Prépa. Agrég. écrit d'Analyse, Annexe n° 5.

Fonctions monotones. Fonctions convexes

Fonctions monotones

Fonctions de sauts

On considère un ensemble dénombrable $\mathrm{D} \subset [0,1]$, que l'on énumère dans une suite de points $(x_n)_{n \in \mathbb{N}}$; pour chaque point $x \in \mathrm{D}$, c'est à dire pour chaque $n \geq 0$, on se donne une hauteur de "saut" $\varepsilon_n > 0$, et on suppose que $\sum_{n=0}^{+\infty} \varepsilon_n = 1$. On pose

$$\forall x \in \mathbb{R}, \quad f(x) = \sum_{x_n \leq x} \varepsilon_n.$$

On définit ainsi une fonction croissante et continue à droite. Si on pense probabilité, cette fonction f est la fonction de répartition de la mesure (de probabilité) purement atomique $\mu = \sum_{n=0}^{+\infty} \varepsilon_n \delta_{x_n}$. On a $f(x) = 0$ si $x < 0$ et $f(x) = 1$ si $x \geq 1$.

Rapport avec le théorème de dérivabilité de Lebesgue

Il ne faut pas croire que la fonction f ait une dérivée nulle en tout point qui n'est pas un point de saut, c'est à dire en tout $x \notin \mathrm{D}$. En effet, on peut vérifier qu'en tout point de l'ensemble

$$\mathrm{A} = \bigcap_m \bigcup_{n \geq m} [x_n - \varepsilon_n, x_n + \varepsilon_n]$$

la dérivée ne peut pas être nulle, et cet ensemble A peut être très riche (un ensemble avec la puissance du continu, genre Cantor : voir un exemple dans le paragraphe suivant).

Cependant, un théorème (assez délicat) de Lebesgue dit que toute fonction croissante est dérivable presque partout ; ce résultat est plus facile à obtenir dans le cas présent, mais il n'est quand même pas trivial, à ma connaissance. On va donc montrer que :

la fonction de sauts f à une dérivée à droite nulle Lebesgue-presque partout.

On a le même résultat à gauche, par une démonstration analogue, donc en fait f est Lebesgue-presque partout dérivable de dérivée nulle. Il est clair que pour tout entier $\mathrm{N} \geq 0$, la fonction f_N définie par $f_\mathrm{N}(x) = \sum_{n=0}^{\mathrm{N}-1} \varepsilon_n \, \mathbf{1}_{x_n \leq x}$ n'a que N sauts, situés aux points $x_0, \ldots, x_{\mathrm{N}-1}$, donc elle a certainement une dérivée nulle partout ailleurs ; on voit ainsi que le problème de la dérivabilité de f se ramène à celui de $g_\mathrm{N} = f - f_\mathrm{N}$, pour un N quelconque ; la seule différence entre f et g_N est que $g_\mathrm{N} = f - f_\mathrm{N}$ est une fonction de sauts pour laquelle la somme des sauts devient petite, puisque cette somme vaut $r_\mathrm{N} = \sum_{n \geq \mathrm{N}} \varepsilon_n$, qui tend vers 0 avec N. Notons que $g_\mathrm{N}(1) - g_\mathrm{N}(0) \leq r_\mathrm{N}$.

Introduisons l'un des quatre *nombres dérivés de Dini* de la fonction g, le nombre $\mathrm{D}^+ g(x)$, qu'on peut appeler la *dérivée supérieure droite* de g au point x,

$$\mathrm{D}^+ g(x) = \limsup_{y \to x, y > x} \frac{g(y) - g(x)}{y - x},$$

valeur $+\infty$ admise. On définit de même une dérivée inférieure droite $\mathrm{D}_+ g(x)$ en remplaçant lim sup par lim inf, et la dérivée à droite ordinaire $g'_d(x)$ existe précisément quand $\mathrm{D}^+ g(x) = \mathrm{D}_+ g(x)$. La fonction $\mathrm{D}^+ g$ est clairement mesurable quand g est croissante et continue à droite (parce qu'on peut limiter le choix de y au cas où $y - x$ est un rationnel

> 0, par exemple). Soit $\alpha > 0$; on va montrer que la dérivée supérieure à droite $\mathrm{D}^+ g$ d'une fonction de sauts g telle que $g(1) - g(0) < \alpha^2$ est plus petite que α sauf sur un ensemble de mesure de Lebesgue $\leq \alpha$.

