Sixième cours

- DM1 en ligne à rendre pour le 6 Octobre (en main propre ou par email: thomas.schatz@univ-amu.fr

- DM2 en ligne mercredi prochain

- Scribes : Rémi Gauchotte et Anas Ouled Sbouria (rendu = fichier .tex)

- Aujourd'hui

    - Fin de l'optimisation début des probabilités

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 $\mathcal{C}^1, s > 0, 0 < \beta < 1, 0 < \sigma < 1.$

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

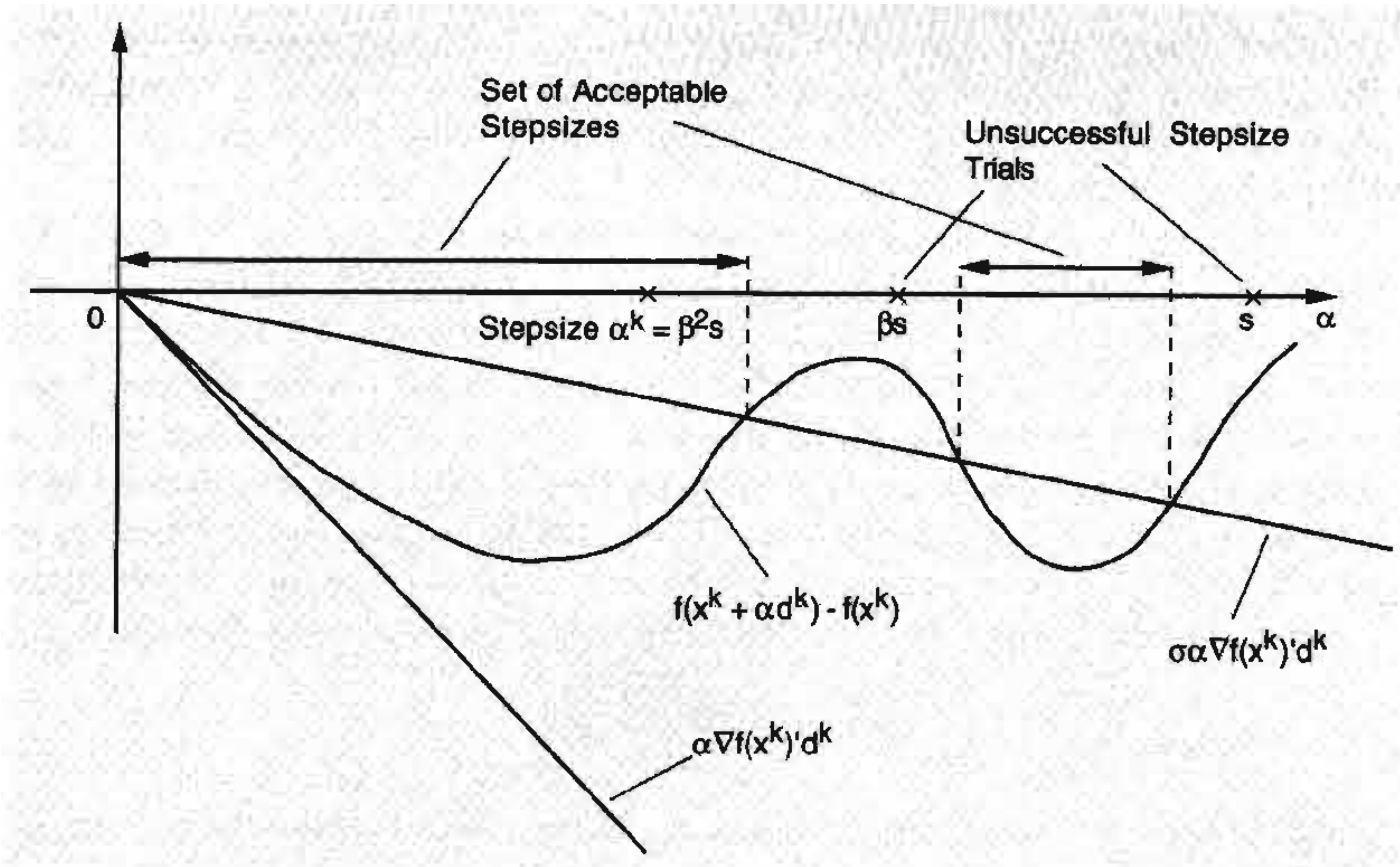
$$\alpha_k = \beta^{m_k} s$$

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

'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

$$f(x_k) - f(x_k - \beta_{m_k} s \nabla f(x_k)) \geq \sigma \beta_{m_k} s \nabla f(x_k)^T \nabla f(x_k)$$

