

ENTROPIE ET TRANSVERSALITÉ QUANTITATIVE

1. Introduction. Une loi mathématique décrivant un phénomène physique, comme la variation d'une valeur y à l'aide des paramètres $x_i \in \mathbb{R}, i \in \{1, \dots, n\}$, est en général d'une des trois formes suivantes :

(a) dans les cas les plus simples (ie exceptionnellement), une équation du type $y = F(x_1, \dots, x_n)$,

(b) ou une équation implicite $G(y, x_1, \dots, x_n) = 0$,

(c) ou plus généralement une équation différentielle :

$$H\left(y, \frac{\partial^{|\alpha_1|} y}{\partial x_{i_1} \cdots \partial x_{i_{\alpha_1}}}, \dots, \frac{\partial^{|\alpha_k|} y}{\partial x_{j_1} \cdots \partial x_{j_{\alpha_k}}}, x_1, \dots, x_n\right) = 0$$

+ initial values.

- Dans le premier cas F représente une quantité d'information trop importante. La connaissance de la valeur y^0 en (x_1^0, \dots, x_n^0) et de $\nabla F_{(x_1^0, \dots, x_n^0)} = \left(\frac{\partial F}{\partial x_1}, \dots, \frac{\partial F}{\partial x_n}\right)(x_1^0, \dots, x_n^0)$ suffit, par approximation linéaire, à prédire le comportement de y pour (x_1, \dots, x_n) proche de (x_1^0, \dots, x_n^0) .

- Dans le cas (b), Il s'agit d'abord de savoir si

$$M = \{(y, x_1, \dots, x_n) \in \mathbb{R}^{n+1}; G(y, x_1, \dots, x_n) = 0\}$$

est localement en $(y^0, x_1^0, \dots, x_n^0)$ le graphe d'une fonction $(x_1, \dots, x_n) \mapsto y = F(x_1, \dots, x_n)$, comme en (a), pour obtenir une approximation linéaire à l'aide de $\nabla F_{(x_1^0, \dots, x_n^0)}$. Si l'on connaît *a priori* $(y^0, x_1^0, \dots, x_n^0) \in M$, cette réduction est possible lorsque :

$$\frac{\partial G}{\partial y}(y^0, x_1^0, \dots, x_n^0) \neq 0.$$

Et : $\nabla F_{(x_1^0, \dots, x_n^0)}$

$$= -\left(\frac{\partial G}{\partial x_1}, \dots, \frac{\partial G}{\partial x_n}\right)(y^0, x_1^0, \dots, x_n^0) / \frac{\partial G}{\partial y}(y^0, x_1^0, \dots, x_n^0).$$

Nos données comportant une imprécision, plus grand est $\left|\left(\frac{\partial G}{\partial x_1}, \dots, \frac{\partial G}{\partial x_n}\right)(y^0, x_1^0, \dots, x_n^0) / \frac{\partial G}{\partial y}(y^0, x_1^0, \dots, x_n^0)\right|$

plus mauvaise est l'approximation de y au voisinage de y^0 .

De plus, si M est le graphe $(x_1, \dots, x_n) \mapsto y = F(x_1, \dots, x_n)$, l'expression de $\nabla F_{(x_1, \dots, x_n)}$ est approchée par un algorithme et plus $\frac{\partial G}{\partial y}(y^0, x_1^0, \dots, x_n^0)$ est proche de 0, plus celui-ci est instable.

- De même dans le cas (c), nous devons éviter les petites valeurs de certains gradients pour obtenir des informations fiables.

Conclusion. Nous devons non seulement éviter l'annulation de gradients mais contrôler leur distance à 0.

L'équivalent géométrique de cette condition analytique est la *transversalité*. En effet $\frac{\partial G}{\partial y}(y^0, x_1^0, \dots, x_n^0) \neq 0$ équivaut à :

$$\nabla G_{(y^0, x_1^0, \dots, x_n^0)} \oplus Ox_1 \cdots x_n = \mathbb{R}^{n+1},$$

ou :

$$T_{(y^0, x_1^0, \dots, x_n^0)} M \oplus Oy = \mathbb{R}^{n+1}.$$

On dit que $\nabla G_{(y^0, x_1^0, \dots, x_n^0)}$ est *transverse* aux paramètres $Ox_1 \cdots x_n$ en $(y^0, x_1^0, \dots, x_n^0)$, ou que M est transverse à Oy en $(y^0, x_1^0, \dots, x_n^0)$.

