Large Dimensions: Mathematics and Applications.(Unfinished and Unedited)

Misha Gromov

December 14, 2018

This is a guide to my lectures in the Fall 2018 at CIMS.

"Dimension" brings to mind the idea of "space", but the word "large", doesn't belong with mathematics. It isn't our business to know *how many grains make a heap.*

Yet, we ask:

Which *spaces* are qualified as *large dimensional*?

To get an idea, look at:

A. Finite dimensional Euclidean spaces $\mathbb{R}^N$.

B. Finite combinatorial power spaces, such as $\{\bullet.\circ\}^N$.

Motivated by the Life on Earth examples (see below), we declare that "large" begins somewhere around $N = 100$ and may go up to 10^{10}-10^{12}, maybe even up to 10^{15}, while dimensions N which are well above 10^{15}, say starting with 10^{18}, are too close to infinity from our perspective to be called "finite".[1]

Then the second question arises:

What is so special about these spaces, what are properties characteristic for these N?

A pronounced feature of N-dimensional power spaces X with large N is

tendency of functions on X to be nearly constant..

For instance,

the angle (function) between two random unit vectors in $\mathbb{R}^N$ is almost constant: close to $90°$ with overwhelming probability for large N.

This is most conspicuous for very large N, greater than 10^{12}, and if you wish $> 10^{18}$, which is common in statistical mechanics, where propensity of functions (macroscopic observables) to be constant is articulated as the identity

average = typical,

[1]Live systems, unlike physical ones, are numberphobic. We shall see later on that virtually all meaningful entities N, be they large or small, enter the scene of Life in non-numerical gowns.

In fact, logic of Life, brings into question the common convention of X^N defined as the set of "strings" $x_1, ...x_i...x_N$, which depends on an ad hoc representation of *numbers* N by *sets* $\{1, ..., N\}$.

which means that the values of functions on these spaces tend to concentrate near their averages.[2]

This concentration of measure phenomenon is ubiquitous in mathematics and in mathematical physics, we return to it in section???, but it is not always there. For instance, in lotteries of Life, where the winners take all, exactly the opposite happens:

instead of concentration of functions f in the domains of their values*, one sees Dirac like δ-concentrations of f in their* domains of definition*. This makes* averages *of such functions f as* far from typical *as it is conceivably possible.*

However, none of these two concentrations is visible, at least not directly, in functions, often called *features* of (almost) anything related to Life. These features come ablaze in billions colors, nothing you can call "constant" or "approximately constant".

And albeit the domains X of definitions and domains Y of values of these functions/features f are associated with power spaces, these X and (not always) Y have intricate internal structures and relevant $f : X \to Y$ are far from being plain and simple.

But what are these spaces? What are functions on these spaces – features of their members?

How about this?

1. The space of all live things on Earth.

But is it a space in any sense? Does it possess any geometry? Can one attach a number to this "space" and justifiably call it *"dimension"*?

A possible (but not the only) way to bring this "space" to the dominion of Math is to view it as a (randomised?) *quotient space* of another space.

2. The space of genomes of individual organisms on Earth,

where the latter can be regarded as

$2_{\#}$ a subset of the set of the space of finite strings in four symbols.[3]

This is still far from the *true definition* (if such a definition exists at all), where the main reason for this is a difficulty with a *proper* mathematical interpretation of "a" in the above "a quotient" and "a set".

But regardless of what this "proper" is – this we shall discuss later – one can safely say that the dimension of the "space" of (significantly different individual) organisms on Earth is in the range 10^4 - 10^9.[4]

The difficulty faced by a mathematician in studying, or even in defining "spaces" of organisms and genomes, besides their size and complexity, is *their accidentally*: these "spaces" are come as end-points of a single, possibly non-representative, branch of a grand random process: biological evolution.

---

[2]Without this the statistics would be inapplicable in physics, since *average* is what is amenable to a mathematical evaluation and *typical* is what is observed in an experiment.

[3]Customary, these are A, C, G, and T for the nucleotides: Adenine, Cytosine, Guanine and Thymine making DNA.

