Optimisation sous contraintes

Fabrice Rossi

TELECOM ParisTech

Décembre 2009/Janvier 2010

Plan

Plan

Forme générale

- un problème d'optimisation $(\mathcal{P})$ est défini par

$$\begin{array}{rl} \text{minimiser sur } \mathbb{R}^n & J(\mathbf{x}) \\ \text{avec} & h_i(\mathbf{x}) = 0\,,\, 1 \leq i \leq p \\ & g_j(\mathbf{x}) \leq 0\,,\, 1 \leq j \leq q \end{array}$$

Forme générale

- un problème d'optimisation $(\mathcal{P})$ est défini par

$$\begin{array}{rl} \text{minimiser sur } \mathbb{R}^n & J(\mathbf{x}) \\ \text{avec} & h_i(\mathbf{x}) = 0\,,\, 1 \leq i \leq p \\ & g_j(\mathbf{x}) \leq 0\,,\, 1 \leq j \leq q \end{array}$$

- rappel de vocabulaire :
  - les h_i sont les **contraintes d'égalité** (notées $\mathbf{h}(\mathbf{x}) = 0$)
  - les g_j sont les **contraintes d'inégalité** (notées $\mathbf{g}(\mathbf{x}) \leq 0$)
  - l'**ensemble des contraintes** est

$$\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^n | h_i(\mathbf{x}) = 0\,,\, 1 \leq i \leq p \text{ et } g_j(\mathbf{x}) \leq 0\,,\, 1 \leq j \leq q\}$$

ensemble des points admissibles ou **réalisables**

Conséquences

- les contraintes changent les conditions d'optimalité
- exemple :
  - $J(x, y) = x^2 + y^2$ à minimiser sous la contrainte $g(x, y) = 4 - x^2 - y^2 \leq 0$
  - sur $\mathbb{R}^2$, on étudierait $\nabla J = 2(x, y)^T$
  - mais ici, le minimum vaut 4 et est atteint sur le cercle $x^2 + y^2 = 4$ sur lequel $\nabla J \neq \mathbf{0}$

Conséquences

- les contraintes changent les conditions d'optimalité
- exemple :
  - $J(x,y) = x^2 + y^2$ à minimiser sous la contrainte $g(x,y) = 4 - x^2 - y^2 \leq 0$
  - sur $\mathbb{R}^2$, on étudierait $\nabla J = 2(x,y)^T$
  - mais ici, le minimum vaut 4 et est atteint sur le cercle $x^2 + y^2 = 4$ sur lequel $\nabla J \neq \mathbf{0}$
- mais pas toujours :
  - $J(x,y) = x^2 + y^2$ à minimiser sous la contrainte $g(x,y) = x^2 + y^2 - 4 \leq 0$
  - le minimum est atteint en $(0,0)$, avec $\nabla J = \mathbf{0}$

Conséquences

- les contraintes changent les conditions d'optimalité
- exemple :
  - $J(x, y) = x^2 + y^2$ à minimiser sous la contrainte $g(x, y) = 4 - x^2 - y^2 \leq 0$
  - sur $\mathbb{R}^2$, on étudierait $\nabla J = 2(x, y)^T$
  - mais ici, le minimum vaut 4 et est atteint sur le cercle $x^2 + y^2 = 4$ sur lequel $\nabla J \neq \mathbf{0}$
- mais pas toujours :
  - $J(x, y) = x^2 + y^2$ à minimiser sous la contrainte $g(x, y) = x^2 + y^2 - 4 \leq 0$
  - le minimum est atteint en $(0, 0)$, avec $\nabla J = \mathbf{0}$
- les contraintes doivent donc apparaître dans les conditions d'optimalité

Existence d'un minimum

- cas général : $(\mathcal{P})$: $\min J(\mathbf{x}), \mathbf{x} \in \mathcal{C} \subset \mathbb{R}^n$
- on suppose :
  - J continue
  - et $\mathcal{C}$ fermé et non vide
- alors :
  - si :
    - $\mathcal{C}$ est borné
    - ou si J est coercitive
  - alors $(\mathcal{P})$ admet au moins une solution

