COURS D'OPTIMISATION
ISIMA – F4 2ème année

Jonas Koko

ISIMA

Plan du cours

Documents

Références Bibliographiques

- J.-B. Hiriart-Urruty, Que sais-je ? no 3184, L'Optimisation.
- D. Luenberger, Introduction to Linear and Nonlinear Programming Addison-Wesley (1984)
- J.F. Bonnans et al., Optimisation Numérique Springer (1997)
- M. Minoux, Programmation Mathématique Lavoisier TEC & DOC (2008)
- M. Bierliaire, Introduction à l'Optimisation Différentiable Presses Poly. Univ. Romandes (2006)

Historique

- Pierre de Fermat (1601-1665)
  Recherche d'extrema de fonctions et de tangentes aux courbes
- Isaac Newton (1643-1727)
  Fonctions dérivables, calcul infinitésimal, Un problème d'aérodynamique.
- Johann Bernoulli (1667-1748)
  Le brachistochrone
- 18ème siècle : Euler et Lagrange
- 19ème siècle : Legendre, Hamilton, Jacobi, Weierstrass,...
- 20ème siècle : Von Neumann, Bellman, Pontryagin,...

Classification des problèmes

- Calcul des Variations
- Commande Optimale
- Programmation Linéaire
- Programmation Non Linéaire
- Optimisation de Formes
- Optimisation Combinatoire
- Optimisation Stochastique
- Optimisation Non Lisse
- Optimisation Multicritère
- Optimisation Globale

Calcul des Variations

Problème de Plateau (Surface minimale)

$$\min_{u\in V} F(u) = \int_\Omega \sqrt{1+|\nabla u|^2}\,\mathrm{d}x$$

Le brachistochrone

$$\min_{u\in V} F(u) = \frac{1}{\sqrt{2g}} \int_{x_1}^{x_2} \frac{\sqrt{1+u'(x)^2}}{\sqrt{u(x)}}\,\mathrm{d}x$$

Problème général

$$\min_{u\in V} F(u) = \int_\Omega f(x, u(x), \nabla u(x))\,\mathrm{d}x$$

Commande Optimale

$$\min_u \int_a^b L(x(t), u(t), t)dt + \Phi(x(b), b)$$

$$\dot{x} = f(x(t), u(t), t)$$

$$x(a) = x_a$$

- $x(t)$ état du système à l'instant t
- $u(t)$, la commande appliquée à l'instant t
- $L(x(t), u(t), t)$ coût de la trajectoire associée

Sous réserve de conditions de contrôlabilité et d'atteignabilité de l'état final il existe :

- il existe une commande optimale u^*,
- une trajectoire optimale x^*
- un état adjoint, noté p^*

Commande Optimale (II)

Fonction de Hamilton (Hamiltonien)

$$H(x, p, u) = L(x, u) + \langle p, f(x, u, t)\rangle$$

Condition d'optimalité

(u^*, x^*, p^*) vérifie :

$$\begin{aligned} \dot{x}^*(t) &= \frac{\partial H}{\partial p}(x^*, p^*, u^*) \\ \dot{p}^*(t) &= -\frac{\partial H}{\partial x}(x^*, p^*, u^*) \\ p^*(b) &= \frac{\partial \Phi}{\partial x}(x^*(b), b) \end{aligned}$$

et La commande optimale u^* maximise la fonction H.

Principe du Maximum de Pontryagin

Questions d'existence dans $\mathbb{R}^n$

Problème sans contrainte

$$(P) \qquad \min_x f(x)$$

Si f continue et $f(x) \to +\infty$ si $\| x \| \to +\infty$ (f coercive)
alors (P) admet une solution

Problème avec contraintes

$$(P_C) \qquad \min_{x \in S} f(x)$$

Si f continue et S compact non vide ou Si f coercive et S fermé
alors (P_C) admet une solution.

Conditions d'optimalité dans $\mathbb{R}^n$ (I)

Problème d'optimisation sans contrainte :

$$\min_x f(x)$$

Hypothèse raisonnable

f de classe $\mathscr{C}^2$ sur $\mathbb{R}$.

Conditions d'optimalité

- $\nabla f(x^*) = 0$
- $\nabla^2 f(x^*)$ semi-définie positive.

Conditions d'optimalité dans $\mathbb{R}^n$ (II)

Problème d'optimisation avec contraintes

$$\min f(x)$$
$$g_i(x) = 0, \quad i = 1, \dots, p.$$

Hypothèses raisonnables

f de classe $\mathscr{C}^2$ sur $\mathbb{R}$, et g_i de classe $\mathscr{C}^1$.

Le Lagrangian : $L(x, \lambda) = f(x) + \langle \lambda, g(x) \rangle$.

Conditions d'optimalité

$$\nabla_x f(x^*) + \sum_{i=1}^{n} \lambda_i \nabla g_i(x^*) = 0, \qquad g_i(x^*) = 0$$

$\langle z, \nabla_x^2 L(x^*, \lambda) z \rangle \geq 0$, $\forall z$ tel que $\nabla g(x^*) z = 0$.

Conditions d'optimalité dans $\mathbb{R}^n$ (III)

Problème d'optimisation avec contraintes (inégalités)

$$\min f(x)$$
$$h_j(x) \leq 0, \quad j = 1, \ldots, q.$$

Hypothèses raisonnables

f de classe $\mathscr{C}^2$ sur $\mathbb{R}$, et g_j de classe $\mathscr{C}^1$.

Conditions d'optimalité (Karush-Kuhn-Tucker)

$$\nabla f(x^*) + \sum_{j=1}^{q} \mu_j \nabla h_j(x^*) = 0, \quad h_j(x^*)\mu_j = 0, \quad \mu_j \geq 0$$

Méthodes de Newton

Convergence linéaire, quadratique

Définition (Convergence linéaire)

x^k converge linéairement vers x^*, à la vitesse $\alpha \in]0,\ 1[$ si $\exists k_\alpha \in \mathbb{N}$ tel que :

$$\| x^{k+1} - x^* \| \leq \alpha \| x^k - x^* \|, \quad \forall k > k_\alpha.$$

Définition (Convergence super-linéaire)

$x^k \to x^*$ super-linéairement si $x^k \to x^*$ linéairement, $\forall \alpha \in]0,\ 1[$, i.e. :

$$\lim_{k \to +\infty} \frac{\| x^{k+1} - x^* \|}{\| x^k - x^* \|} = 0$$

Définition (Convergence quadratique)

$x^k \to x^*$ quadratiquement si $\exists c > 0$ tel que :

$$\| x^{k+1} - x^* \| \leq c \| x^k - x^* \|^2, \quad \forall k \in \mathbb{N}.$$

Equation à une inconnue

f fonction de R dans R différentiable

$$f(x) = 0$$

Modèle (approximation) linéaire en $\bar{x}$

$$m_{\bar{x}}(x) = f(\bar{x}) + f'(\bar{x})(x - \bar{x})$$

Racine du modèle linéaire en $\bar{x}$

$$x^+ = \bar{x} - \frac{f(\bar{x})}{f'(\bar{x})}$$

Méthode de newton

x^0 donné

$$x^{k+1} = x^k - \frac{f(x^k)}{f'(x^k)}$$

Arrêt si $|f(x^k)| < \varepsilon$

Exemple I : convergence

$f(x) = x^2 - 2,\ f'(x) = 2x,\ x^* = \sqrt{2}$

$\varepsilon = 10^{-10},\ x^0 = 4$

iter	x^k	$f(x^k)$
1	2.2500000000	14.0000000000
2	1.5694444444	3.0625000000
3	1.4218903638	0.4631558642
4	1.4142342859	0.0217722067
5	1.4142135625	0.0000586155
6	1.4142135624	0.0000000004
7	1.4142135624	0.0000000000

Table: Historique des itérations

Exemple II : non convergence

$f(x) = \mathrm{atan}(x)$, $f'(x) = 1/(1+x^2)$, $x^* = 0$

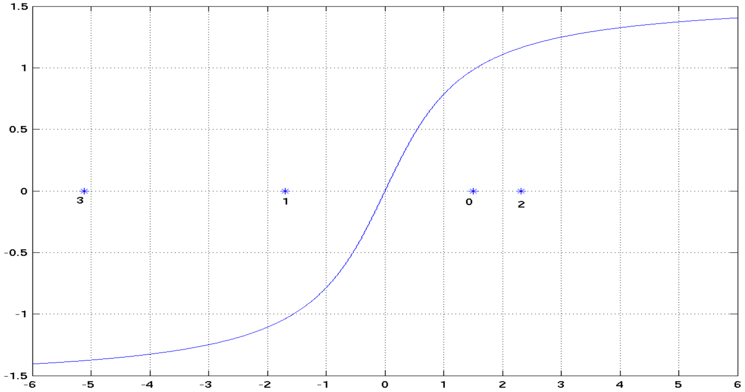
$\varepsilon = 10^{-10}$, $x^0 = 1.5$

iter	x^k	$f(x^k)$	$f'(x^k)$
0	1.5000000000e+00	9.8279372325e-01	3.0769230769e-01
1	-1.6940796006e+00	9.8279372325e-01	3.0769230769e-01
2	2.3211269614e+00	-1.0375463591e+00	2.5840422980e-01
3	-5.1140878368e+00	1.1640020424e+00	1.5655257770e-01
4	3.2295683914e+01	-1.3776945287e+00	3.6827130031e-02
5	-1.5753169508e+03	1.5398423269e+00	9.5784413088e-04
6	3.8949760078e+06	-1.5701615340e+00	4.0296185091e-07
7	-2.3830288974e+13	1.5707960701e+00	6.5915936439e-14
8	8.9202801611e+26	-1.5707963268e+00	1.7609271216e-27
9	-1.2499045994e+54	1.5707963268e+00	1.2567329759e-54
10	2.4539946375e+108	-1.5707963268e+00	6.4009770143e-109

