MIKHAEL GROMOV

Permanent Professor at the Institut des Hautes Etudes Scientifiques, Bures-sur-Yvette (near Paris)

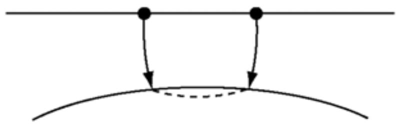
1999 BALZAN PRIZE for MATHEMATICS

Local and global in geometry

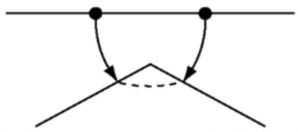
The main paradigm of the classical physics is non-existence of non-local interactions: one object cannot influence another one removed from it without involving intermediate agents located one next to another and making a continuous chain joining the two objects in space. Non-locality would be perceived as a magic, by René Thom's expression, and we do not accept magic in science (reluctantly making an exception for the quantum mechanics paying the price by modifying the ideas of causality and correlation). Thus the fundamental laws of classical physics (including relativity) are local, pertaining to infinitesimally small parts of the space where microscopic rules determine the global features of the macroscopic world. For example, the motion of planets around the sun is determined by gravitational forces acting on all material points in space. A flow of a viscous fluid is bound by the friction between small parts of the fluid across common boundaries. Movements of our own body start with electrochemical currents in the neurons in the brain, expressing our will to move, and the resulting spatial motion propagates from one atom to another via the electromagnetic interactions between the atoms constituting the body. Non-surprisingly, we expect observable physical patterns, e.g. the positions of moving particles after a given time interval, to be predictable in terms of the local laws and a presence of a particular microscopic law should be manifested by a specific global behaviour. Conversely, if a certain medium exhibits all possible kinds of shapes and motions, we conclude it is not restricted by any conceivable microscopic rule. If we see that various particles, for instance, propagate in the same direction, e.g. vertically down, we look for a force field oriented in the direction of the motion. On the other hand, if there is no predominant direction (or directions) of moving particles, then we may assume that there is no field acting on the particles in the observable region of space. (This is an oversimplification: a force field determines not a direction but rather the acceleration vector of a system.)

It happens sometimes that the local rule does not completely determine the global behaviour but just excludes certain possibilities. For example, many objects on the surface of Earth move in the horizontal plane, with no preferred direction, with the (local) constraint on the vertical displacement.

In geometry, the locality comes under the name of the "local-to-global principle". Start with the notion of the *length of a curve* in space. If we divide a curve into small pieces, we can recapture the global length by just adding the lengths of these pieces, where the pieces can be microscopically (or *infinitesimally*, as mathematicians say) small. Thus the familiar notion of length incorporates the local-to-global principle. Next, imagine the curve moves in space and we record the positions of points on the curve at different moments. If the curve is represented by a thin string made of elastic rubber the length does not have to be preserved: two very close points on the curve may come closer or further apart than they originally were. On the other hand, if we think of a flexible thread that can be bent but not stretched or compressed, then the distance in space between infinitesimally close points remains constant in so far as the curve remains *smooth*, i.e. does not develop creases.



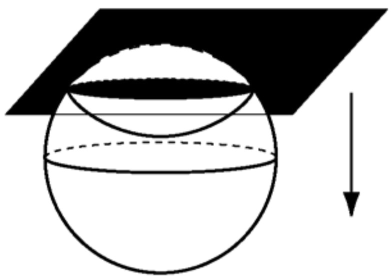
infinitesimal distance is unchanged



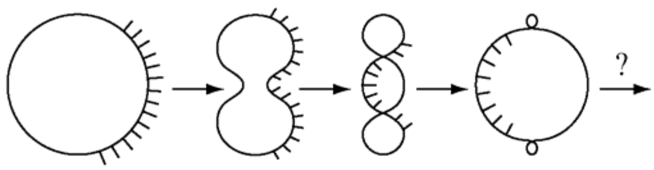
infinitesimal distance significantly decreases

All this is simple and you do not have to learn any mathematics to get a grasp of the picture. But if you turn to the similar problem of *bending surfaces in space* (important in the actual engineering) you face difficult mathematical questions (some of them remaining unsolved at the present time). We think of a surface as a thin film positioned in space that cannot be internally stretched or compressed, yet can be freely bent. A sheet of paper is an example of such a surface. The paper we usually deal with comes flat and then can be bent, if we wish, but there are limitations to the bending. We cannot, for example, perfectly wrap a ball into a flat paper sheet without making creases, though we can wrap a flat paper around a cylinder. On the other hand, it is possible to manufacture paper following the shape of a given surface, e.g. the round sphere can be made of paper. If you take such a paper sphere in your hand you will feel it resists your (not too vigorous) attempts to bend it, remaining perfectly spherical under the (not too strong) pressure, very much unlike a flat sheet or a small piece of the sphere paper which can be bent a little.