Soit K un compact contenu dans $\{\mathrm{D}^+ g > \alpha\}$; posons $k(x) = \alpha \int_0^x \mathbf{1}_\mathrm{K}$, c'est à dire $k(x) = \big|\mathrm{K} \cap [0, x]\big|$, où $|\mathrm{A}|$ désigne la mesure de Lebesgue de $\mathrm{A} \subset [0, 1]$; posons ensuite $h(x) = g(x) - k(x)$; on va montrer que

($*$) *pour tout* $x \in [0, 1[$, *il existe* $y > x$ *tel que* $h(y) \geq h(x)$;

comme g est croissante et k continue, un petit coup de borne supérieure impliquera que $h(1) \geq h(0)$, d'où $\alpha\,|\mathrm{K}| = k(1) - k(0) \leq g(1) - g(0) \leq \alpha^2$, l'inégalité que nous cherchions : considérons en effet l'ensemble B des $x \in [0, 1]$ tels que $h(x) \geq h(0)$; cet ensemble contient 0 et si $(x_n) \subset \mathrm{B}$ tend en croissant vers x, on aura $g(x) \geq \lim_n g(x_n) \geq \lim_n k(x_n) = k(x)$, donc $x \in \mathrm{B}$. Ceci montre que B contient sa borne supérieure, et la propriété ($*$) pour h implique que la borne supérieure de B est égale à 1. Montrons donc cette propriété ($*$) : si $x \notin \mathrm{K}$, la fonction k reste constante au voisinage de x, donc h varie comme g dans ce voisinage V, c'est à dire en croissant et on prend pour y n'importe quel $y > x$ qui est dans V ; si $x \in \mathrm{K}$, alors $\mathrm{D}^+ g(x) > \alpha$, ce qui donne l'existence d'un $y > x$ tel que $g(y) - g(x) > \alpha(y - x)$; comme la fonction k ne peut pas varier de plus de $\alpha(y - x)$ entre x et y, il en résulte que $h(y) - h(x) > 0$.

Puisque tout compact K contenu dans $\{\mathrm{D}^+ g > \alpha\}$ a une mesure de Lebesgue $\leq \alpha$, il en résulte que l'ensemble $\{\mathrm{D}^+ g > \alpha\}$ est de mesure de Lebesgue $\leq \alpha$. D'après nos considérations sur le passage de f à g_N, on déduit que $\{\mathrm{D}^+ f > \alpha\}$ est de même mesure que $\{\mathrm{D}^+ g_\mathrm{N} > \alpha\}$ pour tout N, donc de mesure $\leq \alpha$ (prendre N tel que $\sum_{n \geq \mathrm{N}} \varepsilon_n < \alpha^2$) pour tout $\alpha > 0$, donc $\mathrm{D}^+ f$ est nulle Lebesgue-presque partout. Comme f est croissante, on a $f'_d(x) = 0$ en tout point où $\mathrm{D}^+ f(x) = 0$.

Un exemple plus spécifique

On prend pour D l'ensemble de tous les nombres dyadiques de $]0, 1[$, c'est à dire les nombres de la forme $j\,2^{-k}$, j, k entiers et $k > 0$ et $0 < j < 2^k$.

Si $x_n \in \mathrm{D}$ s'écrit $x_n = j\,2^{-k}$ avec j impair, posons $\varepsilon_n = 3^{-k}$; on vérifie que $\sum_{n=0}^{+\infty} \varepsilon_n = 1$; en effet, il y a 2^{k-1} dyadiques de la forme $j\,2^{-k}$, j impair et donc

$$\sum_{n=0}^{+\infty} \varepsilon_n = \sum_{k=1}^{+\infty} \frac{2^{k-1}}{3^k} = 3^{-1}\,\frac{1}{1 - 2/3} = 1.$$

La fonction f est *strictement croissante* sur $[0, 1]$ (parce que D est dense dans $[0, 1]$), ce qui permet de définir une fonction inverse g par la formule

$$\forall y \in]0, 1[, \quad g(y) = \sup\{x : f(x) < y\} = \inf\{x : f(x) > y\}.$$