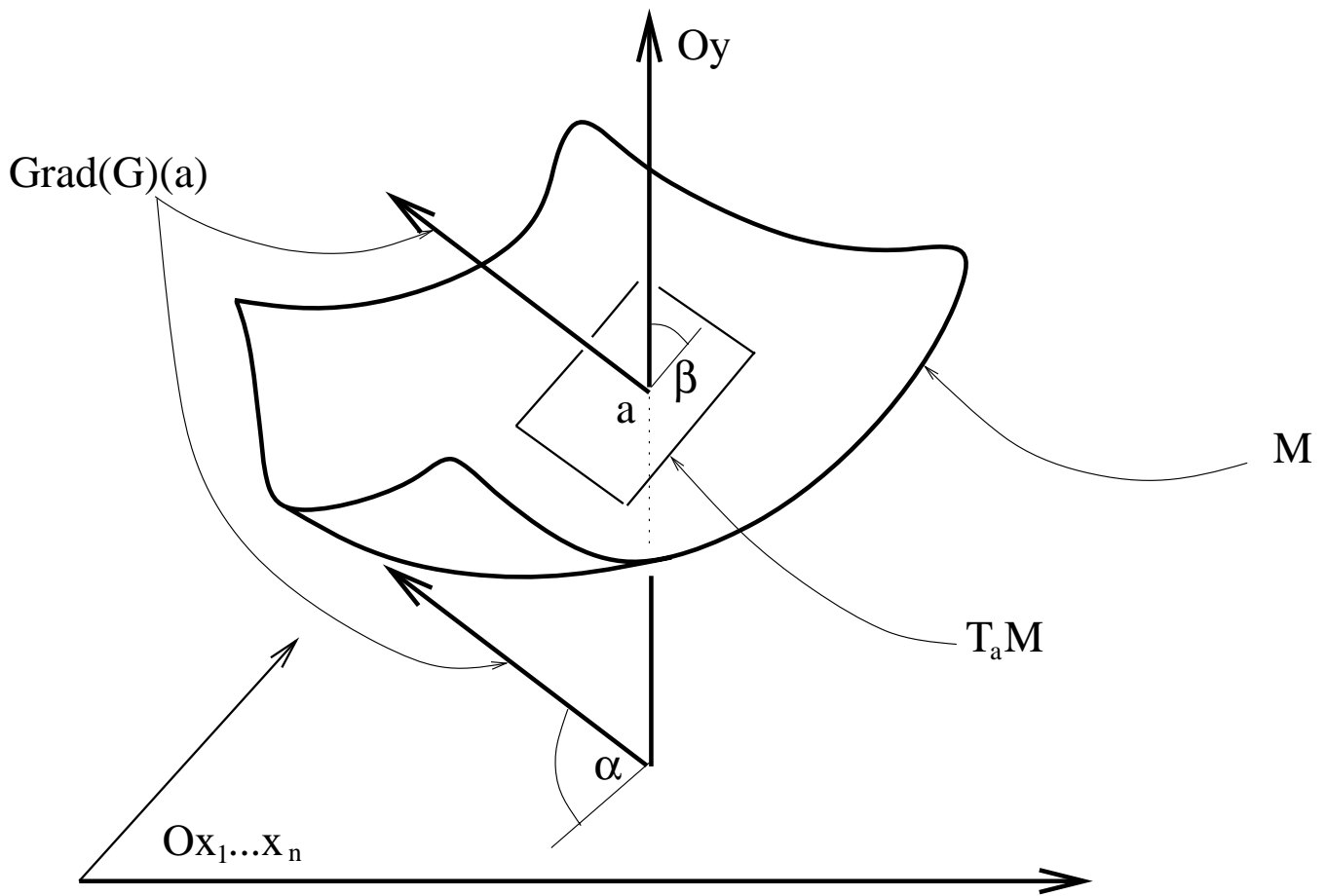


fig. 1 Transversalité de la variété M et de Oy

La condition :

$$\left| \left(\frac{\partial G}{\partial x_1}, \dots, \frac{\partial G}{\partial x_n} \right) (y^0, x_1^0, \dots, x_n^0) / \frac{\partial G}{\partial y} (y^0, x_1^0, \dots, x_n^0) \right| \geq 1/\epsilon$$

signifie que l'angle $\alpha = \widehat{(\nabla G_{(y^0, x_1^0, \dots, x_n^0)}, Ox_1 \cdots x_n)}$

ou que $\beta = (T_{(y^0, x_1^0, \dots, x_n^0)} \widehat{M}, Oy)$ est plus petit que ϵ (*fig. 1*).

