

Algorithme d'optimisation par pénalisation mixte: Lagrangien augmenté et barrière logarithmique

Paul Armand
collaboration avec Riadh Omheni

Université de Limoges
Laboratoire XLIM

Séminaire du LARIS
Université d'Angers
6 janvier 2015

Plan

- 1 Introduction et motivations
- 2 Contraintes d'égalité
- 3 Contraintes générales

Plan

1 Introduction et motivations

2 Contraintes d'égalité

3 Contraintes générales

Problème d'optimisation non linéaire

- On veut résoudre numériquement un problème de la forme

$$\begin{array}{ll} \min & f(x) \\ \text{s.c.} & c_E(x) = 0 \\ & c_I(x) \leq 0 \end{array}$$

où $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $c_E : \mathbb{R}^n \rightarrow \mathbb{R}^m$ et $c_I : \mathbb{R}^n \rightarrow \mathbb{R}^p$ sont différentiables.

- En introduisant des variables d'écart $s \in \mathbb{R}^p$, le problème se reformule sous la forme d'un problème avec contraintes d'égalité et de borne :

$$\begin{array}{ll} \min & f(x) \\ \text{s.c.} & c_E(x) = 0 \\ & c_I(x) + s = 0 \\ & s \geq 0 \end{array}$$

Conditions d'optimalité (KKT)

- Pour simplifier la présentation, on considère un problème sous la forme

$$\begin{array}{ll} \min & f(x) \\ \text{s.c.} & c(x) = 0 \\ & x \geq 0 \end{array}$$

où $f : \mathbb{R}^n \rightarrow \mathbb{R}$ et $c : \mathbb{R}^n \rightarrow \mathbb{R}^m$.

- Conditions de Karush-Kuhn-Tucker (KKT):

$$\begin{array}{rcl} \nabla_x \mathcal{L}(x, y, z) & = & 0 \\ c(x) & = & 0 \\ 0 \leq x \perp z & \geq & 0 \end{array}$$

où $\mathcal{L}(x, y, z) := f(x) + y^\top c(x) - z^\top x$ est le Lagrangien du problème.

- x : variable primale (y, z) : variables duales

Méthodes numériques standards

- Les algorithmes standards consistent à appliquer une méthode de type Newton sur les conditions de KKT ou une suite de KKT perturbées.
 - **Programmation Quadratique Successive (SQP)**
 - **Points Intérieurs (IP)** ou **Primal-Dual**

~ Linéarisations successives du système KKT ou d'une suite de systèmes approchés.
- Ces deux approches se déclinent selon plusieurs variantes
 - La solution du système linéaire peut être
 - **directe** (factorisation de Cholesky, LDL^T , ...)
 - **itérative** (gradient conjugué, GMRES, ...)
 - La technique de globalisation peut être par
 - **recherche linéaire** (SNOPT)
 - **région de confiance** (KNITRO),
 - **filtre** (IPOPT, FILTER).
- Convergence rapide : superlinéaire ou quadratique.

Lagrangien augmenté

- Lagrangien augmenté pour un problème avec contraintes d'égalité:
$$\mathcal{L}_\sigma(x, \lambda) = f(x) + \lambda^\top c(x) + \frac{1}{2\sigma} \|c(x)\|^2$$
- Propriété : $x \mapsto \mathcal{L}_\sigma(x, \lambda^*)$ est une pénalisation exacte pour $0 < \sigma \leq \bar{\sigma}$.
- Algorithme : suite de minimisations approchées de $\mathcal{L}_\sigma(\cdot, \lambda)$ avec mise à jour des paramètres λ et σ .
- $\sigma \not\rightarrow 0$, évite le mauvais conditionnement.
- Convergence lente (linéaire)
- Robuste, peu sensible à une jacobienne des contraintes dégénérée.
- Codes : LANCELOT¹, ALGENCAN²

¹Conn, Gould Toint

²Birgin, Martínez

Pourquoi combiner un Lagrangien augmenté et une barrière logarithmique dans un algorithme primal-dual ?

- Pour sa propriété de régularisation de la Jacobienne du système de Newton pour la résolution des conditions de KKT et obtenir ainsi une méthode robuste ayant de bonnes propriétés de convergence globale.
- Pour s'affranchir de la notion de trajectoire pour un algorithme primal-dual destiné à la résolution d'un problème avec contrainte d'égalité et obtenir ainsi une vitesse de convergence quadratique.
- L'idée est de concevoir un algorithme qui se réduit localement à une méthode de Newton régularisée appliquée sur le système de KKT (perturbé) associé au problème initial.

Regularization Theorem ³

Let $\{M_k\}$ be a sequence of $n \times n$ real symmetric matrices, $\{A_k\}$ be a sequence of $m \times n$ real matrices, and $\{\delta_k\}$ be a sequence of positive numbers. Let $\{x_k\}$ and $\{z_k\}$ be two sequences of \mathbb{R}^n with positive components. Define for all $k \in \mathbb{N}$ the matrices $X_k = \text{diag}(x_k)$, $Z_k = \text{diag}(z_k)$ and

$$J_k = \begin{pmatrix} M_k & A_k^\top & -I \\ A_k & -\delta_k I & 0 \\ Z_k & 0 & X_k \end{pmatrix}.$$

Assume that the following properties are satisfied.

- (i) The sequences $\{M_k\}$ and $\{A_k\}$ are bounded.
- (ii) The sequence $\{\delta_k\}$ is bounded away from zero.
- (iii) There exists $\mu > 0$ such that for all $k \in \mathbb{N}$ and all $i \in \{1, \dots, n\}$, $[x_k]_i [z_k]_i \geq \mu$.
- (iv) There exists $\lambda > 0$ such that for all $k \in \mathbb{N}$ and all $d \in \mathbb{R}^n$, $d^\top H_k d \geq \lambda \|d\|^2$, where $H_k = M_k + X_k^{-1} Z_k + \frac{1}{\delta_k} A_k^\top A_k$.

Then the sequence $\{J_k^{-1}\}$ is well defined and bounded.

³Armand-Benoist, Math Prog 2013

Cadre primal-dual pour les contraintes d'égalité

- Pénalisation quadratique ($\sigma > 0$)

$$\min_{x \in \mathbb{R}^n} f(x) + \frac{1}{2\sigma} \|c(x)\|^2$$

- CN1 : $\nabla f(x) + \frac{1}{\sigma} \nabla c(x) c(x) = 0$
- Substitution $y = \frac{1}{\sigma} c(x)$:

$$\begin{aligned} \nabla f(x) + \nabla c(x) y &= 0 \\ c(x) &= \sigma y \end{aligned}$$

- Linéarisation :

$$\begin{pmatrix} \nabla_{xx}^2 \mathcal{L}(x, y) & \nabla c(x) \\ \nabla c(x)^\top & -\sigma I \end{pmatrix} \begin{pmatrix} d_x \\ d_y \end{pmatrix} = - \begin{pmatrix} \nabla f(x) + \nabla c(x) y \\ c(x) - \sigma y \end{pmatrix}$$

SPDOPT-QP⁴

- **SPDOPT-QP**: Strongly Primal-Dual OPTimization - Quadratic Penalty approach.
- Méthode de suivi de trajectoire.
- Globalisation par recherche linéaire avec une fonction de mérite primale-duale:

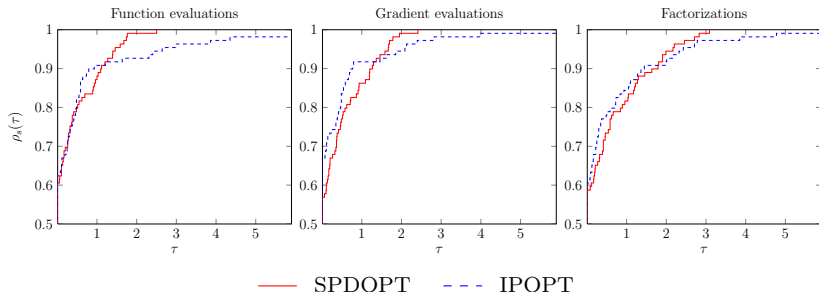
$$f(x) + \frac{1}{2\sigma} \|c(x)\|^2 + \frac{\nu}{2\sigma} \|c(x) - \sigma y\|^2.$$

- Convergence globale et superlinéaire.
- Efficace pour résoudre des problèmes avec jacobienne singulière.