Existence et unicité

- remarque : si

$$\mathcal{C} = \left\{ \mathbf{x} \in \mathbb{R}^n | h_i(\mathbf{x}) = 0\,, 1 \leq i \leq p \text{ et } g_j(\mathbf{x}) \leq 0\,, 1 \leq j \leq q \right\}$$

  avec des h_i et g_j continues, alors $\mathcal{C}$ est fermé
- si J est strictement convexe et $\mathcal{C}$ est convexe, alors $(\mathcal{P})$ admet **au plus** une solution
- problème convexe :
  - J est convexe
  - les h_i sont affines
  - les g_j sont convexes
  - et donc $\mathcal{C}$ est convexe

TELECOM ParisTech

Condition du premier ordre

- si J est Gâteaux-différentiable en x^* solution de $(\mathcal{P})$ et si $\mathcal{C}$ est convexe, alors :

$$\forall \mathbf{x} \in \mathcal{C},\ \langle \nabla J(\mathbf{x}^*), \mathbf{x} - \mathbf{x}^* \rangle \geq 0$$

- remarques :
  - intuitivement : on ne peut s'éloigner du minimum que dans une direction de montée
  - généralisable : notion de direction admissible
  - si $\mathbf{x}^*$ est un point intérieur de $\mathcal{C}$ alors $\nabla J(\mathbf{x}^*) = 0$
- si J est convexe la condition est nécessaire et suffisante

Égalités et inégalités

- Conditions nécessaires non qualifiées
- cas particulier $h_i(\mathbf{x}) = 0$ et $g_j(\mathbf{x}) \leq 0$ où tout est C^1 (J inclus)
- soit $\mathbf{x}^*$ une solution de $(\mathcal{P})$, alors il existe $\boldsymbol{\lambda}^* = (\lambda_1^*, \ldots, \lambda_p^*)$ et $\boldsymbol{\mu}^* = (\mu_0^*, \mu_1^*, \ldots, \mu_q^*)$ tels que
  - $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \neq \mathbf{0}$
  - $h_i(\mathbf{x}^*) = 0,\ 1 \leq i \leq p$ (admissibilité en égalité)
  - $g_j(\mathbf{x}^*) \leq 0,\ 1 \leq j \leq q$ (admissibilité en inégalité)
  - $\mu_j^* \geq 0,\ 0 \leq j \leq q$
  - $\mu_j^* g_j(\mathbf{x}^*) = 0,\ 1 \leq j \leq q$ (**conditions de complémentarité**)
  - et

$$\mu_0^* \nabla J(\mathbf{x}^*) + \sum_{i=1}^{p} \lambda_i^* \nabla h_i(\mathbf{x}^*) + \sum_{j=1}^{q} \mu_j^* \nabla g_j(\mathbf{x}^*) = 0$$

Qualification

- condition utile si $\mu_0 \neq 0$
- problème de qualification des contraintes :
  - conditions (locales) sur le problème qui garantissent que $\mu_0 \neq 0$
  - très nombreuses variantes plus ou moins sophistiquées

Qualification

- condition utile si $\mu_0 \neq 0$
- problème de qualification des contraintes :
  - conditions (locales) sur le problème qui garantissent que $\mu_0 \neq 0$
  - très nombreuses variantes plus ou moins sophistiquées
- contrainte active : g_j est **active** (ou saturée) en $\mathbf{x}^*$ si $g_j(\mathbf{x}^*) = 0$; $I(\mathbf{x}^*)$, ensemble des indices des contraintes actives en $\mathbf{x}^*$

Qualification

- condition utile si $\mu_0 \neq 0$
- problème de qualification des contraintes :
  - conditions (locales) sur le problème qui garantissent que $\mu_0 \neq 0$
  - très nombreuses variantes plus ou moins sophistiquées
- contrainte active : g_j est **active** (ou saturée) en $\mathbf{x}^*$ si $g_j(\mathbf{x}^*) = 0$; $I(\mathbf{x}^*)$, ensemble des indices des contraintes actives en $\mathbf{x}^*$
- régularité : $\mathbf{x}^*$ est régulier pour $\mathbf{g}$ et $\mathbf{h}$ si
  - $\mathbf{x}^*$ est admissible
  - les $\nabla h_i(\mathbf{x}^*)$ sont linéairement indépendants
  - il existe $d \neq 0$ tel que $\langle \nabla h_i(\mathbf{x}^*), d \rangle = 0$ pour tout i et $\langle \nabla g_j(\mathbf{x}^*), d \rangle < 0$ pour tout $j \in I(\mathbf{x}^*)$ (ou $\langle \nabla g_j(\mathbf{x}^*), d \rangle = 0$ si g_j est affine)
  - régularité de Mangasarian-Fromowitz