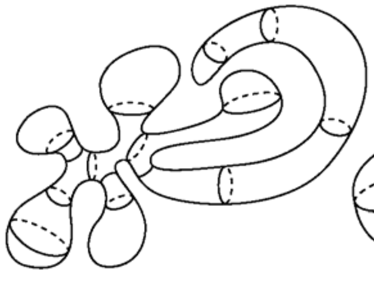
Let us make the notion of bending mathematically precise. A deformation of a surface in space is called a *bending*, with *no* internal compression and stretching, if *every* curve positioned on the surface does not change its length under the deformation. In other words, if we imagine a thin hair glued to the surface, the deformation of this hair following the surface will be as what we have seen for a thread and not as for a rubber string. There is a mathematical theorem, having a long history starting from 1813, claiming, in particular, that the full spherical surface is *rigid*, i.e. cannot be bent in the space, provided it remains *convex* in the course of deformation. This is quite different from a circular curve in the plane – this can be freely bent in many ways without distorting the length – but the only conceivable deformation of the sphere keeping the length of *all* curves on the surface intact comes from reflecting a spherical cup cut by a moving plane



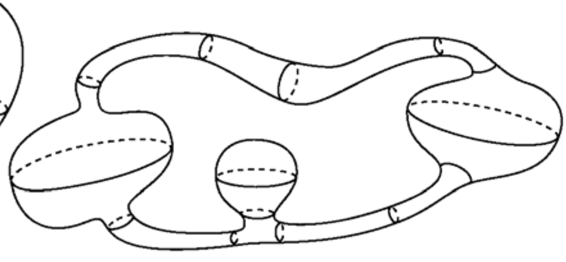
as is shown in the above picture. Notice that the so "bent" sphere has a circular crease at the edge of the reflected cup and, intuitively, such a crease seems unavoidable for any bending. This was a long-standing conjecture: there is *no smooth* bending of a sphere (with no *a priori* convexity assumption). Amazingly, this happened to be false, smooth non-convex bendings of spheres do exist! They came via an intricate high-dimensional geometric construction invented by John Nash in 1954 and implemented for surfaces in the 3-space by Nico Kuiper in 1955. The mathematical essence of Nash's construction depends on the idea of "smooth". There are various grades of smoothness in geometry reflecting what we intuitively perceive as being "truly smooth". Nash's result showed that different kinds of smoothness may lead to quite unsimilar mathematical models of surfaces in the physical space and indicated unsufficiency of the traditional "local-to-global" philosophy. But it remained unclear if it was an accidental phenomenon, and if not, what could be a general framework accounting for the "Nash paradox". Conceptual help came from another branch of mathematics, differential topology, where the story started from something even more perplexing than Nash's bending. Look again at the round spherical surface and observe that this surface (unlike the Möbius band) has *two* sides, the interior and the exterior ones. Let us paint our sphere from outside, deform the sphere in a complicated way and keep track of which side is which by looking at the color. We are not concerned now with preservation of the length of curves on the surface; yet we insist on smoothness, admitting no creases, as earlier. But we do admit appearance of selfintersection in the course of deformations as sketched below with curves representing surfaces.



The above deformation *almost* turnes the circle inside out making most of the outside of the circle to look inside. But there is no way to complete this process and remove "almost" without creating two point creases in the last moment. One can *not* turn the circle in the plane inside out. However, as was discovered by Steve Smale in 1958, the sphere *can* be turned inside out in the 3-space. This was achieved not by exhibiting a particular deformation, this is prohibitively complicated, but by essentially using the principles of differential and algebraic *topology* – a branch of mathematics concerned with most general qualitative properties of geometric objects invariant under all kinds of *continuous* transformations.



