

IFT 2505

Méthodes de points intérieurs

Fabian Bastin
DIRO
Université de Montréal

Approche primale

Considérons le programme

$$\begin{aligned} \min_x \quad & c^T x \\ \text{s.à.} \quad & Ax = b \\ & x \geq 0. \end{aligned}$$

Définissons les ensembles

$$\mathcal{F}_P \stackrel{\text{def}}{=} \{x \mid Ax = b, x \geq 0\}$$

$$\mathring{\mathcal{F}}_P \stackrel{\text{def}}{=} \{x \mid Ax = b, x > 0\}$$

Problème barrière

On suppose que $\mathring{\mathcal{F}}_P$ est non vide et que l'ensemble de solutions optimales pour ce problème est borné.

Soit $\mu \geq 0$. Problème barrière (PB) :

$$\begin{aligned} \min_x \quad & c^T x - \mu \sum_{j=1}^n \log x_j \\ \text{s.à.} \quad & Ax = b \\ & x > 0. \end{aligned}$$

Problème barrière vs Problème linéaire

Si $\mu = 0$, on retrouve le problème original en permettant à x d'avoir des composantes nulles :

$$\begin{aligned} \min_x \quad & c^T x - \mu \sum_{j=1}^n \log x_j \\ \text{s.à.} \quad & Ax = b \\ & x \geq 0. \end{aligned}$$

Chemin central

Soit $x(\mu)$, la solution au PB étant donné μ . En faisant varier μ continûment vers 0, nous obtenons le chemin central primal.

Si $\mu \rightarrow \infty$, la solution s'approche du centre analytique de la région réalisable, lorsque celle-ci est bornée : le terme barrière prédomine alors dans l'objectif.

Comme $\mu \rightarrow 0$, ce chemin converge vers le centre analytique de la face optimale $\{x \mid c^T x = z^*, Ax = b, x \geq 0\}$, où z^* est la valeur optimale du programme linéaire.

L'idée de base est de résoudre une succession de problèmes barrières pour des valeurs décroissantes de μ .

Centre analytique

Considérons un ensemble \mathcal{S} dans un sous-ensemble \mathcal{X} de E^n , défini comme

$$\mathcal{S} = \{\mathbf{x} \in \mathcal{X} : g_j(\mathbf{x}) \geq 0, j = 1, 2, \dots, m\},$$

et supposons que les fonctions g_j sont continues.

Nous supposons aussi que \mathcal{S} a un intérieur non-vide :

$$\mathring{\mathcal{S}} = \{\mathbf{x} \in \mathcal{X} : g_j(\mathbf{x}) > 0, \forall j\} \neq \emptyset.$$

Définissons aussi sur $\mathring{\mathcal{S}}$ la fonction potentiel

$$\psi(\mathbf{x}) = - \sum_{j=1}^m \log g_j(\mathbf{x}).$$

Centre analytique

Le centre analytique de \mathcal{S} est le vecteur (ou l'ensemble de vecteurs) qui minimise le potentiel.

En d'autres termes, c'est le vecteur (ou l'ensemble de vecteurs) qui résout

$$\min_{\mathbf{x}} \psi(\mathbf{x}) = \min_{\mathbf{x}} \left\{ - \sum_{j=1}^m \log g_j(\mathbf{x}) : \mathbf{x} \in \mathcal{X}, g_j(\mathbf{x}) > 0 \ \forall j \right\}.$$

Exemple : un cube

Considérons l'ensemble \mathcal{S} défini par $x_i \geq 0$, $(1 - x_i) \geq 0$, pour $i = 1, 2, \dots, n$:

$$\mathcal{S} = [0, 1]^n,$$

le cube unité dans E^n .

Cherchons à minimiser la fonction potentiel correspondante

$$\psi(\mathbf{x}) = -\sum_{i=1}^n \log x_i - \sum_{i=1}^n \log(1 - x_i).$$

Exemple : un cube

Pour ce faire, annulons le gradient de $\psi(\mathbf{x})$

$$\nabla_{\mathbf{x}}\psi(\mathbf{x}) = \begin{pmatrix} -\frac{1}{x_1} + \frac{1}{1-x_1} \\ -\frac{1}{x_2} + \frac{1}{1-x_2} \\ \vdots \\ -\frac{1}{x_n} + \frac{1}{1-x_n} \end{pmatrix}$$

(Note : $\psi(\mathbf{x})$ est ici une fonction convexe.)

On en tire $x_i = 1/2$ pour tout i . Dès lors, le centre analytique est identique à l'idée usuelle du centre du cube unité.

Toutefois, en général, le centre analytique dépend de la manière dont l'ensemble est défini, en particulier les inégalités utilisées dans la définition.

Exemple : un cube

Par exemple, nous pourrions aussi définir le cube unité avec les inégalités

$$\begin{aligned}x_i &\geq 0, \quad i = 1, \dots, n \\ (1 - x_i)^d &\geq 0, \quad i = 1, \dots, n,\end{aligned}$$

avec $d > 1$, et d impair. Dans ce cas, la solution est

$$x_i = \frac{1}{d+1}, \quad \forall i.$$

Pour un grand d , ce point est proche du coin intérieur, i.e. du point $(0, 0, \dots, 0)$ du cube unité.

De plus, l'ajout d'inégalités additionnelles redondantes peut aussi modifier la position du centre analytique. Répéter par exemple une inégalité donnée changera la position du centre.

Centre analytique

Il y a plusieurs ensembles associés avec des programmes linéaires pour lesquels le centre analytique est d'un intérêt particulier.

Un tel ensemble est la région réalisable elle-même. Un autre est l'ensemble des solutions optimales. Il y a aussi les ensembles associés avec les formulations duale et primale-duale. Tous ceux-ci sont en fait reliés.

Considérons par exemple le centre analytique associé à un **polytope borné** Ω dans \mathcal{R}^m , représenté par n ($> m$) inégalités linéaires :

$$\Omega = \{\mathbf{y} \in \mathcal{R}^m : \mathbf{c}^T - \mathbf{y}^T \mathbf{A} \geq \mathbf{0}\},$$

où $\mathbf{A} \in \mathcal{R}^{m \times n}$ et $\mathbf{c} \in \mathcal{R}^n$ sont données, et A est de rang m .