Conditions qualifiées

- Conditions nécessaires qualifiées du 1er ordre de KKT (Karush, Kuhn et Tucker)
- Hypothèses :
  - J, $\mathbf{h}$ et $\mathbf{g}$ C^1
  - $\mathbf{x}^*$ solution de $(\mathcal{P})$
  - $\mathbf{x}^*$ est régulier pour $\mathbf{g}$ et $\mathbf{h}$
- Alors il existe $\boldsymbol{\lambda}^* = (\lambda_1^*, \ldots, \lambda_p^*)$ et $\boldsymbol{\mu}^* = (\mu_1^*, \ldots, \mu_q^*)$ tels que
  - $h_i(\mathbf{x}^*) = 0, \ 1 \leq i \leq p$
  - $g_j(\mathbf{x}^*) \leq 0, \ 1 \leq j \leq q$
  - $\mu_j^* \geq 0, \ 1 \leq j \leq q$
  - $\mu_j^* g_j(\mathbf{x}^*) = 0, \ 1 \leq j \leq q$
  - et

$$\nabla J(\mathbf{x}^*) + \sum_{i=1}^{p} \lambda_i^* \nabla h_i(\mathbf{x}^*) + \sum_{j=1}^{q} \mu_j^* \nabla g_j(\mathbf{x}^*) = 0$$

Cas convexe

- si le problème $(\mathcal{P})$ est convexe, les conditions de KKT sont nécessaires et suffisantes en un point $\mathbf{x}^*$ régulier
- remarque : le caractère suffisant ne nécessite pas la régularité
- conditions de qualification plus simples (de Slater) : il existe au moins un point strictement admissible $\mathbf{g}(\mathbf{x}) < 0$

Lagrangien

- le Lagrangien du problème $(\mathcal{P})$ est la fonction

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = J(\mathbf{x}) + \sum_{i=1}^{p} \lambda_i h_i(\mathbf{x}) + \sum_{j=1}^{q} \mu_j g_j(\mathbf{x})$$

- quand J, $\mathbf{h}$ et $\mathbf{g}$ sont C^1 les conditions de KKT s'expriment par $\nabla_{\mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = 0$
- les λ_i et les μ_j sont les multiplicateurs de Lagrange associés aux contraintes

Dualité

- fonction duale de Lagrange

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$$

- g est toujours concave
- pour $\boldsymbol{\mu} \geq 0$

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}) \leq \inf_{\mathbf{x} \in \mathcal{C}} J(\mathbf{x})$$

- problème dual $(\mathcal{Q})$ associé au problème primal $(\mathcal{P})$

$$\begin{array}{rl} \text{maximiser sur } \mathbb{R}^{p+q} & g(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ \text{avec} & \mu_j \geq 0\,,\, 1 \leq j \leq q \end{array}$$

- saut de dualité : $\inf_{\mathbf{x} \in \mathcal{C}} J(\mathbf{x}) - \max_{\boldsymbol{\mu} \geq 0} g(\boldsymbol{\lambda}, \boldsymbol{\mu})$

Point selle

- symétrisation du problème : $(\mathcal{P})$ est équivalent à

$$\inf_{\mathbf{x}} \sup_{\boldsymbol{\lambda},\boldsymbol{\mu}\geq 0} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$$

- le problème dual $(\mathcal{Q})$ est

$$\sup_{\boldsymbol{\lambda},\boldsymbol{\mu}\geq 0} \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$$

Point selle

- symétrisation du problème : $(\mathcal{P})$ est équivalent à

$$\inf_{\mathbf{x}} \sup_{\boldsymbol{\lambda},\boldsymbol{\mu}\geq 0} L(\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\mu})$$

- le problème dual $(\mathcal{Q})$ est

$$\sup_{\boldsymbol{\lambda},\boldsymbol{\mu}\geq 0} \inf_{\mathbf{x}} L(\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\mu})$$