A topologist's sphere



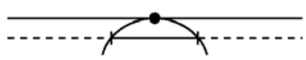
deformed bagel surface – non-sphere

I learnt Smale's theory, and its generalization applicable to all manifolds of arbitrary dimension due to Hirsch, from topologists, Sergey Novikov, and my advisor, professor Rochlin, in the mid-sixties. Around this time my friend and geometry teacher Yura Burago explained to me Nash's construction. I was enchanted by the magical power of topological reasoning of Smale-Hirsch deriving intuitively inconceivable consequences out of basic laws of continuity. Eventually, I found a way to extend these laws beyond topology and apply them to other geometric structures. (I was not the only one trying to push further the Smale-Hirsch theory. The first step was made by Tony Phillips and the most general topological results were obtained by Eliashberg in 1972.) Later on, I realized that Nash's geometric constructions (as well as his incredible analytic techniques developed in 1956 for the study of bending-like problems of surfaces and more general Riemannian spaces) can be incorporated into the topological network of Smale-Hirsch serving there as a powerdrive for the topological machinery. Thus enriched topology took over several domains of geometry and analysis replacing the classical "local-to-global" paradigm by the *homotopy principle*: the infinitesimal structure of a medium, abiding by this principle does not effect the global geometry but only the topological behaviour of the medium. The class of infinitesimal laws subjugated by the homotopy principle is wide, but it does not include most partial differential equations (expressing infinitesimal laws) of physics with a few exceptions in favour of this principle leading to unexpected solutions. In fact, the presence of the *h*-principle would invalidate the very idea of a physical law as it yields very limited global information effected by the infinitesimal data. But this may be desirable in certain situations where we need to navigate in a high-dimensional space with a limited number of control mechanisms. For example, a position of a car on the street is determined by *three* parameters: the location of the center of the car given by two Cartesian coordinates, and the orientation of the car expressed by an additional angular coordinate. On the other hand there are only *two* control parameters: the position of the accelerator pedal and the turn of the wheel. Yet we can bring the car to an arbitrary position, albeit non-straightforwardly as for the parallel parking in a limited space, achieving which we prove (without being aware of it) the homotopy principle for the (non-holonomic) constraint limiting the freedom of movement from three to two dimensions. (The proofs of both Nash's and Smale's theorem use rather contorted geometric moves essentially similar to those we employ to fit a car into a small parking slot.)

Summing up, we see that geometric infinitesimal laws come in (at least) two flavours: those exhibiting rigid classical "local-to-global" features and quite different ones following the relaxed homotopy principle.

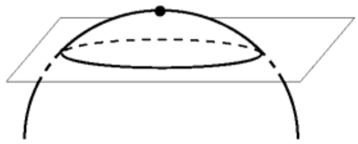
There are several exceptional situations where the two domains meet with most beautiful mathematic sparkling at the contact points. One can arrive at one of such meeting-points by returning to the non-bending theorem for spheres and more general curved convex surfaces in space. How do we measure the curvature? If it is a curved line in the plane, we intersect it with a straight line parallel to the tangent at the point where we wish to measure the curvature. The inverse size of the intersection (properly normalized) gives you the curvature.

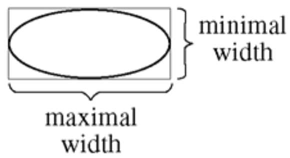
small curvature



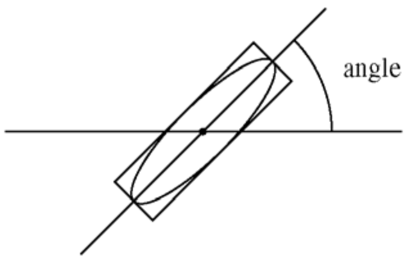
large curvature

Turning to convex surfaces and intersecting them with slightly shifted tangent planes we obtain tiny ellipses.

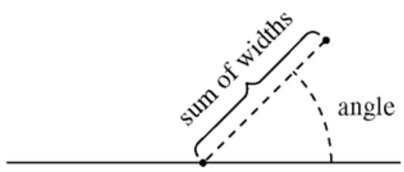




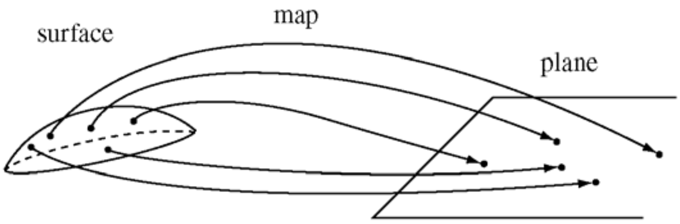
Such an ellipse is characterized by *two* numbers corresponding to its maximal and minimal widths expressing the curvatures at our point. The fundamental (and not at all obvious) Theorema Egregium of Gauss tells us that the *product* of these two numbers remains constant under (smooth convex) bendings: this is the starting-point of the proof of the rigidity which amounts to showing that not only the product of the widths but the curvature ellipses themselves do not change under bendings, provided we deal with the *full* spherical surface or a *full* (mathematicians say "closed") convex surface in general. For example, if the surface is the round sphere, then all ellipses are circular and one must show they remain such under bending. On a general convex surface a (curvature) ellipse is characterized by three parameters: two widths (called "principal axes") and the orientation, i.e. the angle between one of the axes with a given tangent line at the point where we measure the curvature.



