Dualité

I) Introduction

$$V(p) = \begin{cases} \inf f(x) \\ C_E(x) + p_E = 0 \\ C_I(x) + p_I \leq 0 \end{cases}$$

A Orientation

- les multiplicateurs optimaux sont apparus

① dans les conditions d'optimalité

— par la relation $N(A)^{\perp} = R(A^T)$ (contr d'=)

— par Farkas $\{y \in F : A^* y \in K^+\}^+ = \overline{A(K)}$ (contr d'≤)

② comme "dérivée" de la fonction valeur en 0 (exercice III.7)

$$\lambda_* = \nabla v(0) \quad \text{(si ce gradient existe)}$$

On va voir aujourd'hui une 3e occurrence

③ λ_* est *parfois* solution d'un problème d'optimisation, appelé *problème dual*

(— il peut y en avoir plusieurs

— ces problèmes n'ont parfois pas d'expression analytique

- *Sujet difficile* — car les multiplicateurs sont des variables *cachées*, qui n'apparaissent pas dans le problème original, dont la signification n'est pas immédiate

— car les liens entre problèmes primal et dual ne se font que par l'intermédiaire de propriétés ; l'intuition est difficile.

- Sujet difficile mais intéressant

  - d'un point de vue théorique : apporte un éclairage, un autre point de vue, sur le problème primal ;

    → souvent utilisé pour obtenir des propriétés non triviales sur le problème primal

  - d'un point de vue algorithmique : permet de définir des algorithmes non triviaux dans lesquels on génère des multiplicateurs $(\lambda_k) \to \lambda_*$ optimal.

    exemple : l'algorithme des multiplicateurs

$$\lambda_{k+1} = \lambda_k + r\, c(x_k) \qquad (I = \phi)$$

[B] Programme de la séance

(1) introduire un problème dual associé à un problème d'optimisation peut se faire de différentes façons

+ dualité min-max
— dualité par perturbation (Rockafellar)
— dualité de Fenchel (problème convexe)
— dualité de Toland

(2) liens entre problème primal et dual

(3) Dualisation de contraintes fonctionnelles

Remarque préliminaire

Soit $f_0 : \mathbb{E} \to \mathbb{R}$ une fonction

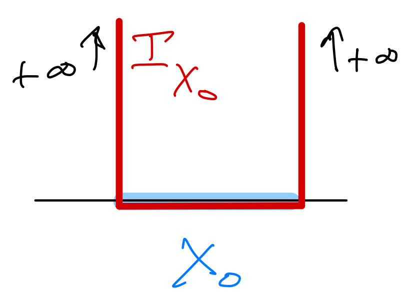
$X_0, X \subset \mathbb{E}$ des ensembles

$$(P_1) \quad \begin{cases} \inf f_0(x) \\ x \in X \cap X_0 \end{cases}$$

$$(P_2) \quad \begin{cases} \inf f_0(x) + I_{X_0}(x) \\ x \in X \end{cases}$$

où I_{X_0} est l'indicatrice de X_0 qui est la fonction $\mathbb{E} \to \mathbb{R} \cup \{+\infty\}$ définie par

$$I_{X_0}(x) = \begin{cases} 0 & \text{si } x \in X_0 \\ +\infty & \text{sinon} \end{cases}$$



On a $(P_1) \iff (P_2)$

- si $X_0 \cap X = \emptyset$, les 2 valeurs optimales $= +\infty$
- si $X_0 \cap X \neq \emptyset$, ―――――― sont identiques

2) Dualité min-max

(remonte à von Neumann, 1928)

A Introduction d'un problème dual

– On considère le problème

$$(P) \quad \inf_{x \in X} f(x) \quad = \quad val(P)$$

→ pas min car (P) peut ne pas avoir de solution

- X un ensemble quelconque (on veut garder la généralité aussi longtemps que possible)
- $f : X \to \overline{\mathbb{R}}$ → attention aux opérations qui n'ont pas de sens ($0 * \infty = ?$)

