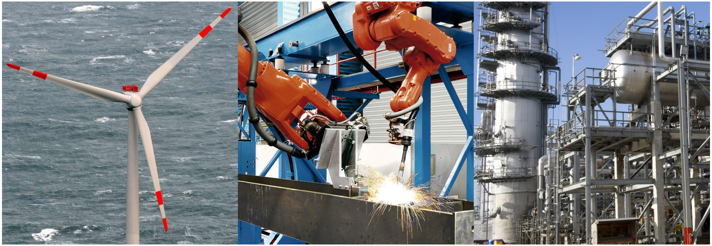
Introduction en optimisation convexe

Roland Hildebrand

LJK, Université Grenoble Alpes / CNRS

80 ans CNRS
Bourget-du-Lac, 23 mai 2019

Applications de l'optimisation



on rencontre des problèmes d'optimisation dans différents domaines

- optimisation de forme,
- contrôle optimal, plannification de mouvements,
- design d'expériences, ...

but: prendre une decision pour minimiser (maximiser) un coût (une performance) sous conditions

Problèmes d'optimisation: formalisation

choisir des variables de decision pour minimiser une fonction coût sous contraintes

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

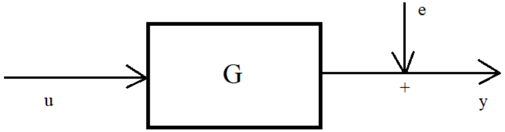
- variables de decision $x \in \mathbb{R}^n$
- fonction coût $f : \mathbb{R}^n \to \mathbb{R}$
- *ensemble faisable* $X \subset \mathbb{R}^n$

souvent la formalisation est elle-même non-triviale

nous devons assurer la solvabilité du problème formalisé

Exemple

design d'expériences



but: choisir le signal d'entrée u pour mieux estimer les paramètres de G à partir de la sortie y mesurée en présence du bruit e

au lieu de u_t on utilise comme variables de décision les *moments* du spectre de u

$$m_k = \int_0^{2\pi} \Phi_u(e^{i\varphi})e^{ik\varphi}\, d\varphi$$

Représentation

l'ensemble faisable peut être représenté par

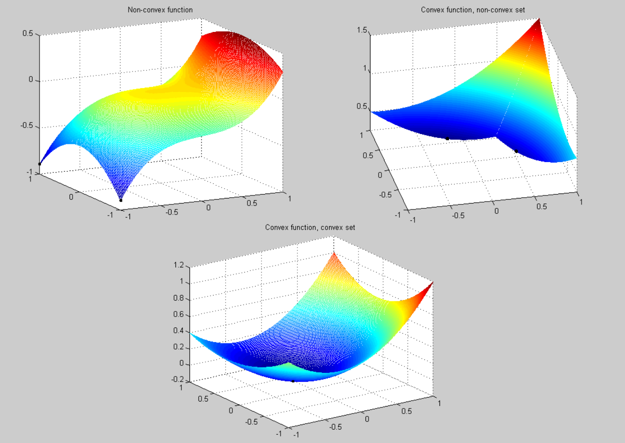
- des égalités scalaires $g_i(x) = 0$,
- des inégalités scalaires $h_j(x) \leq 0$,
- des inégalités matricielles $A_k(x) \succeq 0$,
- une boîte noire (oracle), ...

la fonction coût peut être représentée

- de manière analytique,
- par une boîte noire,
- avec (sous-)gradient,
- avec hessienne, ...

- les données peuvent être incertaines / aléatoires

Convexe vs. non-convexe

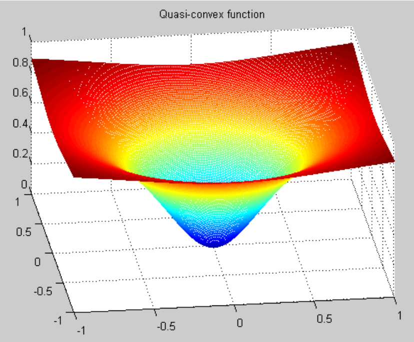


si f et X sont *convexes*, le problème d'optimisation est *convexe*

Convexe vs. non-convexe

problèmes convexes	problèmes non-convexes
soluble	non soluble
algorithmes efficaces polynomiaux	algorithmes exponentiels, heuristiques
minimum global	minima locaux
convergence globale vers le minimum global	convergence locale vers un point stationnaire
théorie développée	souvent pas de garanties

(Non-)convexité cachée



peut être minimisée par des algorithmes polynomiaux

le cône copositif

$$\mathcal{C}_n = \{A \in \mathcal{S}_n \,|\, x^T A x \geq 0 \,\forall\, x \geq 0\}$$

est un ensemble convexe

[Murty, Kabadi, 1987]: de décider l'appartenance au cône copositif est un problème co-NP-complet

Complexité d'un algorithme

l'algorithme produit une séquence $\{x_i\}$ qui converge vers la solution

l'investissement de calcul est une fonction de la précision ϵ

- complexité analytique: le nombre d'iterations k
- complexité arithmétique: le nombre d'opérations arithmétiques, k fois le coût de l'itération

dépend de la classe de problèmes

- classe de différentiabilité
- conditionnement
- constantes de Lipschitz, ...

