Théorème d'Euler/Péno/Kantorovitch, hypothèse de qualification et preuve du théorème de KKT

Guillaume Lecué[1]

Le but de cette section est de présenter les outils de géométrie différentielle qu'on utilise pour démontrer des théorèmes d'optimisation tels que celui des extrema liés ou celui de KKT. On commence par un résultat très général, celui de Euler/Péano/Kantorovitch qui permet de retrouver tous les autres théorèmes en optimisation avec ou sans contrainte.

1 Approximation du premier ordre d'un sous-ensemble de $\mathbb{R}^n$

But : Un problème d'optimisation implique deux objets mathématiques : la fonction objectif f et la contrainte K. Dans le deuxième chapitre de ce cours, on a donné une approximation locale d'une fonction par une fonction affine. Le gradient est l'outil clef pour construire cette approximation. Cela nous sera utile pour approcher f localement. Dans cette section, on construit une approximation locale de l'ensemble K. Décrire localement f et K va permettre d'identifier des conditions (du premier ordre) nécessaires d'optimalité d'un point $x^* \in K$ pour le problème $\min_{x\in K} f(x)$ (voir Figure 2).

On commence par donner une description locale d'une contrainte K. La question qu'on souhaite résoudre ici est de savoir à quoi ressemble K si on regarde cet ensemble à partir de l'un de ces points $x \in K$. Pour cela, on rappelle la notion de vecteur tangent à une surface.

Définition 1.1 *Soit K un sous-ensemble de $\mathbb{R}^n$ non vide. Soit $x \in K$. On dit que $v \in \mathbb{R}^n$ est **tangent à** K **en** x quand il existe deux suites $(\lambda_k)_k \subset \mathbb{R}_+^*$ et $(v_k)_k \subset \mathbb{R}^n$ telles que $(\lambda_k)_k \downarrow 0$, $x + \lambda_k v_k \in K$ et $v = \lim_k v_k$.*

L'ensemble des vecteurs tangents à K en x est ce à quoi ressemble K quand on regarde K en se plaçant en x. C'est la raison pour laquelle, on a longtemps cru que la terre était plate ; en effet, quand on se place à la surface de la terre (comme nous), la terre ressemble à une vaste étendue plate (c'est parce que l'ensemble des vecteurs tangent à une boule en un point de sa surface est un demi-espace). C'est donc cet ensemble qu'on va utiliser pour décrire localement une contrainte K. Il se trouve que cet ensemble à une propriété algébrique : il est positivement homogène, càd si x est dans cet ensemble alors λx aussi pour tout $\lambda \geq 0$. Un ensemble ayant cette propriété est appelé un **cône**. On va alors introduire la notion suivante.

Définition 1.2 *Soit K un sous-ensemble de $\mathbb{R}^n$ non vide. Soit $x \in K$. Le **cône tangent à** K **en** x, noté $T_K(x)$, est l'ensemble de tous les vecteurs tangents à K en x. En d'autres termes, $T_K(x)$ est défini par*

$$T_K(x) = \left\{ v \in \mathbb{R}^n : \exists (\lambda_k)_k \subset \mathbb{R}_+^*, (v_k)_k \subset \mathbb{R}^n \text{ tel que } (\lambda_k)_k \downarrow 0, x + \lambda_k v_k \in K \text{ et } v = \lim_k v_k \right\}.$$

1. CREST, ENSAE. Bureau 3029, 5 avenue Henry Le Chatelier. 91 120 Palaiseau. Email : guillaume.lecue@ensae.fr.

On montre d'abord que $T_K(x)$ est bien un cône. On montre aussi qu'il est fermé.

Proposition 1.3 *Soit $K \subset \mathbb{R}^n$ non vide et $x \in K$. Alors $T_K(x)$ est un cône fermé.*

Preuve. On montre d'abord que $T_K(x)$ est bien un cône : soit $v \in T_K(x)$ et $\lambda \geq 0$, montrons que $\lambda v \in T_K(x)$. Si $\lambda = 0$, on a bien $\lambda v = 0 \in T_K(x)$. Sinon, on écrit $v = \lim_k v_k$ où $(v_k)_k \subset \mathbb{R}^n$ tel que $x + \lambda_k v_k \in K$ et $(\lambda_k)_k \downarrow 0$. On a alors, $\lambda v = \lim_k \lambda v_k$ et $x + (\lambda_k/\lambda)\lambda v_k \in K$ et $(\lambda_k/\lambda)_k \downarrow 0$. Donc $\lambda v \in T_K(x)$.

Montrons que $T_K(x)$ est fermé. Soit $(v^{(k)})_k \subset T_K(x)$ une sous-suite convergeant vers un point $v \in \mathbb{R}^n$. Montrons que $v \in T_K(x)$. Pour tout $k \in \mathbb{N}$, non note $(\lambda_m^{(k)})_m \subset \mathbb{R}^n$ et $(v_m^{(k)})_m \subset \mathbb{R}^n$ tel que $v^{(k)} = \lim_m v_m^{(k)}$, $(\lambda_m^{(k)})_m \downarrow 0$ et $x + \lambda_m^{(k)} v_m^{(k)} \in K$ pour tout $m \in \mathbb{N}$. On va faire une méthode d'extraction triangulaire itérative : on suppose construits $k_{\ell-1}$ et $m_{\ell-1}$ à l'étape $\ell \in \mathbb{N}^*$, on note $k_\ell \in \mathbb{N}$, tel que $\left\| v^{(k_\ell)} - v \right\|_2 \leq 1/\ell$ et $k_\ell > k_{\ell-1}$ et $m_\ell > m_{\ell-1}$ tel que $\left\| v_{m_\ell}^{(k_\ell)} - v^{(k_\ell)} \right\|_2 \leq 1/\ell$ et $\lambda_{m_\ell}^{(k_\ell)} \leq \min(\lambda_{m_{\ell-1}}^{(k_{\ell-1})}, 1/\ell)$. On a alors $(\lambda_{m_\ell}^{(k_\ell)})_\ell \downarrow 0$, $x + \lambda_{m_\ell}^{(k_\ell)} v_{m_\ell}^{(k_\ell)} \in K$ et

$$v = \lim_\ell v_{m_\ell}^{(k_\ell)}.$$

On a donc bien $v \in T_K(x)$. ■

Exemples : On donne quelques exemples de cônes tangents à un sous-ensemble $K \subset \mathbb{R}^n$ non vide :

1) si $K = \mathbb{R}^n$ alors $T_K(a) = \mathbb{R}^n$ pour tout $a \in K$,
2) si $K = \{x \in \mathbb{R}^n : \langle x, w \rangle + b \leq 0\}$ alors $T_K(a) = \{x \in \mathbb{R}^n : \langle x, w \rangle \leq 0\}$ quand $\langle a, w \rangle + b = 0$ et $T_K(a) = \mathbb{R}^n$ si $\langle a, w \rangle + b < 0$,
3) si $K = \{x \in \mathbb{R}^n : \langle x, w \rangle + b = 0\}$ alors $T_K(a) = \{x \in \mathbb{R}^n : \langle x, w \rangle = 0\}$ pour tout $a \in K$,
4) si $K = \{(x, y) \in \mathbb{R}^2 : x \leq 0, y \leq 0\}$ alors pour $a = (x, y)$, $T_K(a) = \mathbb{R}^n$ si $x < 0$ et $y < 0$, $T_K(a) = \{(x, y) \in \mathbb{R}^2 : x \leq 0\}$ si $x = 0$ et $y < 0$, $T_K(a) = \{(x, y) \in \mathbb{R}^2 : y \leq 0\}$ si $x < 0$ et $y = 0$ et $T_K(a) = \{(x, y) \in \mathbb{R}^2 : x \leq 0, y \leq 0\}$ si $x = 0$ et $y = 0$.
5) si $K = \{a\}$ alors $T_K(a) = \{0\}$.

