Gradient conjugué – fonction quelconque
Algorithme de Fletcher-Reeves

- Départ : x_0 et $d_0 = - \partial f / \partial x_0$

  $x_{k+1} = x_k + \lambda_k d_k$

  avec λ_k minimisant $g(\lambda) = f(x_k + \lambda_k d_k)$

  (défini par minimisation suivant d_k)

  puis $d_{k+1} = - \partial f / \partial x_{k+1} + \beta_k d_k$

  avec $\beta_k = \| \partial f / \partial x_{k+1} \|^2 / \| \partial f / \partial x_k \|^2$ tant que (cdt d'arrêt non vérifiée)

  Variante de Polak – Ribière :

$$\beta_k = (\partial f / \partial x_{k+1})^T (\partial f / \partial x_{k+1} - - \partial f / \partial x_k) / \| \partial f / \partial x_k \|^2$$

Méthodes du gradient conjugué: Conclusions

•Polak-Ribière équivalent au Fletcher-Reeves pour les fonctions quadratiques, mais donne des résultats différents pour des fonctions quelconques

•Nécessite le stockage de très peu d'information :

d_k, $\partial f / \partial x_k$, $\partial f / \partial x_{k+1}$ (3 vecteurs de dim. «n»)

• Vitesse de convergence très supérieure à celle des algorithmes du gardient classiques

=> Méthode à considérer pour des problèmes de grande taille (n>100)

Méthodes de Newton

- Idée : utiliser l'algorithme de Newton pour résoudre le système $\partial f / \partial x = 0$
- L'algorithme correspondant est donc

$$x_{k+1} = x_k - (\partial^2 f / \partial x^2_k)^{-1} \partial f / \partial x_k$$

- Dans le cas de fonctions quadratiques, l'algorithme converge en une seule itération
- Demeure le pb de la convergence globale si f(x) est quelconque => on apporte les modifications suivantes

$$x_{k+1} = x_k - \lambda_k (\partial^2 f / \partial x^2_k)^{-1} \partial f / \partial x_k$$

avec λ_k minimisant f(x) dans la direction $(\partial^2 f / \partial x^2_k)^{-1} \partial f / \partial x_k$

- Il se peut que localement $(\partial^2 f / \partial x^2_k)$ ne soit pas définie positive, alors

$$M_k = \mu_k I + \partial^2 f / \partial x^2_k \text{ avec } \mu_k \text{ scalaire positif}$$

Méthodes de quasi-Newton

•Idée: Il s'agit d'une généralisation, permettant de ne pas avoir à calculer le Hessien

L'algorithme correspondant est donc

$$x_{k+1} = x_k - \lambda_k H_k \, \partial f / \partial x_k$$

avec λ_k minimisant f(x) dans la direction $H_k \, \partial f / \partial x_k$

• On veut faire *aussi bien* que Fletcher -Reevs =>

H_k sera une matrice dont on impose la convergence en « n » itérations vers l'inverse du Hessien dans le cas où f(x) est quadratique

• A l'itération « n » on se retrouve dans la configuration de la méthode de Newton et donc on converge en un coup

• D'une manière plus générale, H devra être une approximation de l'inverse du hessien

•Différentes formules peuvent être adaptées

$$H_{k+1} = H_k + \Delta_k$$

• Δ_k est une matrice de rang 1 ou 2

•Dans le cas d'une fonction quadratique, on rappelle

$$A^{-1}(\partial f/\partial x_k - \partial f/\partial x_{k-1}) = A^{-1}(Ax_k + b - Ax_{k-1} - b) = x_k - x_{k-1}$$

•On souhaite imposer la relation

$H_{k+1}(\partial f/\partial x_k - \partial f/\partial x_{k-1}) = x_k - x_{k-1}$ soit $H_{k+1}\gamma_k = \delta_k$

•D'autre part, pour respecter la symétrie on définit

$\Delta_k = \alpha_k u_k u_k^T$ correction de rang 1

•En posant

$\delta_k = x_k - x_{k-1}$ et $\gamma_k = \partial f/\partial x_k - \partial f/\partial x_{k-1}$