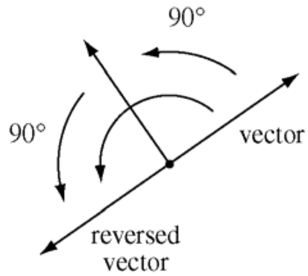
Theorema Egregium removes one degree of freedom leaving us with two numbers at every point of the surface: the above angle and one of the width where it is better to use the sum of the widths as this is more symmetric. We think of these two numbers geometrically, as the, so-called, *polar* coordinates in the plane, and thus we associate to each point on the surface the point in the plane with the polar coordinates corresponding to the curvature.



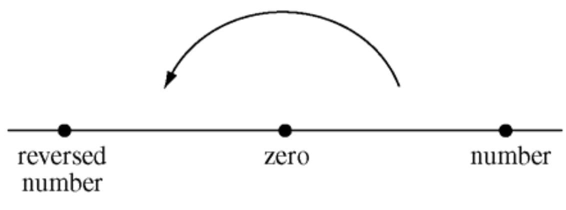
This gives us a geometric transformation, a *map* from the surface to the plane and we need to understand how much the map can change (if at all) when we bend our surface. What happens – and this needs an analytic calculation – is that this map is very much similar to *conformal maps* in the plane, pertaining to *complex*, sometimes called *imaginary* numbers.



Imaginary numbers appear in algebra when we try to take square roots of negative numbers. These were introduced by Gerolamo Cardano (1501-1576) in the study of solutions of algebraic equations, with no apparent relation to geometry. Geometric interpretation consists in observing that two consecutive rotations of the plane by 90° around a fixed point reverse the directions of the

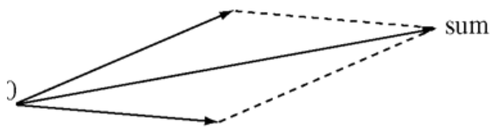


vectors. If we think of the 180°-degree rotation reversing vectors as the geometric counterpart to the multiplication of numbers by –1 reversing the sign, then we are inclined to accept the 90°-rotation (of the plane containing the line of real numbers) as the square root of –1.

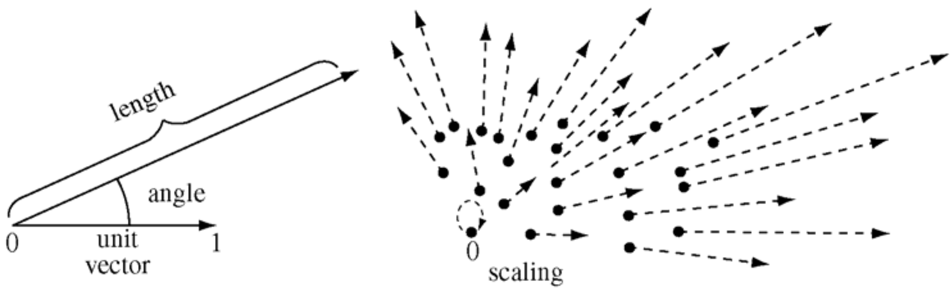


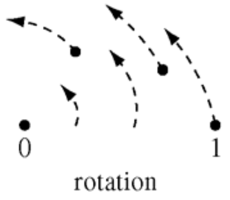
All this look childishly simple, why do mathematicians make such fuss around it? How can one dare to compare this plain idea to profound philosophical pronouncements, such as "Cogito ergo sum" of Descartes? But look (as my colleague David Ruelle once suggested) from another perspective. "Cogito ergo sum" stayed unperturbed for more than three centuries, like a monument, a Greek statue, a magnificent piece of art, impervious to the flow of time, whilst the little speck of dust, the square root of –1, have been growing and developing for hundreds of years in the minds of mathematicians, geniuses like Cauchy, Gauss, and Riemann, and turned into an evergreen intensely alive vibrant tree supporting in its branches our sacred knowledge – quantum mechanics – ruling everything we see (and do not see) in this world. Whenever we trace even a shade of this tree we expect a presence of a mathematical structure.

