Optimisation

1. Optimisation lisse sans contraintes
2. Optimisation lisse avec contraintes
3. Analyse convexe et dualité Lagrangienne
4. Ouverture vers l'optimisation non-lisse
5. Algorithmes d'optimisation

Optimisation

1. Optimisation lisse sans contraintes
2. Optimisation lisse avec contraintes
3. Analyse convexe et dualité Lagrangienne
4. Ouverture vers l’optimisation non-lisse
5. Algorithmes d’optimisation

Convexité

L'ensemble X est convexe ssi pour tout x, y dans X, le segment $[x, y] := \{tx + (1-t)y \mid t \in [0,1]\}$ est inclus dans X

Une fonction $f : X \subset \mathbf{R}^n \to R$ est convexe ssi X est convexe et pour tout x, y dans X et pour tout $t \in [0,1]$,

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

Si l'inégalité est stricte pour $t \in]0,1[$, la fonction est dite *strictement* convexe.

Convexité

Reconnaître une fonction convexe

- les fonctions linéaires sont convexes
- les combinaisons linéaires positives de fonctions convexes sont convexes
- le maximum (point par point) d'un ensemble de fonctions convexes est convexe
- …

Si f est de classe C^2, f est convexe si et seulement si $\nabla^2 f$ est semi-définie positive sur l'intérieur de X. Si $\nabla^2 f$ est définie positive sur l'intérieur de X, f est strictement convexe. $(X = \mathrm{dom} f)$

Hessienne

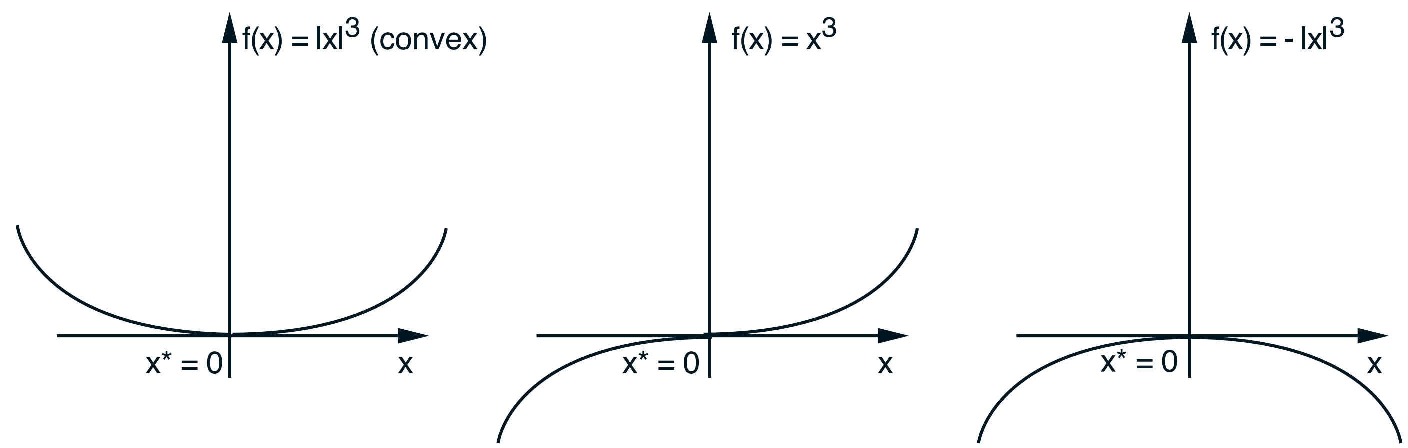
$$f : U \subset \mathbf{R}^n \mapsto \mathbf{R}$$

$$\nabla^2 f = H(f) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1} & \dfrac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Conditions d'optimalité

Conditions nécessaires, premier ordre

Si x^* est un minimum local et f est de classe $\mathcal{C}^1$ sur un ouvert $U \subset \mathbf{R}^n$ contenant x, alors $\nabla f(x^*) = 0$



Conditions nécessaires, second ordre

Si x^* est un minimum local et f est de classe $\mathcal{C}^2$ sur un ouvert $U \subset \mathbf{R}^n$ contenant x, alors $\nabla f(x^*) = 0$ et $\nabla^2 f(x^*)$ est semi-définie positive

Conditions suffisantes, second ordre

Si f est de classe $\mathcal{C}^2$ sur un ouvert $U \subset \mathbf{R}^n$ contenant x, si $\nabla f(x^*) = 0$ et si $\nabla^2 f(x^*)$ est définie positive, alors x^* est un minimum local.