Centre analytique

Dénotons l'intérieur de Ω par

$$\mathring{\Omega} = \{\mathbf{y} \in \mathcal{R}^m : \mathbf{c}^T - \mathbf{y}^T \mathbf{A} > 0\}.$$

La fonction potentiel pour cet ensemble est

$$\psi_{\Omega}(\mathbf{y}) = - \sum_{j=1}^n \log(c_j - \mathbf{y}^T \mathbf{a}_j) = - \sum_{j=1}^n \log s_j,$$

où $\mathbf{s} = \mathbf{c}^T - \mathbf{y}^T \mathbf{A}$ est un vecteur d'écart.

Dès lors, dans le cas présent, la fonction potentiel est l'opposé de la somme des logarithmes des variables d'écart, ou encore le produit de leur inverse :

$$- \sum_{j=1}^n \log s_j = \log \prod_{j=1}^n \frac{1}{s_j}$$

Centre analytique

Le centre analytique de Ω est le point de Ω qui minimise la fonction potentiel. Ce point, dénoté \mathbf{y}^a est associé à $\mathbf{s}^a = \mathbf{c} - \mathbf{A}\mathbf{y}^a$. La paire $(\mathbf{y}^a, \mathbf{s}^a)$ est définie de manière unique, vu que la fonction potentiel est strictement convexe sur un ensemble convexe borné.

Notons aussi que minimiser la fonction potentiel revient ici à maximiser le produit des variables d'écart.

Comme,

$$\frac{d}{dy_i}(c_j - \mathbf{y}^T \mathbf{a}_j) = -a_{ij},$$

nous avons

$$\frac{d}{dy_i} \psi(\mathbf{y}) = \sum_{j=1}^n \frac{a_{ij}}{c_j - \mathbf{y}^T \mathbf{a}_j}.$$

Centre analytique

Annuler le gradient de $\psi(\mathbf{y})$ par rapport à \mathbf{y} donne

$$\sum_{j=1}^n \frac{a_{ij}}{c_j - \mathbf{y}^T \mathbf{a}_j} = 0, \quad \forall i,$$

ou encore

$$\sum_{j=1}^n \frac{a_{ij}}{s_j} = 0, \quad \forall i,$$

Définissons $x_j = 1/s_j$ pour tout j . Nous pouvons réécrire l'équation comme

$$\sum_{j=1}^n a_{ij} x_j = 0, \quad \forall i,$$

soit

$$\mathbf{A} \mathbf{x} = \mathbf{0}.$$

Centre analytique

Notons la multiplication composante par composante comme suit

$$\mathbf{x} \circ \mathbf{s} = (x_1 s_1, x_2 s_2, \dots, x_n s_n)^T.$$

Le **centre analytique** est alors défini par les conditions

$$\mathbf{x} \circ \mathbf{s} = \mathbf{1}$$

$$\mathbf{A}\mathbf{x} = \mathbf{0}$$

$$\mathbf{A}^T \mathbf{y} + \mathbf{s} = \mathbf{c}.$$

Centre analytique

Le centre analytique peut aussi être défini quand l'intérieur est vide ou que des égalités sont présentes, par exemple avec

$$\Omega = \{\mathbf{y} \in \mathcal{R}^m : \mathbf{c}^T - \mathbf{y}^T \mathbf{A} \geq 0, \mathbf{B}\mathbf{y} = \mathbf{b}\}$$

Dans ce cas, le centre analytique est choisi sur la surface linéaire $\{\mathbf{y} : \mathbf{B}\mathbf{y} = \mathbf{b}\}$ pour maximiser le produit des variables d'écart.

Dans ce contexte, l'intérieur de Ω réfère à l'intérieur de l'orthant positif des variables d'écart :

$$R_+^n := \{\mathbf{s} : \mathbf{s} \geq 0\}.$$

Cette définition d'intérieur dépend seulement de la région des variables d'écart.

Même s'il y a seulement un point dans Ω avec $\mathbf{s} = \mathbf{c} - \mathbf{A}^T \mathbf{y}$ pour un certain \mathbf{y} où $\mathbf{B}\mathbf{y} = \mathbf{b}$ avec $\mathbf{s} > 0$, nous continuerons à dire que $\hat{\Omega}$ est non vide.

Fonction lagrangienne

Réécrivons la contrainte $Ax = b$ sous la forme $Ax - b = 0$, et introduisons un vecteur y , en associant une composante y_i à la i^e contrainte $\sum_{j=1}^n a_{ij}x_j = b_i$.

y_i joue à peu près le même rôle que les variables duales précédentes, mais est appelé à présent multiplicateur de Lagrange.

Lagrangien :

$$L(x) = c^T x - \mu \sum_{j=1}^n \log x_j - y^T (Ax - b).$$

On cherche à minimiser cette fonction en annulant son gradient.

Minimisation du lagrangien

$$\begin{aligned}\nabla_x L(x) &= 0 \\ \Leftrightarrow c_j - \frac{\mu}{x_j} - y^T a_j &= 0, \quad j = 1, \dots, n \\ \Leftrightarrow \mu X^{-1} \mathbf{1} + A^T y &= c,\end{aligned}$$

où $X = \text{diag}(x)$.

En notant $s_j = \mu/x_j$, l'ensemble des conditions d'optimalité (primales, duales, minimisation du lagrangien) s'écrit :

$$\begin{aligned}\mathbf{x} \circ \mathbf{s} &= \mu \mathbf{1} \\ \mathbf{A} \mathbf{x} &= \mathbf{b} \\ \mathbf{A}^T \mathbf{y} + \mathbf{s} &= \mathbf{c}.\end{aligned}$$

Lien avec le dual

y est une solution duale réalisable et $c - A^T y > 0$.