=> $\Delta_k = \alpha_k u_k u_k^T = (\delta_k - H_k\gamma_k)(\delta_k - H_k\gamma_k)^T / \gamma_k^T(\delta_k - H_k\gamma_k\ )$

avec $\delta_k = x_k - x_{k-1}$ et $\gamma_k = \partial f / \partial x_k - \partial f / \partial x_{k-1}$

•On peut démontrer que $H_{n+1} = A^{-1}$

(convergence de la méthode de quasi Newton dans le cas quadratique en "n" itérations)

Algorithme de Fletcher-Powell

- Correction de rang 2

$H_{k+1} = H_k+\delta_k\delta_k{}^T/\delta_k{}^T\gamma_k-H_k\gamma_k\gamma_k{}^T H_k/\gamma_k{}^T H_k\gamma_k$

Algorithme:

$$x_{k+1}=x_k + \lambda_\kappa d_k$$

avec λ_k minimisant f(x) dans la direction $d_k=-H_k\, \partial f/ \partial x_k$

$H_{k+1} = H_k+\delta_k\delta_k{}^T/\delta_k{}^T\gamma_k-H_k\gamma_k\gamma_k{}^T H_k/\gamma_k{}^T H_k\gamma_k$

et $\delta_k = x_{k+1}-x_k$ et $\gamma_k = \partial f/ \partial x_{k+1}- \partial f/ \partial x_k$

Algorithme de Broyden,Fletcher, Goldfrab, Shanno (BFGS)

- On sait que $H_{k+1}\gamma_\kappa = \delta_\kappa$
- On intervertit les rôles de γ_κ et δ_κ
- $G_{k+1} = G_k+\gamma_k\gamma_k^T/\gamma_k^T\delta_k-G_k\delta_k\delta_k^TG_k/\delta_k^TG_k\delta_k$
- On obtient $G_{k+1}\delta_k = \gamma_\kappa$ à comparer avec

$$\gamma_\kappa = H_{k+1}^{-1}\delta_\kappa$$

Donc G_{k+1} approxime le hessien directement

La récurrence sur G_{k+1} s'inverse donc

$$(G_{k+1})^{-1}=(G_k)^{-1}+[1+\gamma_k^T(G_k)^{-1}\gamma_k/\delta_k^T\gamma_k]*\delta_k\delta_k^T/\delta_k^T\gamma_k-$$

$$- (\delta_k\gamma_k^T(G_k)^{-1}+(G_k)^{-1}\gamma_k\delta_k^T)/\delta_k^T\gamma_k$$

Algorithme de Broyden,Fletcher, Goldfrab, Shanno (BFGS)

Et finalement

$$H_{k+1}=H_k+[1+\gamma_k{}^T H_k\gamma_k/\delta_k{}^T\gamma_k]*\delta_k\delta_k{}^T/\delta_k{}^T\gamma_k-$$
$$-(\delta_k\gamma_k{}^T H_k+H_k\gamma_k\delta_k{}^T)/\delta_k{}^T\gamma_k$$

- Moins sensible que DFP (David Fletcher Powell) aux imprécisions de la procédure de recherche unidimensionnelle => permet l'utilisation de méthodes unidimensionnelles « économiques » qui nécessitent un petit nombre d'évaluations de la fonction

Grad. Conjugué : Fletcher-Reeves
Quasi-Newton : BFGS

Que choisit-on?

- BFGS est beaucoup plus coûteuse en calcul

(construction matricielle à chaque itération)

N.B. On n'inverse pas la matrice, on récure directement sur H^{-1}

- Si la valeur du vecteur qui minimise la fonction n'est pas importante => Fletcher-Reeves
- Si on veut avoir une confiance en l'estimation => BFGS (matrice des dérivées secondes)

Comparaison des différentes méthodes. Résultats sur la convergence.

- Fait au tableau

- Conclusion:
  - Avantage des méthodes de quasi-Newton sur les méthodes du gradient conjugué en ce qui concerne le nombre de calculs de la fonction; par contre, présence de produits de matrices (calculs en n^3)

Minimisation multivariable avec contraintes. Aspects théoriques.