- point selle : minimal par rapport à une variable, maximal par rapport à l'autre
- $(\mathbf{x}^*,\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*)$ est un point selle du Lagrangien si pour tout $(\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\mu})$ $(\boldsymbol{\mu}^* \geq 0$ et $\boldsymbol{\mu} \geq 0)$

$$L(\mathbf{x}^*,\boldsymbol{\lambda},\boldsymbol{\mu}) \leq L(\mathbf{x}^*,\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*) \leq L(\mathbf{x},\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*)$$

Théorème de dualité

- $(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ est un point selle avec $\boldsymbol{\mu}^* \geq 0$ ssi $\mathbf{x}^*$ est une solution de $(\mathcal{P})$, $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ est une solution de $(\mathcal{Q})$ et le saut de dualité est nul
- intérêt : pour résoudre le problème, on peut donc chercher un point selle du Lagrangien

Théorème de dualité

- $(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ est un point selle avec $\boldsymbol{\mu}^* \geq 0$ ssi $\mathbf{x}^*$ est une solution de $(\mathcal{P})$, $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ est une solution de $(\mathcal{Q})$ et le saut de dualité est nul
- intérêt : pour résoudre le problème, on peut donc chercher un point selle du Lagrangien
- remarque : un point selle du Lagrangien vérifie les conditions de KKT (sans hypothèse autre que J, **h** et **g** C^1)
- si le problème est convexe : point selle $\Leftrightarrow$ KKT

Régularité

- Condition plus forte que celle de Mangasarian-Fromowitz :
  - $\mathbf{x}^*$ est admissible
  - $\nabla h_i(\mathbf{x}^*)$ et les $\nabla g_j(\mathbf{x}^*)$ sont linéairement indépendants pour $j \in I(\mathbf{x}^*)$
- Contraintes fortement actives :

$$I^+(\mathbf{x}^*) = \{j \mid g_j(\mathbf{x}^*) = 0 \text{ et } \mu_j^* > 0\}$$

- si $I(\mathbf{x}^*) = I^+(\mathbf{x}^*)$ on a complémentarité stricte

Conditions nécessaires du 2ème ordre

- Hypothèses :
  - J, $\mathbf{h}$ et $\mathbf{g}$ C^2
  - $\mathbf{x}^*$ solution de $(\mathcal{P})$ et fortement régulier
- alors il existe $\boldsymbol{\lambda}^* = (\lambda_1^*, \ldots, \lambda_p^*)$ et $\boldsymbol{\mu}^* = (\mu_1^*, \ldots, \mu_q^*)$ tels que
  - les conditions de KKT sont vérifiées
  - et pour tout d vérifiant :
    - $\langle \nabla h_i(\mathbf{x}^*), d \rangle = 0$ pour $1 \leq i \leq p$
    - $\langle \nabla g_j(\mathbf{x}^*), d \rangle = 0$ pour $j \in I^+(\mathbf{x}^*)$
    - $\langle \nabla g_j(\mathbf{x}^*), d \rangle \leq 0$ pour $I(\mathbf{x}^*) \setminus I^+(\mathbf{x}^*)$

  on a
  $$\left\langle \nabla^2_{\mathbf{xx}} L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) d, d \right\rangle \geq 0$$

Conditions suffisantes

- Hypothèses :
  - J, $\mathbf{h}$ et $\mathbf{g}$ C^2
  - $(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ vérifie les conditions KKT
- si la matrice $\nabla^2_{\mathbf{xx}} L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ est définie positive sur

$$\Big\{ d \in \mathbb{R}^n \quad | \quad \langle \nabla h_i(\mathbf{x}^*), d \rangle = 0,\ 1 \leq i \leq, p \\ \text{et } \big\langle \nabla g_j(\mathbf{x}^*), d \big\rangle = 0,\ j \in I^+(\mathbf{x}^*) \Big\}$$

alors $\mathbf{x}^*$ est un minimum local de J sur $\mathcal{C}$

Plan

Classes d'algorithmes

- quelques grandes classes d'algorithmes :
  - gradient projeté :
    - descente de gradient
    - projection sur $\mathcal{C}$ à chaque étape
  - pénalisation :
    - optimisation sans contrainte de J+pénalité
    - méthodes extérieures : on ramène progressivement le candidat minimum dans $\mathcal{C}$
    - méthodes intérieures : on relâche progressivement les pénalités
  - programmation quadratique successive : résoudre des approximations quadratiques du problème
- principe sous-jacent : résoudre une série de problèmes sans contrainte (ou plus simple)