En effet, le dual est

$$\begin{aligned} \max_y \quad & y^T b \\ \text{s.à.} \quad & A^T y \leq c \end{aligned}$$

Comme

$$\begin{aligned} A^T y + s &= c \\ s_j &= \frac{\mu}{x_j} \end{aligned}$$

nous avons

$$A^T y < c.$$

Exemple

Considérons le programme

$$\begin{array}{ll}\max & x_1 \\ \text{s.à.} & 0 \leq x_1 \leq 1 \\ & 0 \leq x_2 \leq 1.\end{array}$$

On réécrit ce système sous forme standard

$$\begin{array}{ll}\min & -x_1 \\ \text{s.à.} & x_1 + x_3 = 1 \\ & x_2 + x_4 = 1 \\ & x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0.\end{array}$$

Conditions d'optimalité

$$x_1 s_1 = \mu$$

$$x_2 s_2 = \mu$$

$$x_3 s_3 = \mu$$

$$x_4 s_4 = \mu$$

$$x_1 + x_3 = 1$$

$$x_2 + x_4 = 1$$

$$y_1 + s_1 = -1$$

$$y_2 + s_2 = 0$$

$$y_1 + s_3 = 0$$

$$y_2 + s_4 = 0$$

Conditions d'optimalité

De

$$y_2 + s_2 = 0$$

$$y_2 + s_4 = 0$$

on a $s_2 = s_4$.

On en déduit aussi

$$x_2 = x_4,$$

et de là

$$x_2 = x_4 = \frac{1}{2}.$$

Conditions d'optimalité

On a aussi

$$\begin{aligned} -1 &= s_1 - s_3 = \frac{\mu}{x_1} - \frac{\mu}{x_3} \\ \Leftrightarrow -1 &= \frac{\mu}{x_1} - \frac{\mu}{1-x_1} \\ \Leftrightarrow -x_1(1-x_1) &= \mu(1-x_1) - \mu x_1 \\ \Leftrightarrow x_1^2 - x_1 &= \mu - 2\mu x_1 \\ \Leftrightarrow x_1^2 - (1-2\mu)x_1 - \mu &= 0. \end{aligned}$$

Le discriminant de cette équation quadratique est

$$\rho = (1-2\mu)^2 + 4\mu = 1 + 4\mu^2,$$

et de là

$$x_1 = \frac{1-2\mu \pm \sqrt{1+4\mu^2}}{2}$$

Chemin central

Pour μ grand, on doit choisir la racine correspondant à '+'.
D'autre part,

$$x \rightarrow \left(1, \frac{1}{2}\right), \text{ comme } \mu \rightarrow 0.$$

On converge vers le centre analytique de la face optimale

$$\{x \mid x_1 = 1, 0 \leq x_2 \leq 1\}$$

plutôt qu'un coin du carré.

Chemin central

De plus,

$$\begin{aligned}\lim_{\mu \rightarrow +\infty} \frac{1 - 2\mu + \sqrt{1 + 4\mu^2}}{2} &= \frac{1}{2} - \frac{1}{2} \lim_{\mu \rightarrow +\infty} \left(2\mu - \sqrt{1 + 4\mu^2} \right) \\ &= \frac{1}{2} - \lim_{\mu \rightarrow +\infty} \left(2\mu - 2\mu \sqrt{\frac{1}{4\mu^2} + 1} \right) \\ &= \frac{1}{2}.\end{aligned}$$

Ainsi $x(\mu) \xrightarrow{\mu \rightarrow +\infty} \left(\frac{1}{2}, \frac{1}{2} \right)$.

Le chemin central est une ligne droite progressant du centre analytique du carré (comme $\mu \rightarrow \infty$) vers le **centre analytique de la face optimale** (comme $\mu \rightarrow 0$).

Chemin dual central

Partons à présent du problème dual

$$\begin{aligned} \max \quad & y^T b \\ \text{s.à.} \quad & y^T A + s^T = c^T \\ & s \geq 0. \end{aligned}$$

Le problème barrière associé est

$$\begin{aligned} \max \quad & y^T b + \mu \sum_{j=1}^n \log s_j \\ \text{s.à.} \quad & y^T A + s^T = c^T \\ & s \geq 0. \end{aligned}$$

Chemin dual central

On suppose que $\mathring{\mathcal{F}}_P = \{(y, s) : y^T A + s^T = c^T, s > 0\}$ est non vide, et l'ensemble des solutions optimales du dual est borné.

On obtient le chemin central dual en faisant tendre μ vers 0.

Lagrangien :

$$L(y, s) = y^T b + \mu \sum_{j=1}^n \log s_j - (y^T A + s^T - c^T)x.$$

Dès lors,

$$\nabla_y L = 0 \Leftrightarrow b_i - a^i x = 0, \forall i,$$

$$\nabla_s L = 0 \Leftrightarrow \frac{\mu}{s_j} - x_j = 0, \forall j.$$

Chemin dual central : conditions d'optimalité

On obtient les conditions d'optimalité

$$\mathbf{x} \circ \mathbf{s} = \mu \mathbf{1}$$

$$\mathbf{Ax} = \mathbf{b}$$

$$\mathbf{A}^T \mathbf{y} + \mathbf{s} = \mathbf{c}.$$

On retrouve les mêmes conditions que pour le chemin central.

Par conséquent, \mathbf{x} est une solution réalisable primale et $\mathbf{x} > 0$.

Considérons l'ensemble de niveau dual

$$\Omega(z) = \{\mathbf{y} \mid \mathbf{c}^T - \mathbf{y}^T \mathbf{A} \geq 0, \mathbf{y}^T \mathbf{b} = z\},$$

avec $z < z^*$, la valeur optimale du dual.

Chemin dual central : conditions d'optimalité

Le centre analytique $(y(z), s(z))$ de Ω coïncide avec le chemin central dual comme z tend vers z^* .

