NOTES TO INFORMATION THEORY II
From February to May 2021

Rastko Vuković
Economic Institute Banja Luka (in preparation)
May 2021

Rastko Vuković:
Notes to Information Theory II – From February to May 2021.
(Economic Institute Banja Luka, in prepraration, 2021)
https://archive.org/details/notes-ii

The first translation from Serbian:

Растко Вуковић:
Прилози теорији информације II – Од фербруара до маја 2021.
(Економски институт Бања Лука, у припреми, 2021)
https://www.scribd.com/document/493741600/Prilozi-II

Preface

The cooks cook according to the best recipes, we are judged by university-educated lawyers, mathematics professors teach due to programs written by the best, in the way prescribed of a third party. What could be wrong with striving for such a perfectly arranged world – people haven't asked me once.

In that sense, information theory is disappointing. The freedom we desire is the result of the surpluses we have in relation to the set of inanimate substances of which we are composed, and the security we hope echoes is their principled minimalism. We strive for calm by fleeing from vitality, we adhere to security against uncertainty, and we surrender our personal freedoms to the organization. It is the law of inertia, the principle of least action, or if you want the principle of economy of information. They are just different expressions for more probable occurrence of more probable outcomes.

That is what is bad in that above sentence – I would say from the point of view of information theory – that a too well-arranged system necessarily becomes obsolete. The world is inexorably changing and moving away. If all the physical phenomena of the universe consist only of information, and the essence of this is uncertainty, then escaping into certainty eventually becomes a bad job.

However, the basic thesis of this philosophy is still only a hypothesis. That is why I write locally and I hope globally, so this collection of articles is also private-public. I thank everyone who pointed out my mistakes, especially those whose remarks inspired me.

Author, February 2021.

Contents

1. Representative Sample

Rastko Vuković[1], January 30, 2021

For a given set of random events, I explain the smallest subset that could be sufficiently representative of it. I discuss briefly the connection between the Secretary's Problem, the Normal Probability Distribution, and the Golden Section, all three within Information Theory.

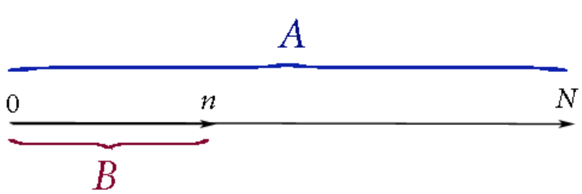
Choice of secretary

In the late 1950s and early 1960s, a simple, partially recreational task of probability theory emerged known as the problem of the secretary, or choice of partner, or dowry, that revolved around the mathematical community. It has a certain appeal, is easy to point out and has an impressive solution[2].

It was immediately taken over and developed by certain eminent probabilists and statisticians, among them Lindley (1961), Dynkin (1963), then Chow, Moriguti, Robins and Samuels (1964), then Gilbert and Mosteller (1966). Since then, the secretary problem has been expanded and generalized in many different directions, so that it can now be said that it is an area of study within mathematics-probability-optimization. From Freeman's work (1983) it can be seen how extensive and vast the area has become; moreover, it has continued to grow exponentially in the years since its text appeared.

The secretary problem in its simplest form has the following characteristics. There is one post available for the secretary. The Commission knows the number of N applicants and interviews them in random order, one by one, not knowing who the next is. The ranking of those interested is detailed enough so that there is no significant duplication of the scores of the best, and the decision on the selection is based only on previous results. After the rejection of the current candidate, it is not possible to call him/her later, and after the acceptance, the further search is suspended.

In the picture on the left is area A with a row in which the respondents are waiting in front of the room



with the Commission which interviews them individually, reviews their applications and awards points. In area B, the candidates were examined until the n-th, after which, we assume, one of the best appeared, with a winning score, which of all N could have (approximately) the highest number of points and be hired as a secretary. This is an idealized situation where B is just a large enough part of the random sequence A ($B \subset A$ and $n < N$) to be a representative sample.

So, we assume that we have one mathematical expectation shown in the figure[3]. In a given series of uniformly distributed (otherwise random) candidates, approximately every n-th is acceptable for the job, so $p = 1/n$ is the probability of finding the “right” one. This means that $q = 1 - p$, respectively

[1] Gimnazija Banja Luka, math prof.

[2] see [10]

[3] Else, this task is solved differently in the literature.

$$q = 1 - \frac{1}{n} \tag{1}$$

the probability of "wrong". In a series of n candidates (area B of the picture), everyone up to the next is "wrong", and the probability of such an event is q^n. However

$$\lim_{n\to\infty} \left(1 - \frac{1}{n}\right)^n = \frac{1}{e}, \tag{2}$$

where $e = 2.71828\ldots$ is Euler number, base of natural logarithm. In the case of large arrays ($N \to \infty$), when the scoring of the candidate is very detailed ($n \to \infty$), then $q^n \to 1/e \approx 0{,}37$, so the substring B makes about 37 percent of the array A.

The conclusion is that we can solve the problem of choosing the best secretary by missing the first 37% of candidates, simply to calibrate the top list of the best, relying on the fact that it is a good enough sample. Then we will declare the first next candidate that has the highest number of points established, or more than that (and if none of them appears, we are left with the last one) as the best choice.

Deviations

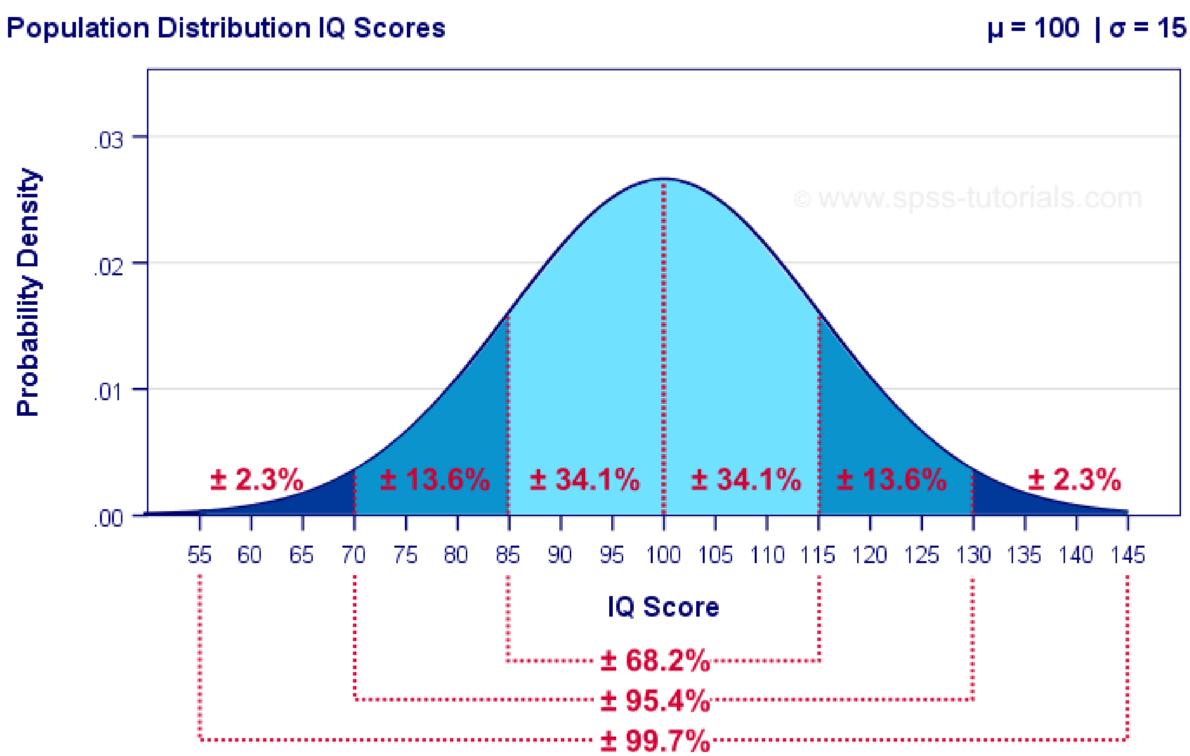
We theorize with the assumed "universe of uncertainty" whose quantities we call information. The laws of conservation and thrift apply to them, so the conclusion that free information is equivalent to physical actions, and then that living beings are physical systems that have information in excess of in relation to the inanimate substance of which they are composed.

In accordance with the principles of least action and information, living beings try to get rid of their surpluses, either directly in an inanimate environment, otherwise already filled and also prone to minimalism, or by incorporating them into the organization of the collective to which they belong. In that sense, society is a physical phenomenon somewhere between living and non-living systems. Hence, for example, the knowledge that an ant colony can have a more intelligent action than its individual ants becomes the subject of information theory.

Unpredictability is at the core of the world in general, so that the ability of society to act and choose with more of it grows, and the amount of uncertainty that we measure with information grows not only with the increase in the number of options but also with their unexpectedness. This leads to the question of the optimal measure of order and vitality of a "community of the living". Namely, regulating, organizing and limiting, is the opposite of vitality, that is, opposite to freedom and the amount of options. In the behavior of the individuals of the community themselves, the emphasis shifts to obedience and disobedience, commonness and unusualness, or passivity and aggression as opposite tendencies.

The greater the dissipation (deviation) in the behavior of individuals, the less organized the community is, and on the other hand, with the increase of compactness, its ability to choose decreases and the society becomes numb in that sense. From this consideration follows the conclusion that there is some optimum between vitality and efficiency, or risk and safety. This is where the previous "approximate third" (2) becomes important again.

In the following figure[4], we see the Normal (Gaussian) distribution of IQ (otherwise arbitrarily taken populations). The intelligence test (IQ score) is set so that the average score of individuals is 100 points and that about a third (34.1 percent) of all have "average intelligence", from 85 to 115 points. The mathematical expectation of the (so-called normal) distribution of points is thus $\mu = 100$, and the standard deviation $\sigma = 15$. Two deviations (2σ) then cover the score interval from 70 to 130, and three (3σ) almost the entire population. In that, I see a similarity with the solution to the previous problem of "choosing a secretary".



When in the first picture (secretary's choice) we consider area A as the vitality of a living being or their organization, then sub-area B is representative enough to expect (statistically) representatives of all (from routine to extreme) qualities of the given area. The second picture of the (normal) distribution of intelligence is an approximate confirmation (2) that deviations will also occur to that extent. At the same time, we do not enter into a discussion about what a particular society considers as "normal" and "deviant" behavior.

The same can be said as follows. When we rearrange those desirable surprises of the system A, which encourage its vitality more, then we will get that they make up a subsystem analogous to B and about a

[4] Taken from the "SPSS tutorials" Facebook, but it could have been from many others.

third of all. In other words, the optimum of “disobedient” individuals of the “living system” is about a third, and the differences are only in the definition of “obedience” in the type and intensity of their aggression. Hence so much diversity of living beings on Earth.

Confirmation (application) of this assessment can be found in the histories of the most successful civilizations, out of about 30 known ones. The life of society as well as civilization is interrupted by violence, suddenly, but if we look only at those who were lucky enough to last, we will notice that each consists of some rise, peak and fall. Their flows are similar to life through youth, maturity and old age, with the first part more prone to risk, and the second in routines.

In the case of civilizations, the decline begins with greater self-restraint. Thus, communism fell behind due to too much regulation, as well as dictatorship, and at the peak of the Ottoman Empire was Suleiman the Magnificent when they called him a “legislator”. It is difficult “on the ground” to measure the “amounts of restrictions” that a society has imposed on itself, for example through legislation, religious, customary or moral norms, in relation to potential “amounts of freedoms”, but from the previous we can assume that in the period of decline it was greater than 2: 1 in favor of the restrictions.

Specifically, when the ratio of total vitality of a “living being” tends to be higher than standard (2 : 1, approximately) in favor of “ordinary” versus “unusual”, then this ratio is re-established in the standard form but with a changed overall vitality of the collective and a new definition normality. In the case when the community becomes more organized, say safer and more efficient, its vitality decreases. Such is at greater risk of lagging behind an environment that would continue to evolve.

In the case of the emergence of an extreme leader who we say by his ability “drives many”, from the point of view of this mathematical certainty it is also a matter of “adjusting the community” which establishes the previous standard (2 : 1). The organization then goes to a greater or lesser vitality depending on the leadership. Not every change, be it euphoric or spontaneous, is a path to betterment or ruin, just as it is not every path that followers believe.

Seemingly a completely different kind of example can be found in today's living organisms on our planet. Greed versus empathy within individual species also stands in roughly the same 1: 2 ratio. With too many aggressive individuals, their organization would break down, and with too many insightful ones, they would become easier prey for others. The diversity of their hierarchies only confirms the theses of this discussion.

Golden ratio

The whole versus the part is treated as that part versus the rest. That is the definition of the “golden mean”. If we denote the size of the whole by 1 and the given part by φ we have the proportion $1 : \varphi = \varphi : (1-\varphi)$, and hence the quadratic equation

$$\varphi^2 + \varphi - 1 = 0, \tag{3}$$

whose solution

$$\varphi = \frac{\sqrt{5}-1}{2} \approx 0{,}62 \quad (4)$$

we call golden number. The second solution $\Phi = \frac{\sqrt{5}+1}{2} \approx 1{,}62$ is a larger golden number and it is reciprocal with the first, $\varphi\Phi = 1$. What we are also interested in here is the remainder, the difference between the whole and the golden value which is $1 - \varphi \approx 0{,}38$ or approximately (2).

Leonardo da Vinci drew his Vitruvius[5] (1487) as a man with ideal proportions, which was later speculated (by Luca Pacioli, in Divina proportione, 1509) with a golden section. Johann Kepler wrote that "the image of a man and a woman comes from the divine proportion (golden section). In my opinion, the reproduction of plants and the offspring of animals are in the same relationship".

Ernő Lendvai analyzes the musical works of Béla Bartók as if they were based on two opposing systems, the golden section and the acoustic scale, although other music scholars reject that analysis. French composer Erik Satie used the golden ratio in several of his works. The golden ratio is also evident in the organization of sections in the music of Claude Debussy (Reflets dans l'eau, 1905).

A geometric analysis of earlier research from the Great Mosque of Kairouan in 2004 (670) reveals the application of the golden section in much of the design. It is assumed that the golden ratio was used by the designers of Naqsh-e Jahan Square (1629) and the neighboring Lotfollah Mosque.

The Swiss architect Le Corbusier, known for his contribution to modern international style, focused his design philosophy on systems of harmony and proportion. His faith in the mathematical order of the universe was closely tied to the golden ratio and the Fibonacci sequence, which he described as "rhythms apparent to the eye and clear in their relations with one another. And these rhythms are at the very root of human activities. They resound in man by an organic inevitability, the same fine inevitability which causes the tracing out of the Golden Section by children, old men, savages and the learned."

The psychologist Adolf Zeising noticed that the golden section appeared in phyllotaxis (arrangement of leaves) and based on these patterns he claimed that the golden section is a universal law. In 1854, he wrote the universal orthogenetic law "striving for beauty and completeness in the realms of both nature and art". In 2010, the journal Science reported that the golden ratio was present on the atomic scale in the magnetic resonance of spins in cobalt niobate crystals. However, some claim that many obvious manifestations of the golden ratio in nature, especially in terms of animal dimensions, are fictional.

Where does so much "beauty in the golden section" come from? From the point of view of our previous considerations, we can claim that a standard relationship (representative sample and whole) was built into our emotions during evolution to make it easier to recognize the systems around us and predict their behavior. Hence, something that is (approximately) in the golden ratio intuitively feels beautiful, or harmonious. Compared to the previous one (normal distribution and the secretary problem), we find

[5] Vitruvian Man, https://en.wikipedia.org/wiki/Vitruvian_Man

that this “harmony” is actually a state of optimum, a balance in which the system does not tilt, does not strive for some revolutions.

Conclusion

Through information theory, we find that the theory of probability behind the normal distribution has some additional (among the already known) deeper roots in social and biological phenomena, as observed in the appendices [1] in the world of physics. It is also in the background of the “beauty” we see in the “golden section”.

2. The Reality of Physics

February 3, 2021

This is an easy conversation about reality and accuracy of interpretations of physical reality. A couple of them ask me privately, and I was free to present the topic publicly and rearrange the answers.

Question: “What is physical reality?”

Answer: What science considers “reality” is actually an uncertain set of fictions, which are constantly changing, supplementing and declaring unscientific.

Q: I'm seriously asking you. I read that you define “real” what can be in physical interaction with something real. Is it true again that it is “fiction”?

A: Yes, of course, our truths about physical reality are always incomplete, and therefore false. In the mathematical sense, if something is “a little incorrect”, it is “not true”.

Q: You write supposedly true and false, emphasizing that it is also suspicious?

A: Yes, but to continue. Take, for example, the flawless geometry of the ancient Greeks and their interpretation of our Milky Way galaxy by the goddess Hera who spilled milk. The first was mathematics, the second was physics. Throughout history, humanity has always tried and is trying not to combine the two, no matter how it seemed to us that physics (not only today) is mathematized.

Q: What do you mean by geometry and spilling milk?

A: The geometry would explain the mutual immobility of the parts of the “Milky Way” by the huge distances, and then by the huge sizes of the “drops” of milk. However, from ancient times until recently, even the smartest among us, but also many great connoisseurs of geometry and mathematics that developed further, gave equal preference to physical lies. They considered their (rare) colleagues who would try to destroy their faith in untruths to be hateful and apostates.

Q: Is there that today?

A: Well, I'm just telling you that it's a constant story. Until the beginning of the 20th century, physics did not believe in molecules, and then it accepted the thermodynamics of Boltzmann (after his death in 1906), then the First Law of Thermodynamics (on energy conservation), then the Second Law of Thermodynamics (on the spontaneous transfer of heat from the body to the environment of lower temperatures), so that all chemistry, biology, and even medicine would slowly become “scientific” only when the phenomena could be broken down into molecules and atoms and thus explained.

Q: And what's stuck there now?

A: When the term “substance” is set from the point of view of “energy”, then “action” (products of energy and time), then “information” – molecules and atoms will look very strange. It will deviate so much from the new “truths” that it will be considered a delusion of 20th century science. They will be like the former phlogiston, or alchemy, or the “four elements” of which the world is supposed to consist (air, water, earth, fire).

Q: Why would molecules and atoms be “unscientific”?

A: It will be, because physics has long since adopted Louis de Broglie thesis (1924) on waves of electron. All matter can be defined only by waves and all its consequences can be derived from wave equations (Schrödinger, 1926). That is why we have a particle-wave structure of matter, and that “particle” part (read my interpretation of the Compton Effect[6]) will be slowly neglected.

Q: I read (The Undamaged Crown of King Syracuse[7]), but what's the point?

A: The Compton Effect was once proof of the corpuscular nature of light (it was known to be a wave). Particles (photons) are scattering in the way described by Compton (1922) that, based on the conservation of momentum and energy, the wavelength of the rejected photon increases (therefore, Compton proved the corpuscular nature by means of wave nature). However, it further follows from the information theory (mine) that this increase in the wavelength of the reflected photon speaks of a greater blurring of the photon on the path, i.e. about the greater uncertainty of its position. It turns into less probable paths (lower probability densities) only under the action of some force, or as in this case due to a collision with another body (electron).

Q: So, what do you mean?

A: From previous misconceptions (mechanistic and materialist understanding of the world), physics has moved to atomistic, then to quantum physics. I believe that in time, it will move to the “information universe”, which, like every previous model, will be with some kind of fault. At the same time, the reality was always wrong, although it was getting wider. It has also become what is too small for our senses (molecules, atoms, quanta), or what is too far away for us as a deep universe, which, by the way, we see only as light that came to us from some distant past.

By adopting that “what interacts with the real” is real, we will also adopt “parallel realities” with which it is possible to interact only indirectly, and by adopting that “interaction” is equivalent to “communication”, it will become “real” although we can perceive it only by logic. And the question then is, aren't molecules a product of our logic more than a matter of immediate sensory perceptions?

Q: How will you explain the world “without molecules”?

A: By the law of conservation, by the principle of least action and communication. For example, I will look at the expression $S = ax + by + cz + \cdots$, which represents “perception information” when the

[6] https://www.academia.edu/40105675/Compton_Effect
[7] http://izvor.ba/rastko-vukovic-neostecena-kruna-kralja-sirakuze/

sequences $(a, b, c, \dots)$ and $(x, y, z, \dots)$ represent two opposing computer quantities. They are components of the vectors; their product is larger when the larger component is multiplied by the larger, and the product is smaller when the larger components of one series are multiplied by the smaller ones of the other, and the smaller with the larger.

In game theory, S would be the “vitality” (intensity of the game), which is higher when the opponent responds to a strong game with a strong one and to a weaker one with a weak one. In the economy, it would be a more dynamic society with good competition, and similarly in politics. In physics we would have less S, more precisely minimal, because of the law of least action; then the subject is less opposed to the stronger obstacle.

Q: More “perception information” means more liveliness and less passivity?

A: Yes, larger S belongs to “living beings” and less to inanimate matter. But we won't talk about it now, it's a broad topic of information theory, still with a lot of speculation.

Q: What are you aiming for with such “explanations”?

A: The assumption that every mathematical truth, every abstraction of it, will eventually become a kind of physical reality. I look back at what I have written many times, in different ways (see [1] and [2]).

3. Special Unitary Group

February 7, 2021

In mathematics, a special unitary group of degree n, denoted SU(n), is a Lie group[8] of type $n \times n$ unitary matrices determinant one. The group SU(2) is closely related to the group SU(3) and plays an important role in quantum physics.

SU(2)

In linear algebra, a complex quadratic matrix $\mathbf{U}$ of order $n \in \mathbb{N}$ is "unitary" if its conjugately transposed matrix $\mathbf{U}^\dagger$ is also inverse to it. In other words, if it is

$$\mathbf{U}^\dagger \mathbf{U} = \mathbf{U}\mathbf{U}^\dagger = \mathbf{I}, \tag{1}$$

where $\mathbf{I}$ is a unit matrix of the same order.

An arbitrary complex number $z = x + iy \in \mathbb{C}$ has real parameters $x, y \in \mathbb{R}$, the first of which is called its "real part", the notation $\mathrm{Re}(z)$, and the second the "imaginary part", the notation $\mathrm{Im}(z)$. For the imaginary unit, $i^2 = -1$ also holds. Conjugated to the number z is the complex number $z^* = x - iy$, so that their product is real, $z^* z = (x + iy)(x - iy) = x^2 + y^2 = |z|^2$. Associated with the matrix $\mathbf{U}$ is the (adjoint, adjugate) matrix $\mathbf{U}^\dagger$, transposed by it and with conjugated corresponding elements.

We call a "group" a structure (G, *) consisting of the set G and a binary operation * that satisfies the following four axioms:

1. (closedness) for each $a, b \in G$ the result of $a * b$ is also in G;
2. (associativity) for each $a, b, c \in G$ being $(a * b) * c = a * (b * c)$;
3. (neutral) there is $e \in G$ such that for every $a \in G$ is $e * a = a * e = a$;
4. (inverse) for every $a \in G$ there is $b \in G$ such that $a * b = b * a = e$, where e is neutral.

It can be shown that the group has exactly one neutral, that the inverse of the given element is unique, and that the left and right inverses are the same elements. When for each pair of elements $a, b \in G$ their product is commutative, $a * b = b * a$, the group is called commutative or Abelian.

If $\det \mathbf{U} = 1$, the square matrix $\mathbf{U}$ of order n is called "unimodular". The set of all complex n-th order matrices that are both unimodular and unitary forms a group if the multiplication of the matrices is taken as a group operation. This group is called the group of unitary unimodular matrices of the n-th order and is denoted by SU(n). In other words, it is a special unitary group of order n.

For quantum physics, the most interesting group SU(2) has a matrix

$$\mathbf{U} = \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix}, \tag{2}$$

[8] https://en.wikipedia.org/wiki/Lie_group

for which $\det \mathbf{U} = 1$ and $\mathbf{U}^\dagger \mathbf{U} = \mathbf{I}$, from which it follows: $u_{11}u_{22} - u_{21}u_{12} = 1$ and

$$\begin{pmatrix} u_{11}^* & u_{21}^* \\ u_{12}^* & u_{22}^* \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{3}$$

where matrix equation (3) replaces four linear equations, three of which are independent:

$$\begin{cases} u_{11}^* u_{11} + u_{21}^* u_{21} = 1 \\ u_{12}^* u_{12} + u_{22}^* u_{22} = 1. \\ u_{11}^* u_{12} + u_{21}^* u_{22} = 0 \end{cases} \tag{4}$$

From there:

$$u_{21}^* = u_{21}^* \cdot 1 = u_{21}^*(u_{11}u_{22} - u_{21}u_{12}) = u_{11}(u_{21}^* u_{22}) - (u_{21}^* u_{21})u_{12},$$

$$u_{21}^* = -(u_{11}u_{11}^* + u_{21}^* u_{21})u_{12} = -1 \cdot u_{12} = -u_{12},$$

$$u_{21}^* = -u_{12}. \tag{5}$$

Too:

$$u_{22} = 1 \cdot u_{22} = (u_{22}^* u_{11}^* - u_{12}^* u_{21}^*)u_{22} = (u_{22}^* u_{22})u_{11}^* - u_{12}^*(u_{21}^* u_{22}),$$

$$u_{22} = (u_{22}^* u_{22})u_{11}^* + u_{12}^*(u_{11}^* u_{12}) = (u_{22}^* u_{22} + u_{12}^* u_{12})u_{11}^* = 1 \cdot u_{11}^*,$$

$$u_{22} = u_{11}^*. \tag{6}$$

Therefore, each $\mathbf{U} \in SU(2)$ matrix (2) has the shape

$$\mathbf{U} = \begin{pmatrix} v & w \\ -w^* & v^* \end{pmatrix}, \quad |v|^2 + |w|^2 = 1, \tag{7}$$

i.e. it is given with two complex parameters, here $v, w \in \mathbb{C}$.

In particular, when the coefficients of the matrix (7) are real numbers it is a classical rotation

$$\mathbf{U} = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}, \tag{8}$$

for the angle φ. As it is known from elementary geometry, all symmetries, more precisely isometric transformations such as translation, reflection (central, axial, mirror) and rotation, can be reduced to rotations themselves. This is the universality of the matrix (8) and the group SU(2) in general.

Special matrices

Second-order matrices are a type of vector space whose base consists of a unit matrix (3) and three quaternions[9], linear operators whose matrices are:

$$\hat{q}_x = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \ \hat{q}_y = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \ \hat{q}_z = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}. \tag{9}$$

The relation $\hat{\sigma}_\xi = -i\hat{q}_\xi$ ($\xi = x, y, z$) defines Pauli matrices (operators), which also form a (new) base of the same space of matrices (operators) of the second order. These are also the bases of the SU(2).

Note that for Pauli operators (matrices) the equations are valid:

$$\hat{\sigma}_\xi \hat{\sigma}_\eta + \hat{\sigma}_\eta \hat{\sigma}_\xi = 2\hat{I}\delta_{\xi\eta}\,, \tag{10}$$

where $\hat{I}$ is a unit operator (matrix representations $\hat{\mathbf{I}}$), and

$$\delta_{\xi\eta} = \begin{cases} 1 & \xi = \eta \\ 0 & \xi \neq \eta \end{cases} \tag{11}$$

is the Kronecker delta symbol. The equations also apply

$$\begin{cases} \hat{\sigma}_x \hat{\sigma}_y - \hat{\sigma}_y \hat{\sigma}_x = 2i\hat{\sigma}_z \\ \hat{\sigma}_y \hat{\sigma}_z - \hat{\sigma}_z \hat{\sigma}_y = 2i\hat{\sigma}_x \\ \hat{\sigma}_z \hat{\sigma}_x - \hat{\sigma}_x \hat{\sigma}_z = 2i\hat{\sigma}_y \end{cases} \tag{12}$$

which is easy to check by directly multiplying the corresponding matrices.

Conservation law

From the needs of physics for SU(2) and quantum evolution in general, which are representations of reversible operators due to which all assumptions can be obtained from the consequences of quantum transformations, which more freely means that quantum processes remember. Hence, one of the proofs of the conservation law for the information.

Namely, the invertibility of quantum operators is the type of symmetry that Noether's theorem[10] speaks of. A system of linear equations written in the matrix $\mathbf{A} = (a_{\mu\nu})$ is

$$\mathbf{Ax} = \mathbf{y} \tag{13}$$

is regular (invertible) if using the values of the copies, the elements of the vector $\mathbf{y} = (y_1, y_2, \dots, y_n)$, we can find out the values of the original, the components of the vector $\mathbf{x} = (x_1, x_2, \dots, x_n)$. Then there exists an inverse matrix $\mathbf{A}^{-1}$ such that

---

[9] see [3], 2.4.6 Generalization

[10] see [2], 1.14 Ammy Noether

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}, \tag{14}$$

where $\mathbf{I}$ is the unit matrix. Comparing with (1), we see that such are all unitary matrices, and then that all quantum processes are such.

Mass and time

The attitude that with the increase of entropy (thermodynamics) its information decreases is not in contradiction with the spontaneous growth of entropy, if we generalize it mostly to the substance. The lost information of matter then becomes space-time information. This is in line with the assumption that space, time and matter are all made of the information.

Then space remembers, and memory as a kind of information also affects something. An example of the action of the past of space on the present is gravity, ie bodies that have (large) mass. I have written about it several times[11], and now I will compare it with the “mechanism” of the recently confirmed Higgs boson (2012) and its field. To understand what I'm comparing it to, look at the attachment[12], or at least part of it, the quote I singled out at the end.

Photons and all particles moving at the speed of light do not have their own (proper) time. They do not have a mass of rest, their own time stands still and they “borrow” time from observers. Particles that travel at the speed of light are trapped in the observer's present and in that sense have only three dimensions (two belong to information and time to the observer). Therefore, they are not able to independently penetrate through the layers of time, through anyone (from the past to the present, or through parallel realities), so by communicating with such, all their other relative subjects (observers) by them can define its present.

The principle of economy of information refers especially to those other particles that have their own time, the more they penetrate the layers of time. The time does not stand for such, and they have a mass of rest. The principle of least action, ie the least information (action and information are equivalents), slows them down, that principle (thrift) makes them inert in a way that is equivalent to “having mass”.

That part of information theory will explain the existence of the Higgs boson, which is also called the particle of God (Leon Lederman, 1993). Finally, here is the promised quote.

“… Physicists at the time (1964) were trying to understand why some particles had more mass than others (to sum the problem another way: We don’t understand why certain particles have mass; it’s believed that all force carrying particles should NOT have mass. To the contrary, we’ve learned that particles that carry weak force do have mass). We needed to know what was the driving force is behind this mechanism.

This is where Peter Higgs stepped in.

---

[11] see [1], 11. Force and Information

[12] https://futurism.com/what-is-the-higgs-field-and-higgs-boson

He was able to come up with a theory which suggested that there was an energy field that all particles in the Universe interacted with. In essence, the more massive the particle, the more it interacted with this field. Conversely, the less massive particles interacted with this field less…

The Higgs boson is the gauge boson (carrier) of the Higgs Field, just as the photon is responsible for Electromagnetic Field, the W and Z boson's are responsible for the Weak Nuclear Force, and the Gluon is responsible for the Strong Nuclear Force ....”

Hermitian matrix

The self-adjoint matrix, or “hermitian matrix”, is a complex square matrix $\mathbf{A} = (a_{\xi\eta})$ that is equal to the conjugate-transposed to itself, $\mathbf{A} = \mathbf{A}^\dagger$. In other words, a complex number that is an element of the ξ-th row and the η-th column of a given Hermitian matrix is equal to the conjugate element of the η-th row and the ξ-th column, $a_{\xi\eta} = a_{\eta\xi}^*$.

That is why the diagonal elements of these matrices must be real numbers, because they are the only ones conjugated to themselves. A square matrix with real coefficients is Hermitian only if it is symmetric. Every Hermitian matrix is normal[13], because obviously $\mathbf{A}^\dagger\mathbf{A} = \mathbf{A}\mathbf{A}^\dagger$.

It is known that the “spectral theorem” applies to finite dimensional vector spaces, which says that any Hermitian matrix can be diagonalized (mapped into diagonal) by means of a unitary matrix and that this diagonal matrix has only real coefficients. Hence, all eigenvalues of the n-dimensional Hermitian matrix are real and it has n linearly independent eigenvectors.

As only real eigenvalues in quantum mechanics can represent observable (measurable physical quantities) hermitic matrices, together with unitary ones, are the basis of quantum physics. Quantum states (particles-waves) are representations of vectors, and the processes over these vectors are representations of these operators. Vectors are superpositions of measurement outcomes, we will also say probability distributions, so the unitarity of operators preserves the unit norm of superposition, and Hermitian operators help predict observables.

Epilogue

This brief overview is, I hope, only a reference to which I can refer in the further interpretation of the “Higgs mechanism” and the application of “inertia due to time” to the masses in general. I have written much more, more extensively and in more detail about unitary and Hermitian operators before, but conciseness also has its value.

[13] normal matrix – commutes with itself conjugated-transposed

4. Big Bang

On the development of the universe from the point of view of information theory

February 9, 2021

This is a more promising version about the universe, considering the new information theory.

Introduction

Space, time and matter consist of information and its essence is uncertainty. Information is equivalent to action[14] (product of energy and time), the law of conservation applies to both, and due to the assumed uncertainty, particles communicate (interact) because they do not have everything. Consistently, we assume that multiplicity and selectivity are properties of the real world. Every subject around us has some information, and so is the universe itself.

What we can prove that can't happen – it doesn't happen, so the information is true. That is why we consider all-time truths, such as mathematical statements, to be information. If their duration is infinite, their energy is zero. In particular, the future is not "written down" in a way inaccessible to us, but is "objectively uncertain"; it arises unpredictably from infinity[15], whereby the finitudes of the world perceptions become connected with the infinite.

Expansion

In 1912, Slipher[16] noticed a redshift of distant galaxies, which was later interpreted as its moving away from Earth. In 1922, Friedmann[17] was the first to use Einstein's field equations to theoretically prove the expansion of the universe, and it is believed that Lemaître[18] came to it independently in 1927, who also calculated the speeds of galaxy distances. Lemaître's estimates were confirmed by Hubble[19] by observation two years later.

Then the "cosmological principle" is assumed, which says that all galaxies are moving away from each other. An imaginary 2-dimensional model of space is the surface of a balloon that we inflate with points representing galaxies that move away in this way. I talked to my colleagues about such official positions of cosmology so that they would eventually ask me about the attitude of "information theory" (mine) about all this. Here is my answer.

What I can tell are the speculations themselves, but there are more likely ones among them. For example, in the aforementioned text "Flows of Events" you could notice that I distinguish two groups of elementary particles, perhaps fermions and bosons, of which the former are at least a little more likely to transform into the latter. Therefore, one should consider the universe of the majority of the first

[14] see [1], 23. Action and Information
[15] see [2], 3.19 Flows of Events
[16] Vesto Slipher (1875-1969), American astronomer.
[17] Alexander Friedmann (1888-1925), Russian physicist and mathematician.
[18] Georges Lemaître (1894-1966), Belgian priest and professor of physics.
[19] Edwin Hubble (1889-1953), American astronomer.

particles, which is slightly but constantly changing into the universe of particles of the second kind. So the universe of matter becomes the universe of space.

In that sense, the spontaneous growth of entropy refers to the substance, and the information that is lost passes into space. To this unusual transition can be added the idea of a space that remembers what I wrote about earlier in various seemingly independent ways. Space grows at the expense of the substance, but it also grows with “biographies of particles” that move through it and do not grow.

The information theory I advocate may seem strange, because it offers an unusual view of the world around us, but it is not illogical. It is such that it requires some information in every free particle, and in its extreme form in every phenomenon, including memories. In order for the law of conservation to survive, for the present and the overall history of the closed physical system we observe, it is necessary to dampen the influences of the aging past on the present and make it equal to the loss of current information. It's an easy task for calculation.

That there is a loss of on-going substance, I said, stems from an increase in entropy and a corresponding decrease in information. On the other hand, the same is also a consequence of the principled economy of information transmission, i.e. more probable occurrence of more probable events, which are otherwise less informative. In other words, the present is evolving towards more likely outcomes.

A special question is where did so much of the present come from? From the point of view of Heisenberg's relations of uncertainty and then the existence of noncommutative operators from which the corresponding principle of uncertainty follows, it is not enough to imagine the future of the universe as a static warehouse of events from which we randomly choose outcomes. That would mean that certainty exists but is not available to us. It would mean that we can deceive the noncommutativity of the operators and Heisenberg's relation of uncertainty. Perhaps this consideration was exactly why I switched from a moderate form of information theory to an extreme one.

The law of conservation is valid because the perceptions (us, the subjects of the universe, the particles) are finite. For infinite sets, such a thing is not possible, because they are by definition such that they can be its own, proper subsets. From infinity, the final parts can be “torn off” indefinitely and it always remains as it was. Therefore, we can imagine that the present arises from parts of some infinity in additionally uncertain ways, and even that these infinities are the mentioned all-time truths, and then special types of information.

Since the free information (that can travel as a separate particle) is equivalent to the action (product of exchanged energy and duration), all-time information will have zero energy. In this way, we find that the very beginning of the universe, which is the “Big Bang”, is actually an unattainable moment when time flowed at infinite speed in relation to ours.

Thus, we enter the second part of this story about the universe, which may seem like a special version of its origin and spread, but which I would not separate from the previous one, at least for now.

Relativity

From the point of view of any (average) moment of the past, we can consider that the time of our present is flowing more and more slowly, for example, as if we are falling[20] into an ever stronger gravitational field. Conversely, from the present point of view, the passage of time of the increasingly older past would seem increasingly faster to us.

In that sense, a traveler who would go back in time with a time machine could need an infinite amount of his own (proper) time to get to the beginning of everything, to the time of the "Big Bang". Together, a relative observer from our (current) present could estimate the duration of the traveler's journey by only 13.8 billion years, or as long as we consider the universe to be old.

Because the relative time of our present is flowing more and more slowly, and radially from us at growing up distances, the lengths seem increasing, we observe the distancing of galaxies. In this way, we could observe the going away of galaxies even if they are static. However, taking into account the relativistic effects, we could get that they move away just as fast as it is necessary to cancel the relatively faster flow of their time in relation to ours and relatively larger units of length.

Epilogue

It is amazing how inspiring this seemingly innocuous version of the universe was for the interlocutors, for connecting and inventing various scenarios of science fiction stories, and believes me or not, it also has mathematically interesting sequels. But about when the time comes.