Exemple II : non convergence

$f(x) = \mathrm{atan}(x)$, $f'(x) = 1/(1 + x^2)$, $x^* = 0$

$\varepsilon = 10^{-10}$, $x^0 = 1.5$



Système non linéaire

f de $\mathbb{R}^n$ dans $\mathbb{R}^n$ différentiable

$$f(x) = 0.$$

Modèle linéaire en $\bar{x}$

$$m_{\bar{x}}(x) = f(\bar{x}) + \nabla f(\bar{x})(x - \bar{x})$$

Solution du modèle linéaire en $\bar{x}$

$$\nabla f(\bar{x})\bar{d} = -f(\bar{x}) \quad \text{et} \quad x^{+} = \bar{x} + \bar{d}.$$

Méthode de newton

- x^0 donné
- tant que $\| f(x^k) \| > \varepsilon$ et $\| x^k - x^{k-1} \| > \varepsilon \| x^k \|^{-1}$
  1. Résoudre $\nabla f(x^k)d^k = -f(x^k)$
  2. Mise à jour $x^{k+1} = x^k + d^k$

Convergence

Soit x^* la solution du système
Si x^0 est «suffisamment» proche de x^*
$\nabla f(x)$ toujours inversible :

$$\| x^{k+1} - x^* \| \leq C \| x^k - x^* \|^2$$

C constante. La convergence de x^k vers x^* est quadratique.

Exercice

On souhaite résoudre par la méthode de Newton l'équation

$$x^2 - a = 0, \quad a > 0.$$

1.– Donner le terme général de la suite générée par la méthode de Newton.
2.– Montrer que la suite générée converge quadratiquement.

Méthodes de Newton locales

Equations d'optimalité

Problème d'optimisation : $\min_{x \in \mathbb{R}^n} f(x)$

Système d'optimalité : $\nabla f(x) = 0$

Méthode de Newton

- x_0 donné
- tant que $\| \nabla f(x^k) \| > \varepsilon$
  1. Résoudre $\nabla^2 f(x_k) d_k = -\nabla f(x_k)$
  2. $x_{k+1} = x_k + d_k$

Remarque

La méthode peut converger vers un point autre qu'un minimum local.

Exemple de «mauvaise» convergence

$f(x) = \frac{1}{2}x_1^2 + x_1 \cos x_2$

$$\nabla f(x) = \begin{bmatrix} x_1 + \cos x_2 \\ -x_1 \sin x_2 \end{bmatrix} \qquad \nabla^2 f(x) = \begin{bmatrix} 1 & -\sin x_2 \\ -\sin x_2 & -x_1 \cos x_2 \end{bmatrix}$$

Il existe une infinité de minima locaux : $k \in \mathbb{Z}$

$$x_k^* = ((-1)^{k+1}, k\pi) \qquad \nabla^2 f(x_k^*) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

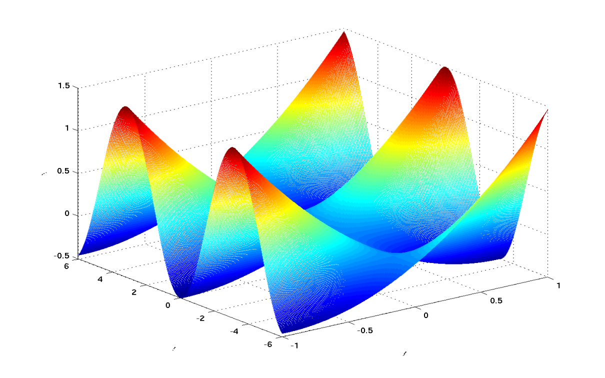
iter	x_k	$\nabla f(x_k)$
0	(1.00000e+00, 1.00000)	(1.5403023059e+00, -8.4147098481e-01)
1	(-2.3384e-01, 1.36419)	(1.5403023059e+00, -8.4147098481e-01)
2	(1.08143e-02, 1.58483)	(-2.8707702717e-02, 2.2887198619e-01)
3	(-2.13237e-06, 1.57079)	(-3.2252480702e-03, -1.0813309420e-02)
4	(1.99048e-17, 1.57079)	(9.2282870662e-07, 2.1323766610e-06)
5	(0.00000e+00, 1.57079)	(8.1137190701e-17, -1.9904850744e-17)

Table: Historique des itérations

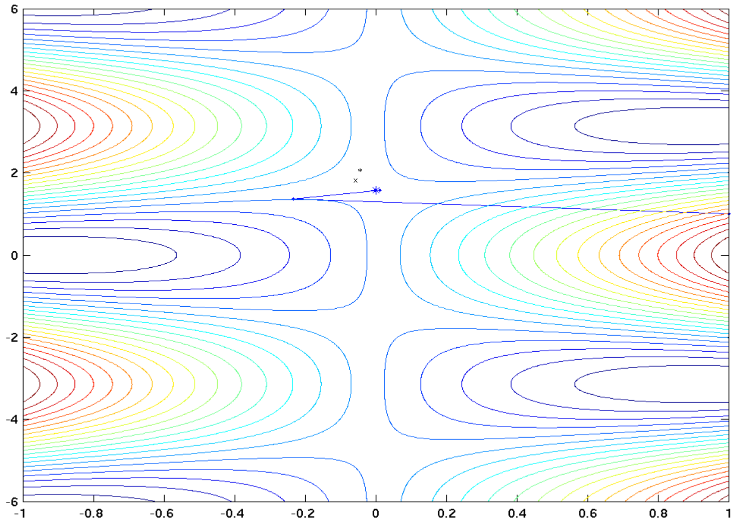
Exemple de «mauvaise» convergence

Matrice Hessienne au point $x^* = (0, 1.5708)$

$$\nabla^2 f(x^*) = \begin{bmatrix} 1 & -1 \\ -1 & 0 \end{bmatrix}$$



Courbes de niveau de la fonction



Approximation quadratique

$$m_{\bar{x}}(x) = f(\bar{x}) + \nabla f(\bar{x})^T(x-\bar{x}) + \frac{1}{2}(x-\bar{x})^T\nabla^2 f(\bar{x})(x-\bar{x})$$

$d := x - \bar{x}$

$$m_{\bar{x}}(\bar{x}+d) = f(\bar{x}) + \nabla f(\bar{x})^T d + \frac{1}{2}d^T\nabla^2 f(\bar{x})d$$

On calcule $\bar{d} = \arg\min m_{\bar{x}}(\bar{x}+d)$:

$$\nabla m_{\bar{x}}(\bar{x}+d) = \nabla^2 f(\bar{x})d + \nabla f(\bar{x}) = 0$$

i.e.

$$\bar{d} = -\left[\nabla^2 f(\bar{x})\right]^{-1}\nabla f(\bar{x}), \qquad x^+ = \bar{x} + \bar{d}$$

Remarque

$\bar{d}$ minimise $m_{\bar{x}}$ si $\nabla^2 f(\bar{x})$ est définie positive

Algorithme de Newton locale

Algorithme de Newton locale

- x^0 donné
- Tant que $\| \nabla f(x_k) \| < \varepsilon$

1. Construire le modèle quadratique

$$m_{x_k}(x_k + d) = f(x_k) + \nabla f(x_k)^T d + \frac{1}{2} d^T \nabla^2 f(x_k) d$$

2. Calculer

$$d_k = \arg\min_d m_{x_k}(x_k + d)$$

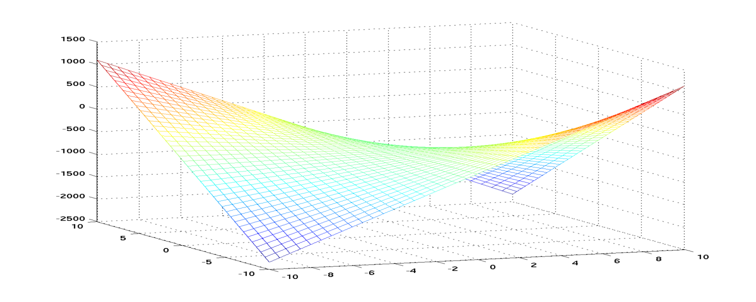
3. $x_{k+1} = x_k + d_k$

Exemple non convergence

$f(x) = \frac{1}{2}x_1^2 + x_1 \cos x_2$: Pour $x_0 = (1,1)$ on a

$$\nabla f(x_0) = \begin{bmatrix} 1.54030 \\ -0.84147 \end{bmatrix} \qquad \nabla^2 f(x_0) = \begin{bmatrix} 1.00000 & -0.84147 \\ -0.84147 & -0.54030 \end{bmatrix}$$

$\mathrm{Sp}\nabla^2 f(x_0) = \{-0.91085,\ 1.37055\}$.



