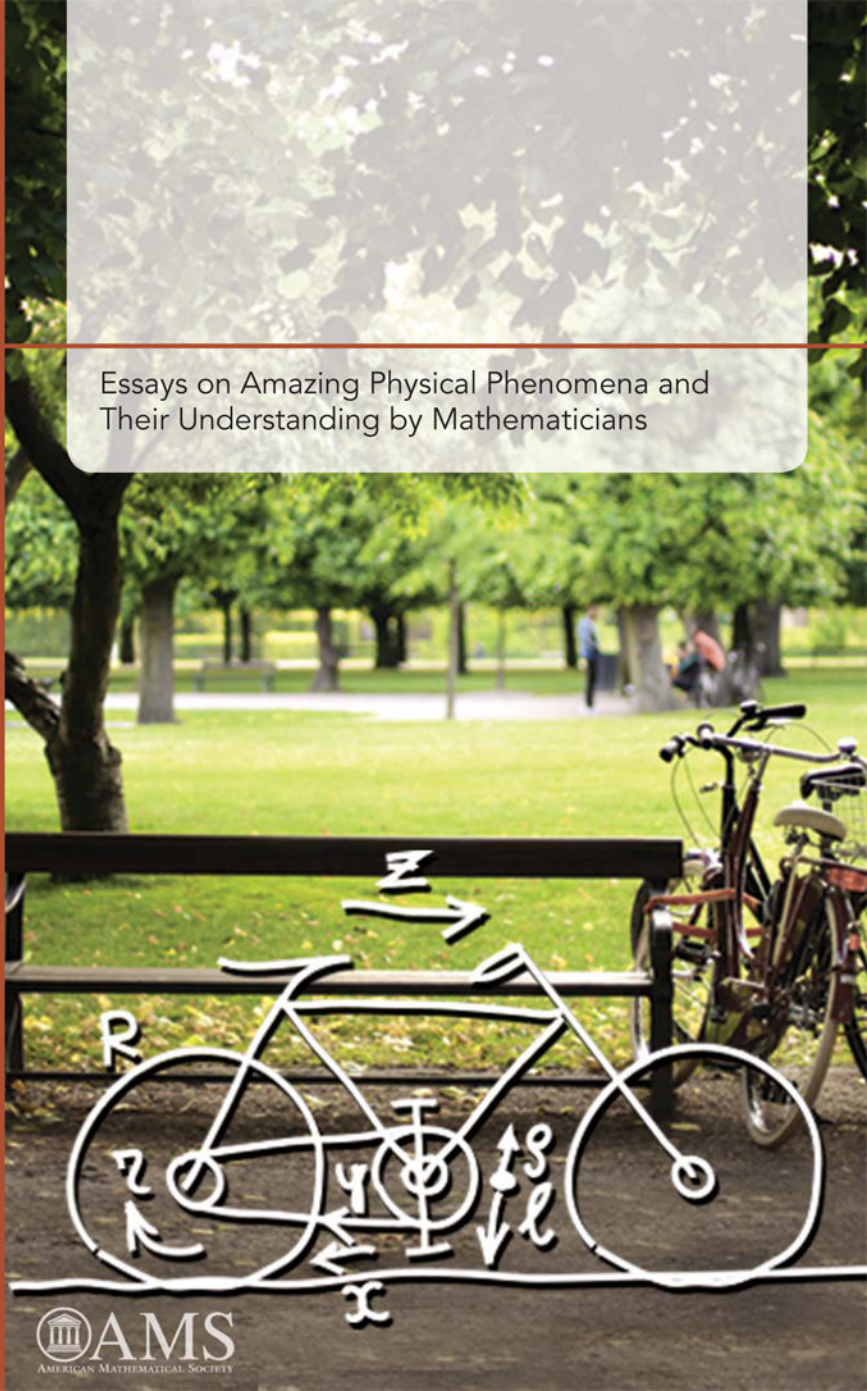
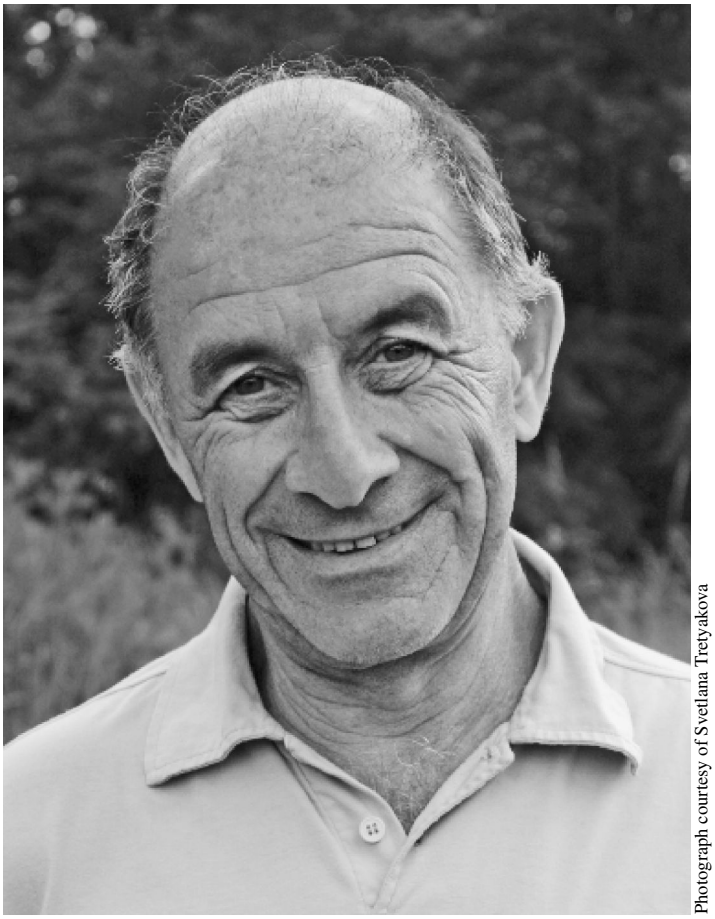
Essays on Amazing Physical Phenomena and Their Understanding by Mathematicians



AMS
American Mathematical Society

Mathematical Understanding of Nature

Essays on Amazing Physical Phenomena and Their Understanding by Mathematicians



Photograph courtesy of Svetlana Tretyakova

Vladimir Igorevich Arnold
June 12, 1937–June 3, 2010

http://dx.doi.org/10.1090/mbk/085

Mathematical Understanding of Nature

Essays on Amazing Physical Phenomena and Their Understanding by Mathematicians

V. I. Arnold

Translated by
Alexei Sossinsky
Olga Sipacheva

AMS
American Mathematical Society
Providence, Rhode Island

This work was originally published in Russian by МЦНМО under the title "Математическое Понимание Природы" © 2011. The present translation was created under license for the American Mathematical Society and is published by permission.

2010 *Mathematics Subject Classification.* Primary 70-01, 76-01, 78-01.

For additional information and updates on this book, visit
www.ams.org/bookpages/mbk-85

Library of Congress Cataloging-in-Publication Data

Arnol′d, V. I. (Vladimir Igorevich), 1937-2010. [Mathematicheskoe ponimanie prirod′y. English]
Mathematical understanding of nature : essays on amazing physical phenomena and their understanding by mathematicians / V.I. Arnold ; translated by Alexei Sossinsky and Olga Sipacheva.
pages cm.

Originally published in Russian by MT̂SNMO, under the title: Mathematicheskoe ponimanie prirod′y, 2011.

Includes bibliographical references.

ISBN 978-1-4704-1701-7 (alk. paper)

1. Mathematics–Popular works. I. Title.

QA39.A7413 2014
510–dc23 2014018911

10 9 8 7 6 5 4 3 2 1 19 18 17 16 15 14

Contents

Foreword

At the early age of eleven, the author of this book participated in the "Children Learned Society", organized at home by prominent Russian mathematician and computer scientist, A. A. Lyapunov (the Russian acronym, ДНО, which means "bottom", can be also interpreted as the "Voluntary Learned Society"). In a "Kvant" interview (1990),[1] Arnold remembers:

> The curriculum included mathematics and physics, along with chemistry and biology, including genetics, that was just recently banned (a son of one of our best geneticists was my classmate; in a questionnaire, he wrote: "my mother is a stay-at-home mom, my father is a stay-at-home dad").

Natalia Lyapunova, a daughter of A. A. Lyapunov, recalls:[2]

> ... And look what were the topics of the talks: "The structure of the solar system", "On comets", "Molecular forces"... One cannot forget the talk "Waves" by Dima Arnold. We had a huge dinner table, extendable to 6 sections. The table was unfolded, an aquarium with water was put into

[1] http://kvant.mccme.ru/1990/07/intervyu_s_viarnoldom.htm, in Russian.
[2] http://oso.rcsz.ru/inf/pp/177, in Russian.

> the hole, and a slide projector was placed underneath. At the time no one had such a projector, but my dad found one somewhere. The light went through the water whose surface projected on the ceiling. Two corks were floating in the aquarium; one needed to give them a push, and the waves started: circular, counter, interference! And all this is projected on the ceiling! Dima is lecturing, and visual demonstrations follow.... I was then in the 4th grade....

The present book is written in the spirit of the "Children Learned Society", and its target audience consists of "young mathematicians of all ages".[3]

The level of sophistication of these essays differs substantially, from being accessible to a high school student to presenting serious challenges for a seasoned researcher. In my opinion, this is a merit of the book: it belongs, equally well, to a high school library and to a faculty lounge.

The philosophy of the author is clearly visible:

> Mathematics is part of physics. Physics is an experimental science, a part of natural science. Mathematics is the part of physics where experiments are cheap.[4]

A popularizer of mathematics finds himself between a rock and a hard place. According to Michael Faraday (one of the greatest popularizers of science),

> Lectures which really teach will never be popular;
> lectures which are popular will never teach.

The present book is a (rare) counter-example to Faraday's maxim: it is eye-opening, open-ended, and never boring.

In the preface, Arnold says:

[3] In his memories of Vladimir Rokhlin, Arnold quotes from Courant: "... a mathematician should be considered young for as long as he is inclined to discuss math at the most inappropriate times".

[4] V. Arnold. "On teaching mathematics".

> Examples teach no less than rules, and errors more than correct but abstruse proofs.

Indeed, there is an error in the essay "What Force Drives a Bicycle Forward?", and the reader is encouraged to ponder what is going on.[5]

There is another special feature of this book that I have to comment upon, its provocative in-your-face style. Arnold was on a crusade against a formalized approach to mathematics or, in his parlance, "criminal Bourbakization". In this fight, he would take no prisoners– consider, for example, his famous 'mathematical duel' with J-P. Serre on Bourbaki at the Institut Henri Poincaré on March 13, 2001.[6]

Equally passionately, Arnold was fighting against the incorrect attribution of mathematical results. I cannot help but quote from Michael Berry's website:[7]

> Three laws of discovery
>
> 1. Arnold's law (implied by statements in his many letters disputing priority, usually in response to what he sees as neglect of Russian mathematicians):
> *Discoveries are rarely attributed to the correct person.*
> (Of course Arnold's law is self-referential.)
>
> 2. Berry's law (prompted by the observation that the sequence of antecedents under law 1 seems endless):
> *Nothing is ever discovered for the first time.*
>
> 3. Whitehead's law (quoted by Max Dresden at the beginning of his biography of Kramers):
> *Everything of importance has been said before by someone who did not discover it.*

I suspect that Arnold used hyperbole and overstated his opinions on purpose; the reader should be ready to take his most extreme claims with a grain of salt.

---

[5] See G. Hart's recent take on this problem at `http://www.simonsfoundation.org/multimedia/mathematical-impressions-multimedia/the-bicycle-pulling-puzzle/`

[6] `http://www.etnoka.fr/qualified/one.tcl?info_id=69919`

[7] `http://michaelberryphysics.wordpress.com/quotations/`

Most of the essays in this little book are quite short; therefore, it is not fitting for this foreword to get any longer. Let me finish with another quotation from Arnold's "Kvant" interview that, in my opinion, well represents both the spirit of this book and of its author:

> The word "Mathematics" means science about truth. It seems to me that modern science (i.e., theoretical physics along with mathematics) is a new religion, a cult of truth, founded by Newton 300 years ago.

Serge Tabachnikov

May 2014

Preface

The investigation of a murder led a movie director (a character of a detective story by Victoriya Tokareva[8]) to the conclusion: "Mathematics is that which can be explained."

The main contribution of mathematics to the natural sciences is not in formal computations (or in other applications of ready-made mathematical achievements), but in the investigation of those non-formal questions where the exact setting of the question (what are we searching for and what specific models must be used) usually constitutes half the matter.

The 39 essays collected below have the same goal: to teach the reader not only to multiply large numbers (which sometimes also has to be done), but to guess about unexpected connections between seemingly unrelated phenomena and facts, at times coming from different branches of the natural and other sciences.

Examples teach no less than rules, and errors, more than correct but abstruse proofs. Looking at the pictures in this book, the reader will understand more than learning by rote dozens of axioms (even together with their consequences about what sea the Volga river falls into and what horses eat).

[8] A Soviet and Russian screenwriter and short story writer.

Boris Pasternak wrote that "the question of the usefulness of poetry arises only in periods of its decline, while in periods of its flowering, no one doubts its total uselessness."

Mathematics is not quite poetry, but in it I try to avoid the feeling of decline preached by the enemies of all natural sciences.

Let me also add that Niels Bohr divided true statements into two classes: the trivial ones and those of genius. Specifically, he regarded a true statement as trivial when the opposite statement is obviously false, and a true statement as genius when the opposite statement is just as non-obvious as the original, so that the question of the truth of the opposite statement is interesting and worth studying.

I take this occasion to thank N. N. Andreev who coerced me into writing this book.

From the editors. Vladimir Arnold died on June 3, 2010. He participated in the preparation of the second edition, but did not see the proofs (in which the only changes were in the essays on pages 37–38 and 51–53).

http://dx.doi.org/10.1090/mbk/085/01

Chapter 1

The Eccentricity of the Keplerian Orbit of Mars

The following problems have the same mathematical model:

A right triangle has the hypotenuse of length 1 *m and a leg of length* 10 *cm. Find the length of the other leg.*

The mathematical "solution"

$$\sqrt{1-(1/10)^2}\ \text{m}$$

given by the Pythagorean theorem is unsatisfactory. The point is that, since

$$(1-a)^2 = 1-2a+a^2 \approx 1-2a$$

(with very small error a^2, provided that a is small), it follows that

$$\sqrt{1-A} \approx 1-A/2.$$

For $A = 1/100$, we obtain $1 - 1/200$ m, that is, 99.5 cm: The length of the long leg cannot be distinguished by eye from that of the hypotenuse, the half-percent difference is indiscernible, although the small angle of the triangle is not that small (about 6°).

The eccentricity of the Keplerian ellipse of Mars is about 0.1. *When Kepler sketched*[1] *the orbit of Mars, he took it for a circle with the Sun off-center. Why did he make such a mistake?*

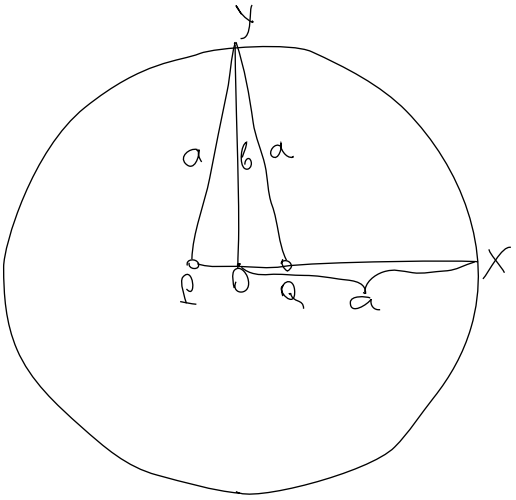
[1] On the basis of visual observations that Kepler's teacher Tycho Brahe made during many decades at the Uranus observatory on an island between Elsinore and Copenhagen, which Brahe owned. Later, Newton sent Halley with a telescope to this observatory in order to prove that telescopic observations may be as precise as those of Tycho Brahe.

Solution. An ellipse is the locus of all points in the plane for which the sum of distances from two fixed points P and Q (called foci) is constant. Let us denote this sum of distances by $2a$. Then, for an ellipse centered at O (the midpoint between the foci) with semi axes OX and OY, we have

$|OX| = a$ (because $|PX| + |QX| = 2a$),

$|QY| = a$ (because $|PY| = |QY|$ and $|PY| + |YQ| = 2a$), and

$|OQ| = ea$ (this is the definition of the eccentricity e).



From the right triangle OYQ we obtain

$$|OY| = \sqrt{|QY|^2 - |OQ|^2} = \sqrt{a^2 - a^2e^2} = a\sqrt{1 - e^2} \approx a(1 - e^2/2).$$

For the eccentricity $e = 0.1$, the distance from each focus to the center is 10% of the semi-major axis, $|OX| = a$, and the minor axis is shorter than the major one by only 0.5% (Kepler did not notice such a small difference at first.)

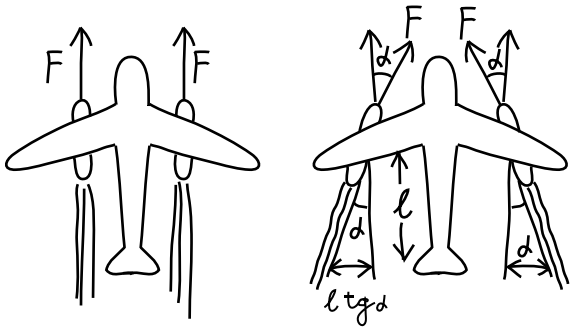
http://dx.doi.org/10.1090/mbk/085/02

Chapter 2

Rescuing the Empennage

The jet stream from the engine of the first jet planes burned their empennages. Engineers suggested slightly turning the engines (by a small angle α). The jet ceased burning the empennage (it moved aside by $l \tan \alpha$, where l is the distance to the empennage).

What fraction of the traction force $2F$ had to be sacrificed for this purpose?



Solution. The resulting traction force is

$$2F \cos \alpha \approx 2F(1 - \alpha^2/2).$$

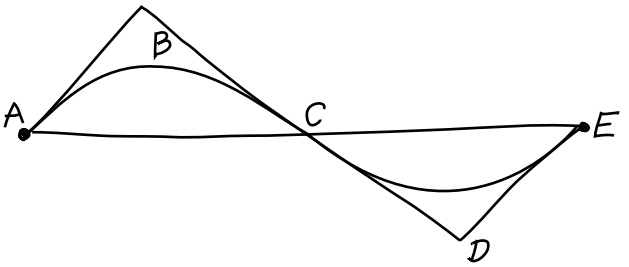
For the quite noticeable deviation of 3°, we have $\alpha \approx 1/20$ radian. Thus, the loss $\alpha^2/2$ is $1/800$ of the traction force, which is negligibly small (the deviation of the jet stream $l \tan \alpha \approx l/20$ amounts to several meters).

http://dx.doi.org/10.1090/mbk/085/03

Chapter 3

Return Along a Sinusoid

Returning home along a sinusoid, a drunkard lengthens his path. By how much?



Solution. By approximately 20%. Most people believe that a sinusoid is twice or at least one and a half times as long as a straight line. But actually, even the sawtooth path $ABCDE$ is longer than the straight one (AE) only by a factor of $\sqrt{2}$; i.e., by approximately 40%.

The sinusoidal path is much shorter. The point is that the part of the sinusoid inclined to AE at an angle α is longer than its projection on the line AE by a factor of about $\sqrt{1+\alpha^2} \approx 1 + \alpha^2/2$. Therefore, even those parts of the sinusoid that are inclined by 20° are longer than their projections by only $(1/3)^2/2 \approx 1/20$ times their length (5%). Only the fragments of the path close to the inflection points (A, C, and E) are lengthened significantly. Because these fragments are short the total lengthening of the path is small. Thus, the lengthening of the major part of the sinusoid is barely noticeable.

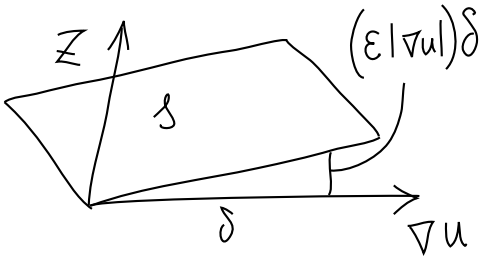
http://dx.doi.org/10.1090/mbk/085/04

Chapter 4

The Dirichlet Integral and the Laplace Operator

The membrane $z = 0$ was slightly bent (in three-dimensional space with Cartesian coordinates (x, y, z)) so that it became the graph of a small function $z = \varepsilon u(x, y)$ (where ε is small).

By how much is the area of the bent membrane greater than that of the initial flat membrane?



Solution. In the first (nonvanishing) approximation, near each point, the membrane stretches along the gradient $\operatorname{grad} u$ of the function u (in the same ratio as that of the hypotenuse in a right triangle in which the tangent of the smallest angle is $\varepsilon|\operatorname{grad} u|$ to the longest leg). Therefore, accurate to ε^2, the increment of the area element s is proportional to the squared deviation:

$$\delta s = \frac{1}{2}\varepsilon^2|\nabla u|^2 = \frac{\varepsilon^2}{2}\Big(\Big(\frac{\partial u}{\partial x}\Big)^2 + \Big(\frac{\partial u}{\partial y}\Big)^2\Big).$$

In other words, the increment of the area of the whole membrane is the (Dirichlet) integral

$$\delta S = \frac{\varepsilon^2}{2} \iint \left(\left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial u}{\partial y} \right)^2 \right) dx\, dy + o(\varepsilon^2).$$

Remark. It can be shown that the Dirichlet integral expresses not only the area increment of a membrane but also its potential energy; i.e., the work required of the force bending the membrane to change the state $z = 0$ to the state $z = \varepsilon u(x, y)$.

A proof of this (nonobvious) fact can be found in, e.g., the book *Lectures on Partial Differential Equations* (Fazis, 1997, pp. 68–70[1])

At the same time, it is proved there that the force bending (and stretching) the membrane is proportional to the Laplacian Δu of the function u (where $\Delta = \operatorname{div} \operatorname{grad}$) and; moreover, that

$$\iint_M (\nabla u)^2\, dx\, dy = -\iint_M u \Delta u\, dx\, dy$$

if $u = 0$ on the boundary of M.

The operator Δ, which takes u to Δu, is expressed (in the Cartesian coordinates x_i of Euclidean space $\mathbb{R}^n$) as

$$(*) \qquad \Delta u = \frac{\partial^2 u}{\partial x_1^2} + \ldots + \frac{\partial^2 u}{\partial x_n^2}.$$

In other coordinate systems in Euclidean space, the expression is different. For example, in polar coordinates (r, φ) in the plane $(x_1 = r\cos\varphi,\ x_2 = r\sin\varphi)$, the Laplace operator of u is given by

$$\Delta u = \frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \varphi^2}.$$

This operator carries over to functions u on any Riemannian manifold as $\Delta u = \operatorname{div} \operatorname{grad} u$. The physical meaning of these expressions is the same as in the example with the Dirichlet integral considered above that dealt with area expansion.

The enemies of physics define the Laplace operator in their mathematical textbooks by relation $(*)$, which renders this physical object

[1] [**Translator's note**] English translation: Vladimir I. Arnold, *Lectures on Partial Differential Equations Springer-Verlag, Berlin–Heidelberg and PHASIS, Moscow, 2004, pp. 57–59.*

relativistically meaningless (it depends not only on the function to which the operator is applied, but also on the choice of coordinate system). On the contrary, the operators div, grad, rot, and Δ depend only on the Riemannian metric and do not depend on the coordinate system.

http://dx.doi.org/10.1090/mbk/085/05

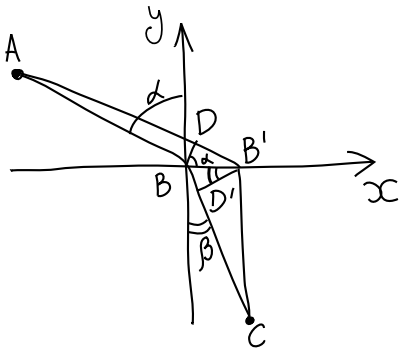
Chapter 5

Snell's Law of Refraction

The velocity of motion (in any direction) in the upper half-plane $y > 0$ (of the plane with Cartesian coordinates (x, y)) is equal to v. In the lower half-plane $y < 0$ the velocity of motion is $w = (3/4)v$ ($v = 1$ and $w = 3/4$ for light propagation in air and water, respectively).

The least-time path ABC from a point A in the upper half-plane to a point C in the lower half-plane is the broken line ABC that breaks at the point B on the interface.

Determine the ratio of the angles α and β made by the paths AB and BC with the normal to the interface.



Solution. Consider an interface point B' close to B: $|BB'| = \varepsilon$. Let the path AB' be longer than AB by the straight line segment $B'D$ of length $\varepsilon \sin\alpha + O(\varepsilon^2)$.

Similarly, the path CB' is shorter than CB by the segment BD' of length $\varepsilon \sin\beta + O(\varepsilon^2)$.

Therefore, the time required to traverse the path $AB'C$ is longer than that required to traverse ABC by

$$\Delta(\varepsilon) = \frac{\varepsilon \sin\alpha}{v} - \frac{\varepsilon \sin\beta}{w} + O(\varepsilon^2).$$

For the travel time along the path ABC to be minimal (for ε of any sign; i.e., for points B' both on the right and on the left of B), it is necessary that $\Delta(\varepsilon) = 0$ (in the first approximation in ε); i.e., that

$$(*) \qquad \frac{\sin\alpha}{v} = \frac{\sin\beta}{w}.$$

The quantity reciprocal to velocity is called the index of refraction and is usually denoted by $n = 1/v$. The above-obtained law $(*)$ of refraction on the interface between media with indices of refraction $n_1 = 1/v$ and $n_2 = 1/w$ can be written in the form of Snell's law as

$$n_1 \sin\alpha_1 = n_2 \sin\alpha_2.$$

Example. For a light beam traveling from air ($n_1 = 1$) into water ($n_2 = 4/3$), the law of refraction takes the form

$$\sin\alpha_1 = \frac{4}{3}\sin\alpha_2.$$

If the angle α_1 between the beam moving from air to water and the vertical normal to the horizontal water surface is small, then the angle α_2 between the refracted ray and the perpendicular is even less, about $(3/4)\alpha_1$.

Above we derived the law of refraction from Fermat's principle, according to which light beams reach their target in the shortest time.

Snellius himself discovered this law of refraction experimentally, by measuring the angles α and β in numerous examples.

The reader familiar with Huygens' principle (which describes wave propagation by using envelopes of families of local wave fronts) will be pleased to observe that Huygens' principle readily implies Snell's law (as a simple special case).

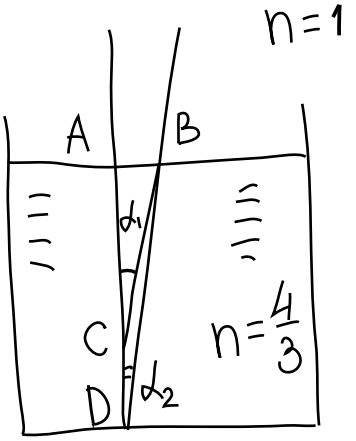
Interestingly, in all of these examples, the nature of propagating waves does not matter much. For example, acoustic and optical beams and fronts behave similarly, and the same mathematics applies to the theory of epidemic dynamics.

http://dx.doi.org/10.1090/mbk/085/06

Chapter 6

Water Depth and Cartesian Science

By how much does the depth of a water-filled pan on a table appear to be less than its true depth to an observer looking from above?



Solution. The triangles BAC and BAD are right, so that

$$|AB| = |AC| \tan \alpha_1 = |AD| \tan \alpha_2.$$

For a small angle of incidence α_1, we have

$$\frac{|AD|}{|AC|} = \frac{\tan \alpha_1}{\tan \alpha_2} \approx \frac{\sin \alpha_1}{\sin \alpha_2} = n = \frac{4}{3};$$

thus, the apparent depth $|AC|$ is one quarter less than the true depth $|AD|$.