Existence et unicité du minimum

Théorèmes

Si f est convexe et minorée, alors f admet un minimum global.

Si f est une fonction *strictement* convexe définie sur un ensemble convexe $X \subset \mathbf{R}^n$, alors f admet au plus un minimum global.

Si f est une fonction convexe définie sur un ensemble convexe $X \subset \mathbf{R}^n$, alors tout minimum local de f est aussi un minimum global de f. Si X est ouvert, alors $\nabla f(x^*) = 0$ est équivaut à x^* est un minimum global de f

Dualité

Problème considéré

$$\min_{x \in X \subset \mathbf{R}^d} f(x) \text{ sous les contraintes } \begin{cases} e_i(x) = 0, \; i = 1..r \\ c_i(x) \leq 0, \; i = 1..s \end{cases}$$

$$f : X \to \mathbf{R}$$
$$r \in \mathbf{N}$$
$$s \in \mathbf{N}$$

$$e_i : X \to \mathbf{R}$$
$$c_i : X \to \mathbf{R}$$

$$\Omega = \{x \in X | \; e_i(x) = 0, \; c_j(x) \leq 0, \; i = 1..r, \; j = 1..s\}$$

$$f^* = \inf_{x \in \Omega} f(x)$$

Hypothèses : $\Omega \neq \emptyset$

$f^* > -\infty$

Dualité

Lagrangien associé au problème

$$L(x, \mu, \lambda) = f(x) + \mu^T e(x) + \lambda^T c(x)$$

avec :

$$x \in X$$
$$c(x) = [c_1(x)\ c_2(x)\ \ \dots\ \ c_s(x)]^T$$
$$e(x) = [e_1(x)\ e_2(x)\ \ \dots\ \ e_r(x)]^T$$
$$\mu \in \mathbf{R}^r$$
$$\lambda \in \mathbf{R}^s$$

Dualité

Fonction duale $q(\mu, \lambda) = \inf_{x \in X} L(x, \mu, \lambda)$

Domaine $D = \{\mu, \lambda \mid q(\mu, \lambda) > -\infty\}$

Problème dual

$$\max_{(\mu,\lambda)\in D} q(\mu, \lambda) \text{ sous les contraintes } \lambda_i \geq 0,\ i = 1..s$$

Résultats

$$D \text{ est convexe et } q \text{ est concave sur } D$$

$q^* \leq f^*$ (dualité faible)

avec $$q^* = \sup_{\mu \in \mathbf{R}^r,\ \lambda \in \mathbf{R}^s, \lambda_i \geq 0,\ i=1..s} q(\mu, \lambda)$$

Dualité

Dualité et convexité

Dualité

Dualité forte, convexité et qualification des contraintes

Théorème 1:

Si X est polyhédral, si f est convexe et si les contraintes e_i et c_j sont affines, alors il n'y a pas d'écart dual: $q^* = f^*$

(Un ensemble polyhédral est un ensemble qui peut être obtenu comme l'intersection d'un ensemble fini de demi-espaces fermés)

Théorème 2 :

Slater's condition. Si X est convexe, si f est convexe, si les contraintes c_i sont convexes, si les contraintes e_j sont affines et s'il existe un point x_0 dans l'intérieur relatif de X tel que $c_i(x_0) < 0$ et $e_j(x_0) = 0$ pour $i = 1..s$ et $j = 1..r$, alors il n'y a pas d'écart dual: $q^* = f^*$

(L'intérieur d'un ensemble et cet ensemble privé de sa frontière. L'intérieur relatif d'un ensemble S est l'intérieur de cette ensemble au sein de son enveloppe affine, c'est à dire au sein du plus petit espace affine contenant S.)