[20] I say to note that the laws of physics and especially the laws of conservation can remain

5. Dwarf Galaxies

February 14, 2021

What is your opinion about dwarf galaxies that might be without "dark matter" in their entourage? – It is one of the questions I get from colleagues regarding new cosmological research. The ones about "gravity through time", which are gaining in importance with new knowledge, and give my earlier speculative answer on weight, are the topic of this story.

Small and large mass

I have a positive expectation about finding weaker masses in space[21] that would not be accompanied by dark matter. It comes to information theory as one of the additional hypotheses worth considering, which follows from the following three settings. The fourth is the explanation.

First, conics (ellipses, parabolas, hyperbolas) are trajectories of motion caused by a constant central force if and only if that force decreases with the square of the distance. This is a theorem that I also proved (see [1]). Then the statement follows that the field of force expands at the speed of light only if the central force decreases with the square of the distance. It is known that the Coulomb force (electromagnetic) has both properties.

Secondly, these are the planets of the solar system that move in ellipses, all except Mercury. Mercury seems to be pulling its ellipse; its perihelion rotates following the direction of Mercury's rotation, which means that its exact path is not exactly an ellipse. This "deviation" is predicted by the general theory of relativity and is interpreted by the proximity of the Sun, that is, by the strong gravitational field torsions.

Third, our information universe has three spatial dimensions and three temporal ones. When we take four of these six (perhaps any) and declare three spatial (x_1, x_2, x_3) and the fourth temporal $(x_4 = ict)$ – the same[22] Einstein's[23] equations of general relativity will apply as Klein-Gordon's (relativistic) quantum mechanics. From this we draw the conclusion that a force acting through space could also act through time (say from the past to the present, and perhaps vice versa) if the carriers of that force (in this case gravitational waves, then their elementary particles – gravitons) had their own (proper) duration.

Fourth, the light (photons – particles of electromagnetic radiation) do not have a rest mass, time stands still. They therefore exist in only three dimensions; say in the planes of their information and the time of the observer. From the upper (first) gravitons also move at the speed of light, except in the vicinity of a strong gravitational field. Hence, a strong gravitational field acts through its own time too, but a weakly is not.

Dwarf galaxies, therefore, will not have their "dark matter" as their companion from the past, if their mass is insufficient for a longer time. Massive galaxies, on the other hand, will leave a trail in the past that will attract them (the action of the past on the present), just as Mercury is more attracted to its

[21] Like: https://phys.org/news/2020-09-physicists-mysterious-dark-deficiency-galaxy.html
[22] see [2], 3.30 Delayed Gravity
[23] Albert Einstein (1879-1955), German-born theoretical physicist.

younger positions that it finds in front of it compared to the older ones behind (stronger action of the closer past to the present), and would make it (the dark matter) to orbit around the present of these galaxies like other bodies (I do not exclude and the possibility of the action of the present on the past).

Penetration through time

Where can I read about the action of gravity over time? – Another interesting question to me recently, not by the same colleague but on a similar topic. It also refers to an earlier hypothesis, derived from an older and as yet unidentified "information theory", in various forms and with different consequences.

For example, read the appendix "3.30 Delayed Gravity" from popular information stories [2], I said. If it seems good, look at more complex representations (in the formulas of general relativity and quantum mechanics). In a short retelling, to see if it was worth reading, the conversation went something like this.

The point is on Einstein's field equations

$$G_{jk} + \Lambda g_{jk} = \kappa T_{jk} \tag{1}$$

where three spatial coordinates can be $x_1 = x$, $x_2 = y$, $x_3 = z$, and the fourth the time is $x_4 = ict$ with imaginary unit ($i^2 = -1$), speed of light (approximately c=300 000 km/s) and duration t, actually the length that light travels in the given time. Another important factor is the Klein-Gordon equation

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} - \frac{\partial^2}{c^2 \partial t^2} - \mu^2\right)\phi = 0 \tag{2}$$

of the quantum mechanics. Here $\phi = \phi(x_1, x_2, x_3, x_4)$ is a (pseudo) scalar function, in the general case complex. When m is the rest mass of the particle, and $h = 6{,}626 \times 10^{-34}$ Js Planck's constant, then we can write this equation with $\mu = 2\pi mc/h$. If ϕ is a real function, the Klein-Gordon equation describes neutral (pseudo) scalar particles, and if ϕ is complex it (2) describes charged particles.

Information theory is needed for a deeper understanding of the symmetry of spatial and temporal coordinates. As a starting point, I suggest you read a short popular text "2.13 Space and Time" from "Information Stories" [2]. This 6D space-time universe should not be viewed as "three spatial and three temporal" coordinates, if it is possible to choose any four of them six, and declare three as "spatial" and the rest one as "temporal".

When we consider one of the lengths as the product of an imaginary unit, the speed of light and duration ($x_4 = ict$), it thus becomes temporal. Due to the first of its factors, "time length" gets a new reality by squaring and appears in the denominator of the gravitational force (which decreases with the square of the distance). For the second, in real time (of our order of magnitudes) that square in the denominator makes the fraction terribly small (a very small number), which approaching to zero quickly and too fast every second. That is why this aspect of gravitational force is difficult to register, and then even more difficult because it occurs only in very strong fields (in the immediate vicinity of the Sun and stronger ones), in which Mercury is barely located.

When asked why Mercury, I say that its orbit is not an exact ellipse, because the perihelion of that "ellipse" is retreating behind the planet, which can now be interpreted as a stronger action closer to its own past. Other celestial bodies farther from the Sun do not have that retreat, and all of that together supports the theses about the non-action of weak gravity through time.

To answer the question of where this (non) action came from, I said that more of the information theory is needed. More precisely, we need the "principle of minimalism" of information, or more loosely, the "principle of least action" (I consider information and action to be formal equivalents) known in physics.

This principled minimalism is the cause of inertia. Mass has its proper (own) time, unlike light, which means that it personally penetrates through the layers of time, and then "stumbles" due to the mentioned principle. In order for gravitational waves (gravitons) to have such penetration, they must move (at least a little) slower than light, and this could only happen within a very strong gravitational field.

Namely, if the waves of the field (constant central) force travel at the speed of light then and only then the force decreases with the square of the distance, and if the force decreases with the square of the distance then and only then the trajectories are forced to be conical. However, we have a deviation from the conic (ellipse) in the case of strong gravity, but not in the case of weak gravity, which means that only gravitons of strong fields travel slower than light and have mass.

Please note, these are still just (my) hypotheses from before, regardless of what recent discoveries in cosmology add to their significance.

Gravitational waves

Gravitational waves are space-time curvature disorders generated by accelerated masses. They were proposed by Poincaré[24] in 1905, and later in 1916 by Albert Einstein on the basis of his general theory of relativity. They transmit energy as gravitational radiation, which Newton's law of universal gravitation does not predict, because in classical mechanics it is based on the assumption that physical interactions propagate instantaneously, at infinite speed. They were experimentally measured directly for the first time on September 14, 2015 as part of the LIGO (Laser Interferometer Gravitational-wave Observatory) project.

As the gravitational wave (at the speed of light) passes by the observer, space-time strains and distorts. The distance between objects rhythmically increases and decreases as the wave passes, and this effect fades with distance. This is a recognized explanation of this phenomenon.

It is predicted that binary neutron stars as they merge, due to the very large acceleration of their masses, can be a powerful source of gravitational waves. Due to the astronomical distance from these sources, the effects measured on Earth become very small, with wrinkles less than 1 to 10^{20}, but they are measured with even more sensitive detectors (accuracy up to 5×10^{22} parts).

[24] Jules Henri Poincaré (1954-1912), French mathematician and theoretical physicist.

They make it possible to observe the merging of black holes and other exotic deep space objects inaccessible by traditional means, optical or radio telescopes. They could be useful to cosmologists for observing a very early universe, before recombination (an era when charged electrons and protons first became attached to electro-neutral hydrogen atoms) when space was opaque to electromagnetic radiation. Accurate measurements of gravitational waves also help in additional testing of the general theory of relativity.

These are familiar things. But in our previous context, when gravitational waves move at the speed of light only in areas of “weak gravity” (solar system), while in areas of “strong gravity” they move (slightly) more slowly, they resemble the reverse movement of sea waves traveling on the surface above great depths when reaching the shore. The lower part of the sea wave that enters the shallows starts to get stuck on the bottom and the wave slows down. Inwardly, it is reminiscent of the process of a gravitational wave which, leaving a strong field where it “gets stuck in time”, accelerates to the speed of light.

I note that the usual “proof” of speed of light of the gravitational waves I consider naive. That which says that the speed of these waves is equal to the speed of light, because in the event of the sudden disappearance of the Sun, the gravity of this star would still act as long as we see its light – for information which travels at the speed of light. This would then also apply to the sudden disappearance of the sound source, because (alleged) information travels at the speed of sound! It is clear, I guess, why I consider this ”method of proving“ wrong.

Instead, I repeat, the gravitational waves move at the speed of light where they cause the movements of celestial bodies along conical trajectories and vice versa, where those trajectories would not be conical their speed would not be the same. This opens up some very interesting possibilities, but which are no longer a matter of popular retelling (for now), so we'll talk about that another time when I check, understand and arrange the formulas.

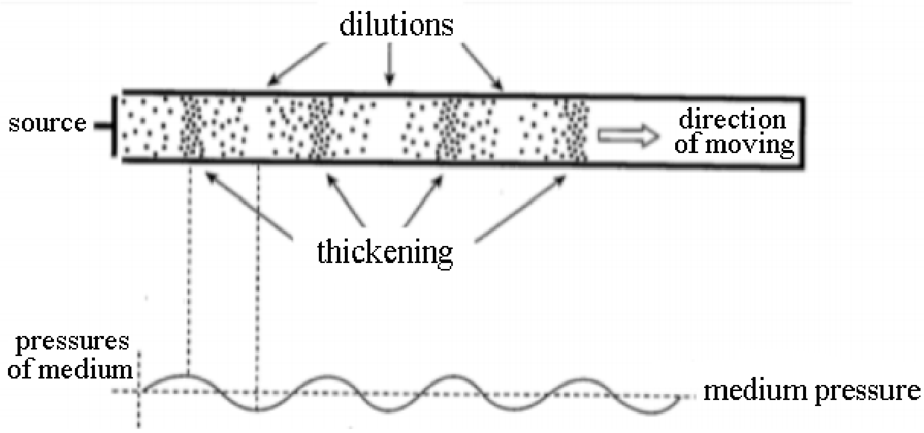
6. Waves

February 18, 2021

Summary of sound, water and boson waves, with some new speculations about the influence of force field waves especially electromagnetic and gravitational.

Sound

Periodic compression and expansion of a substance in a certain direction defines a longitudinal sound wave. Its direction is from the source to the middle of the oscillation, as seen in the picture on the left.



This vibration energy is normally transmitted through space at speed

$$v = \frac{\lambda}{T} = \lambda f$$

Where λ and T are respectively the wave-length and the period of local oscillations, dilutions and thickening (in the figure). The frequency $f = 1/T$ is the reciprocal of the period. These local pressures are shown on the bottom line in the form of a sinusoid.

From Hooke's[25] law of elastic force ($F = -kx$) and Newton's[26] second law ($F = m\ddot{x}$), and considering the acceleration $\ddot{x} = d^2x/dt^2$ of mass m, we obtain the differential equation

$$m\frac{d^2x}{dt^2} + kx = 0, \tag{1}$$

movements of the abscissa x during time t. The solution of this equation (1) is a sinusoid

$$x(t) = A\sin(\omega t), \tag{2}$$

where A is the amplitude (of the oscillations) and $\omega = f/2\pi$ so-called circular frequency.

The circular frequency defines the number of dilution and thickening cycles per unit time, and the amplitude of such waves is equal to the wavelength ($A = \lambda$), so for the kinetic energy of oscillation of a particle of mass m we can write

$$E_k = \frac{mv^2}{2} = \frac{mA^2\omega^2}{2}. \tag{3}$$

[25] Robert Hooke (1635-1703), English scientist.
[26] Isaac Newton (1642-1727), English mathematician and physicist.

The sound source by compresses do work $W = \frac{1}{2}Fx = -\frac{1}{2}kx^2 = -\frac{1}{2}kA^2$, and this elastic state then turns into stretching and transfers energy (3) from which we get $\omega^2 = k/m$. The sound lasts as the source consumes energy while working and the transmitted energy of the oscillations is replenished.

The amplitude of a sound wave determines what we perceive as loudness. Sound intensity, I, is the average speed of energy transfer per unit area perpendicular to the direction of wave propagation, so

$$I = \frac{A^2}{2\rho v}, \tag{4}$$

where ρ is the density of air (kilogram per cubic meter), and v is the speed of sound. Intensity is the amount of energy emitted by a sound source in one second through an area of one square meter perpendicular to the direction of propagation and is measured in watts per square meter.

Due to the analogy with action, or information, we can write that the volume (intensity) felt by the human ear is proportional to $L = \log\frac{I}{I_0}$, where in the numerator (numerus logarithm) is the intensity (4), and the denominator is the audibility threshold. The volume increases with the density ρ of the medium by which the sound is transmitted and with the amplitude A. However, in spreading in all directions, spherically, the density of sound energy decreases with the square of the radius, distance from the source, and thus the volume decreases.

Water wave

Waves propagate through the surface of water due to its tension, gravity and the forces of restoration (which pull the body into equilibrium). In the picture on the right, there is a surface disturbance that spreads in concentric circles, oscillating transversely (perpendicularly) to the wave directions with noticeably smaller amplitudes and slightly changed wavelengths.

Again, the angular frequency is $\omega = 2\pi/T$, where T is the period of oscillation, the wave number $k = 2\pi/\lambda$, where λ is the wavelength, so putting $\theta = kx - \omega t$, means the angle in radians, where x is the path of the wave during t, we have

$$\eta = A\sin\theta, \tag{5}$$

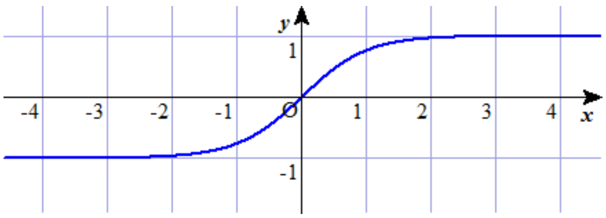
a sinusoid representing wave propagation. The amplitude A is now perpendicular to the wave propagation direction, and the (phase) wave velocity is $v = \omega/k$.

You can find further performances of the properties of water waves in various places[27], and here I will just retell some results. One way or another[28], for the square of the velocity v of waves on waters of various depths h and gravitational acceleration $g = 9.8\ \mathrm{m/s^2}$, we usually find approximately

$$v^2 = \frac{g\lambda}{2\pi}\tanh\frac{2\pi h}{\lambda}. \tag{6}$$

The following expressions apply to the hyperbolic tangent[29]:

$$\tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}} = x - \frac{x^3}{3} + \frac{2x^5}{12} - \frac{17x^7}{315} + \cdots, \quad \text{for } |x| < \frac{\pi}{2}, \tag{7}$$

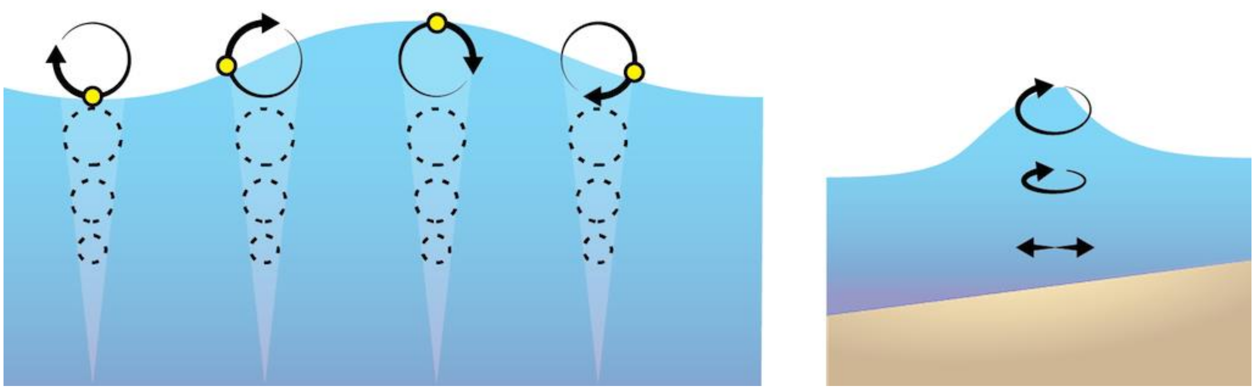


and the graph of that function is in the picture on the left. As can be seen, the hyperbolic tangent is an increasing function, from -1 to +1, so it stands as the fraction coefficient in formula (6) where it achieves a reduction of the first factor, the main expression of velocity.

In waters of great depth in relation to the wavelength λ, for the square of the wave velocity can be obtained and

$$v^2 = \frac{g\lambda}{2\pi} + \frac{2\pi\gamma}{\lambda\rho}, \tag{8}$$

where γ is the surface tension and ρ is the density of water. These differences of (approximate) expressions for wave velocity arise due to different influences that we can ignore and the mechanics of water waves at different depths. In both of the following images, the flow of the wave is from left to right and both show the circulation of water particles. On the left is a wave in deep waters and on the right in shallow ones where the destruction of that circulation can be seen.



[27] Coastal Wiki, http://www.coastalwiki.org/wiki/Main_Page
[28] Wawes in Water, http://web.mit.edu/1.138j/www/material/chap-4.pdf
[29] The Serbian expression is also th x.

Formula (8) is more interesting to us in the following three cases: the speed of the waves: when it comes to the ripple of water in the shallows, when we have long waves in deep water and when long waves reach the shore. Then additional approximations are useful.

The first is the case of small water waves whose speed depends on the wavelength λ and is dominated by surface tension forces that move these waves along. There are also gravitational forces on these small humps of water, but they are negligible. For such, the speed is approximate

$$v = \sqrt{\frac{2\pi\gamma}{\lambda\rho}}. \tag{9}$$

Thus, in ripples, water of shorter wavelengths, the second collection (8) dominates.

In the second case, for long waves in deep water, the first term (8) is more important and the speed is approximately

$$v = \sqrt{\frac{g\lambda}{2\pi}}. \tag{10}$$

In deep water the surface tension γ is too small to be important. The density ρ is also irrelevant, because when it increases – the force acting and the mass moving the wave grow together, without affecting the response time of the water in front of the wave front.

In shallows, when the wavelength is much greater than the depth of the water, and this much greater than the amplitude ($\lambda \gg h \gg A$), the speed of the wave is approximately

$$v = \sqrt{gh}. \tag{11}$$

Bores[30] are a special case of shallow water waves. The bore can be easily made in a long narrow trough of water by sweeping the water at a steady speed using a wide paddle.

The energy of water waves is provided by the kinetic energy of the wind, mostly. According to (6), with higher energy input, higher velocities and higher wavelengths occur, and further, higher wave velocity is affected by higher depth, because the kinetic energy

$$E_k = \frac{mv^2}{2} = \frac{mg\lambda}{4\pi}\tanh\frac{2\pi h}{\lambda}, \tag{12}$$

where m is the mass of water in the wave.

Waves of higher energy will push those with less energy, on average, so it can happen that sea currents go with their own ways, and that surface waves go from greater depths to smaller ones, slowing down. This is especially true where the depth h is not too great (see the graph of the hyperbolic tangent) and the water suddenly becomes shallower.

[30] Tricker, R. A. R. (1964), Bores, Breakers, Waves and Wakes or Barber, N. F. & Whey, G. (1969), Water Waves.

Wave refraction

It is a general phenomenon that waves turn from environments of higher speeds to environments in which they move more slowly. In the case of water, we understood that waves of higher energy push those with less energy[31] and that from greater depths they tend to smaller ones, so we can expect them to turn towards environments where their speed would be lower.

In a more general case, we could refer to the well-known "least action principle" and confirm the above "energy reason" along the way. Physical action is a product of energy and time, so situations of equal time will spontaneously lead to the unfolding of lower energies and, when it comes to kinetic energy – to lower speeds.

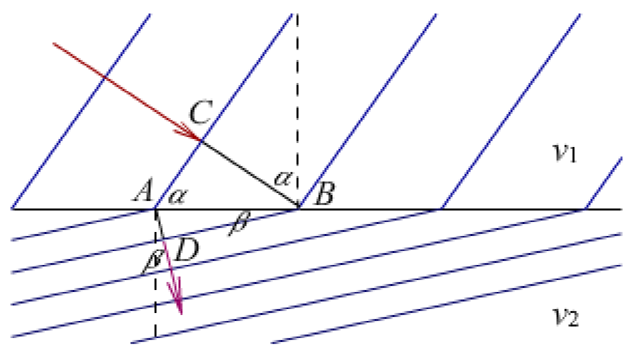
However, the classical ways of determining the refraction of waves at the boundary of the midpoints of their different velocities, v_1 and v_2, are mainly reduced to geometric and common analytical methods of proving Snell's law:

$$v_1 : v_2 = \sin\alpha : \sin\beta . \tag{13}$$

The angles α and β are the inclination of the directions of wave motion towards the normal to the boundary of the midpoints before and after the deflection. I have dealt with similar proves[32] and I will demonstrate something similar here.

In the figure on the right, the assumed wave velocities are $v_1 > v_2$, so from (13) follows $\sin\alpha > \sin\beta$, and hence $\alpha > \beta$. So says Snell's law, which we shall now prove.

In the case of light, when, as in the given picture, $v_1 > v_2$, where in the lower medium, which we call "optically denser", its movement is slower.



From the upper medium, from the direction $\overrightarrow{CB}$ with the speed v_1 the wave reaches the boundary line AB and continues the movement in the lower medium with the direction $\overrightarrow{AD}$ with the speed v_2. The normal on the horizontal boundary are two dashed vertical (mutually parallel) lines, with which the directions of motion, $\overrightarrow{CB}$ and $\overrightarrow{AD}$, form angles α and β respectively.

Long parallel lines, such as AC or BD, represent wave fronts, with spacing between parallels corresponding to wavelengths. As the figure shows, the refraction of a wave from a larger to a shorter wavelength and the same frequency, according to the formula $v = \lambda f$ the wave passes into the medium where it has a lower speed.

---

[31] unofficial interpretation

[32] Р. Вуковић: Преламање таласа (20. januar 2017), https://www.academia.edu/31013581/

The acute angle at the vertex A of the right triangle ABC is also α, and the acute angle at the vertex B of the right triangle BAD is β – because the angles with mutually perpendicular arms are equal (or are supplementary, which is obviously not the case here). However, the same wave sweeps the same path AB both as the upper and as the lower, at the same time t, where when the upper crosses the path $\overline{CB}$ with speed v_1 then the lower crosses $\overline{AD}$ with speed v_2. From:

$$\overline{CB} = \overline{AB} \cdot \sin\alpha, \quad \overline{AD} = \overline{AB} \cdot \sin\beta,$$

$$\overline{CB} : \overline{AD} = \sin\alpha : \sin\beta,$$

$$v_1 t : v_2 t = \sin\alpha : \sin\beta,$$

and after reduction the time t we get the Snell's law (13).

That Snell's law can be obtained on the basis of the principle of least time consumption, the wave that passes from the upper to the lower medium, which you can see in the same my article. This evidence should be distinguished from, only mentioned here, by the principle of least action, but it should also be distinguished from the "energy reason" with which this subtitle was started.

Boson waves

Unlike fermions (e.g. electrons, protons, neutrons, muons), bosons (e.g. photons, gluons, W and Z, gravitons) have an integer spin. The Pauli principle[33] does not apply to them either, which says that two identical fermions cannot be found simultaneously in the same quantum state. Bosons are therefore not pushed as sound-transmitting molecules and are generally pickier in interactions.

Like light, bosons interfere but do not communicate directly. By interference, bosons can increase visibility (observability), but they do not suppress each other or directly exchange energy in the way of a wave of a substance.

For example, the electric field of an electron induces a magnetic and vice versa by defining concentric spheres around the electron and vibration which we call virtual photons. As the radius of the sphere increases, its amplitude and the probability of the action of the corresponding virtual photon on a possible second charge decreases when the wavelength and momentum remain unchanged.

If the spin of a given electron is $+\frac{1}{2}$, the spin of a virtual photon is $+1$, and the spin of another electron is $-\frac{1}{2}$, then an interaction can occur, which causes changes in spins, momentums and energies. By exchanging, the virtual photon becomes real, and the given second electron also receives a spin of $-\frac{1}{2}$ and $+\frac{1}{2}$, respectively. They bounce due to the transmission of momentums, like two boats on the water when we throw a bag of sand from one to the other. The spherical shape of virtual photons corresponds better to Coulomb's law than the linear of the Feynman diagrams.

[33] [2], 3.11 Pauli Principle

In the following, a given electron, now spin $-\frac{1}{2}$, can emit as real only photons of spin -1 that could react only with the electron spin $+\frac{1}{2}$, changing its spin to $-\frac{1}{2}$, and the spin of the given electron to $+\frac{1}{2}$. This returns the spin situation to the previous one, with additional electron repulsion with each new momentum exchange. It is vaguely whether all virtual photons of one electron have the same spin, the same wavelength and momentum, or whether their emissions are also random with all the more limited possibilities of interaction.

The reverse course of time of some particles was seriously considered by some founders of quantum mechanics, such as Dirac with positron (1928), and this idea is still considered open by many. What we can add, according to information theory, is that we are in a 6D universe of parallel realities. The possibilities are a continuum of many, but their realizations cannot be more than countless infinitely many. Following this strange plot, the idea of time inversion becomes even stranger.

I remind you that due to the assumption of the universality of information, and uncertainty as its essence, the universe of possibilities is not a static warehouse of 6D events through which we randomly travel in the 4D world of reality. That this “journey” has additional profound unpredictability is dictated to us, among other things, by Bell's theorem[34] (1964) according to which the very idea of causality believers about the "hidden parameters" of quantum mechanics is contradictory, even if such “insufficiently causal” causes were not approachable to us in any case.

In so many coincidences[35], the old idea of the reverse flow of time is less vulnerable than usual. The particles of the opposite flow of time that we see exist in countless realities around and wherever we go we come across their realizations. In other words, the positron of our world is part of an uninterrupted series of positrons that we encounter in whatever future we turn to.

Additionally, the probability of randomly selecting the same particle from a continuum twice is zero. Moreover, there is a zero probability that we will re-select the same particle from countless infinitely many attempts – from the continuum of possibilities. So many times the continuum is greater than countable infinity that even “returning” to a given event of the past would be an incredible event. On the other hand, if something is already being chosen, it must be chosen.

The basis of this story is in set theory. Countably infinite sets are natural, integer and rational numbers. The set of real numbers is innumerably infinite – the continuum infinite. The chance of randomly extracting the previously mentioned rational number from the continuum of real numbers is zero, and the probability of hitting it at least once in (countable) infinitely many repeated attempts is zero. But with each attempt we will draw some number.

Positrons, like other elementary particles, are just so dumb (impersonal) that we don't even notice their differences from some future there, or parallel realities. That is why this idea (of opposite currents of time) is as resistant to criticism as it is fantastic.

---

[34] [2], 2.23 EPR Paradox

[35] [2], 2.15 Dimensions of Time

Therefore, thinking consistently further, for example about the electromagnetic field and its photons, in the supposed way we engage in speculations about the future and the past. The two electrons have the same flow of time and the photons that mediate their communication define their comparative present. On the contrary, an electron with an elementary particle of opposite charge (positron) has the opposite course of time and the previous reason for the repulsion of two electrons becomes the explanation for the attraction of the electron and positron.

A different speculation[36] will lead us to similar results. Let's say it's based on the assumption that the speed of light becomes just as smaller as the universe gets bigger. Due to the limited speed of light, electrons are always seen with a time delay; looking at each other's past the older they get farther, so trying to get more away they actually turn into an optically denser medium. For now, I cannot say for sure that such an idea will not become part of a broader theory with the previous one, as I do not single it out in a separate story. Also, I do not want to pay more attention to it for now.

The next and last of today's speculations about boson waves is just as strange as the previous two and, at least at first glance, just as independent of them. It is especially interesting in the case of graviton[37], which should have a spin of $+2$ and such be completely unusable for interactions with, say, electrons (or any fermions of the spin $\frac{1}{2}$). By merging such, none of them would be formed, because the resulting particle would have to have a spin of 1.5 or 2.5 due to the law of conservation of spin, which makes gravity a macro phenomenon.

A well-known lesson of the theory of relativity is that mass defines the geometry of space and vice versa. The fact that gravity is a universal phenomenon (in the macro world) only obscures one fact that we are now discovering, that each individual field of gauge bosons (those that define a field of forces) could be attributed a special metric that would apply to its charge-reactive particles.

The “space-time” metric of the electromagnetic field that make the Coulomb force would be such that the geodesics of the electric charge would be trajectories corresponding to the movements due to that force, and in the case of a combination of different forces we would calculate the resultants.

This idea is based on the well-known views of higher algebra and functional analysis. Hence, we know that on the (same) vector space, different normed can be defined, and on these, different appropriate metrics can be obtained, and then the different metric spaces can be obtained too, as well as vice versa. Such an addition to the theory of forces would explain reality to us at least as much as it would initially complicate the previously known calculations of motion. But it would not contradict correct theories because neither algebra nor analysis does it.

Finally, we notice that the second of the speculations mentioned here also applies (perhaps) to gravity. Namely, if the speed of gravitational waves in a strong gravitational field is at least a little lower than in

[36] Big Bang, https://www.academia.edu/45088060/Big_Bang
[37] [2], 3.27 Graviton

a weaker one[38], then the masses could be "deceived" and turned towards the "optically denser" medium, because they are waves themselves, and gravity is a universal macro phenomenon.

[38] as discussed in [1] or Dwarf Galaxies, https://www.academia.edu/45118779/Dwarf_Galaxies

7. Size of Cosmos

February 25, 2021

On modern cosmology and from the point of view of future information theory.

Introduction

The cosmos is an unlimited[39] space around us. The second name is universe. It is considered to be built from space, time and matter, perhaps from ideas such as mathematics [11], or from information whose essence is uncertainty [1].

With the naked eye we can see about 5,000 stars, glowing celestial bodies similar to the Sun. The distances between the stars are huge and are measured in light years (ly), roads approximately 9.6 billion kilometers long that light travels in a year at a speed of $c = 300$ thousand kilometers per second. By the way, the astronomical unit (au) is the average distance of the Earth from the Sun (about 149,600 thousand km). Parsec (1 pc, about 206 thousand au) is the distance from which one au is seen at an angle of one arc second, and mega parsec (Mpc) is one million parsecs.

We see 88 constellations in the sky (Big and Little Bear, Scorpio, Sagittarius and others), but we group the stars more naturally into clusters (constellations) which are parts of galaxies, and these again form their clusters.

The sun is on the edge of one of the spirals of our Galaxy called the Milky Way. We are about 26 thousand light-years from the center of the Milky Way whose diameter is estimated at 100 to 180 thousand ly with 100 to 400 billion stars and at least as many planets. The matter of the Galaxy in its wider scope orbits at a speed of about 220 km/s with a uniformity that is not in accordance with Kepler's[40] laws. With stars around us that are close to 13.8 billion years old, as much as the universe itself, we are moving at a speed of 600 km/s in relation to extragalactic references.

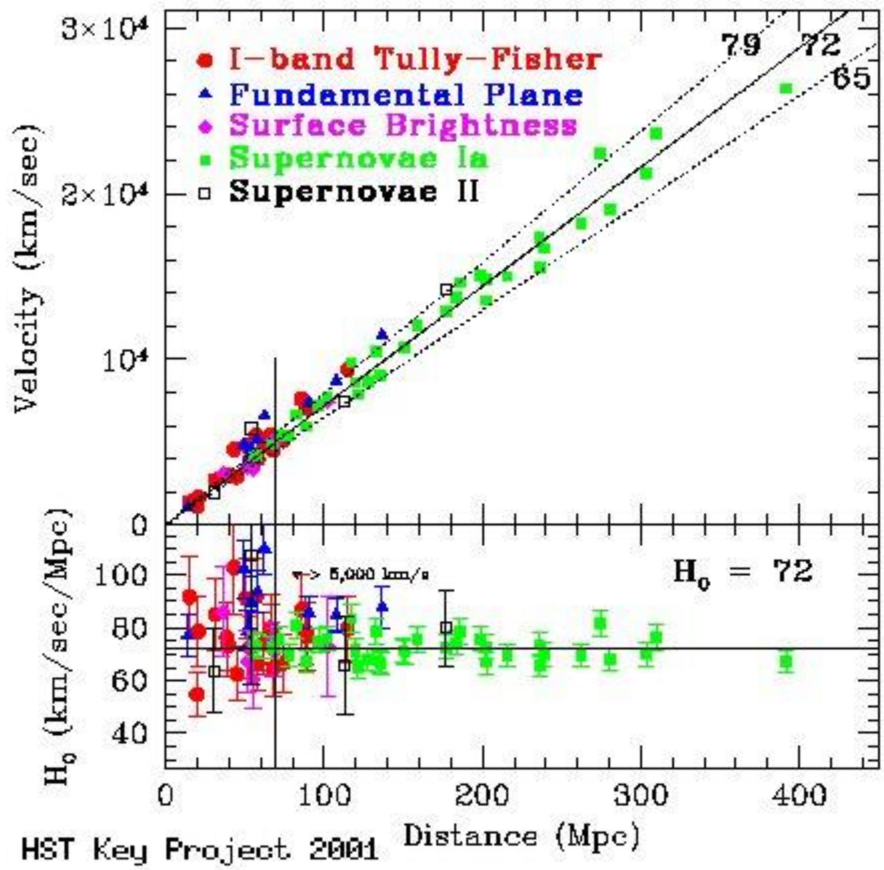
In 1916, Einstein[41] predicted the expansion of the universe with the general theory of relativity, than in 1924 Hubble[42] noticed that there were galaxies other than ours, and with Lemaître[43] (1929) he participated in observing and defining the expansion of the universe, what is called “Hubble's law”[44]

$$v = H_0 r. \tag{1}$$

This v is the average speed of galaxies moving away from us expressed in km/s, r is the distance from us in kilometers, and $H_0 = 67.4$ km/s/Mpc (kilometers per second per mega parsec) is the Hubble constant derived from recent measurements[45] of cosmic microwave radiation. When the Hubble constant is

[39] Zeilik & Gregory 1998
[40] Johannes Kepler (1571-1630), German astronomer and mathematician.
[41] Albert Einstein (1879-1955), German-born theoretical physicist.
[42] Edwin Hubble (1889-1953), American astronomer.
[43] Georges Lemaître (1894-1966), Belgian Catholic priest, mathematician and astronomer.
[44] https://en.wikipedia.org/wiki/Hubble's_law
[45] https://www.space.com/hubble-constant-measurement-universe-expansion-mystery.html

estimated based on the redshift, slightly larger numbers are obtained[46], shown in the following graph. These differences confuse modern cosmology.



In 1933, Zwicky[47] was the first to discover "dark matter", noting that the stars themselves did not provide enough attractive gravitational forces for the rigid rotation of galaxies. The differences between theory and observation that were then observed in other galaxies were supplemented by dust and generally known dark celestial bodies that we do not see with telescopes. But, with better technology and astronomy it was realized that this was insufficient.

Distances

Due to the limited speed of light, the two people talking across the table are never in exactly the same present. We look at the even older past of distant galaxies with a telescope. These perceived pseudo-realities can be represented by some S_p the "system of perceived" (commonly called "proper") coordinates, because fictions also belong to the world of information[48].

[46] https://www.cfa.harvard.edu/~dfabricant/huchra/hubble/
[47] Fritz Zwicky (1898-1974), Bulgarian-Swiss-American astronomer.
[48] [2], 3.21 Fiction

In addition, we believe that the times of distant celestial bodies that some astronomer observes run parallel to the present of him, and that all events, which he cannot see because they escaped into the future of others, belong to a different system S_c of "comparative coordinates" (commonly called "comoving"). This additional system is also a pseudo-reality, inaccessible to direct perception.

The third assumption we need is the "cosmological principle", also the official hypothesis, that an observer to whom the universe is isotropic[49] and homogeneous[50] can be imagined anywhere in the universe. We call them comoving (here also "comparative"). In addition, cosmology defines a "comoving coordinate system" (S_c) in which these observers rest. Cosmological time is added to each commoving observer. With this in mind, we define two types of distances[51].

The "perceived" or proper distance R_p is the remoteness between two regions of space that a given astronomer observes. As the universe expands, the perceivable (proper) distance between two comoving observers grows over time, and what they see is the older past of the other. For official cosmology, this treatment is also[52] theoretical, because we do not know the current (present), i.e. comparative or comoving state of the objects. Light needs time to travel from somewhere, during which the universe expanded, and based on observations, we calculate the appropriate comoving R_c.

Comparative distance is the distance expressed by comoving coordinates. The comoving distance between the two regions of the universe remains constantly the same $R_c = \text{const}$, while the observed (proper) changes with time t. Hence the equation

$$R_p(t) = a(t)R_c \tag{2}$$

where the time function $a(t)$ is the scale factor. At the time of the "Big Bang" it was $a = 0$, and it is assumed that in the current cosmological time $a = 1$. In other words, at present both distances have the same value.