⁴Armand, Benoist, Omheni, Pateloup: Study of a primal-dual algorithm for equality constrained minimization, COAP 2014.

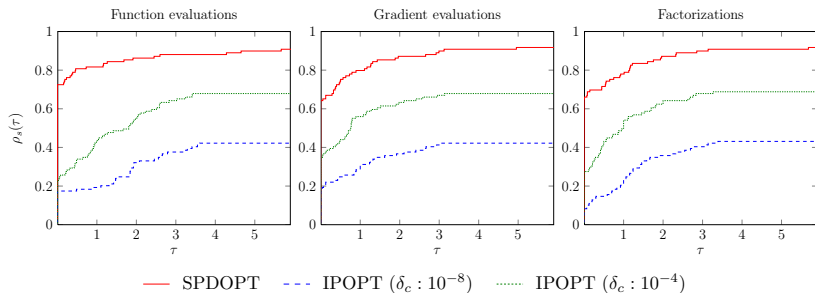
SPDOPT-QP versus IPOPT

- Performances de SPDOPT-QP et IPOPT sur 109 problèmes avec contraintes d'égalité.



SPDOPT-QP versus IPOPT

- Performances de SPDOPT-QP et IPOPT sur 109 problèmes dégénérés avec contraintes d'égalité.
- La première contrainte de chaque problème est dupliquée $c_1 = c_1^2$.



Extension au cas général: $\min f(x)$, s.c. $c(x) = 0, x \geq 0$

- Pénalisation quadratique + barrière logarithmique

$$f(x) + \frac{1}{2\sigma} \|c(x)\|^2 - \frac{1}{\mu} \sum_{i=1}^n \log x_i$$

- Linéarisation des CN1:

$$\begin{pmatrix} \nabla_{xx}^2 \mathcal{L}(x, y) & \nabla c(x) & -I \\ \nabla c(x)^\top & -\sigma I & 0 \\ Z & 0 & X \end{pmatrix} \begin{pmatrix} d_x \\ d_y \\ d_z \end{pmatrix} = - \begin{pmatrix} \nabla f(x) + \nabla c(x)y \\ c(x) - \sigma y \\ XZe - \mu e \end{pmatrix}$$

- Conception de l'algorithme délicate.
- Analyse de convergence compliquée.

Travaux antérieurs

- L'idée d'utiliser une pénalisation mixte est assez ancienne [Fiacco-McCormick – 1968], mais abandonnée à cause des problèmes de mauvais conditionnement.
- Méthode reconsidérée sous l'angle “primal-dual” :
 - Benchakroun, Dussault and Mansouri 1995
 - Forsgren and Gill – SiOpt 1998
 - Gertz and Gill – Math Prog 2004
 - Yamashita and Yabe – SiOpt 2003
 - Chen and Goldfarb – Math Prog 2006 – Math Prog 2009
 - Gill and Robinson – COAP 2010, SiOpt 2013

Plan

- 1 Introduction et motivations
- 2 Contraintes d'égalité**
- 3 Contraintes générales

Lagrangien augmenté : version primal-dual

- Lagrangien augmenté ($\lambda \in \mathbb{R}^m$, $\sigma > 0$)

$$\min_{x \in \mathbb{R}^n} f(x) + \lambda^\top c(x) + \frac{1}{2\sigma} \|c(x)\|^2$$

- CN1 : $\nabla f(x) + \nabla c(x)(\lambda + \frac{1}{\sigma}c(x)) = 0$
- Substitution $y = \lambda + \frac{1}{\sigma}c(x)$:

$$\begin{aligned} \nabla f(x) + \nabla c(x)y &= 0 \\ c(x) + \sigma(\lambda - y) &= 0 \end{aligned}$$

- Linéarisation :

$$\begin{pmatrix} \nabla_{xx}^2 \mathcal{L}(x, y) & \nabla c(x) \\ \nabla c(x)^\top & -\sigma I \end{pmatrix} \begin{pmatrix} d_x \\ d_y \end{pmatrix} = - \begin{pmatrix} \nabla f(x) + \nabla c(x)y \\ c(x) + \sigma(\lambda - y) \end{pmatrix}$$

Lagrangien augmenté : version primal-dual

- Lagrangien augmenté ($\lambda \in \mathbb{R}^m$, $\sigma > 0$)

$$\min_{x \in \mathbb{R}^n} f(x) + \lambda^\top c(x) + \frac{1}{2\sigma} \|c(x)\|^2$$

- CN1 : $\nabla f(x) + \nabla c(x)(\lambda + \frac{1}{\sigma}c(x)) = 0$
- Substitution $y = \lambda + \frac{1}{\sigma}c(x)$:

$$\begin{aligned} \nabla f(x) + \nabla c(x)y &= 0 \\ c(x) + \sigma(\lambda - y) &= 0 \end{aligned}$$

- Linéarisation : si $\lambda = y$

$$\begin{pmatrix} \nabla_{xx}^2 \mathcal{L}(x, y) & \nabla c(x) \\ \nabla c(x)^\top & -\sigma I \end{pmatrix} \begin{pmatrix} d_x \\ d_y \end{pmatrix} = - \begin{pmatrix} \nabla f(x) + \nabla c(x)y \\ c(x) \end{pmatrix}$$

Newton régularisé sur le système de KKT initial.

Algorithme - Itération externe

$$w := (x, y) \quad \Phi(w, \lambda, \sigma) = \begin{pmatrix} \nabla f(x) + \nabla c(x)y \\ c(x) + \sigma(\lambda - y) \end{pmatrix} \quad a \in (0, 1)$$

- 1 Choisir $\zeta_k \geq 0$, $r_k > 0$ t.q. $\{\zeta_k\} \rightarrow 0$ et $\{r_k\} \rightarrow 0$. Poser $\eta_k = \|c(x_k)\| + \zeta_k$. Si $\|c(x_k)\| \leq a \max\{\eta_{ij} : (k - \ell)^+ \leq j \leq k\}$, aller en 2, sinon aller en 3.
- 2 Choisir $\sigma_k^+ \leq \sigma_k$. Poser $s_k = \max\{\sigma_k^+, r_k\}$, $\lambda_{k+1} = y_k$, $i_{k+1} = k$ et aller en 4.
- 3 Choisir $\sigma_k^+ \leq \min\{\tau\sigma_k, r_k\}$. Poser $s_k = r_k$, $\lambda_{k+1} = \lambda_k$ et $i_{k+1} = i_k$.
- 4 Choisir H_k t.q. $\ln(J_k) = (n, m, 0)$, où $J_k = \begin{pmatrix} H_k & \nabla c_k \\ \nabla c_k^\top & -\sigma_k^+ I \end{pmatrix}$. Résoudre

$$J_k (w_k^+ - w_k) = -\Phi(w_k, \lambda_{k+1}, \sigma_k^+).$$

- 5 Choisir $\epsilon_k > 0$ t.q. $\{\epsilon_k\} \rightarrow 0$. Si $\|\Phi(w_k^+, \lambda_{k+1}, \sigma_{k+1})\| \leq \epsilon_k$ alors $w_{k+1} = w_k^+$, $\sigma_{k+1} = \sigma_k^+$ sinon appliquer des *itérations internes* pour calculer w_{k+1} et $\sigma_{k+1} \in [\sigma_k^+, s_k]$ t.q.

$$\|\Phi(w_{k+1}, \lambda_{k+1}, \sigma_{k+1})\| \leq \epsilon_k$$

Algorithme - Itération interne

- Fonction de mérite : $\varphi(w) = f(x) + \lambda^\top c(x) + \frac{1}{2\sigma} \|c(x)\|^2 + \frac{\nu}{2\sigma} \|c(x) + \sigma(\lambda - y)\|^2$
- Initialisations : $\lambda = \lambda_{k+1}$, $\sigma_k^+ \leq \sigma \leq s \leq s_k$, $\omega \in (0, 1)$

- 1 Choisir H symétrique t.q. $\ln(J) = (n, m, 0)$ où $J = \begin{pmatrix} H & \nabla c \\ \nabla c^\top & -\sigma I \end{pmatrix}$ et résoudre

$$Jd = -\Phi(w, \lambda, \sigma)$$

- 2 Partant de $\alpha = 1$, appliquer une recherche linéaire par rebroussement pour trouver $\alpha \in (0, 1]$ t.q.

$$\varphi(w + \alpha d) \leq \varphi(w) + \alpha \omega \nabla \varphi(w)^\top d.$$

Faire $w = w + \alpha d$.