Dualité

Retrouver la solution du primal à partir de la solution du dual en cas de dualité forte

En cas de dualité forte ($q^* = f^*$), les conditions :

- $x^* \in X$ et $e_i(x^*) = 0$ et $c_j(x^*) \leq 0$ (faisabilité primale)
- $\lambda^* \geq 0$ (faisabilité duale)
- $x^* = \arg\min_{x \in X} L(x, \mu^*, \lambda^*)$ (optimalité Lagrangienne)
- $\mu_j^* c_j(x^*) = 0,\ \ j = 1..s$ (complémentarité)

sont des conditions nécessaires et suffisantes pour que x^* soit une solution optimale du problème primal et (μ^*, λ^*) une solution optimale du problème dual.

Optimisation

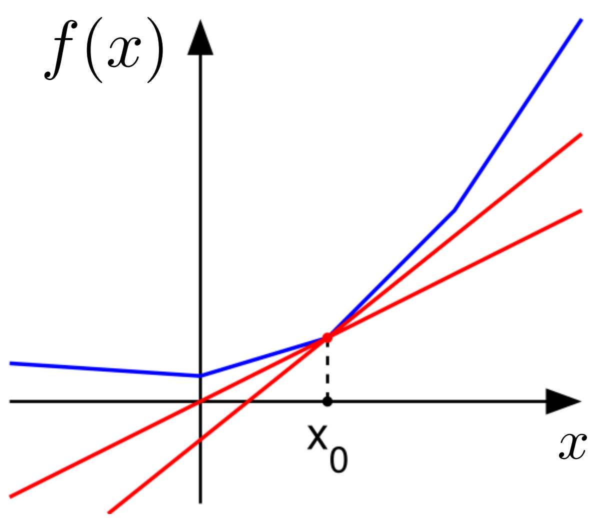
1. Optimisation lisse sans contraintes
2. Optimisation lisse avec contraintes
3. Analyse convexe et dualité Lagrangienne
4. Ouverture vers l'optimisation non-lisse
5. Algorithmes d'optimisation

Optimisation

1. Optimisation lisse sans contraintes
2. Optimisation lisse avec contraintes
3. Analyse convexe et dualité Lagrangienne
4. Ouverture vers l'optimisation non-lisse
5. Algorithmes d'optimisation

Un mot sur l'optimisation non lisse

Convexe



Sous-gradient (subgradient) $\partial f(x_0)$

Condition nécessaire de minimum local pour une fonction convexe (pas forcément lisse) :
Soit $f : \mathbf{R}^d \to \mathbf{R}$ une function convexe. Si x^* est un minimum local de f, alors : $0 \in \partial f(x^*)$

Version KKT disponible pour le cas convexe non-lisse sous contraintes

Non-convexe

Domaine de recherche ouvert

Cas particulier utile en apprentissage automatique :

Condition nécessaire de minimum local pour la somme d'une fonction lisse (pas forcément convexe) et d'une fonction convexe (pas forcément lisse) :
Soit $f : \mathbf{R}^d \to \mathbf{R}$ de classe $\mathcal{C}^1$ et $g : \mathbf{R}^d \to \mathbf{R}$ une function convexe. Si x^* est un minimum local de $f + g$, alors :

$$0 \in \nabla f(x^*) + \partial g(x^*) = \{\nabla f(x^*) + x \mid x \in \partial g(x^*)\}.$$

Optimisation

1. Optimisation lisse sans contraintes
2. Optimisation lisse avec contraintes
3. Analyse convexe et dualité Lagrangienne
4. Ouverture vers l’optimisation non-lisse
5. Algorithmes d’optimisation

Optimisation

1. Optimisation lisse sans contraintes
2. Optimisation lisse avec contraintes
3. Analyse convexe et dualité Lagrangienne
4. Ouverture vers l'optimisation non-lisse
5. Algorithmes d'optimisation

Descente de gradient

Pas fixe

Input: $x_0 \in \mathbf{R}^n$, $f : \mathbf{R}^n \to \mathbf{R}$ de classe $\mathcal{C}^1$, pas fixe $\gamma > 0$.