Remark. Descartes should have looked into that pan before claiming that light propagates in water 30% faster than in air.

He made this conclusion because he knew that sound propagates in water faster than in air (about five times as fast).

Deductive inferences based on such analogies are very dangerous; they must always be tested experimentally. But Descartes solemnly announced that science is a sequence of derivations of deductive consequences from arbitrary axioms, and the experimental verification of these axioms does not belong to science (although may be useful for a market economy).

Of the several dozen of Descartes' similar "principles" the most dangerous is the following one: "The government should immediately prohibit all other methods of teaching except mine, because only this method is politically correct: following my course, any dimwit will advance as quickly as any genius, whereas under any other method of teaching, talent benefits learners."

Following his principles, Descartes expelled figures from geometry, which are, on the one hand, traces of experiments involving drawing straight lines and circles, and on the other hand, a niche for imagination, which Descartes endeavored to eliminate from science.

The former French president Jacques Chirac told me (on June 12, 2008, in the Kremlin) that it is for these features of Cartesian science that he hated mathematics since childhood. But then he added (in Russian): "Although, probably, this refers only to French, Bourbaki's, mathematics, while here I understand all you say. But not for nothing did your Fedor Ivanovich[1] say:

> Who would grasp Russia with the mind?
> For her no yardstick was created:
> Her soul is of a special kind,
> By faith alone appreciated.[2]

In Russia, nobody believes in Descartes' theory that light propagates in water faster than in air; in return, his remarkable theory of the rainbow is better known here than in France.

[1] [**Translator's note**] "Fedor Ivanovich" is the Russian poet F. I. Tyuchev (1803–1873) .

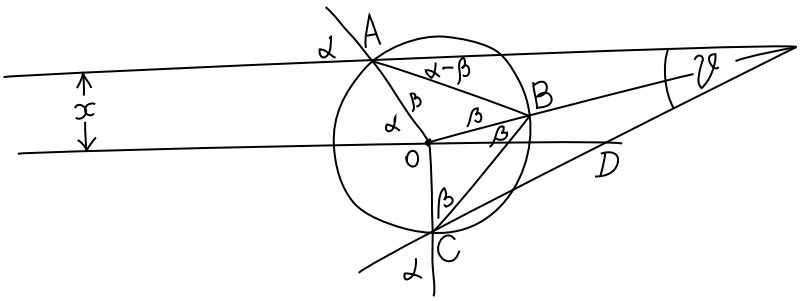
[2] Translated by John Dewey.

http://dx.doi.org/10.1090/mbk/085/07

Chapter 7

A Drop of Water Refracting Light

By what angle θ does a beam incident on a spherical water drop of radius r at a distance x from the ray OD passing through the center of the drop and parallel to the beam deviate from returning along the incidence direction?



Solution. The angle BOD is equal to $\beta - (\alpha - \beta) = 2\beta - \alpha$.

The deviation angle θ is twice as large (due to the symmetry in the OB axis): $\theta = 4\beta - 2\alpha$.

According to the law of refraction, we have $\sin\alpha = n\sin\beta$, and by the definition of the incident ray, we have $r\sin\alpha = x$. Therefore, $\alpha = \arcsin(x/r)$, $\beta = \arcsin(x/(nr))$, and

$$\theta(x) = 4\arcsin\frac{3x}{4r} - 2\arcsin\frac{x}{r}.$$

Although this expression answers the question, its meaning becomes clear only after the graph of the calculated function θ is constructed. This investigation explains both the amazing glitter of dewdrops and rainbows in a rainy sky.

http://dx.doi.org/10.1090/mbk/085/08

Chapter 8

Maximal Deviation Angle of a Beam

Which of the beams incident on a spherical water drop deviates from returning along itself by a maximal angle $\theta_{\max}$ (and by what angle exactly)?

Solution. Let us denote $3x/(4r)$ by u, then $x/r = nu$ and

$$\frac{\theta}{2} = 2\arcsin u - \arcsin nu.$$

The derivative of $\theta/2$ with respect to u must vanish for a beam with maximal deviation angle $\theta_{\max}$: for this beam, we have

$$\frac{2}{\sqrt{1-u^2}} = \frac{n}{\sqrt{1-n^2u^2}}, \qquad \frac{4}{1-u^2} = \frac{n^2}{1-n^2u^2},$$

so that

$$u_{\max}^2 = \frac{4-n^2}{3n^2}, \qquad \frac{\theta_{\max}}{2} = 2\arcsin u_{\max} - \arcsin nu_{\max}.$$

For $n = 4/3$, we find

$$u_{\max}^2 = 5/12, \qquad u_{\max} = \frac{\sqrt{5/3}}{2}, \qquad nu_{\max} = \sqrt{5/3}\cdot 2/3.$$

Since $5/3 \approx 1.666$, we readily calculate

$$\sqrt{5/3} \approx \sqrt{166.6}/10 \approx 1.29.$$

Thus,

$$u_{\max} \approx 0.645, \qquad nu_{\max} \approx 0.86.$$

Since

$$\sin(\pi/6) = 0.5, \qquad \sin(\pi/4) \approx 0.707, \qquad \sin(\pi/3) \approx 0.86,$$

it follows that

$$\arcsin nu_{\max} \approx \pi/3, \qquad \arcsin u_{\max} \approx \pi/4 - \pi/40,$$

whence

$$\frac{\theta_{\max}}{2} \approx \frac{\pi}{2} - \frac{\pi}{20} - \frac{\pi}{3}, \qquad \theta_{\max} \approx \frac{\pi}{3} - \frac{\pi}{10},$$

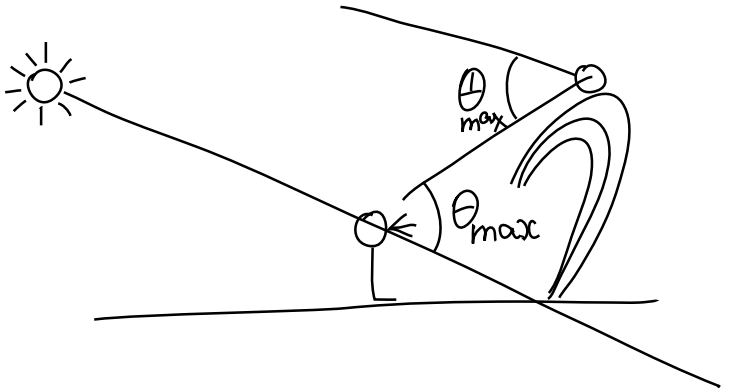
which is about 42°.

http://dx.doi.org/10.1090/mbk/085/09

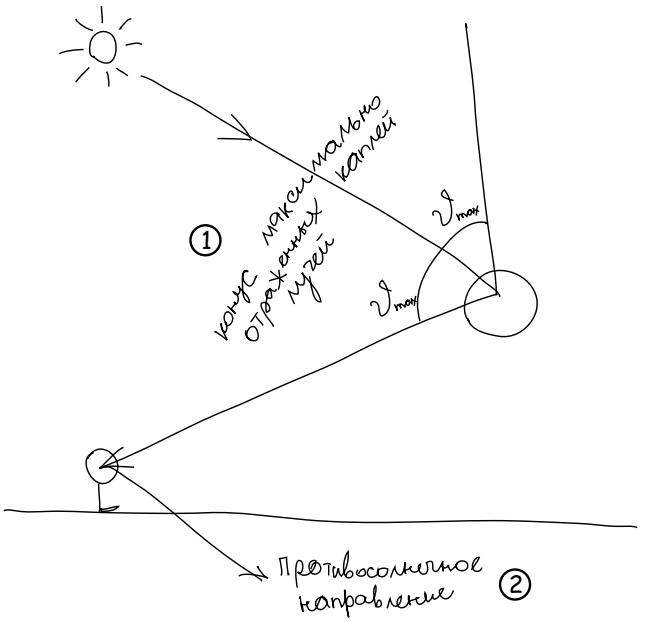
Chapter 9

The Rainbow

Why does an observer see a rainbow as an arc centered at the antisolar point at an angle of about 42° *?*

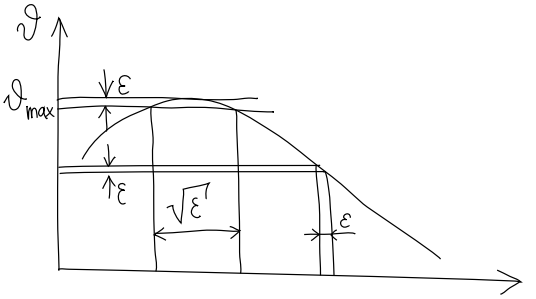


Solution. The beams most bent by refraction carry the maximum energy:



① *The cone of beams reflected most by a droplet.* ② *The antisolar direction.*

The energy of the beams at angles from θ to $\theta+\varepsilon$ of nonmaximal deviation is proportional to ε, while the energy of the beams within an angle of the same width ε, around $\theta_{\max}$, is much higher (it is proportional to $\sqrt{\varepsilon}$).

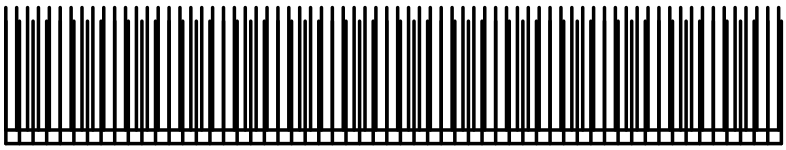


For this reason, these beams are visible, and they are seen as a rainbow. The point is that the refraction indices of light beams of different colors are somewhat different, so that the maximum angle of

deviation $\theta_{\max}$ differs between beams of different colors. This is why a rainbow is multicolored.

Remark. A second rainbow (inside the primary one) is caused by beams reflected more than once on the rear surface of droplets. For these beams, the maximum angle of deviation in somewhat less than 42°.

The blueness of the sky also has a mathematical explanation: looking at a phonograph record from one side, one can observe iridescent colors, which are explained by the interference of light waves in the grating of the record groove (this phenomenon is similar to a moire, that is, to the long-period pattern obtained when a fence is projected on another fence).



The blueness of the sky is due to the moire-like interference of sunbeams caused by density fluctuations in the rarefied atmosphere.

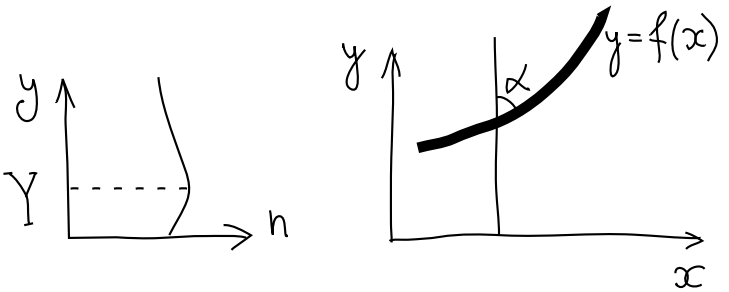
http://dx.doi.org/10.1090/mbk/085/10

Chapter 10

Mirages

The index of refraction $n(y)$ of air at altitude y over a desert is maximal at a certain altitude Y (at which the air density is maximal: The heat of the desert drives lower layers up, and at high altitude, the density of the atmosphere drops down to zero).

Explain the mirage phenomenon in view of such a behavior of the index of refraction.



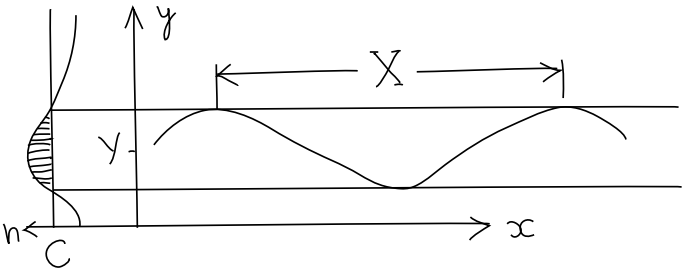
Solution. Let us study the motion $y = f(x)$ of light beams by using the law of refraction $n \sin \alpha = \text{const}$, where α is the angle between the beam and the vertical.

We obtain a (differential) equation for beams of the form

$$\alpha(y) = \arcsin \frac{C}{n(y)}.$$

The parameter C is determined by the choice of the beam under examination. We conclude that a beam (with fixed C close to Y) is entirely contained in the strip where $n(y) \geqslant C$ (and oscillates between

its boundaries):



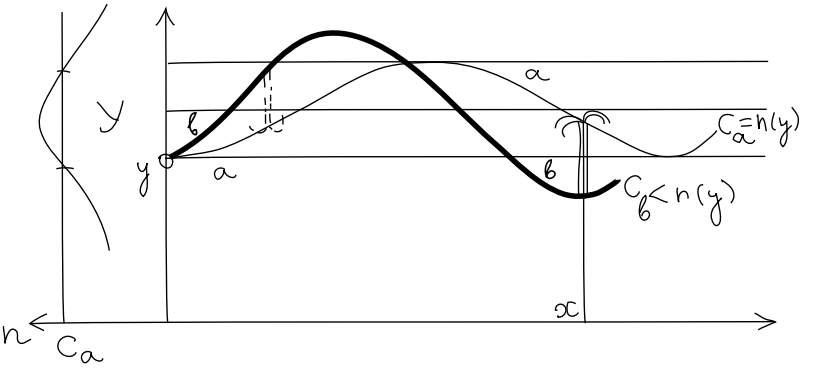
These oscillations render the beam wave-shaped (with wavelength X depending on the constant C).

The value $X(C)$ is finite if C is not a critical value for the index of refraction n: if $dn/dy \neq 0$ at the points where $n(y) = C$.

As the constant C increases to the critical value $n(Y)$ of the index of refraction, the wavelength $X(C)$ grows to infinity, the wave amplitude tends to zero, and the beam propagates along the line $y = Y$.

To understand how the tortuosity of beams affects the images of remote palms, let us look from the point $(x = 0, y)$ at a palm growing at a distance x.

Let us draw rays a and b from the top and the bottom of the palm to the observation point $(x = 0, y)$.



At the observation point $(0, y)$, the ray a, which issues from the top of the palm, is below the ray b, which issues from its bottom. Therefore, the image of the palm turns upside down—that is what the mirage phenomenon is all about.

Remark. To comprehend all this, we must clearly understand how the geometry of beams of light is related to the images (of the emitting objects by these beams for the observer).

This relation ("image construction") is explained when lenses are described in high-school physics courses, but only a few students understand this theory. (To see a mirage, it is not necessary to go to a desert: in summer, looking along the platform while awaiting a commuter train, it is easy to see puddles at a distance, although the platform is perfectly dry; noticing this, smart kids come around to the theory described above, but they are few.)

http://dx.doi.org/10.1090/mbk/085/11

Chapter 11

Tide, Gibbs Phenomenon, and Tomography

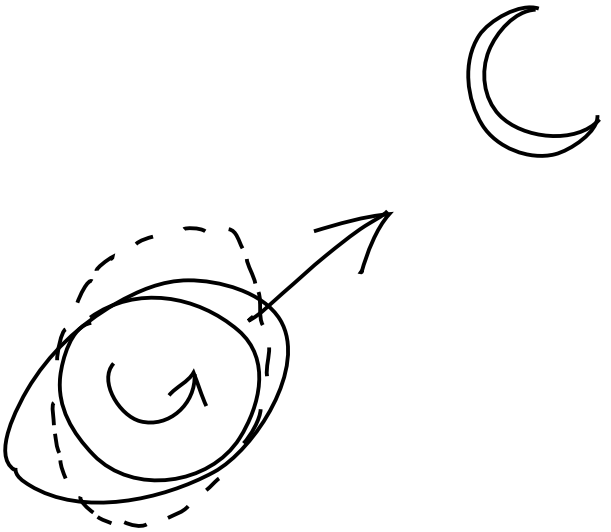
In the city C, the tide occurred at noon today. When will it occur tomorrow?

Solution. Tides are explained by the attraction of the Moon: roughly speaking, this attraction causes the formation of two bulges (one directed toward the Moon and the other in the opposite direction) on the equipotential surface of the Earth's gravity field.[1] Under the influence of this field, ocean water tends to occupy a position in which its surface is aligned with the equipotential (i.e., it is "horizontal")

This is what causes tides: since Earth rotates about its axis once in 24 hours, it follows that the vertex of the bulge directed to (or

[1]Both Kepler and Copernicus discussed two possibilities for the gravitational attraction force; they thought that the decrease in this force with the increase of the distance is inversely proportional to either the distance or the squared distance. The conclusion from this discussion was that the inverse proportionality to the squared distance is more likely, because otherwise tides would be many times higher.

from) the Moon moves with respect to Earth's continents.



As is known, the Moon moves around Earth and performs a complete rotation in a month (about 28 days), in the plane (of the ecliptic) inclined (not too strongly) to the plane of Earth's equator, in the same direction as Earth rotates about its axis (from the West to the East as viewed from the North).

During one day, the Moon shifts by about 1/28 of its orbit with respect to Earth, approximately in the direction of its own rotation. The bulge attracted by the Moon will be directed toward the new position of the Moon tomorrow at noon, and Earth must move by 1/28 of its full rotation for the city C to reach the bulge. Since Earth performs a complete rotation in 24 hours, it must additionally rotate 24/28 hours, which is approximately 50 minutes, for the city C to fall under the tidal bulge.

Thus, the tide will occur in the city C at 50 minutes after midday tomorrow.

Remark. Of course, we used a highly simplified model for the complex phenomenon of tides, assuming that water has time to follow the equipotential bulges. In reality, it is somewhat behind (in different

cities, the lags are different); our model would be more accurate if Earth rotated slower. The attraction of the Sun creates tides as well (they are lower than the lunar ones but are particularly noticeable during the spring and autumn equinoxes, when solar tides are added to lunar ones rather than subtracted from them).

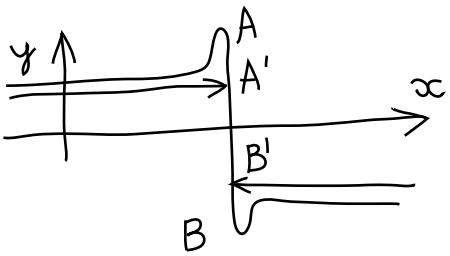
But the answer, a 50-minute delay, agrees well with observations. Apparently, this is caused by the fact that the lag of water behind bulges today and tomorrow is approximately the same.

A detailed prediction of tides in certain geographic zones requires a significant amount of mathematical computation.

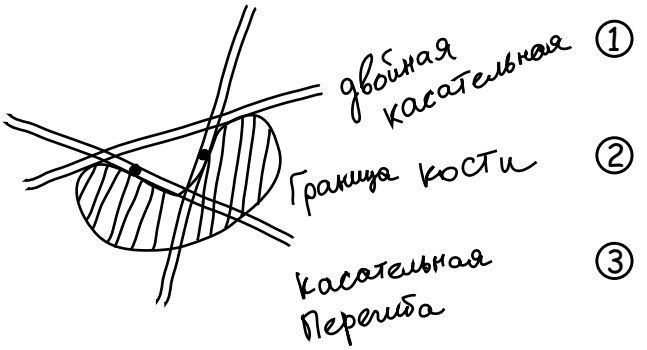
Working on these computations, J. Gibbs experimentally discovered the following amazing fact (which is now known as the Gibbs phenomenon but, unfortunately, is not included in calculus courses):

The limit of the graphs of functions which form a convergent sequence may strongly differ from the graph of the limit function.

Of course, the point is that the sequence may converge nonuniformly. Gibbs noticed this when expanding a discontinuous function in its Fourier series. Near the point of discontinuity (of the simplest type), the limit of the graphs of partial sums contains, in addition to the interval joining the left and right limit values, its extension (AB is about 9% longer than $A'B'$).



Nowadays, this Gibbs phenomenon is used in tomography in order to explain the "artifacts" observed on tomograms: the increase in brightness of the plane section of the body at points of double tangents and tangents at inflection points to the cross section of the bone.



①*Double tangent* ②*Bone boundary* ③*Inflection tangent*

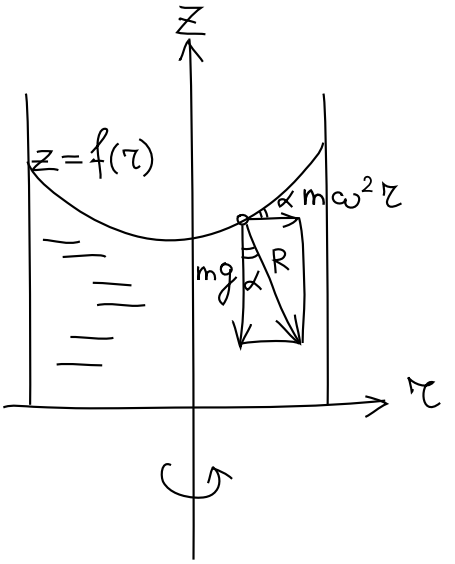
http://dx.doi.org/10.1090/mbk/085/12

Chapter 12

Rotation of a Liquid

A glass of water is put on a uniformly rotating phonograph record (e.g., at the center of this record, so that its axis of rotation coincides with the axis of symmetry of the glass).

What shape does the water surface acquire?



Solution. It is clear from the symmetry that this will be a surface of revolution with an equation of the form $z = f(r)$, where r is the distance to the axis of rotation and z is the height of the water.

The centrifugal force acting on a mass m at distance r from the axis of rotation with angular velocity ω is $m\omega^2 r$, and the force of gravity is mg.

The condition that the resultant force R is orthogonal to the water surface is that the tangent of the angle α between this surface and a horizontal radius of the glass equals

$$\frac{m\omega^2 r}{mg} = cr,$$

(where the constant $c = \omega^2/g$ does not depend on the point of the water surface but rapidly increases with the angular velocity ω of rotation).

For the function f, we obtain the following differential equation (which specifies the slope of the graph of this function):

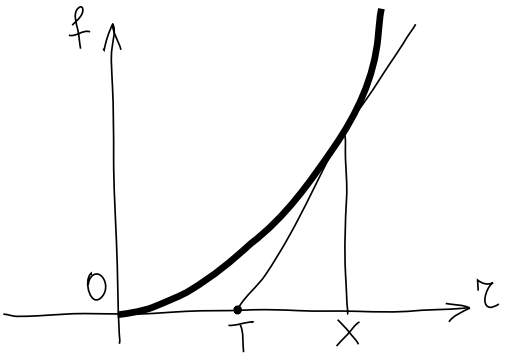
$$\frac{df}{dr} = cr.$$

Its solution

$$f(r) = f(0) + \frac{c}{2}r^2$$

shows that the water surface has the shape of a paraboloid of revolution.

Remark. Our differential equation means that a tangent to a parabola bisects the corresponding segment of the axis of abscissas: $|OT| = |OX|/2$, because $(cr^2/2)/(cr) = r/2$.



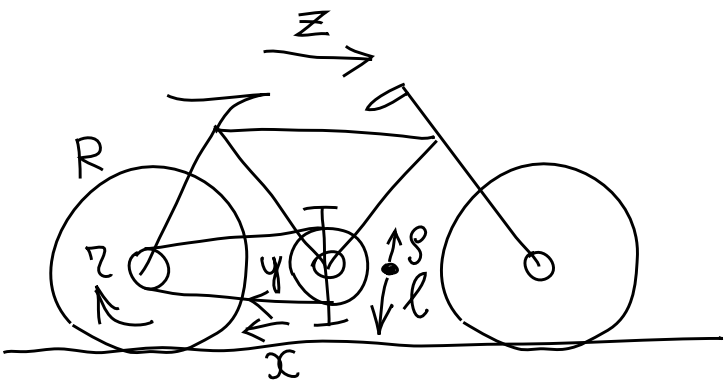
For a parabola of degree a, we have (as Archimedes already knew) $|TX| = |OX|/a$; thus, for a cubic parabola, T is two times closer to X than to O.

http://dx.doi.org/10.1090/mbk/085/13

Chapter 13

What Force Drives a Bicycle Forward?

The lower pedal of a bicycle standing still on a horizontal floor is pulled back. Which way does the bicycle go, and in what direction does the pulled back lower pedal move with respect to the floor?



Solution. Let us denote the length of the crank arm (from the pedal to the axle) by l, the radii of the front and rear sprockets (toothed wheels) by ρ and r, and the radius of the rear wheel by R.

Let x be the (backward) displacement of the pedal with respect to the axle. The lowest tooth of the front (and, hence, the rear) sprocket moves back a distance of $y = x(\rho/l)$.

Therefore, the rear wheel turns by an angle such that its point of tangency with the floor covers a distance of

$$z = y\Big(\frac{R}{r}\Big) = x\Big(\frac{\rho}{l}\Big)\Big(\frac{R}{r}\Big).$$

Looking at the bicycle, we estimate the parameter values as

$$l \approx 2\rho, \qquad R \approx 10r.$$

Therefore, the displacement z of the bicycle with respect to the floor is

$$z \approx 5x \qquad \text{(forward!)}.$$

This is the displacement of the axle; the pedal moves backward by x relative to the axle crank arm and forward by $4x$ relative to the floor.

Answer. The bicycle moves forward, and the lower pedal pulled back moves forward as well but 20% less than the whole bicycle.

Remark. It seems surprising that a force directed back (applied to the pedal) forces the bicycle to move forward. But the rotation of the rear wheel creates at its point of tangency with the floor a forward friction force, which wins.

Original Editor's Note. After the first edition of this book was published, some readers correctly noticed that the model considered above is inaccurate.

Corrections were being preliminarily discussed with the author, and it was planned to finalize them before publishing a new edition of the book. The sudden death of Vladimir Igorevich Arnold on June 3, 2010, prevented this.

Considering it wrong to change the original text, we leave the construction of a correct model to the reader. It is assumed that the force is applied to the pedal (rigidly connected with the wheel) by a rider sitting on the bicycle's saddle.

http://dx.doi.org/10.1090/mbk/085/14

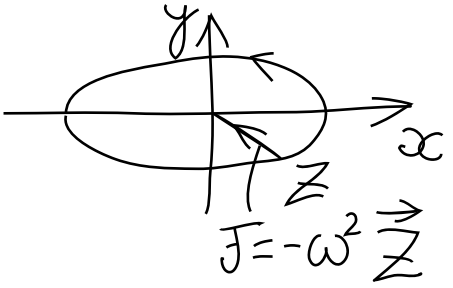
Chapter 14

Hooke and Keplerian Ellipses and Their Conformal Transformations

A point of the Euclidean plane attracted to the origin by a force proportional to the distance from this point to the origin (“Hooke’s law” $\ddot{\vec{z}} = -\omega^2 \vec{z}$) moves along a Hooke ellipse centered at the origin, which is given by

$$(*) \qquad x = a\cos(\omega t), \qquad y = b\sin(\omega t)$$

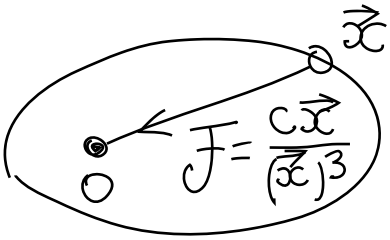
under an appropriate choice of Cartesian coordinates x and y in the plane of motion.



The gravitational field (of attraction with strength inversely proportional to the distance from the origin) causes the attracted point (if

its initial velocity is not too large) to move along a Keplerian ellipse with the attracting center at one of its two foci.

Prove that if the Euclidean plane is treated as the complex line ($z = x + iy$), then the squared points z of any Hooke ellipse form a Keplerian ellipse.



