Extrema sous contrainte

[1.] Soit U, un ouvert (non vide) d'un espace vectoriel E de dimension finie. On considère deux applications de classe $\mathscr{C}^1$

$$f : U \to \mathbb{R}, \qquad g : U \to \mathbb{R}$$

et on s'intéresse aux extrema éventuels de la fonction f restreinte à la partie

$$\Sigma = \big[g(M) = 0\big] \subset U.$$

En particulier, les applications linéaires tangentes

$$df(M_0) : E \to \mathbb{R}, \qquad dg(M_0) : E \to \mathbb{R}$$

sont des formes linéaires, quel que soit le point $M_0 \in U$.

[2.] Il s'agit d'illustrer la condition nécessaire vue en cours : si f atteint un extremum en un point $M_0 \in \Sigma$ tel que la forme linéaire tangente $dg(M_0)$ ne soit pas identiquement nulle, alors la forme linéaire tangente $df(M_0)$ est proportionnelle à $dg(M_0)$.

❧ Si l'espace E est muni d'une structure euclidienne, cela signifie que si le gradient $\nabla g(M_0)$ n'est pas nul, alors le gradient $\nabla f(M_0)$ est proportionnel à $\nabla g(M_0)$.

❧ Cette propriété ne permet pas de prouver que l'application f atteint un extremum au point $M_0 \in \Sigma$, mais permet, si elle atteint un extremum en un point de Σ, de déterminer les points où elle est susceptible d'atteindre cette valeur extrémale.

Présentation de la contrainte

[3.] On suppose ici que $U = E = \mathbb{R}^2$ et que la contrainte Σ est définie par la fonction g définie par

$$\forall\, M = (x, y) \in E, \qquad g(M) = x^2 + y^2 - 1$$

ou par, en considérant la structure euclidienne canonique de E,

$$\forall\, u \in E, \qquad g(u) = \|u\|^2 - 1.$$

❧ La fonction g est de classe $\mathscr{C}^1$ en tant que composée d'une application linéaire, d'une application bilinéaire et d'une application affine.

$$\begin{array}{ccccccc} E & \longrightarrow & E \times E & \longrightarrow & \mathbb{R} & \longrightarrow & \mathbb{R} \\ u & \longmapsto & (u, u) & \longmapsto & \langle u \,|\, u \rangle & \longmapsto & \|u\|^2 - 1 \end{array}$$

On en déduit l'expression de la forme linéaire tangente à g en un point $u_0 \in E$

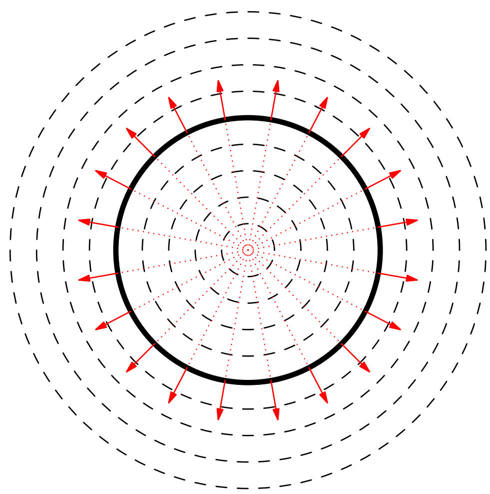
$$\forall\, h \in E, \qquad dg(u_0)(h) = 2 \,\langle u_0 \,|\, h \rangle$$

et donc l'expression du gradient :

$$\forall\, u_0 \in E, \qquad \nabla g(u_0) = 2 \cdot u_0.$$

❧ Le seul point critique pour g est donc l'origine $O = (0, 0)$ de E et comme $O \notin \Sigma$, le gradient de g est différent du vecteur nul en chaque point de la contrainte Σ.

On peut donc considérer la contrainte Σ comme une "courbe" de E et, comme on sait, en tout point $M_0 \in \Sigma$, le gradient de g est orthogonal à Σ.



Les courbes noires sont les lignes de niveau de la fonction g (c'est-à-dire les ensembles $[g(M) = K]$ pour différentes valeurs de $K \in \mathbb{R}$).
La courbe épaisse est la contrainte Σ (ligne de niveau $K = 0$).
Les flèches rouges représentent le gradient $\nabla g(M)$ en différents points M de Σ.

