

MTH 6416 Optimisation avancée: Optimisation de boîtes noires

Charles Audet

École Polytechnique de Montréal
Département de mathématiques et de génie industriel

Août 2006

MTH 6416 – Optimisation de boîtes noires – sommaire

- 1 Introduction
- 2 Recherche par coordonnées
- 3 Optimisation non-lisse
- 4 Recherche par motifs
- 5 Contraintes générales non-lisses
- 6 Recherche sur treillis adaptatifs
- 7 Discussion

MTH 6416 – Optimisation de boîtes noires – sommaire

- 1 Introduction
 - Famille de problèmes d'optimisation
 - Optimisation multi-disciplinaire
 - Conditions d'optimalité
- 2 Recherche par coordonnées
- 3 Optimisation non-lisse
- 4 Recherche par motifs
- 5 Contraintes générales non-lisses
- 6 Recherche sur treillis adaptatifs
- 7 Discussion

MTH 6416 – Optimisation de boîtes noires – sommaire

- 1 Introduction
 - Famille de problèmes d'optimisation
 - Optimisation multi-disciplinaire
 - Conditions d'optimalité
- 2 Recherche par coordonnées
 - Algorithme de recherche par coordonnées
 - Trois variantes algorithmiques
 - Convergence de la recherche par coordonnées dans le cas différentiable
 - Traitement de contraintes
- 3 Optimisation non-lisse
 - Limitations de la recherche par coordonnées
 - Dérivée et sous-gradient
 - Calcul de Clarke: dérivée et gradient généralisés
 - Analyse de convergence de la recherche par coordonnées
- 4 Recherche par motifs
 - La sonde locale et recherche globale
 - Algorithme de recherche par motifs
 - Analyse de convergence de la recherche par motifs
 - Fonctions substitués
- 5 Contraintes générales non-lisses
 - Limitations de la recherche par motifs
 - Quatre types de cônes
 - Conditions nécessaires d'optimalité
- 6 Recherche sur treillis adaptatifs
 - Algorithme de recherche sur treillis adaptatifs - barrière
 - Analyse de convergence - barrière
 - Algorithme de recherche sur treillis adaptatifs - filtre
 - Analyse de convergence - filtre
- 7 Discussion

Famille de problèmes d'optimisation

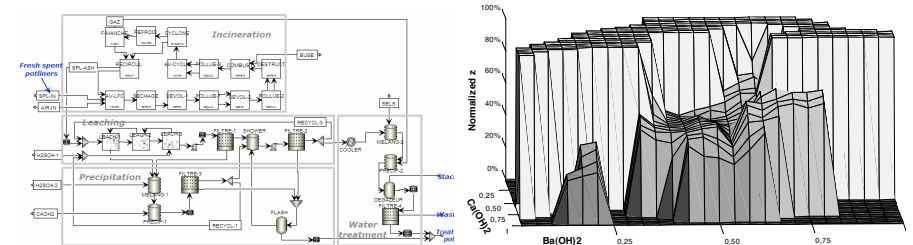
$$(NLP) \quad \min_x f_p(x) \\ \text{s.c. } x \in \Omega,$$

où $\Omega = \{x \in X : C_p(x) \leq 0\}$ et $X \subset \mathbb{R}^n$ est le domaine, et $f_p : X \rightarrow \mathbb{R} \cup \{\infty\}$, $C_p : X \rightarrow (\mathbb{R} \cup \{\infty\})^m$.

- p est le vecteur des paramètres contextuels. p est fixe.
- x est le vecteur de variables à déterminer. [variables de design, d'optimisation, de contrôle].
- X est défini par les contraintes devant nécessairement être satisfaites. X est fréquemment simple, un polyèdre par exemple. Ce sont les contraintes fermées.

Exemple : traitement des brasques – production d'aluminium

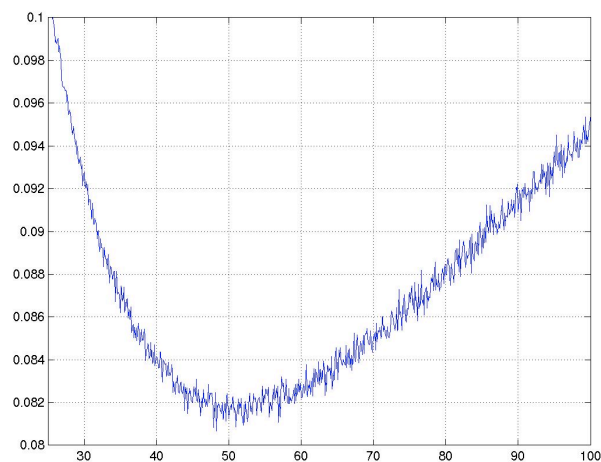
L'évaluation de f_p et Ω requiert l'exécution du logiciel ASPEN qui simule un procédé de production.



7 variables, 4 contraintes de boîtes noires. Aspen échoue sur 43% des évaluations (plancher du graphique).

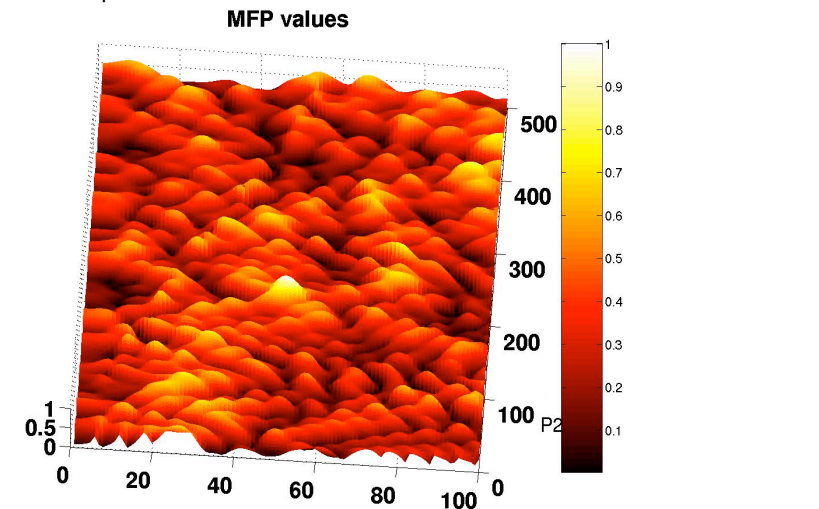
Exemple: Planification de maintenance

Une fonction bruitée.

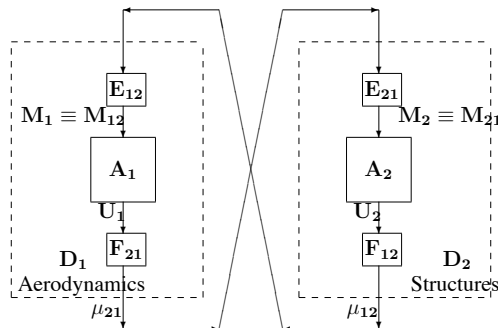


Exemple: Acoustique sous-marine

Plusieurs optima locaux.



MDA = Analyse Multidisciplinaire



- Disciplines D_1, D_2 avec des codes A_1, A_2
- μ_{ij} est l'information de D_j requise par D_i (Exemple, transformer un flot en pressions)

Formulation Mathématique

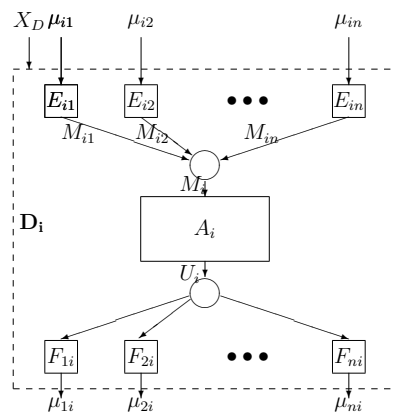
Écrire $F(x, u) = 0$, mais résoudre $F(x, M_1, M_2) = 0$ pour M , où

$$F_1(x, M_1, M_2) = E_{21} \circ F_{21} \circ A_1(M_1) - M_2,$$

$$F_2(x, M_1, M_2) = E_{12} \circ F_{12} \circ A_2(M_2) - M_1$$

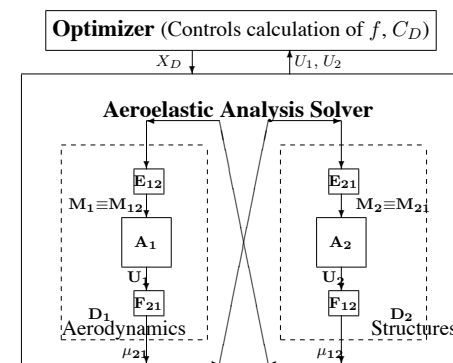
- A_1 et A_2 sont les fonctions coûteuses à évaluer

Une discipline dans un problème MDO



Les routines E et F sont difficiles à écrire

Formulation *boîte noire*



- Requiert MDA complète à chaque évaluation

Conditions d'optimalité

Condition nécessaire d'optimalité

Si $\hat{x} \in \Omega$ est un minimum local d'une fonction différentiable f sur un ensemble convexe $\Omega \subset \mathbb{R}^n$, alors

$$f'(\hat{x}; d) \geq 0 \quad \forall d \in T_{\Omega}(\hat{x}),$$

où $f'(\hat{x}; d) = \lim_{t \rightarrow 0} \frac{f(\hat{x} + td) - f(\hat{x})}{t} = d^T \nabla f(\hat{x})$
et $T_{\Omega}(\hat{x})$ est le cône tangent de Ω en $\hat{x}$.