On peut aussi faire le lien entre cône tangent et gradient d'une fonction quand on regarde les lignes et ensemble de niveau de cette fonction et de sa meilleure approximation au premier ordre. Idéalement, on voudrait que le cône tangent à $\mathcal{L}_f(f(x)) = \{y \in U : f(y) = f(x)\}$ ou à $\mathcal{L}_f(\leq f(x)) = \{y \in U : f(y) \leq f(x)\}$ en x soit égal au cône tangent à $\mathcal{L}_{F_x}(f(x))$ ou à celui de $\mathcal{L}_{F_x}(\leq f(x))$ en x. Ici, on note par F_x la meilleure approximation affine de f en x donnée par

$$F_x : v \in U \to f(x) + \langle \nabla f(x), v - x \rangle. \tag{1.1}$$

On a en effet, quand $h \to 0$, $f(x+h) = F(x+h) + o(\|h\|)$ et F est affine. On commence par donner les cônes tangents à $\mathcal{L}_{F_x}(f(x))$ et $\mathcal{L}_{F_x}(\leq f(x))$ en x.

Proposition 1.4 *Soit U un ouvert de $\mathbb{R}^n$ et $f : U \to \mathbb{R}$. Soit $x \in U$. On suppose que f est différentiable en x. On considère l'approximation affine F_x de f en x. Les cônes tangents à la ligne et l'ensemble de niveau $f(x)$ de f en x sont*

$$T_{\mathcal{L}_{F_x}(f(x))} = \{y \in \mathbb{R}^n : \langle y, \nabla f(x) \rangle = 0\} = \mathrm{vect}(\nabla f(x))^\perp$$

et

$$T_{\mathcal{L}_{F_x}(\leq f(x))} = \{y \in \mathbb{R}^n : \langle y, \nabla f(x) \rangle \leq 0\}.$$

Preuve. Par définition de F_x et de sa ligne et de son ensemble de niveau $f(x)$, on voit que

$$\mathcal{L}_{F_x}(f(x)) = \{y \in U : F_x(y) = f(x)\} = \big\{y \in U : \big\langle \nabla f(x), y - x\big\rangle = 0\big\} = \Big(x + \mathrm{vect}(\nabla f(x))^{\perp}\Big) \cap U$$

et

$$\mathcal{L}_{F_x}(\leq f(x)) = \{y \in U : F_x(y) \leq f(x)\} = \big\{y \in U : \big\langle \nabla f(x), y - x\big\rangle \leq 0\big\}.$$

Comme U est un ouvert et $x \in U$, on a $B_2^d(x, \epsilon) \subset U$ pour un certain $\epsilon > 0$ (où $B_2^d(x, \epsilon) = \{y \in \mathbb{R}^n : \|x - y\|_2 \leq \epsilon\}$). Comme la notion de cône tangent à un sous-ensemble en un point est une notion locale, on voit que le cône tangent à $\mathcal{L}_{F_x}(f(x))$ en x coïncide avec celui de l'espace affine $x + \mathrm{vect}(\nabla f(x))^{\perp}$, c'est donc l'espace vectoriel qui le dirige càd $\mathrm{vect}(\nabla f(x))^{\perp}$.

Le cône tangent à l'ensemble de niveau $\mathcal{L}_{F_x}(\leq f(x))$ est le demi-espace $\{y : \big\langle \nabla f(x), y\big\rangle \leq 0\}$ car comme U est ouvert et $x \in U$, que localement $\mathcal{L}_{F_x}(\leq f(x))$ coïncide avec $x + \{y : \big\langle \nabla f(x), y\big\rangle \leq 0\}$ et donc leurs cônes tangents sont égaux. Celui du demi-espace affine $x + \{y : \big\langle \nabla f(x), y\big\rangle \leq 0\}$ est donné par son demi-espace linéaire qui le dirige, càd $\{y : \big\langle \nabla f(x), y\big\rangle \leq 0\}$. ∎

La Proposition 1.4 donne les descriptions locales au premier ordre de la ligne de niveau $\mathcal{L}_{F_x}(f(x))$ et de l'ensemble de niveau $\mathcal{L}_{F_x}(\leq f(x))$. Comme F_x est une application affine (restreinte à U), ces lignes de niveau et ensemble de niveaux sont des respectivement des espace affines et demi-espace affines (restreints à U). Ce sont donc des objets déjà affines. Donc leurs descriptions locales coïncident avec les ensembles qu'ils décrivent (à la translation en 0 près). Idéalement, on aimerait pouvoir retrouver ces cônes (obtenus pour F_x) pour les lignes et ensembles de niveau de f. On commence par étudier le cône tangent à la ligne de niveau $f(x)$ de f en x.

Proposition 1.5 *Soit U un ouvert de $\mathbb{R}^n$ et $f : U \to \mathbb{R}$. Soit $x \in U$. On suppose que f est différentiable en x. On a*

$$T_{\mathcal{L}_f(f(x))}(x) \subset \mathrm{vect}(\nabla f(x))^{\perp}.$$

Preuve. On note $S = \mathcal{L}_f(f(x))$ (on vérifie bien que $x \in S$). Soit v un vecteur tangent à S en x. Montrons que $\big\langle v, \nabla f(x)\big\rangle = 0$. Si $v = 0$ c'est bien vérifié. On suppose que $v \neq 0$. On note $(v_k)_k \subset \mathbb{R}^n$ tel que $v_k \to v$ et $(\lambda_k)_k \subset \mathbb{R}_+^*$ tel que $\lambda_k \downarrow 0$ et $x_k := x + \lambda_k v_k \in S$. Quand $k \to \infty$, on a $x_k \to x$ (car $\lambda_k v_k \to 0$) et donc

$$0 = f(x_k) - f(x) = \big\langle \nabla f(x), x_k - x\big\rangle + o(\|x_k - x\|_2) = \|x_k - x\|_2 \left(\Big\langle \nabla f(x), \frac{v_k}{\|v_k\|_2}\Big\rangle + o(1)\right).$$

Alors, quand $k \to \infty$, $\big\langle \nabla f(x), \frac{v_k}{\|v_k\|_2}\big\rangle + o(1) = 0$. Par, ailleurs, la continuité de la norme $\|\cdot\|_2$ nous dit que $\|v_k\|_2 \to \|v\|_2 \neq 0$. Alors pour tout k assez grand, $\|v_k\|_2 \geq \epsilon$ pour $\epsilon > 0$. En en déduit que $\big\langle \nabla f(x), v_k\big\rangle \to 0$ et donc

$$\big\langle \nabla f(x), v\big\rangle = \lim_k \big\langle \nabla f(x), v_k\big\rangle = 0.$$

Donc $T_S(x) \subset \mathrm{vect}(\nabla f(x))^{\perp}$. ∎

Réciproquement, on aimerait pouvoir décrire exactement et simplement $T_S(x)$ par $\mathrm{vect}(\nabla f(x))^{\perp}$ comme on a pu le faire pour F_x dans la Proposition 1.4. Mais, on n'a pas toujours $T_{\mathcal{L}_f(f(x))}(x) = \mathrm{vect}(\nabla f(x))^{\perp}$. On peut par exemple le voir avec la fonction $f : (x, y) \to x^2 + y^2$. On a $S = \mathcal{L}_f(0) = \{(0, 0)\}$ alors $T_S((0, 0)) = \{(0, 0)\}$ mais $\nabla f((0, 0)) = (0, 0)$ donc $\mathrm{vect}(\nabla f(x))^{\perp} = \mathbb{R}^2$; on a donc $T_S((0, 0)) \neq \mathrm{vect}(\nabla f(x))^{\perp}$. Pour avoir $T_S(x) = \mathrm{vect}(\nabla f(x))^{\perp}$, il faut faire l'hypothèse que l'ensemble $S = \{x : f(x) = f(x^*)\}$ est **qualifiée en** x (on verra en détails cette hypothèse plus tard). C'est le but de l'hypothèse de qualification (et même sa définition) de donner une description simple du cône tangent à une contrainte. Pour le moment, on ne souhaite pas devoir faire ce type d'hypothèse portant sur f. Pour 'éviter' cette hypothèse, on travail plutôt avec l'ensemble de niveau $f(x)$ plutôt que la ligne de niveau $f(x)$. En effet, dans ce cas, on retrouve bien une description simple du cône tangent à cet ensemble comme on le montre dans la proposition suivante.