L'approximation quadratique n'a pas de minimum

Méthodes de descente

Algorithme de descente

Algorithme en deux étapes :

1. Recherche d'une direction de descente
2. calcul d'un pas de déplacement

Algorithme de descente

- Recherche d'une direction de descente : $\nabla f(x_k)^T d_k < 0$
- Recherche linéaire : t_k tel que $f(x_k + t_k d_k) < f(x_k)$
- Mise à jour : $x_{k+1} \leftarrow x_k + t_k d_k$

Exemple de direction de descente :

$$d_k = -\nabla f(x_k)$$

Exemple de pas de déplacement :

$$t_k = \arg\min_{t>0} f(x_k + td_k)$$

Choix direction et pas

Direction suffisamment descendante

$$\begin{aligned} \nabla f(x_k)^T d_k &\leq -\gamma_0 \| \nabla f(x_k) \|^2 && (1) \\ \| d_k \| &\leq \gamma_1 \| \nabla f(x_k) \| && (2) \end{aligned}$$

Pas de déplacement admissible (Conditions de Wolfe)

$$\begin{aligned} f(x_k + t_k d_k) - f(x_k) &\leq \tau_0 t_k \nabla f(x_k)^T d_k, \quad \tau_0 \in]0, 1/2[&& (3) \\ \nabla f(x_k + t_k d_k)^T d_k &\geq \tau_1 \nabla f(x_k)^T d_k, \quad \tau_1 \in]\tau_0, 1[. && (4) \end{aligned}$$

(3) : Diminution suffisante
(4) : Progression suffisante.

Convergence

Théorème

Si $\{x_k\}$ généré suivant (1)-(4), alors tout point d'accumulation x^* de la suite est un point stationnaire de f.

$f(x_k) > f(x_{k+1}) > f(x^*)$, $\forall k$ $\{f(x_k)\}$ décroissante et minorée.

$$\lim_{k\to+\infty} f(x_{k+1}) - f(x_k) = \lim_{k\to+\infty} f(x_k + t_k d_k) - f(x_k) = 0.$$

1ère condition de Wolfe $\Longrightarrow \lim_{k\to+\infty} \tau_0 t_k \nabla f(x_k)^T d_k = 0$

Soit $t_k \to 0$, soit $\nabla f(x_k)^T d_k \to 0$

Si $t_k \to 0$, (2) $\Longrightarrow \| t_k d_k \| \to 0$

$$\lim_{k\to+\infty} \nabla f(x_k + t_k d_k)^T d_k = \lim_{k\to+\infty} \nabla f(x_k)^T d_k \geq \lim_{k\to+\infty} \tau_1 \nabla f(x_k)^T d_k$$

Il faut que $\nabla f(x_k)^T d_k \to 0$ et (1) $\Longrightarrow \nabla f(x_k) \to 0$

Recherche linéaire I : Règle d'Armijo

Règle d'Armijo

- $0 < \alpha < 1$ fixé
- Prendre t_k comme le premier terme de la suite $\{1, \alpha, \alpha^2, \ldots, \}$ qui vérifie

$$f(x_k + td_k) - f(x_k) \leq t\tau_0 \nabla f(x_k)^T d_k, \quad \tau_0 \in]0, 1/2[.$$

Hypothèse principale d'Armijo : le pas n'est jamais trop petit.

Exercice

Montrer qu'avec $f(x) = x^4$, la méthode de la plus forte pente ne converge pas avec la règle d'Armijo.

Recherche linéaire II

Algorithme de recherche linéaire

1. $i = 0$, $\tau_0 = 10^{-4}$, $\tau_1 = 0.1$, $\lambda = 2$, $t_L = 0\,;\ t_R = +\infty$
2. Si t_i vérifie les conditions de Wolfe : $t_k = t_i$, STOP
3. Si $f(x_k + t_i d_k) - f(x_k) > \tau_0 t_i \nabla f(x_k)^T d_k$ (le pas est trop long)

   $t_R = t_i$, $t_{i+1} = (t_L + t_R)/2$. Aller en 2.
4. Si $t_i\ \nabla f(x_k + t_i d_k)^T d_k < \tau_1 \nabla f(x_k)^T d_k$ (le pas est trop court)

   $t_L = t_i$, $t_{i+1} = (t_L + t_R)/2$ si $t_R < +\infty$ $t_{i+1} = \lambda t_i$. Aller en 2.

Théorème

Si f est $\mathscr{C}^1$ et coercive, alors l'algorithme se termine en un nombre fini d'itérations

Newton avec recherche linéaire

On suppose $\nabla^2 f(x_k)$ définie positive :

$$x_{k+1} = x_k - t_k \left[\nabla^2 f(x_k)\right]^{-1} \nabla f(x_k)$$

t_k vérifiant les conditions de Wolfe (3)-(4) (ou Armijo)

Exercice

Montrer que si $\nabla^2 f(x_k)$ est définie positive, la direction de Newton est suffisamment descendante.

On suppose que $\nabla^2 f(x_k)$ n'est pas définie positive :

$$d_k = -S_k \nabla f(x_k)$$

S_k modification de $\nabla^2 f(x_k)^{-1}$:

$$S_k = \left[\nabla^2 f(x_k)\right]^{-1} + (\varepsilon - \lambda_{\min})\mathbb{I}$$

Algorithme de Newton avec recherche linéaire

Itération $k \geq 0$

1. Calculer R_k tel que

$$R_k R_k^T = \nabla^2 f(x_k) + \tau \mathbb{I}, \quad \tau \geq 0.$$

2. Résoudre $R_k z_k = \nabla f(x_k)$
3. Résoudre $R_k^T d_k = -z_k$
4. Recherche linéaire avec $t_0 = 1$
5. $x_{k+1} = x_k + t_k d_k$

Remarques

- Le calcul de $\nabla^2 f(x_k)$ peut être coûteux ou impossible
- La modification avec factorisation de Choleski est coûteuse.

Méthodes quasi-Newton

On calcule la direction de Newton approchée :

$$d_k = -S_k \nabla f(x_k) \text{ ou } G_k d_k = -\nabla f(x_k)$$

- S_k approximation (définie positive) de $[\nabla^2 f(x_k)]^{-1}$
- G_k approximation (définie positive) de $\nabla^2 f(x_k)$

$g_k \leftarrow \nabla f(x_k)$, $p_k \leftarrow x_{k+1} - x_k$, $q_k \leftarrow \nabla f(x_{k+1}) - \nabla f(x_k)$

Mise à jour de Broyden-Fletcher-Goldfarb-Shanno (BFGS)

$$\begin{aligned} G_{k+1} &= G_k + \frac{q_k q_k^T}{\langle p_k, q_k \rangle} - \frac{G_k p_k p_k^T G_k}{\langle p_k, G_k p_k \rangle} \\ S_{k+1} &= (I - \frac{p_k q_k^T}{\langle p_k, q_k \rangle}) S_k (I - \frac{p_k q_k^T}{\langle p_k, q_k \rangle}) + \frac{p_k p_k^T}{\langle p_k, q_k \rangle} \end{aligned}$$

Méthodes de régions de confiance

Principe de la méthode

Modèle quadratique en $\bar{x}$

$$m_{\bar{x}}(\bar{x}+d) = f(\bar{x}) + \nabla f(\bar{x})^T d + \frac{1}{2} d^T \nabla^2 f(\bar{x}) d$$

Principe de la méthode

Faire confiance au modèle quadratique dans une boule de rayon $\bar{r}$, i.e.

$$\| x - \bar{x} \| \leq \bar{r}$$

Sous-problème de région de confiance

$$\min_d m_{\bar{x}}(\bar{x}+d) \tag{5}$$

$$\| d \| \leq \bar{r}. \tag{6}$$

Questions

- Comment résoudre (5)-(6) ?
- Comment déterminer $\bar{r}$?

Un peu de dualité !

On remplace (6) par $(1/2)(\| d \|^2 - \bar{r}^2) \leq 0$.

Conditions d'optimalité de Kuhn-Tucker

d^* solution du problème de région de confiance $\Longrightarrow \exists \mu^* \geq 0$

$$\begin{aligned} \nabla f(\bar{x}) + \nabla^2 f(\bar{x}) d^* + \mu^* d^* &= 0 \\ \mu^* (\| d \|^2 - \bar{r}^2) &= 0 \end{aligned}$$

Si $\| d^* \| < \bar{r} \Longrightarrow \mu^* = 0$ et $\nabla^2 f(\bar{x}) d^* = -\nabla f(\bar{x})$ (Newton locale)

Si $\| d^* \| = \bar{r} \Longrightarrow \mu^* > 0$ et

$$\left(\nabla^2 f(\bar{x}) + \mu^* \mathbb{I}\right) d^* = -\nabla f(\bar{x})$$

La matrice hessienne de f en $\bar{x}$ est corrigée par $\mu^* \mathbb{I}$.