[4]The length of genomes of certain viruses goes below 10^3, humans have almost $3 \cdot 10^9$-long genomes and the genomes of some amoeba like creatures may reach close to 10^{11}. But most of "information" carried by genomes, especially by the long ones, is, apparently, erased by the "quotient map" from genomes to organisms. Probably, every organism (class of organisms?) can be identified and adequately described – modulo stochastic variations – by 10^4-10^6 (not necessarily numerical) parameters – the (mainly physiological) features of this organism.

Thus, a mathematician should either turn to the larger "space" of *all conceivable*[5] organisms and/or genomes or, on the contrary, to focus at representative fragments of these spaces.

The most studied such fragment, which is located near the boundary of Life with the physical world, is

3. The space of proteins,

This, similarly to the "space of organisms", can be seen as a quotient of a larger but more accessible space:

4. The space of polypeptide chains,

that is a sequence space in 20 letters:

polypeptides which make proteins are (hetero)polymeric chains of length N, roughly, between 30 and 30 000 [6] *composed of 20 (sometimes 21) basic amino acids,*[7]

The arrow

polypeptides → proteins[8]

is physically implemented by the process of *protein folding*, which takes place in the *polypeptide configuration spaces.*

Mathematically, the configuration space C_P for a polypeptide P, is a domain in the torus of dimension $(2+\sigma)N$, where N is the length of P and σ is, roughly, the average number of the *side chains* in the amino acids in P.

Folding of a polypeptide to a protein in a water environment can be modelled by a randomised gradient descent for the energy function E_P on C_P defined by the mutual physical/chemical interactions between the residues in P as well as their interaction with the water molecules.

Albeit the principles of the protein folding (essentially, the shape of the energy landscape in C_P), unlike how it is with the arrow genomes *to* organisms, are, at least in general terms, understood, the protein folding problem in most respects remains unresolved.

But the true biological problem, which is more subtle and more interesting than the (essentially physical/mathematical) folding problem, concerns not individual spaces (C_P, E_P), but their *totality* parametrised by the space $\mathcal{P} \ni P$ of the polypeptide (sequences) P, where the present day $\mathcal{P}_{now}$ can be seen as a set of quasistationary points of the evolutionary dynamics acting on $\mathcal{P}$, which, up to some extent, may be represented by a *protein fitness landscape* in $\mathcal{P}$.[9] (We

---

[5] "Conceivable" and "mathematically expressible" are are synonimous for a mathematician.

[6] Short polypeptide chains, even if they serve some functions in cells, are, somewhat arbitrarily, called *peptides.*

[7] In the course of polymerisation – synthesis of polypeptides – amino acids are slightly curtailed; what remains of them in polypeptides are called *amino acid residues.*

[8] There is also an opposite arrow as well, proteins → polypeptides, since proteins "remember" the order of amino acids in them: strictly speaking the protein space is bigger than that of polypeptides. But a working protein is as little aware of this order as an organism of the order of nucleotides in its chromosomal DNA.

[9] According to the orthodox Darwinism, evolution is adequately described by the fitness function defined as the *relative reproduction rate*, similarly to how a physical system is run by a single energy function. But when you look at this "relative" with an open mind ready to accept the ubiquitous numberphobicity of evolutionary biology, you realise that what come out of "rate" is not a mere number but an elaborate structural entity and you will see a new mathematical picture of the evolutionary landscape much richer in colors than what is offered by the model(s) of the (neo)classical Darwinism.

shall return to this in section ???.)

"Space" is an attractive concept, but does it apply to all Life's children.

Are, for instance :

5. the space of states of the mind

and

6. the space of states of the brain

which are so close and dear to us, true spaces? There is no simple answer (some non-answers are given in section???) but there are several bona fide spaces which contains traces of the above which we shall discuss in detail later. Among these are:

7. memory spaces, including *Kanerva model* as an example.

and

8. weight spaces of neural networks with composed functions on them.

Also Life has several beautiful grandchildren spaces, such as

9. spaces of natural languages and sentences in these

and

10. space of mathematical ideas,

where even the tiny fragments of the latter:

11. spaces of chess positions and chess games,

hide more charming surprises (we shall demonstrate them later), than a traditional mathematical picture of these spaces shows.

We do not expect close similarities between these spaces but we want to develop a mathematical language applicable to all of them.

That would help in overcoming the most serious difficulty in approaching unknown: our disability to ask questions.

A good language would allow *articulations of questions* about B inspired by certain knowledge about A.