Proposition 1.6 *Soit U un ouvert de $\mathbb{R}^n$ et $f : U \to \mathbb{R}$. Soit $x \in U$. On suppose que f est différentiable en x. On a*

$$T_{\mathcal{L}_f(\leq f(x))}(x) = \{y \in \mathbb{R}^n : \left\langle y, \nabla f(x) \right\rangle \leq 0\}.$$

Preuve. On note $S = \mathcal{L}_f(\leq f(x))$ (on vérifie bien que $x \in S$). Soit v un vecteur tangent à S en x. Montrons que $\left\langle v, \nabla f(x) \right\rangle \leq 0$. Si $v = 0$ c'est bien vérifié. On suppose que $v \neq 0$. On note $(v_k)_k \subset \mathbb{R}^n$ tel que $v_k \to v$ et $(\lambda_k)_k \subset \mathbb{R}_+^*$ tel que $\lambda_k \downarrow 0$ et $x_k := x + \lambda_k v_k \in S$. Quand $k \to \infty$, on a $x_k \to x$ (car $\lambda_k v_k \to 0$) et donc

$$0 \leq f(x_k) - f(x) = \left\langle \nabla f(x), x_k - x \right\rangle + o(\|x_k - x\|_2) = \|x_k - x\|_2 \left(\left\langle \nabla f(x), \frac{v_k}{\|v_k\|_2} \right\rangle + o(1) \right).$$

Alors, quand $k \to \infty$, $\left\langle \nabla f(x), \frac{v_k}{\|v_k\|_2} \right\rangle + o(1) \geq 0$. Par, ailleurs, la continuité de la norme $\|\cdot\|_2$ nous dit que $\|v_k\|_2 \to \|v\|_2 \neq 0$. Alors pour tout k assez grand, $\|v_k\|_2 \geq \epsilon$ pour $\epsilon > 0$. En en déduit que

$$\left\langle \nabla f(x), v \right\rangle = \lim_k \left\langle \nabla f(x), v_k \right\rangle \geq 0.$$

Donc $T_S(x) \subset \{y \in \mathbb{R}^n : \left\langle y, \nabla f(x) \right\rangle \leq 0\}$.

Pour la réciproque, on prends un vecteur $y \in \mathbb{R}^n$ tel que $\left\langle y, \nabla f(x) \right\rangle \leq 0$. Comme on sait que les cônes tangents sont fermés, il suffit de prendre y tel que $\left\langle y, \nabla f(x) \right\rangle < 0$. On veut montrer que y est un vecteur tangent à $\mathcal{L}_f(\leq f(x))$ en x. On prend $\lambda_k = 1/k$ et $v_k = y$. Montrons que pour tout k assez grand on a $x + \lambda_k v_k \in \mathcal{L}_f(\leq f(x))$. Cela impliquera le résultat car $v_k \to y$. On veut donc montrer que $f(x + \lambda_k y) \leq f(x)$ pour k assez grand. On a quand $k \to \infty$,

$$\frac{f(x + \lambda_k y) - f(x)}{\lambda_k} = \left\langle y, \nabla f(x) \right\rangle + o(1).$$

Alors, comme $\left\langle y, \nabla f(x) \right\rangle < 0$, pour k assez grand, on a bien $f(x + \lambda_k y) \leq f(x)$. On a donc bien y qui est tangent.

Par donner précisément l'argument de fermeture qu'on utilise ici, on observe qu'on a montré

$$\{y \in \mathbb{R}^n : \left\langle y, \nabla f(x) \right\rangle < 0\} \subset T_{\mathcal{L}_f(\leq f(x))}(x).$$

En passant à la fermeture dans cette inclusion et vue que $T_{\mathcal{L}_f(\leq f(x))}(x)$ est fermé, on a obtient bien la réciproque. ■

Contrairement à la Proposition 1.5 qui ne permet pas de décrire le cône tangent à $\mathcal{L}_f(f(x))$ exactement en fonction de son cône linéaire $\mathrm{vect}(\nabla f(x))^\perp$ (càd le cône qu'on obtient quand on change f en sa meilleure approximation affine F_x en x), dans le cas des ensembles de niveaux le cône tangent à $\mathcal{L}_f(\leq f(x))$ coïncide avec celui de $\mathcal{L}_{F_x}(\leq f(x))$, càd avec son cône linéaire $\{y \in \mathbb{R}^n : \left\langle y, \nabla f(x) \right\rangle \leq 0\}$. C'est ce que nous apprend la Proposition 1.6. Ce qui est une bonne chose car cela nous évitera une hypothèse de qualification supplémentaire dans les théorèmes comme KKT (qu'on verra plus loin). Au-delà de l'aspect purement mathématique, les deux propositions 1.5 et 1.6, nous montrent que la bonne manière géométrique de voir les résultats d'optimalité en optimisation est que les points x^* solutions du problème d'optimisation sont ceux pour lesquels l'ensemble (et non pas la ligne) de niveau $f(x^*)$ rencontre la contrainte K pour la première fois. Cette rencontre de $\mathcal{L}_f(\leq f(x^*))$ avec K aura des implications sur les descriptions au premier ordre de $\mathcal{L}_f(\leq f(x^*))$ (donnée exactement par son cône linéaire dans Proposition 1.6) et celle de K. Malheureusement, décrire localement K au premier ordre nécessitera cette hypothèse supplémentaire de qualification mais cette fois uniquement sur K car comme on l'a déjà vu pour $\mathcal{L}_f(f(x))$ dans la Proposition 1.5, le cône tangent ne se décrit pas toujours facilement à l'aide du cône linéaire.

On finit sur les cônes tangents avec une propriété des cônes tangents à un ensemble K convexe qui nous sera utile en (OCD).

Proposition 1.7 *Soit K un ensemble convexe non vide de $\mathbb{R}^n$. Soit $x \in K$. Le cône tangent à K en x est donné par*
$$T_K(x) = \overline{\{\lambda(y-x) : y \in K, \lambda \geq 0\}}.$$

Preuve. Soit $y \in K$ et $\lambda \geq 0$. Montrons que $\lambda(y-x) \in T_K(x)$. Pour tout $k \in \mathbb{N}$ tel que $k \geq \lambda$, on a
$$x + \frac{1}{k}\lambda(y-x) = \left(1 - \frac{\lambda}{k}\right)x + \left(\frac{\lambda}{k}\right)y \in K$$
car $0 \leq \lambda/k \leq 1$ et K est convexe. Donc $\lambda(y-x) \in T_K(x)$. Par ailleurs, $T_K(x)$ est un ensemble fermé (voir Proposition 1.3), donc
$$\overline{\{\lambda(y-x) : y \in K, \lambda \geq 0\}} \subset T_K(x).$$

Montrons maintenant l'autre inclusion. En fait, l'autre inclusion est vraie même quand K n'est pas convexe. Soit v un vecteur tangent à K en x. On veut montrer que $v \in \overline{\{\lambda(y-x) : y \in K, \lambda \geq 0\}}$. Soient $(\lambda_k)_k \subset \mathbb{R}_+^*$ et $(v_k)_k \subset \mathbb{R}^n$ telles que $(\lambda_k)_k \downarrow 0$, $x + \lambda_k v_k \in K$ et $v = \lim_k v_k$. On a
$$v = \lim_k \lambda_k(x_k - x) \text{ où } x_k = x + \lambda_k v_k$$
et $x_k \in K$ pour tout $k \in \mathbb{N}$. Alors la suite $(\lambda_k(x_k - x))_k$ est une suite de $\{\lambda(y-x) : y \in K, \lambda \geq 0\}$ qui converge vers v. Donc, $v \in \overline{\{\lambda(y-x) : y \in K, \lambda \geq 0\}}$. ■

Il existe une manière **duale** de décrire un cône. On utilise cette notion de cône dual ici car le cône tangent de $\mathcal{L}_f(\leq f(x^*))$ – l'ensemble de niveau $f(x^*)$ de f – est $\{y : \langle y, \nabla f(x^*)\rangle \leq 0\}$ (voir Proposition 1.6). On peut donc décrire simplement ce cône tangent par son vecteur normal qui est donné ici par $\nabla f(x^*)$. Or passer d'un demi-espace (ici $\{y : \langle y, \nabla f(x^*)\rangle \leq 0\}$) à son vecteur normal (ici $\nabla f(x^*)$) est un argument de dualité. On va donc préférer décrire localement une fonction par son gradient, ça à dire par le vecteur normal à son cône tangent plutôt que par son cône tangent dont la description est moins aisée. On va faire de même pour la contrainte K : au lieu de décrire localement une contrainte par son cône tangent, on va le décrire par le cône dual de son cône tangent. On introduit cette notion maintenant de cône dual maintenant.