Condition nécessaire d'optimalité

Si $\hat{x} \in \Omega$ est un minimum local d'une fonction f sur un ensemble $\Omega \subset \mathbb{R}^n$, alors

$$f^{\circ}(\hat{x}; d) \geq 0 \quad \forall d \in T_{\Omega}^H(\hat{x}),$$

où $f^{\circ}(\hat{x}; d)$ est une généralisation de la dérivée directionnelle,
et $T_{\Omega}^H(\hat{x})$ est une généralisation du cône tangent.

Sommaire de la présentation

- 1 Introduction
- 2 Recherche par coordonnées
 - Algorithme de recherche par coordonnées
 - Trois variantes algorithmiques
 - Convergence de la recherche par coordonnées dans le cas différentiable
 - Traitement de contraintes
- 3 Optimisation non-lisse
- 4 Recherche par motifs
- 5 Contraintes générales non-lisses
- 6 Recherche sur treillis adaptatifs

Optimisation sans contraintes

Considérons le problème

$$\min_{x \in \mathbb{R}^n} f(x)$$

où $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$.

Hypothèse

Nous disposons d'un point initial $x_0 \in \mathbb{R}^n$.

Cas possibles :

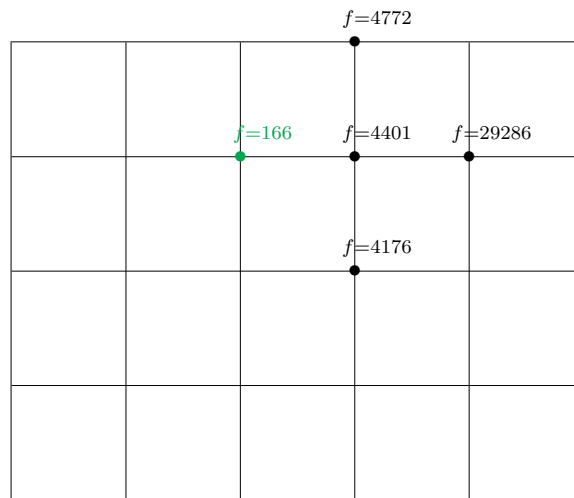
- x_0 est la solution courante du système.
- x_0 est à déterminer.
- $x_0 = 0$

Recherche par coordonnées (ancêtre de pattern search).

- INITIALISATION:
 - x_0 : point initial de $\mathbb{R}^n$
 - $\Delta_0 > 0$: taille initiale du pas.
- SONDE LOCALE (POLL STEP): pour $k = 0, 1, \dots$
 - Si $f(t) < f(x_k)$ pour un $t \in P_k := \{x_k \pm \Delta_k e_i : i \in N\}$,
alors poser $x_{k+1} = t$
et $\Delta_{k+1} = \Delta_k$;
 - sinon x_k est un minimum local par rapport à P_k ,
poser $x_{k+1} = x_k$
et $\Delta_{k+1} = \frac{\Delta_k}{2}$.

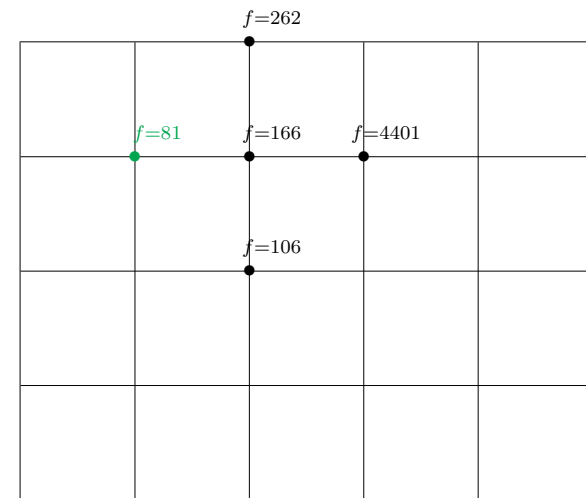
Recherche par coordonnées : algorithme de base

$$x_0 = (2, 2)^T, \Delta_0 = 1$$



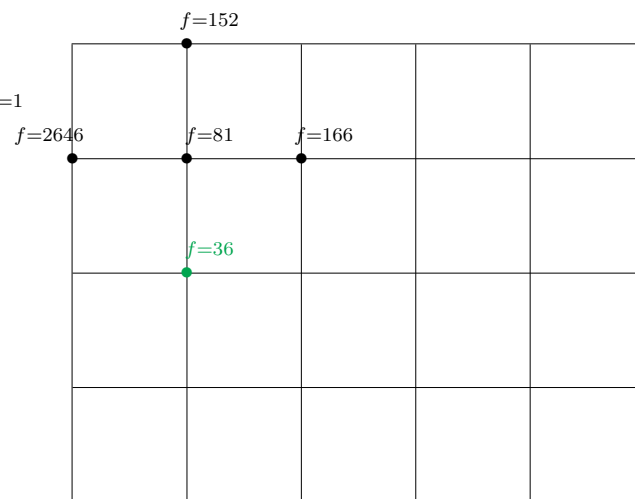
Recherche par coordonnées : algorithme de base

$$x_1 = (1, 2)^T, \Delta_1 = 1$$



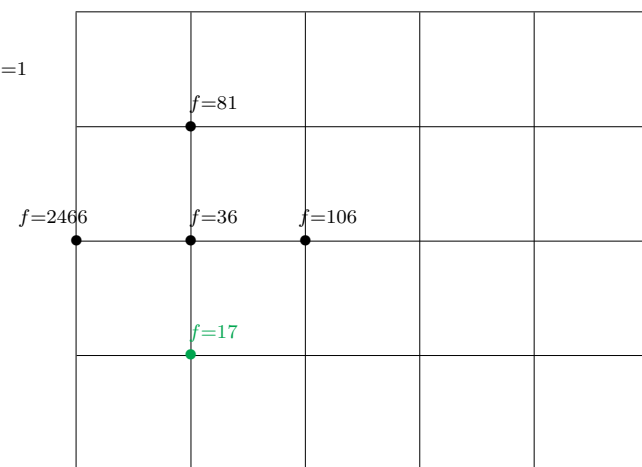
Recherche par coordonnées : algorithme de base

$$x_2 = (2, 2)^T, \Delta_2 = 1$$



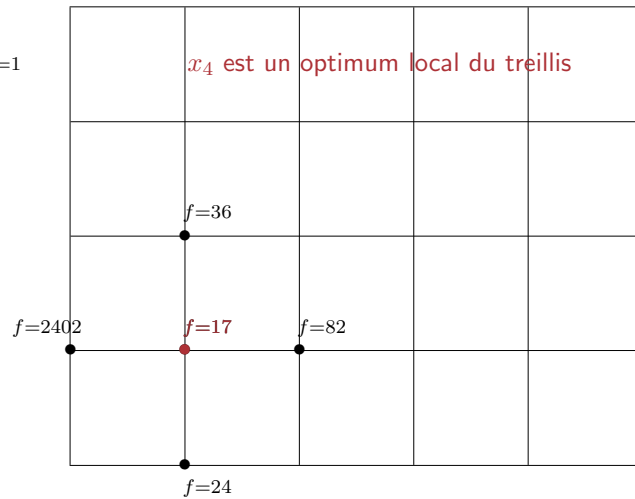
Recherche par coordonnées : algorithme de base