– Convention très importante ici

$$\inf_{x \in \emptyset} f(x) = +\infty$$

– exemple : $f(x) = +\infty$ permet de prendre en compte des contraintes

$$\begin{cases} \inf f_0(x) \\ c(x) = 0 \\ x \in X \end{cases} \iff \begin{cases} \inf \left(f(x) = f_0(x) + I_{X_E}(x) \right) \\ x \in X \end{cases}$$

où $X_E = \{x \in X : c(x) = 0\}$

indicatrice $I_{X_E}(x) = \begin{cases} 0 & \text{si } x \in X_E \\ +\infty & \text{sinon} \end{cases}$

en effet

- si $X \cap X_E \neq \emptyset$: les 2 problèmes sont équivalents (même valeur optimale, mêmes solutions)

- si $X \cap X_E = \emptyset$: les 2 valeurs optimales sont $+\infty$

– On suppose que l'on peut écrire $f(x)$ comme un supremum

$$f(x) = \sup_{y \in Y} \varphi(x, y)$$

où $\begin{cases} Y = \text{ensemble quelconque} \\ \varphi : X \times Y \to \overline{\mathbb{R}} \quad \text{fonction de couplage} \end{cases}$

$\begin{cases} \min f_0(x) \\ c(x) = 0 \\ x \in X \end{cases}$

$X_E = \{x \in E : c(x) = 0\}$

• dans l'exemple précédent

$$f(x) = f_0(x) + I_{X_E}(x)$$

$$= \sup_{y \in F} \underbrace{f_0(x) + \langle y, c(x) \rangle}_{\varphi(x,y)}$$

$$\left(\text{car } \sup_{y \in F} \langle y, c(x) \rangle = \begin{cases} +\infty & \text{si } c(x) \neq 0 \\ \cancel{0} & \text{sinon} \end{cases} \right.$$

$t \|c(x)\|^2$

$$\left. \text{on prend } y = t\, c(x),\ t \to \infty \right)$$

– le problème primal s'écrit donc

$$\boxed{(P) \quad \inf_{x \in X} \underbrace{\sup_{y \in Y} \varphi(x, y)}_{\text{c'est 1 fonction de } x}}$$

⚠ bien comprendre ce que cela veut dire

— Problème dual

On inverse l'inf et le sup :

$$\boxed{(D) \quad \sup_{y\in Y} \underbrace{\left(\inf_{x \in X} \varphi(x,y) \right)}_{= -\delta(y)}} = \mathrm{val}(D)$$

$\delta \equiv$ la fonction duale

Remarques

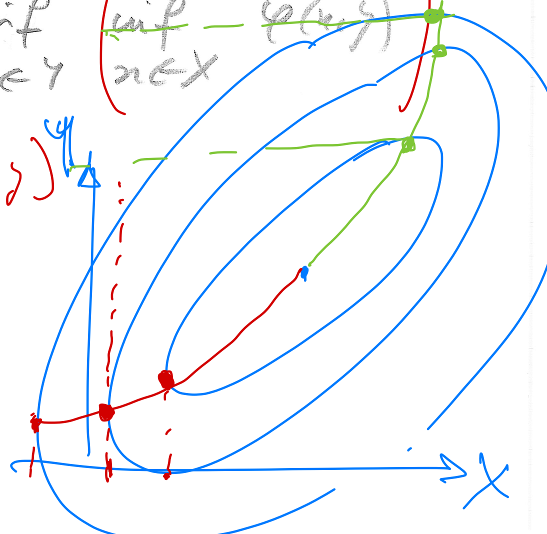
- (P) est un problème en x et (D) un problème en y $\to$ lien entre les 2 n'est pas trivial
- (P) et (D) ne sont pas nécessairement "équivalents" dans le sens où

$$\left.\begin{array}{l} \bar{y} \in \mathrm{sol}(D) \\ \bar{x} \in \operatorname{arg\,min}_{x\in X} \varphi(x,\bar{y}) \end{array}\right) \Rightarrow \bar{x} \in \mathrm{sol}(P)$$