Définition 1.8 *Soit T un cône de $\mathbb{R}^n$. On appelle* ***cône dual à*** *T le cône*
$$T^\circ = \{z \in \mathbb{R}^n : \langle z, v\rangle \leq 0, \forall v \in T\}.$$
Si K est un sous-ensemble non-vide de $\mathbb{R}^n$ et $x \in K$, le ***cône normal à*** *K* ***en*** *x est noté $N_K(x)$, c'est le cône dual de $T_K(x)$. Il est donc défini par*
$$N_K(x) = (T_K(x))^\circ = \{z \in \mathbb{R}^n : \langle z, v\rangle \leq 0, \forall v \in T_K(x)\}.$$

On vérifie d'abord que le cône dual à T est bien un cône et aussi qu'il est fermé.

Proposition 1.9 *Soit T un cône de $\mathbb{R}^n$. Le cône dual T° est un cône convexe fermé et convexe.*

Preuve. On montre que T° est un cône. Si $z \in \mathbb{R}^n$ est tel que $\big\langle z, v\big\rangle \leq 0$ pour tout $v \in T$ alors $\big\langle \lambda z, v\big\rangle \leq 0$ pour tout $v \in T$ et $\lambda \geq 0$. Donc $\lambda z \in T^\circ$ et donc T° est bien un cône.

Montrons que T° est convexe. Si $z_1, z_2 \in T^\circ$ et $0 \leq \lambda \leq 1$ alors pour tout $v \in T$, on a $\big\langle \lambda z_1 + (1-\lambda) z_2, v\big\rangle = \lambda\big\langle z_1, v\big\rangle + (1-\lambda)\big\langle z_2, v\big\rangle \leq 0$. Donc $\lambda z_1 + (1-\lambda) z_2 \in T^\circ$.

Montrons que T° est un fermé de $\mathbb{R}^n$. Soit $(z_k)_k$ une suite convergente de T° de limite notée $z \in \mathbb{R}^n$. Montrons que $z \in T^\circ$. Soit $v \in T$, on a pour tout k, $\big\langle z_k, v\big\rangle \leq 0$ et donc en passant à la limite, on a $\big\langle z, v\big\rangle \leq 0$. Ceci étant vrai pour tout $v \in T$, on a bien $z \in T^\circ$. ■

On donne quelques exemples de cônes duaux en reprenant les exemples de cônes tangents ci-dessus (voir aussi la Figure 1) :

1) si $K = \mathbb{R}^n$ alors $T_K(a) = \mathbb{R}^n$ pour tout $a \in K$ et $N_K(a) = \{0\}$
2) si $K = \{x \in \mathbb{R}^n : \big\langle x, w\big\rangle + b \leq 0\}$ alors $T_K(a) = \{x \in \mathbb{R}^n : \big\langle x, w\big\rangle \leq 0\}$ et $N_K(a) = \{\lambda w : \lambda \geq 0\}$ quand $\big\langle a, w\big\rangle + b = 0$ et $T_K(a) = \mathbb{R}^n$ et $N_K(a) = \{0\}$ si $\big\langle a, w\big\rangle + b < 0$,
3) si $K = \{x \in \mathbb{R}^n : \big\langle x, w\big\rangle + b = 0\}$ alors $T_K(a) = \{x \in \mathbb{R}^n : \big\langle x, w\big\rangle = 0\}$ et $N_K(a) = \mathrm{vect}(w)$ pour tout $a \in K$,
4) si $K = \{(x, y) \in \mathbb{R}^2 : x \leq 0, y \leq 0\}$ alors pour $a = (x, y)$, $T_K(a) = \mathbb{R}^n$ et $N_K(a) = \{0\}$ si $x < 0$ et $y < 0$, $T_K(a) = \{(x, y) \in \mathbb{R}^2 : x \leq 0\}$ et $N_K(a) = \{(\lambda, 0) : \lambda \geq 0\}$ si $x = 0$ et $y < 0$, $T_K(a) = \{(x, y) \in \mathbb{R}^2 : y \leq 0\}$ et $N_K(a) = \{(0, \lambda) : \lambda \geq 0\}$ si $x < 0$ et $y = 0$ et $T_K(a) = \{(x, y) \in \mathbb{R}^2 : x \leq 0, y \leq 0\}$ et $N_K(a) = \{(\mu_1, \mu_2) : \mu_1 \geq 0, \mu_2 \geq 0\}$ si $x = 0$ et $y = 0$.
5) si $K = \{a\}$ alors $T_K(a) = \{0\}$ et $N_K(a) = \mathbb{R}^n$.

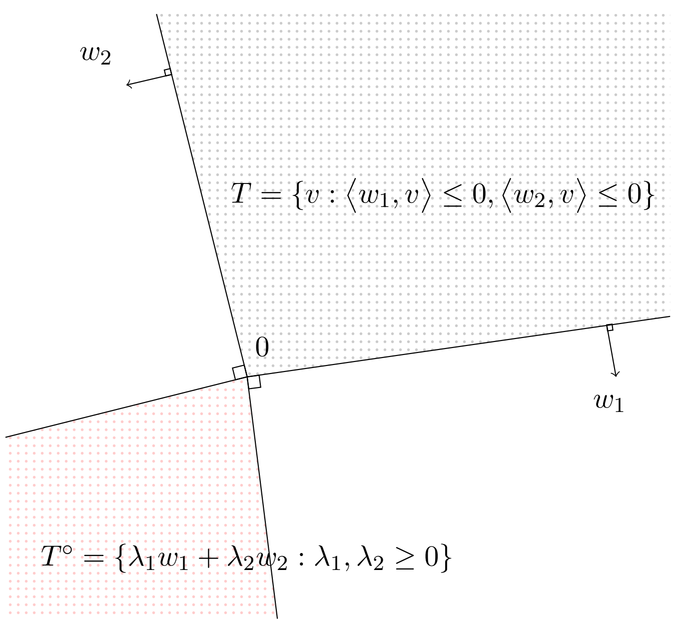

Figure 1 – Un cône T et son cône dual T°.

Dans la Proposition 1.7, on a décrit le cône tangent en un point $x \in K$ d'un ensemble convexe K. On peut aussi décrire son cône normal. Ce résultat nous sera utile pour les théorèmes en optimisation convexe différentiable.

Proposition 1.10 *Soit $K \subset \mathbb{R}^n$ un ensemble convexe non vide. Soit $x \in K$. Le cône normal à K en x est*

$$N_K(x) = \left\{z \in \mathbb{R}^n : \left\langle z, y - x\right\rangle \leq 0, \forall y \in K\right\}.$$

Preuve. Par définition du cône normal à K en x et par la Proposition 1.7, on a

$$\begin{aligned} N_K(x) &= \{z \in \mathbb{R}^n : \left\langle z, v\right\rangle \leq 0, \forall v \in T_K(x)\} \\ &= \left\{z \in \mathbb{R}^n : \left\langle z, v\right\rangle \leq 0, \forall v \in \overline{\{\lambda(y-x) : y \in K, \lambda \geq 0\}}\right\} \\ &= \left\{z \in \mathbb{R}^n : \left\langle z, v\right\rangle \leq 0, \forall v \in \{\lambda(y-x) : y \in K, \lambda \geq 0\}\right\} \\ &= \left\{z \in \mathbb{R}^n : \left\langle z, y - x\right\rangle \leq 0, \forall y \in K\right\} \end{aligned}$$

où l'avant-dernière égalité est due à la continuité de la fonction $v \to \left\langle z, v\right\rangle$ pour tout $z \in \mathbb{R}^n$. ■

En conclusion, on décrira localement une fonction par son gradient qui donne la description duale de la meilleure approximation affine de f en un point. De même on décrira une contrainte localement par son cône normal. Aux points solutions du problème $\min_{x\in K} f(x)$, on verra que ces deux descriptions locales vont donner une condition nécessaire qu'on appelle condition d'Euler/Péano/Kantorovitch.