Solution. Consider the complex number of modulus r with argument ωt, that is,

$$\zeta = re^{i\omega t} = r\cos(\omega t) + ir\sin(\omega t).$$

The reciprocal complex number has modulus r^{-1} and argument $-\omega t$:

$$\zeta^{-1} = r^{-1}\cos(\omega t) - ir^{-1}\sin(\omega t).$$

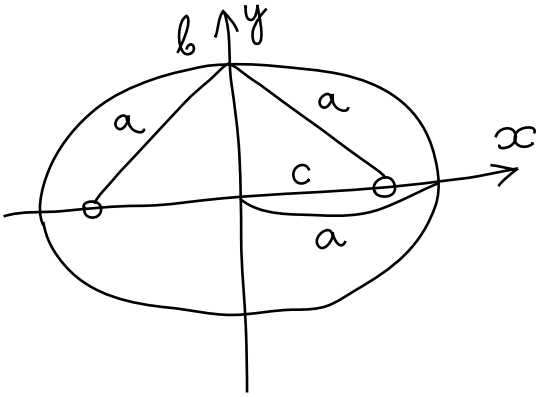
Therefore, the sum

$$Z = \zeta + \zeta^{-1} = (r + r^{-1})\cos(\omega t) + i(r - r^{-1})\sin(\omega t)$$

of these two complex numbers belongs to the Hooke ellipse with

$$a = r + r^{-1}, \qquad b = r - r^{-1}.$$

For simplicity, we assume that $r \geqslant 1$. Then a is the semi-major axis of this Hooke ellipse and b is its the semi-minor axis. A point moving according to Hooke's law describes this ellipse once while the point ζ performs one full rotation on the circle $|\zeta| = r$.



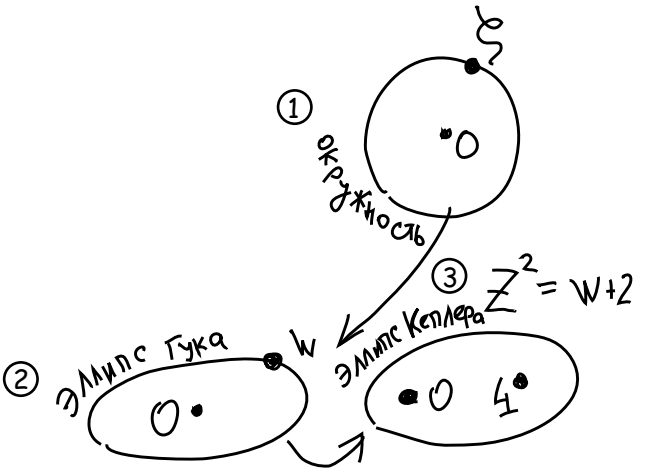
For the foci of the ellipse thus obtained, the Pythagorean theorem gives $c^2 = a^2 - b^2 = 4$, so that the distance between the center and each focus is equal to 2.

Any ellipse with the same distance between its foci can be constructed in this way (for a suitable radius r of the initial circle). Moreover, any ellipse (centered at the origin) can be constructed in this way under a suitable choice of the direction of the coordinate axes and the scale.

Squaring the complex numbers $Z = \zeta + \zeta^{-1}$, that is, the points of the Hooke ellipse, we obtain

$$Z^2 = \zeta^2 + \frac{1}{\zeta^2} + 2.$$

As a point $\zeta = re^{i\omega t}$ describes a circle of radius r once, the point $\zeta^2 = r^2 e^{2i\omega t}$ describes a circle of radius r^2 twice.



① *Circle.* ② *Hooke ellipse.* ③ *Keplerian ellipse.*

Thus, the point

$$W = \zeta^2 + \frac{1}{\zeta^2} = X + iY$$

twice describes the Hooke ellipse centered at 0 whose foci (the points $c = \pm 2$) are at distance 2 from the center:

$$X = A\cos(2\omega t), \qquad Y = B\sin(2\omega t).$$

Recall that $Z^2 = W+2$. Therefore, the point Z^2 describes (twice) an ellipse with foci 0 and 4 (i.e., a Keplerian ellipse), as required.

Remark 1. Taking an appropriate initial ellipse (centered at the origin $Z = 0$), we can obtain any ellipse with a focus at 0 as the set of points Z^2. This follows from the above consideration of special ellipses of the form $(*)$ under a suitable choice of the directions of the coordinate axes and the scale.

Remark 2. A motion along a Hooke ellipse governed by Hooke's law does not transform into a motion along a Keplerian ellipse governed by Kepler's law when the points of Hooke's ellipse are squared. Indeed, constant areal velocities cease to be constant when complex points are squared (because areal velocities only double, while squared distances to the origin are multiplied by different numbers).

Remark 3. Surprisingly, raising complex numbers to an appropriate power transforms the orbits of motion in a central field whose force of attraction is proportional to the αth power of the distance from the center into the orbits of motion in a central field whose force of attraction is proportional to the βth power of the distance from the center.

Here the exponents α and β in the dual laws of attraction are related by

$$(\alpha + 3)(\beta + 3) = 4.$$

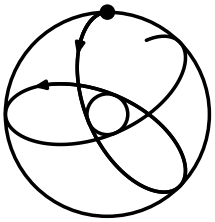
Example. Hooke's law corresponds to $\alpha = 1$, and the law of universal gravitation corresponds to $\beta = -2$.

The power to which the points of an orbit in an α-field should be raised in order to obtain points of an orbit in a β-field is $\gamma = (\alpha+3)/2$.

Thus, for $\alpha = 1$, we have $\gamma = 2$; i.e., Hooke ellipses transform into Keplerian ellipses when complex numbers are squared.

For $\alpha = -2$, we have $\gamma = 1/2$; i.e., Hooke ellipses are obtained from Keplerian ellipses by taking the square root of complex numbers.

Interestingly, the (dual) laws of Hooke and of universal gravitation describe two unique central fields in which all orbits close to circular ones are closed; in all other cases, their shapes are similar to the epicycloid (in the annuli between the apocenters and pericenters).



Remark 4. The transformation of orbits of motion in a central field with exponent α into orbits of motion in a central field with dual exponent β is the same in quantum mechanics: the solutions of the Schrödinger equation corresponding to the first attracting field are mapped under this transformation of the plane into solutions of the Schrödinger equation corresponding to the second one (R. Faure,

Transformations conformes en mécanique ondulatoire, C. R. Acad. Sci. Paris, vol. 237, pp. 603–605 (1953)).

Although this conclusion can also be made from direct calculations (similar to those given above), it is much easier to transform Lagrangians of variational principles equivalent to the Schrödinger equation (by appropriately transforming the total energy value and the eigenvalues) rather than its solutions.

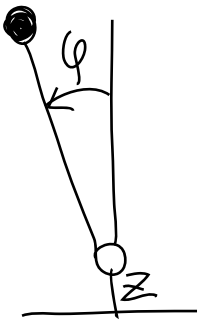
Interestingly, the whole theory of duality between fields with potentials $|\partial w/\partial z|$ and $|\partial z/\partial w|$ described above carries over (both in classical mechanics and for the Schrödinger equation) from the case of the conformal mapping $w = z^2$ of the Hooke–Kepler duality not only to the case $w = z^\gamma$ (as shown above, for forces with exponents α and β satisfying the condition $(\alpha+3)(\beta+3) = 4$, we have $\gamma = (\alpha+3)/2$) but also to the case $\gamma = \infty$, which corresponds to the conformal mapping $w = e^z$, $z = \ln w$. (The strange relation $w^0 = \ln w$ is explained later on in the essay "Mathematical notion of potential" on pp. 115–126.)

http://dx.doi.org/10.1090/mbk/085/15

Chapter 15

The Stability of the Inverted Pendulum and Kapitsa's Sewing Machine

Suppose that the pivot point of a pendulum oscillates in the vertical direction so that $z = a\cos(\Omega t)$. *If the frequency* Ω *of these oscillations is sufficiently high, then the inverted pendulum will remain steady in its upward position (for* $\varphi = 0$ *in the figure).*



Solution. We pass to the (noninertial) coordinate system in which the pivot point is fixed. To the force of gravity acting on the pendulum we must add the inertia force, which is proportional to the acceleration of the coordinate system, that is, to

$$\ddot{z} = \Omega^2 a\cos(\Omega t).$$

This is equivalent to oscillations of the gravitational constant g in the usual equation

$$\ddot{\varphi} = (l/g)\sin\varphi$$

of an (inverted) pendulum of length l.

The study of the second-order differential equation thus arising (in which the coefficients vary periodically with time) is provided for by KAM theory (see; e.g., the 1967 book *Ergodic Problems of Classical Mechanics* by V. Arnold and A. Avez reprinted in 1999 in Izhevsk by the journal *Regulyarnaya i Khaoticheskaya Dinamika*, pp. 87–90, 245–263).[1]

Replacing the difficult nonlinear equation of motion of the pendulum by its linearization, we obtain the linear equation

$$\ddot{\varphi} = (l/g)\varphi$$

"of small oscillations" for the inverted pendulum.

The eigenvalues of the monodromy operator of this linear equation with periodic coefficients can be calculated, at least approximately, by integrating the equation over the period ($0 \leqslant t \leqslant T = 2\pi/\Omega$) on a computer or by means of perturbation theory (as described; e.g., in V. Arnold's book *Ordinary Differential Equations*, on pages 281–289 of the fourth 2000 Izhevsk edition.[2]

From these calculations of the eigenvalues of the monodromy operator it follows that, for an inverted pendulum of length $l = 20$ cm whose pivot point oscillates with amplitude 1 cm, the equilibrium position $\varphi = 0$ of the linearized equation is stable when the pivot point performs vertical oscillations with frequency more than 30 oscillations per second.

The fact that this stability is preserved for nonlinear pendulums is not as obvious, but it is true.

Remark. This problem arose in the theory of accelerators. One of the projects was based on the stability of an inverted pendulum

[1] **[Translator's note]** English translation: V. I. Arnold and A. Avez, *Ergodic Problems of Classical Mechanics* (Benjamin, New York, 1968), pp. 88–90, 250–269.

[2] **[Translator's note]** English translation: V. I. Arnold, *Ordinary Differential Equations* (MIT Press, Cambridge, Mass., 1973), pp. 199–207.

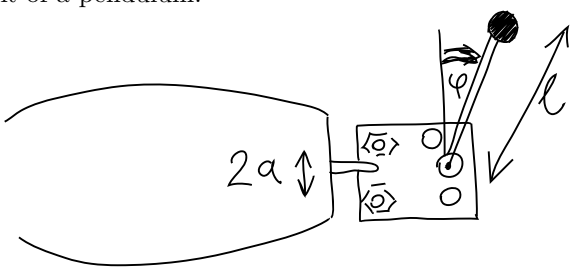
with a vertically oscillating pivot point (the problem on the stability of the circular motion of accelerated particles reduces to the same equation).

P. L. Kapitsa suggested that, before millions are spent for building an accelerator, the conclusion about pendulums should be checked experimentally. He rebuilt an electric sewing machine so that its rotation caused vertical oscillations of the pivot point of the pendulum.

The pendulum stood steadily upward, and when slightly displaced sideways, it began to swing about this vertical position in the same way as an ordinary pendulum swings about its lower equilibrium position.

When Kapitsa was the chairman of the organizing committee of a physics olympiad for school students and Arnold was the chairman of the organizing committee of a math olympiad (the committees sat together at the Institute for Physical Problems), P. L. demonstrated his sewing machine-cum-pendulum, which was kept in the next room as a relic, to the members of both committees.

Arnold, who did not have an electric sewing machine, adapted a *Neva* vibrating electric shaver to create vertical oscillations of the pivot point of a pendulum.[3]



The upper equilibrium position turned out to be unstable, because the length $l = 20$ of the pendulum was too large. Arnold had to perform the (linearized) calculations whose results are presented above.

After the pendulum was shortened to 10 cm, its oscillations (about the upper equilibrium position) became stable, and then Arnold proved

[3] A video record of the operation of this electric shaver is stored at the site "Mathematical Etudes" (`http://etudes.ru`).

this stability by using KAM theory (this theory includes a general theorem on the stability of elliptic fixed points, which substantiated the possibility of judging the stability of a nonlinear system from its linearization, as early as in 1961).

The accelerators had been already constructed at that time, because physicists were satisfied by the experimental verification of stability in Kapitsa's experiments with his sewing machine (notwithstanding that they did not yet possess mathematical KAM theory rigorously proving this nontrivial nonlinear stability).

http://dx.doi.org/10.1090/mbk/085/16

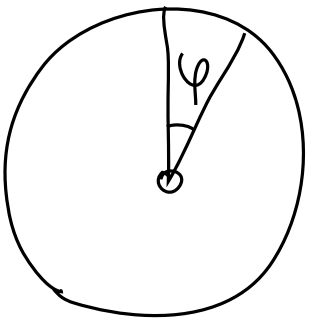
Chapter 16

Space Flight of a Photo Camera Cap

An astronaut in a spaceship flying along a circular orbit threw a photo camera lens cap to Earth (*say with velocity* 10 *m/sec*). *Where will it fly?*

Describe the orbit of the cap relative to the spaceship (*in the plane of the orbit*).

Solution. Let r denote the distance to the center of Earth, and let φ be the central angle (counted from some point on the orbit). We take the radius of the orbit for the unit of length and choose the unit of time so that the orbital period is 2π.



The differential equation for the law of universal gravitation is written in this coordinate system as

$$\ddot{\vec{r}} = -\frac{\vec{r}}{r^3}.$$

Let us investigate the solutions of this equation close to the circular motion ($r_0 = 1$, $\varphi_0 = t$). We seek them in the form ($r = r_0 + r_1$, $\varphi = \varphi_0 + \varphi_1$); after linearization, we obtain the variational equations

$$\ddot{r}_1 = 3r_1 + 2\dot{\varphi}_1, \qquad \ddot{\varphi}_1 = -2\dot{r}_1$$

for the perturbations r_1 and φ_1.

We solve these equations under the initial conditions determined by the problem (these are $r_1(0) = \varphi_1(0) = \dot{\varphi}_1(0) = 0$ and $\dot{r}_1(0) = -1/800$, because the first cosmic velocity, which is equal to 1 in our coordinate system, is about 8 km/sec).

The variational equations imply $\dddot{r}_1 = -\dot{r}_1$, whence, taking into account the initial conditions, we obtain

$$r_1 = -\frac{1}{800}\sin t, \qquad \varphi_1 = \frac{2}{800}\cos t.$$

This means that the cap moves along an ellipse with major axis about 32 km and minor axis about 16 km whose center is 16 km ahead of the ship on the orbit. In about an hour and a half (which is the rotation period of the ship), the lens cap will describe its one-hundred-kilometer elliptic orbit around the ship and return to the ship from above, passing at a distance of a few dozen meters from the ship, because its true motion differs from the first approximation considered above.

Remark. The flight of the lens cap described above indeed occurred during the walk of astronaut Leonov in outer space (it was Leonov's narrative which had led V. V. Beletskii to the above calculations).

Leonov, however, said that the lens cap, which he had thrown to Earth, "flew there": he did not expect that it would return back (in an hour and a half). His description of the experiment was correct: the first kilometer of the one-hundred-kilometer ellipse is very nearly a segment of the straight path to Earth; at greater distances, the lens cap was indiscernible.

http://dx.doi.org/10.1090/mbk/085/17

Chapter 17

The Angular Velocity of a Clock Hand and Feynman's "Self-Propagating Pseudoeducation"

On a horizontal table in St. Petersburg, a watch lies face-up. Where does the angular velocity vector of the hour hand point?

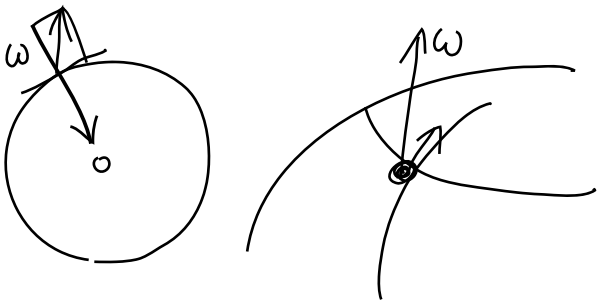
Solution. The angular velocity of the hand relative to the watch case is directed downward (because the hand moves clockwise), and in magnitude, it is double that of Earth's rotation about its axis (because the hand rotates once every 12 hours, while Earth rotates once every 24 hours).

The angular velocity vector of Earth's rotation points to the northern pole star (because Earth rotates counterclockwise as viewed from that star).

The angular velocity of the overall motion is the sum of these two vectors (the angular velocity of Earth's motion and the angular velocity of the hand relative to Earth).

The latitude of St. Petersburg is 60°; therefore, the components of a vector of the angular velocity of Earth (of length ω) that are directed East along a parallel, North along a meridian, and upward

are $(0, \omega/2, \omega\sqrt{3}/2)$.



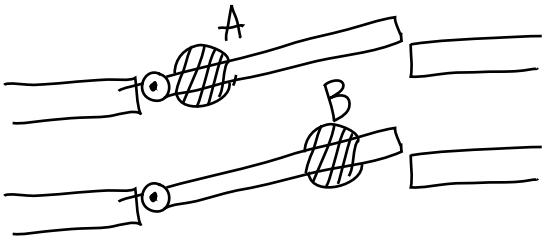
The components of the angular velocity vector of the hour hand with respect to Earth (in the same orthogonal coordinate system) are $(0, 0, -2\omega)$.

Thus, the components of the angular velocity vector of the hand are $(0, \omega/2, -\omega(4-\sqrt{3})/2)$: it is coplanar with the meridian but points in a direction different from that of the pole star (whose angle of elevation is 60° in St. Petersburg), specifically, to a point below the horizon with angle of elevation $\arctan(-(4-\sqrt{3})) \approx -66°$.

Everyone is taught angular velocity at school, but a few gain the understanding needed to solve the above problem.

Richard Feynman mentions (in his book *Surely, You're Joking, Mr. Feynman!*) that education, even university education, gets students into "this funny state of self-propagating 'education'," in which a student properly understands nothing but can successfully pass exams.

For example, according to Feynman, the definition of the moment of inertia of a point mass with respect to an axis and the squared distance to the axis makes no sense for students so long as there is there is no discussion about how much easier it is to push a door open when you put a heavy weight A near the hinge as compared to when you put the same weight B on the edge opposite the hinge:



But, unfortunately, professors confine themselves to the (correct) statement $I = \sum mr^2$, and it is this statement that students are supposed to know at the exam.

At an optics exam, Feynman asked a student what would happen to the image of an examination program lying under a piece of glass if he tilted the glass by an angle α. The student, naturally, answered that the image would turn by 2α (although he had just been answering about Snell's law, he did not know the relationship between the geometry of beams and the position of the visible image). Feynman's question as to whether the student had confused a plane-parallel glass plate with a mirror did not help at all.

http://dx.doi.org/10.1090/mbk/085/18

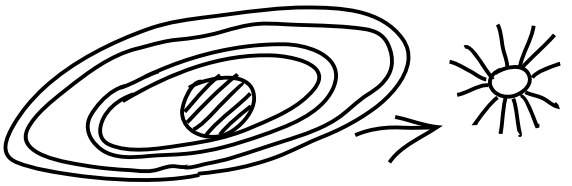
Chapter 18

Planetary Rings

Orbiting around the Sun, the planet Uranus obscured a star far from Earth (for a short time). Astronomers prepared for this event well in advance, but, on the due night, the star became invisible earlier than expected. Then it appeared, disappeared again, and there were four such disappearances observed until Uranus "occulted the disk of the star."

After that, the star hid behind Uranus, as predicted by astronomers, was obscured for the expected time, and appeared again; then, it disappeared four more times (for a short time).

How can these disappearances be explained?



Solution. The most natural guess is that Uranus, like Saturn, is surrounded by rings. Four concentric rings separated (like those of Saturn) by gaps must obscure the star four times before and four times after Uranus passes in front of it. Observations provide the sizes of the rings and gaps.

Remark. The gaps between Saturn's rings are explained by the perturbing influence of the attraction of the ice blocks constituting the rings by the satellites of this planet. Such perturbations render

the orbiting of a block unstable if the distance between the orbit and the planet is such that the rotation on the orbit is in resonance with the rotation of the satellite (say, the period of rotation is half that of the satellite: it is the rational ratios of periods that are dangerous for stability).

The knowledge of the size of the gaps between Uranus' rings observed during the passage of Uranus in front of the star has enabled astronomers (A. M. Fridman and others) to predict the radii of the orbits of Uranus' five perturbing satellites, which then were not known (but were discovered during the subsequent flight of *Voyager* past Uranus).

Interestingly, an international astronomical journal rejected the predictions of Soviet astronomers, motivating this by the fact that "the journal is published in a country where a different theory of gaps between Saturn's rings prevails."

This "different theory" predicted Uranus' satellites as well, but in reality, these predicted satellites were not in their places, and the American *Voyager* expedition did not find them.

I believe that Nobel prizes were created for crowning scientific discoveries confirmed by subsequent experiments or observations, such as the theory of Uranus' rings described above.

But American astronomers with whom I later discussed this objected that "their purpose is to support American theories rather than Russian ones."

Fortunately, neither Nobel prizes, Fields Medals, nor other similar distinctions exert substantial influence on the onward development of the natural sciences, which progress not so much by the decisions of various academies, as by the curiosity of explorers, to whom I address this book.

Zel'dovich used to say, chuckling, that he, and I, and Sakharov, and Kolmogorov—we are all members of the ChVAN (from the verb "chvanit'sya"[1]) society, which means literally "Chlen Vsekh Akademii Nauk"[2] But Kolmogorov valued only one such distinction: being in

[1] [**Translator's note**] The Russian verb "chvanit'sya" means "to swagger".
[2] [**Translator's note**] "Member of All Academies of Sciences".

the list of Honorary Members of the London Mathematical Society. Of course, he was also a member of the London Royal Society (the Academy of Sciences of England), but he did not value this so highly: the first Russian member, proposed to that society by the president, Newton, was Aleksandr Danilovich Menshikov[3] (who was illiterate and signed his own consent to join (written for him by Shafirov) by four crosses, which were shown to me when the record of my election as a member was put in the same folder with other Russian members of the Royal Society).

[3] [**Translator's note**] *A.D. Menshikov* and *P.P. Shafirov*, dignitaries at the court of Peter the Great, were sent on a mission by the Tsar to England.

http://dx.doi.org/10.1090/mbk/085/19

Chapter 19

Symmetry (and Curie's Principle)

Draw a straight line through the center of a homogeneous cube so that the moment of inertia of the cube with respect to this line is maximal.

Solution. Consider the ellipsoid of inertia of the cube (with respect to its center). The cube has four axes of symmetry of order 3 (a rotation through $2\pi/3$ about any of the space diagonals takes the cube to itself).

Therefore, the ellipsoid of inertia of the cube has the same four axes of symmetry of order 3.

But an ellipsoid has an axis of symmetry of order 3 only if this is an ellipsoid of revolution (about this axis).

It follows that the ellipsoid of inertia of a homogeneous cube (with respect to its center) has four axes of revolution and hence is a sphere.

Thus, *the moments of inertia of a homogeneous cube with respect to all straight lines passing through its center are the same.*

Remark. Instead of a homogeneous cube, we could take a cube with any system of masses having the same symmetries. For example, we could put eight equal masses at its vertices. Thus, *the sum of the eight squared distances from the eight vertices of a cube to a straight line passing through its center is the same for all such lines.*

In this form, the problem considered above appeared in *Quantum Mechanics* by Landau and Lifshits;[1] instead of moments of inertia, they considered self-oscillation axes of symmetric molecules: for a molecule with the symmetry of a cube, a self-oscillation axis is any straight line (passing through the center).

Pierre Curie believed that his main discovery was the following principle (cited today as "Curie's principle"): The symmetries of consequences reflect the symmetries of causes; thus, observing the former, one should always inquire into the latter (e.g., observing the symmetries of a crystal, look for their causes in the structure of the corresponding molecules).

[1] [**English Translation**] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Vol. 3 of *A Course in Theoretical Physics*), Pergamon Press, 1965.

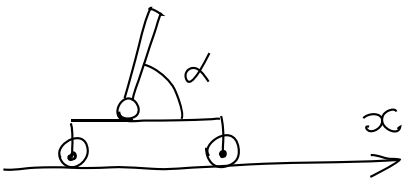
http://dx.doi.org/10.1090/mbk/085/20

Chapter 20

Courant's Erroneous Theorems

A platform stands on horizontal rails, on it the horizontal pivot of an "inverted pendulum" is fixed perpendicularly to the rails. The pendulum can freely swing in the vertical plane parallel to the rails. The platform moves according to a law $x = f(t)$ (where f is a function of time smooth in some interval $[0, T]$).

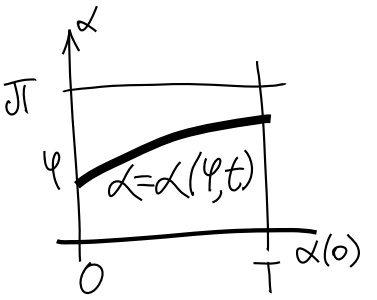
Prove the existence of an initial state of the pendulum ($\alpha(0) = \varphi$, $(d\alpha/dt)(0) = 0$) such that the pendulum never hits the platform during the travel time T.



Solution (Courant's). If $\varphi = 0$, then we always have $\alpha(t) = 0$, and if $\varphi = \pi$, then $\alpha(t) = \pi$ for all t.

Since a solution of a (smooth) differential equation continuously depends on the initial condition φ, it follows by the intermediate value theorem that there is a value $\alpha(0) = \varphi$ between the initial conditions $\alpha(0) = 0$ and $\alpha(0) = \pi$ such that $\alpha(t)$ is strictly between 0 and π for

$0 \leqslant t \leqslant T$, so that the pendulum does not fall down.



Remark. Many people disputed this (incorrect) proof, because, even if a continuous function $\alpha(\cdot, T)$ of the initial position φ were defined, its difference from 0 and π under the initial condition $\cdot = \varphi$ would not imply that the angle α differs from 0 and π at all intermediate moments of time $0 < t < T$.

Probably, there is a reasonable generalization of Courant's argument, in which $\alpha(\varphi, t)$ is naturally extended beyond the moments of hitting the platform (when $\alpha(\varphi, t) = 0$ or π). But such a generalization is missing from the literature, and no rigorous proof of Courant's above-stated conjecture is known.

Various objections to various attempts to substantiate Courant's conclusion were published by J. Littlewood and other mathematicians (some of the "counterexamples" to one of such substantiations turned out to be invalid if the speed df/dt is slower than the speed of light).

But I have not seen a reasonable analysis of this problem with hits taken into account.

Courant included the above theorem in the remarkable elementary textbook *What is Mathematics?* by Courant and Robbins with a reference to Whitney.

Another erroneous theorem was included by Courant in the famous book *Methods of Mathematical Physics* by Courant and Hilbert. This theorem provides a topological description for the linear combinations of the first n eigenfunctions of the Laplace operator: their zeros split an oscillating manifold into no more than n parts.

Courant proved this correctly for the nth eigenfunction, but for its combinations with the preceding eigenfunctions, this is not always true. In the case of a one-dimensional oscillating body (a string), Courant's statement to me is probably correct.

I. M. Gelfand explained the idea of his proof of Courant's statement. It is based on the replacement of Bose's Laplace equation by Fermi's (for n electrons, say on a fixed circle).

Applying Courant's theorem on the number of pure eigenfunctions to the first skew-symmetric eigenfunction of this n-dimensional problem and fixing the positions of all but one electron, Gelfand promised to obtain, on the fixed circle, any linear combination of the first n eigenfunctions of the one-dimensional problem.

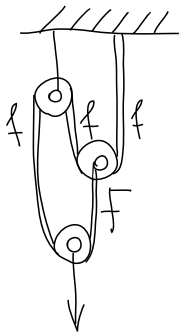
But nobody has published a complete proof so far.

http://dx.doi.org/10.1090/mbk/085/21

Chapter 21

Ill-Posed Problems of Mechanics

Three (weightless) ideal pulleys are joined as shown in the picture. Find the acceleration of the mass suspended from the bottom pulley.