[4.] On considère une forme quadratique $f : \mathbb{R}^2 \to \mathbb{R}$:

$$\forall\, u \in \mathbb{R}^2, \qquad f(u) = \langle u \,|\, Au \rangle$$

où $A \in \mathfrak{M}_2(\mathbb{R})$. Cette application est de classe $\mathscr{C}^1$ (pour les mêmes raisons que g) et

$$\forall\, u \in \mathbb{R}^2, \qquad \nabla f(u) = 2 \cdot Au.$$

En particulier, le gradient $\nabla f(u)$ est proportionnel à u si, et seulement si, u est un vecteur propre de A (à moins que u ne soit le vecteur nul).

❧ Quitte à remplacer la matrice A par la matrice symétrique $\frac{1}{2}(A + A^\top)$, ce qui ne modifie pas l'application f, on peut supposer que $A \in \mathcal{S}_2(\mathbb{R})$ et donc que A est diagonalisable et qu'il existe une base orthonormée $(\varepsilon_1, \varepsilon_2)$ de $\mathbb{R}^2$ constituée de vecteurs propres pour A.

❧ Notons (X, Y), les coordonnées du vecteur $u \in \mathbb{R}^2$ relatives à la base orthonormée $(\varepsilon_1, \varepsilon_2)$:

$$u = X \cdot \varepsilon_1 + Y \cdot \varepsilon_2.$$

On en déduit que

$$f(u) = \lambda_1 X^2 + \lambda_2 Y^2$$

où λ_1 (resp. λ_2) est la valeur propre de A associée au vecteur propre ε_1 (resp. ε_2).

❧ On peut ainsi représenter la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ par son graphe

$$\Gamma = \big\{ \big(u, f(u)\big),\ u \in \mathbb{R}^2 \big\} \subset \mathbb{R}^2 \times \mathbb{R}$$

qui est aussi la surface d'équation cartésienne

$$z = \lambda_1 X^2 + \lambda_2 Y^2$$

dans la base orthonormée

$$\big((\varepsilon_1, 0), (\varepsilon_2, 0), e_3 = (0, 0, 1)\big).$$

On distingue alors quatre cas.

— Si les valeurs propres λ_1 et λ_2 sont de même signe (non nulles), alors la surface Γ est un **paraboloïde elliptique**.
— Si les valeurs propres λ_1 et λ_2 sont de signes opposées (non nulles), alors la surface Γ est un **paraboloïde hyperbolique**.
— Si une seule valeur propre est nulle, alors la surface Γ est un **paraboloïde cylindrique**.
— Si les deux valeurs propres sont nulles, alors la fonction f est constante et son graphe Γ est un plan horizontal.

❧ Il est plus simple de représenter la fonction f par ses **lignes de niveau**

$$\forall\, a \in \mathbb{R}, \qquad C_a = \big[f(u) = a\big]$$

qui sont des "courbes" planes.

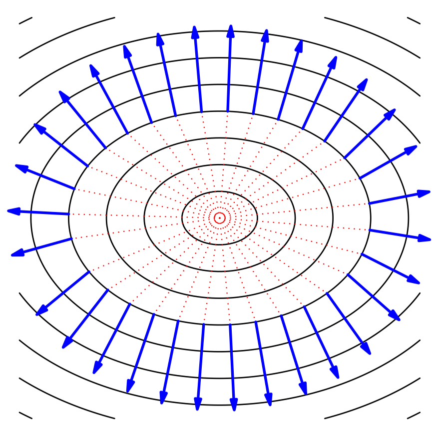
En éliminant le cas idiot où la forme quadratique f est l'application nulle, on distingue trois types de courbes planes.

— Si les valeurs propres sont de même signe, alors les lignes de niveau sont des **ellipses**.

$$C_a = \big[\lambda_1 X^2 + \lambda_2 Y^2 = a\big]$$

Si a, λ_1 et λ_2 sont de même signe, la ligne de niveau est une vraie ellipse :

$$C_a = \Big[\Big(\frac{X}{\alpha}\Big)^2 + \Big(\frac{Y}{\beta}\Big)^2 = 1\Big].$$



Le plan étant rapporté à une base orthonormée de vecteurs propres de A, les axes de coordonnées sont les axes de symétrie des ellipses.
Dans le cas présent, les deux valeurs propres de A sont strictement positives, donc la forme quadratique f atteint un minimum strict à l'origine (ce que confirme le sens du vecteur gradient).
On voit que le gradient, vecteur bleu colinéaire à Au, n'est en général pas colinéaire à u (dirigé par les pointillés rouges).