Chemin dual central : exemple

Reprenons le problème

$$\min -x_1$$

$$\text{s.à. } x_1 + x_3 = 1$$

$$x_2 + x_4 = 1$$

$$x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0.$$

Le dual s'écrit

$$\max y_1 + y_2$$

$$\text{s.à. } y_1 \leq -1$$

$$y_2 \leq 0$$

$$y_1 \leq 0$$

$$y_2 \leq 0.$$

Chemin dual central : exemple

On peut réécrire le dual comme

$$\begin{aligned} \max \quad & y_1 + y_2 \\ \text{s.à.} \quad & y_1 + s_1 = -1 \\ & y_2 + s_2 = 0. \end{aligned}$$

Les conditions d'optimalité sont les mêmes que pour le primal, aussi

$$x_2 = x_4 = \frac{1}{2},$$

d'où

$$\begin{aligned} s_2 &= 2\mu, \\ y_2 &= -2\mu. \end{aligned}$$

Chemin dual central : exemple

Nous avons également, en se rappelant de la résolution du primal,

$$\begin{aligned}y_1 &= -1 - s_1 \\&= -1 - \frac{\mu}{x_1(\mu)} \\&= -1 - \frac{2\mu}{1 - 2\mu \pm \sqrt{1 + 4\mu^2}}\end{aligned}$$

Comme $\mu \rightarrow 0$, $y_1 \rightarrow -1$, $y_2 \rightarrow 0$.

Il s'agit de l'unique solution du problème linéaire (les deux contraintes sont actives)

Si $\mu \rightarrow \infty$, y est non borné, car l'ensemble dual réalisable est non borné.

Chemin central primal-dual

Hypothèse : la région réalisable du problème (primal) de programmation linéaire a un intérieur non vide et un ensemble borné de solutions optimales.

Le dual a un intérieur réalisable non vide, en vertu des conditions (d'optimalité) sur le lagrangien.

Le chemin primal-dual est l'ensemble des vecteurs $(x(\mu), y(\mu), s(\mu))$ satisfaisant

$$\mathbf{x} \circ \mathbf{s} = \mu \mathbf{1}$$

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

$$\mathbf{A}^T \mathbf{y} + \mathbf{s} = \mathbf{c}$$

$$\mathbf{x} \geq 0, \mathbf{s} \geq 0$$

$$0 \leq \mu < \infty.$$

Proposition

Si les ensembles réalisables primal et dual ont des intérieurs non vides, alors le chemin central $(x(\mu), y(\mu), s(\mu))$ existe pour tout μ , $0 \leq \mu < \infty$.

De plus, $x(\mu)$ est le chemin central primal, $(y(\mu), s(\mu))$ est le chemin central dual.

$x(\mu)$ et $(y(\mu), s(\mu))$ convergent vers les centres analytiques des faces des solutions optimales primales et duales, respectivement, quand $\mu \rightarrow 0$.

Saut de dualité

Soit $(x(\mu), y(\mu), s(\mu))$ sur le chemin central primal-dual. Nous avons

$$\begin{aligned}c^T x - y^T b &= (A^T y)^T x + s^T x - y^T b \\&= y^T A x + s^T x - y^T b \\&= y^T b + s^T x - y^T b \\&= s^T x \\&= n\mu.\end{aligned}$$

Comme pour la dualité faible, $c^T x \geq y^T b$, et $n\mu$ est appelé le saut de dualité.

Soit $g = c^T x - y^T b$.

Comme $y^T b \leq z^*$, $z^* \geq c^T x - g$, et donc, étant donné (x, y, s) , on peut mesurer la qualité de x comme $c^T x - z^* \leq g$.

Stratégies de solution en points intérieurs

Trois grandes approches, suivant les différences dans les définitions du chemin central :

1. barrière primale, méthode de poursuite de chemin,
2. méthode primale-duale de poursuite de chemin,
3. méthode primale-duale de réduction de potentiel.

		R-P	R-D	Saut nul
Caractéristiques	simplexe primal	X		X
	simplexe dual		X	X
	barrière primale	X		
	poursuite de chemin primale-duale	X	X	
	réduction de potentiel primale-duale	X	X	

R : réalisabilité, P : primal, D : dual

Méthode primale barrière

Nous partons du problème primal barrière

$$\begin{aligned} \min_x \quad & c^T x - \mu \sum_{j=1}^n \log x_j \\ \text{s.à.} \quad & Ax = b \\ & x \geq 0. \end{aligned}$$

Nous voudrions le résoudre pour μ petit.

Par exemple, $\mu = \epsilon/n$ permet d'obtenir un saut de dualité inférieur à ϵ .

Souci : difficile de résoudre pour μ proche de 0.

Méthode primale barrière

Une stratégie générale est de commencer avec μ modérément large (p.e. $\mu = 100$) et de résoudre le problème approximativement.

La solution correspondante est approximativement sur le chemin central primal, mais probablement assez loin du point correspondant à $\mu \rightarrow 0$.

Ce point ne servira que de point de départ pour le problème avec un μ plus petit.

Typiquement, on mettra à jour μ de l'itération k à l'itération $k + 1$ comme

$$\mu_{k+1} = \gamma \mu_k,$$

pour $0 < \gamma < 1$ fixé.

Méthode primale barrière

Si on commence avec une valeur μ_0 , à l'itération k ,

$$\mu_k = \gamma^k \mu_0,$$

Dès, réduire μ_k/μ_0 sous ϵ requiert

$$k = \left\lceil \frac{\log \epsilon}{\log \gamma} \right\rceil.$$

Souvent, une variante de la méthode de Newton est utilisée pour résoudre les sous-problèmes ainsi construits :

$$\mathbf{x} \circ \mathbf{s} = \mu \mathbf{1}$$

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

$$\mathbf{A}^T \mathbf{y} + \mathbf{s} = \mathbf{c}$$

Algorithme général

Etape 1 Choisir μ_0 et un point de départ réalisable (x_0, y_0, s_0) , tel que

$$\begin{aligned}Ax_0 &= b \\ A^T y_0 + s_0 &= c^T\end{aligned}$$

et $s_0 \geq 0$, $x_0 > 0$. Poser $k = 0$, et choisir $\epsilon > 0$.