Projection

- outil important : projection sur un convexe fermé
  - soit C un convexe fermé et non vide de $\mathbb{R}^n$
  - pour tout $\mathbf{x}$ alors

$$\pi_C(\mathbf{x}) = \arg\min_{\mathbf{y}\in C} \|\mathbf{x} - \mathbf{y}\|^2$$

    existe et est unique
- propriétés :
  - $\pi_C(\mathbf{x})$ est l'unique élément de C tel que

$$\forall \mathbf{y} \in C,\ \langle \mathbf{x} - \pi_C(\mathbf{x}), \mathbf{y} - \pi_C(\mathbf{x})\rangle \leq 0$$

    ou encore tel que

$$\forall \mathbf{y} \in C,\ \langle \pi_C(\mathbf{x}) - \mathbf{y}, \mathbf{y} - \mathbf{x}\rangle \leq 0$$

Projection

- π_C est une contraction :

$$\forall \mathbf{x}, \mathbf{y}, \ \|\pi_C(\mathbf{x}) - \pi_C(\mathbf{y})\| \leq \|\mathbf{x} - \mathbf{y}\|$$

- preuve simple :
  - on a

$$\begin{aligned} \langle \mathbf{x} - \pi_C(\mathbf{x}), \pi_C(\mathbf{y}) - \pi_C(\mathbf{x}) \rangle &\leq 0 \\ \langle \mathbf{y} - \pi_C(\mathbf{y}), \pi_C(\mathbf{x}) - \pi_C(\mathbf{y}) \rangle &\leq 0 \end{aligned}$$

  - soit

$$\langle \mathbf{x} - \mathbf{y}, \pi_C(\mathbf{y}) - \pi_C(\mathbf{x}) \rangle + \|\pi_C(\mathbf{x}) - \pi_C(\mathbf{y})\|^2 \leq 0$$

  - et on termine par Cauchy-Schwartz

Gradient projeté

- minimisation de J sur C convexe fermé
- algorithme :
  1. point initial $\mathbf{x}_0$
  2. pour $k \geq 1$ croissant
     2.1 calculer $\mathbf{y}_{k+1} = \mathbf{x}_k - \rho_k \nabla J(\mathbf{x}_k)$
     2.2 puis $\mathbf{x}_{k+1} = \pi_C(\mathbf{y}_{k+1})$
     2.3 tester la convergence et quitter la boucle le cas échéant (par ex. $\|\mathbf{x}_{k+1} - \mathbf{x}_k\| < \epsilon$)

Gradient projeté

- minimisation de J sur C convexe fermé
- algorithme :
  1. point initial $\mathbf{x}_0$
  2. pour $k \geq 1$ croissant
     2.1 calculer $\mathbf{y}_{k+1} = \mathbf{x}_k - \rho_k \nabla J(\mathbf{x}_k)$
     2.2 puis $\mathbf{x}_{k+1} = \pi_C(\mathbf{y}_{k+1})$
     2.3 tester la convergence et quitter la boucle le cas échéant (par ex. $\|\mathbf{x}_{k+1} - \mathbf{x}_k\| < \epsilon$)
- formulation simple mais mise en oeuvre potentiellement délicate :
  - si C est simple (par ex., $l_i \leq x_i \leq u_i$), pas de problème
  - sinon le calcul de $\pi_C(\mathbf{y}_{k+1})$ est un problème d'optimisation sous contraintes

