Optimisation continue SD

E. Flayac

14 septembre 2026

Organisation du cours

- Optimisation continue (1h CM + 2.5h BE)
- Optimisation discrète (3.5h)
- Algorithmes évolutionnaires (2h)
- Evaluation (1h) : QCM sur les 3 parties

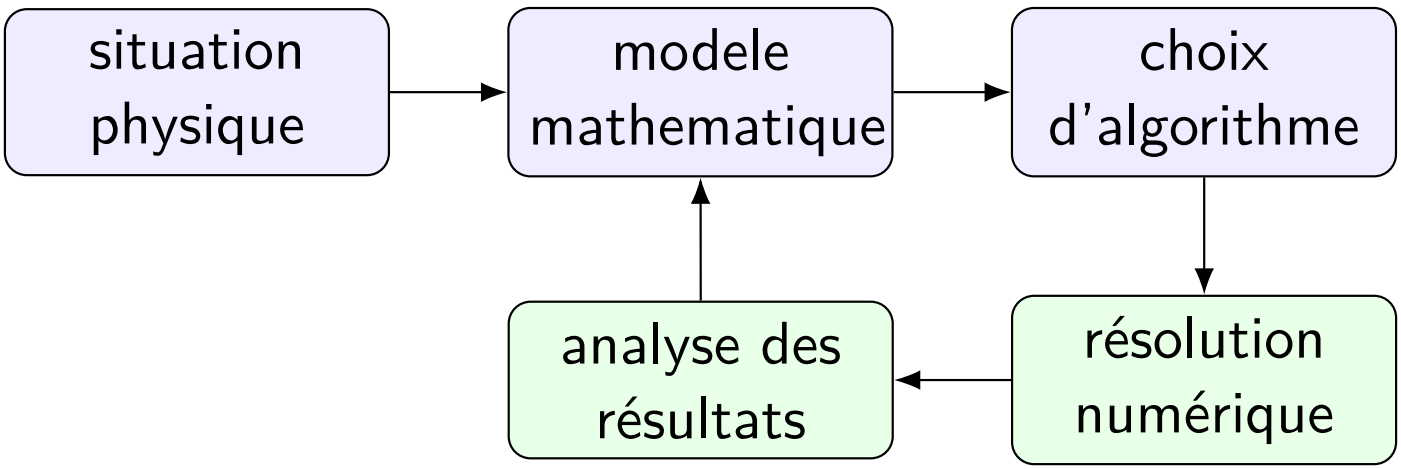
Plan du cours

Introduction

Pourquoi formaliser un problème ?

Objectif du cours

Identifier la structure du problème d'optimisation, choisir un cadre théorique, puis une méthode de résolution adaptée.



Définition générale d'un problème

On considère

- une variable de décision $x \in \mathbb{R}^n$,
- un coût $f : \mathbb{R}^n \to \mathbb{R}$,
- des contraintes d'égalité $h : \mathbb{R}^n \to \mathbb{R}^m$
- des contraintes d'inégalité $g : \mathbb{R}^n \to \mathbb{R}^p$.

qui définissent le problème :

$$\min_{x \in \mathbb{R}^n} f(x) \qquad \text{sous} \qquad h(x) = 0, \quad g(x) \leq 0. \qquad (P)$$

Définition : Ensemble admissible

L'ensemble admissible est $\mathcal{A} = \{x \in \mathbb{R}^n \ : \ h(x) = 0, \ g(x) \leq 0\}$.

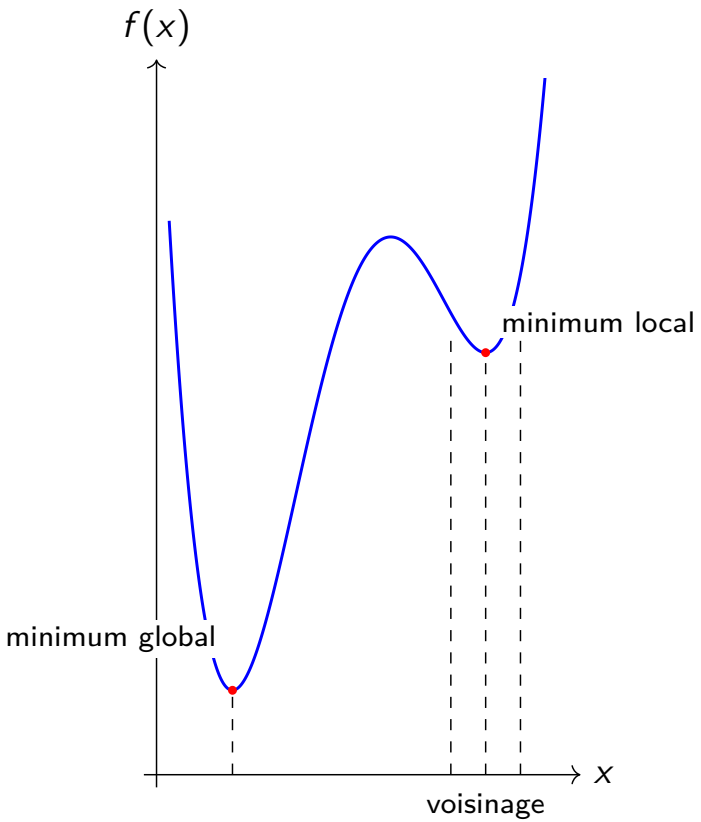
Vocabulaire : minimum global et minimum local

Définition : Solution globale

$x^\star \in \mathcal{A}$ est une solution globale si $f(x^\star) \leq f(x)$ pour tout $x \in \mathcal{A}$.

Définition : Solution locale

$x^\star \in \mathcal{A}$ est locale s'il existe $\rho > 0$ tel que $f(x^\star) \leq f(x)$ pour tout $x \in \mathcal{A} \cap B(x^\star, \rho)$.



Optimisation sans contraintes différentiable

Problème sans contrainte différentiable

Cadre

On considère $f : \mathbb{R}^n \to \mathbb{R}$ différentiable et le problème

$$\min_{x \in \mathbb{R}^n} f(x). \qquad (P_{sc})$$

Gradient

Le gradient $\nabla f(x) \in \mathbb{R}^n$ répresente une approximation linéaire de f vérifie

$$f(x+d) = f(x) + \nabla f(x)^T d + o(\|d\|).$$

Hessienne

Si f est deux fois différentiable, $\nabla^2 f(x) \in \mathbb{R}^{n \times n}$ décrit la courbure locale et vérifie

$$f(x+d) = f(x) + \nabla f(x)^T d + \frac{1}{2} d^T \nabla^2 f(x) d + o(\|d\|^2).$$

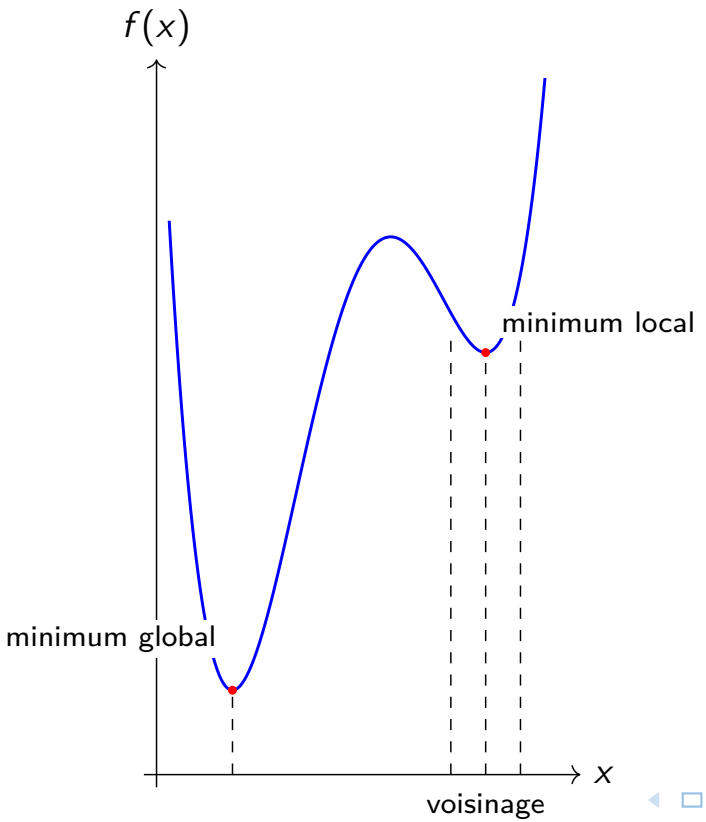
Solutions globales et locales sans contrainte

Définition : Globalité

Une solution globale minimise f sur tout $\mathbb{R}^n$. C'est une propriété forte, souvent difficile à certifier en non convexe.

Définition : Localité

Une solution locale minimise seulement dans une boule euclidienne $B(x^\star, \rho)$ avec $\rho > 0$. Les algorithmes locaux visent typiquement ce type de point.



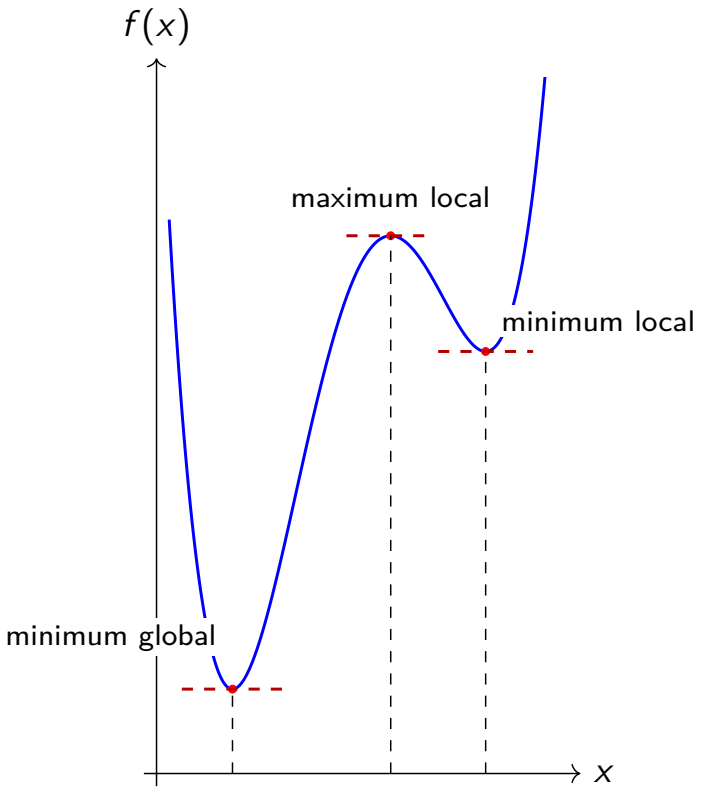
Condition nécessaire du premier ordre

Proposition : Fermat différentiable

Si f est différentiable et si $x^\star$ est un minimum local, alors $\nabla f(x^\star) = 0$.