Etape 2 Projeter x_k sur le chemin central avec la méthode de Newton : calculer le vecteur (d_x, d_y, d_s) et poser

$$x_{k+1} = x_k + d_x$$

$$y_{k+1} = y_k + d_y$$

$$s_{k+1} = s_k + d_s$$

Si $\mu_k < \epsilon$, arrête. Sinon, définir $\mu_{k+1} = \gamma \mu_k$, poser $k := k + 1$, et retourner au début de l'étape 2.

Méthode primale barrière

Nous pouvons en particulier appliquer la méthode de Newton pour trouver un zéro du gradient associé au lagrangien du problème de barrière logarithmique primale.

Rappelons que le lagrangien du problème barrière logarithmique primale est

$$L(x) = c^T x - \mu \sum_{j=1}^n \log x_j - y^T (Ax - b)$$

Il est facile d'exprimer le gradient et la matrice hessienne du lagrangien :

$$\begin{aligned}\nabla L(x) &= c - \mu X^{-1} \mathbf{1} - A^T y \\ H(x) &= \mu X^{-2}\end{aligned}$$

Système d'équation non linéaires : méthode de Newton

Considérons le système d'équations non linéaires

$$F(\mathbf{x}) = 0$$

avec $F : \mathcal{R}^n \rightarrow \mathcal{R}^m$.

En partant d'un point \mathbf{x}_0 , la méthode de Newton construit une suite d'itérés à partir de la récurrence

$$\mathbf{x}_{k+1} = \mathbf{x}_k - J^{-1}(\mathbf{x}_k)F(\mathbf{x}_k)$$

où $J(\mathbf{x}_k)$ est la matrice jacobienne de f :

$$J(\mathbf{x}) = \begin{pmatrix} \nabla_x^T f_1(x) \\ \nabla_x^T f_2(x) \\ \vdots \\ \nabla_x^T f_n(x) \end{pmatrix}$$

La méthode converge si on est suffisamment proche d'un zéro de la fonction.

Méthode de Newton en optimisation

Supposons que $f \in C^2$. On peut appliquer la méthode de Newton à la résolution du système non-linéaire

$$\nabla f(\mathbf{x}) = 0.$$

Dans ce cas, la matrice jacobienne de f n'est rien d'autre que la matrice hessienne :

$$\begin{aligned} J(\mathbf{x}) &= \begin{pmatrix} \nabla_{\mathbf{x}}^T \frac{d}{dx_1} f(\mathbf{x}) \\ \nabla_{\mathbf{x}}^T \frac{d}{dx_2} f(\mathbf{x}) \\ \vdots \\ \nabla_{\mathbf{x}}^T \frac{d}{dx_n} f(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} \frac{d^2}{dx_1^2} f(\mathbf{x}) & \frac{d^2}{dx_1 dx_2} f(\mathbf{x}) & \cdots & \frac{d^2}{dx_1 dx_n} f(\mathbf{x}) \\ \frac{d^2}{dx_1 dx_2} f(\mathbf{x}) & \frac{d^2}{dx_2^2} f(\mathbf{x}) & \cdots & \frac{d^2}{dx_2 dx_n} f(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{d^2}{dx_1 dx_n} f(\mathbf{x}) & \frac{d^2}{dx_2 dx_n} f(\mathbf{x}) & \cdots & \frac{d^2}{dx_n^2} f(\mathbf{x}) \end{pmatrix} \\ &= H(\mathbf{x}) \end{aligned}$$

Méthode de Newton

La récurrence de Newton devient dès lors

$$\mathbf{x}_{k+1} = \mathbf{x}_k - H^{-1}(\mathbf{x}_k) \nabla f(\mathbf{x}_k).$$

On suppose ici que $H(\mathbf{x}_k)$ est définie positive.

Note : nous ne calculons pas $H^{-1}(\mathbf{x}_k)$ pour des questions de précision numériques, mais nous obtenons le pas \mathbf{d}_k en résolvant le système linéaire

$$H(\mathbf{x}_k) \mathbf{d}_k = -\nabla f(\mathbf{x}_k).$$

La récurrence de Newton se réécrit

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{d}_k.$$

Méthode de Newton : difficultés

La convergence n'est assurée que si le point de départ est suffisamment proche de la solution, et est alors quadratique.

On peut obtenir la convergence globale en utilisant la direction de Newton comme direction de descente avec une méthode de recherche linéaire : on cherche une longueur de pas appropriée α_k pour produire le nouvel itéré

$$\mathbf{x}_{k+1} := \mathbf{x}_k + \alpha_k \mathbf{d}_k,$$

en s'assurant que α_k puisse prendre la valeur 1 afin d'assurer une convergence quadratique au cours des dernières itérations.

Méthode primale barrière

Dès lors, la méthode de Newton pour la méthode primale barrière conduit à calculer

$$\mathbf{d}_x = -H^{-1}\nabla L(x)$$

ou encore

$$\begin{aligned} -H\mathbf{d}_x &= c - \mu X^{-1}\mathbf{1} - A^T y \\ \Leftrightarrow H(-\mathbf{d}_x) + A^T y &= c - \mu X^{-1}\mathbf{1} \end{aligned}$$

Nous devons cependant aussi nous assurer que le nouvel itéré $\mathbf{x} + \mathbf{d}_x$ est réalisable, i.e.

$$A(\mathbf{x} + \mathbf{d}_x) = b$$

Ceci nous mène à imposer $A\mathbf{d}_x = 0$.