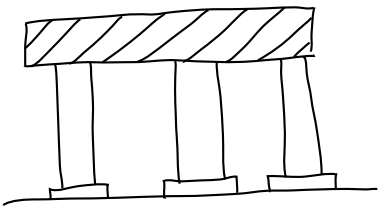
Solution. Let f denote the tension of the part of the rope between the central pulley and the ceiling. Then, the tension F of the part of the rope between the top pulley and the bottom one is $F = 2f$ because the second part of the rope (between the top pulley and the central one) pulls the central pulley upward with force f (since this pulley is ideal).

Similarly, the force with which the rope between the bottom pulley and the top one pulls the former upward is f (because the top pulley is ideal). Finally, since the bottom pulley is ideal, we have $f = F$ (the equality of the forces of tension of both parts of the rope on which this pulley hangs).

Thus, we have obtained two relations between tension forces, $F = 2f$ and $F = f$. They imply $f = F = 0$. This means that *the third pulley is not suspended from anything and falls, together with its load, with free-fall acceleration g.*

Remark. Accounting for the masses of pulleys (and their inertia of rotation) significantly complicates the problem. We do not justify here the fact that the acceleration of the bottom pulley tends to g as the masses of pulleys tend to zero (which is, mathematically speaking, required for substantiating the "physical" solution of the problem given above).

The number of similar "ill-posed" problems in diverse applied areas is enormous, even if I mention only "statically indeterminate" situations, like the distribution of the weight of a beam between three pillars supporting this beam.



There are hundreds of papers presenting "algorithms" for solving such problems, and some of the mathematical theories constructed for this purpose (e.g., in L. Nirenberg's recent papers) are very beautiful. But the question of their practical applicability is quite different.

A. N. Krylov recalled that Volterra gave a rigorous mathematical proof of the stability of a railway bridge crossed by a train of mass M with velocity v, provided that M is not too large.

But he points out that the mass M (calculated in practically interesting examples) turns out to be 10 grams: "the theorem is correct, and so is its proof, but it is entirely beside the point." He was able to calculate realistic limits for M above, which the bridge would collapse, but there was no rigorous proof of stability for smaller values of M.

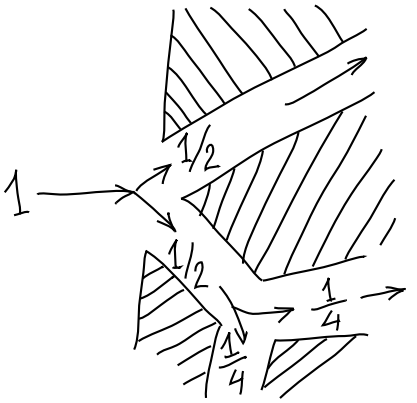
Krylov's student S. Timoshenko designed and calculated many of (the most famous) bridges in the United States, including the reconstruction of the Tacoma Narrows Bridge, which had collapsed because of flutter, but he proceeded from a correct understanding of the essence of the matter rather than from the bounds obtained by Volterra.

http://dx.doi.org/10.1090/mbk/085/22

Chapter 22

Rational Fractions of Flows

A splitter (of a crowd passing through a corridor) sends one person to the left and the next one to the right, so that the flow splits into parts of the same intensity going in different directions:

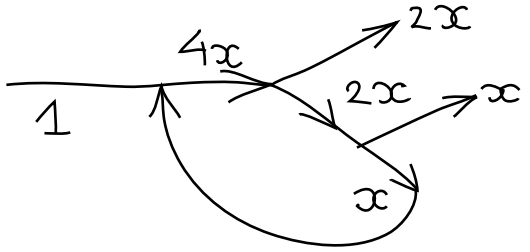


By using several such splitters, it is possible to isolate 1/4 or 3/8 of the flow.

Is it possible to isolate 1/3 of the flow?

Solution. Let us combine two splitters so that the first splits the entire incoming flow and the second, one of the resulting halves of the flow. We send one of the two isolated quarters of the flow back

to the first splitter, so as to include it in the incoming flow:



Suppose that the initial incoming flow in our system was of intensity 1 (say 100 people passing through per minute). Let us denote the intensity of the returned quarter of the flow passing through the first splitter by x.

Then, the entire incoming flow (at the first splitter) has intensity

$$1 + x = 4x;$$

i.e., $x = 1/3$, which solves the problem.

Remark. By using a larger number of standard splitters of flows into two equal parts, it is possible to isolate any rational fraction ($x = p/q$) of the initial flow.

These mathematical theorems were discovered by experts in fire-fighting in search for the optimal use of the subway after a nuclear bombing.

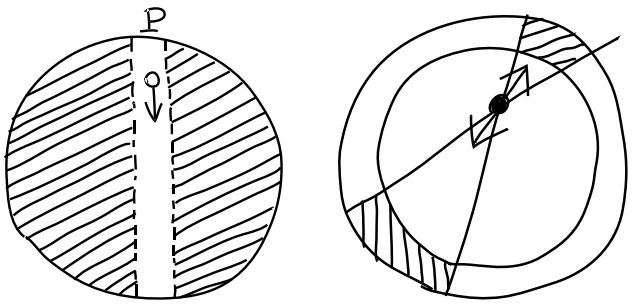
http://dx.doi.org/10.1090/mbk/085/23

Chapter 23

Journey to the Center of the Earth

A stone falls in a well (*without initial velocity*) *passing diametrically through a whole spherical planet.*

Investigate its motion under the action of the gravitational field (*assuming that the planet is homogeneous, i.e., of constant density*).



Solution. According to Newton's theorem, the already passed (homogeneous) spherical layers do not attract the stone, while the layers not yet passed attract it as if their mass were concentrated at the center of the planet.

Let us denote the distance from the stone to the center of the planet by r. Then, the volume (and, hence, the mass M) of the layers not yet passed is proportional to r^3. According to the law of universal gravitation, the force of attraction by such a mass located at the center of the planet decreases with increasing r as $M/r^2 = r$.

Therefore, the motion of the stone in such a well is governed by the force field of Hooke's law:

$$\ddot{r} = -\omega^2 r, \qquad r = R\cos(\omega t).$$

Here, the amplitude R is the radius of the planet.

Thus, the stone performs harmonic oscillations about the center of the planet. It returns to the initial point P after the period $T = 2\pi/\omega$ (and visits the antipodes in the middle of this period).

To avoid the burdensome calculations of the coefficient ω^2 in the equation of Hooke's field, consider a nearby satellite orbiting the planet along the great circle passing through P. The orthogonal projection of the orbit of this satellite on the diameter of the planet oscillates harmonically with amplitude R. At the point P, the gravitational field acting on the stone is the same as that acting on the satellite (because the stone has not yet passed any spherical layers).

Therefore, the oscillation period T of the stone in the well is equal to the period during which the nearby satellite makes a full circular orbit (for planet Earth, this is approximately an hour and a half).

These Newton laws explain the amazing composition of Saturn's rings: the size of the blocks of ice from which they are made is 10 to 20 meters on average.

Naturally, the blocks moving randomly along Kepler's orbits (which are not quite circular) may collide, and the mean collision velocity is calculated from the mean size of a block: it depends on the difference between the speeds of motion along close Keplerian orbits.

The splinters resulting from a collision move faster the higher the collision speed. Calculations show that, for blocks of size more than 20 m, the speed of the splinters is higher than the escape velocity (needed to escape far away from the maternal block) so that such blocks become smaller after collisions.

If the size of a block is less than 10 meters, then the splinters fly away with slower speeds and return back so that at least one of the two colliding blocks grows.

It is this dynamic that leads to the occupation of each ring by blocks that are not too large and not too small (this phenomenon was

discovered, after the calculations described above were performed, during the *Voyager* mission).

http://dx.doi.org/10.1090/mbk/085/24

Chapter 24

Mean Frequency of Explosions (according to Ya. B. Zel'dovich) and de Sitter's World

A process involving explosions is described by (Bernoulli's) evolution law

$$\frac{dx}{dt} = a(t)x^2 + b(t)x + c(t).$$

The example of the equation $\dot{x} = x^2$ *shows that a solution may go away to infinity in finite time; such a solution describes an explosion*:

$$(*) \qquad x(t) = \frac{x(0)}{1 - tx(0)}.$$

But this solution can be extended beyond the moment of explosion (by passing round the pole $t = 1/x(0)$ *in its complex neighborhood)*

Shortly before his death, Ya. B. Zel'dovich stated the following conclusion of his study of the asymptotic behavior of the above equation describing explosion processes (primarily, he had in mind cosmology) in large time.

Suppose that the coefficients (a, b, c) *are periodic smooth functions of time* t*. Then the number* N *of explosions during a long time* T *averaged over time is*

$$\lim_{T\to\infty} \frac{N(T)}{T} = \overline{N}.$$

If this mean number of explosions during the period is rational, then the process is periodic; otherwise, it is not periodic.

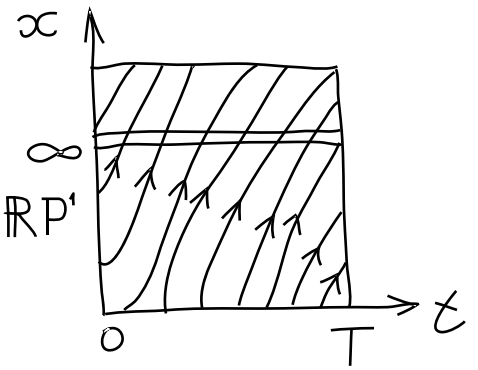
Solution. The appropriate phase space $\{x\}$ is the projective line, that is, the circle $\mathbb{R}\mathrm{P}^1 \approx S^1_x$ (including the point $x = \infty$) rather than the real line $\mathbb{R}$.

For example, relation $(*)$ means that the phase flow transformation (which takes $x(0)$ to $x(t)$) is a projective transformation of the phase space.

The time axis $\{t\}$ must be assumed to be a circle as well; this is the circle $\mathbb{R}/(T\mathbb{Z}) \approx S^1_t$ of phase variation of the coefficients.

This transforms the differential equation of evolution into a smooth direction field on the product torus

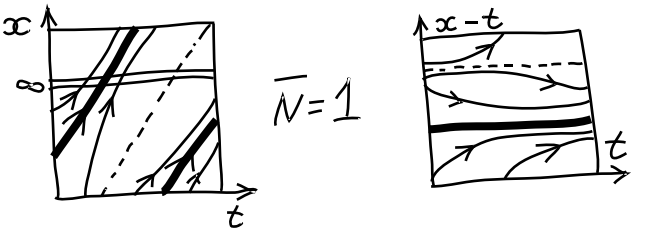
$$T^2 = S^1_t \times S^1_x.$$



Applying Poincaré's theory (see, e.g., *Additional Chapters of the Theory of ODEs*)[1] to this dynamical system on the torus, we see that its Poincaré rotation number is $\overline{N}$.

[1] [**Translator's note**] V. I. Arnold, *Additional Chapters of the Theory of Ordinary Differential Equations* (Nauka, Moscow, 1978) [in Russian]; English translation: *Geometrical Methods in the Theory of Ordinary Differential Equations* (Springer-Verlag, New York–Heidelberg–Berlin, 1983).

When this number is rational, the system has an attractor, which is a closed curve on the torus; this curve describes the periodic evolution of the process.



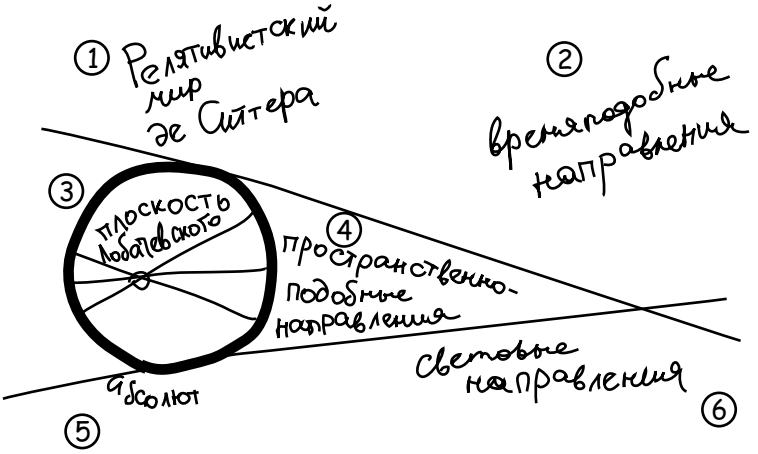
Remark. Ya. B. Zel'dovich solved this problem a week before his death. He did not know Poincaré's theory and simply invented it. But he had no time to publish his theory.

Interestingly, none of the mathematicians had mentioned these applications of Poincaré theory before. The point is that Zel'dovich's theory was based on a daring change of the topological structure of the phase space: he replaced the affine line $\mathbb{R}$ by the projective line $\mathbb{R}\mathrm{P}^1$, which is diffeomorphic to the circle.

Mathematicians avoid such changes of models (except, perhaps, when they pass from the Euclidean to the hyperbolic plane); they prefer to investigate questions already formulated in precise mathematical terms.

On the contrary, in physics, it has been widely believed for a long time that "homology and cohomology are the same old physical fields, only with singularities of a certain form at infinity."

But let me present, incidentally, an example that even physicists are not familiar with: the hyperbolic space whose points at infinity are extended beyond the absolute (of the Cayley–Klein model) is a (relativistic) de Sitter space.



① *De Sitter relativistic universe.* ② *Time-like directions*
③ *Hyperbolic plane.* ④ *Space-like directions.* ⑤ *Absolute*
⑥ *Directions of light*

Each point of the de Sitter space complementing the disk of the Cayley–Klein model to the projective plane is the intersection of two tangents to the absolute bounding the hyperbolic space in the model. These two tangents determine light geodesics of the de Sitter relativistic space: they separate time-like and space-like directions.

Topologically, this de Sitter space (complementary to the hyperbolic plane) is the Möbius strip (which Möbius discovered in precisely this way, as the complement of a neighborhood of a point in the projective plane).

http://dx.doi.org/10.1090/mbk/085/25

Chapter 25

The Bernoulli Fountains of the Nikologorsky Bridge

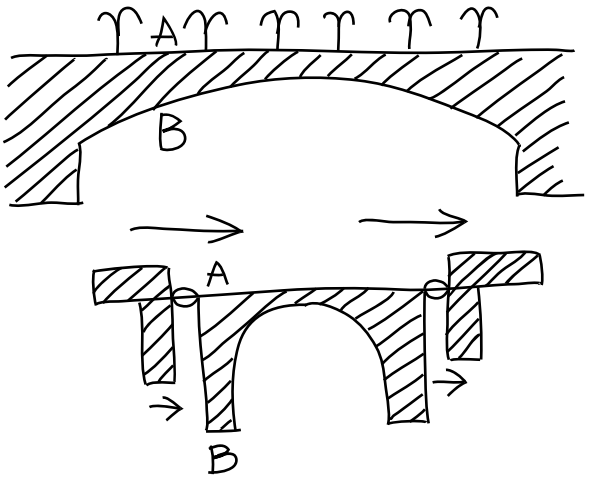
In the bridge near the village Uspenskoe there were (until April 2007*) twelve drainage holes for the outflow of the rain water that may accumulate on the roadway near the sidewalks. In* 1980*, during a powerful thunderstorm, a bicycle rider (V.I.A.) crossing that bridge decided to have a look at one of the holes to observe how water was flowing out. But nothing was flowing out, on the contrary, from each hole a fountain was sprouting up to nearly three meters.*

How can one explain these twelve fountains?

Solution. The Bernoulli law "more speed = less pressure" explains what was going on. A strong wind was blowing along the river above the bridge, but there was practically no wind under it because the bridge itself blocked the moving air. Thus, the wind speed at the top A of the hole was greater than that at its bottom B.

The greater pressure at the bottom created the thrust that caused the fountains to spurt.

Remark 1. The above story was narrated by V.I.A. during a ride from Nikolina Gora to Novodarino as we passed over the bridge in a car driven by N. N. Andreev (who initiated the writing of the present book). A TV interviewer, A. N. Marutyan, was also listening.



The bridge and the view unfolding from it (actually depicted by the Russian painter Levitan) were shown in the resulting film, called *Ostrova* (Islands) on the TV channel *Kul'tura*. There one could see that V.I.A. is explaining something, but the sound at that moment was turned off so that the undesirable references to the unfamiliar Bernoulli law were successfully avoided by the TV people.

In 1980 few people believed in the existence of the observed fountains on the bridge (they are visible only during strong gusts of wind blowing along the river), and now they no longer occur: The holes were filled up when the bridge was repaired in the spring of 2007.

Remark 2. A few minutes after we crossed the bridge, we observed, to the right of the road, another object, the interesting story about which that followed was also excluded from the *Kul'tura* program by the TV people.

Namely, a bit above the place where the road crosses the small river Sleznya, there is a pond in the village Uspenskoe, which at the time belonged to *Sovmin*, the Soviet Council of Ministers. Along its nicely equipped shore, there is a planked walkway where people like to tan in the sun. The story is this: This is the very walkway from which Boris Yeltsin fell in the water (although the legend claims

that "he was thrown into the Moscow river from the high bridge near Uspenskoe").

The height of the above-described bridge is about 10 meters, so the fall would not have been particularly pleasant, whereas there is less than one meter between the walkway and the water of the Sleznya.

Unfortunately, no picture of that walkway appears in the film *Ostrova*. But then Marutyan found (using a map that I had drawn) the cranberry swamp, located some 15 kilometers from the Sleznya (near the village Dmitrovskoe, close to the palace where the Patriarch Alexy resides and the church where Lzhedmitry[1] met Marina on her way from Poland). The ice-age lake in the middle of the swamp, surrounded by shrubs, provides me with several pails of cranberries every year, and in Marutyan's film it looks just like a lake from Karelia.

In that swamp, besides cranberry shrubs, there is an abundance of carnivorous sundew—a kind of swamp grass that feeds on live insects that stick to its leaves. The leaf then rolls up (like a mousetrap) and digests the trapped insect.

Some 40 years ago I could share the pleasure of swimming in that ice-age lake with moose and boars, but now the boars have all been eaten up, while the moose wait till I ride away on my bicycle with my pails of cranberries to take their dip in the lake.

[1] [**Translator's note**] *Lzhedmitry* (False Dmitry) was an usurper (supported by Poland) of the Russian throne in 1605–1606; Marina Mnishek was a Polish noblewoman who became his wife.

http://dx.doi.org/10.1090/mbk/085/26

Chapter 26

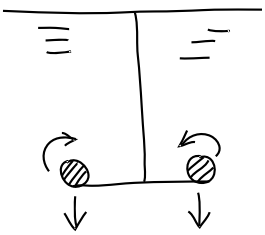
Shape Formation in a Three-Liter Glass Jar

On the surface of a three-liter glass jar filled with water, place a drop of ink (*or India ink*) *avoiding an initial push* (*i.e., by carefully "hanging up" the drop at surface level.*)

How will the drop drown?

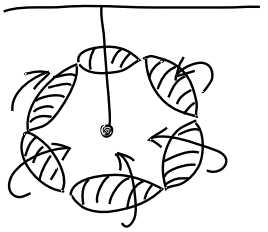
Solution. At first, the "hanging" drop spreads out on the water's surface as a small flat disk, "collects" to its center, and then sinks down a couple of centimeters.

The resistance of the water causes the sinking drop to flatten out until it becomes a colored torus rotating along its meridian. From below, the torus is bounded by a thin film of ink, from the center of which a trace of ink stretches up along the path of the drop.

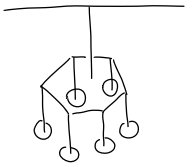


Then the motion of the sinking torus becomes unstable, and it loses its symmetry. Usually the symmetric "doughnut" falls apart into a ring of attached "sausages," each of which is like the initial

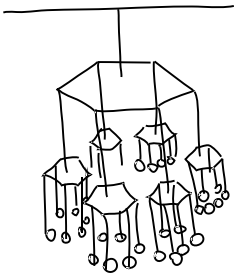
drop (but continues to rotate).



Soon, some of these six "rotating droplets" sink further down just as the initial drop did. A kind of "chandelier," consisting of six hanging tori, is then formed.



The further evolution of these tori is the same as that of the first one: The chandelier now has two "stories."



If all this is done carefully, making sure that the water in the jar has settled and become quite still, not moving your hands near the receptacle so as to avoid inducing the motion of water due to the gradient of temperature, we can then observe, in an ordinary three-liter glass jar, the formation of a six-story chandelier before hundreds of tiny tori reach the bottom.

Apparently, the described picture does not yet have a mathematical proof, but it can be clearly observed in experiments.

For instance, René Thom, who told me about this, read a description of these phenomena in D'Arcy Thompson's ancient book *On Growth and Form* (in which these "chandeliers" appear next to a description of the growth of different corals).[1]

[1] [**Translator's note**] Actually, D'Arcy Thompson mentions medusas (jellyfish), not corals (see p.395 of the 1942 edition of *On Growth and form*).

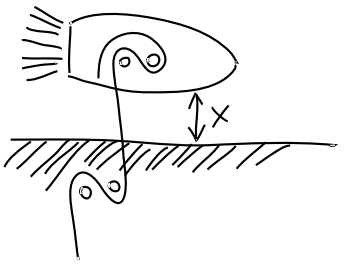
http://dx.doi.org/10.1090/mbk/085/27

Chapter 27

Lidov's Moon Landing Problem

The technique of mooring a ship to a pier is this: a sailor throws a cable on the shore, then jumps off the ship, wraps the cable around a bollard, and hauls in the ship by hand, pulling in a meter or two of cable.

Explain why such a hand-controlled mooring is due to a uniqueness theorem, which works against us here.



Solution. The thing is, the integral curves of the differential equation $dx/dt = -x$ with initial conditions $x(0) = 1$ and $x(0) = 0$ obviously intersect on any computer-produced picture: for $t = 30$ (or even 10), one cannot even fit an atom between the curves.

The usual principles of the control theory of motion require regulating the speed dx/dt of approach to the shore by the feedback loop; i.e., to choose the speed in dependence on the remaining distance, $dx/dt = f(x)$.

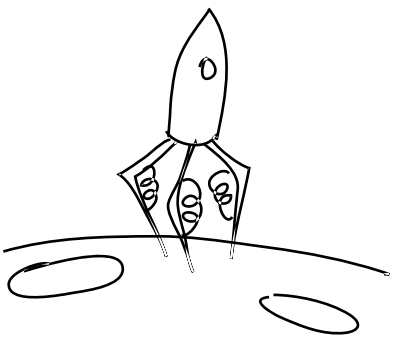
Having this in mind and assuming that the function f is smooth (or at least Lipschitz), we see that the uniqueness theorem implies that the time needed for mooring is infinite.

Or, one must maintain a nonzero speed until the ship reaches the pier and bumps into it (for this reason automobile tires hanging on the edge of the pier are needed, even in the case when the mooring is hand-controlled).

Remark. My good friend M. L. Lidov was a leading expert in ballistics, calculated the orbits of artificial space bodies, sputniks, lunar expeditions and so on.

Once, around 1960, he told me: "The uniqueness theorem in your ordinary differential equations course is completely wrong, although it has a perfectly rigorous proof" ("which", he added, "I don't doubt"). As a confirmation, he communicated the following problem to me.

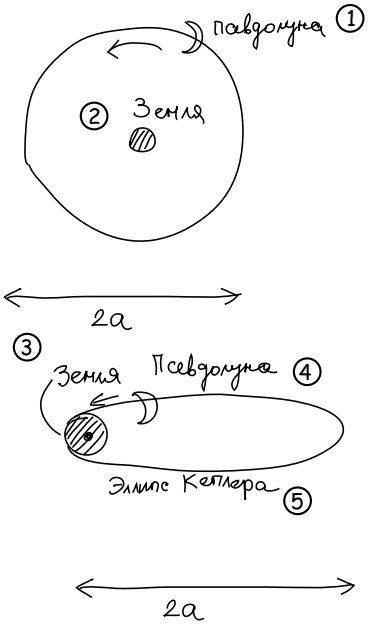
Lidov knew all about the mooring of ships because he had to land spaceships on the Moon. The controlled soft landing there also contradicts the uniqueness theorem. The chosen practical method consists in damping the final collision by brief oscillations about the knees of the three "legs" of the rocket.



A number of remarkable achievements in space ballistics are due to Lidov. For instance, he studied the evolution of "pseudomoons," Earth satellites with orbit the size of that of the Moon, under the condition that the inclination of the orbit to the plane of the ecliptic (in which Earth rotates around the Sun) is not small (unlike that of

the Moon, for which it is approximately five degrees), but, on the contrary, large.

When the inclination is 80°, Lidov's analysis led to the conclusion that such a pseudomoon would fall on Earth in some four years or so (because of the rapid growth of the orbit's eccentricity due to the perturbing influence of the Sun).



①, ④ *Pseudomoon.* ②, ③ *Earth.* ⑤ *Keplerian ellipse.*

This surprising result of Lidov does not contradict Laplace's theorem about the invariance of the mean distance a of the evolving trajectory of a perturbed (here, by the Sun) Keplerian ellipse to the attracting center (here, Earth).

Even when the pseudomoon falls on Earth, its *mean* distance from Earth remains the same ($a \approx 380000$ km). But the eccentricity of its Keplerian ellipse increases in four years to such an extent that the ellipse begins to intersect the earth (which is not a material point, but has a radius of nearly 6400 km).

http://dx.doi.org/10.1090/mbk/085/28

Chapter 28

The Advance and Retreat of Glaciers

According to Lagrange's theory of secular perturbations of planetary orbits in the planar problem of n planets, the Laplace vector z, joining the Sun with the center of the instantaneous Keplerian ellipse of each planet, is the sum of n uniformly rotating (with angular velocity ω_k) vectors of fixed lengths a_k, $(r = 1, \ldots, n)$ in the plane of motion:

$$z = \sum_{k=1}^{n} a_k e^{i\omega_k t}$$

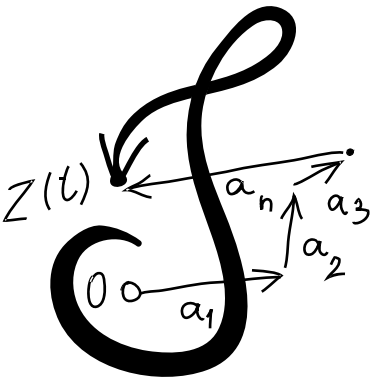
(*expressed in terms of the complex coordinate z in the plane of motion*).

The length of the eccentricity, proportional to $|z|$, and the direction to the perigee, $\varphi(t) = \arg z$, vary in a complicated way over time.

Lagrange assumed that there exists a mean value of its motion,

$$\Omega = \lim_{t\to\infty} \frac{\varphi(t)}{t},$$

and proposed to calculate the mean value of the angular velocity of the perigee motion.

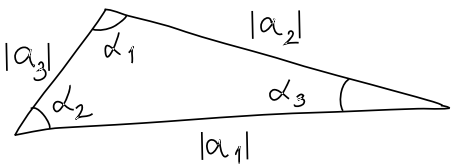


The evolution of the Laplace vector $z(t)$ is one of the causes of the advance of glaciers because, when the vector is large, the Keplerian ellipse has a large eccentricity. In that case, the planet spends more time far from the Sun, and its average temperature drops.

For example, the amount of energy provided by solar rays at the latitude of Saint Petersburg varies (for a dozen centuries or so) between the amounts provided on the latitude of Taimyr[1] and the latitude of Kiev.

Remark. If one of the summed vectors is longer than the sum of the others (i.e., $|a_j| > \sum_{k\neq j} |a_k|$), then the mean angular velocity of the perigee will be the angular velocity of that summand, a_j (Lagrange).

In our real solar system, this turns out to be precisely the case for most planets. But for Earth and Venus, there are several summands of approximately the same length, so that the problem already becomes meaningful for three planets, where the numbers $|a_1|, |a_2|, |a_3|$ are the lengths of the sides of a triangle:



[1] [**Translator's note**] The *Taimyr* peninsula is located at 75° North latitude.