Contents

1 Probability, Symmetry, Concentration.

1.1 The Law of Large Numbers, Pythagorean $\sqrt{N}$-Rule and Exponential Tail Bounds.

The fist (?) recorded instance of what is now called "measure concentration" was the *Law of Large Numbers* stated by Cardano (1501 – 1576) in qualitative form:

the typical value of the sum of many independent random variables is (relatively) close to the sum of expectations of these variables,

which was was proved by Jacob Bernoulli almost two hundred years later for independent equidistributed $(0,1)$ variables. [10]

In this case the sum of Σ_N of these takes the values $0,1,2,..,N$, where the probability p_i of $\Sigma_N = i$, for $i = 0,1,...N$, equals the ith binomial coefficient[11] normalises by 2^{-N} in order to have $p_0 + p_1 + ... + p_N = 1$. Thus, p_i are defined by

$$(1+x)^N = 2^N(p_0 + p_1 x + ... + p_N x^N), \text{ i.e. } p_i = 2^{-N}\binom{N}{i}.$$

They say, it took 20 years for Bernoulli to prove his theorem, but today it is effortlessly derived from the (Hilbertian) *Pythagorean theorem* as follows.

Let $\mathcal{F}$ be the (Hilbert) space of functions f on a measure space $\mathcal{M}$ (e.g. a finite set of atoms with unit weights), where the distance is defined by the formula

$$dist^2(f,g) = \int_{\mathcal{M}} (f-g)^2 d\mu.$$

Let $f_1, f_2, ...f_N$ be *mutually orthogonal* (i.e. $\int_{\mathcal{M}}(f_i \cdot f_j) = 0$) functions.

[10] I haven't check what and how was originally proven by Bernoulli.

[11] However simple, this representation of the numbers p_i, which are the values of the N-th *convolution power* of the dyadic measure on the set of integers, by the coefficients of the *algebraic power* $(1+x)^N$ is an instance of the *Fourier transform* that, in the present case, establishes an isomorphism between the group algebra of $\mathbb{Z}$ and the algebra of Laurent polynomials.

Then *the norm (i.e the distance from the zero function) of the sum of these functions satisfies*

$$||f_1 + f_2 + ... + f_N||^2 = ||f_1||^2 + ||f_2||^2 + ... + ||f_N||^2.$$

(It is hard to appreciate the greatness of this formula – *familiarity breeds contempt.* Yet, try to rationally explain why nothing of the kind holds true if the exponent "2" is replaced by any other number.[12])

Back to Bernoulli, think of random variables as functions f on a probability space $\mathcal{M}$.

This may (justifiably) strike you as artificial but there is a distinguished space $\mathcal{M}$ where the action takes place, namely the set $\{0,1\}^N \subset \mathbb{R}^N$ with the 2^{-N} weights assigned to all points in this set, where the ith random variable is given by the projection of $\{0,1\}^N$ to the ith Euclidean coordinate.

These, of course, are non-orthogonal, but we can render them such by subtracting the constant functions equal 1/2 from all of them and, to save notation, we multiply each of them by 2.

Now the resulting, (still independent!) variables, call them f_i, $i = 1, 2, ..., N$, take values -1 and 1 and, since they are independent and have zero means, they are orthogonal. (This can be seen directly for our f_i.) Therefore, by the Pythagorean theorem the norm of their average

$$A_N = \frac{1}{N}(f_1 + ... + f_N)$$

satisfies

$$||A_N|| = \frac{||f_1 + ... + f_N||}{N} = \sqrt{\frac{||f_1||^2 + ... + ||f_N||^2}{N^2}} = \frac{1}{\sqrt{N}},$$

since

$$||f_1| = ||f_2|| = ... = ||f_N|| = 1.$$

Now, obviously, if a random variable A, seen (naturally or or unnaturally) as a function on a probability space $\mathcal{M}$, have small norm it must be small on the most part of $\mathcal{M}$.

Indeed if A were $> c$ on a subset in $\mathcal{M}$ of measure $> \varepsilon$ then its norm would satisfy

Markov (Chebyshev-Bienaymé) Inequality (in reverse).

$$||A|| > c\sqrt{\varepsilon}.$$

Thus,

the probability of A_N being $> c$ is bounded by by

Bernoulli Inequality

$$\mathbf{Prob}\{|A_N| > c\} < \frac{1}{Nc^2}.$$

[12]A physicist would say this is so because the Nature has chosen the exponent "2" to relate energy to velocity, but a mathematician would maintain that this choice was forced on Nature by the Pythagorean theorem.