Complexité

vitesse de convergence

- linéaire: $\log \epsilon \sim -k$
- sous-linéaire: p.ex. $\epsilon \sim k^{-\alpha}$
- quadratique: $\log \epsilon \sim -2^k$

précision

- valeur: $\epsilon = f(x_k) - f(x^*)$
- distance: $\epsilon = ||x_k - x^*||$
- gradient: $\epsilon = ||\nabla f(x_k)||$

Example: descente de gradient

on minimise une fonction $f(x)$ sans contraintes

itération:

$$x_{k+1} = x_k - \gamma_k f'(x_k)$$

$$\alpha\gamma_k \leq \frac{f(x_k) - f(x_{k+1})}{||f'(x_k)||^2} \leq \beta\gamma_k$$

$0 < \alpha < \beta < 1$ (règle d'Armijo)

convergence locale sur de classes différentes:

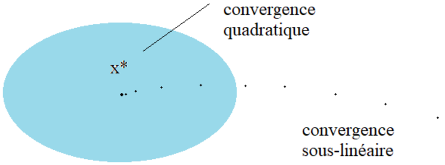
- gradient Lipschitz avec constante L: $||f'(x_k)||^2 \sim L/k$
- C^2 strictement convexe: $\log||x_k - x^*|| \sim -k$

Méthode de Newton

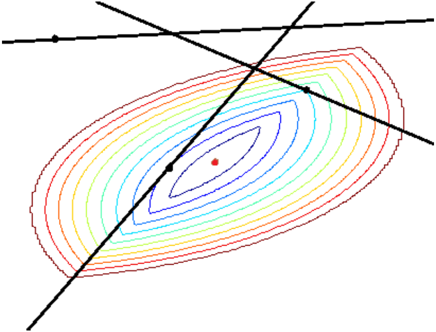
méthode de 2ème ordre

$$x_{k+1} = x_k - \gamma_k (f''(x_k))^{-1} f'(x_k), \qquad \gamma_k \leq 1$$

convergence sous-linéaire loin du minimum avec $\gamma_k \ll 1$
convergence rapide proche du minimum avec $\gamma_k = 1$



Hyperplan séparant



supposition: pour chaque point donné un hyperplan peut être calculé à partir d'information *locale* qui sépare le point du minimum *global*

exemple: $C = \{X \mid X \succeq 0\}$
$\hat{X} \not\succeq 0$, $u^T \hat{X} u < 0$

$\Rightarrow \langle uu^T, X \rangle > \langle uu^T, \hat{X} \rangle$ pour tout $X \succeq 0$

Méthode d'ellipsoïdes

on minimise une fonction convexe sur un convexe compact C

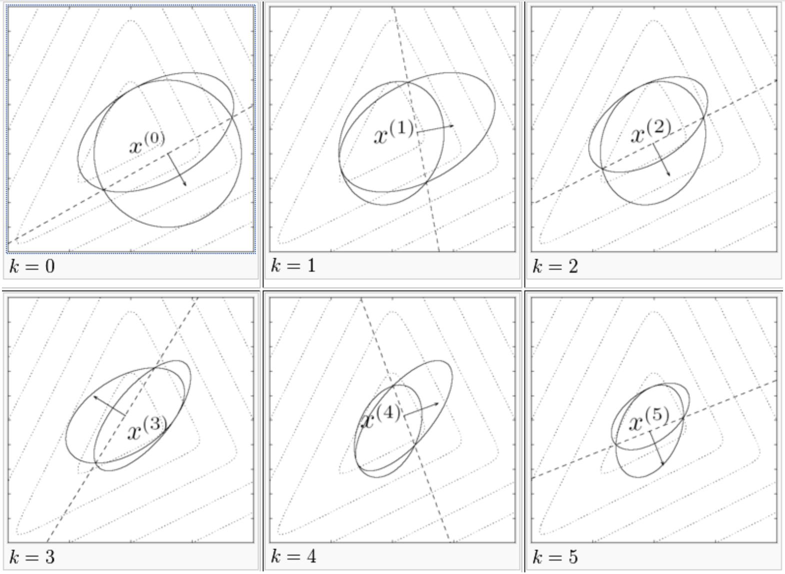
données initiales: ellipsoïde
$E_0 = \{x \in \mathbb{R}^n \,|\, (x - x_0)^T P_0^{-1}(x - x_0) \leq 1\}$ contenant l'ensemble C

iteration k:
on calcule un (co-)vecteur g_{k+1} tel que $\langle g_{k+1}, x_k - x^* \rangle \geq 0$
on calcule un nouvel ellipsoïde E_{k+1} par

$$x_{k+1} = x_k - \frac{1}{n+1} \frac{P_k g_{k+1}}{\sqrt{g_{k+1}^T P_k g_{k+1}}},$$