Méthode du gradient conjugué (Rappel)

$$\min_x q(x) = \frac{1}{2} x^T Q x + b^T x$$

On suppose Q $(n \times n)$ définie positive

Algorithme

- $k = 0$: x_0, $\nabla q(x_0) = Qx_0 + b$, $d_0 = -\nabla q(x_0)$
- $k \geq 0$
  1. $t_k = -\nabla q(x_k)^T d_k / d_k^T Q d_k$
  2. $x_{k+1} = x_k + t_k d_k$
  3. $\beta_k = \| \nabla q(x_{k+1}) \|^2 \| \nabla q(x_k) \|^{-2}$
  4. $d_{k+1} = -\nabla q(x_{k+1}) + \beta_k d_k$

Convergence (quadratique) en n itération au plus

Méthode du gradient conjugué modifié (Steihaug-Toint)

- Si le modèle est non convexe dans la direction d, aller au bord de la région de confiance en suivant d. STOP
- Si $x^+ = \bar{x} + d$ est à l'extérieur de la région de confiance, aller au bord de la région de confiance en suivant d. STOP
- Dans les autres cas : Gradient Conjugué

Sous-problème à résoudre

Déterminer λ tel que $\| \bar{x} + \lambda d \| = \bar{r}$

Solution

$a = \| d \|^2$, $b = 2x^T d$, $c = \| x \|^2 - \bar{r}^2$:

$$\lambda = \frac{-b + \sqrt{b^2 - 4ac}}{2a}$$

Méthode du gradient conjugué modifié (Steihaug-Toint) II

Algorithme

- $k \leftarrow 0$: $r > 0$, $x_0 = 0$, $d_0 = -b$
- $k \geq 0$
  1. Si $d_k^T Q d_k \leq 0$ (test de convexité), $x^* = x_k + \lambda_k d_k$. STOP
  2. Pas de déplacement : $t_k = -d_k^T(Qx_k + b)/d_k^T Q d_k$
  3. Itéré suivant : $x_{k+1} = x_k + t_k d_k$
  4. Si $\| x_{k+1} \| > r$ (test de région), $x^* = x_k + \lambda_k d_k$. STOP
  5. $\beta_k = \| \nabla q(x_{k+1}) \|^2 \| \nabla q(x_k) \|^{-2}$
  6. Nouvelle direction du gradient conjugué : $d_{k+1} = -\nabla q(x_{k+1}) + \beta_k d_k$

Arrêt normal : $\| \nabla q(x_k) \| = \| Qx_k + b \| < \varepsilon$

Qualité du modèle

Prédiction de la diminution :

$$\Delta m^k = m_{x_k}(x_k) - m_{x_k}(x_k + d^*) = f(x_k) - m_{x_k}(x_k + d^*)$$

Diminution réelle de la fonction :

$$\Delta f^k = f(x_k) - f(x_k + d^*)$$

Ratio :

$$\rho = \frac{\Delta f^k}{\Delta m^k}$$

Trois cas possibles :

- ρ proche de 1, le modèle est bon
- ρ proche de 0, le modèle est mauvais
- ρ prend des valeurs intermédiaires

Calcul pratique du rayon de confiance

Choix de 2 constantes $0 < \eta_1 \leq \eta_2 < 1$

- Si $\rho > \eta_2$ l'adéquation modèle/fonction est très bonne

  Diminution prédite par le modèle presque atteinte par le fonction

  On double le rayon

- Si $\eta_1 \leq \rho < \eta_2$ l'adéquation modèle/fonction n'est pas parfaite

  La fonction a quand même diminiué

  On garde le rayon inchangé

- Si $\rho < \eta_1$ l'adéquation modèle/fonction est mauvaise

  Le rayon est réduit à $\frac{1}{2} \parallel d^* \parallel$

Algorithme de Newton avec région de confiance

- $k = 0$: x_0, $\eta_1 = 0.01$, $\eta_2 = 0.9$
- $k \geq 0$

1. Construire le modèle quadratique $m_{x_k}(x_k + d)$, $d \in \mathbb{R}^n$

   Résoudre le problème de région de confiance (PRC)

   d_k la solution

2. Calculer le ratio

$$\rho = \frac{\Delta f^k}{\Delta m^k}$$

3. Si $\rho < \eta_1$, alors $x_{k+1} = x_k$ et $r_k = \frac{1}{2} \parallel d^* \parallel$

4. Si $\rho \geq \eta_1$, alors $x_{k+1} = x_k + d_k$

   Si $\rho \geq \eta_2$, alors $r_{k+1} = 2r_k$

   Sinon $r_{k+1} = r_k$

Arrêt dès que $\parallel \nabla f(x_k) \parallel < \varepsilon$

Convergence

Théorème (Convergence de l'algorithme)

Si $f \in \mathscr{C}^2$

$\{x_k\} \subset B \subset \mathbb{R}^n$, B *borné*

$\| \nabla^2 f(x) \|_2 \leq \Lambda, \forall x \in B.$

Alors il existe un point d'accumulation x^* *de la suite tel que :*

$$\begin{aligned} \nabla f(x^*) &= 0 \\ s^T \nabla^2 f(x^*) s &\geq 0, \quad \forall s \in \mathbb{R}^n. \end{aligned}$$

Point de Cauchy - Point de Newton

Modèle quadratique en x_k $(g_k = \nabla f(x_k))$

$$m_{x_k}(x_k + d_k) = f(x_k) + g_k^T d + \frac{1}{2} d^T \nabla^2 f(x_k) d$$

Définition (Point de Cauchy)

$$x_C = x_k - t_k g_k, \qquad \bar{t} = \arg\min_{t>0} m_{x_k}(x_k - t g_k).$$

calcul direct : $t_k = \dfrac{g_k^T g_k}{g_k^T \nabla^2 f(x_k) g_k}$

Définition (Point de Newton)

$$\begin{aligned} x_N &= x_k + d_k \\ \nabla^2 f(x_k) d_k &= -g_k \end{aligned}$$

Principe de la méthode dogleg

Sous-problème de région de confiance (PRC)

$$\min_{d} m_{\bar{x}}(\bar{x} + d)$$

$$\| d \| \leq \bar{r}.$$

Principe de la méthode

- Région de confiance très petite $\Longrightarrow$ point de Cauchy
- Région de confiance très large $\Longrightarrow$ point de Newton
- Combinaison des deux : $x^+ = \bar{x} + p_\alpha$

$$p_\alpha = \begin{cases} \alpha d_C & \text{si } 0 \leq \alpha \leq 1 \\ d_C + (\alpha - 1)(x_d - x_C) & \text{si } 1 \leq \alpha \leq 2 \\ (\eta(3 - \alpha) + \alpha - 2) d_N & \text{si } 2 \leq \alpha \leq 3 \end{cases}$$

$x_C = \bar{x} + d_C$, $x_d = \bar{x} + \eta d_N$.

Newton vs Région de confiance

Newton

- \- Coût du calcul de $\nabla^2 f(x)$
- \- Absence de convergence globale
- \+ Convergence quadratique

Région de confiance

- \- Coût du calcul de $\nabla^2 f(x)$
- \+ Convergence globale
- \+ Convergence quadratique à l'approche de x^*

Problème de moindres carrés

Problème de moindres carrés non linéaires

Modèle non linéaire : $y = y(x; a)$, paramètre $a = (a_1, \ldots, a_n)$
Mesures sur une fonction $F : \mathbb{R}^n \to \mathbb{R}^m$

$$F_i(a) = y(x_i; a) - y_i, \quad i = 1, \ldots, m$$

Erreur entre mesures et modèle :

$$E(a) = \frac{1}{2} F(a)^T F(a) = \sum_{i=1}^{m} |y(x_i; a) - y_i|^2 = \| F(a) \|_2^2$$

Problème de moindres carrés non linéaires

$$\min_{a \in \mathbb{R}^n} E(a)$$

Différentiation

$F = (F_1, \ldots, F_m)$ multi-application
Jacobienne de F

$$J_F(a) := \nabla F(a) = \left(\frac{\partial F_i}{\partial a_j}(a)\right)_{i,j}$$

Gradient de E :

$$\nabla E(a) = J_F(a)^T F(a) = \sum_{i=1}^{m} F_i(a) \nabla F_i(a)$$

Matrice Hessienne :

$$\nabla^2 E(a) = J_F(a)^T J_F(a) + \sum_{i=1}^{m} F_i(a) \nabla^2 F_i(a)$$

On suppose $F_i(a)$ «petit»

$\nabla^2 E(a) \approx J_F(a)^T J_F(a)$

Méthode de Gauss-Newton

Gauss-Newton $\equiv$ Newton avec la matrice hessienne approchée :

Algorithme de Gauss-Newton

a_0 donné

1. Résoudre $J_F(a_k)^T J_F(a_k) d_k = -J_F(a_k) F(a_k)$
2. $a_{k+1} = a_k + d_k$

On arrête dès que $\| F(a) \| < \varepsilon$

Méthode de Levenberg-Marquardt

Principe : $a_{k+1} = a_k + d_k$

- Loin de l'optimum : plus forte pente

$$d_k = -\lambda_k \nabla E(a_k), \quad \lambda_k > 0$$

- Près de l'optimum : Newton

$$J_F(a_k)^T J_F(a_k) d_k = -\nabla E(a_k)$$

Formule unique

$$\left(J_F(a_k)^T J_F(a_k) + \lambda_k \mathbb{I}_n\right) d_k = -\nabla E(a_k)$$