Dans les autres cas, la ligne de niveau est une ellipse dégénérée.
— Si $a = 0$, la ligne de niveau C_a est réduite à un point (l'origine).
— Si a est du signe opposé à celui de λ_1 et λ_2, alors la ligne de niveau C_a est vide :

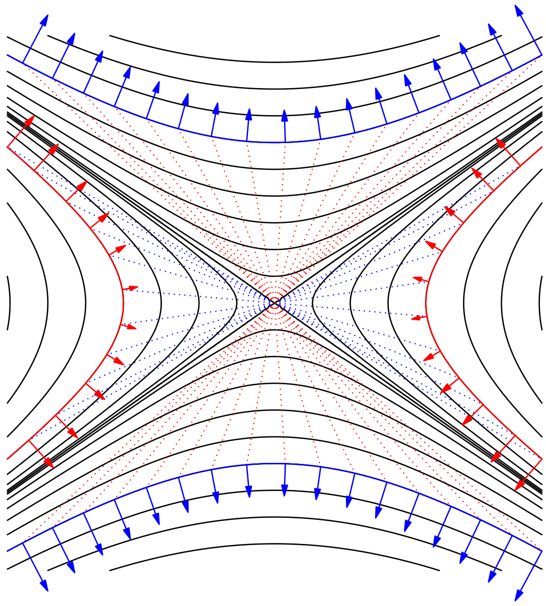
$$C_a = \left[\left(\frac{X}{\alpha}\right)^2 + \left(\frac{Y}{\beta}\right)^2 = -1\right].$$

— Si les valeurs propres sont de signes opposés, alors les lignes de niveau sont des **hyperboles**

$$C_a = \left[\left(\frac{X}{\alpha}\right)^2 - \left(\frac{Y}{\beta}\right)^2 = \pm 1\right]$$

sauf pour $a = 0$: la ligne de niveau C_0 est une paire de droites (les deux asymptotes de la famille d'hyperboles).

$$C_0 = \left[\left(\frac{X}{\alpha}\right)^2 - \left(\frac{Y}{\beta}\right)^2 = 0\right] = \left[\left(\frac{X}{\alpha} - \frac{Y}{\beta}\right)\left(\frac{X}{\alpha} - \frac{Y}{\beta}\right) = 0\right] = \left[\frac{X}{\alpha} - \frac{Y}{\beta} = 0\right] \cup \left[\frac{X}{\alpha} + \frac{Y}{\beta} = 0\right].$$

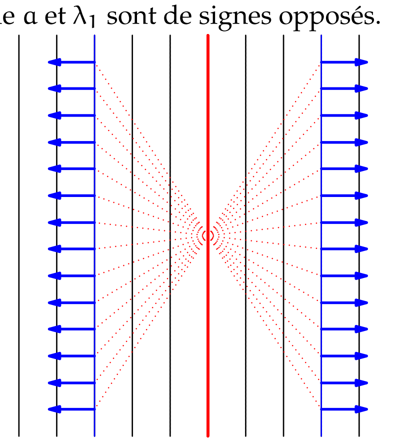


Les deux asymptotes forment la ligne de niveau 0.
En suivant la direction du gradient, on voit que l'hyperbole rouge est une ligne de niveau négatif et que l'hyperbole bleu est une ligne de niveau positif.

— Si seule la valeur propre λ_2 est nulle, alors les lignes de niveau sont des paires de droites parallèles lorsque a est du signe de λ_1 :

$$C_a = [\lambda_1 X^2 = a] = \left[X = \sqrt{\frac{a}{\lambda_1}}\right] \cup \left[X = -\sqrt{\frac{a}{\lambda_1}}\right]$$

ou une seule droite : $C_0 = [X = 0]$ ou l'ensemble vide (lorsque a et λ_1 sont de signes opposés.



Sur cette figure, la valeur propre λ_1 est strictement positive et la forme quadratique f est minimale sur l'axe des ordonnées (droite rouge).
On doit noter que le gradient de f est nul le long de cette droite.

[5.] Pour visualiser le comportement de la forme quadratique f lorsqu'on se restreint à la courbe Σ, on superpose la contrainte Σ (en trait noir épais) à quelques lignes de niveau de f.

En différents points M de Σ, on peut alors tracer le gradient de la fonction f (vecteur bleu) et celui de la contrainte g (vecteur rouge).

On constate que les deux gradients sont colinéaires aux points où la courbe Σ rencontre les lignes de niveau extrêmes (minimum et maximum de f sur la courbe Σ).

La colinéarité des gradients se traduit par le fait que, en ces points, les lignes de niveau de f et de g soient tangentes.