- 3 Si $\lambda - y \neq 0$, $\hat{\sigma} = \|c\|/\|\lambda - y\|$, sinon $\hat{\sigma} = +\infty$. Si $\hat{\sigma} \leq s$, $\sigma = \max\{\hat{\sigma}, \sigma\}$.

Convergence des itérations internes

Hypothèses

- $f, c \in \mathcal{C}^1$.
- L'algorithme d'itérations internes génère des suites $\{w^i\}$ et $\{\sigma^i\}$.
- $\{\nabla c(x^i)\}$ et $\{H^i\}$ sont bornées.
- $\{H^i + \frac{1}{\sigma^i} \nabla c(x^i) \nabla c(x^i)^\top\}$ uniformément définies positives.

Theorem

Au moins une des deux conclusions a lieu:

- 1 $\{f(x^i)\}$ non bornée inférieurement.
- 2 $\liminf \|\Phi(w^i, \lambda, \sigma^i)\| = 0$.

Convergence des itérations externes

Hypothèses

- $f, c \in \mathcal{C}^1$.
- L'algorithme génère une suites $\{w_k\}$ (itérés internes et externes).
- $\{x_k\}$ est contenue dans un compact.
- $\{H_k\}$ bornée.
- $\{H_k + \frac{1}{\sigma_k} \nabla c(x_k) \nabla c(x_k)^\top\}$ uniformément définies positives.

Theorem

Une des deux conclusions a lieu.

- ① *Si le problème est non réalisable, alors tout point limite de $\{x_k\}$ est un point stationnaire de $x \mapsto \|c(x)\|$, $\{y_k\}$ est non bornée et $\{\sigma_k\} \rightarrow 0$.*
- ② *Si le problème est réalisable, alors*
 - *soit $\{y_k\}$ est non bornée et $\{x_k\}$ a un point limite \bar{x} tel que $\nabla c(\bar{x})$ n'est pas de plein rang,*
 - *soit $\{y_k\}$ est bornée et tout point limite de $\{w_k\}$ est un point stationnaire du problème initial.*

Comportement asymptotique

Hypothèses

- $f, c \in \mathcal{C}^{2,1}$, LICQ et SOC en w^* .
- L'algorithme génère une suite $\{w_k\} \rightarrow w^*$.
- $\{\sigma_k\} \rightarrow 0$
- $\|H_k - \nabla_{xx}^2 \mathcal{L}(w_k)\| = O(\sigma_k^+)$
- $\zeta_k = \Omega(\sigma_k)$
- $\sigma_k^+ = \Theta(\|F(w_k)\|)$ où $F(w) = (\nabla_x \mathcal{L}(w), c(x))$
- $\varepsilon_k = \Omega(\sigma_k^+)$

Theorem

Pour tout k assez grand, $w_{k+1} = w_k^+$, $\sigma_{k+1} = \sigma_k^+$, $\lambda_{k+1} = y_k$ et $\|w_{k+1} - w^*\| = O(\|w_k - w^*\|^2)$.

Comportement asymptotique

Hypothèses

- $f, c \in \mathcal{C}^{2,1}$, LICQ et SOC en w^* .
- L'algorithme génère une suite $\{w_k\} \rightarrow w^*$.
- $\{\sigma_k\} \rightarrow 0$
- $\|H_k - \nabla_{xx}^2 \mathcal{L}(w_k)\| = O(\sigma_k^+)$
- $\zeta_k = \Omega(\sigma_k)$
- $\sigma_k^+ = \Theta(\|F(w_k)\|)$ où $F(w) = (\nabla_x \mathcal{L}(w), c(x))$
- $\varepsilon_k = \Omega(\sigma_k^+)$

Theorem

Pour tout k assez grand, $w_{k+1} = w_k^+$, $\sigma_{k+1} = \sigma_k^+$, $\lambda_{k+1} = y_k$ et $\|w_{k+1} - w^*\| = O(\|w_k - w^*\|^2)$.

- $\|w_k^+ - w^*\| = O(\|w_k - w^*\|^2) + O(\sigma_k^+ \|w_k - w^*\|) + O(\sigma_k^+ \|\lambda_{k+1} - y_k\|)$.
- $\|\Phi(w_k^+, \lambda_{k+1}, \sigma_k^+)\| = o(\sigma_k^+) \leq \varepsilon_k \Rightarrow \|w_{k+1} - w^*\| = o(\|w_k - w^*\|)$
- $\|c_k\| = o(\|w_{k-1} - w^*\|) = o(\sigma_{k-1}^+) = o(\sigma_k) \leq \zeta_k \Rightarrow \lambda_{k+1} = y_k$

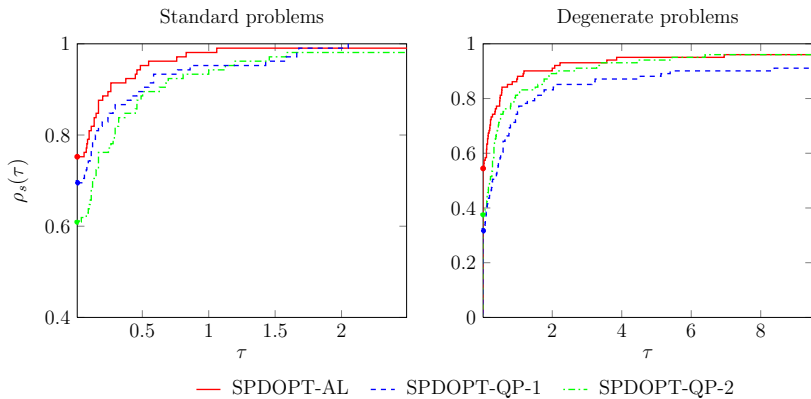
Implementation

- **SPDOPT-AL** est implémenté en **C**.
- Interfaces AMPL, CUTEr, Matlab.
- $\sigma_k^+ = \tau \min\{\sigma_k, \|F(w_k)\|\}, \tau = 0.1 \text{ ou } 0.2$
- $\zeta_k = 10\sigma_k$.
- $r_k = \min\{1/k, 10^4\|F(w_k)\|\}$.
- $H_k = \nabla_{xx}^2 \mathcal{L}(w_k) + \delta I$, avec $\delta \geq 0$.
- Solution du système linéaire : MA57 (MUMPS).
- Test d'arrêt des itérations internes

$$\|\Phi(w_{k+1}, \lambda_{k+1}, \sigma_{k+1})\|_\infty \leq \varepsilon_k := 0.9 \max_{(k-4)^+ \leq i \leq k} \|\Phi(w_i, \lambda_i, \sigma_i)\|_\infty + 10\sigma_k$$

