

Métodos de Punto Interior

Oscar S. Dalmau Cedeño
dalmau@cimat.mx

4 de octubre de 2018

- ① Problema de Programación Lineal
Simplex: Algoritmo de dos Fases
Punto Interior: Punto Inicial
- ② Mehrotra Predictor-Corrector
- ③ Optimización del Algoritmo

Forma estándar

Forma estándar de un problema de programación lineal (PL).

$$\min c^T x, \quad \text{s.a.} : \quad Ax = b, \quad x \succeq 0. \quad (1)$$

Simplex: Algoritmo de dos Fases

- Supongamos el problema de PL

$$\min \quad c^T x, \quad (2)$$

$$s.a : \quad Ax = b, \quad (3)$$

$$x \succeq 0. \quad (4)$$

donde $b \succeq 0$.

Simplex: Punto básico factible

- Consideremos el *problema artificial* con $b \succeq 0$

$$\min_{x,y} \quad 1^T y, \quad (5)$$

$$s.a : \quad [A, I][x; y] = b, \quad (6)$$

$$x, y \succeq 0. \quad (7)$$

- Al vector y se le conoce como *vector de variables artificiales* y a $1^T y$ *función objetivo artificial*
- Se puede notar que el problema anterior tiene una solución básica factible trivial, ie, $[x; y] = [0; b]$
- Si la solución del problema artificial es $y^* = 0$ entonces el óptimo correspondiente $[x^*; y^*]$ es un punto básico factible del problema artificial y como $y^* = 0$ entonces x^* es un punto básico factible en el problema original.

Simplex: Algoritmo de dos Fases

- Es decir, usando variables artificiales podemos resolver el problema de programación lineal, aplicando el *método simplex en dos fases*.
- **Fase I:** Se introducen las variables artificiales, la función objetivo artificial y se encuentra una solución básica factible (si hay solución)
- **Fase II:** Se usa la solución básica factible de la Fase I para inicializar el algoritmo simplex para resolver el problema original.

big-M

- Se añade una variable artificial, y se obtiene el nuevo problema

$$\min \quad c^T x + M u, \quad (8)$$

$$s.a : \quad A x + (b - A x^0) u = b, \quad (9)$$

$$x, u \succeq 0. \quad (10)$$

con $M = \infty$ y seleccionando $x^0 \succ 0$, por ejemplo $x^0 = e$, y $u \in \mathbb{R}$.

- Se puede ver que $[x; u] = [x^0; 1] \in \mathbb{R}^{n+1}$ es un punto interior factible. Si M es suficientemente grande, el problema original puede ser resuelto a través del problema anterior.

Metodo de Dos Fases

El problema original se resuelve en dos fases.

- Fase I: Se resuelve un problema con el objetivo de encontrar un punto interior factible del problema original
- Fase II: Se resuelve el problema original

Metodo de Dos Fases

- Fase I: Sea x^0 un vector positivo. Consideremos el siguiente problema

$$\min_{x,u} \quad u, \quad (11)$$

$$s.a : \quad Ax + (b - Ax^0)u = b, \quad (12)$$

$$x, u \succeq 0. \quad (13)$$

con $u \in \mathbb{R}$. Es obvio que $[x; u] = [x^0; 1] \in \mathbb{R}^{n+1}$ es un punto interior factible.

- Si $u^* > 0$ el problema original es no factible. Si $u^* = 0$ entonces $x^* \succeq 0$ es factible en el problema original, aunque no necesariamente estrictamente factible.
- Fase II: Si hay solución x^* se puede usar como punto inicial del problema original

Mehrotra Predictor-Corrector: Introducción

- Los algoritmos prácticos de punto interior tratan de seguir la idea de cumplir con las restricciones de positividad de los algoritmos estudiados
- En cada iteración se calcula un paso tipo Newton que considera un componente de centrado
- Sin embargo, la mayoría de las implementaciones usan puntos iniciales e iteraciones no factibles
- Varios aspectos teóricos se ignoran y se incorporan mejoras con el objetivo de tener desempeños prácticos

Mehrotra Predictor-Corrector: Introducción

En resumen, los MPI siguen los siguientes pasos

- Se calcula el paso $(\Delta x, \Delta \lambda, \Delta s)$ usando Newton
- Se selecciona el tamaño de paso α de modo que la nueva iteración $(x_n, \lambda_n, s_n) = (x, \lambda, s) + \alpha(\Delta x, \Delta \lambda, \Delta s)$ cumpla $(x_n, s_n) > 0$.
- El punto inicial (x_0, λ_0, s_0) puede ser no factible y se selecciona satisfaciendo la condición $(x_0, s_0) \succ 0$

Mehrotra Predictor-Corrector: Introducción

Las dos características que diferencian este algoritmo del Algoritmo *Primal-Dual* son:

- Adiciona un paso corrector
- Selección adaptativa del parámetro de centrado σ .