I. Généralités, propriétés caractéristiques des optimums constraints

I.1. Position du problème – définition

$$\begin{cases} \text{Minimiser } f(x) \\ x \in \mathcal{D} \text{ (domaine)} \end{cases} \qquad \begin{cases} f(x) \text{ scalaire} \\ \underline{x} \text{ vecteur en } \Re^n,\ \mathcal{D} \subset \Re^n \end{cases}$$

- **Contraintes (définissant $\mathcal{D}$)**

Inégalité : $g_i(x) \leq 0 \quad i=1 \ldots m$ $\begin{cases} \text{Saturées en } x_0, \text{ si } g_i(x_0) = 0 \text{ (x0 est sur les bords du domaine } \mathcal{D}) \\ \text{Non saturées en x0, si } g_i(x_0) < 0 \end{cases}$

Egalité : $h_j(x) = 0 \quad j=1 \ldots p \leq n-1$

- **Minimum absolu x***

Défini par $\forall\, x \in \mathcal{D}$, $f(x) \geq f(x^*)$ strict si l'égalité est stricte.

- **Minimum relatif x***

$\exists$ un voisinage $\mathcal{U}(x^*)$ tel que $\forall\ x \in \mathcal{D} \cap \mathcal{U}(x^*)$, $f(x) \geq f(x^*)$

- **Point admissible**

Direction admissible en x_0

Soit x_0 admissible et une direction d
Soit $x(\alpha) = x_0 + \alpha d$, avec $\alpha \geq 0$

d est admissible $\Leftrightarrow$ $[\exists\ \alpha^*>0,\ \forall\ \alpha < \alpha^*,\ x(\alpha) \in \mathcal{D}]$

Remarque: Si x_0 est strictement intérieur à $\mathcal{D}$, toutes les directions sont admissibles.

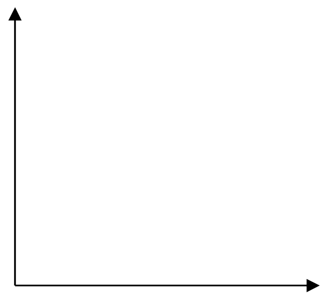


Fig.1 Direction admissible

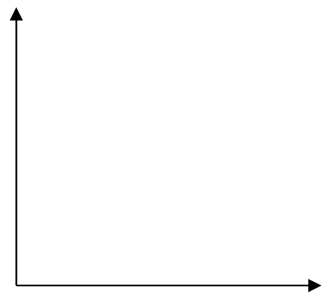


Fig.2 Point admissible régulier

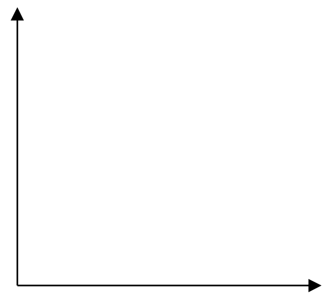


Fig.3 Point admissible non régulier

- Point admissible régulier (qualification des contraintes)

x_0 est admissible régulier si les gradients des contraintes inégalité sont saturées en x_0 et ceux des contraintes égalité sont indépendants
Le gradient des contraintes est dirigé vers l'extérieur du domaine (vers les g>0)

•Directions limite (ou tangente) en x_0 admissible

$\zeta(x_0)$=ensemble des directions d telles que $[\partial h_j / \partial x_0]^T.d=0$ (pour toutes les contraintes égalité)

$[\partial d_j / \partial x_0]^T.d=0$ (pour toutes les contraintes inégalité saturées en x_0)

Fig.4 Direction tangente

But recherché:

La recherche d'un mini absolu est difficile: L'algotrithme est « aveugle »

<u>D'où recherche d'un minimum local.</u>

Hypthèses où le mini local est un mini absolu (convexité, unimodalité)

Conditions nécessaires du 1er ordre

CN1 (sans hypothèse sur la différentialité des contraintes)

Soit f(x) de classe C_1

<u>x* mini local => $\exists$ d admissible en x*, $f_{x^*}^T.d \geq 0$ où f_{x^*} est le gradient de f</u>

Interprétation géométrique:

Les déplacements dans une direction admissible ne peuvent pas faire décroître la fonction (f(x) *dans le sens des niveaux croissants*)

Conditions nécessaires du 1er ordre - suite

On peut chercher des propriétés en précisant la façon dont est défini le domaine (Kuhn & Tucker)…

Le plus évident étant $f_{x^*}=-\mu_i\, g_{ix^*}$ où g_{ix^*} est le gradient de g saturée et $\mu_i > 0$

