Optimisation en Mécanique

1. Introduction

- Formulation d'un problème d'optimisation :
  Terminologie : variables, fonction objectif, limitations
  Ecriture des problèmes sans/avec limitations
  Applications en Mécanique : exemples de problèmes académiques et industriels, fonctions coût et limitations
- Représentation graphique d'un problème d'optimisation à 2 variables
- Rappels mathématiques (gradient, convexité, formule de Taylor, conditions d'optimalité dans le cas sans limitations…)
- Différentes classes de problèmes : fonctions linéaires, quadratiques, non linéaires, continues, non continues, non partout définies, différentiables ou non, convexes ou non , unimodales, multimodales

2. Notion de plan d'expériences

- Définitions : intérêt, relation avec l'optimisation, notion de facteurs, effets, interactions, ..
- Plans factoriels complets, fractionnaires, ...
- Exemple

3. Méthodes d'approximation

- Approximations locales : linéarisation , linéarisation convexe, approximations quadratiques
- Exemple
- Approximations globales, surfaces de réponse

4. Minimisation à 1 dimension

- localisation du minimum (dichotomie, ..),
- Méthodes de Newton-Raphson, de la sécante, interpolation polynômiale, section d'or
- Exemple

5. Minimisation en dimension n sans limitations :

- sans dérivées : simplexe (Nelder-Mead)
- avec dérivées 1ères : relaxation, plus forte pente, gradient conjugué
- Méthodes de Newton et quasi-Newton
- Problèmes numériques : mise à l'échelle, conditionnement du Hessien
- Exemple : Minimisation de l'énergie potentielle d'une poutre cantilever

6. Minimisation avec limitations

- aspects théoriques : conditions d'optimalité, Lagrangien, point-col, interprétation
- Exemple
- Méthodes primales : directions admissibles, gradient réduit, gradient projeté
- Méthodes de pénalité
- Méthodes duales : Lagrangien augmenté
- Critères d'optimalité, exemple du FSD

7. Application au calcul des structures : calcul des sensibilités, mise en œuvre, logiciels

- calcul des sensibilités par différences finies
- calcul des sensibilités par dérivation des équations d'équilibre en éléments finis: état direct et adjoint
- optimisation de forme : remaillage, estimation d'erreur
- Mise en œuvre de l'optimisation : couplages CAO-calcul, utilisation de bibliothèques, logiciels E.F. commerciaux intégrant de l'optimisation, logiciels de gestion de calculs, logiciels de différentiation automatique
- optimisation multi-critères, multi-disciplinaire

8. Identification

- Principe de l'identification, exemple
- méthode de Levenberg-Marquardt
- exemples (système de 2 ressorts, problème de thermique)

9. Méthodes pour l'optimisation globale et cas des variables discrètes

- recuit simulé
- algorithmes évolutionnaires
- exemple (section d'une poutre)

10. Optimisation topologique

- Principe
- Aspects théoriques : homogénéisation
- Application en élasticité 2D : méthode de Ole et Sigmund, méthode 'ESO', algorithme évolutionnaire.

Chapitre I - Introduction

1. Formulation d'un problème d'Optimisation

Conception optimale :

Optimiser la conception d'une pièce, d'une structure =

rechercher la ***meilleure*** structure ***possible***,

c'est-à-dire celle qui assure le prix de revient minimum (coûts de l'étude, de la fabrication, des matières premières, de la maintenance, …)

⇒ **fonction coût (ou objectif)** { poids ; volume ; fonction des contraintes ; énergie de déformation ; déformée, … }

définie à partir de paramètres à ajuster :

⇒ **Variables de conception, (ou paramètres d'optimisation, variables de décision)** { dimensions ; forme ; type de matériau ; orientation de fibres, … }

et telle qu'elle vérifie certaines restrictions :

⇒ **limitations, (ou contraintes d'optimisation**) { contraintes ; déplacements ; fréquences propres ; charges critiques ; dimensions, faisabilité, … }

Ce problème se formule de manière mathématique tout à fait générale :

$$\underset{x}{\text{Minimiser}}\ f(x)$$

avec les conditions :

$$g_j(x) \leq 0 \qquad j = 1, \ldots, m$$

Terminologie :
f : **fonction coût**
$x = (x_1 , x_2 , \ldots , x_n)$: vecteur des **variables de conception** ($x \in R^n$)
$g_j(x)$, $j = 1, \ldots, m$ = **limitations**

L'ensemble A des points vérifiant toutes les limitations est appelé le **domaine admissible**

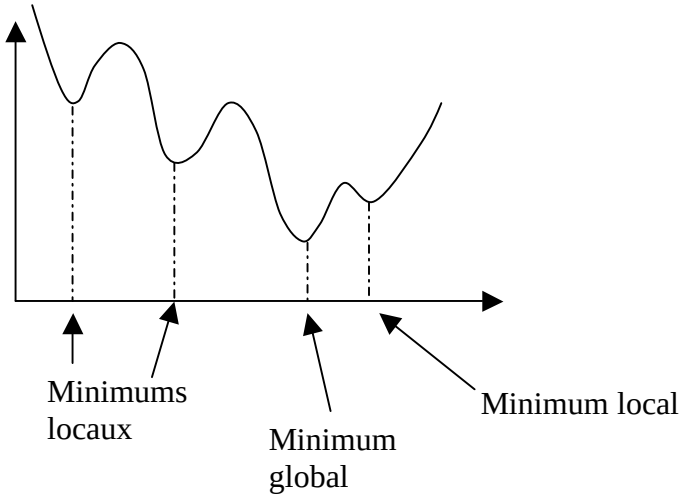
L'**optimum global** x*, s'il existe vérifie :

$$f(x^*) \leq f(x) \ \forall \ x \in A \subset R^n$$

On dira que x* est un **optimum local** de f s'il est un minimum dans une région limitée seulement, c'est-à-dire s'il existe $\eta \geq 0$ tel que :

$$f(x^*) \leq f(x) \ \forall \ x \in A \text{ et vérifiant } || x - x^* || \leq \eta$$

Le minimum global est le minimum local pour lequel la valeur de f est la plus faible, comparée à celle des autres minimums.



Minimum global et minimums locaux à 1 dimension, sans limitations.

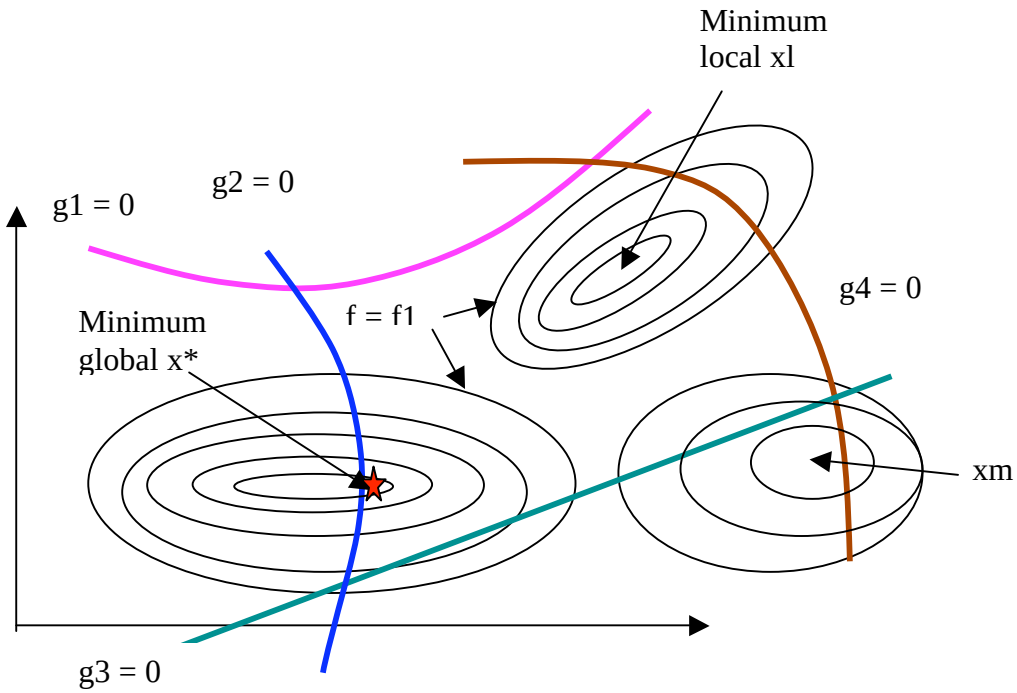
Généralement, l'optimum d'un problème avec contraintes est atteint sur la frontière du domaine admissible (sinon, cela signifie que les contraintes sont inutiles …). Si en x*, gj(x*) = 0, on dit que x* sature gj. , ou que gj. est **saturée** en x*.

Représentation graphique d'un problème à 2 variables.

Lorsque x est dans R^2 , on peut représenter graphiquement les problèmes d'optimisation en utilisant les isovaleurs (ou lignes de niveau, analogues aux altitudes sur les cartes géographiques) de la fonction coût et des limitations :

➡ Les isovaleurs « 0 » des contraintes délimitent le domaine admissible A (intersection pour tout j des domaines où l'on a $g_j(x) \leq 0$)

➡ Les isovaleurs de f (tracé des courbes + valeur associée) permettent de localiser graphiquement le minimum de f à l'intérieur du domaine admissible (ou les minimums locaux), c'est-à-dire le point de plus faible altitude à l'intérieur de A.
La figure suivante donne un exemple d'une telle représentation.



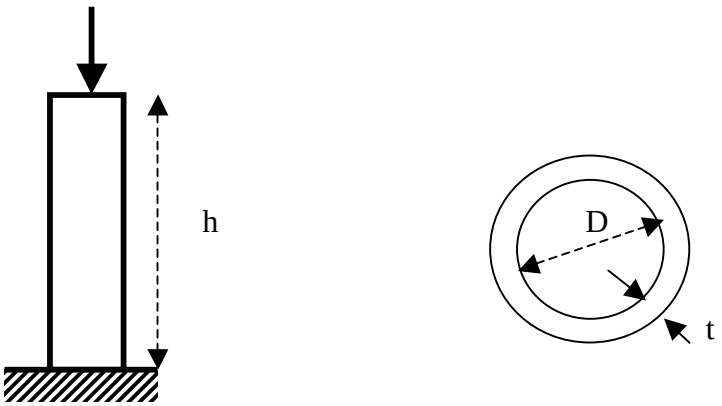
Exemple : problème avec 4 limitations, fonction avec plusieurs minima locaux :
Le but de l'optimisation est de rechercher x* A tel que f(x*) ≤ f(x) pour tout x ∈ A

si on suppose f(x*) < f(xl), x* est le minimum global
xl est un minimum local de f admissible,
xm est un minimum local de f non admissible (contrainte g3 non satisfaite)

Le minimum global x* est atteint sur la frontière du domaine admissible : la contrainte g2 est saturée en x*.

Mais dans le cas général d'un problème de conception, même avec 2 variables seulement, on ne dispose pas de cette représentation graphique, qui nécessite le calcul de la fonction coût et des limitations en un très grand nombre de points sur tout le domaine. Nous utiliserons ce type de représentation essentiellement pour comprendre comment fonctionnent les algorithmes de minimisation. Notons cependant que si on n'a que 2 variables, et si un calcul rapide de toutes les fonctions est possible, on peut avoir recours à une résolution graphique.

Exemple d'optimisation d'une structure : conception d'une colonne tubulaire en compression



Données : la charge P, la hauteur h, le matériau
On minimise la masse :

$$F = W = \rho\pi\, D\, t\, h$$

Les variables de conception sont le diamètre moyen D et l'épaisseur t.

La colonne doit résister en flambement global ; en flambement local, en rupture statique.
La contrainte axiale de compression :

$$\sigma = \frac{P}{\pi D t}$$

doit donc vérifier :

$$\sigma \le \sigma_R \qquad \text{(rupture)}$$

$$\sigma \le \sigma_{fg} = \frac{4\pi^2 E I}{\pi D t h} \qquad \text{(flambement global)}$$

$$\sigma \le \sigma_{fl} = \frac{2\ Et}{D\sqrt{3(1-\nu)^2}} \qquad \text{(flambement local)}$$

Les conditions de faisabilité sont :

$D \ge t$
$D \ge 10^{-6}$
$t \ge 10^{-6}$

Sous forme standard , le problème s'écrit :

$$\text{Minimiser } f(D, t) = \frac{W(D,t)}{W_0} = \frac{\rho\pi D t h}{W_0}$$

(D, t)

avec :

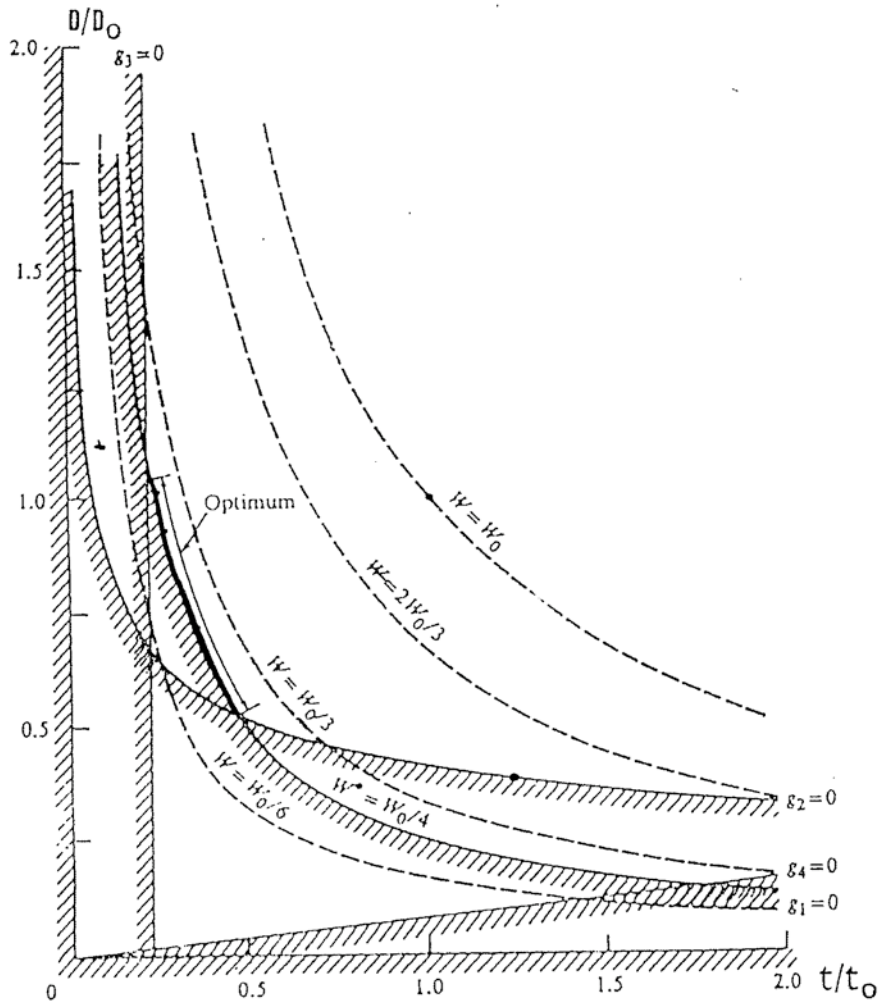
$$g_1(D, t) = \frac{\sigma}{\sigma_R} - 1 \le 0$$

$$g_2(D, t) = \frac{\sigma}{\sigma_{fg}} - 1 \le 0$$

$$g_3(D, t) = \frac{\sigma}{\sigma_{fl}} - 1 \le 0$$

$$g_4(D, t) = \frac{t}{D} - 1 \le 0$$

Ce problème est représenté graphiquement comme suit :



L'espace de conception

(Exemples d’applications)

Remarques générales sur l'application des méthodes d'optimisation.

- On peut utiliser divers choix de variables : aires de sections, dimensions, forme, topologie. Tout dépend de la « liberté » dont disposent les concepteurs.

- Plusieurs formulations différentes peuvent donner des résultats équivalents, mais :

- Plusieurs formulations théoriquement équivalentes peuvent être très différentes sur le plan de leur efficacité numérique. Le choix d'une formulation efficace est souvent le premier problème à résoudre.

- Les méthodes de calcul des fonctions sont multiples : analytiques, différences finies, éléments finis, équations intégrales

- Il est toujours préférable d'écrire le problème sous forme adimensionnelle

- Les temps calcul et la nature de la solution obtenue dépendent fortement de l'estimation initiale (utilisation des connaissances acquises sur le sujet, plans d'expériences numériques dans le domaine de conception, …)

- On ne peut que rarement affirmer que l'on atteint un minimum global ; souvent, on se contente d'améliorer une conception préexistante.

- Il n'existe pas de logiciel « boîte noire » universel, résolvant automatiquement tout problème d'optimisation; bien souvent, les logiciels proposent différents modules parmi lesquels il faut choisir l'algorithme le mieux adapté (éventuellement plusieurs méthodes successivement). On distingue généralement deux grandes catégories de méthodes : les méthodes qui recherchent un minimum **local** en utilisant par exemple les dérivées des fonctions, et les méthodes **globales** qui font souvent appel à des processus aléatoires.

2. Notations et rappels mathématiques

- Vecteurs, matrices

x: vecteur colonne de R^n , x^T =(x1, x2, … xn)
I : matrice identité

H : matrice (n, n), H^T sa transposée
H H^T est symétrique car $(H H^T)^T = (H^T)^T H^T = H H^T$

Dét(H) = déterminant de H
Si dét(H) = 0, H est une matrice singulière
Si H est non singulière, il existe une matrice unique H^{-1} = matrice inverse de H telle que :
$H H^{-1} = H^{-1} H = I$

Produit scalaire de 2 vecteurs : $x^T x$ noté aussi (x, x)
Norme d'un vecteur x de R^n : $|| x || = (x^T x)^{1/2} = (\Sigma xi^2)^{1/2}$

Des vecteurs x^1, x^2…, x^p sont dits orthogonaux si $x^{iT} x^j = 0$ pour tout $i \neq j$
Ils sont orthonormés si, en plus, $x^{iT} x^i = 1$ pour tout i

Une matrice carrée symétrique est dite semi-définie positive si l'on a :
$x \in R^n$, $x^T H x \geq 0$ pour tout $x \neq 0$,
(définie positive si on a inégalité stricte >)

- Différentiabilité

Soit f une fonction dérivable, f : $R^n \rightarrow R$; on appelle gradient de f en x° le vecteur :

$$\nabla f(x°) = \text{gradf}\,(x°) = \begin{Bmatrix} \frac{\partial f}{\partial x_1}(x°) \\ \vdots \\ \vdots \\ \vdots \\ \frac{\partial f}{\partial x_n}(x°) \end{Bmatrix}$$

Si f admet aussi des dérivées secondes, on peut définir le hessien de f en x° :

$$\nabla^2 f(x°) = Hf\,(x°) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x°) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x°) & . & . & . \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x°) & \vdots & \vdots & . & . \\ \vdots & \vdots & \vdots & . & . \\ \vdots & \vdots & \vdots & . & . \\ \vdots & \vdots & \vdots & . & \frac{\partial^2 f}{\partial x_n^2}(x°) \end{bmatrix}$$

- Formule de Taylor

Pour une fonction d'une seule variable r fois dérivable, la formule de Taylor donne une approximation de f au voisinage d'un point x :

$$f(x+s) = f(x) + sf'(x) + \frac{s^2}{2} f''(x) + \ldots + \frac{s^{r-1}}{(r-1)!} f^{(r-1)}(x) + O(s^r)$$

où $O(s^r) \to 0$ quand $s \to 0$

Pour les fonctions de n variables, nous utiliserons seulement le développement d'ordre 2 :

$f(x+s) = f(x) + s^T \nabla f(x) + ½ s^T H(x) s + ||s||^2 O(s)$ où $O(s) \to 0$ quand $s \to 0$

- Directions de descente

$s \in R^n$ est une direction de descente pour la fonction f au point x s'il existe un intervalle $[0, \eta] \subset R$ tel que l'on ait:

$$f(x+r s) < f(x) \text{ pour tout } r \in [0, \eta]$$

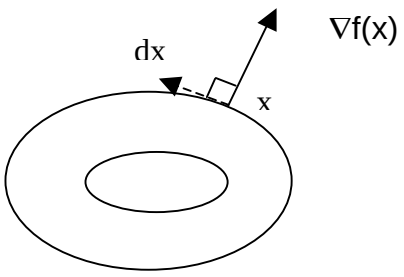
Caractérisation des directions de descente :
s est une direction de descente si et seulement si : $s^T \nabla f(x) < 0$
En effet :
$f(x+rs) = f(x) + r s^T \nabla f(x) + ½ r^2 s^T H(x) s + \ldots$
Pour r suffisamment petit, on aura :
$f(x+rs) < f(x)$ si et seulement si $s^T \nabla f(x) < 0$

Interprétation graphique en 2D:
Toutes les directions qui sont situées dans le demi-plan contenant $-\nabla f(x)$ sont des directions de descente.
En particulier, $s = -\nabla f(x)$ est la direction de plus grande descente (ou plus forte pente) ; en effet, parmi les s normés, $-\frac{\nabla f(x)}{\|\nabla f(x)\|}$ est le vecteur qui minimise la valeur $s^T \nabla f(x)$

- Propriété du gradient

Le gradient de f en x est orthogonal à la ligne de niveau passant par x :
f = constante sur la ligne de niveau, donc :

$$df = \nabla f^T(x)\, dx = 0$$



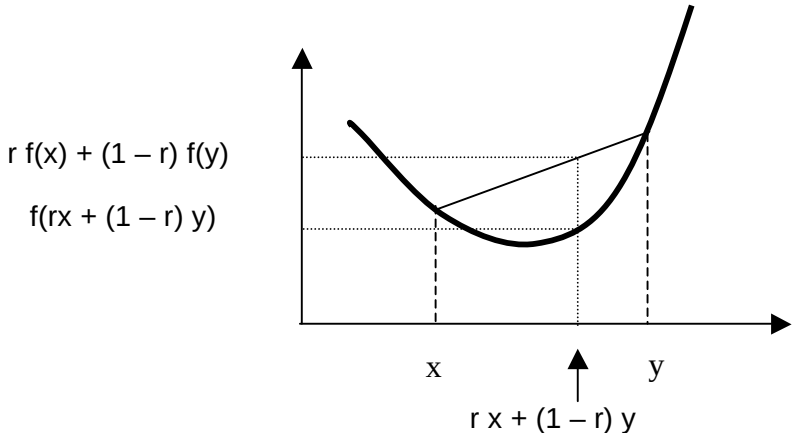
- Convexité

C'est une propriété importante dans le contexte de l'optimisation, car une bonne partie des résultats théoriques sur lesquels s'appuient les méthodes de minimisation sont valables seulement pour des fonctions convexes, et des domaines convexes.

On dit qu'une fonction f : $R^n \rightarrow R$ est convexe si :

$\forall x \in R^n, \forall y \in R^n$, $\forall r$ tel que $0 \leq r \leq 1$, on a :
$f(rx + (1 – r)\, y) \leq r\, f(x) + (1 – r)\, f(y)$

f est dite strictement convexe si on a l'inégalité stricte pour $x \neq y$ et $0 < r < 1$
Si l'inégalité est en sens inverse, la fonction est dite concave ($\geq$) ou strictement concave (>).



Fonction convexe

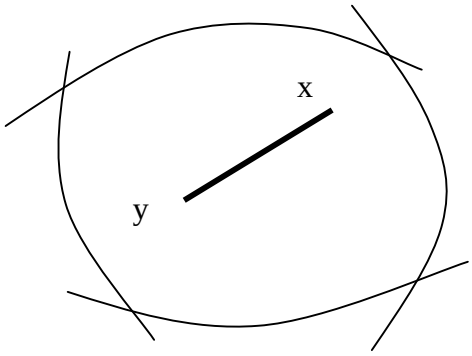
Soit f : $R^n \rightarrow R$. Les propriétés suivantes sont équivalentes :

- f est une fonction convexe
- $\forall x \in R^n, \forall y \in R^n, f(y) \geq f(x) + \nabla f^T(x)\,(y-x)$
- (si f est 2 fois continûment différentiable) la matrice hessienne de f est semi-définie positive.

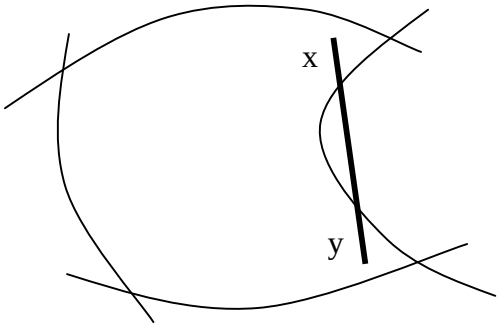
Ensemble convexe :

On dit qu'un ensemble $A \subset R^n$ est convexe si le segment joignant 2 points quelconques de A est contenu dans A :

$\forall x \in A, \forall y \in A$, $\forall r$ tel que $0 \leq r \leq 1$, on a : $rx + (1 – r)\, y \in A$



Ensemble convexe



Ensemble non convexe

3. Conditions d'optimalité

- Conditions nécessaires d'optimalité locale :

> Soit $f \in C^2(R^n)$ (2 fois continûment différentiable)
> Si x* est un minimum local de f, alors :
>
> $\nabla f(x^*) = 0$, et
> $H(x^*) = \nabla^2 f(x^*)$ est semi- définie positive

preuve :
Soit x* un minimum local de f ; le développement de Taylor en x* donne pour tout x:

$$f(x) = f(x^*) + (x - x^*)^T \nabla f(x^*) + \tfrac{1}{2} (x - x^*)^T H(x^*) (x - x^*) + ||(x - x^*)||^2 O(x - x^*)$$
où $O(x - x^*) \to 0$ quand $x - x^* \to 0$

Supposons que $\nabla f(x^*) \neq 0$; en prenant $x = x^* - r \nabla f(x^*)$ avec $r > 0$, on a :

$$f(x) = f(x^*) - r \nabla f(x^*)^T \nabla f(x^*) + \tfrac{1}{2} r^2 \nabla f(x^*)^T H(x^*) \nabla f(x^*) + r^3 ||\nabla f(x^*)|| O(\nabla f(x^*))$$

Pour r suffisamment petit, la somme des 3 derniers termes est négative, d'où une contradiction. Donc :

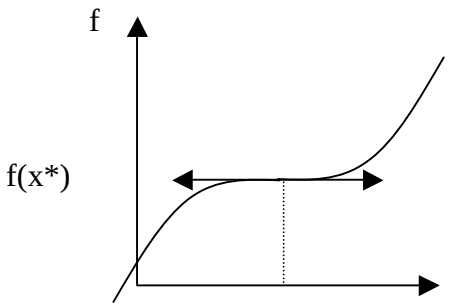
$$f(x) = f(x^*) + \tfrac{1}{2} (x - x^*)^T H(x^*) (x - x^*) + ||(x - x^*)^2|| O(x - x^*)$$

Supposons H(x*) non semi-définie positive. Il existe alors s tel que $s^T H(x^*) s < 0$
Pour $x = x^* + r s$, on obtient :

$$f(x) = f(x^*) + \frac{r^2}{2} s^T H(x^*) s + r^3 ||s||^2 O(s)$$

Pour r suffisamment petit, la somme des 2 derniers termes est négative, d'où une contradiction. Donc H(x*) est semi-définie positive.

Remarque : cette condition n'est pas suffisante ; en 1 dimension :



$f'(x^*) = 0$
$f''(x^*) = 0$

- Conditions suffisante d'optimalité locale

> Soit $f \in C^2(R^n)$ (2 fois continûment différentiable)
> Si un point x* vérifie :
>
> $\nabla f(x^*) = 0$,
> et $H(x^*) = \nabla^2 f(x^*)$ est définie positive

alors x* est un minimum local de f.

Preuve :
Soit $x = x^* + r\,s$ où $||\,s\,|| = 1$
On a alors, pour tout r suffisamment petit:

$$f(x + r\,s) = f(x^*) + \frac{r^2}{2}\, s^T H(x^*)\, s + r^2\, ||\,s\,||^2 O(rs)$$

Pour tout $r \neq 0$, le second terme est > 0, ce qui implique :

$f(x + r\,s) > f(x^*)$ pour tout r suffisamment petit, ce qui signifie bien que x* est minimum local de f.

- Conditions nécessaires et suffisantes d'optimalité globale dans le cas de fonctions convexes

> Si f est une fonction convexe continûment différentiable, x* est un optimum global de f si et seulement si $\nabla f(x^*) = 0$

Remarque : c'est la seule CNS d'optimalité dont on dispose.

Chapitre II – Notion de plan d'expériences

1. Introduction

A l'origine, les P.E. s'appliquent à l'expérimentation (non numérique) et concernent la stratégie de recherche à suivre pour répondre à un certain nombre de questions ; l'expérimentateur cherche comment sélectionner les expériences à faire, quelle est la meilleure stratégie pour :

- Conduire le plus rapidement possible aux résultats espérés,
- Eviter des expériences inutiles,
- Apporter une bonne précision,
- Conduire à la modélisation et à l'optimisation des phénomènes étudiés

Une littérature abondante existe sur les P.E., mais dans le cas d' "expérimentation" numérique, tous les aspects liés aux erreurs de mesure sont sans objet (répétitions de la même expérience par exemple)

Un plan d'expériences peut être utilisé comme une méthode d'optimisation, pour trouver une ou des solutions au problème posé, mais aussi comme une étape préliminaire à l'optimisation et a alors pour objectif le choix des variables à optimiser et des fonctions à prendre en compte dans une formulation mathématique classique pour résoudre le problème par une méthode de gradient par exemple.

Terminologie :

Réponse : grandeur à laquelle on s'intéresse (y) ; son choix est un autre problème.
Facteurs : variables qui peuvent être continues ou discrètes (x_1, x_2 , … x_n), qualitatives (pas de logique de classement) ou quantitatives.
Niveaux : valeurs prises par un facteur dans les expériences

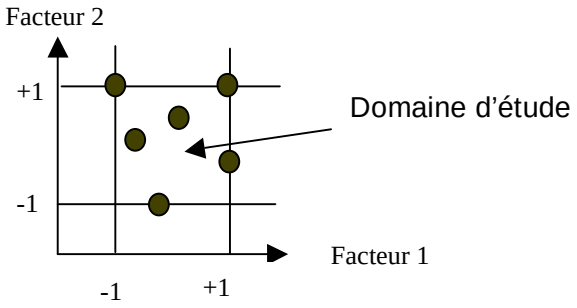
Méthode classique (cf relaxation, ou directions alternées) : fixer les niveaux de toutes les variables sauf une (ex : x_1) à laquelle on donne plusieurs valeurs successivement ; la mesure de la réponse donne une courbe $y=f(x_1)$. Recommencer pour chaque facteur …Mais :

⇒ 7 facteurs, 5 niveaux donne 5^7 = 78125 expériences
7 facteurs, 3 niveaux donne 3^7 = 2187 expériences

Pour diminuer le nombre d'expériences, il faut faire varier les niveaux de tous les facteurs à la fois à chaque expérience. Ceci permettra d'étudier un grand nombre de facteurs, de détecter les interactions entre facteurs, et de trouver les facteurs influents. Pour cela, au départ, le nombre de niveaux sera généralement fixé à 2 seulement par facteur:

Niveau bas d'un facteur: borne inférieure, notée par –1
Niveau haut : borne supérieure, notée par +1

Représentation graphique (géométrique): domaine de variation d'un facteur, espace expérimental, domaine d'étude, 1 expérience=1 point



Plan d'expériences : liste des combinaisons de niveaux de facteurs pour lesquels la réponse du modèle doit être observée. (cf. liste des points du domaine de conception à partir desquels une approximation est construite)

Surface de réponse : ensemble des réponses correspondant à tous les points du domaine, associée à un modèle mathématique f (modèle postulé, approximation) de la réponse :

$y = f(x_1, x_2 , \ldots x_n)$

(avec 2 variables, la représentation géométrique du P.E. et de la surface de réponse associée est possible dans un espace de dimension 2 ou 3, sous forme de lignes ou de surface d' isovaleurs)

Les plans d'expériences ne font pas tous appel à des surfaces de réponse : les plus simples utilisent seulement 2 niveaux (réponse linéaire), et la notion de surface de réponse n'apparaît pas explicitement. Les approximations les plus utilisées sont des polynômes de degré 1 ou 2. Si un modèle est insuffisant (pas assez précis), il sera possible d'ajouter des termes au polynôme et des points au P.E. pour obtenir une meilleure approximation.

Différents types de plans :

Plans de criblage : pour trouver les facteurs les plus influents sur une réponse.