Nous allons quantifier le défaut de transversalité pour presque-génériquement éviter les mauvaises situations décrites.

2. Transversalité quantifiée.

Soit une application linéaire $L : \mathbb{R}^n \rightarrow \mathbb{R}^p$. L'image par L de $B^n(0, 1)$ est un r -ellipsoïde in \mathbb{R}^p de demi-axes $l_1(L) \geq \dots \geq l_r(L)$, où r est le rang de L . Si $r < p$, on note $l_{r+1}(L) = 0, \dots, l_p(L) = 0$.

Soient M et N deux sous-variétés de \mathbb{R}^n , $x \in M \cap N$, $\pi : \mathbb{R}^n \rightarrow T_x N^\perp$, $p = n - \dim(N)$ et $\pi|_M$ la restriction de π à M .

Définitions. On dit que (M, N) est *transverse* à x , et on note $M \pitchfork_x N$, ssi $\pi|_M$ est une submersion en x , ie $D\pi|_{M(x)} : T_x M \rightarrow T_x N^\perp$ est surjective.

Pour $\Lambda = (\epsilon_1, \dots, \epsilon_p)$, $\epsilon_1 \geq \dots \geq \epsilon_p$, on dit que (M, N) est Λ -*non-transverse* en x , et on note $M \not\pitchfork_x^\Lambda N$, ssi $l_i(D\pi|_{M(x)}) \leq \epsilon_i, \forall i \in \{1, \dots, p\}$.

$M \not\pitchfork_x N$ ssi $x \notin M \cap N$ ou $M \not\pitchfork_x^\Lambda N$ pour Λ avec $\epsilon_p = 0$. Plus (M, N) est Λ -non-transverse avec Λ proche

de $(\epsilon_1, \dots, \epsilon_{p-1}, 0)$, moins M et N semblent transverses en $x \in M \cap N$ (*fig. 2*).

