Sur quelques métamorphoses de la méthode des centres

Jean-Philippe Vial

ORDECSYS & Université de Genève

Journées en l'honneur de Pierre Huard
Paris

20ème journée JFRO
24-26 novembre 2008

Première manifestation

La méthode des centres

P. Huard (1967)

Problème convexe

$$\lambda^* = \max\{f(x) \mid g_i(x) \geq 0,\ i \in J, x \in B\}$$

- $f(x)$ et g_i sont des fonctions concaves
- $B \subset \mathbb{R}^n$ est un ensemble convexe, fermé, borné
- $\text{int}(A) \cap B \neq \emptyset$ où $A = \{x \mid g_i \geq 0, i \in J\}$
- Un point $x^o \in A \cap B$ avec $\lambda^0 = f(x^o) \leq \lambda^*$

Outils de la méthode

– Tronçon $E(\lambda) = \{x \in A \mid f(x) \geq \lambda\}$

– F-distance $d(x, E(\lambda)) = \begin{cases} \min\{f(x), g_i(x)\ i \in J\} \\ \\ (f(x) - \lambda)^{\alpha} \prod_{i \in J} g_i(x) \\ \\ \alpha \ln(f(x) - \lambda) + \sum_{i \in J} \ln g_i(x) \end{cases}$