Plans de modélisation = plans pour surfaces de réponse: modèles du 1^{er} ou $2^{ème}$ degré

Plans de mélanges : adaptés aux facteurs dépendants

Plans complets/fractionnaires :

plans factoriels complets : toutes les combinaisons des niveaux de facteurs sont présentes

plans factoriels fractionnaires : tous les niveaux de chaque facteur sont présents , mais pas toutes les combinaisons possibles de facteurs

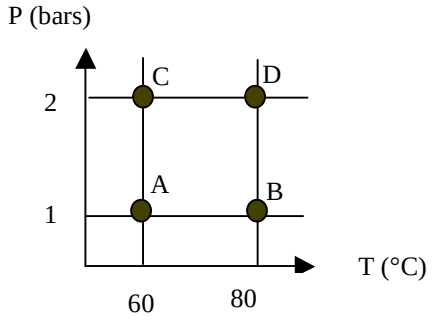
Historique : travaux de Fisher (1925) pour des essais combinant plusieurs facteurs, applications en agronomie (très grand nombre de paramètres), puis aspects théoriques étudiés par les statisticiens (Plackett et Burman, 1946), ensuite applications par les industriels et universitaires (Box, Behnken, Taguchi, …).

2. Plans factoriels complets à deux niveaux : 2^k

Ce sont les plus simples, mais ils ont de nombreuses applications (complets ou fractionnaires)

2.1 Plan complet à 2 facteurs : 2^2

Exemple : étude du rendement d'une réaction chimique
2 facteurs : température et pression



Variables centrées réduites : changement d'unité de mesure et d'origine pour avoir niveau bas= -1 et niveau haut = +1 pour tous les facteurs :

T	60°C	70°C	80°C
P	1 bar	1.5 bar	2 bars
Var.c..réd.	-1	0	1

La variable centrée réduite prend la valeur 0 au milieu du domaine.
Pas : valeur en unités d'origine correspondant à 1 en unités réduites. (pour T, pas = 10°C)

Matrice d'expérience (ou matrice des essais) :

N°	Température Facteur 1	Pression Facteur2	Rendement (réponse)
1	-1	-1	60
2	+1	-1	70
3	-1	+1	80
4	+1	+1	95

Effet global et effet moyen d'un facteur :

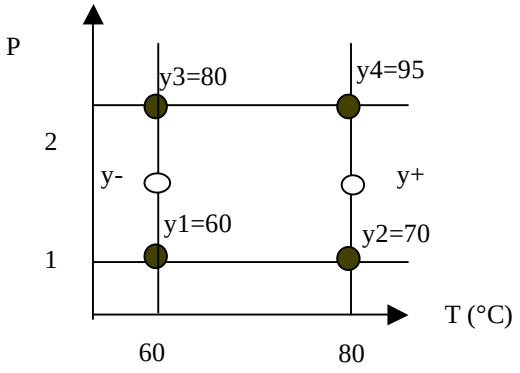
Moyenne des réponses au niveau +1 de T:

$y_+ = ½ (y2 + y4) = 82.5$

Moyenne des réponses au niveau -1 de T:

$y_- = ½ (y1 + y3) = 70$

Effet global de la température :



$E_{Tg} = (y_+ - y_-) = 12.5$

Effet moyen de la température :

$E_T = ½ E_{Tg} = ½ (y_+ - y_-) = 6.25$

soit :

$E_T = ½ (- y_1 + y_2 - y_3 + y_4)$

Et de même pour l'effet moyen de la pression :

$E_P = ½ (- y_1 - y_2 + y_3 + y_4)$

Remarque : les réponses sont numérotées dans l'ordre des essais (cf. matrice d'expériences) ; les signes dans E_T et E_P sont ceux qui apparaissent dans la colonne du facteur étudié. Cette règle se généralise à tous les plans factoriels à 2 niveaux quel que soit le nombre de facteurs.

Moyenne de toutes les réponses :

$I = ¼ (y_1 + y_2 + y_3 + y_4) = ¼ (60 +70 +80 +95) = 305/4 = 76.25$

(valeur de la réponse au centre du domaine expérimental)

Interaction entre les facteurs : lorsque l'effet d'un facteur dépend du niveau des autres facteurs :

Au niveau bas de T, l'effet de la pression est : $E_p^- = ½ (y_3 - y_1) = ½ (80 – 60) = 10$

Au niveau haut de T : $E_p^+ = ½ (y_4 - y_2) = ½ (95 – 70) = 12.5$

L'interaction de la température sur la pression est la moitié de la différence entre E_p^+ et E_p^-

$E_{tp} = ½ \{ ½ (y_4 - y_2) - ½ (y_3 - y_1) \} = ¼ (y_1 - y_2 - y_3 + y_4)$

$E_{tp} = ½ (12.5 – 10) = 1.25$

Interaction de la pression sur la température :

$½ (½ (95 – 80) – ½ (70 – 60)) = ½ (7.5 – 5) = 1.25$

On a: interaction T sur p = interaction p sur T (résultat général)

$E_{Tp} = E_{pT} = ¼ (y_1 - y_2 - y_3 + y_4)$

Remarque:

Interaction: couplage des effets des facteurs sur la réponse

Indépendance: les niveaux d'un facteur ne sont pas liés aux niveaux des autres (toutes les combinaisons sont réalisables)

Pour retrouver ces formules dans un cas général , on construit la **matrice des effets** :

Matrice des effets

N°	Moyenne	Température Facteur 1	Pression Facteur2	Interaction 12		Rendement (réponse)
1	+1	-1	-1	+1		60
2	+1	+1	-1	-1		70
3	+1	-1	+1	-1		80
4	+1	+1	+1	+1		95

Div.	4	4	4	4

Effet	76.25	6.25	11.25	1.25

Signe d'une réponse dans la formule d' interaction = (signe facteur1) * (signe facteur 2)

> Pour les plans factoriels à 2 niveaux, les effets et les interactions se calculent de cette manière quel que soit le nombre de facteurs.

conclusion de l'exemple : le meilleur rendement est obtenu en augmentant à la fois la pression et la température.

2.2 Plan complet à 3 facteurs : 2^3

Exemple : étude d'une émulsion de bitume .
Problème : recherche de l'influence d'un acide gras et de l'acide chlorhydrique sur la stabilité de l'émulsion, pour 2 bitumes A et B.

Facteur 1 : acide gras : faible (-) et forte (+) concentration
Facteur 2 : acide chlorhydrique : très dilué (-) et peu dilué (+)
Facteur 3 : nature du bitume : A (-) et B (+) (c'est une variable discrète)

Réponse : indice de stabilité de l'émulsion (en points)

Objectif : émulsion stable (indice faible)

Domaine expérimental : cube

Matrice des effets :

Essai	Moy.	Fac.1 (ac.gr)	Fac.2 (HCl)	Fac.3 (bitume)	Int. 12	Int. 13	Int. 23	In. 123		Réponses
1	+	-	-	-	+	+	+	-		38
2	+	+	-	-	-	-	+	+		37
3	+	-	+	-	-	+	-	+		26
4	+	+	+	-	+	-	-	-		24
5	+	-	-	+	+	-	-	+		30
6	+	+	-	+	-	+	-	-		28
7	+	-	+	+	-	-	+	-		19
8	+	+	+	+	+	+	+	+		16

Div.	8	8	8	8	8	8	8	8

Effets	27.25	-1	-6	-4	-0.25	-0.25	0.25	0

Interprétation :

- L'effet du facteur 2 (E2) est le plus important ; il est < 0 : la valeur de la réponse diminue quand le facteur passe du niveau –1 à +1.
- E1 est faible (la concentration d'acide gras est sans importance)
- La nature du bitume est également importante (meilleur résultat avec B)
- Pas d'interactions significatives

Notation de Box : E1 → **1**, E123 → **123**

2.3 Construction des plans factoriels complets à k facteurs : 2^k

Construction de la matrice d'expérience et numérotation
Numérotation dite « classique » : les essais sont numérotés de 1 à n. Dans les colonnes, les valeurs sont les suivantes :
Suite des signes du facteur 1 : - + - + - + …
Suite des signes du facteur 2 : - - + + - - …
Suite des signes du facteur 3 : - - - - + + + + …
Et ainsi de suite (8-, 8+ ; 16-, 16 + ; …)

2^k essais à faire, k effets principaux, 2^k -k- 1 interactions

Représentation matricielle : soit X la matrice formée des lignes et colonnes des moyennes, facteurs, interactions :

X =

Moy.	Fac1	Fac2	In12
+1	-1	-1	+1
+1	+1	-1	-1
+1	-1	+1	-1
+1	+1	+1	+1

Vecteur des réponses : $Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$

Vecteur des effets : $E = \begin{bmatrix} I \\ E_1 \\ E_2 \\ E_{12} \end{bmatrix}$

On a alors : $E = \frac{1}{4} X^t Y$

Et pour tout plan factoriel complet à 2 niveaux :

$$E = \frac{1}{n} X^t Y \qquad \text{avec n = nombre des essais = } 2^k$$

- On peut considérer la transformation du vecteur des réponses Y par la matrice X^t comme le moyen de faire apparaître l'effet de chaque facteur séparément, alors que les influences ne sont pas visibles directement dans Y
- X^t caractérise l'emplacement des points expérimentaux, d'où l'importance de bien placer ces points.
- Les matrices X^t sont des matrices d'Hadamard, qui ont des propriétés mathématiques particulières :

$$X^{-1} = \frac{1}{n} X^t$$

Et on a :

$Y = X E$ (calcul des réponses à partir des effets)

$X^t X = n I$ (I = matrice identité)

- On peut démontrer que cette valeur de $X^t X$ est celle qui donne la meilleure précision possible

- Unités : la moyenne, les effets et les interactions sont exprimés dans les mêmes unités que celles utilisées pour mesurer les réponses.

3. Notion de plan optimal

Soit X la matrice d'expériences.
$X^t X$ = matrice d'information
$(X^t X)^{-1}$ = matrice de dispersion

Critères d'optimalité d'un plan d'expériences :

a) critère de la matrice unité : $X^t X = n I$ où n = nombre d'expériences est le meilleur critère, mais ne peut être satisfait que pour n = 4 p

b) déterminant maximal : dét (X^t X) le plus grand possible

c) Trace minimale : tr(X^t X) $^{-1}$ la plus faible possible (trace = somme des éléments de la diagonale principale)

d) L'élément maximal de la diagonale principale doit être le plus faible possible.

Remarques : ces critères ne sont pas incompatibles ; a) implique b), c) et d).
Si (X^t X)$^{-1}$ est diagonale, la matrice d'expériences est dite orthogonale

Il existe un très grand nombre de matrices d'expériences, pour des domaines cubiques ou sphériques, des modèles du 1er ou 2ème degré (Box-Behnken, Doehlert, …); pour choisir un P.E., on dispose de tables donnant les matrices d'expériences possibles en fonction du nombre d'expériences envisagé, de niveaux des variables, …

Références :

Introduction aux plans d'expériences, Jacques Goupy, Dunod, Paris, 2001,
Plans d'expériences pour surfaces de réponse, Jacques Goupy, Dunod, 1999,
Plans d'expériences, applications à l'entreprise, Jean-Jacques Droesbeke, Jeanne Fine, Gilbert Saporta Editeurs, Technip
Pratique des plans d'expériences, Paul Schimmerling, Jean-Claude Sisson, Ai Zaïdi, Tec&Doc

Chapitre III - Méthodes d'Approximation

1.Principe

Soit le problème d'optimisation suivant :

$$\min_{x} f(x)$$

avec les conditions :

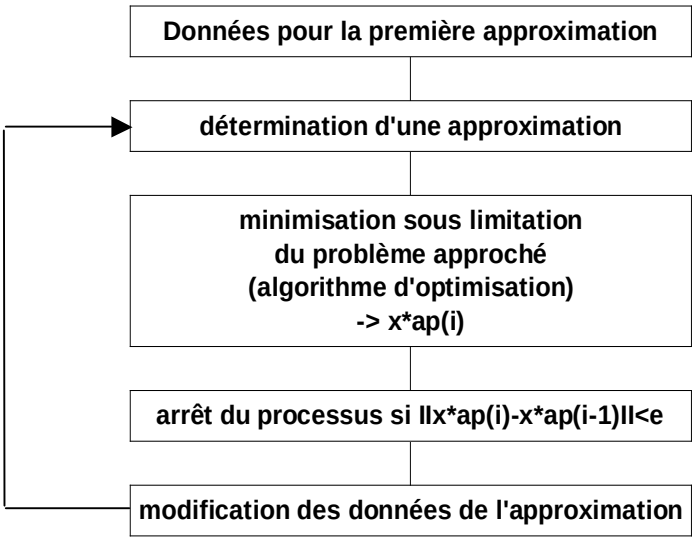
$$g_j(x) \leq 0 \qquad j = 1, \ldots, m$$

En mécanique, on n'a pas en général de relations explicites liant la fonction coût f et les limitations g_j aux paramètres x. Ces relations sont implicites et font intervenir des résolutions de systèmes algébriques issus de modélisations par éléments finis.

L'utilisation d'algorithmes d'optimisation nécessitant un processus itératif peut faire intervenir un grand nombre de ces résolutions. De plus, le processus itératif peut diverger, laissant l'utilisateur sans résultat significatif.

En fait, on traite souvent un problème d'optimisation en résolvant des approximations successives du problème, établies à partir d'un faible nombre de calculs de f et g_j. L'algorithme utilisé est alors le suivant :

Optimisation par approximations successives



Cette notion d'approximation est aussi connue sous le nom de "surfaces de réponse"; elle est largement utilisée dans le contexte des plans d'expériences considérés comme outils d'optimisation, lorsqu'il y a plus de 2 niveaux par facteur (approximations linéaires sinon);

De plus, les approximations permettent de relier analytiquement les fonctions aux paramètres ce qui permet :

- de résoudre des problèmes approchés de minimisation de manière plus robuste
- de capitaliser les connaissances relatives au problème étudié quand on utilise des approximations globales (surfaces de réponse)

Les approximations peuvent être de deux types : les approximations globales et les approximations locales.

2.Approximations locales

Soit le problème d'optimisation suivant :

Minimiser f(x)
x
avec les conditions :
$g_j(x) \leq 0$ j = 1, …, m

Une approximation locale de ce problème au point x_0 consiste à approcher la fonction coût f et les limitations g_j par des approximations de Taylor $f^{Ap}_{x_0}$ et $g^{Ap}_{jx_0}$ autour de ce point x_0. On minimise alors une suite de problèmes approchés :

Minimiser $f^{Ap}_{x_0}(x)$
x
avec les conditions :
$g^{Ap}_{jx_0}(x) \leq 0$ j = 1, …, m

On considère en général pour x0 la solution du problème approché précédent. On arrête le processus lorsque deux problèmes approchés successifs donnent des solutions très proches. Comme ces approximations ne sont valables que localement, on imposera en général à la solution du problème approché de rester dans un certain voisinage du point de départ (techniques de « région de confiance»), sinon des solutions successives peuvent osciller entre plusieurs points.

2.1 Linéarisation directe

On utilise un développement limité d'ordre 1 par rapport à x_i et on considère les approximations suivantes:

$$f^L(x) = f(x_0) + \sum_{i=1}^{n} (x_i - x_{0i}) \frac{\partial f}{\partial x_i}(x_0)$$

$$g_j^L(x) = g_j(x_0) + \sum_{i=1}^{n} (x_i - x_{0i}) \frac{\partial g_j}{\partial x_i}(x_0)$$

2.2 Linéarisation inverse (réciproque)

On utilise un développement limité d'ordre 1 par rapport à l'inverse des variables $y_i = \frac{1}{x_i}$ et on transforme le problème en considérant :

$$f^R(x) = f(x_0) + \sum_{i=1}^{n} (x_i - x_{0i}) \frac{x_{0i}}{x_i} \frac{\partial f}{\partial x_i}(x_0)$$

$$g_j^R(x) = g_j(x_0) + \sum_{i=1}^{n} (x_i - x_{0i}) \frac{x_{0i}}{x_i} \frac{\partial g_j}{\partial x_i}(x_0)$$

2.3 Linéarisation convexe

La linéarisation convexe consiste à linéariser les fonctions par rapport aux variables ou par rapport à leur inverse (choix fait variable par variable) de manière à obtenir un domaine admissible approché le plus « conservatif » possible (si possible inclus dans le domaine admissible réel, mais ce n'est pas systématiquement vérifié).

Dans le cas où les limitations imposées sont de la forme :

$$g_j(x) \leq 0 \qquad j = 1, \ldots, m$$

l'approximation convexe g_j^C est conservative par rapport aux deux précédentes si elle surestime la valeur de la contrainte, c'est-à-dire si on a :

$$g_j^C \geq g_j^L \text{ et } g_j^C \geq g_j^R$$

On prendra donc, pour chaque variable x_i, le terme le plus grand entre :

$$(x_i - x_{0i}) \frac{\partial g_j}{\partial x_i}(x_0) \qquad \text{(lin. directe)}$$

et :

$$(x_i - x_{0i}) \frac{x_{0i}}{x_i} \frac{\partial g_j}{\partial x_i}(x_0) \qquad \text{(lin. réciproque)}$$

Il faut donc choisir en fonction du signe de chaque terme de la somme:

$$g_j^L(x) - g_j^R(x) = \sum_{i=1}^{n} \frac{(x_i - x_{0i})^2}{x_i} \frac{\partial g_j}{\partial x_i}(x_0)$$

Si les paramètres sont positifs (c'est souvent le cas), on définit donc la linéarisation convexe par :

$$g_j^C(x) = g_j(x_0) + \sum_{i=1}^{n}(x_i - x_{0i})C_i \frac{\partial g_j}{\partial x_i}(x_0)$$

$$\text{où : } C_i = \begin{cases} 1 & \text{si } \dfrac{\partial g_j}{\partial x_i} \geq 0 \\ \dfrac{x_{0i}}{x_i} & \text{si } \dfrac{\partial g_j}{\partial x_i} \leq 0 \end{cases}$$

Remarque : si les contraintes sont de la forme :

$g_j(x) \geq 0 \qquad j = 1, \ldots, m$

et/ou si les x_i peuvent être ≤ 0, ce choix n'est plus valable.

2.4 Approximation quadratique

On utilise des développements de Taylor d'ordre 2 en x_0.

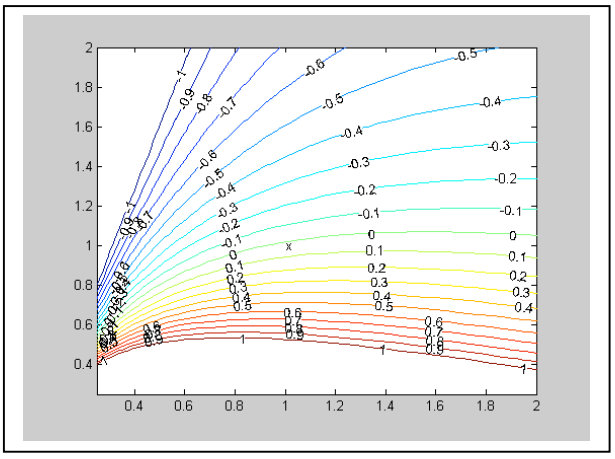
$$f^Q(x) = f(x_0) + \sum_{i=1}^{n}(x_i - x_{0i})\frac{\partial f}{\partial x_i}(x_0) + \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{n}\frac{\partial^2 f}{\partial x_i \partial x_k}(x_0)\ (x_i - x_{0i})(x_k - x_{0k})$$

$$g_j^Q(x) = g_j(x_0) + \sum_{i=1}^{n}(x_i - x_{0i})\frac{\partial g_j}{\partial x_i}(x_0) + \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{n}\frac{\partial^2 g_j}{\partial x_i \partial x_k}(x_0)\ (x_i - x_{0i})(x_k - x_{0k})$$

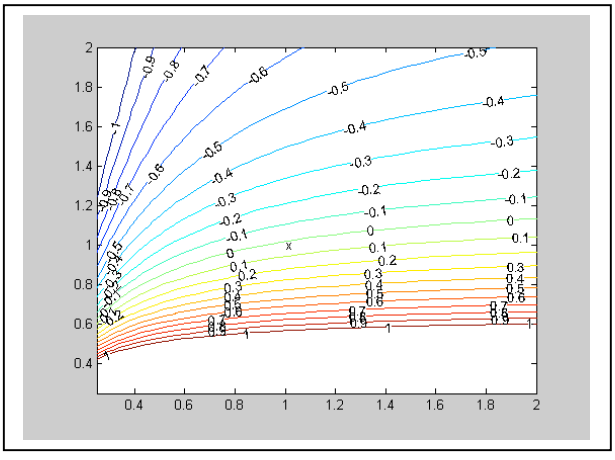
Exercice :

Calculer les approximations linéaires directe, réciproque, convexe et l'approximation quadratique de la fonction g au point (1,1)

$$g(x_1, x_2) = \frac{2}{x_2 + 0.25\,x_1} - \frac{\sqrt{3}}{3\,x_1} - 1$$



fonction g exacte



linéarisation convexe de g

3. Approximations globales

Les approximations globales relient les fonctions aux variables de conception en s'appuyant sur un quadrillage ou sur un ensemble de points répartis dans tout ou partie du domaine admissible. Ces approximations sont généralement d'autant meilleures ou plus sûres (critères de qualité à définir) que l'approximation est plus riche (dépendant d'un grand nombre de paramètres) et qu'elle est définie à partir d'un grand nombre de points représentatifs. Cependant, les approximations polynômiales de degré élevé présentent un risque d'oscillations ; en pratique, ce sont les degrés 2 ou 3 les plus utilisés.

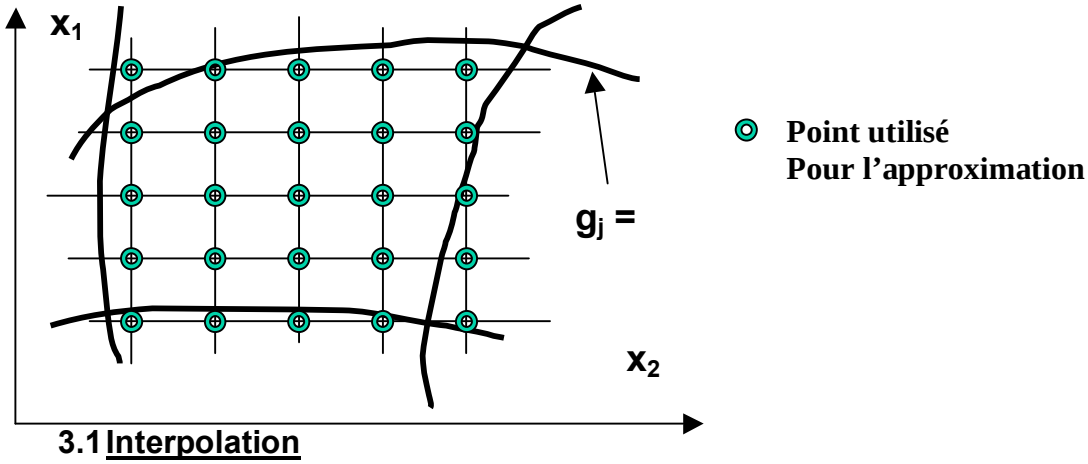
Supposons que l'on cherche à approcher la fonction $\mathbf{g(x_1, x_2)}$ (f ou l'un des g_j). On choisit d'approcher cette fonction sur l'espace de conception sous la forme suivante :

$$\mathbf{g_{ap}}\left(x_1, x_2\right) = \left\langle \mathbf{P}\left(x_1, x_2\right)\right\rangle \{\alpha\}$$

où : $\left\langle \mathrm{P}\left(x_1, x_2\right)\right\rangle$ est un vecteur $(1, \mathrm{N_p})$ définissant la forme de l'approximation (par exemple, base polynomiale)

$\{\alpha\}$ est le vecteur de paramètres $(\mathrm{N_p}, 1)$ de l'approximation

On peut, à ce niveau, différencier une approximation obtenue par interpolation de celle obtenue par minimisation d'écarts en moindres carrés.



3.1 Interpolation

Si le nombre de points utilisés pour définir l'approximation $\mathbf{N_{ap}}$ **est égal à** $\mathbf{N_p}$, alors on interpole la fonction g sur l'espace de conception par la formule :

$$\{\alpha\} = [\mathrm{Pn}]^{-1}\{\beta\}$$

où : $[\mathrm{P_n}] = \begin{bmatrix} \left\langle \mathbf{P}\left(x_1, y_1\right)\right\rangle \\ \left\langle \mathbf{P}\left(x_2, y_2\right)\right\rangle \\ \cdots \\ \left\langle \mathbf{P}\left(x_{N_{ap}}, y_{N_{ap}}\right)\right\rangle \end{bmatrix}, \{\beta\} = \begin{pmatrix} \mathbf{g}\left(x_1, y_1\right) \\ \mathbf{g}\left(x_2, y_2\right) \\ \cdots \\ \mathbf{g}\left(x_{N_{ap}}, y_{N_{ap}}\right) \end{pmatrix}$

Remarque : il faut que les points sont placés de telle sorte que la matrice Pn soit inversible

La solution s'écrit alors directement :

$$g_{ap}(x_1, x_2) = \langle P(x_1, x_2) \rangle [P_n]^{-1} \{\beta\}$$

3.2 Approximation par moindres carrés

Si le nombre de points utilisés pour définir l'approximation **N_{ap} est supérieur à N_p**, alors Pn est rectangulaire, et on approche la fonction g sur l'espace de conception en minimisant une fonction écart en moindres carrés définie par :

$$J(\{\alpha\}) = \frac{1}{2}([P_n]\{\alpha\} - \{\beta\})^T ([P_n]\{\alpha\} - \{\beta\})$$

La minimisation de cette fonction écart par rapport à {a} permet d'obtenir le vecteur de paramètres :

$$\{\alpha\} = ([A])^{-1}\{\gamma\} \ , \ [A] = [P_n]^T [P_n] \ , \{\gamma\} = [P_n]^T \{\beta\}$$

Exercice :

Déterminer une approximation de : $g(x_1, x_2) = \frac{2}{x_2 + 0.25 x_1} - \frac{\sqrt{3}}{3 x_1} - 1$ sous la forme approchée suivante :

$$g(x_1, x_2) = \langle 1, x_1, x_2 \rangle \{\alpha\}$$

à partir des points d'évaluation suivants: $(0.5, 0.5)$, $(0.5, 1.5)$, $(1.5, 0.5)$, $(1.5, 1.5)$.

Chapitre IV - Minimisation à une dimension

1. Introduction

Le cas particulier de la minimisation unidirectionnelle, ou unidimensionnelle est très particulier car il concerne deux situations :

- une seule variable est à optimiser : Min f(x) $x \in R$
- en dimension n, dans cas où la direction de descente s^k est connue:

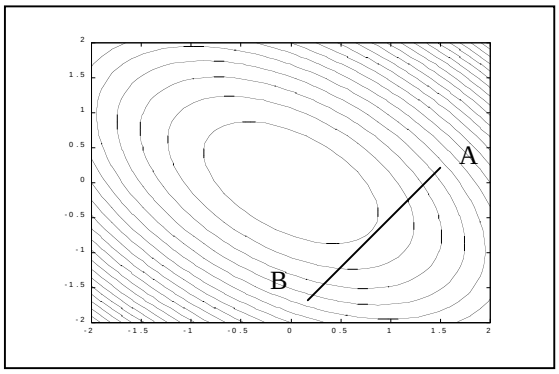
En effet, tout point dans la direction s^k s'écrit :

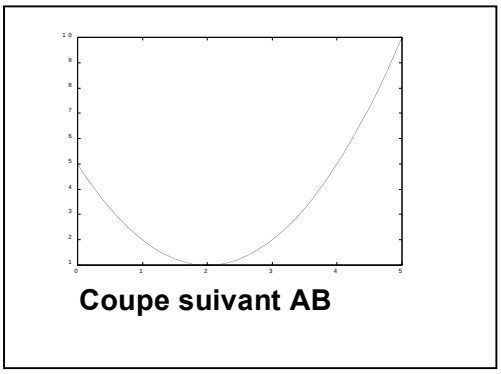
$$x = x^k + \alpha s^k$$

x^k , $s^k \in R^n$
et on peut définir la fonction $F(\alpha) = f(x^k + \alpha s^k)$ avec $\alpha > 0$, x^k et s^k connus.

La recherche du pas α^* qui minimise $F(\alpha)$ est un problème de minimisation à 1 dimension (recherche 1D, « linéaire »).

Représentation graphique dans le cas de 2 dimensions :





La minimisation 1D comporte généralement deux étapes :

a) Recherche d'un intervalle $[0, \overline{\alpha}]$ encadrant le minimum

b) Réduction de cet intervalle : plutôt que de minimiser f avec précision dans une direction, on réduit l'intervalle initial (α_g , α_d) jusqu'à une longueur finale donnée.

On aura, dans certains cas, besoin de supposer f **unimodale** pour justifier les algorithmes: cela signifie qu'elle est décroissante jusqu'à α^* , puis croissante :

$\alpha^1 < \alpha^2 < \alpha^* \Rightarrow f(\alpha^1) > f(\alpha^2)$
$\alpha^* < \alpha^1 < \alpha^2 \Rightarrow f(\alpha^1) < f(\alpha^2)$

f a alors 1 seul minimum, mais n'est pas forcément continue ni dérivable.

2. Méthode d'encadrement du minimum

Cette méthode suppose f unimodale.
Soit un point de départ α^0 donné. L'objectif est de trouver 3 points α^{k-2}, α^{k-1} et α^k tels que l'on ait :

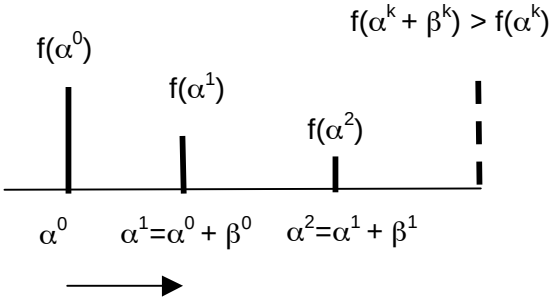
$$f(\alpha^{k-2}) > f(\alpha^{k-1}) \text{ et } f(\alpha^{k-1}) < f(\alpha^k)$$

On définit le paramètre d'expansion r > 1, le pas β^0 et t $(0 < t < 1/r)$

l'algorithme est le suivant :

a) calculer $f(\alpha^0)$ et $f(\alpha^0 + \beta^0)$

b) Si $f(\alpha^0 + \beta^0) > f(\alpha^0)$, aller en d)
Sinon,
soit $\alpha^1 = \alpha^0 + \beta^0$ et $\beta^1 = r\,\beta^0$
(expansion, $\beta^1 > \beta^0$)
Calculer $f(\alpha^1 + \beta^1)$



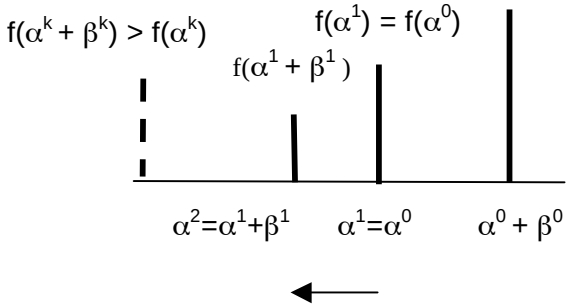
c) Si $f(\alpha^1 + \beta^1) < f(\alpha^1)$,
soit $\alpha^2 = \alpha^1 + \beta^1$ et $\beta^2 = r\,\beta^1$
et continuer à calculer α^k et β^k ainsi
jusqu'à ce que l'on ait:

$f(\alpha^k + \beta^k) > f(\alpha^k)$; aller alors en h).

d) on a $f(\alpha^0 + \beta^0) > f(\alpha^0)$
soit $\alpha^1 = \alpha^0$ et $\beta^1 = -\,t\,\beta^0$
Calculer $f(\alpha^1 + \beta^1)$

e) Si $f(\alpha^1 + \beta^1) > f(\alpha^1)$, aller en g)

sinon,
f) Soit $\alpha^2 = \alpha^1 + \beta^1$ et $\beta^2 = r\,\beta^1$
calculer α^k et β^k de cette manière
jusqu'à ce que l'on ait:



$f(\alpha^k + \beta^k) > f(\alpha^k)$; aller alors en h).

g) on a : $f(\alpha^1 + \beta^1) > f(\alpha^1)$
Le minimum est encadré par $(\alpha^0 - t\,\beta^0)$ et $(\alpha^0 + \beta^0)$; aller en i)

h) Les 3 derniers points vérifient :

$$f(\alpha^{k-2}) > f(\alpha^{k-1}) \text{ et } f(\alpha^{k-1}) < f(\alpha^k)$$

Le minimum est encadré par α^{k-2} et α^k

i) Réduire l'intervalle par une des méthodes suivantes ...