$$x_3 = (0, 1)^T, \Delta_3 = 1$$



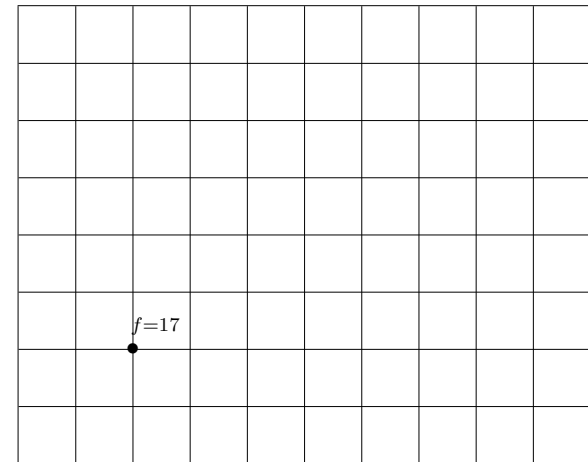
Recherche par coordonnées : algorithme de base

$$x_4 = (0,0)^T, \Delta_4 = 1$$



Recherche par coordonnées : algorithme de base

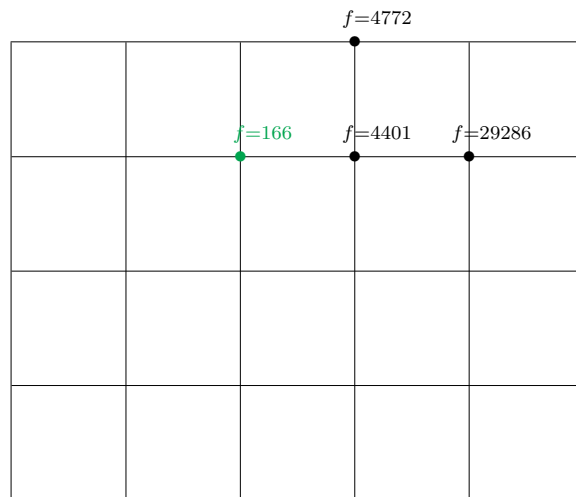
$$x_5 = (0,0)^T, \Delta_5 = \frac{1}{2}$$



Avec un budget de 20 évaluations on obtient $x = (0,0)^T$ avec $f(x) = 17$. Peut-on faire mieux ?

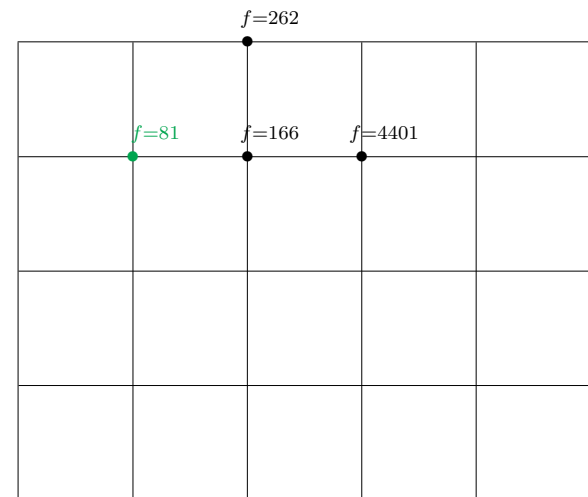
Recherche par coordonnées : algorithme opportuniste

$$x_0 = (2,2)^T, \Delta_0 = 1$$

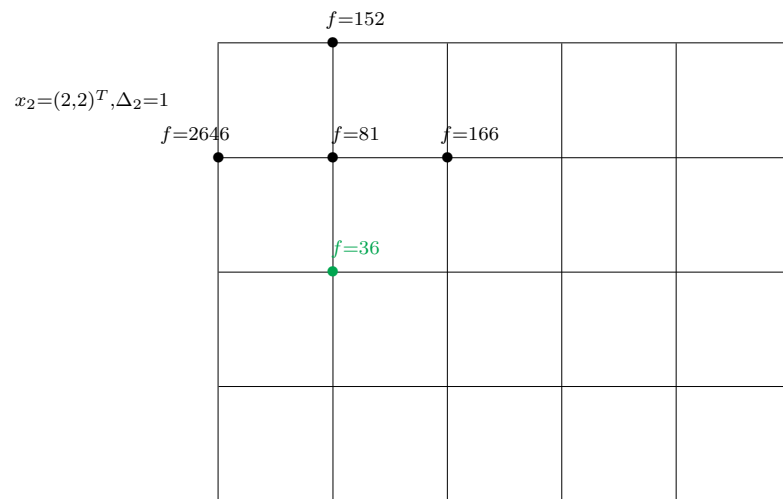


Recherche par coordonnées : algorithme opportuniste

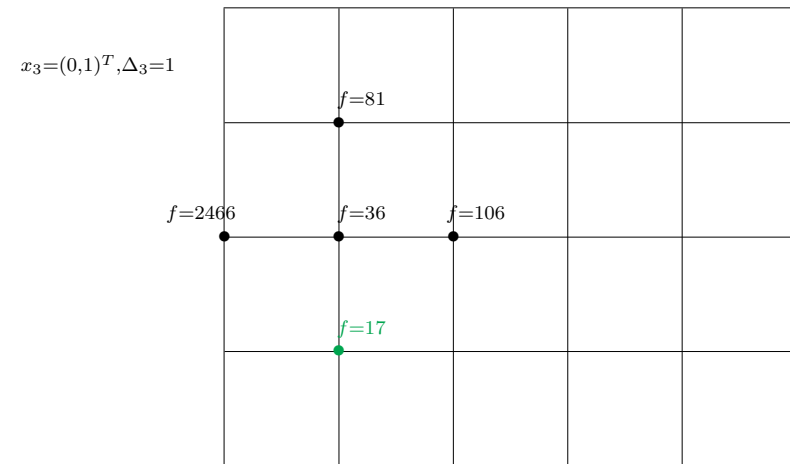
$$x_1 = (1,2)^T, \Delta_1 = 1$$



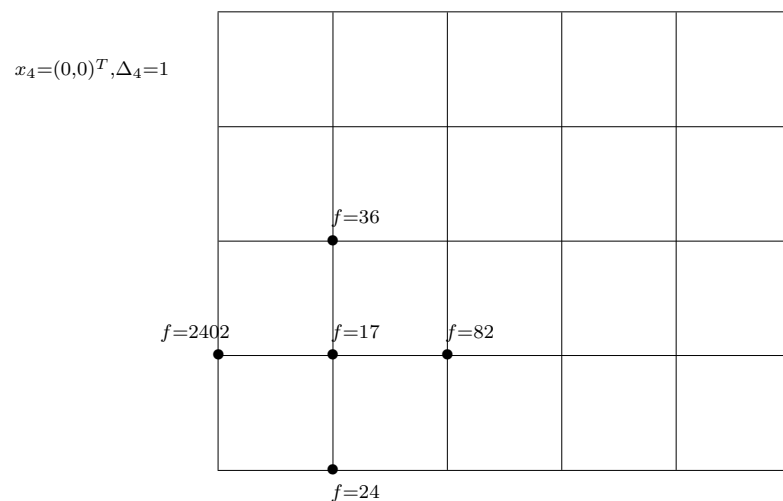
Recherche par coordonnées : algorithme opportuniste



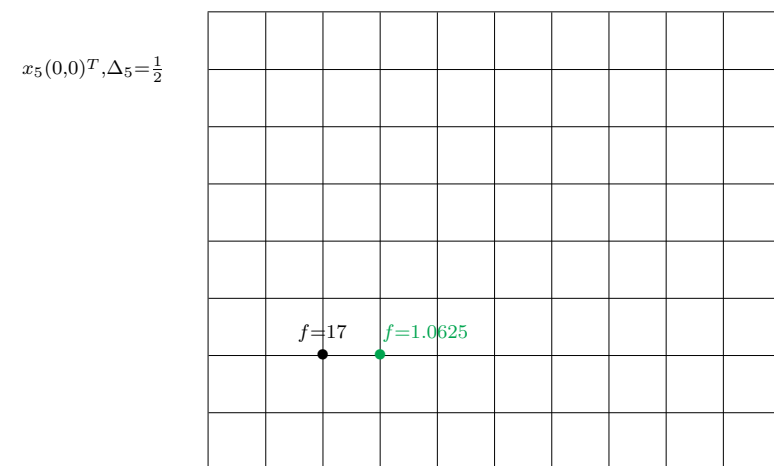
Recherche par coordonnées : algorithme opportuniste