Paso Predictor

Para calcular la dirección $(\Delta x, \Delta \lambda, \Delta s)$, primero se calcula el paso predictor $(\Delta x^{aff}, \Delta \lambda^{aff}, \Delta s^{aff})$ considerando $\sigma = 0$. Para lo cual se resuelve el sistema

$$\begin{bmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x^{aff} \\ \Delta \lambda^{aff} \\ \Delta s^{aff} \end{bmatrix} = \begin{bmatrix} -r_c \\ -r_b \\ -XSe \end{bmatrix} \quad (14)$$

donde $r_c = A^T \lambda + s - c$ y $r_b = Ax - b$.

Paso Predictor

Luego se determina a calidad de dicha dirección, a través de α_{aff}^{pri} y α_{aff}^{dua}

$$\alpha_{aff}^{pri} = \min \left(1, \min_{i: \Delta x_i^{aff} < 0} -\frac{x_i}{\Delta x_i^{aff}} \right) \quad (15)$$

$$\alpha_{aff}^{dua} = \min \left(1, \min_{i: \Delta s_i^{aff} < 0} -\frac{s_i}{\Delta s_i^{aff}} \right) \quad (16)$$

Es decir, se calculan los maximos tamaños de pasos permitidos a lo largo de la dirección affine scaling.

Paso Predictor

Luego se define μ_{aff} , que es el paso que lleva a la frontera

$$\mu_{aff} = (x + \alpha_{aff}^{pri} \Delta x^{aff})^T (s + \alpha_{aff}^{dua} \Delta s^{aff}) / n \quad (17)$$

$$\sigma = \left(\frac{\mu_{aff}}{\mu} \right)^3 \quad (18)$$

Paso Corrector

Para el paso corrector se resuelve el sistema de ecuaciones.

$$\begin{bmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \lambda \\ \Delta s \end{bmatrix} = \begin{bmatrix} -r_c \\ -r_b \\ -XSe - \Delta X^{aff} \Delta S^{aff} e + \sigma \mu e \end{bmatrix}$$

Paso Corrector: Comentario I

Si se da el paso completo, entonces la condición de complementareidad sería

$$(x_i + d_{x_i}^{aff})(s_i + d_{s_i}^{aff}) = \tau$$

entonces

$$s_i d_{x_i}^{aff} + x_i d_{s_i}^{aff} = \tau - x_i s_i - d_{x_i}^{aff} d_{s_i}^{aff}$$

para $i = 1, 2, \dots, n$.

Paso Corrector: Comentario II

Luego

$$\begin{aligned} s_1 d_{x_1}^{aff} + x_1 d_{s_1}^{aff} &= \tau - x_1 s_1 - d_{x_1}^{aff} d_{s_1}^{aff} \\ s_2 d_{x_2}^{aff} + x_2 d_{s_2}^{aff} &= \tau - x_2 s_2 - d_{x_2}^{aff} d_{s_2}^{aff} \\ &\dots \quad \dots \quad \dots \\ s_n d_{x_n}^{aff} + x_n d_{s_n}^{aff} &= \tau - x_n s_n - d_{x_n}^{aff} d_{s_n}^{aff} \end{aligned}$$

que podemos reescribir como sigue

$$S d_x^{aff} + X d_s^{aff} = \tau e - X S e - \Delta X^{aff} \Delta S^{aff} e$$

haciendo $\tau = \sigma \mu$ obtenemos la corrección, (lado derecho)

$$\begin{bmatrix} S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \lambda \\ \Delta s \end{bmatrix} = \begin{bmatrix} -X S e + \underbrace{\sigma \mu e}_{\text{centrado}} - \underbrace{\Delta X^{aff} \Delta S^{aff} e}_{\text{corrección}} \end{bmatrix}$$

Paso Corrector: Comentario III

Otra forma de motivar la corrección

$$\begin{bmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x^{aff} \\ \Delta \lambda^{aff} \\ \Delta s^{aff} \end{bmatrix} = \begin{bmatrix} -r_c \\ -r_b \\ -XSe \end{bmatrix}$$

Luego

$$S\Delta x^{aff} + X\Delta s^{aff} = -XSe$$

es decir, ..

$$\begin{aligned} s_i d_{x_i}^{aff} + x_i d_{s_i}^{aff} &= -x_i s_i \\ x_i s_i + s_i d_{x_i}^{aff} + x_i d_{s_i}^{aff} &= 0 \end{aligned}$$