3. Méthode de dichotomie

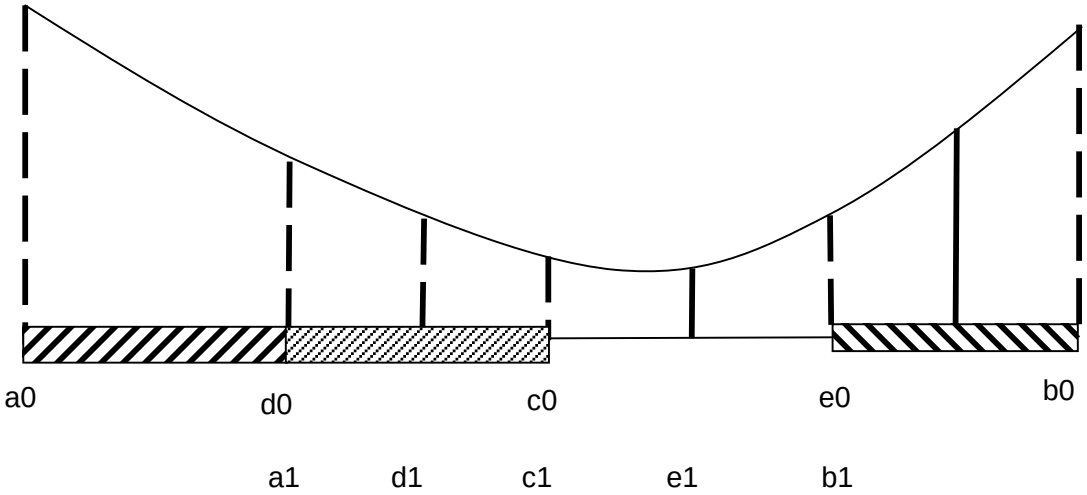
L'intervalle, (a°, b°) contenant le minimum est supposé connu.
Principe : diviser par 2, à chaque pas, la longueur de l'intervalle en calculant la fonction en 2 points.

- f unimodale, (a°, b°) donnés,

- calculer c°=(a°+b°)/2, d° milieu de [a°, c°], e° milieu de [c°, b°]

- Calculer f en ces 5 points. f unimodale, donc on peut toujours éliminer 2 des 4 sous-intervalles : on garde les 2 sous-intervalles entourant la valeur minimale parmi les 4.

- Calculer 2 nouveaux points (milieux des 2 sous-intervalles)
- Test d'arrêt : sur la longueur de l'intervalle, ou sur le nombre d'itérations (= sa réduction).



Sous-intervalles éliminés à la 1ère itération

Sous-intervalle éliminé à la 2ème itération

Il faut 17 calculs de f pour réduire l'intervalle de 100,
29 pour le réduire de 10000.

- il existe une version utilisant les dérivées

4. Méthode de la section d'or

Dans cette méthode, les intervalles successifs sont réduits dans un rapport constant qui permet de réutiliser 3 des 4 points de l'itération précédente.

Principe :

a) intervalle initial [0, c], soient a et b 2 points entre 0 et c

b) éliminer un des sous-intervalles [0, a] ou [b, c], suivant les valeurs de f

c) si longueur de l'intervalle < ε , fin

d) sinon, calculer la fonction en UN seul nouveau point qui permet de découper le nouvel intervalle (réduit) en 3 segments de manière analogue au découpage initial, et aller en b)

Pour que 3 points sur 4 soient réutilisés, a et b doivent donc être tels que :

symétrie : $c - b = a$

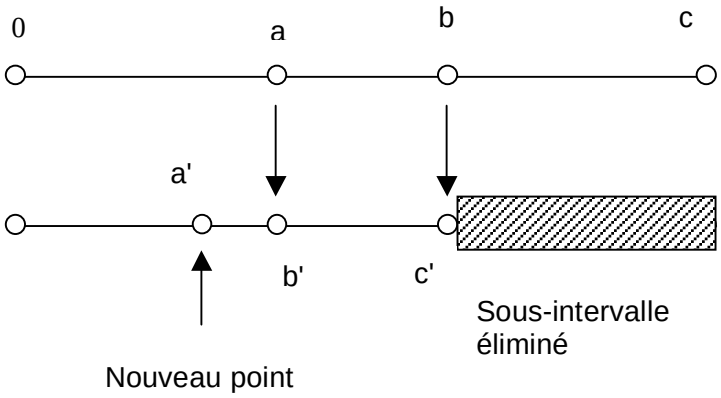
rapport constant : $(c- b)/ c = (b - a)/ b$

$(c- b)/ c = (b - (c-b)) / b$

$bc – b^2 – bc + c^2 - bc = 0$

$-1 + (c/b)^2 - (c/b) = 0$

$\Delta = 5$, et $c/b = (1 \pm \sqrt{5})/2$
$c/b > 0 \Rightarrow c/b = (1 + \sqrt{5})/2 = 1.618034..$= nombre d'or



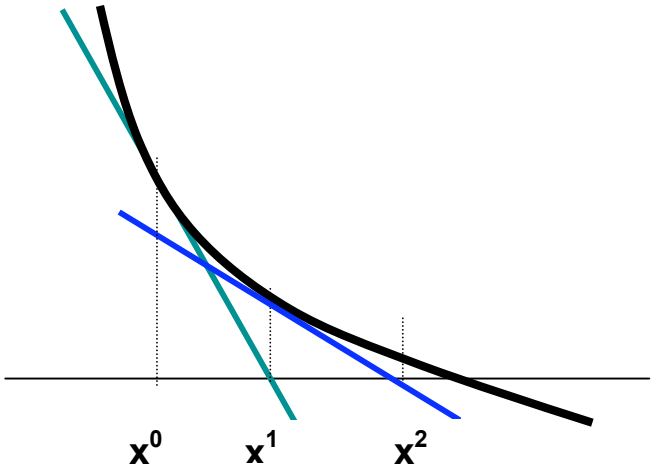
Il faut 13 calculs de f pour réduire l'intervalle initial par 100,
22 pour le réduire par 10000

5. Méthodes de Newton Raphson et de la sécante

C'est une méthode très générale et très utilisée pour la résolution d'équations non linéaires. Pour trouver le minimum de f(x), on l'applique à la recherche d'un point stationnaire de f c'est-à-dire à la résolution de: $f'(x^*) = 0$

La méthode consiste à linéariser f' :
Au point x^k, on considère l'approximation:

$$f'(x) \cong f'(x^k) + (x - x^k)\, f''(x^k) = f_1(x)$$



et on prend comme prochain point l'intersection de f_1 avec l'axe des x :

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

- Taux de convergence d'ordre 2 : Lim $\dfrac{\left|x^{k+1} - x^*\right|}{\left|x^k - x^*\right|^2} = \beta$ $(\beta > 0)$ $k \to \infty$

- La méthode revient à **supposer f quadratique,** il y a donc convergence en **un seul pas si f est quadratique**, mais :

- il faut f' et la **dérivée seconde** de f, et il faut $f''(x^k) > 0$

- et **pas de convergence globale**

Variante : la méthode de la sécante permet d'éviter le calcul de la dérivée seconde de f ; elle est analogue à la méthode de Newton, mais on remplace f'' par une approximation du type différences finies:

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

Interpolation polynomiale

(cf chapitre sur les méthodes d'approximations)
Le principe est d'approcher f par une fonction dont on sait calculer facilement le minimum, par exemple une parabole ou une cubique. Pour déterminer les coefficients du polynôme, on utilisera plusieurs valeurs de f, ou bien des valeurs de f et de sa dérivée, même de sa dérivée seconde si on peut la calculer. On itérera sur l'approximation polynômiale, en remplaçant par exemple un des points par le dernier minimum estimé.

- Approximation quadratique : on doit calculer les coefficients a, b, c tels que :
  $f_q(x) = a\,x^2 + b\,x + c$

soit une approximation de f(x) ; il faut donc 3 données : la valeur de f en 3 points, ou f en 2 points et sa dérivée en 1 point.

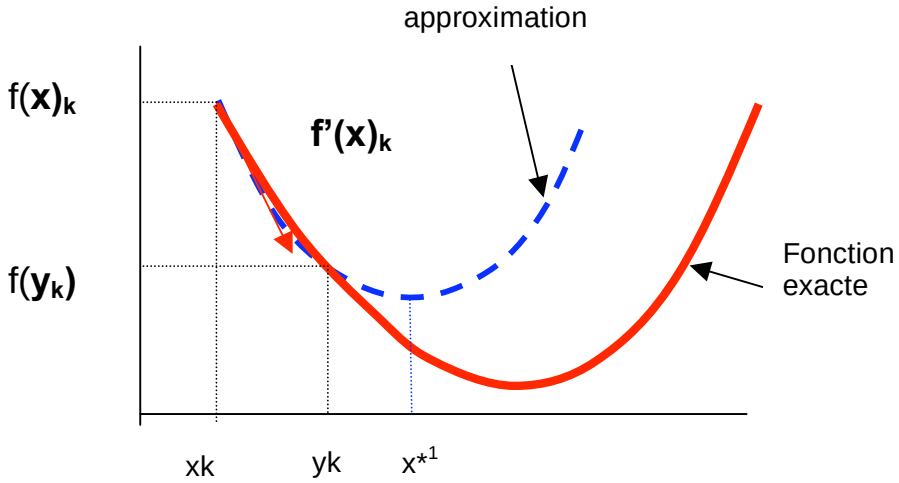
Soient x_k et y_k 2 points de l'intervalle contenant le minimum, ou bien défini par les bornes sur les variables ; on calcule par exemple $f(x_k)$, $f'(x_k)$, $f(y_k)$. On a alors :

$$f_q(x_k) = a\,x_k^2 + b\,x_k + c = f(x_k)$$

$$f_q(y_k) = a\,y_k^2 + b\,y_k + c = f(y_k)$$

$$f'_q(x_k) = 2\,a\,x_k + b = f'(x_k)$$

d'où a, b, c, et si f est convexe, le minimum de f_q est $x^{*1} = -b/2a$
Ce nouveau point est ensuite utilisé pour refaire une interpolation ; on calcule $f(x^{*1})$, (éventuellement $f'(x^{*1})$) et on fait une nouvelle approximation en utilisant 3 des quantités connues.



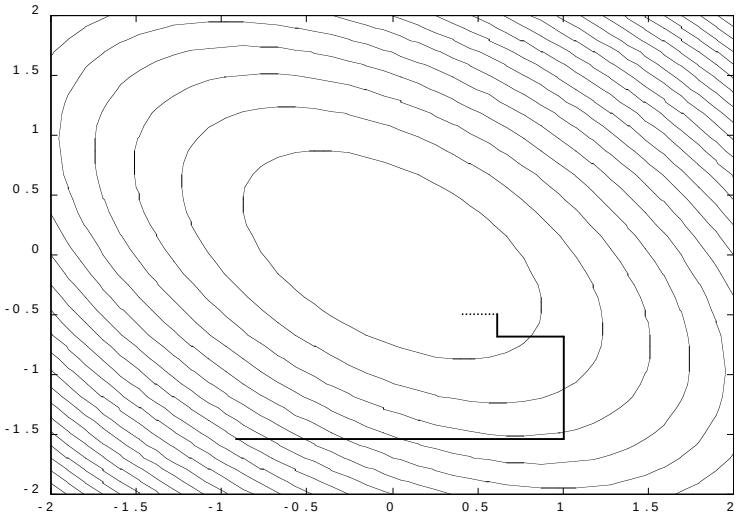
- approximation cubique : de manière analogue au cas quadratique, les coefficients du polynôme sont calculés à partir de f en 4 points , ou de f et son gradient en 2 points.

- si on veut prendre en compte la non-convexité de f dans l'intervalle initial, il existe des méthodes plus sophistiquées

Chapitre V – Minimisation locale sans limitations

1. Méthode de relaxation

On optimise une seule variable à la fois, les autres restant fixes ⇒ ne nécessite qu'une méthode 1D, mais pas très efficace.



2. Méthode du simplexe (Nelder-Mead)

C'est une méthode qui n'utilise pas les dérivées de la fonction. (méthode d'ordre 0).
En dimension n, un simplexe est un polygone à n + 1 sommets (en 2D : triangle)
Le principe est de déplacer dans le domaine un simplexe en remplaçant itérativement le plus mauvais point, (c'est-à-dire le point du simplexe où f atteint sa plus grande valeur), par un point meilleur.

Il y a 3 étapes : réflexion, contraction, expansion. Dans ses grandes lignes, l'algorithme est le suivant :

a) Soit un simplexe : $x_0, x_1, \ldots x_n$, qui peut être construit au départ en prenant par exemple :

$$\mathbf{x_j = x_0 + p\, e_j + \sum_{\substack{k=1 \\ k\neq j}}^{n} q\, e_k} \qquad j = 1, \ldots, n$$

avec : x_0: point initial

$$p = \frac{a}{n\sqrt{2}}(\sqrt{n+1} + n - 1) \qquad q = \frac{a}{n\sqrt{2}}(\sqrt{n+1} - 1)$$

e_k: vecteur de base dans la direction k

b) soit : x_s le point du simplexe où f est maximale $f(x_s) = f_s$

x_g celui où elle prend sa 2ème plus grande valeur

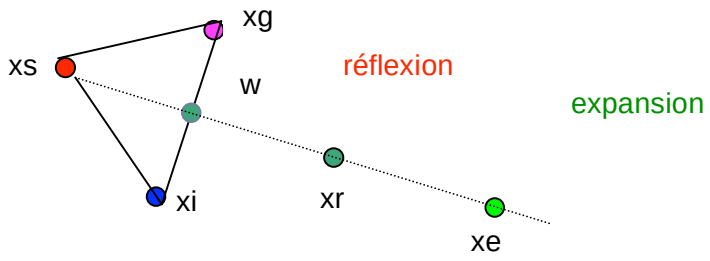
x_i celui où elle est minimale $f(x_i) = f_i$

w : barycentre des points x_j pour $j \neq s$ (sauf le max)

c) réflexion (s'éloigner de x_s): créer un point x_r sur la droite x_s w :

$x_r = w + \alpha(w - x_s)$ avec $\alpha > 0$ (α=1)

calculer $f(x_r) = f_r$



d) si $f_i < f_r < f_g$ remplacer x_s par x_r et aller en b) (i, g, r)

e) si $f_r < f_i$ (x_r meilleur que le min actuel) : essayer d'aller plus loin (expansion) en calculant :

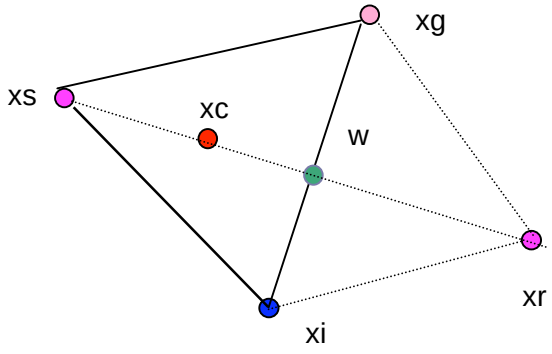
$x_e = w + \beta(x_r - w)$ avec $\beta > 0$ (β=2)

f) si $f(x_e) < f_r$ remplacer x_s par x_e et aller en b) avec (i, g, e)

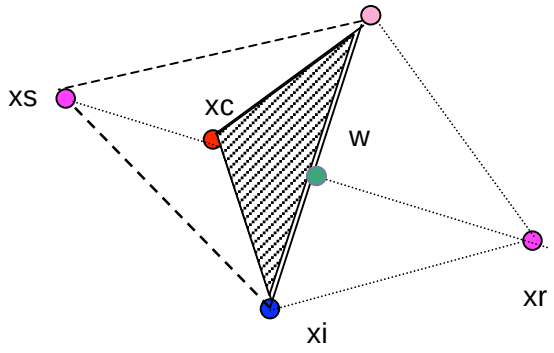
g) sinon ($f(x_e) \geq f_r$) remplacer x_s par x_r et aller en b) avec (i, g, r)

h) si $f_g < f_r < f_s$ (x_r pas meilleur que x_g), se rapprocher de x_i (contraction): en calculant x_c:

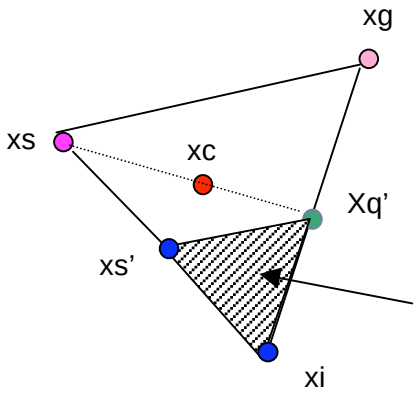
$x_c = w + \gamma (w - x_s)$ avec $0 < \gamma < 1$ (γ=0.5)



i) si $f_c \leq f_s$, remplacer x_s par x_c et aller en b) avec (c, g, i)



j) si $f_c > f_s$ (x_c encore moins bon que le pire des points, se rapprocher du meilleur): remplacer tous les points par : $x_j' = x_i + 0.5\ (x_i - x_j)$ et aller en b) avec le simplexe de taille réduite (s', i, q')



Nouveau simplexe
après contraction,

Critère de convergence :

$$\left\{ \frac{1}{1+n} \sum_{j=0}^{n} (fj - f(w))^2 \right\}^{\frac{1}{2}} < \varepsilon$$

(Arrêt lorsque toutes les valeurs de la fonction sur les points du simplexe sont proches en moyenne)

C'est une méthode assez efficace et robuste lorsqu'il y a peu de variables (< 10), à utiliser lorsqu'on ne dispose pas des dérivées de f. Il faut vérifier que la qualité géométrique des simplexes ne se dégrade pas.

3. Méthode de la plus forte pente

Nécessite les dérivées 1$^{\text{ères}}$

Cette méthode est attribuée à Cauchy (1847) ; elle utilise la propriété du gradient : en tout point x, $-\nabla f(x)$ est la direction de plus grande descente, et suit le schéma général des algorithmes de minimisation: construire une suite de points $x^1, x^2, \ldots, x^k$,
tels que $f(x^{k+1}) < f(x^k)$
et qui converge vers un point stationnaire de f

La méthode de la plus forte pente consiste à prendre : $x^{k+1} = x^k - \alpha^k \nabla f(x^k)$
avec α = pas optimal obtenu par une recherche 1d
Le test d'arrêt pourra porter sur :

- la variation de f : $|f(x^{k+1}) - f(x^k)| < \varepsilon_f$

- ou la variation des variables x : $||x^{k+1} - x^k|| < \varepsilon_x$

- ou la: norme du gradient : $||\nabla f(x^{k+1})|| < \varepsilon_x$

Le taux de convergence est linéaire ; c'est une méthode peu efficace car il y a un risque de « zig-zag » sur les fonctions mal conditionnées; en effet :

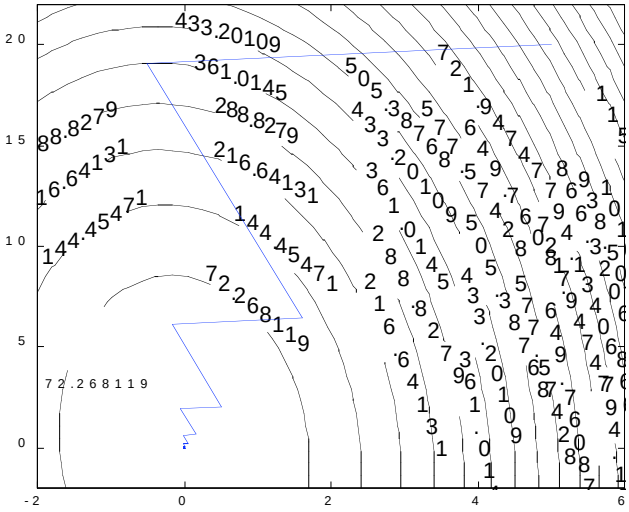
α^k = pas optimal, donc $F(\alpha) = f(x^k - \alpha \nabla f(x^k))$ est minimale en α^k

$F'(\alpha^k) = 0 = \nabla f(x^k)^T \nabla f(x^k - \alpha \nabla f(x^k)) = \nabla f(x^k)^T \nabla f(x^{k+1})$

⇒ avec le gradient à pas optimal, 2 directions consécutives sont orthogonales, donc la suite des points (x^k) oscille si f mal conditionnée (de type « vallée étroite »)

Exemple : le graphique ci-dessous représente les itérés obtenus avec la méthode de plus forte pente sur la fonction :

$f(x) = \frac{1}{2} x^T A x$ avec $A = \begin{bmatrix} 50 & 1 \\ 1 & 2 \end{bmatrix}$



Orthogonalité des directions successives : attention, les axes du graphique ne sont pas orthonormés

4. Méthodes du gradient conjugué et de Fletcher-Reeves

Caractéristique des méthodes de **directions conjuguées** : elles convergeront en **n itérations** sur des **fonctions quadratiques de n variables**. Pour une fonction quelconque (non quadratique), ceci signifie qu'au voisinage du minimum la convergence sera rapide, car on sait (Taylor) qu'on peut approximer localement toute fonction par une fonction quadratique.

Définition de directions conjuguées:

Soit A une: matrice symétrique définie positive ; on dit que s^0 , s^1 , …, $s^{n-1} \in R^n$ sont **conjuguées** par rapport à A si on a :

s^{iT} A s^k = 0 pour tous i, k, i ≠ k (si A = I, c'est l'orthogonalité)

Algorithme du gradient conjugué :

Soit $q(x)= ½ x^T A x + b^T x + c$ une fonction quadratique à minimiser, avec
A : matrice symétrique définie positive, $b \in R^n$, $c \in R$

a) k=0, point initial x^0 , $g^0 = \nabla f(x^0)=A x^0 + b$, $s^0 = - g^0$

b) Calcul du pas optimal dans la direction s^k : $\alpha^k = - \dfrac{s^{kT} g^k}{s^{kT} A s^k}$

c) Point suivant : $x^{k+1} = x^k + \alpha^k s^k$

d) Gradient et direction suivants : $g^{k+1} = \nabla f(x^{k+1})$, $\beta^k = \dfrac{g^{k+1T} A s^k}{s^{kT} A s^k}$

$s^{k+1} = -g^{k+1} + \beta^k s^k$

e) k=k+1 ; si k<n aller en b), sinon fin

Justification du calcul de α^k: dans la direction s^k on a : $F(\alpha) = f(x^k + \alpha s^k)$
Le pas optimal est cherché comme solution de $F'(\alpha) = 0$, soit :

$F'(\alpha) = s^{kT} \nabla f(x^k + \alpha s^k) = s^{kT} \{A (x^k + \alpha s^k) + b\} = s^{kT} \{(A x^k + b) + \alpha A s^k\}$
$= s^{kT} \{ \nabla f(x^k) + \alpha A s^k \}$

$F'(\alpha) = 0$ donne donc bien: $\alpha^k = - \dfrac{s^{kT} \nabla f(x^k)}{s^{kT} A s^k}$

La méthode de **Fletcher-Reeves** est une adaptation du gradient conjugué pour des fonctions **non quadratiques**. Les différences sont les suivantes :

- Calcul du pas optimal α^k par recherche 1D

- Calcul de β : $\beta^k = ||\nabla f(x^{k+1})||^2 / ||\nabla f(x^k)||^2$

- Test d'arrêt nécessaire, puisqu'on n'a plus la propriété de convergence en n itérations.

Une autre variante, la méthode de Polak-Ribière utilise :

$$\beta^k = (\nabla f(x^{k+1}) - \nabla f(x^k))^T \nabla f(x^{k+1}) \ / \ || \nabla f(x^k) ||^2$$

5. Méthode de Newton

C'est une méthode d'ordre 2 (nécessite les dérivées secondes de la fonction à minimiser), généralisation à n dimensions de la méthode de Newton-Raphson.
En un point x^k donné, on calcule le prochain point ($x^{k+1} = x^k+s$) en utilisant l'approximation quadratique locale de f au point x^k :

$$f(x^k+s) \cong f(x^k) + s^T \nabla f(x^k) + \tfrac{1}{2} s^T H(x^k)\, s = q(s)$$

avec $H(x) = \nabla^2 f(x)$

On prend comme point suivant le minimum de q(s) , qui est défini par:

$$\nabla q(s) = \nabla f(x^k) + H(x^k)\, s = 0$$

soit : $s^k = - (H(x^k))^{-1} \nabla f(x^k)$

La méthode de Newton consiste donc à itérer sur la formule de mise à jour suivante :

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

Elle converge en un seul pas si f est quadratique, mais présente plusieurs inconvénients :

Risque de divergence (si non convexité, $H(x^k)$ n'est pas définie positive)

Risque de convergence vers un point stationnaire qui n'est pas un minimum si $H(x^k)$ n'est pas définie positive (maximum local, point de selle)

Pour tenter d'éviter ces problèmes, on peut ajouter une étape de minimisation unidirectionnelle ; s^k est alors considéré comme une direction de descente; on doit faire une recherche unidirectionnelle, et on a la formule de Newton plus générale ::

$$x^{k+1} = x^k - \alpha^k (H(x^k))^{-1} \nabla f(x^k)$$

nécessité de calculer les **dérivées secondes** et d'inverser le Hessien

Les méthodes de **quasi-Newton** (ou « métrique variable ») vont permettre de remédier à ces difficultés.

6. Méthodes de métrique variable ou de quasi-Newton

Leur principe est de remplacer, dans la formule de la méthode de Newton, $(H(x^k))^{-1}$ par une matrice B^k calculée à chaque itération de telle sorte que la suite (B^k) converge vers $(H(x^k))^{-1}$. Puisque l'on n'a plus le Hessien exact, il est cette fois indispensable d'utiliser une recherche unidirectionnelle :

Newton : $x^{k+1} = x^k - \alpha^k (H(x^k))^{-1} \nabla f(x^k)$

Quasi-Newton : $x^{k+1} = x^k - \alpha^k B^k \nabla f(x^k)$

Il reste à définir comment construire la suite (B^k). Plusieurs algorithmes existent pour cela. On utilise une mise à jour du type :

$B^{k+1} = B^k + \Delta^k$ avec Δ^k telle que B^{k+1} soit symétrique définie positive.

et on impose la condition, dite équation de quasi-Newton :

$B^{k+1} (\nabla f(x^{k+1}) - \nabla f(x^k)) = x^{k+1} - x^k$

notée : $B^{k+1} \gamma^k = \delta^k$ avec :

$\delta^k = x^{k+1} - x^k$ et $\gamma^k = \nabla f(x^{k+1}) - \nabla f(x^k)$

Remarque : cf. l' approximation de f'' par différences finies en dimension 1, ou méthode de la sécante :

$f''(x) \cong (f'(y) - f'(x)) / (y - x)$; ici B^k doit approcher $(H(x^k))^{-1}$.

L'algorithme d'une méthode de quasi-Newton se déroule selon les étapes suivantes:

a) k=0 ; soient x° point initial, B° une matrice symétrique définie positive (par exemple la matrice identité)

b) Calculer la direction de descente : $s^k = - B^k \nabla f(x^k)$

c) Recherche unidirectionnelle : calculer α^k minimisant $f(x^k + \alpha^k s^k)$

d) Calculer :

le point suivant	$x^{k+1} = x^k + \alpha^k s^k$
le gradient en ce point :	$\nabla f(x^{k+1})$
la variation de x :	$\delta^k = x^{k+1} - x^k$
la variation du gradient :	$\gamma^k = \nabla f(x^{k+1}) - \nabla f(x^k)$

e) Calculer B^{k+1} : ce calcul diffère selon l'algorithme choisi ; nous donnons les formules correspondant aux deux méthodes qui sont le plus couramment utilisées :

♦ la méthode de **Davidon-Fletcher-Powell** :

$$B^{k+1} = B^k + \frac{\delta^k \delta^{kT}}{\delta^{kT} \gamma^k} - \frac{B^K \gamma^k \gamma^{kT} B^K}{\gamma^{kT} B^k \gamma^k}$$

♦ la méthode **BFGS** (Broyden-Fletcher-Goldfarb-Shanno)

$$B^{k+1} = B^k + \left[1 + \frac{\gamma^{k} B^{k} \gamma^{k}}{\delta^{kT} \gamma^{k}}\right] - \frac{\delta^{k} \gamma^{kT} B^{k} + B^{k} \gamma^{k} \delta^{kT}}{\delta^{kT} \gamma^{k}}$$

f) k = k + 1 ; test d'arrêt : fin ou aller en b)

Remarques:

Il faut procéder à des réinitialisations de la matrice B^k périodiquement pour assurer la convergence globale.

BFGS est moins sensible que DFP aux imprécisions de la recherche unidirectionnelle. Elle a un taux de convergence superlinéaire, et est souvent considérée comme une des méthodes les plus efficaces. BFGS est aussi utilisée pour résoudre des systèmes d'équations.

Chapitre VI - Minimisation locale avec limitations

1. Introduction

Problème posé : Minimiser f(x)
avec : $g_j(x) \leq 0$ j=1, ..., m (Q)
$h_l(x) = 0$ l=1, ..., p
$x = (x_1, ..., x_n)$
ici, on suppose :

x : variables réelles, continues
g_j et h_l : *contraintes*, *ou limitations*, *ou restrictions*,
continues, (différentiables, convexes ?)
égalités et/ou inégalités (égalités seulement ?)

Point admissible : qui vérifie les limitations

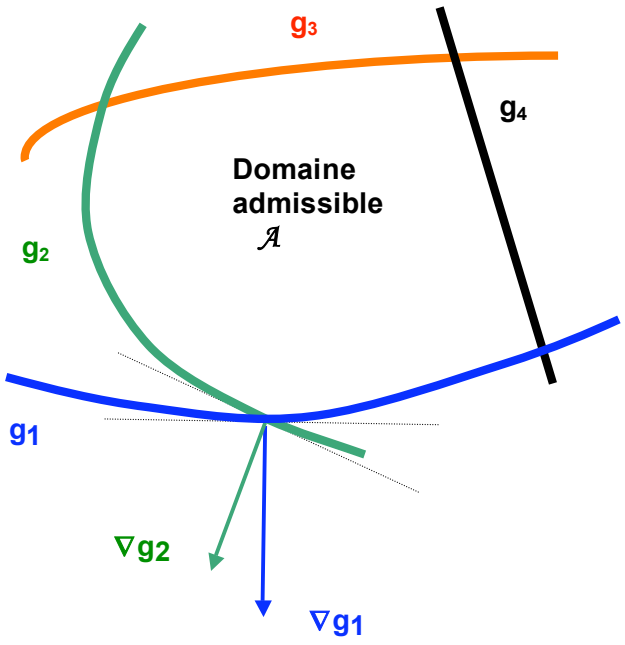
domaine *admissible* : ensemble $\mathcal{A}$ de tous les points x où toutes les limitations sont satisfaites (convexe ?).
(domaine *non admissible* : ensemble de tous les points x où au moins une des limitations n'est pas satisfaite)

(Q) s'écrit aussi : Minimiser f(x)
$x \in \mathcal{A}$

Généralement, la solution x* se trouve sur la frontière du domaine admissible (sinon les limitations sont inutiles, problème mal posé …).

Si une limitation d'inégalité est nulle en un point x elle est dite *active* ou *saturée* en x.

Une des difficultés, d'un point de vue algorithmique, est que l'on ne sait pas a priori quelles limitations seront actives à l'optimum.



Cas particuliers (résultats théoriques, méthodes associées):

f, g_j et h_l linéaires
f, g_j et h_l linéaires ou quadratiques
f quelconque, g_j et h_l linéaires
f, g_j et h_l convexes

Dans le cas général : recherche d'un minimum local

2. Résultats théoriques

Il faut déjà définir le cadre dans lequel des conditions d'optimalité peuvent être établies :

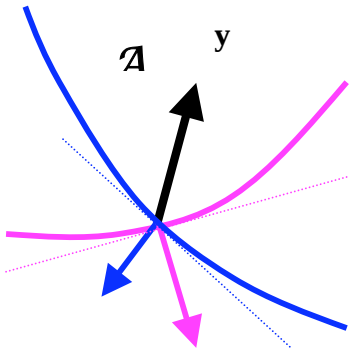
2.1 Qualification des limitations

Définitions : soit x situé sur la frontière de $\mathcal{A}$. Un *arc de courbe* est *admissible* en x si il part de x et est contenu dans $\mathcal{A}$ au voisinage de x.

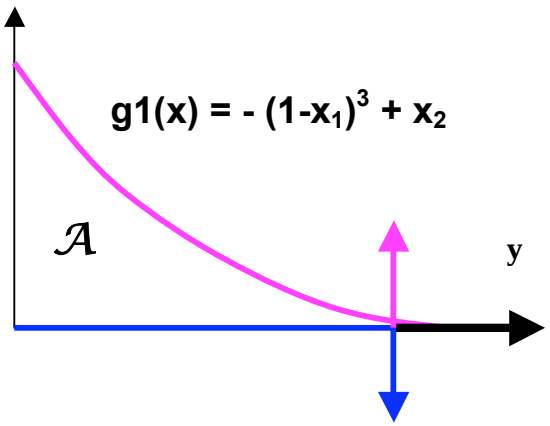
Une direction admissible est un vecteur tangent à un arc de courbe admissible en x. (on peut se déplacer sur ce vecteur en restant dans le domaine admissible, au moins dans un voisinage de x)

Propriété :
Toute direction y admissible en x° vérifie : $\nabla g_j(x°)^T . y \leq 0$
pour toute limitation g_j saturée en x° (parce que g_j doit être ≤ 0)

Mais la réciproque n'est pas toujours vraie (voir figure).