The change in the observed (perceived, proper) distance over time is called the recession rate v, for which we find:

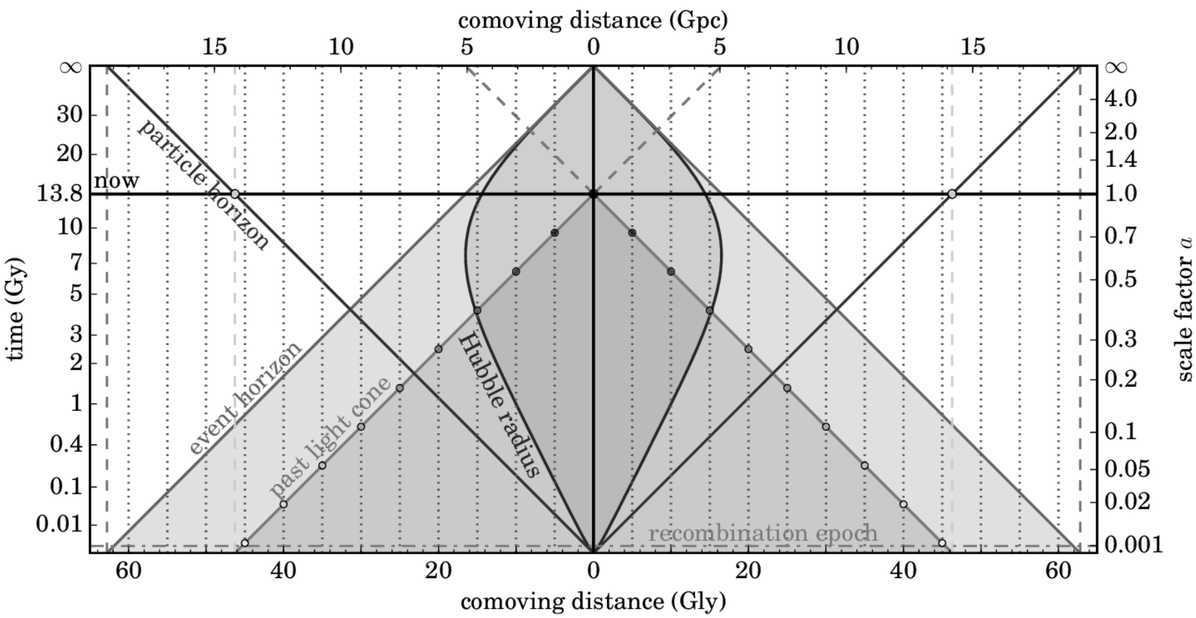
$$v(t) = \dot{a}(t)R_c = H(t)R_p(t) \tag{3}$$

where $H(t) = \dot{a}(t)/a(t)$ is placed. It is Hubble's law (1), now less static.

The following figure[53] shows, on the lower line (comoving distance) the radius in giga light years (Gly) and on the upper line in giga parsecs (Gpc). Cosmological times are shown on the left vertical axis, and the corresponding scale factors on the right. Oblique lines at an angle of 45° represent the rays of light we observe. The vertical line in the middle is our world line (comoving observer in our place), and the

[49] Isotropy (Greek: ἴσος "equal", τρόπος, "path") – equality in all directions.
[50] Homogeneity – equality through volume, at all points.
[51] https://en.wikipedia.org/wiki/Comoving_and_proper_distances
[52] so she recognizes it as a pseudo-reality
[53] https://arxiv.org/pdf/astro-ph/0310808.pdf

horizontal (marked with “now”) means the current cosmological time (13.8 billion years after the Big Bang). We are now at the intersection of those two.



Dotted verticals are “comoving regions” that should be drawn from the point of view of observed (proper) distances to expand, to disperse in relation to us (away from the middle vertical). Here they are shown from the point of view of “comoving distances” and therefore they are all vertical. The rounded line that goes around our world line is the “Hubble radius”, and the areas inside are recession velocities less than the speed of light c, areas outside are speeds greater than c.

The upper line of the image shows that the observable universe is about 14.26 giga parsecs, or the same at the lower 46.5 in billions of light years, or $4{,}40 \times 10^{26}$ meters, in all directions.

Cooling

More likely events are more frequent and less informative. The first is indisputable (I guess), and the second I believe can be understood without talking. In this, we notice the universal “saving of uncertainty” of nature, which I call the “principle of information”. Then we see it in the aspiration of entropy for growth. The spontaneous transfer of heat from a higher body to a neighboring lower temperature body, known as the Second Law of Thermodynamics, is the effort of molecules to reduce the “amount of uncertainty” in their oscillations.

By transferring the energy of oscillation of its molecules to the environment, the substance gains entropy, cools, and now we notice that it also loses information. This general frugality with options and unexpectedness is an effort for non-communication, or non-action, for sluggishness, and I call it the principle of information.

It is also present in the pursuit of growth entropy! Namely, the spontaneous transfer of heat from a higher to a neighboring lower temperature body, known as the Second Law of Thermodynamics, is actually directing the molecules towards reducing the "amount of uncertainty" in their oscillation. By transferring the energy of vibration of its own molecules to the environment, the substance entropy grows, substance cools down, and now we notice that it also loses information.

The universe is cooling. It is already at only a few degrees Kelvin, so only a little above absolute zero – when (so to speak) the movement stops. Modern physics holds that every movement stops at absolute zero (not exactly, but let's say they are close). In the so-called The Bose-Einstein condensate, which is considered to be the fifth state of matter, a very dilute gas at a temperature close to absolute zero, reduces the kinetic energy of the atom. Getting such is becoming a routine thing in better laboratories, and what is important here is that the light there can be significantly slowed down, that its speed can drop to below 20 meters per second!

Well, the universe is expanding and its entropy is growing, but we can say the opposite is also true. Moreover, we can equate statements, consider them equivalent, that the universe is cooling, the speed of light (present) is declining, and that we are seeing the universe getting bigger! We can also say this: the entropy of a substance increases because the information of the substance decreases, and that is because the missing information (in equal amounts) goes into space. Also, because the information of space is higher (they are particles too) the space is getting bigger, the paths between galaxies are getting longer. Or, the universe is becoming rarer (substance).

Epilogue

This is a sequel to the Big Bang story[54], and a further sequel should be perhaps the metrics of the cosmos. I say intentionally in the plural because it is already clear from the attached that there are more of them. For example, these include space-time comparative coordinates, observable, comparable from the point of view of a given astronomer. It is noticeable that modern cosmology looks more like science fiction than dry science, but that is exactly part of its charm. We shouldn't touch her hastily; we shouldn't remove its magic, at least until we are very sure of what we are talking about.

[54] Big Bang, https://www.academia.edu/45088060/Big_Bang

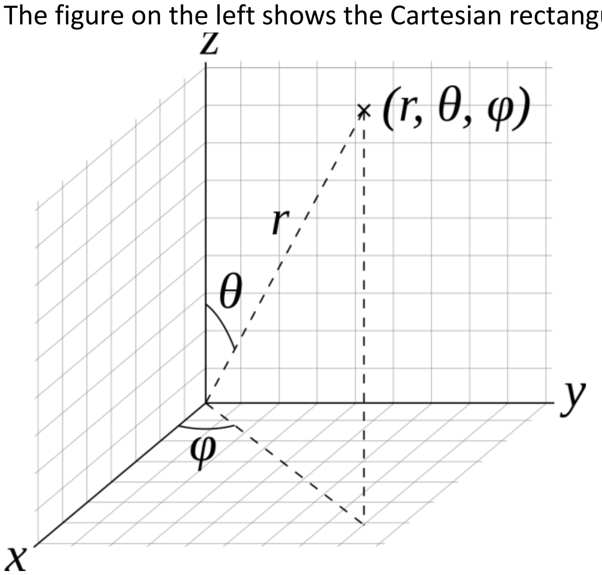
8. Metrics of Cosmos

March 1, 2021

The basic space-time metrics of the theory of relativity, Minkowski and Schwarzschild, are presented, with the intention that the text be an introduction to the sequels. Comments from the standpoint of (my) information theory have also been added.

Introduction

The figure on the left shows the Cartesian rectangular coordinate system (xyz) and in relation to it the



Spherical ($x_1 = r, x_2 = \theta, x_3 = \varphi$) with ties:

$$x = r\sin\theta\cos\varphi,$$
$$y = r\sin\theta\sin\varphi,$$
$$z = r\cos\theta.$$

The square of a linear element, dl^2, is calculated according to (Pythagorean theorem) formulas:

$$dl^2 = dx^2 + dy^2 + dz^2,$$
$$dl^2 = dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\varphi^2).$$

The same result of this infinitesimal length (dl) is obtained by using consistently any of the two given formulas. We introduce the fourth, $x_4 = ict$, time coordinate as the path that passes light at the speed of approximately $c = 300$ thousand kilometers per second (km/s) during time t multiplied by the imaginary unit ($i^2 = -1$).

The corresponding element of flat space-time thus becomes:

$$ds^2 = dl^2 - c^2dt^2 \tag{1}$$

and such was first used in the special theory of relativity. That is Minkowski's[55] metric.

For example, let S' be an inertial coordinate system moving uniformly rectilinear (at a constant) velocity v along the abscissa (x' and x axes) with respect to the reference system S. The linear elements, ds' and ds, are equal respectively to the observer's (which rests in S') and to relative (which rests in S). Hence:

$$ds'^2 = ds^2,$$

$$dl'^2 - c^2dt'^2 = dl^2 - c^2dt^2.$$

As proper one is in rest ($dl' = 0$), and the relative one is moving at speed $v = dl/dt$, it will be further:

$$-c^2dt'^2 = dl^2 - c^2dt^2,$$

$$dt'^2 = dt^2 - dl^2/c^2,$$

[55] Minkowski space, https://en.wikipedia.org/wiki/Minkowski_space

and from there

$$dt = \frac{dt'}{\sqrt{1-\frac{v^2}{c^2}}}. \tag{2}$$

This is a well-known formula of the special theory of relativity[56] for the dilation of time. As long as dt' of its own (proper) time passes, dt of relative time passes, and as $dt > dt'$ follows from (2) whenever $v > 0$, the proper clock (of one's own) is late for the relative observer.

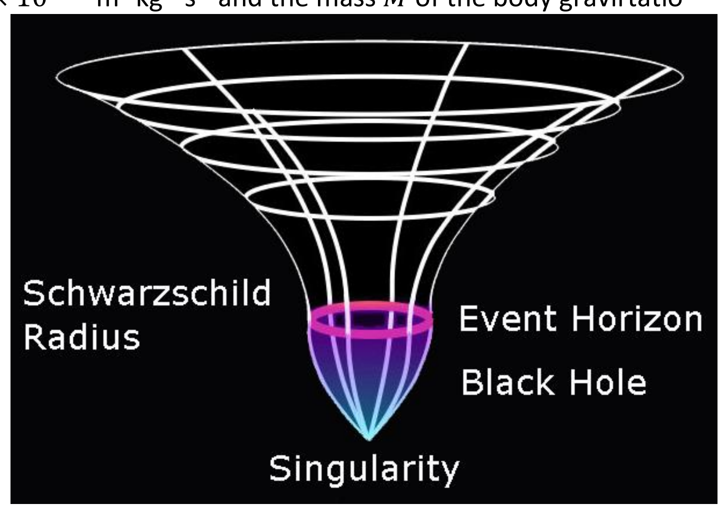
It can be shown that the relative lengths in the direction of motion are shorter than their proper as many times as time slows down. Units of length perpendicular to the direction of motion do not change, so space-time remains “straight”, zero Gaussian curves[57]. Namely, the longitudinal line of motion is a circle of infinite radius ($r \to \infty$) and its reciprocal value is zero ($\kappa = 1/r \to 0$).

Schwarzschild metric

Solving Einstein's general equations for the central symmetric[58] gravitational field, Schwarzschild (1915) found an expression for the distance

$$ds^2 = \frac{dr^2}{1-\frac{r_s}{r}} + r^2(d\theta^2 + \sin^2\theta\, d\varphi^2) - \left(1 - \frac{r_s}{r}\right)c^2 dt^2, \tag{3}$$

where $r_s = 2GM/c^2$ is so-called Schwarzschild radius, in the following figure[59] on the right, with the universal gravitational constant $G = 6{,}674 \times 10^{-11}$ m^3 kg^{-1} s^{-2} and the mass M of the body gravirtationally attracted.



As can be seen from the given expre-ssion, if the observed object were at a distance of the Schwarzschild radius from the center of force, $r = r_s$, in the denominator (first addition) it would be zero, so expression (3) would not make sense. Therefore, the distance $r = r_s$ is the radius of the sphere, called the “event horizon”, with the “black hole” inside. At that boundary, the event horizon, relative radial lengths and time disappear.

In general[60], suppose we have one remote observer in system S' with a flat metric (1) and one in system S within a gravitational field with a Schwarzschild metric (3), it will be:

[56] [8], 1.1.7 Special relativity
[57] Gravitation Multiplicity, https://www.academia.edu/44936839/Gravitation_Multiplicity
[58] [8], 1.2.10. Schwarzschild solution
[59] https://universe-review.ca/R15-17-relativity03.htm
[60] https://universe-review.ca/R15-17-relativity03.htm

$$ds'^2 = ds^2,$$

$$dl'^2 - c^2 dt'^2 = \left(1 - \frac{r_s}{r}\right)^{-1} dr^2 + r^2 d\Omega^2 - \left(1 - \frac{r_s}{r}\right) c^2 dt^2,$$

where the substitution $d\Omega^2 = d\theta^2 + \sin^2\theta\, d\varphi^2$ was introduced, otherwise common for shorter writing. When we assume that both observers are at rest, all lengths will be zero ($dl' = dr = d\Omega = 0$) and we get

$$dt = \frac{dt'}{\sqrt{1 - \frac{r_s}{r}}}. \tag{4}$$

The gravitational field slows down the passage of time. If the observer in the field (in system S) is circling around the center at the speed $v = r d\Omega/dt$, then an additional deceleration follows from the same equality

$$dt = \frac{dt'}{\sqrt{1 - \frac{r_s}{r} - \frac{v^2}{c^2}}}. \tag{5}$$

In contrast to the movement within the field, the movement of the remote observer (system S') with the speed $v = dl'/dt'$ and the rest of the internal ($dr = d\Omega = 0$), gives

$$\left(1 - \frac{v^2}{c^2}\right) dt'^2 = \left(1 - \frac{r_s}{r}\right) dt^2, \tag{6}$$

which means that any deceleration of time by a gravitational field has some equivalent in inertial rectilinear motion with a corresponding velocity $v = c\sqrt{r_s/r}$ outside the field.

Here it is also possible to show that the gravitational field shortens the radial lengths (in direction to center of force) as many times as time (4) slows them down. The lengths perpendicular to the radials are the same and that is why we say that the space-time of the gravitational field is curved.

Creating space

Those random events that happen more often are more likely. It is a description of the frequency, nothing special, but the more probable the event, the less informative it is. When we know that something is going to happen and it is happening, then it is not some big news. This is the origin of nature's principled tendency towards less informative random events, its need not to emit information if it does not have to, and to avoid communication where it can.

The idea of the development of the cosmos towards greater certainty requires that we see its beginnings in greater uncertainty, and also to ask the question why the overall information has not disappeared for a long time ago. The answers to this question are the law of conservation[61] and the (hypo) thesis that information is the basic elements of space, time and matter.

Information is a growing function of the amount of uncertainty to which the law of conservation applies, and in the choice of measure there remains the freedom to associate information with entropy. The first decreases in principle, while the second increases spontaneously (we know from thermodynamics) and it is possible to tighten their connection[62], but that does not matter for now. The more important issue is the paradox of the disappearance of information with the law of its conservation.

Hartley (1928) defined information on equally probable random events as the logarithm of their number ($H = \log N$). It turns out that the state of uncertainty before realization and information after – are equal, which is not new for the mathematical theory of information and communication. It is not a novelty that uncertainty is formally a type of information also, but it is new statement that the principle of minimalism gives priority to potential over active information.

That not realizing the uncertainty will have at least a slightly higher probability than its realization, ie inaction over doing, leads us to understand the law of inertia using the mentioned principle of information, and in general to notice the equivalence of information and physical action[63]. This is a good basis for the following two important theses. Spontaneous growth of entropy refers to the substance, and the information that the substance loses goes to space-time. That is why space and time grow.

Another way to draw conclusions about the growth of space would be to “accumulate biographies” of elementary particles that travel through space, which “do not grow” and their histories have nowhere else to go in the “information universe”. I wrote about it earlier[64]. The half-period of decay of some elementary particles, especially fermions into bosons, would again be the third story.

Deep space

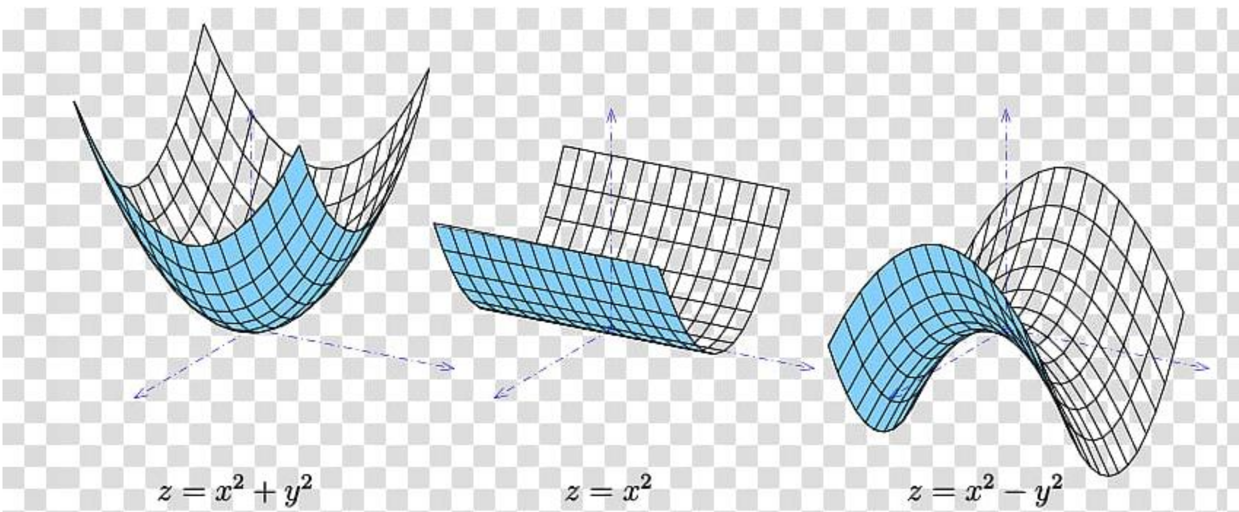
What the astronomer views as a “deep universe” consists of layers of concentric spheres of older histories that have larger radii. The development of the universe, which is otherwise in line with information theory, makes the cosmos not symmetrical in time if it is already in space. It is a feature that makes the geometry of long distances different from the local, the Schwarzschild’s.

The distances from astronomer to the galaxies are considered isotropic (uniform in all directions) and let’s say that’s acceptable. But the universe is expanding and distances are growing. The journey of an object through an imaginary circle, through places visible to the astronomer, in an effort of the traveler to maintain an equal distance to his observer, would end in a spiral similar to Archimedes. It defines the geometry of the observable (proper) universe as hyperbolic.

[61] [2], 1.14 Ammy Noether
[62] [2], 2.24 Entropy Generalization
[63] [1], 23. Action and Information
[64] [2], 2.18 Pilling of History

Namely, the circle that could the astronomer "perceiving" around (imagined) would have a greater length, which means that the ratio of circumference and diameter of such is greater than pi ($\pi \approx 3{,}14$) and that the geometry of the deep space we can look at – has a negative Gaussian curve. In the next picture, in the third figure ($z = x^2 - y^2$), the "saddle surface" is shown. On such a surface, a drawn circle (geometric place of points equidistant from the center) would be twisted and stretched. The ratio of its length (circumference) and radius would also be greater than pi.

All Lobachevsky geometries, which are called hyperbolic[65], are like that. They have a negative Gaussian curve, because the radii of the circles inscribed in the mutually perpendicular planes of the saddle surface are finite and in opposite directions. In the middle figure of the image, regardless of the fact that the surface is wrapped in a roll ($y = x^2$), the space is Euclidean. The circle on it that would represent one of the axes (ordinate) has an infinite radius, which makes the Gaussian curve zero. That is why such a surface could always be unfolded into a plane, with circles whose periphery and radius ratio is pi.



Finally, on the first figure ($z = x^2 + y^2$) is a spherical geometry[66] which we also call Riemannian. The circle described on the sphere would be shorter than the Euclidean one and the ratio of its circumference and radius would be smaller than pi. Similarly would be with a circle drawn around a centrally symmetric gravitational field, the aforementioned Schwarzschild solution of Einstein's general equations. Radial units of length (directions through the center) are shortened and the amount of radius becomes larger, but the vertical lengths (circular around the center of the field) are the same, so the ratio of the perimeter and radius of the circle is less than pi. The Gaussian curvature of such geometries is a positive number.

Epilogue

Only now are the most interesting parts coming (and the most difficult calculations), but in order not to be too extensive, I cut the story. In the continuation, there should have been at least Friedmann and

[65] Hyperbolic geometry, https://en.wikipedia.org/wiki/Hyperbolic_geometry
[66] Spherical geometry, https://en.wikipedia.org/wiki/Spherical_geometry

Gödel's metrics, as two historically most important solutions of Einstein's general equations for the cosmos, and then comments from the standpoint of information theory. However, cosmology has developed in the meantime, and with the theory of information, new moments are emerging, so these classic publications require more attention. Once we discuss all their differences (if ever) then we could run through similar issues.

9. Metrics of Cosmos II

March 6, 2021

Question: The universe is expanding, more and more galaxies are leaving us faster and faster, but are that why[67] time flows slower than ours?

Answer: I don't know for sure, I'm waiting to see some measurements and working on assumptions. If it were that in distant galaxies moving away from us time flows (approximately) as fast as ours and that their units of length are equal to ours, and that the theory of relativity also applies there, then if those galaxies are at rest in relation to us their time flowed faster than ours and the radial (in the directions from us) units of length would be greater. This difference would be such that it can be reversed with relativistic (special theories) time dilatations and length contractions.

Question: On what basis do you draw such an unusual conclusion?

Answer: It is a long story. First read “1.2.8 Vertical Fall” from the book “Space-Time”[2] and notice that the same procedure can be applied to deep space.

Then see “Example 1.2.12. Derive the Schwarzschild metric from (Einstein's) field equations, starting from

$$ds^2 = -e^{2B(r)}c^2dt^2 + e^{2A(r)}dr^2 + r^2\sin^2\theta\, d\varphi^2 + r^2d\theta^2, \tag{1}$$

where $A(r)$ and $B(r)$ are unknown distance functions r”, in the following. Notice how widely the initial conditions are set in the example, and that they still give the Schwarzschild metric in the end.

From the point of view of the distant past of the universe, we leave endlessly, our present in relation to a very old one behaves as when we look from the outside at a body falling into a strong gravitational field, someone that would constantly sink to the horizon of a black hole and never, in our time, it would not reach the edge at which (in relation to us) time stands and radial lengths disappear, become zero.

Conversely, from the standpoint of our present, the distant past of the cosmos is on the edges of the space visible to us and further, to which we could not travel because it would constantly flee from passengers at the speed of light. Even with an imaginary time machine, which would take our time backwards, it would never be possible to reach the beginning of the “big bang”, because the units of the length of one's own (proper) time machine would change so that the journey in it would last indefinitely, although for us the cosmos is only 13.8 billion years old.

Question: What would an imaginary traveler find in the past of the cosmos?

Answer: Space, time and matter consist of information, and their essence is the unexpected, that is, novelty and change. That is why the cosmos is expected to change. Going back to the past, there is less

[67] Because of the Special theory of relativity.

and less space, the substance is denser and there are more black holes. Black holes are integrators of space and time, and on the other hand they are like anchors[68] stationed in space-time as the world around them moves.

Question: And what are the chances that "black holes" do not exist at all?

Answer: Weak. They are predicted by Einstein's equations of general relativity, and those equations also are derived from the principle of least action[69], from which all the equations of theoretical physics known today follow. A lot of "repairs" should be made to physics, which would hardly be possible, by possibly rejecting the idea of the black holes.

Question: Have you heard of Gödel's model of the universe?

Answer: Yes, it is one of the earliest derived from the theory of relativity [12]. Gödel searched for the symmetries of the cosmos of both space and time and found an interesting and instructive example of metrics [13] which, consistent with his assumptions, also allows travel to the past. However, the principle of information requires an asymmetry of time. It is more principled to develop towards less informative, that is, more probable events, which the Gödel's approach can make only one curiosity.

Question: Are you familiar with Friedmann's metrics?

Answer: Yes. Mitra [14] in his "exercise" starts from the assumption of isotropic and homogeneous space and a general form of metrics, more general than (1), to which he applies Einstein's field equations in order to obtain this known metric. Otherwise, the very idea of Friedmann's metrics seems to be a strong competition to other proposals, at least now while we are still waiting for key observations.

Simply put, these (Friedmann – Lemaître – Robertson – Walker) metrics start from

$$r^2 d\Omega^2 = r^2 \sin^2\theta \, d\varphi^2 + r^2 d\theta^2, \tag{2}$$

the part of the spatial expression

$$dl^2 = \frac{dr^2}{1-kr^2} + r^2 d\Omega^2, \tag{3}$$

to obtain for $k = 0$ a flat, Euclidean space, for $k = 1$ a curved and closed spherical, and for $k = -1$ a curved and open hyperbolic. These third cases are close to the above (1).

Question: Space is diluted out, and is the information lost in the process?

Answer: Yes, but there are two ways to save the law of conservation of information during the expansion of the universe. Both could be topical.

---

[68] metal objects that a ship throws into the water when it stops or docks along the shore to maintain balance and prevent movement

[69] [4], 2.5 Einstein's general equations

The first is that in accordance with the principled minimalism of communication, i.e. interaction, only the information of the substance decreases and that such goes into space. It is consistent with the spontaneous growth of (generalized) entropy and the attitude (also of information theory) that more entropy of a given system means less information.

The second is that “space remembers” and that its “memory” can affect our present. The latter arises again from the information theory according to which space, time and matter consist of it, from information, including biographies of elementary particles that are formed while particles travel through space. As such do not grow, we do not notice that they become bigger and bigger during their journey, so we cannot say that they remember something and thus accumulate information within themselves, so the space and time grow on their way. Elementary particles leave space-time to their history.

The second is harmonized with the first by the assumption that the present receives exactly as much information from the past as it loses due to the principle of minimalism of information, i.e. increase of entropy.

Question: Did I understand you to say that time also remembers, or is it a slip?

Answer: It is not a slip; you have noticed well, they both remember, the space and time, because they are symmetrical concepts in the broader view of the universe, as a 6-dimensional continuum.

One time dimension would mean a “deterministic universe,” which (my) information theory does not assume. But if you add only one more dimension of time, it will be an insufficient compromise with the assumed “objectivity of chance” and you will need another one.

More precisely, for each spatial dimension (however many there are in perhaps some future theory such as "string theory") there is one temporal dimension. The symmetries between them are also valid, where $x = ict$, where on the left side of the equation is the spatial length corresponding to the right of the product of the imaginary unit ($i^2 = -1$), the speed of light (approximately $c = 300$ km/s) and time t the light needed to pass the given length.

Well, the time remembers and that's why cosmos have it more and more and it flows to us more slowly. We cannot see the slowdown of the “creation of the present” directly (as well as many things around us are that we do not actually see directly), but we sense and calculate. Namely, if there is less and less information (of substance), there are less and less random events, and the speed of time is a measure of the amount of such events. I say all this from the standpoint of “information theory”, which is still not considered in official physics, moreover, which is an unknown (hypo) thesis there.

Question: Is there a particular problem with “only one dimension of time”, or is this assumption completely arbitrary?

Answer: Yes, there is one difficult problem with such an assumption, and therefore with determinism itself. Some processes are not commutative. I have written this many times.

For example, give a turn signal and turn the car to the left, and turn and signal with a turn signal – they do not have to lead to the same. In general, the order of execution is important for some operations, such as: "double the number" and "add three to the number". In the first composition it would be $2x + 3$, and in the second $(x + 3)2$, so for $x = 5$ the first gives 13, and the second 16. Quantum processes, the so-called quantum evolution, are the representations of the operators (abstract algebras) and not all are commutative. The noncommutative ones lead to the "uncertainty principle", a special case of which is the better known Heisenberg's relations of uncertainty.

Due to the existence of noncommutativity, the theory that there is only one flow of time, that space-time is only 4-dimensional, and that one could agree that "there is no time" (which is seriously considered by some physicists), whole models of the modern physics (relativity theory and quantum mechanics) become questionable.

The exclusive one-linearity of the flow of time leads us into contradiction with the existence of noncommutative time processes and seeks a revision not only of modern physics, but it also challenges the algebra in which noncommutative operators exist. That is why I once gave up on the idea of causal reality, among others, and just not to always fall into the same endless discussions with the followers of such, I say that I use the "hypothesis" of coincidence.

10. Quantity of Options

March 9, 2021

Abbreviated easier answers to more frequent questions related to information theory. About logarithm, additivity to the "quantity of options", fragmentation of equal chances, and physical action associated with information.

Logarithm

The real exponential function $y = b^x$, bases $b > 0$ and $b \neq 1$, will take only positive values $y > 0$; to it the inverse mapping is the logarithm $x = \log_b y$. Conversely, if the logarithmic function $y = \log_b x$, base $b > 0$, $b \neq 1$ and numerus $x > 0$ is given, then its inverse is the exponential function $x = b^y$. In short:

$$b^{\log_b x} = x, \quad \log_b b^x = x. \tag{1}$$

The exponential and logarithmic functions of the same base are mutually inverse.

Since $b^0 = 1$ and $b^1 = b$, it follows from the above:

$$\log_b 1 = 0, \quad \log_b b = 1. \tag{2}$$

Also from (1), due to $b^{u+v} = b^u b^v$ we find $\log_b b^{u+v} = u + v = \log_b b^u + \log_b b^v$, i.e.

$$\log_b(xy) = \log_b x + \log_b y. \tag{3}$$

The logarithm of a product is equal to the sum of the logarithms, if the logarithms are of the same bases and when all three are defined. The immediate consequence is

$$\log_b x^n = n \log_b x. \tag{4}$$

Similar to making (3), from $b^{u-v} = b^u : b^v$ follows $\log_b b^{u-v} = u - v = \log_b b^u - \log_b b^v$, i.e.

$$\log_b \frac{x}{y} = \log_b x - \log_b y. \tag{5}$$

The logarithm of the quotient is equal to the difference of the logarithms, if all three logarithms are defined and the same base.

It can be seen that (4) does not only apply to natural numbers ($n \in \mathbb{N}$), but also to all real numbers ($n \in \mathbb{R}$). Prove of such a generalization is often found in the lessons of real functions and I omit it here. Next we find $\log_c a = \log_c b^{\log_b a} = (\log_b a) \cdot \log_c b$, and hence

$$\log_b a = \frac{\log_c a}{\log_c b}. \tag{6}$$

It is a handy formula for transforming logarithmic bases.

I note that in texts where we use only one base, say $e = 2.71828\ldots$, there is no need to list it in every logarithm. Second, using Euler's formula

$$e^{i\varphi} = \cos\varphi + i\sin\varphi, \tag{7}$$

where $i^2 = -1$ holds for the imaginary unit i, it is possible to define analogous (1) complex logarithms. Such are the periodic functions, such as the cosine and sine.

Additivity

Information arises from uncertainty. More options before realizing a random event will give more information after. Let $L(x)$ be a, for now unknown, function that represents a "measure of uncertainty", a special real amount of equally probable outcomes $x \in \mathbb{N}$. Basically, this function is positive: for any $x > 1$ it is $L(x) > 0$.

Let's say that the mentioned positivity is the first feature of this measure of uncertainty. In particular, with only one outcome, $x = 1$, there is no uncertainty and $L(1) = 0$. Another equally obvious property of the function $L(x)$ is that from $x < y$ follows $L(x) < L(y)$, which means that it is increasing. The third feature is additivity (collectability) and its explanation is a bit more extensive.

Let A and B be equally probable outcomes and let them imply a further choice of equal m and n different options ($m, n \in \mathbb{N}$). The realization of both outcomes, A and B, would contain many options, quantities $L(mn)$, while the quantity of individual realizations A or B would be $L(m)$ or $L(n)$. The additivity of the function $L(x)$ arises from the expectation that the total uncertainty of the outcome will be equal to the sum of individual uncertainties.

For example, A and B are throwing fair coins and dice. The coin itself has the options "tails" and "heads", let's mark them with T and H, and the dice with six numbers from 1 to 6. Throwing both will simultaneously produce $2 \cdot 6 = 12$ results, equal elements of the set {T1,T2,...,T6,H1,H2,...,H6}. Consistent with conservation of uncertainty, it would be required to be $L(12) = L(2) + L(6)$.

In general, for the function $L(x)$ we have stated three requirements: I. that $L(x) > 0$ for every $x > 1$, so that it is positive; II. That from $x < y$ follows $L(x) < L(y)$, that it is increasing; III. That it is additive

$$L(xy) = L(x) + L(y), \tag{8}$$

for each $x, y \in \mathbb{R}$ when $x, y > 1$. Then $L(x)$ is a logarithmic function (arbitrary bases). I borrow the proof from the book [15].

For every $x > 1$ and every $r > 0$ there exists a natural number $k \in \mathbb{N}$ such that $x^k \le 2^r < x^{k+1}$. It is based on this and property II that

$$L(x^k) \le L(2^r) < L(x^{k+1}),$$

whence because of III

$$k \cdot L(x) \le r \cdot L(2) < (k+1) \cdot L(x).$$

As according to I it is possible to divide these inequalities by $r \cdot L(x)$, we have

$$\frac{k}{r} \le \frac{L(2)}{L(x)} < \frac{k}{r} + \frac{1}{r}.$$

The function $\log x$, given bases $b > 0$ and $b \neq 1$, has properties I - III, so the above is valid for it, i.e.

$$\frac{k}{r} \le \frac{\log 2}{\log x} < \frac{k}{r} + \frac{1}{r}.$$

Based on that is

$$\left|\frac{\log 2}{\log x} - \frac{L(2)}{L(x)}\right| < \frac{1}{r},$$

for each $r > 0$. Due to the arbitrariness r the expression in parentheses is zero. So it is

$$L(x) = \frac{L(2)}{\log 2} \cdot \log x = a \cdot \log x. \tag{9}$$

This proves that (8) is a logarithmic function.

However, when the "amount of uncertainty" is considered in more detail or more broadly, an improved information function, such as $L(x)$, will not quite consistently meet the mentioned conditions I, II and III, which means that the logarithmic form alone is not sufficient to represent it.

Shredding

Extending the previous complexity of divisions, we can imagine that we have $n = 1,2,3,\dots$ equally probable choices such that each of them has n_k, $k = 1,2,\dots,n$, mutually equal extensions, and that each k-th the continuation can have the following n_{kj}, $j = 1,2,\dots,k$, mutually equal options, and so on. Then the questions arise, how can we explain these fragments with the help of the mentioned measure of uncertainty, and how far can we get with them?

The information is a special "quantity of options". The logarithmic formula (9) served well for its first approximation, and we further explain it by binary search. The number of binary questions, answered "yes" or "no", required to discover the option, also defines the same (above) information. Namely, we divide the group with $M = 2^n$ equal possibilities into two sets with 2^{n-1} elements each and choose the one in which the choice is requested. After n steps of division and dialing, working with smaller and smaller groups, a hidden option emerges. The number of steps $n = \log_2 M$ is declared information, then binary in bits.

For example, let the number "7" be an unknown imaginary, among the first $M = 8$ natural numbers. Let us divide the group of possible (eight) numbers into two parts, the first set {1,2,3,4} and the set of others {5,6,7,8}. Let's ask the question "Is the requested number in the first set?". According the answer "no" we divide the second, again into two parts, the first {5,6} and the second part {7,8}. We ask the same question "Is the requested number in the first group?", to which the answer is again "no". We divide the new second group into two parts, the first set {7} and the second {8}. The answer to the same question is now "yes". As there are no more divisions, the required number is "7", and the number of divisions is $n = 3$. The number of possibilities is $M = 8$, and the information they carry is $3 = \log_2 8$.

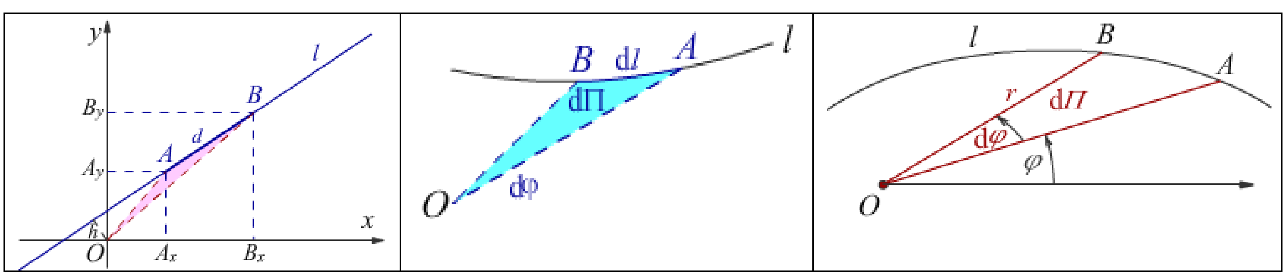
In order for the information ($n = \log M$) not to diverge, it is necessary that the number of possibilities (M) of the given situation be limited. In that sense, our communications (interactions) are finite, with the smallest (quantities) of free information. These smallest packets of information travel along with the physical action, which is also atomized (quantized).

Action

The fact that information is "held" together with physical action makes it a physical phenomenon. The quantum of action is the Planck constant ($h \approx 6{,}626 \times 10^{-34}$ m^{2}kg/s) which appears in the expression for the energy $E = hf$ of the electromagnetic radiation particle (photon), where $f = 1/\tau$ is the frequency of the wave, and τ is the time of one period. Hence, the elemental action is $h = E\tau$, a product of energy and duration.

In short, information is the equivalent of what[70] we can consider a "surface", a generalized product of energy and time. The fact that communication is formally equal to the exchange of interactions makes information in free form (physical particle) an accomplice of any energy change. This, among other things, means that there are carriers of gravitational force – gravitons, as we know that there are carriers of electromagnetic force – photons.

In addition, the path of a particle from point A to B moving under the action of some constant central force O is equivalent to the area swept by its radius vector, from position $\overrightarrow{OA}$ to position $\overrightarrow{OB}$, for a given time. The established connection between objects A, B and O is communication. During that time, an energetic interaction takes place, due to which we say that body A, or B, moves under the action of force O.



In the three images above, from left to right, the (curved) line l moves the particle (body) from point A to point B so that at equal times the area of the (curvilinear) triangle ABO is constant:

70 [1], 3. Potential Information

1. The line l is straight and the force from O is zero.
2. The line is a hyperbola, and the force is repulsive.
3. The ine is an ellipse and the force is attractive.

If the trajectory is a conic (line, circle, ellipse, hyperbola, parabola), then from a given point O the charge is driven by a force that decreases with the square of the distance and its force carriers (gauge bosons) move at the speed of light. Conversely, if the path l is not conical then the force does not decrease with the square of the distance and its carriers do not travel at the speed of light. I discussed this in [1].

The constancy of the ABO surface (area) in the images comes from the influence of force. It is clear that a change of force would change this area, so now we know that it then changes the action, the information, and consistently the probability. In other words, instead of saying that such charges (A, B) move under the action of the corresponding force (O), we can say that they travel in their trajectories (l) because such movements are most probable from the point of view of the charges.