Recherche par coordonnées : algorithme opportuniste

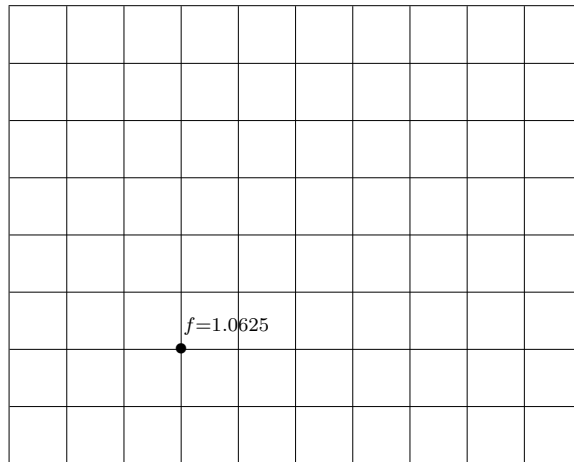


Recherche par coordonnées : algorithme opportuniste



Recherche par coordonnées : algorithme opportuniste

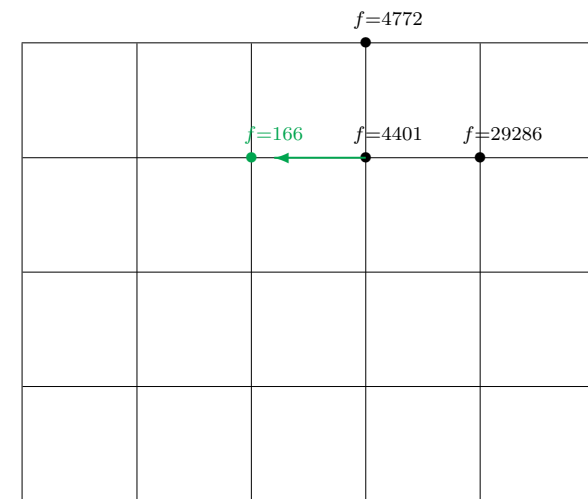
$$x_6 = (0, 0.5)^T, \Delta_6 = \frac{1}{2}$$



Avec un budget de 20 évaluations on obtient $x = (0.5, 0)^T$ avec $f(x) = 1.0625$. Peut-on faire mieux ?

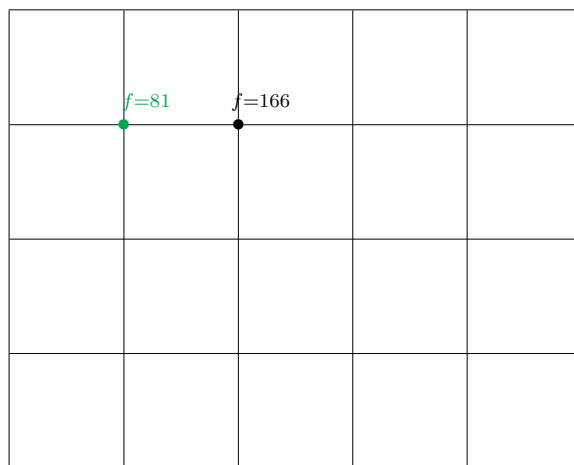
Recherche par coordonnées : algorithme dynamique

$$x_0 = (2, 2)^T, \Delta_0 = 1$$



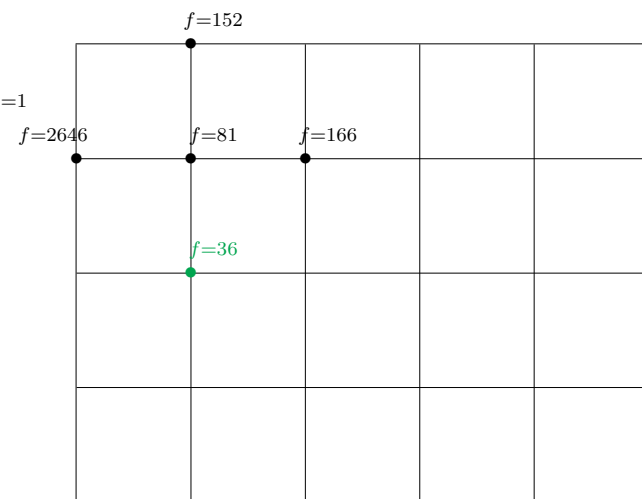
Recherche par coordonnées : algorithme dynamique

$$x_1 = (1, 2)^T, \Delta_1 = 1$$



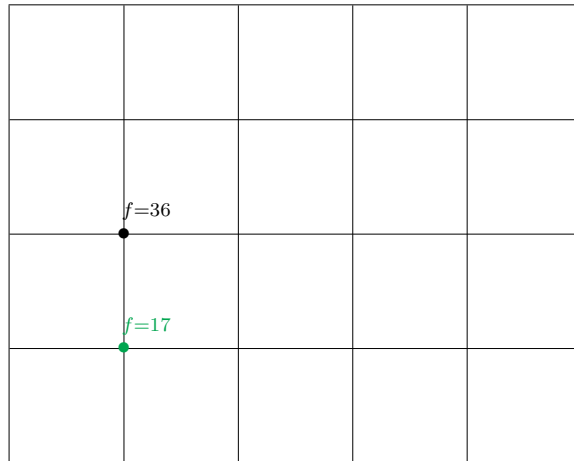
Recherche par coordonnées : algorithme dynamique

$$x_2 = (2, 2)^T, \Delta_2 = 1$$



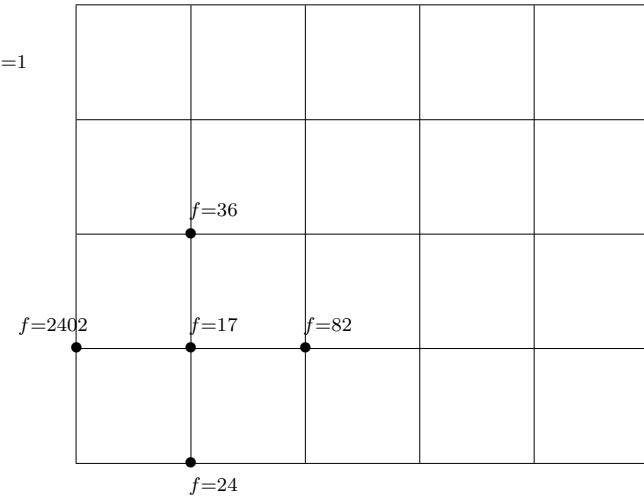
Recherche par coordonnées : algorithme dynamique

$$x_3 = (0, 1)^T, \Delta_3 = 1$$



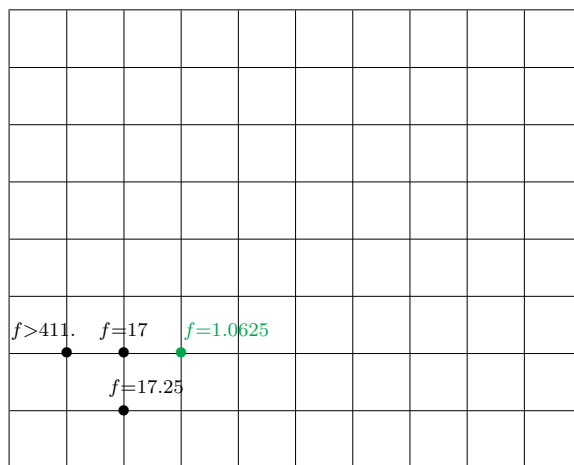
Recherche par coordonnées : algorithme dynamique

$$x_4 = (0, 0)^T, \Delta_4 = 1$$



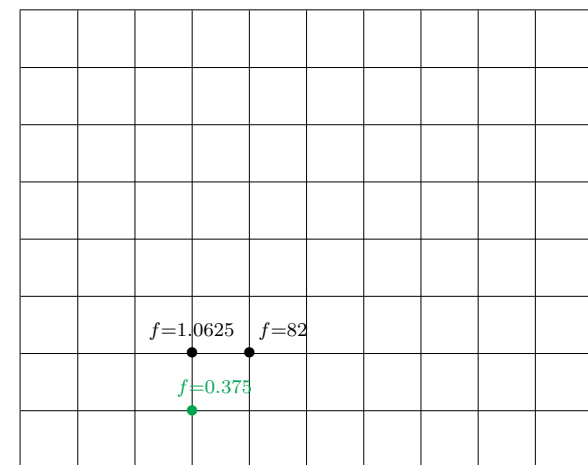
Recherche par coordonnées : algorithme dynamique

$$x_5 = (0, 0)^T, \Delta_5 = \frac{1}{2}$$



Recherche par coordonnées : algorithme dynamique

$$x_6 = (0, 0.5)^T, \Delta_6 = \frac{1}{2}$$



Recherche par coordonnées : 3 stratégies

Complète		Opportuniste		Dynamique	
x^T	$f(x)$	x^T	$f(x)$	x^T	$f(x)$
Après 20 évaluations de f					
(0, 0)	17	(.5, 0)	1.0625	(.5, -.5)	0.375
Après 50 évaluations de f					
(.375, -.375)	1.8e-2	(.375, -.312)	5.7e-3	(.375, -.344)	3.1e-3

Convergence de la recherche par coordonnées

Si la suite d'itérés $\{x_k\}$ appartient à un ensemble borné, alors

- $\lim_k \Delta_k = 0$
- il existe $\hat{x} \in \mathbb{R}^n$, qui est la limite d'une sous-suite d'optima locaux du treillis, sur des treillis devenant infiniment fins
- si f est continûment différentiable en $\hat{x}$, alors $\nabla f(\hat{x}) = 0$.