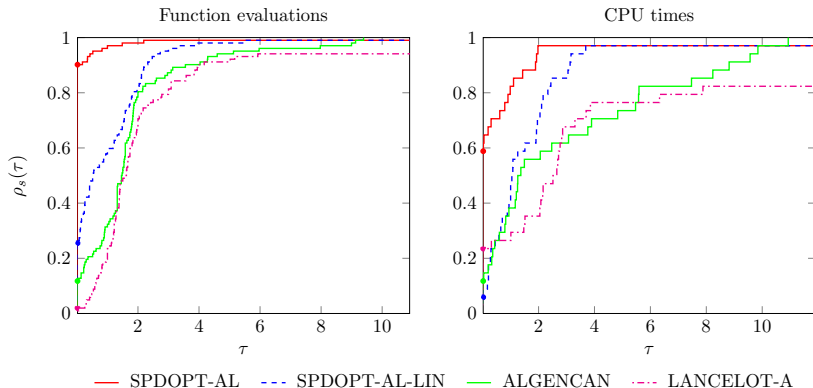
Comparaison SPDOPT-AL versus SPDOPT-QP

- Comparaison du nombre d'évaluations de f pour SPDOPT-AL, SPDOPT-QP sur des problèmes standards et dégénérés



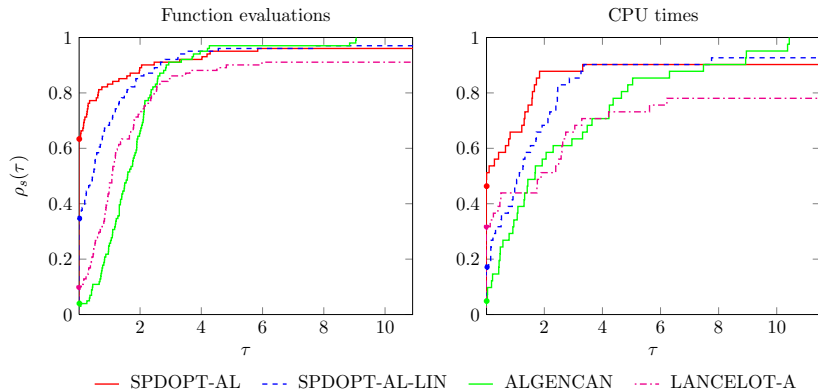
Comparaison SPDOPT-AL versus ALGENCAN et LANCELOT

- Comparaison du nombre d'évaluations de f et du CPU pour SPDOPT-AL, ALGENCAN et LANCELOT sur des problèmes standards avec contraintes d'égalité.



Comparaison SPDOPT-AL versus ALGENCAN et LANCELOT

- Comparaison du nombre d'évaluations de f et du CPU pour SPDOPT-AL, ALGENCAN et LANCELOT sur des problèmes dégénérés avec contraintes d'égalité.



Plan

- 1 Introduction et motivations
- 2 Contraintes d'égalité
- 3 Contraintes générales

Lagrangien augmenté + Log-barrière

- Pénalisation mixte:

$$\min_{x \in \mathbb{R}^n} f(x) + \lambda^\top c(x) + \frac{1}{2\sigma} \|c(x)\|^2 - \mu \sum_{i=1}^n \log x_i,$$

- CN1 :

$$\nabla f(x) + \nabla c(x) \left(\lambda + \frac{1}{\sigma} c(x) \right) - \mu X^{-1} e = 0.$$

- Introduisons les variables duales $y = \lambda + \frac{1}{\sigma} c(x)$ et $z = \mu X^{-1} e$.
- KKT $_{\mu, \sigma}$:

$$\Phi(w, \lambda, \sigma, \mu) := \begin{pmatrix} \nabla f(x) + \nabla c(x)y - z \\ c(x) + \sigma(\lambda - y) \\ XZe - \mu e \end{pmatrix} = 0.$$

Système linéaire et fonction de mérite

- Linéarisation de $\text{KKT}_{\mu,\sigma}$, **symétrisation** et **régularisation** :

$$\begin{pmatrix} \nabla_{xx}^2 \mathcal{L}(x, y) + \delta I & \nabla c(x) & -I \\ \nabla c(x)^\top & -\sigma I & 0 \\ -I & 0 & -Z^{-1}X \end{pmatrix} \begin{pmatrix} d_x \\ d_y \\ dz \end{pmatrix} = - \begin{pmatrix} \nabla_x \mathcal{L}(w) \\ c(x) + \sigma(\lambda - y) \\ x - \mu Z^{-1}e \end{pmatrix}$$

- Fonction de mérite **primal-duale**

$$\begin{aligned} \Theta(w) = & f(x) + \lambda^\top c(x) + \frac{1}{2\sigma} \|c(x)\|^2 - \mu \sum_{i=1}^n \log x_i \\ & + \frac{\nu}{2\sigma} \|c(x) + \sigma(\lambda - y)\|^2 \\ & + x^\top z - \mu \sum_{i=1}^n \log(x_i z_i). \end{aligned}$$

Système linéaire et fonction de mérite

- Linéarisation de $\text{KKT}_{\mu,\sigma}$, symétrisation et régularisation :

$$\begin{pmatrix} \nabla_{xx}^2 \mathcal{L}(x, y) + \delta I & \nabla c(x) & -Z^{\frac{1}{2}} \\ \nabla c(x)^\top & -\sigma I & 0 \\ -Z^{\frac{1}{2}} & 0 & -X \end{pmatrix} \begin{pmatrix} d_x \\ d_y \\ Z^{-\frac{1}{2}} dz \end{pmatrix} = - \begin{pmatrix} \nabla_x \mathcal{L}(w) \\ c(x) + \sigma(\lambda - y) \\ -XZ^{\frac{1}{2}} e + \mu Z^{-\frac{1}{2}} e \end{pmatrix}$$

- Fonction de mérite primale-duale

$$\begin{aligned} \Theta(w) = & f(x) + \lambda^\top c(x) + \frac{1}{2\sigma} \|c(x)\|^2 - \mu \sum_{i=1}^n \log x_i \\ & + \frac{\nu}{2\sigma} \|c(x) + \sigma(\lambda - y)\|^2 \\ & + x^\top z - \mu \sum_{i=1}^n \log(x_i z_i). \end{aligned}$$

