
 - +) Controlar $x_K, s_K > 0$ para cada K
 - +) Cortar el paso en la longitud de paso: α_K
 - +) Elección estática/dinámica de α_K
 - +) Actualización de M_K

$$\text{En PC} \quad \begin{array}{l} \min_{\substack{x \in \mathbb{R}^n \\ s.a. \quad Ax = b \\ x \geq 0}} \frac{1}{2} x^T H x + c^T x \\ \text{con } H \in \mathbb{R}^{n \times n} \text{ simétrica} \\ \text{positiva definida (s.p.d)}, A \in \mathbb{R}^{m \times n} \\ \text{de rango completo} \end{array}$$

Definimos $\min_{\substack{x \in \mathbb{R}^n \\ s.a. Ax = b}} \frac{1}{2} x^T H x + c^T x - \mu \sum_{i=1}^n \log(x_i)$ el problema de

barrera logarítmica cuyas condiciones KKT perturbadas son:

$$c^T H x - A^T \lambda - s = 0 \quad \text{con } \lambda \in \mathbb{R}^m, s \in \mathbb{R}^n, s \geq 0$$

$$Ax = b$$

$$\sum s = \mu e$$