- Pour avoir équivalence il est presque nécessaire d'avoir $\mathrm{val}(P) = \mathrm{val}(D)$, ce qui n'est pas assuré
- Il n'y a pas de problème avec deux inf

$$\inf_{x\in X} \left(\inf_{y\in Y} \varphi(x,y) \right) = \inf_{y\in Y} \left(\inf_{x\in X} \varphi(x,y) \right)$$

$$= \inf_{(x,y)\in X\times Y} \varphi(x,y)$$



B Liens entre problèmes primal et dual

① Dualité faible

On a toujours

$$\boxed{\mathrm{val}(D) \le \mathrm{val}(P)}$$

Démo

$$\sup_{y\in Y}\inf_{x\in X}\varphi(x,y) \le \inf_{x\in X}\underbrace{\sup_{y\in Y}\varphi(x,y) \quad \forall x_0\in X \ \forall y_0\in Y}_{\text{indép. de } y_0}$$

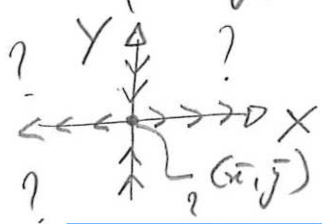
- Lien avec $\bigcup_j \bigcap_i A_{ij} \subset \bigcap_i \bigcup_j A_{ij}$

- On dit qu'il y a un saut de dualité si $\mathrm{val}(D) < \mathrm{val}(P)$.
  Alors le saut de dualité $= \mathrm{val}(P) - \mathrm{val}(D)$

② Dualité forte et point-selle

$(\bar{x},\bar{y})$ est un point-selle de φ sur $X\times Y$ si

$$\varphi(\bar{x},y) \le \varphi(\bar{x},\bar{y}) \le \varphi(x,\bar{y}), \quad \forall (x,y)\in X\times Y$$



on n'a de l'information que sur la croix

Théo

$$\boxed{\begin{array}{l} (\bar{x},\bar{y}) \text{ est un point-selle de } \varphi \\ \iff \quad \bar{x} \text{ est solution de } (P) \\ \qquad\ \ \bar{y} \text{ est solution de } (D) \\ \qquad\ \ \text{il n'y a pas de saut de dualité} \end{array}}$$

$$\underbrace{\inf_x \varphi(x,\bar y) \overset{②}{\leq} \sup_y \Big(\inf_x \varphi(x,y)\Big)}_{(D)} \overset{③}{\leq} \underbrace{\inf_x \Big(\sup_y \varphi(x,y)\Big)}_{(P)} \overset{①}{\leq} \sup_y \varphi(\bar x, y)$$

$$\varphi(\bar x, y) \overset{\forall y}{\leq} \varphi(\bar x,\bar y) \overset{\forall x}{\leq} \varphi(x,\bar y)$$

(⇒)

$$\inf_x \varphi(x,\bar y) = \varphi(\bar x,\bar y) \qquad \sup_y \varphi(\bar x, y) = \varphi(\bar x,\bar y)$$

$\Rightarrow$ on a egalité partout

$=$ en ① $\Rightarrow$ $\bar x \in \mathrm{Sol}\,(P)$

$=$ en ② $\Rightarrow$ $\bar y \in \mathrm{Sol}\,(D)$

$=$ en ③ $\Rightarrow$ il n'y a pas de saut de dualité

(⇐) $=$ en ①②③ $\Rightarrow$

$$\varphi(\bar x,\bar y) \geq \inf_x \varphi(x,\bar y) = \sup_y \varphi(\bar x, y) \geq \varphi(\bar x,\bar y)$$

$\Rightarrow$ $=$ partout

$$\varphi(\bar x, y) \leq \varphi(\bar x,\bar y) \leq \varphi(x,\bar y) \quad , \forall x \in X \ \forall y \in Y$$

$\Rightarrow (\bar x,\bar y) \in PS$ □

Coroll 1 (produit cartésien des points-selles)

Si φ a un point-selle

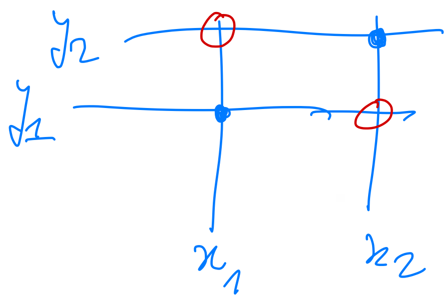
alors 1) l'ensemble des points-selles de φ est un produit cartésien $\bar{X} \times \bar{Y}$ avec $\bar{X} \subset X$ et $\bar{Y} \subset Y$