2 Théorème d'Euler/Péano/Kantorovitch

But : Le but de cette section est de donner une condition que toute solution à un problème de la forme $\min_{x\in K} f(x)$ doit satisfaire à partir des approximations locales du premier ordre de f et K. Tous les théorèmes d'optimisation différentiable du premier ordre découlent de ce théorème (comme KKT, extrema lié ou la condition du premier pour les problèmes d'optimisation sans contrainte). Ce résultat fera aussi ressortir la nécessité de donner une description simple du cône normal (qu'on obtiendra en faisant une description simple du cône tangent). C'est-à-dire à l'introduction de l'hypothèse de qualification.

Théorème 2.1 *Soit U un ouvert non vide de $\mathbb{R}^n$ et $f : U \to \mathbb{R}$ différentiable. Soit $K \subset U$ et $x^* \in K$. Si x^* est un minimum local de $f_{|K}$ alors $-\nabla f(x^*) \in N_K(x^*)$.*

Preuve. Comme x^* un minimum local de $f_{|K}$, il existe $\epsilon_0 > 0$ tel que pour tout $x \in K \cap B_2(x^*, \epsilon_0), f(x) \geq f(x^*)$. Comme f est différentiable en x^*, on a quand $x \to x^*$,

$$f(x) = f(x^*) + \left\langle \nabla f(x^*), x - x^*\right\rangle + o(\|x - x^*\|_2).$$

Ainsi pour tout $x \in K \cap B_2(x^*, \epsilon_0)$ tel que $x \to x^*$, on a $\left\langle \nabla f(x^*), x - x^*\right\rangle + o(\|x - x^*\|_2) \geq 0$ et donc

$$\left\langle \nabla f(x^*), \frac{x - x^*}{\|x - x^*\|_2}\right\rangle + o(1) \geq 0. \tag{2.1}$$

Soit $v \in T_K(x^*)$, montrons que $\left\langle -\nabla f(x^*), v\right\rangle \leq 0$. On suppose $v \neq 0$ (le cas $v = 0$ est immédiat). On considère $(\lambda_k)_k \subset \mathbb{R}_+^*$ et $(v_k)_k \subset \mathbb{R}^n$ telles que $x_k := x^* + \lambda_k v_k \in K$, $v = \lim_k v_k$ et $(\lambda_k)_k \downarrow 0$. Comme $(v_k)_k$ converge, elle est bornée et comme $(\lambda_k)_k \downarrow 0$ on a que $(\lambda_k v_k)_k$ tend vers 0 et donc $x_k \to x^*$ quand $k \to \infty$. On peut alors appliquer (2.1) : quand $k \to \infty$,

$$\left\langle \nabla f(x^*), \frac{x_k - x^*}{\|x_k - x^*\|_2}\right\rangle + o(1) \geq 0$$

et comme $((x_k - x^*)/\|x_k - x^*\|_2)_k$ tend vers $v/\|v\|_2$. On a $\left\langle \nabla f(x^*), v\right\rangle \geq 0$ càd $\left\langle -\nabla f(x^*), v\right\rangle \leq 0$. Ceci étant vrai pour tout $v \in T_K(x^*)$, on en déduit que $-\nabla f(x^*) \in N_K(x^*)$.

■

Commentaires sur le Théorème d'Euler/Péano/Kantorovitch :

1) On ne suppose rien d'autre sur K sauf que $K \subset U$. En particulier, K n'est pas supposé être un fermé de $\mathbb{R}^n$. On peut donc prendre $K = U$ dans le Théorème 2.1. Dans ce cas, tout point $x \in U$ est tel que $B_2(x, \epsilon_0) \subset U$ donc $T_K(x) = \mathbb{R}^n$ et alors $N_K(x) = \{0\}$. La condition d'Euler/Péano/Kantorovitch s'écrit alors "$\nabla f(x^*) = 0$". C'est-à-dire, x^* est un point critique ! On retrouve donc la condition du premier pour les problèmes d'optimisation sans contrainte qu'on a rencontrés au chapitre précédent.

2) Le théorème d'Euler/Péano/Kantorovitch s'interprète géométriquement comme dans la Figure 2 : le premier ensemble de niveau de f en contact en un point x^* avec la contrainte K doit satisfaire que ses vecteurs normaux comme $\nabla f(x^*)$ sont dans l'opposé du cône normal à K en x^*.

3) On ne sait appliquer le théorème d'Euler/Péano/Kantorovitch uniquement quand on sait décrire facilement le cône normal à K en x^*. C'est le but de l'hypothèse de qualification d'assurer que $N_K(x^*)$ s'écrit facilement en fonction des fonctions $g_1, \ldots, g_r, h_1, \ldots, h_l$ décrivant K.

4) Contrairement à ce que peut laisser penser la Figure 2, on n'a pas $\nabla f(x^*) \in T_K(x^*)$. En fait, d'un point de vue conceptuel, $\nabla f(x^*)$ est un objet dual (c'est le vecteur qui engendre l'espace normal à l'hyperplan tangent à f en x^*) alors que $T_K(x^*)$ est un objet primal, c'est l'approximation au premier ordre de K en x^*. Ce ne sont donc pas des objet de même type même si ici ils sont tous les deux sous-ensemble ou élément de $\mathbb{R}^n$. En effet, dans ce cours on ne travaille qu'en dimension finie, ces deux objets vivent donc le même espace, on peut donc être tenté de supposer que $\nabla f(x^*) \in T_K(x^*)$ au vu de la Figure 2. Cependant, on peut voir sur un exemple que ce n'est pas le cas. On prend la fonction objectif $f(x, y) = x + y$ et la contrainte $K = \mathbb{R}^+ \times \{0\} \cup \{0\} \times \mathbb{R}^+$. On voit que $\min_{(x,y)\in K} f(x, y)$ est atteint en l'unique point $x^* = (0, 0)$. Mais $\nabla f(x^*) = (1, 1)^\top$ et $T_K(x^*) = \mathbb{R}^+ \times \{0\} \cup \{0\} \times \mathbb{R}^+$ donc $\nabla f(x^*) \notin T_K(x^*)$.

