Entropy and Quantitative Transversality

G Comte, Université de Nice Sophia Antipolis, Nice, France

Introduction

A mathematical law for a physical phenomenon, describing the variation of a value $y(\in \mathbb{R})$ in terms of parameters $x_i \in \mathbb{R}$, $i \in \{1, \dots, n\}$, is usually given:

1. in the simplest cases (and hence in exceptional cases), by an explicit functional equation $y = F(x_1, \dots, x_n)$, or
2. by an implicit equation $G(y, x_1, \dots, x_n) = 0$, or
3. more generally, by a partial differentiable equation,

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

In the first case, the exact equation $y = F(x_1, \dots, x_n)$ fully describes the behavior of y as $(x_1, \dots, x_n)$ vary, but in practice this information is too substantive: using the Taylor formula, knowledge of the value y^0 at some point $(x_1^0, \dots, x_n^0)$ and of the value of

$$\nabla F_{(x_1^0, \dots, x_n^0)} = \left(\frac{\partial F}{\partial x_1}, \dots, \frac{\partial F}{\partial x_n}\right)(x_1^0, \dots, x_n^0)$$

is enough to predict, with controlled accuracy, by linear approximation, the behavior of y for parameters $(x_1, \dots, x_n)$ close to $(x_1^0, \dots, x_n^0)$.

In the case (2), both the parameters $(x_1, \dots, x_n)$ and the value y belong to the set $M = \{(y, x_1, \dots, x_n) \in \mathbb{R}^{n+1}; G(y, x_1, \dots, x_n) = 0\}$, and we would like to know whether or not this set may be (at least locally around one of its point $(y^0, x_1^0, \dots, x_n^0)$) a graph of some function $(x_1, \dots, x_n) \mapsto y = F(x_1, \dots, x_n)$, as in the case (1). Using the implicit function theorem, we may try to reduce our equation to the explicit equation of (1), and then perform a linear approximation involving $\nabla F_{(x_1^0, \dots, x_n^0)}$. Assuming that *a priori* we know a value y^0 such that for $(x_1^0, \dots, x_n^0)$, $(y^0, x_1^0, \dots, x_n^0) \in M$, this reduction is possible, locally around $(y^0, x_1^0, \dots, x_n^0)$, under the condition that

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

In this situation

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

Now, as it is normally the case, when they come from observation, the variables $x_1, \dots, x_n$ are known with an estimate and one sees that the larger

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

is, the worse the estimate on y near y^0.

Furthermore, assuming that M is locally a graph of a function $(x_1, \dots, x_n) \mapsto y = F(x_1, \dots, x_n)$, for a given $(x_1, \dots, x_n)$, the exact expression of $y = F(x_1, \dots, x_n)$ and consequently the exact value of $\nabla F_{(x_1, \dots, x_n)}$ is not possible to obtain; we have to approach it using an algorithm (classically the Newton algorithm), and closer

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

is to 0, the more such an algorithm is unstable.

Finally, in the case (3), skipping technical details, we encounter the same type of difficulties: we have to avoid small values for some gradient functions at a given point, in order to obtain, locally at some point $(x_1^0, \dots, x_n^0)$, in a stable way, reliable information on y in terms of $(x_1, \dots, x_n)$.

To sum up, the prediction of a physical phenomenon by a mathematical law greatly depends not only on the noncancellation of some gradient functions, but, as we deal with approximations and algorithms, on how different those gradient functions are from zero.

This principle, of course, extends directly to applied problems (see the last of our examples in the final section): being close to singular values essentially means that the control (e.g., of the positions of some device by a manipulator) is poor.