Illustration de l'algorithme - Globalisation

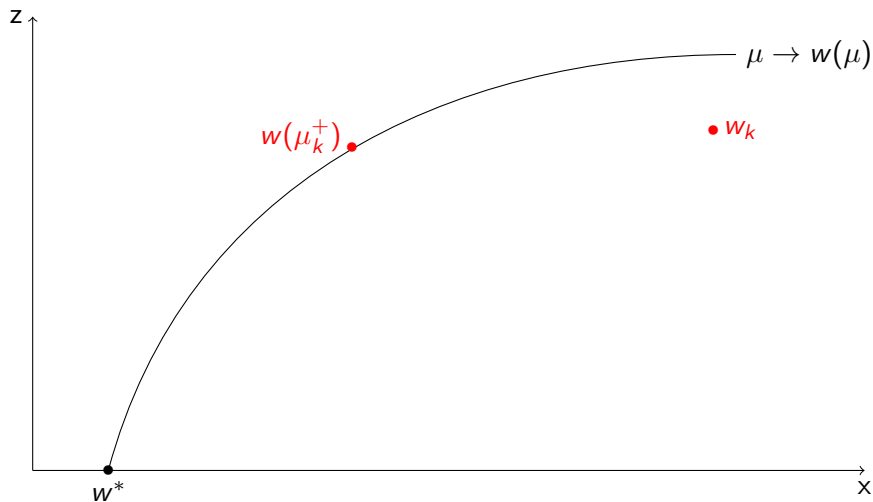


Illustration de l'algorithme - Globalisation

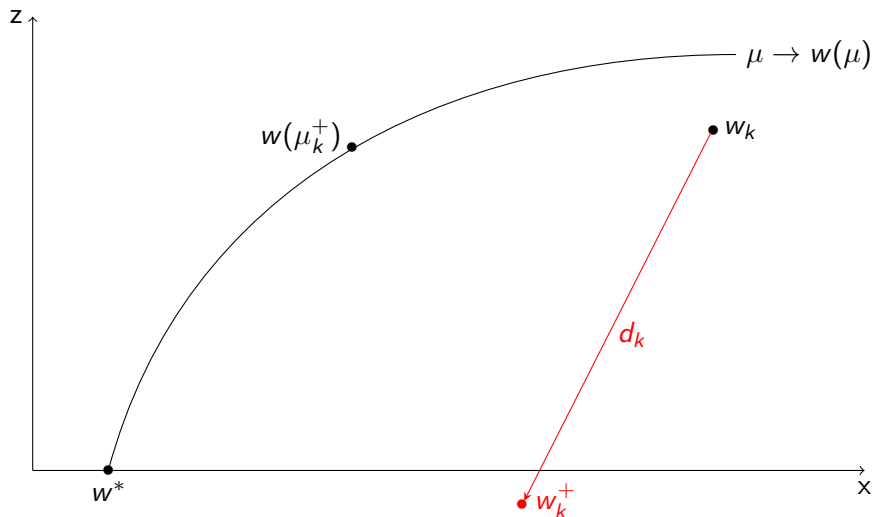


Illustration de l'algorithme - Globalisation

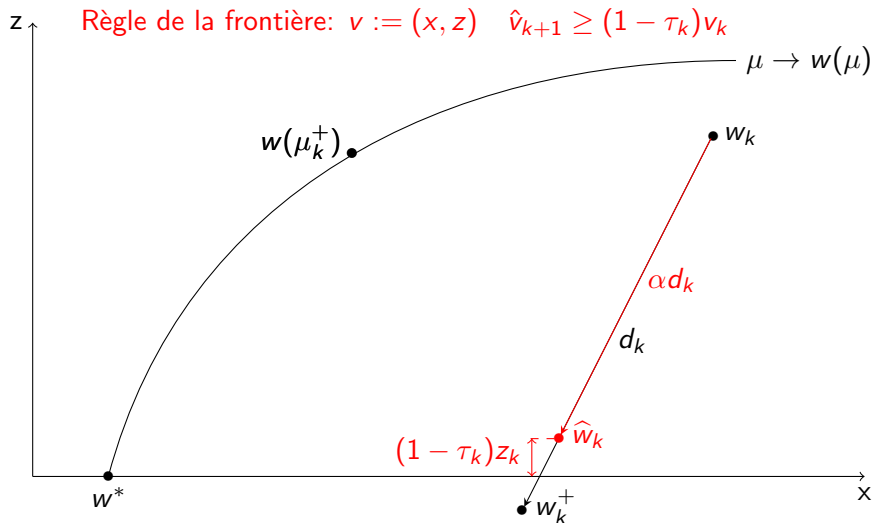


Illustration de l'algorithme - Globalisation

$$\|\Phi(w_{k+1}, \lambda_{k+1}, \sigma_{k+1}, \mu_{k+1})\| \leq \epsilon_k$$

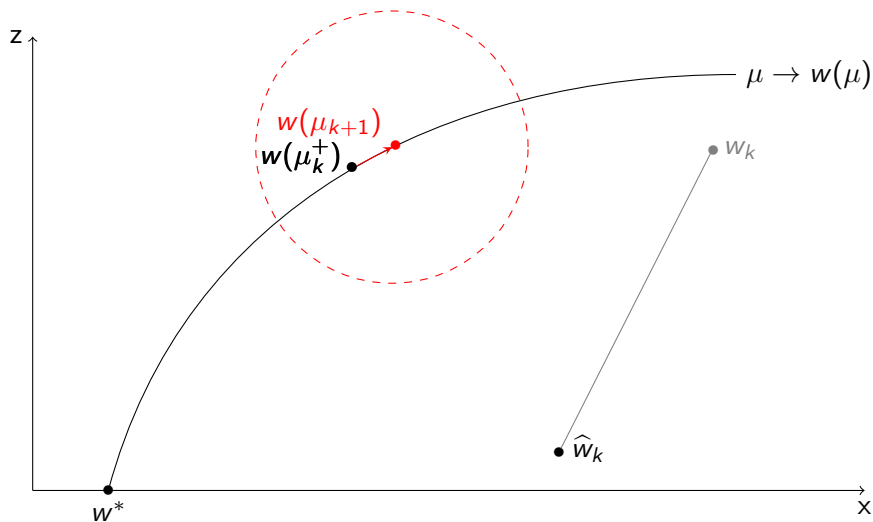


Illustration de l'algorithme - Globalisation

$$\|\Phi(w_{k+1}, \lambda_{k+1}, \sigma_{k+1}, \mu_{k+1})\| \leq \epsilon_k$$

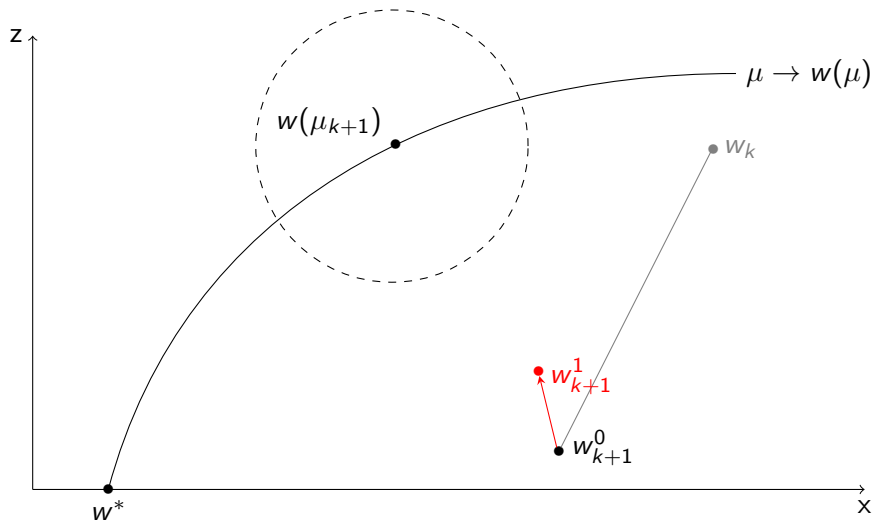


Illustration de l'algorithme - Globalisation

$$\|\Phi(w_{k+1}, \lambda_{k+1}, \sigma_{k+1}, \mu_{k+1})\| \leq \epsilon_k$$

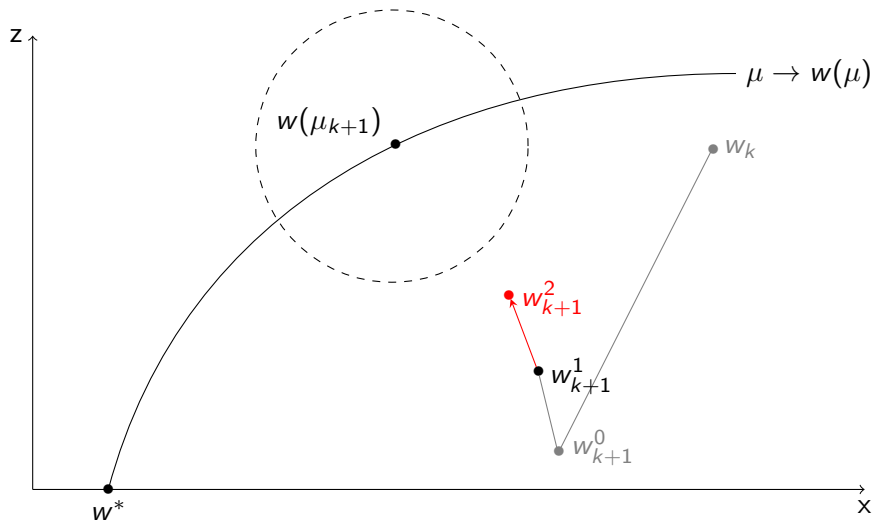


Illustration de l'algorithme - Globalisation

$$\|\Phi(w_{k+1}, \lambda_{k+1}, \sigma_{k+1}, \mu_{k+1})\| \leq \epsilon_k$$

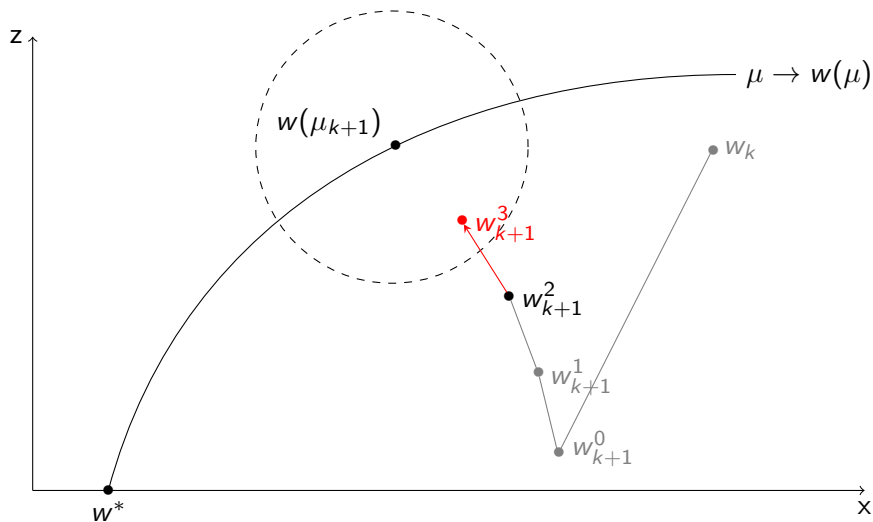


Illustration de l'algorithme - Globalisation