The impossibility of changing the probability, information and action without changing the force indicates to the conservation laws of the mentioned phenomena, and the observation that in the absence of force we have a rectilinear motion (the first of three images) – about inertia and minimalism of each of these quantities. On the other hand, we see the principled minimalism of information in the maximalism of probability, in the fact that more probable phenomena are more frequent and less informative.

A special issue is particles that do not move at the speed of light. They have their own (proper) time and hence the mass of rest, i.e. additional inertia – again due to the principled minimalism of communication. However, that is a special topic.

11. Quantity of Options II

March 11, 2021

Excerpts from conversations about compliance with language, free networks, event direction, and everything related to information perception.

Language

I explain the scalar product of vectors using probabilities and one type of language statistics.

Question: Can you explain to me the "information of perception" on a specific example?

Answer: Yes, let's take English and say computer data processing. Let's form a list of $n = 3000$ the most frequently used words in a wide sample of written texts, movies, speeches. Each word has a certain frequency, the probability of appearing in a given sample, the k-th word ($k = 1,2,\dots,n$) has the probability $p_k \in (0,1)$. We can arrange the words in descending order, so that they more often have a smaller index k, but the sum of all p_k is one. Thus, we defined the sequence, ie the probability distribution vector $\mathbf{P} = (p_1, p_2, \dots, p_n)$.

Then let's take an individual who speaks English and do the same with his knowledge and use of English, his vocabulary as a sample. We get the vector $\mathbf{Q} = (q_1, q_2, \dots, q_n)$. Information perception is then the so-called scalar (inner) product of these vectors

$$S = \mathbf{P} \cdot \mathbf{Q} = p_1 q_1 + p_2 q_2 + \dots + p_n q_n. \tag{1}$$

Some of the well-known preferences of people who speak languages are further shown to be the result of mathematical premises about vectors.

For example, one premise[71] says that the scalar product of the vectors is not greater than the product of the intensities of the vectors. That product can therefore (formally) be treated as probability (which is a number between zero and one), and then apply the "probability principle" (I invented the name), that more probable events are more common. Then comes the conclusion that "general" and "personal" language will strive to increase that result, the scalar product, that is, "information of perception".

As the (scalar) product is, the higher the "alignment" of the vector components, ie coefficients of the sequence, and the largest is when both are arranged in the same order (descending or ascending), it will adapt to the "language" the "person", if the former can't be sufficiently adapted to the latter.

Note that the same can be applied (calculated) to the behavior of an individual in relation to the behavior of a group. This creates a "spontaneous synchronicity", otherwise a long before noticed but still insufficiently understood phenomenon of individual adaptation. However, the theory of information perception is much more universal. Its formal model works well in various examples of social and natural sciences, and especially in quantum mechanics (which is full of intuitively difficult to digest results).

[71] [16], Lemma 1.2.49 (Cauchy–Schwarz inequality), p. 132

Free networks

Nodes with equally probable connections create large networks[72] whose degree probability distribution follows a power law, at least asymptotically. They are characteristic for internet connections, electronic lines of larger regions, acquaintances among people in general, and even free markets. Such spontaneously grow into a relatively small number of positions (concentrators) with many connections and the rest of them many with few connections. We call them scale-free networks.

Question: Why are there few on the free market who are much richer than others, they say you know?

Answer: Yes, I am not alone, it is known. It is a spontaneous process of creating the so-called free networks. When the connections of the network (read flows of money and goods) are equal, then the few nodes (exchange centers) become the owners of more and more lines. Nodes with more of them thus get even more, and the poor ones remain short-sleeved.

The well-known explanation goes by probability. When you add a new link, and they are all equally credible, it is more likely that it will belong to a node that has more of them. I forwarded him an additional attachment with details (irrelevant here).

Q: Okay, I flipped through, that's something clear with probability, but they said you know a different explanation, supposedly with some new theory. Is it true?

A: Maybe it could be "information theory", but it is still my private matter. I believe I know several ways and here is one, in short, using "information perception".

Consider (1) when the number of additions (n) is large. Let it be the norm again ($p_k = x_k^2$)

$$\|\mathbf{x}\| = x_1^2 + x_2^2 + \dots + x_n^2 = 1 \tag{2}$$

and for the vector $\mathbf{y}$ (that is $\mathbf{Q}$) the similar, and let all these probabilities be uniform, constant. Then, when approximately $p_k = q_k = 1/n$, the sum S is smaller as the number n is larger. Namely, from

$$S = \frac{1}{n^2} + \frac{1}{n^2} + \dots + \frac{1}{n^2} = \frac{n}{n^2} = \frac{1}{n} \to 0, \ \ n \to \infty, \tag{3}$$

it follows that the probability of the scalar product S decreases with increasing number of additions. If the first probabilities (components of vector $\mathbf{P}$) represent the existing state of the network, and the second (vector $\mathbf{Q}$) new members, then the network will "avoid" growth that leads to uniformities (3).

A small number of high probability aggregates and a large number of low probability aggregates will be formed spontaneously – to increase product S – which then becomes a well-known statement about the growth of these networks into a small number of nodes with many connections and a large number of nodes with few connections.

[72] [2], 1.6 Equality

That the scalar product (1) will indeed increase each time its two equal sums, $S = x^2 + x^2 + w$, are replaced by two unequal sums, $S' = (x + a)^2 + (x - a)^2 + w$, where the remainder of the sum w remains the same in both, we see from the following:

$$S' = (x + a)^2 + (x - a)^2 + y = 2x^2 + 2a^2 + w > 2x^2 + w = S. \tag{4}$$

For example, $0.6^2 + 0.4^2 > 0.5^2 + 0.5^2$, or $0.5^2 + 0.3^2 + 0.2^2 > 0.33^2 + 0.33^2 + 0.33^2$, which is easy to check. Additionally, we note that the difference $S' - S = 2a^2$ crescent with increasing sum of inequalities, which means that the spontaneity of this growth continues.

When it comes to free networks, unfettered association, the united system will try to be formed so that there is no equality of nodes, and inequality will strive for further development. By interpretation, we will understand that there are few very "rich" versus many "poor" and that there is a natural tendency of the "free market" to further increase the differences between them.

Routing

Information theory starts from the assumed coincidence. However, although it is not deterministic in the true sense, the theory I advocate predicts the development of the cosmos towards greater causality – whether we start from the existence of random events in rare situations, or deny the ultimate, strict certainty in any case. This is a consequence of the "probability principle" according to which more probable outcomes are sought.

In that sense, determinism (causality) can be considered a consequence of chance. The opposite is difficult, except, for example, when the basis of "coincidence" is the limitation of our perceptions to infinite causes. However, this reverse case is also a topic of information theory.

Question: What is better for development, clutter or organization?

Answer: It's like asking me what's better for water (H_2O) oxygen or hydrogen, and the answer would be the same, it doesn't work without both. I will explain this from the point of view of "information of perception" in its basic expression similar to (1).

We consider intelligence $I = S/H$ as the ability of an individual that is proportional to its ability to choose S and inversely proportional to the surrounding constraints H. Hence $S = IH$, and the value is related to a special situation, a special problem that the subject faces and solves .

If we assume that an individual perceives $n \in \mathbb{N}$ different situations that he can see as a problem and solve them, then we have so many ($k = 1,2, \dots, n$) of his special abilities $S_k = I_k H_k$. Their sum

$$S = S_1 + S_2 + \dots + S_n \tag{5}$$

is the total information of a given person's perception.

On the other hand, due to the law of conservation[73], it makes sense to define the intensity, the overall measure of both intelligence and the hierarchy involved in perception information (5). It does not have to be a simple sum, neither the “intelligence” I_k nor the “hierarchy” H_k, nor the “Pythagorean theorem” (the root of the sum of squares), but anything from the rich treasury of mathematical theory of measure.

Intelligence is a more adaptable size; it is more plastic than hierarchy in most cases, so it is the person who adapts to be more successful, and less often the other way around. In order to keep the amount of intelligence in question more or less unchanged, it can be organized by organizing its daily routines (time to get up and go to bed, meals, schedule of tools it works with, contact with other subjects, work methods) in order to release excess skills for the most important in his career.

The same goes for the collective. Intelligence is then a feature of the group, and hierarchy is again something outside that subject. Due to the plasticity of intelligence versus the surrounding limitations, the term “efficiency” in the military, enterprise, local commune-government, has the weight we know it has. I compared it earlier[74] to a sausage squeezed on one side to explode on the other.

Unfortunately, the environment is changing. It also changes in unpredictable ways, so an efficient system that tries to follow these changes sooner or later comes to a standstill. As in the saying that even roads paved with good intentions can lead us to hell. The hunter will outwit the beast with his excess of intelligence, typical of originality that is not on the planned paths, unlike routines.

Q: Is there any certainty that an efficient system will sooner or later become obsolete?

A: Yes, and look for proof of this in Gödel's theorem of impossibility. An organization that would be so effective that it does not become obsolete would be able to adapt to any change. But such would not be possible, because changes can be unpredictable for any pre-given system of constraints, such as truths that cannot fit into any concept of logic (e.g. arithmetic, algebra, mathematics), so any way of organizing will eventually show failure.

The material world is slowly losing information. The entropy of the substance of the cosmos increases as its information decreases[75] and goes into space which is increasing. We are becoming an environment whose interior design is improving, certainties are growing and we look more and more like someone who closes windows and doors away from the outside. At the same time, external uncertainties will not disappear, they will only concern us less, but that process does not go to the very end, until the complete disappearance of information.

Assuming that space, time and matter consist of the information itself, and that the essence of this is uncertainty, and all this together with the laws of conservation, there is no conclusion that in the end unpredictability can completely disappear. With such, any secure, safe and efficient system of rules, precisely because they will have more limitations than the outside world, will have to become obsolete.

[73] information and probabilities
[74] [2], 2.3 Information of Perception I
[75] official physics would not agree with me on this

Q: If we imagine a system as big as the whole universe. Could it be that it does not become obsolete?

A: No. No matter how the universe is, she is also some information and, therefore, uncertainty considering her environment. Therefore, it is not possible to fully predict and describe her, nor is it possible to design her relationships in advance. In that sense, the universe itself is not and cannot be a sufficiently organized system that could "not become obsolete".

Perception information

We can draw different interpretations from the scalar product of the vector, and among them we will find different explanations of the information of perception. Confusion about such content is most often caused by our prejudices in understanding information.

Question: How is it possible to use the term "perception information" for something that should be called "perception probability"?

Answer: The question is clear to me, it is well asked. The absurdity of replacing the words "probability" and "information" comes from their supposed "clear and different" meaning. More likely events are less informative; the first implies the realization of larger values, the second smaller ones. But not everything[76] about them is so black and white.

When an event A is almost certain and the probability $p = P(A)$ that it will happen is approximately one ($p \to 1$), then the opposite event $\overline{A}$ is almost impossible and the probability that it will not happen $q = P(\overline{A})$ is approximately zero ($q \to 0$). In general it is $p + q = 1$, so based on the functional analysis (by developing the logarithmic function in series) we get the relation

$$q = -b \cdot \log p, \tag{6}$$

where $b > 0$ is a constant that determines the unit of information.

Therefore, in the summands of information perception (1) the second factors may be probabilities (that the given events will not happen), thus they become information of their realization. The more "impossible" such events are, the more likely they are to approximate Hartley's information (the logarithm of probability) of their negation, and such are mostly the prohibitions of natural laws and solid hierarchies.

So, when we talk about "perception information" in the basic sense, as a scalar product of "intelligence" and "hierarchy", then we aim at such expressions.

Another example is Shannon's information

$$S = -p_1 \log p_1 - p_2 \log p_2 - \cdots - p_n \log p_n, \tag{7}$$

[76] [1], 14 Uncertainty

where p_k is the probability distribution, and S is the mean value (mathematical expectation) of their individual information. I leave the choice of the logarithm base to the reader.

The logarithms of the numbers $p \in (0,1)$ are negative and Shannon's information is a positive real number. When the first factors make a descending sequence then the second make a growing one, so the expression S represents a kind of minimum. If these factors were matched so that both form a descending sequence, or both ascending, the sum of the items would be maximal. On the contrary, this expresses the minimalism of the mean value of the probability distribution information.

If the number of choices of probability distribution (7) were very large ($n \to \infty$), and the probabilities were more or less uniform, then according to the previous one we could write $p_k = -\log q_k$, respectively for each $k = 1,2, \dots, n$, where $q_k \to 1$ are the probabilities that the k-th event will not occur. Then “perception information” could be said to represent a scalar product of vectors of components that are not probabilities, but information.

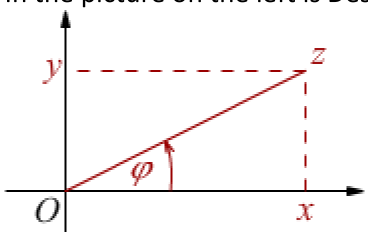
12. Quantity of Options III

March 14, 2021

The text is a continuation of the topic "quantity of options" now on complex numbers, vector multiplication and correlation.

Complex numbers

In the picture on the left is Descartes' rectangular coordinate system Oxy and in it the point $z = (x, y)$.

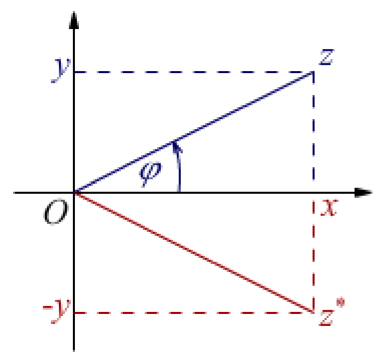


We put

$$z = x + iy,$$

where is also the imaginary unit i for which $i^2 = -1$, where z becomes a complex number and the plane Oxy a complex. Unlike the set of real numbers $\mathbb{R}$, for a set of complex the mark $\mathbb{C}$ is the same as for a complex plane. Note, $x, y \in \mathbb{R}$, or $z \in \mathbb{C}$.

The imaginary unit has the position $i = (0,1)$, while the coordinates of the real unit are $(1,0)$. As we know, the conjugate complex number of the number $z = x + iy$ is the number $z^* = x - iy$, in the following figure on the right.



The abscissa (x-axis) is the axis of symmetry of conjugate complex numbers, so $f^*(z)$ and $f(z)$ will be axially symmetric with respect to it, for each real function $f(x)$, and hence $f^*(z) = f(z^*)$. That is why f^*f is a real number, it lies on the abscissa, because $(f^*f)^* = f^*f$.

In particular we see this from $z^*z = (x + iy)(x - iy) = x^2 + y^2 \in \mathbb{R}$. It follows that the square of the modulus of a complex number, its so-called absolute values, can be defined by the distance of a point, which represents it in a complex plane, from the origin:

$$|z| = \overline{Oz} = \sqrt{z^*z}.$$

That definition consistently applies to real numbers as well.

In the polar coordinates $(Or\varphi)$, the notation of the point $z = x + iy$ becomes $z = r(\cos\varphi + i\sin\varphi)$, and its conjugate complex $z^* = r(\cos\varphi - i\sin\varphi)$, where $r = |z|$. This is usually and easily obtained from the transformation of Cartesian coordinates into polar ones, for the derivation of which the trigonometry of a right triangle is sufficient. A little more complex is the proof of the equality of the so-called $\text{cis}\,\varphi = \cos\varphi + i\sin\varphi$functions

$$e^{i\varphi} = \cos\varphi + i\sin\varphi\,, \tag{1}$$

which was first found by Euler[77] and named after him. Euler is known for his great contribution to the development of power series, displaying functions in the form of sums of an infinite number of summands, such as:

[77] Leonhard Euler (1707-1783), Swiss-German-Russian mathematician.

$$\begin{cases} e^w = 1 + \frac{w}{1!} + \frac{w^2}{2!} + \cdots + \frac{w^n}{n!} + \cdots \\ \cos\varphi = 1 - \frac{\varphi^2}{2!} + \frac{\varphi^4}{4!} - \cdots (-1)^n \frac{\varphi^{2n}}{(2n)!} + \cdots \\ \sin\varphi = \frac{\varphi}{1!} - \frac{\varphi^3}{3!} + \frac{\varphi^5}{5!} - \cdots (-1)^n \frac{\varphi^{2n+1}}{(2n+1)!} + \cdots \end{cases} \quad (2)$$

whence by multiplying the sine by the imaginary unit and adding the cosine, then putting $w = i\varphi$ we get (1). Euler's number $e = 2.71828 \dots$ is irrationally transcendent.

Starting from (1), the definition of logarithm[78] and Hartley information, it is possible to define "generalized information", moreover, to recognize it in the solution of Schrödinger's equation

$$-\frac{\hbar^2}{2m}\nabla^2\ \psi(\boldsymbol{r}, t) = i\hbar\frac{\partial}{\partial t}\psi(\boldsymbol{r}, t), \quad (3)$$

where ψ is the wave function of the particle at place $\boldsymbol{r}$ at time t. The solution for a particle of momentum $\boldsymbol{p}$ or wave vector $\boldsymbol{k}$, with angular frequency ω or energy E, is given by a complex wave plane

$$\psi(\boldsymbol{r}, t) = Ae^{i(\boldsymbol{k}\cdot\boldsymbol{r} - \omega t)} = Ae^{i(\boldsymbol{p}\cdot\boldsymbol{r} - Et)/\hbar}, \quad (4)$$

with amplitude A and $\omega = \frac{\hbar k^2}{2m}$ (or equivalent $E = \frac{p^2}{2m}$) in case the particle has mass m, or $\omega = kc$ for particle without mass (in rest).

Multiplication

Each sequence of $n \in \mathbb{N}$ numbers can be represented by the vector $\boldsymbol{a} = (a_1, a_2, \dots, a_n)$ of some S_1 vector space ($Ox_1x_2 \dots x_n$). In the Euclidean sense, the intensity, or norm, of the vector is its length

$$a = |\boldsymbol{a}| = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2}, \quad (5)$$

and in general it is a quantity from the rich treasury of measure theory. Certainly, a vector is zero when its intensity is zero, and a "larger" vector is of greater intensity. Changing the system does not change the norm.

Every two non-zero vectors, such as the mentioned $\boldsymbol{a}$ and some $\boldsymbol{b} = (b_1, b_2, \dots, b_n)$, determine only one plane $O\boldsymbol{ab}$, as in the following figure. They belong to the new S_2 coordinate system Oxy, where:

$$\overline{OA} = |\boldsymbol{a}| = a,\ \overline{OB} = |\boldsymbol{b}| = b;\ A = (A_x, A_y),\ B = (B_x, B_y).$$

The angle between the given vectors is preserved

$$\varphi = \sphericalangle(AOB) = \beta - \alpha;\ \alpha = \sphericalangle(xOA),\ \beta = (xOB).$$

[78] 10. Quantity of Options

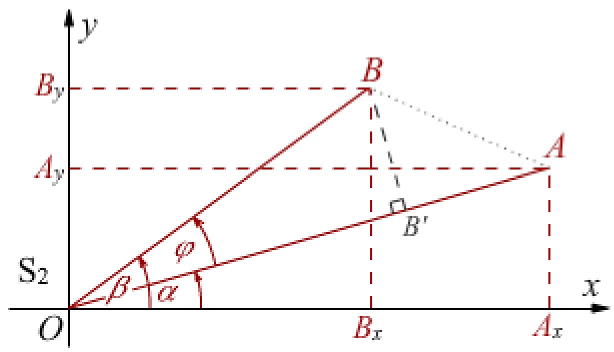
The vertical projection of point B on the line OA is point B', and the length of the segment $\overline{OB'} = \overline{OB} \cdot \cos\varphi$. The scalar (inner) product of the given vectors has the same value in both systems

$$\boldsymbol{a} \cdot \boldsymbol{b} = |\boldsymbol{a}||\boldsymbol{b}| \cos\varphi. \tag{6}$$

In the initial, using the property of orthogonality of each pair of n coordinate axes and because $\cos 90° = 0$, we easily obtain

$$\boldsymbol{a} \cdot \boldsymbol{b} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n, \tag{7}$$

and that is a scalar (number) equal to (6).



Note that the area Π of the triangle OAB in the figure is equal to the pseudoscalar[79] product of the base $\overline{OA}$ and the height $\overline{BB'}$. Since $\overline{BB'} = \overline{OB} \sin\varphi$, we define a pseudoscalar product

$$\boldsymbol{a} \therefore \boldsymbol{b} = |\boldsymbol{a}||\boldsymbol{b}| \sin\varphi \tag{8}$$

and thus the area $\boldsymbol{a} \therefore \boldsymbol{b} = 2\Pi(OAB)$. The same area of triangle OAB can be obtained by subtracting the area of triangle AOA_x from the sum of the areas of triangle BOB_x and the trapezoid ABB_xA_x , so:

$$\Pi(OAB) = \frac{1}{2}B_xB_y + \frac{1}{2}(B_y + A_y)(A_x - B_x) - \frac{1}{2}A_xA_y = \frac{1}{2}(A_xB_y - B_xA_y).$$

If we introduce the label

$$[A, B] = A_xB_y - B_xA_y, \tag{9}$$

which I call the "commutator" of points $A(A_x, A_y)$ and $B(B_x, B_y)$, will be $\boldsymbol{a} \therefore \boldsymbol{b} = [A, B]$. The commutator of points, ie the pseudoscalar product, is actually the intensity of the vector product of the vectors $\overrightarrow{OA} \times \overrightarrow{OB}$.

We understand the information of perception[80] in the way of a scalar product (6), for example, when the coefficients of both vectors are probabilities, in the extreme case when the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ represent

[79] [1], 2.4.4 Pseudoscalar product, p. 91.
[80] 11. Quantity of Options II

probability distributions, so their larger product means a more probable occurrence. In contrast, perception information[81] is better understood in the form of a pseudoscalar product (8) when it represents, for example, “equal surfaces that the radius vectors of the planets erase at equal times orbiting the Sun” (Kepler's Second Law).

Correlation

In larger cities, people walk faster, where people walk faster, there is a higher percentage of cardiovascular diseases. This is a recent example of a more difficult correlation. The more well-known ones are, the more we run the more calories we burn, the longer the hair the more shampoo we spend on it, or by salting food the more we have higher blood pressure. Note that the opposite is not always the case, for example, you may have high blood pressure without a salty diet.

In statistics textbooks, when correlation is studied and the possibility of drawing erroneous conclusions based on it is pointed out, the phenomenon observed in Copenhagen a few years after the Second World War is often mentioned. A positive correlation was noticed between the number of newborn children and the number of storks that nested in that city. However, it does not mean that storks bring children, but it is a consequence of the relocation of the population from the countryside to the city in the time after the war. With the arrival of the new population, there was an increased number of newborn children and the construction of houses with chimneys on which additional storks could place their nests.

In general, correlation (lat. con = with, relatio = relationship) is the interrelationship or mutual connection of two different phenomena represented by the values of two variables. Let the first and the second be given by their data series, $\boldsymbol{a} = (a_1, a_2, \dots, a_n)$ and $\boldsymbol{b} = (b_1, b_2, \dots, b_n)$. For each of the series we find the mean value separately, their own centers of gravity:

$$\overline{a} = \frac{a_1 + a_2 + \dots + a_n}{n}, \ \overline{b} = \frac{b_1 + b_2 + \dots + b_n}{n}. \tag{10}$$

Subtract their mean values from the components of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. We obtain new coefficients $a'_j = a_j - \overline{a}$ and $b'_j = b_j - \overline{b}$ for the indices $j = 1,2, \dots, n$ orderly, the two new vectors:

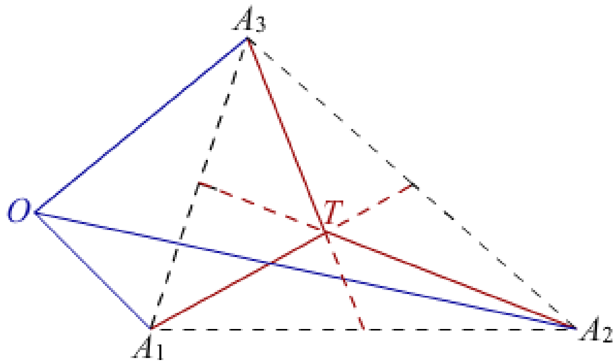
$$\boldsymbol{a}' = (a_1 - \overline{a}, \ a_2 - \overline{a}, \dots, a_n - \overline{a}), \ \boldsymbol{b}' = (b_1 - \overline{b}, b_2 - \overline{b}, \dots, b_n - \overline{b}). \tag{11}$$

What happens then is shown simplified ($n = 3$) in the following figure on the left.

Triangle $A_1A_2A_3$ defines a plane and a coordinate system. Point T is the center of gravity of the triangle, and point O is the origin of the coordinate system. Only one of the axes of that system and the projection of all points (from the picture) on that axis are important to us. The components of the first of the above vectors, vector $\boldsymbol{a}$, are in order:

$$a_1 = \overline{OA_1}, \ a_2 = \overline{OA_2}, \ a_3 = \overline{OA_3}.$$

[81] [3], 2.4.7 Potential information, p. 95.



Оне постају компоненте тежишних вектора:

$$a'_1 = \overline{TA_1},\ a'_2 = \overline{TA_2},\ a'_3 = \overline{TA_3}.$$

The explanation for vector $\boldsymbol{b}$ is similar.

We calculate the intensities of these vectors:

$$\begin{cases} a' = |\boldsymbol{a}'| = \sqrt{a'^2_1 + a'^2_2 + \dots + a'^2_n} \\ b' = |\boldsymbol{b}'| = \sqrt{b'^2_1 + b'^2_2 + \dots + b'^3_n} \end{cases} \tag{12}$$

According to the Cauchy-Schwartz inequality, the scalar product of two vectors is not greater than the product of their intensities ($\boldsymbol{a}' \cdot \boldsymbol{b}' \le |\boldsymbol{a}'||\boldsymbol{b}'|$), so the correlation coefficient

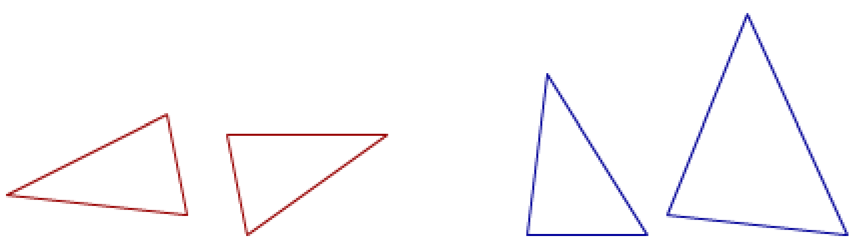
$$r = \frac{\boldsymbol{a}\prime \cdot \boldsymbol{b}\prime}{|\boldsymbol{a}\prime||\boldsymbol{b}\prime|}, \tag{13}$$

is real number from the interval $(-1, +1)$. When $-0.20 \le r \le +0.20$ the correlation is nil or negligible. When $0.20 \le |r| \le 0.40$ the connection is easy. When $0.40 \le |r| \le 0.70$ the connection is real or significant. When $0.70 \le |r| \le 1.00$ the correlation is high or very high.

Calculating the correlation coefficient (13) is a typical task of statistics. The data are then observed paired:

$$(a_1, b_1),\ (a_2, b_2), \dots,\ (a_n, b_n), \tag{14}$$

and the connection with the “center of gravity” like the one in the previous picture is not important. However, when we observe two such centers of gravity (n-tuple of points) with corresponding center of gravity vectors (strokes from center of gravity to points, vertices) and vertices (points), we can notice their different orientations. The orientation of the correlation can be mapped to shapes, not just numbers, as shown in the following figure.



The first two triangles (red) are negatively correlated, and the next two (blue) are positively correlated. In this kind of visual comparison, analogous to the calculation of the correlation, the size of the triangles is not important, but only their geometry. A special way of understanding correlation, of course, is the information of perception.

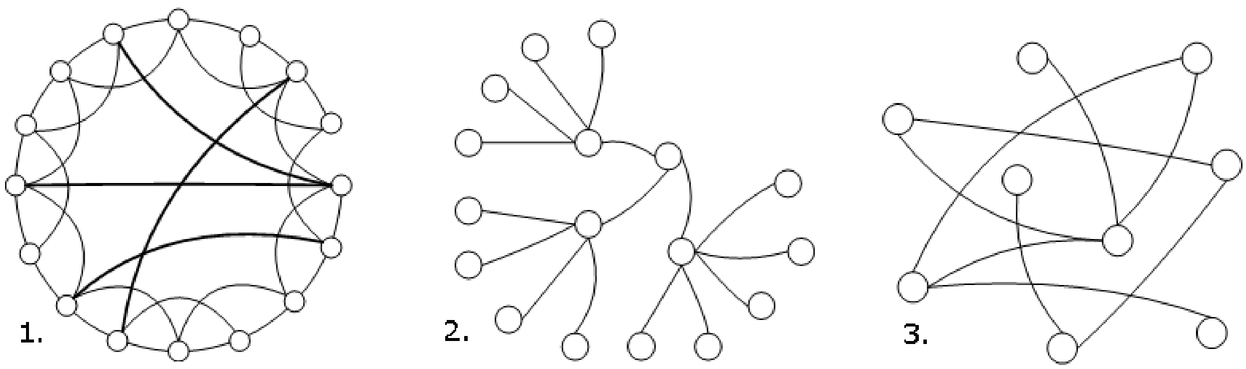
13. Shape recognition

About three types of networks and correlation

March 17, 2021

Free networks

Graph theory is a very common field of mathematics in computer science as well. Its goal is to study the properties of nodes, i.e. points or vertices, connected by lines or branches. The three types of graphs that are most important to us are the "networks" shown in the following figure.



The first on the left is the "small-world network", a type of graph in which most nodes are not immediate neighbors, but their neighbors are mostly. The predominant number of peaks of this type is loaded with a large number of lines due to shorter journeys between nodes. Expressed by the total number N of points of the network, the average number of links between the two is

$$L_1 \propto \log\log N. \tag{1}$$

For example, in the base of the natural logarithm $\mathrm{e} = 2.71828\ldots$, for the number of nodes $N = 10^8$, we find $\log N = 18{,}4207$ and $L_1 = \log 18{,}4207 = 2{,}91347$. So, it takes (on average) less than three steps to an arbitrary another place in the "small world" network with one hundred million positions.

The second, in the middle of that picture is "scale-free network". The hub, i.e. center, concentrator or "pile" of the network is a node whose number of connections significantly exceeds the average. Such a network is some optimum of the load by the links and the length of the paths between the nodes

$$L_2 \propto \frac{\log N}{\log\log N}. \tag{2}$$

For example, for the natural logarithm and number of nodes from the previous case (bases e and $N = 10^8$) we now get L_2=6.32258. Through a free network with one hundred million nodes, you can reach another arbitrary node with about six steps.

A new free network node is more likely to bind to a multi-link node. This is precisely because the probabilities of the links are equal. Due to the equality of the branches, the new peaks are referred to the existing, with the probabilities proportional to the number of the current

$$p_i = \frac{k_i}{\sum_j k_j}, \tag{3}$$

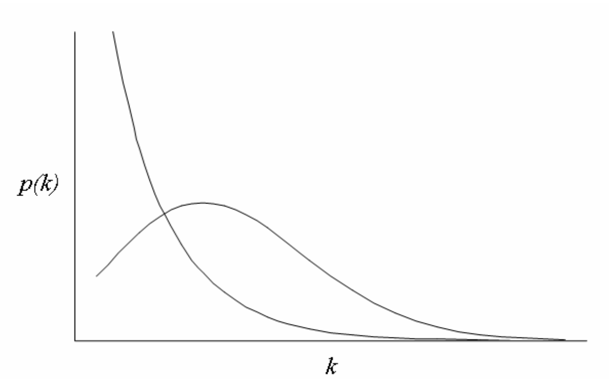
where p_i is the probability of connecting the new node with the i-th existing one, and k_j is the number of connections of the j-th node.

If the number of links per topic were balanced, as in the picture of the "random network" above right, where k_j is approximately constant, then the probability distribution would be uniform and of the order of $1/N$ for each topic. However, the free network follows the degree distribution, at least asymptotically, with probability

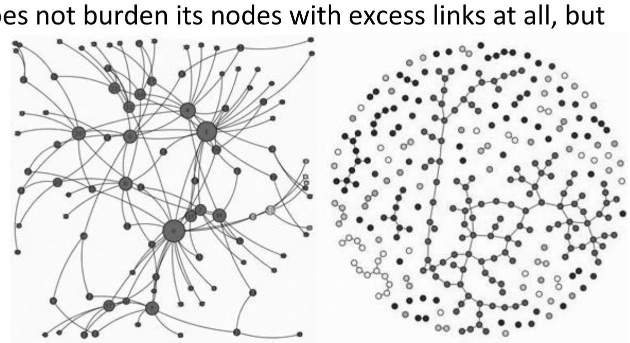
$$p(k) \propto k^{-\gamma} \tag{4}$$

of the nodes with k connections. The parameter γ is usually a number from 2 to 3.

That is why a small number of large intersections stand out against a large number of small ones in the free-growing road network; only some celebrities are very famous unlike many; special books are the bestsellers; about 80 percent of web links point to 15 percent of web pages. All of them form the so-called free networks. Pareto[82] was the first to notice the mentioned legality in the enrichment of the few on the free market of goods and money.



The figure on the left shows the degree function of decreasing the participation of nodes with the number of connections k, along with the bell (Gaussian) distribution of approximately uniform distribution which, it is said, has the third type of network from the previous image. This third type does not burden its nodes with excess links at all, but at the cost of having very long paths between the nodes.



Details on the path lengths through the nets (1) and (2) can be found in the appendices [17] or [18]. The following figure on the right shows once again examples of "free network" on the left and "random" on the right. Even more complex, these networks can be so vague that it becomes difficult to assess which type they belong to.

[82] Vilfredo Pareto (1848-1923), Italian scientist.

If it seems to you that distinguishing a "free" from a "small" or "random" network should be an easy task, I recommend that you look at, for example: The Advantages of Attention Surplus Condition, more commonly known as Attention Deficit Hyperactivity Disorder[83]. The question is whether these advantages can be understood (modeled) by any of the mentioned networks and by which?

In the mentioned case, these are the characteristics: 1. Ability to find alternate paths to overcome obstacles; 2. Being able to see the big picture; 3. Can create order from chaos; 4. Dedicated; 5. Energetic; 6. Flexible – changes as the situation requires; 7. Good in a crisis; 8. Hands-on workers; 9. Idea generator; and other groups listed there. I took the example at random, but the idea is general. Recognition of shape (physical appearance, behavior or character, person or appearance) is required.

Correlation

Let's count the links of the nodes of the first of the networks from the picture at the beginning of the text and arrange the results in sequence in ascending order. It is a "small world network", and the sequence is obtained: $a_0 = 0$, $a_1 = 0$, $a_2 = 0$, $a_3 = 4$, $a_4 = 8$, $a_5 = 4$ and $a_6 = 0$. The notation $a_k = n$ means that the first (a-th) network has n nodes with k connections.

We calculate the average $\overline{a} = \frac{16}{7}$, so we form the sequence $a'_k = a_k - \overline{a}$ to find the correlation: : $a'_0 = a'_1 = a'_2 = a'_6 = -\frac{16}{7}$, $a'_3 = a'_5 = \frac{12}{7}$, $a'_4 = \frac{40}{7}$. The intensity, ie the norm of that sequence is $a' = \sqrt{a'^2_0 + \dots + a'^2_6} = 4\sqrt{26/7} \approx 7.70899$. Concisely:

$$a_0 = 0, \quad a_1 = 0, \quad a_2 = 0, \quad a_3 = 4, \quad a_4 = 8, \quad a_5 = 4, \quad a_6 = 0, \quad \sum = 16, \quad \overline{a} = \frac{16}{7},$$

$$a'_0 = -\frac{16}{7}, a'_1 = -\frac{16}{7}, a'_2 = -\frac{16}{7}, a'_3 = \frac{12}{7}, a'_4 = \frac{40}{7}, a'_5 = \frac{12}{7}, a'_6 = -\frac{16}{7}, a' = 7.70899.$$

We do the same with the number of links of the other, the so-called "free" networks: $b_0 = b_2 = b_4 = 0$, $b_1 = 13$, $b_3 = 1$, $b_5 = 2$ and $b_6 = 1$. Thus $b_5 = 2$ means that the five links have two nodes of the second in a row, the b-th network (scale-free network), which is in the first image. The arithmetic mean of this sequence is $\overline{b} = \frac{17}{7}$, so for $b'_k = b_k - \overline{b}$ we get: $b'_0 = b'_2 = b'_4 = -\frac{17}{7}$, $b'_1 = \frac{74}{7}$, $b'_3 = b'_6 = -\frac{10}{7}$ and $b'_5 = -\frac{3}{7}$. The intensity is $b' = \sqrt{b'^2_0 + \dots + b'^2_6 6} = 6\sqrt{26/7} \approx 11.5635$. Concisely:

$$b_0 = 0, \quad b_1 = 13, \quad b_2 = 0, \quad b_3 = 1, \quad b_4 = 0, \quad b_5 = 2, \quad b_6 = 1, \quad \sum = 17, \quad \overline{b} = \frac{17}{7},$$

$$b'_0 = -\frac{17}{7}, b'_1 = \frac{74}{7}, b'_2 = -\frac{17}{7}, b'_3 = -\frac{10}{7}, b'_4 = -\frac{17}{7}, b'_5 = -\frac{3}{7}, b'_6 = -\frac{10}{7}, b' = 11.5635.$$

For the third ("random" network) nodes, in ascending order by number of connections, there are: $c_0 = c_5 = c_6 = 0$, $c_1 = 3$, $c_2 = 4$, and $c_3 = c_4 = 1$. The mathematical expectation is $\overline{c} = \frac{9}{7}$, so the correlation sequence is: $c'_0 = c'_5 = c'_6 = -\frac{9}{7}$, $c'_1 = \frac{12}{7}$, $c'_2 = \frac{19}{7}$ и $c'_3 = c'_4 = -\frac{2}{7}$. Intensity $c' = 6\sqrt{3/7} \approx 3.92792$. Concisely:

[83] Positives Of ADHD, https://addcoach4u.com/151-positives-of-adhd/

$$c_0 = 0, \quad c_1 = 3, \quad c_2 = 4, \quad c_3 = 1, \quad c_4 = 1, \quad c_5 = 0, \quad c_6 = 0, \quad \sum = 9, \quad \bar{c} = \frac{9}{7},$$

$$c'_0 = -\frac{9}{7}, c'_1 = \frac{12}{7}, c'_2 = \frac{19}{7}, c'_3 = -\frac{2}{7}, c'_4 = -\frac{2}{7}, c'_5 = -\frac{9}{7}, c'_6 = -\frac{9}{7}, c' = 3{,}92792.$$

The correlation coefficients[84] of these networks (arrays) are:

$$r_{ab} = \frac{a'_0 b'_0 + a'_1 b'_1 + \dots + a'_6 b'_6}{a' b'} = \frac{\left(-\frac{16}{7}\right)\left(-\frac{17}{7}\right) + \dots + \left(-\frac{16}{7}\right)\left(-\frac{10}{7}\right)}{7.70899 \times 11{,}5635} = -0.301282.$$

$$r_{ac} = -0.28307 \quad \text{and} \quad r_{bc} = 0.399442.$$

Thanks to the absence of correlation between the networks, we can capture a wider range of cases, applications. Let's look at one example with the distribution of wealth.