In that case, the problem was solved by H. Weyl, and the answer has the form of a weighted arithmetic mean

$$(*) \qquad \Omega = p_1\omega_1 + p_2\omega_2 + p_3\omega_3$$

with weights proportional to the angles of the above-mentioned triangle:

$$p_j = \alpha_j/\pi.$$

This value of the mean angular velocity (which does not depend on the initial positions of the rotating vectors) is obtained for almost all (in the sense of Lebesgue measure) values of the angular velocities ω_k (their arithmetic independence suffices; i.e., the absence of resonances

$$m_1\omega_1 + m_2\omega_2 + m_3\omega_3 = 0$$

with integer coefficients $m \neq 0$).

In the presence of resonances, the answer may depend on the initial conditions, but remains the same (if one averages over initial positions of the rotating vectors).

When there are more than three not too long summands (i.e., if $|a_j| < \sum_{k\neq j} |a_k|$), the answer is also given by an arithmetic mean of angular velocities similar to $(*)$, but the role of the angles of the triangle is played by the following "generalized angles" p_k.

Consider the $(n-1)$-dimensional torus T^{n-1} with angular coordinates φ_k, $k \neq j$. Let us construct the vector

$$\xi(\varphi) = \sum_{k\neq j} a_k e^{i\varphi_k}$$

depending on the point φ of the torus.

For some j-dangerous points φ, the length of this vector is less than the number $|a_j|$: $|\xi(\varphi)| < |a_j|$.

The weight p_j is the measure of the set of j-dangerous points φ (scaled so that the measure of the entire torus is 1).

In the case $n = 3$, this definition leads to the weights $h_j = \alpha_j/\pi$, as one can easily calculate.

The assertion that the sum of weights $p_1 + \cdots + p_n$ equals 1 for any n is true, but is not immediately obvious (although it suffices to

take independent angular velocities ω_j close to ω to prove the above identity for the described weights p_j).

A detailed solution on the mean motion of the perigees is described by H. Weyl in two long articles, and below we only sketch the idea. For simplicity, we consider the case $n = 3$.

Consider the three-dimensional torus T^3 with angular coordinates φ_1, φ_2, φ_3 and on it, the vector field of the rotations with angular velocities ω_1, ω_2, ω_3:

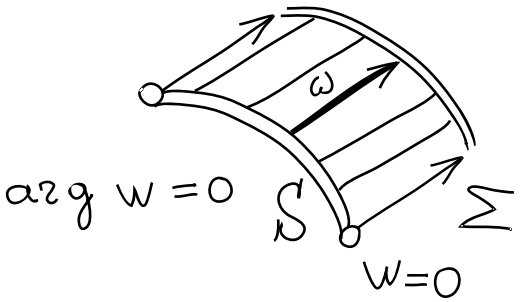
$$\dot{\varphi}_1 = \omega_1, \qquad \dot{\varphi}_2 = \omega_2, \qquad \dot{\varphi}_3 = \omega_3.$$

To each point of the torus let us assign the complex number

$$w(\varphi) = a_1 e^{i\varphi_1} + a_2 e^{i\varphi_2} + a_3 e^{i\varphi_3}.$$

The argument of this complex number is well defined (up to an integer multiple of 2π) whenever $\omega(\varphi) \neq 0$.

The points φ of the torus where $\omega(\varphi) = 0$ form a closed curve on it, bounding a surface S on which the complex number ω is positive (and where $\arg \omega(\varphi) = 0$).



This surface is cooriented (by the direction of increase of the argument). To each point of the surface S, let us attach the vector of angular velocities

$$\omega = \omega_1 \frac{\partial}{\partial \varphi_1} + \omega_2 \frac{\partial}{\partial \varphi_2} + \omega_3 \frac{\partial}{\partial \varphi_3}.$$

All these attached vectors form a 3-chain Σ on the torus T^3 (possibly with self-intersections and non smooth). On the torus T^3, define a function f whose value at each point φ is equal to the number of

times the chain passes through that point (in the picture, $f = 0$ everywhere except on the hashed domain, where $f = 1$).

The mean value Ω is now (up to an integer multiple of 2π) the mean in time of the function f constructed above along the orbits of the dynamical system $\varphi(t) = \varphi(0) + \omega t$.

The independence of the frequencies implies that this (Lebesgue measure preserving) dynamics is ergodic, and so the temporal mean coincides with the mean in space.

The spatial mean depends of the frequency vector ω that was used to construct the chain and the averaged function f on the torus.

But the dependence on the vector ω is linear (for example, this follows from the representation of the integral of f in the form of the flow of the field ω through the surface S).

Therefore, in order to compute the spatial mean, it suffices to calculate its value on three basis vectors $\omega = \partial/\partial\varphi_k$ (no longer paying attention to the ergodicity of the flow corresponding to the field).

This last computation is easy; for instance, we can use the fact that if only one summand is rotating, then the mean rotation of the sum is 0 provided that this summand is shorter than the sum of the other two.

Now, if the rotating summand A_3 is longer than that sum, then the mean rotation coincides with the angular velocity ω_3 of the rotating summand.

Hence, the mean (over all directions of the motionless summands A_1 and A_2) coincides with that fraction of the directions φ_1 and φ_2 for which $|A_1| + |A_2| < |A_3|$), and that fraction is α/π, i.e., the ratio of the angle opposite to the side $|A_3| = |a_3|$ of the triangle with sides $|a_1|$, $|a_2|$, and $|a_3|$ to the sum of angles of the triangle.

A remark about ergodic theory

The computation sketched above is carried out by H. Weyl using ideas of "ergodic theory," which allow reducing the computation to finding spatial means of an appropriate function in the phase space T^3 of the dynamical system $\varphi \mapsto \varphi + \omega t$ on the torus.

Spatial means are often easier to compute than temporal means. All of statistical physics is based on that idea. The coincidence of temporal means (along the orbits of chaotic dynamical systems) with spatial means (over the entire phase space) is regarded by physicists as self-evident, and they have always applied this "ergodic" conjecture (since the time of L. Boltzmann).

But mathematicians know that this coincidence of different types of means does not always take place: the system must be "ergodic," and even ergodicity does not guarantee the coincidence of the two means for all initial values of the dynamics (although it ensures it for almost all the initial conditions).

In the particular case of uniform motion $\varphi \mapsto \varphi + \omega t$ on the torus T^n (with independent frequencies ω_k), H Weyl proved the coincidence of the temporal and spatial means for any continuous (or at least Riemann integrable) function $f : T^n \to \mathbb{R}$ on the torus of volume 1:

$$\lim_{T\to\infty}\left(\frac{1}{T}\int_0^T f(\varphi+\omega t)\,dt\right) = \int\cdots\int_{T^n} f(\varphi)\,d\varphi.$$

Example. Let f be the characteristic function of a (Riemann measurable) domain X on the torus

$$\begin{cases} f(\varphi) = 1 & \text{for } \varphi \in X, \\ f(\varphi) = 0 & \text{for } \varphi \in T^n \setminus X. \end{cases}$$

Then the integral on the left-hand side expresses the time during which the segment $0 \le t \le T$ of the orbit of the dynamical system $\varphi \mapsto \varphi + \omega t$ spends in the domain X.

The limit on the left-hand side is therefore the fraction of time that the orbit spends in X during the whole infinite period $t \ge 0$.

As to the integral on the right-hand side, it is equal, for the characteristic function of the domain X, to the volume of this domain. If the whole phase space (T^n in our case) is of volume 1, then this integral also expresses the fraction of the volume of the phase space occupied by X.

The coincidence of the temporal means of the characteristic function (appearing on the left-hand side) with its spatial mean (appearing on the right-hand side) is known as the equidistribution of the orbits of the dynamical system under study in phase space. If the dynamics of the system possesses this property of uniform distribution (for any Riemann measurable domain X), then the time spent by the mobile point in various parts of the phase space is proportional to their volume.

The uniform distribution of the orbits of the dynamical system in phase space implies the coincidence of the temporal and spatial means of all (at least Riemann measurable) functions. This follows from the fact that such a function can be approximated by a linear combination of characteristic functions of several domains.

Together with the proof of the uniform distribution of the dynamical system $\varphi \mapsto \varphi + \omega t$, H. Weyl obtained a similar theorem on the uniform distribution for dynamical systems with discrete time

$$g\colon T^n \to T^n, \qquad g(\varphi) = \varphi + \lambda.$$

The orbits of the dynamics are uniformly distributed on the torus T^n, if the rotation vector λ has the following independence property of its components: A linear combination with integer coefficients

$$m_1\lambda_1 + \ldots + m_n\lambda_n + m_0$$

vanishes if and only if the vector $m \in \mathbb{Z}^n$ is zero.

Example. For $n = 1$, the transformation g is a rotation of the circle, while the independence condition is that the rotation angle must be incommensurable with the flat angle (i.e., with 2π).

http://dx.doi.org/10.1090/mbk/085/29

Chapter 29

The Ergodic Theory of Geometric Progressions

Consider the first digits of the terms of the geometric progression 2^t, $t = 0, 1, 2, \ldots$

$$1,\ 2,\ 4,\ 8,\ 1,\ 3,\ 6,\ 1,\ 2,\ 5,\ 1,\ 2,\ \ldots$$

What fraction of the first digits of the numbers constituting this sequence are ones? In the first 10 terms there are three, and we must find the limit $p_1 = \lim_{T\to\infty}$ (*number of terms among the first* T *terms* 2^t *of the progression that begin with the digit 1*).

Solution. Consider the logarithms to the base 10 of the terms of the progression: $\lg 2^t = t\lambda$, where $\lambda = \lg 2$.

The first digit of a positive number z equals k if the number z lies in the interval $k10^a \leqslant z < (k+1)10^a$, where a is an integer. In other words, $a + \lg k \leqslant \lg z < a + \lg(k+1)$; i.e., the fractional part of the number $\lg z$ lies in the half interval of length

$$p_k = \lg(k+1) - \lg k = \lg(1 + 1/k).$$

The number $\lambda = \lg 2$ is irrational, (otherwise we would have $10^{p/q} = 2$; i.e., $10^p = 2^q$, which is impossible for positive p since 2^q is not divisible by 5).

By H. Weyl's theorem, the sequence $\{t\lambda\}$, $t = 0, 1, 2, \ldots$ of points on the circle of fractional parts $\mathbb{R}/\mathbb{Z}$ is uniformly distributed. The fraction of time that the orbit spends in the interval $[0, \lg 2)$ of the

circle of fractional parts is equal to the length $p_1 = \lg 2 \approx 0.30$ of this interval, because the distribution of orbits is uniform.

And so approximately 30% of the terms 2^t of our geometric progression begin with the digit 1.

Remark. The same argument with spatial means yields the fraction of twos, p_2, the fraction of threes, p_3, etc.:

$$p_k = \lg(k+1) - \lg k = \lg\left(1 + \frac{1}{k}\right):$$

k	1	2	3	4	5	6	7	8	9
$100p_k$	30	18	12	10	8	7	6	5	4

Instead of the geometric progression with common ratio 2, we can take other geometric progressions, a^t (say, taking $a = 3$), the fraction p_k of numbers beginning with k will still be the same as for the progression 2^t. It is only important that the common ratio a must not equal any rational power of 10: The shift $\lambda = \lg a$ in the dynamical system on the circle must be an irrational number in order to ensure the uniformity of the distribution of orbits.

Remark. In the U.S., Weyl's theorem is usually called "Bedford's Law" in honor of the physicist who noticed (around 1930) that the first pages of logarithm tables in libraries are dirtier than the last ones. He explained this by saying that "random numbers" begin by "1" more often than by other digits, so that one needs to find logarithms of numbers from the first pages more often.

But Bedford was wrong: for example, in the statistics of the lengths of rivers or altitudes of mountains, just as many numbers begin with a 1 as with a 9.

Here the usual "eponimical principle" works: No discovery bears the name of its first discoverer, everything is ascribed to friends of those who give the name (for example, America is not called Columbia).

The British physicist M. Berry called this eponymic principle "Arnold's Principle," adding a second one to it: "Berry's Principle."

Berry's eponymic principle asserts that "Arnold's Principle is applicable to itself" (i.e., it wasn't Arnold who invented it).

http://dx.doi.org/10.1090/mbk/085/30

Chapter 30

The Malthusian Partitioning of the World

Let us consider all the countries of the world and count in how many of them the first digit of their population is k.

Justify that the fraction of such countries is $p_k = \lg(1 + 1/k)$, just as for the geometric progression 2^t and for the table appearing on p. 100.

Solution. According to Malthus' law, the numbers expressing the population of a country in subsequent years form a geometric progression. Therefore the fraction of these numbers beginning with k is p_k.

According to the ergodic principle, the temporal means (averaging the situation over the years in each country) is equal to the spatial means (averaging the situation for the present year over all the countries of the world).

Remark. Replace the numbers expressing the population of countries by some other sequences, say the altitudes of mountains or the lengths of rivers, or the number of pages of the books on your favorite bookshelf.

In those cases all digits appear in the first position with approximately the same frequency $p_1 = p_2 = \cdots = p_9 = 1/9$ (whereas the

numbers for the population whose first digit is a 1 appear 7–8 times more often than those beginning with a 9).

The thing is that neither rivers, nor mountains, nor books grow in geometric progression, while the distribution $p_k = \lg(1 + 1/k)$ is characteristic precisely for geometric progressions.

It is amazing that the areas of countries (be they measured in square kilometers, or square miles, or square inches) also produce the same distribution of first digits as geometric progressions.

This phenomenon may be explained by the fact that countries unite from time to time (which leads to an increasing geometric progression with common ratio 2 when countries of similar size merge), and split in half from time to time (which creates a decreasing geometric progression, for which the distribution of the fractional parts of the logarithms is also uniform).

For the simplest models of such partitions of the world, the appearance of the distribution of first digits described above can be proved, but computer experiments[1] show that they also occur in more complicated models (e.g., when a country can merge only with neighbors), although no one has proved theorems justifying the appearance of the distribution $p_k = \lg(1 + 1/k)$ in such cases.

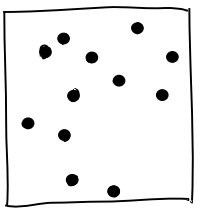
[1] Carried out by F. Aicardi in Sistiana (Italy) and M. Khesina in Toronto (Canada).

http://dx.doi.org/10.1090/mbk/085/31

Chapter 31

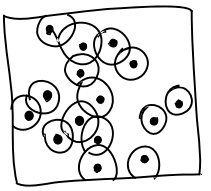
Percolation and the Hydrodynamics of the Universe

Consider N points in some domain of Euclidean space (for example, in the unit cube $I^n \subset \mathbb{R}^n$, say the square in the Euclidean plane).



If r is small enough, then the balls of radius r centered at these points don't intersect.

If the radius is larger, not only do some of the balls intersect, but certain intersecting balls form chains of order 1, along which one can move from one side of the cube to the opposite side.



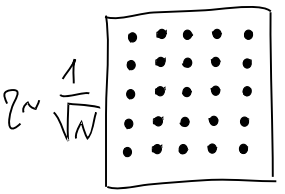
In that situation, we say that percolation has occurred: if the given domain is filled with the substance of the receptacle having N sources of faults, and each fault has grown to the size of a spherical hole of radius r, then, as percolation occurs, the receptacle begins to leak.

The percolation radius of a system of points is the least radius r of balls centered at these points for which percolation occurs.

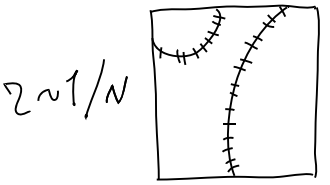
The percolation radius depends not only on the number of points but also on the geometry of their positions.

The problem that will now be discussed is how the percolation radius r decreases as the number of points grows for different positions of the percolation centers in the substance of the receptacle.

For a filling of the cube by N points of a regular lattice, the distance between neighboring points will be of the order of $1/\sqrt[3]{N}$, so that the percolation radius is of the order of $N^{-1/3}$.



This conclusion remains true for less regular positions of the points, even for those randomly thrown into the cube: the percolation radius of a system of N points in I^n decreases as $N \to \infty$, as a rule, like $C/N^{1/n}$.



Now if the n percolation centers are not chaotically positioned in the cube I^3, but, say, lie along a smooth curve, then the percolation radius will be much smaller, namely C/N.

When the n percolation centers are positioned on a smooth surface embedded in the cube I^3, the percolation radius will decrease as $N \to \infty$ at an intermediate level between the two previously described cases: $r \sim C/N^{1/2}$ (for N centers positioned on a k-dimensional submanifold in I^k, the same argument yields a percolation radius of the order of $1/N^{1/k}$).

Rigorous mathematical proofs of all the results listed above are not easy, mainly because one must define what a "random filling" of a submanifold by N centers of percolation is, and what submanifolds are admissible.

But physicists, experts in chemistry, and astronomers bravely use such a "stochastic geometry" without much care about rigorous justification—and obtain spectacular nontrivial results.

For example, in cosmology, it is important to understand how the galaxies are distributed in the Universe: do they tend to position themselves along some surfaces or lines, or do galaxies accumulate near separate points, or are they uniformly distributed everywhere, like the N points randomly thrown into the cube in the previous example?

The answers to these questions on the accumulation of galaxies may shed light on the extremely difficult problems of their origin.

The first peculiarity of the distribution of galaxies observed by astronomers was the presence of huge empty places between them, holes where no galaxies appear.

These holes led to the idea that for some reason galaxies, instead of placing themselves randomly, prefer being situated along certain special two-dimensional surfaces or one-dimensional curves (which can intersect, forming networks).

Astronomers and cosmologists have computed the percolation radius of the system of thousands of observed galaxies. The dimension of the manifold along which they accumulate was obtained by comparing the percolation radius with the number N of observed galaxies. The percolation radius turned out to be of the order of magnitude of C/N^{α}, where $1/2 < \alpha < 1$.

This means that the manifold is "of dimension one-and-a-half": apparently it is a not very smooth surface (dim = 2) near which the density of the galaxies is higher than in the complementary "empty" domains; however, on that surface there are lines (dim = 1), where the density is greater than on the surface (an additional increase of density near singular points (dim = 0) of these lines is not to be excluded).

All these consequences of the value of the computed percolation radius have been borne out by a detailed analysis of the spatial distribution of galaxies (and that of the "hydrodynamical Universe," which explains the origin of these density singularities by nonuniformities of the velocity field in parts of the Universe after the "Big Bang").

The advantage of the mathematical approach based on the percolation radius over the direct viewing of the observed spatial distribution is in that a human being tends to unite objects, accidentally close to each other, into more convenient structures (for instance, by dividing the starry sky into subjectively chosen constellations: in China the seven stars of the Big Dipper had been split long ago into two constellations—the Horse and the Carriage).

The percolation approach replaces these subjectively determined structures by objective characteristics of the studied objects, not depending on the investigator's bias.

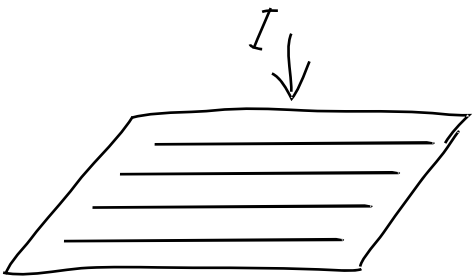
http://dx.doi.org/10.1090/mbk/085/32

Chapter 32

Buffon's Problem and Integral Geometry

Let us randomly throw a needle of length 1 *on a horizontal piece of paper lined by parallel lines so that the distance between neighboring lines is* 1.

Repeat this experiment many times ($N \to \infty$). *How will the number* M *of thrown needles that intersect one of the lines grow with* N*?*



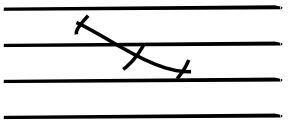
Solution. The answer is surprising:

$$\lim_{N\to\infty} \frac{M}{N} = \frac{2}{\pi},$$

so that, having thrown the needle a million times, we can obtain a fairly good approximation to the number π.

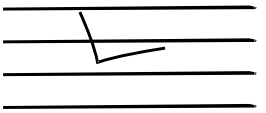
The explanation to this surprising answer is as follows. Clearly, as $N \to \infty$, the number of intersections $M(N)$ is cN (for a certain

constant c equal to the probability of falling on a line in one throw).



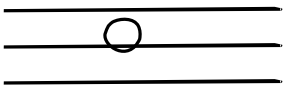
Let us replace the needle of length 1 by one twice as long. Then the probability of intersection will also double (on the average) because the additional half of the needle is of length 1 and also falls randomly so that it will produce as many intersections as did the original needle of length 1.

There is no need for the needle of length 2 to be straight. We can bend it at the midpoint in the form of a poker, both halves will give the same number of intersections, and together, twice as many as before.



It follows from the above arguments that, throwing any "crooked needle" of length l, we obtain asymptotically, as $N \to \infty$, the value cNl for the number of intersections.

In particular, we can throw a circle of diameter 1. The length of this circle is π. Asymptotically, it will produce $cM\pi$ intersection points after N throws.



But such a circle produces two intersection points, no matter how it is thrown.

Thus we have shown that

$$cN\pi = 2N;$$

i.e., $c = 2/pi$, as claimed.

The Buffon problem described here gave rise to a whole new branch of mathematics—so-called integral geometry.

The source of this science was not the study of consequences of some axioms, but the desire to understand simple experiments, invented, one might add, by researchers distant from mathematics.

Today, integral geometry is one of the most active branches of theoretical mathematics. But it is constantly applied to other branches of science, for instance, to study the complicated geometric structures of crystals, plants, or animals on the basis of the statistics of their two-dimensional cross-sections or tomograms (including the study of random projections, shadows produced by randomly placed sources of light, or the reflections of a light beam randomly falling on the object under study).

http://dx.doi.org/10.1090/mbk/085/33

Chapter 33

Average Projected Area

Find the area of the orthogonal projection of a cube with edge length 1 *on a random plane.*

Solution. Arguing as in Buffon's problem, we come to the conclusion that this mean area of projection does not depend on the shape of the (convex) projected body, but depends only on the area of its surface.

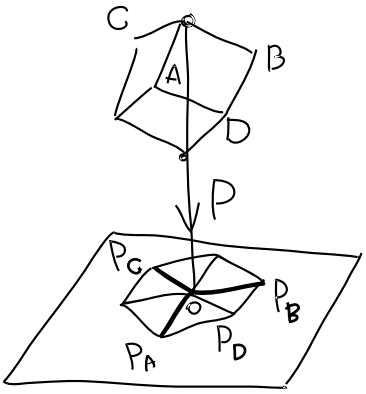
Therefore, the average projected area of a cube is as many times smaller than its surface area as the area of the equatorial section of a ball is smaller than the area of its surface.

For a ball of radius 1, the area of its equatorial section (which is a disk of radius 1) is π. The surface of a ball of radius 1 has area 4π.

Therefore, the average projected area of a cube is four times smaller than its surface area, which is 6. Thus, the average projected area of the unit cube is 3/2.

Remark. The projection area is minimal (equal to 1) when the projection is along an edge. The (hexagonal) projections along the

diagonals of the cube have maximum area.



The diagonals of the cube's faces perpendicular to the projection direction retain their lengths under projection: $|P_AP_B| = |AB| = \sqrt{2}$.

Considering the equilateral triangle OP_AP_D, we obtain $|OP_D| = 2(|P_AP_B|/2)/\sqrt{3} = \sqrt{2/3}$ (because $\tan 60^\circ = \sqrt{3}$).

Therefore, the area of this triangle is

$$\frac{1}{2} \cdot \frac{\sqrt{2}}{2} \cdot \sqrt{\frac{2}{3}} = \frac{1}{2\sqrt{3}}.$$

The area of the entire projection P of the cube is equal to $6/(2\sqrt{3}) = \sqrt{3}$.

Thus, the average projected area $1\frac{1}{2}$ that we have found is contained between the minimum projected area (equal to 1) and the maximum projected area (equal to $\sqrt{3}$).

This confirms the answer found above in the essay *Buffon's Problem and Integral Geometry*. In physics, such test comparisons of means with extreme cases are a necessary element of any study, and mathematicians, too, should remember to carry them out.

Given a smooth boundary of a surface domain in Euclidean space $\mathbb{R}^n$, consider the k-dimensional volume S_k of its orthogonal projection on a random k-plane.

It turns out that this average (over all k-planes, which are considered equiprobable) value exists; e.g., for any surface in $\mathbb{R}^3$, its average projected area and average projected length are defined.

These average k-volumes turn out to be equal to the mean values of symmetric functions of principal curvatures of the surface averaged over the entire surface.

They also participate in the (surprising) expression for the volume of an h-neighborhood of a surface:

$$V(h) = V_0 + V_1 h + V_2 h^2 + \ldots + V_n h^n$$

(here V_0 is the volume of the surface, V_1 is the $(n-1)$-volume of its boundary, which is proportional to the mean value at 1, and V_k is proportional to S_k and can be expressed in terms of the mean value of the product of the k principal curvatures).

In the case of $n = 3$; i.e., a two-dimensional smooth surface in three-dimensional Euclidean space, from the principal curvatures k_1 and k_2 at each point we can produce the *mean curvature* $k_1 + k_2$ and the *Gaussian curvature* $K = k_1 k_2$.

In this case, the volume of an h-neighborhood is

$$V(h) = V_0 + hS + h^2 V_2 + h^3 V_3,$$

where V_2 is proportional to the integral of the mean curvature over the entire surface and V_3 is proportional to the integral of the Gaussian curvature:

$$V_3 = \frac{4}{3}\pi\Big(\iint K\, dS\Big).$$

Thus, for a sphere of radius R, we have

$$V(h) = \frac{4}{3}\pi(R+h)^3 = \frac{4}{3}\pi R^3 + h(4\pi R^2) + h^2(4\pi R) + \frac{4}{3}\pi h^3.$$

Here

$$k_1 = k_2 = 1/R, \quad k_1 + k_2 = 2/R, \quad k_1 k_2 = 1/R^2,$$

$$\iint (k_1 + k_2)\, dS = 8\pi R,$$

$$\iint (k_1 k_2)\, dS = 4\pi \quad \text{(this is the Gauss–Bonnet formula).}$$

The coefficient V_3 does not depend on details of the surface; it depends only on the Euler characteristic. The discovery of this fact led Hermann Weyl to the theory of characteristic classes and characteristic numbers, which generalize the Gauss–Bonnet formula, in his work *On the Volume of Tubes*[1] dealing with $V(h)$.

[1] [**Translator's note**] H. Weyl, "On the volume of tubes," Amer. J. of Math. **61** (2), 461–472(1939).

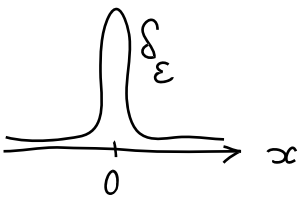
http://dx.doi.org/10.1090/mbk/085/34

Chapter 34

The Mathematical Notion of Potential

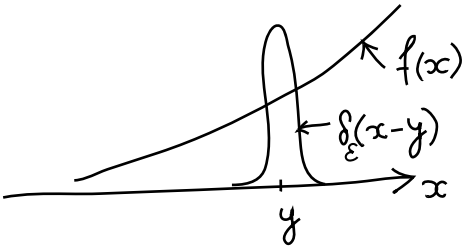
The mathematical model of physical "material points" and "point charges" is known as the *δ-function*.

Physicists say that $\delta(x) = 0$ for any $x \neq 0$, while $\int_{-\infty}^{\infty} \delta(x)\,dx = 1$. Certainly, there exist no such functions in mathematics. Mathematically, they are understood as follows: if a formula contains a δ-function, then we must render it meaningful by replacing the δ-function by its "smoothed version" $\delta_\varepsilon(x)$, where δ_ε is a smooth non-negative function vanishing everywhere outside the ε-neighborhood of the point 0 and having integral 1, and then pass to the limit as $\varepsilon \to 0$.



Example. Let f be a continuous function on the real line. Evaluate $\int_{-\infty}^{\infty} f(x)\delta(x-y)\,dx$.

Solution. The function $\delta_\varepsilon(x-y)$ of x is a smoothed δ-function translated from 0 to y.