Exponential Tail Bound. There is a sharper bound on this $A_N = \frac{1}{N}\sum_i f_i$,

$$\mathbf{Prob}\{|A_N| > c\} < 2\exp -\frac{Nc^2}{2}.$$

Proof. [Abraham de Moivre 1733] Rewrite this inequality as the following bound on the sum of the first k binomial coefficients $\binom{N}{i} = \frac{N!}{i!(N-i)!}$.

$$2^{-N}\sum_{i=0}^{k}\binom{N}{i} \leq \exp -2\alpha^2 N$$

for all $k \leq N/2$ and

$$\alpha = \frac{1}{2} - \frac{k}{N},$$

and evaluate binomial coefficients with *Stirling's approximation formula*[13]: $n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$.

Remarks.(a) It is instructive to think of the numbers $p_i = 2^{-N}\binom{N}{i}$ as probability weights of the $N+1$ atoms of the quotient space of the binary power set $\{0,1\}^N$ divided by the permutation group Π_N. In fact, other interesting probability spaces come this way, when a (homogeneous or non-homogeneous) space is divided by a group of its automorphisms. Examples of these are the spaces of unitary matrices with the Haar measures divided by conjugation and *moment maps* resulting from factorisation $\frac{symplectic\ manifolds}{their\ symmetry\ groups}$.

(b) Bernoulli's inequality say in geometric terms that
the majority of vertices of a high dimensional N-cube

$$\square^N = [0,1]^N$$

is located near the hyperplane H_0 and which passes through the center of the cube,

$$\underbrace{\left(\frac{1}{2}, ..., \frac{1}{2}\right)}_{N} \in [0,1]^N,$$

and which is normal to the principal diagonal, that is the line between the opposite vertices

$$\underbrace{(0,...,0)}_{N} \text{ and } \underbrace{(1,...,1)}_{N}$$

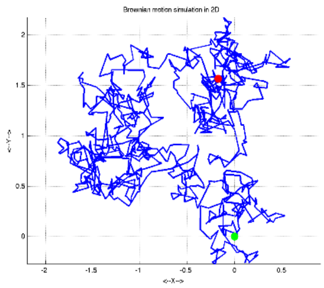
in the cube.

In fact, by the exponential tail bound, the percentage of points that lie within the Euclidean distance $> c\sqrt{N}$ from H_0 *exponentially* decays for all $c > 0$ and $N \to \infty$

(c) On Random Walk and Random Spread. Bernoulli's Pythagorean $\sqrt{N}$ is most vividly demonstrated by *random walks*, e.g in the integer lattice $\mathbb{Z}^k \subset \mathbb{R}^k$, where a typical path of length N has the (Euclidean) diameter about $\sqrt{N}$.[14]

[13] Stirling in his 1763 paper proved a sharper formula.

[14] Accordingly, *Fourier's law* implies that the time needed for a noticeable amount of heat to propagate through a wall of thickness l is proportional to l^2. (What *is* proportional to l is the rate of the steady heat flow, for which you have to wait time $t \sim \kappa l^2$.) This is what helps the Earth core to keep its heat.



This suggests that a *typical configuration* of a molecular chain C_N of N monomers, e.g. a polypeptide chain freely floating in (hot) water, must have the diameter about N^α for $\frac{1}{2}+\varepsilon_1 \leq \alpha \leq 1-\varepsilon_2$.

The rational behind this is that C_N looks like a path of a *self avoiding* random walk in $\mathbb{Z}^3$, where the self-avoidance (or self-repulsion) condition[15] would make C_N spread/stretch on the average more than for the unrestricted random walk – conjecturally $\varepsilon_1 > 0$ – yet, not enough to make C_N virtually straight, i.e. ε_2 should be > 0.[16]

Amazingly, none of the two inequalities $\varepsilon_{1,2} > 0$ has been proved,[17] and neither one knows (this seems easier) if the typical diameter of C_N is equal to the average one.[18]

(d) On General Random Variables. The Pythagorean proof of the law of large numbers applies to the sums of independent random variables $f_i(x)$ with *bounded first and second moments*, that are the integrals $\int f_i d\mu$ and $\int f_i^2 d\mu$. In fact, what one needs for this is mere orthogonality of f_i.