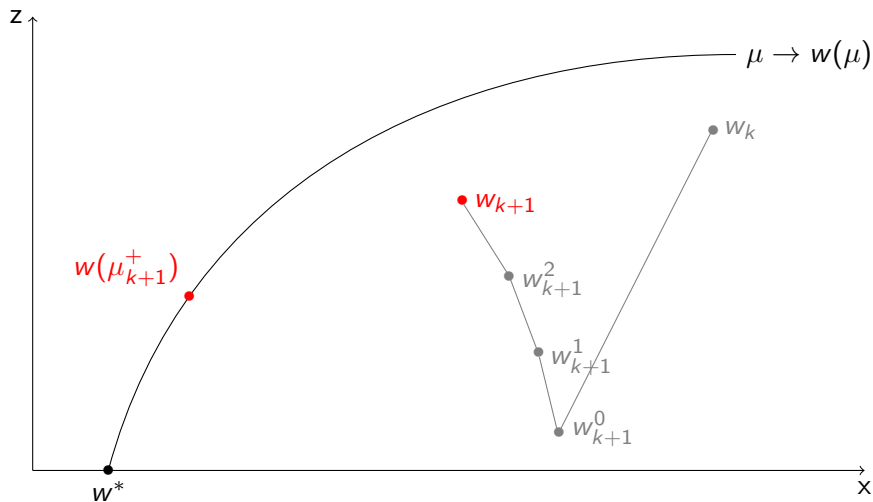


Illustration de l'algorithme - Globalisation

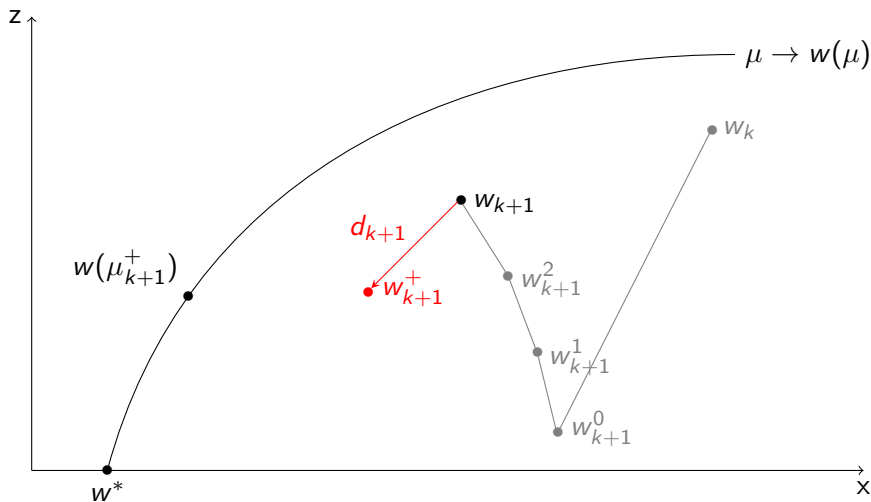


Illustration de l'algorithme - Globalisation

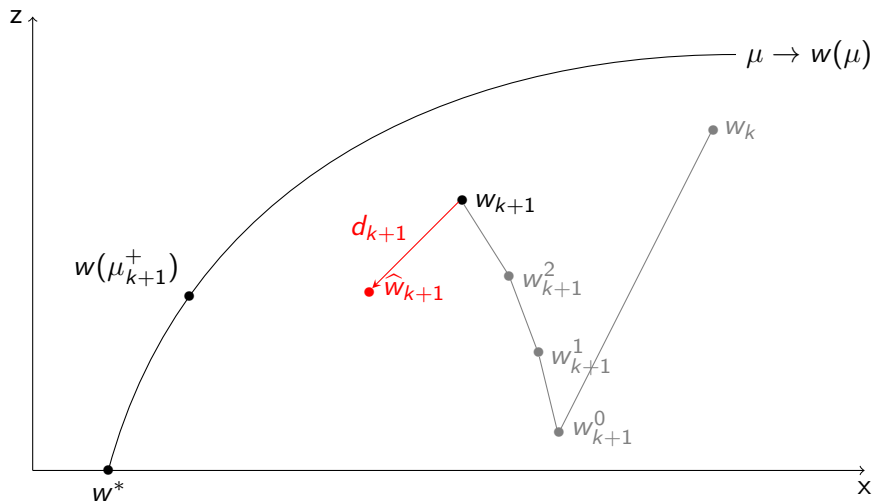


Illustration de l'algorithme - Globalisation

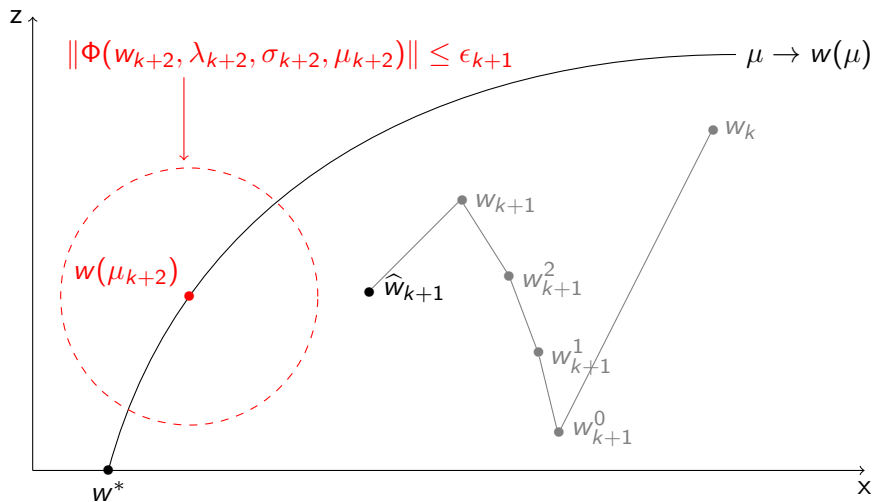


Illustration de l'algorithme - Globalisation

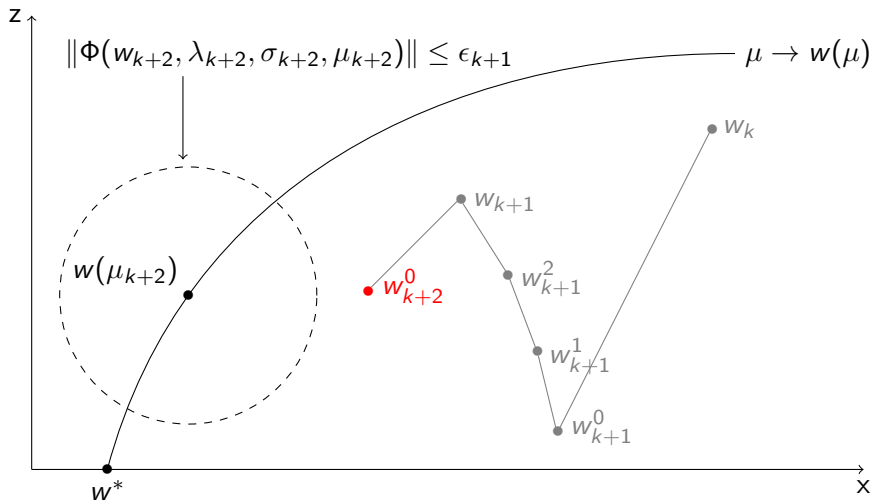


Illustration de l'algorithme - Globalisation

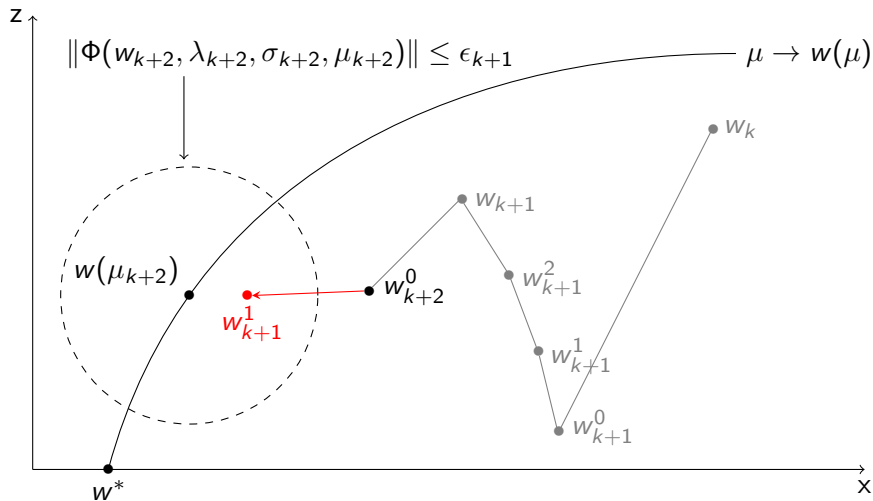


Illustration de l'algorithme - Globalisation

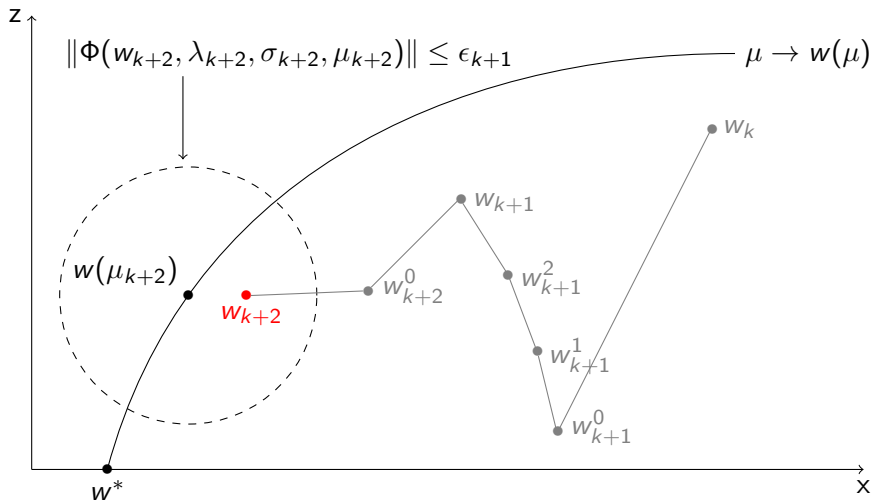
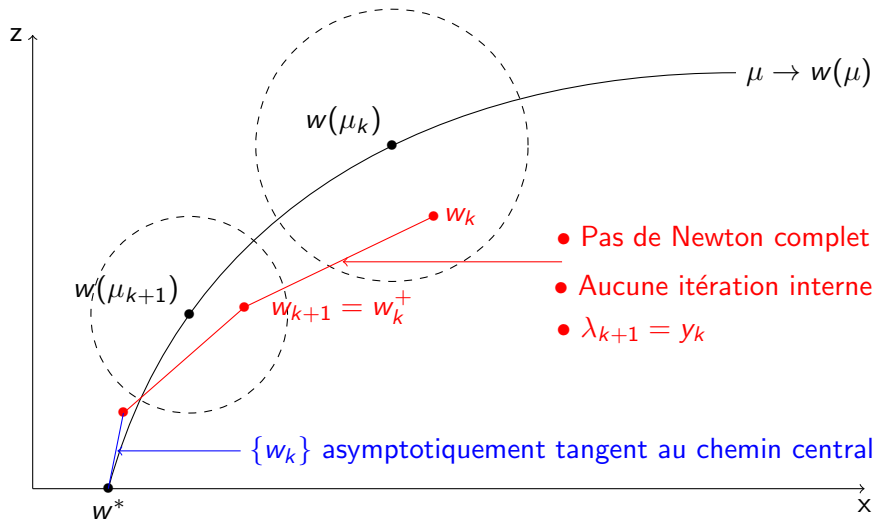


Illustration de l'algorithme - Comportement asymptotique



Convergence globale des itérations internes

Hypothèses

- ① La suite $\{x^i\}$ des itérés internes est dans un compact de \mathbb{R}^n .
- ② Les matrices $\{\nabla_{xx}^2 \mathcal{L}(w^i) + \theta^i I + \frac{1}{\sigma} \nabla c(x^i) \nabla c(x^i)^\top + (X^i)^{-1} Z^i\}$ sont uniformément définies positives.

Theorem