Remarque : Attention

La réciproque est fausse en général : un point stationnaire peut être un maximum local ou un point selle.

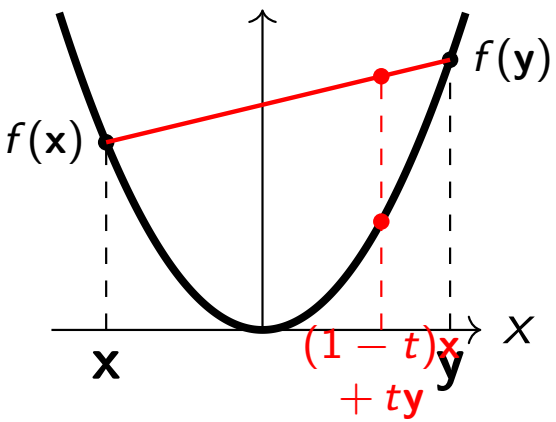


Fonctions convexes

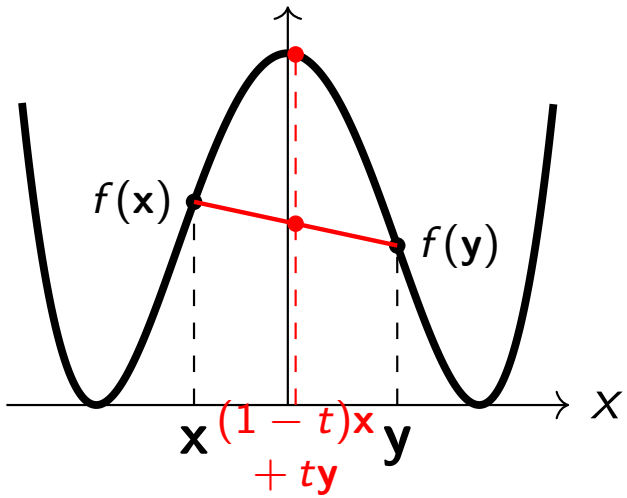
Définition

Une fonction $f : \mathbb{R}^n \to \mathbb{R}$est dite **convexe** si $\forall (x, y) \in (\mathbb{R}^n)^2, \quad \forall t \in [0, 1],$

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y).$$



Fonction convexe



Fonction non convexe

Interprétation géométrique

Une fonction est convexe si son graphe est situé **en dessous de toute corde** reliant deux points de son graphe.

Conditions d'optimalité convexe

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

Minima d'une fonction convexe

Soit $x^* \in \mathbb{R}^n$. Si f est convexe alors x^* est une solution locale de (P_{sc}) ssi x^* est une solution globale

Conditions d'optimalité convexe

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

Minima d'une fonction convexe

Soit $x^* \in \mathbb{R}^n$. Si f est convexe alors x^* est une solution locale de (P_{sc}) ssi x^* est une solution globale

Condition suffisante d'optimalité d'ordre 1 convexe (CS1 convexe)

Soit $x^* \in \mathbb{R}^n$. Si f est convexe et différentiable en x^* et si x^* est tel que :

$$\nabla f(x^*) = 0,$$

alors x^* est une solution locale (donc globale) de (P_{sc}).

Synthèse théorique sans contrainte

Cadre	Condition utile	Garantie
Différentiable quelconque	$\nabla f(x^\star) = 0$	nécessaire locale seulement
Convexe différentiable	$\nabla f(x^\star) = 0$	nécessaire et suffisante d'optimalité globale

Point clé

- les algorithmes d'optimisation sont itératifs
- On ne peut que satisfaire $\nabla f(x^\star) = 0$ en pratique

Directions de descente : principe commun

Définition : Direction de descente

Au point x_k, une direction d_k est de descente si $\nabla f(x_k)^T d_k < 0$. Alors

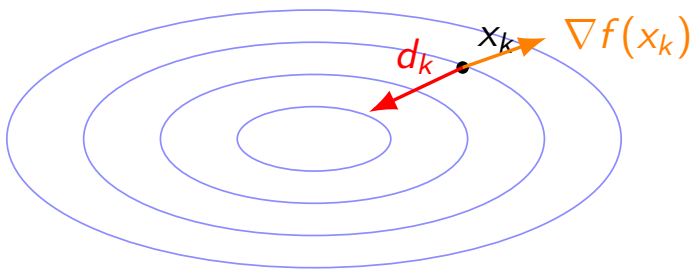
$$f(x_k + \alpha d_k) < f(x_k)$$

pour $\alpha > 0$ assez petit.

Schéma générique

Choisir d_k, choisir un pas $\alpha_k > 0$, puis poser

$$x_{k+1} = x_k + \alpha_k d_k.$$



$\nabla f(x_k)^T d_k < 0$

Recherche linéaire

Schéma générique

Choisir d_k, choisir un pas $\alpha_k > 0$, puis poser

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

La recherche du pas α_k s'appelle la recherche linéaire

Pas constant

on choisit $\alpha_k \equiv \alpha > 0$.

Pas optimal

on choisit α_k tel que $\varphi_k(\alpha) = f(x_k + \alpha d_k)$ et on cherche un pas acceptable.

Règle d'Armijo

On accepte α_k si $f(x_k + \alpha_k d_k) \leq f(x_k) + c\alpha_k \nabla f(x_k)^T d_k$, avec $c \in (0, 1)$.

Ordres d'information

Ordre 0

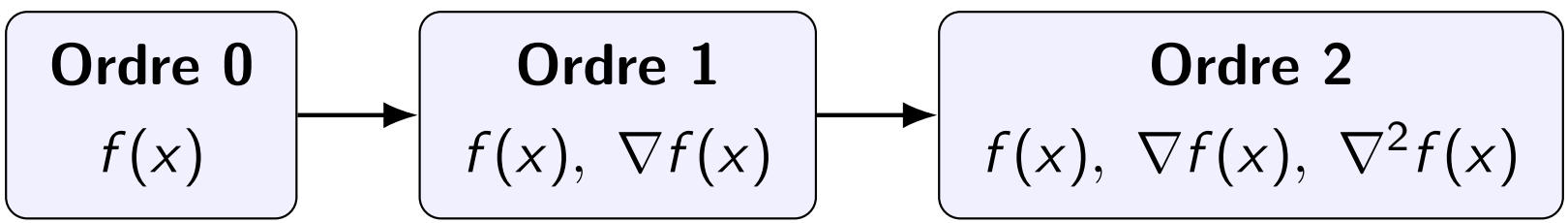
Seules les valeurs de f sont utilisées. Ces méthodes sont utiles si les dérivées sont indisponibles ou bruitées.

Ordres 1 et 2

L'ordre 1 utilise ∇f. L'ordre 2 utilise aussi $\nabla^2 f$ ou une approximation de cette matrice.

Compromis

Plus l'information locale est riche, plus l'itération peut être efficace, mais plus chaque itération peut être coûteuse.



Algorithme du gradient

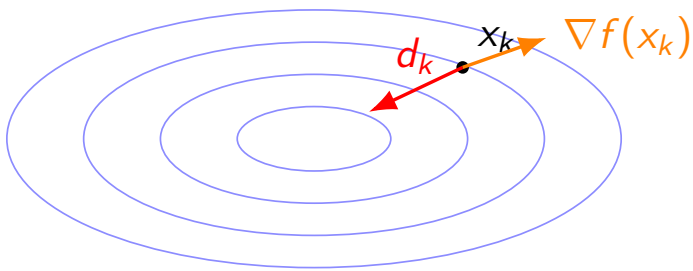
Définition : Direction

Le choix $d_k = -\nabla f(x_k)$ donne la direction de plus forte descente pour la norme euclidienne.

Itération

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k),$$

La méthode est robuste mais peut être lente.



$$\nabla f(x_k)^T d_k < 0$$

Newton : modèle quadratique local

Définition : Problème quadratique local

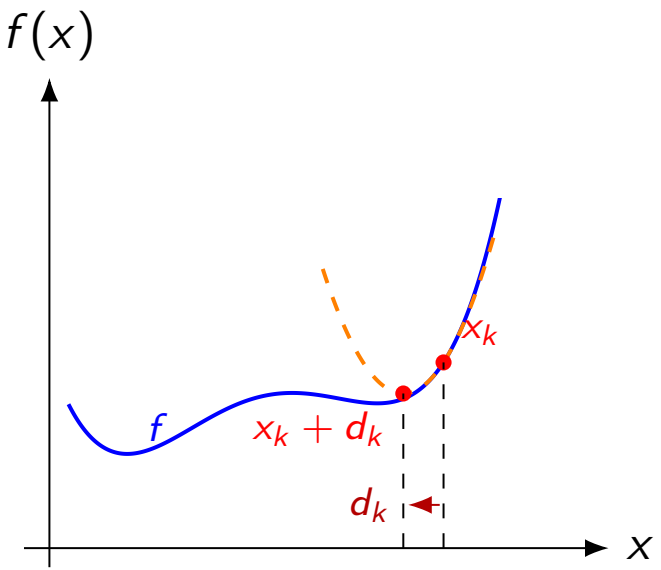
La direction de Newton est obtenue en résolvant le problème

$$\min_{d \in \mathbb{R}^n} f(x_k) + \nabla f(x_k)^T d + \frac{1}{2} d^T \nabla^2 f(x_k) d.$$

Système linéaire associé

Si $\nabla^2 f(x_k)$ est définie positive, ce problème admet une unique solution d_k, caractérisée par

$$\nabla^2 f(x_k) d_k = -\nabla f(x_k).$$



Newton : forces et fragilités

Force

Près d'un minimiseur régulier, la convergence peut être quadratique.