Paso Corrector: Comentario IV

Si se da el paso completo, entonces la condición de complementareidad sería

$$\begin{aligned} (x_i + d_{x_i}^{aff})(s_i + d_{s_i}^{aff}) &= \overbrace{x_i s_i + s_i d_{x_i}^{aff} + x_i d_{s_i}^{aff}}^0 + d_{x_i}^{aff} d_{s_i}^{aff} \\ (x_i + d_{x_i}^{aff})(s_i + d_{s_i}^{aff}) &= d_{x_i}^{aff} d_{s_i}^{aff} \end{aligned}$$

para $i = 1, 2, \dots, n$.

Luego, la condición de complementareidad en la nueva iteración es $d_{x_i}^{aff} d_{s_i}^{aff}$ en lugar de ser igual a 0.

Paso Corrector: Comentario V

Una idea es corregir esta desviación, introduciendo un paso corrector,

$$S\Delta x^{corr} + X\Delta x^{corr} = -\Delta X^{aff} \Delta S^{aff} e$$

ie, se resuelve

$$\begin{bmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x^{corr} \\ \Delta \lambda^{corr} \\ \Delta s^{corr} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -\Delta X^{aff} \Delta S^{aff} e \end{bmatrix}$$

Finalmente se combinan ambos pasos,...

$$[\Delta x, \Delta \lambda, \Delta s] = [\Delta x^{aff}, \Delta \lambda^{aff}, \Delta s^{aff}] + [\Delta x^{corr}, \Delta \lambda^{corr}, \Delta s^{corr}]$$

que en la práctica, funciona (casi siempre) mejor que el paso afin. Además, se incluye la estrategia de centrado (ie, se incluye el término del camino central $+\sigma\mu e$).

Paso Corrector

Luego se calcula el máximo paso a lo largo de la dirección calculada, antes de que se incumplan las restricciones.

$$\alpha_{max}^{pri} = \min_{i: \Delta x_i < 0} -\frac{x_i}{\Delta x_i} \quad (19)$$

$$\alpha_{max}^{dua} = \min_{i: \Delta s_i < 0} -\frac{s_i}{\Delta s_i} \quad (20)$$

Paso Corrector

Luego se determinan las longitudes de los pasos primal y dual:

$$\alpha_k^{pri} = \min(1; \eta \alpha_{max}^{pri}) \quad (21)$$

$$\alpha_k^{dua} = \min(1; \eta \alpha_{max}^{dua}) \quad (22)$$

Donde $\eta \in [0, 1)$ Si $\eta \rightarrow 1$ cerca de la solución se acelera la convergencia.

Algoritmo

Dado (x^0, λ^0, s^0) con $(x^0, s^0) > 0$

Para $k = 1, 2, \dots$

Poner $(x, \lambda, s) = (x^k, \lambda^k, s^k)$

Calcular $(\Delta x^{aff}, \Delta \lambda^{aff}, \Delta s^{aff})$

Calcular α_{aff}^{pri} , α_{aff}^{dua} y μ_{aff}

Calcular el parámetro de centrado σ

Resolver el sistema del paso corrector para $(\Delta x, \Delta \lambda, \Delta s)$

Calcular α_k^{pri} y α_k^{dua}

Poner

$$x^{k+1} = x^k + \alpha_k^{pri} \Delta x$$

$$\lambda^{k+1} = \lambda^k + \alpha_k^{dua} \Delta \lambda$$

$$s^{k+1} = s^k + \alpha_k^{dua} \Delta s$$

Fin del ciclo

Punto inicial

- La selección de un punto inicial (x^0, λ^0, s^0) tal que $(x^0, s^0) \succ 0$ es muy importante para la convergencia del algoritmo
- Se vera una posible heurística para obtener un punto inicial. Para ello se resolverían problemas de norma minima
- Uno para la variable x que garantice la igualdad $Ax = b$ y el otro, para las variables duales (λ, s) que garantice la igualdad $A^T \lambda + s = c$.
- Luego, las soluciones anteriores son modificadas para garantizar que $x, s \succeq 0$

Punto inicial: Problemas de mínima norma

$$\begin{aligned} \min_x \quad & \frac{1}{2} \|x\|^2; \text{ s.a: } Ax = b \\ \min_{\lambda, s} \quad & \frac{1}{2} \|s\|^2; \text{ s.a: } A^T \lambda + s = c \end{aligned}$$

La solución de los problemas anteriores es

$$\begin{aligned} \tilde{x} &= A^T (AA^T)^{-1} b \\ \tilde{\lambda} &= (AA^T)^{-1} A c \\ \tilde{s} &= c - A^T \tilde{\lambda} \end{aligned}$$

Punto inicial

Algunas de las componentes de \tilde{x} \tilde{s} puede ser negativa, para garantizar que sean no negativas, se definen

$$\delta_x = \max(0, -(3/2) \min_i \tilde{x}_i)$$

$$\delta_s = \max(0, -(3/2) \min_i \tilde{s}_i)$$

Luego, los nuevos vectores

$$\hat{x} = \tilde{x} + \delta_x e$$

$$\hat{s} = \tilde{s} + \delta_s e$$

cumplen $\hat{x}, \hat{s} \geq 0$.