The geometric counterpart of this analytic phenomenmon is called "transversality," the condition for some function G to have a nonzero partial derivative

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

is equivalent to the condition

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

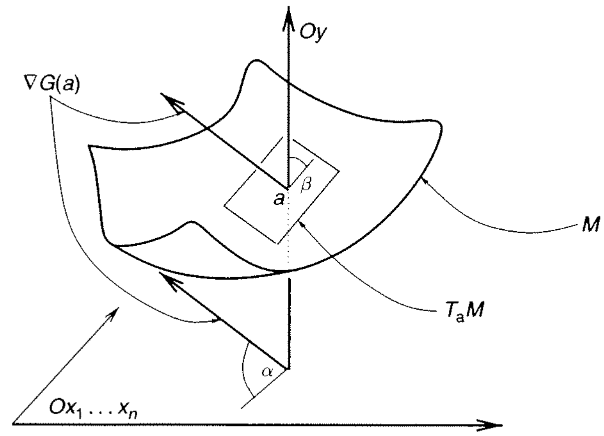

Figure 1 Transversality of the manifold M and Oy.

or to the condition

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

where T_aM is the tangent space of M at $a \in M$.

We say that $\nabla G_{(y^0,x_1^0,\ldots,x_n^0)}$ is transverse to the space of parameters $Ox_1 \cdots x_n$ at $(y^0, x_1^0, \ldots, x_n^0)$, or that M is transverse to Oy at $(y^0, x_1^0, \ldots, x_n^0)$.

For some quantity $\epsilon > 0$, the condition that

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

means that the angle $\alpha = (\nabla G_{(y^0,x_1^0,\ldots,x_n^0)}, \widehat{Ox_1 \cdots x_n})$ or the angle $\beta = (T_{(y^0,x_1^0,\ldots,x_n^0)} \widehat{M, Oy})$ is smaller than ϵ (see **Figure** 1).

Our purpose in the sequel is to indicate how we can quantify the situations described above (the defect of transversality), in order to generically or almost generically avoid them with quantified accuracy.

Quantifying Transversality

Given two submanifolds M and N of the Euclidean space $\mathbb{R}^n$, we can measure the transversality defect of (M, N) at $x \in \mathbb{R}^n$ with a differential criterion, both analytical and geometric.

Let us first introduce some notations. For a given linear map $L : \mathbb{R}^n \to \mathbb{R}^p$, the image by L of the unit ball of $\mathbb{R}^n$ is an r-dimensional ellipsoid in $\mathbb{R}^p$ with semi-axes denoted as $l_1(L) \geq \cdots \geq l_r(L)$, where r is the rank of L. For $r < p$, we denote $l_{r+1}(L) = 0, \ldots, l_p(L) = 0$.

Now, let $x \in M \cap N$; let $\pi : \mathbb{R}^n \to T_xN^\perp$ be the projection onto the orthogonal space of T_xN, $p = n - \dim(N)$ and $\pi_{|M}$ the restriction of π to M.

Definitions

We say that (M, N) is transverse at x, and we denote it by $M \pitchfork_x N$, if and only if $\pi_{|M}$ is a submersion at x, that is, $D\pi_{|M(x)} : T_xM \to T_xN^\perp$ is onto.

For a given $\Lambda = (\epsilon_1, \ldots, \epsilon_p)$, $\epsilon_1 \geq \cdots \geq \epsilon_p$, we say that (M, N) is Λ-nontransverse at x, and we denote it by $M \not\pitchfork_x^\Lambda N$, if and only if $l_i(D\pi_{|M(x)}) \leq \epsilon_i, \forall\, i \in \{1, \ldots, p\}$.

With these notations, we have: $M \not\pitchfork_x N$ (*i.e.*, (M, N) nontransverse at x) if and only if $x \notin M \cap N$ or $M \not\pitchfork_x^\Lambda N$, for some Λ with $\epsilon_p = 0$, and the more (M, N) is Λ-nontransverse, with Λ close to $(\epsilon_1, \ldots, \epsilon_{p-1}, 0)$, the less the manifolds M and N seem transverse at $x \in M \cap N$ (see Figure 2).

The final step in our formalism to give a convenient quantitative approach of transversality is the following: let X, Y be two (real) Riemannian manifolds, $f : X \to Y$ a (smooth) mapping, $N \subset Y$ a submanifold of Y with codimension p in N, $y \in N$, and $\Phi : \mathcal{O} \to \mathbb{R}^p$ a submersion, where $\mathcal{O}$ is an open neighborhood of x in Y, such that $\Phi^{-1}(\{0\}) = N \cap \mathcal{O}$. Then we say that (f, N) is transverse at x, and we denote it by $f \pitchfork_x N$, if and only if $f \circ \Phi$ is submersive in x.