According to the Organization for Economic Cooperation and Development (OECD), in 2012 there was 0.6 percent of the world’s adult population with assets of more than one million US dollars, more precisely 42 million of the richest people in the world held 39.3% of the world's wealth. The next 4.4%, or 311 million people, held 32.3% of the world's wealth. So, the lower 95% held only 28.4% of the world's wealth. In particular, the lowest 60 percent of the world's population, in 2012, had the same wealth as the 1,226 richest people on the Forbes list.

It is noticeable that the cited distribution of wealth has the form of the “free network” described here, with the coefficient $b_1 = 13$ which is significantly higher than the others in the b-sequence. This means that the distribution of wealth in 2012 does not have the form of an a-sequence or c, i.e. the network of “small world” or “random”, because the distributions of the three series are significantly different. This further indicates “equality” in capital flows (goods, services, money), ie that the world is mostly a “free market”.

Additional and more precise conclusions that we would draw by interpreting the theorems of “free networks” might be more interesting, but they are not the topic here. Extending the method to social phenomena, to psychology, the physiology of the living world, or physics, one should take into account the limitations of estimation by statistical correlation of sequences. If there is a deeper connection, the correlation will give confirmation, but it will give false “confirmation” sometimes even where there is no connection.

---

[84] 12. Quantity of Options III, Correlation

14. Democracy Evolution

March 21, 2021

Parts of discussions on information perception[85] and consequences on nature and society.

Perception information

- Can you explain the perception information to me with a simple example?

- Yes, on the example of the strength of chess. For example, the first player is good (grade 3) in the opening, very good (4) in the center game, and excellent (5) in the endgame. We arrange the order of these estimates with the vector $\vec{a} = (3,4,5)$. The power of the game is the intensity or norm of the vector, which is here the root of the sum of the squares. Our player has the root of the sum of the squares of his three grades, $\sqrt{50}$, approximately 7, or more precisely:

$$a = |\vec{a}| = \sqrt{\vec{a} \cdot \vec{a}} = \sqrt{3^2 + 4^2 + 5^2} \approx 7{,}071. \tag{1}$$

The second player $\vec{b} = (4,3,2)$ in the opening is very good (4), good (3) is in the center and enough (2) in the endgame. The vigor of his game is the root of 29, approximately 5, or:

$$b = \left|\vec{b}\right| = \sqrt{\vec{b} \cdot \vec{b}} = \sqrt{4^2 + 3^2 + 2^2} \approx 5{,}385. \tag{2}$$

The third player $\vec{c} = (2,3,4)$ has the same "player power" as $\vec{b}$, but the "game power" $\vec{a}$ against $\vec{b}$ and $\vec{a}$ against $\vec{c}$ are not equal. Namely, in the competition of the first and second, the information of perception is:

$$\vec{a} \cdot \vec{b} = 3 \cdot 4 + 4 \cdot 3 + 5 \cdot 2 = 34, \tag{3}$$

and in the competition of the first and third the information of perception is:

$$\vec{a} \cdot \vec{c} = 3 \cdot 2 + 4 \cdot 3 + 5 \cdot 4 = 38. \tag{4}$$

In case (3) the level of the game was lower than in case (4) when we can say that the game was stronger, fiercer, livelier, or more vicious. Heavier play generally has more information of perception, because it is the result of multiplying strings of length $n \in \mathbb{N}$, or vectors:

$$\vec{u} \cdot \vec{v} = (u_1, u_2, \dots, u_n) \cdot (v_1, v_2, \dots, v_n) = u_1 v_1 + u_2 v_2 + \dots + u_n v_n \tag{5}$$

that is greatest if the series are of equal monotony (both increasing or both decreasing). It is the smallest in the case of opposite monotony, one series increasing and the other decreasing, both of the same length. It can be noticed that we measure the strength of the player himself in an analogous way, multiplying the string with itself

$$|\vec{u}|^2 = (u_1, u_2, \dots, u_n) \cdot (u_1, u_2, \dots, u_n) = u_1^2 + u_2^2 + \dots + u_n^2 \tag{6}$$

[85] see [9]

and that there are a lot of linear algebras in these estimates.

- Therefore, the product of arrays is less than or equal to the product of their intensities?

- Yes, bravo, it is the known Schwartz inequality[86] $\vec{a} \cdot \vec{b} \leq |\vec{a}||\vec{b}|$, where the equality is valid if and only if the vectors ($\vec{a}$ and $\vec{b}$) are collinear (parallel) and therefore proportional.

- So in application we need to have some way of evaluating players, and the rest is a calculation?

- That's right, but "evaluation" is not a simple matter, and after "calculation" it is neither an easy task nor the interpretation of the results. In fact, interpreting computational results can be more difficult than computing it. We only need to remember the "problem tasks" from elementary school mathematics, when it was harder for us to form an equation than to solve it, or to look a little further at quantum mechanics. Even after decades of calculations, no result was found there that would not agree with the experiment, and even today we are struggling with the interpretations.

Order and disorder

- What would be the main interpretations of information perception?

- The higher level of the game is "strong on strong, and weak on weak" than the reverse "sharp with weak and humble with strong". That is why the state is in principle stronger than the mafia when it behaves "economically illogically", for example by giving subsidies to beginners and taxing the rich, unlike (thieving) behavior in which it is kidnapped where possible, mostly from the weak, and stays away from the strong. That is why competition on the market is good for society, because it encourages the competition of the powerful, and it would be similar with competition in politics, but I would not talk about that.

- And what about the saying that you shouldn't fight with a horn?

- The information of perception in its basic form ($S = \vec{I} \cdot \vec{H}$) is proportional to the (scalar) product of the vector (series) of "intelligence" and "hierarchy". Here, the former refers to a person's ability, and the latter to external, objective limitations. In the magnitude of perception (S), both factors are equally important, the relative unpredictability ($\vec{I}$) and constraints ($\vec{H}$). The chess master does not win a dunce by hitting him on the head with a wooden board, but with a subtle game.

- What is "relative unpredictability"?

- If a hunter prepares a trap and catches game with it, he is "relatively unpredictable" when he knows what he is doing and what will be, unlike the catch. That is the essence of intelligence, to be one step away from deeper unpredictability – which is essentially information. The theory of information holds that the greater "strength" of the player, i.e. the game, is expressed against more difficult obstacles. It is also intuitively acceptable.

[86] [7], Lemma 1.2.49 (Cauchy–Schwarz inequality), p. 132.

Due to the law of conservation (information), that power should be saved, so we come to something like the former bushido code: “with the weak sofly (cool) and fighting with the strong”. Emphasizing the importance of the factor $\vec{H}$ we will say, the focused power is greater, and it is then directed, that is, organized.

- Hardness (hierarchy) is then as important a contribution to the information of perception as softness (intelligence)?

- That's right. The value of perception information ($S = \vec{I} \cdot \vec{H}$) grows equally with both values, in slang for softness and hardness, i.e. $\vec{I}$ and $\vec{H}$, because the law of commutation, the change of order in this (so-called scalar) multiplication of vectors and in general due to of equal importance two vectors in their product.

Information is expressed by channeling through constraints. Therefore, it is possible to have an organization with a surplus of information in relation to the simple sum of its members[87], because the organization releases latent information. We will use vehicles more efficiently if the traffic in the city is better regulated, or, for example, the ant colony can behave “inexplicably” intelligently with regard to its individuals.

I wrote earlier, if we did not have any wisdom (abilities, intelligence, power of choice) in a situation with a lot of limitations, then not only would nothing be clear to us, but we would not be able to perceive the world around us. The information of perception perceives our world as virtual, formal, or non-essential and dependent on the observer. On the other hand, someone would be extremely smart, but in vain, if he didn't have (descry) obstacles, he wouldn't see anything either.

- I understand. These are the positive sides, and which are the negative ones?

- Greater information of perception means greater vitality, greater “amount of options” of the player or game (depending on what is calculated), and the negative is that we all (living and non-living matter) tend to calm down. In nature, there is a mild, constant and ubiquitous tendency towards less information, communication, and that then means action (because the smallest amounts of information are packed into quantum action, products of energy and time).

This is because more likely outcomes carry less information, and the more likely is more common. Our future is evolving towards more probable conditions, and that means less informative. The world is evolving towards more order! I emphasize this because modern physics believes otherwise.

Namely, the entropy is interpreted as a “mess”, which comes from observing a glass that falls from the table and breaks in heart that flies all over the floor and the housewife then has to pick up that mess. However, the theory of information views the same process in the opposite way; it views the molecules that tend to be evenly distributed as soldiers on the lookout. With uniformity, they become impersonal and thus lose (emit less) information!

[87] Latent Information, https://www.academia.edu/44725794/Latent_Information

The growth of entropy is a loss (of emission) of information, and only then does its growth not contradict the "probability principle" – that more probable phenomena are more frequent.

So, by increasing the "information of perception", vitality increases, but it is an unnatural process. It is like a geyser or a volcano that spews out its contents in defiance of the otherwise mild, constant and ubiquitous gravity of the earth, but which sooner or later goes down again. This opposite tendency creates tensions and the danger that too strong players will tear themselves or the game system. That is how Yugoslavia was broken up, or the World Boxing Federation was divided into three, and the world was constantly trying to be divided into two parts, East and West.

Not only we humans, but also all living beings, even inanimate physical substances, obey the "principle of least action", which is the opposite of "liveliness". Because of the same, we like to get rid of our excesses of freedom (amounts of uncertainty, options), so to submit, unite, and organize. We love safety or efficiency, because in addition to the desire to live (to participate in a good game), there is also the desire not to live (to calm down).

Modeling

- What about computer simulations, can something be learned from them about the "evolution of democracy"?

- Perhaps, when the notion of "democracy" is specified, then the parameters and causes of its "development". Is it better to evolve into the brave or the obedient, into suffering or into dullness, can the state of equality be maintained "by grace or force". For the sake of simplicity, let us limit ourselves to something like the latter, to the modern model of society that we want to strive for equality, abundance, peace and order. We will be surprised how much you can get from such a "small" framework.

When, on the one hand, we have equality, which maximizes information, and on the other hand, principled minimalism, it is clear that they must strengthen tensions over time and stratification. The belief that the state can constantly weigh down the system of equality and human rights is equal to the belief that with a healthy lifestyle and regular visits to the doctor, we can live indefinitely.

It is already clear from the logic of free networks that a society dedicated to preserving equality and human rights will sooner or later fail because of other values. When communications (goods, money, power, initiative) are free, then rare concentrators are formed with many links, versus many nodes that have few links, with a growing separation of the two classes. This is similar to the free market situation today.

However, capital and the power that money carries cannot be the only threat of division, nor does it have to be the main one, as it seems to us today, but it is a general view that the status quo is unsustainable in the long run. A legally regulated society of "people of equal chances" will shoot at the seams because of other importance. It will stratify in accordance with the "universe of uncertainty" and, therefore, perhaps in ways unknown at this time.

A society with strong internal cohesion, like an imagined future global and well-connected, will stratify in depth. That would be the path to some form of “slave-owning system” from which we have already seen similar ones.

For example, the classes of rich and poor today, or previous mass feudal divisions, or those less classical divisions into slaves and masters known to us from the United States, South Africa, all the way to ancient Sparta, will still have problems with duration. In the end, every organization faces a lag in relation to the rest of the world, which is moving away, be it poverty, external threats, the white plague, because the organization means a surplus of information and thus an unstable situation that tends to decrease.

A society with weak internal cohesion will stratify like the evolution of living species on earth.

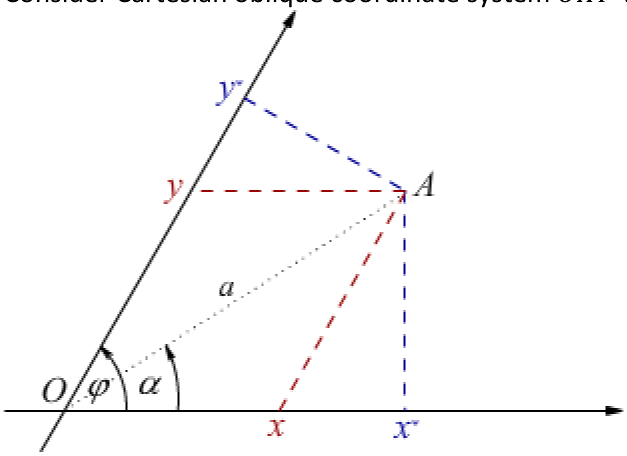
15. Variant Vectors

March 24, 2021

Simple explanations of covariant and contravariant vectors and their significance, using an oblique system of inclination φ, rotation by angle θ, multiplication of vectors and bases.

Коси систем

Consider Cartesian oblique coordinate system OXY with angle φ between abscissa and ordinate and in it point A. Its distance from the origin is $a = \overline{OA}$, parallel projections of A fall on "contravariant" coordinates (x, y), and the orthogonal on "covariant" (x^*, y^*).



Point A does not have to be on the angle bisector φ, but the angles with parallel arms are equal, ie. $\sphericalangle(x^*xA) = \sphericalangle(y^*yA) = \varphi$, so we find:

$$x^* - x = y\cos\varphi,\ \ y^* - y = x\cos\varphi.$$

Hence the projection transformations:

$$\begin{cases} x^* = x + y\cos\varphi \\ y^* = x\cos\varphi + y \end{cases} \tag{1}$$

These are transformations of contravariant coordinates into covariant ones. The reverse is calculated:

$$x = \frac{x^* - y^*\cos\varphi}{\sin^2\varphi},\ \ y = \frac{-x^*\cos\varphi + y^*}{\sin^2\varphi}. \tag{2}$$

The triangle OxA and the cosine theorem give the square of the distance of the point from the origin

$$a^2 = x^2 + y^2 + 2xy\cos\varphi, \tag{3}$$

in contravariant coordinates. By substituting (2) into (3) we get

$$a^2 = [(x^*)^2 + (y^*)^2 - 2x^*y^*\cos\varphi]/\sin^2\varphi. \tag{4}$$

which is a square of the same length (3), now expressed by covariant coordinates.

The angle between the abscissa and the given point seen from the origin $\sphericalangle(x^*OA) = \alpha$. The triangle Ox^*A is right-angled and $\cos\alpha = x^*/a$, or $\cos(\varphi - \alpha) = y^*/a$. From there:

$$x^* = a\cos\alpha,\ \ y^* = a\cos(\varphi - \alpha). \tag{5}$$

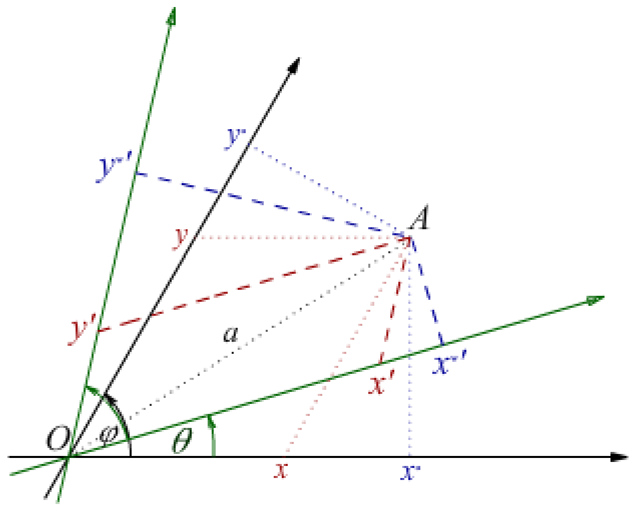
The double area of the triangle OxA are $ax\sin\alpha$ and $ay\sin(\varphi - \alpha)$. They are exactly equal to the area of the parallelogram $OxAy$ which is $xy\sin\varphi$. Equalizations give:

$$x = \frac{a\sin(\varphi - \alpha)}{\sin\varphi},\ \ y = \frac{a\sin\alpha}{\sin\varphi}, \tag{6}$$

and these are the expressions for the contravariant coordinates, these slightly more complex than the covariant ones (5).

Rotation

Let us now consider the rotation of a given system OXY into the system $OX'Y'$ around the origin O for the angle θ, as in the figure on right.



Denote $\alpha' = \sphericalangle(x'OA)$, so (6) gives:

$$x' = a\frac{\sin(\varphi - \alpha')}{\sin\varphi}$$
$$= a\frac{\sin[(\varphi - \alpha) + \theta]}{\sin\varphi}$$
$$= a\frac{\sin(\varphi - \alpha)\cos\theta + \cos(\varphi - \alpha)\sin\theta}{\sin\varphi}$$
$$= a\frac{\sin(\varphi - \alpha)}{\sin\varphi}\cos\theta + a\frac{\cos(\varphi - \alpha)\sin\theta}{\sin\varphi}$$
$$= x\cos\theta + a\frac{\cos\varphi\cos\alpha + \sin\varphi\sin\alpha}{\sin\varphi}\sin\theta.$$

Addition formulas were used:

$$\sin(u + v) = \sin u\cos v + \cos u\sin v,$$
$$\cos(u - v) = \cos u\cos v + \sin u\sin v.$$

From (5) and (1) follows $a\cos\alpha = x + y\cos\varphi$, so we calculate further[88]:

$$x' = x\cos\theta + [(x + y\cos\varphi)\operatorname{ctg}\varphi + a\sin\alpha]\sin\theta$$
$$= x\cos\theta + (x\operatorname{ctg}\varphi + y\cos\varphi\operatorname{ctg}\varphi)\sin\theta + y\sin\varphi\sin\theta$$
$$= x(\cos\theta + \operatorname{ctg}\varphi\sin\theta) + y(\cos\varphi\operatorname{ctg}\varphi + \sin\varphi)\sin\theta$$
$$= x\frac{\sin\varphi\cos\theta + \cos\varphi\sin\theta}{\sin\varphi} + y\frac{\cos^2\varphi + \sin^2\varphi}{\sin\varphi}\sin\theta$$
$$= x\frac{\sin(\varphi + \theta)}{\sin\varphi} + y\frac{\sin\theta}{\sin\varphi}$$

therefore

$$x'\sin\varphi = x\sin(\varphi + \theta) + y\sin\theta. \qquad (7)$$

Similarly, starting from the second equation (6) for the system $OX'Y'$, we obtain:

$$y' = \frac{a\sin\alpha'}{\sin\varphi} = a\frac{\sin(\alpha - \theta)}{\sin\varphi} = a\frac{\sin\alpha\cos\theta - \cos\alpha\sin\theta}{\sin\varphi}$$
$$= y\cos\theta - (x + y\cos\varphi)\frac{\sin\theta}{\sin\varphi} = -x\frac{\sin\theta}{\sin\varphi} + y\frac{\sin\varphi\cos\theta - \cos\varphi\sin\theta}{\sin\varphi}$$

[88] $\operatorname{ctg}\phi = \cot\phi$ – cotangent angle ϕ

$$= -x\frac{\sin\theta}{\sin\varphi} + y\frac{\sin(\varphi-\theta)}{\sin\varphi}$$

therefore

$$y'\sin\varphi = -x\sin\theta + y\sin(\varphi-\theta). \tag{8}$$

Equations (7) and (8) are written briefly in matrix

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \frac{1}{\sin\varphi}\begin{pmatrix} \sin(\varphi+\theta) & \sin\theta \\ -\sin\theta & \sin(\varphi-\theta) \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}, \tag{9}$$

that is, $A' = R'(\varphi,\theta)A$. These are transformations $OXY \to OX'Y'$ of contravariant coordinates of the Cartesian oblique system, slope $\varphi = \sphericalangle(XOY)$, on rotation around the origin O for the angle θ.

Of course, formulas (9) can also be derived geometrically, from the given figure at the top right.

Example 1. For a rectangular system, with $\varphi = \frac{\pi}{2}$ and $\sin\varphi = 1$, $\sin(\varphi \pm \theta) = \cos\theta$, from (9) we get:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}. \tag{10}$$

These are well-known formulas for the rotation of Cartesian rectangular 2D coordinate system. □

Example 2. By rotating the system, the distance of a given point from the origin remains the same and the contravariant expression (3) retains the form:

$$a^2 = (x')^2 + (y')^2 + 2x'y'\cos\varphi$$

$$= \left[\frac{x\sin(\varphi+\theta)+y\sin\theta}{\sin\varphi}\right]^2 + \left[\frac{-x\sin\theta+y\sin(\varphi-\theta)}{\sin\varphi}\right]^2 + 2\cdot\frac{x\sin(\varphi+\theta)+y\sin\theta}{\sin\varphi}\cdot\frac{-x\sin\theta+y\sin(\varphi-\theta)}{\sin\varphi}\cdot\cos\varphi$$

$$= \frac{x^2}{\sin^2\varphi}[\sin^2(\varphi+\theta) + \sin^2\theta - 2\sin(\varphi+\theta)\sin\theta\cos\varphi] +$$

$$+\frac{y^2}{\sin^2\varphi}[\sin^2\theta + \sin^2(\varphi-\theta) + 2\sin\theta\sin(\varphi-\theta)\cos\varphi] +$$

$$+\frac{2xy}{\sin^2\varphi}[\sin(\varphi+\theta)\sin\theta - \sin\theta\sin(\varphi-\theta) + \sin(\varphi+\theta)\sin(\varphi-\theta)\cos\varphi - \sin^2\theta\cos\varphi]$$

however, the factor next to x^2 in the numerator is:

$$\sin^2(\varphi+\theta) + \sin^2\theta - 2\sin(\varphi+\theta)\sin\theta\cos\varphi =$$

$$= [\sin^2(\varphi+\theta) - \sin(\varphi+\theta)\sin\theta\cos\varphi] + [\sin^2\theta - \sin(\varphi+\theta)\sin\theta\cos\varphi]$$

$$= \sin(\varphi+\theta)[\sin(\varphi+\theta) - \sin\theta\cos\varphi] + [\sin^2\theta - \sin(\varphi+\theta)\sin\theta\cos\varphi]$$

$$= \sin(\varphi+\theta)\sin\varphi\cos\theta + [\sin^2\theta - \sin(\varphi+\theta)\sin\theta\cos\varphi]$$

$$= \sin(\varphi+\theta)\,[\sin\varphi\cos\theta - \sin\theta\cos\varphi] + \sin^2\theta$$

$$= \sin^2\varphi\cos^2\theta - \cos^2\varphi\sin^2\theta + \sin^2\theta$$

$$= \sin^2\varphi\cos^2\theta + (1-\cos^2\varphi)\sin^2\theta = \sin^2\varphi$$

and this is shortened with the denominator and only x^2 remains. Similarly, the coefficient with y^2 is one, so it remains to calculate the coefficient with mixed xy. It is easy to check that in angular brackets the sum is reduced to $\sin^2\varphi\cos\varphi$, the first factor (sine square) is shortened by the denominator and expression (3) remains. □

Example 3. Prove the transformations:

$$x'^* = x\cos\theta + y\cos(\varphi-\theta), \tag{11}$$

$$y'^* = x\cos(\varphi+\theta) + y\cos\theta. \tag{12}$$

These equations express the covariant coordinates of the rotated system using the contravariant non-rotated ones, followed, for example, by the corresponding (1) by including (9). We prove the first:

$$x'^* = x' + y'\cos\varphi =$$

$$= \frac{x\sin(\varphi+\theta) + y\sin\theta}{\sin\varphi} + \frac{-x\sin\theta + y\sin(\varphi-\theta)}{\sin\varphi}\cos\varphi$$

$$= x\frac{\sin(\varphi+\theta) - \sin\theta\cos\varphi}{\sin\varphi} + y\frac{\sin\theta + \sin(\varphi-\theta)\cos\varphi}{\sin\varphi}$$

$$= x\frac{\sin\varphi\cos\theta}{\sin\varphi} + y\frac{\sin[(\theta-\varphi)+\varphi] - \sin(\theta-\varphi)\cos\varphi}{\sin\varphi}$$

$$= x\cos\theta + y\frac{\cos(\theta-\varphi)\sin\varphi}{\sin\varphi}$$

and hence (11). In the corresponding second equation (1) we include (9) and find:

$$y'^* = x'\cos\varphi + y' =$$

$$= \frac{x\sin(\varphi+\theta) + y\sin\theta}{\sin\varphi}\cos\varphi + \frac{-x\sin\theta + y\sin(\varphi-\theta)}{\sin\varphi}$$

$$= x\frac{\sin(\varphi+\theta)\cos\varphi - \sin\theta}{\sin\varphi} + y\frac{\sin\theta\cos\varphi + \sin(\varphi-\theta)}{\sin\varphi}$$

$$= x\frac{\sin(\varphi+\theta)\cos\varphi - \sin[(\varphi+\theta)-\varphi]}{\sin\varphi} + y\frac{\sin\varphi\cos\theta}{\sin\varphi}$$

$$= x\frac{\cos(\varphi+\theta)\sin\varphi}{\sin\varphi} + y\cos\theta$$

thus proving (12). □

Example 4. Prove the transformations of covariant coordinates:

$$x'^* = x^*\frac{\sin(\varphi-\theta)}{\sin\varphi} + y^*\frac{\sin\theta}{\sin\varphi} \tag{13}$$

$$y'^* = -x^*\frac{\sin\theta}{\sin\varphi} + y^*\frac{\sin(\varphi+\theta)}{\sin\varphi} \tag{14}$$

using examples 3. and (2).

Substituting (2) into (11) we get:

$$x'^* = \frac{x^* - y^*\cos\varphi}{\sin^2\varphi}\cos\theta + \frac{-x^*\cos\varphi + y^*}{\sin^2\varphi}\cos(\varphi-\theta)$$

$$= x^*\frac{\cos\theta - \cos\varphi\cos(\varphi-\theta)}{\sin^2\varphi} + y^*\frac{-\cos\varphi\cos\theta + \cos(\varphi-\theta)}{\sin^2\varphi}$$

$$= x^*\frac{\cos[\varphi-(\varphi-\theta)] - \cos\varphi\cos(\varphi-\theta)}{\sin^2\varphi} + y^*\frac{\sin\varphi\sin\theta}{\sin^2\varphi}$$

$$= x^*\frac{\sin\varphi\sin(\varphi-\theta)}{\sin^2\varphi} + y^*\frac{\sin\theta}{\sin\varphi}$$

$$= x^*\frac{\sin(\varphi-\theta)}{\sin\varphi} + y^*\frac{\sin\theta}{\sin\varphi}$$

thus proving (13). By changing (2) to (12) we get:

$$y'^* = \frac{x^* - y^*\cos\varphi}{\sin^2\varphi}\cos(\varphi+\theta) + \frac{-x^*\cos\varphi + y^*}{\sin^2\varphi}\cos\theta$$

$$= x^*\frac{\cos(\varphi+\theta) - \cos\varphi\cos\theta}{\sin^2\varphi} + y^*\frac{-\cos\varphi\cos(\varphi+\theta) + \cos\theta}{\sin^2\varphi}$$

$$= x^*\frac{-\sin\varphi\sin\theta}{\sin^2\varphi} + y^*\frac{-\cos\varphi\cos(\varphi+\theta) + \cos[(\varphi+\theta)-\varphi]}{\sin^2\varphi}$$

$$= -x^*\frac{\sin\theta}{\sin\varphi} + y^*\frac{\sin(\varphi+\theta)\sin\varphi}{\sin^2\varphi}$$

and hence (14). □

Example 5. In the equality of example 4, we include (1) and derive (11) and (12).

Starting from (13) and including (1) we get:

$$x'^* = x^* \frac{\sin(\varphi - \theta)}{\sin\varphi} + y^* \frac{\sin\theta}{\sin\varphi}$$

$$= (x + y\cos\varphi)\frac{\sin(\varphi - \theta)}{\sin\varphi} + (x\cos\varphi + y)\frac{\sin\theta}{\sin\varphi}$$

$$= x\frac{\sin(\varphi - \theta) + \cos\varphi\sin\theta}{\sin\varphi} + y\frac{\cos\varphi\sin(\varphi - \theta) + \sin\theta}{\sin\varphi}$$

$$= x\frac{\sin\varphi\cos\theta}{\sin\varphi} + y\frac{\cos\varphi\sin(\varphi - \theta) + \sin[\varphi - (\varphi - \theta)]}{\sin\varphi}$$

$$= x\cos\theta + y\frac{\sin\varphi\cos(\varphi - \theta)}{\sin\varphi}$$

$$= x\cos\theta + y\cos(\varphi - \theta)$$

and thus formula (11) is obtained. Starting from (14) and substituting (1), we find:

$$y'^* = -x^* \frac{\sin\theta}{\sin\varphi} + y^* \frac{\sin(\varphi + \theta)}{\sin\varphi}$$

$$= -(x + y\cos\varphi)\frac{\sin\theta}{\sin\varphi} + (x\cos\varphi + y)\frac{\sin(\varphi + \theta)}{\sin\varphi}$$

$$= x\frac{-\sin\theta + \cos\varphi\sin(\varphi + \theta)}{\sin\varphi} + y\frac{-\cos\varphi\sin\theta + \sin(\varphi + \theta)}{\sin\varphi}$$

$$= x\frac{-\sin[(\varphi + \theta) - \varphi] + \cos\varphi\sin(\varphi + \theta)}{\sin\varphi} + y\frac{\sin\varphi\cos\theta}{\sin\varphi}$$

$$= x\frac{\cos(\varphi + \theta)\sin\varphi}{\sin\varphi} + y\cos\theta$$

$$= x\cos(\varphi + \theta) + y\cos\theta$$

and that is formula (12). □

Product

The square of the distance of a given point from the origin, $a = \overline{AO}$, is equal to the scalar product of its covariant and contravariant coordinates

$$a^2 = x^* x + y^* y. \tag{15}$$

Namely, the equation (3) follows from:

$$x^*x + y^*y = (x + y\cos\varphi)x + (x\cos\varphi + y)y = x^2 + y^2 + 2xy\cos\varphi = a^2.$$

However, multiplying the contravariant with the contravariant we get a simple sum of squares which in the case of an oblique system does not correspond to the square of the distance, $x^2 + y^2 \neq a^2$, but the product of the covariant does not correspond to it either.

The distance from point $A(x, y)$ to point $B(x + \Delta x, y + \Delta y)$ is

$$\overline{AB}^2 = (\Delta x^*)(\Delta x) + (\Delta y^*)(\Delta y), \tag{16}$$

where x^* and y^* denote the covariant coordinates. Namely, by translating the system for the vector $\overrightarrow{OB}$, so that its starting point is point B, the equation (16) becomes (15). The same is true for infinitesimal lengths.

Due to this peculiarity of coordinate multiplication, and due to the way of matrix multiplication, we write covariant vectors as a matrix-row, and contravariant as a matrix-column. Thus (15) becomes:

$$a^2 = x^*x + y^*y = (x^* \quad y^*)\begin{pmatrix} x \\ y \end{pmatrix} = (x^* \quad y^*)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}. \tag{17}$$

That this also applies in a rotated system is confirmed by applying (13), (14) and (9):

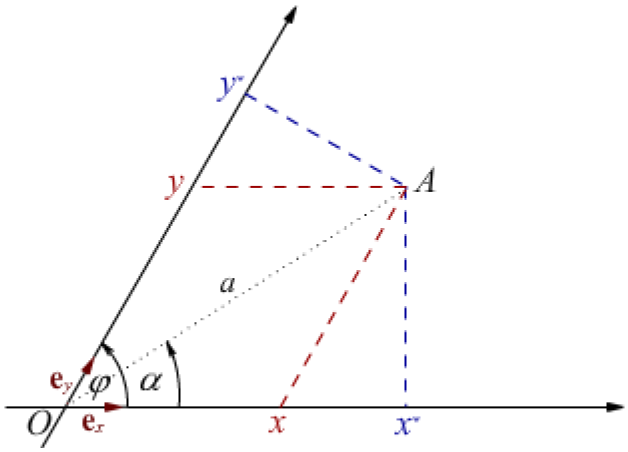
$$(x'^* \quad y'^*)\begin{pmatrix} x' \\ y' \end{pmatrix} = \tag{18}$$

$$= \left[(x^* \quad y^*)\begin{pmatrix} \frac{\sin(\varphi-\theta)}{\sin\varphi} & -\frac{\sin\theta}{\sin\varphi} \\ \frac{\sin\theta}{\sin\varphi} & \frac{\sin(\varphi+\theta)}{\sin\varphi} \end{pmatrix}\right]\left[\begin{pmatrix} \frac{\sin(\varphi+\theta)}{\sin\varphi} & \frac{\sin\theta}{\sin\varphi} \\ -\frac{\sin\theta}{\sin\varphi} & \frac{\sin(\varphi-\theta)}{\sin\varphi} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}\right]$$

$$= (x^* \quad y^*)\left[\begin{pmatrix} \frac{\sin(\varphi-\theta)}{\sin\varphi} & -\frac{\sin\theta}{\sin\varphi} \\ \frac{\sin\theta}{\sin\varphi} & \frac{\sin(\varphi+\theta)}{\sin\varphi} \end{pmatrix}\begin{pmatrix} \frac{\sin(\varphi+\theta)}{\sin\varphi} & \frac{\sin\theta}{\sin\varphi} \\ -\frac{\sin\theta}{\sin\varphi} & \frac{\sin(\varphi-\theta)}{\sin\varphi} \end{pmatrix}\right]\begin{pmatrix} x \\ y \end{pmatrix}$$

$$= (x^* \quad y^*)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = a^2$$

from which follows (17) for a rotated system. Matrix multiplication is not commutative but is associative, and the latter was used in the proof.



Let $\mathbf{e}_x$ and $\mathbf{e}_y$ be the unit vectors of abscissa and ordinate, shown in the figure on the left, i.e. $\vec{e}_x$ and $\vec{e}_y$. Then the vector $\overrightarrow{OA} = x\vec{e}_x + y\vec{e}_y$ represents a point $A(x, y)$ in its vertex.

The translations defined by the base vectors $\vec{e}_x$ and $\vec{e}_y$ behave contravariantly, while the projections of the vector $\overrightarrow{OA}$ are covariant.

Therefore, transformations (9) are used to change the base of the system (vectors such as $\mathbf{e}_x$ and $\mathbf{e}_y$), and transformations (13) and (14) from example 4, are for transformations of vector components (such as x and y). The first ones are contravariant, unlike the other covariant ones, and the differences between them disappear in a rectangular system, when $\varphi = 90°$.

In other words, working as in proof (18), we can show that on the rotation (of the same oblique system, slope φ for angle θ) the vector remains unchanged. Namely, after the rotation:

$$\overrightarrow{OA'} = x'\vec{e}_{x'} + y'\vec{e}_{y'} = (x' \quad y')\begin{pmatrix}\vec{e}_{x'} \\ \vec{e}_{y'}\end{pmatrix} \tag{19}$$

$$= (x \quad y)\begin{pmatrix}\frac{\sin(\varphi-\theta)}{\sin\varphi} & -\frac{\sin\theta}{\sin\varphi} \\ \frac{\sin\theta}{\sin\varphi} & \frac{\sin(\varphi+\theta)}{\sin\varphi}\end{pmatrix}\begin{pmatrix}\frac{\sin(\varphi+\theta)}{\sin\varphi} & \frac{\sin\theta}{\sin\varphi} \\ -\frac{\sin\theta}{\sin\varphi} & \frac{\sin(\varphi-\theta)}{\sin\varphi}\end{pmatrix}\begin{pmatrix}\vec{e}_x \\ \vec{e}_y\end{pmatrix}$$

$$= (x \quad y)\begin{pmatrix}1 & 0 \\ 0 & 1\end{pmatrix}\begin{pmatrix}\vec{e}_x \\ \vec{e}_y\end{pmatrix} = (x \quad y)\begin{pmatrix}\vec{e}_x \\ \vec{e}_y\end{pmatrix} = x\vec{e}_x + y\vec{e}_y = \overrightarrow{OA}$$

so $\overrightarrow{OA'} = \overrightarrow{OA}$, which should have been shown.

It is easy to understand that the information of perception, $S = ax + by + cz + \cdots$, should be treated as a multiplication of a covariant and a contravariant sequence, if we want its value not to depend on the choice of the coordinate system.

Epilogue

This was my attempt to use the methods of elementary mathematics (which is taught in high schools) to explain perhaps the too difficult problem of "variant vectors" (co-variant and counter-variant transformations), otherwise of tensor calculus. I hope that this unusual introduction to differential calculus and multilinear algebra will be interesting to connoisseurs as well.

16. Variant Vectors II

March 31, 2021

Introduction

- What is the difference between your information science and the classical materialist concept of the world, apart from information, or the amount of options, as the ontological basis of the world[89]?

- Okay, I understand and I'll try to answer. But in order to interpret something typical for information perception philosophically and to some extent qualitatively, several explanations from linear algebra are necessary. The first is a description of the vector and its dual vector space.

When we add series of numbers (scalars) of the same length n (some natural number) so that the first is added to the first, the second to the second and then the n-th with the n-th, then we can consider them as vectors of n-dimensional space. When the image of the sum of vectors is equal to the sum of the images of these vectors, the mapping is called linear. It is clear that such images are also some vectors, and if they are the same type of numbers that make up the components of the original vectors, then the mapping is called a functional[90].

Thus, a string x of given length n forms a vector of the space X. A linear functional y^* maps it into a string of the same length $y^*(x)$ which belongs to some dual, say *adjoint* (associated) vector space X^* (any functional among more them, in a special adjuvant space). We call the first vectors (X) countervariant, the second (X^*) covariant, and they, in relation to a given functional, stand as two sides of the same coin, as a figure and an image in a mirror. It is a matter of pure mathematics, that is, a part that is usually called "higher linear algebra". The good thing is that this field is very developed, it has enough theorems, it is very accurate and, at least basically, it is not the most difficult in mathematics.