Punto inicial

Finalmente el punto inicial se calcula como sigue

$$\begin{aligned}x^0 &= \hat{x} + \hat{\delta}_x e \\s^0 &= \hat{s} + \hat{\delta}_s e \\\lambda^0 &= \tilde{\lambda}\end{aligned}$$

donde $\hat{\delta}_x, \hat{\delta}_s$ son promedios pesados

$$\hat{\delta}_x = \frac{1}{2} \frac{\hat{x}^T \hat{s}}{1^T \hat{s}}; \quad \hat{\delta}_s = \frac{1}{2} \frac{\hat{x}^T \hat{s}}{1^T \hat{x}}$$

para asegurar que no x^0, s^0 no sean muy diferentes y no cercanas a cero.

- En el algoritmo Mehrotra hay que resolver dos sistemas de ecuaciones en cada iteración, uno en el *paso predictor* y el otro en el *paso corrector* lo que hace más costosa cada iteración.
- En el caso del sistema a resolver en el paso predictor es el mismo que se explicó en la tarea pasada, pero en este caso para $\sigma = 0$, de modo que nos sirve lo ya explicado en dicha tarea.

Vamos a simplificar el sistema a resolver en el *paso corrector*.

$$\begin{bmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \lambda \\ \Delta s \end{bmatrix} = \begin{bmatrix} -r_c \\ -r_b \\ -r_{xs} \end{bmatrix}$$

con $r_{xs} = XSe + \Delta X^{aff} \Delta S^{aff} e - \sigma \mu e$

Efectuado el producto e igualando:

$$A^T \Delta \lambda + \Delta s = -r_c \quad (23)$$

$$A \Delta x = -r_b \quad (24)$$

$$S \Delta x + X \Delta s = -r_{xs} \quad (25)$$

Despejando Δs en (25), tenemos:

$$\Delta s = -X^{-1}r_{xs} - X^{-1}S\Delta x \quad (26)$$

definifiendo $D = S^{-1/2}X^{1/2}$, entonces

$$\Delta s = -X^{-1}r_{xs} - D^{-2}\Delta x \quad (27)$$

De (23) y (25) se tiene que:

$$\begin{aligned}A^T \Delta\lambda + \Delta s &= -r_c \\ -D^{-2} \Delta x + A^T \Delta\lambda &= -r_c + X^{-1} r_{xs}\end{aligned}$$

con $D = S^{-1/2} X^{1/2}$.

El resultado anterior y la ecuación (24) conducen al sistema

$$\begin{bmatrix} -D^{-2} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \lambda \end{bmatrix} = \begin{bmatrix} -r_c + X^{-1}r_{xs} \\ -r_b \end{bmatrix} \quad (28)$$

$$\Delta s = -X^{-1}r_{xs} - D^{-2}\Delta x \quad (29)$$

con $D = S^{-1/2}X^{1/2}$.

Ecuaciones normales

Finalmente

$$AD^2A^T\Delta\lambda = -r_b - AXS^{-1}r_c + AS^{-1}r_{xs} \quad (30)$$

$$\Delta s = -r_c - A^T\Delta\lambda \quad (31)$$

$$\Delta x = -S^{-1}r_{xs} - D^2\Delta s \quad (32)$$

que son las Ecuaciones normales

- Como se observa es más simple y mucho más eficiente que resolver el sistema de ecuaciones original.
- Para resolver las ecuaciones normales, la idea sería:
 - 1 Resolver el sistema (30) para $\Delta\lambda$,
 - 2 Luego, se sustituye $\Delta\lambda$ en (31) para obtener Δs
 - 3 Este último resultado, Δs , se sustituye en (32) y calculamos Δx .

- Se puede usar sparse Cholesky para descomponer la matriz AD^2A^T .
- Se debe usar un sparse Cholesky de propósito general, pues la matriz puede estar mal condicionada o puede ser singular.
- El algoritmo puede encontrar valores cercanos a cero, cero o incluso negativos debido al redondeo.