... mais on ne connaît rien de f .

On doit travailler avec des hypothèses de différentiabilité moins restrictives.

Optimisation sous contraintes de bornes: Approche barrière

Soient $\ell \in (\{-\infty\} \cup \mathbb{R})^n$ et $u \in (\mathbb{R} \cup \{\infty\})^n$ des vecteurs de bornes.

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

Lancer la recherche par coordonnées sur

$$\min_{x \in \mathbb{R}^n} f_X(x)$$

$$\text{où } f_X(x) = \begin{cases} f(x) & \text{si } x \in X \\ +\infty & \text{sinon.} \end{cases}$$

Optimisation sous contraintes: combinaison de contraintes.

Soit $\Omega = \{x \in X : c_j(x) \leq 0, j = 1, 2, \dots, m\}$ le domaine réalisable.

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

- Lancer la recherche par coordonnées sur

$$\min_{x \in \mathbb{R}^n} h_X(x)$$

$$\text{où } h_X(x) = \begin{cases} \sum_{j=1}^m (\max\{0, c_j(x)\})^2 & \text{si } x \in X \\ +\infty & \text{sinon.} \end{cases}$$

- Interrompre dès qu'une solution réalisable est obtenue.
- Lancer la recherche par coordonnées sur

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

MTH 6416 – Optimisation de boîtes noires – sommaire

1 Introduction

2 Recherche par coordonnées

3 Optimisation non-lisse

- Limitations de la recherche par coordonnées
- Dérivée et sous-gradient
- Calcul de Clarke: dérivée et gradient généralisés
- Analyse de convergence de la recherche par coordonnées

4 Recherche par motifs

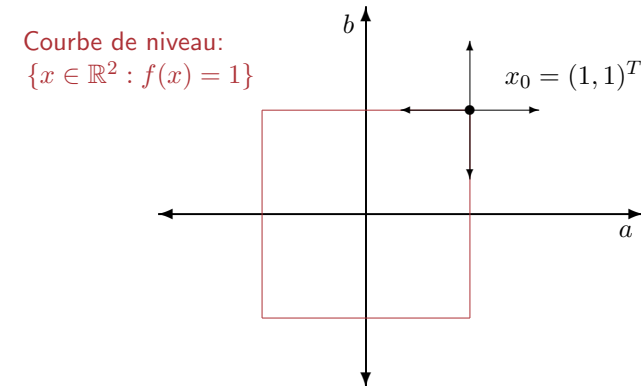
5 Contraintes générales non-lisses

6 Recherche sur treillis adaptatifs

7 Discussion

Limitations de la recherche par coordonnées

$$f(x) = \|x\|_\infty \text{ avec } x_0 = (1, 1)^T.$$



Un algorithme directionnel

Calul de Clarke

Si f est Lipschitz¹ près de $\bar{x} \in \mathbb{R}^n$, alors la dérivée généralisée de Clarke en $\bar{x}$ dans la direction $v \in \mathbb{R}^n$ est

$$f^\circ(\bar{x}; v) = \limsup_{y \rightarrow \bar{x}, t \downarrow 0} \frac{f(y + tv) - f(y)}{t}.$$

¹il existe un scalaire K tel que

$$|f(x) - f(x')| \leq K\|x - x'\|$$

pour tout x, x' dans un voisinage de $\bar{x}$.

Calul de Clarke

- Le *gradient généralisé* de f en x est l'ensemble

$$\partial f(x) := \{\zeta \in \mathbb{R}^n : f^\circ(x; v) \geq v^T \zeta \text{ pour tout } v \in \mathbb{R}^n\}.$$

- Si f est Lipschitz près de x , alors

$$\partial f(x) = \text{co}\{\lim \nabla f(x_i) : x_i \rightarrow x \text{ et } \nabla f(x_i) \text{ existe}\}.$$

- La dérivée généralisée peut être obtenue à partir du gradient généralisé : $f^\circ(x; v) = \max\{v^T \zeta : \zeta \in \partial f(x)\}$.
- Si x est un minimum de f , et f est Lipschitz près de x , alors $0 \in \partial f(x)$. Ceci généralise la condition nécessaire d'optimalité de premier ordre du cas continûment différentiable : $0 = \nabla f(x)$.
- Si f est différentiable (Hadamard, Gâteaux, ou Fréchet) en x , alors le gradient de f en x appartient au gradient généralisé $\partial f(x)$.

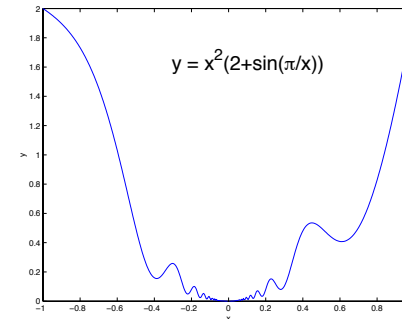
Calul de Clarke

- Si f est convexe, $\partial f(x)$ = sous-différentiel.
- f est *régulier* en x si pour tout $v \in \mathbb{R}^n$, la dérivée directionnelle $f'(x; v)$ existe et est égale à $f^\circ(x; v)$.
- f est *strictement différentiable* en x si pour tout $v \in \mathbb{R}^n$,

$$\lim_{y \rightarrow x, t \downarrow 0} \frac{f(y + tv) - f(y)}{t} = \nabla f(x)^T v.$$

- Si f est Lipschitz près de x et $\partial f(x)$ se réduit à un singleton $\{\zeta\}$, alors f est strictement différentiable en x et $\nabla f(x) = \zeta$.

Differentiable, mais pas strictement différentiable



f est Lipschitz et différentiable près de 0:

$$y'(0) = 0 \text{ et } y' = 2x(2 + \sin(\frac{\pi}{x})) - \pi \cos(\frac{\pi}{x})$$

La dérivée n'est pas continue en $x = 0$:

$$y'(\frac{1}{2k}) = \frac{2}{k} - \pi$$

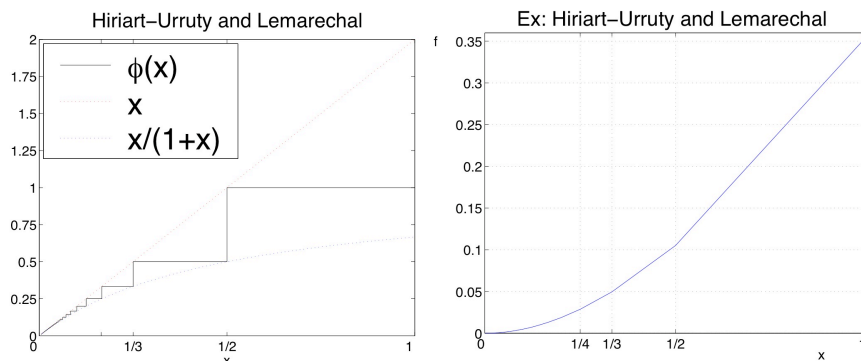
f n'est pas strictement différentiable :

$$\partial f(0) = [-\pi, \pi]$$

f n'est pas régulière :

$$f^\circ(0, \pm 1) = \pi \neq f'(0, \pm 1) = 0$$

Strictement, mais pas continûment différentiable



$$f(x) = \int_0^x \varphi(u) du \quad \text{où} \quad \varphi(u) = \begin{cases} u & \text{si } u \leq 0 \\ \frac{1}{1+\kappa} & \text{si } \kappa + 1 > \frac{1}{u} \geq k \end{cases}$$

f est Lipschitz près de $\hat{x} = 0$, et $\partial f(\frac{1}{\kappa}) = [\frac{1}{\kappa+1}, \frac{1}{\kappa}]$

f n'est pas strictement différentiable, ni continûment différentiable, dans tout voisinage de $\hat{x} = 0$.