Démonstration: On se déplace à partir de x*. Soit $x(\alpha) = x^* + \alpha d$ avec $d>0$, d admissible

Développement de Taylor

$f[x(\alpha)]=f(x^*)+ \alpha[\partial f/ \partial x^*]^T.dx/d\alpha + \alpha\, \varepsilon(\alpha)$

$f[x(\alpha)]=f(x^*)+ \alpha[f_{x^*}^T d] + \alpha\, \varepsilon(\alpha)$

x^* mini => $f[x(\alpha)] \geq f(x^*)$ => $f_{x^*}^T.d \geq 0$

Cas sans contrainte

x^* mini => $f_{x^*}=0$

Introduction du Lagrangien pour des contraintes C_1

Cas des contraintes égalité (Conditions de Lagrange)

$$\min_x f(x) \text{ contraint par } h_j(x) = 0 \quad j=1\ldots p$$

On définit une fonction de pénalité $\Phi(x)$ = {0 si x est admissible, + ∞ si non}

Le problème contraint : {Min f(x), $h_j(x) = 0$} est équivalent au problème non contraint

$$\min_x [f(x) + \Phi(x)]$$

Choix de la fonction de pénalité $\Phi(x)$

$$\Phi(x) = \max_\lambda \left[\sum_{j=1}^{p} \lambda_j h_j(x)\right]$$

Si x est admissible $\Phi(x)=0$

Si x est non admissible, il existe *j* tel que $h_j(x) \neq 0$ D'où $\max_\lambda \left[\sum_{j=1}^{p} \lambda_j h_j(x)\right] = +\infty$

Le problème non contraint auquel on est ramené s'écrit donc

$$\min_x \left[f(x) + \max_\lambda \sum_{j=1}^{p} \lambda_j h_j(x)\right]$$

Comme f(x) est indépendant de λ, on peut écrire $\min_x \left[\max_\lambda \left(f(x) + \sum_{j=1}^{p} \lambda_j h_j(x)\right)\right]$

Le Lagrangien est défini par

$$L(x,\lambda)=f(x)+\sum_{j=1}^{p}\lambda_j h_j(x)$$

On appellera problème primal $\quad \underset{x}{\mathrm{Min}}\Big[\underset{\lambda}{\mathrm{Max}}\big(f(x)+\sum_{j=1}^{p}\lambda_j h_j(x)\big)\Big]$

On appellera problème dual $\quad \underset{\lambda}{\mathrm{Max}}\Big[\underset{x}{\mathrm{Min}}\big(f(x)+\sum_{j=1}^{p}\lambda_j h_j(x)\big)\Big]$

Soient (x^*,λ_j^*) la solution (si elle existe) du problème (primal ou dual) sous condition de dualité forte

x^* optimum => Il existe $\lambda_1^* \ldots \lambda_p^*$ tels que $\partial f/\partial x^* + \sum_{j=1}^{p}\lambda_j^*\, \partial h_j/\partial x^* = 0$

(Conditions de Lagrange du 1er ordre)

<u>Cas de contraintes inégalité</u> (Conditions de Kuhn & Tucker)

$\underset{x}{\mathrm{Min}}\ f(x),\ g_i(x) \leq 0\ ,\quad i=1 \ldots m$

On choisit la fonction de pénalité

$\Phi(x) = \underset{\lambda_i}{\mathrm{Max}} \sum_{i=1}^{m}\lambda_i g_i(x))] \quad \lambda_i \geq 0$

Si x est admissible $g_i(x) \leq 0$ et $\Phi(x)=0$

Si x est non admissible, $\exists\ i$ tel que $g_i(x) > 0$ D'où $\mathrm{Max}[\sum \lambda_i g_i(x)] = +\infty \quad \lambda_i \geq 0$