Méthode primale barrière

Les deux équations précédentes donnent, sous forme matricielle,

$$\begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} -\mathbf{d}_x \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} c - \mu X^{-1} \mathbf{1} \\ 0 \end{pmatrix}.$$

Reprenons la première ligne du système

$$-H\mathbf{d}_x + A^T \mathbf{y} = c - \mu X^{-1} \mathbf{1}$$

et multiplions de part et d'autre part AX^2 . Comme $H = \mu X^{-2}$, nous obtenons

$$-\mu A\mathbf{d}_x + AX^2 A^T \mathbf{y} = AX^2 c - \mu AX \mathbf{1}.$$

Comme $A\mathbf{d}_x = 0$, nous obtenons un système linéaire permettant de trouver \mathbf{y} :

$$AX^2 A^T \mathbf{y} = AX^2 c - \mu AX \mathbf{1}.$$

Méthode primale barrière

Repartons de

$$-H\mathbf{d}_x + A^T \mathbf{y} = c - \mu X^{-1} \mathbf{1}$$

avec \mathbf{y} connu. Nous avons dès lors

$$\begin{aligned} \mathbf{d}_x &= \mu H^{-1} X^{-1} \mathbf{1} - H^{-1} c + H^{-1} A^T \mathbf{y} \\ &= \mathbf{x} + \frac{1}{\mu} X^2 (A^T \mathbf{y} - c). \end{aligned}$$

Nous avons dès lors obtenu la direction \mathbf{d}_x . Nous nous déplacerons le long de cette direction d'une longueur de pas α , obtenue en minimisant le problème barrière le long de \mathbf{d}_x :

$$\min_{\alpha} c^T \mathbf{x}^+ - \mu \sum_{j=1}^n \log x_j^+$$

avec $\mathbf{x}^+ = \mathbf{x} + \alpha \mathbf{d}_x$

Méthode primale barrière

La méthode marche relativement bien si μ est modérément grand, ou si l'algorithme est démarré avec un point proche de la la solution.

Poursuite de chemin primal-dual

Une autre approche consiste à suivre le chemin central à partir d'une paire initiales de solutions primale et duale. À nouveau, considérons la paire

$$\begin{aligned} \min_x \quad & c^T x \\ \text{s.c.} \quad & Ax = b \\ & x \geq 0, \end{aligned} \tag{P}$$

$$\begin{aligned} \max_y \quad & b^T y \\ \text{s.c.} \quad & A^T y + s = c \\ & s \geq 0. \end{aligned} \tag{D}$$

Nous supposons que $\mathring{\mathcal{F}} \neq \emptyset$, ou de manière équivalente

$$\begin{aligned} \mathring{\mathcal{F}}_P &= \{x \mid Ax = b, x > 0\} \neq \emptyset, \\ \mathring{\mathcal{F}}_D &= \{(y, s) \mid s = c - A^T y > 0\} \neq \emptyset. \end{aligned}$$

Conditions d'optimalité

$$A^T y + s = c$$

$$Ax = b$$

$$Xs = 0$$

$$x \geq 0, s \geq 0$$

Les méthodes de points intérieurs primales-duales diffèrent des méthodes primales strictes en un point-clé : la direction de recherche est calculé tant pour x que pour (y, s) , mais numériquement, ce changement a révolutionné la programmation linéaire, et l'approche est maintenant à la base de nombreux solveurs.

Points intérieurs

Nous perturbons les conditions de complémentarité pour obtenir

$$A^T y + s = c$$

$$Ax = b$$

$$XS\mathbf{1} = \mu\mathbf{1}$$

$$x \geq 0, s > 0$$

On cherche dès lors un zéro du système

$$F(x, y, s) = \begin{pmatrix} A^T y + s - c \\ Ax - b \\ XS\mathbf{1} - \mu\mathbf{1} \end{pmatrix}$$

avec $x, s > 0$.

Points intérieurs

Appliquons la méthode de Newton pour résoudre ce système.

Notons $J(Ax) = A$. En effet, la i^e composante de Ax vaut

$$(Ax)_i = \sum_{j=1}^n a_{ij}x_j.$$

Dès lors

$$\nabla_x(Ax)_i = a^i.$$

On peut aussi le voir en remarquant que chaque colonne de la matrice jacobienne s'obtient en dérivant par rapport à la même variable, or

$$Ax = \sum_{i=1}^n x_i a_i.$$

Points intérieurs

En étendant le raisonnement précédent,

$$J(F(x, y, s)) = \begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{pmatrix}$$

Le système à résoudre pour trouver la direction de recherche est à présent

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{pmatrix} \begin{pmatrix} \mathbf{d}_x \\ \mathbf{d}_y \\ \mathbf{d}_s \end{pmatrix} = - \begin{pmatrix} A^T y + s - c \\ Ax - b \\ XS\mathbf{1} - \mu\mathbf{1} \end{pmatrix}$$

Direction de recherche

En prenant x et (y, s) réalisables pour le problème primal et le problème dual, respectivement, le système se simplifie comme

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{pmatrix} \begin{pmatrix} \mathbf{d}_x \\ \mathbf{d}_y \\ \mathbf{d}_s \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \mu \mathbf{1} - XS\mathbf{1} \end{pmatrix}$$

Le nouveau point (x^+, y^+, s^+) est obtenu en se déplaçant le long de $(\mathbf{d}_x, \mathbf{d}_y, \mathbf{d}_s)$ d'une longueur de pas α :

$$(x^+, y^+, s^+) = (x, y, s) + \alpha(\mathbf{d}_x, \mathbf{d}_y, \mathbf{d}_s).$$

Squelette d'algorithme

Étape 0 Choisir $x_0 \in \mathring{\mathcal{F}}_P$, $(y_0, s_0) \in \mathring{\mathcal{F}}_D$, $\mu_0 > 0$, $\gamma \in (0, 1)$, $\epsilon > 0$. Poser $k = 0$, $X_k = \text{diag}(x_k)$, $S_k = \text{diag}(s_k)$.

Étape 1 Si $\mu_k < \epsilon$, arrêt. Sinon, passer à l'étape 2.

Étape 2 Résoudre

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S_k & 0 & X_k \end{pmatrix} \begin{pmatrix} \mathbf{d}_x \\ \mathbf{d}_y \\ \mathbf{d}_s \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \mu \mathbf{1} - X_k S_k \mathbf{1} \end{pmatrix}$$

Étape 3 Calculer α_k tel que $x_k + \alpha_k \mathbf{d}_x > 0$, $s_k + \alpha_k \mathbf{s}_x > 0$ et permettant une réduction suffisante de l'objectif. Poser

$$\begin{aligned} \mu_{k+1} &= \gamma \mu_k, & x_{k+1} &= x_k + \alpha_k \mathbf{d}_x, \\ y_{k+1} &= y_k + \alpha_k \mathbf{d}_y, & s_{k+1} &= s_k + \alpha_k \mathbf{d}_s. \end{aligned}$$

Incrémenter k de 1, et retourner à l'étape 1.