For a given $\Lambda = (\epsilon_1, \ldots, \epsilon_p)$, $\epsilon_1 \geq \cdots \geq \epsilon_p$, we say that (f, N) is (Φ, Λ)-nontransverse at x, and we denote it by $f \not\pitchfork_x^{(\Phi,\Lambda)} N$, if and only if $l_i(D[f \circ \Phi]_{(x)}) \leq \epsilon_i, \forall\, i \in \{1, \ldots, p\}$.

Clearly, we recognize the definition of transversality and of Λ-nontransversality of two submanifolds M, N of $\mathbb{R}^n$ by letting $f : M \to \mathbb{R}^n$ be the inclusion and $\Phi = \pi_{|M}$ (for more details on transversality and stability, see, e.g., Golubitski and Guillemin (1973)).

With the definitions and notations above, our general problem may be posed as follows:

> For a C^k-regular ($k \in \mathbb{N} \cup \{\infty\}$) mapping $f : \mathbb{R}^n \to \mathbb{R}^p$ and a given $\Lambda = (\epsilon_1, \ldots, \epsilon_p)$, how large is the set $\Delta(f, B_r, \Lambda) = f(\Sigma(f, B_r, \Lambda))$, where $\Sigma(f, B_r, \Lambda) = \{x \in B_r \subset \mathbb{R}^n;\ l_i(Df_{(x)}) \leq \epsilon_i,\ \forall i \in \{1, \ldots, p\}\}$ and B_r is a ball of radius r in $\mathbb{R}^n$?

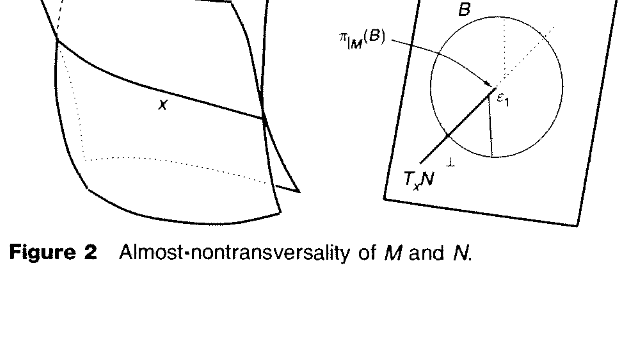

Figure 2 Almost-nontransversality of M and N.

The "bad" set $\Delta(f, B_r, \Lambda)$ is called the set of Λ-almost critical values of f (restricted to B_r). Our purpose is to show that one can control its size in terms of k and Λ. However, before explicitly stating quantitative results, let us precise what we understand by "big set" or by "size of a set."

Measure and Dimensions

We have a very natural way to measure a subset A of a metric space. To do this, we consider $\alpha \geq 0$ a real number and we denote

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

where $|D_i|$ is the diameter of D_i,

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

and

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

$\mathcal{H}^\alpha(A)$ is called the α-dimensional Hausdorff measure of A. It appears that when $\mathcal{H}^\alpha(A) \neq \infty$, $\mathcal{H}^{\alpha'}(A) = 0$ for $\alpha' > \alpha$, and when $\mathcal{H}^\alpha(A) \neq 0$, $\mathcal{H}^{\alpha'}(A) = \infty$ for $\alpha' < \alpha$. This gives rise to the following definition of the Hausdorff dimension of A:

$$\begin{aligned} \dim_{\mathcal{H}}(A) &= \inf\{\alpha; \mathcal{H}^\alpha(A) = 0\} \\ &= \sup\{\alpha; \mathcal{H}^\alpha(A) = \infty\} \end{aligned}$$

The Hausdorff dimension generalizes the classical notions of dimension, for instance, when A is a subset of $\mathbb{R}^n$, $\dim_{\mathcal{H}}(A) \leq n$, a d-dimensional manifold has Hausdorff dimension d, and $\mathcal{H}^n(A)$ is the same as the Lebesgue measure $\mathcal{L}_n$ of A (for a very large class of subset A, which we do not describe here. For more details on geometric measure theory, see Falconer (1986) and Federer (1969)).