$$\text{Iteration: } x_{k+1} = x_k - \gamma \nabla f(x_k)$$

Descente de gradient

'Backtracking' avec la règle d'Armijo

Input: $x_0 \in \mathbf{R}^n$, $f : \mathbf{R}^n \to \mathbf{R}$ de classe $\mathcal{C}^1$, $s > 0, 0 < \beta < 1, 0 < \sigma < 1$.

$$\text{Iteration : } \; x_{k+1} = x_k - \alpha_k \nabla f(x_k)$$

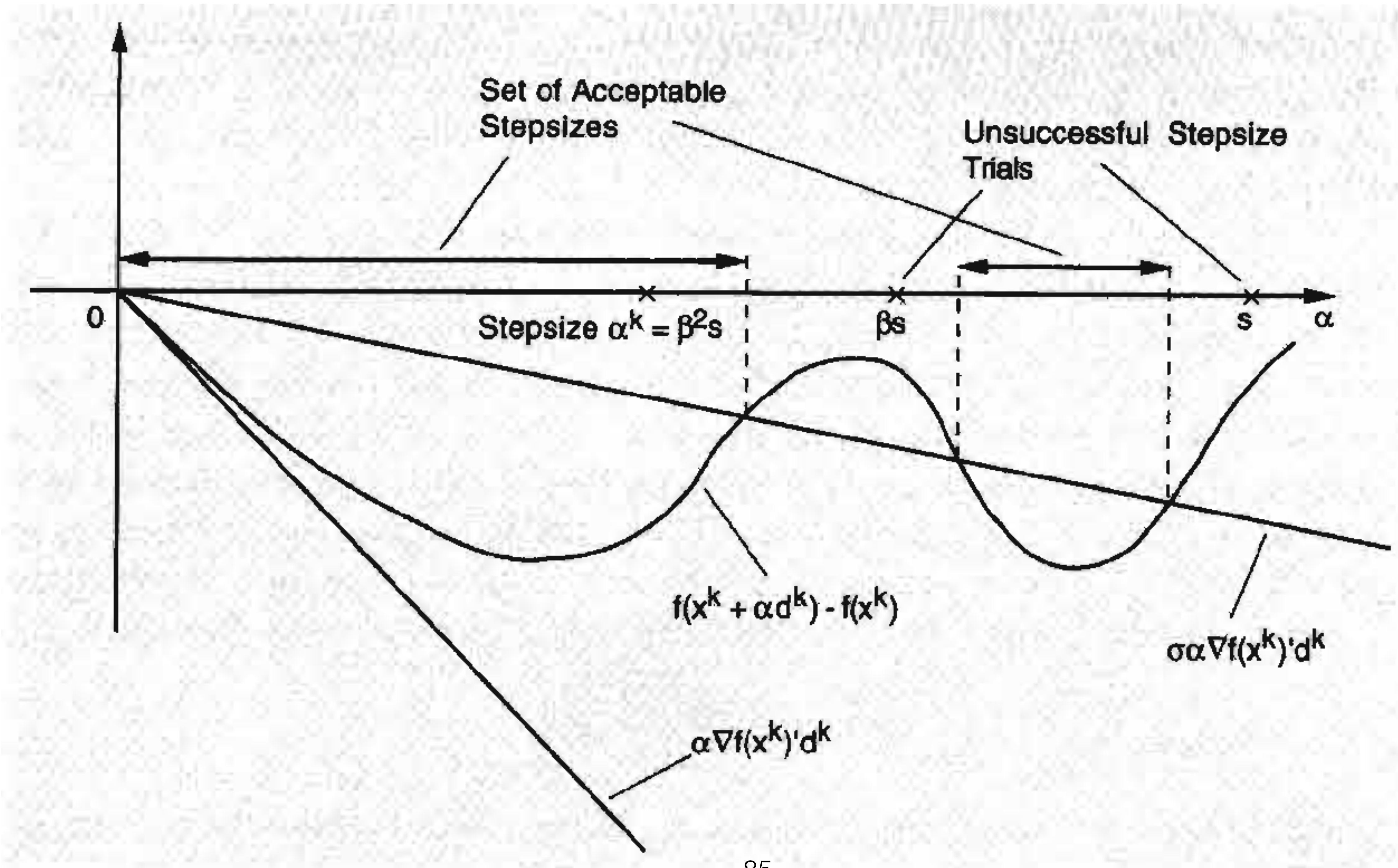
$$\alpha_k = \beta^{m_k} s$$

m_k plus petit entier positif tel que

$$f(x_k) - f(x_{k+1}) \geq \sigma \beta^{m_k} s \nabla f(x_k)^T \nabla f(x_k)$$

Descente de gradient

'Backtracking' avec la règle d'Armijo



Descente de gradient

'Backtracking' avec la règle d'Armijo

Input: $x_0 \in \mathbf{R}^n, f : \mathbf{R}^n \to \mathbf{R}$ de classe $\mathcal{C}^1, s > 0, 0 < \beta < 1, 0 < \sigma < 1.$