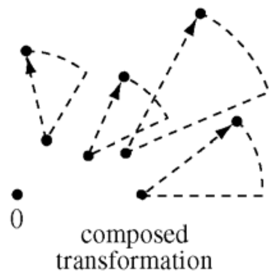
So encouraged, we turn to the *complex plane*: this is the ordinary plane with two marked points taken for *zero* and *one*. Granted these, we can interpret points in the plane as numbers – complex numbers – and perform the basic arithmetic operations: addition and multiplication. The addition does not need the marking of one, it is the familiar sum of vectors making sense in a space of any dimension.

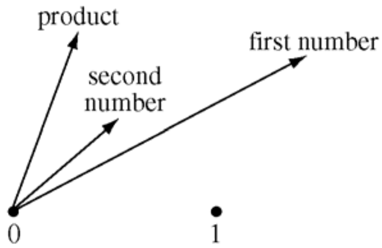


But the multiplication needs 1 and it is defined in several steps. First, given any complex number represented by a vector stemming from zero, we measure its *length* as well as the *angle* it makes with the unit vector. Then we consider two transformations of the plane: one is the scaling of all vectors by the above length and the second making rotation by our angle. If we apply one transformation followed by the other we obtain a new transformation of the plane. The effect of the latter on a point, i.e. a number, is what we call the product of this number with the original one employed for making up the whole transformation. (Look at the picture and try to see why the product is *commutative*, i.e. does *not* depend on the order of the numbers.)

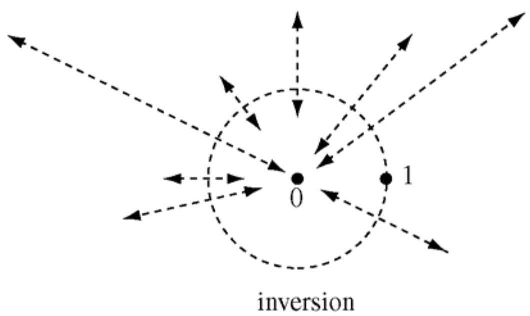




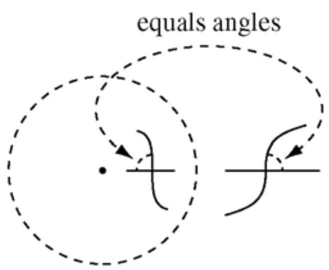




Given multiplication we can speak of division, in particular, we may go from each number to its inverse. The resulting transformation, called *inversion*, of the plane at zero (or rather of the plane minus zero as there is no room for the inverse to zero in the plane) preserves the unit circle and exchange points inside and outside.



What one cannot see in the picture – and this is *the* essential property of the inversion – is *conformality*, i.e. preservation of *angles* between curves transformed by the inversion.



Then, one can produce further maps by *adding together* inversions at different points, where inversions are regarded as complex functions of a complex variable. Since conformality can be expressed by a *linear* differential equation (this needs a proof, but it is easy) finite (as well as infinite) sums of inversions are conformal. This is the starting-point of the classical complex analysis created in the last century by Gauss, Cauchy, and Riemann.

In order to comprehend the geometric nature of conformal maps and complex functions we mimic what we do with more familiar real functions: we "look" at the *graphs* of these functions. In the real case these are curves in the plane but the complex numbers themselves are two-dimensional and 2 + 2 = 4. So we need a 4-dimensional space to contain such graphs appearing as surfaces rather than curves. The equations expressing conformality, called *Cauchy-Riemann* equations, can be seen in the geometry of (the tangent planes to) these surfaces and so reexpressed Cauchy-Riemann equations can be written down and geometrically studied on every 4-space.

Now we can return to our point of depart, convex surfaces. The relevant 4-space is built out of the original surface with extra two parameters coming from the curvature. Altogether we have a 4-dimensional counterpart of a cylinder. If we bend our surface, we distort curvatures and the "graph" of this distortion given by a complex function (encoding two real functions) can be seen as a surface in our cylinder. Furthermore, the "preservation of length" property, when translated to the language of curvature, becomes an equation imposed on the surface having the same (Cauchy-Riemann) structure as we had for graphs of complex functions. (Strictly speaking, there is no true complex function in our case but the graph makes sense all the same.) When all this is fully formalized and clarified, we arrive at a general theory encompassing a variety of geometric situations where bent surfaces appear on an equal footing with conformal maps and where rigidity comes from an extension of the theorem about d roots of algebraic equations of degree d.