The product $f(x)\delta_\varepsilon(x-y)$ vanishes when $|x-y| \geqslant \varepsilon$, so that it suffices to take the integral over the ε-neighborhood of the point y on the x-axis. But the function f differs little from the number $f(y)$ in this neighborhood. Therefore,

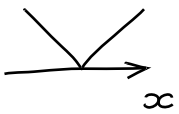
$$\lim_{\varepsilon\to 0}\int_{-\infty}^{\infty} f(x)\delta_\varepsilon(x-y)\,dx = f(y)\int_{-\infty}^{\infty}\delta_\varepsilon(x-y)\,dx = f(y).$$

Thus, we have proved the equality

$$f(y) = \int_{-\infty}^{\infty} f(x)\delta(x-y)\,dx.$$

Remark. If we were physicists (and identified integrals with sums), we might read this equality as "any (continuous) function f of an argument y is a 'linear combination' of δ-functions of y translated to all points x of the y-axis (the translated function of y is $\delta(x-y)$). The coefficients in this linear combination are the values $f(x)$ at all points of the function being expanded.

Problem. *Calculate the second derivative of the function $|x|$ of x.*



Solution. The first derivative $\operatorname{sgn} x$ takes the value 1 at $x > 0$ and -1 at $x < 0$. But this is the integral of the second derivative.

Therefore, the second derivative vanishes at $x \neq 0$, and its integral (from any $a < 0$ to any $b > 0$) is equal to 2 (the increment of $\operatorname{sgn} x$). Therefore,

$$\frac{d^2|x|}{dx^2} = 2\delta(x).$$

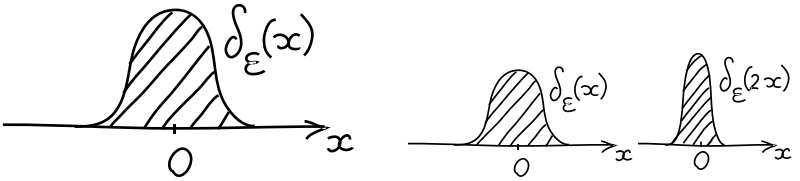
Problem. *Is the δ-function homogeneous?*

A function f is homogeneous of degree k if

$$f(cx) = c^k f(x) \quad \text{for any } c > 0.$$

Example. The function $1/x$ is homogeneous of degree $k = -1$.

Solution of the problem. Consider an approximation $\delta_\varepsilon(2x)$ of $\delta(2x)$. The graph of this "δ-shaped" function of x is the graph of the function δ_ε (of x) compressed by a factor of 2:



We see that $\delta(2x)$ vanishes at any $x \neq 0$, and the integral of this function (over the entire real line) is half the integral of the δ-function (the latter integral is equal to 1).

Therefore,

$$\delta(2x) = \frac{1}{2}\delta(x), \qquad \delta(cx) = \frac{1}{c}\delta(x),$$

so that the δ-function is homogeneous of degree -1.

Problem. *Is the δ-function of n variables (which vanishes everywhere on $\mathbb{R}^n$ except at 0 and has integral 1) homogeneous?*

Answer. This function is homogeneous of degree $-n$.

We can prove this by the same argument as we used above for $n = 1$. But we can also apply the useful easy-to-prove identity

$$\delta(x_1, x_2, \dots, x_n) = \delta(x_1)\delta(x_2)\dots\delta(x_n)$$

and the fact that the product of homogeneous functions of degrees k and l is a homogeneous function of degree $k+l$.

The Laplace operator. Let f be a smooth function on Euclidean space (of dimension n). Consider a sphere of small radius r centered at x. The mean value of f over this sphere is close to its value $f(x)$ at the center but does not exactly coincide with it.

Problem. *What is the order of the difference between the mean value and the value at the center as* $r \to 0$?

Solution. For $n = 1$, the mean value is

$$\hat{f}(r) = \frac{f(x+r) + f(x-r)}{2}.$$

Expanding the function f in the Taylor series $f(x+r) = f(x) + rf'(x) + \frac{r^2}{2}f''(x) + \ldots$, we obtain

$$\hat{f}(r) = f(x) + \frac{r^2}{2}f''(x) + o(r^2);$$

therefore, the difference

$$\hat{f}(r) - f(x) = \frac{r^2}{2}f''(x) + o(r^2)$$

is of the *second* order of magnitude with respect to the radius r of the sphere.

For an arbitrary n, the argument is almost the same. The linear term of the Taylor series averaged over the sphere is 0 because this term takes opposite values at opposite points of the sphere.

The cubic and higher-degree terms add a small (in comparison with r^2) correction $o(r^2)$. Therefore, the difference under examination is of the second order of smallness:

$$\hat{f}(r) - f(x) = Kr^2 + o(r^2).$$

The coefficient K is called the value (at the central point x) of the Laplacian, denoted Δf, under an appropriate normalization.

Problem. *Express the coefficient* K *in terms of the second partial derivatives of the function* f.

Solution. We must average the terms of the Taylor series, which contain different second-degree monomials (in the increments of the arguments), over a sphere of radius r.

For simplicity, we assume that the sphere is centered at 0 and denote the Cartesian coordinates of the increment vector by (x, y) (assuming that $n = 2$).

The mean values of the functions x^2 and y^2 on the sphere (circle) are the same, and the mean value of xy over this sphere is 0 (because the change of the sign of the x coordinate changes the sign of the function).

But the mean value of $x^2 + y^2$ over the sphere (circle) $\{x^2 + y^2 = r^2\}$ is r^2. Therefore, the mean values of the functions x^2, y^2, and xy over this sphere are $r^2/2$, $r^2/2$, and 0. The Taylor formula gives the quadratic contribution

$$\frac{\partial^2 f}{\partial x^2}\frac{x^2}{2} + \frac{\partial^2 f}{\partial y^2}\frac{y^2}{2} + \frac{\partial^2 f}{\partial x\,\partial y}xy$$

with mean value

$$\frac{\partial^2 f}{\partial x^2}\frac{r^2}{4} + \frac{\partial^2 f}{\partial y^2}\frac{r^2}{4},$$

whence we obtain (for $n = 2$) the required expression for the coefficient:

$$K = \frac{1}{4}\Big(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}\Big).$$

In the case of an arbitrary number n of Cartesian orthonormal coordinates $(x_1, \ldots, x_n)$ in Euclidean n-space, the same argument gives the answer

$$K = \frac{1}{2n}\Big(\frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} + \ldots + \frac{\partial^2 f}{\partial x_n^2}\Big)$$

(because the mean value of the function x_1^2 over the sphere $x_1^2 + \ldots + x_n^2 = r^2$ is equal to r^2/n).

Problem. *Find a spherically symmetric solution of the equation* $\Delta u = \delta$ *in Euclidean space* $\mathbb{R}^n$.

Solution. Any spherically symmetric function u on $\mathbb{R}^n \setminus 0$ has the form

$$u(x_1, \ldots, x_n) = f(r).$$

Considering $\Delta u = 0$ as an equation for f, we obtain a second-order ordinary differential equation. This equation has two linearly independent solutions; one of them is (obviously) $f \equiv 1$. As the second solution, we shall now find a homogeneous function u for which $\Delta u = \delta$.

If a function u is homogeneous of degree k, then Δu is homogeneous as well, but of degree $k - 2$.

Since the δ-function on $\mathbb{R}^n$ is homogeneous of degree $-n$, it follows that u must be homogeneous of degree $2 - n$. Thus, for $n \neq 2$, we obtain

$$f(r) = cr^{2-n}.$$

Let us calculate the constant c. To this end, note that $\Delta u = \operatorname{div} \operatorname{grad} u$. The gradient of the function r^{2-n} is a spherically symmetric field whose components have homogeneity degree $1 - n$:

$$\operatorname{grad} r^{2-n} = c_1 x r^{-n}.$$

The coefficient c_1 is determined by the behavior of this field on the x_1 coordinate axis, on which $r^{2-n} = x_1^{2-n}$ and

$$\frac{dx_1^{2-n}}{dx_1} = (2-n)x_1^{1-n} = (2-n)xr^{-n},$$

so that $c_1 = 2 - n$.

The flux of the vector field $\operatorname{grad} r^{2-n}$ through the sphere $x_1^2 + \ldots + x_n^2 = r^2$ of any radius r is equal to the product of $(2-n)r^{1-n}$ by the volume of an $(n-1)$-sphere of radius r; thus, this flow is

$$(2-n) \cdot \omega(n-1),$$

where $\omega(n-1)$ is the volume of an $(n-1)$-sphere of radius 1:

n	1	2	3
$\omega(n-1)$	2	2π	4π

By Stokes' theorem, this flux is equal to the integral of the divergence of the field under consideration over the ball bounded by the sphere.

Therefore,

$$\int_{\mathbb{R}^n} \operatorname{div}\operatorname{grad} r^{n-2}\, dx_1 \dots dx_n = (2-n)\omega(n-1).$$

It follows that, in $\mathbb{R}^n$, we have

$$\Delta r^{n-2} = (2-n)\omega(n-1)\delta.$$

Thus, the equation $\Delta u = \delta$ in $\mathbb{R}^n$ with $n \neq 2$ has the spherically symmetric solution

$$u = cr^{2-n}, \quad \text{where} \quad c = \frac{1}{(2-n)\omega(n-1)}.$$

This solution is called the fundamental solution (of the Laplace equation).

Example. In three-dimensional space ($n = 3$), the fundamental solution is

$$u = \frac{c}{r}, \text{where } c = -\frac{1}{4\pi}.$$

This is the law of the gravitational field and of the electrostatic Coulomb field.

For $n = 1$, we obtain the fundamental solution $u = cr$, where $c = 1/2$ (that is, $u = |x|/2$).

Problem. *Investigate the fundamental solution of the Laplace equation $\Delta u = \delta$ in the Euclidean plane.*

Solution. If we were physicists, we would say that the function $u = cr^{2-n}$, $n = 2$, must be understood as the limit of the prelimit functions with $n = 2 + \varepsilon$ as $\varepsilon \to 0$.

We would obtain

$$r^{2-n} = e^{(2-n)\ln r} = e^{-\varepsilon \ln r} = 1 - \varepsilon \ln r + o(\varepsilon).$$

The constant 1 contributes the harmonic function ($\Delta u = 0$), and the ε-linear term contributes the function $\ln r$, which is harmonic in $\mathbb{R}^2$ for $r \neq 0$.

Arguing as above, we can easily verify that the fundamental solution has the form

$$u = c \ln \frac{1}{r}$$

(we leave the determination of the constant c by calculating the divergence of the gradient of this function to the reader).

Remark. More "mathematically," the above argument consists in considering solutions of the equation $\Delta u = 0$ (for $u = f(r)$) as eigenfunctions (corresponding to the eigenvalue 0) of the operator Δ on a function space.

Of course, the above analysis of fundamental solutions for the Laplace operator on $\mathbb{R}^n$ provides formulas for the *potentials* of the corresponding force fields (gravitational and electrostatic).

The very fields ($\operatorname{grad} u$) are of the form, respectively,

$$F \sim r^{1-n} \sim \frac{\vec{x}}{|x|^n},$$

both for $n \neq 2$ and in the exceptional case $n = 2$.

In the case $n \neq 2$, this operator has the double eigenvalue 0 with two eigenvectors (studied above).

But the operator depends on the parameter n so that, for $n = 2$, there arises a Jordan block of size 2: the eigenvalue 0 is double, but the eigendirection is unique.

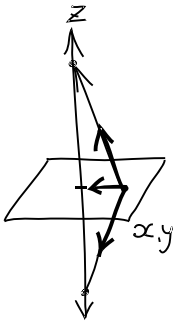
In this case, the planes spanned by both eigenvectors (corresponding to the case $n \neq 2$ of the operators) tend to a limit position as $n \to 2$. This limit plane is spanned by an eigenvector and a generalized eigenvector of the Jordan block.

It was this generalized eigenvector ($\ln r$) that we calculated in the "physical" solution given above.

A gravitational (or electrostatic) field in the plane can be obtained from a field in 3-space by considering cylindrically arranged attracting masses or charges (with density $\rho(x, y)$ not depending on the orthogonal coordinate z).

In other words, it is required to calculate the attraction of the points of the homogeneous line $x = y = 0$ in $\mathbb{R}^3$ by integrating the forces of attraction of various points on this line.

We leave this simple calculation to the reader; it is convenient to use at once the symmetry $(x, y, z) \mapsto (x, y, -z)$ and consider the net attractions of symmetric masses.



Physicists state the theorem proved above concerning the fundamental solutions of the Laplace operator on the Euclidean plane in the form (mathematically erroneous without additional explanations) of the relation $r^0 = \ln(1/n)$.

The *Dirac δ-function* described above is the simplest particular case of a "generalized function," whose theory was constructed by N. M. Gunter in 1916 under the title "theory of functions of domains": these "generalized functions" are not defined by their values at points, but by their integrals over all possible domains.

Gunter constructed this theory in order to prove existence (and uniqueness) theorems for solutions of the equations of hydrodynamics, Navier–Stokes. Gunter was then accused of producing an "anti-proletarian" aristocratic theory. To defend himself, he organized a seminar for communists and Young Communist League members. One of the participants, Gunter's pupil S. L. Sobolev, used Gunter's method of generalized solutions to study the linear wave equation (where discontinuous generalized solutions are needed for the proletariat; e.g., in seismology).

Sobolev's papers were translated from the French to the American language by L. Schwartz, who constructed, in this way, his "theory of distributions" for which he was awarded the Fields Medal.

In 1965 Laurent Schwartz told me that he got the Fields Medal for correcting errors in Sobolev's paper. I had read that paper and found no errors, so I asked Schwartz to point them out to me.

Schwartz answered: "Sobolev published his results in a language no one understands, in a city where nobody is interested in science, and also in a journal that nobody reads."

Although I knew where Sobolev's paper was published, I asked Schwartz to name the language, the city, and the journal; Schwartz answered: "In French, in Paris, in the journal *Comptes Rendus de l'Académie des Sciences*."

After returning in 1966 to Moscow, when I had pulled Sergei Lvovich Sobolev out from a rut into which he had driven in his car near the Zvenigorod market (to buy milk), I told him about Schwartz's theory.

Sergei Lvovich replied: "Laurent is a wonderful person, and he likes us both, but he lied to you: in the paper he was given the Fields Medal for, he not only translated my article, but also added his own theorems about the Fourier transforms of my generalized solutions, which I did not know!"

The question of the relationship between the work of Schwartz and Sobolev was to be settled by Hadamard, who came to Moscow for that purpose to talk to Sobolev. But in this he did not succeed, because S. L. Sobolev was then in Los-Arzamas (Sarov) as Kurchatov's Deputy Director.[1] Hadamard sought the advice of Kolmogorov, who said that to both he prefers the "true author"—Gunter (whose work on "functions of domains" motivated Kolmogorov's cohomology theory).

Dirac introduced his δ-function around 1930. He wrote that the only right way to create a new physical theory is "to forget about all the physical considerations, which are actually only a polite pseudonym for the prejudices of previous generations."

[1] [**Translator's note**] *I. V. Kurchatov* was the head of the Soviet atomic bomb ("Manhattan") project.

In his own words, one must begin with some meaningful mathematical theory: "if it is really beautiful, then its results will find, today or tomorrow, useful physical applications."

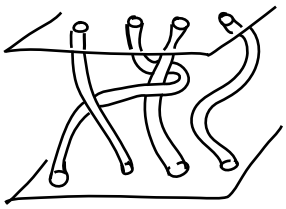
In his construction of the theory of spins of electrons, Dirac met with the following difficulty: physicists could not understand why those spins take two values ($+1/2$ and $-1/2$) although they describe the same "rotation of the electron."

The essence here is in a meaningful topological theorem: The fundamental group of the rotation group of three-dimensional space consists of two elements; i.e., $\pi_1(\mathrm{SO}(3)) \simeq \mathbb{Z}_2$.

This means that a rotation by 360° does not return the corresponding physical characteristic to its initial state. To return to that state, the rotation must be continued so that angle of rotation becomes 720°, not 360°.

This difficult theorem was not understood by physicists and led them to distrust spin theory.

Then Dirac found its consequence (also not at all obvious) in mathematical braid theory: he constructed a "spherical braid of four hairs," which is a second order element in the spherical braid group. Ordinary braids are "planar," their group is the fundamental group of the configuration space of n points in the plane (for braids on n strands).



In that group of planar braids there are no elements of finite order: we cannot unravel such a braid by attaching another one just like it to its end.

But for spherical braids, Dirac was able to demonstrate such an unraveling to physicists in an experiment (its strands were attached to three concentric spheres and unraveled when he burned the middle one).

To invent this physical experiment, Dirac used the (beautiful and nontrivial) mathematical theory of elliptic functions, which he understood quite well.

The idea is to consider four distinct points on the Riemann sphere $\mathbb{C}P^1$. Then the two-fold covering of the sphere branching at these points is the two-dimensional torus (the Riemann surface of the function $y = \sqrt{x^4 + ax^2 + bx}$; i.e., an elliptic curve). This circumstance determines a representation of the group of spherical braids on four strands in the automorphism group $\mathbb{Z} \oplus \mathbb{Z}$ of the 1-homology of the elliptic curve.

It is by computing these automorphisms that Dirac found a spherical braid on four strands of order two in the spherical braid group.

If Dirac had not been enamored in that sort of mathematics, physicists would never have been given the spin theory of electrons.

http://dx.doi.org/10.1090/mbk/085/35

Chapter 35

Inversion in Cylindrical Mirrors in the Subway

Everyone has seen his/her reflection in plane mirrors: The reflection of a left-hander is right-handed, but, otherwise, the image is similar to the original.

But those who have seen their reflections in a curved mirror know how funny they are.

For simplicity, consider a cylindrical mirror. How do the reflections of various objects in it look?

There are many cylindrical mirrors (vertical poles and horizontal handrails) in each subway car. The images of the surrounding world in these cylindrical mirrors are quite unusual. What are they like?

Hint. It is easiest to consider the reflection of a single point source of light. Its reflection in a cylindrical mirror is closely related to mathematical inversion, that is, the operation taking each point A of the Euclidean plane (in which a circle of radius r centered at O is fixed) to the point B "symmetric with respect to this circle"; the point B belongs to the same ray from O as A, but its distance from O is larger the closer A is to the center:

$$|OA| \cdot |OB| = r^2.$$

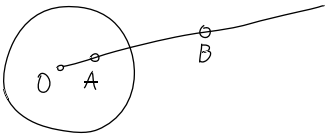


Fig. 1. Inversion takes A to B

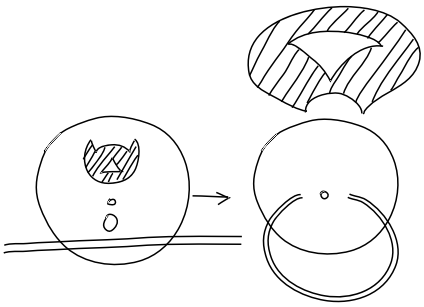


Fig. 2. Inversion releases the cat from the cage and rounds out the straight line

Solution. Each ray issuing from A and intersecting the given circle is reflected from the circle according to the law "the angle of incidence is equal to the angle of reflection" (see Fig. 3).

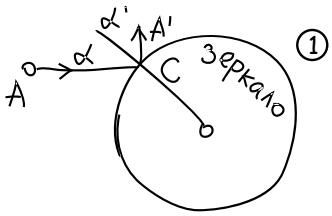


Fig. 3. The reflection CA' of the ray AC (the angles α and α' are equal) ① *Mirror.*

If the mirror is plane, then all rays from the source A are reflected as rays whose extensions pass through the same point A^* behind the mirror. Thus, the reflected rays form the pencil $\{A^*A'\}$, and that is why we see the reflection of the point A at the point A^* behind the mirror (see Fig. 4).

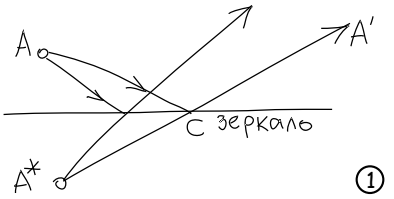


Fig. 4. The reflection CA' of the ray AC in a plane mirror and the point A^* behind the mirror. ① *Mirror.*

If the mirror is curved, then rectilinear rays reflected at different points do not necessarily pass through a common point, even when extended behind the mirror.[1] To understand this, it suffices to consider an example, say the reflection in a circular mirror of a pencil of parallel rays issuing from the same point A at infinity.

The explicit calculation of rays reflected at different points of a circular mirror is not very hard (for those who knows trigonometry). But it is even easier to draw these rays (see Fig. 5). The arcs CD

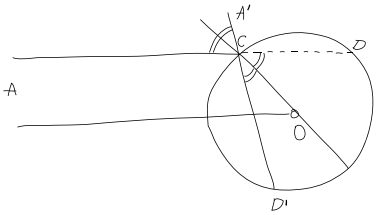


Fig. 5. Construction of a ray CA' reflected in a circle

[1] The only exception is that of rays parallel to the axis of a parabolic mirror: after reflection in the parabola, they meet at one point.

and CD' are of the same length (because of the law "the angle of incidence is equal to the angle of reflection" at the point of reflection C). This allow us to quickly construct the reflected rays.

Drawing these reflected rays neatly enough, I obtained the following picture (see Fig. 6).

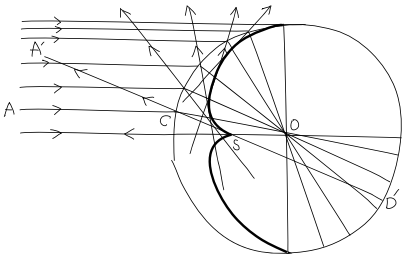


Fig. 6. A family of rays issuing from infinity reflected in a circle and its envelope

The resulting one-parameter family of reflected straight lines has an envelope (the heavy line in Fig. 6). This is the curve at whose points the straight lines from the reflected family of rays intersect the infinitely close reflected lines from the same family of rays. These lines (extended reflected rays) are tangent to the envelope. We can also say that this curve is formed by the "focal points" of the reflected family of rays (in optics, a focal point is an intersection point of extensions of infinitely close rays from a family).

The envelope of such a family of rays is called a *caustic* ("burning"), because the light carried by the family is concentrated (focused) on it, and so the energy on this curve is higher than at other places. According to the legend, it is the caustic of a system of mirrors which Archimedes used to burn down enemy ships besieging Syracuse.[2]

[2]In *The Clouds*, Aristophanes attributed an even earlier use of caustics for business purposes to Socrates, who advised his client to buy a lens in a pharmacy and, at the court session, choose a sunny spot, wait until his opponent would show his promissory note to the court, and burn it down by means of a caustic of solar rays. Aristophanes, however, mentions that it was this applied mathematics which had led Socrates to the capital sentence pronounced by his fellow citizens.

In any case, most of the reflected rays travel as if they issue from points of the caustic, so that the image of our initial point A at infinity appears to be a line spread over the caustic rather than a point.

However, the matter is even more complicated, because the brightness of the image along the caustic is not at all constant; some parts of the caustic are brighter (and it is these parts which Archimedes used for his system of rays).

Namely, the caustic of the family of reflected rays in Fig. 6 is not a smooth curve: it has a singular point S (it is easy to calculate that it is the midpoint of the radius).

Near this point, the (extended) rays concentrate even more than at the other points of the caustic.[3] Therefore, although the image of a shining (infinitely remote) point A is spread over the caustic, the singular point S is particularly bright (while the other points may remain unnoticed by a spectator not observant enough).

As a consequence of all this, the image of the point A observed by an experimenter is, rather than a line, the single point S of maximum concentration of the reflected rays extended beyond the mirror.

Trigonometric calculations, which I leave to the reader, confirm these conclusions and their stability: for a light source A at a different location, we again obtain a caustic of rays extending beyond the mirror with a singular point of return, which is perceived by an observer as the image A^* of A in the curved mirror.

This point A^*, as well as the point S in the above example with an infinitely remote source A, lies on the same ray from the center O of the mirror as the source A. But the position of this point on the corresponding radius of the circle depends on the distance of A from the center (when this distance is infinite, the reflected point bisects the radius, and when A is on the reflecting circle, the point A^* degenerates into A).

[3] It can be calculated that this singularity is a semicubical return point (in its neighborhood, the caustic is determined by the equation $y^2 = x^3$ in an appropriate curvilinear coordinate system). Such a singularity is typical (of generic systems of rays) and stable (it does not disappear under small perturbations of the family), and it was the one used by Socrates and Archimedes.

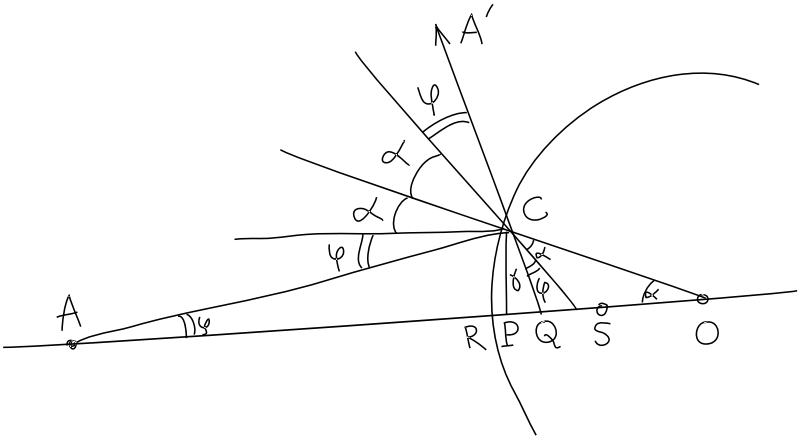


Fig. 7. The intersection of infinitely close straight lines AR and AC behind the mirror

The calculation of the position of the image A^* on the ray OA for a given distance $|OA| = X \cdot R$ is shown in Fig. 7.

The radii of the reflecting circle have the lengths

$$|OR| = |OC| = R.$$

The small central angle α determines the legs of triangle OCP:

$$|OP| = R\cos\alpha, \qquad |CP| = R\sin\alpha.$$

From the right triangle ACP, we obtain an asymptotic expression for the small angle φ:

$$\tan\varphi = \frac{|CP|}{|AP|} = \frac{R\sin\alpha}{R(X-\cos\alpha)} \sim \frac{\alpha}{X-1}, \qquad \varphi \sim \frac{\alpha}{X-1}.$$

The right triangle OCP yields the expression

$$\gamma = (\pi/2) - (\varphi + 2\alpha), \text{ where } \varphi + 2\alpha \sim \frac{2X-1}{X-1}\alpha,$$

for the angle PCQ.

The length of the leg opposed to γ in the right triangle CPQ is

$$|PQ| = |CP|\tan\gamma = |CP|\,\frac{\cos(\varphi+2\alpha)}{\sin(\varphi+2\alpha)}.$$

The asymptotics for $|CP|$ and $\varphi + 2\alpha$ found above determine the behavior of the distance between P and Q as $\alpha \to 0$, namely,

$$|PQ| \sim \frac{R \sin \alpha}{\frac{2X-1}{X-1}\alpha} \to R\frac{X-1}{2X-1}.$$

Thus, the distance between the reflected point Q and the midpoint S of the radius OR tends to

$$|QS| = |PS| - |PQ| \to \frac{R}{2} - R\frac{X-1}{2X-1} = \frac{R}{2(2X-1)}.$$

The distance between the source A and the midpoint S of OR is

$$|AS| = |AO| - |SO| = R(X - 1/2) = \frac{2X-1}{2}R.$$

We conclude that the distances from S to the source A and to its reflection Q are reciprocal in the sense that

$$|QS| \cdot |AS| = R^2/4.$$

Thereby, we have obtained the following (amazing) result.

In a cylindrical mirror, an observer sees the inverse of the surrounding world with respect to the cylinder tangent to the axis of the reflecting cylinder and half as thick (in our plane notation, this is the inverse with respect to the circle of radius $R/2$ centered at S).