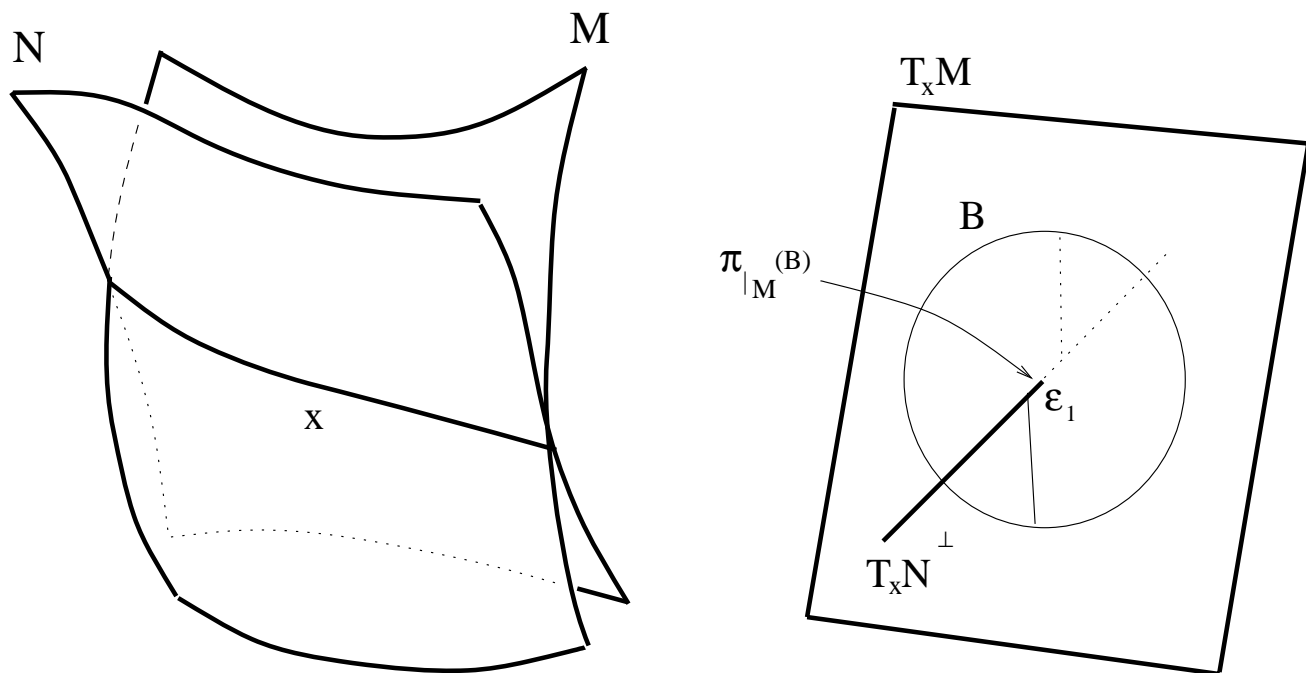


fig. 2 Presque-non-transversalité de M et N

Soient X, Y deux variétés, $f : X \rightarrow Y$ lisse, $N \subset Y$ une sous-variété de Y de codimension p , $y \in N$, et $\Phi : \mathcal{O} \rightarrow \mathbb{R}^p$ submersive, où \mathcal{O} est un voisinage ouvert de x dans Y , tel que $\Phi^{-1}(\{0\}) = N \cap \mathcal{O}$. On dit que (f, N) est transverse en x , et on note $f \pitchfork_x N$, ssi $f \circ \Phi$ est submersive en x .

Pour $\Lambda = (\epsilon_1, \dots, \epsilon_p)$, $\epsilon_1 \geq \dots \geq \epsilon_p$, on dit que (f, N) est (Φ, Λ) -non-transverse en x , et on note $f \not\pitchfork_x^{(\Phi, \Lambda)} N$, ssi $l_i(D[f \circ \Phi]_{(x)}) \leq \epsilon_i$, $\forall i \in \{1, \dots, p\}$.

On reconnaît la définition de la transversalité et de la Λ -non-transversalité de deux sous-variétés M, N de \mathbb{R}^n en posant $f : M \rightarrow \mathbb{R}^n$ l'inclusion et $\Phi = \pi$.

Notre problème est alors :

“Pour une application \mathcal{C}^k -régulière ($k \in \mathbb{N} \cup \{\infty\}$) $f : \mathbb{R}^n \rightarrow \mathbb{R}^p$ et pour $\Lambda = (\epsilon_1, \dots, \epsilon_p)$, quelle est la taille de $\Delta(f, B_r, \Lambda) = f(\Sigma(f, B_r, \Lambda))$, où $\Sigma(f, B_r, \Lambda) = \{x \in B_r \subset \mathbb{R}^n; l_i(Df_{(x)}) \leq \epsilon_i, \forall i \in \{1, \dots, p\}\}$?”

Le “mauvais” ensemble $\Delta(f, B_r, \Lambda)$ est appelé l'ensemble des valeurs Λ -presque critiques de f .

On peut contrôler sa taille en fonction de k et Λ .

3. Mesure et dimensions. Soit A un sous-ensemble d'un espace métrique et $\alpha \geq 0$. Posons :

$$\mathcal{A}_\nu = \{(D_i)_{i \in \mathbb{N}}; A \subset \bigcup_{i \in \mathbb{N}} D_i \text{ and } |D_i| \leq \nu\},$$

$$\mathcal{H}_\nu^\alpha(A) = \inf \left\{ \sum_{i \in \mathbb{N}} |D_i|^\alpha; (D_i)_{i \in \mathbb{N}} \in \mathcal{A}_\nu \right\}$$

et

$$\mathcal{H}^\alpha(A) = \lim_{\nu \rightarrow 0} \mathcal{H}_\nu^\alpha(A) \in \mathbb{R} \cap \{\infty\}.$$