Calcul du centre analytique

Rappelons que le centre analytique peut se calculer en considérons les conditions

$$\mathbf{x} \circ \mathbf{s} = \mathbf{1}$$

$$\mathbf{A}\mathbf{x} = \mathbf{0}$$

$$\mathbf{A}^T \mathbf{y} + \mathbf{s} = \mathbf{c}.$$

De ce qui précède, le système de Newton à considérer pour résoudre ce système est

$$\begin{pmatrix} 0 & \mathbf{A}^T & \mathbf{I} \\ \mathbf{A} & 0 & 0 \\ \mathbf{S} & 0 & \mathbf{X} \end{pmatrix} \begin{pmatrix} \Delta \mathbf{x} \\ \Delta \mathbf{y} \\ \Delta \mathbf{s} \end{pmatrix} = - \begin{pmatrix} \mathbf{A}^T \mathbf{y} + \mathbf{s} - \mathbf{c} \\ \mathbf{A}\mathbf{x} \\ \mathbf{X}\mathbf{S}\mathbf{1} - \mathbf{1} \end{pmatrix}$$

Calcul du centre analytique

Nous pouvons réorganiser le système comme

$$\begin{pmatrix} 0 & 0 & A \\ 0 & X & S \\ A^T & I & 0 \end{pmatrix} \begin{pmatrix} \Delta y \\ \Delta s \\ \Delta x \end{pmatrix} = - \begin{pmatrix} Ax \\ XS\mathbf{1} - \mathbf{1} \\ A^T y + s - c \end{pmatrix}$$

La contrainte $x \circ s = \mathbf{1}$ peut se réécrire comme

$$x - S^{-1}\mathbf{1} = 0$$

Utilisant cette nouvelle expression, similairement à ce qui précède, nous pouvons construire le système de Newton

$$\begin{pmatrix} 0 & 0 & A \\ 0 & S^{-2} & I \\ A^T & I & 0 \end{pmatrix} \begin{pmatrix} \Delta y \\ \Delta s \\ \Delta x \end{pmatrix} = - \begin{pmatrix} Ax \\ x - S^{-1}\mathbf{1} \\ A^T y + s - c \end{pmatrix}$$

Calcul du centre analytique : pas de Newton

En notant

$$r_p = s + A^T y - c, \quad r_d = \begin{pmatrix} Ax \\ x - S^{-1}\mathbf{1} \end{pmatrix}, \quad H = S^{-2}$$

le système devient

$$\begin{pmatrix} 0 & 0 & A \\ 0 & H & I \\ A^T & I & 0 \end{pmatrix} \begin{pmatrix} \Delta y \\ \Delta s \\ \Delta x \end{pmatrix} = - \begin{pmatrix} r_d \\ r_p \end{pmatrix},$$

Après réorganisation des équations, nous pouvons obtenir

$$\Delta y = (AHA^T)^{-1}(Ax - AHr_p)$$

$$\Delta s = -A^T \Delta y - r_p$$

$$\Delta x = -H\Delta s - x + S^{-1}\mathbf{1}$$

Calcul du centre analytique : pas de Newton

L'approche est expliquée en détails dans Goffin et Sharifi-Mokhtarian, "Primal–Dual–Infeasible Newton Approach for the Analytic Center Deep-Cutting Plane Method", Journal of Optimization Theory and Applications 101(1), 1999, pp. 35–58, disponible à l'adresse <https://link.springer.com/article/10.1023/A:1021714926231>.

Calcul du centre analytique : algorithme

Soient $y_0, \epsilon > 0, \alpha \in (0, 1/2), \beta \in (0, 1), s_0 > 0, k = 0, k_{\max}$.

s_0 peut être défini comme précédemment, et on prendra $\alpha = 0.01, \beta = 0.5, k_{\max} = 50$.

Posons $x_0 = 0$, et calculons r_p, r_d .

Répéter

1. Calculer le pas de Newton $(\Delta y, \Delta s, \Delta x)$.
2. Déterminer une longueur de pas t .
3. Mettre à jour la solution :

$$\begin{pmatrix} y_{k+1} \\ s_{k+1} \\ x_{k+1} \end{pmatrix} = \begin{pmatrix} y_k \\ s_k \\ x_k \end{pmatrix} + t \begin{pmatrix} \Delta y_k \\ \Delta s_k \\ \Delta x_k \end{pmatrix}$$

4. Poser $k := k + 1$.

Jusqu'à $s = c - A^T y$ et $\|r(y, s, \lambda)\|_2 \leq \epsilon$, ou $k > k_{\max}$.

Calcul du centre analytique : longueur du pas

Cela se fait par *backtracking* (recherche arrière) sur $\|r\|_2$.

Posons $t = 1$.

Tant que des composantes de $s + t\Delta s$ sont négatives ou nulles, faire $t := \beta t$.

Tant que

$$\|r(y + t\Delta y, s + t\Delta s, \lambda + t\Delta \lambda)\|_2 > (1 - \alpha t)\|r(y, s, \lambda)\|_2$$

faire $t := \beta t$.

Un exemple d'algorithme

Source : Sanjay Mehrotra, “On the Implementation of a Primal-Dual Interior Point Method”, SIAM Journal on Optimization 2(4), pp. 575–601.

Avec l'aide d'Arnaud L'Heureux.

La méthode nécessite un point de départ satisfaisant les contraintes de non-négativité, mais pouvons violer les autres contraintes.