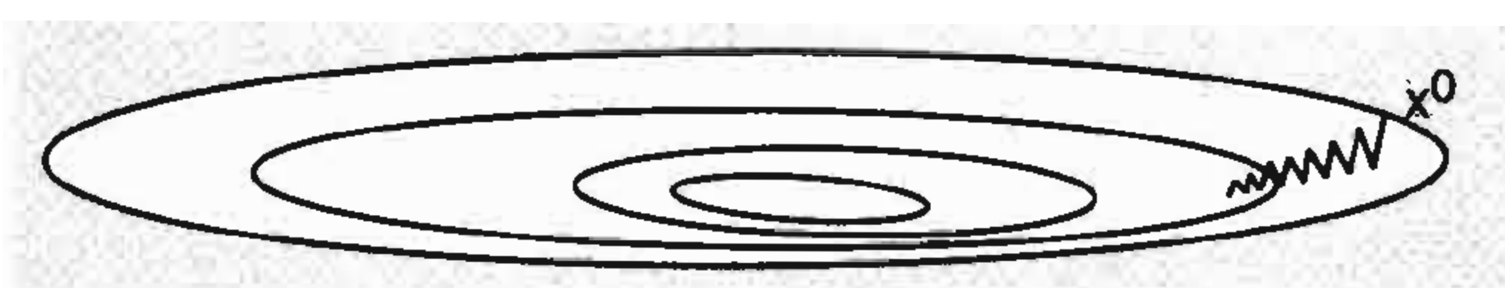
Descente de gradient

Convergence

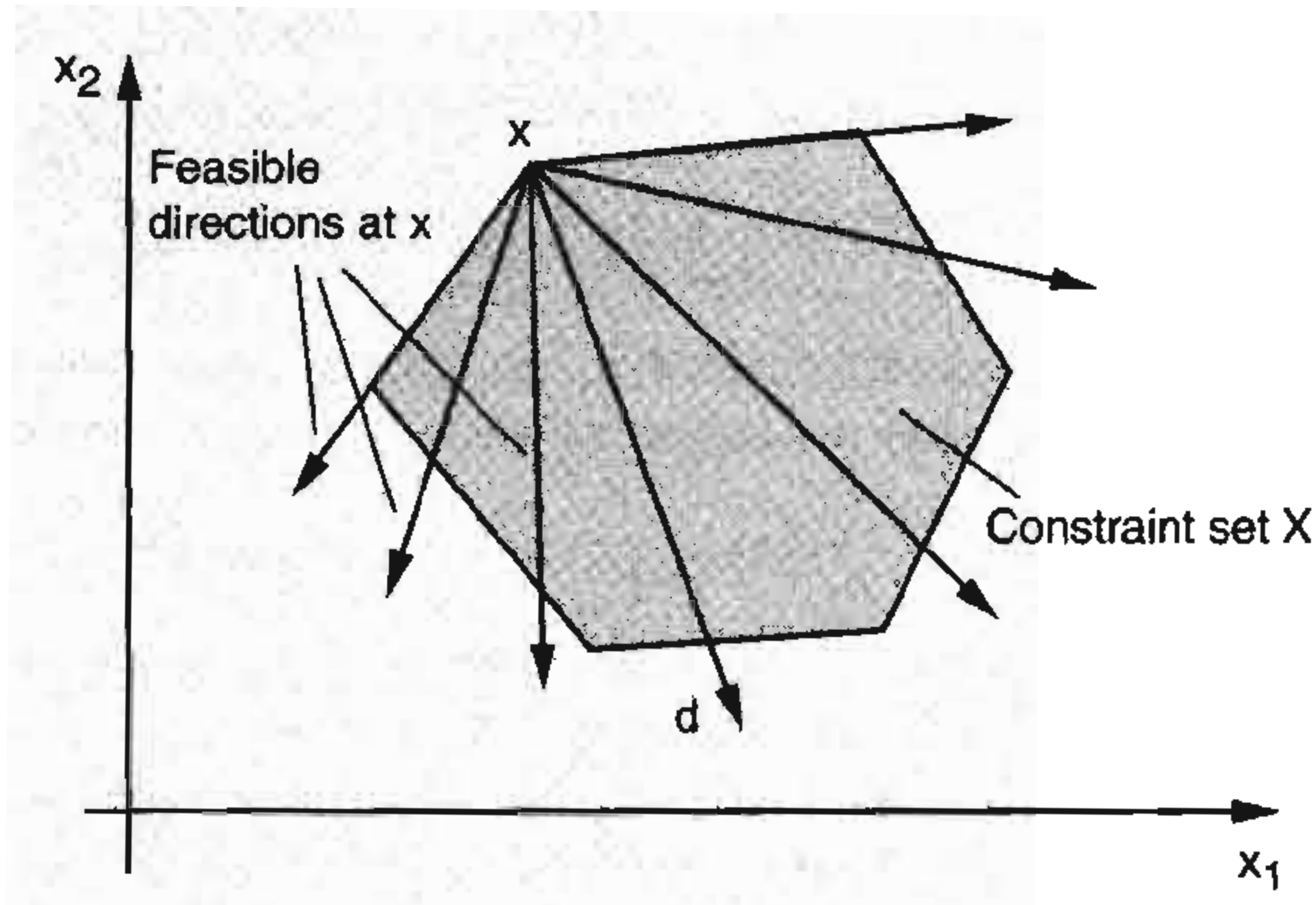
Soit (x_k) la suite des points générés par l'algorithme de descente de gradient avec pas choisi par la règle d'Armijo. Alors, tout point limite (i.e. valeur d'adhérence) de (x_k) est un point stationnaire.