Fragilité

Loin de la solution, le modèle quadratique peut être mauvais ou la direction peut ne pas être de descente.

Remèdes

Recherche linéaire, régions de confiance ou modification de la hessienne globalisent la méthode.

Conditionnement : deux messages

Numérique

Résoudre un système linéaire mal conditionné amplifie les erreurs numériques.

Algorithmique

Le gradient est ralenti par les longues vallées ; Newton compense mieux les échelles si la hessienne est fiable.

Modélisation

Une mauvaise mise à l'échelle des variables peut transformer un problème raisonnable en problème difficile.

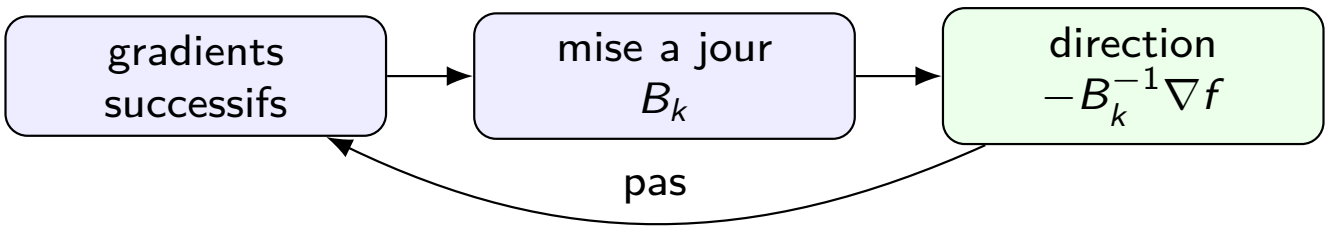
Quasi-Newton

Définition : Principe

Les méthodes quasi-Newton utilisent $d_k = -B_k^{-1}\nabla f(x_k)$, où B_k approxime la hessienne à partir des gradients successifs.

Exemples

BFGS et L-BFGS évitent de calculer explicitement $\nabla^2 f$; L-BFGS limite en plus le stockage mémoire.



Moindres carrés non linéaires

Définition : Structure

On minimise $f(x) = \frac{1}{2} \|r(x)\|_2^2$ avec $r : \mathbb{R}^n \to \mathbb{R}^m$ et jacobienne $J_r(x)$.

Gradient

$\nabla f(x) = J_r(x)^T r(x)$. La hessienne contient $J_r(x)^T J_r(x)$ plus des termes utilisant les dérivées secondes des résidus.

Gauss–Newton

Approximation

On remplace la hessienne par $J_r(x_k)^T J_r(x_k)$.

Direction

On résout $J_r(x_k)^T J_r(x_k) d_k = -J_r(x_k)^T r(x_k)$.

Intérêt

Les dérivées secondes des résidus sont évitées ; la méthode est adaptée aux problèmes de calibration et d'ajustement.

Méthodes sans dérivées : quand les utiliser ?

Pertinent

Fonction boîte noire, dérivées indisponibles, simulations non différentiables ou bruitées.

Risque

Un nombre élevé d'évaluations peut être nécessaire, surtout lorsque n augmente.

À distinguer

Les algorithmes génétiques et métaheuristiques relèvent d'un autre cours ; ils ne sont pas détaillés ici.

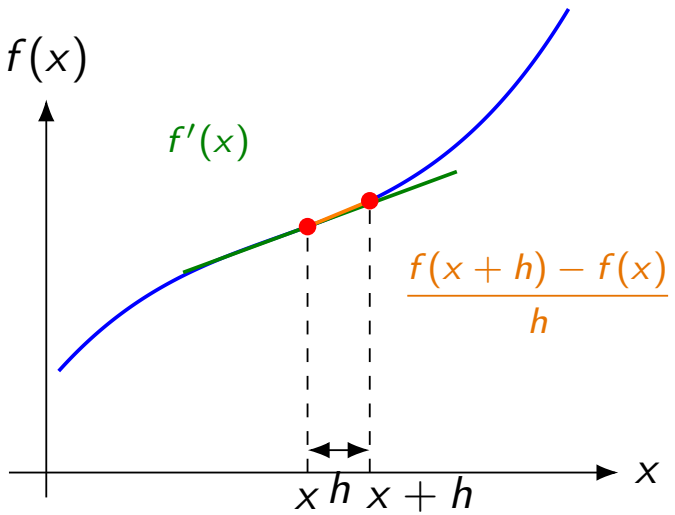
Différences finies : approximation de dérivées

Définition : Approximation

Pour $h > 0$ et le vecteur de base e_i, $\partial_i f(x) \simeq \big(f(x+he_i) - f(x)\big)/h$. Des formules centrées améliorent souvent la précision.

Remarque : Pas h

Un pas trop grand crée une erreur de troncature ; un pas trop petit amplifie les erreurs d'arrondi ou le bruit.



Synthèse des méthodes différentiables

Méthode	Info	Atout	Limite
Nelder–Mead	valeurs	simple boîte noire	peu adaptée aux grandes dimensions
Gradient	gradient	robuste, peu coûteux	sensible au conditionnement
Newton	hessienne	rapide localement	coûteux, fragile loin de la solution
BFGS/L-BFGS	gradients	compromis efficace	dépend de la recherche linéaire
Gauss–Newton	jacobienne	idéal moindres carrés	modèle moins fiable si résidus grands

Règle pratique

Commencer par identifier l'information disponible : valeurs seules, gradients, hessiennes ou jacobiennes.

Optimisation convexe non différentiable

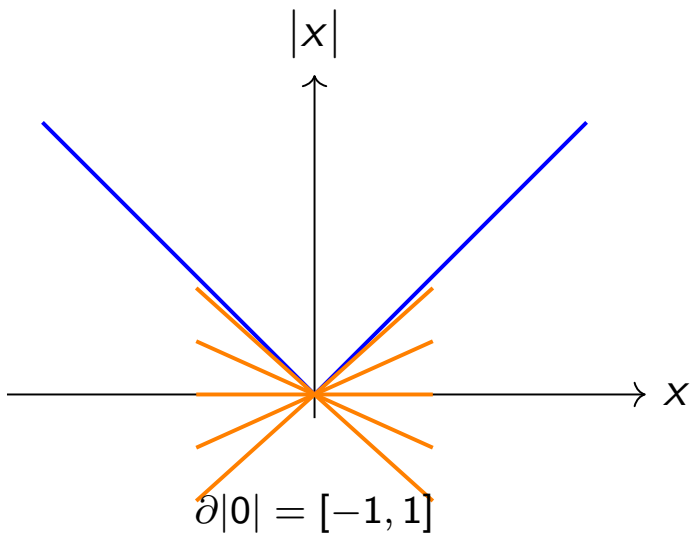
Pourquoi un cadre non différentiable ?

Exemples

Les fonctions $|x|$, $\|x\|_1$ et $\max_i f_i(x)$ sont convexes dans de nombreux cas, mais ne sont pas différentiables partout.

Remarque : Idée

Le gradient est remplacé par un ensemble de pentes de minorantes affines : le sous-différentiel.



Sous-gradient convexe

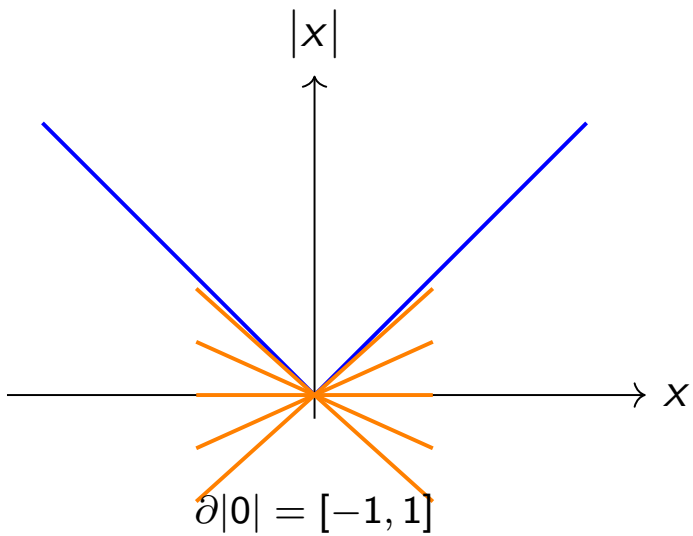
Définition : Sous-gradient

Pour une fonction convexe $f : \mathbb{R}^n \to \mathbb{R}$, un vecteur v est un sous-gradient en $x \in \mathbb{R}^n$ si pour tout $y \in \mathbb{R}^n$:

$$f(y) \geq f(x) + v^T(y - x).$$

Définition : Sous-différentiel

L'ensemble de tous les sous-gradients en x est noté $\partial f(x)$.



Lien avec le gradient

Fonction lisse

Si f est convexe et différentiable en x, alors

$$\partial f(x) = \{\nabla f(x)\}.$$

Non lisse

Au point anguleux, plusieurs minorantes affines peuvent être exactes ; le sous-différentiel contient plusieurs vecteurs.

Géométrie

Chaque $v \in \partial f(x)$ définit un hyperplan d'appui au graphe de f.

Sous-différentiel de la valeur absolue

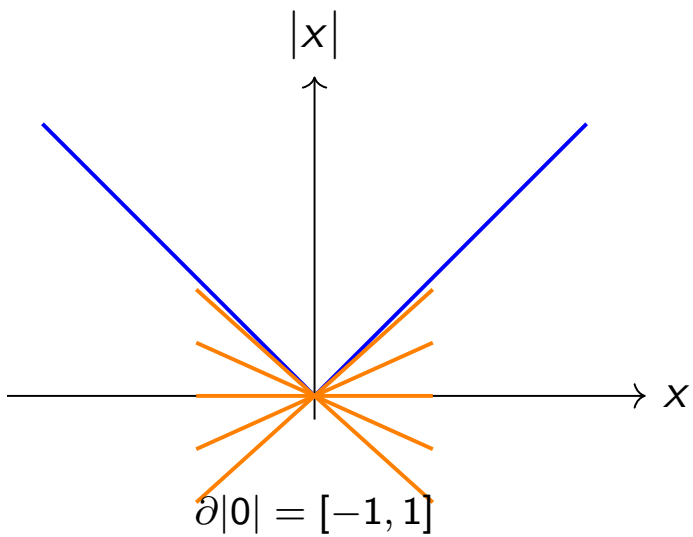
Définition : Calcul

Pour $f(t) = |t|$, on a

- $\partial f(t) = \{1\}$ si $t > 0$,
- $\partial f(t) = \{-1\}$ si $t < 0$
- $\partial f(0) = [-1, 1]$.