Pour des paramètres σ , μ et λ fixés, les itérations internes convergent au sens où

$$\lim_{i \rightarrow \infty} \|\Phi(w^i, \lambda, \sigma, \mu)\| = 0.$$

Convergence globale des itérations externes

Hypothèses

- ① $\{\nabla_{xx}^2 \mathcal{L}(w_k) + \theta_k I\}$, $\{c(x_k)\}$, $\{\nabla f(x_k)\}$ et $\{\nabla c(x_k)\}$ bornées.
- ② $\{\nabla_{xx}^2 \mathcal{L}(w_k) + \theta_k I + \frac{1}{\sigma} \nabla c(x_k) \nabla c(x_k)^\top + (X_k)^{-1} Z_k\}$ uniformément définie positive.
- ③ $\mu_k^+ \leq a \mu_k$, pour $a \in (0, 1)$ et k assez grand.
- ④ $\sigma_k = \Omega(\mu_k)$.
- ⑤ $\{\varepsilon_k\} \rightarrow 0$.

Convergence globale des itérations externes

Résultat

Theorem

Les suites $\{\nabla_x \mathcal{L}(w_k)\}$, $\{X_k Z_k e\}$ et $\{\mu_k\}$ convergent vers zéro. La suite $\{x_k\}$ est asymptotiquement stationnaire pour le problème de réalisabilité

$$\begin{array}{ll} \min & \frac{1}{2} \|c(x)\|^2, \\ \text{s.c.} & x \geq 0. \end{array}$$

De plus, une des deux conclusions suivantes se produit.