Fig. 8. A reflection in a cylindrical mirror is the inverse with respect to a (heavy) circle. ① *Observer.* ② *Mirror.* ③ *Center of inversion.*

One might think that, looking at a cylindrical mirror (e.g., at handrails in the subway), we see the inverse image of the surrounding objects.

That this cannot be so is clear already from the description of the position of the circle (or cylinder) of inversion with respect to the reflecting circle in the plane (or the cylindrical mirror in space). Namely, the cylinder of inversion is shifted off in a certain direction from the axis of the reflecting cylinder, while, because of the symmetry of the reflecting cylinder with respect to rotations about its axis, all directions from the axis of rotation must enjoy equal rights, and none of them can be preferred.

In reality, the above calculations show that the reflection of each source of light can be obtained by applying the inversion described above to the source point only for points of the ray passing through the center of the reflecting circle and the observer's eye (in calculations, this is formalized by the assumption that the angle φ is small).

On this central ray of view, the images A^*, B^*, C^*, and D^* of the points A, B, C, and D (see Fig. 9) are indeed inverse to the points;

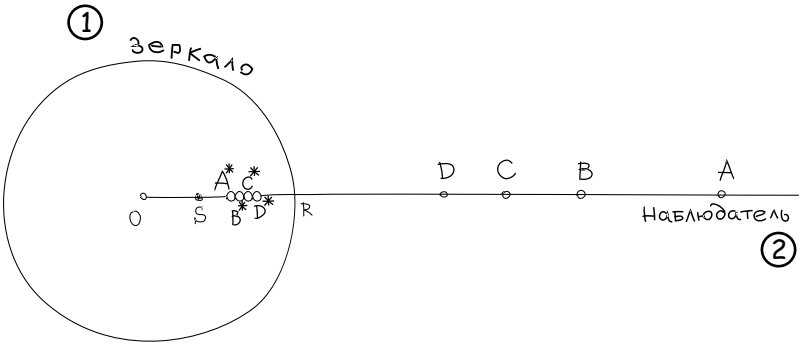


Fig. 9. The images A^*, B^*, C^*, and D^* of points A, B, C, and D on a central ray. ① *Mirror.* ② *Observer.*

therefore, near the central ray of view, the reflection is approximately described by an inversion. But as a point moves away from the central ray of view, the circle of inversion that describes its reflection is rotated, so that the whole reflection does not reduce to a single inversion.

Appendix: On Properties of Inversions. Here I briefly describe several remarkable facts, although many readers may already know about them.

Theorem. *An inversion transforms the circles not passing through its center to circles and those passing through its center to straight lines* (see Fig. 10).

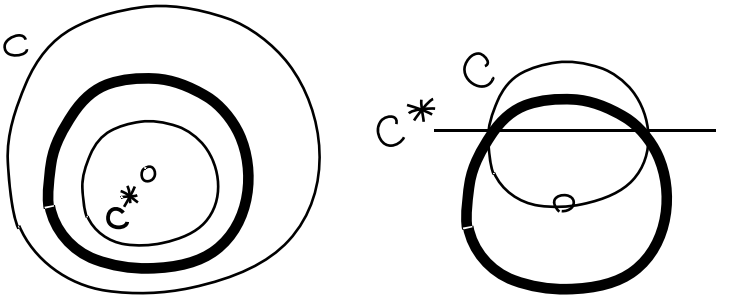


Fig. 10. The inverse of the circle c is the circle c^*, and the inverse of the circle C is the straight line C^*

The proof of the second assertion is particularly simple when C intersects the circle of inversion (see Fig. 11).

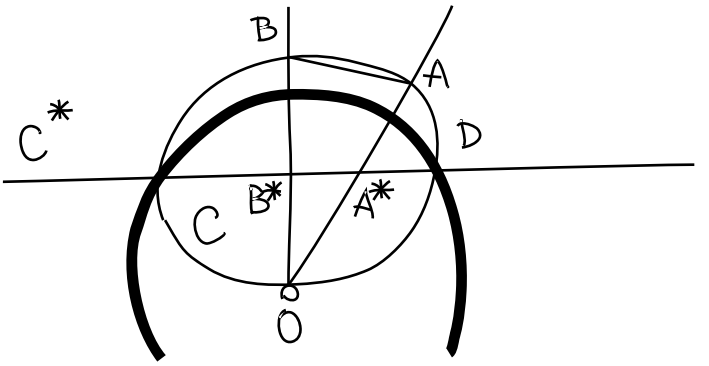


Fig. 11. The inversion of the circle C passing through the center O of the (heavy) circle of inversion

The right triangles OB^*A^* and OAB are similar; therefore, we have $|OB^*|/|OA^*| = |OA|/|OB|$, so that $|OA| \cdot |OA^*| = |OB| \cdot |OB^*|$.

In the case $A = D$, we obtain $|OB| \cdot |OB^*| = R^2$. This proves the coincidence of the inverse of the circle C with the straight line C^*, which joins the intersection points of C with the circle of inversion.

The case in which the circle C is too small to intersect the circle of inversion is reduced to the case considered above by applying a dilation (a homothety centered at O). When C undergoes such a homothety (is dilated by a factor a), its inverse also undergoes a homothety centered at O (is contracted by a factor a).

Since the contracted inverse is a straight line, it follows that the true (noncontracted) inverse is a straight line as well (but does not intersect the circle of inversion).

The assertion of the theorem about the image of a circle c not passing through the center of inversion is particularly easy to prove in the case when the disk enclosed by this circle c does not contain the center of inversion O (see Fig. 12).

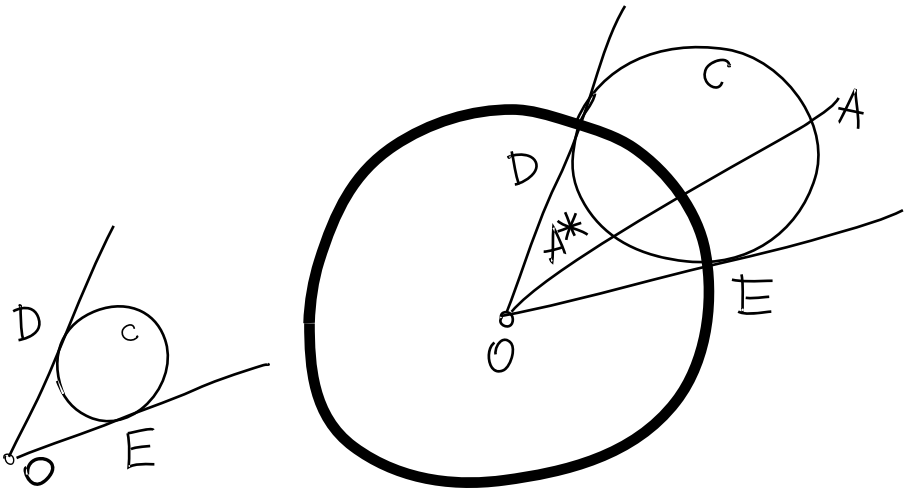


Fig. 12. An inversion of a circle c not enclosing the center of inversion O

In this case, we can draw two tangents to the circle c from O. They have equal lengths: $|OD| = |OE|$. Dilating (or contracting) the plane by a homothety centered at O, we can transform the circle c into a homothetic special circle for which the lengths $|OD| = |OE| = R$ of the tangents coincide with the radius R of the circle of inversion (so that the special circle intersects the heavy circle of inversion in points D and E at right angles).

Applying the secant theorem to the secant OA^*A of the special circle c, we obtain

$$|OA^*| \cdot |OA| = |OD|^2 = R^2.$$

This identity means that the points A and A^* of the special circle c are inverse to each other, so that the image of the special circle coincides with this circle itself.

Returning to the initial circle by contracting the special circle, we see that the inverse of this contracted (initial) circle is obtained from the special circle by a homothetic dilation. Therefore, this image c^* is a circle as well.

In the case when the circle c encloses the center O, the theorem remains valid. But I do not know a proof which is as simple as that given above.

Remark. The special circle is orthogonal to the circle of inversion. The inversion takes these circles to themselves and, therefore, preserves the angle between them.

It turns out that an inversion preserves angles between any curves (up to sign). This is seen; e.g., from Fig. 13, where a circle C passing through the center of inversion O intersects the circle of inversion at the point D (and inverts to the straight line DE).

The normals OD (to the circle of inversion) and OB (to the inverted curve C^*) at the intersection point O form the angle $\alpha = \angle DOB$.

The tangents at the points O and D to the circle being inverted form an isosceles triangle; therefore, angles DOM and ODM are equal to $\pi/2 - \alpha$.

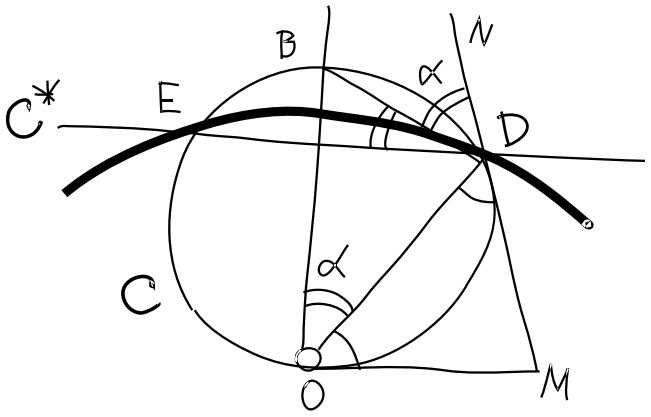


Fig. 13. Preservation under inversion of the angle with the circle of inversion

The tangent DB to the circle of inversion at D passes through the endpoint B of the diameter OB of the circle C being inverted because angle BDO (between the tangent and a radius of the circle of inversion) is right.

Therefore, the angle BDN (between the tangents to the circle of inversion and the circle C at their intersection point D) is equal to the angle $DOB = \alpha$ between the normals to the circle of inversion and the inverted curve C^* (it is equal to $\pi - \pi/2 - (\pi/2 - \alpha) = \alpha$).

Thus, at the point D, the angle which the circle being inverted makes with its inverse is equal to the angle which this circle makes with the circle of inversion.

It follows that inversion preserves the angle between any curves passing through D and the circle of inversion; hence, it also preserves the angle between any two curves passing through D.

Of course, the orientation of angles is not preserved: Like an ordinary reflection, an inversion changes the orientation of the plane and takes "positive" angles to "negative" ones (of the same magnitude).

Our considerations prove the preservation of all (undirected) angles at the points of the circle of inversion. But we can place any

(different from O) point of the plane on this circle (of radius R) by applying an appropriate homothety (centered at O).

Homotheties preserve angles; therefore, applying homothetic dilations and contractions of the plane, we can derive the preservation of the angles of intersection of any curves at any point (other than the center of inversion O) from their preservation at points of the circle of inversion, which we have already proved.

Transformations which preserve angles are said to be conformal. Thus, inversion is a conformal transformation of the plane (minus the point O) changing orientation.

Problem. *Let $f\colon \mathbb{C} \to \mathbb{C}$ be any polynomial treated as a (self-)map of the Euclidean plane $\mathbb{C} \approx \mathbb{R}^2$ with Cartesian orthonormal coordinates (every point $z = x + iy$ has coordinates (x, y)).*

Prove that the map f is conformal (at any noncritical point of the polynomial f; i.e., at any point where the derivative of f is different from 0).

Solution. Begin with a linear polynomial and use the Taylor formula to reduce any map to its (linear) differential.

In these terms, an inversion is given by

$$f(z) = \frac{1}{\overline{z}},$$

where $\overline{z} = x - iy$, and its conformality follows from differentiability:

$$\frac{d(1/z)}{dz} = -\frac{1}{z^2}.$$

Problem. *Is the map taking $z \in \mathbb{C}$ to z^2 conformal at all points of the plane?*

Solution. The real axis $\{y = 0\}$ and the imaginary axis $\{x = 0\}$ in the plane $\mathbb{C} = \{z\}$, which are perpendicular straight lines, are mapped to the half-lines of positive and negative values of z^2, which are not at all orthogonal.

This violation of conformality strongly distorts shapes of figures (see Fig. 14).

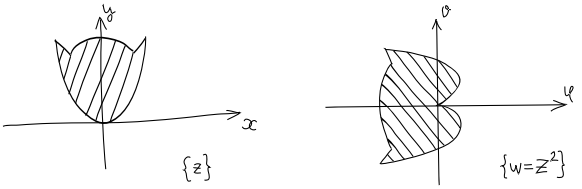


Fig. 14. Under a nonconformal transformation, the smooth chin of a cat became nonsmooth

An inversion is a conformal transformation, and the inverted figures resemble the originals more closely.

Problem. *Do the inversion transformations* (*with various circles of inversion*) *form a group?*

Solution. Any inversion changes orientation, and there are only two orientations in the plane. Therefore, the product of two inversions (which preserves the orientation of the plane) cannot be an inversion.

The orientation-preserving products of inversions (with even number of multipliers) do form a group. This is the group of linear-fractional transformations

$$f(z) = \frac{az + b}{cz + d},$$

which is fundamental for hyperbolic geometry: all f with real a, b, c, and d for which $ad - bc = 1$ form the group of motions in the Poincaré model.

This is a model of the hyperbolic plane in the upper half-plane $\operatorname{Im} z > 0$; in this model, in contrast to the Cayley–Klein disk model discussed above, the role of straight lines is played by all Euclidean

straight lines and circles perpendicular to the absolute $\operatorname{Im} z = 0$ (see Fig. 15), rather than by the Euclidean straight lines.

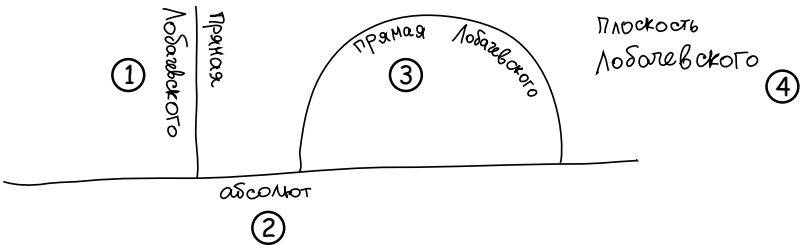


Fig. 15. ①, ③ *Hyperbolic line.* ② *Absolute.* ④ *Hyperbolic plane.*

The Poincaré model has the remarkable property that the hyperbolic angles in this model are equal to the Euclidean angles between the corresponding curves in the upper half-plane.

It is also amazing that the Poincaré and Cayley–Klein models are equivalent: these are simply different charts of the same hyperbolic plane.

Problem. *Find a diffeomorphism of the upper half-plane to the interior of the unit disk which maps the Poincaré model to the Cayley–Klein model.*

http://dx.doi.org/10.1090/mbk/085/36

Chapter 36

Adiabatic Invariants

The theory of adiabatic invariance is a strange example of a physical theory that apparently contradicts mathematical facts that seem easy to verify.

Despite such an unpleasant property, this "theory" has led to remarkable physical discoveries by those who were not afraid to use its conclusions (although they were not justified mathematically).

The development of science over a couple of centuries finally led to an agreement of sorts between mathematicians and physicists: mathematicians proved the "theorem on the conservation of adiabatic invariants" under certain (precisely specified) assumptions.

Conjectures on the possibility of substantially weakening these assumptions have also acquired more or less rigorous mathematical formulations today (but they are still awaiting proofs). The present essay presents only a few examples, it is hardly exhaustive even for already proved theorems on adiabatic invariance. (For a review of these theorems, see the book *Additional Chapters of the Theory of Differential Equations*[1] §20.)

We will be dealing with systems of differential equations with coefficients variable in time and depending on a point x in the phase

[1] [**Translator's note**]. English translation: V. I. Arnold, *Geometrical Methods in the Theory of Ordinary Differential Equations* (Springer-Verlag, New York–Heidelberg–Berlin, 1983).

space M:

$$\frac{dx}{dt} = v(x,t), \quad x \in M. \tag{1}$$

The assertion that some quantity $I(x,t)$ is an adiabatic invariant means that, although it is not an exact first integral of the equations of motion (1), its changes are nevertheless small even for large values of $x(t)$, under the condition that the right-hand side of equation (1) varies "slowly enough" in "fast time" t.

In order to give a mathematically precise definition of the "slowness" needed here, let us consider, instead of the nonautonomous system (1), a family of dynamical systems (with the same phase space M) depending on a parameter λ (that ranges over some manifold Λ):

$$\frac{dx}{dt} = v(x,\lambda), \quad x \in M,\ \lambda \in \Lambda.$$

The condition of slowness of change of the system can now be stated in terms of the dependence of the parameter λ on time:

$$\lambda = f(t).$$

In order to make the variations in the values of λ small, let us also consider, together with the "fast time" t, the "slow time" $\tau = \varepsilon t$ (where ε is a small parameter, which will later tend to 0).

The variation of the parameter λ in time is now regarded as given by its dependence on slow time,

$$\lambda = f(t) = F(\tau), \quad \tau = \varepsilon t,$$

where F *is a fixed dependence of the parameter on slow time.*

The *adiabatic invariance* of the quantity $I(x,\lambda)$ is defined by the difference

$$|I(x(0),\lambda(0)) - I(x(t),\lambda(t))| < \kappa$$

being small for $0 \le \tau \le 1$; i.e., during a long interval of fast time $0 \le t \le 1/\varepsilon$ (as the point in phase space shifts by $|x(t) - x(0)| \sim 1$), provided that the parameter varies slowly enough:

$$\lambda = F(\varepsilon t),$$

where ε is sufficiently small ($\varepsilon < \varepsilon_0(x)$).

The difficulty here is in that the described condition $\varepsilon < \varepsilon_0(x)$ of smallness in the speed of variation of the parameter (although it is indeed required) does not always guarantee the smallness in the change of the quantity I (that physicists still insist on calling an "adiabatic invariant") in time $1/\varepsilon$, which is, you see, quite long.

The way out found today is that, in many cases, only a small increment of the quantity I can occur if the dependence of F on slow time is a sufficiently smooth function (say, $F \in C^2$). The role of smoothness here is to replace the physical notion of "absence of knowledge."

Physicists say that "the person who changes the value of the parameter λ at time t should not have any knowledge of the position of the point $x(t)$ in phase space."

It is difficult to give a mathematical formulation of the "absence of knowledge." But it turns out that it can be replaced by the requirement that the function F be smooth. If there is no such smoothness, then, by choosing appropriate jumps or breaks in the dependence of λ on time, one can achieve large changes in I, while smoothness excludes such counterexamples.

Other attempts to give sufficient conditions for adiabatic invariance are based on the following: although the changes in the value of $I(x,t)$ in large time $t \sim 1/\varepsilon$ may be not be small, they occur rarely (i.e., are observed only for a small-measure set of low probability initial points $x(0)$ of the trajectories $\{x(t)\}$ in phase space).

In what follows I shall use the term "adiabatic invariance" in the sense specified above and based on the smoothness of the function F, which leads to the smallness of the changes in the values of the quantity I along the studied trajectory in time $1/\varepsilon$ for all initial conditions.

Example 1. The equation of small oscillations of the mathematical pendulum.

Consider the equation

$$\frac{d^2x}{dt^2} = -\lambda x, \quad \lambda > 0. \tag{2}$$

For a fixed value of the parameter λ, the phase curve is the ellipse

$$\frac{p^2}{2} + \lambda\frac{q^2}{2} = E$$

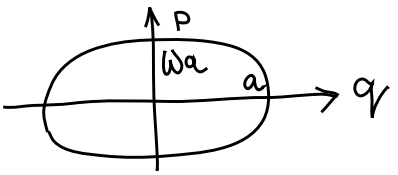
that comes from the law of conservation of energy

$$H(p, q; \lambda) = \frac{p^2}{2} + \lambda\frac{q^2}{2}$$

(here, as usual, $p = dx/dt$, $q = x$, and H is the Hamiltonian of the system $\dot{q} = \partial H/\partial p$, $\dot{p} = -\partial H/\partial q$).

The solution of system (2) with constant $\lambda = \omega^2$ is a harmonic oscillation

$$q = a\sin(\omega t), \qquad p = \omega a\cos(\omega t).$$



This elliptic-phase-curve bounds (on the symplectic plane with coordinates (p, q)) the area $S = \pi a \cdot (\omega a) = \omega \cdot (\pi a^2)$. The angular amplitude a of this oscillation and its energy $E = \omega^2 a^2$ are linked by the "Planck relation" $E = \frac{\omega}{\pi} S$. It turns out that *the value of the phase area*

$$I(p, q; \omega) = \frac{\pi H(p, q)}{\omega} = \pi a^2 \omega$$

is an adiabatic invariant of system (2).

The invariance of the product I means, in particular, that if the length l of the mathematical pendulum (for which we actually have $\omega^2 = \lambda = l/g$) slowly doubles, then the factor ω increases $\sqrt{2}$ times, and so the factor a^2 in the product $I = \pi a^2 \omega$ decreases $\sqrt{2}$ times.

In other words, the maximal deviation angle decreases $\sqrt{2}$ times when the length of the pendulum slowly doubles. And if the length of the pendulum returns to its initial value, the amplitude of oscillations also returns to its initial value.

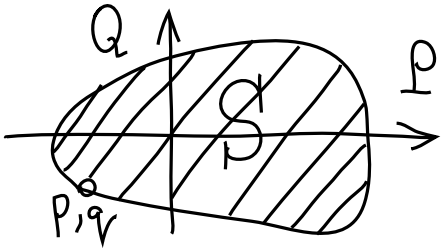
The amazing thing about this theorem is that the result absolutely does not depend on the law according to which the lengthening

of the pendulum occurred: it is only required that the function F which determines the change of $\lambda = F(\varepsilon t)$ be smooth.

Thus, in the "adiabatic limit", two physically independent quantities (a and l) become functionally dependent. This unusual physical phenomenon immediately distinguishes the adiabatic theory among many others.

The proof of the theorem on the adiabatic invariance of the "action variable" $S(p, q; \lambda)$, which expresses, in terms of the initial point in phase space, both the value of the parameter λ and the area bounded by the phase curve

$$S(p, q; \lambda) = \iint_{H(P,Q) \leqslant H(p,q)} dP\, dQ.$$



can be found in the textbook *Mathematical Methods of Classical Mechanics*[2] (for any Hamiltonian system with one degree of freedom), as well as in *Geometric Methods of the Theory of Ordinary Differential Equations*.

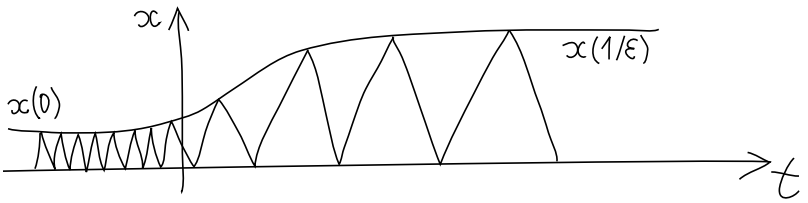
The examples below provide other similar cases (where the proof of adiabatic invariance can be carried out in a similar way)—they can also be derived from the already considered case of Hamiltonian system with Hamiltonian function $H(p, q; \lambda)$ by its appropriate generalization, allowing, say, collisions with rigid walls instead of potential force fields.

Example 2. Consider a "billiard ball" moving between two parallel walls whose distance from each other is x. Denote the velocity of the ball by v and assume that at the moment of impact the velocity reverses with respect to the wall.

[2] [**Translator's note**] English translation: V. I. Arnold, *Mathematical Methods of Classical Mechanics*, 2nd ed. (Springer, New York, 1989).

In this case the adiabatic invariant is the product $I = x|v|$ (which is of course proportional to the area of the corresponding phase curve for a fixed value of the parameter λ). Adiabatic invariance here means that the product $I = x|v|$ changes very little (in large time of order $1/\varepsilon$).

In other words, when the distance between the walls doubles, the velocity of the ball between them decreases by half (whatever smooth law $x = F(\varepsilon t)$ governs how the distance increases in time $t \sim 1/\varepsilon$):



The fact that moving the walls apart decreases the velocity of the ball bouncing between them is understandable, but the theory of adiabatic invariance of the product $x|v|$ provides us with a remarkably precise description of this decrease.

Remark. Although this theory is applicable only for $t \sim 1/\varepsilon$, in the case of an analytic dependence ($x = F(\tau)$, $\tau = \varepsilon t$) of the distance between the walls on time, we can even study the increment of the adiabatic invariant in infinite time

$$I(x(+\infty), |v(+\infty)|) - I(x(-\infty), |v(-\infty)|)$$

The description of this increment (which turns out to be exponentially small as $\varepsilon \to 0$) is obtained by studying the behavior of the holomorphic function F at complex points τ (it was found by A. M. Dekhne, Zh. Eksp. Teor. Fiz., **38**, No. 2, 1960, 570–578).

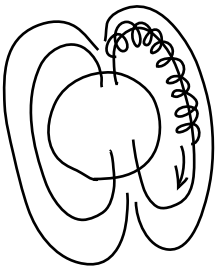
Example 3. In three-dimensional Euclidean space, consider a magnetic field B and a charged particle moving with velocity v. Denote by $v_\perp$ the velocity component perpendicular to B.

Were the field constant, the particle would move around a straight line of force along a Larmor spiral, rotating at a fixed distance r from

the line of force (called the Larmor radius and depending on the vectors B and v at the initial point of the trajectory).

In this case the adiabatic invariant is $v_\perp^2$.

By the adiabatic limit here, we can understand either the limit as $|v| \to 0$ or as $|B| \to \infty$ (it is only important that the Larmor radius tend to 0). For a smooth field B, one can prove that the adiabatic invariant indicated above does not change much in large time.



In particular, all this can be applied to explain Polar Auroras: charged particles in their spiraling motion around the lines of force of Earth's magnetic field near the magnetic poles reach the region of large values of the tension of the magnetic field $|B|$. The conservation of the adiabatic invariant in this case leads to the reflection of the moving particle from the "magnetic plug," and the particle returns along a (different) magnetic line to the second pole (in microseconds).

The particles awaiting reflection accumulate near the plugs, and it is these "clouds" of charged particles that are observed as auroras.

A mathematically precise description of this situation is rather long and I do not present it here. In contrast, the next version of a similar theory is easy to formulate precisely.

Example 4. Consider a smooth surface M with a fixed Riemann metric. On this surface, consider a curve of constant geodesic curvature κ.

In the case of the Euclidean plane M, the curve will be a closed (for $\kappa > 0$) circle of radius $r = 1/\kappa$. Now, if the geodesic curvature varies along the curve ($\kappa = B(x)$, $x \in M$), then the function $B : M \to \mathbb{R}$ determines "Larmor circles" on the surface M; i.e., a spiraling

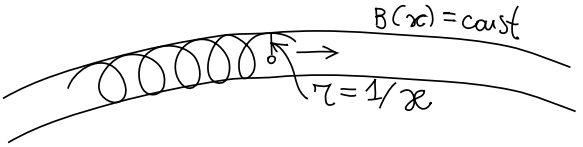
motion along a varying "Larmor circle" whose center moves along M and whose radius changes.

Physically this motion can be called motion in M of a charged particle (in a magnetic field B "perpendicular" to M).

This motion admits an adiabatic description when radii of the "Larmor circles" are small (physically we can either consider strong magnetic fields, $B \to \infty$, or small initial velocities, $|v| = |dx/dt|$.

If the function B is smooth, then the adiabatic invariant is simply the geodesic curvature of its corresponding Larmor circle of varying radius (in physical terms, it is $|v|^2/|B|$).

In particular, such a curve with large geodesic curvature depending on the point of the surface M oscillates between two neighboring level lines of the nonconstant function B, provided the curvature is large.



But if the function B is constant, then the corresponding Larmor circles of large geodesic curvature κ still oscillate in an annulus between two level lines of the adiabatic invariant, except that instead of the function $B : M \to \mathbb{R}$ we must consider the Gaussian curvature $G : M \to \mathbb{R}$.