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

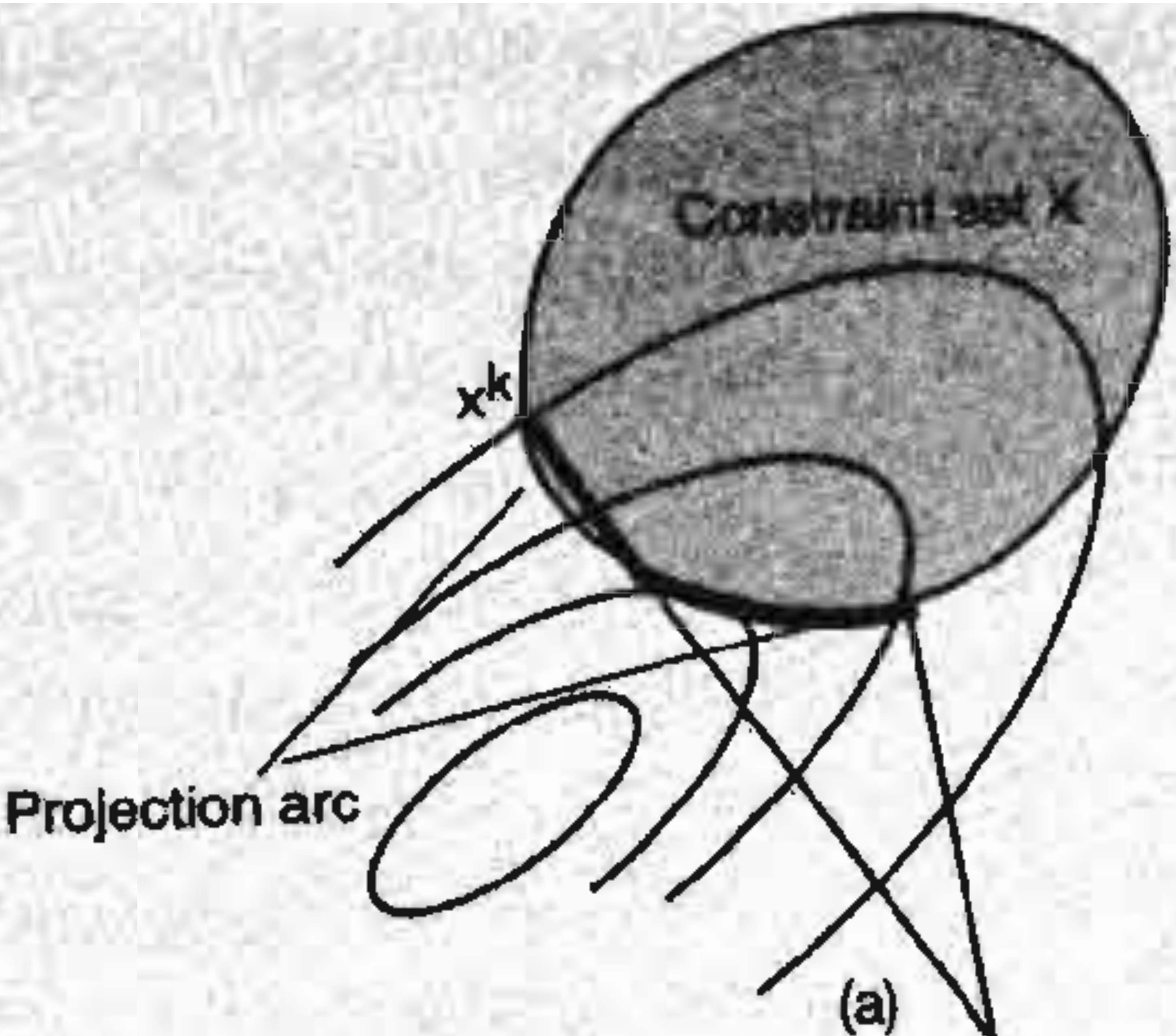
Iteration: $x_{k+1} := p_k(\beta^{m_k} s)$