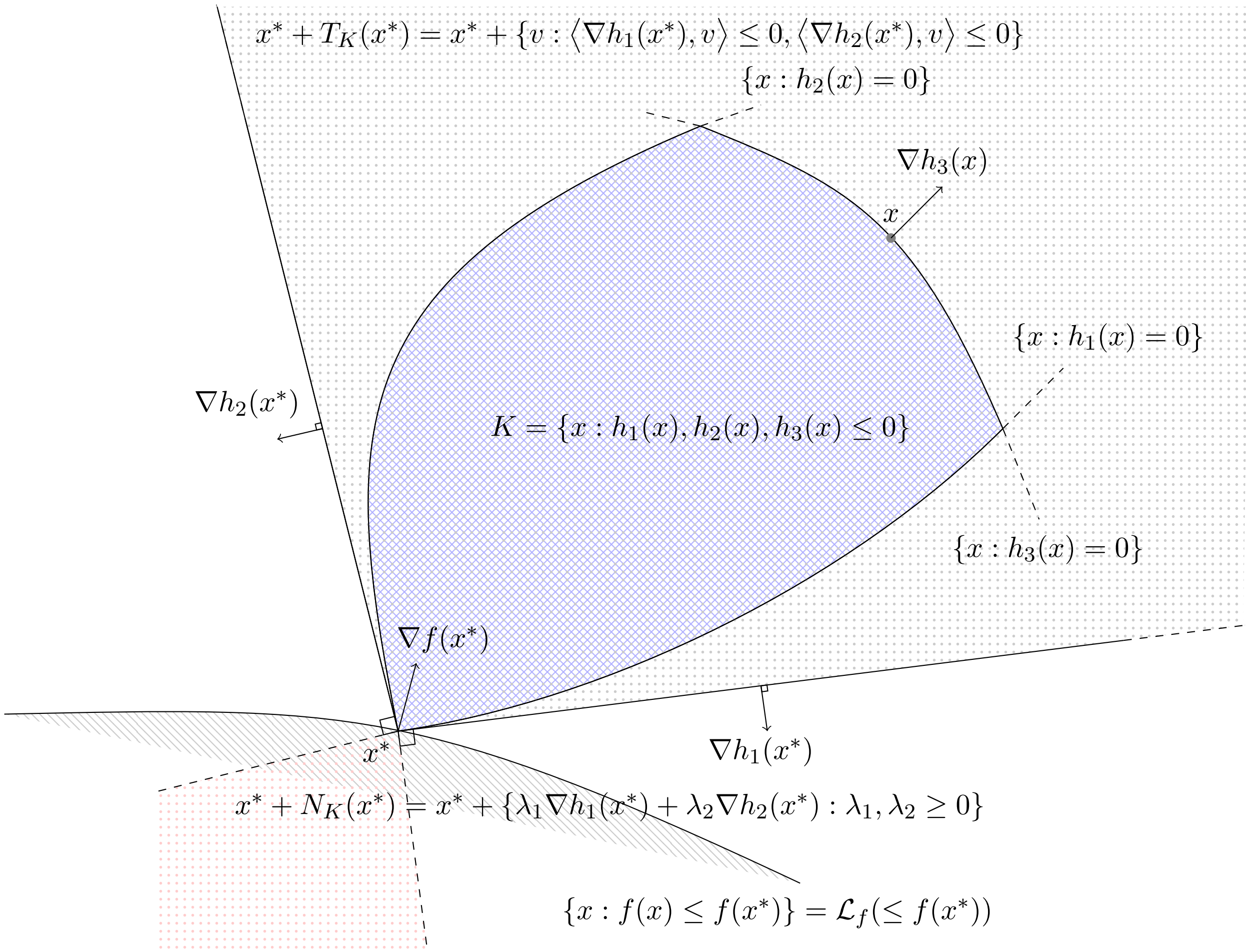

FIGURE 2 – Interprétation géométrique du théorème d'Euler/Péano/Kantorovitch : $-\nabla f(x^*) \in N_K(x^*)$. Quand la contrainte K est qualifiée en x^*, on a en plus $N_K(x^*) = \{\lambda_1 \nabla h_1(x^*) + \lambda_2 \nabla h_2(x^*) : \lambda_1, \lambda_2 \geq 0\}$ et donc E/P/K donne dans ce cas KKT : $\nabla f(x^*) + \lambda_1 \nabla h_1(x^*) + \lambda_2 \nabla h_2(x^*) = 0$ pour certains $\lambda_1, \lambda_2 \geq 0$

3 Qualification d'une contrainte K en un point $x \in K$

But : Le but de l'hypothèse de qualification est de décrire simplement le cône normal pour pouvoir appliquer le Théorème d'Euler/Péano/Kantorovitch. Ici, on donne une définition de la qualification d'une contrainte en un point via une description simple du cône tangent qui implique une description simple du cône normal.

Étant donné un ouvert U de $\mathbb{R}^n$, on considère un ensemble de contrainte K de la forme

$$K = \left\{ x \in U : \begin{array}{c} g_1(x) = \cdots = g_r(x) = 0 \\ h_1(x) \leq 0, \ldots, h_l(x) \leq 0 \end{array} \right\} \tag{3.1}$$

pour $g_1, \ldots, g_r : U \to \mathbb{R}$ définissant les contraintes d'égalité et $h_1, \ldots, h_l : U \to \mathbb{R}$ définissant les contraintes d'inégalité.

On s'intéresse dans cette section à décrire le cône tangent à K en un point $x \in K$. On commence avec la proposition suivante.

Proposition 3.1 *On considère K comme dans (3.1). On suppose que $g_1, \dots, g_r, h_1, \dots, h_l$ sont différentiables. Soit $x \in K$. On note $J(x) = \{j \in \{1, \dots, l\} : h_j(x) = 0\}$. On a*

$$T_K(x) \subset \left\{ v \in \mathbb{R}^n : \begin{array}{c} \langle \nabla g_i(x), v \rangle = 0, \forall i = 1, \dots, r \\ \langle \nabla h_j(x), v \rangle \leq 0, \forall j \in J(x) \end{array} \right\}.$$

Preuve. Soit $v \in T_K(x)$. On écrit $v = \lim_k v_k$ où $(v_k)_k \subset \mathbb{R}^n$, $x + \lambda_k v_k \in K$ pour $(\lambda_k)_k \downarrow 0$. On a pour tout $k \in \mathbb{N}$, tout $i = 1, \dots, r$ et $j \in J(x)$,

$$h_j(x + \lambda_k v_k) \leq 0 \text{ et } g_i(x + \lambda_k v_k) = 0. \tag{3.2}$$

Comme $(v_k)_k$ est une suite convergente, elle est bornée. Par ailleurs, $(\lambda_k)_k \downarrow 0$ donc $(\lambda_k v_k)_k \to 0$. On a alors quand $k \to +\infty$, pour tout $j \in J(x)$,

$$0 \geq h_j(x + \lambda_k v_k) = h_j(x) + \langle \nabla h_j(x), \lambda_k v_k \rangle + o(\|\lambda_k v_k\|_2) = \langle \nabla h_j(x), \lambda_k v_k \rangle + o(\|\lambda_k v_k\|_2)$$

car $h_j(x) = 0$ pour $j \in J(x)$ et pour tout $i = 1, \dots, l$, $g_i(x) = 0$ donc

$$0 = g_i(x + \lambda_k v_k) = g_i(x) + \langle \nabla g_i(x), \lambda_k v_k \rangle + o(\|\lambda_k v_k\|_2) = \langle \nabla g_i(x), \lambda_k v_k \rangle + o(\|\lambda_k v_k\|_2).$$

On a alors quand $k \to +\infty$, pour tout $j \in J(x)$,

$$\langle \nabla h_j(x), v_k \rangle + o(1) \leq 0 \text{ et donc } \langle \nabla h_j(x), v \rangle \leq 0$$

et pour tout $i = 1, \dots, l$,

$$\langle \nabla g_i(x), v_k \rangle + o(1) = 0 \text{ et donc } \langle \nabla g_i(x), v \rangle = 0.$$

■

Exemple d'inclusion stricte : On considère l'ensemble $K = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 \geq 0, x_2 \geq 0, x_1 x_2 = 0\}$. On voit que $\bar{x} = (0, 0) \in K$ et on veut comparer les deux cônes $T_K(\bar{x})$ et le cône "linéarisé" de la Proposition 3.1. On écrit K sous la forme standard : $K = \{x \in \mathbb{R}^2 : g_1(x) = 0, h_1(x) \leq 0, h_2(x) \leq 0\}$ où $g_1(x) = x_1 x_2$, $h_1(x) = -x_1$ et $h_2(x) = -2$ pour $x = (x_1, x_2) \in \mathbb{R}^2$. On a :

$$T_K(\bar{x}) = \mathbb{R}_+ \times \{0\} \cup \{0\} \times \mathbb{R}_+$$

et

$$L(\bar{x}) := \{ v \in \mathbb{R}^n : \langle \nabla g_1(\bar{x}), v \rangle = 0, \langle \nabla h_1(\bar{x}), v \rangle \leq 0 \text{ et } \langle \nabla h_2(\bar{x}), v \rangle \leq 0 \} = \mathbb{R}_+ \times \mathbb{R}_+.$$

On a donc bien $T_K(\bar{x}) \subset L(\bar{x})$ comme annoncé dans la Proposition 3.1, mais l'inclusion est stricte ; il n'y a pas égalité. Néanmoins, on peut voir que les cône duaux sont eux bien égaux : $T_K(\bar{x})^\circ = L(\bar{x})^\circ = \mathbb{R}_- \times \mathbb{R}_-$. ■

L'inclusion réciproque n'étant pas vraie en générale dans la Proposition 3.1 (voir le contre-exemple ci-dessus ou celui en dessous de la Proposition 1.5), on va devoir supposer qu'elle l'est pour pouvoir décrire $T_K(x)$ simplement. C'est le but de l'hypothèse de qualification dont on donne la définition maintenant.