– Centre $x(\lambda) = \arg\max\{d(x, E(\lambda)) \mid x \in B\ \}$

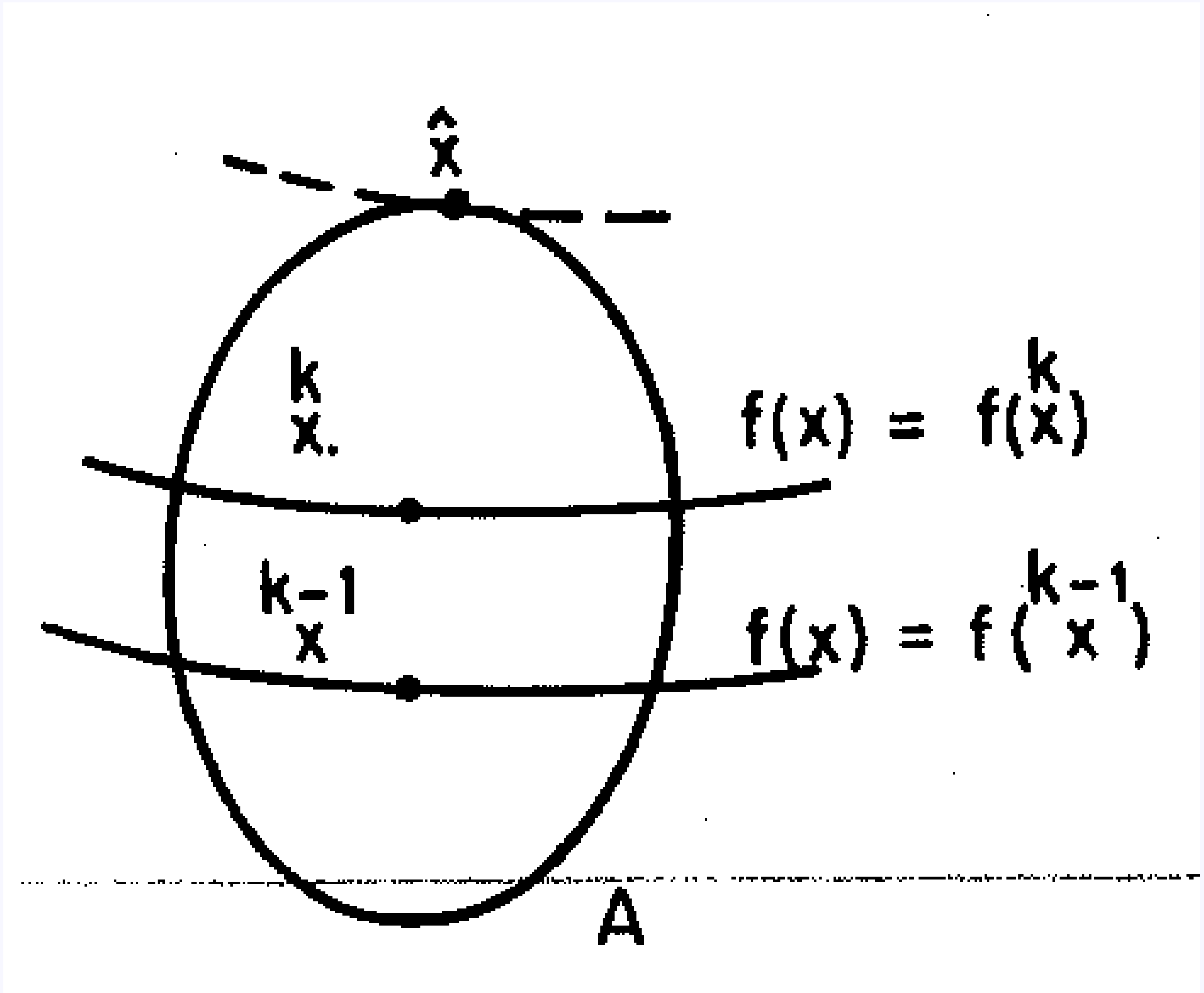
Algorithme théorique

– A l'étape k, $f(x^k) = \lambda^k < \lambda^*$

– Pas de base

- $x^{k+1} = x(\lambda^k) = \arg\max\{d(x, E(\lambda_k)) \mid x \in B\}$
- $\lambda^{k+1} = f(x^{k+1})$

Convergence : Les points d'accumulation de la suite générée appartiennent à l'ensemble des solutions optimales


$\hat{x}$
$\overset{k}{x}$
$\overset{k-1}{x}$
$f(x) = f(\overset{k}{x})$
$f(x) = f(\overset{k-1}{x})$
A