Direction y admissible
$\nabla g_j(x°)^T . y \leq 0$



Réciproque fausse :
on a : $\nabla g_j(x)=(3(1-x_1)^2, 1)$

En x0=(1, 0) : $\nabla g_j(x°)^T . y \leq 0$ (= 0) avec le vecteur y=(1, 0)
et pourtant : y non admissible (Contraintes non qualifiées)

Si la réciproque est vraie, c'est-à-dire si :

y est telle que : $\nabla g_j(x°)^T . y \leq 0$ pour toute g_j saturée en x°

est équivalent à :

y est une direction admissible

on dit alors que A vérifie l'hypothèse de qualification des contraintes (Q.C.) en x° (les limitations sont *qualifiées* en x°)

Exemples d'ensembles admissibles vérifiant (QC) :

- **quand les fonctions g_j sont toutes linéaires,**
- **quand les gradients des contraintes actives sont linéairement indépendants,**
- **quand toutes les fonctions s g_j sont convexes (et que A n'est pas vide)**
- **...**

Dans les autres cas, on ne peut pas vérifier (QC) , et on ne connaît pas de conditions d'optimalité (donc on n'a pas de garantie de convergence des méthodes basées sur ces conditions)

2.2 Condition nécessaire d'optimalité (Karush, Kuhn et Tucker)

2.2.a cas des contraintes d'égalité (fonction à 2 variables)

Minimiser $f(x_1, x_2)$
$h(x_1, x_2)=0$ (Qe)

La contrainte d'égalité exprime une relation implicite entre x_1 et x_2 que l'on peut écrire sous la forme explicite (même si elle n'est pas à déterminer) $x_2=\theta(x_1)$.
Le problème revient alors à déterminer le minimum de la fonction d'une seule variable ϕ définie par : $\phi(x_1)= f(x_1, \theta(x_1))$. Une condition nécessaire pour obtenir ce minimum est :

$$\frac{d\phi}{dx_1}(x_1)=\frac{\partial f}{\partial x_1}(x_1,\theta(x_1))+\frac{\partial f}{\partial x_2}(x_1,\theta(x_1))\frac{d\theta}{dx_1}(x_1)=0$$

D'autre part, la relation d'égalité (Qe) implique :

$$h(x_1,\theta(x_1))=0 \quad \rightarrow \quad \frac{\partial h}{\partial x_1}(x_1,\theta(x_1))+\frac{\partial h}{\partial x_2}(x_1,\theta(x_1))\frac{d\theta}{dx_1}(x_1)=0$$

Des deux dernières équations, on tire :

$$\frac{\partial f}{\partial x_1}(x_1,\theta(x_1))+\frac{\partial f}{\partial x_2}(x_1,\theta(x_1))\frac{d\theta}{dx_1}(x_1)+\lambda\left(\frac{\partial h}{\partial x_1}(x_1,\theta(x_1))+\frac{\partial h}{\partial x_2}(x_1,\theta(x_1))\frac{d\theta}{dx_1}(x_1)\right)=0$$

$\forall\lambda\in\Re$

En regroupant différemment on peut écrire :

$$\frac{\partial f}{\partial x_1}(x_1,\theta(x_1))+\lambda\frac{\partial h}{\partial x_1}(x_1,\theta(x_1))+\frac{d\theta}{dx_1}(x_1)\left(\frac{\partial f}{\partial x_2}(x_1,\theta(x_1))+\lambda\frac{\partial h}{\partial x_2}(x_1,\theta(x_1))\right)=0$$

$\forall\lambda\in\Re$

Cette relation est vérifiée pour tous les extremums liés de la fonction f et quelque soit λ.¨
La valeur particulière de λ qui est telle que le système suivant est vérifié représente le multiplicateur de Lagrange associé à la limitation $h(x_1, x_2)=0$:

$$\begin{cases} \dfrac{\partial f}{\partial x_1}\left(x_1,\theta(x_1)\right)+\lambda\dfrac{\partial h}{\partial x_1}\left(x_1,\theta(x_1)\right)=0 \\ \dfrac{\partial f}{\partial x_2}\left(x_1,\theta(x_1)\right)+\lambda\dfrac{\partial h}{\partial x_2}\left(x_1,\theta(x_1)\right)=0 \end{cases}$$

La résolution du problème (P) revient ainsi à déterminer (x_1^*, x_2^*, λ) tels que :

$$\begin{cases} \dfrac{\partial f}{\partial x_1}\left(x_1^*,x_2^*\right)+\lambda\dfrac{\partial h}{\partial x_1}\left(x_1^*,x_2^*\right)=0 \\ \dfrac{\partial f}{\partial x_2}\left(x_1^*,x_2^*\right)+\lambda\dfrac{\partial h}{\partial x_2}\left(x_1^*,x_2^*\right)=0 \\ h\left(x_1^*,x_2^*\right)=0 \end{cases}$$

On introduit la fonction de Lagrange : $L(x_1, x_2, \lambda) = f(x_1, x_2) + \lambda\, h(x_1, x_2)$

Les relations précédentes correspondent aux conditions de stationnarité du Lagrangien L :

$$\begin{cases} \dfrac{\partial L}{\partial x_1}\left(x_1^*,x_2^*,\lambda\right)=0 \\ \dfrac{\partial L}{\partial x_2}\left(x_1^*,x_2^*,\lambda\right)=0 \\ \dfrac{\partial L}{\partial \lambda}\left(x_1^*,x_2^*,\lambda\right)=0 \end{cases}$$

Ces conditions de Lagrange caractérisent un point stationnaire (minimum, maximum, ou point de selle) de f vérifiant les égalités :

$$h_l(x) = 0 \qquad l = 1, p$$

Elles sont nécessaires mais ne sont pas suffisantes. En effet, elles peuvent correspondre soit à un minimum, à un maximum ou bien encore à un point-selle.

Exemple 2.2.1 :

Minimiser $2x_1 + x_2$
Avec $h(x) = x_1^2 + x_2^2 - 1 = 0$

$L(x, \lambda) = 2x_1 + x_2 + \lambda(x_1^2 + x_2^2 - 1)$

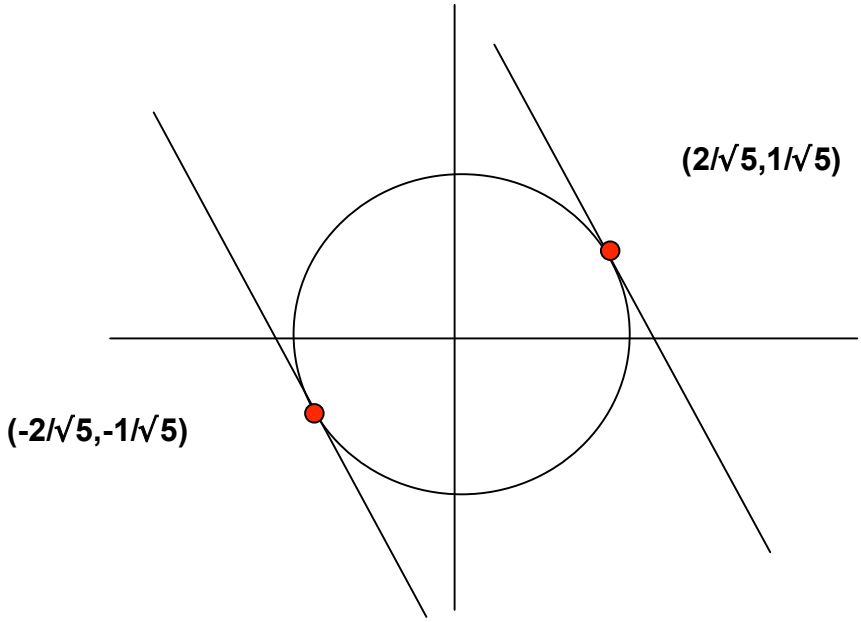
$\dfrac{\partial L}{\partial x_i} = 0$ **donne :**

$2 + 2\lambda x_1 = 0$
$1 + 2\lambda x_2 = 0$

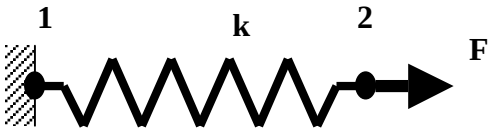
$x_1 = -1/\lambda$
$x_2 = -1/(2\lambda)$

et $h(x) = 0 \Rightarrow (1/\lambda)^2 + 1/(2\lambda)^2 = 0$
$5 = 4\lambda^2$
$\lambda = \pm\sqrt{5}/2$

2 points vérifient les conditions de Lagrange, mais un seul est minimum : ces conditions sont nécessaires, mais pas suffisantes pour qu'un point soit optimum.



<u>Exemple 2.2.2</u> . Prise en compte des conditions aux limites à l'aide des multiplicateurs de Lagrange :
a. On considère le cas élémentaire de l'équilibre d'un ressort de rigidité k dont l'une des extrémités est fixée, une force F étant appliquée à l'autre extrémité.



La solution en déplacement du système est obtenue en recherchant le minimum lié suivant :

$$\underset{u_1,u_2}{\text{Min}}\ E(u_1,u_2)=\frac{1}{2}\langle u_1,u_2\rangle\begin{bmatrix}k & -k\\ -k & k\end{bmatrix}\begin{pmatrix}u_1\\ u_2\end{pmatrix}-Fu_2 \quad \text{(Energie potentielle totale)}$$

$u_1=0$ **(contrainte d'égalité)**
Le Lagrangien associé s'écrit :

$$L(u_1,u_2,\lambda)=\frac{1}{2}\langle u_1,u_2\rangle\begin{bmatrix}k & -k\\ -k & k\end{bmatrix}\begin{pmatrix}u_1\\ u_2\end{pmatrix}-Fu_2+\lambda(u_1-0)$$

La stationnarité du Lagrangien induit la résolution du système suivant :

$$\begin{bmatrix}k & -k & 1\\ -k & k & 0\\ 1 & 0 & 0\end{bmatrix}\begin{pmatrix}u_1\\ u_2\\ \lambda\end{pmatrix}=\begin{pmatrix}0\\ F\\ 0\end{pmatrix}$$

La résolution du système donne :

$$u_1=0, u_2=\frac{F}{k}, \lambda=F$$

On peut remarquer d'une part, que le système algébrique est inversible, alors que la matrice de rigidité non modifiée par les conditions aux limites ne l'est pas et d'autre

part, que le multiplicateur de Lagrange a une interprétation physique(c'est la réaction au signe près).

Dans le cas général, si l'on impose à un système issu d'une modélisation par éléments finis des conditions aux limites sur certains degrés de liberté du type :

$u_{i_p} = \bar{u}_{i_p}, \quad p = 1,...n_c$

le Lagrangien s'écrit :

$$L((u),\lambda) = \frac{1}{2}\langle u\rangle[K](u) - \langle u\rangle(F) + \sum_{p=1}^{n_c} \lambda_p \left(u_{i_p} - \bar{u}_{i_p}\right)$$

où : (u) est le vecteur des degrés de liberté de dimension n, [K] la matrice de rigidité assemblée de dimension n*n.
La stationnarité de L induit le système à résoudre suivant :

$$\begin{bmatrix} [K] & [B] \\ [B]^T & [0] \end{bmatrix} \begin{pmatrix} (u) \\ (\lambda) \end{pmatrix} = \begin{pmatrix} (F) \\ (\beta) \end{pmatrix}$$

$B_{i_p p} = 1, \quad p = 1,...n_c \quad$ sinon $B_{ij} = 0$

où : $\beta_p = \bar{u}_{i_p}$

$\langle\lambda\rangle = \langle\lambda_1,...\lambda_{n_c}\rangle$

2.2.b cas des contraintes d'égalité (général)

Minimiser f(x)
$h_1(x)=0$
…. **(Qe)**
$h_p(x)=0$

On introduit la fonction de Lagrange : $L(x, \lambda) = f(x) + \sum_{l=1}^{p} \lambda_l h_l(x)$

où les nombres λ_l sont les multiplicateurs de Lagrange
$\lambda_l \in R$ pour l =1, 2, …, p

Les conditions de Lagrange caractérisent un point stationnaire (minimum, maximum, ou point de selle) de f vérifiant les égalités :

$h_l(x) = 0 \qquad l = 1, p$

> **Si x est solution de (Qe), alors il existe des nombres $\lambda_l \in R$, l =1,2, …, p tels que :**
>
> $$\frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i} + \sum_{l=1}^{p} \lambda_l \frac{\partial h}{\partial x_l} = 0 \qquad \text{(ou encore : } \nabla_x L = 0\text{)}$$
>
> $h_l(x) = 0 \qquad l = 1, p \qquad$ (ou encore : $\nabla_\lambda L = 0$)

ce qui peut s'exprimer aussi par : une condition nécessaire pour que x soit solution de (Qe) est qu'il existe un λ tel que (x, λ) soit un point stationnaire de L.

La méthode des multiplicateurs de Lagrange consiste à rechercher la solution de (Qe) en résolvant la condition de stationnarité de L:

$\nabla_x L = 0$
$\nabla_\lambda L = 0$

2.2.c Problème avec limitations d'inégalité : Conditions nécessaires d'optimalité (KKT, Karush, Kuhn, Tucker) :

On suppose que :
• f, g_j j=1, ..., m sont continues et différentiables
• l'hypothèse de Q.C. est vérifiée en $x° \in \mathcal{A}$

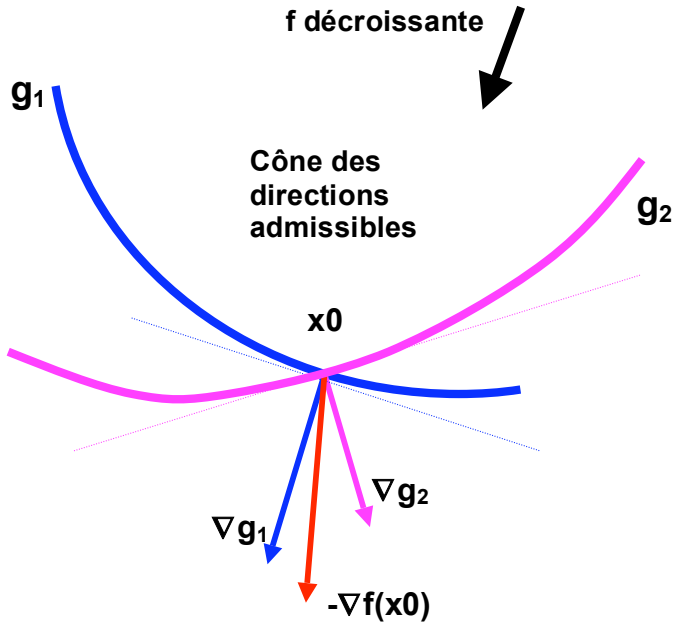
Alors, si x° est un optimum local, il existe des nombres $\lambda_j \geq 0$ tels que :

$$\nabla f(x°) + \sum_{j=1}^{m} \lambda_j \nabla g_j(x°) = 0$$

$$\lambda_j \, g_j(x°) = 0 \quad j=1, ..., m$$

Et si x° est régulier (gradients des limitations linéairement indépendants), alors les λ_j sont uniques.

Caractérisation de l'optimum : interprétation graphique des conditions de KKT :



• l'ensemble des directions admissibles est l'intersection des demi-espaces définis par :
{y tel que $\nabla g_j(x°)^T \cdot y \leq 0$ pour toute g_j saturée en x°}

• x° est un optimum (local) : $-\nabla f(x°)$ fait un angle obtus avec chaque direction admissible (on ne peut plus diminuer f sans sortir du domaine admissible).

• $-\nabla f(x°)$ est une combinaison linéaire, à coefficients $\lambda_j > 0$, des $\nabla g_j(x°)$, pour toute g_j saturée en x°

• on a soit $\lambda_j = 0$, soit $g_j(x°) = 0$: les λ_j associés aux limitations non actives sont nuls.

Conditions de KKT pour un problème avec limitations d'égalité et d'inégalité

On suppose que :
- **f, g_j , j=1, ..., m , h_l, l =1, ..., p sont continues et différentiables l**
- **l'hypothèse de Q.C. est vérifiée en x°∈ $\mathcal{A}$**

Alors, si x° est un optimum local, il existe des nombres $\lambda_j \geq 0$ et μ_l de signe quelconque tels que :

$$\nabla f(x^\circ) + \sum_{j=1}^{m} \lambda_j \nabla g_j(x^\circ) + \sum_{l=1}^{p} \mu_l \nabla h_l(x^\circ) = 0$$

$$\lambda_j \, g_j(x^\circ) = 0 \quad j=1, ..., m$$

Remarques :
- **On suppose x°∈ $\mathcal{A}$, donc on a $g_j(x^\circ) \leq 0$ $h_l(x^\circ) = 0$**
- **Si on a des limitations d'égalité seulement, on retrouve les conditions de Lagrange**

2.3 Condition suffisante d'optimalité (pour des inégalités)

Définition : (x*, λ*) est un *point-selle* *(point-col)* de la fonction de Lagrange:

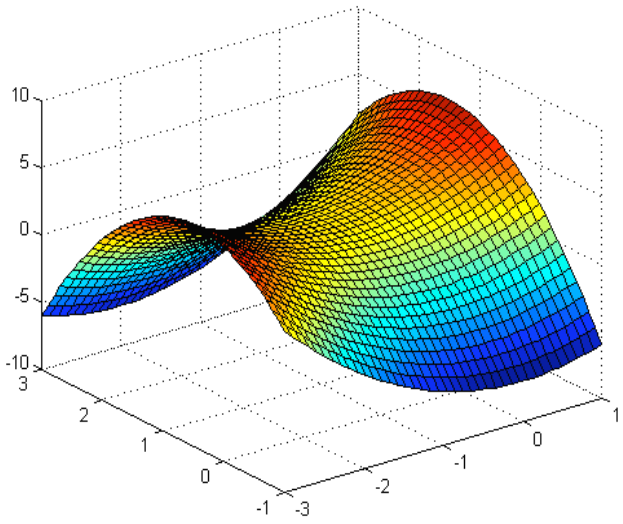
$$L(x, \lambda) = f(x) + \sum_{j=1}^{m} \lambda_j g_j(x)$$

$$\lambda_j \geq 0 \; j=1, ..., m$$

si on a :

$$L(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L(x, \lambda^*)$$

pour tout x ∈R^n, pour tout λ ∈R^m



Ceci signifie qu'en (x*, λ*), L a un minimum en fonction de la variable x et un maximum en fonction de λ

CS d'optimalité : si (x*, λ*) est un point-col de L(x, λ), alors x* est l'optimum global du problème (Q).

C'est un résultat très général (fonctions non convexes, non différentiables, variables discrètes,…) mais pour certains problèmes, il n 'existe pas de point-col...

CS d'optimalité du second ordre:
si on suppose que f, g_j pour j=1, ..., m, h_l pour l=1, …, L sont 2 fois continûment différentiables,

x* est un minimum local de (Q) s'il existe des multiplicateurs de Lagrange λ et μ tels que :

- **Les conditions nécessaires de KKT sont vérifiées en x***
- **La matrice Hessienne du Lagrangien :**

$$\nabla^2 L(x^*) = \nabla^2 f(x^*) + \sum_{j=1}^{m} \lambda_j \nabla^2 g_j(x^*) + \sum_{l=1}^{p} \mu_l \nabla^2 h_l(x^*) = 0$$

est définie positive sur le sous-espace de R^n tangent aux contraintes actives en x*.

$y^T \nabla^2 L(x^*)\ y > 0$ pour tout vecteur $y \neq 0$ qui vérifie :
$\nabla h_l^T (x^*)\ y = 0$ pour l=1, …, p
et $\nabla g_j^T (x^*)\ y = 0$ pour tout j pour lequel $g_j (x^*) = 0$ avec $\lambda_j < 0$

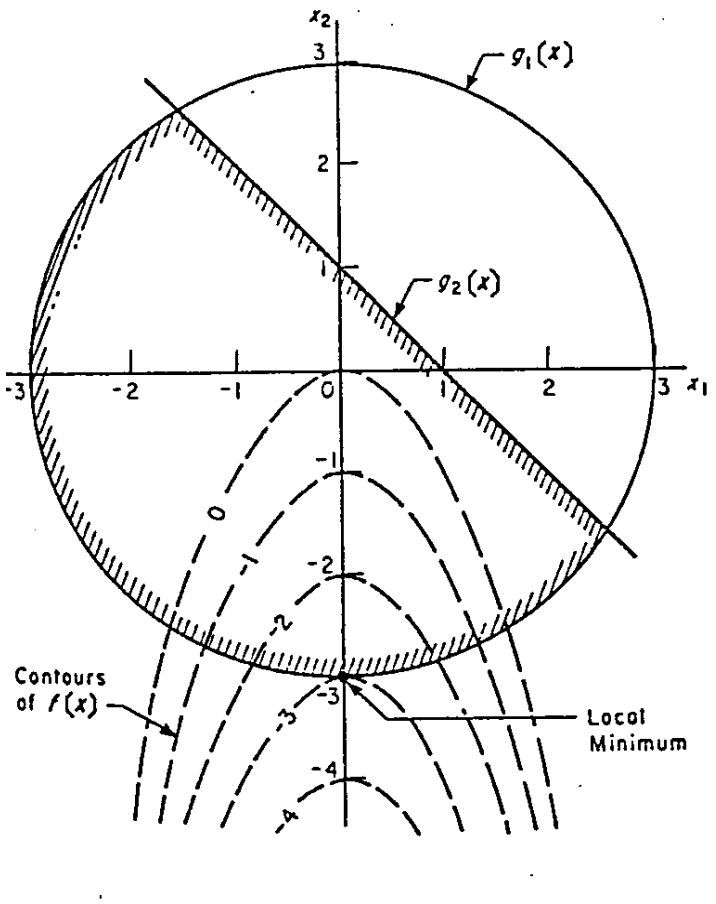
2.4 CNS dans le cas convexe

Si on suppose que f et g_j pour j=1, ..., m sont convexes et différentiables, alors :

x* optimum global de (Q) $\Leftrightarrow$ conditions de K.K.T. vérifiées en x*

Remarque : en pratique, il est très rare de savoir si les conditions de convexité sont satisfaites. Cela n'empêche pas d'utiliser des algorithmes qui s'appuient sur ces conditions d'optimalité, mais on obtiendra généralement au mieux un minimum local. Pour être fiable et robuste, un algorithme doit inclure des vérifications qui permettent d'exclure les points-selle (en x) et les maxima (où les conditions de stationnarité sont aussi satisfaites).

Exemple : **Minimiser $f(x)=x_1^2 + x_2$**
Sous les contraintes :
$g_1 (x) = x_1^2 + x_2^2 - 9 \leq 0$
$g_2 (x) = x_1 + x_2 - 1 \leq 0$



On va montrer que la résolution des conditions de KKT donne les valeurs du minimum x et des multiplicateurs λ associés :

f , g_1 , g_2 sont différentiables , f et g_1 sont convexes ($\nabla^2 f = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$, $\nabla^2 g_1 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$), g_2 est linéaire

Les conditions de KKT s'écrivent :

$$\nabla f(x) + \lambda_1 \nabla g_1(x) + \lambda_2 \nabla g_2(x) = \begin{Bmatrix} 2x_1 \\ 1 \end{Bmatrix} + \lambda_1 \begin{Bmatrix} 2x_1 \\ 2x_2 \end{Bmatrix} + \lambda_2 \begin{Bmatrix} 1_1 \\ 1 \end{Bmatrix} = 0 \qquad (1)$$

$$\lambda_1 g_1(x) = \lambda_1 (x_1^2 + x_2^2 - 9)$$
$$\lambda_2 g_2(x) = \lambda_2 (x_1 + x_2 - 1) \qquad (2)$$

$$\lambda_1 \geq 0$$
$$\lambda_2 \geq 0 \qquad (3)$$

Pour résoudre, on étudie les différents cas possibles *(ce n'est pas une méthode générale).*

Si $\lambda_1 = \lambda_2 = 0$, (2) est vérifiée, mais pas (1)

Si $\lambda_1 \neq 0$ et $\lambda_2 = 0$, (1) $\Rightarrow 2 x_1 + 2 x_1 \lambda_1 = 0$,
$\lambda_1 = -1$ ne vérifie pas (3), donc $x_1 = 0$

(2) s'écrit $\lambda_1 (x_2^2 - 9) = 0$, et comme $\lambda_1 \neq 0$, $x_2 = \pm 3$

si $x_2 = 3$, (1) $\Rightarrow 1 + 6\lambda_1 = 0$,impossible avec (3)
donc $x_2 = -3$, et (1) $\Rightarrow 1 - 6\lambda_1 = 0$, d'où $\lambda_1 = 1/6$

La résolution des conditions de KKT donne donc :
$x_1 = 0$ $x_2 = -3$ $\lambda_1 = 1/6$ $\lambda_2 = 0$

Interprétation (voir figure) : la solution est en (0, -3), la limitation g_2 est inactive, le gradient de f est parallèle au gradient de g_1

2.5 Problème dual

Soit le problème : Minimiser f(x)
avec : $g_j(x) \leq 0$ $j=1, ..., m$ (Q)
$x = (x_1, ..., x_n)$

D'après la CS d'optimalité, trouver un point-col du Lagrangien associé à (Q) permet de résoudre le problème:

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

Point col : (x^*, λ^*) tel que :
$L(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L(x, \lambda^*)$ pour tout $x \in R^n$, tout $\lambda \in R^m$

ou encore :
$(x^*, \lambda^*) = \max_{\lambda} \min_{x} (L(x, \lambda))$

On définit la fonction duale :

$w(\lambda) = \min_{x \in \mathcal{A}} \{L(x,\lambda)\}$

et on appelle problème dual la recherche du maximum de la fonction duale w :

Maximiser $w(\lambda)$ (D)
$\lambda \geq 0$

(Q) = problème primal, (D) = problème dual

Sous certaines conditions (convexité, matrice $\nabla_x^2 L(x^*,\lambda)$ non singulière), ces 2 problèmes sont équivalents (on a : $f(x^*) = w(\lambda^*)$) ; les algorithmes de minimisation avec limitations peuvent être classées en :

- **méthodes primales : résolvent (P)**
  ex : directions admissibles, gradient projeté, gradient réduit, …

- **méthodes duales : résolvent (D)**
  ex : méthode d'Uzawa, de pénalités , Lagrangien augmenté

- **méthodes primales-duales : (P) et (D)**
  ex : méthode SQP

remarques :

- **w et (D) sont définis même s'il n'existe pas de point-col.**
- **w est une fonction concave (en λ)**
- **Le problème dual est souvent plus facile à résoudre que le primal**
- **Dans le cas particulier où le problème est séparable, la formulation duale est particulièrement intéressante :**

Problème séparable : f et les contraintes s'écrivent :

$$f(x) = \sum_{i=1}^{n} f_i(x_i)$$

$$g_j(x) = \sum_{j=1}^{m} g_{ji}(x_i) \qquad j = 1, \ldots, m$$

On a alors aussi:

$$L(x, \lambda) = \sum_{i=1}^{n} L_i(x_i, \lambda)$$

et la minimisation de L par rapport à x peut être faite indépendamment dans chaque direction x_i . Les méthodes d'approximation telles que la linéarisation convexe et la MMA conduisent à des problèmes séparables.

Exemple de résolution du problème dual :

$$\text{Min } f(x) = x_1^2 + x_2^2 + x_3^2$$
Sous les contraintes :
$$g_1(x) = -x_1 - x_2 + 10 \le 0$$
$$g_2(x) = -x_2 - 2x_3 + 8 \le 0$$

$$L(x,\lambda) = x_1^2 + x_2^2 + x_3^2 + \lambda_1(-x_1 - x_2 + 10) + \lambda_2(-x_2 - 2x_3 + 8)$$
$$= L_1(x,\lambda) + L_2(x,\lambda) + L_3(x,\lambda)$$
$$L_1(x,\lambda) = x_1^2 - \lambda_1 x_1$$
$$L_2(x,\lambda) = x_2^2 - \lambda_1 x_2 - \lambda_2 x_2$$
$$L_3(x,\lambda) = x_3^2 - 2\lambda_2 x_3$$

$L_0(x,\lambda) = 10\lambda_1 + 8\lambda_2$

On cherche le minimum de L en x, c'est-à-dire un point stationnaire de L :

$/L//x_1 = /L_1//x_1 = 2\,x_1 - \lambda_1 = 0 \Rightarrow x_1 = \lambda_1/2$
$/L//x_2 = 0 \quad \Rightarrow x_2 = (\lambda_1 + \lambda_2)/2$
$/L//x_3 = 0 \quad \Rightarrow x_3 = \lambda_2$

On obtient la fonction duale en remplaçant x_1, x_2, x_3 par ces valeurs dans $L(x,\lambda)$; on trouve :

$w(\lambda) = -(1/2)\lambda_1^2 - (5/4)\lambda_2^2 - (1/2)\lambda_1\,\lambda_2 + 10\lambda_1 + 8\lambda_2$
et w sera maximale pour :

$/w//\lambda_1 = 0 = -\lambda_1 - ½\,\lambda_2 + 10$
$/w//\lambda_2 = 0 = -(5/2)\lambda_2 - ½\,\lambda_1 + 8$

soit : $\lambda_1^* = 10 - \lambda_2/2$

$\lambda_2^* = 4/3$
$\lambda_1 = 28/3$

$w(\lambda_1^*, \lambda_2^*) = 182.66$

Les valeurs de x associées sont :

$x_1 = 14/3 = 4.66$
$x_2 = 16/3 = 5.33$
$x_3 = 4/3$

et $f(x^*) = (14^2 + 16^2 + 4^2)/9 = 468/9 = 52$

3. Méthodes primales

3.1 Changement de variables

- Pas de règle générale
- Attention aux non-linéarités introduites

3.2 Directions admissibles

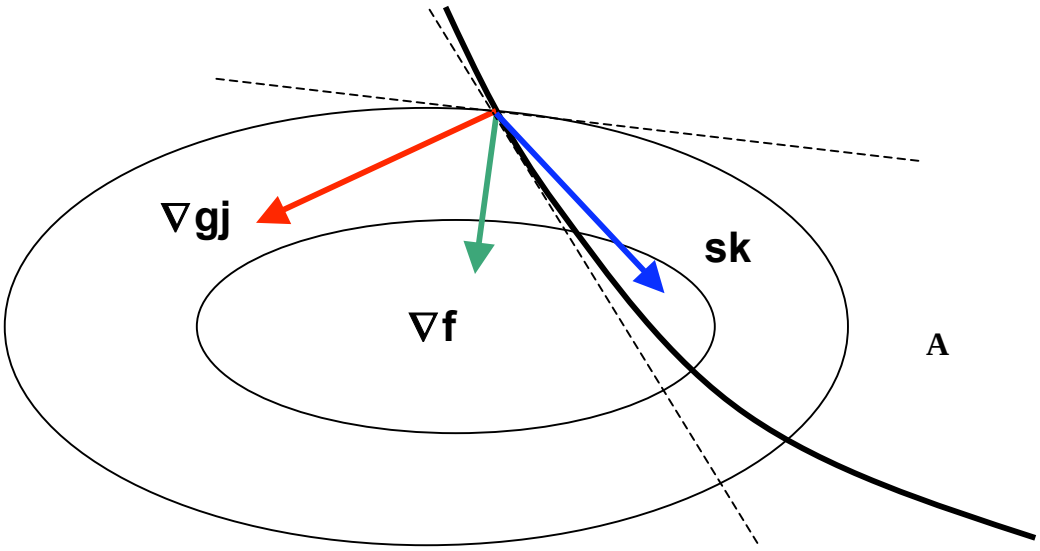
Principe : à l'itération k, chercher une direction s^k telle que, en un point où g_j est saturée ($j \in J°$):

- s^k admissible : $x^k + t\, s^k \in \mathcal{A}$ pour $t \in [0, \theta]$

$$\nabla g_j{}^T (x^k)\, s^k \leq 0 \qquad j \in J°$$

- s^k dir. de descente : f décroît dans la dir. s^k

$$\nabla f^T (x^k)\, s^k < 0$$



Méthode de Zoutendijk : chercher s^k comme solution de :

$$\begin{array}{ll} & \text{Min } \nabla f^T (x^k)\, s^k \\ (L) & \nabla g_j{}^T (x^k)\, s^k \leq 0 \qquad j \in J° \quad \text{(contraintes saturées)} \\ & s^{kT} s^k = 1 \qquad \text{ou bien} \qquad \sum |s_i| = 1 \quad \text{(normalisation)} \end{array}$$

C'est un problème de programmation linéaire.