$p_k(r) = [x_k - r\nabla f(x_k)]_U$ 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

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

Dualité

minimize $f(x)$

subject to $x \in X, \qquad g_j(x) \leq 0, \quad j = 1, \ldots, r,$

$$f : \mathbf{R}^n \to \mathbf{R}$$
$$g_j : \mathbf{R}^n \to \mathbf{R}$$
$$X \subset \mathbf{R}^n$$

$$f^* = \inf_{\substack{x \in X \\ g_j(x) \leq 0,\, j=1,\ldots,r}} f(x).$$

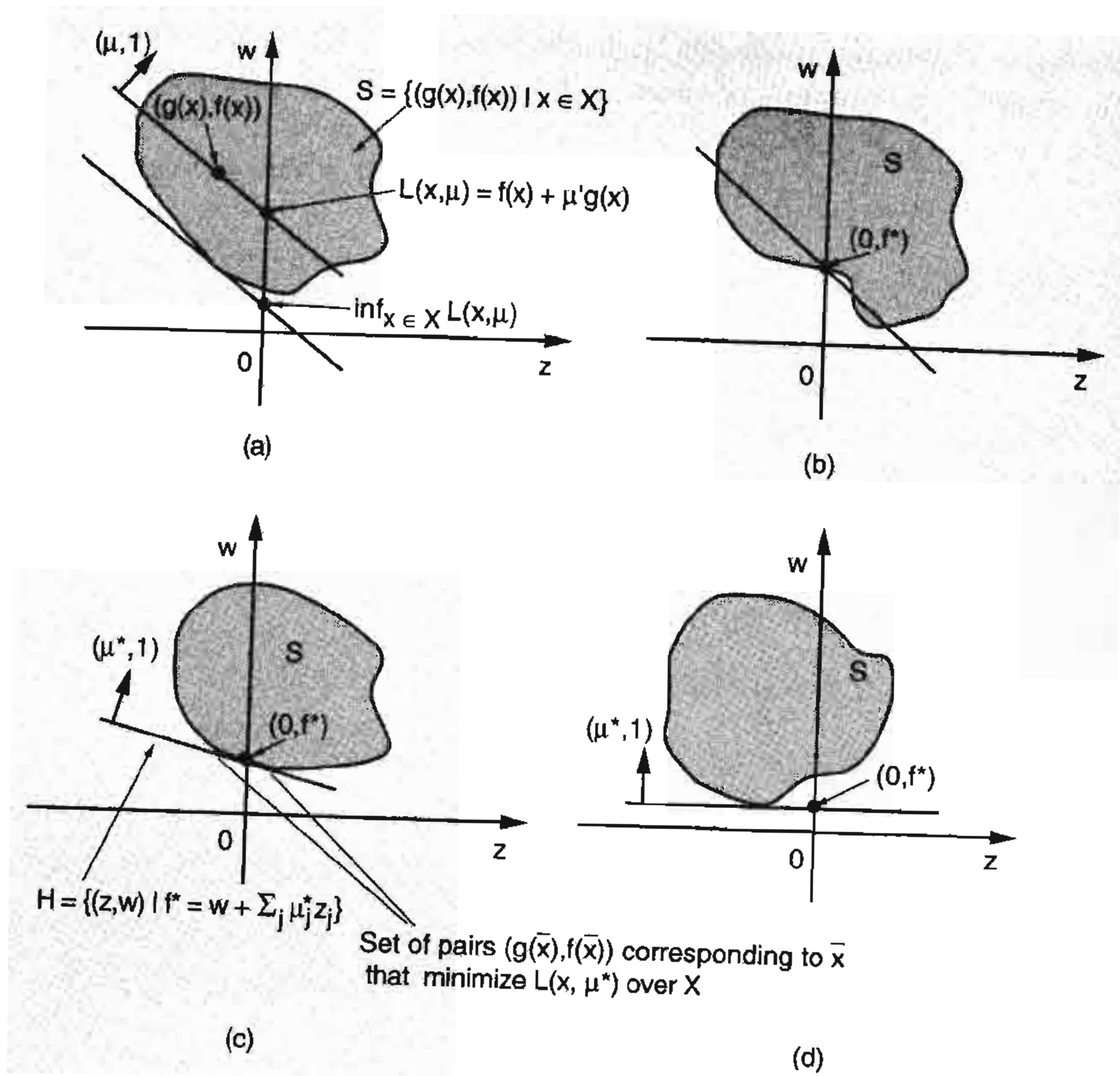
$$-\infty < f^* < +\infty$$

Il existe x^*, tel que $f(x^*) = f^*$

Dualité

$$L(x, \mu) = f(x) + \sum_{j=1}^{r} \mu_j g_j(x) = f(x) + \mu' g(x).$$

Dualité



(μ,1)
w
(g(x),f(x))
S = {(g(x),f(x)) | x ∈ X}
L(x,μ) = f(x) + μ'g(x)
inf$_{x \in X}$ L(x,μ)
0
z
(a)
w
S
(0,f*)
0
z
(b)
w
(μ*,1)
S
(0,f*)
0
z
H = {(z,w) | f* = w + Σ$_j$ μ$_j^*$z$_j$}
Set of pairs (g(x̄),f(x̄)) corresponding to x̄
that minimize L(x, μ*) over X
(c)
w
S
(μ*,1)
(0,f*)
0
z
(d)

Dualité

$$q(\mu) = \inf_{x \in X} L(x, \mu).$$