Algorithme ≪ pratique ≫

1. Linéariser

$$f_L(x;y) = f(y) + f'(y)(x-y) \;\; \geq \;\; f(x), \; \forall x$$

$$g_{iL}(x;y) = g_i(y) + g_i'(y)(x-y) \;\; \geq \;\; g_i(x), \; \forall x$$

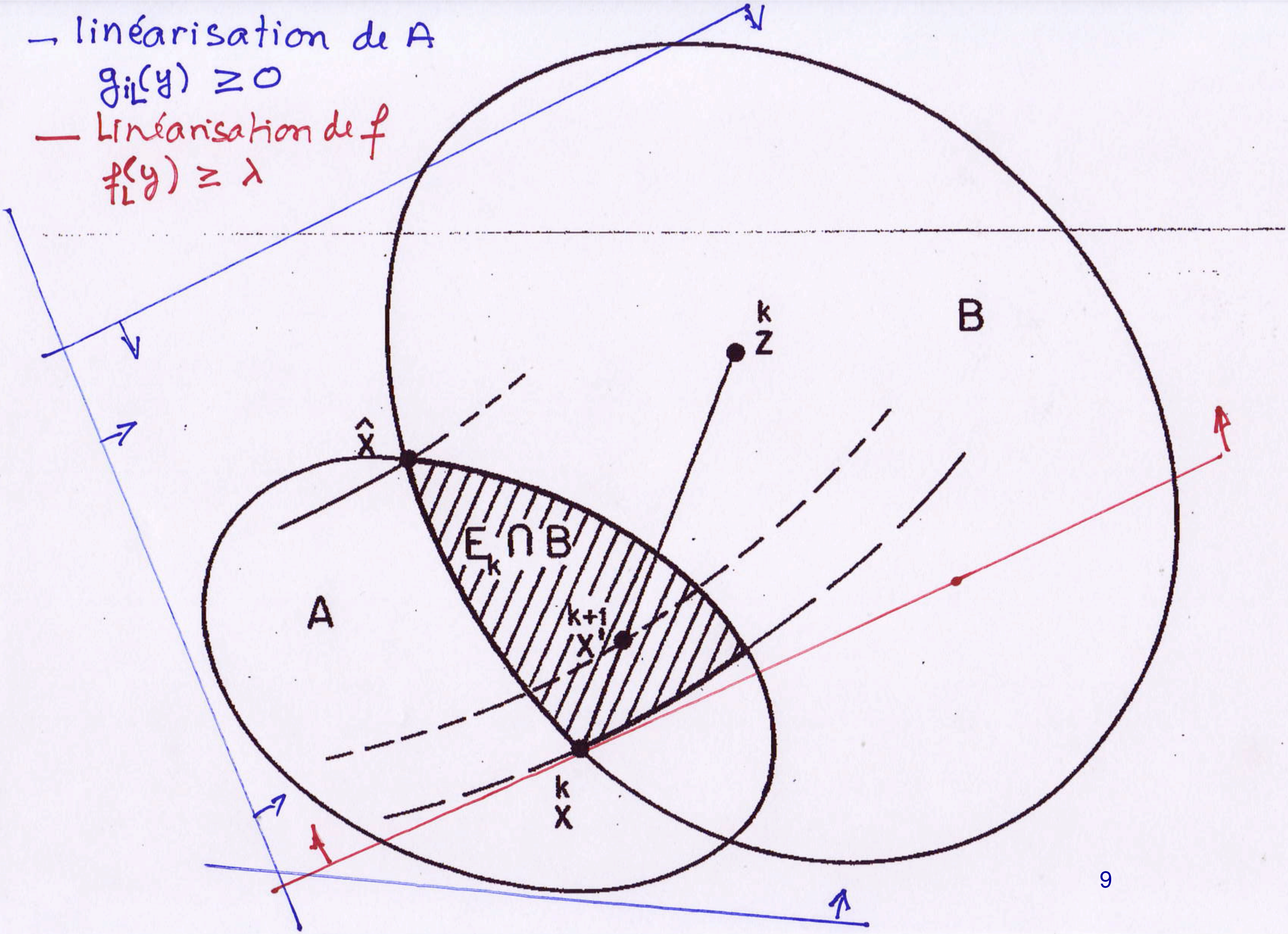
2. Utiliser une F-distance approximante

$$d_L(x, y, E_L(\lambda)) = \min \{ f_L(x;y) - \lambda, g_{iL}(x;y) \; i \in J(\lambda) \}$$

Algorithme de la méthode des centres linéarisés

- **Initialisation** choisir $y^0 \in B$ et faire $k = 0$ et $\lambda_0 = -\infty$
- **Pas de base**

1. $z^k = \arg\max_{x \in B}\ d(x, y^k, E_L(\lambda_k))$
2. $y^{k+1} = y^k + \gamma(z^k - y^k)$ avec
   $\gamma = \arg\max_\delta \{d(x, E(\lambda^k)) \mid x = y^k + \delta(z^k - y^k), 0 < \delta \leq 1\}$
3. $\lambda_{k+1} = f(y^{k+1})$


— linéarisation de A
$g_{iL}(y) \geq 0$
— Linéarisation de f
$f_L(y) \geq \lambda$
B
$\overset{k}{z}$
$\hat{x}$
$E_k \cap B$
A
$\overset{k+1}{x}$
$\overset{k}{x}$

Premier avatar (métamorphose)

Méthode duale de points intérieurs pour la programmation linéaire

Renegar (1988)