Point de départ

Nous commençons par calculer

$$\bar{x} = A^T(AA^T)^{-1}b$$

$$\bar{\lambda} = (AA^T)^{-1}c$$

$$\bar{s} = c - A^T\bar{\lambda}$$

puis

$$\delta_x = \max\{-1.5 \min\{\bar{x}_i\}, 0\}$$

$$\delta_s = \max\{-1.5 \min\{\bar{s}_i\}, 0\}$$

Finalement, nous calculons

$$\overline{\delta_x} = \delta_x + 0.5 \frac{(\bar{x} + \delta_x \mathbf{1})^T (\bar{s} + \delta_s \mathbf{1})}{\sum_{i=1}^n (\bar{s}_i + \delta_s)}$$

$$\overline{\delta_s} = \delta_s + 0.5 \frac{(\bar{x} + \delta_x \mathbf{1})^T (\bar{s} + \delta_s \mathbf{1})}{\sum_{i=1}^n (\bar{x}_i + \delta_x)}$$

Point de départ

Le point de départ de l'algorithme est alors calculé come

$$x_0 = \bar{x} + \overline{\delta_x} \mathbf{1}$$

$$\lambda_0 = \bar{\lambda}$$

$$s_0 = \bar{s} + \overline{\delta_s} \mathbf{1}$$

Les constantes 1.5 et 0.5 sont arbitraires. La même heuristique fonctionne en remplaçant 1.5 par $k > 1$ et 0.5 par $\ell > 0, \ell \in \mathbb{R}$.

Système KKT

Rappelons les conditions KKT perturbées :

$$Ax = b$$

$$A^T \lambda + s = c$$

$$XS\mathbf{1} = \mu\mathbf{1}$$

$$x \geq 0, s \geq 0$$

où $X = \text{diag}(x)$, $S = \text{diag}(s)$.

Soit $u = (x, \lambda, s)$. Si u n'est pas optimal, nous appliquons la méthode de Newton au système

$$F_\mu(u) = \begin{pmatrix} A^T \lambda + s - c \\ Ax - b \\ XS\mathbf{1} - \mu\mathbf{1} \end{pmatrix} = 0, (x, s) \geq 0$$

où $\mu > 0$ est le paramètre de barrière.

Direction de Newton

La direction Δu est donnée par

$$J_\mu(u)\Delta u = F_\mu(u)$$

soit

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta \lambda \\ \Delta s \end{pmatrix} = \begin{pmatrix} A^T \lambda + s - c \\ Ax - b \\ XS - \mu \mathbf{1} \end{pmatrix}$$

Multipliant la dernière ligne par la gauche par S^{-1} ,

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ I & 0 & S^{-1}X \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta \lambda \\ \Delta s \end{pmatrix} = \begin{pmatrix} r_d \\ r_p \\ S^{-1}r_c \end{pmatrix}$$

où r_d , r_p et r_c sont respectivement les résidus du dual, de primal et de complémentarité.

Direction de Newton

Notant $M = AXS^{-1}A^T$ et $t_d = r_p - AS^{-1}(r_c - Xr_d)$, nous pouvons montrer que le pas est

$$\Delta\lambda = M^{-1}t_d$$

$$\Delta s = r_d - A^T \Delta\lambda$$

$$\Delta x = S^{-1}(r_c - X\Delta s)$$

La solution est mise à jour avec

$$\begin{pmatrix} \lambda^+ \\ s^+ \\ x^+ \end{pmatrix} = \begin{pmatrix} \lambda \\ s \\ x \end{pmatrix} - \alpha \begin{pmatrix} \Delta\lambda \\ \Delta s \\ \Delta x \end{pmatrix}$$

où α est choisi pour garantir $(x^+, s^+) \geq 0$.

Terminaison

μ est calculé suivant le saut de dualité à chaque itération :

$$\mu = \frac{x^T s}{n}.$$

L'algorithme s'arrête lors que $\max(\mu, \|r_p\|, \|r_d\|) < \epsilon$, avec $\epsilon > 0$ choisi à l'avance. Autrement dit, nous exigeons que la solution satisfasse, à une tolérance près, la faisabilité primale, duale, et les conditions de complémentarité.

Algorithme de Mehrotra

L'algorithme est divisé en deux étapes :

1. étape prédictive,
2. étape correctrice.

Étape prédictive

L'étape prédictive consiste à calculer le pas de Newton classique, en résolvant

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ I & 0 & S^{-1}X \end{pmatrix} \begin{pmatrix} \Delta x_p \\ \Delta \lambda_p \\ \Delta s_p \end{pmatrix} = \begin{pmatrix} r_d \\ r_p \\ S^{-1}r_c \end{pmatrix}$$

puis en limitant le longueur de pas pour satisfaire les contraintes de non-négativité.

Calculons d'abord

$$\alpha_p^p = \min \left\{ 1, \min_{\Delta x_i^p > 0} \frac{x_i}{\Delta x_i^p} \right\}$$
$$\alpha_d^p = \min \left\{ 1, \min_{\Delta s_i^p > 0} \frac{s_i}{\Delta s_i^p} \right\}$$

Étape prédictive

Le solution n'est pas nécessairement proche du chemin primal-dual. Nous nous en rapprochons en calculant tout d'abord le paramètre

$$\sigma = \left(\frac{(x - \alpha_p^p \Delta x_p)^T (s - \alpha_d^p \Delta s_p)}{n\mu} \right)^3,$$

comparant la solution prédite avec le saut dualité théorique.

Nous résolvons ensuite le système

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ I & 0 & S^{-1}X \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta \lambda \\ \Delta s \end{pmatrix} = \begin{pmatrix} r_d \\ r_p \\ S^{-1}(Xs - \sigma\mu\mathbf{1} + \Delta x_p \Delta s_p) \end{pmatrix}$$

Étape de correction

Nous calculons finalement le pas primal et dual. Soit

$$\alpha_p = \min \left\{ 1, \eta \min_{\Delta x_i > 0} \frac{x_i}{\Delta x_i} \right\}$$
$$\alpha_d = \min \left\{ 1, \eta \min_{\Delta s_i > 0} \frac{s_i}{\Delta s_i} \right\}$$

où $\eta = \max\{0.995, 1 - \mu\}$, le nouveau point est

$$x^+ = x - \alpha_p \Delta x$$

$$\lambda^+ = \lambda - \alpha_d \Delta \lambda$$

$$s^+ = s - \alpha_d \Delta s$$