λ_k ajusté pour que d_k direction de descente

Méthode de Levenberg-Marquardt

Algorithme

$k = 0$ $\boldsymbol{a}_0$, $\lambda_0 = 0.001$

$k \geq 0$

1. Calculer d_k solution de
$$\left(J_F(a_k)^T J_F(a_k) + \lambda_k \mathbb{I}_n\right) d_k = -J_F(a_k)^T F(a_k)$$
2. Si $E(a_k + d_k) < E(a_k)$ (descente)
$$\lambda_{k+1} = \lambda_k/10$$
$$a_{k+1} = a_k + d_k,\ k \leftarrow k+1$$
3. Si $E(a_k + d_k) \geq E(a_k)$ (pas de descente)
$$\lambda_{k+1} = 10\lambda_k$$
$$a_{k+1} = a_k,\ k \leftarrow k+1$$

On arrête dès que $\| \nabla E(a_k) \| < \varepsilon$

Méthodes quasi-Newton

Principe de la méthode

On remplace la matrice des dérivées secondes par une approximation définie positive

Équation non linéaire dans $\mathbb{R}$

Approximation en $\bar{x}$: $m_{\bar{x}}(x) = f(\bar{x}) + \dfrac{1}{s}(f(\bar{x}+s) - f(\bar{x}))(x - \bar{x})$

Dérivée approchée : $a_s(\bar{x}) = \dfrac{1}{s}(f(\bar{x}+s) - f(\bar{x}))$

Itération : $x^+ = \bar{x} - \frac{f(\bar{x})}{a_s(\bar{x})}$

Méthode de la sécante

$s = x_k - x_{k-1}$ et $a_k = (f(x_k) - f(x_{k-1})/(x_k - x_{k-1})$

Algorithme

- $k = 0$, x_0, $a_0 = 1$
- $k \geq 0$
  1. $x_{k+1} = x_k - \dfrac{f(x_k)}{a_k}$
  2. $a_{k+1} = \dfrac{f(x_{k+1}) - f(x_k)}{x_{k+1} - x_k}$

Système non linéaire

On cherche une matrice A_k telle que : $A_k(x_{k+1} - x_k) = f(x_{k+1}) - f(x_k)$

On pose $d_{k-1} = x_k - x_{k-1}$, $y_{k-1} = f(x_{k+1}) - f(x_k)$

Équation de la sécante : $A_k d_{k-1} = y_{k-1}$

On pose : $\mathcal{S} = \{A \mid Ad_{k-1} = y_{k-1}\}$

On cherche A_k comme solution du problème

$$\min_{A \in \mathcal{S}} \| A - A_{k-1} \|_2^2 \quad \text{ou} \quad \min_{A \in \mathcal{S}} \| A - A_{k-1} \|_F^2$$

Théorème (Formule de Broyden)

$$A_k = A_{k-1} + \frac{(y_{k-1} - A_{k-1} d_{k-1}) d_{k-1}^T}{d_{k-1}^T d_{k-1}}$$

Algorithme quasi-Newton (Broyden)

Algorithme

- $k = 0,\ x_0,\ A_0 = \mathbb{I}_n$
- $k = 1, x_1 = x_0 - A_0^{-1} f(x_0)$
- $d_0 = x_1 - x_0,\ y_0 = f(x_1) - f(x_0)$
- $k \geq 1$
  1. $A_k = A_{k-1} + \dfrac{(y_{k-1} - A_{k-1} d_{k-1}) d_{k-1}^T}{d_{k-1}^T d_{k-1}}$
  2. Résoudre : $A_k d_k = -f(x_k)$
  3. $x_{k+1} = x_k + d_k$
  4. $d_k = x_{k+1} - x_k,\ y_k = f(x_{k+1}) - x(f_k)$

On arrête dès que $\| f(x_k) \| < \varepsilon$

Principe

Pour le problème

$$\min_x f(x)$$

On applique la méthode de la sécante à l'équation

$$\nabla f(x) = 0$$

$d_{k-1} := x_k - x_{k-1}$ et $y_{k-1} := \nabla f(x_k) - \nabla f(x_{k-1})$

Matrice hessienne approchée (Broyden)

$$H_k = H_{k-1} + \frac{(y_{k-1} - H_{k-1}d_{k-1})d_{k-1}^T}{d_{k-1}^T d_{k-1}}$$

Propriétés de H_k

- H_k vérifie l'équation de la sécante : $H_k d_{k-1} = y_{k-1}$
- H_k n'est pas symétrique, ni définie positive

Principe de la mise à jour BFGS

Au lieu de $H_k d_{k-1} = y_{k-1}$
On cherche le facteur de Choleski R_k tel que

$$R_k R_k^T d_{k-1} = y_{k-1}$$

i.e.

$$\begin{aligned} R_k z_{k-1} &= y_{k-1} \quad (7) \\ R_k^T d_{k-1} &= z_{k-1} \quad (8) \end{aligned}$$

Broyden (7) (R le plus proche possible de R_{k-1})

$$R_k = R_{k-1} + \frac{(y_{k-1} - R_{k-1} z_{k-1}) z_{k-1}^T}{z_{k-1}^T z_{k-1}}$$

Dans (8)

$$z_{k-1} = \alpha R_{k-1}^T d_{k-1} \quad \text{et} \quad \alpha^2 = \frac{y_{k-1}^T d_{k-1}}{d_{k-1}^T H_{k-1} d_{k-1}}$$

Formule BFGS

Théorème (BFGS)

f différentiable x_{k-1}, x_k
$d_{k-1} := x_k - x_{k-1}$ et $y_{k-1} := \nabla f(x_k) - \nabla f(x_{k-1})$ tels que $d_{k-1}^T y_{k-1} > 0$
La mise à jour BFGS est :

$$H_k = H_{k-1} + \frac{y_{k-1} y_{k-1}^T}{y_{k-1}^T d_{k-1}} - \frac{H_{k-1} d_{k-1} d_{k-1}^T H_{k-1}}{d_{k-1}^T H_{k-1} d_{k-1}}$$

H_k est définie positive si H_{k-1} est définie positive.
BFGS$\equiv$Broyden-Fletcher-Goldfarb-Shanno

Remarque (Validité de la formule)

Il faut que $y_{k-1}^T d_{k-1} > 0$

Algorithme

Algorithme BFGS

- $k = 0,\ x_0,\ H_0 = \mathbb{I}_n$
- $k \geq 0$
  1. Résoudre $H_k d_k = -\nabla f(x_k)$
  2. Calculer t_k par recherche linéaire de Wolfe ($t_0 = 1$)
  3. $x_{k+1} = x_k + t_k d_k$
  4. Mise à jour de la matrice H_k

$$H_{k+1} = H_k + \frac{y_k y_k^T}{y_k^T \bar{d}_k} - \frac{H_k \bar{d}_k \bar{d}_k^T H_k}{\bar{d}_k^T H_k \bar{d}_k}$$

$$\bar{d}_k = x_{k+1} - x_k = t_k d_k,\ y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$$

Arrêt si $\| \nabla f(x_k) \| < \varepsilon$

Remarque

$\mathrm{rang}(H_k - H_{k-1}) = 2$ *BFGS méthode (mise à jour) de rang 2.*

Approximation de l'inverse

Au lieu d'approcher $\nabla^2 f(x)$ puis résoudre le système
on peut directement approcher $[\nabla^2 f(x)]^{-1}$

Formule de Sherman-Morrison

A non singulière $(n \times n)$, U, V des matrices $(n \times p)$
$B = A + UV^T$ est inversible et

$$B^{-1} = A^{-1} - A^{-1}(I + V^T A^{-1} U)^{-1} V^T A^{-1}$$

BFGS inverse

$$S_{k+1} = \left(\mathbb{I} - \frac{d_k y_k^T}{d_k^T y_k}\right) S_k \left(\mathbb{I} - \frac{d_k y_k^T}{d_k^T y_k}\right) + \frac{d_k d_k^T}{d_k^T y_k}$$

- Si $d_k^T y_k > 0$ et S_k définie positive, alors S_{k+1} est définie positive
- La matrice S_k^* converge vers $[\nabla^2 f(x^*)]^{-1}$

Convergence

Théorème (Convergence)

Appliqué à une fonction quadratique avec une matrice hessienne A définie positive on a à l'itération k :

- $d_i^T A d_j = 0,\ 0 \leq i < j \leq k$
- $H_{k+1} A d_i = d_i,\ 0 \leq i \leq k$

BFGS$\equiv$ méthode de directions conjuguées$\equiv$ convergence en n iter. au plus

Fonction non quadratique

Comme pour la méthode du gradient conjugué, il faut réinitialiser la méthode BFGS toutes les n itérations pour assurer la convergence