Cette fonction inverse est la fonction g de Cantor-Lebesgue, fonction croissante continue, qui vérifie $g(0) = 0$ et $g(1) = 1$, et qui est constante sur chaque intervalle complémentaire du triadique de Cantor. Le procédé de construction précédent apparaît dans un article de Ludwig Scheeffer, élève de Cantor, en 1885. Il est donc certainement abusif d'appeler cette fonction g la “fonction de Lebesgue”, comme font certains. Cette fonction intéressait les mathématiciens de l'époque car elle montrait que la dérivée peut être nulle en dehors d'un fermé de mesure nulle (notion accessible à l'intégrale de Riemann) sans que la fonction soit constante, contrairement à ce que certains avaient écrit un peu vite, à peu près à la même époque.

Montrons comme promis que la fonction de sauts f admet une infinité non dénombrable de points où sa dérivée à droite est non nulle (la même chose est vraie pour la dérivée à gauche). Choisissons une suite strictement croissante d'entiers $(m_k)_{k\geq 0}$ telle que $2^{m_{k+1}} > 3^{m_k}$ pour tout $k \geq 0$. Pour chaque suite $\mathbf{a} = (a_k)$ formée de 0 et de 1 posons

$$y(\mathbf{a}) = 1 - \sum_{k=0}^{+\infty} \frac{a_k}{2^{m_k}} \in [0,1].$$

Fixons $\mathbf{a}$ qui change une infinité de fois de valeur, posons $y = y(\mathbf{a})$ et pour tout $k \geq 0$

$$y_k = 1 - \sum_{j=0}^{k} \frac{a_j}{2^{m_j}} > y.$$

Lorsque $a_k = 1$, on voit que y_k est de la forme $j\, 2^{-m_k}$ avec j impair, donc $y_k = x_n \in \mathrm{D}$ pour un certain n, et $\varepsilon_n = 3^{-m_k}$. Alors $f(y_k) - f(y) \geq 3^{-m_k}$ et

$$y_k - y = \sum_{j>k} \frac{a_j}{2^{m_j}} \leq 2^{1-m_{k+1}}$$

donc $\bigl(f(y_k) - f(y)\bigr)/(y_k - y) \geq 1/2$. Il en résulte que la dérivée à droite de f au point y ne peut pas être nulle. Et il y a une infinité non dénombrable de tels points $y = y(\mathbf{a})$, quand $\mathbf{a}$ varie dans $\{0,1\}^{\mathbb{N}}$.

Fonctions convexes

Il est commode d'admettre qu'une fonction convexe peut prendre la valeur $+\infty$, ce qui permet de considérer qu'elle est définie sur l'espace entier. En revanche pour faire simple on excluera la valeur $-\infty$. Les inégalités de convexité avec valeurs $+\infty$ admises sont à traiter avec les conventions utilisées pour l'intégration des fonctions ≥ 0.

Continuité et caractère localement lipschitzien

Théorème. *Si* F *est convexe et finie sur un ouvert convexe* U *de* $\mathbb{R}^d$, *elle est localement lipschitzienne dans* U.

La démonstration passe par les étapes qui suivent.

1. *Si* F *est finie dans* U, *elle est majorée au voisinage de tout point* $x \in \mathrm{U}$ *(cette étape utilise la finitude de la dimension).*

Pour $x \in \mathrm{U}$, on choisit un voisinage convexe fermé V de x tel que $\mathrm{V} \subset \mathrm{U}$, et tel que V soit l'enveloppe convexe d'un ensemble fini S (on peut prendre pour V un (hyper)-cube de centre x). La convexité entraîne que $\mathrm{F}(y) \leq \mathrm{M} = \max\{\mathrm{F}(z) : z \in \mathrm{S}\}$ pour tout $y \in \mathrm{V}$.