Another convenient notion of dimension is the (metric) entropy dimension. Let us briefly define it. For a bounded subset A in some metric space and a real number $\alpha > 0$, we denote $M(\alpha, A)$ the minimal number of closed balls of radius $\leq \alpha$, covering A. $H_\alpha(A) = \log_2(M(\alpha, A))$ is called the α-entropy of the set A. This terminology was introduced in Kolmogorov and Tihomirov (1961) and reflects the fact that $H_\alpha(A)$ is the amount of information needed to digitally memorize A with accuracy α. The entropy dimension of A, $\dim_e(A)$, is the order of $M(\alpha, A)$ as $\alpha \to 0$. Precisely,

$$\begin{aligned} \dim_e(A) &= \limsup_{\alpha \to 0} \frac{\log(M(\alpha, A))}{\log(1/\alpha)} \\ &= \inf\{\delta; M(\alpha, A) \leq (1/\alpha)^\delta, \\ &\qquad \text{for sufficiently small } \alpha\} \end{aligned}$$

We clearly have

$$\dim_{\mathcal{H}}(A) \leq \dim_e(A)$$

For any bounded set A in $\mathbb{R}^n$, we can bound $M(\alpha, A)$ from above by a polynomial in $1/\alpha$ (see Ivanov (1975) and Yomdin and Comte (2004)):

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

where $c(n)$ only depends on n and $V_i(A)$ (the ith variation of the set A) is the mean value, with respect to P (for a suitable measure), of the number of connected components of $A \cap P$, with P an affine $(n-i)$-dimensional space of $\mathbb{R}^n$.

Since for A contained in a d-dimensional manifold, $V_i(A) = 0$ for $i > d$, we deduce from this inequality that in this case $M(\alpha, A)$ is bounded from above by a polynomial of degree $\leq d$ in $1/\alpha$.

Our goal is to explain that we can be more precise than this general inequality when A is a set of critical or almost-critical values of a C^k mapping.

Transversality Is a Generic Situation

The results in this section concern critical values, and not almost-critical values. They show that a "generic" point of the target space is not a critical value, and the more regular, the mapping the smaller the set of critical values. Such theorems relating the regularity of a mapping and the size of its critical values are called Morse–Sard type theorems (see Sard (1942, 1958, 1965)). The simplest theorem in this direction is the following:

Theorem 1 (C^∞ Morse–Sard theorem) (Morse 1939, Sard 1942, Holm 1987). *Let $f: \mathbb{R}^n \to \mathbb{R}^p$ be a C^∞-regular mapping. Then $\mathcal{H}^p(\Delta(f, B_r)) = 0$, where $\Delta(f, B_r) = f(\Sigma(f, B_r))$ and $\Sigma(f, B_r)$ is the set of points $x \in B_r$ where* $\operatorname{rank}(Df_{(x)}) < p$.

The set $\Delta(f, B_r)$ is the image, under f, of the points of the ball B_r in the source space at which f is not submersive, that is, the set of critical values of f. Consequently, the Morse–Sard theorem ensures that for almost all points y in the target space, $f^{-1}(\{y\})$ is either empty or a smooth submanifold of the source space of dimension $n - p$.

Note that $\Delta(f, B_r) = \Delta(f, B_r, \Lambda)$ for some convenient $\Lambda = (\epsilon_1, \ldots, \epsilon_p)$ with $\epsilon_p = 0$, because $x \mapsto l_i(Df_{(x)})$ is bounded on B_r, for all $i \in \{1, \ldots, p\}$.