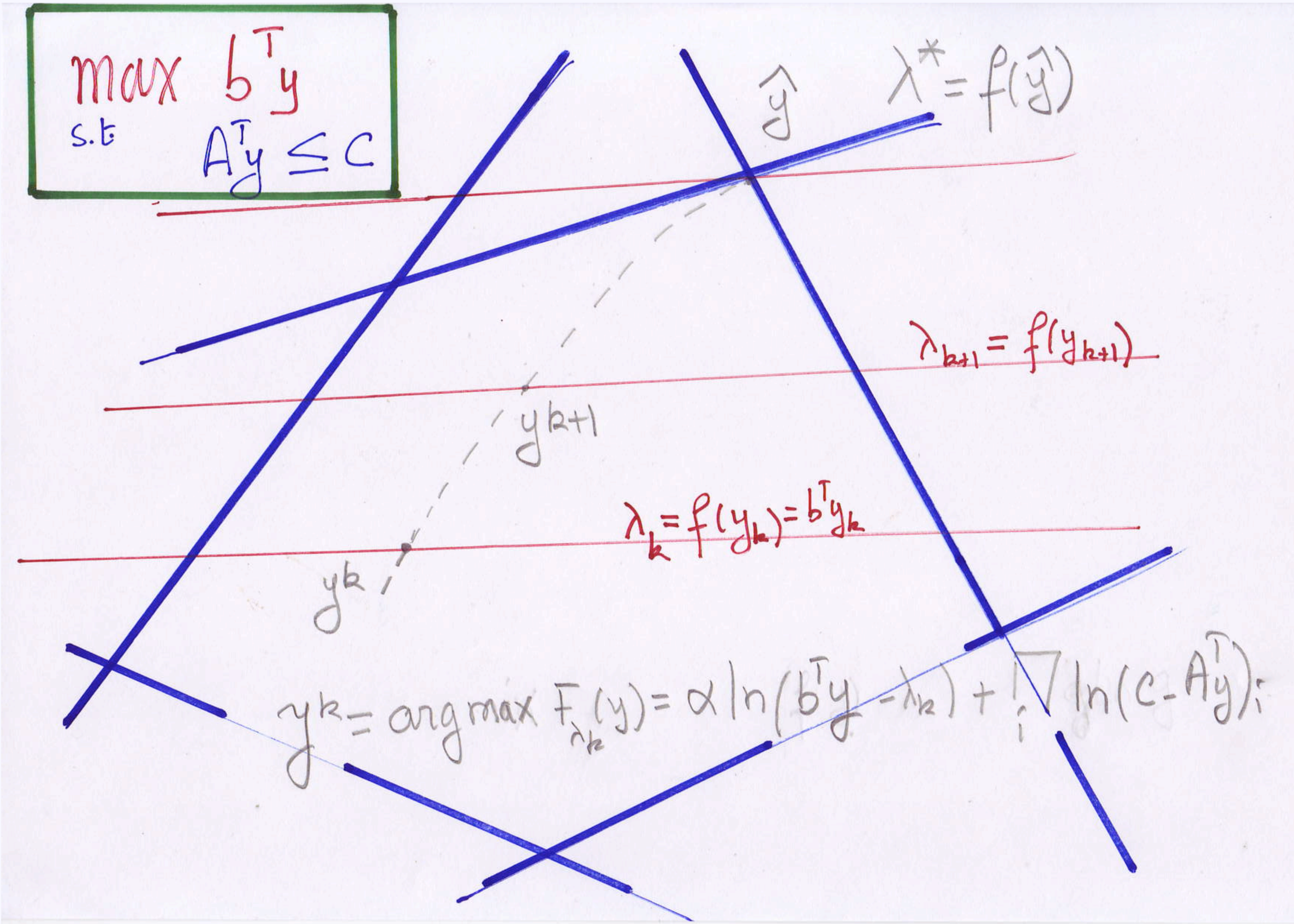
Problème linéaire

$$\lambda^* = \min\{c^T x \mid Ax = b, x \in \mathbb{R}^m_+\}$$
$$\lambda^* = \max\{b^T y \mid A^T y + s = c, s \in \mathbb{R}^n_+\}$$

Potentiel pour le problème dual

$$F_\lambda(y) = \alpha \ln(b^T y - \lambda) + \prod_{i \leq n} (c - A^T y)_i$$

Défini pour $y \in \{y \mid A^T y < c\}$ et $\lambda < \lambda^*$


$\max\ b^T y$
s.t. $A^T y \le c$
$\hat{y}$
$\lambda^* = f(\hat{y})$
$\lambda_{k+1} = f(y_{k+1})$
y^{k+1}
$\lambda_k = f(y_k) = b^T y_k$
y^k
$y^k = \arg\max F_{\lambda_k}(y) = \alpha \ln(b^T y - \lambda_k) + \sum_i \ln(c - A^T y)_i$

Algorithme théorique

- **Initialisation** choisir y^0 tel que $A^T y^0 < c$ et faire $k = 0$ et $\lambda_0 = b^T y^0$. Précision demandée $\epsilon > 0$.
- **Pas de base**
  1. $y^{k+1} = \arg\max F_{\lambda_k}(y)$
  2. $\lambda_{k+1} = b^T y^{k+1}$
  3. Si $\lambda_{k+1} - \lambda_k < \epsilon \frac{1}{1+n/\alpha}$
- **Convergence** Le nombre d'itérations est borné par

$$k^* = \left\lceil \frac{n}{\alpha} \ln \frac{\lambda^* - \lambda^0}{\epsilon} \right\rceil$$

Propriété du centre analytique

Gradient du potentiel :

$F'_\lambda(y) = \alpha s_0^{-1} b - A s^{-1}$ avec $s_0 = b^T y - \lambda$ et $s = c - A^T y > 0.$

Centre analytique $y_\lambda = \arg\max F_\lambda(y)$ défini par $F'_{y_\lambda = 0}$.