$$\text{Iteration : } \; x_{k+1} = x_k - \alpha_k \nabla f(x_k)$$

$$\alpha_k = \beta^{m_k} s$$

m_k plus petit entier positif tel que

$$f(x_k) - f(x_{k+1}) \geq \sigma \beta^{m_k} s \nabla f(x_k)^T \nabla f(x_k)$$

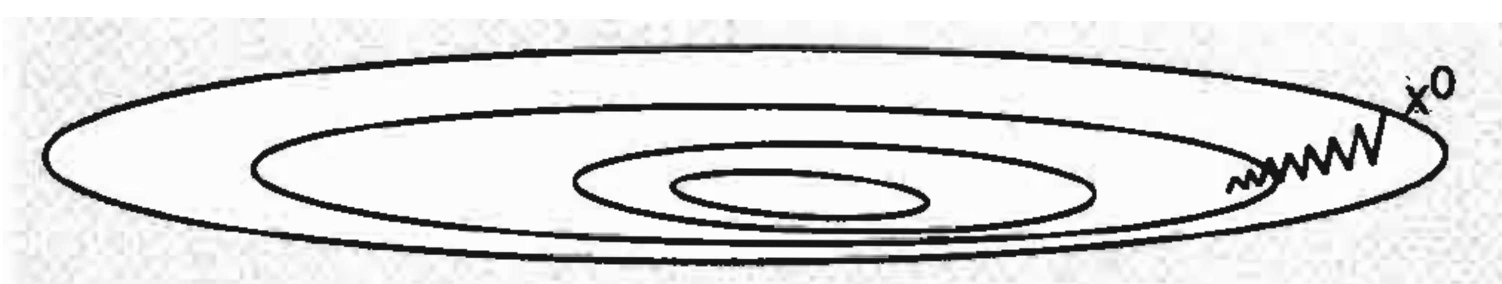
Descente de gradient

Convergence

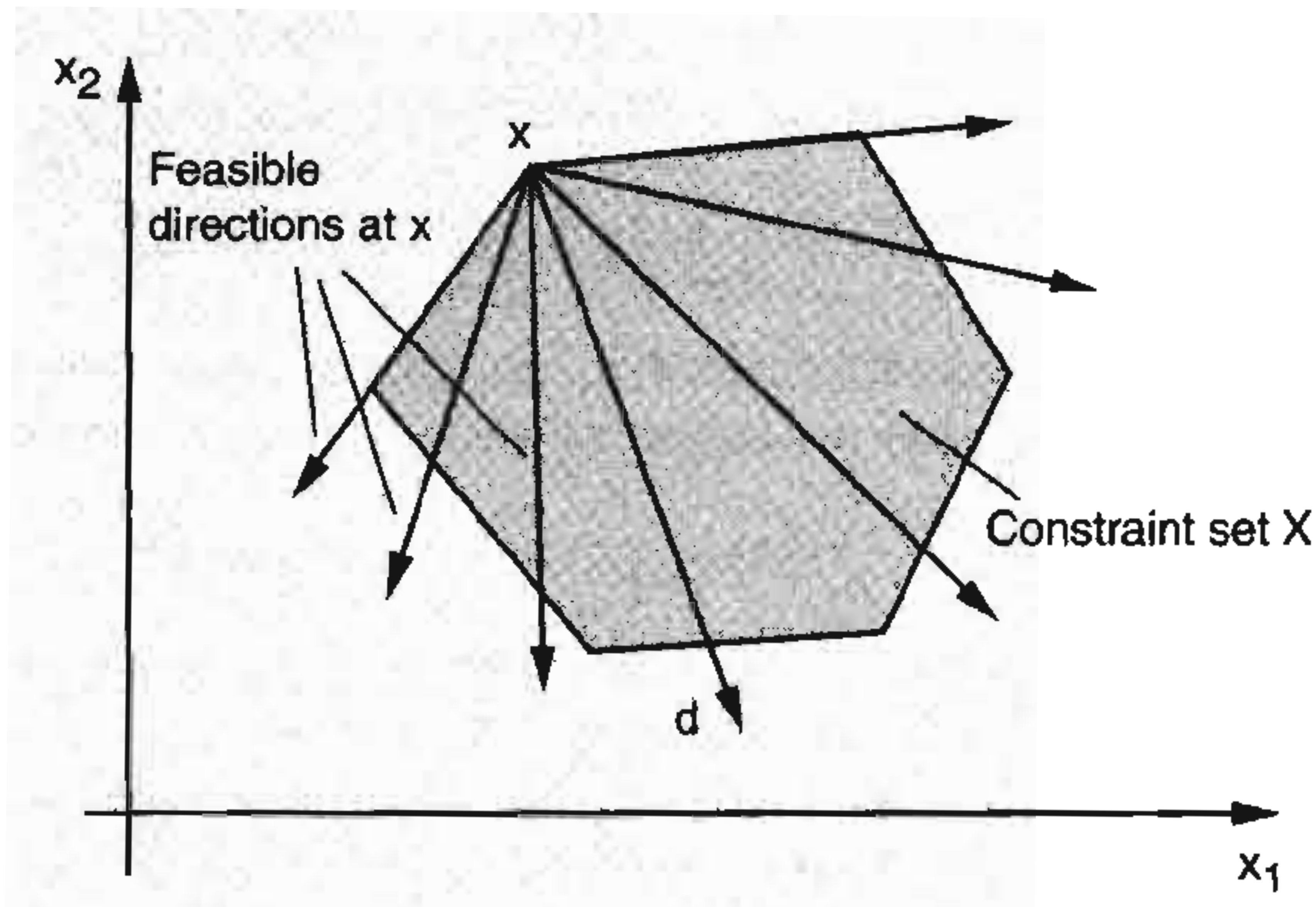
Soit (x_k) la suite des points générés par l'algorithme de descente de gradient avec pas choisi par la règle d'Armijo. Alors, tout point limite (i.e. valeur d'adhérence) de (x_k) est un point stationnaire.

De plus, si x^* est le seul point stationnaire de f dans un ensemble ouvert, il existe un ensemble ouvert S contenant x^* tel que si il existe k_0 tel que $x_{k_0} \in S$, alors $x_k \in S$ pour tout $k \geq k_0$ et $x_k \to x^*$.

Vitesse de convergence ?



Présence de contraintes: Descente de gradient projeté



Présence de contraintes: Descente de gradient projeté

'Backtracking' avec la règle d'Armijo le long de l'arc de projection

Input: $x_0 \in \mathbf{R}^n, f : U \subset \mathbf{R}^n \to \mathbf{R}$ de classe $\mathcal{C}^1, s > 0, 0 < \beta < 1, 0 < \sigma < 1.$