Now, we can concentrate our attention on more singular points than the critical ones, those at which the rank ρ of f is prescribed. Let us denote such points by $\Delta^\rho(f, B_r)$, for $\rho < p$. By definition, $\Delta^\rho(f, B_r) = f(\Sigma^\rho(f, B_r))$, where $\Sigma^\rho(f, B_r) = \{x \in B_r \subset R^n; \operatorname{rank}(Df_{(x)}) \leq \rho\}$. With these notations, the result for rank-r critical values is the following:

Theorem 2 (C^k Morse–Sard theorem for rank-r critical values) (Federer 1969). *Let $f: \mathbb{R}^n \to \mathbb{R}^p$ be a C^k-regular mapping. Then $\mathcal{H}^{\rho+(n-\rho)/k}(\Delta^\rho(f, B_r)) = 0$. In particular,*

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

One can produce examples showing that the bound of Theorem 2 is the sharpest one (see Comte (1996), Whitney (1935), Grinberg (1985), and Yomdin and Comte (2004)).

We note that Theorem 1 is a corollary of Theorem 2 (just replace k by ∞ and ρ by $p-1$ in Theorem 2). This result tells nothing about the entropy dimension of $\Delta^\rho(f, B_r)$; in the next section, we will bound the growth of entropy of almost-critical values.

Almost-Transversality Is Almost Generic

In this section, $f: \mathbb{R}^n \to \mathbb{R}^p$ is a C^k mapping. We denote by K a Lipschitz constant of $D^{k-1}f$ on B_r and by $R_k(f)$ the quantity $(K/(k-1)!) \cdot r^k$. We have:

Theorem 3 (C^k quantitative Morse–Sard theorem) (Yomdin 1983 Yomdin and Comte 2004). *Let $f: \mathbb{R}^n \to \mathbb{R}^p$ be a C^k mapping, $\Lambda = (\epsilon_1, \ldots, \epsilon_p)$, $\epsilon_1 \geq \cdots \geq \epsilon_p$, and let us denote $\epsilon_0 = 1$. We have (for $\alpha \leq R_k(f)$):*

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

where C is a constant depending only on n, p, and k.

As a corollary, one can bound the entropy dimension of $\Delta^\rho(f, B_r)$ by $\rho + (n-\rho)/k$, and hence its Hausdorff dimension, again finding Theorem 2: we just have to put $\epsilon_{\rho+1} = 0$ and $\epsilon_1, \ldots, \epsilon_\rho$ large enough, that is, $\epsilon_i \geq \lambda_i(Df_{(x)})$, for all $x \in B_r$, in Theorem 3, to obtain:

Theorem 4 (C^k entropy Morse–Sard theorem) (Yomdin 1983 Yomdin and Comte 2004). *Let $f: \mathbb{R}^n \to \mathbb{R}^p$ be a C^k mapping, let us denote $\epsilon_0 = 1$ and $\epsilon_i = \sup\{\lambda_i(Df_{(x)}); x \in B_r\}$, for $i \in \{1, \ldots, \rho\}$. We have (for $\alpha \leq R_k(f)$):*

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

where C is a constant depending only on n, p, and k. In particular,

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

Again we have examples showing that this bound is sharp (see Yomdin and Comte 2004).

Furthermore, the mapping f in Theorems 2–4 may be of real differentiability class (Hölder smoothness class C^k), with the same conclusions in these theorems. That is, k may be a real number written as $k = p + \beta$ with $\beta \in [0, 1]$, $p \in \mathbb{N} \setminus \{0\}$, and f is C^k means that f is p times differentiable and there exists a constant $C > 0$ such that for all $x, y \in B_r$, $\|D^p f_{(x)} - D^p f_{(y)}\| \leq C \cdot \|x - y\|^\beta$ (see Yomdin and Comte (2004)).

Examples

Let us denote by A the set of real polynomial mappings of degree d and of the following type:

$$x \mapsto Q(a, x) = 1 + \sum_{j=1}^{d} a_j x^j$$

with $a = (a_1, \ldots, a_d)$ and $\|a\| \leq 1$ (where $\|\|$ is the Euclidean norm of $\mathbb{R}^d$). We identify the set A with $B^d(0, 1) = \{a \in \mathbb{R}^d; \|a\| \leq 1\}$.