Soit

$$xs = \frac{s_0}{\alpha} e$$

qui définit un point intérieur $x > 0$ primal réalisable $Ax = b$.

(e, vecteur avec composantes égales à 1.)

Calcul de la borne sur le nombre d'itérations

$$\lambda^* - b^T y \leq c^T x - b^T y = x^T s = \frac{n}{\alpha}(b^T y - \lambda) \qquad (1)$$

$$\lambda^* - \lambda = (\lambda^* - b^T y) + (b^T y - \lambda) \leq (b^T - \lambda)(1 + \frac{n}{\alpha}) \qquad (2)$$

Conclusion 1 :

$$b^T y - \lambda \leq \frac{\alpha}{n}\epsilon \quad \Rightarrow \lambda^* - b^T y \leq \epsilon$$

Conclusion 2 : Si $\lambda^+ = b^T y_\lambda$

$$\begin{aligned}\lambda^* - \lambda^+ &= (\lambda^* - \lambda) + (\lambda - b^T y_\lambda) \leq (\lambda^* - \lambda)(\frac{1}{1 + \alpha/n}) \\ \lambda^* - \lambda^k &\leq (\lambda^* - \lambda^0)(\frac{1}{1 + \alpha/n})^k\end{aligned}$$

The “missing link”

La borne sur le nombre d’itération suppose un calcul exact des centres analytiques y_λ.

– Peut-on borner le nombre d’itérations avec des centres approchés ?

– Quel effort de calcul pour obtenir un centre analytique approché ?

Le jury répond OUI aux deux questions, grâce aux bonnes propriétés du potentiel F vis-à-vis de la méthode de Newton. Mais il faut le payer par un $\alpha = \sqrt{n}$ pas trop grand, ce qui donne la borne

$$k^* = O(\sqrt{n} \ln \frac{\lambda^* - \lambda^0}{\epsilon})$$

Une affaire de voisinage au pays de Newton

$$N_\lambda(y,\eta) = \{y \mid ||F'_\lambda(y)||_{F''_\lambda(y)} \leq \eta\}$$

Théorème 1 *Pour tout* $0 < \eta \leq \bar{\eta} < 1$ *tel que* $||F'_\lambda(y)||_{F''_\lambda(y)} \leq \eta$

1. $y_\lambda \in N_\lambda(y,\eta)$
2. *Le pas de Newton* $y^+ = y + [F''_\lambda(y)]^{-1}F'_\lambda(y)$ *garantit* $||F'_\lambda(y^+)||_{F''_\lambda(y^+)} \leq \eta^2$.

Théorème 2 *Soit* $\alpha = \sqrt{n}$ *dans la définition du potentiel.*
Si $||F'_\lambda(y)||_{F''_\lambda(y)} \leq \eta^2$ *et* $\lambda^+ = \lambda + \theta(b^T y - \lambda)$ *pour* $0 < \theta < 1$ *alors* $||F'_{\lambda^+}(y)||_{F''_{\lambda^+}(y)} \leq \eta$.

La méthode des petits pas

Hypothèse : on dispose d'un point initial qui se trouve dans un voisinage η^2 du point y_λ de la trajectoire des centres.

Iteration de base

1. Changer λ en λ^+ suivant la règle du théorème 2. Le point courant se trouve dans un voisinage η du point y_{λ^+}.
2. Faire un pas de Newton comme décrit dans le théorème 1. Le nouveau point y^+ se trouve dans un voisinage η^2 du point y_{λ^+} de la trajectoire des centres.

Extension à une classe de problèmes non-linéaires

Nesterov et Nemirovski (1994) ont introduit la classe de fonctions convexes non-linéaires, dites fonctions auto-concordantes, qui ont les mêmes propriétés (théorème 1) vis-à-vis de la méthode de Newton.

Cela permet de donner des bornes sur la minimisation de fonctions convexes auto-corcordantes et sur la résolution de programmes linéaires sur les cônes du second-ordre et des matrices semi-définies positives.

Second avatar

Méthodes de plans sécants pour la programmation convexe non-différentiable

	Centre analytique	Centre de Tchebichev
proposition	Goffin-Haurie-Vial (1992)	Elzinga-Moore (1975)
convergence	Nesterov (1995) Goffin-Luo-Ye (1996) Nesterov-Vial (1999)	Ouorou (2008)
	borne sur # d'itérations	point d'accumulation

Du différentiable au non différentiable

Les fonctions linéaires $f_L(y,z)-\lambda$ et $g_{iL}(y,z), \forall i \in J$ introduites dans la méthode des centres linéarisée définissent des supports globaux pour E :

$$E \subset \{y \mid f_L(y,z)-\lambda \geq 0, g_{iL}(y,z), \forall i \in J\}$$

La propriété reste vraie si on utilise un sous-différentiel lorsque les fonctions sont simplement convexes (peut-être non différentiables).