2) φ prend une valeur constante $\bar{\varphi}$ sur $\bar{X} \times \bar{Y}$

3) $\bar{X} = \bigcap_{y \in Y} \{x \in X : \varphi(x,y) \leq \bar{\varphi}\}$

$\bar{Y} = \bigcap_{x \in X} \{y \in Y : \varphi(x,y) \geq \bar{\varphi}\}$.

Dém 1)



Soient (x_1,y_1) et $(x_2,y_2) \in PS$

$\Rightarrow (x_1,y_2) \in PS$

$x_1 \in Sol(P)$
$y_2 \in Sol(D)$
Il n'y a pas de saut de dualité

$\Rightarrow (x_1,y_2) \in PS$

2)
3) à faire (voir le syllabus)

$\square$

2 <u>Coroll 2</u> : Comment retrouver une solution primale à partir d'une solution duale $\bar{y}$. On a bien envie de résoudre le problème

$$\inf_{x \in X} \varphi(x, \bar{y}) \qquad (P_{\bar{y}})$$

<u>si</u> • φ a un point-selle

• $\bar{y} \in Sol(D)$

<u>alors</u> $\emptyset \neq Sol(P) \subset \operatorname{Argmin}_{x \in X} \varphi(x, \bar{y})$

pas égalité !

Donc $(P_{\bar{y}})$ peut avoir des <u>solutions parasites</u> (non pertinentes, qui ne sont pas solutions de (P))

<u>Dém</u>

1) $\exists$ point-selle $(\hat{x}, \hat{y}) \Rightarrow \hat{x} \in Sol(P)$
et donc $Sol(P) \neq \emptyset$

2) Soit $\bar{x} \in Sol(P)$ $\overset{?}{\Rightarrow}$ $\bar{x} \in Sol(P_{\bar{y}})$
$\bar{y} \in Sol(D)$
pas de saut de dualité

$\Rightarrow (\bar{x}, \bar{y})$ est point-selle

$\Rightarrow \varphi(\bar{x}, \bar{y}) \leq \varphi(x, \bar{y}), \quad \forall x \in X$

$\Rightarrow \bar{x} \in Sol(P_{\bar{y}})$

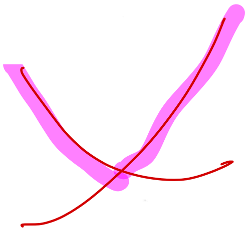
$\square$

C Schéma des algorithmes de dualité

Stratégie : générer une suite $(y_k) \subset Y$ qui résout (D) c-à-d qui minimise la fonction duale

$$y \longmapsto \delta(y) = - \inf_{x \in X} \varphi(x,y)$$

On passe de y_k à $y_{k+1} \in Y$ par



1) on résout le problème interne qui donne la valeur de la fonction duale

$$\delta(y_k) = - \inf_{x \in X} \varphi(x, y_k) \qquad (1)$$

2) on met à jour y_k en utilisant des informations issues de la résolution de (1) (voir plus loin)

$$y_{k+1} = y_k + \dots$$

Pour que cette stratégie soit intéressante il faut que

(D_1) (P) et (D) sont liés (en part, pas de saut de dualité)

(D_2) le problème interne primal (1) doit être beaucoup plus facile à résoudre que (P)

$\Rightarrow$ dualisation des contraintes gênantes (voir plus loin)

(D_3) le problème interne n'a pas de solutions parasites.