We want to bound the α-entropy of the set of such polynomials for which the real roots are multiple or almost multiple.

We denote by V the set $V = \{(a, x) \in \mathbb{R}^{d+1}; Q(a, x) = 0\}$. At points (a, x) of V with $\nabla Q_{(a,x)} \neq 0$, V is a C^∞ manifold of codimension 1 of $\mathbb{R}^{d+1}$. We denote by $V^{\text{reg}} = \{(a, x) \in V; \nabla Q_{(a,x)} \neq 0\}$ and by $V^{\text{sing}} = \{(a, x) \in V; \nabla Q_{(a,x)} = 0\} = V \setminus V^{\text{reg}}$. By Whitney (1957), V^{sing} is a union of smooth manifolds of dimension $\leq d-1$.

A root x of a polynomial $Q(a, \cdot)$ is multiple if and only if

$$Q(a, x) = \frac{\partial Q}{\partial x}(a, x) = 0$$

Consequently, the set A^Σ of polynomials of A with multiple roots is $\pi(V^{\text{sing}}) \cup \Delta(\pi_{|V^{\text{reg}}})$, where $\pi: \mathbb{R}^{d+1} \to \mathbb{R}^d$ is the standard projection $\pi(a, x) = a$, and $\Delta(\pi_{|V^{\text{reg}}})$ is the set $\{(a, x) \in V^{\text{reg}}; Ox \subset T_{(a,x)} V^{\text{reg}}\}$ of critical values of $\pi_{|V^{\text{reg}}}$. By Sard's theorem

(Theorem 2), $\dim_{\mathcal{H}}(\Delta(\pi_{|V^{\text{reg}}})) \leq d-1$. Since $\dim_{\mathcal{H}}(\pi(V^{\text{sing}})) \leq d-1$, we obtain: $\dim_{\mathcal{H}}(A^{\Sigma}) \leq d-1$: thus, having distinct roots is a generic property.

Let, as above, $\Lambda = (\epsilon_1, \ldots, \epsilon_d)$ with $\epsilon_1 \geq \cdots \geq \epsilon_d$ and $\epsilon_0 = 1$. A root x of a polynomial $Q(a, \cdot) \in A$ is said to be Λ-almost multiple if and only if $Q(a,x) = 0$ and $V \not\pitchfork^{\Lambda}_{x} Ox$, that is, $(a,x) \in V^{\text{sing}}$ or $\sin(T_{(a,x)}V^{\text{reg}}, Ox) \leq \epsilon_d$. This condition only concerns ϵ_d and we can take $\epsilon_1 = \cdots = \epsilon_{d-1} = 1$. We denote $A^{\Sigma,\Lambda}$ to be the set of polynomials of A with (at least) a Λ-almost multiple root. By Theorem 3,

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

But $\pi(V^{\text{sing}})$ being a finite union of manifolds of dimension at most $d-1$, we finally obtain

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

Thus, having no Λ-almost multiple root is Λ-almost a generic property. In Figure 3, we represent V for $d = 3$ and $a_3 = 1$,

$$W = \left\{ (a,x) \in \mathbb{R}^{d+1};\ \frac{\partial Q}{\partial x}(a,x) = 0 \right\}$$