Remarque : Interprétation

Au point 0, toutes les pentes entre -1 et 1 donnent une minorante affine exacte.



Sous-différentiel de la norme ℓ_1

Définition : Norme ℓ_1

La norme ℓ_1 est définie par :

$$f(x) = \|x\|_1 = \sum_{i=1}^{n} |x_i| \,.$$

On a $v = (v_1, \ldots, v_n) \in \partial f(x)$ ssi

- $v_i = \text{sign}(x_i)$ lorsque $x_i \neq 0$;
- $v_i \in [-1, 1]$ lorsque $x_i = 0$.

Condition d'optimalité convexe non lisse

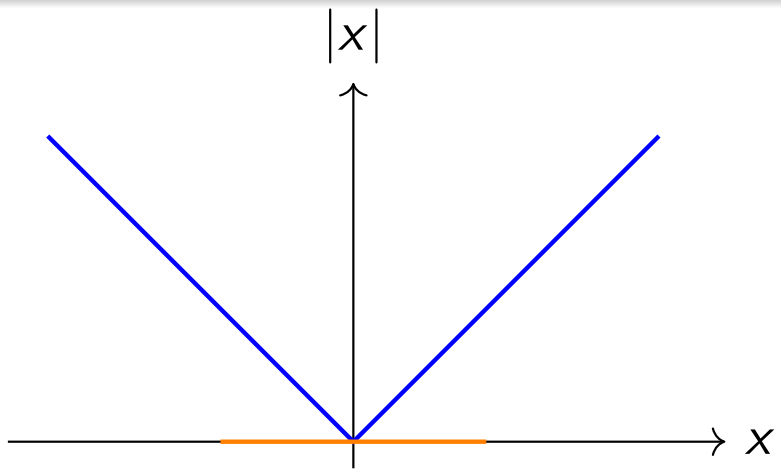
Proposition

Pour f convexe, $x^\star$ minimise f sur $\mathbb{R}^n$ si et seulement si

$$0 \in \partial f(x^\star).$$

Analogie

C'est la version non lisse de $\nabla f(x^\star) = 0$.



$$0 \in \partial|0| = [-1, 1]$$

Fonctions composites

Forme

On considère $F(x) = f(x) + g(x)$, avec f convexe différentiable et g convexe éventuellement non différentiable.

Optimalité

Une condition naturelle est $0 \in \nabla f(x^\star) + \partial g(x^\star)$.

Exemple TP

$f(a) = \frac{1}{2} \|Da - y\|_2^2$ et $g(a) = \lambda \|a\|_1$.

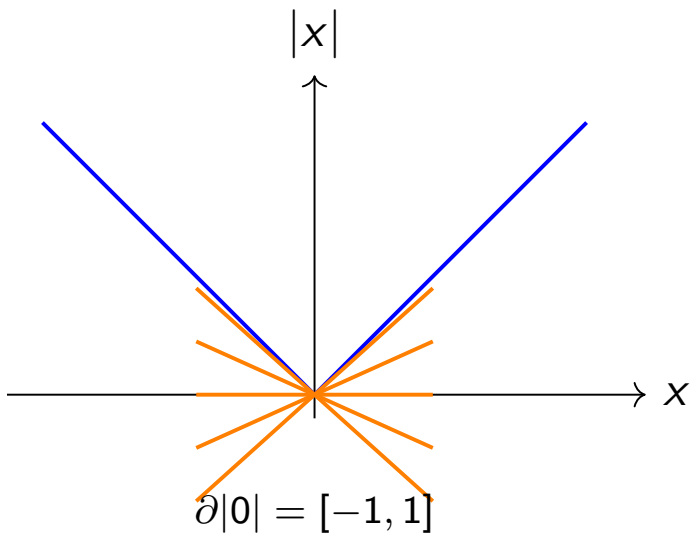
Algorithme du sous-gradient

Définition : Itération

Choisir $v_k \in \partial f(x_k)$ puis poser $x_{k+1} = x_k - \alpha_k v_k$.

Remarque : Différence avec le gradient

La direction $-v_k$ n'est pas toujours une direction de descente. Les pas décroissants jouent donc un rôle central.



Sous-gradient : avantages et limites

Avantage

Très général pour les fonctions convexes non lisses.

Limite

Convergence souvent lente et trajectoire parfois oscillante près des coins.

Usage

Utile comme méthode de référence ou lorsque l'opérateur proximal n'est pas disponible.

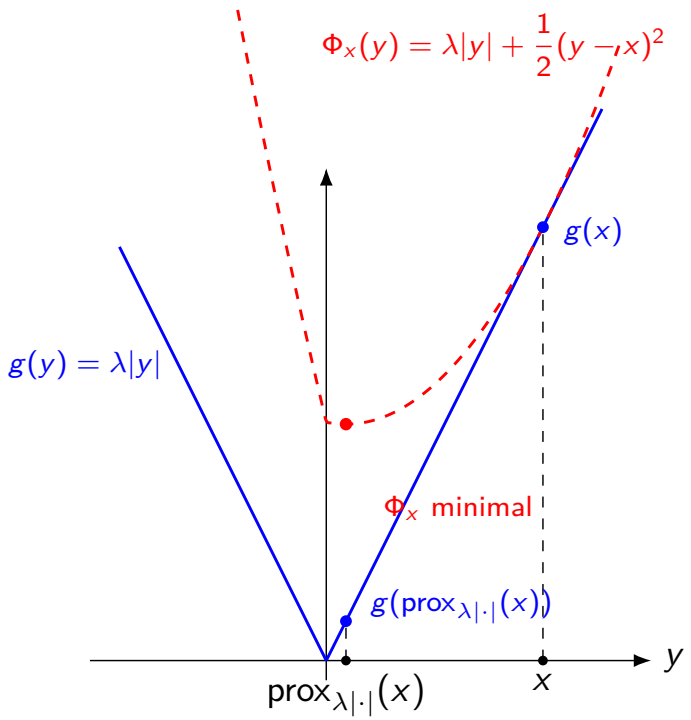
Opérateur proximal

Définition : Définition

Pour g convexe et $\gamma > 0$, $\text{prox}_{\gamma g}(z) = \arg\min_x \left(g(x) + \frac{1}{2\gamma} \|x - z\|_2^2 \right)$.

Interprétation

L'opérateur proximal résout le compromis entre minimiser g et rester proche du point x.



Proximal de la norme ℓ_1

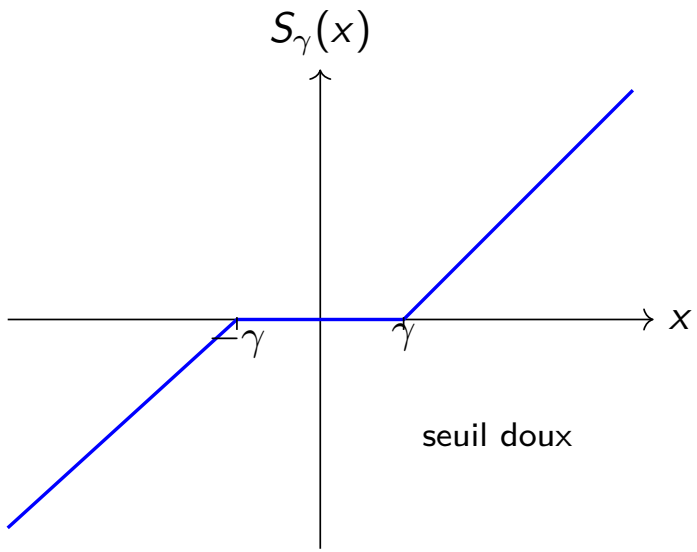
Définition : Seuil doux

Pour $g(x) = \lambda \|x\|_1$, le proximal agit composante par composante :

$$S_{\gamma\lambda}(z_i) = \text{sign}(z_i) \max(|z_i| - \gamma\lambda, 0).$$

Effet

Les petites composantes sont annulées ; les autres sont rapprochées de zéro. C'est le mécanisme numérique de parcimonie.



ISTA : problème visé

Objectif

Minimiser :

$$F(x) = f(x) + g(x),$$

avec f convexe différentiable à gradient lipschitzien et g convexe proximable.

Étape gradient

On descend d'abord sur la partie lisse :

$$z_k = x_k - \gamma \nabla f(x_k),$$

où γ doit respecter la régularité de ∇f.

Étape proximale

On applique ensuite

$$x_{k+1} = \text{prox}_{\gamma g}(z_k).$$

Le terme non différentiable est traité par l'opérateur proximal.

Optimisation sous contraintes

Problème avec contraintes d'égalité et d'inégalité

Formulation

Minimiser $f(x)$ sous $h(x) = 0$ et $g(x) \leq 0$, où $h : \mathbb{R}^n \to \mathbb{R}^m$ et $g : \mathbb{R}^n \to \mathbb{R}^p$.

Admissibilité

$\mathcal{A} = \{x \in \mathbb{R}^n : h(x) = 0,\ g(x) \leq 0\}$. Les comparaisons de coût se font uniquement sur $\mathcal{A}$.

Solutions

Les notions deglobale et locale sont identiques, mais restreintes aux points admissibles.

Épigraphe : principe

Définition : Épigraphe

Pour une fonction f, $\text{epi}(f) = \{(x, t) : f(x) \leq t\}$.