2. *Si* F *est majorée dans une boule* $\mathrm{B}(x,r)$, *elle est minorée dans la même boule (valable pour un espace normé de dimension infinie).*

Supposons F majorée par M dans $\mathrm{B}(x,r)$, et soit $x+v$ un point quelconque de $\mathrm{B}(x,r)$. On écrit $\mathrm{F}(x+v) + \mathrm{F}(x-v) \geq 2\mathrm{F}(x)$, d'où résulte que $\mathrm{F}(x+v) \geq 2\mathrm{F}(x) - \mathrm{M}$.

3. *Si* F *est bornée dans la boule fermée* $\overline{\mathrm{B}(x_0,r)}$, *elle est lipschitzienne dans* $\mathrm{B}(x_0,s)$ *pour tout* $s < r$ *(valable pour un espace normé de dimension infinie).*

Supposons $|f| \leq \mathrm{M}$ sur $\overline{\mathrm{B}(x_0,r)}$. Soient $x \neq y$ deux points de $\mathrm{B}(x_0,s)$; on va écrire $y = x + tu$, avec u vecteur de norme un et $t = \|y-x\|$. On prolonge la demi-droite issue de x, de vecteur directeur u, qui passe par le point $y = x + tu \in \mathrm{B}(x_0, s)$, jusqu'à ce qu'elle rencontre le bord de $\overline{\mathrm{B}(x_0, r)}$ en un point $x + \theta u$; on a nécessairement $t < \theta$, et

$$\theta = \|\theta u\| \geq \|(x + \theta u) - x_0\| - \|x - x_0\| = r - \|x - x_0\| \geq r - s.$$

La fonction convexe $g(s) = \mathrm{F}(x + su)$ est bornée par M sur l'intervalle $[0, \theta]$ puisque F est bornée par M sur $\overline{\mathrm{B}(x_0, r)}$. Si on utilise la croissance de la pente de la fonction convexe g (sachant que $0 < t < \theta$), on obtient

$$\frac{\mathrm{F}(y) - \mathrm{F}(x)}{\|y - x\|} = \frac{g(t) - g(0)}{t} \leq \frac{g(\theta) - g(0)}{\theta} \leq \frac{2\mathrm{M}}{r - s}$$

ce qui montre que F est lipschitzienne de constante $2\mathrm{M}(r - s)^{-1}$ dans la boule $\mathrm{B}(x_0, s)$.

Une forme du théorème de Hahn-Banach

Théorème. *Si f est convexe, à valeurs $\mathbb{R} \cup \{+\infty\}$, définie sur un espace affine réel* E *de dimension finie, et si x_0 est un point au voisinage duquel f est finie, il existe une fonction affine a sur* E *telle que $a(x_0) = f(x_0)$ et $a \leq f$ sur* E.

Le résultat découle d'un résultat de prolongement, que l'on appliquera avec le sous-espace affine $\mathrm{F} = \{x_0\}$ et la fonction affine sur F égale à $f(x_0)$.

Si f est convexe, à valeurs $\mathbb{R} \cup \{+\infty\}$, définie sur un espace affine réel E *de dimension finie, si x_0 est un point au voisinage duquel f est finie, si* F *est un sous-espace affine de* E *contenant x_0, et si a est une fonction affine sur* F, *telle que $a(y) \leq f(y)$ pour tout $y \in \mathrm{F}$, il existe un prolongement $\widetilde{a}$ de a à* E *tout entier, qui est affine sur* E *et qui vérifie l'inégalité $\widetilde{a} \leq f$ sur* E.

L'étape cruciale est de prolonger à une dimension de plus, tant que $\mathrm{F} \neq \mathrm{E}$. Si $\mathrm{F} \neq \mathrm{E}$, on peut trouver un vecteur $\overrightarrow{v}$ tel que $x_0 + \overrightarrow{v} \notin \mathrm{F}$. L'ensemble des points $y + t\overrightarrow{v}$, où y varie dans F et t dans $\mathbb{R}$, est un sous-espace affine $\widetilde{\mathrm{F}}$ strictement plus grand que F. Si $\widetilde{a}$ est un prolongement affine de a à $\widetilde{\mathrm{F}}$, on aura

$$\forall y \in \mathrm{F},\ \forall t \in \mathbb{R}, \quad \widetilde{a}(y + t\overrightarrow{v}) = a(y) + t\left(\widetilde{a}(x_0 + \overrightarrow{v}) - a(x_0)\right);$$