$\partial f(0) = \{0\}$, et donc f est strictement différentiable en $\hat{x} = 0$.

Convergence de la recherche par coordonnées

Si la suite d'itérés $\{x_k\}$ appartient à un ensemble borné, alors

- $\lim_k \Delta_k = 0$
- il existe $\hat{x} \in \mathbb{R}^n$, qui est la limite d'une sous-suite d'optima locaux du treillis, sur des treillis devenant infiniment fins :
 $f(x_k \pm \Delta_k e_i) \geq f(x_k)$ pour tout e_i .

Théorème

Si f est Lipschitz près de $\hat{x}$, alors $f^\circ(\hat{x}; \pm e_i) \geq 0$ pour tout e_i .

PREUVE:
$$f^\circ(\hat{x}; e_i) := \limsup_{y \rightarrow \hat{x}, t \downarrow 0} \frac{f(y + te_i) - f(y)}{t} \geq \limsup_{k \in K} \frac{f(x_k + \Delta_k e_i) - f(x_k)}{\Delta_k} \geq 0.$$

Théorème

Si f est Lipschitz près de $\hat{x}$, alors $f^\circ(\hat{x}; \pm e_i) \geq 0$ pour tout e_i .

Si f est régulier près de $\hat{x}$, alors $f'(\hat{x}; \pm e_i) \geq 0$ pour tout e_i .

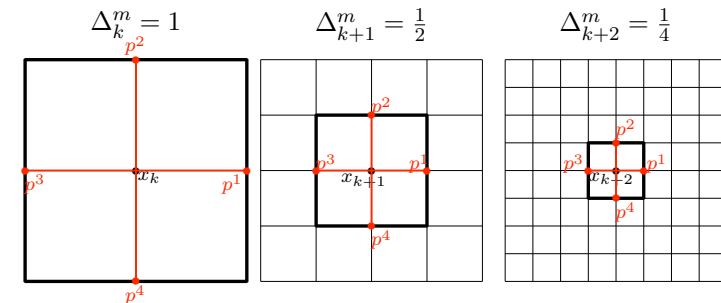
Si f est strictement différentiable près de $\hat{x}$, alors $\nabla f(\hat{x}) > 0$.

MTH 6416 – Optimisation de boîtes noires – sommaire

- 1 Introduction
- 2 Recherche par coordonnées
- 3 Optimisation non-lisse
- 4 Recherche par motifs
 - La sonde locale et recherche globale
 - Algorithme de recherche par motifs
 - Analyse de convergence de la recherche par motifs
 - Fonctions substituts
- 5 Contraintes générales non-lisses
- 6 Recherche sur treillis adaptatifs
- 7 Discussion

Recherche par coordonnées – Directions de sonde

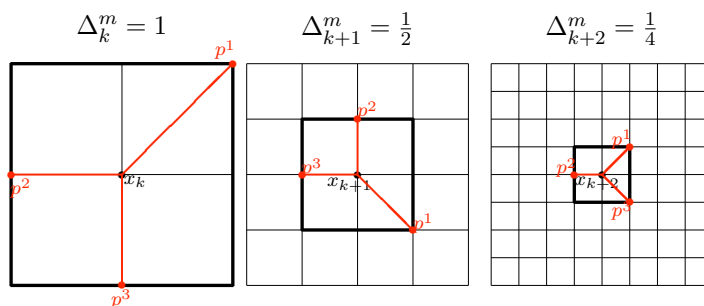
$P_k = \{x_k \pm \Delta_k^m e_i\}$; $2n$ points du treillis adjacents à x_k .



Il y a toujours les mêmes $2n = 4$ directions.

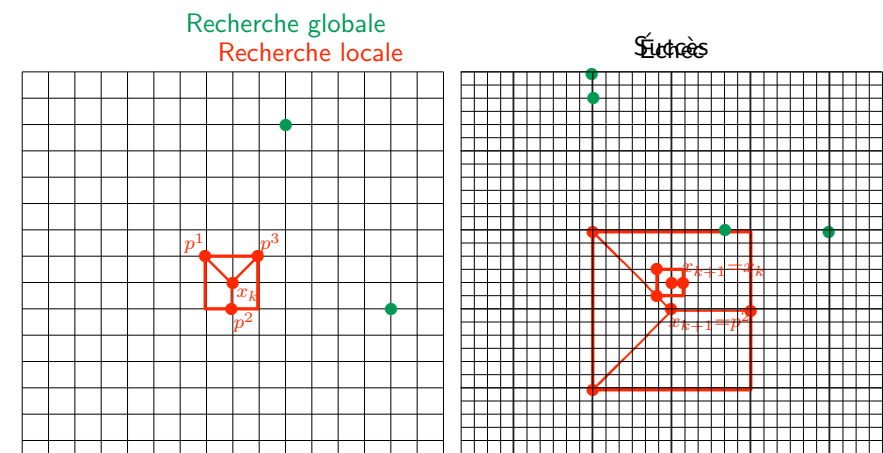
Recherche par motifs – Directions de sonde

$P_k = \{x_k + \Delta_k^m d : d \in D_k\}$; $n + 1$ points du treillis adjacents à x_k .



Il y a 14 différentes façons de définir D_k sur ce treillis.

Recherche par motifs



Algorithme de recherche par motifs

L'itération k est initiée à partir de $x_k \in \Omega$

Le but visé par l'itération est de trouver une solution meilleure que x_k .

Recherche

globale: Recherche finie sur le treillis M_k .

Sonde locale: Si la **recherche globale** échoue, alors f est évalué sur P_k .

Mise à jour: Si la **recherche globale** ou la **sonde locale** réussit, alors augmenter Δ_k .
Sinon, réduire Δ_k et poser $\Delta_{k+1} = \Delta_k$.

Les points non-réalisables sont simplement rejetés.

Analyse de convergence de la recherche par motifs

Si la suite d'itérés $\{x_k\}$ appartient à un ensemble borné, alors

- $\lim_k \Delta_k = 0$
- il existe $\hat{x} \in \mathbb{R}^n$, qui est la limite d'une sous-suite d'optima locaux du treillis, sur des treillis devenant infiniment fins :
 $f(x_k + \Delta_k d) \geq f(x_k)$ pour tout $d \in D_k$.

Théorème

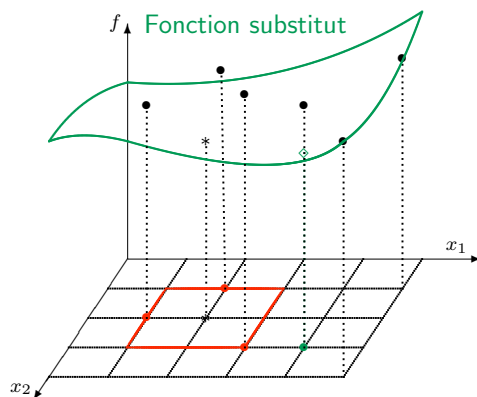
Si f est Lipschitz près de $\hat{x}$, alors $f^\circ(\hat{x}; d) \geq 0$ pour toutes les directions $d \in D$ utilisées infiniment souvent.

PREUVE:
$$f^\circ(\hat{x}; d) := \limsup_{y \rightarrow \hat{x}, t \downarrow 0} \frac{f(y + td) - f(y)}{t}$$
$$\geq \limsup_{k \in K} \frac{f(x_k + \Delta_k d) - f(x_k)}{\Delta_k} \geq 0.$$

Note : L'ensemble de ces directions constitue un ensemble générateur positif.

Théorème

Utilisation d'une fonction substitut dans la recherche globale



* est la meilleure solution à date
Sonde locale est évalué au point candidat
* est toujours la meilleure solution à date

MTH 6416 – Optimisation de boîtes noires – sommaire

- 1 Introduction
- 2 Recherche par coordonnées
- 3 Optimisation non-lisse
- 4 Recherche par motifs
- 5 Contraintes générales non-lisses
 - Limitations de la recherche par motifs
 - Quatre types de cônes
 - Conditions nécessaires d'optimalité
- 6 Recherche sur treillis adaptatifs
- 7 Discussion

Limitations de la recherche par motifs

- Pour que la recherche par motifs quitte un point non-optimal par la sonde locale, elle doit générer des directions de descente dans le cône tangent par rapport à Ω .
- Il est facile de générer celles-ci lorsque le domaine est polyédrique: $\{T_\Omega(x) : x \in \Omega\}$ contient un nombre fini de cônes.
- Dans le cas de contraintes générales, $\{T_\Omega(x) : x \in \Omega\}$ contient habituellement un nombre infini de cônes : Il est alors impossible d'identifier tous les générateurs des cônes tangents.