Algorithme :

(a) point initial x^0, k=0

(b) Résoudre le problème linéaire (L) pour trouver s^k

(c) si $\nabla f^T(x^k)\, s^k < 0$,

déterminer par une recherche unidirectionnelle :

- le pas maximal possible dans la direction y^k , c'est-à dire le pas α_m tel que
$x^k + \alpha_m$ y soit encore dans le domaine admissible.

- le pas optimal
Faire k=k+1 et aller en (b)

sinon ($\nabla f^T(x0)\ s^k = 0$), arrêt

Remarque : **Très simple, mais :**

- **les points sortent du domaine admissible ⇒ imposer à s^k de s'écarter de la frontière**
- **changement de contrainte saturée ?**

Variante (Topkis et Veinott):

Min v
$\nabla f^T(x^k)\ s^k - v \leq 0$
$g_j(x^k) + \nabla g_j^T(x^k)\ s^k - u_j\ v \leq 0$ $j \in J$ (toutes les contr.)

1 variable v et la direction s^k inconnus, $u_j > 0$ fixés (=1 ?)
Si $v^* < 0$, <u>toutes</u> les g_j restent < 0 pour des « petits » déplacements

4. Méthodes de pénalité et du Lagrangien Augmenté

4.1 Principe

Remplacer le problème contraint :

Minimiser f(x)
avec : $g_j(x) \leq 0$ j=1, ..., m (Q)

par une suite de problèmes sans contraintes

Si l'on définit la fonction h telle que :

$$h(y) = \begin{cases} 0 \text{ si } y \leq 0 \\ +\infty \text{ si } y > 0 \end{cases}$$

la résolution du problème d'optimisation revient à résoudre le problème de minimisation sans limitation suivant :

$$\text{Min} \quad \Pi(x)=f(x)+H(x)$$

où $H(x)= \sum_{j=1}^{m} h_j(g_j(x))$

La fonction H étant discontinue et 'infinie' en dehors du domaine admissible, on ne peut utiliser cette méthode telle quelle. Une méthode de pénalité consiste à minimiser une fonction (pénalisée) de la forme :

$$\underset{x}{\text{Min}} \; P(x,\alpha) = f(x) + \Gamma(x,\alpha)$$

où la fonction Γ est telle :

qu'elle force l'optimum à saturer approximativement les limitations qui doivent l'être

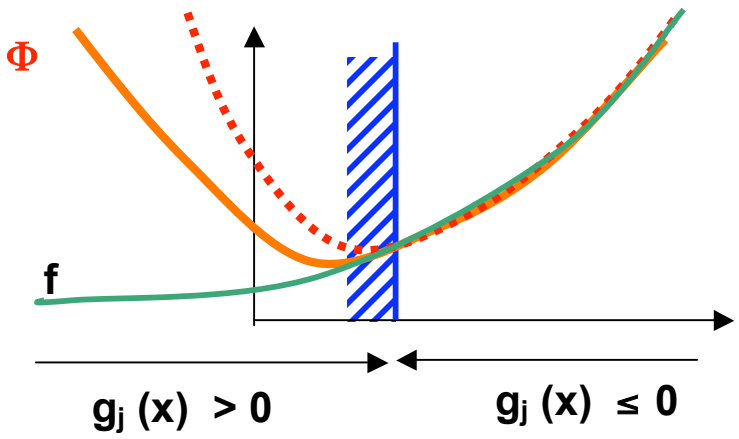
qu' elle n'altère pas trop la solution approchée obtenue

Le paramètre de pénalisation α est à caler pour que la solution approchée soit précise sans que le système à minimiser soit trop mal conditionné.

4.2 Pénalité extérieure

Modifier la fonction à minimiser pour qu'elle prenne de « grandes » valeurs en dehors du domaine admissible.

Exemple avec 1 variable, 1 limitation :



Pour cela, on définit :

$$< g_j(x) >^+ = \begin{cases} 0 \text{ pour } g_j(x) \leq 0 \\ g_j(x) \text{ pour } g_j(x) > 0 \end{cases}$$

et on considère la fonction obtenue en ajoutant à la fonction f(x) à minimiser les termes :

$$r < g_j >^{+2}$$

où r est un scalaire > 0 (facteur de pénalité), ceci pour chaque contrainte g_j .

On obtient la fonction :

$$\Phi(x, r) = f(x) + r \sum_{j=1}^{m} <g_j(x)>^{+2}$$

et on considère le problème :

Minimiser $\Phi(x, r) = f(x) + r \sum_{j=1}^{m} <g_j(x)>^{+2}$ **(P')**

Soit (x*) la solution de (P'). Si on a choisi r « assez grand », les termes $<g_j(x^*)>^{+2}$ seront « petits », ce qui signifie que x* sera « presque » admissible.

Mais en pratique, si r est très grand, Φ est mal conditionnée. C'est pourquoi on procède itérativement, en augmentant r progressivement, et en initialisant chaque recherche par la solution précédente.

Algorithme général :

(a) **k=1, choix de x_1, r_1(proportion coût/pénalité)**

(b) **Résoudre (par une méthode sans limitations) :**

Minimiser $\Phi(x, r_k) = f(x) + r_k \sum_{j=1}^{m} <g_j(x)>^{+2}$ **(P_k')**

avec pour point de départ x_k

→ solution x_k^*

(c) **Tester si les limitations sont « suffisamment » satisfaites :**

Max $(g_j(x)) \leq \varepsilon$?

j=1,…,m

Si oui, fin, $x^* = x_k$

Si non,

(d) **augmenter le facteur de pénalité :**

$r_k = r_k \times c$ (par ex. $1 < c < 10$)

mettre à jour le point de départ : $x_{k+1} = x_k^*$

k=k+1

aller en (b)

- **facile à implanter, mais pas très performant (conditionnement), solution toujours légèrement non admissible**

Remarque : Dans le cas de limitations d'égalité ($h_l(x) = 0$ l = 1, p), la fonction pénalisée à considérer est :

$$\Phi(x, r) = f(x) + r \sum_{j=1}^{p} (h_j(x))^2$$

SI on considère l'équilibre du ressort du paragraphe 2.2.2, on doit minimiser :

$$\underset{u_1,u_2}{\text{Min}}\ \Phi(u_1,u_2,r) = \frac{1}{2}\langle u_1,u_2\rangle \begin{bmatrix} k & -k \\ -k & k \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} - Fu_2 + r\,u_1^2$$

Ce qui donne : $u_1 = \frac{1}{2r}F$, $u_2 = \frac{k+2r}{2rk}F$

Ces valeurs tendent vers la solution exacte quand $r \to +\infty$.

4.3 Pénalité intérieure

Même principe que la pénalité extérieure, mais pour obtenir un minimum approché par l'intérieur du domaine (donc toujours admissible). On définit :

$$\Phi(x, r) = f(x) - r \sum_{j=1}^{m} \frac{1}{g_j(x)}$$

ou bien :

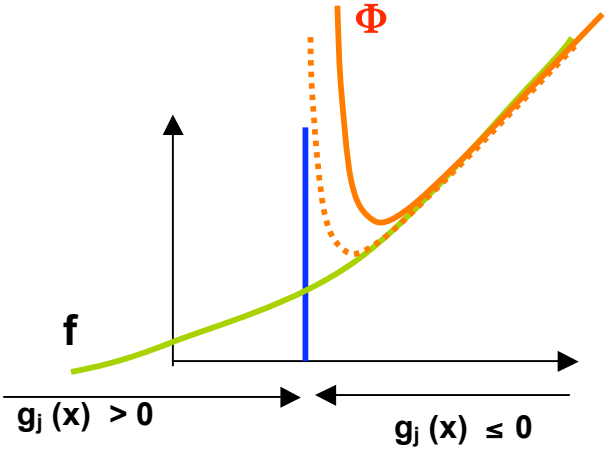
$$\Phi(x, r) = f(x) - r \sum_{j=1}^{m} \log(-g_j(x))$$

ou le facteur de pénalité r doit cette fois tendre vers 0 pour que x s'approche de la frontière du domaine admissible. On résout donc une succession de problèmes sans contraintes :

$$\Phi(x, r_k) = f(x) + r_k \sum_{j=1}^{m} \frac{1}{g_j(x)} \qquad k=1, \ldots$$

avec :

$r_{k+1} = r_k \times c$ **(par ex. 0.1 < c < 1)**



Il faut connaître un point de départ admissible

- **cette pénalisation n'est pas valable sur la frontière ni à l'extérieur du domaine admissible.**

4.4 Approximation des multiplicateurs de K.K.T. à l'optimum

Les méthodes de pénalité permettent d'obtenir les valeurs des multiplicateurs à l'optimum. Exemple avec la pénalité extérieure :

La solution x_k^* du k-ième problème sans contraintes :

$$\text{Minimiser } \Phi(x, r_k) = f(x) + r_k \sum_{j=1}^{m} \langle g_j(x) \rangle^{+2} \qquad (P_k')$$

est un point stationnaire de Φ :

$$\nabla\Phi(x_k^*, r_k) = \nabla f(x_k^*) + 2\, r_k \sum_{j=1}^{m} \langle g_j(x_k^*) \rangle^{+} \nabla g_j(x_k^*) = 0 \qquad (1)$$

Par ailleurs, soit x* la solution du problème initial (Q) ; si les contraintes sont qualifiées en x* , les conditions nécessaires de KKT s'appliquent, et il existe des nombres $\lambda_j^* \geq 0$ j=1,…,m uniques tels que :

$$\nabla f(x^*) + \sum_{j=1}^{m} \lambda_j^* \nabla g_j(x^*) = 0 \qquad (2)$$

$$\lambda_j^*\, g_j(x^*) = 0$$

Soit J° l'ensemble des indices des contraintes saturées en x*. Pour les autres indices $j \notin J°$, $g_j(x^*) < 0$. On peut supposer qu'à partir d'une certaine itération K, les points x_k^* vérifieront aussi $g_j(x_k^*) < 0$
Les termes associés : $\langle g_j(x_k^*) \rangle^+$ sont donc nuls pour $j \notin J°$,et il reste dans (1) :

$$\nabla f(x_k^*) + 2\, r_k \sum_{j \in J^0}^{m} \langle g_j(x_k^*) \rangle^{+} \nabla g_j(x_k^*) = 0 \quad \text{pour } k > K$$

∇f et ∇g_j sont continues, les λ_j sont uniques donc, en passant à la limite :

$$\lambda_j^* = \lim\ 2\, r_k \langle g_j(x_k^*) \rangle^{+} \quad \text{pour } j \in J°$$

Les valeurs des λ_j^* permettent, par exemple, d'estimer l'influence des limitations sur la valeur de f à l'optimum.

4.5 Méthode du Lagrangien augmenté

Les deux méthodes précédentes permettent d'obtenir des solutions approchées qui convergent vers la solution théorique avec les facteurs de pénalisation (vers l'infini pour la pénalité extérieure, vers 0 pour la pénalité intérieure). Toutefois le processus de convergence est confronté à des problèmes de conditionnement pour les valeurs limites des paramètres. La méthode du lagrangien augmenté mixe la stationnarité du lagrangien et la méthode des pénalités extérieures. Elle est associée à un processus itératif permettant d'obtenir des solutions précises même pour des valeurs finies du paramètre de pénalisation.

On considère la fonction :

$$A(x, \lambda, r) = f(x) + r \sum_{j=1}^{m} \left\langle \frac{\lambda_j}{2r} + g_j \right\rangle^{+2} \qquad \text{(Lagrangien augmenté)}$$

La stationnarité de A implique :

$$\frac{\partial A}{\partial x_i} = \frac{\partial f}{\partial x_i} + 2r \sum_{j=1}^{m} \left\langle \frac{\lambda_j}{2r} + g_j \right\rangle^{+} \frac{\partial g_j}{\partial x_i} = 0$$

Les relations de Kuhn et Tucker s'écrivent :

$$\frac{\partial f}{\partial x_i} + \sum_{j=1}^{m} \lambda_j \frac{\partial g_j}{\partial x_i} = 0, \quad \lambda_j g_j = 0, \quad \lambda_j \geq 0$$

En comparant les deux dernières relations, on peut introduire un processus itératif, afin d'améliorer l'estimation des multiplicateurs de Lagrange, en prenant :

$$\lambda_j = \max\left\langle \lambda_j + 2\, r\, g_j, 0 \right\rangle$$

L'algorithme de résolution associé est alors :

(a) k=1, choix de x_1, r, λ_1 (souvent nul)

(b) Résoudre (par une méthode sans limitations) :

$$\underset{x}{\text{Min}} \quad A(x,\lambda_k,r) = f(x) + r\sum_{j=1}^{m}\left\langle \frac{\lambda_{jk}}{2r} + g_j \right\rangle^{+2}$$ avec pour point de départ x_k

→ solution x_k^*

(c) Actualiser les multiplicateurs :

$$\lambda_{jk+1} = \max\left\langle \lambda_{jk} + 2\, r\, g_j(x_k^*), 0 \right\rangle$$

(d) Tester si les limitations sont « suffisamment » satisfaites :

Max $(g_j(x)) \leq \varepsilon$?
j=1,…,m

Si oui, fin, $x^* = x_k$, $\lambda^* = \lambda_{k+1}$

Si non,

(e) augmenter <u>éventuellement</u> le facteur de pénalité :

$r_k = r_k \times c$ (par ex. $1 < c < 10$)

mettre à jour le point de départ : $x_{k+1} = x_k^*$

k=k+1

aller en (b)

♦ **La méthode est assez faiblement dépendante du facteur de pénalisation r.**

4.6. <u>Exemple à une variable</u>

Soit à minimiser la fonction f(x)=0.5 x sous la limitation $g(x)=4-x \leq 0$ par les différentes méthodes de pénalité.

a. <u>Pénalités extérieures</u>.

La fonction pénalisée s'écrit :

$$P_{ext}(x,r) = 0.5x + r\left\langle 4 - x \right\rangle^{+2}$$

Sa minimisation induit les solutions approchées dépendant de r :

$$x_r = 4 - \frac{0.5}{2r}$$

b. <u>Pénalités intérieures</u>

La fonction pénalisée s'écrit :

$$P_{int}(x,t) = 0.5x - \frac{t}{4-x}$$

La solution approchée associée est :

$$x_t = 4 + \sqrt{\frac{t}{0.5}}$$

c. <u>Lagrangien augmenté</u>

$$A(x,\lambda,r) = 0.5x + r\left\langle \frac{\lambda_j}{2r} + 4 - x \right\rangle^{+2}$$

La stationnarité de A par rapport à x donne :

$$\frac{\partial A}{\partial x}(x,\lambda,r) = 0.5 - 2\,r\left\langle \frac{\lambda_j}{2r} + 4 - x \right\rangle^{+} = 0 \quad \rightarrow \quad x^* = 4 + \frac{\lambda_j}{2r} - \frac{0.5}{2r}$$

On part de l'estimation initiale :

$\lambda_1 = 0$

La minimisation de A(x,0,r) donne

$$x_2 = 4 - \frac{0.5}{2r}$$

La nouvelle estimation du multiplicateur est :

$$\lambda_2 = \max\left\langle \lambda_1 + 2rg_j(x_2^*), 0 \right\rangle = 0.5$$

La minimisation de A(x, λ_2,r) donne :

$$x_2 = 4$$

On a donc obtenu la solution exacte au bout de 2 itérations, ceci pour r quelconque.

Chapitre VII – Lien optimisation - mécanique : calcul des sensibilités, mise en œuvre, logiciels

1. Couplage calcul – optimisation

1.1 Introduction

Contexte :

- analyse par éléments finis (ou équations intégrales, …)
- pour l'optimisation de produits ou de procédés, ou l'identification de paramètres

Variables :

- épaisseurs, sections, distances, dimensions, coordonnées de points, pentes, rayons de courbure,
- Orientation de fibres, densité, …
- Propriétés de matériau (problèmes d'identification de lois de comportement) , forces, coefficients de frottement, …
- Nombre de raidisseurs, de rainures, de couches ou de plis dans un composite (variables discrètes)…
- Présence ou absence de matière (variables binaires 0/1, optimisation topologique)

Etc …

Les variables sont associées à des données du modèle E.F. : des nœuds, des éléments ou des entités géométriques, technologiques (groupes d'éléments)

→ Toute donnée d'un code de simulation numérique peut être prise comme une variable à identifier ou optimiser.

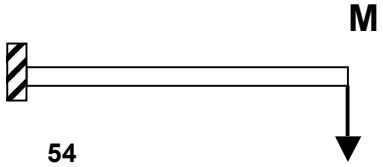
Fonctions : explicites, ou à partir des « réponses » = résultats du calcul mécanique

- poids, surface, volume,
- Contrainte, déplacement, déformation, température,…
- Energie de déformation, travail,
- Fréquences propres, modes propres,
- Norme de l'écart calcul – expérience (identification)

Etc …

Valeurs globales ou ponctuelles, intégrales, moyennes, valeurs maximales, …

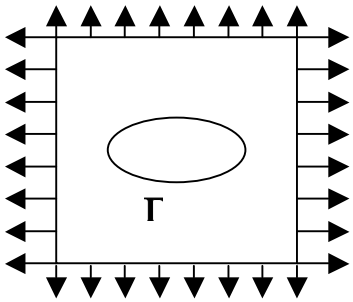
Exemples :



Minimiser w_M

Minimiser (maximum (σ))
Γ

Minimiser $\int_{\Gamma} \sigma \; d\gamma$



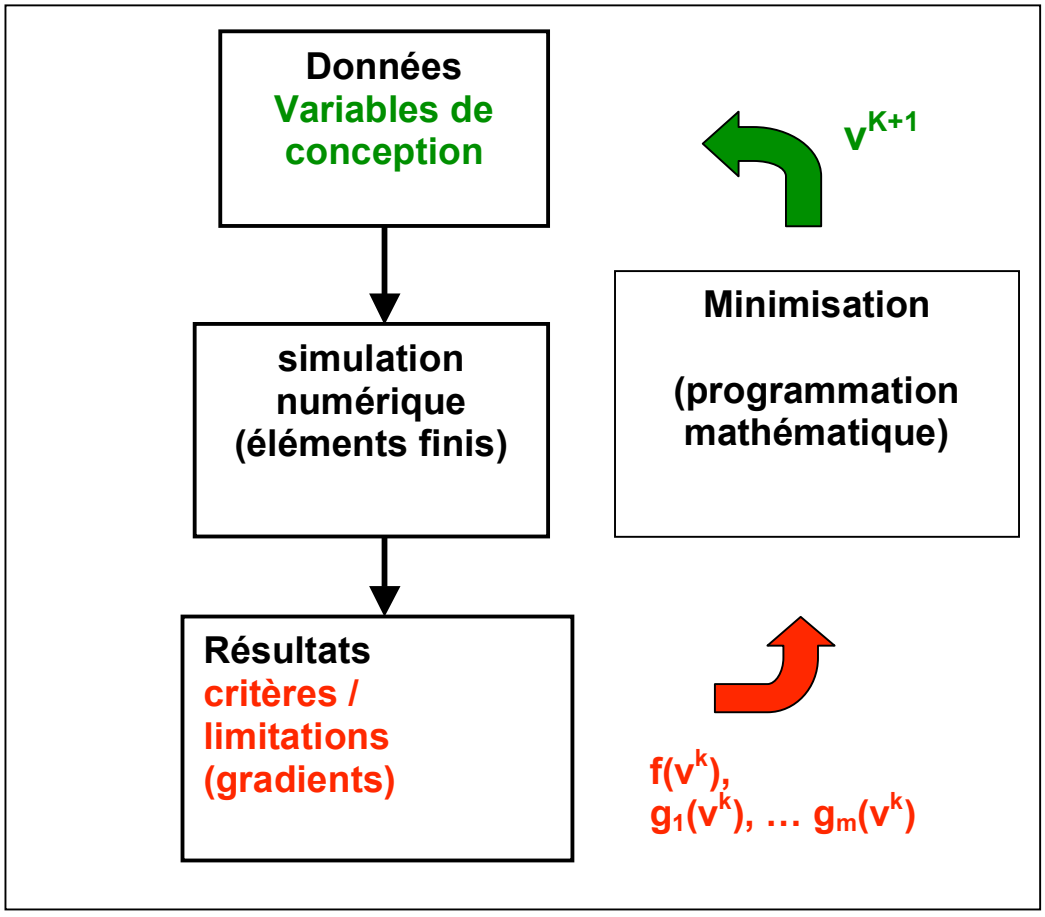
➡ **Tout résultat de la simulation numérique peut être pris comme fonction objectif ou limitation.**

➡ *** Problèmes multi-objectifs, multidisciplinaires**

1.2 Couplage « direct »

Principe : le code de minimisation calcule itérativement une suite de valeurs des variables (v^1, v^2 , …, $v^k \rightarrow v^*$)

et fait appel au code de simulation pour évaluer les fonctions $f(v^k)$, $g_1(v^k)$, … $g_m(v^k)$, ceci à chaque itération k.



➡ **Nécessité d'un interfaçage pour l'utilisation itérative simulation / optimisation**

Points clés / difficultés de l'interfaçage :

Développeurs

Choix des variables : quel paramétrage disponible (forme), restrictions pour cohérence du modèle

Mise à jour des variables (générer tout le modèle à chaque itération ?)

Optimisation de forme : remaillage, estimateurs d'erreur

Accès aux sources ?

ou utilisation des fichiers d'échange ?

Utilisateurs :

Formulation représentative du problème réel

Temps de calcul disponible

Choix d'un algorithme

Mise à l'échelle, tests d'arrêt

Contrôle de la précision au cours des itérations,

Minimum local ou global

Validation de la solution trouvée

1.3 Méthodes d'approximation (surfaces de réponse)

Principe :

- **a) Effectuer p simulations numériques pour p valeurs « bien choisies » des variables $v^1, v^2, \ldots, v^p$**

$v^1=(v_1^1, \ldots, v_n^1)$

- **b) On obtient : $f^1=f(v^1)$, $f^2=f(v^2)$, …, $f^p=f(v^p)$**

et éventuellement ses dérivées.

De même pour chaque limitation.

c) Construire à partir de ces valeurs une approximation de la fonction objectif : f_a et des limitations : g_{1a}, …, g_{ma}
(ex : polynômes)

En fonction du nombre de points p et de leur choix, l'approximation peut être locale ou globale (surfaces de réponse)

d) Résoudre le problème approché : Minimiser f_a sous $g_{1a} \leq 0$, …, $g_{ma} \leq 0$

où f_a, g_{1a} ,… g_{ma} , sont des fonctions explicites des variables ; la minimisation est donc très rapide car elle n'implique pas de calculs E.F.
On obtient une solution v^*_a

e) Calculer les fonctions exactes $f(v^*_a)$, $g_1(v^*_a)$, …, $g_m(v^*_a)$
(une simulation E.F.)

f) Test de convergence → Fin ou aller en a)
La convergence est obtenue lorsque les solutions successives sont proches et que l'écart entre les valeurs approchées et les valeurs exactes des fonctions est suffisamment faible.

Intérêt : sépare (≈…) étapes simulation E.F. et minimisation, parallélisme possible pour les simulations, contrôle du nombre de calculs des fonctions.

En optimisation de structures, cette approche est souvent utilisée avec la méthode de linéarisation convexe, ou des asymptotes mobiles (approximations locales). Ces deux méthodes permettent d'avoir des problèmes approchés séparables, qui pourront être traités très efficacement par des algorithmes duaux.

2. Calcul des sensibilités

La plupart des méthodes de descente efficaces nécessitent les gradients de la fonction objectif et des limitations, et certaines méthodes d'approximation également. Ce calcul n'est pas immédiat lorsque les fonctions à dériver ne sont pas exprimées explicitement en fonction des variables, et c'est le cas par exemple avec une modélisation par éléments finis.

Exemple : élasticité linéaire

Soit : $v = (v_1, \ldots, v_n)$ **les variables**
$U(v)$ **la réponse considérée (ex : déplacements)**
$f(v)$ **la fonction considérée (ex : norme des déplacements)**

On doit calculer $\nabla f = \left\{ \begin{array}{c} \frac{\partial f}{\partial v_1} \\ \\ \frac{\partial f}{\partial v_n} \end{array} \right\}$ **avec** $f(v) = F(v, U(v))$ **(1)**

2.1) Calcul des sensibilités par différences finies

d.f. avant : $$\frac{\partial f(v)}{\partial v_i} \approx \frac{f(v_i + \partial v_i) - f(v_i)}{\partial v_i} \quad (2)$$

d.f. centrées : $$\frac{\partial f(v)}{\partial v_i} \approx \frac{f(v_i + \partial v_i) - f(v_i - \partial v_i)}{2\partial v_i}$$

➡ **une simulation supplémentaire par variable**

➡ **implantation numérique immédiate**

➡ **comment choisir la perturbation** δv_i **?**
Problèmes numériques liés aux approximations des dérivées par différences finies

$f : R \rightarrow R$

Erreur de troncature : termes négligés dans le développement de Taylor :

$f(x+h) = f(x) + h\, f'(x) + \frac{1}{2} h^2 f''(y)$ **avec** $x \leq y \leq x+h$

$f'(x) - (f(x+h)-f(x)) / h = \frac{1}{2} h\, f''(y)$ **= troncature,**

croissante en h, ➡ **diminuer h pour diminuer l'erreur**

Erreur de condition (arrondi, dépend de la précision machine)

soustraction de 2 nombres trop proches f(x+h)- f(x)

décroissante en h, ➡ **augmenter h pour diminuer l'erreur**

erreur sur un calcul par d .f. = somme d'une erreur linéaire en h et d'une erreur en 1/h

Trouver un compromis (par une étude de convergence préalable) :

- valeur valable sur tout le domaine

- pour toutes les variables et les fonctions

ou utiliser plusieurs pas

En cas de difficultés de convergence, vérifier les gradients (toujours vrai, quelle que soit la méthode de calcul des dérivées…)
Exemple 1 variable :

$f2(x) = x^2$ f2'(x=1) = 2 $f4(x) = x^4$ f4'(x=1) = 4

Tracé des valeurs obtenues pour f2'(1) et f4'(1)par diff. finies avant et centrées en fonction du pas.

Valeurs du pas de perturbation :
1.0000 0.2000 0.0400 … (divisé par 5) … 1.67e-017

2.2) Calcul analytique des sensibilités

Remarque : calcul « analytique » signifie qu'on établit analytiquement les expressions à calculer ; elles sont ensuite discrétisées et leur valeur est calculée par le code d'éléments finis.

U dépend de v car c'est la solution des équations d'équilibre pour une valeur de v donnée, donc :

$$\frac{\partial f(v)}{\partial v_i} = \frac{\partial F(v,U(v))}{\partial v_i} + \frac{\partial F(v,U(v))}{\partial U}\frac{\partial U(v)}{\partial v_i} \qquad (3)$$

(a) (b) (c)

Les termes (a) et (b) ne posent généralement pas de problème :
(a) correspond à la dérivation des termes de F où les variables apparaissent explicitement,

Pour le terme (b), si on se place dans le cadre de l'élasticité, F est explicite en fonction du vecteur des déplacements. D'une manière générale, les fonctions à minimiser ou les limitations s'expriment en fonction de la variable d'état.

Par contre, le vecteur d'état U dépend implicitement des variables, puisque la matrice K n'est jamais inversée.

➡ **comment calculer le terme (c) ?**

Deux approches principales pour le calcul des sensibilités par dérivation des équations d'équilibre : méthode directe et état adjoint

Avec U calculé par résolution linéaire (élasticité)

K(v) U(v) = Q(v) (4)

Dérivation par rapport à v_i :

$$\frac{\partial K}{\partial v_i}U + K\frac{\partial U}{\partial v_i} = \frac{\partial Q}{\partial v_i} \qquad (5)$$

si on peut calculer $\frac{\partial K}{\partial v_i}, \frac{\partial Q}{\partial v_i}$, on peut obtenir $\frac{\partial U}{\partial v_i}$ par une résolution linéaire supplémentaire :

méthode directe, ou pseudo-charges :

$$K\frac{\partial U}{\partial v_i} = \frac{\partial Q}{\partial v_i} - \frac{\partial K}{\partial v_i}U \qquad (6)$$

La matrice de rigidité K a déjà été triangularisée :

➡ **Le calcul du gradient de f nécessite la résolution de cas de charge seulement (1 résolution par variable)**

$\frac{\partial U}{\partial v_i}$ **pourra servir au calcul du gradient d'autres fonctions**

ou bien méthode de l'état adjoint (charges virtuelles)

on a (3): $\frac{\partial f(v)}{\partial v_i} = \frac{\partial F(v,U(v))}{\partial v_i} + \frac{\partial F(v,U(v))}{\partial U} \boxed{\frac{\partial U(v)}{\partial v_i}}$

et (6): $K \boxed{\frac{\partial U}{\partial v_i}} = \frac{\partial Q}{\partial v_i} - \frac{\partial K}{\partial v_i} U$

d'où : $\frac{\partial f(v)}{\partial v_i} = \frac{\partial F}{\partial v_i} + \boxed{\frac{\partial F}{\partial U} K^{-1}} \left(\frac{\partial Q}{\partial v_i} - \frac{\partial K}{\partial v_i} U \right)$

on définit : λ^T **= vecteur d'état adjoint**

On peut calculer directement le produit $\lambda^T = \frac{\partial F}{\partial U} K^{-1}$ **, en résolvant :** $K^T \lambda = \frac{\partial F}{\partial U}$

λ dépend de la fonction qu'on dérive f, mais pas de la variable par rapport à laquelle on dérive

⇒ **Le calcul du gradient de f nécessite un seul cas de charge à résoudre**

Pour un problème donné, autant de résolutions que de fonctions à dériver (1 f.objectif + m limitations)

Pour le calcul des gradients, on choisira la méthode directe ou l'état adjoint selon le nombre de variables et de fonctions.

Schéma d'implantation numérique des calculs analytiques de sensibilités

Assemblage des matrices de rigidité Ke

Triangularisation

Assemblage du chargement Qe dans Q

Résolution → U

Calcul des gradients par la méthode directe

Pour chaque variable vi (n)

Dérivées de Ke et Qe

2nd mb élémentaire :

$$We = \frac{\partial Qe}{\partial v_i} - \frac{\partial Ke}{\partial v_i} Ue$$

Assemblage We → Wi

Résolution → $\frac{\partial U}{\partial v_i}$

Pour chaque fonction gj (f, gj) (m+1)

Termes explicites $\frac{\partial g_j}{\partial v_i}$

$$\frac{dg_j}{dv_i} = \frac{\partial g_j}{\partial v_i} + \frac{\partial g_j}{\partial U}\frac{\partial U}{\partial v_i}$$

Calcul des gradients par l'état adjoint

Pour chaque fonction gj (f et gj) (m+1)

2nd mb élém. état adjoint :

$\frac{\partial g_j}{\partial U}$ **(assemblage)**

Résolution → λj

Pour chaque variable vj (n)

Dérivées de Ke et Qe

$$We = \frac{\partial Qe}{\partial v_i} - \frac{\partial Ke}{\partial v_i} Ue$$

Assemblage We → W_i

Termes explicites $\frac{\partial g_j}{\partial v_i}$

$$\frac{dg_j}{dv_i} = \frac{\partial g_j}{\partial v_i} + \lambda_j{}^T W_i$$

2.3) Calcul semi-analytique

C'est un calcul basé sur les expressions analytiques (méthode directe ou état adjoint) mais utilisant des différences finies pour les calculs dérivées des matrices de rigidité élémentaires.
Cette méthode évite d'avoir besoin des programmes sources des modèles éléments finis utilisés

2.4) Logiciels de différentiation automatique

A partir des programmes calculant les fonctions, de la désignation des variables, … génèrent les programmes donnant les dérivées

Exemples de logiciels: ADOL_C, ADIFOR, ODYSSEE,…

A suivre…

http://www.inria.fr/rrrt/rr-3555.html

http://www.sc.rwth-aachen.de/Research/AD/

http://euler.mcs.utulsa.edu/~class_diaz/cs8243/

3.Optimisation de forme

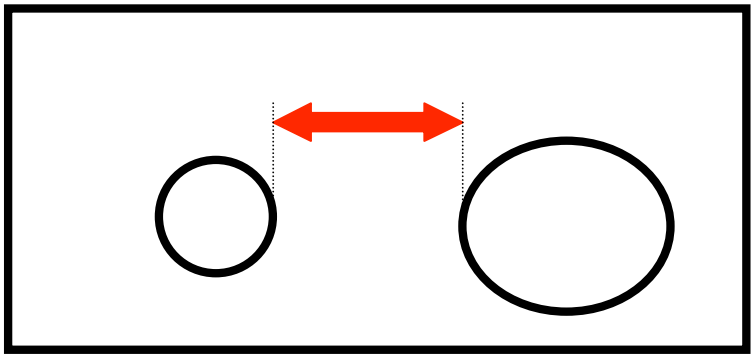
3.1 Choix des variables de conception, contrôle du modèle géométrique

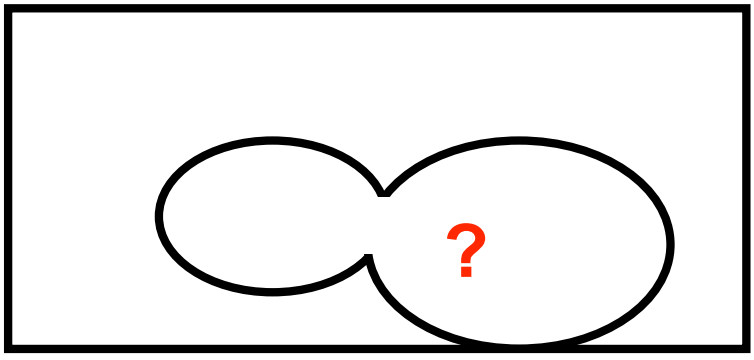
- **Nœuds du maillage en éléments finis : trop nombreux, ne garantissent pas la régularité des contours**

- **Modèles CAO basés sur : courbes de Bézier, splines, B-splines, NURBS, …**
  Les variables définissent les positions de pôles, qui ne sont pas forcément situés sur la courbe à optimiser. Il faut privilégier des approximations de faible degré afin d'éviter les oscillations.