Définition 3.2 *Soit K une contrainte définie comme dans (3.1) et $x \in K$. On suppose que $g_1, \dots, g_r, h_1, \dots, h_l$ sont différentiables. On dit que K* ***est qualifiée en*** *x quand*

$$T_K(x) = \left\{ v \in \mathbb{R}^n : \begin{array}{c} \langle \nabla g_i(x), v \rangle = 0, \forall i = 1, \dots, r \\ \langle \nabla h_j(x), v \rangle \leq 0, \forall j \in J(x) \end{array} \right\}.$$

Commentaires sur l'hypothèse de qualification :

1) La contrainte de qualification revient donc à supposer que l'inclusion réciproque dans la Proposition 3.1 est vraie.

2) En pratique, l'hypothèse de qualification est très difficile à vérifier. Cependant, on peut donner des conditions suffisantes qui impliquent la qualification. On a déjà vu deux conditions remplissant ce rôle : la condition de Mangasarian-Fromovitz et la condition QC-A. On en verra encore d'autres.

3) L'hypothèse de qualification de K est une propriété des fonctions $(g_i)_i$ et $(h_j)_j$ et non de K en tant que sous-ensemble de $\mathbb{R}^n$. On peut, par exemple, trouver des exemples où une contrainte K se décrit de manière équivalente par deux systèmes de fonctions $(g_i)_i$ et $(h_j)_j$ et $(g_i')_i$ et $(h_j')_j$ pour lesquels K peut être qualifiée pour le premier système mais pas pour le deuxième. Par exemple, on a

$$K = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 0\} = \{(x,y) \in \mathbb{R}^2 : x = 0, y = 0\} = \{(0,0)\}$$

qui est une contrainte qui se décrit par une contrainte d'égalité $g_1(x,y) = x^2 + y^2$ dans le premier cas et par deux contraintes d'égalité dans le deuxième cas $g_1'(x,y) = x$ et $g_2'(x,y) = y$. On a $T_K((0,0)) = \{(0,0)\}$ et dans le premier cas, le cône linéaire à K en $(0,0)$ est $L((0,0)) = \{v \in \mathbb{R}^2 : \langle \nabla g_1((0,0)), v \rangle = 0\} = \mathbb{R}^2$ et dans le deuxième cas, le cône linéaire à K en $(0,0)$ est $L'((0,0)) = \{v \in \mathbb{R}^2 : \langle \nabla g_1'((0,0)), v \rangle = 0 \text{ et } \langle \nabla g_2'((0,0)), v \rangle = 0\} = \{(0,0)\}$. On a $T_K((0,0)) = L_2((0,0))$ donc la contrainte K est qualifiée en $(0,0)$ quand K est décrit par les deux contraintes d'égalité g_2 et g_3 mais $T_K((0,0)) \neq L_1((0,0))$ donc K n'est pas qualifiée quand K est décrite par l'unique contrainte d'égalité g_1. Pourtant K est exactement le même ensemble, c'est le singleton $\{(0,0)\}$, mais les deux descriptions mènent à des conclusions différentes concernant sa propriété de qualification. En conclusion, la propriété de qualification est donc bien une propriété dépendante de la description de K par les fonctions $(g_i)_i$ et $(h_j)_j$ et non une propriété géométrique de l'ensemble K en soit.

4) Dire que K est qualifiée en x revient à dire que le cône tangent à K en x est un polytope dont les faces ont pour vecteurs normaux les gradients

$$\nabla g_1(x), \dots, \nabla g_r(x), -\nabla g_1(x), \dots, -\nabla g_r(x), (\nabla h_j(x))_{j \in J(x)}.$$

5) Il est possible de définir l'hypothèse de qualification de manière plus minimale que celle de la Définition 3.2. Par exemple, dans le contre-exemple suivant la Proposition 3.1, on a bien l'inclusion stricte $T_K(\bar{x}) \subset L(\bar{x})$ mais $T_K(\bar{x}) \neq L(\bar{x})$ alors que les cônes duaux sont eux égaux : $T_K(\bar{x})^\circ = L(\bar{x})^\circ = \mathbb{R}_- \times \mathbb{R}_-$. On peut, en fait, définir la qualification d'une contrainte en supposant que le cône normal à K en x est égal au cône dual du cône linéarisé de K en x. C'est-à-dire, on peut définir la qualification de K en x par "K est qualifié en $x \in K$" quand

$$N_K(a) = \left\{ \sum_{i=1}^r \lambda_i^{(+)} \nabla g_i(a) + \sum_{i=1}^r \lambda_i^{(-)} (-\nabla g_i(a)) + \sum_{j \in J(a)} \mu_j \nabla h_j(a) : \lambda_i^{(+)}, \lambda_i^{(-)}, \mu_j \geq 0 \right\} \tag{3.3}$$

où, on le verra dans la suite, que

$$\left\{ \sum_{i=1}^r \lambda_i^{(+)} \nabla g_i(a) + \sum_{i=1}^r \lambda_i^{(-)} (-\nabla g_i(a)) + \sum_{j \in J(a)} \mu_j \nabla h_j(a) : \lambda_i^{(+)}, \lambda_i^{(-)}, \mu_j \geq 0 \right\}$$

est le cône dual de

$$\left\{ v \in \mathbb{R}^n : \begin{array}{c} \langle \nabla g_i(x), v \rangle = 0, \forall i = 1, \dots, r \\ \langle \nabla h_j(x), v \rangle \leq 0, \forall j \in J(x) \end{array} \right\}.$$

On peut montrer tous les résultats sous l'hypothèse de qualification telle que définie dans (3.3) à la place de celle de la Définition 3.2. La relation (3.3) étant impliquée par celle de la Définition 3.2, elle est donc plus faible et, en particulier, couvre le cas du contre-exemple suivant la Proposition 3.1. Néanmoins, dans le cadre de ce cours, on définira la qualification d'une contrainte grâce à la Définition 3.2 qui est celle la plus communément utilisée même si dans toute nos preuves on n'utilisera que la relation (3.3).

4 Preuve du théorème de KKT

Dans cette section, on démontre le théorème de KKT à partir du théorème d'Euler/Péano/Kantorovitch. D'une manière générale, pour pouvoir appliquer Euler/Péno/Kantorovitch, on a besoin de savoir décrire facilement le cône normal $N_K(x^*)$. C'est le but de l'hypothèse de qualification. Néanmoins, cette hypothèse porte sur le cône tangent et non directement sur le cône normal. Comme sous hypothèse de qualification, le cône tangent est un cône polyédral, on va d'abord devoir décrire les cônes duaux aux cônes polyédraux.

4.1 Cône dual d'un cône polyédral

On commence par rappeler la définition d'un polyèdre (ou polytope).

Définition 4.1 *Un **polyèdre** est une intersection de demi-espaces affines. C'est un ensemble de la forme*

$$\{v \in \mathbb{R}^n : Av \leq b\}$$

où $A \in \mathbb{R}^{\ell \times n}$, $b \in \mathbb{R}^\ell$ et "$Av \leq b$" signifie que $(Av)_j \leq b_j$ pour tout $j = 1, \ldots, \ell$.

Les contraintes d'égalité de K mènent aux conditions "$\langle \nabla g_i(v), v \rangle = 0$" pour le cône tangent qui définissent des hyper-plans affines de $\mathbb{R}^n$ et non des demi-espaces affines. Néanmoins, on peut réécrire de manière équivalente l'équation "$\langle \nabla g_i(x), v \rangle = 0$" sous la forme de deux équations "$\langle \nabla g_i(x), v \rangle \leq 0$" et "$-\langle \nabla g_i(x), v \rangle \leq 0$" définissant chacune un demi-espace affine. Ainsi, sous l'hypothèse de qualification, le cône tangent à K en $x \in K$ s'écrit

$$T_K(x) = \left\{ v \in \mathbb{R}^n : \begin{array}{c} \langle \nabla g_i(x), v \rangle = 0, \forall i = 1, \ldots, r \\ \langle \nabla h_j(x), v \rangle \leq 0, \forall j \in J(x) \end{array} \right\}$$

qui est donc bien un polyèdre au sens de la Définition 4.1. On est donc bien amené à décrire le cône dual d'un cône polyédral. C'est le but de la section suivante.