$\mathcal{H}^\alpha(A)$ est la mesure de Hausdorff α -dimensionnelle de A . Quand $\mathcal{H}^\alpha(A) \neq \infty$, $\mathcal{H}^{\alpha'}(A) = 0$ pour $\alpha' > \alpha$ et quand $\mathcal{H}^\alpha(A) \neq 0$, $\mathcal{H}^{\alpha'}(A) = \infty$ pour $\alpha' < \alpha$. On définit :

$$\dim_{\mathcal{H}}(A) = \inf\{\alpha; \mathcal{H}^\alpha(A) = 0\} = \sup\{\alpha; \mathcal{H}^\alpha(A) = \infty\}.$$

Notons $M(\alpha, A)$ le nombre minimal de boules fermées de rayon $\leq \alpha$, nécessaire pour couvrir A . $H_\alpha(A) = \log_2(M(\alpha, A))$ est l' α -entropie de A .

$$\dim_e(A) = \limsup_{\alpha \rightarrow 0} \frac{\log(M(\alpha, A))}{\log(1/\alpha)}$$

$$= \inf\{\delta; M(\alpha, A) \leq (1/\alpha)^\delta, \text{ pour } \alpha \ll 1\}.$$

Clairement : $\dim_{\mathcal{H}}(A) \leq \dim_e(A)$ et ([Iva], [Yom-Com]):

$$M(\alpha, A) \leq c(n) \sum_{i=0}^n V_i(A)(1/\alpha)^i,$$

où $c(n)$ ne dépend que de n et $V_i(A)$ (la i -ème variation de A).

4. La transversalité est la situation générique.

Theorem 1 (Morse-Sard C^∞) [Mors], [Sar 1], [Hol]. — Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}^p$ une application C^∞ . Alors $\mathcal{H}^p(\Delta(f, B_r)) = 0$, où $\Delta(f, B_r) = f(\Sigma(f, B_r))$ et $\Sigma(f, B_r)$ est l'ensemble des points de $x \in B_r$ où $\text{rang}(Df_{(x)}) < p$.

Soit $\Delta^\rho(f, B_r) = f(\Sigma^\rho(f, B_r))$ avec $\Sigma^\rho(f, B_r) = \{x \in B_r \subset \mathbb{R}^n; \text{rang}(Df_{(x)}) \leq \rho\}$.

Theorem 2 (Morse-Sard C^k pour les points de rang r) [Fed], [Mor]. — Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}^p$ une application C^k . Alors $\mathcal{H}^{\rho + \frac{n-\rho}{k}}(\Delta^\rho(f, B_r)) = 0$.

EN particulier : $\dim_{\mathcal{H}}(\Delta^\rho(f, B_r)) \leq \rho + \frac{n-\rho}{k}$.

Cette borne est la meilleure (voir [Com], [Whi 1], [Gri], [Yom-Com]). Ces résultats ne disent rien sur la dimension d'entropie de valeurs critiques.

5. Almost-transversality is almost generic.

Ici $f : \mathbb{R}^n \rightarrow \mathbb{R}^p$ est une application \mathcal{C}^k . On note K une constante de Lipschitz de $D^{k-1}f$ sur B_r et $R_k(f) = \frac{K}{(k-1)!} \cdot r^k$.

Théorème 3 (\mathcal{C}^k Morse-Sard quantitatif) [Yom], [Yom-Com]. — Soit $\Lambda = (\epsilon_1, \dots, \epsilon_p)$, $\epsilon_1 \geq \dots \geq \epsilon_p$, et $\epsilon_0 = 1$. On a (pour $\alpha \leq R_k(f)$):

$$M(\alpha, \Delta(f, B_r, \Lambda)) \leq C \cdot \sum_{i=0}^p \epsilon_0 \cdots \epsilon_i \left(\frac{r}{\alpha}\right)^i \left(\frac{R_k(f)}{\alpha}\right)^{\frac{n-i}{k}},$$

où C est une constante ne dépendant que de n , p et k .