On appellera problème primal $\min_{x} [\max_{\lambda \geq 0} L(x,\lambda)]$

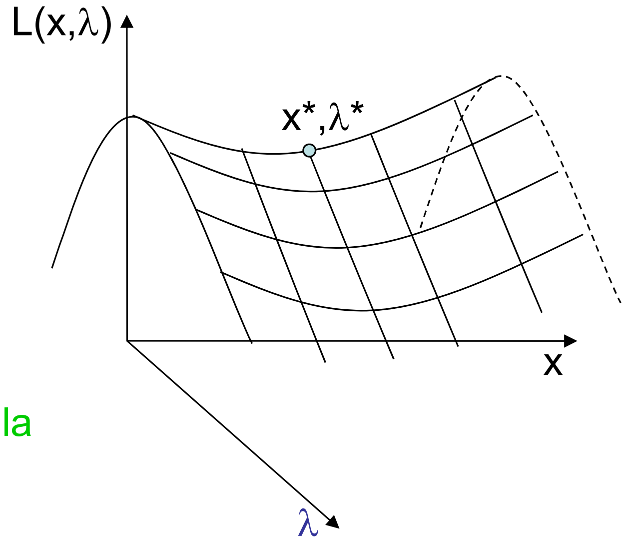
On appellera problème dual $\max_{\lambda \geq 0} [\min_{x} L(x,\lambda)]$

Soient (x^*,λ_j^*) la solution (si elle existe) du problème (primal ou dual)

x^* optimum => il existe $\lambda_i^* \geq 0$ tels que $\partial f / \partial x^* + \sum_{i=1}^{m} \lambda_i^* \partial g_j / \partial x^* = 0$

Interprétation: $\partial f/\partial x^* = - \Sigma \lambda_i \, \partial g_i/\partial x^*$

(Gradient de f = combinaison linéaire négative des contraintes)



Point selle du Lagrangien

$\min_{x} [\max_{\lambda \geq 0} L(x,\lambda)]$ primal

$\max_{\lambda \geq 0} [\min_{x} L(x,\lambda)]$ dual

Si la Lagrangien admet un point col, alors la solution du pb dual= solution du pb primal

Exemple (cas convexe)   Utilisation du problème dual

- Min $f(x) = x_1^2 + x_2^2$
  avec $2x_1 + x_2 \leq -4$

Lagrangien:

$L(x, \lambda) =$

$\underset{x}{\text{Min}}\ L(x, \lambda) \Rightarrow$

Fonction duale: Max $W(\lambda) = L[x^*(\lambda), \lambda] =$

d'où $x_1^* =$, $x_2^* =$ et $f(x^*) = W(\lambda^*) =$

Le problème dual est parfois beaucoup plus simple à résoudre

Conditions nécessaires du 2e ordre

CN2 (pas d'hypothèse sur les contraintes)

Soit f(x) de classe C_2

x^* mini local => $\forall$ d admissible en x^*, $f_{x^*}^T d \geq 0$ où f_{x^*} est le gradient de f

Et si $f_{x^*}^T d = 0$, alors $d^T F^*_{xx} d \geq 0$ (F^*_{xx} traduit la courbure vers le haut ou le bas)

Démonstration:

x^* mini local => $\forall$d admissible $f_{x^*}^T d \geq 0$ déjà démontré en CN1

Si $f_{x^*}^T d = 0$, alors

$f[x(\alpha)]=f(x^*)+ \alpha.0 + ½ \alpha^2 d^T F^*_{xx} d+ \alpha^2 \varepsilon(\alpha)$

$f[x(\alpha)] \geq f(x^*)$ => $d^T F^*_{xx} d \geq 0$ pour tout d admissible

On peut chercher des propriétés en précisant la façon dont est défini le domaine (Kuhn & Tucker)…

<u>Cas sans contraintes</u>

x^* opt => $f_{x^*}=0$, $F^*_{xx} \geq 0$

Conditions suffisantes du 2nd ordre (fonction et contraintes de classe C_2)

Conditions suffisantes du 2^e ordre

Introduction du Lagrangien du problème

Minimiser f(x) contraint par

$g_i(x) \leq 0 \quad i = 1 \ldots m$

$h_j(x) = 0 \; j = 1 \ldots p \leq n-1$

$L(x, \mu_i, \lambda_i) = f(x) + \Sigma \mu_i g_i(x) + \Sigma \lambda_j h_j(x)$

C.S.2 Il existe un lot de λ_j un lot de μ_i positionnels

tels que

(1) $\Sigma \mu_i g_i(x^*)=0$ (2) $L_{x^*}= (x^*,\mu_i , \lambda_j)=0$ (3) $d^T L^*_{xx} d \geq 0$ pour tout $d \in \zeta(x^*)$

Alors x* est minimum local.