Posons $\mathrm{M} = \widetilde{a}(x_0 + \overrightarrow{v}) - a(x_0)$; la seule question est de montrer qu'on peut trouver M de façon que $\widetilde{a} \leq f$ sur $\widetilde{\mathrm{F}}$, c'est à dire $a(y) + t\mathrm{M} \leq f(y + t\overrightarrow{v})$ pour tous $y \in \mathrm{F}$ et t réel. On va diviser ces conditions en deux cas, selon que $t = t_1 > 0$ ou bien $t = -t_2 < 0$ (le cas $t = 0$ est vrai par hypothèse : $a \leq f$ sur F) ; on veut que

$$\mathrm{M} \leq t_1^{-1}\big(f(y_1 + t_1\overrightarrow{v}) - a(y_1)\big) \quad \text{et} \quad \mathrm{M} \geq t_2^{-1}\big(a(y_2) - f(y_2 - t_2\overrightarrow{v})\big),$$

pour tous $y_1, y_2 \in \mathrm{F}$ et $t_1, t_2 > 0$; autrement dit M doit être coincé entre l'inf I des $t_1^{-1}\big(f(y_1 + t_1\overrightarrow{v}) - a(y_1)\big)$ et le sup S des $t_2^{-1}\big(a(y_2) - f(y_2 - t_2\overrightarrow{v})\big)$; le fait que f soit finie au voisinage de $x_0 \in \mathrm{F}$ permet de dire que l'inf n'est pas $+\infty$ et le sup n'est pas $-\infty$ (prendre $y_1 = y_2 = x_0$ et $t_1 = t_2 = t > 0$ petit). Si on montre que $\mathrm{S} \leq \mathrm{I}$, ces deux nombres seront finis et on pourra prendre pour M n'importe quel nombre de l'intervalle $[\mathrm{S}, \mathrm{I}]$ (qui est peut-être réduit à un point). Pour montrer que $\mathrm{S} \leq \mathrm{I}$, il suffit de voir que tout nombre du type 1 est plus grand que tout nombre du type 2, ce qui revient à demander si on a bien, pour tous $y_1, y_2 \in \mathrm{F}$ et $t_1, t_2 > 0$

$$t_1^{-1}a(y_1) + t_2^{-1}a(y_2) \leq t_1^{-1}f(y_1 + t_1\overrightarrow{v}) + t_2^{-1}f(y_2 - t_2\overrightarrow{v})\ ??$$

On multiplie les deux côtés par $t_1 t_2/(t_1+t_2) > 0$ et on écrit

$$\frac{t_2}{t_1+t_2}\, a(y_1) + \frac{t_1}{t_1+t_2}\, a(y_2) = a\Big(\frac{t_2}{t_1+t_2}\, y_1 + \frac{t_1}{t_1+t_2}\, y_2\Big) \le$$
$$\le f\Big(\frac{t_2}{t_1+t_2}\, y_1 + \frac{t_1}{t_1+t_2}\, y_2\Big) = f\Big(\frac{t_2}{t_1+t_2}\,(y_1 + t_1 \vec{v}) + \frac{t_1}{t_1+t_2}\,(y_2 - t_2 \vec{v})\Big) \le$$
$$\frac{t_2}{t_1+t_2}\, f(y_1 + t_1 \vec{v}) + \frac{t_1}{t_1+t_2}\, f(y_2 - t_2 \vec{v})$$

ce qu'il fallait démontrer.

Remarque. Une hypothèse est nécessaire pour qu'on puisse trouver une fonction affine a égale à f au point x_0 et $a \le f$ sur E ; la moindre des choses est évidemment que $f(x_0)$ soit fini, mais ça n'est pas suffisant : pour la fonction f convexe sur $\mathbb{R}$ égale à 0 pour $x > 0$, $f(0) = 1$ et $f(x) = +\infty$ pour $x < 0$, il n'existe pas de fonction a affine sur $\mathbb{R}$ telle que $a(0) = f(0)$ et $a \le f$ sur $\mathbb{R}$.