An inconvenient feature of more linear algebra is that, for now, there are not enough serious interpretations in information theory. I fill that gap.

Anyway, when arrays of $(a, b, \dots, c)$ with a given n components form a vector space X, then images of these components $(x, y, \dots, z)$, numbers of the same type obtained by linear mapping, form vectors of dual space X^*. Due to the given mapping, the originals are domain (set of departure) vectors, the copies are codomain vectors, the former are the properties of the individual and the latter are its constraints.

It is easily proved that the action of the linear functional y^* on an arbitrary vector x of the space X is completely known, if we know how it acts on some base of that space. If we know the numbers (scalars) that are images of the base vectors of the space X then we know to represent each vector of the dual space X^*. Also, knowing the images of the base, the functional y^* on X is completely and unambiguously defined.

[89] A colleague asked me a question anonymously.
[90] a function from a vector space into the scalar field

Further things become “simple” about perception information. The space X of the individual's ability is associated with the dual space X^* of the constraint. The effect of the hierarchy y^* on arbitrary responses of intelligence x is completely known, if we know how it affects some basic options. The base could be any complete set of independent situations.

Unlike the usual materialistic view, the observation of the world through these dual vectors becomes subjective. Since dual of the dual vector space is the vector space again, it does not matter whether we join intelligence (I) to the hierarchy (H) to calculate perception information, get it as a scalar product of two vectors ($S = I \cdot H$), or do the opposite, first to have some hierarchy to which we will then join intelligence as dual. It is also possible, thirdly, to associate a given intelligence of a person with the intelligence of another person, which is dual in contrast to it, or fourthly, to oppose a hierarchy to a hierarchy. The consistency of the attitudes of algebra is a guarantee of the consistency of these combinations of interpretations.

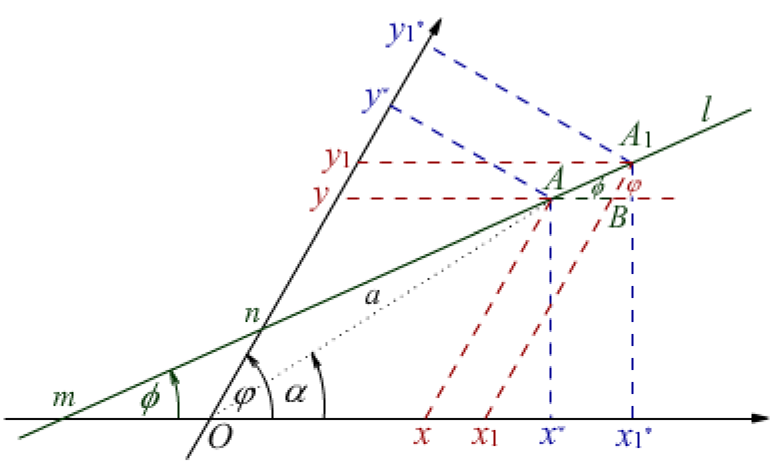
On the other hand, the scalar product of contravariant and covariant vectors is invariant. See what exactly this means in the previous appendix[91], where before the end it says: “It is easy to understand that the information of perception, $S = ax + by + \dots + cz$, should be treated as multiplying a contravariant and covariant sequence, if we want its value not to depend on system choice coordinate.” This in all its subjectivity of this world (seen through the “theory of information perception”) means certain objectivity.

Complex plane

If X and Y are two vector spaces over a scalar field Φ, then the *linear operator* $L : X \to Y$ means a function defined on X with values in Y, if

$$L(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 L(x_1) + \alpha_2 L(x_2) \quad (x_1, x_2 \in X;\quad \alpha_1, \alpha_2 \in \Phi). \tag{1}$$

As usual, we say that the originals (X) form the domain, and the images (Y) the codomain mappings (L).



A typical linear function is $y(x) = kx + n$. Its graph is line l, which is shown in the picture on the left. We continue the previous applica-tion, so the picture is a continuation of the first picture of that appendix. The points $A(x, y)$ and $A_1(x_1, y_1)$ are on the line l and with the point $B(x_1, y)$ form an oblique triangle.

It follows from $y(0) = n$ that n is the ordinate of the intersection of the l and y-axes. However, from:

$$y_1 - y = k(x_1 - x), \quad \Delta y = k\Delta x$$

---

[91] 15. Variant Vectors

substitute $\Delta l = \overline{AA_1}$ and by cosine law we find $\Delta l^2 = \Delta x^2 + \Delta y^2 + 2\Delta x \Delta y \cos\varphi$, and that is (3) of the mentioned appendix, the distance between two points in contravariant coordinates.

Note that the coefficient $k \ne \tan\phi$ of the line l is not a tangent[92] of the angle between the x-axis and the given line, except when the Cartesian coordinate system OXY is rectangular. In this special case, if $\varphi = 90°$, the cosine law, ie the expression for the square of the distance Δl^2, becomes Pythagoras' theorem. But, from the picture we read, $\Delta x^* = \Delta l \ \cos\phi$, where $\Delta x^* = x_1^* - x^*$ is the difference of the covariant coordinates of the abscissa, and the angle between the abscissa and the given line is $\phi = \sphericalangle(xOl)$.

The normalized form of the line l in a given oblique system is

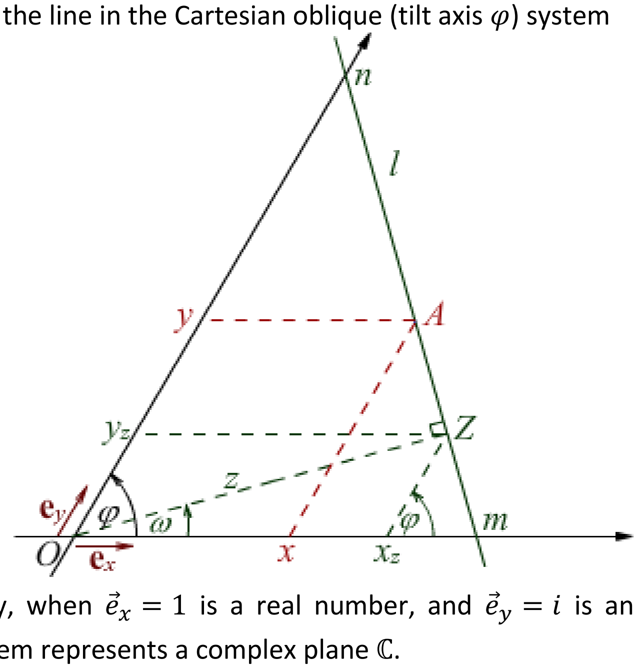
$$\frac{x}{m} + \frac{y}{n} = 1, \tag{2}$$

where the numbers $m, n \in \Phi$ are the intersection points with the x and y axes, as seen in the previous figure. In the following figure (right), the line l intersects the first quadrant ($\phi > 90°$) and leg $z = \overline{OZ}$ is perpendicular to it, where $Z(x_z, y_z) \in l$, $\omega = \sphericalangle(x_z OZ)$. The triangle mOZ is right-angled with a right angle at the vertex Z, so $z = m\cos\omega$. Similarly, observing the triangle nOZ, we find $z = n\cos(\varphi - \omega)$. Hence (2) is:

$$\frac{1}{m} = \frac{\cos\omega}{z}, \ \frac{1}{n} = \frac{\cos(\varphi-\omega)}{z},$$

$$x\cos\omega + y\cos(\varphi - \omega) = z. \tag{3}$$

This is the normal (vertical) form of the equation of the line in the Cartesian oblique (tilt axis φ) system OXY coordinates. The points $A(x, y)$ and $Z(x_z, y_z)$ are in contravariant, they are reduced to axes by parallel displacement, translation.



Let $\mathbf{e}_x$ and $\mathbf{e}_y$ be unit vectors of the axes, as in the figure, or the notations $\vec{e}_x$ and $\vec{e}_y$. Then it is

$$\overrightarrow{OA} = x\vec{e}_x + y\vec{e}_y$$

and that is what we saw in the previous article. The novelty here is that the point $A(x, y)$, on the line l distant by $t = \overline{OT}$ from the origin, with the normal to it from O, the angle ω to the abscissa, satisfies equation (3).

When $\varphi = 90°$ equation (3) becomes the known normal form of the equation of the line in the Cartesian rectangular coordinates. Then especially, when $\vec{e}_x = 1$ is a real number, and $\vec{e}_y = i$ is an imaginary unit ($i^2 = -1$), the given coordinate system represents a complex plane $\mathbb{C}$.

[92] $\tan\phi = \text{tg}\,\phi$ – tangent of the angle fi.

Rotation of oblique systems

Let two Cartesian oblique systems OXY and $OX'Y'$ be given, with the same origin O, with angles $\varphi = \sphericalangle(XOY)$, $\varphi' = \sphericalangle(X'OY')$ and $\theta = \sphericalangle(XOX')$. The coordinates (x, y) and (x', y') of one point in two systems are connected by the relations:

$$\begin{cases} x = \frac{x' \sin(\varphi-\theta)+y' \sin(\varphi-\varphi'-\theta)}{\sin\varphi} \\ \quad y = \frac{x' \sin\theta+y' \sin(\varphi'+\theta)}{\sin\varphi} \end{cases} \tag{4}$$

This is the rotation $(x', y') \to (x, y)$ for the angle θ between the abscissa, arbitrary oblique systems of inclination φ and φ'. I cite it only to supplement the previous attachment.

The inverse rotation $(x, y) \to (x', y')$ is obtained from (4) by changing θ with $-\theta$, when, for $\varphi = \varphi'$, it becomes the transformation (9) from the previous appendix. By a similar change of the sign of the angle, we translate the other rotations from that appendix into inverse, into shapes more convenient for changing the variables of equation (3) here.

Functional

When a function together with the coefficients of a vector takes values from the same scalar field Φ it is called a *functional*. The values of each individual linear functional, such as (1), taken on the vectors of a given domain, satisfy the properties of the vector spaces. Moreover, the set of all linear functional defined on an n-dimensional vector space X, with common addition and multiplication by a scalar, is also a vector space. We denote the latter by X^* and call it an *adjoint* (associated) or dual space of the space X.

Using the base $e_1, e_2, \dots, e_n \in X$, we write an arbitrary vector $x \in X$

$$x = \sum_{k=1}^{n} \xi_k e_k$$

so for an arbitrary functional $y^* \in X^*$ we get:

$$y^*(x) = y^*\left(\sum_{k=1}^{n} \xi_k e_k\right) = \sum_{k=1}^{n} \xi_k y^*(e_k)$$

from which it can be seen that the action of the linear functional y^* on an arbitrary vector x is completely known, if we know how it acts on a base (different bases can define the same space), if we know scalars

$$\eta_k = y^*(e_k). \tag{5}$$

We write with these scalars

$$y^*(x) = \sum_{k=1}^{n} \xi_k \eta_k$$

which becomes the general form of a linear functional on X. Conversely, if $\eta_1, \eta_2, \dots, \eta_n \in \Phi$ are given scalars, then this equality in the base $e_1, e_2, \dots, e_n \in X$ completely and unambiguously defines the functional y^* on X.

We interpret the same equality by information of perception[93]. The space X of the individual's ability is associated with the dual space X^* of the constraint. The effect of the hierarchy y^* on arbitrary responses of intelligence x is completely known, if we know how it affects some basic options. The base could be any complete set of independent situations.

Comparing this with the previous (eponymous) appendix, the contravariant coordinates of the projection of a point on the axes are obtained by its parallel displacement (translation) in relation to the opposite axis, and the covariant coordinates form the vertical projections of the point on the axes. Both the first and the second, these coordinates change by changing the angle φ between the axes.

Note that by decreasing the angle φ between the axes and maintaining the same contravariant coordinates, the data point moves away from the origin, and by maintaining the covariant coordinates the data point would approach the origin. These two movements are such that their *scalar* product (hereinafter referred to as the canonical product) would remain unchanged, and that is the topic of the sequel.

Canonical product

Other names for multiplying coordinates with functionals $y^*(x)$ are scalar, or internal product of the vector, here $(\xi_1, \xi_2, \dots, \xi_n)$ and $(\eta_1, \eta_2, \dots, \eta_n)$, and especially the *canonical product* of the vector x from the space X with the vector y^* from the dual space X^*. We mark it

$$y^*(x) = \langle x|y^*\rangle \tag{6}$$

in mathematics and similarly (different order) in physics. We write it in a matrix form

$$(\eta_1 \quad \cdots \quad \eta_n)\begin{pmatrix}\xi_1 \\ \cdots \\ \xi_n\end{pmatrix} = \sum_{k=1}^{n} \xi_k \eta_k$$

where we multiply the matrix-row (type 1 × n) by the matrix-column (type n × 1) and we get a scalar (matrix of type 1 × 1).

Example 1. Consider what is said in an example similar to rotation (4).

[93] [9]

Let $A : X' \to X$ be a linear mapping of coordinates $\xi_k = \sum_{j=1}^{n} a_{kj}\xi_k{}'$, or shorter $x = A(x')$, or $x = Ax'$, which maps a vector space of dimension n to some again n-dimensional one. Then the operator A is regular, invertible, there is an inverse mapping $A^{-1} : X \to X'$.

The same operator maps the base in reverse ($A : e \to e'$). Namely, $x = \sum_{k=1}^{n} \xi_k e_k = \sum_{k=1}^{n} \xi_k{}' e_k{}'$, which means that the operator A that maps the coordinates cancels the operator A^{-1} which maps the bases. We know that $AA^{-1} = A^{-1}A = I$, where I is the unit operator ($Ix = x$, for all x), and the inverse operator A^{-1} is unique, is uniquely determined given by the regular operator A.

Finally, from (5) and (6) we see that the functional y^* maps the bases and is mapped inversely of coordinates. In other words, the canonical product (6) is invariant to linear mappings of space, which includes rotation (4). That is the meaning and relations (19) of the last article. □

Example 2. The canonical product (6) is a linear functional of the first and second arguments.

Namely, $\langle \alpha x + \beta y | z^* \rangle = z^*(\alpha x + \beta y) =$ (because z^* is a linear functional)

$$= \alpha z^*(x) + \beta z^*(y) = \alpha\langle x|z^*\rangle + \beta\langle y|z^*\rangle,$$

which means that (6) is a linear function of the first argument.

Next, $\langle z|\alpha x^* + \beta y^*\rangle = (\alpha x^* + \beta y^*)(z) = (\alpha x^*)(z) + (\beta y^*)(z) =$

$$= \alpha x^*(z) + \beta y^*(z) = \alpha\langle z|x^*\rangle + \beta\langle z|y^*\rangle,$$

which means that (6) is a linear functional of the second argument. □

For two vectors $x \in X$ and $y^* \in X^*$ we say that they are perpendicular to each other, which is written $x \perp y^*$, when $\langle x|y^*\rangle = 0$. The set $S_1 \subseteq X$ is perpendicular to the set $S_2 \subseteq X^*$, written $S_1 \perp S_2$, is if each element of the first set is perpendicular to each element of the second set.

Example 3. If $x^* \in X^*$ is perpendicular to the space X, then $x^* = 0$. Also, if $x \in X$ is perpendicular to the space X^*, then $x = 0$.

Indeed, a zero-functional ($x^* \in X^*$) is a functional that translates each vector into zero, i.e. $x^*(y) = 0$ for every $y \in X$, which is written using the canonical product $\langle y|x^*\rangle = 0$ for all $y \in X$. Hence the first statement is true.

Let $x \perp X^*$, ie. $\langle x|y^*\rangle = 0$ for every y^* from X^*. It will be in the base (e)

$$y^*(x) = \sum_{k=1}^{n} \xi_k \eta_k = 0$$

where $\eta_k = y^*(e_k)$, and ξ_k are the coordinates of that fixed vector in the solid base. From the arbitrariness of the vector y^* follows the arbitrariness of the numbers η_k and the above sum disappears for all η_k only if all ξ_k are zeros, i.e. if $x = 0$, which is the second statement. □

These examples are also mentioned in algebra textbooks as theorems, and especially the third one. The third is used for a broader definition of duality, here's how.

For two vector spaces X and Y over the field of scalars (numbers) Φ we say that they are dual in a broader sense, if there is a *bilinear* (linear on both arguments) functional $B(x,y)$ with the following two properties:

1. if $B(x,y) = 0$ for all $x \in X$ then $y = 0$;
2. if $B(x,y) = 0$ for all $y \in Y$ then $x = 0$.

Consistent with Example 3, we see that $B(x,y^*) = \langle x|y^* \rangle$ is a bilinear functional ($x \in X$, $y^* \in X^*$) and that the spaces X и X^* are dual in the above the broader sense.

However, the corresponding "canonical product" of duality in a broader sense does not have to be invariant. For example, the scalar product of covariant with covariant coordinates (as well as contravariant with contravariant) will be bilinear functional, and we know that such changes by mapping space.

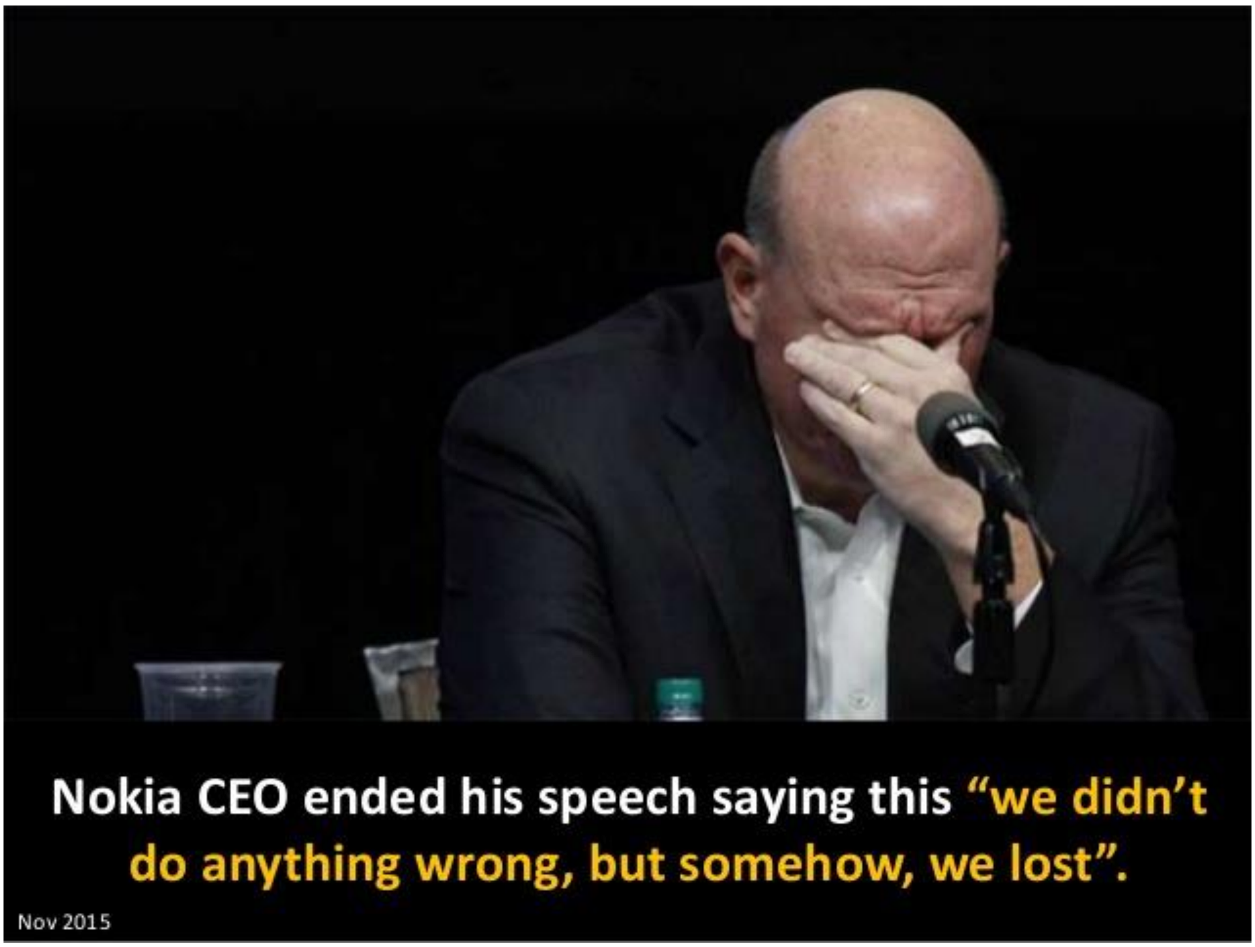
17. Quadratic Form

April 3, 2021

Introduction

Question: Can you tell me why Nokia failed?

Answer: Ha, ha, ha ... you're seriously asking me, did it fail, who do you say Nokia? ... (I searched and found something). Ok, I don't know much about that company, but I can give you the general principles of (my) information theory, so see if it fits.



Perception information $S = ax + by + cz + ...$ measures information, or the amount of options of a given system, in a way that can represent the vitality of the company. Let's say we only have three summands. The first factors are some participations in shares, for example a = 20% of old owners, those who contributed to the growth of the company, or owners whose ideas are similar to those that could have raised the company. I note that we often do it unaware of what it is that can "lift up" a company, but after a while we can still somehow evaluate that work. Let the value of the work of such x = 10 points - whatever the "points", but the more means a higher value.

Let b = 35% be some small shareholders who usually grab the shares of a successful company wanting higher profits, looking for success at all costs, order, work and discipline, in short, "efficiency". From the point of view of (my) information theory, it has less value than intelligence and its lucidity. Risk avoidance reduces information, its vitality and aggressiveness. Let us give y = 7 points to such.

Let c = 45% of competing companies that bought shares wanting to slow down a given company rather than help it. Let's give them c = 3 points.

The total information of perception is now S = 20 × 10 + 35 × 7 + 45 × 3 = 580. However, when Nokia was of successful composition, or mode of operation, the structure of its shares was different, say: 20% were competing companies, 35% small shareholders, 45% the composition of the old forge, so the then information of perception $S_0 = 45 \times 10 + 35 \times 7 + 20 \times 3 = 755$. It was more vital than in the end.

***

Question: ... these examples with "perception information" of you are great. Do you ever tell students about it?

Answer: Never, I don't think it happens, maybe involuntarily (Thank you! - is implied in communication). I give them books for free, whoever wants, and that's all, mostly. I do not retell my "theories" to connoisseurs or beginners.

Q: And would any of this be possible somewhere in the school curriculum?

A: Of course yes. Here, for example, in the final grade of high school, high school graduates who are learning probability these days can find one such task.

Assignment. Three factories produce the same item. The first produces twice as much as the second, and the second and third produce the same number, over a period of time. However, 2% of the products of the first and second factories are defective, and 4% of the products are defective in the third. All manufactured items arrive at the same warehouse, from where we randomly pick up one. Find the probability that it is defective.

The solution. Of the total quantity of products in stock from the first, second and third factories, they are: a = 1/2, b = 1/4 and c = 1/4. Their damage share is x = 2%, y = 2% and z = 4%. Perception information, then a measure of possible damage, is $S = ax + by + cz$ = 2% / 2 + 2% / 4 + 4% / 4 = 1% + 0.5% + 1% = 2.5%. That is a coefficient of 2.5 / 100 = 0.025. □

There is a 0.025 probability that a randomly selected item from the warehouse is defective. This solution is "IT" and is not done at school. Of course, the same result was obtained in other (correct) ways.

Bilinear functional

A bilinear mapping is a function that combines elements of two vector spaces to give an element of a third and is linear in each of its arguments. Examples are matrix multiplication, bilinear forms and perception information.

The function of two variables $f(x,y)$ is bilinear if:

$$\begin{cases} f(\alpha_1 x_1 + \alpha_2 x_2, y) = \alpha_1 f(x_1, y) + \alpha_2 f(x_2, y) \\ f(x, \beta_1 y_1 + \beta_2 y_2) = \beta_1 f(x, y_1) + \beta_2 f(x, y_2) \end{cases} \tag{1}$$

for each $x \in X$ and $y \in Y$. Domains X and Y can be different vector spaces, but when the coefficients of the vector, and with them the codomains of the function are scalars (numbers) from the same body Φ, then the function is called a *functional*.

The computational operation between the variables of the bilinear functional in the first and second place, in case (1) between x and y, does not have to be defined, but when it comes to "perception information" it is a scalar product of the vector. Even that is "too general", as some colleagues tell me, because they are "confused" by my "too free" use of it. I'll explain again.

From the very principles that more probable events are less informative, and then that more probable ones are realized more often, and that physical states tend to be less informative, it follows that we can treat probability and information equally. They are like positive and negative numbers for which we have a common name "numbers", or inverse functions such as exponential and logarithmic for which we do not (yet) have a common name.

When the probability p that D will happen is a small number (event D will almost certainly not happen), then $p = -\log(1-p)$, so the probability itself is then some information[94]. Examples of these connections can be "unexpected" in various ways. The wave function $\psi = A \exp i(kx - \omega t)/\hbar$, the solution of the Schrödinger equation for a free particle, represents the "probability wave" and its argument in the exponent as the logarithm of some probability can be treated as slightly generalized Hartley information (logarithm of the number equally likely options).

Therefore, care should be taken when interpreting the form of information perception, but this is not new to algebra. Thus, in the following definition, it is assumed that X is a vector space of dimension $n \in \mathbb{N}$ and much more in connection with notations and rules, and only then the next it said.

The mapping $f : X^2 \to \mathbb{C}$ is defined by the equation

$$f(x,y) = (Ax, y) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_j y_i^*, \tag{2}$$

where $A = (a_{ij})_n$ is a square matrix of type $n \times n$, and $y^* \in \mathbb{C}$ is a conjugate number of the number $y \in \mathbb{C}$, it is called a *bilinear form* or bilinear functional[95]. In particular, when $y = x$, the mapping $f : X \to \mathbb{C}$ is defined by

$$f(x) = (Ax, x) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_j x_i^*, \tag{3}$$

it is called a *quadratic form* or a quadratic functional.

---

[94] [1], 14. Uncertainty, p. 50.

[95] Functional is a function from a vector space into the scalar field.

The quadratic matrix $A = (a_{ij})_n$ is Hermitian[96] if it is equal to its conjugate transposed, i.e. if $a_{ij} = a_{ji}^*$ for all indices $i, j = 1,2, \dots, n$, which we write $A^\dagger = A$, or shorter $A^* = A$ denoting conjugation and implying transposition. For example, a matrix

$$A = \begin{pmatrix} 1 & 2-i & 3+2i \\ 2+i & 2 & 1-2i \\ 3-2i & 1+2i & 3 \end{pmatrix}$$

is Hermitian. The diagonal elements of the Hermitian matrix are always real numbers. Unless otherwise stated, the square-shaped matrix (3) is assumed to be Hermitian.

The values of the square form (Hermitian matrix) are real. Namely, conjugation (3) follows:

$$f^*(x) = \left(\sum_{ij=1}^{n} a_{ij} x_j x_i^*\right)^* = \sum_{ij=1}^{n} \left(a_{ij} x_j x_i^*\right)^* = \sum_{ij=1}^{n} a_{ij}^* x_j^* x_i = \sum_{ij=1}^{n} a_{ji} x_i x_j^* = f(x)$$

because $a_{ij} = a_{ji}^*$, and only real numbers remain the same by conjugation.

A Hermitian matrix A is called *positive definite* if for every $x \neq 0$ $(Ax, x) > 0$ holds, and then $a_{ii} > 0$ for all $i = 1,2, \dots, n$. A symmetric positive definite matrix is called a *normal* matrix. The application of these matrices in quantum mechanics is enormous.

Perception information

In information theory, we come to the functional in a few steps. We first note that “information” is a “quantity of options”, and we must take into account that not all solitary options have the same “information weight”. Less likely events are more informative. Not all events have the same number of options (regardless of whether they have the same or different “weight”) and not all options need to have the same probability, and then they can lean towards special realizations (the more likely are more frequent).

Thus, we come to the second observation, that the “information of the system” is proportional to some kind of “intelligence” of the system, by random plus more frequent outcomes. When trying to separate equally and unequally probable outcomes, in order to multiply their quantities, we will notice that there is an even better determination of the information system using personal “abilities” and external “limitations”. The first are, say, the values of subjective properties, the second are objective.

Finally, based on the observation that in the “world of information” not everyone communicates with everyone, we will understand the mentioned measure of the system as “information of perception”. Some systems, living or non-living beings, have different perceptions of the same environment. A neutron does not react to an electric force, and bats and birds do not use the same senses in hunting. This is how we come to the separation even after special, independent random events.

[96] Hermite, Charles (1822-1901), French mathematician.

Extending the notion of "intelligence", $I(\omega)$, to the living and non-living world, from the point of view of a particular experiment $\omega \in \Omega$ from a set of events Ω, we consider it proportional to "perception information" $S(\omega)$ and external environments, let's call them "hierarchy" $H(\omega)$. Thus, from $I = S/H$, for a given event ω, $S = IH$ follows, so for n-tuple of separate events $\omega_1, \dots, \omega_n \in \Omega$ we get $S = S_1 + \dots + S_n$, or

$$S = I_1H_1 + I_2H_2 + \dots + I_nH_n. \tag{4}$$

We recognize this expression as a bilinear functional (1), and in case of need in quantum physics, and especially as a quadratic form (4).

Let us note the first "unexpected" result of the "information of perception" defined in this way, that the reduced information of the perception of a given subject means his reduced ability to get out of difficulties, or reduced perception of difficulties, or both. A stupid creature, we would say, neither sees the problems around him nor knows how to solve them.

18. Variant Vectors III

April 7, 2021

I continue to interpret functional, here dual (biorthogonal) bases of vector spaces and the application of that algebra to quantum states, all from the point of view of information perception.

Introduction

The basic meanings of “vitality” are will, enthusiasm, liveliness, and “information of perception” is its measure. The point is that according to the assumption of the “information universe”, the matter, space and time consist of information, and it is a measure of the options and uncertainty, such that it can represent various concepts. That only, of course, if the very formula of that perception proves to be correct.

First of all, Shannon's definition of information (mean value) is its special case. In short, if a (complete) sequence of probabilities $a, b, c, \dots$ that we will obtain information respectively valued with $x, y, z, \dots$ then we will obtain information $S = ax + by + cz + \cdots$, which is actually the type of the “perception information“. The application is similar.

For example, if 100 cars are exhibited on a plot, 100 from one and 40 from another factory, and among those from the first with 5% damaged and the second with 3%, then the chance that a randomly taken car from the plot will be damaged is $0.60 \times 5\% + 0.40 \times 3\% = 4.2\%$, i.e. probability 0.042. It is a model that is easy to generalize and, admittedly, a little harder to interpret.

When in bigger group more cars are damaged in an even more significant percentage, it is more likely that a randomly selected specimen from the plot will be damaged. Conversely, if in the bigger group more cars have the percentage lower, then more correct ones will be represented in the total mass. We understand the same intuitively; it is part of the information of perception, but not Shannon's.

The same form of computation is transferable to seemingly completely different situations. Let's say into the “difficulties” of a player, society, economy, state with two basic threats, size of x and y. Let the subject who is dealing with threats distribute his forces to them in relation to $a\!:b$. When it is 100 cars, the ratio is a : b = 60 : 40. The amount of subjects participation in “work on difficulties” is precisely the information of perception $S = ax + by$. If he threw more at a smaller threat, the number S is smaller and we say that the ratio of countering threats is more sluggish.

Many laws of these relations, dualisms, can be found in (multi) linear algebra, in quadratic and bilinear forms in general. This is the topic of the articles on variant vectors. It is surprising how many from ordinary life is among the well-known theorems, but I am not saying that they are easy to interpret.

The laws of conservation information perception (probability and uncertainty too) make it final, for example, because only infinity can be its rightful part. This further means divisibility to the smallest portions of “pure uncertainty”, after which the continuation of the removal of uncertainty would lead to greater certainty. We will focus on the existence of elements, the smallest portions of information perception, the quantum of action, that is, the final base of vector spaces.

Dual base

Unlike the previous sequel[97] of the same name, where we had one functional in the entire vector space, here we will observe them at once as much as the base of the space. If $e_1, \dots, e_n$ is the base of the space X, then each vector $x \in X$ can be uniquely written in the form

$$x = \xi_1 e_1 + \dots + \xi_n e_n. \tag{1}$$

To find the coordinates ξ_k, respectively for the indices $k = 1, \dots, n$, we look for linear functional[98]

$$e_k^*(e_j) = \delta_{kj} = \begin{cases} 1 & k = j, \\ 0 & k \neq j. \end{cases} \tag{2}$$

The symbol δ_{kj} is called Kronecker[99]. For functionalities (2) the following applies:

$$e_k^*\left(\sum_{j=1}^n \xi_j e_j\right) = \sum_{j=1}^n \xi_j e_k^*(e_j) = \sum_{j=1}^n \xi_j \delta_{kj} = \xi_k$$

therefore

$$\xi_k = e_k^*(x). \tag{3}$$

That the functional e_k^*, if any, is linear when joining the k-th coordinate to the vector (1) in the base e, follows from $(\lambda x)_k = \lambda (x)_k$ and $(x + y)_k = (x)_k + (y)_k$, where the k-th coordinate of the vector is $(x)_k$. The following theorem is important for the proof of the existence of the functional e_k^*.

Theorem 1. Let $e_1, \dots, e_n$ be a base in the space X. If $A, B : X \to Y$ are two linear operators and $Ae_k = Be_k$ for $k = 1, \dots, n$ then $A = B$. For arbitrary vectors $f_1, \dots, f_n \in Y$ there is one and only one linear operator $A : X \to Y$ such that $f_k = Ae_k$ for all $k = 1, \dots, n$.

Proof. The first statement follows from:

$$Ax = A\sum_{k=1}^n \xi_k e_k = \sum_{k=1}^n \xi_k A e_k = \sum_{k=1}^n \xi_k B e_k = B\sum_{k=1}^n \xi_k e_k = Bx$$

for each $x \in X$, i.e. $A = B$.

To prove the second statement, we define the operator $Ax = A\sum_{k=1}^n \xi_k e_k = \sum_{k=1}^n \xi_k f_k$. Such a vector:

[97] 16. Variant Vectors II
[98] The functional is mapping of vectors into scalars.
[99] Kronecker delta, https://en.wikipedia.org/wiki/Kronecker_delta

$$ax + by = a\sum_{k=1}^{n} \xi_k e_k + b\sum_{k=1}^{n} \eta_k e_k = \sum_{k=1}^{n} \left(a\xi_k + b\eta_k\right) e_k$$

maps to the vector:

$$A(ax + by) = \sum_{k=1}^{n} \left(a\xi_k + b\eta_k\right) f_k = a\sum_{k=1}^{n} \xi_k f_k + b\sum_{k=1}^{n} \eta_k f_k = aA(x) + bA(y)$$

and since it is a linear operator and $Ae_k = f_k$, the existence of operator A is proved. Unambiguity follows from the first statement of this theorem. □

Thus, the existence of linear operators from X to Y is proved. From the point of view of the representation of these spaces in quantum mechanics, the consistency of the behavior of quantum states (vectors) in a quantum system (vector space) based on their elementary properties (base), then the existence of the processes (operators) that map one quantum state (X) into another (Y) – is proved.

The above theorem is known and its mentioned interpretation is almost known. The following positions are also known, the second, then the third theorem, which I state without proof. Any system of linearly independent vectors of a finite dimensional space can be supplemented to the base of that space, i.e. each such vector system is a subset of at least one space base. This is the "second theorem".

The next is the "third theorem" which says that there is one and only one matrix $\mathbf{A}$ of type $m \times n$ (with m types and n columns) associated with the linear operator $A : X^m \to Y^n$, i.e. mapping m into n dimensional vector space. These spaces, X and Y, have m and n base vectors, respectively, and the stated association, between the operator A and its matrix $\mathbf{A}$, is an *isomorphism*[100] of vector spaces.

Evidence is known that $m \times n$ is a dimension of both the operator A and the isomorphic matrix $\mathbf{A}$, so I do not list them. Also, in the case of the mapping composition $A: X^m \to Y^n$, $B: Y^n \to Z^p$ and $C: X^m \to Z^p$, $\mathbf{C} = \mathbf{BA}$ is valid for their isomorphic matrices. Then, that the laws of distribution and association apply to the multiplication of matrices, but not commutation.

This long-established connection between matrices and operators is also considered to be a representation of linear operators using "concrete" elements – matrices. As we know, at the beginning of the 20th century, it expanded towards the representation of linear operators, together with their matrix representatives, to quantum states. Because of the connection between information and action, we also continue to expand on the information itself.

The existence of the functional e_k^* follows from the first theorem, substituting $Y = \Phi$ and $f_1 = \dots = f_n = 1$ in Φ. In addition to the already proven property (2), it is shown that the vectors $e_1^*, \dots, e_n^*$ form an independent and complete set, so that they represent the base of a space of dimension n.

Namely, from the assumption $\alpha_1 e_1^* + \dots + \alpha_n e_n^* = 0$ for each $x \in X$ there will be $\alpha_1 e_1^*(x) + \dots + \alpha_n e_n^*(x) = 0$, so substituting $x = e_k$ in order for $k = 1, \dots, n$ we get $\alpha_k = 0$, which shows the

[100] Isomorphism is a mutually unique mapping (bijection) between two mathematical structures.

independence of the vector e_k^*. That they span an n-dimensional space, that is, that they can define any vector $x^* \in X^*$ of such a space, follows from

$$\left[x^* - \sum_{j=1}^{n} x^*(e_j)e_j^*\right](e_k) = x^*(e_k) - \sum_{j=1}^{n} x^*(e_j)e_j^*\,(e_k) = 0$$

for $k = 1, \dots, n$ so it is the functional in angular bracket of a zero-functional. Thus

$$x^* = \sum_{j=1}^{n} x^*(e_j)e_j^* \tag{4}$$

is an arbitrary vector $x^* \in X^*$ and is a linear compound of the vectors $e_1^*, \dots, e_n^*$. Hence, and on the basis of the mentioned second and third theorems, it follows that X^* is an n-dimensional vector space.

The base e^* from X^* is biorthogonal, or the dual base to e from X, if

$$\langle e_k | e_j^* \rangle = \delta_{kj}, \quad k, j = 1,2, \dots, n. \tag{5}$$

Here again, the Kronecker symbol (2) is used. The existence of the biorthogonal base follows from example 3 of the last appendix[101] of the same name, and the benefit from it comes from the ease of obtaining the coordinates of the vector x in the base e. For an arbitrary vector $x \in X$ will be

$$x = \sum_{k=1}^{n} \langle x | e_k^* \rangle e_k \tag{6}$$

because $x = \sum_{k=1}^{n} \xi_k e_k$. By applying the functional e_k^* we get $\xi_k = e_k^*(x)$, and this is by definition exactly what gives (5).