Optimisation avec Contraintes

Contraintes d'égalité

$$\min f(x)$$
$$h_j(x) = 0, \quad j = 1, \ldots, q$$
$$x \in \mathbb{R}^n.$$

Définition (Contraintes régulières)

h_j sont dites régulières en a si les $\nabla h_j(a)$ sont linéairement indépendants

$$\nabla h(x) = \begin{bmatrix} \nabla h_1(x)^T \\ \vdots \\ \nabla h_q(x)^T \end{bmatrix}$$

x régulier :

$$C_x = \{d \in \mathbb{R}^n \mid \nabla h(x) d = 0\} = \left\{d \in \mathbb{R}^n \mid \nabla h_j(x)^T d = 0, \ \forall j\right\}$$

Conditions d'optimalité

Théorème (Conditions (nécessaires) d'optimalité du 1er ordre)

Si x^* est un minimum local et point régulier pour les h_j, alors $\exists \lambda^* \in \mathbb{R}^q$:

$$\nabla f(x^*) + \nabla h(x^*)\lambda^* = \nabla f(x^*) + \sum_{j=1}^{q} \lambda_j^* \nabla h_j(x^*)^T = 0$$

λ_j *multiplicateur de Lagrange*

On pose : $L = \nabla^2 f(x^*) + \sum_{j=1}^{q} \lambda_j^* \nabla^2 h_j(x^*)$

Théorème (Conditions (nécessaires) d'optimalité du 2ème ordre)

Si x^* est un minimum local et point régulier pour les h_j, alors L est semi-définie positive sur C_{x^*} :

$$d^T L d \geq 0, \quad \forall d \in C_{x^*}$$

Equations d'optimalité

Si x^* est un minimul local :

$$\begin{aligned} \nabla f(x^*) + \nabla h(x^*)\lambda^* &= 0 \\ h_j(x^*) &= 0,\ j = 1, \ldots, q \end{aligned}$$

Système de $n + q$ équations à $n + q$ inconnues

Fonction de Lagrange $\mathscr{L} : \mathbb{R}^n \times \mathbb{R}^q \to \mathbb{R}$

$$\mathscr{L}(x, \lambda) = f(x) + h(x)^T \lambda$$

Equations d'optimalité $\Longleftrightarrow \nabla \mathscr{L}(x^*, \lambda^*) = 0$

$\mathscr{L}$ fonction de Lagrange ou Lagrangien

Un exemple

$$\min f(x) = x_1^2 + x_2^2$$
$$x_1 + x_2 = 1$$

$$\nabla f(x) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} \quad \nabla h(x) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Le système

$$\begin{array}{rcl} 2x_1 + \lambda & = & 0 \\ 2x_2 + \lambda & = & 0 \\ x_1 + x_2 & = & 1 \end{array}$$

La solution $x^* = (1/2, 1/2)$, $\lambda^* = -1$

Contrainte d'inégalité

$$(PI) \qquad \begin{array}{l} \min f(x) \\ g_i(x) \leq 0, \quad i = 1, \ldots, p; \; x \in \mathbb{R}^n. \end{array}$$

Définition (Contrainte active ou saturée)

$I(x) = \{i \in \{1, 2, \ldots, p\} \mid g_i(x) = 0\}$

Proposition

Si x^* solution optimale de (PI), alors x^* est solution optimale de

$$\begin{array}{l} \min f(x) \\ g_i(x) = 0, \quad i \in I(x^*) \end{array}$$

Définition (Cône de directions admissibles)

$$C_{x^*} = \left\{ d \in \mathbb{R}^n \mid \nabla g_i(x^*)^T d \leq 0 \right\}$$

Conditions d'optimalité

Théorème (Karush-Kuhn-Tucker (KKT))

Soit x^* *tel que* $\nabla g_i(x^*)$, $i \in I(x^*)$, *linéairement indépendants*
x^* *minimum local pour* (PI), *alors* $\exists \mu^* \in \mathbb{R}^p$, $\mu_i^* \geq 0$ *:*

$$\begin{aligned} \nabla f(x^*) + \sum_{i \in I(x^*)} \mu_i^* \nabla g_i(x^*) &= 0 \\ \mu_i^* g_i(x^*) &= 0 \end{aligned}$$

μ_i^* *multiplicateur de Kuhn-Tucker*

Équation de complémentarité

$$\mu_i^* g_i(x^*) = 0 \Longleftrightarrow g_i(x^*) \leq 0 \text{ ou } \mu_i^* \geq 0.$$

Fonction de Lagrange (Lagrangien)

Fonction de Lagrange $\mathscr{L} : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$

$$\mathscr{L}(x,\mu) = f(x) + \mu^T g(x)$$

Gradient du Lagrangien

$$\nabla \mathscr{L}(x,\lambda) = \begin{pmatrix} \nabla f(x) + \sum \mu_i \nabla g_i(x) \\ g_i(x) \end{pmatrix}$$

$\mu_i = 0$ si $g_i(x^*) < 0 \Longrightarrow$ contraintes non actives ne comptent pas

Dérivées secondes :

$$L = \nabla^2 f(x^*) + \sum_{i \in I(x^*)} \mu_i^* \nabla g_i(x^*)$$

Théorème (Condition d'optimalité du 2nd ordre)

x^* minimum local pour (PI), alors $d^T L d \geq 0$, $\forall d \in C_{x^*}$.

Problème primal - problème dual

Contraintes :
$X = \{x \in \mathbb{R}^n \mid g_i(x) \leq 0,\ i = 1, \ldots, p,\ h_j(x) = 0, j = 1 \ldots, q\}$

Problème primal

$$(P) \qquad m = \inf_{x \in X} f(x)$$

Fonction de Lagrange (Lagrangien) : $\mathscr{L} : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$

$$\mathscr{L}(x, \mu, \lambda) = f(x) + \mu^T g(x) + \lambda^T h(x)$$

Problème dual

$$(P') \qquad \alpha = \sup_{\lambda, \mu \geq 0} \mathscr{L}(x, \mu, \lambda)$$

Point-selle

Définition (Point-selle)

(x^*, μ^*, λ^*) *est un point selle (ou point col) si*

1. x^* *solution optimale de* (P)
2. (μ^*, λ^*) *solution optimale de* (P')
3. $m = \alpha = L(x^*, \mu^*, \lambda^*)$

Si (x^*, μ^*, λ^*) point-selle :

$$\mathscr{L}(x^*, \mu, \lambda) \leq \mathscr{L}(x^*, \mu^*, \lambda^*) \leq \mathscr{L}(x, \mu^*, \lambda^*)$$

ou

$$\sup_{\mu \geq 0, \lambda} \inf_{x \in X} \mathscr{L}(x, \mu, \lambda) = \inf_{x \in X} \sup_{\mu \geq 0, \lambda} \mathscr{L}(x, \mu, \lambda)$$

En général

$$\sup_{\mu \geq 0, \lambda} \inf_{x \in X} \mathscr{L}(x, \mu, \lambda) \leq \inf_{x \in X} \sup_{\mu \geq 0, \lambda} \mathscr{L}(x, \mu, \lambda)$$

Existence d'un point-selle

Théorème (Point-selle)

- *f convexe et $\mathscr{C}^1$*
- *g_i convexes et $\mathscr{C}^1$, h_j affines*
- *$\exists x_0$ tel que $g_i(x_0) < 0$ et $h_j(x_0) = 0$, $\forall i, j$*

Si (P) admet une solution optimale alors un point selle de $\mathscr{L}$ existe.

Méthodes de résolution

- Méthodes basées sur x : Méthodes primales
- Méthodes basées sur (μ, λ) : Méthodes duales
- Méthodes basées sur x et (μ, λ) : Méthodes primales-duales

Méthode du gradient projété

Principe de la méthode

$$\begin{aligned} &\min f(x) \\ &a_i^T x \leq b_i, \quad i \in I_1 \\ &a_i^T x = b_i, \quad i \in I_2. \end{aligned}$$

Contraintes actives en x : $I_0 = \{i \in I_1 \mid a_i^T x = b_i\} \cup I_2$

Matrice des contraintes actives en x :

$$A_0 = [a_i^T \mid i \in I_0], \quad p \times n, \quad \mathrm{rg} A_0 = p$$

Cône des directions admissibles en x :

$$S_0 := C_x = \{d \in \mathbb{R}^n \mid A_0 d = 0\}$$

Idée principale

Projéter $-\nabla f(x)$ sur S_0

Projection

Problème de projection sur $S_0 = \ker A_0$

$$\min \nabla f(x)^T d$$
$$A_0 d = 0, \quad \| d \| = 1$$

Projection sur $\ker A_0 \equiv$ Projection sur $(\mathrm{Im} A_0^T)^\perp$

Projection de $\nabla f(x)$ sur $\mathrm{Im} A_0^T$: $y = A_0^T (A_0 A_0^T)^{-1} A_0 \nabla f(x)$