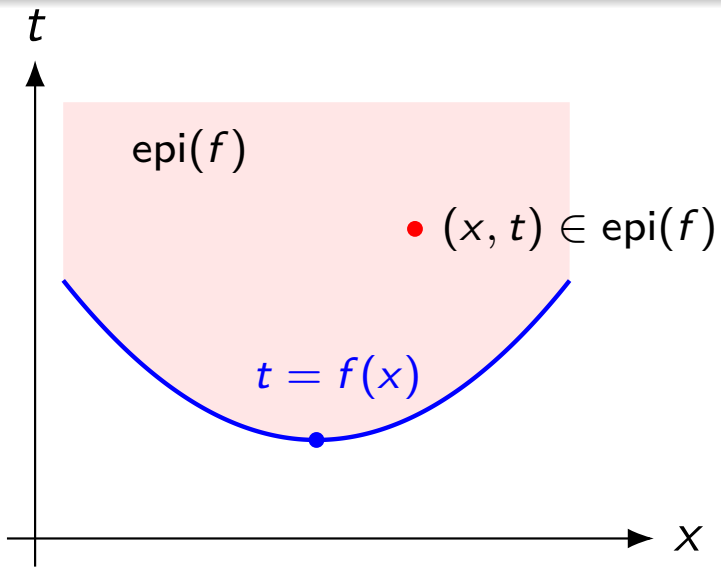
Reformulation

Minimiser $f(x)$ revient à minimiser t sous la contrainte $f(x) \leq t$, c'est sà dire :

$$\min_x f(x) = \min_{x,t} \{t \mid f(x) \leq t\}.$$

Propriété

f est convexe ssi $epi(f)$ est convexe



Épigraphe de la norme ℓ_1

Définition : Reformulation exacte

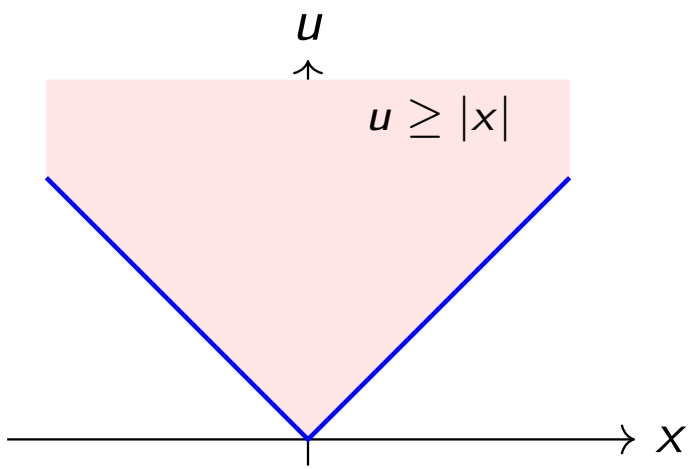
Le problème $\min_x q(x) + \lambda \|x\|_1$ se réécrit

$$\min_{x,u} q(x) + \lambda \sum_i u_i$$

sous $-u_i \leq x_i \leq u_i$.

Remarque

La valeur absolue disparaît de l'objectif : le problème devient lisse avec des contraintes linéaires.



Deux lectures de la pénalisation ℓ_1

Composite

$q(x) + \lambda \|x\|_1$ mène à l'opérateur proximal et à ISTA.

Contraint lisse

L'épigraphe mène à des contraintes linéaires et à un solveur de programmation non linéaire.

Choix pratique

La meilleure écriture dépend de la taille, des contraintes supplémentaires, du besoin de certificats et des outils disponibles.

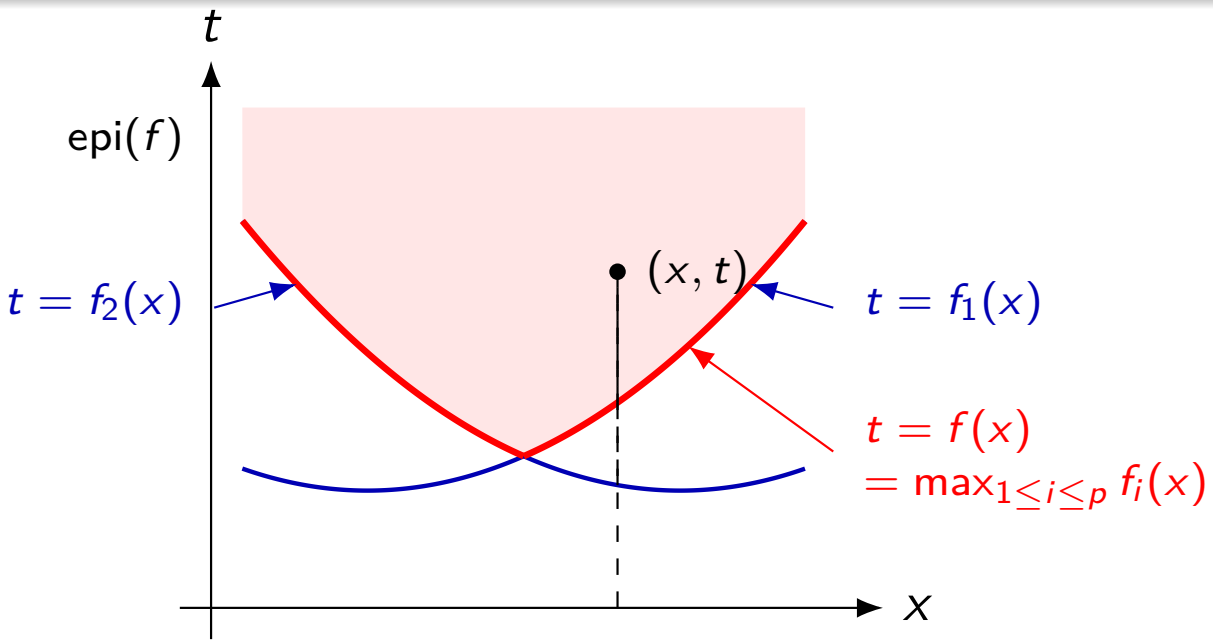
Épigraphe d'une fonction maximum

Définition : Maximum

Pour $f(x) = \max_{1 \leq i \leq p} f_i(x)$, on introduit t et les contraintes $f_i(x) \leq t$.

Équivalence

$\min_x \max_i f_i(x) \iff \min_{x,t} t$ sous $f_i(x) \leq t$ pour tout i.



Relaxation : principe

Définition : Relaxation

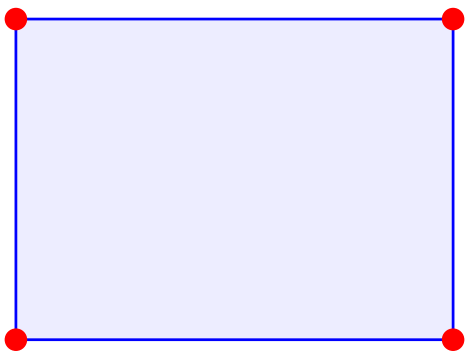
Une relaxation remplace $\mathcal{A}$ par un ensemble $\widetilde{\mathcal{A}}$ tel que $\mathcal{A} \subset \widetilde{\mathcal{A}}$.

Borne

En minimisation, la valeur relâchée est inférieure ou égale à la valeur du problème initial :

$$\min_{x \in \widetilde{\mathcal{A}}} f(x) \leq \min_{x \in \mathcal{A}} f(x)$$

Exemple : $\mathcal{A} = \{0,1\}^2$ et $\widetilde{\mathcal{A}} = [0,1]^2$



relaxation $\{0,1\}^2 \to [0,1]^2$

Relaxation convexe et PLNE

Tractabilité

Une relaxation convexe est souvent plus facile à résoudre et possède des certificats globaux.

Borne inférieure

Pour un problème de minimisation avec variables entières, la relaxation continue fournit une borne inférieure.

Utilité

En programmation linéaire en nombres entiers, ces bornes guident les méthodes de branchement et coupures.

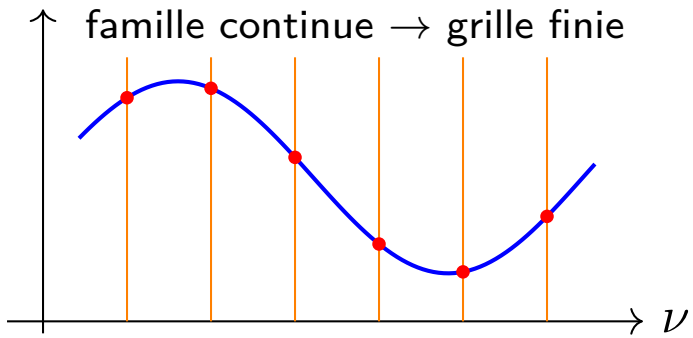
Approximation par discrétisation

Définition : Discrétisation

Une variable continue $\nu \in [\nu_{\min}, \nu_{\max}]$ peut être remplacée par une grille $\{\nu_1, \ldots, \nu_p\}$.

Remarque : Attention

La solution du problème discret n'est pas automatiquement solution du problème continu ; une erreur hors grille apparaît.



Reformuler, approximer, relaxer

Reformuler

On cherche à garder le même problème tout en révélant une structure exploitable.

Approximer

On accepte une erreur contrôlée pour obtenir un problème fini ou moins coûteux.

Relaxer

On agrandit l'ensemble admissible pour calculer une borne et guider la résolution.

Lagrangien

Problème général

Minimiser $f(x)$ sous $h(x) = 0$ et $g(x) \leq 0$, où $h : \mathbb{R}^n \to \mathbb{R}^m$ et $g : \mathbb{R}^n \to \mathbb{R}^p$.

Définition

Le lagrangien est $\mathcal{L}(x, \lambda, \mu) = f(x) + \lambda^T h(x) + \mu^T g(x)$.

Multiplicateurs

$\lambda \in \mathbb{R}^m$ est libre ; $\mu \in \mathbb{R}^p$ vérifie $\mu \geq 0$ pour des contraintes $g_i(x) \leq 0$.

Interprétation

La stationnarité du coût est compensée par les normales aux contraintes actives.

Conditions de Karush–Kuhn–Tucker

Proposition : KKT, forme de premier ordre

Sous une qualification de contraintes adaptée, si $x^\star$ est un minimum local régulier, alors il existe $\lambda^\star \in \mathbb{R}^m$ et $\mu^\star \in \mathbb{R}^p_+$ tels que

$$\nabla_x \mathcal{L}(x^\star, \lambda^\star, \mu^\star) = 0, \quad h(x^\star) = 0, \quad g(x^\star) \leq 0, \quad \mu_i^\star g_i(x^\star) = 0.$$

Remarque : Lecture

La complémentarité impose qu'une contrainte inactive ait un multiplicateur nul.