Le modèle doit être remis à jour à partir des valeurs des variables seulement :

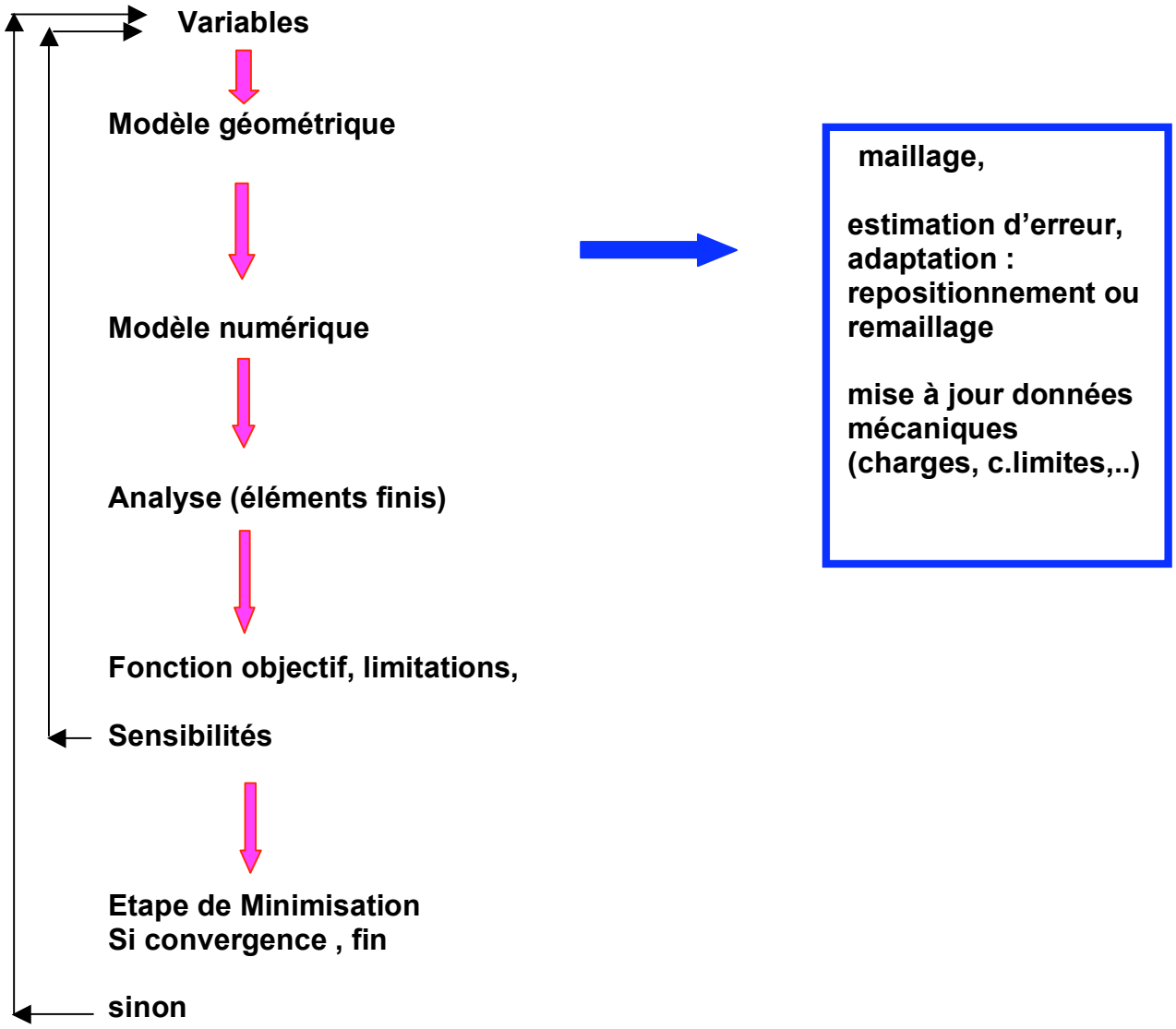
- **indépendance des variables entre elles,**
- **définition de contraintes pour garantir la cohérence du modèle.**





Dans certains cas, les limitations d'égalité peuvent permettre d'éliminer certaines variables et de réduire le problème.

3.2 Actualisation du modèle éléments finis (analyse de sensibilité, mise à jour des variables)



Il faut disposer d'outils d'actualisation du modèle éléments finis tels que :

- **repositionnement des nœuds, par exemple par lissage laplacien ou repositionnement élastique, ceci pour les modifications de forme de faible amplitude**
- **remaillage pour les modifications plus importantes**
- **estimateur d'erreur et maillage adaptatif, afin de contrôler que l'erreur commise reste constante au cours de l'optimisation.**

4. Logiciels d'optimisation et de gestion de calculs

Bibliothèques générales: Nag, IMSL, Optimization Toolbox de Matlab,

Logiciels (commercialisés ou libres)
Solvopt http://www.kfunigraz.ac.at/imawww/kuntsevich/solvopt/

Minlp (integer prog.) http://www.gamsworld.org/minlp/

DOT,VisualDOC, BIGDOT,.. http://www.vrand.com/

CONMIN, NEWSUMT, IDESIGN, et d'autres ...

Codes d'éléments finis incluant de l'optimisation: NASTRAN, I-DEAS, SAMCEF, ABAQUS, GENESIS , OPTISTRUCT(optimisation topologique), …

Plate-formes de gestion de calculs :

Motivé par besoin de couplage entre plusieurs simulations (fluide-structure, mécanique-acoustique,…)
optimisation multi-critères, multi-disciplinaire

Boss-Quattro http://www.samtech.fr/

LMS-Optimus http://www.lms.be

iSIGHT http ://www.engineous.com

modeFRONTIER http://esteco1.area.trieste.it

Un site concernant la Recherche (revue, congrès,…) :

http://www.aero.ufl.edu/~issmo/

Calcul de sensibilités par différences finies

1. **Principe**

Calcul de gradients par perturbation des variables

d.f. avant : $$\frac{\partial f(v)}{\partial v_i} \approx \frac{f(v_i + \partial v_i) - f(v_i)}{\partial v_i}$$

d.f. centrées : $$\frac{\partial f(v)}{\partial v_i} \approx \frac{f(v_i + \partial v_i) - f(v_i - \partial v_i)}{2\partial v_i}$$

2. **Sources d'erreur**

Développement limité :

$$f(v + \partial v_i) = f(v) + \frac{\partial f}{\partial v_i}(v)\Delta v_i + \frac{1}{2}\frac{\partial^2 f}{\partial v_i^2}(v + \theta \partial v_i)\Delta v_i^2 \quad , \ 0 \le \theta \le 1$$

où le vecteur des paramètres perturbé ∂v_i s'écrit :

$$\langle \partial v_i \rangle = \langle 0, 0, 0 ... \Delta v_i, 0, ... 0 \rangle$$ **(composantes nulles sauf la ième)**

Δv_i est la valeur de la perturbation associée à la ième variable.
Mais le calcul de f est déjà entaché d'erreur. Ceci est dû à la précision machine, à des erreurs d'arrondi lors d'inversion de système,… En fait, sur un aspect numérique, on ne calcule pas f(v) mais $\hat{f}(v) = f(v) + \varepsilon_c(v)$ où $\varepsilon_c(v)$ est l'erreur de condition. L'équation précédente s'écrit alors :

$$\hat{f}(v + \partial v_i) + \varepsilon_c(v + \partial v_i) = \hat{f}(v) + \varepsilon_c(v) + \frac{\partial f}{\partial v_i}(v)\Delta v_i + \frac{1}{2}\frac{\partial^2 f}{\partial v_i^2}(v + \theta \partial v_i)\Delta v_i^2 \quad , \ 0 \le \theta \le 1$$

On en déduit alors l'expression de la dérivée :

$$\frac{\partial f}{\partial v_i}(v) = \frac{\hat{f}(v + \partial v_i) - \hat{f}(v)}{\Delta v_i} + \frac{\varepsilon_c(v + \partial v_i) - \varepsilon_c(v)}{\Delta v_i} + \frac{1}{2}\frac{\partial^2 f}{\partial v_i^2}(v + \theta \partial v_i)\Delta v_i \quad , \ 0 \le \theta \le 1$$

Le calcul de sensibilité par différence finie avant utilise l'approximation suivante :

$$\frac{\partial f}{\partial v_i}(v) \approx \frac{\hat{f}(v + \partial v_i) - \hat{f}(v)}{\Delta v_i}$$

Ce calcul est entaché d'erreur de condition $e_c(\Delta v_i)$ et de troncature $e_t(\Delta v_i)$:

$$\frac{\partial f}{\partial v_i}(v) = \frac{\hat{f}(v + \partial v_i) - \hat{f}(v)}{\Delta v_i} + e_c(\Delta v_i) + e_t(\Delta v_i)$$

avec : $$e_c(\Delta v_i) = \frac{\varepsilon_c(v + \partial v_i) - \varepsilon_c(v)}{\Delta v_i}, \quad e_t(\Delta v_i) = \frac{1}{2}\frac{\partial^2 f}{\partial v_i^2}(v + \theta \partial v_i)\Delta v_i$$

L'erreur de condition croît avec la perturbation, tandis que l'erreur de troncature décroît avec la perturbation.
L'erreur de condition peut-être très importante dans le cas de systèmes mal conditionnés.

Exemple 1 :
Une fonction f a une valeur théorique pour un jeu de paramètres donné de :
f(v) = 19.273428
Sa valeur théorique pour une perturbation d'un des paramètres de 0.0001 est de :
$f(v+\delta v_i)$=19.273491
En fait, les valeurs sont calculées avec 6 chiffres significatifs.
Les valeurs des sensibilités obtenues par différences finies avant sont de :
0.63 en théorie et de 1. en calcul numérique.

Exemple 2
On considère un système en (u,v) dépendant d'un paramètre x :

$$\begin{bmatrix} 101 & x \\ x & 100 \end{bmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 10 \\ 10 \end{pmatrix}$$

Calculons la sensibilité de u par rapport à x théorique:

$$u(x) = \frac{-10x + 1000}{10100 - x^2}$$

Pour x=100

Chapitre VIII - Identification en mécanique (problème inverse)

1. Introduction.

1.1.Problème direct

La résolution d'un **problème direct** en thermomécanique du solide consiste à déterminer :

les champs de température, de flux thermiques, de déplacements, de contraintes, de déformations

connaissant :

- les propriétés du matériau telles que la conductivité et la capacité calorifique, la loi de comportement liant contraintes, déformations et température (module de Young, coefficient de Poisson, courbe d'écrouissage dépendants de la température)
- les conditions initiales et aux limites : températures initiales, conditions d'échange thermique (convection, radiation,…) et conditions aux limites mécaniques (déplacements ou efforts imposés, coefficients de frottement)

Au plan mathématique, la résolution d'un problème direct revient à résoudre analytiquement ou de manière approchée(éléments finis) un système d'équations différentielles associées à des conditions initiales et des conditions aux limites ,tous les paramètres régissant le système étant connus. On peut représenter les opérations associées à un calcul direct en thermoélasticité par la figure 1.

1.2.Problème inverse

Par opposition , la résolution d'un problème inverse consiste à déterminer un ou plusieurs paramètres définissant le système (paramètres matériaux ou conditions aux limites) en s'appuyant sur des résultats expérimentaux obtenus à partir d'un ou plusieurs essais. Le principe est de 'caler' ces paramètres de telle manière que des calculs analytiques ou par éléments finis réalisés avec ces paramètres identifiés donnent des résultats s'approchant au mieux des résultats expérimentaux.

Pour un problème de thermoélasticité (figure 2), les inconnues pourraient être les facteurs d'échange h et λ ainsi que la loi de comportement définie en fonction de la température par les coefficients $C_{ijkl}(T)$ et les efforts $\overline{F_i}$. Parmi ces paramètres, on peut considérer que h, λ et les efforts $\overline{F_i}$ dépendent du système étudié, alors que les coefficients $C_{ijkl}(T)$ sont intrinsèques au matériau.

On pourra par exemple utiliser des expériences complémentaires (tractions simples à différentes températures) pour les déterminer.

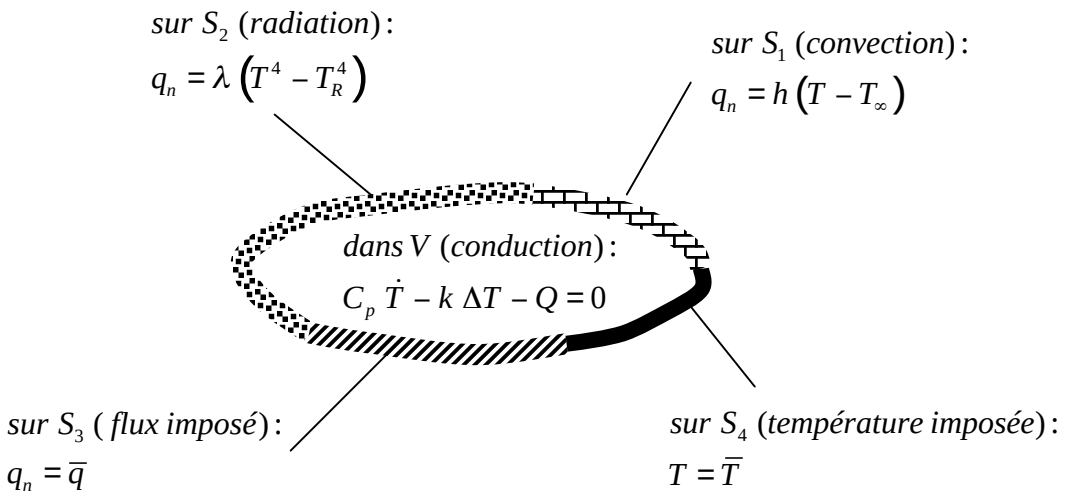
On cherchera donc à minimiser une fonction écart que l'on considère en moindre carrés sous la forme :

$$J\left(h,\lambda,C_{ijkl},\alpha,\overline{F_i}\right)= A\sum_{\substack{thermo\\couples}}\left(T_{j\,cal}\left(h,\lambda\right)-T_{j\,\exp}\right)^2 + B\sum_{jauges}\left(\varepsilon_{j\,cal}\left(h,\lambda,C_{ijkl},\alpha,\overline{F_i}\right)-\varepsilon_{j\,\exp}\right)^2 + C\sum_{tractions}\left(\sigma_{j\,cal}\left(C_{ijkl},\alpha,\overline{F_i}\right)-\sigma_{j\,\exp}\right)^2$$

A, B, C sont des constantes utilisées pour 'adimensionner' la fonction coût ou pour donner plus d'importance à certaines expériences.

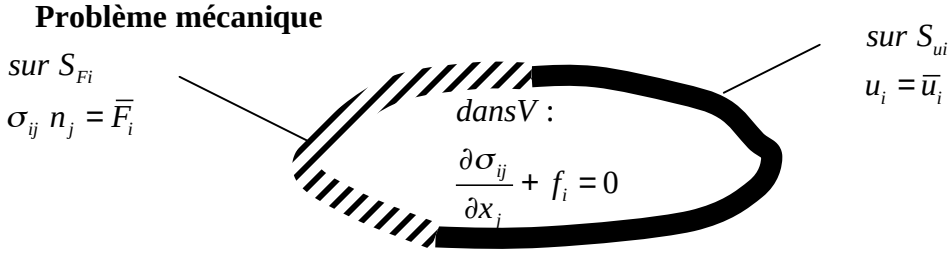
Problème direct en thermoélasticité

Problème thermique



$$q_n = -k\,\frac{\partial T}{\partial n} = -k\,\vec{\nabla} T\,.\vec{n}$$

Problème mécanique



Loi de Hooke :

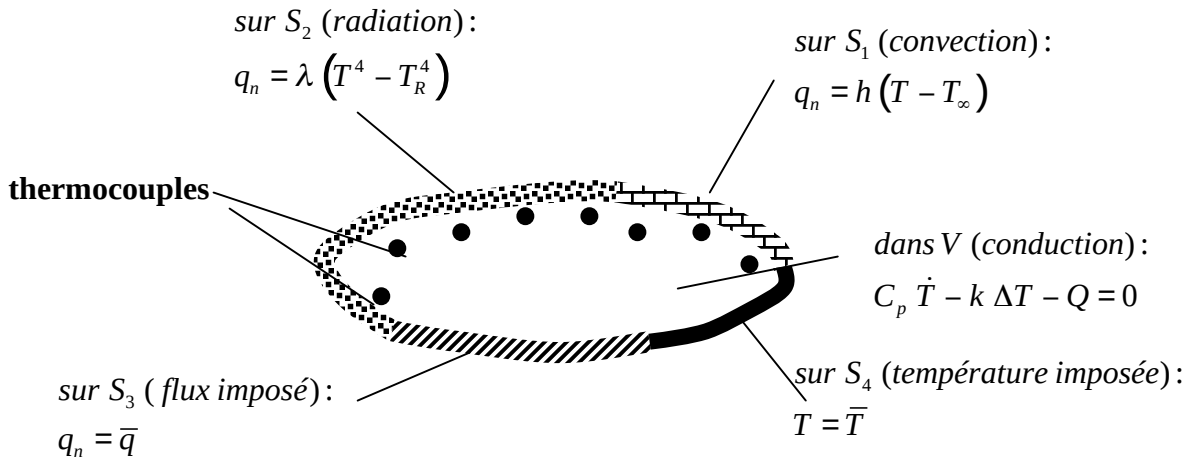
$$\sigma_{ij} = C_{ijkl}\,(T)\,(\varepsilon_{kl} - \alpha\, T\, \delta_{kl})$$

Connu:	Inconnu:	moyen
C_p , k , λ , h Q T_R , T_∞ , $\overline{T}$, $\overline{q}$	T , $\vec{q} = -k\,\vec{\nabla} T$	$[K]_T\,(T) = (F)_T$
$C_{ijkl}\,(T)$, $\alpha(T)$ $\vec{f}$ $\overline{F}_i$, $\overline{u}_i$	$\vec{u}$, σ_{ij}	$[K(T)]_u\,(u) = (F(T))_u$

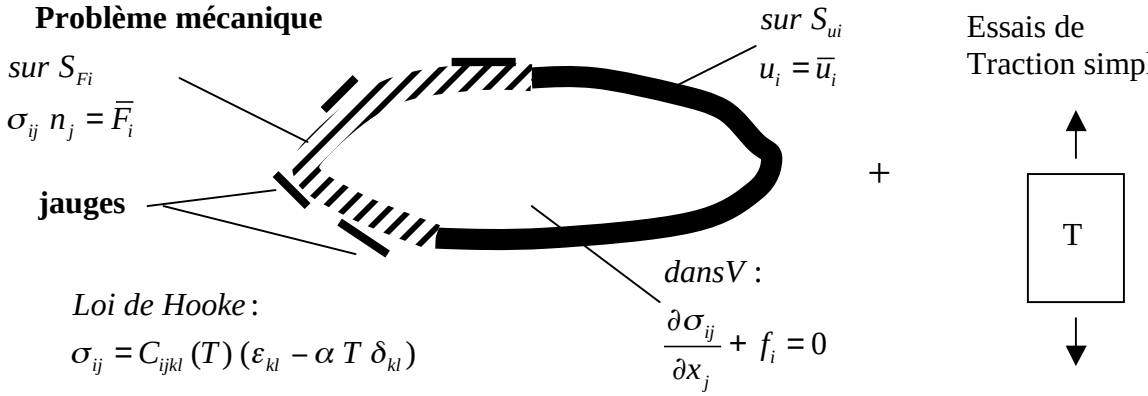
Figure 1.

Problème inverse en thermoélasticité

Problème thermique



Problème mécanique



+

Connu:	Inconnu:	moyen
C_p , k Q , $T(t=0,x)$ T_R , T_∞ , $\overline{T}$, $\overline{q}$ *températures aux thermocouples* $T_{i\exp}$	h , λ T , $\vec{q} = -k\,\vec{\nabla} T$	$\left[K(h,\lambda)\right]_T \left(T\right)_{cal} = \left(F(h,\lambda)\right)_T$ $\left[K(T(h,\lambda), C_{ijkl}, \alpha)\right]_u \left(u\right)_{cal} = \left(F(T(h,\lambda), C_{ijkl}, \alpha, \overline{F}_i)\right)_u$ *Courbes de traction* : $\sigma_{ij} = C_{ijkl}(T)\left(\varepsilon_{kl} - \alpha\, T\, \delta_{kl}\right)$
$\vec{f}$ $\overline{u}_i$ *déformations aux jauges* $\varepsilon_{i\exp}$ *Courbes de traction*	$C_{ijkl}(T)$, $\alpha(T)$ $\overline{F}_i$ $\vec{u}$, σ_{ij}	$\min_{h,\lambda,C_{ijkl},\alpha,\overline{F}_i} J\left(h,\lambda,C_{ijkl},\alpha,\overline{F}\right)$ $J\left(h,\lambda,C_{ijkl},\alpha,\overline{F}_i\right) = \sum_{\substack{thermo\\couples}} \left(T_{j\,cal}\left(h,\lambda\right) - T_{j\exp}\right)^2 + \sum_{jauges} \left(\varepsilon_{j\,cal}\left(h,\lambda,C_{ijkl},\alpha,\overline{F}_i\right) - \varepsilon_{j\exp}\right)^2 + \sum_{tractions} \left(\sigma_{j\,cal}\left(C_{ijkl},\alpha,\overline{F}_i\right) - \sigma_{j\exp}\right)^2$

Figure 2.

Ce problème d'optimisation est à priori sans limitation, sauf si on impose aux paramètres d'appartenir à un intervalle fixé à l'avance ou si on impose certaines relations d'égalité entre paramètres (pente à l'origine,…).

2. Mise en œuvre des procédures d'identification.

2.1.Interpolation et identification de relation entre deux paramètres

Soit un phénomène physique liant deux grandeurs (contrainte/déformation par exemple) que l'on note x et y. Des expériences nous ont fourni nexp couples (x_i, y_i). On peut envisager de relier x à y par une paramètrisation interpolante, c'est à dire que l'approximation est définie par nexp paramètres déterminés en faisant passer l'approximation par tous les points expérimentaux. Cette méthode peut provoquer de mauvaises approximations si l'on considère un grand nombre de points et une approximation polynomiale dont le degré croît avec nexp. Dans ce cas, la fonction interpolée peut subir de grandes variations entre les points expérimentaux. C'est pourquoi il vaut mieux considérer des approximations de degré peu élevé et plus de points expérimentaux que de paramètres et les identifier par minimisation de fonction écart (figure 3).

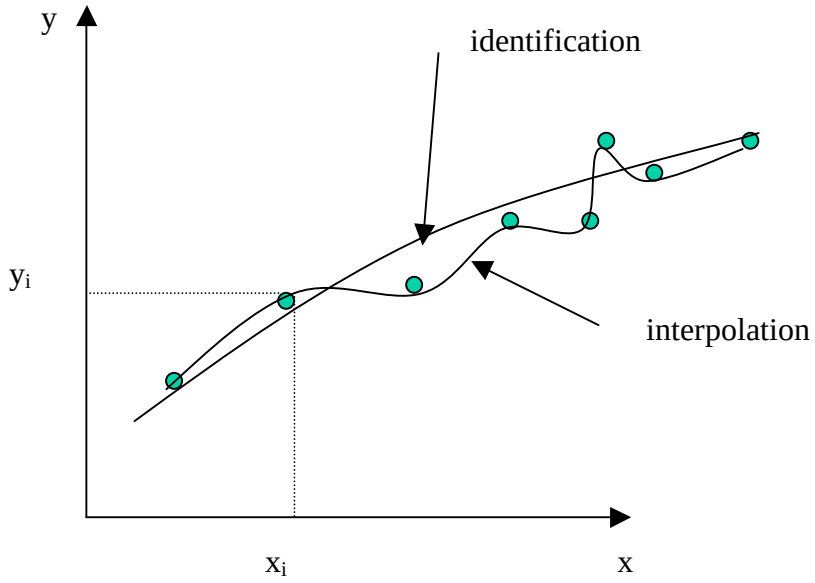


Figure 3.

Soit l'approximation

$$y = f(x, p_1,\ldots, p_n)$$

$p_1,\ldots, p_n$ étant les paramètres à identifier.
La fonction écart à minimiser est alors :

$$J(p_1, ..., p_n) = \sum_{n\exp} \left(f(x_i, p_1, ..., p_n \right) - y_i)^2$$

Si l'approximation est telle que :

$$y = \langle P(x) \rangle \left(p \right)$$

où $\langle P(x) \rangle$ *est une base polynomiale*
$\langle p \rangle$ *est le vecteur des paramètres*

La fonction écart s'écrit alors comme la norme du vecteur (V) :

$$(V)=[A](p)-(B)$$

où : $[A]=\begin{bmatrix}\langle P(x_1)\rangle \\ ... \\ \langle P(x_{n\exp})\rangle\end{bmatrix}$, $(B)=\begin{pmatrix}y_1 \\ ... \\ y_{n\exp}\end{pmatrix}$

La minimisation de J par rapport aux paramètres fournit alors le système à résoudre suivant :

$$[A]^T[A](p)=[A]^T(B)$$

Exemple : relation linéaire entre deux paramètres.

Dans ce cas, on a (figure 4) :

$$y=\langle P(x)\rangle(p)$$

où :

$$\langle P(x)\rangle=\langle 1 \quad x\rangle \text{ , } \langle p\rangle=\langle p_1 \quad p_2\rangle$$

Alors :

$$[A]=\begin{bmatrix}1 & x_1 \\ ..... & \\ 1 & x_{n\exp}\end{bmatrix}$$

On doit donc résoudre :

$$\begin{bmatrix}n\exp & \sum_{i=1}^{n\exp} x_i \\ \sum_{i=1}^{n\exp} x_i & \sum_{i=1}^{n\exp} x_i^2\end{bmatrix}\begin{pmatrix}p_1 \\ p_2\end{pmatrix}=\begin{pmatrix}\sum_{i=1}^{n\exp} y_i \\ \sum_{i=1}^{n\exp} x_i\, y_i\end{pmatrix}$$

Remarque : Si l'on désire définir une loi de comportement du type(figure 5) :

$$\sigma=C\,\varepsilon^m$$

où C et m sont deux paramètres à identifier, on se ramène au cas précédent en effectuant le changement de variables suivant :

$$y=L_n(\sigma), x=L_n(\varepsilon), p_1=L_n(C), p_2=m$$

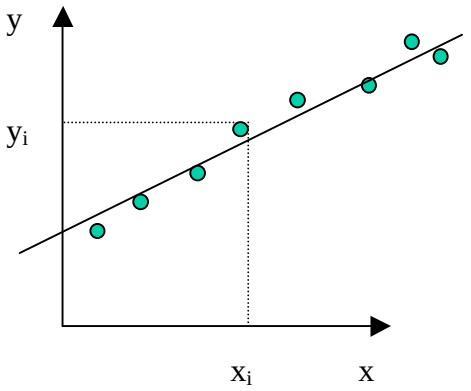


Figure 4.

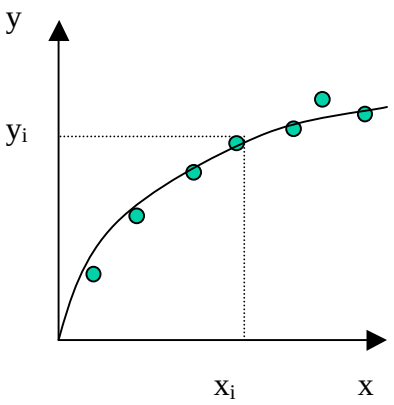


Figure 5.

2.2 Généralités sur le calcul des sensibilités

Le calcul des sensibilités $\frac{\partial J}{\partial p_i}$ de la fonction écart J par rapport aux paramètres p_i peut à priori être effectué par une des méthodes suivantes :

- méthode des différences finies
- méthode directe
- méthode de l'état adjoint

En général, on identifie plusieurs paramètres, aussi on privilégiera la méthode des différences finies (toujours possible si les systèmes à inverser sont bien conditionnés) ou la méthode de l'état adjoint si l'on accès aux sources du code de calcul. En effet la méthode directe requiert, par itération d'optimisation, autant de résolutions qu'il y a de paramètres, tandis que la méthode de l'état adjoint requiert autant de résolution qu'il y a de fonctions dont on recherche les sensibilités (une dans le cas de l'identification).

Les calculs associés à la méthode de l'état adjoint sont :

$$\frac{\partial J}{\partial p_i} = \left\langle \frac{\partial J}{\partial (d)} \right\rangle \frac{\partial (d)}{\partial p_i} \qquad (1)$$

où *(d)* est le vecteur des degrés de liberté du système. *(d)* est le vecteur des températures aux nœuds *(T)* et/ou le vecteur des déplacements aux nœuds *(u).*
(d) est obtenu par résolution du système suivant :

$$[K](d) = (F) \qquad (2)$$

où *[K]* et *(F)* sont les matrices de rigidité et vecteur second membre globaux.
Par dérivation de (2), on obtient :

$$\frac{\partial ([K](d))}{\partial p_i} = \frac{\partial ([K])}{\partial p_i}(d) + [K]\frac{\partial (d)}{\partial p_i} = \frac{\partial (F)}{\partial p_i} \qquad (3)$$

$\frac{\partial (d)}{\partial p_i}$ est ainsi obtenu par résolution du système suivant :

$$[K]\frac{\partial (d)}{\partial p_i} = \frac{\partial (F)}{\partial p_i} - \frac{\partial ([K])}{\partial p_i}(d) \qquad (4)$$

Introduisant (4) dans (1), on obtient :

$$\frac{\partial J}{\partial p_i} = \left\langle \frac{\partial J}{\partial (d)} \right\rangle [K]^{-1} \left(\frac{\partial (F)}{\partial p_i} - \frac{\partial ([K])}{\partial p_i}(d) \right)$$

Si l'on note par *(λ)* le vecteur état adjoint obtenu par résolution de :

$$[K](\lambda) = \left(\frac{\partial J}{\partial (d)} \right)$$

la sensibilité de *J* par rapport à chacun des paramètres est obtenue par :

$$\frac{\partial J}{\partial p_i} = \langle \lambda \rangle \left(\frac{\partial (F)}{\partial p_i} - \frac{\partial ([K])}{\partial p_i}(d) \right)$$

2.3. Exemple d'identification en thermique unidimensionnelle.

On considère le phénomène d'échange thermique dans un barreau de section circulaire. Celui-ci est soumis à ses extrémités à des conditions de température et de flux imposé. Sa surface latérale est soumise à des échanges par convection. On fait l'hypothèse que la température est uniforme dans une section.

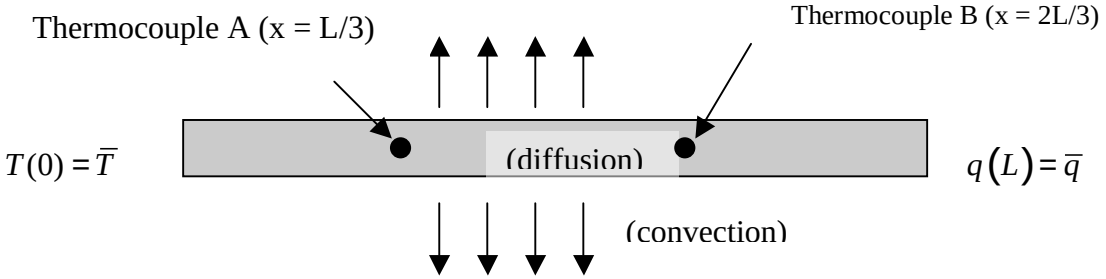
Dans ce cas, l'équation différentielle gouvernante appliquée au champ de température T(x) s'écrit :

$$-k\frac{d^2T}{dx^2}+\alpha\,T-Q=0$$

Les conditions aux limites s'écrivent :

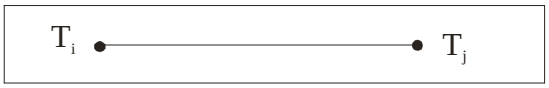
$$T(0)=\bar{T}$$

$$q(L)=-k\frac{dT}{dx}(L)=\bar{q}$$

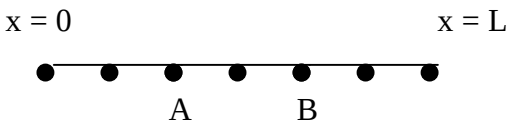


On cherche à identifier les valeurs de α et de $\bar{q}$ à l'aide de mesures de température en A et B aux cotes x = L/3 et x = 2L/3.