Si nous supposons que a existe, la fonction a sera lipschitzienne sur E, d'une certaine constante C et on devra avoir $f(y) - f(x_0) \ge a(y) - a(x_0) \ge -\mathrm{C}\,\|y - x_0\|$ pour tout $y \in \mathrm{E}$. On peut vérifier que la condition qu'il existe une constante C telle que

$$\forall y \in \mathrm{E}, \quad f(y) \ge f(x_0) - \mathrm{C}\,\|y - x_0\|$$

est la condition nécessaire et suffisante pour l'existence de a ; en effet, on introduit la fonction g définie par $g(x) = \inf_{y \in \mathrm{E}}\big(f(y) + \mathrm{C}\,\|x - y\|\big) \le f(x)$ (prendre $y = x$) ; sous la condition précédente on a $g(x_0) = f(x_0)$ et on vérifie que cette fonction g est partout finie et convexe sur E. Le théorème déjà vu s'applique à g et donne une fonction affine $a \le g$, vérifiant $a(x_0) = g(x_0) = f(x_0)$.

Transformée de Legendre-Fenchel

Ici E est l'espace $\mathbb{R}^d$ muni du produit scalaire usuel $(x, y) \to x \,.\, y$. La fonction F est convexe sur E, à valeurs $\mathbb{R} \cup \{+\infty\}$, non identiquement $+\infty$. On pose

$$\forall y \in \mathrm{E}, \quad \mathrm{F}^*(y) = \sup_{x \in \mathrm{E}} \big(y \,.\, x - \mathrm{F}(x)\big)$$

(la valeur trouvée peut être $+\infty$, mais elle n'est pas $-\infty$). La fonction F^* est la *fonction conjuguée* de F ; c'est évidemment une fonction convexe (sup de fonctions affines en y).

Si F *est une fonction de classe* C^1 *sur* E*, strictement convexe, et telle que* $\mathrm{F}(x)/\|x\|$ *tende vers* $+\infty$ *lorsque* $\|x\| \to +\infty$*, alors la fonction conjuguée* F^* *possède les mêmes propriétés. La fonction* $x \to \nabla \mathrm{F}(x)$ *est un homéomorphisme de* E *sur* E*, et la bijection réciproque est l'application gradient de* F^*.

Montrons que $\nabla \mathrm{F}$ est surjectif. Soit $y_0 \in \mathrm{E}$; la fonction $x \to y_0 \,.\, x - \mathrm{F}(x)$ est continue sur E, et elle tend vers $-\infty$ lorsque $\|x\| \to +\infty$ (mettre $\|x\|$ en facteur et utiliser l'hypothèse sur F). Cette fonction atteint donc un maximum sur E, en un point x_0, unique parce que F est strictement convexe ; il en résulte que son gradient s'annule en x_0 (et seulement en x_0), ce qui donne $y_0 - \nabla \mathrm{F}(x_0) = 0$. On a montré que $\nabla \mathrm{F}$ est une bijection continue de E sur E.

On montre ensuite que l'application inverse est continue ; si (y_n) converge vers y, il faut montrer que les points x_n tels que $\nabla \mathrm{F}(x_n) = y_n$ convergent vers le point x tel que $\nabla \mathrm{F}(x) = y$. Pour cela il suffit de savoir que (x_n) reste borné ; si nous admettons ce point, nous noterons que pour toute sous-suite (x_{n_j}) convergente vers une limite x', nous aurons $\nabla \mathrm{F}(x') = y$, donc $x' = x$, et on en déduit que (x_n) tend vers x (on vient de refaire la démonstration du fait qu'une bijection continue entre compacts est bicontinue ; on aurait pu se ramener à appliquer ce résultat classique).

La croissance de la dérivée de la fonction convexe g (de classe C^1) définie sur $\mathbb{R}$ par $g(t) = \mathrm{F}(tx)$ implique que

$$\mathrm{F}(x) - \mathrm{F}(0) = g(1) - g(0) = \int_0^1 g'(s)\, ds \le g'(1) = \nabla \mathrm{F}(x) \,.\, x \le \|\nabla \mathrm{F}(x)\| \, \|x\|$$

et $\|\nabla \mathrm{F}(x)\| \ge (\mathrm{F}(x) - \mathrm{F}(0))/\|x\|$, qui tend vers $+\infty$ par hypothèse. Il en résulte bien que (x_n) sera bornée si $(\nabla \mathrm{F}(x_n))$ est bornée, ce qu'il nous fallait au paragraphe précédent.