Illustration des conditions de KKT

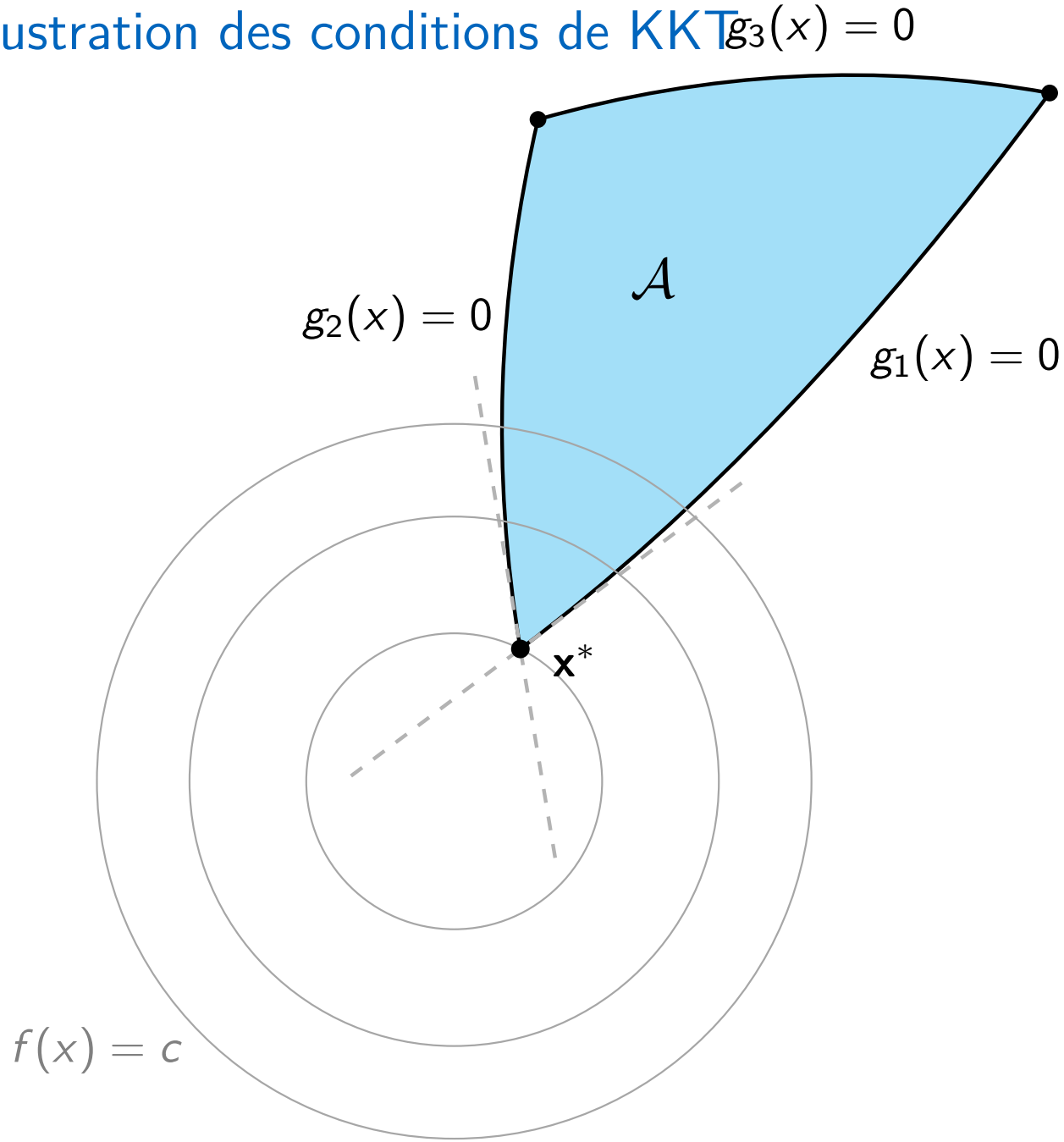


Illustration des conditions de KKT

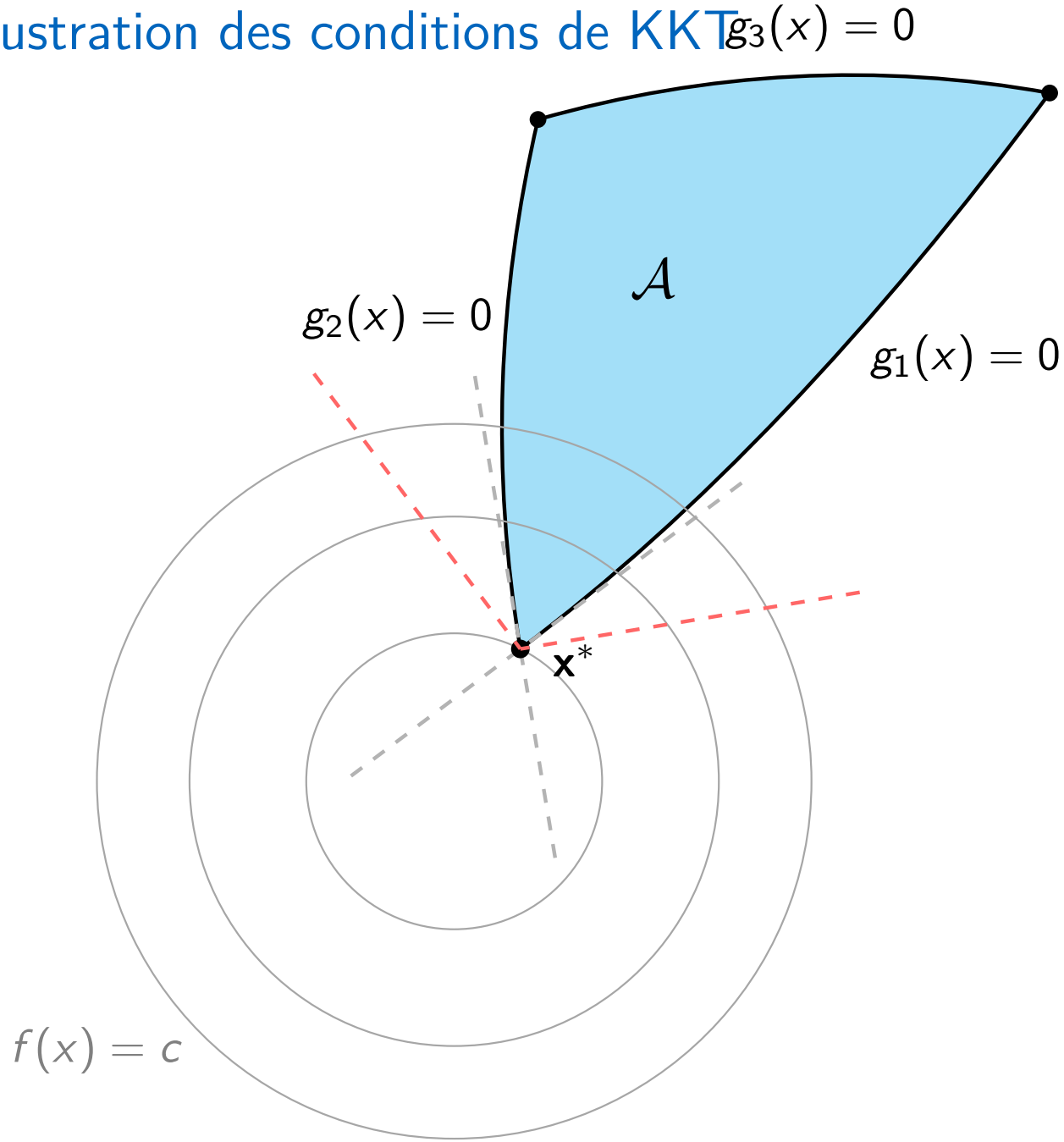


Illustration des conditions de KKT

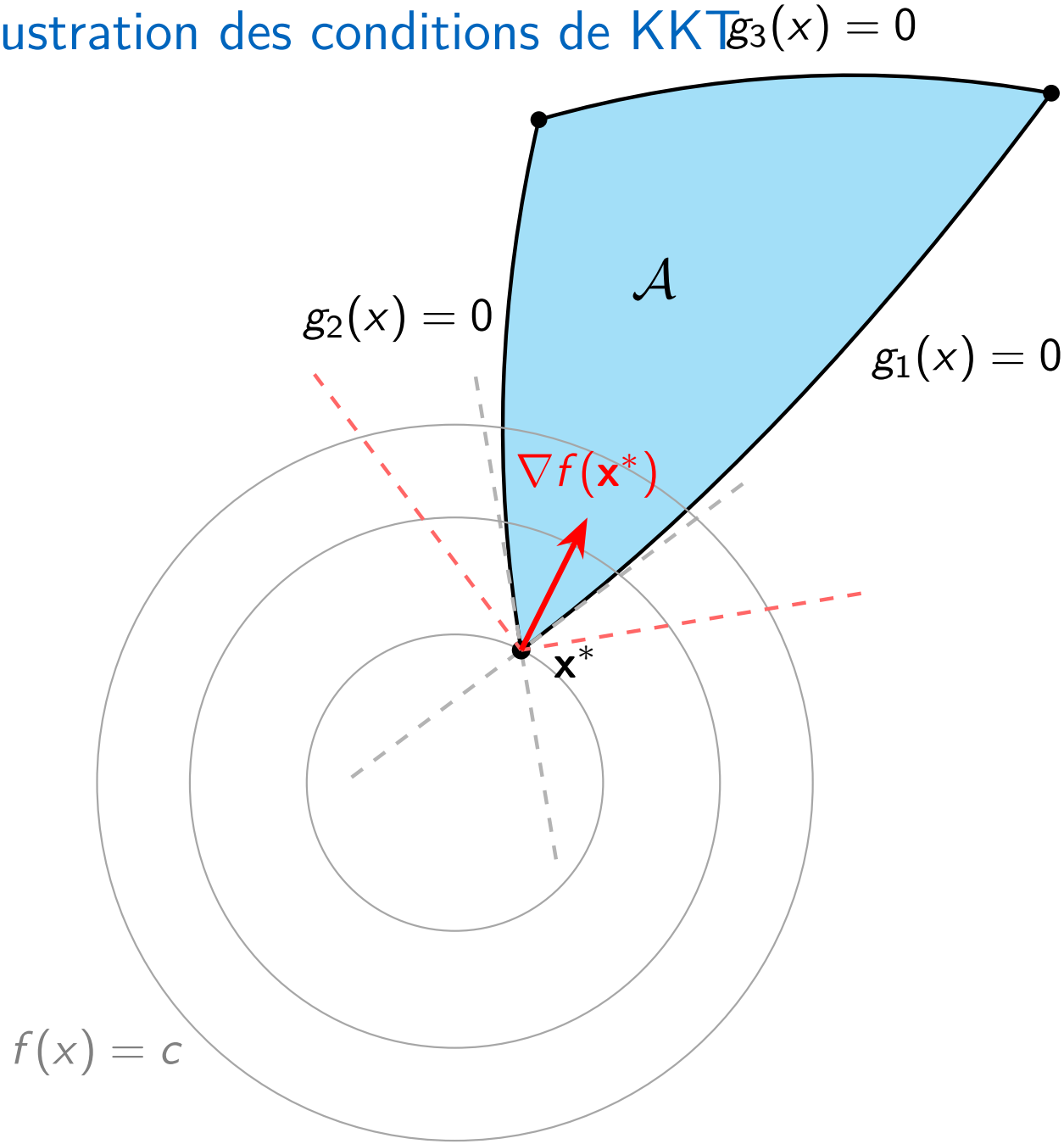


Illustration des conditions de KKT

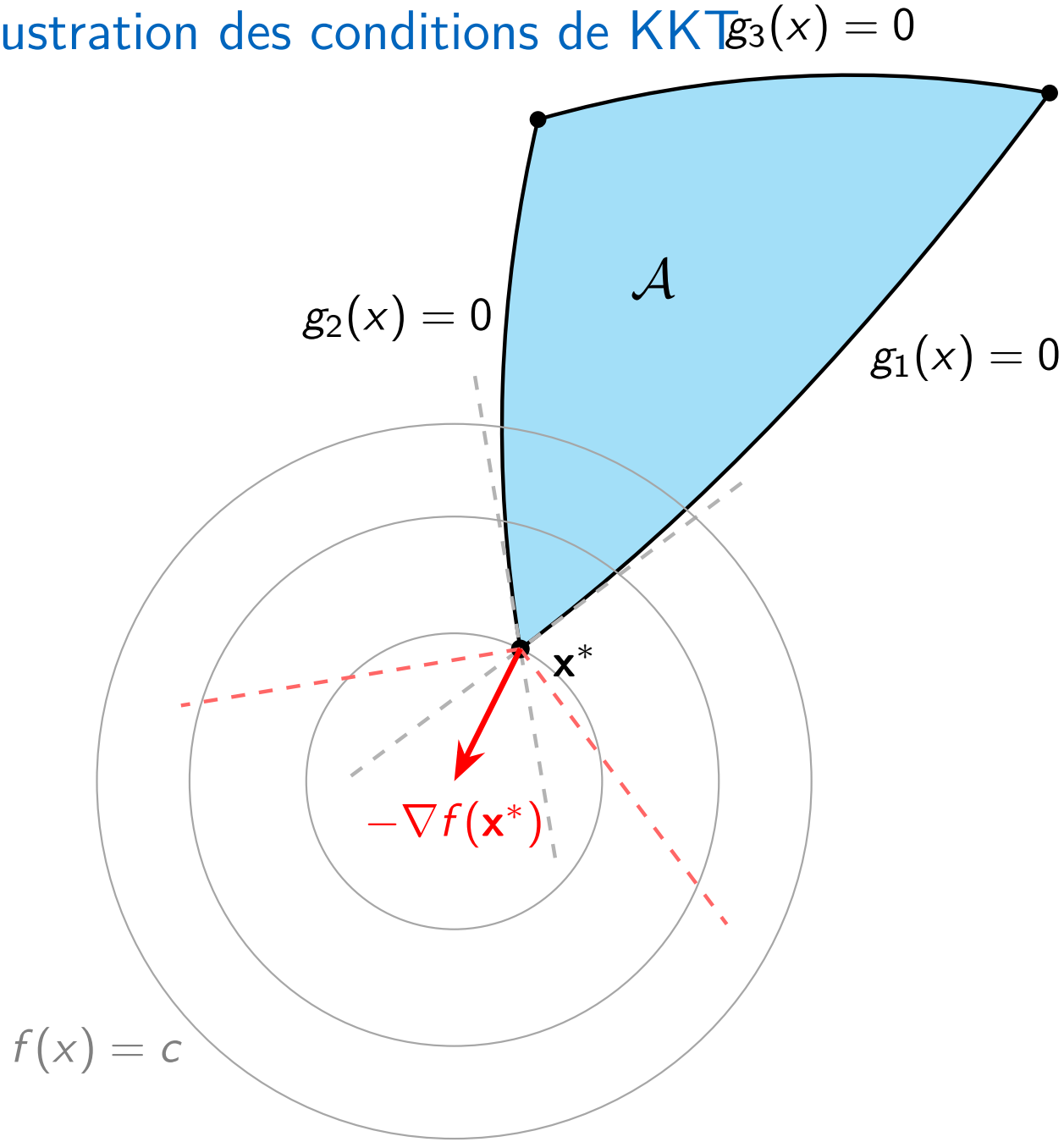
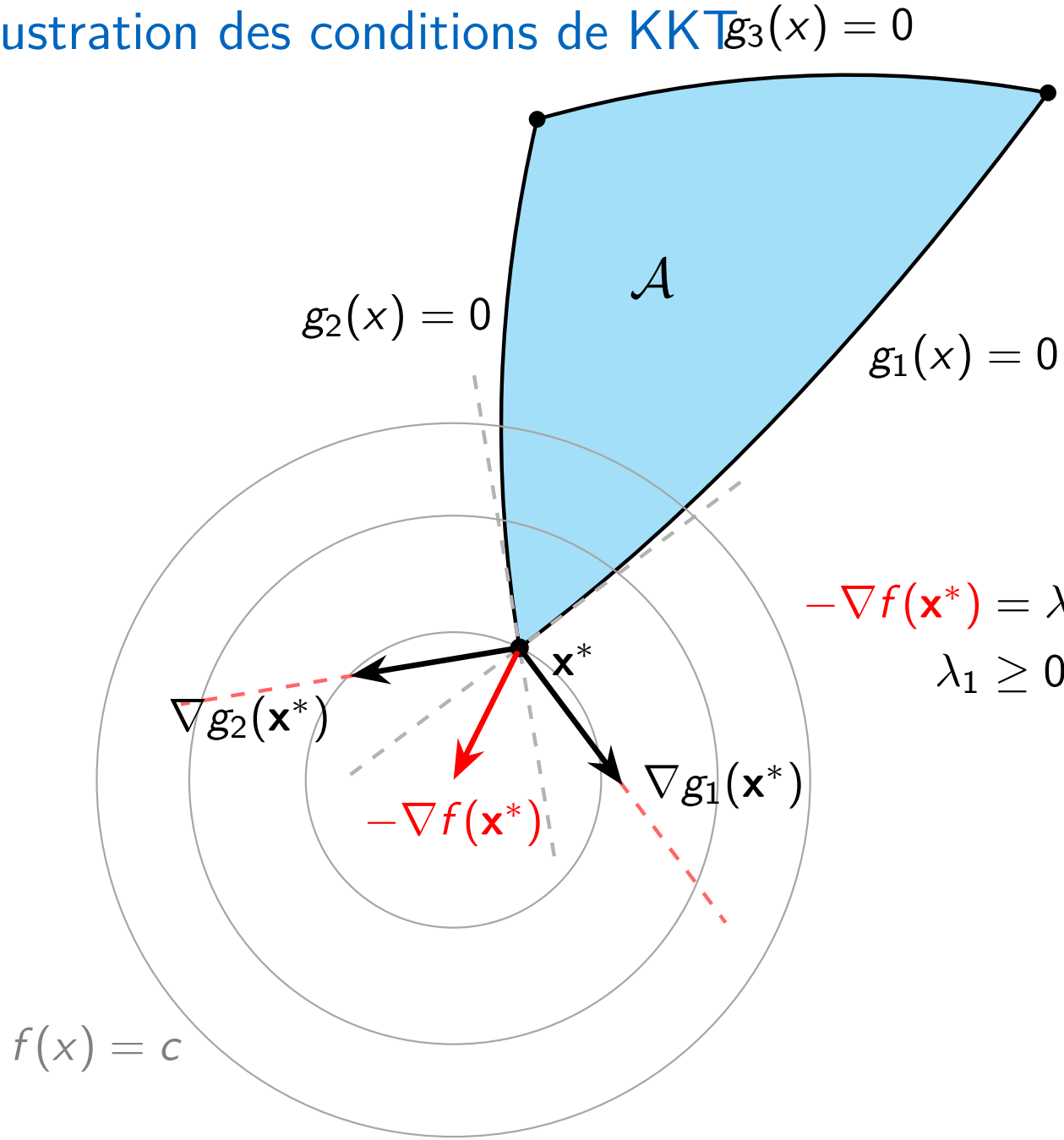


Illustration des conditions de KKT



$$-\nabla f(\mathbf{x}^*) = \lambda_1 \nabla g_1(\mathbf{x}^*) + \lambda_2 \nabla g_2(\mathbf{x}^*)$$
$$\lambda_1 \geq 0,\ \lambda_2 \geq 0$$

Résidus de KKT

Faisabilité

Mesurer $\|h(x)\|$ et les violations positives $\max(g_i(x), 0)$.

Stationnarité

Mesurer $\|\nabla_x \mathcal{L}(x, \lambda, \mu)\|$.

Complémentarité

Mesurer $|\mu_i g_i(x)|$ et vérifier $\mu_i \geq 0$.