Quantum state

Quantum state is the distribution of probabilities of all possible measurement outcomes of a given quantum system. It is represented by vectors like (6), with slight adjustments. The first is, conversely to the algebra, that covariant vectors are denoted by conjugation x^*, and contravariant[102] vectors are unconjugated, x. Here we use both methods and we will emphasize which is in progress, if necessary. The second is that the vectors (quantum states) be normalized to the unit, $\|x\| = 1$.

In addition to (my) information theory, we will consider the scalar product of vectors, quantum states, as the probability of their association. How is the square of the norm of the vector $\|x\|^2 = \langle x | x^* \rangle =$

$$= \left(\sum_{k=1}^{n} \langle x | e_k^* \rangle e_k\right) \cdot \left(\sum_{j=1}^{n} \langle x | e_j^* \rangle e_j\right)^* = \sum_{k,j=1}^{n} \langle x | e_k^* \rangle \langle x | e_j^* \rangle^* \delta_{kj} = \sum_{k=1}^{n} |\langle x | e_k^* \rangle|^2$$

[101] 16. Variant Vectors II
[102] 15. Variant Vectors

well

$$\|x\|^2 = \sum_{k=1}^{n} |\xi_k|^2 \tag{7}$$

and on the other hand, $\|x\| = 1$, then the components ξ_k of the vector $x = (\xi_1, \xi_2, \dots, \xi_n)$ define the probability of its finding by their square of the module $|\xi_k|^2 = \xi_k \xi_k^*$ in state e_k. We interpret the projection of vectors on vectors by measurement, measurement by interaction, and interaction by communication. This is consistent with the usual choosing the observables (measurable quantities) for base vectors.

In Dirac's[103] bra-ket notation, covariant (bra) and contravariant (ket) vectors are written, for example:

$$\langle\psi| = (a \quad b \quad c), \quad |\phi\rangle = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \tag{8}$$

so that their (matrix) product

$$\langle\psi|\phi\rangle = ax + by + cz. \tag{9}$$

If it is a scalar product of two vectors of quantum states, the result will be a real number. Instead of ψ and ϕ, quantum physics uses and other notations to determine the state more closely. When $\psi = \phi$, then (9) represents the square of the norm. In order for these vectors to be physical, they must be normalized to unity, so for arbitrary $a, b, \dots, z \in \mathbb{C}$ we have:

$$\langle\psi| = \frac{(a,b,c)}{\sqrt{|a|^2+|b|^2+|c|^2}}, \quad |\phi\rangle = \frac{1}{\sqrt{|x|^2+|y|^2+|z|^2}} \begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

$$\langle\psi|\psi^*\rangle = \frac{(a,b,c)}{\sqrt{|a|^2+|b|^2+|c|^2}} \frac{1}{\sqrt{|a|^2+|b|^2+|c|^2}} \begin{pmatrix} a^* \\ b^* \\ c^* \end{pmatrix} = \frac{aa^*+bb^*+cc^*}{\left(\sqrt{|a|^2+|b|^2+|c|^2}\right)^2} = 1,$$

$$\langle\phi^*|\phi\rangle = \frac{(x^* \quad y^* \quad z^*)}{\sqrt{|x|^2+|y|^2+|z|^2}} \frac{1}{\sqrt{|x|^2+|y|^2+|z|^2}} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \frac{x^*x+y^*y+z^*z}{\left(\sqrt{|x|^2+|y|^2+|z|^2}\right)^2} = 1,$$

$$\langle\psi|\phi^*\rangle = \frac{ax^* + by^* + cz^*}{\sqrt{|a|^2+|b|^2+|c|^2} \cdot \sqrt{|x|^2+|y|^2+|z|^2}} \tag{10}$$

It is understood that the number n components $\langle x|e_k^*\rangle$ of the vector (6) can be different, that the components themselves, the coefficients of the vector can be complex numbers, but the scalar products (10) are required to be real. These products represent interactions, and only those are observable, physically real, whose products are real numbers.

[103] Paul Dirac (1902-1984), English theoretical physicist.

The quantum state before the interaction is called superposition. It represents the uncertainty of the same amount of information as any of its eventual outcomes. For example, the state of a fair coin before tossing and any of the two outcomes (tails, heads) has the information $\log_2 2 = 1$ bit. The state of the dice before the roll and each of the six outcomes have the same information of $\log_2 6 \approx 2.58$ bits.

Superposition includes all the possibilities that lead to known measurements, but unlike throwing coins or dice, in the micro world there are (sometimes unknown) options of real uncertainty that precede the (known) measurement. There we discover assumptions (superposition) by means of outcomes (measurements), and that can confuse us. I will illustrate this with a familiar[104] example.

Denote the base vectors, observable, by $|0\rangle$ and $|1\rangle$, and the possible particle paths by the vectors:

$$|\alpha\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle, \quad |\beta\rangle = \frac{1}{\sqrt{2}}|0\rangle - \frac{1}{\sqrt{2}}|1\rangle.$$

The quantum mechanical state $|\psi\rangle$ is a superposition of the trajectories α and β, and we define it as their (normalized) sum:

$$|\psi\rangle = \frac{1}{\sqrt{2}}|\alpha\rangle + \frac{1}{\sqrt{2}}|\beta\rangle = \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle\right) + \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}|0\rangle - \frac{1}{\sqrt{2}}|1\rangle\right) = |0\rangle.$$

However, the quantum state $|\psi\rangle$ instead of the sum, could be the difference of the trajectories α and β, so we would measure $|1\rangle$. Both paths as well as both outcomes have the same probabilities $\frac{1}{2}$, which can be seen from the amplitudes $\frac{1}{\sqrt{2}}$, and in the superposition we add the amplitudes. It is interference.

Particles of quantum mechanics interfere like waves, and probabilities define measurements, so quantum states are particles-waves of probability. Conservation laws limit these interactions and this also sometimes confuses us. Like losing the clarity of individual images in a video when we want to show motion in more detail, or increasing the vagueness of a particle's position when we want to more accurately determine its momentum (uncertainty relations), declaring one property of a particle will lead to the loss of another.

For example, in the known double-slit experiment[105], when one of the two holes is closed, the motion of the particle-wave is declared and its behavior is different, it is more particle than when both are open and its behavior is wave. When we pass only one particle-wave through these two slits in long periods in between, interferences of its possible paths occur and the "particle" behaves more like a wave. So much for superposition.

[104] https://www.quantiki.org/wiki/states
[105] https://en.wikipedia.org/wiki/Double-slit_experiment

Epilogue

The topic has only just begun, but in anticipation of the continuation, I hope, some basics have been clarified. Next, we need to determine that the dual of the dual vector is again the starting vector and explain the differences between pure and mixed states, first of all, and then something else.

19. Natural law

April 10, 2021

This is a discussion of “immutable, objective and eternal rules of human behavior which in this sense are considered similar to natural laws”, from the standpoint of information theory.

Question: Can modern law achieve the logical consistency of geometry?

Answer: No, the assumption of equality does not allow it. Even in theory, there are not enough “equal” persons or situations to be considered such an assumption correct. The problem is further with deduction, which gives unreliable conclusions from an incorrect assumption; the derived consequences are truthfully irrelevant.

Q: Then how can two electrons be equal?

A: The elementary particles of physics cannot be absolutely equal either. The cosmos is constantly changing, galaxies are moving away, matter is rearranging, and those changes are partly unpredictable. As it travels through space, the photon constantly enters at least a slightly different environment than everything before. Always in a new state with its environment, it is different at any two moments.

Some particle without an environment does not exist, depending on the environment it makes a phenomenon. This is a consequence of the theory of information perception ($S = I \cdot H = ax + by + \dots + cz$) which defines “information” only when we have a particle, vector $I(a, b, \dots, c)$, with its environment, vector $H(x, y, \dots, z)$. Consistent with the theory of relativity but also with quantum mechanics, where what is observed depends on the observer, in the information theory I advocate, there is no “particle information” independent of the rest of the universe. On the other hand, information is the basis of space, time and matter.

Q: Is it then possible to slightly change the assumptions of legal science to obtain a logically perfect system of social laws?

A: No. I know the answer is confusing and we are reluctant to believe it, but the truth does not matter what we believe. It is not possible to consistently use any “legal science” (jurisprudence) so that its deductions always and always derive exact consequences, consistently like geometry.

Namely, we already know a part of reality through the laws of physics. Some of its laws we will discover, and we will never find out all, either because we will not be able to do so subjectively or because they are objectively unknowable (due to Gödel's theorems on incompleteness). However, what does not exist in physical reality at all, or is provably impossible – cannot exist[106].

[106] The world is evolving towards the more probable, less informative, with fewer unexpected events.

Looking at it that way, information is always some truth, open or concealed. Even explicit truths like theorems cannot be clear to everyone, and especially not messages from lies, or silence itself. Some of this is demonstrated by the following, otherwise well-known example.

Three boxes are shown with exactly one containing the treasure (prize). You can keep the prize if you choose the right box. There is a statement on each box, but only one is true.

If the first is true, then the second is also true, so the first statement is not true and the treasure is not in the first box. If the treasure is in the second box, the second statement is not true, and one and only one of the statements of the first and third is true. Therefore, the treasure may be in the second box. If the treasure is in the third box, the third statement is true, but the second is also true, so the third statement is not true, which means that "the prize is in box 1", which contradicts the claim that the treasure is in the third box and that it cannot be in some other. So the treasure is in the second box!

That the "world of untruth" is isomorphic (equivalent) to the "world of truth" I also pointed out earlier. There is a mutual unambiguous mapping (bijection) of the truth tables of the algebra of logic with those that we would get from these by replacing "true" ($\top$) with "false" ($\perp$) and vice versa. The world of pure untruths (contradictions) with this double substitution of true-false would become the world of pure truths (tautology), which in its own way was understood earlier in mathematics, in its method of proving by "reduction to contradiction".

Q: A lie is a hidden truth?

A: That's right. In principle, probability theory holds that more probable events are more frequent, and hence more frequent are less informative. This is the "principle of least information", or if you want the "principle of least action", because we can assume that action and information are equivalent. Nature does not like the show of information, it prefers to hide it, wherever it can, and some of those ways are lies and misinformation.

Q: What can stop us from at least trying to put a system in a state of equality, what can happen then?

A: The force changes the probabilities. The trajectories along which physical particles move, from their point of view, are most likely their trajectories and it remain so until their perceptions change. The glass is on the table because at that moment it is it's the most probable position. Due to the law of information conservation, so remains the same – until another body (hand) or force moves the glass.

Force occurs by changing the uncertainty and relationship of surplus (lack) of environmental information.

The answer to this question, then, we find by noting that the state of equality carries the maximum information. There is more uncertainty in throwing a fair coin than a phallic one. When we know that a coin is deformed so that the heads is more likely to fall than the tails and the heads falls, it is an expected, less informative event. The principle of minimalism of information will oppose equality.

In (my) information theory, equality is in itself a force that would change relations into unequal ones, which we intuitively feel and harness when we put competitors in equal positions in sport to make the game as lively as possible, or exploit it by simulating democracy to achieve greater development.

Q: So you say that equality generates conflicts?

A: Yes. We can always recognize tendencies as attractive or repulsive. Let's say that this is the first case, let's work with probabilities and with only two possibilities ($S = ax + by$). If for each of the two participants (environment, society, individual) the first option is more attractive so that there are more of them in the offer ($a > b$ and as $x > y$), their coupling (S) will show a greater[107] amount, a higher probability than that the first option is more attractive to the first participant and more repulsive to the second ($a > b$ and as $x < y$). It is more likely that the “same” will unite and increase the amount of the first possibilities.

This is how political fractions and hatred towards those who are “not ours” arise. Contrary to “ours”, there is antagonism towards “theirs”, so working in the first mentioned case (with probabilities), as an echo we find repulsive forces and work with, say, Hartley's information, the logarithms of probabilities.

Q: We come to a similar point with the free network model?

A: Yes, the correct theories are consistent. In the “free networks” model, we it is the equal connections that start (end) with nodes. For simplicity, imagine that we have only two nodes, the first with ten links and the second with only one, and that we need to add a new link (and node) to them. As all 11 are equally probable sequels, and the first node has 10, the chance that the connection to the first node will occur is 10 times higher than to the second. The chance is so much higher that the first node will then have 11 links and the second will still have only one. More likely events are more common, so nodes with more links increase more often!

A free network is evolving into one with very few nodes rich with links versus a large number of poor ones. Wherever we have the principle of equality, that kind of inequality grows. For example, in the free money market, the few very rich stand out. Along with the “principled equality” of travel, opinions, the right to freedom, a class of those who can travel more, whose opinion is more valued, who are freer, develops. They do not have to be of the same class, but again among equal of them – the same watershed continues to develop.

---

[107] If $\varepsilon\delta > 0$ then $(a + \varepsilon)(x + \delta) + ax > (a + \varepsilon)x + a(x + \delta)$, which is easy to check.

This method is equivalent to the above, using “information perception”, interpreted using “attractive” and “repulsive” tendencies. In the case of the first we work with probabilities (more probabilities – more inclinations), in the case of the second with information (more information – less inclinations). Then it helps to imagine inclinations in general, for example as “loving squares”, and reluctance to “loving circles”. An environment that “loving squares” grows, attracting others with the same inclinations. Whoever stays in the middle of the road is run over, proverb says. That is how every substance is determined, every form of life, and that is how it will be with automata. It will not be possible to create an artificial intelligence that could ignore this regularity!

Q: What if we maintain equality by force?

A: That is what we are trying to do with the legal system, and we are getting more and more lawyers, laws, costs. The need for intervention grows (not exponentially as some write hastily, but with a slower power-law of scale-free networks) and at some stage the mightiness of the “principle of least information” overcomes. In the end, our force of law is defeated by the “pressure of equality”.

Q: The greater that “pressure” it manages to hold, the more vital society is?

A: Yes, that's how it should be, I've already written. That is the power and evil fate of democracy. It is necessary to keep the information of the system as large as possible, ie with more freedoms, or at least more ones of the “more important” ones, as long as possible, in order for the development to continue. Admittedly, that is how the explosiveness of the situation will grow and equality will be prevailed in the end by some oligarchy, but in a sense, the success was greater with a longer prevention of the disintegration of the initial state.

20. Reflexive space

April 17, 2021.

The isomorphism of information perception of systems (beings) with an equal number of senses, using a form known from linear algebra, is explained.

Introduction

Perception information, $S = A \cdot B$, is the backbone of the information theory I am researching. It is a coupling of options and limitations, the possibilities of the individual and the proscription of the environment; we can say "intelligence" and "hierarchy". When these properties, actions A and reactions B, are manifested on a series of phenomena $(\omega_1, \omega_2, \dots, \omega_n)$, on the k-th element ($k = 1,2,\dots,n$) with special values a_k and b_k, then the factors of perception information are sequences $A(a_1, a_2, \dots, a_n)$ and $B(b_1, b_2, \dots, b_n)$, and it is the canonical product

$$S = a_1 b_1 + a_2 b_2 + \dots + a_n b_n. \tag{1}$$

The perception of what is perceived is in conjunction with what is perceived, that is, who perceives.

The structure (1) is formed by two factors, the components of the subject and the components of the object. These are the ability to solve problems and the ability to perceive problems, or possibilities and impossibilities. These factors are like Newton's action and reaction, but also the waves (light, particles, water) whose amplitude recede and pull back as an oscillation of the elastic properties of itself against its environment, or as a change against the state. I will explain the latter.

The operator representing the quantum process acts on the vector representing the quantum state and transforms it. Both structures, operators and vectors, are algebraic forms of mutually dual (co- and counter-variant) Hilbert vector spaces[108]. The multiplication of their respective representatives gives "information of perception", better known as a scalar or canonical product. This is one way of generalizing formula (1).

Processes and states are dual structures, whose components have the form of a past attachment[109] as well. Their product is invariant, stable, that is, the most probable compound is at a given moment in a given place. This type of invariance was known even before Einstein, who used it in basing his General Equations, and can be traced through the properties of tensors. The same property of vectors is transferable to states and processes of quantum mechanics.

Thinking within the theory of information in a classical (materialistic) way, the point I stresses out will escape us, although its elements can be found in the most mechanistic known theories.

The change of an electron into the electron again by a process that changes it, and which alternates with the same process that "changes electron into electron", is a dual occurrence of states between

[108] https://royalsocietypublishing.org/doi/10.1098/rsta.2016.0393
[109] 15. Variant Vectors, formula (19).

processes. Replacing a state with the same state allows the process to be changed by the same process. In the case of “particle decay”, we also have the corresponding “process decay”. Consistently, we expand the information of perception ($S = A \cdot B$) to multiply the vectors by the operator, ie map the state by the process, which is an easier part of the work that awaits us, because the (abstract) mathematics of these operators is already very developed. The harder part of the job will be interpretation.

For example, the multiplying of a matrix row by the vector column results in the same position of the

$$\begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a_1x+b_1y+c_1z \\ a_2x+b_2y+c_2z \\ a_3x+b_3y+c_3z \end{pmatrix}$$

new vector as in the figure on the left. It is the phenomenon that “everything depends on everything” that is actually ostensible, because there are also independent vectors in vector spaces. This will prove convenient to clarify the multiplicity, intertwining, and particularity that is required by information theory, which are only at first glance contradictory.

Isomorphism

An isomorphism is a mutually unique mapping between two mathematical structures. Like any bijection, the isomorphism $f : X \to Y$ is also invertible, there is a unique inverse function $f^{-1} : Y \to X$. The existence of (at least one) isomorphism is written $X \cong Y$.

In particular, when X and Y are vector spaces of the same dimension $n = 1,2,3,\dots$ over the same scalar field Φ, we say that they are *algebraically isomorphic* if there is a bijection f with property

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y) \tag{2}$$

for any pair $x, y \in X$ and any $\alpha, \beta \in \Phi$. Each (algebraic) statement about the space X on that occasion turns into the corresponding statement about the space Y.

Thus the linearly independent vectors $x_1, \dots, x_m$ of the space X become linearly independent vectors $y_1 = f(x_1), \dots, y_m = f(x_m)$ of the space Y. Indeed, from

$$\sum_{k=1}^{m} \alpha_k f(x_k) = 0$$

follows

$$f\left(\sum_{k=1}^{m} \alpha_k x_k\right) = 0$$

and $f(x) = 0$ pulls $x = 0$, because f translates zero into zero. Namely, from $f(0) = f(0+0) = f(0) + f(0)$ follows $f(0) = 0$, and due to invertibility we also have $f^{-1}(0) = 0$. Further, from $x = \sum_{k=1}^{m} \alpha_k x_k = 0$ and the independence of x_k follows $\alpha_1 = \dots = \alpha_m = 0$. Therefore, the spaces X and Y are of the same dimension n.

It is easy to check that isomorphism is a relation of equivalence, i.e. valid:

(Reflexivity) $X \cong X$;
(Symmetry) if $X \cong Y$ then $Y \cong X$;
(Transitivity) if $X \cong Y$ and $Y \cong Z$ then $X \cong Z$.

It follows from such considerations

Statement. Any two n-dimensional spaces with the same Φ are mutually isomorphic.

For example, imagine a particle that can be in one of three places on the x-axis, the abscissas -1, 0, and +1. This can be a larger, mean, and smaller distance of a particle from a place (say $x = 2$), or a ranking of the energy states of an arbitrary quantum system, or any successive values. When these interpretations start from isomorphic vector spaces, then the corresponding statements are equivalent.

Assume that the quantum state is $|1\rangle$ where the particle is located, i.e. is at position 1, and the state $|0\rangle$ where the particle is located at position 0. By measuring the position of the particle in the state $|1\rangle$ we find $x = 1$ with probability 1. Similarly, we find $x = 0$ with probability 1 for the particle represented by the vector states $|0\rangle$. These are two (different) representations of the same vector space, but we can also say representations of two isomorphic vector spaces.

These representations should be distinguished from the state $|\psi\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$ which is also an acceptable quantum state. By measuring the position of the particle state $|\psi\rangle$, the outcome will be $x = 0$ or $x = 1$ with 50% probability. It does not allow any other value.

Analogous isomorphic information perception are of the creatures with the same number of (independent) senses. We would achieve an even greater degree of equivalence with the same types of senses, but we still have a lot in common for research. Let us note further that the mapping of the intensity of perception of the mentioned isomorphic creatures, of the same or different senses, represents conjugations which are the subject of the continuation.

Conjugations

We know that the functional on the vectors from X (over the scalars Φ) form a new but equal-dimensional vector space X^* (over the same scalars). From X^* again in a similar way X^{**} is formed, then from this X^{***} and so on, each of the same dimension $n = \dim X$. All these spaces have scalars from the same Φ, are adjugate (adjoint) and mutually isomorphic (Statement).

Let us then observe a pair of spaces X and X^{**}, and single out one, the so-called *natural isomorphism*, defined by a given $x \in X$ and an arbitrary $y^* \in X^*$. The element $y^*(x)$ is completely defined in Φ, and the association $y^* \to y^*(x)$ is linear, because:

$$\alpha_1 y_1^* + \alpha_2 y_2^* \to (\alpha_1 y_1^* + \alpha_2 y_2^*)(x) = \alpha_1 y_1^*(x) + \alpha_2 y_2^*(x).$$

Thus, by given $x \in X$ we obtain a linear functional on X^*. Let $\hat{x}(y^*) = y^*(x)$, so $\hat{x} \in X^{**}$, and different space vectors X correspond to different space vectors X^{**}.

Namely[110], from $x_1 \neq x_2$ follows the existence of the functional $y^* \in X^*$ with the property $y^*(x_1 - x_2) \neq 0$, or $y^*(x_1) \neq y^*(x_2)$. Hence $\hat{x}_1(y^*) \neq \hat{x}_2(y^*)$, i.e. $\hat{x}_1 \neq \hat{x}_2$.

As the relation $\hat{x}(y^*) = y^*(x)$ joins to each $x \in X$ one and only one element $\hat{x} \in X^{**}$, then with $\hat{x} = \varphi \cdot x$, where $\varphi \in \Phi$, we can define an operator on X with values in X^{**}. The corresponding mapping, $\varphi : X \to X^{**}$, is linear, translates different elements into different ones, so it is an isomorphism between the space X and the space $\varphi(X) \subseteq X^{**}$. It is constructed without the use of any basis, and these two spaces have the same algebraic structures, so we can equate them in the sense that instead of the first space we study the second. Equating X and $\varphi(X)$, or x and $\hat{x} = \varphi x$, we can consider X as a subspace of X^{**}.

These considerations have long been known, they are basically algebra. A space X for which $\varphi(X) = X^{**}$, or $X = X^{**}$, is called a *reflexive space*. Every n-dimensional space is reflexive, but this does not apply to infinite-dimensional spaces. Moreover, if the space is reflexive, then it is necessarily finally dimensional.

The difference between the finitely and infinitely dimensional base of spaces arises by equating the linear combination of independent vectors with zero, $x = \xi_1 e_1 + \xi_2 e_2 + \dots + \xi_n e_n = 0$, where in the case of a finite number $n \in \mathbb{N}$ must be $\xi_k = 0$ for each $k = 1,2,\dots,n$. However, in the case of an infinite base, in the limit case $n \to \infty$, it must be $\xi_k = 0$ for "almost all" indices, which means that for a finite number of indices k can be $\xi_k \neq 0$.

Epilogue

Two immediate conclusions emerge after this presentation. The first is that it is difficult, and perhaps impossible, to separate form from the supposed essence of information perception and that there are "countless" layers of equivalent forms. The second is that all these forms are within finitude, although infinity seems to be "at their fingertips".

Thinking a little more freely, in information theory, it is possible to allow infinity under certain conditions. It is essential to note that by subtracting a finite subset from an infinite one, this larger one always remains infinite. Even more, we can extract an infinite subset from an infinite set and the starting subset can remain infinite again, such as separating negative ones from a set of integers.

It is also possible to subtract an infinite number of infinite subsets from an infinite set, while the starting all the time remains infinite. Such is the separation of the product of a given prime number with integers, and then the subtraction of all such infinite subsets by taking in order every second prime number from an unlimited series 3, 5, 7, 11, 13, 17, 19, 23, ..., after which it remains infinite again number (every second) of the prime numbers.

110 16. Variant Vectors II, Example 3

This possibility of infinity, together with its non-contradiction, allows the existence of worlds larger than finite. I have already written[111] that such are actually necessary in the information theory. If we adhere only to the static "contingent of possibilities" through whom we supposedly travel through our own present, we would deny the objectivity of uncertainty. Uncertainty relations, for example, would be a consequence only of our inability to know the final causes, which in that case are tacitly assumed to exist. But if there were "ultimate causes" then there would be no noncommutative operators whose representations are quantum evolution, so we would come to the unacceptable conclusion that algebra is not correct.

[111] [2], 3.17 Present

21. Conjugated Space

April 21, 2021

The conjugated matrix is performed and step by step the unitary matrix of the second order is derived.

Introduction

Question: Why do you call conjugation a "reflexive space", what does that have to do with complex numbers?

Answer: To make it easier for newcomers to understand what it is about. Conjugation, transit, or coupling, in algebra means a change in the sign of an imaginary unit (i – whose square is -1). In a complex plane it is the reflection of a point (representing a complex number) around the x-axis.

Functions that map vectors into scalars (numbers) are called functionals, and when complex numbers make up the body of scalars of vector space, then by conjugation we get reflections (axis-symmetric images). Double reflection is the initial original and that is the essence of the evidence in that article[112].

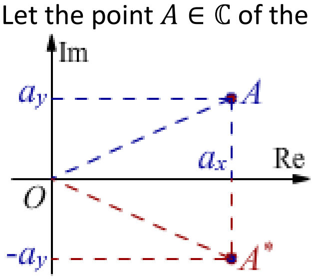
Q: Why don't you say it in such a simplistic way that everyone can understand, but you mathematicians are constantly complicating things?

A: Because it is not told accurately enough. A complex plane is not something special in vector spaces. If I were to reduce the "proof" of that theorem (that a twice conjugate gives a starting space) to this story, it would be as "proof" that the Earth is a flat plate because a meadow is flat.

Enumerating other vector spaces where the statement is valid would not help, as you cannot prove the summation table by listing individual objects and pairs of numbers with such as "two apples plus three apples are five apples". Mathematics is it what it is (extremely accurate), because it does not fish for half-truths. On the contrary, after that "higher" proof, we still know that the same is true "down" for complex numbers, but I will try to think of something.

Conjugated matrix

Let the point $A \in \mathbb{C}$ of the complex plane be given, as in the figure on the left. Its coordinates are



$A(a_x, a_y)$. It represents a complex number $A = a_x + ia_y$ with the real and imaginary part $a_x = \mathrm{Re}\, A$ and $a_y = \mathrm{Im}\, A$ which are real numbers, and $i^2 = -1$ is valid for the imaginary unit i. The point $A^*(a_x, -a_y)$ conjugated to the point A is reflected (axially mapped) around the abscissa, here the "real axis".

We know that the product of conjugate complex numbers is a real number:

$$|A|^2 = AA^* = (a_x + ia_y)(a_x - ia_y) = a_x^2 + a_y^2 \in \mathbb{R}$$

so the question arises how to transfer it to the matrix space.

112 20. Reflexive Space

The matrices are representations of linear operators, and both are types of vector spaces. Quantum states and quantum processes are also interpretations of vectors, where the former are tacitly considered to be common vectors and column matrices, and the latter as unitary operators and quadratic matrices. States and corresponding processes are mutually dual representations of vectors.

Analogously to $A(a_x, a_y)$ we denote the coordinates of the points $B(b_x, b_y)$, $C(c_x, c_y)$, $D(d_x, d_y)$ and define a square matrix of the second order (type 2×2):

$$\mathbf{M} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{1}$$

In order for the product of such conjugated matrices to give a real *trace* (the sum of the coefficients of the main diagonal), it is necessary to transpose the matrix by conjugation. So we get:

$$\mathbf{MM}^* = \begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} A & B \\ C & D \end{pmatrix}^* = \begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} A^* & C^* \\ B^* & D^* \end{pmatrix} = \begin{pmatrix} AA^* + BB^* & AC^* + BD^* \\ CA^* + DB^* & CC^* + DD^* \end{pmatrix}$$

$$\mathbf{M}^*\mathbf{M} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}^*\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} A^* & C^* \\ B^* & D^* \end{pmatrix}\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} A^*A + C^*C & A^*B + C^*D \\ B^*A + D^*C & B^*B + D^*D \end{pmatrix}$$

with the same trace in the both cases:

$$\mathrm{Tr}(\mathbf{MM}^*) = \mathrm{Tr}(\mathbf{M}^*\mathbf{M}) = |A|^2 + |B|^2 + |C|^2 + |D|^2 \tag{2}$$

which is, therefore, a real number. It is special:

$$\begin{aligned} AC^* &= (a_x + ia_y)(c_x - ic_y) = (a_x c_x + a_y c_y) - i(a_x c_y - a_y c_x) = A \cdot C - i[A, C] \\ A^*C &= (a_x - ia_y)(c_x + ic_y) = (a_x c_x + a_y c_y) + i(a_x c_y - a_y c_x) = A \cdot C + i[A, C] \end{aligned}$$

where scalar and pseudo-scalar product, i.e. commutator[113]:

$$\begin{cases} A \cdot C = a_x c_x + a_y c_y = |A||C| \cos \sphericalangle(A, C) \\ [A, C] = a_x c_y - a_y c_x = |A||C| \sin \sphericalangle(A, C) \end{cases} \tag{3}$$

or:

$$\begin{cases} AC^* = |A||C| e^{i\sphericalangle(C,A)} \\ A^*C = |A||C| e^{i\sphericalangle(A,C)} \end{cases} \tag{4}$$

taking into account the direction of the angle. We find similar for B, D and further.

For the elements of the secondary diagonal of the product of the conjugate matrices (1) to be real numbers, it is sufficient that the zeros of both elements of one of the diagonals (main or secondary) of the initial matrix (1). If $B = C = 0$, then, with $|A| = |D| = 1$, we have:

$$\mathbf{MM}^* = \mathbf{M}^*\mathbf{M} = \mathbf{I} \tag{5}$$

wherein

$$\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{6}$$

[113] [1], 3. Potential Information

the unit matrix is also of the second order. The second option is $|A| = |D| = 0$, with $|B| = |C| = 1$, when (5) holds again. In addition to the unit matrix (6), examples of the above are:

$$\boldsymbol{\sigma}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \ \boldsymbol{\sigma}_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \ \boldsymbol{\sigma}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{7}$$

otherwise known *Pauli matrices*[114], or:

$$\mathbf{q}_x = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \ \mathbf{q}_y = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \ \mathbf{q}_z = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \tag{8}$$

which are *quaternions*[115].

Of course, the diagonal elements of a matrix (1) do not have to be zero for its conjugate products to give a real matrix. It is sufficient that the numbers $AC^* + BD^*$ and $A^*B + C^*D$ are real, i.e.

$$(AC^* + BD^*)^* = AC^* + BD^*, \ \ (A^*B + C^*D\,)^* = A^*B + C^*D, \tag{9}$$

$$A^*C + B^*D = AC^* + BD^*, \ \ AB^* + CD^* = A^*B + C^*D.$$

Next, starting with (4), we develop only the first equation:

$$|A||C|e^{i\sphericalangle(A,C)} + |B||D|e^{i\sphericalangle(B,D)} = |A||C|e^{i\sphericalangle(C,A)} + |B||D|e^{i\sphericalangle(D,B)},$$

$$|A||C|\left[e^{i\sphericalangle(A,C)} - e^{i\sphericalangle(C,A)}\right] = |B||D|\left[e^{i\sphericalangle(D,B)} - e^{i\sphericalangle(B,D)}\right],$$

$$2|A||C|e^{i\sphericalangle(A,C)} = 2|B||D|e^{i\sphericalangle(D,B)},$$

$$|A||C| = |B||D|e^{i[\sphericalangle(D,B)+\sphericalangle(C,A)]}$$

The left side of this equation is a real number, so it must be the right one. We work similarly with the right equation (9) and get:

$$\begin{cases} \sphericalangle(D,B) + \sphericalangle(C,A) = 0 \\ \sphericalangle(A,B) + \sphericalangle(C,D) = 0 \end{cases} \tag{10}$$

as far as angles are concerned.

Regarding the intensity of these complex numbers, there will be:

$$|A||C| = |B||D|, \ \ |A||B| = |C||D|. \tag{11}$$

When $|B||C| \neq 0$, multiplying these two equations we get $|A| = |D|$, and by dividing $|B| = |C|$.

Example 1. Here is a matrix that meets conditions (11), but not (10):

$$\mathbf{M} = \begin{pmatrix} 2+i & 3-5i \\ 5+3i & 1+2i \end{pmatrix}, \ \mathbf{M}^* = \begin{pmatrix} 2-i & 5-3i \\ 3+5i & 1-2i \end{pmatrix},$$

---

[114] [3], formula (2.56)
[115] [3], formula (2.58)

$$\mathbf{MM}^* = \begin{pmatrix} 39 & 6-12i \\ 6+12i & 39 \end{pmatrix}, \ \ \mathbf{M}^*\mathbf{M} = \begin{pmatrix} 39 & 12-6i \\ 12+6i & 39 \end{pmatrix}.$$

As we can see, not all coefficients of conjugated products are real numbers. □

Example 2. The following matrix satisfies conditions (10), but not (11):

$$\mathbf{M} = \begin{pmatrix} 6+3i & 1+2i \\ 2+4i & 2+i \end{pmatrix}, \ \mathbf{M}^* = \begin{pmatrix} 6-3i & 2-4i \\ 1-2i & 2-i \end{pmatrix},$$

$$\mathbf{MM}^* = \begin{pmatrix} 50 & 28-15i \\ 28+15i & 25 \end{pmatrix}, \ \ \mathbf{M}^*\mathbf{M} = \begin{pmatrix} 65 & 20+3i \\ 20-3i & 10 \end{pmatrix}.$$

These coefficients of conjugated products are not real numbers either. □

Example 3. The following matrix meets both conditions (10) and (11):

$$\mathbf{M} = \begin{pmatrix} -2+i & 1+2i \\ 1+2i & -2+i \end{pmatrix}, \ \ \mathbf{M}^* = \begin{pmatrix} -2-i & 1-2i \\ 1-2i & -2-i \end{pmatrix},$$

$$\mathbf{MM}^* = \begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix}, \ \ \mathbf{M}^*\mathbf{M} = \begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix}.$$

All coefficients of conjugated products are now real. □

Unitary matrix

A square matrix is unitary if its conjugate (and transposed) is equal to its inverse. In the last example (Example 3) it would be a matrix $\mathbf{U} = \frac{1}{\sqrt{10}}\mathbf{M}$. Because the product of the matrix and the conjugate is a real (identity) matrix, so the unitary matrix meets conditions (10) and (11). And now we will derive its general form on the basis of previous considerations.

We define arbitrary coefficients of the matrix (1), and use conditions (10):

$$A = pe^{i\alpha}, \ \ B = qe^{i\beta}, \ \ C = re^{i\gamma}, \ \ D = se^{i\delta}, \tag{12}$$

$$\sphericalangle(D,B) + \sphericalangle(C,A) = 0, \ \ \sphericalangle(A,B) + \sphericalangle(C,D) = 0,$$

$$(\delta-\beta) + (\gamma-\alpha) = 0, \ \ (\alpha-\beta) + (\gamma-\delta) = 0,$$

$$\gamma = \beta, \ \ \delta = \alpha. \tag{13}$$

Then we use conditions (11):

$$|A||C| = |B||D|, \ \ |A||B| = |C||D|,$$

$$|pe^{i\alpha}||re^{i\beta}| = |qe^{i\beta}||se^{i\alpha}|, \ \ |pe^{i\alpha}||qe^{i\beta}| = |re^{i\beta}||se^{i\alpha}|,$$

$$|pr| = |qs|, \ \ |pq| = |rs|,$$

$$|r| = |q|, \quad |s| = |p|. \tag{14}$$

It follows from (12), (13) and (14)

$$\mathbf{M} = \begin{pmatrix} pe^{i\alpha} & qe^{i\beta} \\ \pm qe^{i\beta} & pe^{i\alpha} \end{pmatrix}, \tag{15}$$

where the (positive) sign of the coefficient at the bottom right can be determined from the condition that the product of the conjugate matrices (15) is a real matrix, multiplying them directly:

$$\mathbf{MM}^* = \begin{pmatrix} pe^{i\alpha} & qe^{i\beta} \\ \pm qe^{i\beta} & pe^{i\alpha} \end{pmatrix}\begin{pmatrix} pe^{-i\alpha} & qe^{-i\beta} \\ \pm qe^{-i\beta} & pe^{-i\alpha} \end{pmatrix} = \begin{pmatrix} p^2 \pm q^2 & 2pq\cos(\alpha-\beta) \\ \pm 2pq\cos(\alpha-\beta) & p^2 \pm q^2 \end{pmatrix},$$

$$\mathbf{M}^*\mathbf{M} = \begin{pmatrix} pe^{-i\alpha} & qe^{-i\beta} \\ \pm qe^{-i\beta} & pe^{-i\alpha} \end{pmatrix}\begin{pmatrix} pe^{i\alpha} & qe^{i\beta} \\ \pm qe^{i\beta} & pe^{i\alpha} \end{pmatrix} = \begin{pmatrix} p^2 \pm q^2 & 2pq\cos(\alpha-\beta) \\ \pm 2pq\cos(\alpha-\beta) & p^2 \pm q^2 \end{pmatrix}.$$

The sign of the coefficient at the bottom left of the matrix (15), in front of q, can be plus or minus.

The matrix (15) is unitary if the conjugated products give a unit matrix. This further defines:

$$p = \cos\omega, \quad q = \sin\omega, \quad \alpha - \beta = \pm\frac{\pi}{2}, \tag{16}$$

where the angle φ is arbitrary, so arbitrary that p and q can be hyperbolic sine and cosine as well. Thus, in the case of the plus sign, the most general form of the unitary matrix becomes

$$\mathbf{U} = e^{i\alpha}\begin{pmatrix} \cos\omega & -\sin\omega \\ \sin\omega & \cos\omega \end{pmatrix}, \tag{17}$$

which means that it consists of a composition of two rotations, complex $e^{i\alpha}$ for the angle α and real for the angle ω. I may talk about the hyperbolic version on another occasion.