3) Analyse convexe

A Enveloppe supérieure de fonctions

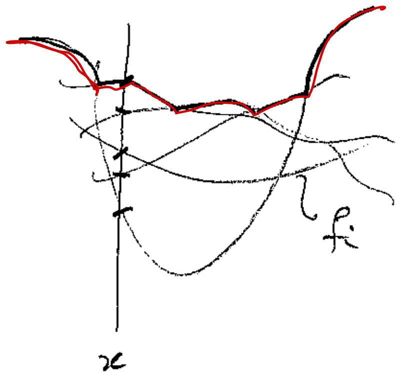
– Soit $(f_i)_{i\in I}$ une famille quelconque de fonctions $f_i : \mathbb{E} \to \overline{\mathbb{R}}$ ($\mathbb{E}$ est un espace vectoriel)

– L'enveloppe supérieure des f_i est la fonction

$$\left(\sup_{i\in I} f_i\right) : \mathbb{E} \to \overline{\mathbb{R}}$$

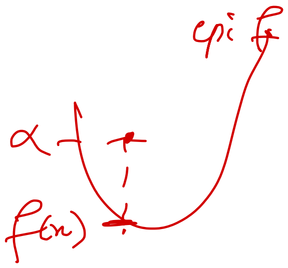
définie en $x \in \mathbb{E}$ par

$$\boxed{\left(\sup_{i\in I} f_i\right)(x) = \sup_{i\in I} f_i(x)}$$



– **Prop**

$$\boxed{\operatorname{epi}\left(\sup_{i\in I} f_i\right) = \bigcap_{i\in I} (\operatorname{epi} f_i)}$$



Dem $(x,\alpha) \in \operatorname{epi}\left(\sup_{i\in I} f_i\right)$

$\Leftrightarrow \left(\sup_{i\in I} f_i\right)(x) \leq \alpha$

$\Leftrightarrow \sup_{i\in I} [f_i(x)] \leq \alpha$

$\Leftrightarrow f_i(x) \leq \alpha, \quad \forall i \in I$

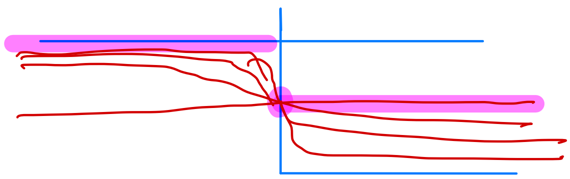
$\Leftrightarrow (x,\alpha) \in \operatorname{epi} f_i, \quad \forall i \in I$

$\Leftrightarrow (x,\alpha) \in \bigcap_{i\in I} (\operatorname{epi} f_i)$ □

– **Coroll**

f_i convexes	$\Rightarrow$	$\left(\sup_{i\in I} f_i\right)$	convexe
f_i fermées	$\Rightarrow$	$\left(\sup_{i\in I} f_i\right)$	fermée

l'enveloppe supérieure de fcts continues n'est pas continue (mais s.c.i) →



4) Dualisation de contraintes fonctionnelles

[A] Dualisation lagrangienne (φ = lagrangien)

$$(P)\ \begin{cases} \inf f(x) \\ c(x) \leq 0 \\ x \in X_0 \end{cases}$$

$f : \mathbb{R}^n \to \mathbb{R}$

$c : \mathbb{R}^n \to \mathbb{R}^m$

$x \in X_0 \to$ les contraintes "simples" qu'on ne veut pas dualiser

On prend $\varphi(x,y) = \ell(x,\lambda)$ $\quad X = \mathbb{R}^n,\ Y = \mathbb{R}^m$

$$= f(x) + \lambda^T c(x)$$

$\downarrow$ on dit que l'on dualise la contrainte $c(x) \leq 0$

$$\boxed{(P) \equiv \inf_{x \in X_0} \sup_{\lambda \geq 0} \ell(x,\lambda)}$$

$\sup_{\lambda \geq 0} \ell(x,\lambda) = f(x) + I_{\{c(x) \leq 0\}}(x)$

en effet

$$\sup_{\lambda \geq 0} \ell(x,\lambda) = \begin{cases} f(x) & \text{si } c(x) \leq 0 \\ +\infty & \text{sinon} \end{cases}$$

Donc :

$$\boxed{(D)\ \sup_{\lambda \geq 0} \overbrace{\inf_{x \in X_0} \ell(x,\lambda)}^{-\delta(\lambda)}} = -\inf_{\lambda \geq 0} \delta(\lambda)$$

$x \in X_0 \to$ on a besoin d'avoir X_0 "simple" car ce problème doit souvent être résolu

$\lambda \geq 0$: les contraintes compliquées "$c(x) \leq 0$" ont été remplacées par les contraintes simples $\lambda \geq 0$.