Generalized analytic or *pseudoholomorphic* functions have been extensively studied by analysts since 1941. This (rather formalistic) theory was used by the Russian geometer Pogorelov for a proof of rigidity of surfaces. Reading his book and defeated by formulas I suddenly realized that the 4-dimensional geometric interpretation of Pogorelov's equations solves the problem without a single line of calculation. Then I looked around and found out that high-dimensional Cauchy-Riemann equations neatly fit into the geometric framework of *symplectic geometry*.

The symplectic geometry, which studies oriented areas of surfaces in spaces (rather than lengths of curves) originated in classical mechanics. Recall that an evolution of a mechanical system, e.g. a configuration of several particles in space, is determined by the initial conditions comprised of the positions and velocities of the particles. These change with time under some force field, e.g. electromagnetic or gravitational field. For example, if we have three celestial bodies, say the Sun, the Earth, and the Moon, then the system is described by 18 coordinates: three points with three coordinates each and three velocity vectors, where the evolution can be encoded by a transformation of the 18-space driven by Newton's inverse square law of gravitation. The two groups of coordinates, 9 corresponding to the positions of the bodies and 9 accounting for their velocities, seem quite different. But there is an invisible symmetry in this 18-space making the coordinates equivalent and this symmetry is preserved under the transformations induced by all possible force fields defined via potentials, referred to as *Hamiltonians* in this context. Geometrically, there is something like area (but not the usual area) assignment to the 2-surfaces in this space, called the *symplectic* structure and this is conserved by all Hamiltonian (sometimes called canonical) transformations. Thus the Hamiltonian systems satisfy a certain hidden infinitesimal geometric constraint and one is faced with the dilemma:

is this constraint of the classical "local-to-global" kind or is it subjugated by the homotopy principle? Prior to the homotopy principle, everybody (tacitly) assumed the existence of non-trivial "local-to-global" symplectic geometry with a specific list of conjectures proposed by Arnold in the early sixties. On the other hand, when a chunk of the symplectic geometry was taken over by this principle, one could equally expect the finite outcome to be in the favour of homotopy rather than geometry. Eventually, geometry won, the symplectic homotopy principle was dislodged by Eliashberg and later on several Arnold's conjectures were established with a new variational method of Rabinowitz. But what has this to do with the complex numbers? One knows from algebra (of Lie groups) that "symplectic" and "complex linear" are in a close kinship; this allows one to "write down" the Cauchy-Riemann equations on each symplectic space and then the solution to these equations reveals symplectic features invisible without the "pseudoholomorphic glasses". (I cannot be more explicit as the symplectic geometry cannot be truly seen below dimension 4.) And as time goes one discovers more and more structure growing at the cross-road of "soft" symplectic and "rigid" holomorphic.

Finally, I want to say a couple of words on more general aspects of the "local-to-global" idea. Any time we encounter a complicated logical, mathematical, physical or biological structure, we regard strongly interacting components as being close to each other with the distance growing as the interaction decreases going through longer and longer chains of intermediaries. A combinatorist will describe this in the language of *graphs*, and a geometer will speak about the distance or *metric*. But still there is no general perspective in the study of such structures. We have a long road ahead of us in developing a flexible language able to encompass and analyse the bewildering variety of structural patterns appearing in science and mathematics.

References

A. Carbone, S. Semmes, *Graphic apology for the symmetry and implicitness* (to appear).

M. Gromov, *Partial Differential Relations*, Springer-Verlag (1986), Ergeb. der Math. 3. Folge, Bd. 9.

M. Gromov, J. LaFontaine, P. Pansu, *Metric Structures for Riemannian and Non-Riemannian Spaces*, based on *Structures Métriques des Variétés Riemanniennes*, edited by J. LaFontaine and P. Pansu. English translation by Sean Michael Bates, Birkhäuser (1999), *Progress in Mathematics* **152** (1999).

H. Hofer, E. Zehnder, *Symplectic invariant and Hamiltonian dynamics*, Birkhäuser, Basel-Boston-Berlin, 1994.

S. Kuksin, *Elements of a qualitative theory of Hamiltonian P.D.E.*, Proc. of ICM-1998, Vol. II, p. 819–830, Berlin, 1998.

D. McDuff, D. Salamon, *J-holomorphic curves and quantum cohomology*, Univ. Lect. Series n° 6, A.M.S. Providence, 1994.

S. Müller, V. Šverák, *Unexpected solutions of first and second order partial differential equations*, Proc. of ICM-1998, Vol. II, p. 691–702, Berlin, 1998.