Théorème 4 (\mathcal{C}^k Morse-Sard pour l'entropie) [Yom], [Yom-Com]. — Soit $\epsilon_i = \sup\{\lambda_i(Df(x)); x \in B_r\}$, pour $i \in \{1, \dots, \rho\}$. On a (pour $\alpha \leq R_k(f)$):

$$M(\alpha, \Delta^\rho(f, B_r)) \leq C \cdot \sum_{i=0}^{\rho} \epsilon_0 \cdots \epsilon_i \left(\frac{r}{\alpha}\right)^i \left(\frac{R_k(f)}{\alpha}\right)^{\frac{n-i}{k}},$$

où C est une constante de dépendant que de n , p et k .

En particulier :

$$\dim_{\mathcal{H}}(\Delta^\rho(f, B_r)) \leq \dim_e(\Delta^\rho(f, B_r)) \leq \rho + \frac{n - \rho}{k}.$$

Cette borne est la meilleure (voir [Yom-Com]).

6. Examples. Soit A l'ensemble des polynômes réels de degré d et du type : $x \mapsto Q(a, x) = 1 + \sum_{j=1}^d a_j x^j$, avec $a = (a_1, \dots, a_d)$ et $\|a\| \leq 1$. On identifie l'ensemble A avec $B^d(0, 1) = \{a \in \mathbb{R}^d; \|a\| \leq 1\}$.

On veut majorer l' α -entropie de l'ensemble des polynômes de A pour lesquels *les racines réelles sont multiples ou presque-multiples*.

Soit $V = \{(a, x) \in \mathbb{R}^{d+1}; Q(a, x) = 0\}$. Aux points (a, x) de V avec $\nabla Q_{(a,x)} \neq 0$, V est une \mathcal{C}^∞ sous-variété de codimension 1 de \mathbb{R}^{d+1} . On note $V^{reg} = \{(a, x) \in V; \nabla Q_{(a,x)} \neq 0\}$ et $V^{sing} = \{(a, x) \in V; \nabla Q_{(a,x)} = 0\} = V \setminus V^{reg}$. Par [Whi 2], V^{sing} est réunion de sous-variétés lisses de dimension $\leq d - 1$.

Une racine x du polynôme $Q(a, \cdot)$ est multiple ssi $Q(a, x) = \frac{\partial Q}{\partial x}(a, x) = 0$. Donc l'ensemble A^Σ des polynômes de A ayant des racines multiples est $\pi(V^{sing}) \cup \Delta(\pi|_{V^{reg}})$, où $\pi : \mathbb{R}^{d+1} \rightarrow \mathbb{R}^d$ est la projection $\pi(a, x) = a$, et $\Delta(\pi|_{V^{reg}})$ est l'ensemble $\{(a, x) \in V^{reg}; Ox \subset T_{(a,x)}V^{reg}\}$ des valeurs critiques

de $\pi|_{V^{reg}}$. Par le théorème de Sard (Théorème 2), $\dim_{\mathcal{H}}(\Delta(\pi|_{V^{reg}})) \leq d - 1$. Comme $\dim_{\mathcal{H}}(\pi(V^{sing})) \leq d - 1$, on obtient : $\dim_{\mathcal{H}}(A^{\Sigma}) \leq d - 1$: avoir des racines distinctes est générique.

Posons $\Lambda = (\epsilon_1, \dots, \epsilon_d)$ avec $\epsilon_1 \geq \dots \geq \epsilon_d$ et $\epsilon_0 = 1$. Une racine x du polynôme $Q(a, \cdot) \in A$ est Λ -presque multiple ssi $Q(a, x) = 0$ et $V \not\perp_x^{\Lambda} Ox$, ie $(a, x) \in V^{sing}$ ou $\sin(T_{(a,x)}V^{reg}, Ox) \leq \epsilon_d$. Cette condition porte sur ϵ_d et on peut prendre $\epsilon_1 = \dots = \epsilon_{d-1} = 1$. Soit $A^{\Sigma, \Lambda}$ l'ensemble des polynômes de A avec (au moins) une racine Λ -presque multiple. Par le Théorème 3,

$$M(\alpha, A^{\Sigma, \Lambda} \setminus \pi(V^{sing})) \leq C \cdot \left[\sum_{i=0}^{d-1} \left(\frac{1}{\alpha}\right)^i + \epsilon_d \cdot \left(\frac{1}{\alpha}\right)^d \right].$$

Mais $\pi(V^{sing})$ est une union finie de variété lisses de dimension au plus $d - 1$, on obtient : $M(\alpha, A^{\Sigma, \Lambda}) \leq$

$$C' \cdot \left[\sum_{i=0}^{d-1} \left(\frac{1}{\alpha}\right)^i + \epsilon_d \cdot \left(\frac{1}{\alpha}\right)^d \right].$$

Avoir une racine Λ -presque multiple est une propriété Λ -presque générique.