The exponential tail bound also extends to general independent random variables f_i under the name Hoeffding's inequality, where one needs f_i themselves be bounded, e.g. i take values in the interval $[-1,1]$. Then average of f_i

$$A_N = \frac{1}{N}\sum_{i=1}^{N} f_i$$

exponentially sharply concentrates near its expectation $\mathbf{E}(A_N)$,

$$\mathbf{Prob}\{|A_N - \mathbf{E}(A_N)| \geq c\} \leq 2\exp-\frac{Nc^2}{2}.$$

where the proof, that *fully* relies on the independence of f_i, proceeds, roughly, as follows.

[15]This condition says, in effect, that the relevant probability measure is *supported on the set of all non-self-intersecting paths* in the lattice graph of $\mathbb{Z}^3$, and all these paths are assigned *equal weights*.

[16]This is expected for the self avoiding random walk in $\mathbb{Z}^2$ and in $\mathbb{Z}^3$, while the higher dimensional random walks are oblivious of the self-avoidance condition.

[17]*Lectures on Self-Avoiding Walks* by Roland Bauerschmidt, Hugo Duminil-Copin, Jesse Goodman, and Gordon Slade,
`https://www.ihes.fr/~duminil/publi/saw_lecture_notes.pdf`

[18]Conceivably (but unlikely) 49% of chains have diameters approximately $\sqrt{N}$ and another 49% are close to N.

Since f_i are independent, the functions $\lambda \exp f_i$ are also independent for all λ, and the multiplicativity property of *expectations* applies:

$$\mathbf{E}\left(\exp \lambda \sum_i f_i\right) = \mathbf{E}\left(\prod_i (\lambda \exp f_i)\right) = \prod_i \mathbf{E}(\lambda \exp f_i).$$

This, in conjunction with the (obvious) Markov inequality yields the proof. (Do it yourself or consult [High Dimensional Probability].)

(e) On Euclidean Cube $\blacksquare^N = [-1,1]^N$. The geometric interpretation of the above for the random variables uniformly distributed on the segment $[-1,1]$, shows that if $d >> 1$ then for all N,

almost all Euclidean volume/measure of the cube lies within distance d from the hyperplane normal to the principal diagonal of the cube, that is $H_0 = \{x_i\}_{\sum_i x_i = 0}$.

Here, it is worth mentioning that

All vertices of the cube $[-1,1]^N$ *lies within distance* $\sqrt{N}$ *from the centre of the cube.*

However obvious it worthwhile observing that it is *exactly* the same Pythagorean $\sqrt{N}$ which underlines the law of large numbers and the average displacements of random walks.

Concentrated? Yes, but where?

There are 2^{n-1} hyperplanes in the cube $[0,1]^N \subset \mathbb{R}^N$ geometrically indistinguishable from H_0: one hyperplane H for each pair of opposite vertices. The band ow width $\varepsilon\sqrt{N}$ around every of one of these H contains most of cube volume, if $N >> \varepsilon^{-1}$.

It follows, that the intersections of such bands around *several* H, if this "several" is significantly smaller than 2^N, contain most of the cube. But if $\varepsilon < e^{-1}$, $e = 2.718...$, then the intersection of *all* these bands in the cube around *all* these H, carries only δ^N-fraction of the volume of the whole cube $[0,1]^N$ for δ only negligibly greater than εe.

In fact this intersection is contained in the "diamond" of "radius" $\varepsilon N/2$ around the center of the cube, that is isometric to the standard $\blacklozenge^N\left(\frac{\varepsilon N}{2}\right) \subset \mathbb{R}^N$, defined by

$$\blacklozenge^N\left(\frac{\varepsilon N}{2}\right) = \{x_1, ...x_i, ...x_N\}_{\sum_i |x_i| \le \frac{\varepsilon N}{2}},$$

the volume of which is $\frac{(\varepsilon N)^N}{N!} \approx (\varepsilon e)^N$.[19] (Isn't it amazing that the diamond of radius $N/6$, which looks pretty large, say for $N > 600$, has negligibly smaller volume than that of the unit cube?)