Convergence

- si J est C^1, α-convexe et de dérivée M-lipschitzienne
- et si $\rho_k \in [\beta_1, \beta_2]$ avec $0 < \beta_1 < \beta_2 < \frac{2\alpha}{M^2}$
- alors l'algorithme du gradient projeté converge :
  - preuve très proche de celle du cas sans contrainte
  - si $\mathbf{x}^*$ est l'optimum, on a pour tout ρ

$$\mathbf{x}^* = \pi_C(\mathbf{x}^* - \rho\nabla J(\mathbf{x}^*))$$

  - donc par contraction de π_C

$$\|\mathbf{x}_{k+1} - \mathbf{x}^*\|^2 \leq \|(\mathbf{x}_k - \rho_k\nabla J(\mathbf{x}_k)) - (\mathbf{x}^* - \rho_k\nabla J(\mathbf{x}^*))\|^2$$

ce qui nous ramène au cas sans projection

Pénalisation

- idée principale : remplacer un problème avec contraintes par un problème sans contrainte dont la fonction objectif « décourage » les points non admissibles
- solution naïve :
  - on définit
$$\alpha(\mathbf{x}) = \begin{cases} +\infty & \mathbf{x} \notin C \\ 0 & \mathbf{x} \in C \end{cases}$$
  - alors trouver $\arg\min_{\mathbf{x}\in C} J(\mathbf{x})$ est équivalent à trouver $\arg\min_{\mathbf{x}\in\mathbb{R}^n} J(\mathbf{x}) + \alpha(\mathbf{x})$
- solutions réalistes : utiliser une fonction α régulière (C^1 au moins) petite sur C et grande en dehors

Méthodes de point extérieur

- α vérifie :
  - α est continue sur $\mathbb{R}^n$
  - $\alpha(\mathbf{x}) \geq 0$
  - $\alpha(\mathbf{x}) = 0 \Leftrightarrow \mathbf{x} \in C$
- Exemples :
  - $\mathbf{h}(\mathbf{x}) = 0$ est représentée par $\alpha(\mathbf{x}) = \|\mathbf{h}(\mathbf{x})\|^2$
  - $\mathbf{g}(\mathbf{x}) \leq 0$ est représentée par $\alpha(\mathbf{x}) = \|\mathbf{g}^+(\mathbf{x})\|^2$
- on considère la famille de problèmes $(\mathcal{P}_r)$ pour $r > 0$ définis par

$$\min_{\mathbf{x}\in\mathbb{R}^n} \left(J(\mathbf{x}) + r\alpha(\mathbf{x})\right)$$

- méthodes de point extérieur : on ne peut pas garantir $\mathbf{x}_r^* \in C$

Méthodes de point intérieur

- même principe, mais avec des points à l'intérieur de C
- pour les contraintes d'inégalité seulement
- α est une barrière et vérifie :
  - α est continue sur $\overset{\circ}{C}$
  - $\mathbf{x} \notin C \Rightarrow \alpha(\mathbf{x}) = \infty$
- pour les contraintes $\mathbf{g}(\mathbf{x}) \leq 0$, on prend généralement

$$-\sum_{j=1}^{q} \log(-g_j(\mathbf{x}))$$

- comme pour les méthodes de point extérieur, on considère les problèmes $(\mathcal{P}_r)$ pour $r > 0$ définis par

$$\min_{\mathbf{x}\in\mathbb{R}^n} (J(\mathbf{x}) + r\alpha(\mathbf{x}))$$

- ici, on a toujours $\mathbf{x}_r^* \in \overset{\circ}{C}$

Pénalisation

- minimisation de J sur C convexe fermé
- algorithme (pour une suite $(r_k)_k$ croissante) :
  1. point initial $\mathbf{x}_0$
  2. pour $k \geq 1$ croissant
     2.1 résoudre le problème $(\mathcal{P}_{r_k})$

     $$\min_{\mathbf{x}\in\mathbb{R}^n} \left(J(\mathbf{x}) + r_k\alpha(\mathbf{x})\right)$$

     en partant de la solution $\mathbf{x}_{k-1}$

     2.2 tester la qualité du point obtenu $\mathbf{x}_k$ et quitter la boucle le cas échéant

Pénalisation

- minimisation de J sur C convexe fermé
- algorithme (pour une suite $(r_k)_k$ croissante) :
  1. point initial $\mathbf{x}_0$
  2. pour $k \geq 1$ croissant
     2.1 résoudre le problème $(\mathcal{P}_{r_k})$

     $$\min_{\mathbf{x} \in \mathbb{R}^n} (J(\mathbf{x}) + r_k \alpha(\mathbf{x}))$$

     en partant de la solution $\mathbf{x}_{k-1}$

     2.2 tester la qualité du point obtenu $\mathbf{x}_k$ et quitter la boucle le cas échéant
- fonctionne très bien en pratique pour le cas du point intérieur (on résout $(\mathcal{P}_{r_k})$ par une méthode de (quasi)Newton avec contraintes d'égalités)
- l'efficacité vient du redémarrage depuis un bon candidat