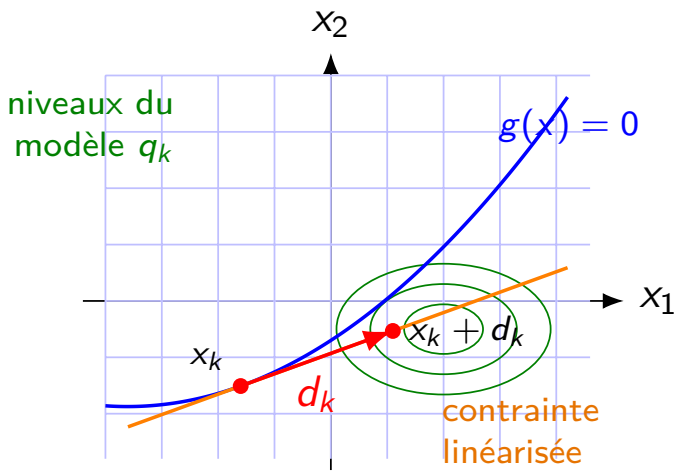
SQP : idée générale

Définition : SQP

La programmation quadratique séquentielle remplace localement le problème contraint par un sous-problème quadratique avec contraintes linéarisées.

Analogie

C'est l'équivalent contraint de Newton : on linéarise les contraintes et on approxime la courbure du lagrangien.



Sous-problème SQP

Modèle

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

sous les contraintes $h(x_k) + \nabla h(x_k) d = 0$ et $g(x_k) + \nabla g(x_k) d \leq 0$, où B_k est à définir

Itération

Après résolution, on pose souvent $x_{k+1} = x_k + \alpha_k d_k$ avec une globalisation.

SQP : avantages et difficultés

Avantages

Bonne efficacité locale, exploitation des dérivées, comportement proche de Newton près d'une solution régulière.

Difficultés

Le sous-problème peut être coûteux ou infaisable ; il faut gérer la faisabilité, la fonction de mérite et le choix du pas.

Solveurs

SLSQP, SNOPT et des variantes de méthodes de points actifs ou de régions de confiance sont couramment utilisées.

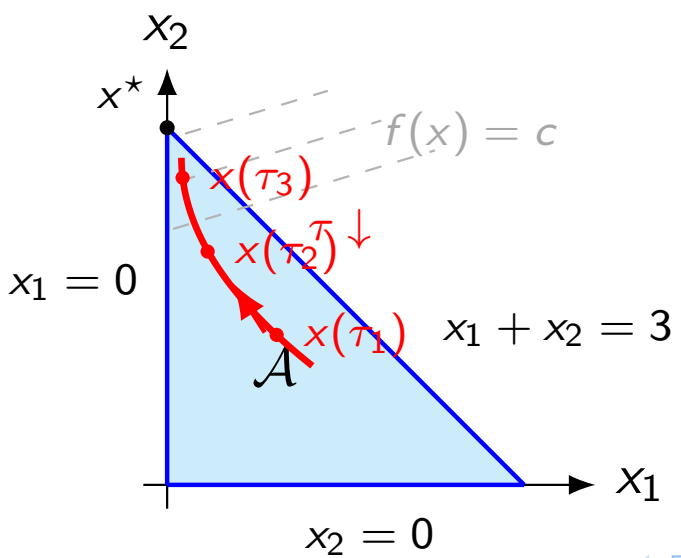
Points intérieurs

Définition : Barrière et Chemin central

Pour $g_i(x) < 0$, on remplace les inégalités par $f(x) - \tau \sum_i \log(-g_i(x))$.
Lorsque τ diminue, les minimiseurs des problèmes barrières approchent la frontière active sans la franchir.

Exemple

$$\begin{aligned} \min_{x_1, x_2} \quad & f(x_1, x_2) = x_1 - x_2 \\ \text{sous les contraintes} \quad & x_1 \geq 0,\ x_2 \geq 0, \\ & x_1 + x_2 \leq 3. \end{aligned}$$



Points intérieurs : forces et limites

Forces

Bonne adaptation aux grands problèmes lisses, exploitation des dérivées et traitement naturel de nombreuses inégalités.

Limites

Nécessite un point intérieur ou des variables d'écart ; les systèmes linéaires internes et l'échelle des variables peuvent dominer le coût.

Solveurs

IPOPT est un solveur libre majeur de programmation non linéaire par points intérieurs.

Utilisations des solveurs

Statut

Le message du solveur doit être lu avec les tolérances et les résidus KKT.

Échelle

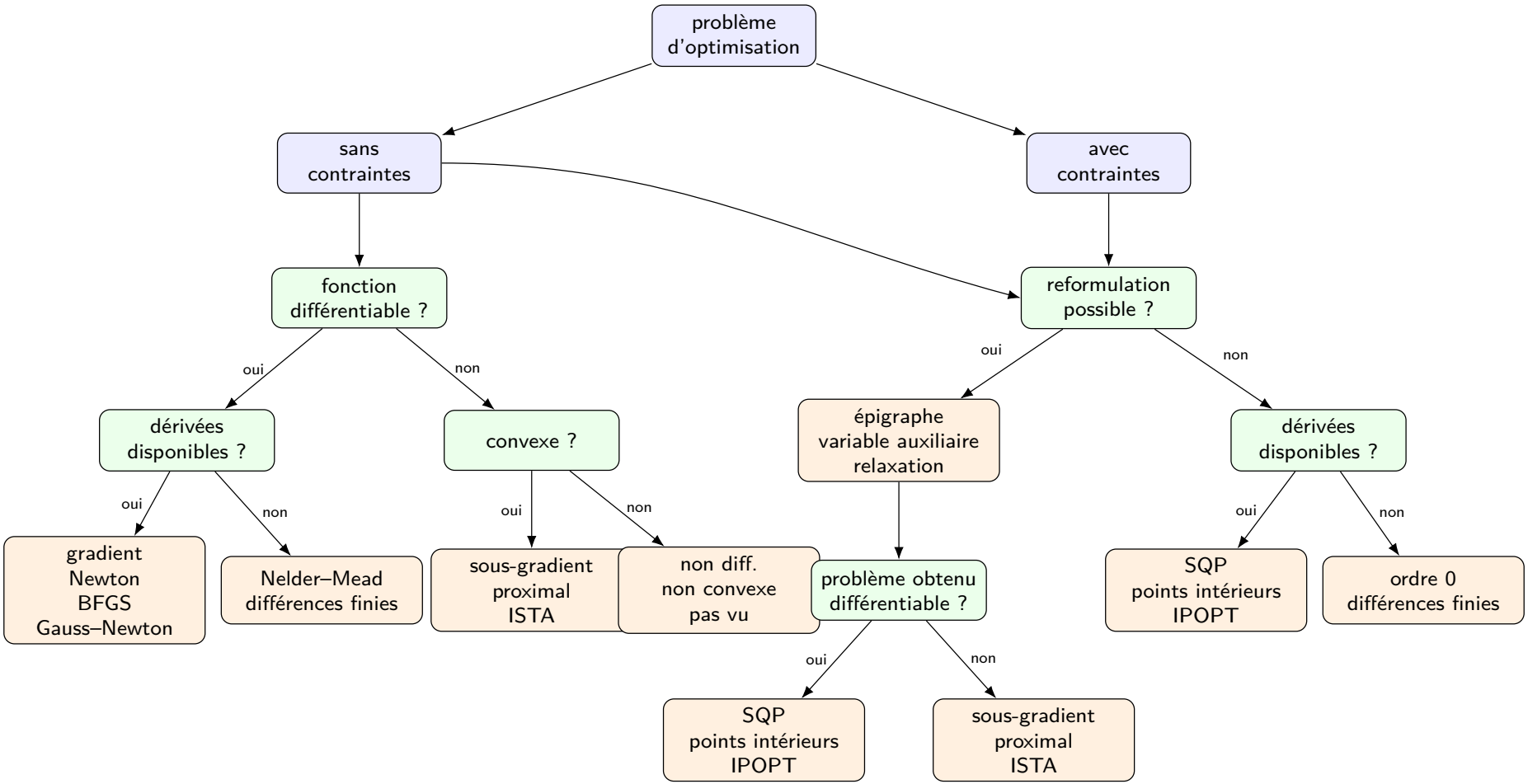
Des variables mal mises à l'échelle dégradent les systèmes internes et les critères d'arrêt.

Initialisation

L'initialisation influence fortement les problèmes non convexes et peut affecter la faisabilité intermédiaire.

Synthèse des cadres et solveurs

Arbre de décision



Choisir une famille d'algorithmes

Question	Oui	Non
Contraintes explicites ?	KKT, SQP, points intérieurs	méthodes sans contrainte
Coût différentiable ?	gradient, Newton, quasi-Newton	sous-gradient, proximal, reformulation
Gradient disponible ?	ordre 1 ou 2	ordre 0 ou différences finies prudentes
Composite $f + g$ proximable ?	ISTA/FISTA	reformulation ou sous-gradient
Convexe ?	certificats globaux possibles	optimum local, initialisation importante

Repères de solveurs

Cadre	Méthodes typiques	Outils à connaître
Boîte noire locale	Nelder–Mead, Powell, COBYLA	SciPy, NLopt
Lisse sans contrainte	BFGS, L-BFGS, Newton-CG	SciPy, Julia Optim, MATLAB
Contraintes lisses	SLSQP, trust-constr, IPOPT	SciPy, CasADi+IPOPT, SNOPT
Composite convexe	ISTA, FISTA, proximal	implémentations dédiées, CVXPY selon problème
PL/PLNE	simplexe, points intérieurs, branchement	HiGHS, GLPK, CBC, Gurobi, CPLEX

Remarque : Prudence

Un nom de solveur ne remplace pas l'analyse : type de problème, dérivées, convexité, échelles et tolérances restent déterminants.

Messages à retenir

Cadre

Convexité, différentiabilité et contraintes déterminent le sens des conditions d'optimalité.

Formulation

Épigraphe, relaxation, discrétisation et pénalisation peuvent changer la famille numérique du problème.

Analyse des résultats

Un bon résultat combine faible coût, faisabilité, stationnarité et pertinence applicative.

Passage au BE

Dictionnaire

La discrétisation transforme une famille continue de fréquences en colonnes de matrice.

Parcimonie

La norme ℓ_1 remplace un comptage combinatoire par une pénalisation convexe non lisse.

ISTA

Le proximal de ℓ_1 fournit une itération simple, interprétable et directement programmable.

Références succinctes


J.-C. Gilbert, *Fragments d'optimisation différentiable – Théories et algorithmes*, HAL, version du 9 janvier 2026.


J. Nocedal et S. Wright, *Numerical Optimization*, Springer, 2006.


S. Boyd et L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.


S. J. Wright, *Primal-Dual Interior-Point Methods*, SIAM, 1997.