- i) *La suite $\{y_k\}$ est non bornée. Alors, les conditions de qualification de Mangasarian-Fromovitz sont asymptotiquement non satisfaites, i.e.*
 $\exists \{(a_k, b_k)\}$ *unitaire, telle que*

$$\liminf_{k \rightarrow \infty} \|\nabla c(x_k) a_k - b_k\| = 0 \quad \text{et} \quad b_k \geq 0.$$

- ii) *La suite $\{y_k\}$ est bornée. Alors, la suite $\{x_k\}$ est asymptotiquement réalisable, i.e., $\lim c(x_k) = 0$.*

Analyse asymptotique

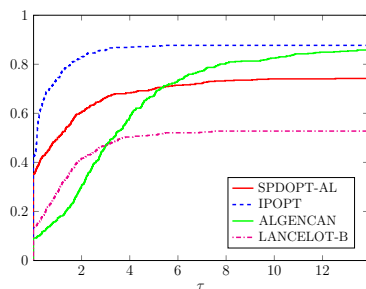
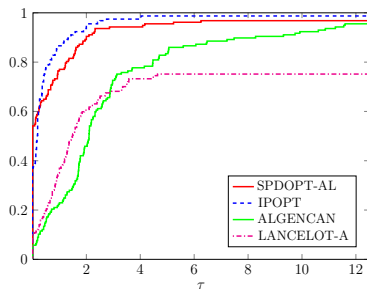
Theorem

Sous des hypothèses standards de régularité en w^ (LICQ, SSOC, SC), si $\varepsilon_k = \Omega(\mu_k)$, $\sigma_k = O(\mu_k)$, $\{\sigma_k\} \rightarrow 0$ et $\{\mu_k\} \rightarrow 0$ avec un taux de convergence au plus superlinéaire, alors*

- ❶ *pour k assez grand, $\alpha_k = 1$ et pas d'itération interne.*
- ❷ *$w_k = w(\mu_k) + o(\mu_k)$ et $\{w_k\}$ converge vers w^* à la même vitesse que $\{\mu_k\}$ vers zéro.*
- ❸ *Pour k assez grand, $\lambda_{k+1} = y_k$.*

Comparaison SPDOPT-AL versus IPOPT, ALGENCAN et LANCELOT sur HS et CUTEst

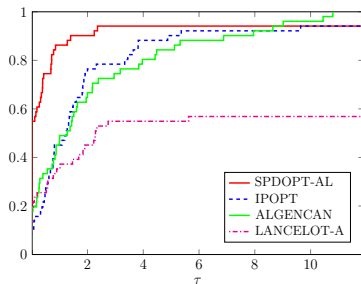
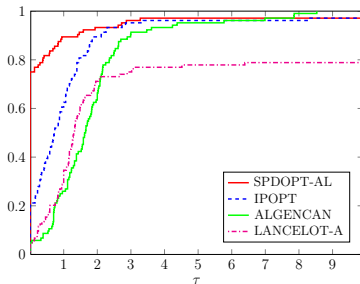
- Nombre d'évaluations de fonctions de SPDOPT-AL, IPOPT, ALGENCAN et LANCELOT sur des problèmes de les collections Hock-Schittkowski et CUTEst.



Comparaison SPDOPT-AL versus IPOPT, ALGENCAN et LANCELOT sur problèmes dégénérés

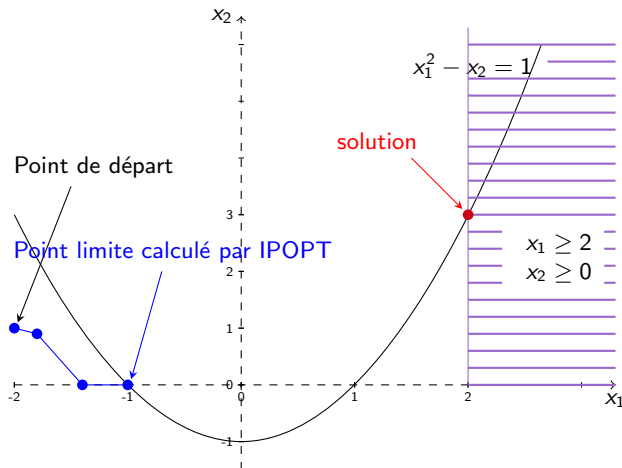
- Nombre d'évaluations de fonctions et temps CPU de SPDOPT-AL, IPOPT, ALGENCAN et LANCELOT sur des problèmes dégénérés:

$$\begin{array}{ll} \min & f(x) \\ \text{s.c.} & g(x) = 0 \end{array} \quad \rightarrow \quad \begin{array}{ll} \min & f(x) \\ \text{s.c.} & g(x) \leq 0 \\ & g(x) \geq 0 \end{array}$$



Non convergence de certaines méthodes de points intérieurs : exemple de Wächter-Biegler

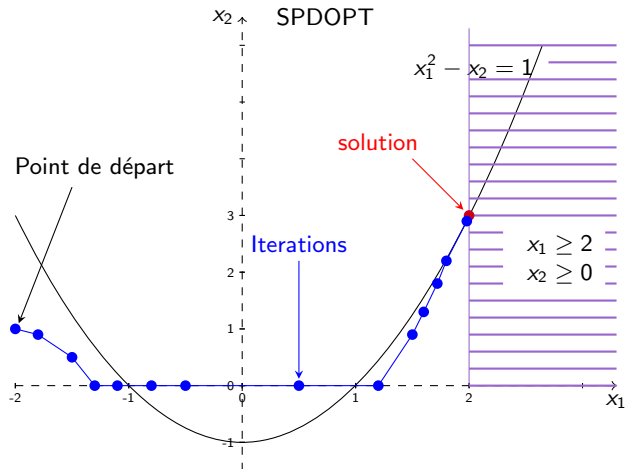
$$\begin{array}{ll}\min & x_1 \\ \text{s.t.} & x_1^2 - x_2 = 1 \\ & x_1 - x_3 = 2 \\ & x_2 \geq 0, x_3 \geq 0\end{array}$$



- Toute méthode qui **linéarise les contraintes** et **maintient les contraintes de positivité**, échoue.

Comportement de SPDOPT sur l'exemple de Wächter-Biegler

$$\begin{array}{ll}
 \min & x_1 \\
 \text{s.t.} & x_1^2 - x_2 = 1 \\
 & x_1 - x_3 = 2 \\
 & x_2 \geq 0, x_3 \geq 0
 \end{array}$$



Extension et questions ouvertes

- La propriété de convergence locale peut s'étendre au cas d'une résolution approchée du système de Newton⁵. Concevoir un algorithme "matrix-free".
- Avec une pénalisation directe de $c(x) \geq 0$,

$$\min f(x) - \sum_{i=1}^m \log c_i(x),$$

a-t-on le même comportement asymptotique ? Extension aux SDP.

- Analyser la convergence dans le cadre d'une approximation du hessien du lagrangien par une méthode de quasi-Newton (BFGS) ? A-t-on une convergence superlinéaire ? Implémentation.
- Le cas "dynamique" $\mu = \|F(w)\|^{1+\delta}$, $0 \leq \delta < 1$, est à analyser.
- Trouver des règles de mise à jour des paramètres qui soient "optimales".

⁵Armand-Benoist-Dussault, COAP 2011

Références

- P. Armand, J. Benoist, R. Omhenni, V. Pateloup. *Study of a primal-dual algorithm for equality constrained minimization*, Computational Optimization and Applications, 59:405–433, 2014
- P. Armand and J. Benoist. *Uniform boundedness of the inverse of a Jacobian matrix arising in regularized interior-point methods*, Mathematical Programming, 137: 587-592, 2013
- P. Armand, R. Omhenni *A globally and quadratically convergent primal-dual augmented Lagrangian algorithm for equality constrained optimization*, to appear in Optimization Methods and Software, 2014
- R. Omhenni. *Méthodes primales-duales régularisées pour l'optimisation non linéaire avec contraintes*, Thèse, Laboratoire XLIM, Novembre 2014