Convergence

- on suppose J continue et coercitive, et C fermé et non vide
- on considère une méthode de point extérieur
- résultat :
  - pour tout $r > 0$, $(\mathcal{P}_r)$ possède au moins une solution
  - toute famille $(\mathbf{x}_r)_{r>0}$ de solutions est bornée
  - les valeurs d'adhérence de toute famille $(\mathbf{x}_r)_{r>0}$ de solutions sont des solutions de $(\mathcal{P})$

Convergence

- on suppose J continue et coercitive, et C fermé et non vide
- on considère une méthode de point extérieur
- résultat :
  - pour tout $r > 0$, $(\mathcal{P}_r)$ possède au moins une solution
  - toute famille $(\mathbf{x}_r)_{r>0}$ de solutions est bornée
  - les valeurs d'adhérence de toute famille $(\mathbf{x}_r)_{r>0}$ de solutions sont des solutions de $(\mathcal{P})$
- preuve ($\mathbf{x}^*$ est une solution de $(\mathcal{P})$) :
  - $J_r(\mathbf{x}) = J(\mathbf{x}) + r\alpha(\mathbf{x})$ est coercitive et continue
  - $\mathbf{x}_r$ solution de $(\mathcal{P}_r)$, $J(\mathbf{x}_r) \leq J_r(\mathbf{x}_r) \leq J_r(\mathbf{x}^*) = J(\mathbf{x}^*)$, donc $(\mathbf{x}_r)_{r>0}$ est bornée
  - comme $\alpha(\mathbf{x}_r) \leq \frac{1}{r}(J(\mathbf{x}^*) - J(\mathbf{x}_r))$, on a $\lim_{r\rightarrow+\infty} \alpha(\mathbf{x}_r) = 0$
  - et donc pour toute valeur d'adhérence $\hat{\mathbf{x}}$, $J(\hat{\mathbf{x}}) \leq J(\mathbf{x}^*)$

Algorithme d'Uzawa

- idée : chercher directement un point selle du Lagrangien
- algorithme (paramètre $\rho > 0$) :
  1. valeurs initiales des multiplicateurs $\boldsymbol{\lambda}^1, \boldsymbol{\mu}^1$
  2. pour $k \geq 1$ croissant

     2.1 trouver $\mathbf{x}_k$ solution du problème (sans contrainte)

     $$\min_{\mathbf{x}\in\mathbb{R}^n} L(\mathbf{x}, \boldsymbol{\lambda}^k, \boldsymbol{\mu}^k)$$

     2.2 mettre à jour $(\boldsymbol{\lambda}^k, \boldsymbol{\mu}^k)$ par

     $$\begin{aligned} \lambda_i^{k+1} &= \lambda_i^k + \rho h_i(\mathbf{x}_k) \\ \mu_j^{k+1} &= \max(0, \mu_j^k + \rho g_j(\mathbf{x}_k)) \end{aligned}$$

     2.3 tester la convergence et quitter la boucle le cas échéant (par ex. $\|\mathbf{x}_{k+1} - \mathbf{x}_k\| < \epsilon$)

Algorithme d'Uzawa

- on peut montrer que l'algorithme d'Uzawa est un gradient projeté sur le problème dual :
  - on montre que $\nabla g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = (\boldsymbol{\lambda}, \boldsymbol{\mu})^T$
  - on maximise $g(\boldsymbol{\lambda}, \boldsymbol{\mu})$, ce qui explique le signe dans la mise à jour des multiplicateurs
  - intéressant parce que la projection est triviale
- Convergence :
  - si J est C^1, α-convexe, **h** affine et **g** convexe et M_g lipschitzienne
  - et si L possède un point selle $(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$
  - alors l'algorithme converge vers $\mathbf{x}^*$ pour tout choix de ρ dans $]0, \frac{2\alpha}{M_g^2 + M_h^2}[$