2 prop $\boxed{\delta \text{ est une fonction convexe et fermée}}$

en effet $\delta(\lambda) = -\inf_{x \in X_0} \underbrace{f(x) + \lambda c(x)}$ fct affines de λ donc convexes fermées $\square$

$= \sup_{x \in X_0} -f(x) - \lambda^T c(x)$

même si (P) n'est pas convexe !!

Prop (CS pour que la dualisation lagrangienne fonctionne)

> **Si** • f et les c_i sont convexes
> • $X_0 = \mathbb{R}^n$
> • $(\bar{x}, \bar{\lambda})$ vérifie (KKT)
>
> **alors** $(\bar{x}, \bar{\lambda})$ est un point selle de ℓ sur $\mathbb{R}^n \times \mathbb{R}^m_+$

⚠

Dem $\ell(\bar{x}, \lambda) \leq \ell(\bar{x}, \bar{\lambda}) \leq \ell(x, \bar{\lambda})$ $\forall x \in \mathbb{R}^n$, $\forall \lambda \geq 0$

1) $x \longmapsto \ell(x, \bar{\lambda})$ est convexe

$$\underbrace{f(x)}_{\text{convexe}} + \sum_{i=1}^{m} \underbrace{\bar{\lambda}_i}_{\geq 0} \underbrace{c_i(x)}_{\text{convexes}}$$

$\nabla_x \ell(\bar{x}, \bar{\lambda}) = 0 \Rightarrow \bar{x}$ minimise $\ell(\cdot, \bar{\lambda})$

2) $\ell(\bar{x}, \lambda) = f(\bar{x}) + \underbrace{\sum_{i=1}^{m} \underbrace{\lambda_i}_{\geq 0} \underbrace{c_i(\bar{x})}_{\leq 0}}_{\leq 0}$

$\forall \lambda \geq 0$

$\leq f(\bar{x})$

$= \ell(\bar{x}, \bar{\lambda})$ car $\bar{\lambda}^T c(\bar{x}) = 0$ par complémentarité

□

Algorithme associé (UZAWA) (relaxation lagrangienne)

On passe de λ_k à λ_{k+1} par

1) $x_k \in \arg\min_{x \in \mathbb{R}^n} \ell(x, \lambda_k)$

2) (x_k, λ_k) satisfaisant $\longrightarrow$ arrêt

3) $\lambda_{k+1} = [\lambda_k + \alpha_k\, c(x_k)]^{+}$ pour assurer $\lambda_{k+1} \geq 0$

(algo du gradient projeté)

Explication

- $-c(x_k)$ est une direction de descente de la fonction duale

Supposons que $\delta(\cdot) = -\inf_{x \in \mathbb{R}^n} \ell(x, \cdot)$ soit différentiable (ce n'est pas le cas !!).

et $\delta(\lambda_k) = -\ell(\xi(\lambda_k), \lambda_k)$

$$\rightarrow \delta'(\lambda_k)\cdot\mu = -\underbrace{\ell'_x(\xi(\lambda_k), \lambda_k)}_{=0} \cdot [\xi'(\lambda_k)\cdot\mu] - \underbrace{\ell'_\lambda(\xi(\lambda_k), \lambda_k)}_{= c(\xi(\lambda_k))} \cdot \mu = -\mu^T c(\xi(\lambda_k))$$

$$\rightarrow \nabla\delta(\lambda_k) = -c(x_k) \qquad (\xi(\lambda_k) = x_k)$$

- $\alpha_k > 0$ est un pas qui assure la décroissance de δ (point délicat)

- $[\cdot]^{+}$ vient de l'algorithme du gradient projeté (on s'assure que $\lambda_{k+1} \geq 0$)

$$\boxed{-\ c\left(\operatorname{Argmin}_{x \in X} \ell(x,d)\right) \subset \partial\delta(d)}$$

X quelconque.