Dans la figure 3 on représente V pour $d = 3$ et $a_3 = 1$, $W = \{(a, x) \in \mathbb{R}^{d+1}; \frac{\partial Q}{\partial x}(a, x) = 0\}$.

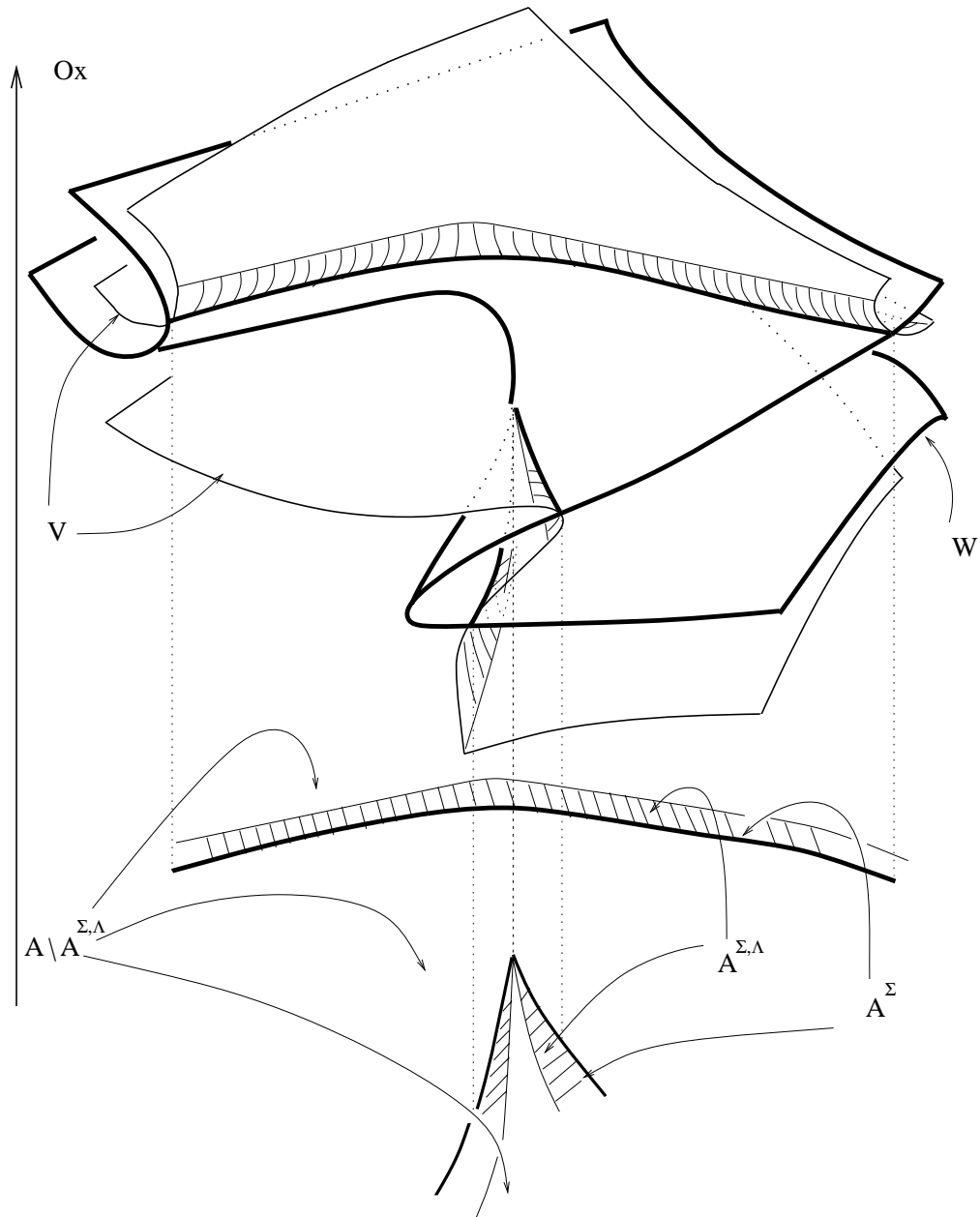


fig. 3 L'espace des polynômes du type $1+a_1x+a_2x+x^3$
avec des racines presque multiples