Cône hypertangent

Définition

Un vecteur $v \in \mathbb{R}^n$ est *hypertangent* à l'ensemble $\Omega \subseteq \mathbb{R}^n$ au point $x \in \Omega$ s'il existe un scalaire $\epsilon > 0$ tel que

$$y + tw \in \Omega \quad \text{pour tout } y \in \Omega \cap B_\epsilon(x), \quad w \in B_\epsilon(v) \quad \text{et } 0 < t < \epsilon.$$

L'ensemble des vecteurs *hypertangents* à Ω en x est le cône hypertangent à Ω en x et s'écrit $T_\Omega^H(x)$.

Cône de Clarke

Définition

Un vecteur $v \in \mathbb{R}^n$ est *Clarke tangent* à l'ensemble $\Omega \subseteq \mathbb{R}^n$ au point x dans la fermeture de Ω si pour toute suite $\{y_k\}$ d'éléments de Ω convergeant vers x et pour toute suite de scalaires positifs $\{t_k\}$ convergeant vers zéro, il existe une suite de vecteurs $\{w_k\}$ convergeant vers v tels que $y_k + t_k w_k \in \Omega$.

L'ensemble des vecteurs *Clarke tangents* à Ω en x est le cône tangent de Clarke à Ω en x et s'écrit $T_\Omega^{Cl}(x)$.

Cône réalisable

Définition

Un vecteur $v \in \mathbb{R}^n$ est une *direction réalisable* pour l'ensemble $\Omega \subseteq \mathbb{R}^n$ au point x de Ω s'il existe un $\epsilon > 0$ tel que $x + tv \in \Omega$ pour tout $0 \leq t < \epsilon$.

L'ensemble des directions réalisables pour l'ensemble $\Omega \subseteq \mathbb{R}^n$ au point x de Ω est le cône réalisable pour Ω en x et s'écrit $T_\Omega^F(x)$.

Cône de Bouligand

Définition

Un vecteur $v \in \mathbb{R}^n$ est Bouligand tangent à l'ensemble $\Omega \subseteq \mathbb{R}^n$ au point x dans la fermeture de Ω s'il existe une suite $\{y_k\}$ d'éléments de Ω convergeant vers x et une suite de scalaires positifs $\{\lambda_k\}$ pour lesquels $v = \lim_k \lambda_k(y_k - x)$.

L'ensemble des vecteurs Bouligand tangents à Ω en x est le cône tangent de Bouligand à Ω en x et s'écrit $T_{\Omega}^{Co}(x)$.

Liens entre les cônes

- L'ensemble Ω est régulier en x si $T_{\Omega}^{Cl}(x) = T_{\Omega}^{Co}(x)$.
- Un ensemble convexe est régulier en tous ses points.
- $T_{\Omega}^{Co}(x)$ et $T_{\Omega}^{Cl}(x)$ sont des cônes fermés.
 $T_{\Omega}^{Cl}(x)$ et $T_{\Omega}^H(x)$ sont des cônes convexes.
- $T_{\Omega}^H(x) \subseteq T_{\Omega}^{Cl}(x) \subseteq T_{\Omega}^{Co}(x)$ $T_{\Omega}^H(x) \subseteq T_{\Omega}^F(x)$.
- $T_{\Omega}^H(x) = \text{int}(T_{\Omega}^{Cl}(x))$ lorsque $T_{\Omega}^H(x)$ est non-vide.
 $T_{\Omega}^{Cl}(x) = \text{cl}(T_{\Omega}^H(x))$ lorsque $T_{\Omega}^H(x)$ est non-vide.

Conditions nécessaires d'optimalité

Définition

Soit f une fonction Lipschitz près de $\hat{x} \in \Omega$. Alors, $\hat{x}$ est un point stationnaire au sens de Clarke ou Bouligand de f sur Ω , si $f^{\circ}(\hat{x}; v) \geq 0$ pour toutes directions v du cône de Clarke ou Bouligand de Ω en $\hat{x}$, respectivement.

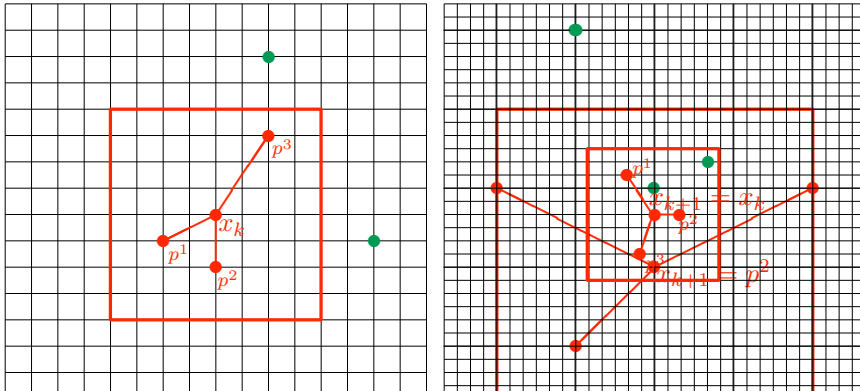
De plus, $\hat{x}$ est un point Clarke, ou Bouligand KKT-stationnaire de f sur Ω , si $-\nabla f(\hat{x})$ existe et appartient au polaire du cône de Clarke ou Bouligand de Ω en $\hat{x}$, respectivement.

MTH 6416 – Optimisation de boîtes noires – sommaire

- 1 Introduction
- 2 Recherche par coordonnées
- 3 Optimisation non-lisse
- 4 Recherche par motifs
- 5 Contraintes générales non-lisses
- 6 Recherche sur treillis adaptatifs
 - Algorithme de recherche sur treillis adaptatifs - barrière
 - Analyse de convergence - barrière
 - Algorithme de recherche sur treillis adaptatifs - filtre
 - Analyse de convergence - filtre
- 7 Discussion

Traitement des contraintes par la barrière

Recherche globale
Recherche locale



Analyse de convergence - barrière

Théorème

Lorsque $k \rightarrow \infty$, l'ensemble des directions de sonde normalisées produit par MADS forme un ensemble dense dans la boule unité avec probabilité 1.

Théorème

Soit $\hat{x}$ la limite d'une sous-suite d'optima locaux du treillis, sur des treillis devenant infiniment fins. Si f est Lipschitz près de $\hat{x}$, alors $f^\circ(\hat{x}, v) \geq 0$ pour tout $v \in T_\Omega^H(\hat{x})$ avec probabilité 1.

Analyse de convergence - barrière

Encore une fois, nous avons une hiérarchie de résultats de convergence "proportionnelle" à la différentiabilité de f et d' Ω .

- f est Lipschitz implique que les dérivées de Clarke sont non-négatives dans le cône hypertangent, et donc dans le cône de Clarke si $T_\Omega^H(\hat{x}) \neq \emptyset$, i.e., Clarke-stationnaire.
- f et Ω réguliers implique que les dérivées directionnelles sont non-négatives dans le cône de Bouligand, i.e., Bouligand stationnaire.
- f est strictement différentiable et Ω régulier implique Bouligand KKT-stationnaire.

Et sous encore plus d'hypothèses:

- Si f est deux fois strictement différentiable et Ω est convexe et $\nabla^2 f(\hat{x})$ est non-singulier, alors $\hat{x}$ est un minimum local strict.

Traitement des contraintes par le filtre (basé sur Fletcher - Leyffer)

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & x \in \Omega \equiv \{x \in X : C(x) \leq 0\}, \\ \text{où} \quad & C = (c_1, c_2, \dots, c_m)^T \text{ et } X \subseteq \mathbb{R}^n \end{aligned}$$

Les contraintes sont partitionnés en deux groupes.

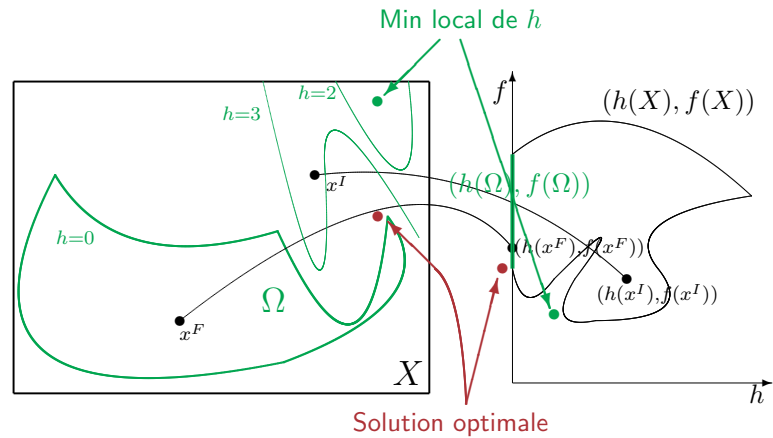
La barrière traite X . Le filtre traite $C(x)$.