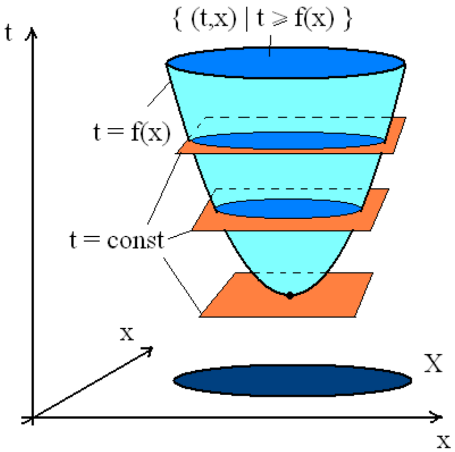
$$P_{k+1} = \frac{n^2}{n^2 - 1} \left(P_k - \frac{2}{n+1} \frac{P_k g_{k+1} (P_k g_{k+1})^T}{g_{k+1}^T P_k g_{k+1}} \right)$$

vitesse de convergence: $\log \; Vol \, E_k \sim -k$



https://en.wikipedia.org/wiki/Ellipsoid_method

Reduction à une fonction coût linéaire



la fonction coût peut être supposée linéaire sans restriction de généralité

si $f(x)$ non-linéaire, on ajoute une dimension et minimise t sur l'épigraphe

Méthodes de point intérieur

notons $X \subset \mathbb{R}^n$ l'ensemble faisable, $\langle c, x\rangle$ la fonction coût

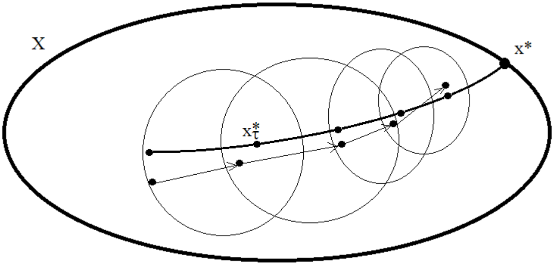
une barrière auto-concordante sur X avec paramètre ν est une fonction $F : X^o \to \mathbb{R}$ de classe C^3 satisfaisante

- $F'' \succ 0$ (convexité forte)
- $F|_{\partial C} = +\infty$ (propriété de barrière)
- $F'''(x)[h, h, h] \leq 2(F''(x)[h, h])^{3/2}$ for all $x \in X^o$, $h \in T_x X$
- $F'(x)[h] \leq \nu\sqrt{F''(x)[h, h]}$ for all $x \in X^o$, $h \in T_x X$

on résout le problème *sans contraintes* $\min_x(F(x) + \tau\langle c, x\rangle)$

pour $\tau \to +\infty$ la solution x_τ^* converge vers x^*

les solutions x_τ^* définissent le chemin central



on alterne une itération de Newton et une augmentation du poids τ, en restant dans la zone de convergence quadratique autour de x_τ^*

les itérées restent dans le voisinage du chemin central

rayon du voisinage et longueur du pas $\sim \nu^{-1/2}$

Programmes coniques

Definition

Un cône convexe régulier $K \subset \mathbb{R}^n$ est un cône fermé, avec intérieur non-vide et sans droites entières.

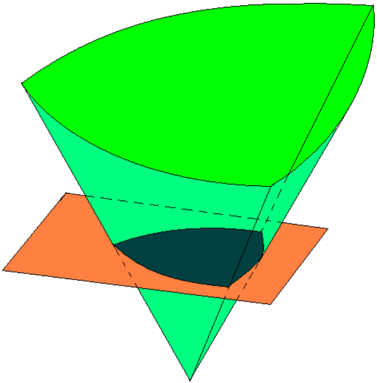
Definition

Un programme conique sur un cône convexe régulier $K \subset \mathbb{R}^n$ est un problème d'optimisation de la forme

$$\min_{x \in K} \langle c, x \rangle : \quad Ax = b.$$

tout problème d'optimisation convexe peut être reformulé comme programme conique

Interprétation géométrique



l'ensemble faisable d'un programme conique est l'intersection du cône K avec un sous-espace affine

Programmes sur de cônes symétriques

des programmes sur de *cônes symétriques* sont résolvables par des méthodes de point intérieur [Nesterov, Nemirovski, 1994]

- programmes linéaires (LP) sur $\mathbb{R}_+^n$ $\sim 10^6$ variables
- programmes coniques quadratiques (SOCP) sur le cône de Lorentz $L_n \sim 10^4$ variables
- programmes semi-définies (SDP) sur le cône de matrices positive semi-définies $\mathcal{S}_+^n \sim 10^2$ variables

la présence de structure peut augmenter le nombre de variables qu'on peut traiter

des outils libres (CLP, LiPS, SDPT3, SeDuMi, ...) et commerciaux (CPLEX, MOSEK, ...) sont disponibles

Merci de votre attention