L'exemple suivant vient de la robotique : considérons un manipulateur dans le plan, constitué de deux axes articulés de longueur a et b , le point P est fonction des angles ϕ et ψ . La distance r de P à l'origine est : $r^2 = \|P\|^2 = a^2 + b^2 + 2ab \cos(\psi)$. Les points critiques de r sont donnés par : $\frac{dr}{d\psi}(\psi) = -2ab \sin(\psi) = 0$ et correspondent au cercle $\psi = 0$. La valeur critique de r est $a + b$. Près des positions critiques le contrôle de r à l'aide de ψ est mauvais; on voudrait éviter ces positions presque-critiques. Soit $\epsilon > 0$, la condition $|\frac{dr}{d\psi}(\psi)| \leq \epsilon$ implique $|\psi| \leq \arcsin(\epsilon/2ab)$, et les valeurs ϵ -presque critiques de r sont : $r_{max}^2 - r^2 \leq 2ab[1 - \cos(\arcsin(\epsilon/2ab))]$, où r_{max} est $a + b$; elles sont donc contenues dans un intervalle de longueur $\leq \frac{c \cdot \epsilon^2}{4ab \cdot r_{max}}$, et $M(\alpha, \Delta(r, \epsilon)) \leq \frac{c \cdot \epsilon^2}{4ab \cdot r_{max} \cdot \alpha}$ (le Théorème 3 donne : $M(\alpha, \Delta(r, \epsilon)) \leq C(1 + \frac{\epsilon}{\alpha})$).

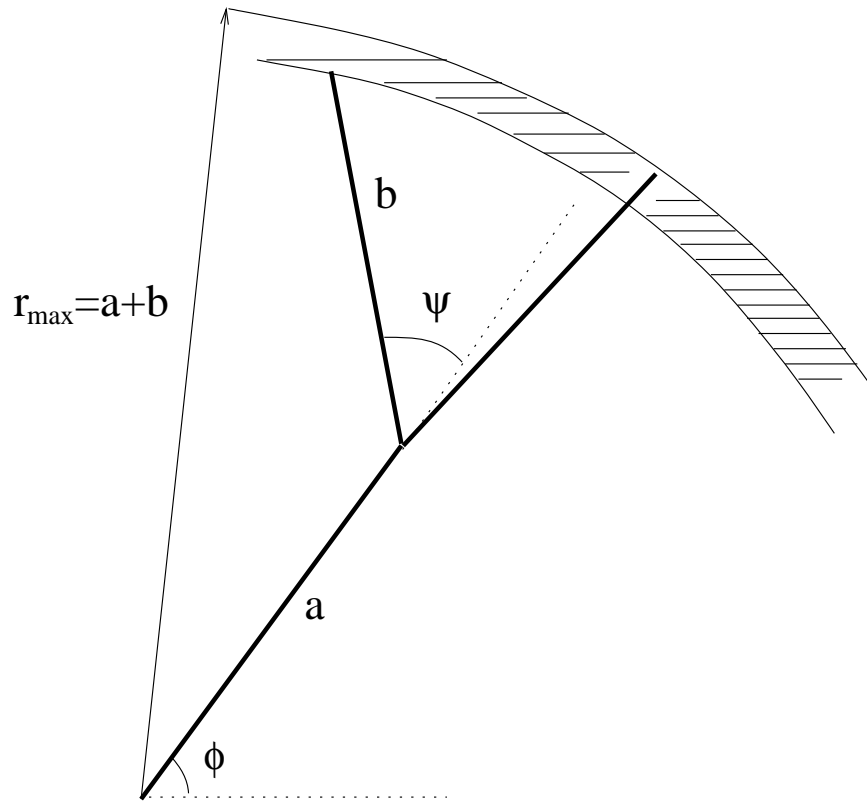


fig. 4 Points presque-critiques de la distance de P à l'origine