On introduit la contrainte d'aggrégation de violations

$$h(x) := \begin{cases} \sum_j \max(0, c_j(x))^2 & \text{si } x \in X, \\ +\infty & \text{sinon.} \end{cases} \quad \begin{aligned} &\Leftarrow \text{contraintes ouvertes} \\ &\Leftarrow \text{contraintes fermées} \end{aligned}$$

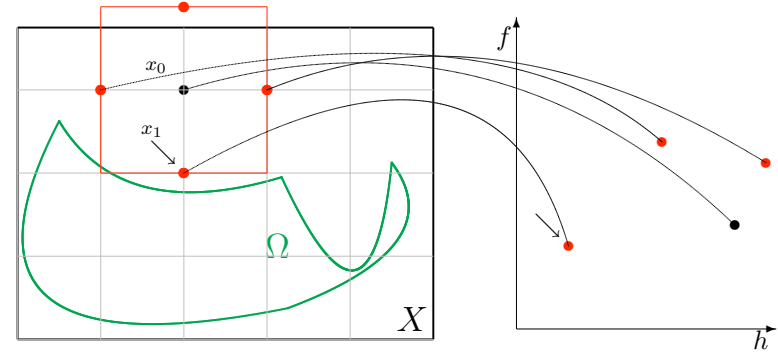
$h(x) = 0$ si et seulement si $x \in \Omega$.

Traitement des contraintes par le filtre



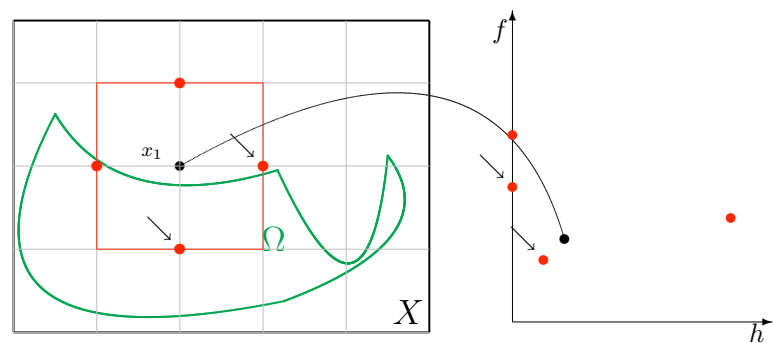
MADS-filtre : Itération 0

Meilleure solution non-réalisable est celle à gauche de la précédente (dans le graphe (h, f)) avec la meilleure valeur de f



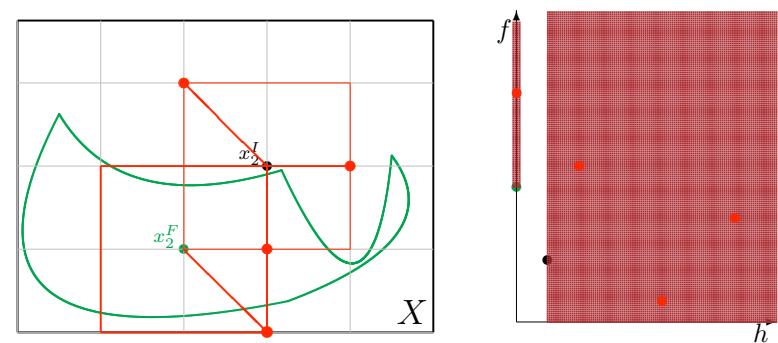
MADS-filtre : Itération 1

Nouvelle meilleure solution réalisable



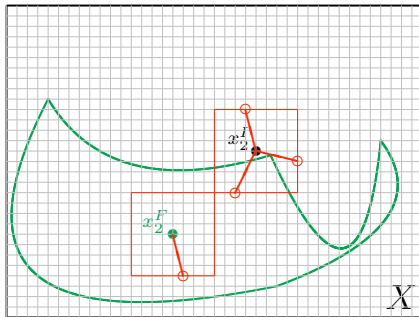
MADS-filtre : Itération 2

Échec : toutes les directions de (du) filtre sont des directions de gradient

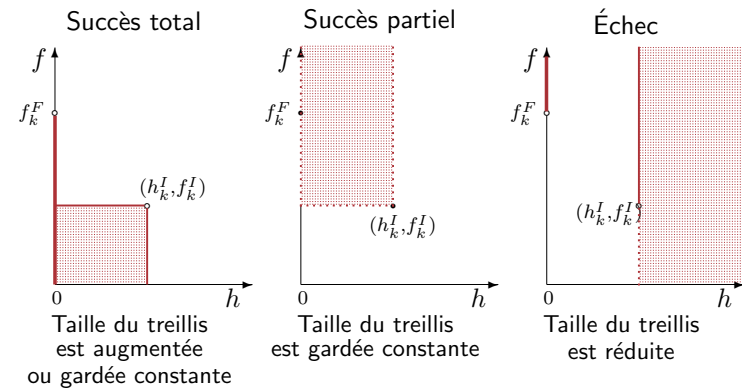


MADS-filtre : Itération 3

Les directions de sondes sont générées avec MADS.

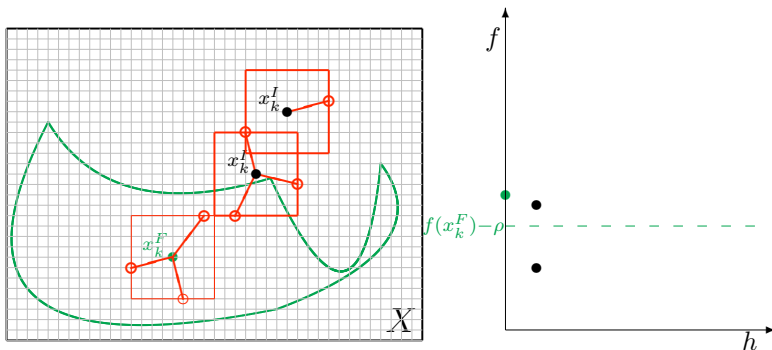


Classification d'une itération



Centres de sonde primaire et secondaire

x_k^F est le centre primaire
lorsque $f(x_k^I) \geq f(x_k^F) - \rho$



Analyse de convergence - filtre

Si $\rho > 0$ alors l'algorithme devient essentiellement identique à MADS-barrière si on trouve une solution réalisable.

Théorème

Si $\hat{x} \in \Omega$ est la limite d'optima locaux du treillis réalisables $\{x_k^F\}$ alors l'analyse de MADS pour le problème $\min_{x \in \Omega} f(x)$ s'applique.

Théorème

Si $\hat{x} \in \Omega$ est la limite d'optima locaux du treillis non-réalisables $\{x_k^I\}$ alors l'analyse de MADS pour le problème $\min_{x \in X} h(x)$ s'applique.

Théorème

Si $\hat{x} \in \Omega$ est la limite d'optima locaux du treillis non-réaliabiles $\{x_k^I\}$ et si

pour tout $v \in T_{\Omega}^H(\hat{x}) \neq \emptyset$, il existe un $\epsilon > 0$ pour lequel $h^{\circ}(x; v) < 0$ pour tout $x \in X \cap B_{\epsilon}(\hat{x})$ tel que $h(x) > 0$.

alors l'analyse de MADS pour le problème $\min_{x \in \Omega} f(x)$ s'applique.

- 1 Introduction
- 2 Recherche par coordonnées
- 3 Optimisation non-lisse
- 4 Recherche par motifs
- 5 Contraintes générales non-lisses
- 6 Recherche sur treillis adaptatifs
- 7 Discussion

Discussion

- Les méthodes de recherche directes travaillent uniquement avec les valeurs des fonctions.
- Elles sont utiles lorsque les dérivées ne sont pas utilisables ou peu fiables.
- MADS avec la barrière ou le filtre traite le cas à contrainte générales.
 - Analyse de convergence hiérarchique basée sur des propriétés locales.
 - NOMAD www.gerad.ca/Nomad c++ Gilles Couture Licence GNU.
 - NOMADm www.afit.edu/en/enc/Faculty/MAbramson/nomadm.html Matlab Mark Abramson's Licence GNU.
 - MADS dans GADS MATHWORKS MATLAB toolbox.