Remarques générales sur CS2

1) $e\mathcal{T}\mu_i g_i(x^*)=0$ => prendre les μ_i nuls pour les contraintes non actives

2) $L_{x^*}= (x^*,\mu_i , \lambda_j)=0$ => $f_x(x^*)= - \Sigma \mu_i g_i(x^*) - \Sigma \lambda_j h_j(x^*)$

Illustration (en l'abscence de h_j)

$-f_x(x^*)$ se décompose dans composantes > 0 sur g_{1x} et g_{2x} (strictement)

descente de f => sortie du domaine

3) Il s'agit bien d'un minimum et non d'un maximum

Remarque sur CS2 quand x* est régulier

- Quand x* est régulier, mes conditions (1) et (2) sont appelés conditions nécessaires de K.T. du 1er ordre

Cas sans contraintes:

$f_x(x^*) = 0$ et $F_{xx}(x^*) > 0$ => x* est minimum local

Méthodes primales

- Les méthodes primales cherchent à résoudre le problème directement, et

1) engendrent une suite de points satisfaisant les contraintres

2) correspondent à une suite décroissante des valeurs de la fonction

- En cas d'arrêt de la procédure, le point courant peut être pris comme approximation de l'optimum recherché.

I.1Méthodes de changement de variable:

$a \leq x \leq b \quad \Rightarrow x=a+(b-a)\sin^2(y)$

On remplace x par une variable non contrainte y

I.2Méthodes des diretions admissibles (Zoutendijk, Topkins et Veinott)

- x_0 le point de départ satisfaisant les contraintes
- I_0 l'ensemble des indices des contraintes saturées en X_0

On cherche une direction admissible d

On doit avoir $(\partial g_j / \partial x_0).d \leq 0$

(afin que le déplacement ne rende aucune contrainte saturée positive)

D'autre part la fonction doit diminuer lors d'un déplacement

$(\partial f / \partial x_0)^T.d < 0$

Idée : Min $(\partial f / \partial x_0)^T.d < 0$

contraint par $(\partial g_j / \partial x_0).d \leq 0 \ i \in I_0$

$\Sigma \left| d \right| = 1$

Inconvénients:

•Tout déplacement suivant d peut faire sortir du domaine
Exemple

•Un petit déplacement peut changer le nombre de contraintes saturées et entraîner un discontinuité brusque de la direction

Exemple (fait au tableau)

=> Raffinements de la méthode précédente (Topkins et Veinott)
Tenir compte de toutes les contraintes (saturées + non saturées)

II.2 Méthodes de pénalité avec paramètres

- Pénalité à point intérieur : <u>la proximité des contraintes est pénalisée avec un poids de moins en moins grand</u>

- Min $f(x)$ avec $g_i(x) > 0$ et X_0 point de départ admissible

- $\underset{x}{\text{Min}}\ \theta\,(x,r_k)=f(x)+r_k \sum 1/g_i(x)$ avec r_k suite décroissante vers zéro

Illustration: (faite au tableau)

II.2 Méthodes de pénalité avec paramètres

- Pénalité à point extérieur : la non satisfaction des contraintes est pénalisée avec un poids de plus en plus grand

- Min $f(x)$ avec $g_i(x) > 0$ et X_0 point de départ, non nécessairement admissible

- Min $F(x,r_k)=f(x)+1/r_k \Sigma \{Min[0,g_i(x)]\}^2$ avec r_k suite décroissante
  x

II.2 Méthodes de pénalité avec paramètres

- Méthodes mixtes
- Min $f(x)$ avec

  $g_i(x) > 0$ $i=1\ldots m$

  $h_j(x) = 0$ $j=1 \ldots p$

On décompose $h_j(x) = 0 \iff h_j(x) > 0, j=1 \ldots p$ et $h_j(x) \leq 0, j=1 \ldots p$

on note $g_i(x) > 0$ $i=m+1, \ldots, m+1+2p$

Soit en posant $q=m+1+2p$

- $$\min_x F(x,r_k)=f(x)+ r_k \sum_{i=1}^{n} 1/g_i(x) + 1/r_k \sum_{i=n+1}^{q}\{\mathrm{Min}[0,g_i(x)]\}^2$$ avec r_k suite décroissante vers 0

1er terme : inégalités impératives 2eme terme : inégalités non impératives (tjs vérifiées) et contraintes égalité