On réalise un modèle aux éléments finis linéaires à deux nœuds régulier dont le nombre d'éléments est un multiple de 3 (les points de mesure A et B coïncident avec des nœuds du maillage).



élément à deux nœuds



discrétisation à 6 éléments

On décrit ensuite le modèle aux éléments finis.

Modèle éléments finis

Forme forte

La forme forte associée à l'équation différentielle s'écrit :

$$W_F = \int_0^L T^* \left(-k \frac{d^2T}{dx^2} + \alpha\, T - Q \right) dx = 0$$

$$\forall T^*, T \text{ vérifiant } T(0) = \overline{T}, q(L) = -k \frac{dT}{dx}(L) = \overline{q}$$

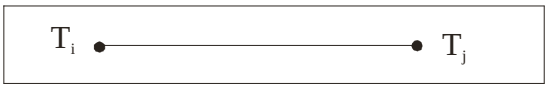
Forme faible

Après intégration par parties, la forme faible s'écrit :

$$W = \int_0^L T^* \left(k \frac{\partial T^*}{\partial x} \frac{\partial T}{\partial x} + \alpha\, T^* T - T^* Q \right) dx + T^*(L)\,\overline{q} = 0$$

$$\forall T^*, T \text{ vérifiant } T^*(0) = 0, T(0) = \overline{T}$$

Construction du modèle éléments finis



Sur un élément, T varie en fonction des degrés de liberté de l'élément $(T)_e$ et des fonctions de forme *(N)* par la relation :

$$T(x) = \langle N(x) \rangle (T)_e$$

où:

$$\langle N(x) \rangle = \left\langle \frac{x_j - x}{l} \quad \frac{x - x_i}{l} \right\rangle \qquad \langle T \rangle_e = \langle T_i \quad T_j \rangle \qquad l = x_j - x_i$$

Introduisant ces dernières relations dans la forme faible, on obtient:

$$W = \sum_e \langle T^* \rangle_e \left\{ \int_e \left(k \frac{\partial (N)}{\partial x} \frac{\partial \langle N \rangle}{\partial x} + \alpha\, (N) \langle N \rangle \right) dx\, (T)_e - Q \int_e (N) dx \right\} + T^*(L)\,\overline{q} = 0$$

$$\forall T^*, T \text{ vérifiant } T^*(0) = 0, T(0) = \overline{T}$$

Cette équation s'écrit pour un maillage de N éléments :

$$W = \sum_e \langle T^* \rangle_e \left\{ [k]_e (T)_e - (f)_e \right\} + T^*_{N+1}\,\overline{q} = 0$$

$$\forall (T^*), (T) \text{ vérifiant } T_1^* = 0, T_1 = \overline{T}$$

où :

$$[k]_e = \int_e \left(k \frac{\partial (N)}{\partial x} \frac{\partial \langle N \rangle}{\partial x} + \alpha\, (N) \langle N \rangle \right) dx = \frac{k}{l} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + \frac{\alpha l}{6} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$ (matrice de rigidité)

$$(f)_e = \int_e Q (N) dx = \frac{Q}{2} \begin{pmatrix} l \\ l \end{pmatrix}$$ (vecteur second membre élémentaire)

La forme faible s'écrit finalement :

$$W = \left\langle T^* \right\rangle \left\{ [K](T) - (F) \right\} = 0$$

$$\forall \left(T^*\right)(T) \text{ vérifiant } T_1^* = 0, T_1 = \overline{T}$$

D'où le système à résoudre :

$$[K](T) = (F)$$

$$avec : T_1 = \overline{T}$$

où :

$$[K] = \sum_{ass} [k]_e \qquad (F) = \sum_{ass} (f)_e + (F)_{cont}$$

$\sum_{ass}$ désigne une somme au sens de l'assemblage de vecteur ou de matrice élémentaires (utilisant les connectivités)

$(F)_{cont} = \begin{pmatrix} 0 \\ \dots \\ 0 \\ -\overline{q} \end{pmatrix}$ (contribution au contour)

Après réduction du système assemblé en introduisant la condition à la limite au nœud 1, on obtient le système à résoudre suivant :

$$\begin{bmatrix} \frac{2k}{l}+\frac{2\alpha l}{3} & -\frac{k}{l}+\frac{\alpha l}{6} & 0 & \dots & 0 & 0 \\ -\frac{k}{l}+\frac{\alpha l}{6} & \frac{2k}{l}+\frac{2\alpha l}{3} & -\frac{k}{l}+\frac{\alpha l}{6} & \dots & 0 & 0 \\ 0 & -\frac{k}{l}+\frac{\alpha l}{6} & \frac{2k}{l}+\frac{2\alpha l}{3} & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \frac{2k}{l}+\frac{2\alpha l}{3} & -\frac{k}{l}+\frac{\alpha l}{6} \\ 0 & 0 & 0 & \dots & -\frac{k}{l}+\frac{\alpha l}{6} & \frac{k}{l}+\frac{\alpha l}{3} \end{bmatrix} \begin{pmatrix} T_2 \\ T_3 \\ T_4 \\ \dots \\ T_N \\ T_{N+1} \end{pmatrix} = \begin{pmatrix} Ql + \left(\frac{k}{l} - \frac{\alpha l}{6}\right)\overline{T} \\ Ql \\ Ql \\ \dots \\ Ql \\ \frac{Ql}{2} - \overline{q} \end{pmatrix}$$

Calcul de sensibilité

Si le nombre d'élément N = 3m, alors le point de mesure A coïncide avec le nœud de numéro m+1 tandis que le point B correspond au nœud de numéro 2m+1.

La fonction écart s'écrit alors :

$$J = \left(T_{m+1} - T_A\right)^2 + \left(T_{2m+1} - T_B\right)^2$$

Le calcul de la sensibilité de J par rapport à α s'écrit d'après les relations du paragraphe 2.2.

$$\frac{\partial J}{\partial \alpha} = \left\langle \lambda \right\rangle \left(\frac{\partial (F)}{\partial \alpha} - \frac{\partial \left([K]\right)}{\partial \alpha}(T) \right)$$

où le vecteur adjoint (λ) est solution de :

$$[K](\lambda)=\left(\frac{\partial J}{\partial (T)}\right)$$

Dans notre cas, le vecteur $\left(\frac{\partial J}{\partial (T)}\right)$ s'écrit :

$$\left(\frac{\partial J}{\partial (T)}\right)=\begin{pmatrix} 0 \\ \dots \\ 2(T_{m+1}-T_A) \\ 0 \\ .... \\ 2(T_{2m+1}-T_B) \\ 0 \end{pmatrix}$$

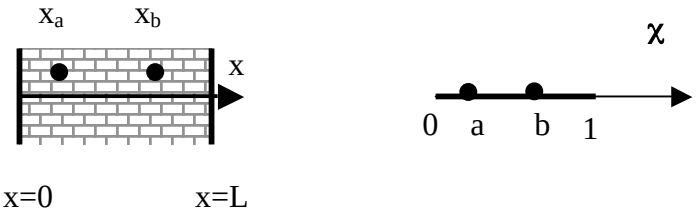
Le vecteur des pseudo-forces $\left(\frac{\partial (F)}{\partial \alpha}-\frac{\partial ([K])}{\partial \alpha}(T)\right)$ s'écrit :

$$\left(\frac{\partial (F)}{\partial \alpha}-\frac{\partial ([K])}{\partial \alpha}(T)\right)=\begin{pmatrix} -\frac{l}{6}\bar{T} \\ 0 \\ 0 \\ \dots \\ 0 \\ 0 \end{pmatrix}-\begin{bmatrix} \frac{2l}{3} & \frac{l}{6} & 0 & \dots & 0 & 0 \\ \frac{l}{6} & \frac{2l}{3} & \frac{l}{6} & \dots & 0 & 0 \\ 0 & \frac{l}{6} & \frac{2l}{3} & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \frac{2l}{3} & \frac{l}{6} \\ 0 & 0 & 0 & \dots & \frac{l}{6} & \frac{l}{3} \end{bmatrix}\begin{pmatrix} T_2 \\ T_3 \\ T_4 \\ \dots \\ T_N \\ T_{N+1} \end{pmatrix}$$

2.4. Exemple de problème mal posé

La résolution de problème inverse peut conduire à des systèmes algébriques mal conditionnés, en particulier en raison d'un mauvais choix de mesures (nature, position des capteurs, …).
Par exemple, considérons le problème académique suivant :
On cherche à évaluer les températures sur les deux faces d'une paroi à partir de 2 mesures de température obtenues par des capteurs disposés dans l'épaisseur de la paroi. Cele-ci a une épaisseur h. On utilise la coordonnée a-dimensionnelle $\chi = \frac{\mathbf{x}}{\mathbf{L}}$.
L'équation différentielle gouvernante (équation de la chaleur en régime établi sans source de chaleur) :

$$\frac{\partial^2 \mathbf{T}}{\partial \chi^2} = \mathbf{0}$$



La solution est une distribution linéaire de T dans l'épaisseur :

$$T(\chi) = (1-\chi)T_0 + \chi T_1$$

Les températures théoriques des capteurs sont :

$$T_a = (1-a)T_0 + a\,T_1 \qquad T_b = (1-b)T_0 + b\,T_1$$

Par inversion des relations précédentes, on a :

$$\begin{bmatrix} 1-a & a \\ 1-b & b \end{bmatrix}\begin{pmatrix} T_0 \\ T_1 \end{pmatrix} = \begin{pmatrix} T_a \\ T_b \end{pmatrix} \quad \rightarrow$$

$$T_0 = \frac{b}{b-a}T_a - \frac{a}{b-a}T_b \qquad T_1 = \frac{1-a}{b-a}T_a - \frac{1-b}{b-a}T_b$$

En fait on a mesuré $\tilde{T}_a = T_a \pm \varepsilon_a$ et $\tilde{T}_b = T_b \pm \varepsilon_b$, où ε_a et ε_b sont les erreurs de mesure. Les températures des faces externes de la paroi effectivement évaluées sont alors :

$$\tilde{T}_0 = \frac{b}{b-a}\tilde{T}_a - \frac{a}{b-a}\tilde{T}_b \qquad \tilde{T}_1 = \frac{1-a}{b-a}\tilde{T}_a - \frac{1-b}{b-a}\tilde{T}_b$$

Les erreurs d'évaluation sont alors :

$$\varepsilon_0 = \tilde{T}_0 - T_0 = \frac{b}{b-a}\varepsilon_a - \frac{a}{b-a}\varepsilon_b \qquad \varepsilon_1 = \tilde{T}_1 - T_1 = \frac{1-a}{b-a}\varepsilon_a - \frac{1-b}{b-a}\varepsilon_b$$

On peut alors faire les deux remarques suivantes :

- si les deux capteurs sont très proches $(b \rightarrow a)$, les erreurs sur les températures évaluées peuvent devenir très importantes. Les deux capteurs ont tendance à donner la même information , ce qui se traduit par l'obtention d'un système mal conditionné.
- Selon les valeurs relatives des erreurs de mesure on peut obtenir des évaluations plus ou moins erronées.
  Par exemple, si $\varepsilon_a = \varepsilon, \varepsilon_b = \varepsilon$, les erreurs sur les évaluations sont : $\varepsilon_0 = \varepsilon_1 = \varepsilon$ (même ordre de grandeur des erreurs).
  Par contre, si $\varepsilon_a = \varepsilon, \varepsilon_b = -\varepsilon$, les erreurs sur les évaluations sont :
  $\varepsilon_0 = \frac{b+a}{b-a}\varepsilon, \varepsilon_1 = \frac{2-b-a}{b-a}\varepsilon$, qui peuvent être beaucoup plus importantes que les erreurs de mesure.

MQ14

Mini projet
Identification en thermique

On considère l'exemple d'identification en thermique unidimensionnelle présenté dans le cours.

1. Etude de convergence du modèle E.F.
   Utiliser le programme de calcul écrit dans l'environnement MATLAB pour effectuer une étude de convergence de la solution en faisant varier la finesse du maillage en considérant les valeurs suivantes (en unités cohérentes) :

L	$\overline{q}$	$\overline{T}$	k	α	Q
1	1	1	1	1 et 5	1

2. Calcul de sensibilités.
   On développe une procédure numérique d'identification d'un ou de plusieurs paramètres du modèle (α ou $\overline{q}$ par exemple) par la résolution de problème inverse. On s'appuiera sur des résultats expérimentaux traduits par des températures mesurées aux points A et B (x=L/3 et x=2L/3)
   a. Déterminer les sensibilités des températures calculées en A et B par rapport à α et $\overline{q}$ par la méthode des différences finies avant pour les jeux de paramètres précédents en effectuant une étude de convergence des sensibilités en fonction des valeurs des perturbations $\Delta\alpha$ et $\Delta\overline{q}$.
   b. Comparer avec des calculs utilisant la méthode directe.
   c.
3. Procédure d'identification
   a. Développer des procédures d'identification de α en couplant le programme élément finis avec une procédure de minimisation en utilisant la méthode de l'état adjoint pour obtenir les sensibilités.
   b. Tester vos développements en utilisant pour valeurs expérimentales celles obtenues numériquement pour α =1. Prendre au début du processus itératif α =0 et vérifier que la procédure d'identification converge vers α =1.

4. Application
   Utiliser votre programme pour identifier α avec les données suivantes :

L	$\overline{q}$	$\overline{T}$	k	Q	T_A	T_B
1	1	1	1	1	0.7959	0.5640

Chap. IX - Optimisation en variables discrètes et méthodes d'optimisation globale

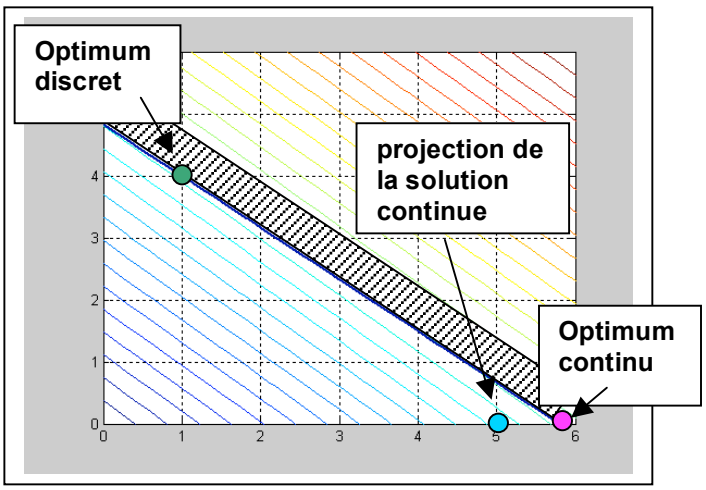
1. Introduction

Le cas des variables discrètes et celui de la recherche d'un minimum global sont deux types de problèmes très différents, indépendants l'un de l'autre, mais qui ont un point commun: ils ne peuvent pas, - ou mal - être résolus par les méthodes de descente classiques abordées dans la première partie du cours (gradient conjugué, quasi-Newton, pénalités, …). Ce chapitre présente des méthodes qui peuvent être utilisées pour résoudre ces deux catégories de problèmes.

Exemple de variables discrètes :
nombres de plis, de couches,
orientation des angles des fibres dans un composite,
épaisseurs, sections à prendre dans un catalogue de valeurs,
raidisseurs (présence ou non),
barres dans un treillis (présence ou non)
trous dans une pièce 2D, 3D,
....

Comme on ne peut pas définir le gradient d'une fonction par rapport à une variable discrète, les méthodes basées sur le gradient ne peuvent pas s'appliquer. Des méthodes d'ordre 0 telles que le simplexe de Nelder-Mead ne sont pas non plus utilisables puisque les points successifs sont calculés par des "pas" à valeurs continues (sommes de vecteurs).

Cependant, quand on doit prendre les valeurs des variables dans un catalogue, mais que les fonctions sont définies pour toutes les valeurs réelles des variables, on peut utiliser les méthodes classiques de descente pour trouver le minimum du problème continu, puis ensuite projeter cette solution sur le domaine discret (prendre les valeurs discrètes les plus proches), mais ce n'est pas toujours la meilleure solution comme le montre l'exemple ci-dessous (réf : M.Minoux, programmation mathématique, Dunod):



Min f(x1, x2) = $-10\ x_1 - 11\ x_2$
Avec : $10\ x_1 + 12\ x_2 \leq 59$
x_1 et x_2 entiers ≥ 0

L'optimisation en variables discrètes semble a priori plus facile qu'en continu, puisque le domaine admissible est "plus petit" . Mais il est disjoint (ensemble de points), les limitations d'égalité sont difficiles à satisfaire, et il n'y a pas forcément d'inégalités actives à l'optimum. Le problème est en fait plus difficile qu'en variables continues.

Il existe des méthodes spécifiques pour la recherche d'optimum en **variables discrètes** :
- calcul combinatoire : « explose » très vite ! (10 variables, 5 valeurs chacune : 5^{10} possibilités, soit environ $9.\ 10^{6}$
- méthodes de coupes,
- méthodes de recherche arborescente (« branch and bound »),

mais elles sont difficiles à généraliser.

Une autre difficulté en optimisation est le cas où il existe plusieurs, voire **un grand nombre de minima locaux** (fonction multimodale) . Même si les variables sont continues et les fonctions dérivables, les méthodes basées sur le calcul du gradient ne peuvent trouver que le minimum local le plus proche du point de départ. Il faut alors exécuter ces algorithmes avec différents points initiaux pour espérer trouver l'optimum global, ou au moins plusieurs optima locaux.

Les deux méthodes présentées dans ce chapitre: **recuit simulé** et **algorithmes évolutionnaires** peuvent en fait servir à résoudre à la fois le cas des variables discrètes et celui de l'optimisation globale. Ce sont deux catégories de méthodes stochastiques (utilisant des processus aléatoires, donc plusieurs exécutions d'un même programme pourront donner des solutions différentes…). Elles nécessitent bien souvent **un très grand nombre d'évaluations** de la fonction coût (des centaines, des milliers, des dizaines de milliers,...). Mais elles permettent d'apporter des solutions à des problèmes pour lesquels aucune autre méthode ne peut s'appliquer.

2. Méthode du recuit simulé

origine : analogie avec le refroidissement d'un métal : on le refroidit lentement (et même on le réchauffe) pour obtenir un métal sans défaut (un état d'énergie stable). La fonction coût est assimilée à une énergie, et on va descendre lentement vers le minimum, en autorisant la fonction à remonter de temps en temps, ou plus précisément: on construit une suite de points qui ne font pas systématiquement diminuer la fonction par rapport au point précédent, en espérant ainsi ne pas s'arrêter à un minimum local.
Cette méthode a été développée initialement par Metropolis et al. (1953), puis appliquée à l'optimisation par Kirkpatrick (1983) ; de nombreuses variantes existent.

2.1 Principe général de la méthode:

Soit $x = (x_1, \ldots, x_n)$ la variable, f(x) la fonction à minimiser, $g(x) = (g_1(x), \ldots, g_m(x))$ les limitations.
L'algorithme utilise un générateur de nombres aléatoires sur [0, 1] (loi uniforme)

- a) initialisations :
  point de départ x^0 admissible ; calculer $f(x^0)$;
  k=1 ; iter=0 ; j=1 ; $T = T_1$;
  choix de : it_max = nb max d'iterations pour T fixé
  j_max = nb max d'itérations sur T

- b) iter=iter+1 ; générer aléatoirement un point x^k ; calculer la fonction $f(x^k)$

- c) Tester :

si le point n'est pas admissible (limitations non satisfaites), le rejeter et aller en b)

si le point est admissible et meilleur que le précédent ($f(x^k) < f(x^{k-1})$) :

accepter x^k ; k = k+1 ; aller en d)

si le point est admissible et moins bon que le précédent ($f(x^k) > f(x^{k-1})$) : "quelquefois" accepter v^k , "quelquefois" le rejeter. Cette décision est prise avec une certaine probabilité : la <u>probabilité d'acceptation</u>, en utilisant le procédé suivant:

- génération d'un nombre z aléatoire, $0 \leq z \leq 1$
- calcul de: $\Delta f = f(x^k) - f(x^{k-1})$ $(\Delta f > 0)$
- calcul de $p_{Tj}(\Delta f) = \exp(\frac{-\Delta f}{T_j})$ ou T est un paramètre dit « température ».
- test (voir 2.2 et figure ci-dessous):

  si $z < p_{Tj}(\Delta f)$ → accepter le point x^k k = k+1 ; aller en d)

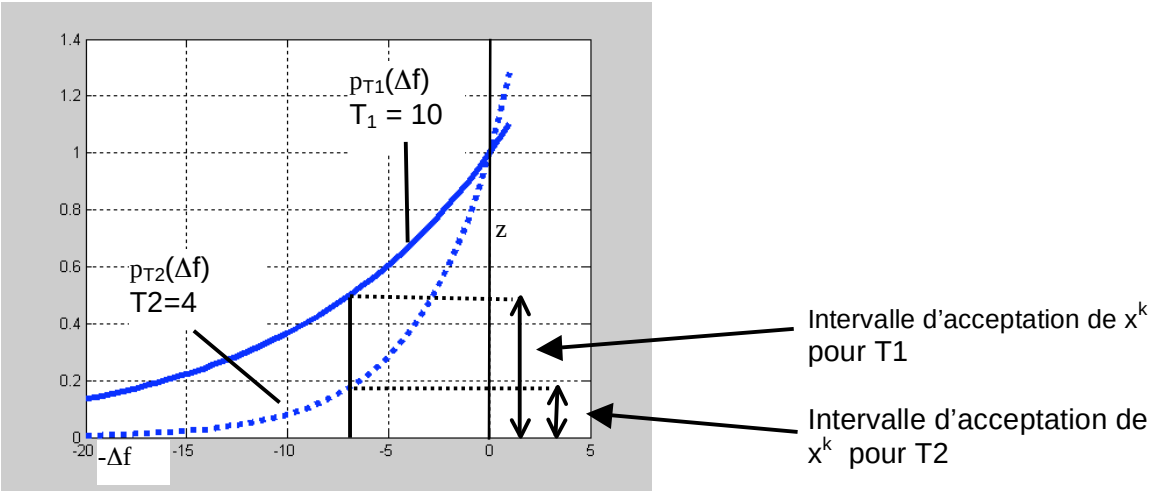
  si $z \geq p_{Tj}(\Delta f)$ → refuser le point x^k ; aller en b)

- d) Test d'arrêt des itérations:

si iter < it_max aller en b) (itérations à T fixée)
sinon,
si j < j_max : diminuer T : $T_{j+1} = \alpha * T_j$; j= j+1 ; aller en b)
sinon, fin ; prendre le meilleur des x^k comme solution x*

<u>2.2 Test d'acceptation d'un point qui fait augmenter f :</u>

T diminue au cours des itérations: ainsi on accepte de moins en moins de valeurs de x qui font augmenter la fonction d'une valeur Δf:



Pour T1 > T2 , une même valeur de Δf donne $p_{T1}(\Delta f) > p_{T2}(\Delta f)$

Un nombre aléatoire z qui satisfait le critère d'acceptation pour T1 : $z < p_{T1}(\Delta f)$ ne le satisfera pas forcément pour T2 < T1 ; un point qui fait augmenter f de Δf aura donc plus de « chances » d'être accepté pour T « grand », c'est-à-dire en début d'itérations sur j.

L'idée est d'éviter de rester près d'un minimum local : au début, la suite de points x^k peut explorer le domaine, mais il faut ensuite converger vers le minimum global : dans ce but, à la fin du processus, on retient essentiellement les points qui font diminuer f.

2.3 Choix des paramètres de l'algorithme

L'algorithme comporte deux niveaux d'itérations : sur la valeur de la température, et à T fixée, sur les perturbations de points.
Le choix de la température initiale T_1 doit permettre pratiquement tous les déplacements dans le domaine de conception.
Dans l'étape b), les points x_k sont générés par perturbations aléatoires d'un point précédent, d'amplitude maximale r_T fixée, soit décroissante au cours des itérations, soit ajustée en tenant compte de la proportion de points acceptés. Chaque variable peut être modifiée composante par composante, ou bien par des vecteurs aléatoires.
La prise en compte des variables à valeurs discrètes se fait lors de la génération des perturbations : ce sont par exemple les indices des valeurs admissibles dans une table qui sont utilisés.

3. Algorithmes évolutionnaires

Il s'agit d'une autre catégorie de méthodes inspirées d'un phénomène naturel : l'évolution des espèces vivantes. C'est initialement John Holland qui développa ce type de méthodes au milieu des années 70. Leur application à l'optimisation est plus récente (milieu des années 80). Ces algorithmes sont basés sur l'idée d'adaptation d'une population à son environnement. On utilise l'analogie :

1 point de l'espace de conception → 1 individu
un ensemble de points → une population, ou une génération
fonction objectif → adaptation au milieu

Ces méthodes ont pour principe de faire évoluer une population par trois opérateurs :

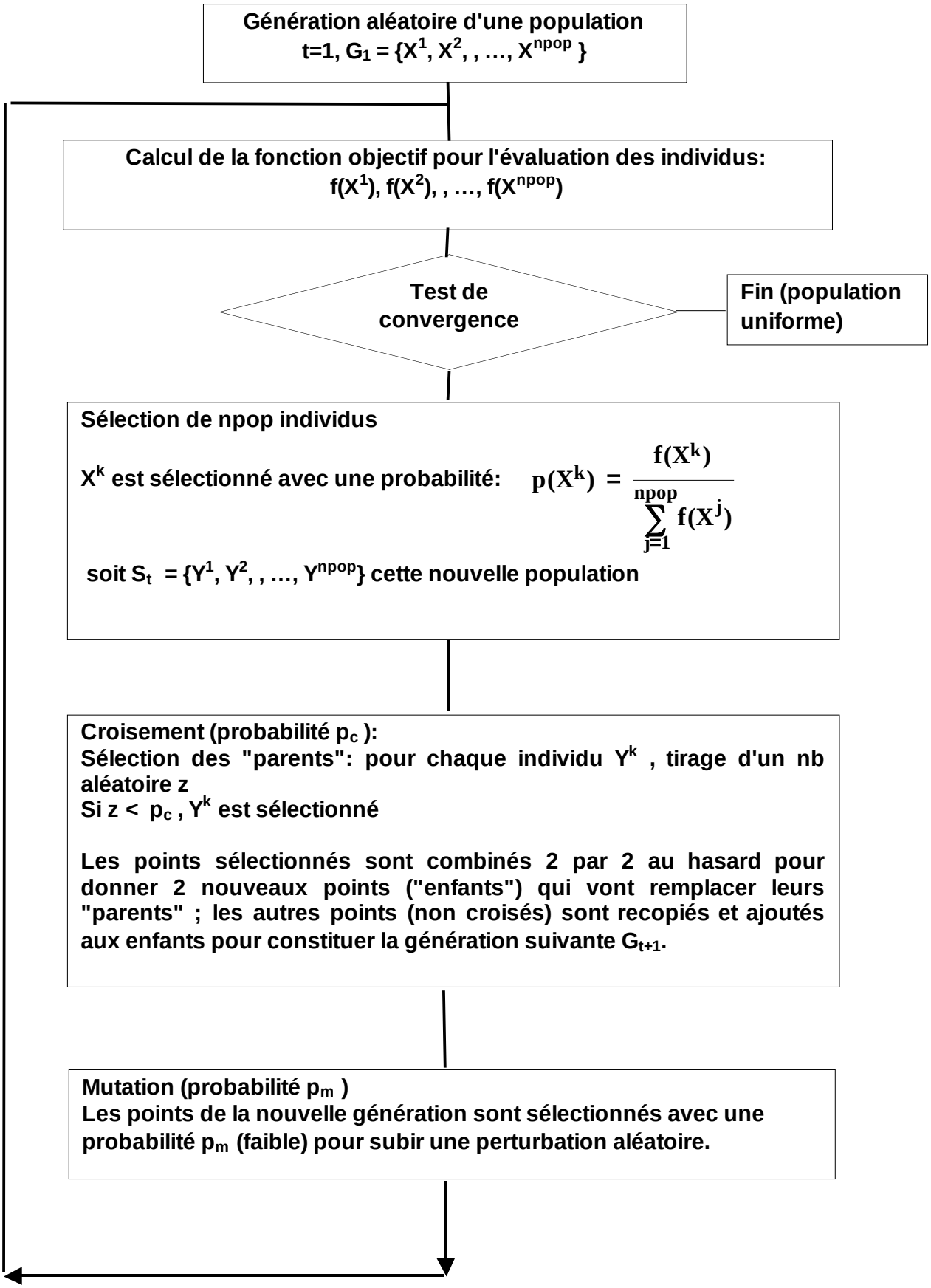
- sélection,
- croisement,
- mutation

de manière à générer une nouvelle population mieux adaptée (dont les individus ont de meilleures valeurs de la fonction objectif), ceci jusqu'à convergence des individus vers le ou les optimum(s). Il existe deux grandes classes de méthodes : algorithmes génétiques (GA) et stratégies d'évolution (ES)

Pour faciliter l'analogie avec la notion d'adaptation, on présente généralement ces algorithmes en supposant que la fonction objectif doit être maximisée et non minimisée ; il faut également qu'elle ne prenne que des valeurs positives (pour les calculs de probabilité de sélection). On peut toujours se ramener à ce cas en prenant F(X) = A - f(X) avec A = constante > 0

Les paragraphes suivants donnent les principes de base de ces méthodes. Il en existe une multitude de variantes…

3.1 Algorithmes génétiques. Un algorithme type peut être représenté comme suit :



Les caractéristiques générales de ces méthodes sont les suivantes :

- elles ne nécessitent que le calcul de la fonction objectif, et peuvent s'appliquer dans un cadre très général : pour des fonctions et/ou limitations non convexes, non dérivables, non continues, des variables discrètes ou continues (la prise en compte des limitations se fait la plupart du temps par une pénalisation.

- Leur convergence est généralement très lente (mais améliorations possibles par hybridation avec d'autres approches, nombreux travaux en cours)

- Il faut réaliser un compromis entre **« exploration »** (pour atteindre l'optimum global) et **« exploitation »** (pour converger)

- Elles sont facilement parallélisables (calcul simultané d'une population)

- Elles peuvent trouver le minimum ou le maximum global, ou plusieurs optima locaux.

Quelques opérateurs classiques sont décrits ci-dessous :

a) Sélection

C'est l'opérateur qui favorise l' « exploitation ». Les probabilités de sélection des points sont d'abord calculées, à partir des fonctions d'évaluation (coût).

$$p(X^k) = \frac{f(X^k)}{\sum_{j=1}^{npop} f(X^j)}$$

Ensuite on procède comme suit :

- au 1er point (quelconque) on associe l'intervalle $[0, p(X^1)]$,
- au suivant, l'intervalle $[p(X^1), p(X^2)]$,
- et ainsi de suite ; l'intervalle [0, 1] est ainsi divisé en segments S^k de longueur $p(X^k)$; S^k est attribué au point X^k .

et pour sélectionner m points, on effectue m fois :

- génération d'un nombre aléatoire z dans l'intervalle [0, 1]
- sélection du point X^k tel que $z \in S^k$

Certains points pourront ainsi être sélectionnés plusieurs fois, avec une probabilité dépendant de leur fonction objectif, et d'autres éliminés. Dans une utilisation classique, il est indispensable de vérifier que le meilleur point de chaque génération a été sélectionné. Dans des stratégies permettant de trouver plusieurs optima, il peut être utile de ne pas sélectionner trop de points proches les uns des autres.

Variantes :

- sélection par le rang : les points sont classés par ordre de leurs fonctions d'évaluation et p(X) est calculée en fonction des rangs et non des valeurs de f,

- sélection par tournoi : les points sont comparés par sous-ensembles, le meilleur de chaque sous-ensemble est sélectionné
- …

b) Croisement

Le but de cet opérateur est l'exploration : créer de nouveaux points à partir des points sélectionnés, qui seront si possible mieux « adaptés » que leurs parents, c'est-à-dire qui auront une meilleure fonction objectif.