The difference between these two theories, however, lies in the fact that when the Larmor radius tends to zero, the small velocity of motion of the center of the Larmor circle along the annulus between two level lines of the adiabatic invariant has a different order of magnitude (for $b \neq$const the velocity is much larger).

In both cases, the adiabatic invariant changes little not only in time $|\varepsilon|$, but in infinite time as well (this follows from "KAM theory").

Returning to the pendulum from Example 1, note that the adiabatic invariance of the ratio of its energy to its frequency seems to contradict the possibility of "pumping up" a swing: when the effective length l of the swing changes with arbitrarily small amplitude, its

lower position becomes unstable in the case of parametric resonance (when the period of changes of l is an integer multiple of the own half periods of oscillation of the unperturbed swing).

This remark provides a "counterexample" to the adiabatic invariance of the ratio of energy to frequency because the frequency, having slowly oscillated, returns to its unperturbed value, while the amplitude of the oscillating swing has increased.

But there is no contradiction here, because in order to achieve the increase in the amplitude of swinging, one needs "feedback"; i.e., one must know whether to increase or decrease the value of the parameter l at the given moment of the phase of the swing's own oscillations.

The smoothness of the law F of variation of the parameter $\lambda = F(\varepsilon t)$, assumed in the statement of the theorem on the adiabatic invariance of the action variable $S(p, q; \lambda)$ excludes the possibility of such feedback. But if this smoothness is not assumed, then the mathematical counterexamples to the physical statement of adiabatic invariance become possible.

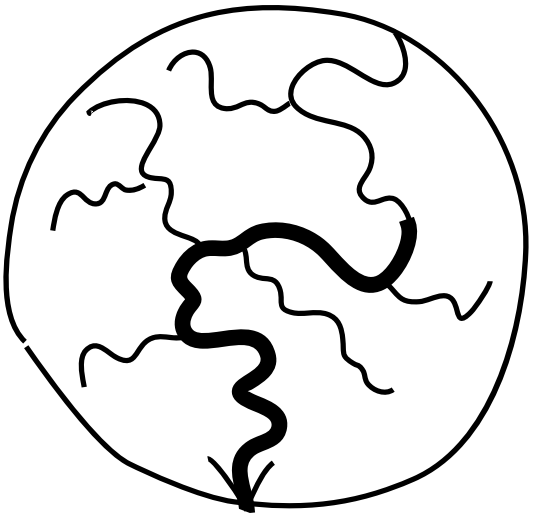
Further generalizations of the theory of adiabatic invariants are described in the book *Geometric Methods in the Theory of Ordinary Differential Equations*, §20, where numerous references to the literature appear.

http://dx.doi.org/10.1090/mbk/085/37

Chapter 37

Universality of Hack's Exponent for River Lengths

Encyclopedia articles on many rivers provide both the length l of the river and the area S of its basin. The question is, how are these two numbers related?



Solution. If the basin of a river were a disk centered at the middle of a straight river, then we would have $l = cS^{1/2}$ (which is also suggested by the dimensions).

According to American data, the statistics (over a large number of rivers, big and small, mountainous and flowing in plains) gives usually a greater length, namely, $l \approx cS^{\alpha}$, where $\alpha \approx 0.58$ ("Hack's exponent").

Hack's exponent α being larger than $1/2$ is explained by the fractal tortuosity of a river, because of which the length of the river is greater than the diameter of its basin.

But why the exponent α is universal and has exactly this value is not clear (although attempts have been made to derive it from the Navier–Stokes equations of hydrodynamics, the instability of whose solutions make rivers wind). The following table contains data for a dozen rivers in the Moscow area:

River	Length (km)	Area (km^2)	$\alpha \approx \ln l / \ln S$
Moscow	502	17640	0.64
Protva	275	4640	0.66
Vorya	99	1160	0.65
Dubna	165	5474	0.56
Istra	112	2120	0.61
Nara	156	2170	0.65
Pakhra	129	2720	0.62
Skhodnya	47	259	0.69
Volgusha	40	265	0.60
Pekhorka	42	513	0.59
Setun'	38	187	0.69
Yauza	41	452	0.59

The average value of the 12 exponents α presented here is 0.63.

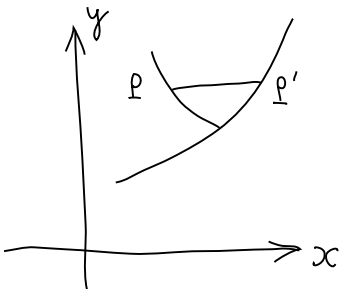
http://dx.doi.org/10.1090/mbk/085/38

Chapter 38

Resonances in the Shukhov Tower, in the Sobolev Equation, and in the Tanks of Spin-Stabilized Rockets

Working on Hilbert's 13th problem in 1958, I studied a representation of a function defined on a plane curve in the form of the sum of two functions, each depending on only one of the coordinates:

$$u(P) = f(x) + g(y).$$



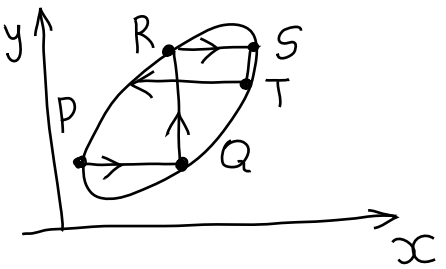
I had succeeded in investigating this problem when the curve is a tree. For example, let us choose a point P on the tree and decompose

the known value $u(P)$ into a sum $f(x) + g(y)$ of any two terms. We know the value of the term g at the point P', and we can find the value of f because the value of the sum is known as well.

And so, it turned out that any tree can be placed in $\mathbb{R}^3$ so that, for any continuous function u on this tree, a similar method yields a representation in the form of a sum of three continuous functions, each of which depends on only one of the coordinates (x, y, z):

$$u(P) = f(x) + g(y) + h(z).$$

Having solved this problem of Hilbert, I decided to generalize the theorem proved for a tree to any curve. If a curve (in the plane) has a cycle, then the dynamical system $P \to Q \to R \to S \to \dots$ arises on this cycle.



And it turned out that the existence of the required representation depends on the properties of this dynamics (a self-map of the cycle): if this dynamical system has periodic orbits, then such a representation does not always exist, and if there are no periodic orbits, then its existence depends both on the smoothness properties of the function u to be decomposed and on the arithmetical properties of Diophantine problems on the given closed curve.

Having proved dozens of theorems in this area (which is equivalent to the study of the Dirichlet problem for the wave equation), I wrote a paper about this. The referees pointed out to me that this problem had earlier been tackled by S. L. Sobolev's students (R. A. Aleksandryan and N. N. Vakhaniya) and by Sobolev himself (whose work, however, still remains classified, because he applied his theorems to study the flight of spin-stabilized projectiles containing liquid).

Sergei Lvovich Sobolev told me at that very time what was known and what was not known; here I give a brief account of his narrative.

Already Cauchy considered the rigidity of convex surfaces. For example, the thin shell of a convex egg persistently retains its shape as long as it has no cracks. But as soon as its integrity is broken along even a very short arc, nontrivial deformations become possible.

However, the surfaces bounding planes and rockets are not convex; for instance, attaching wings to the fuselage necessarily requires hyperbolic transition regions.

Therefore, the rigidity problem in the hyperbolic case is practically important.

The simplest model (for linear, that is, very small, deformations) is exactly the Dirichlet problem for the wave equation

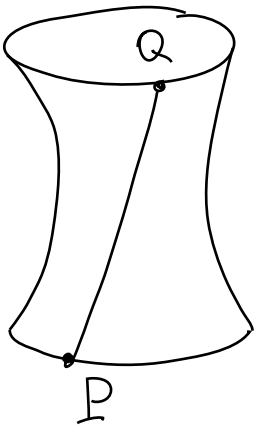
$$\frac{\partial^2 u}{\partial x\,\partial y} = 0, \qquad u(x,y) = f(x) + g(y)$$

(and its multidimensional generalizations).

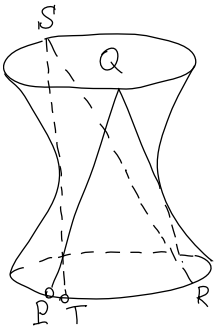
The 1943 work of Sobolev was still classified, but his paper on the so-called Sobolev equation, which generalized this work, had already appeared, and he procured me a permit to watch the corresponding experiments on the rigidity of hyperbolic surfaces at a classified department of the Institute of Mechanics.

Such surfaces were cylinders, a kind of can. I have seen hundreds of thin-wall cylinders with hyperbolic curvature of various sizes; some of them firmly retain their shape when pressed, while others really breathe in one's hands (although the difference, which amounts to 1–2%, is unnoticeable by sight).

It turned out that the following resonances are of importance. From a point P on the bottom base an asymptotic curve PQ issues (if the surface is hyperbolic, like the Shukhov tower on Shabolovka, then this asymptotic line is a straight line segment; in the Shukhov tower, it is made of steel).



Starting from the point Q on the top base along the second asymptotic line, we return to the bottom base (at the point R).



The self-map $P \to R$ of the bottom base thus arising may have a periodic point $T = P$ (for the "Shukhov tower," this corresponds to a closed polygonal chain $PQRST$ formed by rods of which the tower is made). It is this "resonance" which leads to instability (namely, the surface bends in a small neighborhood of the polygonal chain with the link-characteristics specified above).

The study of such instabilities resembled both my work on the Dirichlet problem for the wave equation and the study of resonances in planetary motions, which I took on at that time (for the sake of investigating the stability of the Solar System, where resonances are

also dangerous and cause, for example, gaps in the rings of Saturn similar to complementary intervals in the Cantor set).

Sobolev's works on resonances between oscillations of a liquid filling the thin-wall tank of a rocket and the natural oscillations of the rocket's body allowed B. I. Rabinovich to propose a method for avoiding these resonances to S. P. Korolev (it suffices to place appropriate obstructions into the fuel cans), and rockets ceased to break down.

Many years have passed since then, and I even received letters from the U.S. reproaching me for unjust praise of Sobolev (in relation to these his works on the "Sobolev equation").

Namely, contemporary American physicists (of Moscow origin, though) pointed out to me that the "Sobolev equation" was already published in 1910 by a mathematician who investigated it by the same method as Sobolev and obtained many interesting results: He wrote this equation not in connection with oscillations of fuel in the tanks of spin-stabilized rockets, but in studying meteorological peculiarities of the atmosphere of Jupiter (where a many-hundred-year-old cyclone appears as a "red spot"), for which the rotation of Jupiter is one of the basic factors.

Surprisingly enough, the Sobolev equation had already been studied by Poincaré.

To construct its theory, Sobolev invented a generalization of the Hilbert function space L_2. In his generalization, the Hermitian form was not positive definite, as in the Hilbert case; rather, it was relativistic and had one square of different sign, as for the Lorentz metric.

Investigating this question during his World War II evacuation to Kazan, Sobolev turned for advice to his neighbor, also evacuated from Moscow, who helped him. But this helper noticed: "Why such a preposterous axiom: One square is of different sign? One should immediately consider any finite number of them!"

When the neighbor wrote a paper about this generalization of the Hilbert space, he asked Sergei L'vovich to give him a precise reference to that work (with only one square of different sign), in order to insert it in his bibliography.

But Sobolev answered: "By no means, this paper is not only unpublished, it is also top-secret."

He told me about this when already fighting for declassifying this work, and now I can talk about it. But meanwhile, the Sobolev spaces received the name "Π-spaces" after the neighbor who generalized the theory of these spaces and published it (without any reference to Sobolev).

The name "Π-spaces" is after Lev Pontryagin.[1]

Although all these old studies have become classical today, I shall mention yet another question from this area; I dreamed of investigating it in the late 1950s, but, as far as I know, it is still unanswered.

Consider a smooth embedding $T^2 \subset \mathbb{R}^3$ of the torus in three-dimensional Euclidean space. Such an embedding is said to be *rigid* if any close (isometric) embedding can be obtained from it by a (small) motion of Euclidean space.

The question is whether there exist nonrigid embeddings (and of which embeddings there are more, rigid or nonrigid ones).

I have heard that the rigidity of the standard embedding of the torus of revolution (between two parallel planes tangent to this torus along circles) has been proved.

But this does not eliminate the nonrigidity of other embeddings (for instance, knotted in some way): as far as I know, this problem has not been solved even for infinitesimal deformations.

[1] [**Translator's note**] The first letter of "Pontryagin" in Russian is Π.

http://dx.doi.org/10.1090/mbk/085/39

Chapter 39

Rotation of Rigid Bodies and Hydrodynamics

Eighteenth-century sailors faced the following difficulty in determining their location on a map: Orientation required measuring the coordinates of stars on the celestial sphere at the moment of location, and these measurements could be used only if the exact time of measurement was known.

Time signals were not radio broadcast in those years, and time was kept by using chronometers. But chronometers, especially during a long voyage, tended to become very wrong. A lot of factors counted, including the ship's roll, Earth's rotation, variations of the gravitational field (which affected the natural frequency of pendulum oscillations), and even climatic conditions (tropical heat expands pendulums, while frost contracts them).

Therefore, the Admiralty of England offered a big prize for anyone who could determine time accurately. Euler invented a clever solution of this problem: to use the Moon as a clock.

At that time, people had already tried to use the motion of Jupiter's four satellites (discovered by Galileo Galilei) as a timekeeper. But this required, in addition to a good theory of the far-from-simple

motion of the satellites, a good telescope, because the "dial" of this clock is very small: Jupiter is far away, and the satellites are not always clearly visible.

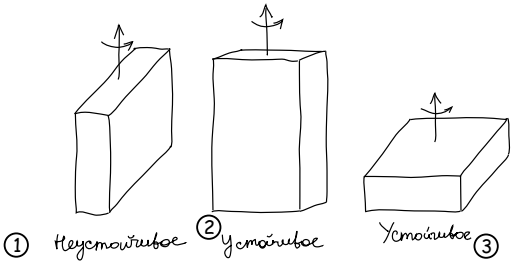
The Moon is much closer, and it is easy to observe; thus, to solve the problem, it was sufficient to construct an accurate enough theory of the Moon's small oscillations about its center of gravity (with account of the perturbations, mostly due to Earth and the Sun, and caused by the complex orbital motions of Earth around the Sun and the Moon around Earth).

It is this theory that Euler decided to create. In 1765, he published a remarkable treatise on this subject, in which he considered not only the Moon, but also the motion of any solid body around its center of gravity, mostly due to inertia and also caused by the perturbing influence of other bodies.

The notable result of Euler's study was, first of all, a complete solution of the problem about the inertial motion of any solid body around its center of gravity. This problem turned out to be a "completely integrable Hamiltonian system," and Euler found the required complete system of first integrals in involution.

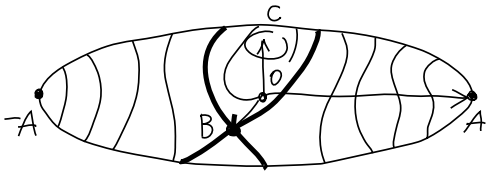
It follows from his results, for example, that there exist stationary rotations about all of the three axes of the inertia ellipsoid of a rigid body, but the rotation about the intermediate axis of inertia is unstable, while both rotations about the long and the short axis of inertia are stable.

This means that, say, a matchbox thrown while rotating about the long or short axis will keep rotating, but if it is thrown when spun about the intermediate axis, then it will chaotically somersault (I demonstrated this more than once to students during my lectures; it is better to throw a wrapped book, rather than a brick, and paint the six faces of the body to be thrown in different colors, in order that the instability be noticeable right away).



① *Unstable.* ②, ③ *Stable.*

Topologically, this difference is caused by the different behaviors of the intersection lines of the ellipsoid with the spheres centered at the origin.



Near the endpoint A of the major semi-axis OA of the ellipsoid, the distance to the center of the ellipsoid is maximal, and the lines along which this distance is slightly smaller than the length of OA, are closed curves enclosing the point of maximum A on the surface ellipsoid. Under a small deviation of the direction of the axis of rotation from OA, the corresponding vector shifts from OA to one of these closed curves near A and begins to perform small oscillations about OA, so that the motion, although ceasing to be a steady rotation, remains close to such a rotation.

Similarly, near the endpoint C of the minor semi-axis OC, the distance to the center O attains its minimum, and the lines where it only slightly exceeds the minimum distance $|OC|$ are closed curves near the point C on the surface of the ellipsoid. The corresponding perturbed rotation remains close to the steady one.

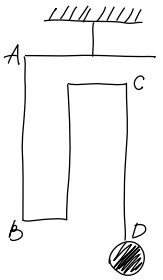
On the contrary, near the endpoint B of the intermediate semi-axis, the distance to the center O of the ellipsoid has a saddle point.

The level set where the distance is exactly $|OB|$ consists of two circles (intersecting at the point B), and each level set where the distance is close to $|OB|$ consists of two closed curves going far away from the point B (which may even nearly reach the opposite endpoint $-B$ of the intermediate axis). A perturbation of the steady rotation about the axis OB results in a "somersault" quite dissimilar to this rotation, and the body may even turn almost upside down.

At present, the Moon safely performs small oscillations, nearly always keeping the same face turned towards Earth and only slightly oscillating about this "pendulum" position. On the contrary, Earth's artificial satellites can perform all of the motions described by Euler, depending on how they are controlled, so that Euler's theory provides a basis for computations aimed at preventing satellites from somersaulting even today.

Euler's theory yields a detailed analysis of the Moon's oscillations about its usual position, so that by observing the phase of these oscillations, one can use it as a clock hand and find out the moment of observation.

However, the Admiralty did not reward Euler but rewarded a watchmaker who solved the time-keeping problem in a fundamentally different way. Namely, he proposed to suspend the pendulum AD by a three-link pendant $ABCD$.



The thermal expansion coefficient of the rods AB and CD is half that of the rod BC joining them. As a result, the thermal expansion of the rods AB and CD lowers the load D by the same distance

as the thermal expansion of the rods BC raises it. Therefore, the effective length AD of the pendulum, as well as its oscillation period, is not affected by the thermal expansion of the rods: the chronometer became insensitive to temperature changes!

Scrutinizing Euler's treatise on the Moon's rotation in 1965 on the occasion of its bicentenary, I noticed that Euler's arguments prove much more than Euler stated. Namely, his whole theory carries over, almost without changes, to the study of geodesic lines on Lie group manifolds endowed with a left- (or right-) invariant Riemannian metric.

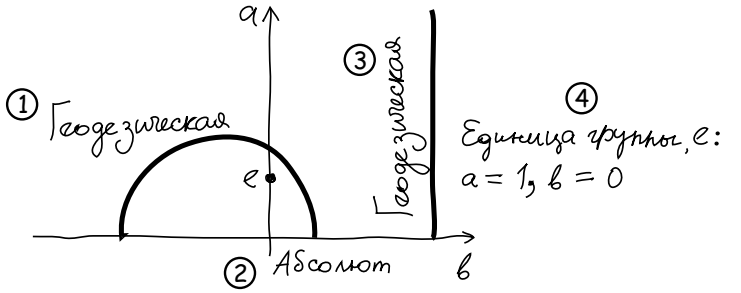
For the group SO(3) of rotations of three-dimensional Euclidean space, these geodesics are provided by Euler's study of the motion of a rigid body relative to its center of gravity. But Euler's theory can also be applied to other groups, and the conclusions suggested by its statements are not at all obvious.

As a very simple example, take the two-dimensional group of affine transformations $x \mapsto ax + b$ of the real line. Assuming the transformation to be orientation-preserving ($a > 0$), we can identify this group with the half-plane $\{a, b\colon a > 0\}$. In this case, the Euler left-invariant metric

$$ds^2 = \frac{da^2 + db^2}{a^2}$$

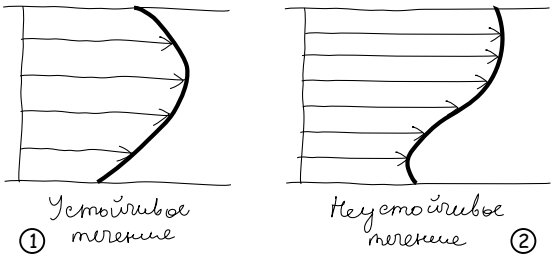
produces precisely the Poincaré model of hyperbolic geometry, so that Euler's theory becomes hyperbolic geometry. The role of Euler's steady rotations is played in this case by those straight lines and circles in the Euclidean half-plane $a > 0$ (with Cartesian coordinates (a, b), which are perpendicular to the "absolute" $a = 0$.

As a much more instructive example of an application of Euler's theory of rigid body rotations, consider the group SDiff M of "incompressible" diffeomorphisms of a manifold M (that is, diffeomorphisms $M \to M$ preserving the volume element τ of M). The geodesics of the right-invariant metric on this group are the (Euler) flows of an incompressible fluid on the manifold M.



①, ③ *Geodesic.* ② *Absolute.* ④ *The identity element of the group.*

Euler's stability theory of steady motions of rigid bodies becomes in this case a generalization of Rayleigh's theorem on the stability of two-dimensional incompressible flows with velocity profiles having no inflection points.



① *Stable flow.* ② *Unstable flow.*

In the case under consideration, flows with inflection points turn out to be similar to steady rotations of a solid body around the intermediate axis of inertia: Euler's general theorem on stability is applied in the same way in both cases, but under the passage from the three-dimensional group SO(3) to the infinite-dimensional group SDiff M, Euler's theorem becomes Rayleigh's (generalized) theorem.

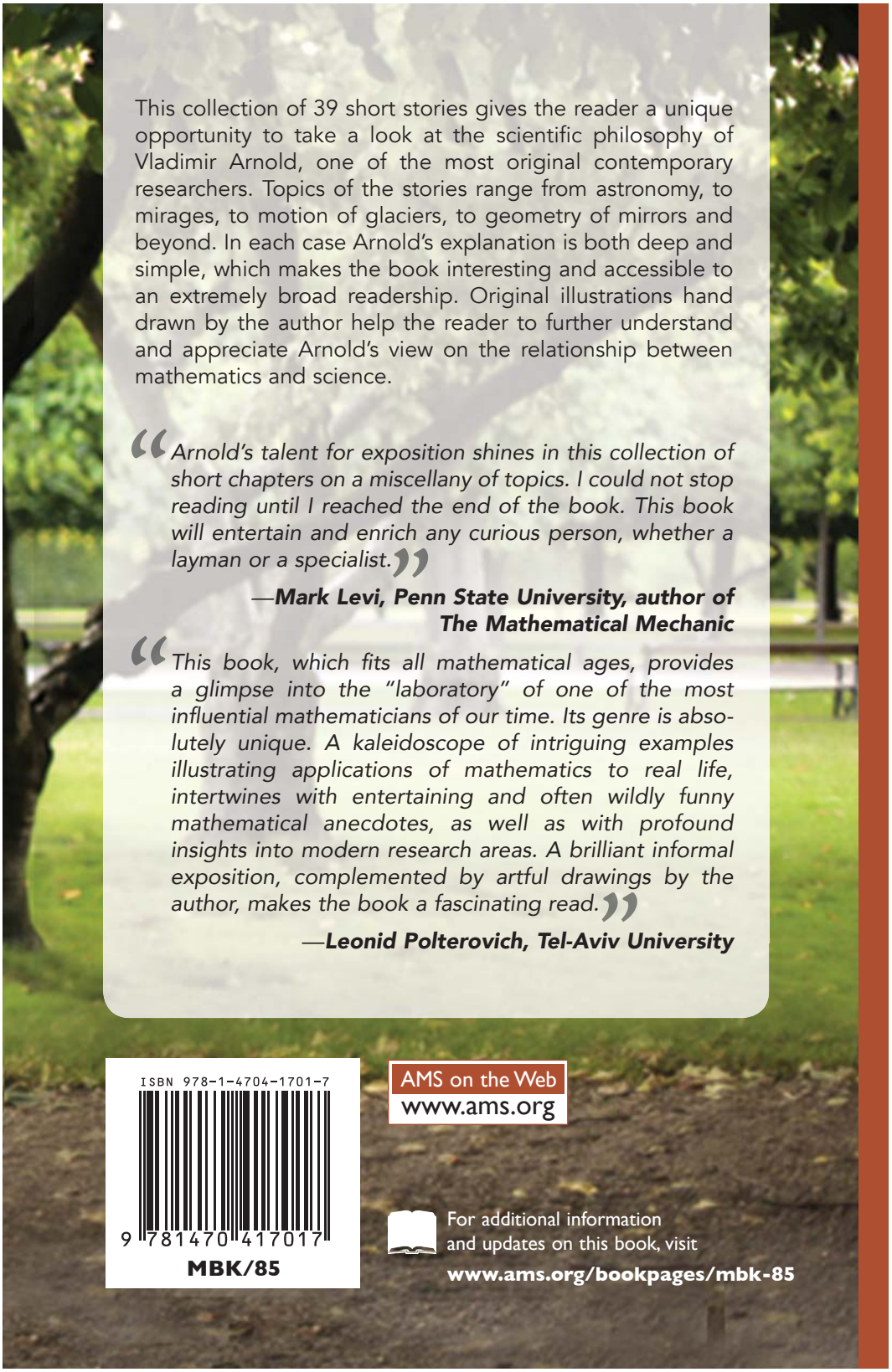
The stability of geodesics on a manifold is strongly affected by the "sectional curvatures in two-dimensional directions" of this manifold. Namely, the negativity of a curvature causes the scattering of

geodesics (with close initial conditions) at a rate exponentially depending on time. Euler's theory makes it possible to calculate these sectional curvatures (for groups with left- or right-invariant metrics).

Applying these calculations to groups of incompressible diffeomorphisms of surfaces, I obtained many two-dimensional directions of strongly negative curvature. For example, applying these estimates to two-dimensional hydrodynamics on the surface of a torus (and to trade-wind type flows), I convinced myself that initially small perturbations of the initial velocity field grow approximately by a factor of 10^5 (from a one-kilometer-wide (thunder)storm to planetary weather changes) during a time period on the order of one month.

This means that dynamically forecasting weather for time periods strongly exceeding one week will forever remain impossible, no matter how strongly computers, computational methods, and meteorological sensors recording the initial weather conditions are improved. Indeed, minute changes in the initial velocities in each cubic kilometer (even such that the mean velocities over the neighboring dozen cubic kilometers remains the same) yield new initial conditions, which sensors cannot tell apart from the old ones and which prevent a typhoon from hitting, in a couple of weeks, New Orleans, as required by the old scenario, but lead it to, say, Bombay.

It only remains to marvel at how substantial the applications of Euler's fundamental theories and ideas are, even in those cases in which Euler himself confined the exposition to the first informative case (of the group SO(3) in our example), while all far-reaching generalizations have been obtained only recently.



This collection of 39 short stories gives the reader a unique opportunity to take a look at the scientific philosophy of Vladimir Arnold, one of the most original contemporary researchers. Topics of the stories range from astronomy, to mirages, to motion of glaciers, to geometry of mirrors and beyond. In each case Arnold's explanation is both deep and simple, which makes the book interesting and accessible to an extremely broad readership. Original illustrations hand drawn by the author help the reader to further understand and appreciate Arnold's view on the relationship between mathematics and science.

"Arnold's talent for exposition shines in this collection of short chapters on a miscellany of topics. I could not stop reading until I reached the end of the book. This book will entertain and enrich any curious person, whether a layman or a specialist."

—Mark Levi, Penn State University, author of The Mathematical Mechanic

"This book, which fits all mathematical ages, provides a glimpse into the "laboratory" of one of the most influential mathematicians of our time. Its genre is absolutely unique. A kaleidoscope of intriguing examples illustrating applications of mathematics to real life, intertwines with entertaining and often wildly funny mathematical anecdotes, as well as with profound insights into modern research areas. A brilliant informal exposition, complemented by artful drawings by the author, makes the book a fascinating read."

—Leonid Polterovich, Tel-Aviv University

ISBN 978-1-4704-1701-7

9 781470 417017

MBK/85

AMS on the Web
www.ams.org

For additional information
and updates on this book, visit
www.ams.org/bookpages/mbk-85