<u>Dém</u> (à faire lorsqu'on aura vu le sous-différentiel d'une fonction convexe)

B Dualisation lagrangienne augmentée (pas vu)

(φ = lagrangien augmenté)

$$(P)\ \begin{cases} \inf f(x) \\ c(x) = 0 \end{cases}$$

$f : \mathbb{R}^n \to \mathbb{R}$

$c : \mathbb{R}^n \to \mathbb{R}^m$

On prend $\varphi(x, y) \equiv \ell_r(x, \lambda)$ $\quad X = \mathbb{R}^n,\ Y = \mathbb{R}^m$

où $\boxed{\ell_r(x,\lambda) = f(x) + \lambda^T c(x) + \frac{r}{2} \| c(x) \|_2^2}$

$$\boxed{(P) \equiv \inf_{x \in \mathbb{R}^n} \ \sup_{\lambda \in \mathbb{R}^m} \ \ell_r(x, \lambda)}$$

en effet $\displaystyle\sup_{\lambda \in \mathbb{R}^m} \ell_r(x,\lambda) = \begin{cases} f(x) & \text{si } c(x) = 0 \\ +\infty & \text{sinon} \end{cases}$

$$\boxed{(D)\ \sup_{\lambda \in \mathbb{R}^m} \ \underbrace{\inf_{x \in \mathbb{R}^n} \ \ell_r(x, \lambda)}_{-\delta_r(\lambda)}}$$

pas de signe sur λ car contraintes d'égalité

prop $\boxed{\delta_r \text{ est une fonction convexe fermée}}$

(même démonstration que pour δ)

! même si le problème (P) n'est pas convexe

<u>Prop</u> (CS pour que la dualisation lagrangienne augmentée fonctionne ; <u>plus besoin de convexité !!</u>)

<u>si</u> • $(\bar{x}, \bar{\lambda})$ vérifie (KKT)

• CS 2

<u>alors</u> $\exists$ voisinage V de $\bar{x}$, $\exists \underline{r} > 0$, $\forall r \geq \underline{r}$,
$(\bar{x}, \bar{\lambda})$ est point-selle de ℓ_r sur $V \times \mathbb{R}^m$

pas sur $\mathbb{R}^n$ à cause de la non-convexité

pas $\mathbb{R}^m_+$ car uniquement des égalités.

Algorithme associé (méthode des multiplicateurs)

On passe de (λ_k, r_k) à (λ_{k+1}, r_{k+1}) par

1) $x_k \in \underset{x \in \mathbb{R}^n}{\arg\min}\ \ell_{r_k}(x, \lambda_k)$

2) si (x_k, λ_k) satisfaisant $\longrightarrow$ arrêt

3) $\boxed{\lambda_{k+1} = \lambda_k + r_k\, c(x_k)}$

4) $r_{k+1} \nearrow$ si nécessaire (heuristique)

Montrons que $\boxed{-c(x_k) = \nabla \delta_{r_k}(\lambda_k)}$ si δ_{r_k} différentiable

(donc l'algorithme est une méthode de gradient sur ℓ_r)

$$
\begin{aligned}
-\delta_{r_k}(\lambda_k+\mu) &= \inf_x \ell_{r_k}(x, \lambda_k+\mu) \\
&\leq \ell_{r_k}(x_k, \lambda_k+\mu) \\
&= \ell_{r_k}(x_k, \lambda_k) + \mu^T c(x_k) \\
&= -\delta_{r_k}(\lambda_k) + \mu^T c(x_k) \\
&= -\delta_{r_k}(\lambda_k) - \nabla\delta_{r_k}(\lambda_k)^T\mu + o(\|\mu\|)
\end{aligned}
$$

$\Rightarrow \nabla\delta_{r_k}(\lambda_k)^T\mu + o(\|\mu\|) \geq -\mu^T c(x_k)$

$\mu \rightsquigarrow t\mu$ et $t \downarrow 0$ donnent

$\nabla\delta_{r_k}(\lambda_k)^T\mu \geq -\mu^T c(x_k)$, $\forall \mu$

linéarité $\Rightarrow \nabla\delta_{r_k}(\lambda) = -c(x_k)$

$\square$