Croisement réel barycentrique :

Soient X^1 et X^2 deux points à croiser ; on génère 2 points sur la droite X^1X^2:

$$Z' = a\, X^1 + (1- a)\, X^2$$

et

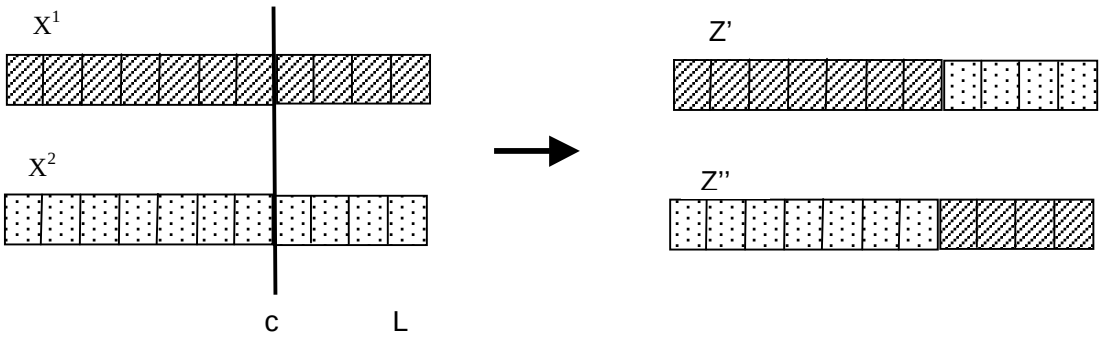
$$Z'' = (1 – a)\, X^1 + a\, X^2$$

où a est un nombre aléatoire
- compris entre 0 et 1 si on veut rester sur le segment X^1X^2
- compris entre - d et 1 + d (par ex. (-0.5, 1.5]) si on veut faciliter l'exploration du domaine.

Croisement binaire.

Les premières versions des algorithmes génétiques utilisaient une représentation binaire des variables, même si leurs valeurs pouvaient être continues dans le problème posé. Les intervalles de variation de chaque variable étaient discrétisés , permettant le **codage** des variables par une chaîne binaire (de longueur liée au nombre de valeurs possibles) appelée « chromosome » pour continuer l'analogie avec la biologie. Si cette phase de codage n'est plus utile pour des variables réelles, elle est conservée pour les variables discrètes : tout point peut être représenté par une liste de valeurs entières (par exemple des indices de valeurs des variables), qui peut être mise aussi sous forme binaire. Tout point est donc assimilé à une chaîne de « 0 » et de « 1 » de longueur L . Le croisement de 2 points consiste à leur faire échanger des parties de leurs « chromosomes » respectifs, coupés au hasard :

Exemple de croisement à 1 point :
Une valeur entière c comprise entre 1 et L est choisie aléatoirement. Les deux points X1 et X2 donnent les 2 « enfants » Z' et Z'' :



Dans ce cas, une variable au plus est modifiée entre les parents et les enfants.
Variantes : croisement à 2 points, croisement uniforme (sur chaque variable)

c) Mutation

Elle est appliquée à un faible pourcentage des points, choisis aléatoirement. Son rôle est d'augmenter la diversité de la population.

Mutation réelle : une perturbation gaussienne est ajoutée à X :

$$X \rightarrow X + N(0, \sigma)$$

L'écart-type σ peut être variable au cours des itérations (décroissant pour exploiter plus, croissant pour explorer plus).
Mutation binaire : transformation d'un 0 en 1 ou inversement, à un emplacement choisi aléatoirement entre 0 et L.

Test de convergence

Il est prudent d'avoir un test sur le nombre maximum de calculs de la fonction objectif, au cas où …
Mais on arrête généralement les itérations lorsqu' il n'y a plus d'amélioration de la fonction objectif depuis plusieurs générations successives ; tous les points de la population ont alors convergé vers la même solution (perte de la diversité, il ne resterait plus que la mutation pour évoluer). La convergence d'un algorithme génétique se représente souvent par le tracé des 2 courbes : meilleur f et f moyen au cours des générations.

Attention : à cause du caractère stochastique de la méthode, le résultat d'une seule exécution n'est pas significatif (la solution pourrait se trouver dans la première génération !) : il faut toujours présenter des moyennes de plusieurs « runs ».

3.2 Stratégies d'évolution (ES).

Les principales différences entre AG et ES sont les suivantes :

- la taille de la population peut être plus faible que dans les AG (jusqu'à moins de 10 individus pour les AE, plutôt 20 à quelques centaines pour AG).

- dans les AG, 2 parents ont 2 enfants, et les enfants remplacent les parents ; dans les ES, μ parents ont λ enfants, avec $\lambda >> \mu$

- la sélection est déterministe : les μ meilleurs points, sans tirage aléatoire, parmi les λ enfants ou bien parmi les $(\lambda + \mu)$ parents + enfants

- Le croisement est un opérateur essentiel pour les AG, peu important pour les ES.

- Au contraire, la mutation est un opérateur essentiel pour les ES, peu important pour les AG. Certaines versions des ES utilisent une mutation adaptative : l'écart-type σ devient alors une variable supplémentaire, ajoutée aux variables de conception ; elle va elle-même être optimisée.

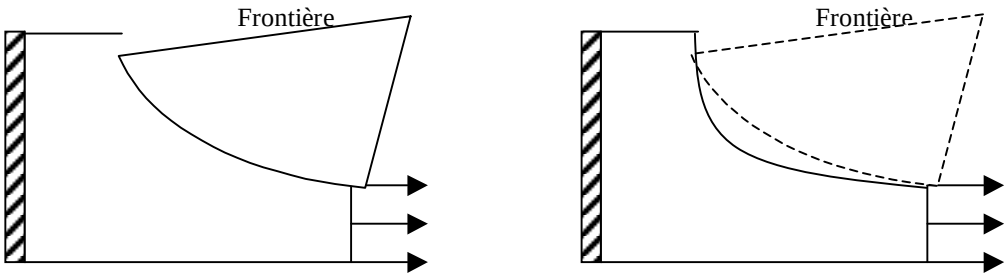
Chapitre X - Optimisation topologique

1. Introduction

En Mécanique, on distingue 3 classes de problèmes d'optimisation selon la nature des variables de conception :

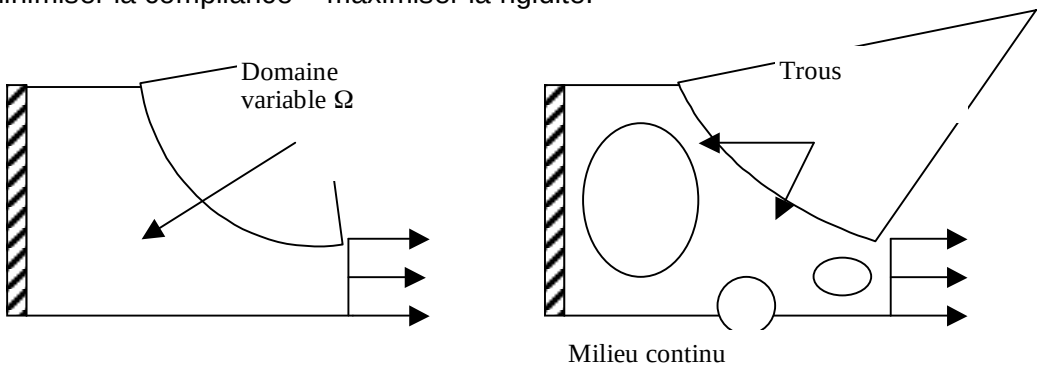
- Le dimensionnement : les variables sont des sections, des épaisseurs, …, toutes quantités qui peuvent évoluer sans nécessiter de modification du maillage en éléments finis

- L'optimisation de forme : l'inconnue est un contour (2D) ou une surface (3D). Par rapport au dimensionnement, une difficulté supplémentaire vient de ce qu'il faut gérer l'évolution d'une géométrie : sa description est paramétrée par les variables, et le maillage doit être mis à jour à chaque nouvelle valeur de ces variables (chaque itération).



- L'optimisation topologique : l'inconnue est bien sûr la topologie de la pièce, ce qui signifie que l'optimisation peut faire apparaître ou disparaître des trous, des renforts, des changements de connexions entre éléments (enlever ou ajouter une barre,…). C'est une méthode intéressante surtout dans la phase de pré-conception d'une pièce nouvelle ou pour laquelle on veut remettre en question la conception habituelle. Actuellement (2004), l'optimisation topologique peut donner des tendances, mais pas directement une forme adaptée à la fabrication, car on ne sait pas encore tenir compte précisément des impératifs technologiques. En pratique, il faut généralement ajouter une étape d'optimisation de forme une fois que la topologie optimale a été déterminée approximativement.

Comme en dimensionnement et en optimisation de forme, la fonction objectif à minimiser et les limitations à satisfaire sont le plus souvent liées à la masse, aux contraintes ou aux déplacements. Le critère le plus utilisé est la compliance (compliance = travail des forces extérieures = énergie élastique), qui constitue une mesure globale de la rigidité de la structure: minimiser la compliance = maximiser la rigidité.

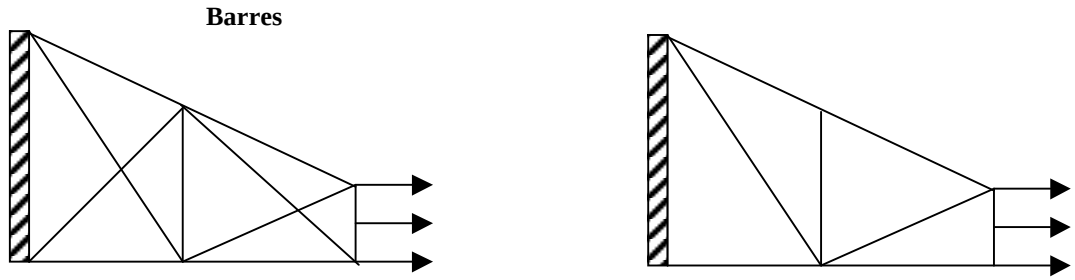


pour un milieu continu :

On se donne :

un domaine qui contiendra la solution,
les chargements,
les conditions aux limites,
éventuellement des restrictions (dimension max d'un trou, ...)

On doit déterminer s'il y a de la matière ou du vide en chaque point de l'espace, c'est-à-dire chercher la position et la forme des trous qui peuvent être introduits pour alléger la structure tout en lui conservant une rigidité suffisante.



pour un treillis

Pour un treillis de poutres ou de barres dont les points sont donnés, le problème consiste à chercher les meilleures connections (rigidité maximale, masse minimale) parmi toutes les connections possibles. C'est donc un problème discret.

2.Position du problème d'optimisation topologique

On ne considère ici que des problèmes d'élasticité linéaire.
La recherche d'une topologie optimale est un problème complexe à formuler et à résoudre, en particulier parce qu'il est dans ce cas plus difficile de représenter les transformations de la structure par des variables continues. En effet, dans le cas d'un milieu continu, la solution possède un nombre de frontières inconnu a priori, et il n'est pas possible de décrire une topologie quelconque avec un nombre donné de courbes paramétrées (de connexions pour des structures constituées barres).

Idéalement, chaque point du domaine peut être constitué de vide ou de matière; lorsque la structure est modélisée par éléments finis, un paramétrage naturel consiste à associer à chaque élément une variable binaire: 1:matière, 0: vide. Malheureusement, le problème ainsi obtenu présente deux difficultés:

- c'est un problème discret de grande taille (nombre d'éléments, nombre de connexions possibles) et on ne connaît pas de méthodes de minimisation efficaces dans ce cas. Les algorithmes évolutionnaires peuvent néanmoins résoudre certains problèmes de faible taille.

- c'est un problème "mal posé", ce qui signifie qu'il n'y a pas forcément convergence lorsque le maillage est affiné. On peut interpréter cela en disant qu'il est plus intéressant mécaniquement de faire un "très grand nombre" de trous "très petits". On voit apparaître des solutions présentant le phénomène de " damier " dont la taille est liée à la discrétisation éléments finis.

Pour surmonter ces difficultés, il existe essentiellement deux approches :

a) intégrer les damiers dans la formulation : on introduit la notion de **problème relaxé** :

- Pour les treillis, on permettra aux sections de varier continûment de 0 à s.
- Pour les milieux continus, on étend le domaine de conception à des matériaux composites, qui contiendront les solutions en damier, en utilisant la théorie de l'**homogénéisation** qui permet de calculer et d'optimiser les propriétés effectives des matériaux composites. Les variables seront alors **continues**. Cette formulation a donné lieu à de nombreux développements et à des logiciels performants; elle est présentée rapidement dans le paragraphe 3.

b) restreindre l'espace de conception pour éviter l'apparition des damiers. On constate que, à surface de trous égale, plus le nombre de trous est élevé, plus le périmètre de la région solide augmente :

Exemple : dans un carré, on compare les périmètres obtenus pour une même surface de matière, avec 1 trou circulaire, 4 trous, …
pour une surface totale des trous = πr^2

1 trou de rayon	r	périmètre = $2 \pi r$
4	r/2	périmètre = $4*2 \pi r/2 = 4\pi r$
…..		
n	$r/\sqrt{n}$	périmètre= $2 \pi r \sqrt{n}$

L'introduction d'une borne supérieure sur le périmètre réduira donc le nombre de trous et s'opposera à l'apparition de damiers. Cette limitation peut être introduite par pénalisation dans un algorithme évolutionnaire.

3. Formulation relaxée: homogénéisation

Origine : la recherche d'un domaine constitué d'une partition macroscopique de vide et de matière est un problème mal posé au sens mathématique. Un moyen de le transformer en problème bien posé est d'introduire la notion de microstructure (matériau composite formé de matière et de vide). La variable est la densité (continue) du matériau de base, et la solution deviendra alors indépendante du maillage éléments finis. Le problème revient à chercher les composantes du tenseur d'élasticité E_{ijkl} parmi un ensemble de tenseurs admissibles correspondant au type de microstructure choisi.

Le problème de minimisation de la compliance peut s'écrire :

Minimiser l(u)= [redacted]

par rapport aux variables E_{ijkl}

avec u solution de : [redacted] pour tout $v \in V$

et pour un volume total de matériau V° fixé :

[redacted] °

avec : F_v = forces de volume, f_s = tractions de surface

u est solution des équations d'équilibre avec une matrice de rigidité définie par les modules E_{ijkl} : $K(E_{ijkl})\ u = F$

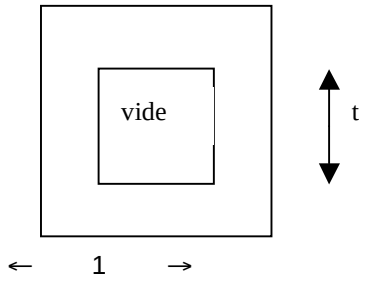
Résolution :

- on choisit tout d'abord le type de microstructure, ce qui détermine les variables au niveau de la cellule de base du matériau composite

- le milieu macroscopique est supposé constitué d'une infinité de cellules qui se répètent périodiquement. Par une procédure d'homogénéisation, on calcule les valeurs des propriétés du matériau au niveau macroscopique E_{ad}

- on résout le problème d'optimisation pour $E \in E_{ad}$ = ensemble des tenseurs de rigidité des composites ayant la microstructure (m) choisie.

Exemple dans le cas 2D, si on choisit une microstructure (m) carrée telle que représentée ci-dessous :



Microstructure (m)

- Au niveau de la microstructure, les variables sont :

  - la densité de matière μ,

    $\mu = 1 - \alpha$, avec $\alpha = t^2$, où t est la taille du trou

  - et l'orientation de la cellule θ dans la structure (modèle anisotrope). Les constantes du matériau dépendent donc de μ et θ :

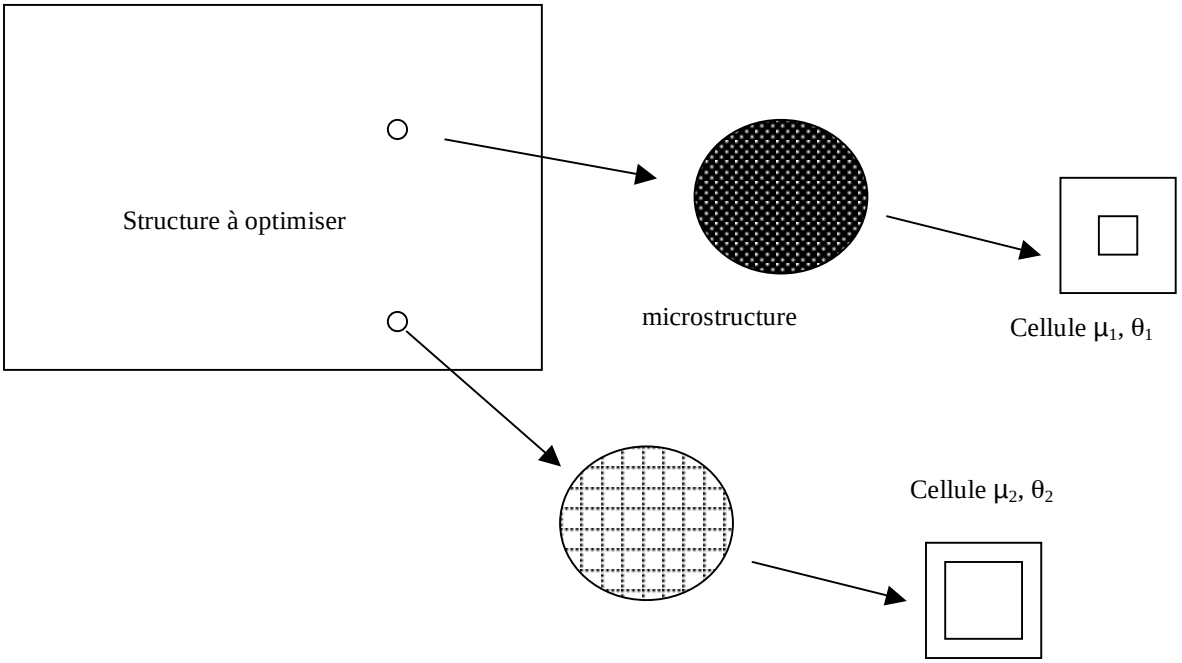
On calcule $E_{ijkl}(\mu)$, par une procédure d'homogénéisation (analyse éléments finis d'une région constituée d'un très grand nombre de cellules) en quelques valeurs de μ entre 0 et 1 et on interpolera pour pouvoir évaluer $E_{ijkl}(\mu)$ en toute valeur.

La dépendance de E_{ijkl} par rapport à la rotation de la cellule θ est donnée par des formules analytiques.

- Au niveau de la structure à optimiser, on définit un maillage, et on associe à chaque élément une variable μ et une variable θ.

$\mu=0$ correspond au vide, $\mu=1$ à de la matière, et $0 < \mu < 1$ à un matériau composite poreux.

Avec cette formulation, l'existence d'une solution est garantie.



Mais on obtient un grand nombre de variables, et la méthode serait difficile à appliquer si on devait passer par un algorithme de minimisation de type gradient. Son intérêt vient de ce qu'il est possible d'établir les conditions d'optimalité (Kuhn et Tucker) du problème et de les **résoudre itérativement, d'une manière explicite** : ainsi chaque variable μ et θ.est mise à jour par une formule simple utilisant les résultats d'une analyse éléments finis ; les valeurs des multiplicateurs de Kuhn et Tucker doivent être eux aussi calculés itérativement par une formule explicite.

La solution obtenue à convergence, qui est donc stable par rapport au maillage, doit ensuite être transformée pour éliminer les densités intermédiaires (quelques itérations de pénalisation).
Cette méthode est utilisée par exemple dans le logiciel OPTISTRUCT ® (ALTAIR, travaux de recherche Bendsoe, Kikuchi) et s'avère assez efficace même sur de gros problèmes. Une approche utilisant divers types de microstructures (laminés séquentiels) a été développée par G. Allaire au CEA.

De nombreuses approches inspirées de l'homogénéisation existent, parmi lesquelles celle proposée par Bensoe et Sigmund. Les variables de conception sont alors les densités des éléments, et les propriétés mécaniques sont calculées explicitement (au lieu de la procédure d'homogénéisation) en fonction des propriétés d'un élément de densité 1. Cette approche simplifiée est décrite dans le paragreaphe suivant.

4. Un exemple de méthode d'optimisation topologique : TOPOPT

(Voir le site http://www.topopt.dtu.dk où une documentation sur cette méthode, des références et le programme " Topopt ", 99 lignes de Matlab, sont disponibles). Le principe est le suivant :

On se place en élasticité 2D, les variables sont des densités x_e des éléments du maillage; on minimise la compliance à volume imposé ; les densités sont comprises entre 2 bornes; l'équation d'équilibre apparaît comme une limitation du problème d'optimisation:

Minimiser $C = F^T U$
$V = f V°$

$0 < xmin \leq x_e \leq xmax$ pour e= 1, … ,N

K U = F

F: chargement

U: déplacements nodaux

f: définit le volume imposé par rapport au volume initial qui est rectangulaire, avec tous les éléments de densité égale à 1

xmin est une valeur faible correspondant à l'élimination d'un élément, (qui, en pratique, ne sera pas détruit afin qu'il n'y ait pas singularité de K).

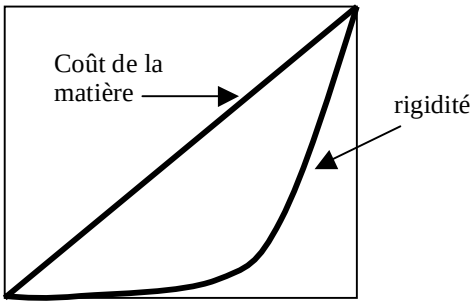
x_e: variables à optimiser, 1 densité pour chaque élément

xmax: densité maximale = 1

Si on prend un domaine rectangulaire, maillé en carrés de côté 1, une seule matrice de rigidité élémentaire K° doit être calculée, pour l'élément de référence de densité 1.
Pour une densité x_e , la matrice de rigidité de l'élément devient :

$$K_e = (x_e)^p K°$$

où K° est la matrice de rigidité d'un élément de densité 1 et Ke celle d'un élément de densité x_e



L'exposant p pénalise les densités intermédiaires.
En effet, avec q=2 ou q=4, les densités intermédiaires (entre 0 et 1) ne sont pas « intéressantes » pour l'algorithme de minimisation de la compliance à volume fixé car elles donnent une rigidité faible pour un coût de matière important. Cette formulation permet d'éliminer de la solution finale les densités qui ne sont ni 0 ni 1.

On utilise les conditions de stationnarité du Lagrangien associé au problème pour établir une formule de mise à jour des variables ; celle-ci équivaut à rendre uniforme sur le domaine la densité d'énergie de déformation.

L'effet de damier est évité par une procédure de lissage des sensibilités : sur chaque élément e, la sensibilité de la compliance par rapport à la variable locale x_e est remplacée par une valeur moyenne pondérée des sensibilités calculées sur les éléments voisins.

Pour stabiliser la convergence et satisfaire les bornes, la variation des variables est limitée à chaque itération (région de confiance, ou « move-limits »).

5. Approche par critère d'optimalité : méthode ESO (evolutionnary structural optimization).

Il s'agit d'un autre type d'approche, que l'on peut rapprocher de la méthode du "fully stressed design" car elle s'appuie sur un concept intuitif.
Principe : en enlevant progressivement la matière inutile, la forme (topologie) évolue vers une forme meilleure (Steven et Xie). Suivant la fonction à minimiser et les limitations à satisfaire, le critère d'élimination de la matière sera différent. Le cas le plus simple est celui qui est basé sur le niveau des contraintes locales : on supprime progressivement les éléments sur lesquels les contraintes (von Misès) sont faibles.

En contraintes planes par exemple, on procédera de la manière suivante :

1. maillage de la structure initiale. La topologie optimale sera à l'intérieur de ce domaine.

2. application des charges et des conditions aux limites

3. analyse : calcul des déplacements, des contraintes. On évalue sur chaque élément la contrainte de von Mises :

$\sigma_e^{vm} =$

et la valeur maximale de σ_e^{vm} sur tout le domaine σ_{max}^{vm}

4. calcul, sur chaque élément, du rapport :

Re =

5. Suppression des éléments où $Re < RR_i$
c'est-à-dire : $\sigma_e^{vm} < RRi * \sigma_{max}^{vm}$

où le paramètre RR_i est le " rapport de rejet " de l'itération courante

6. Tests de convergence :

- si on vient de détruire des éléments, aller en 3) pour refaire une analyse avec le nouveau maillage

sinon,

- si aucun élément n'a une contrainte inférieure d'au moins p% (25% par exemple) à la contrainte maximale, arrêter. Ceci signifie que tous les éléments restant sont utiles.

- sinon, modifier le coefficient RR_i :

$RR_{i+1} = RR_i + ER$ avec ER = " evolution rate "

et aller en 3.

Deux paramètres sont à définir pour exécuter cette procédure très simple : RR_{i0} et ER ; des valeurs de 1% conviennent généralement ; sinon, elles sont à déterminer par quelques essais.

Exemple : structure à 2 barres, structure avec contraintes de surface uniforme

Généralisation de la méthode ESO :

Cas de charge multiples : la méthode s'étend facilement. L'étape 3 comprend plusieurs résolutions avec des seconds membres différents, mais la même matrice de rigidité. L'élimination des éléments se fait lorsque leurs " rapports de rejet " pour chaque charge est inférieur au seuil fixé.

Conditions aux limites multiples : idem, mais il faut cette fois autant d'analyses éléments finis que de cas de conditions aux limites.

Limitations sur la rigidité :

On utilise la compliance :

$$C=1/2\ \{F\}^T \{U\} \qquad \text{avec U solution de } [K]\{U\} = \{F\}$$

Minimiser C revient à maximiser la rigidité globale de la structure ; on imposera une limite sur C :

$$C < C°$$

Lorsqu'on enlève l'élément i, de rigidité K^i, on a une matrice de rigidité globale K* et:

$$\Delta K = K^* - K = - K^i$$

On suppose que le chargement ne change pas quand on enlève l'élément i :

$$K^* (U + \Delta U) = F \qquad \text{soit :} \quad (K + \Delta K) (U + \Delta U) = F$$

$$K U + \Delta K\ U + K \Delta U + \Delta K \Delta U = F$$

en négligeant le terme du 2ème ordre $\Delta K\ \Delta U$, on a :

$$\Delta U = -K^{-1} \Delta K\ U$$

et la variation de l'énergie de déformation s'écrit :

$$\Delta C= ½\ F^T \Delta U = -1/2\ F^T K^{-1} \Delta K\ U$$

$U = K^{-1}\ F$ et donc $F^T K^{-1} = U^T$
$\Delta K = - K^i$

donc on a :

$$\Delta C = ½\ U^T\ K^i\ U = ½\ U^{iT} K^i\ U^i$$

avec U^i = vecteur déplacement du ième élément

Pour rechercher la structure la plus légère possible ayant une rigidité suffisante, on définit

$\alpha i = \frac{1}{2} U^{i\,T} K^{i} U^{i}$

(énergie de déformation de l'élément) comme un indicateur de sensibilité pour choisir les éléments à éliminer. La procédure est analogue au cas précédent.

Limitations sur les déplacements :

Pour imposer une condition telle que :

[redacted] $< U°$

on suit le même raisonnement que dans le cas d'une limite sur la rigidité, mais cette fois on calcule les indicateurs de sensibilité en considérant des charges unitaires :

Fj =(0 , 0, …, 0, 1, 0, ….) (j ème composante =1)

$Fj^{T} \Delta U = \Delta Uj = Fj^{T} K^{-1} Ki\ U = Uj^{T} Ki\ U = Uji^{T} Ki\ Ui$

($\Delta U = -K^{-1} \Delta K\ U$ avec $\Delta K = Ki$ = matrice de l'élément i, 0 ailleurs)

où Uj est le déplacement dû à la charge Fj,
Ui et Uij sont les composantes de U et de Uj correspondant au i ème élément.

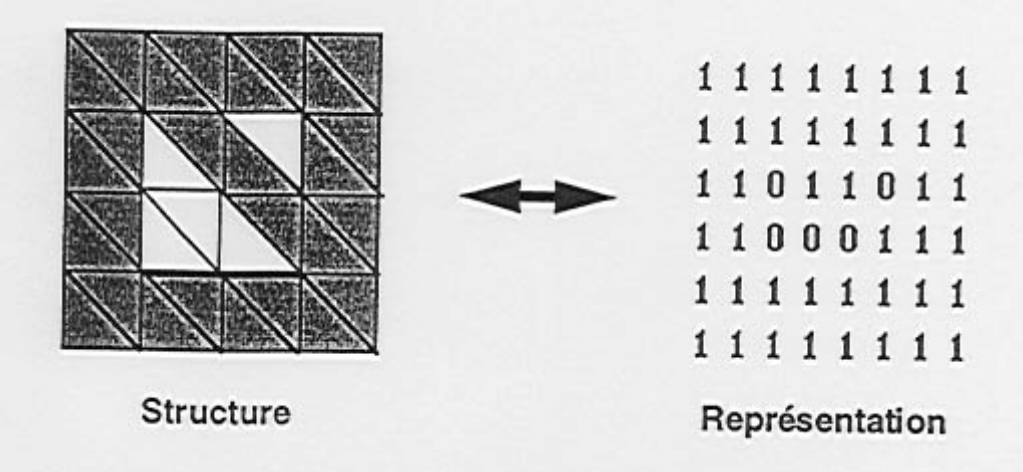
L'indicateur de sensibilité (changement de Uj quand on enlève le ième élément) est ici :

$\alpha i = Uji\ Ki\ Uj$

6.Optimisation topologique par algorithmes évolutionnaires (AG, ES)

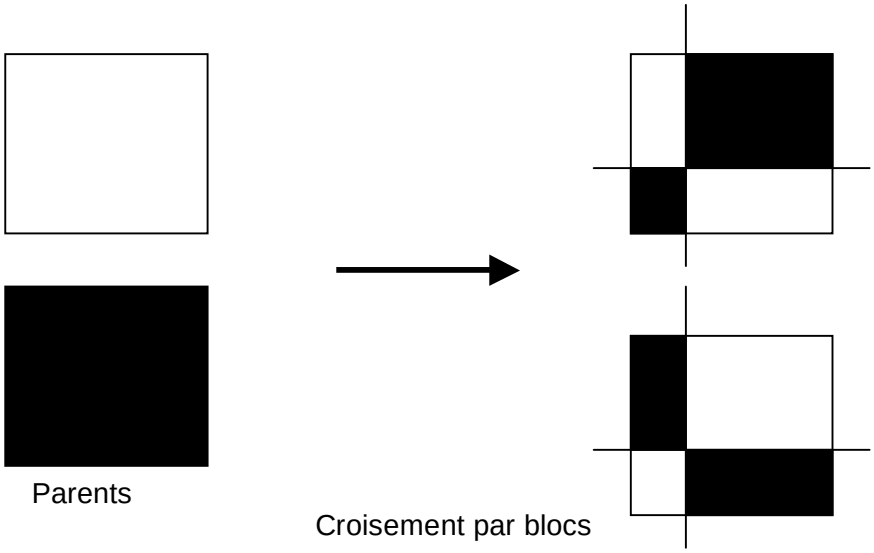
Pour simplifier la présentation, on se place dans le cas de structures planes (plaque). On peut généraliser aux coques ou au 3D, ou au cas de poutres ou de barres bien qu'il soit un peu différent.

La méthode la plus simple consiste à représenter le domaine par un modèle discret (un maillage fixe, Ne éléments) ; les éléments seront présents ou non (une variable binaire 0/1 associée à chaque élément). Certains éléments portant les charges ou les conditions aux limites doivent être toujours présents pour que la structure optimisée réponde bien aux spécifications imposées (Ni éléments). Un " individu " dans la terminologie AG représente une topologie particulière, et est constitué d'une chaîne binaire dont la longueur (Ne - Ni) est proche du nombre d'éléments du maillage (voir figure).



La fonction d'évaluation est construite à partir de la fonction objectif, et les limitations s'il y en a, sont traitées par pénalité.

L'opérateur de croisement a pour but de générer de nouvelles topologies en combinant deux topologies déjà évaluées et sélectionnées comme " parents ". Le croisement binaire standard peut être appliqué, mais il existe des opérateurs mieux adaptés (croisement diagonal ou par blocs pour des problèmes de plaques par exemple, voir figure)



Le code d'éléments finis associé doit être capable de gérer l'assemblage sélectif de la matrice de rigidité et du vecteur chargement, c'est-à-dire de prendre en compte les seuls éléments présents dans l'individu (le nombre de degrés de liberté et donc les dimensions de la matrice changent). Il se pose aussi le problème des parties non connexes : dans la version de base des AG, les opérateurs de croisement et de mutation peuvent faire apparaître des éléments isolés. Il faut avoir une procédure de correction, pour les éliminer ou bien les " reconnecter " au reste de la structure.