En revanche, le pas de la méthode des centres linéarisée

$$y^{k+1} = y^k + \gamma(z^k - y^k)$$

avec

$$\gamma = \arg\max_\delta \{d(x, E(\lambda^k)) \mid x = y^k + \delta(z^k - y^k), 0 < \delta \leq 1\}$$

n'est plus garanti car la direction $(z^k - y^k)$ n'est plus nécessairement une direction de descente $\Rightarrow \delta = 0$.

Méthode des centres par plans sécants

- **Initialisation** Choisir $y^0 \in B$. Faire $k = 0$, $\lambda_0 = -\infty$ et $H_0 = \mathbb{R}^n$
- **Pas de base**

1. Cas $z^k \notin A$, c.-à.d. $g_i < 0$ for some $i \in J$
   $h_k = \{y \mid g_i(z^k) + g_i'(z^k)(y - z^k) \geq 0\}$ (plan strictement sécant)
   $H_{k+1} = H_k \cap h_k$ et $\lambda_{k+1} = \lambda_k$
2. Cas $z^k \in A$, c.-à.d. $g_i \geq 0$ for all $i \in J$
   (a) Cas $f(z^k) \leq \lambda_k$
   $h_k = \{y \mid f(z^k) - \lambda_k + f'(z^k)(y - z^k) \geq 0\}$ (plan sécant)
   $H_{k+1} = H_k \cap h_k$ et $\lambda_{k+1} = \lambda_k$
   (b) Cas $f(z^k) > \lambda_k$
   $H_{k+1} = H_k$ et $\lambda_{k+1} = f(z^k)$

Convergence : méthode des centres analytiques, version homogène

Problème	Hypothèse	Convergence
	Problème de réalisabilité	
Trouver $y \in Y^*$	Y^* convexe $B(\bar{y}, \varepsilon) \subset Y^* \subset B(0, R)$	$k \leq \frac{(1+R^2)^2}{\varepsilon^2} M$
	Problème d'inégalité variationnelle	
Trouver $y^* \in Q$ tel que $h_y^T(y - y^*) \geq -\epsilon$ $\forall y \in Q,$ and all $h_y \in H(y)$	Q convexe avec s.-c. barrière $H(y)$ monotone sur Q $\|\|h_y\|\| \leq L$ sur Q	$\frac{k}{\sqrt{k+\nu}} \leq \frac{L(1+R^2)}{\epsilon\theta_3} e^{\theta_2 \sqrt{\nu}}$
	Problème d'optimisation convexe sous contrainte	
$\min\{f(y) \mid y \in Q\}$	Q convexe avec s.-c. barrière $Q \subset \{y \mid \|\|y\|\| \leq R\}$ f convexe et $\|\|\partial f\|\| \leq L$	$f(\hat{y}) - f^* \leq$ $\frac{L}{\sqrt{k+\nu}} \left[\sqrt{\nu} + \frac{e}{\theta_3}\left(1 + \frac{\nu}{k}\right)\right] [1 + R^2]^{1+\frac{\nu}{2k}}$

Quelques exemples récents d'applications

[1] BABONNEAU, F., ET VIAL, J.-P. Proximal-ACCPM with a nonlinear constraint and an active set strategy to solve nonlinear multicommodity flow problems. (2006). A paraître dans *Mathematical Programming*.

[2] BABONNEAU, F., ET VIAL, J.-P. Traffic equilibrium with elastic demands. Working paper (2006). A paraître dans in *Transportation Science*.

[3] L. DROUET, A. HAURIE, J.-P. VIAL ET M. VIELLE. A coupled game solved with the homogeneous version of OBOE to model Post Kyoto international climate policy. Document de travail ORDECSYS (2008). Soumis à *Annals of Operations Research*

Les sources de la suite OBOE (dernier avatar de la méthode ACCPM) sont disponibles sur le site de COIN-OR ou depuis
`http ://www.ordecsys.com/fr/produits`