Projection de $\nabla f(x)$ sur $(\mathrm{Im} A_0^T)^\perp$:

$$p = \nabla f(x) - y = (\mathbb{I} - A_0^T (A_0 A_0^T)^{-1} A_0) \nabla f(x)$$

Matrice de projection : $P_0 = (\mathbb{I} - A_0^T (A_0 A_0^T)^{-1} A_0)$

Gradient projeté

$$p = -P_0 \nabla f(x) \qquad d = \frac{p}{\| p \|}$$

Optimalité

Théorème

d est une direction de descente pour f

$$\nabla f(x)^T p = - \| \nabla f(x) \|^2 - (A_0 \nabla f(x))^T (A_0 A_0^T)^{-1} A_0 \nabla f(x) < 0$$

$$S_0^\perp = \mathrm{Im} A_0^T \implies \mathbb{R}^n = S_0 \oplus S_0^\perp \implies -\nabla f(x) = p + A_0^T u, \; p \in S_0$$

Si $p = -P_0 \nabla f(x) = 0$

$$\nabla f(x) + A_0^T u = 0$$

Moindres carrés : $u = -(A_0 A_0^T)^{-1} A_0 \nabla f(x)$

Optimalité

- Si $u \geq 0$, On a les conditions KKT
- S'il existe $u_i < 0$, il faut supprimer certaines contraintes actives

Algorithme du gradient projeté

$k = 0$ Initialisation : x^0 donné

$k \geq 0$

1. Calculer $I_0(x^k)$
2. Former A_0
3. $P_0 = \mathbb{I} - A_0^T (A_0 A_0^T)^{-1} A_0$
4. $p^k = -P_0 \nabla f(x^k)$, $d^k = p^k / \parallel p^k \parallel$
5. $u^k = -(A_0 A_0^T)^{-1} A_0 \nabla f(x^k)$, $\quad u_i^k = \min_j u_j^k$
6. Si $p^k = 0$ et $u_i^k \geq 0$ STOP
7. Si $p^k = 0$ et $u_i^k < 0$ supprimer la ligne i de A_0

   Retourner en 2.
8. Calculer $\bar{t} = \max_{t>0} \{x^k + td^k \in X\}$
9. $t_k = \arg \min_{t \in (0, \bar{t})} f(x^k + td^k)$
10. $x^{k+1} = x^k + t_k d^k$

Programmation Quadratique I : Contraintes d'égalité

Problème quadratique

$$\min q(x) = \frac{1}{2}x^T Qx + x^T c$$
$$Ax = b$$
$$x \in \mathbb{R}^n.$$

- A matrice $m \times n$, $m \leq n$, $\mathrm{rang} A = m$
- Q semi-définie positive (q convexe)

Lagrangien : $\mathscr{L}(x, \lambda) = q(x) + \lambda^T (Ax - b)$

Conditions d'optimalité du 1er ordre (KKT)

$$\begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \lambda^* \end{bmatrix} = \begin{bmatrix} -c \\ b \end{bmatrix}$$

Existence et unicité

z_i, $i = 1, \dots, n-m$ base de $\ker A$

$Z = [z_1 \ \cdots z_{n-m}]$ $(AZ = 0)$

Lemme

Si $Z^T QZ$ définie positive, alors

$$K = \begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix}$$

est non signulière et (x^*, λ^*) est solution unique de (KKT).

Théorème

Si A de rang plein et $Z^T QZ$ définie positive, alors
(x^*, λ^*) solution $(KKT) \Longrightarrow x^*$ unique solution de (P)

Factorisation

Les factorisations LU et LDL^T ne sont pas adaptées à la matrice K

Factorisation symétrique indéfinie (fonction `ldl` dans matlab)

$$PKP^T = LBL^T$$

- P matrice de permutation
- B matrice bloc-diagonale (1×1 ou 2×2)

Résolution du système factorisé :

1. $Lz = P \begin{bmatrix} -c \\ b \end{bmatrix}$
2. $B\hat{z} = z, \;\; L^T \bar{z} = \hat{z}$
3. $\begin{bmatrix} x^* \\ \lambda^* \end{bmatrix} = P\bar{z}$

Complément de Schur

Elimination de Gauss sur (KKT)

$$\begin{bmatrix} Q & A^T \\ 0 & -AQ^{-1}A^T \end{bmatrix} \begin{bmatrix} x^* \\ \lambda^* \end{bmatrix} = \begin{bmatrix} -c \\ b - AQ^{-1}c \end{bmatrix}$$

La matrice $S = AQ^{-1}A^T$ est appelée complément de Schur de Q.

Résolution de (KKT) à l'aide du complément de Schur :

1. Résoudre $S\lambda^* = AQ^{-1}c - b$
2. Résoudre : $Qx^* = A^T\lambda^* - c$

Remarque

Pour éviter l'inversion explicite de Q, on peut utiliser le gradient conjugué pour le système en λ^*

Gradient conjugué projeté

Remarque

PCG n'est pas adapté pour résoudre K
Utiliser les méthodes de type krylov : GMRES, LSQR, etc.

Méthode du gradient conjugé projété

- Initialisation : $k = 0$, x_0 tel que $Ax_0 = b$, $r_0 = Qx_0 + c$
  $AA^T v_0 = Ar_0$
  $g_0 = r_0 - A^T v_0$, $d_0 = -g_0$
- $k \geq 0$
  1. $t_k = r_k^T g_k / d_k^T Q d_k$
  2. $x_{k+1} = x_k + t_k d_k$, $r_{k+1} = Qx_{k+1} + c$
  3. $AA^T v_{k+1} = Ar_{k+1}$
  4. $g_{k+1} = r_{k+1} - A^T v_{k+1}$
  5. $\beta_k = r_{k+1}^T g_{k+1} / r_k^T g_k$
  6. $d_{k+1} = -g_{k+1} + \beta_k d_k$

Contraintes d'égalité et d'inégalité : (Primal) Active Set Méthod

Problème quadratique

$$\min q(x) = \frac{1}{2}x^T Qx + c^T x$$
$$a_i^T x = b_i, \quad i \in \mathcal{E}$$
$$a_i^T x \leq b_i, \quad i \in \mathcal{I}.$$

Q sémi-définie positive.
Contraintes actives en x^* : $\mathcal{A}^* = \{i \in \mathcal{E} \cup \mathcal{I} \mid a_i^T x = b_i\}$

Conditions d'optimalité (KKT)

$$Qx^* + \sum_{i \in \mathcal{A}^*} \lambda^* a_i = 0$$
$$a_i^T x^* = b_i, \quad i \in \mathcal{A}^*$$
$$a_i^T x^* \leq b_i, \quad i \in \mathcal{I} \setminus \mathcal{A}^*$$
$$\lambda_i^* \geq 0, \quad i \in \mathcal{I} \cap \mathcal{A}^*$$

Principe

x_k point courant et W_k contraintes actives en x_k

$x_k \rightarrow x = x_k + p$:

$$q(x) = q(x_k + p) = \frac{1}{2}p^T Qp + g_k^T p + c_k$$

Sous-problème quadratique

$$\min_p \frac{1}{2}p^T Qp + g_k^T p$$
$$a_i^T p = 0, \quad i \in W_k$$

Solution p_k :

- $a_i^T(x_k + \alpha p_k) = a_i^T x_k = b_i$, $\forall i \in W_k$, $\forall \alpha \geq 0$
- Si $a_i^T p_k \leq 0$, $a_i^T(x_k + \alpha p_k) \leq a_i^T x_k \leq b_i$, $\forall i \notin W_k$, $\forall \alpha \geq 0$
- Si $a_i^T p_k > 0$

$$\alpha \leq \rho_k = \frac{b_k - a_i^T x_k}{a_i^T p_k}$$

Pas de déplacement

$$\rho_k = \min_{i \notin W_k,\ a_i^T p_k > 0} \frac{b_k - a_i^T x_k}{a_i^T p_k}$$

On prend comme pas optimal :

$$\alpha_k = \min\{1, \rho_k\}$$

- Si $\alpha_k = 1$, aucune contrainte ne bloque le déplacement
  $W_{k+1} = W_k$
- Si $\alpha_k < 1$, il y a au moins une contrainte qui bloque
  Rajouter la contrainte qui bloque à W_k pour former W_{k+1}

Optimalité

Si $p_k = 0$, (KKT) pour le sous-problème quadratique :

$$\sum_{i \in W_k} \hat{\lambda}_i a_i = -g_k = -(Qx_k + c)$$

- Si $\hat{\lambda}_i \geq 0$, $i \in W_k \cap \mathcal{I}$, les conditions d'optimalité sont vérifiées
  x_k solution optimale pour le problème (QP)
- Si $\exists i \in W_k \cap \mathcal{I}$ tel que $\hat{\lambda}_i < 0$
  supprimer la contrainte i de l'ensemble W_k.