22. Rotations

April 25, 2021

Several explanations of complex numbers, rotations and their interpretations in information theory are given in the appendix.

Introduction

Question: What is the “catch” with these unitary operators?

Answer: All isometric transformations (rigid, which do not change the distances between points) are some rotations: translation, reflections (plane, line, central symmetry), and of course rotations. This has been noticed in geometry for a long time, and only quantum mechanics, created at the beginning of the 20th century, gave it importance in physics. All quantum processes of representation known today are of some unitary operators, and they consist of two rotations[116]. All the processes of macro-physics, we imagine further, are complex compositions of these more elementary ones.

Q: It's a little harder to follow these formulas, can you clarify? What does information theory have to do with this?

A: This unitary operator consists of one real rotation (in 3D space) and one imaginary (complex). When the imaginary one appears in a real form (it is not often) then it is interpreted as an observable (physically measurable quantity). With information theory, I try to understand those “imaginary” states, at least as “not presence”, and I hope (I am testing for now) to interpret it even more, as “presence” of the phenomenon in question, but not in “our reality” but in pseudo-reality. She is at the starting point of my information theory. I'm talking about a parallel universe, or multiverse or Everett's “many worlds” of quantum mechanics.

Q: What do we get from this interpretation?

A: I guess the truth. We get the rational explanation of an imaginary (complex) number. This area of mathematics has long been rationalized (e.g. for easier, faster and safer control of air traffic of larger airports), but still not as directly as it may soon be.

***

Question: Can you give me a simpler explanation of the “double-slit” experiment from the point of view of “information theory”?

Answer: I have several[117], and here is one that is perhaps closest to the official observation of things in quantum physics. First of all, I remind you, it is paradoxical in that old Jung's experiment from 1801, by which he proved the wave nature of light, that the interference of a particle (photons, electrons, any) occurs even when we let only one particle through two openings with arbitrarily long breaks. It even

[116] 21. Conjugated Space
[117] [2], 2.17 Double Slit

then interferes with itself like real waves, but this does not happen when we pass the wave or the particle itself through only one opening.

The appearance is a consequence of superposition, a state of more possibilities before interaction with the screen (measurement). These are the possibilities of the dice before the roll, the future outcomes of some of the numbers from 1 to 6, which are equally “real” to each other in the time before the realization. The path of the electron becomes defined only after the measurement – some founders of quantum mechanics noticed at the time. The trajectory of a particle is in a much greater state of objective uncertainty before than after measurement or interaction, I will add from the point of view of information theory – because by interaction it transmits part of its information (indeterminacy) to the measuring device.

Thus, traveling through the double slot towards the screen where it will be measured (observed), the particle is really in all possible states that its uncertainty allows. Such can interfere simply because that uncertainty is objective, assuming information theory. It physically exists in its uncertain states even when we do not measure it.

By closing one of the openings, the superposition loses options and the path of the particle before the measurement becomes more certain, which affects the shape of its appearance on the screen at the end of the road. Interaction is a similar loss of options also, then due to the transition to a state of less information, or more probability by conjunction with the environment.

Q: Isn't “uncertainty” a total, uncertain thing?

A: If I understood correctly, the dilemma is that before rolling, the dice can have more possibilities than the real ones, that in addition to the given six numbers, say 13 is on the list of its possibilities too. However, this is wrong thinking, not only the outcomes but the uncertainties also have their limitations. Losing a part of it, the uncertainty remains with more certainty (I remind you, the information is quantity, and uncertainty is its type). The cube has six, and the coin has only two.

Q: How is it possible to have multiple explanations of the same? Which of them is wrong?

A: Theorems in geometry can also be proved using algebra (analytical geometry). Not all proves are equally simple or understandable, although they are equally accurate. So it is with physics, and so it is with information theory.

Q: Where is the “information of perception” in this story?

A: In the probability of coupling the particle we are measuring and the device we are measuring it with. These are two quantum states, two series of numbers whose product defines the probability of interaction and information which that state then possesses.

Rotation matrix

The rotation matrix of the Oxy coordinate plane about the origin O for the angle ω is

$$R_\omega = \begin{pmatrix} \cos\omega & -\sin\omega \\ \sin\omega & \cos\omega \end{pmatrix} \tag{1}$$

We know how to prove it in various ways, and here are some non-standard ones that I have used for years.

For the scalar and vector product of vectors $A(a_x, a_y, 0)$ and $B(b_x, b_y, 0)$ the equations apply:

$$A \cdot B = a_x b_x + a_y b_y = |A||B|\cos\omega, \tag{2}$$

$$\begin{cases} A \times B = (0 \quad 0 \quad a_x b_y - a_y b_x) \\ [A, B] = a_x b_y - a_y b_x = |A||B|\sin\omega \end{cases} \tag{3}$$

where $|A| = \sqrt{a_x^2 + a_y^2}$ and $|B| = \sqrt{b_x^2 + b_y^2}$ are the intensities of the vectors, and $\omega = \sphericalangle(AOB)$ is the angle between them. In particular, if the vector $B = R_\omega A$, then the vectors A and B satisfy equations (1) and (2).

Example 1. Using (2) we show that (1) is the rotation for the angle $\omega = \sphericalangle(AOB)$.

Solution. Let $p^2 + q^2 = 1$ and $A(p, q)$. Then:

$$(p \quad q)\begin{pmatrix} \cos\omega & -\sin\omega \\ \sin\omega & \cos\omega \end{pmatrix}\begin{pmatrix} p \\ q \end{pmatrix} = (p \quad q)\begin{pmatrix} p\cos\omega - q\sin\omega \\ p\sin\omega + q\cos\omega \end{pmatrix} =$$

$$= p^2\cos\omega - pq\sin\omega + qp\sin\omega + q^2\cos\omega =$$

$$= (p^2 + q^2)\cos\omega = \cos\omega,$$

which due to (2) means that the vector A multiplied by (1) is rotated for ω. ■

Example 2. Using (2) and (3) derive the rotation (1).

Solution. Let $|A| = |B|$. Then, first:

$$b_x = \frac{(a_x^2 + a_y^2)b_x}{|A||B|} = \frac{a_x^2 b_x + a_x a_y b_y - a_x a_y b_y + a_y^2 b_x}{|A||B|}$$

$$= a_x \frac{a_x b_x + a_y b_y}{|A||B|} - a_y \frac{a_x b_y - a_y b_x}{|A||B|} = a_x\cos\omega - a_y\sin\omega$$

then:

$$b_y = \frac{(a_x^2 + a_y^2)b_y}{|A||B|} = \frac{a_x^2 b_y - a_x a_y b_x + a_x a_y b_x + a_y^2 b_y}{|A||B|} =$$

$$= a_x \frac{a_x b_y - a_y b_x}{|A||B|} + a_y \frac{a_x b_x + a_y b_y}{|A||B|} = a_x\sin\omega + a_y\cos\omega$$

that is

$$B = \begin{pmatrix} b_x \\ b_y \end{pmatrix} = \begin{pmatrix} \cos\omega & -\sin\omega \\ \sin\omega & \cos\omega \end{pmatrix} \begin{pmatrix} a_x \\ a_y \end{pmatrix} = R_\omega A$$

which should have been shown. ■

Formulas (2) and (3) of the scalar and vector product of the vector otherwise, as well as

$$\cos\omega = \frac{a_x b_x + a_y b_y}{|A||B|}, \quad \sin\omega = \frac{a_x b_y - a_y b_x}{|A||B|}, \tag{4}$$

we see and use on various occasions. Let's look at one of their lesser known uses. Let $A(a_x, a_y)$ and $B(b_x, b_y)$ points be in complex planes, such that $A = a_x + ia_y$ and $B = b_x + ib_y$_y are complex numbers, where $i^2 = -1$ holds for the imaginary unit.

For the product of conjugate complex numbers we get:

$$A^*B = (a_x + ia_y)^*(b_x + b_y) = (a_x - ia_y)(b_x + ib_y) =$$

$$(a_x b_x + a_y b_y) + i(a_x b_y - a_y b_x) = |A||B|(\cos\omega + i\sin\omega) = |A||B|e^{i\omega}$$

that is

$$A^*B = A \cdot B + i[A, B]. \tag{5}$$

The real part (scalar product $A \cdot B$) remains in the plane of the given complex numbers ($A, B \in \mathbb{C}$) while the imaginary part (vector product $A \times B$) can be considered perpendicular to that plane. Using (4), we also see that (5) is the rotation of a complex plane around that perpendicular for the angle ω.

Example 3. Show that (5) is the rotation for the angle ω.

The solution. Let 1 and i are the base vectors of the complex plane $\mathbb{C}$ with the given points $A(a_x, a_y)$ and $B(b_x, b_y)$, and let $1'$ and i' are the base vectors of the complex plane $\mathbb{C}'$, then from:

$$\begin{cases} 1'\dfrac{a_x b_x + a_y b_y}{|A||B|} - i'\dfrac{a_x b_y - a_y b_x}{|A||B|} = 1'\cos\omega - i'\sin\omega = 1 \\ 1'\dfrac{a_x b_y - a_y b_x}{|A||B|} + i'\dfrac{a_x b_x + a_y b_y}{|A||B|} = 1'\sin\omega + i'\cos\omega = i \end{cases}$$

follows:

$$(1' \quad i')\begin{pmatrix} \cos\omega & -\sin\omega \\ \sin\omega & \cos\omega \end{pmatrix} = (1 \quad i)$$

and that is the rotation of the base vectors for the angle ω. ■

When the scalar product represents the product (distribution) of probabilities, and the vector product the information, then (5) can represent the "information of perception". The real part (products of conjugate numbers A^*B) is as larger as the "closer" the distributions are. When the angle $\omega \to 0$, then $\cos\omega \to 1$ and the probability of association is higher, and the imaginary part (surface $|A \times B|$) is then smaller $\sin\omega \to 0$. In accordance with the more frequent occurrence of more probable outcomes and also "parsimony of information", when A и B represent states, then (5) evaluates the realization of their coupling.

Characteristic values

The products of the corresponding co- and counter-variant vectors determine the "information of perception". It represents the vitality of the coupling of two states (multiplied vectors) and the chance of association, and its value thus defined remains the same (it is invariant) when the coordinate system changes. The strangest thing about such products is that linear operators (processes) are dual vectors of state vectors. In that context, the characteristic values are especially interesting.

Let a given linear transformation be represented by a matrix $M \in X^*$ of type $n \times n$ and its eigenvector $u \in X$ of the same order (further for simplicity we take $n = 2$)

$$Mu = \lambda u, \tag{6}$$

where scalar $\lambda \in \Phi$ is the eigenvalue of a given vector. In the case of rotation (1) we get:

$$\begin{pmatrix} \cos\omega & -\sin\omega \\ \sin\omega & \cos\omega \end{pmatrix}\begin{pmatrix} u_x \\ u_y \end{pmatrix} = \lambda\begin{pmatrix} u_x \\ u_y \end{pmatrix},$$

$$\begin{pmatrix} \cos\omega & -\sin\omega \\ \sin\omega & \cos\omega \end{pmatrix}\begin{pmatrix} u_x \\ u_y \end{pmatrix} = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}\begin{pmatrix} u_x \\ u_y \end{pmatrix},$$

$$\begin{pmatrix} \cos\omega - \lambda & -\sin\omega \\ \sin\omega & \cos\omega - \lambda \end{pmatrix}\begin{pmatrix} u_x \\ u_y \end{pmatrix} = 0. \tag{7}$$

We have obtained a homogeneous system of equations that we know can have a nontrivial solution (other than zero-vector) if the determinant of the system is zero:

$$\det\begin{pmatrix} \cos\omega - \lambda & -\sin\omega \\ \sin\omega & \cos\omega - \lambda \end{pmatrix} = 0,$$

$$(\cos\omega - \lambda)^2 + \sin^2\omega = 0,$$

$$\lambda^2 - 2\lambda\cos\omega + 1 = 0,$$

$$\lambda_{12} = \frac{2\cos\omega \pm \sqrt{4\cos^2\omega - 4}}{2} = \begin{cases} \cos\omega + i\sin\omega \\ \cos\omega - i\sin\omega \end{cases}$$

Thus, there are two eigenvalues of rotation, both complex numbers:

$$\lambda_\pm = e^{\pm i\omega}. \tag{8}$$

Knowing the eigenvalues, continuing (7) we calculate the corresponding eigenvectors:

$$\begin{pmatrix} \cos\omega - e^{\pm i\omega} & -\sin\omega \\ \sin\omega & \cos\omega - e^{\pm i\omega} \end{pmatrix} \begin{pmatrix} u_x \\ u_y \end{pmatrix} = 0,$$

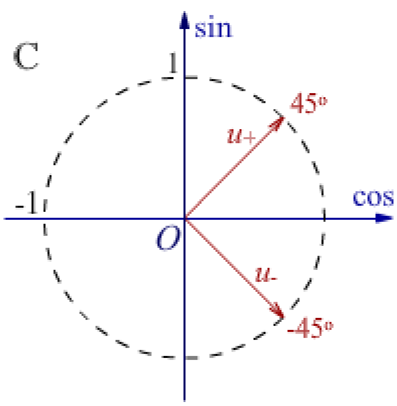
$$\begin{pmatrix} -i\sin\omega & -\sin\omega \\ \sin\omega & -i\sin\omega \end{pmatrix}_+ \begin{pmatrix} u_x \\ u_y \end{pmatrix} = 0, \quad \begin{pmatrix} i\sin\omega & -\sin\omega \\ \sin\omega & i\sin\omega \end{pmatrix}_- \begin{pmatrix} u_x \\ u_y \end{pmatrix} = 0,$$

$$u_x = \pm i u_y$$

so the eigenvectors are:

$$u_+ = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix}, \quad u_- = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix}. \tag{9}$$

They form the orthonormal base of these rotations, because $|u_+| = |u_-| = 1$ and $u_+ \perp u_-$, ie. $u_+ \cdot u_- = 0$.



The figure on the left shows two eigenvectors of rotation in a complex plane. The vertices of both are located on a unit circle with the center at the origin. The first (u_+) is the angle $\frac{\pi}{4}$ radians (45°) with the abscissa, and the second $-\frac{\pi}{4}$. Another way to write the same is:

$$u_+ = \cos\frac{\pi}{4} + i\sin\frac{\pi}{4}$$
$$u_- = \cos\frac{\pi}{4} - i\sin\frac{\pi}{4}$$

because the cosine is an even function, so $\cos(-x) = \cos x$, and the sine is odd and $\sin(-x) = -\sin x$. They are the base of the rotation space vectors. In addition to the mentioned conditions of orthonormalization, for these vectors it is valid that each vector of rotation is some linear combination of them.

Indeed, for arbitrary coordinates a and b, from

$$\begin{pmatrix} a \\ ib \end{pmatrix} = \frac{x}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix} + \frac{y}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix}$$

a regular system of linear equations follows

$$\begin{cases} \frac{x}{\sqrt{2}} + \frac{y}{\sqrt{2}} = a \\ \frac{x}{\sqrt{2}} - \frac{y}{\sqrt{2}} = b \end{cases}$$

and hence the solution to the unknown

$$x = \frac{a+b}{\sqrt{2}}, \quad y = \frac{a-b}{\sqrt{2}}.$$

Thus, for each pair (a, b) there is a pair (x, y) such that each complex number $z = a + ib$ can be written $= xu_+ + yu_-$.

Conclusion

Every isometric transformation (every physical process to which the law of conservation applies) can be decomposed into some rotations (representations of rotations). This also applies to the representation of complex numbers. Given the property of the product of conjugate complex numbers (5), the coupling of the states they represent, will be more certain if the real part of that product is larger, i.e. the imaginary part is smaller.

Due to the laws of large numbers of probability theory, this further means that more elementary (physically simpler) systems can more easily be in an imaginary, unreal state. Otherwise, the existence of the vector space base, as well as the evidence of the existence of the smallest portions of (free) information, supports this conclusion.

23. Vector Projections

April 29, 2020

Measurement in quantum physics, the Hermitian operator, and the position of information theory are explained.

Measurement

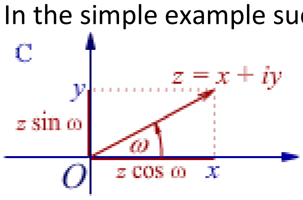
Question. Can you explain the measurement in quantum mechanics to me with a simple example?

Answer. For example, let's consider the throwing (fair or unfair) coins. Possibilities are "heads" or "tails", and the outcome is one of them. Observable, the physically measurable quantities are represented by the rectangular Cartesian coordinates Oxy. The state before tossing a coin is a vector with two components, and the realizations are projections of the vector on the abscissa and the ordinate (x and y axis).

Q. Is it a quantum-mechanical or IT explanation?

A. The beginning is common. Official quantum physics considers only real outcomes. It does this by means of vectors with complex coefficients, but is limited to Hermitian operators whose eigenvalues are only real numbers. It observes only real observables, those that belong to a given invariant observer. The IT explanation goes a step further and allows complex outcomes too, and sees the same situation as a combination of two states, object and subject, both changeable.

In the simple example such as coin toss, the vector of the (binary) state of uncertainty can also be a



complex number $z = x + iy \in \mathbb{C}$, where $i^2 = -1$ again holds for the imaginary unit. In the picture on the left, there is a general one, with an abscissa and an ordinate on which there are real and imaginary parts of a complex number, $\mathrm{Re}\, z = x$ and $\mathrm{Im}\, z = y$, and both are real numbers. They determine the probabilities of observing the "heads" and the "tails". The square of the state projection is the probability of realization so that the sum of the squares of the outcome is one, because the outcome of one of the two possibilities is a certain event:

$$|z|^2 = z^* z = (x + iy)^*(x + iy) = (x - iy)(x + iy) = x^2 + y^2 = 1. \qquad (1)$$

The number $|z|$ is called modulus, absolute value, or intensity of a complex number $z = x + iy$. The number $z^* = x - iy$ is conjugated to the number z, and the product $z^* z$ remains unchanged after the rotation of the complex plane $\mathbb{C}$ about the origin O for an arbitrary angle φ. Namely,

$$z \to z e^{i\varphi},\ z^* \to z^* e^{-i\varphi},\ \text{then } z^* z \to \left(z^* e^{-i\varphi}\right)\left(z e^{i\varphi}\right) = z^* z.$$

We still stick to the classical quantum-mechanical interpretation.

In the case of more than two possibilities, six when rolling the dice, we consider all individual outcomes as special observables and assign each of them one coordinate axis. The squares of the projections of the vectors on these axes (the intensity of the vector multiplied by the cosine of the angle to the axis) are the probabilities of realization, so that the sum of the squares of all projections is again one.

We distinguish between states before and after the realization of a random event. In quantum mechanics, it is common to call possibilities “superposition” of the quantum state and say that they “collapse” into one of the outcomes. In the IT understanding, we add a more equal treatment of possible outcomes.

In addition, we distinguish what is perceived from the one who perceives, for each individual (possible) realization. The first is the state of the object $z = x + iy$, the second is the state of, say, the k-th subject $z_k = x_k + iy_k$ of the parallel realities[118]. The combination of these two is the “information of perception” of the event, the product of the first conjugate complex number with the second:

$$z^* z_k = (x + iy)^*(x_k + iy_k) = (x - iy)(x_k + iy_k) =$$

$$= (xx_k + yy_k) + i(xy_k - yx_k) = z \cdot z_k + i|z \times z_k|$$

$$= |z||z_k| \cos\omega + i|z||z_k| \sin\omega$$

where $\omega = \measuredangle(z, z_k)$. We get

$$z^* z_k = |z||z_k|(\cos\omega + i\sin\omega)$$

what can be written

$$z^* z_k = |z||z_k|e^{i\omega}. \tag{2}$$

This is a well-known result of complex number theory, but its equivalent and less known forms in derivation are now considered more important. One of these more important is the commutator, or pseudo-scalar product

$$[z, z_k] = |z \times z_k| = xy_k - yx_k = |z||z_k| \sin\omega, \tag{3}$$

which is larger with the coupling information of the states z и z_k, opposed to the scalar product

$$z \cdot z_k = xx_k + yy_k = |z||z_k| \cos\omega, \tag{4}$$

which expresses the probability of coupling z and z_k. These are the results of the previous attachment[119].

In IT interpretation, it is possible to view each of the possible outcomes as a separate “pseudo reality” in which the realization takes place (the possibilities are objective), so that's how we work. The index k, and the state z_k of a given subject, should take as many values as possible outcomes, and these outcomes are equivalent to the observables of quantum mechanics.

[118] [2], 2.13 Space and Time.
[119] 22. Rotations.

Suppose that a random event was a roll of the dice and let the "third reality" ($k = 3$) be the one where the "three" fell. There, the "objective" and "subjective" states coincide ($z = z_k$), so the result (2) becomes $z^* z = 1$, and each of the other five where the "three did not fall" is not real. The angle (state vector) of the realized "three" with such $\omega \neq 0$, so $\sin\omega \neq 0$. Therefore, the imaginary part of the coupling, the product (2), does not disappear, and from the point of view of this outcome ($k = 3$) any other outcome ($s \neq k$) is not real.

Hermitian operator

A given matrix M is said to be a Hermitian conjugated matrix that has been transposed to it and has conjugated coefficients. We denote it shorter only by conjugating M^* and call it conjugated, implying transposing. We will mention if a different need arises. A square matrix is Hermitian if it is equal to its (Hermitian) conjugate. A linear operator is Hermitian if its matrix is Hermitian.

These are the operators of "quantum evolution". From the point of view of information theory, I will add, the use of Hermitian operators to treat the process implies the equality of objective and subjective, that is the only one realized outcome. For this reason alone, the development of this theory will have to abandon the only use of Hermitian operators.

Matrices and operators are types of vectors, so the definition of "hermit" can be easily transferred to vectors. The vector-row (Hermitian) conjugate passes into the vector-column and vice versa. With Dirac's bra-ket notation, writing is simplified, for example:

$$\langle\phi| = \begin{pmatrix} \phi_1^* & \cdots & \phi_n^* \end{pmatrix}, \quad |\phi\rangle = \begin{pmatrix} \phi_1 \\ \cdots \\ \phi_n \end{pmatrix}, \tag{5}$$

so it is

$$\langle\phi|\psi\rangle = \phi_1^*\psi_1 + \cdots + \phi_n^*\psi_n, \quad |\phi\rangle\langle\psi| = \begin{pmatrix} \phi_1\psi_1^* & \cdots & \phi_1\psi_n^* \\ \cdots & \cdots & \cdots \\ \phi_n\psi_1^* & \cdots & \phi_n\psi_n^* \end{pmatrix}, \tag{6}$$

in the usual interpretations of the scalar (inner) product, and matrix.

Therefore, $|\psi\rangle^* = \langle\psi|$ and $\langle\psi|^* = |\psi\rangle$, and in general the eigenvalue, otherwise the scalar $\lambda \in \mathbb{C}$, of the eigenvector $|\psi\rangle$ of the matrix (operator) M, we write

$$M|\psi\rangle = \lambda|\psi\rangle. \tag{7}$$

Let us now prove one well-known theorem of linear algebra.

Theorem. The eigenvalues of the Hermitian operator are real. Mutually perpendicular are eigenvectors corresponding to different eigenvalues of the Hermitian operator.

Proof. Multiplying on the left side (7) by the vector $\langle\psi|$ we get $\langle\psi|M|\psi\rangle = \langle\psi|\lambda|\psi\rangle = \lambda\langle\psi|\psi\rangle$, and that is λ multiplied by the intensity of the vector ψ. When the matrix is Hermitian, by conjugating (7) we get $\langle\psi|M = \langle\psi|\lambda^*$, so we find similarly $\langle\psi|M|\psi\rangle = \langle\psi|\lambda^*|\psi\rangle$ and equality $\lambda = \lambda^*$, which means that the eigenvalue of the Hermitian matrix (operator) is a real number $\lambda \in \mathbb{R}$.

That the arbitrary eigenvectors ϕ и ψ to which different eigenvalues belong, respectively a and b, of the Hermitian matrices M are perpendicular, we prove as follows:

$$a\langle\phi|\psi\rangle = \langle a\phi|\psi\rangle = \langle\phi|M|\psi\rangle = \langle\phi|b\psi\rangle = b\langle\phi|\psi\rangle$$

and subtracting the end values

$$(a-b)\langle\phi|\psi\rangle = 0$$

so since by the assumption $a \ne b$, it is $\langle\phi|\psi\rangle = 0$, so these vectors are mutually perpendicular, $\phi \perp \psi$. ■

With this theorem, it becomes clearer why in information theory we should talk about different dimensions of space-time, where the realized event happened and did not happen. Measurements based on different eigenvalues originate from states from different dimensions of the vector space, and that structure must then correspond to the physical structure up to the isomorphism.

Different eigenvectors that correspond to the same eigenvalue do not have to be mutually perpendicular (orthogonal), because the same interaction (measurement) can be given by different states of the same reality[120]. Also, the orthonormal base vectors (unit and mutually perpendicular) do not have to be eigenvectors of any corresponding linear operator; they do not have to be the outcomes of the same event. But, as we have seen, there are operators whose eigenvectors are orthogonal and, therefore, it is possible to have a base that is orthonormal and consists of eigenvectors.

For example, an identical matrix I, which has units on the main diagonal and all other zeros, commutes with any matrix ($MI = IM$) and its eigen-vector is every vector of a given space, and no other matrix has this property. Therefore, it is not true that two commutative Hermitian operators necessarily share a common eigenbase[121], but if two operators share one eigenbase, in other words, if they can be mutually diagonalized, they must be commutative. That is the explanation from the point of view of algebra.

Recall that noncommutativity indicates Heisenberg's relations of uncertainty[122]. By increasing the definiteness of one physical quantity (momentum of a particle, or energy), the definiteness of another (position, or time of the measurement) is lost. Add to this the principled tendency of the state towards less information (higher probability), which interferes with commutativity, so we notice that all the amount of uncertainty, information before the realization of a random event, passes into the information of the one of outcomes. The law of conservation in that case, that two operators share one own (eigen) base, is not an obstacle to commutativity. And that is the explanation of the previous, now from the point of view of physics and information.

---

[120] Here I interpret the otherwise well-known theorems of linear algebra.
[121] I have seen this alleged claim in the texts.
[122] [2], 3.3 Uncertainty Relations

It is also well known in mathematics that matrices associated with Hermitian operators can always be diagonalized, from which it follows that a Hermitian operator on an n-dimensional vector space has n linearly independent eigenvectors. That is why in quantum mechanics it is always possible to take observables for representations of such (base vectors).

When the matrix M is associated with some (orthogonal) operator of the n-dimensional space of the orthonormal base, then the columns (rows) of the matrix are mutually orthonormal vectors. Also, $M^*M = I$, as well as $M^{-1}M = I$, from which it follows that its determinant is ± 1 and that its hermit conjugate is equal to its inverse matrix, $M^* = M^{-1}$. These matrices (operators) are actually rotations (reflections can be reduced to rotations).

When the matrix M is associated with a unitary operator in the orthonormal base of a vector n-dimensional space, then the columns (also rows) are orthonormal vectors and $M^*M = I$ is valid again, as well as $M^{-1}M = I$, from which it follows that the hermitically conjugated matrix is equal to its inverse.

Similar to Hermitian operators, the eigenvectors of unitary and orthogonal operators are orthogonal. Proves of these statements are easy to find in many textbooks of higher algebra and there is no need to repeat them here.

Epilogue

The information theory I am developing agrees well with the idea of "parallel realities", which I have written about in various ways on several occasions. In this paper, you will recognize an attempt to elaborate this idea on the method of projecting a vector (quantum state) onto coordinate axes (observable), which in quantum physics today is routinely treated by describing the process of physical measurement.

I note that "measurement" here is an interaction with measuring devices, i.e. a kind of "information of perception" between two states, special only in that the other state is the base vector.

24. Various questions

May 2, 2021

Here are rests, unnecessary repetitions or unfinished questions, but interesting in their own way.

Question: How do you explain noncommutativity?

Answer: In mathematics, noncommutativity is the impossibility of changing the order of computational operations without changing the results. If A means "add number one" and B is the operation "doubling the number", then $BA(x) = B(x + 1) = 2x + 2$, and $AB(x) = A(2x) = 2x + 1$. These two compositions of operations are noncommutative, because $2x + 2$ is not equal to $2x + 1$ for at least one x (here not even for one).

Q: What about noncommutativity in quantum physics and information theory?

A: In physics, starting with examples like $(BA - AB)(x) = 1$ for each number x, we would further find that the noncommutativity of some processes is in the very essence of Heisenberg's relations of uncertainty[123]. Processes are representations of operators. It is a thesis of quantum mechanics that resists all attempts at destruction with extraordinary tenacity, including by some of the founders and greats of that branch of physics, such as Einstein. In other words, by disputing the "objectivity of indeterminacy" (whatever that means), we will hit our head against this wall, that there are noncommutative processes and that their negation implies the denial of the accuracy of algebra itself.

In information theory, we understand noncommutativity by means of spontaneous increase of entropy[124] (second law of thermodynamics), then spontaneous decrease of information (increase of entropy is decrease of information), or the principle of least action otherwise undoubted in classical theoretical physics (and information is equivalent to action), and also by means of "probability principle": that more probable random events happen more often, which are more likely to be less informative. So we have rounded up all four positions.

Because of these four laws, and then because there is a direction of the outcome of the process on the roads that can be disastrous, going around in a different order becomes important. It is not the same to give a turn signal to a vehicle and go to the left, as to do it in reverse order, turn and signal, at least as far as the number of victims in traffic is concerned. If process A is fatal (no further) and B is not, then $BA(x)$, where x is now an event, will not have the phase B, while $AB(x)$ will flow to the end. Therefore, in a broader sense, "non-commutativity" is a consequence of the "principle of information minimalism".

***

Question: Do scientists behave "like sheep"?

---

[123] [2], 3.3 Uncertainty Relations

[124] [2], 2.24 Entropy Generalization

Answer: There is a formal connection of all forms of transfer of personal freedoms to the collective. Organized behavior, which means a greater degree of participation of the whole in management (increase of options and operability of the group), due to the law of information conservation, must draw that excess of options (information) from somewhere. They are almost always individuals who make up a group, who as living beings have information in excess, and because of the principle of information minimalism they want to get rid of that excess. That is why, among other things, we hand over our freedoms to the state.

Mark Twain noticed similarly with his famous question about three cows in a meadow. If two cows look in one direction, where does the third cow look? – he wondered to answer – Where those two are looking! The movement of starlings[125] as they fly in their flocks is analogous and confusing, or fantastic. It is the same with scientists, especially in well-organized institutions, but also with those “free shooters” who depend on them.

By surrendering his freedom to the institution, that is, some surplus of his own information, the individual scientist is all the more stupid, but also greater because of the synergistic influence of the greater power of the group on his rating. We can now write volumes of psychology books about these relationships, but I told you “bottom bottom”. That is the essence, and everything else is nuance.

***

Question: How far have it comes with “dark matter”?

Answer: Here read[126].

Q: I thought where is the “information theory” in that topic?

A: In the “information theory” that I am developing (exclusively independently), the mass could leave a trace in the past from which it could (weakened) gravitationally act on the present.

Q: Are there similar hypotheses in physics today?

A: Yes and no. Appropriate and similar contributions are kindly sent to me by the editors of the portal where (locally) I upload my texts. Some attachments are sent to me to review and express my opinion.

For example, I liked the recent idea of one author that “dark matter” is a kind of former Maxwell's ether. I don't know if he read my texts, but he could also draw it from “Mach's principle”, that the total mass of the universe affects local inertia and gravity. Mach explained this principle by spilling water from a washbowl that rotates around its axis (it does not matter whether the washbowl or the universe is at rest), and which became part of Einstein's general theory of relativity.

[125] Flock of starlings, https://www.youtube.com/watch?v=rDbGdc7L-qA
[126] What Is Dark Matter? https://www.space.com/20930-dark-matter.html

Another interesting recent contribution (these days) has emerged with the idea that the periodic table also applies to “dark matter”, except that the observable are not “orthogonal” (a common requirement in quantum mechanics) and therefore show different properties from “normal matter”.

***

Question: Why would it be dangerous for us humans to encounter an advanced alien civilization?

Answer: Well-known physicists such as Michio Kaku, or Stephen Hawking draw these conclusions from experience and history. For the Inca civilization, it was a fatal acquaintance with Kotez and the civilization of Europe at that time. Their reasons should not be underestimated, but I have additional ones.

Namely, if this world is all about information, and the information is a measure of uncertainty, then there is no construction or destruction without aggression. The term “aggression” (generalised) is one to which freedoms and restrictions, intelligence and hierarchy, or problem solving and recognition, can be reduced. When we talk about biological species, the domesticated ones will need a smaller brain; they will be less curious, less invasive.

A possible alien civilization that would be far ahead of us in terms of science and technology, if it were far more intelligent than us, would probably be a danger to us.

Q: If it is already scientifically and technologically far ahead of us, how come “if it were more intelligent”, is it a slip?

A: It is not a mistake, because the effects of opposite tendencies are at stake. An increase in comfort is the opposite of an increase in intelligence. Scientific and technological development is encouraged by the “principled minimalism of information”, the desire for security, orderliness and escape from uncertainty. It is basically the aspiration for inaction and inertia. After some level of development (intelligence, science and application), the development of consequences continues but its generator (intelligence) declines. That is one of the most important reasons (the other is self-destruction) because I wrote at the time “that very intelligent species in space could be a rarity if they exist at all”.

Q: At what stage is human civilization?

A: That is difficult to say from the current “frog perspective” of us. Here we are talking about eons of development, or at least thousands of years of history. The sample that Europe had from 1500 to 2000 is insufficient, it is far from reliable. The statistics will be weak in this; its “truths” are otherwise local and deceptive, and they are the only ones to which I should “dare” refer today. The best random hit might be if we were in the “maturity” phase. We have gone through the “madness of youth”, and before us are the “slowdowns of old age“.

Q: A very advanced and very old civilization then could be tame?

A: Yes.

Afterword

And that would be it. If this “information theory” proves to be true, these texts would be ahead of their time, I hope, and all will be worth the effort. Otherwise well again, we won’t cry over the miss.

At the time of my writing, the alleged corona pandemic lasted for a year, and I, like the others in the high school, Banja Luka Gymnasium, did math classes a little “off line” teaching 20 minutes in classes, every about 30 students divided into two groups, and two terms, and a little “on line”. What is separated here is the same as my private and local as in the previous texts of this “theory” and it has little to do with anyone, including other participants in the school, as well as us in the picture.

Photographing the graduates is a tradition in Republika Srpska as well. These are the students of the IV-4 classroom of the general course of the “Gimnazija Banja Luka” with their class teacher and two favorite professors, in April 2021.

Author, May 2021.

Literature

[1] Rastko Vuković: *NOTES TO INFORMATION THEORY* – From October 2020 to January 2021; (Economic Institute, on hold, 20211). https://archive.org/details/notes-to-information-theory/

[2] Rastko Vuković: *INFORMATION STORIES* – of physics, biology, law; (Economic Institute Banja Luka, on hold, 2020). https://archive.org/details/stories_202008

[3] Rastko Vuković: *ACTION of INFORMATION* – energy, time and communication; Economic Institute Banja Luka, 2020. https://www.scribd.com/document/463637180/Action-of-Information

[4] Rastko Vuković: *MINIMALISM of INFORMATION* – physical information and application; Economic Institute Banja Luka, 2020. https://www.scribd.com/document/421999486/Minimalism-of-Information

[5] Rastko Vuković: *PHYSICAL INFORMATION* – in social phenomena and mathematics; Economic Institute Banja Luka, 2019. https://www.scribd.com/document/408301124/Physical-Information

[6] Rastko Vuković: *MULTIPLICITIES* – Choices for animate and inanimate; Economic Institute Banja Luka, summer 1918. https://archive.org/details/Multiplicity_Rastko

[7] Rastko Vuković: *QUANTUM MECHANICS* – the representation of continuum (of metrical and vector spaces); (Economic Institute Banja Luka, on hold, April 2018). https://www.scribd.com/document/377641503/Quantum-Mechanics

[8] Rastko Vuković: *SPACE-TIME*, principles physics of chances; Economic Institute Banja Luka, May 2017. https://www.scribd.com/document/349336253/Principles

[9] Rastko Vuković: *INFORMATION of PERCEPTION* – freedom, democracy and physics; Economic Institute Banja Luka, 2016. https://www.scribd.com/document/360939276/Information-of-Perception

[10] Ferguson, Thomas S. "*Who Solved the Secretary Problem*?" Statistical Science, vol. 4, no. 3, 1989, pp. 282–289. JSTOR. http://www.jstor.org/stable/2245639

[11] Max Tegmark (2007): *The Mathematical Universe*. Dept. of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139. https://arxiv.org/pdf/0704.0646.pdf

[12] M Buser, E Kajari, W P Schleich: *Visualization of the Gödel universe*. Published 30 January 2013. New Journal of Physics 15 (2013) 013063 (32pp). https://iopscience.iop.org/article/10.1088/1367-2630/15/1/013063

[13] A. Aguirre-Astrain, H.N. Núñez-Yépez, L. Salas-Brito: *Geodesic time travel in Gödel's universe* ; Rev. mex. fís. E vol.62 no.2 México dic. 2016; ISSN 1870-3542; http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1870-35422016000200135

[14] Abhas Mitra: *Deriving Friedmann Robertson Walker metric and Hubble's lawfrom gravitational collapse formalism*; Results in Physics,Volume 2, 2012, Pages 45-49, ISSN 2211-3797; https://doi.org/10.1016/j.rinp.2012.04.002

[15] Растко Вуковић: *МАТЕМАТИЧКА ТЕОРИЈА ИНФОРМАЦИЈЕ И КОМУНИКАЦИЈЕ*; Друштво математичара Републике Српске, Бања Лука, 1995.

[16] Rastko Vuković: *QUANTUM MECHANICS* – the representation of continuum (of metrical and vector spaces). Economic Institute Banja Luka, 2019. https://www.scribd.com/document/377641503/Quantum-Mechanics

[17] Reuven Cohen, Shlomo Havlin, and Daniel ben-Avraham: *Structural Properties of Scale-Free Networks*; WILEY-VCH Verlag Berlin GmbH, January 2003. http://havlin.biu.ac.il/PS/handbook446.pdf

[18] Reuven Cohen, Shlomo Havlin: *Scale-Free Networks are Ultrasmall*; arXiv:cond-mat/0205476, 4 February 2003. https://arxiv.org/pdf/cond-mat/0205476.pdf