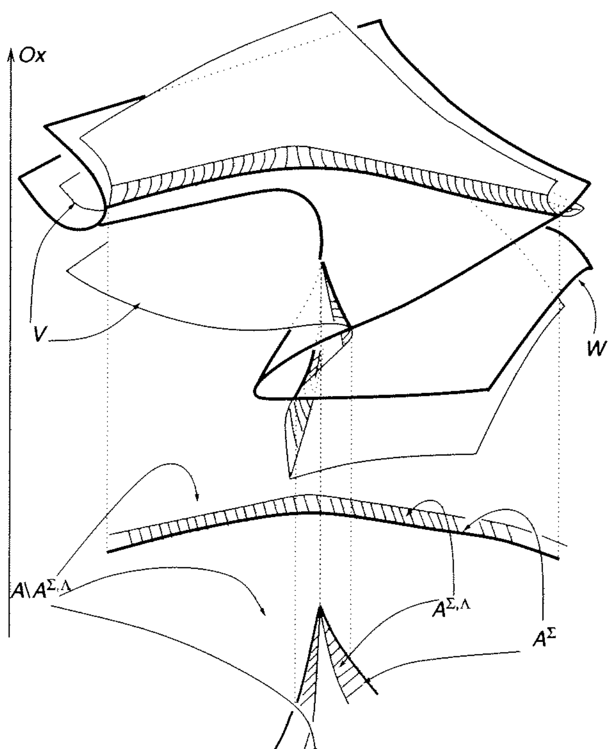

Figure 3 The space of polynomials of type $1 + a_1 x + a_2 x + x^3$ with almost-multiple roots.

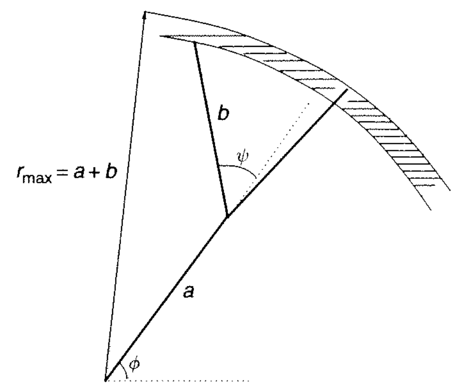

Figure 4 Almost-critical points of the distance function of P to the origin.

The next example comes from robotics: let us consider a planar robotic manipulator consisting of two jointed bars of length a and b, as presented in Figure 4. We may parametrize the positions of the endpoint P of this device by the angles ϕ and ψ (see Figure 4). Now the distance r from the origin to P is $r^2 = \|P\|^2 = a^2 + b^2 + 2ab\cos(\psi)$. The critical points of r are given by

$$\frac{\mathrm{d}r}{\mathrm{d}\psi}(\psi) = -2ab\sin(\psi) = 0$$

and correspond to the circle $\psi = 0$. The critical value of r is $a+b$. Near these critical positions, the control of r with respect to ψ is poor; we would like to avoid those near-critical values. Given $\epsilon > 0$, the condition

$$\left| \frac{\mathrm{d}r}{\mathrm{d}\psi}(\psi) \right| \leq \epsilon$$

implies $|\psi| \leq \arcsin(\epsilon/2ab)$, and the ϵ-near-critical values of r are

$$r_{\max}^2 - r^2 \leq 2ab[1 - \cos(\arcsin(\epsilon/2ab))]$$

where $r_{\max}$ is $a+b$; thus, they are contained in an interval of length $\leq c \cdot \epsilon^2/(4ab \cdot r_{\max})$, and $M(\alpha, \Delta(r,\epsilon)) \leq c \cdot \epsilon^2/(4ab \cdot r_{\max} \cdot \alpha)$ (Theorem 3 gives $M(\alpha, \Delta(r,\epsilon)) \leq C(1 + \epsilon/\alpha)$).

See also: Entanglement; Entanglement Measures; Quantum Entropy; Singularity and Bifurcation Theory.

Further Reading

Comte G (1996) Sur les hypothèses du théorème de Sard pour le lieu critique de rang 0. *C. R. Academy of Science Paris* 323(I): 143–146.

de Araujo Moreira CGT (2001) Hausdorff measures and the Morse–Sard theorem. *Publication Mathemàtiques* 45(1): 149–162.