Si x_1 est fixé, on a que $\mathrm{F}^*(y) \ge y \,.\, x_1 - \mathrm{F}(x_1)$ pour tout y par définition, avec au point $y_1 = \nabla \mathrm{F}(x_1)$ l'égalité $\mathrm{F}^*(y_1) = y_1 \,.\, x_1 - \mathrm{F}(x_1)$, ce qui donne

$$\forall y \in \mathrm{E}, \quad (y - y_1) \,.\, x_1 \le \mathrm{F}^*(y) - \mathrm{F}^*(y_1)$$

et si $y = \nabla \mathrm{F}(x)$ on aura aussi en inversant les rôles de y et y_1

$$(y - y_1) \,.\, x_1 \le \mathrm{F}^*(y) - \mathrm{F}^*(y_1) \le (y - y_1) \,.\, x.$$

Quand y tend vers y_1, le point x tend vers x_1 et l'encadrement précédent montre que F^* est différentiable au point y_1, avec $x_1 = (\nabla \mathrm{F})^{-1}(y_1)$ comme gradient. D'après ce qui précède F^* est de classe C^1, et son application gradient est injective.

Puisque le gradient de F^* est injectif l'hyperplan d'appui en un point ne touche le graphe qu'en ce point : si $y_0 = \nabla x_0$, la fonction $\mathrm{G}(y) = \mathrm{F}^*(y) - \mathrm{F}^*(y_0) - (y - y_0) \,.\, x_0$ est ≥ 0 partout, nulle en y_0 et de classe C^1 ; cette valeur $\mathrm{G}(y_0) = 0$ est donc le minimum absolu de G sur E, et le gradient $\nabla \mathrm{G}(y) = \nabla \mathrm{F}^*(y) - x_0 = \nabla \mathrm{F}^*(y) - \nabla \mathrm{F}^*(y_0)$ s'annulerait en tout autre point y où $\mathrm{G}(y) = 0$, ce qui est impossible puisque $\nabla \mathrm{F}^*$ est injective ; on a donc $\mathrm{G}(y) > 0$ en tout point $y \ne y_0$; il en résulte en appliquant ceci au milieu $y_0 = (y_1 + y_2)/2$ d'un segment que F^* est strictement convexe.

Pour finir il reste à vérifier que $\mathrm{F}^*(y)/\|y\|$ tend vers $+\infty$ lorsque $\|y\| \to +\infty$. L'image $\mathrm{K_R}$ par $\nabla \mathrm{F}$ de la boule fermée de rayon R est un compact, donc il existe une constante $\mathrm{M(R)}$ telle que $\|y_0\| \le \mathrm{M(R)}$ et $\mathrm{F}^*(y_0) \le \mathrm{M(R)}$ pour tout $y_0 \in \mathrm{K_R}$. Choisissons y de norme $> \mathrm{M(R)}$, posons $x_0 = \mathrm{R}\, y/\|y\|$ et $y_0 = \nabla \mathrm{F}(x_0)$. On sait que

$$\mathrm{F}^*(y) \ge \mathrm{F}^*(y_0) + (y - y_0) \,.\, x_0 \ge \mathrm{R}\, \|y\| - |\mathrm{F}^*(y_0)| - \|y_0\| \, \|x_0\| \ge \mathrm{R}\, \|y\| - (\mathrm{R} + 1)\, \mathrm{M(R)}$$

ce qui montre le résultat voulu.

On aurait pu noter depuis longtemps que, sous nos hypothèses, F est la fonction conjuguée de F^*, ce qui complète le tableau de la réversibilité de la situation. On sait déjà que $\mathrm{F}(x) \ge x \,.\, y - \mathrm{F}^*(y)$ pour tous $x, y \in \mathrm{E}$, par définition de F^*, mais on a vu qu'il y a égalité si on prend $y = \nabla \mathrm{F}(x)$, donc

$$\mathrm{F}(x) = \sup_{y \in \mathrm{E}} \big(x \,.\, y - \mathrm{F}^*(y) \big) = (\mathrm{F}^*)^*(x).$$