But : Décrire simplement le cône normal $N_K(x)$ à K en $x \in K$ quand la contrainte K est qualifiée en x. Sous cette hypothèse de qualification, le cône tangent est un polyèdre ; c'est donc à la fois un cône et un polyèdre : c'est un **cône polyédral** pour lequel on peut décrire le cône dual et ainsi avoir une description du cône normal $N_K(x)$ sous l'hypothèse de qualification. On commence d'abord par décrire les cône polyédraux.

Lemme 4.2 *Si C est un cône polyédral alors il existe une matrice $A \in \mathbb{R}^{\ell \times n}$ telle que $C = \{v \in \mathbb{R}^n : Av \leq 0\}$.*

Preuve. Comme C est un polyèdre il existe $A \in \mathbb{R}^{\ell \times n}$ et $b \in \mathbb{R}^\ell$ tels que $C = \{v \in \mathbb{R}^n : Av \leq b\}$. Comme C est un cône, on a pour tout $v \in C$ et tout $\mu \geq 0$, $\mu v \in C$ càd $A(\mu v) \leq b$. Alors $Av \leq b/\mu$ et ceci étant vrai pour tout μ, on a $Av \leq 0$. Par ailleurs, $0 \in C$, donc $0 = A0 \leq b$, alors si $Av \leq 0$ on a aussi $Av \leq b$. On a donc bien $C = \{v \in \mathbb{R}^n : Av \leq 0\}$. ∎

Théorème 4.3 *Soit $A \in \mathbb{R}^{\ell \times n}$. Soit $C = \{v \in \mathbb{R}^n : Av \leq 0\}$ un cône polyédral non vide. On note $a_1, \ldots, a_\ell \in \mathbb{R}^n$ les vecteurs lignes de A. Le cône dual de C est*

$$C^\circ = \left\{ \sum_{j=1}^{\ell} \lambda_j a_j \in \mathbb{R}^n : \lambda_1, \ldots, \lambda_\ell \geq 0 \right\}.$$

Preuve. On note

$$L = \left\{ \sum_{j=1}^{\ell} \lambda_j a_j \in \mathbb{R}^n : \lambda_1, \ldots, \lambda_\ell \geq 0 \right\}.$$

1) On montre la première inclusion "$L \subset C^\circ$" : Soit $\lambda_1, \ldots, \lambda_\ell \geq 0$. Montrons que $\sum_{j=1}^{\ell} \lambda_j a_j \in C^\circ$. Pour cela, il suffit de montrer que pour tout $v \in C$ on a $\left\langle \sum_{j=1}^{\ell} \lambda_j a_j, v \right\rangle \leq 0$. Soit $v \in C$. On a

$$\left\langle \sum_{j=1}^{\ell} \lambda_j a_j, v \right\rangle = \left\langle A^\top \lambda, v \right\rangle = \lambda^\top A v = \sum_{j=1}^{\ell} \lambda_j \left\langle a_j, v \right\rangle$$

or $Av \leq 0$ ce qui est équivalent à dire que $\left\langle a_j, v \right\rangle \leq 0$ pour tout $j = 1, \ldots, \ell$, on a donc bien $\left\langle \sum_{j=1}^{\ell} \lambda_j a_j, v \right\rangle \leq 0$.

2) On montre maintenant la deuxième inclusion "$C^\circ \subset L$". Pour cela, on montre l'inclusion des complémentaires : $\bar{L} \subset \bar{C^\circ}$. Soit $z \in \mathbb{R}^n$ tel que $z \notin L$. Montrons que $z \notin \bar{C^\circ}$. On utilise une version du théorème de Hahn-Banach : si L_0 est un cône convexe fermé de $\mathbb{R}^n$ et $z \notin L_0$ alors il existe un hyper-plan séparant strictement L_0 et z passant par l'origine. En d'autres termes : il existe $u \in \mathbb{R}^n$ tel que $\left\langle u, z \right\rangle > 0$ et pour tout $x \in L_0, \left\langle u, x \right\rangle \leq 0$.

Il est facile de voir que L est un cône convexe fermé. On applique le théorème de Hahn-Banach à L et z : il existe $u \in \mathbb{R}^n$ tel que $\left\langle u, z \right\rangle < 0$ et pour tout $x \in L$, on a $\left\langle u, x \right\rangle \geq 0$. On veut montrer que $z \notin C^\circ$ càd qu'il existe un $v \in C$ pour lequel on a $\left\langle z, v \right\rangle > 0$. On va montrer que $u \in C$. Comme on a $\left\langle z, u \right\rangle > 0$ cela finira la preuve. On a pour tout $x \in L, \left\langle u, x \right\rangle \leq 0$ donc pour tout $\lambda \in (\mathbb{R}_+)^\ell, \left\langle u, A^\top \lambda \right\rangle \leq 0$ càd $\left\langle Au, \lambda \right\rangle \leq 0$. Cela implique que $Au \leq 0$. On a donc bien $u \in C$ et comme $\left\langle z, u \right\rangle > 0$, on en déduit bien que $z \notin C^\circ$. ■

4.2 Preuve du théorème de KKT à partir du théorème d'Euler/Péano/Kantorovitch

On rappelle le cadre et le théorème de KKT. On étudie les problèmes de la forme

$$\min_{x \in K} f(x) \tag{4.1}$$

où $f : U \to \mathbb{R}$, U est un ouvert de $\mathbb{R}^n$ et K est une contrainte de la forme

$$K = \left\{ x \in U : \begin{array}{c} g_1(x) = \cdots = g_r(x) = 0 \\ h_1(x) \leq 0, \ldots, h_l(x) \leq 0 \end{array} \right\}$$

pour $g_1, \ldots, g_r : U \to \mathbb{R}$ définissant les contraintes d'égalité et $h_1, \ldots, h_l : U \to \mathbb{R}$ définissant les contraintes d'inégalité.

Théorème 4.4 (Théorème de Karush-Kuhn-Tucker (KKT)) *On suppose que les fonctions $f, g_1, \ldots, g_r, h_1, \ldots, h_l$ sont différentiables. Soit $a \in K$. On suppose que K est qualifiée en a. Si a est un minimum local de f restreint à K alors il existe $\lambda_1, \ldots, \lambda_r \in \mathbb{R}$ et $\mu_1, \ldots, \mu_l \in \mathbb{R}$ tels que :*

a)

$$\nabla f(a) + \sum_{i=1}^{r} \lambda_i \nabla g_i(a) + \sum_{j=1}^{l} \mu_j \nabla h_j(a) = 0$$

b) $\mu_j \geq 0$ *pour tout* $j = 1, \ldots, l$

c) $\mu_j h_j(a) = 0$ *pour tout* $j = 1, \ldots, l$.

Preuve. D'après le Théorème d'Euler/Péano/Kantorovitch (voir Théorème 2.1) : si a est un minimum local de f restreint à K alors

$$-\nabla f(a) \in N_K(a). \tag{4.2}$$

Par hypothèse, K est qualifiée en a donc

$$T_K(a) = \left\{ v \in \mathbb{R}^n : \begin{array}{c} \langle \nabla g_i(a), v \rangle = 0, \forall i = 1, \ldots, r \\ \langle \nabla h_j(a), v \rangle \leq 0, \forall j \in J(a) \end{array} \right\}$$

où $J(a) = \{j \in \{1, \ldots, l\} : h_j(a) = 0\}$. Comme $T_K(a)$ est un cône polyédral, son cône dual, qui est le cône $N_K(a)$, est décrit dans le Théorème 4.3 : on a

$$N_K(a) = (T_K(a))^\circ = \left\{ \sum_{i=1}^{r} \lambda_i^{(+)} \nabla g_i(a) + \sum_{i=1}^{r} \lambda_i^{(-)} (-\nabla g_i(a)) + \sum_{j \in J(a)} \mu_j \nabla h_j(a) : \lambda_i^{(+)}, \lambda_i^{(-)}, \mu_j \geq 0 \right\}.$$

qui peut se réécrire sous la forme

$$N_K(a) = \left\{ \sum_{i=1}^{r} \lambda_i \nabla g_i(a) + \sum_{j \in J(a)} \mu_j \nabla h_j(a) : \lambda_i \in \mathbb{R}, \mu_j \geq 0 \right\}.$$

Ainsi la condition d'Euler/Péano/Kantorovitch de (4.2) est ici équivalente aux trois condition de KKT. ■