Active Set Method

$k = 0$ x_0 solution réalisable

W_0 contraintes actives en x_0

$k \geq 0$

1. Résoudre le sous-problème quadratique (p_k la solution)
2. Si $p_k = 0$, calculer $\hat{\lambda}_i$

   Si $\hat{\lambda}_i \geq 0$, pour tout $i \in W_k \cap \mathcal{I}$, STOP $x^* = x_k$

   Sinon $j = \arg\min_{i \in W_k \cap \mathcal{I}} \hat{\lambda}_i$

   $$W_{k+1} = W_k \setminus \{j\} \text{ ; retourner en 1.}$$
3. Si $p_k \neq 0$, Calculer α_k

   $x_{k+1} = x_k + \alpha_k p_k$; Si $\alpha_k < 1$,
   $W_{k+1} = W_k \cup \{\text{nouvelles contraintes actives}\}$

Remarques

- La condition «x_0 solution réalisable» n'est pas triviale
  A partir d'un $\widetilde{x}$, on détermine x_0 par PL ou pénalisation.
- La stratégie de faire entrer ou sortir de W une contrainte à la fois peut-être pénalisante :
  Si en x_0, $W_0 = \emptyset$ et en x^*, $\operatorname{card} W^* = 200$, il faut au moins 200 itérations pour atteindre l'optimum.

Contraintes d'égalité et d'inégalité : Primal-Dual Active Set Méthod

Principe de la méthode

Problème modèle

$$\min q(x) = \frac{1}{2} x^\top Q x - c^\top x$$
$$x \leq b.$$

Pour le problème général

$$\min q(x) = \frac{1}{2} x^\top Q x - c^\top x$$
$$Ax \leq b.$$

Remplacer par

$$\min q(x, z) = \frac{1}{2} x^\top Q x + \frac{\varepsilon}{2} z^\top z - c^\top x$$
$$Ax - z = b$$
$$-z \leq 0.$$

Primal-Active Set Method

1. Construire les ensembles

$$\begin{aligned} \mathcal{A}_k &= \left\{ i : \; x_i^k > b_i \text{ ou } x_i^k = b_i \text{ et } \mu_i^k \geq 0 \right\} \\ \mathcal{I}_k &= \left\{ i : \; x_i^k < b_i \text{ ou } x_i^k = b_i \text{ et } \mu_i^k < 0 \right\} \end{aligned}$$

2. Poser $x_i^{k+1} = b_i$ pour $i \in \mathcal{A}_k$
3. Résoudre

$$\sum_{j \in \mathcal{I}_k} Q_{ij} x_j^{k+1} = c_i - \sum_{j \in \mathcal{A}_k} Q_{ij} b_j, \quad i \in \mathcal{I}_k$$

4. Calculer les multiplicateurs des contraintes actives

$$\mu_i^{k+1} = b_i - \sum_j Q_{ij} x_j^{k+1}, \quad \forall i \in \mathcal{A}_k$$

5. Si $x^{k+1} \leq b$ et $\mu_i^{k+1} \geq 0$ pour tout $i \in \mathcal{A}_k$, STOP

Commentaires

- + Convergence rapide : 2 ou 3 itérations si $\mathcal{A}^* = \emptyset$
- + Mise en oeuvre facile
- - Matrice à restructurer à chaque itération
- - Augmentation du nombre de variables dans le cas général

Algorithme primale-dual 1 : Contraintes d'égalité

Principe

$$\min f(x)$$
$$h_i(x) = 0, i = 1, \ldots, m$$
$$x \in R^n, \ m < n.$$

$A = A(x) = \nabla h(x)$ jacobienne $(m \times n)$ de h

Conditions d'optimalité

$(x^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^m$

$$\nabla f(x^*) + A_*^T \lambda^* = 0$$
$$h(x^*) = 0$$

Principe

Résoudre les équations d'optimalité par la méthode de Newton

Approximations

Notations : $f_k := f(x_k)$, $h_k := h(x_k)$, $A_k := A(x_k)$

$$\nabla \ell_k := \nabla_x \mathscr{L}(x_k, \lambda_k), \;\; L_k := \nabla_x^2 \mathscr{L}(x_k, \lambda_k)$$

Dérivation

$$\mathscr{L}(x_k, \lambda_k) = f_k + \lambda_k^T h_k$$

$$\nabla_x \mathscr{L}(x_k, \lambda_k) = \nabla f(x_k) + A_k \lambda_k, \quad \nabla_{x\lambda}^2 \mathscr{L}(x_k, \lambda_k) = A_k$$

$$\nabla_\lambda \mathscr{L}(x_k, \lambda_k) = h_k, \quad \nabla_{\lambda x}^2 \mathscr{L}(x_k, \lambda_k) = A_k^T$$

Coniditions d'optimalité : $\nabla \mathscr{L}(x^*, \lambda^*) = 0$

Acroissement de Newton : $(x_k, \lambda_k) \to (x_k + d_k, \lambda_k + \mu_k)$

Approximation

$$\nabla \mathscr{L}(x_k + d_k, \lambda_k + \mu_k) \approx \nabla \mathscr{L}(x_k, \lambda_k) + \nabla^2 \mathscr{L}(x_k, \lambda_k) \begin{pmatrix} d_k \\ \mu_k \end{pmatrix}$$

Problème linéarisé

Système linéarisé

$$\begin{pmatrix} L_k & A_k^T \\ A_k & 0 \end{pmatrix} \begin{pmatrix} d_k \\ \mu_k \end{pmatrix} = - \begin{pmatrix} \nabla \ell_k \\ h_k \end{pmatrix} \qquad (SL)$$

Problème tangent

(SL) C.N. du problème quadratique

$$\min \nabla \ell_k^T d + \frac{1}{2} d^T L_k d$$
$$h_k + A_k d = 0$$

Comme $\nabla \ell_k$ linéaire par rapport à λ_k on préfère :

Problème tangent simplifié

$$\min \nabla f_k^T d + \frac{1}{2} d^T L_k d$$
$$h_k + A_k d = 0$$

Algorithme

Algorithme de Newton

$k = 0$ Initialisation (x_0, λ_0) donné

$k \geq 0$ h_k, ∇f_k, A_k, L_k disponibles

1. Déterminer $(d_k, \bar{\lambda}_k)$ point-selle du problème
$$\min \nabla f(x_k)^T d + \frac{1}{2} d^T L_k d$$
$$h_k + A_k d = 0$$
2. $x_{k+1} = x_k + d_k$, $\lambda_{k+1} = \bar{\lambda}_k$

Théorème (Convergence)

Si f et h sont $\mathscr{C}^2$ dans le voisinage de x^* (point stationnaire) de multiplicateur λ^*
Alors $\exists V$ vois de (x^*, λ^*) tel que si $(x_0, \lambda_0) \in V$ l'algorithme de Newton converge.

Algorithme primale-dual 2 : Contraintes d'inégalité

Principe

$$
\begin{aligned}
&\min f(x)\\
&c_i(x) = 0, \quad i \in \mathcal{E}\\
&c_i(x) \le 0, \quad i \in \mathcal{I}.
\end{aligned}
$$

Conditions d'optimalité (KKT)

$$
\begin{aligned}
&\nabla f(x^*) + \sum_{\mathcal{E}\cup\mathcal{I}^*} \lambda_i^* \nabla c_i(x^*) = 0\\
&c_i(x^*) = 0, \quad i \in \mathcal{E}\\
&\lambda_i^* c_i(x^*) = 0, \quad i \in \mathcal{I}.
\end{aligned}
$$

Principe

Linéariser les équations (KKT) en passant de $(x_k, \lambda_k) \to (x_k + d_k, \lambda_k + \mu_k)$

Problème linéarisé

Equations (KKT) linéarisées :

$$L_k d_k + \sum_{i\in\mathcal{E}} \lambda_i^k(\nabla c_i(x_k) + c_i(x_k)) + \sum_{i\in\mathcal{I}} \lambda_i^k(\nabla c_i(x_k) + c_i(x_k)) = -\nabla f(x_k)$$

$$\nabla c_i(x_k)^T d_k + c_i(x_k) = 0, \quad i \in \mathcal{E}$$

$$\lambda_i^k(\nabla c_i(x_k)^T d_k + c_i(x_k)) = 0, \quad i \in \mathcal{I}$$

Problème tangent simplifié

$$\min \frac{1}{2} d^T L_k d + \nabla f(x_k) d$$

$$\nabla c_i(x_k)^T d + c_i(x_k) = 0, \quad i \in \mathcal{E}$$

$$\nabla c_i(x_k)^T d + c_i(x_k) \leq 0, \quad i \in \mathcal{I}.$$

Algorithme (PQS)

$k = 0$ Initialisation (x_0, λ_0) donné

$k \geq 0$ h_k, ∇f_k, $c_i(x_k)$, $\nabla c_i(x_k)$, L_k disponibles

1. Déterminer $(d_k, \bar{\lambda}_k)$ point-selle du problème

$$\min \nabla f(x_k)^T d + \frac{1}{2} d^T L_k d$$
$$\nabla c_i(x_k)^T d + c_i(x_k) = 0, \quad i \in \mathcal{E}$$
$$\nabla c_i(x_k)^T d + c_i(x_k) \leq 0, \quad i \in \mathcal{I}.$$

2. $x_{k+1} = x_k + d_k$, $\lambda_{k+1} = \bar{\lambda}_k$

Théorème (Convergence)

Si f et c sont $\mathscr{C}^2$ dans le voisinage de x^* (point stationnaire) de multiplicateur λ^*. On suppose les CSO du 2nd ordre vérifiées. Alors $\exists V$ vois de (x^*, λ^*) tel que si $(x_0, \lambda_0) \in V$ l'algorithme de PQS converge.