Falconer KJ (1986) *The Geometry of Fractal Sets*, Cambridge Tracts in Mathematics, vol. 85. Cambridge: Cambridge University Press.
Federer H (1969) *Geometric Measure Theory*, Grundlehren Math. Wiss., vol. 153. Springer.
Golubitski V and Guillemin V (1973) *Stable Mappings and Their Singularities*, Graduate Texts in Mathematics, vol. 14. Springer.
Grinberg EL (1985) On the smoothness hypothesis in Sard's theorem. *American Mathematical Monthly* 92(10): 733–734.
Holm P (1987) The theorem of Brown and Sard. *L'enseignement Mathématiques* 33: 199–202.
Ivanov LD (1975) In: Vituskin AG (ed.) *Variazii množhestv i funktsii (Russian) (Variations of Sets and Functions)*. Moscow: Nauka.
Kolmogorov AN and Tihomirov VM (1961) ε-Entropy and ε-capacity of sets in functional space. *American Mathematical Society Translations* 17: 277–364.
Morse AP (1939) The behavior of a function on its critical set. *Annals of Mathematics* 40(1): 62–70.
Sard A (1942) The measure of the critical values of differentiable maps. *Bulletin of the American Mathematical Society* 48: 883–890.
Sard A (1958) Images of critical sets. *Annals of Mathematics* 68: 247–259.
Sard A (1965) Hausdorff measure of critical images on Banach manifolds. *American Journal of Mathematics* 87: 158–174.
Whitney H (1935) A function not constant on a connected set of critical points. *Duke Mathematical Journal* 1: 514–517.
Whitney H (1957) Elementary structure of real algebraic varieties. *Annals of Mathematics* 66: 545–556.
Yomdin Y (1983) The geometry of critical and near-critical values of differentiable mappings. *Mathematische Annalen* 264(4): 495–515.
Yomdin Y and Comte G (2004) *Tame Geometry with Application in Smooth Analysis*, Lecture Notes in Mathematics, vol. 1834. Berlin: Springer.

Equivariant Cohomology and the Cartan Model

E Meinrenken, University of Toronto, Toronto, ON, Canada

Introduction

If a compact Lie group G acts on a manifold M, the space M/G of orbits of the action is usually a singular space. Nonetheless, it is often possible to develop a "differential geometry" of the orbit space in terms of appropriately defined equivariant objects on M. This article is mostly concerned with "differential forms on M/G." A first idea would be to work with the complex of "basic" forms on M, but for many purposes this complex turns out to be too small. A much more useful complex of equivariant differential forms on M was introduced by Cartan (1950). In retrospect, Cartan's approach presented a differential form model for the equivariant cohomology of M, as defined by A Borel (1960). Borel's construction replaces the quotient M/G by a better-behaved (but usually infinite-dimensional) homotopy quotient M_G, and Cartan's complex should be viewed as a model for forms on M_G.

One of the features of equivariant cohomology are the localization formulas for the integrals of equivariant cocycles. The first instance of such an integration formula was the "exact stationary phase formula," discovered by Duistermaat and Heckman. This formula was quickly recognized by Berline and Vergne (1983) and Atiyah and Bott (1984), as a localization principle in equivariant cohomology. Today, equivariant localization is a basic tool in mathematical physics, with numerous applications.

This article begins with Borel's topological definition of equivariant cohomology, then proceeds to describe H Cartan's more algebraic approach, and concludes with a discussion of localization principles.

As additional references for the material covered here, we particularly recommend books by Berline, Getzler, and Vergne (1992) and Guillemin and Sternberg (1999).

Borel's Model of $H_G(M)$

Let G be a topological group. A G-space is a topological space M on which G acts by transformations $g \mapsto a_g$, in such a way that the action map

$$a : G \times M \to M \qquad [1]$$

is continuous. An important special case of G-spaces are principal G-bundles $E \to B$, that is, G-spaces locally isomorphic to products $U \times G$.

Definition 1 A classifying bundle for G is a principal G-bundle $EG \to BG$, with the following universal property: for any principal G-bundle $E \to B$, there is a map $f : B \to BG$, unique up to homotopy, such that E is isomorphic to the pullback bundle f^*EG. The map f is known as a "classifying map" of the principal bundle.

To be precise, the base spaces of the principal bundles considered here must satisfy some technical condition. For a careful discussion, see Husemoller (1994). Classifying bundles exist for all G (by a construction due to Milnor (1956)), and are unique up to G-homotopy equivalence.

It is a basic fact that principal G-bundles with contractible total space are classifying bundles.