U convexe, fermé, non-vide

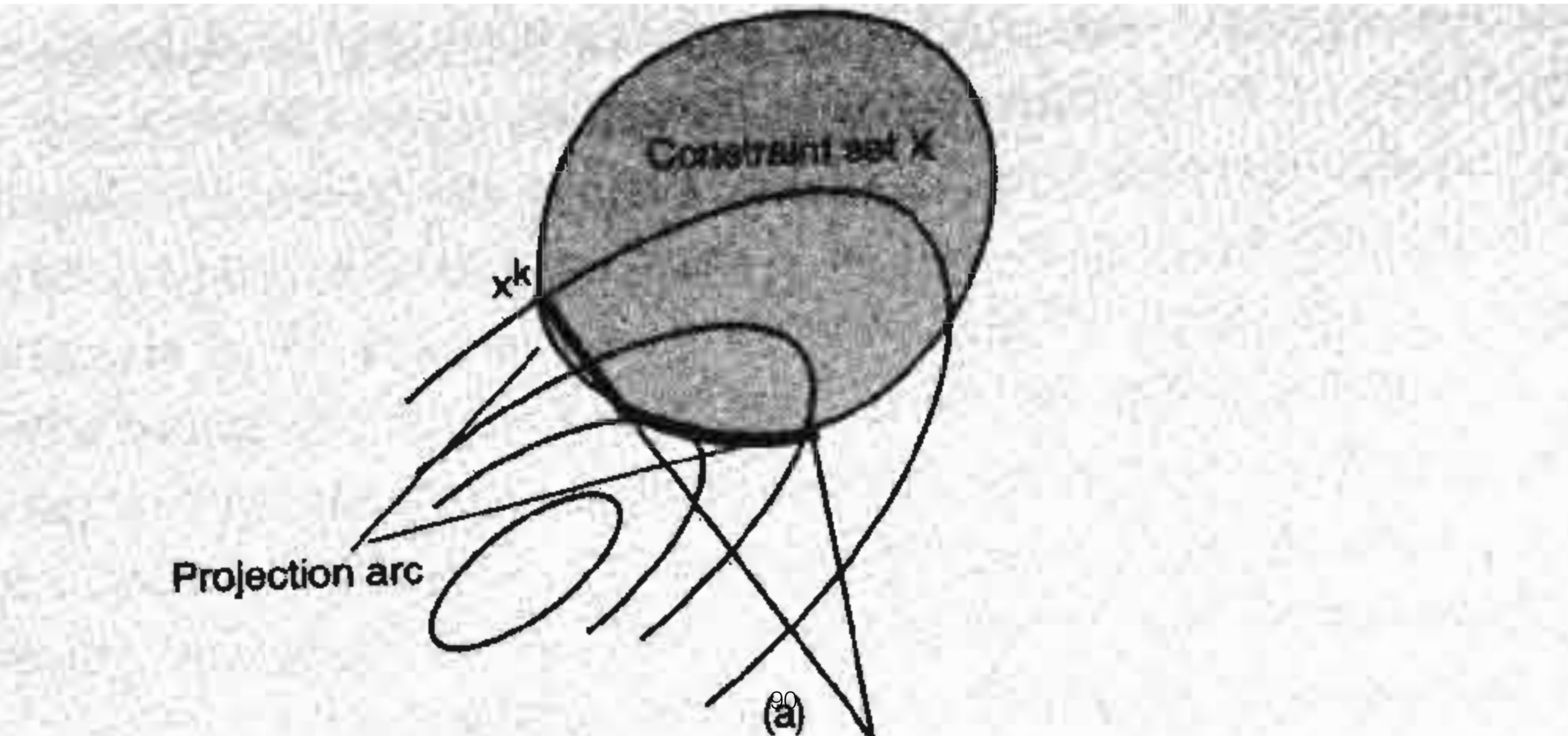
Iteration: $x_{k+1} := p_k(\beta^{m_k} s)$

$p_k(r) = [x_k - r\nabla f(x_k)]_U$ et m_k plus petit entier m tel que

$$f(x_k) - f(x_{k+1}) \geq \sigma \nabla f(x_k)^T (x_k - x_{k+1})$$

Présence de contraintes: Descente de gradient projeté

‘Backtracking’ avec la règle d’Armijo le long de l’arc de projection



Présence de contraintes: Descente de gradient projeté

'Backtracking' avec la règle d'Armijo le long de l'arc de projection

Input: $x_0 \in \mathbf{R}^n, f : U \subset \mathbf{R}^n \to \mathbf{R}$ de classe $\mathcal{C}^1, s > 0, 0 < \beta < 1, 0 < \sigma < 1.$

U convexe, fermé, non-vide

Iteration: $x_{k+1} := p_k(\beta^{m_k} s)$

$p_k(r) = [x_k - r\nabla f(x_k)]_U$ et m_k plus petit entier m tel que

$$f(x_k) - f(x_{k+1}) \geq \sigma \nabla f(x_k)^T (x_k - x_{k+1})$$