$$\begin{aligned} \text{maximize} \quad & q(\mu) \\ \text{subject to} \quad & \mu \geq 0, \\ & \mu \in D, \end{aligned}$$

$$D = \{\mu \mid q(\mu) > -\infty\}.$$

Proposition 5.1.2: The domain D of the dual function q is convex and q is concave over D.

Proposition 5.1.3: (Weak Duality Theorem) We have

$$q^* \leq f^*.$$

Dualité

Dualité forte et qualification des contraintes

Example :

Si f est convexe et les contraintes g_j sont linéaires, alors il n'y a pas d'écart dual: $q^* = f^*$

Dualité

Proposition 5.1.5: (Optimality Conditions) (x^*, μ^*) is an optimal solution-Lagrange multiplier pair if and only if

$$x^* \in X, \quad g(x^*) \leq 0, \qquad \text{(Primal Feasibility)}, \tag{5.4}$$

$$\mu^* \geq 0, \qquad \text{(Dual Feasibility)}, \tag{5.5}$$

$$x^* = \arg\min_{x \in X} L(x, \mu^*), \qquad \text{(Lagrangian Optimality)}, \tag{5.6}$$

$$\mu_j^* g_j(x^*) = 0, \quad j = 1, \ldots, r, \qquad \text{(Complementary Slackness)}. \tag{5.7}$$

Crédits

Certaines figures de

- Bertsekas, nonlinear programming
- Wikipedia