But despite probability shouting in you ear that:

THE MEASURE OF THE N-CUBE IS STUCK TO ITS BOUNDARY

you can't dismiss the center of the cube as something non-essential.

What one probabilistically perceives of a high dimensional object is sensitive to the position of the observer.

---

[19] Since only a small part of our "diamond" is contained in the cube, the volume of this part must be significantly (?) smaller than $\frac{(\varepsilon N)^N}{N!} \approx (\varepsilon e)^N$.

Exercise. Figure out how much of the mass of the unit N-dimensional cube $[0,1]^N$ is concentrated near its k-dimensional faces. That is, evaluate the volumes of the ρ-neighbourhoods $U^k(\rho) \subset [0,1]^N$ of the unions of the k-faces in $[0,1]^N$. (Relevant ρ are $\varepsilon\sqrt{N}$)

Hint. It is instructive to start with such an evaluation for the cubical lattice, where is also useful to look at the dual lattice and where the computation simplifies if instead of the Euclidean/Pythagorean norm $||x|| = \sqrt{\sum_i x_i^2}$ one takes $\sup_i |x_i|$ for the norm.

1.2 Hamming Geometry and Kanerva Memory

The Hamming distance between two elements in a product space[20]

$$X = \bigtimes_{i \in I} F_i,$$

such as the power spaces F^I, e.g. for $F = \{0,1\}$, equals, by definition,

the number of those i in the set I, where

$$a_i \neq b_i.$$

For example , if I is a one point set and X has no nontrivial product structure, then

$$dist_{Ham}(x,y) = 1 \text{ whenever } x \neq y.$$

Exercise. Check the triangle inequality for the Hamming distance, observe that the Hamming diameter of $X = F^I$ equals the cardinality of I,

$$\sup_{x,y \in X} dist_{Ham}(x,y) = |I|,$$

and show that every two points $x, y \in X = F^I$ with $dist_{Ham}(x,y) = d$ can be joined by a chain of immediate neighbour points $z_k \in X$,

$$x = z_1, ...z_k, ...z_d = y, \ dist_{Ham}(z_k, z_{k+1}) = 1.$$

Hamming Concentration. The most essential feature of this metic is the *concentration of the Hamming distance function* which follows from the Bernoulli (exponential tail) inequality which, in terms of $dist_{Ham}$, says that

for majority of the pairs $(x,y) \in X \times X$, the distance $dist_{Ham}(x,y)$ is close to the mean distance that is half cardinality of the set I.

For instance, the distances of 98% of "$(0,1)$-strings" $x \in \{0,1\}^{1000}$ from a given $x_0 \in X$ are in the range:

$$dist_{Ham}(x, x_0) = 500 \pm 37 \approx 500 \pm 0.135 \cdot 500,$$

where there are only 70 points out of 1000 within this distance from x_0.[21]

[20]The Hamming metric is most commonly (but not exclusively) used for binary spaces F^I, where F is a two element set and where $dist_{Ham}$ reasonably well *quantifies the concepts of similarity* between "binary strings" $\{f_i\} \in X$.

[21]Here, $A = B \pm C$ means $|A - B| \leq C$.

Hamming Connectivity. The balls of radii 250 contain only negligible proportions of X, namely less than $\exp -125$ <0.000 000 000 000 000 000 000 000 000 000 000 1 of all 2^{1000} points in X.

Yet all point x in $X = \{0,1\}^{1000}$ can be reached from any given point x_0 in four 250-long steps

$$x_0 \underset{250}{\longleftrightarrow} x_1 \underset{250}{\longleftrightarrow} x_2 \underset{250}{\longleftrightarrow} x_3 \underset{250}{\longleftrightarrow} x_4 = x.$$

Exercise. Show that the there are more than 2^{1500} and less than 2^{2000} triples (x_1, x_2, x_3) that can serve in such chains between given x_0 and x in X.

Imagine, conceivable (potential?) *memory items* x being described by their features ϕ the list Φ of which is known to you beforehand.[22] (This is unrealistic, but let it go.)

This means, our x are represented by $\{yes, no\}$-valued functions on a (fixed) set Φ and the set X of all conceivable memory items is equated with the binary $\{yes, no\}$-space,

$$X = \{yes, no\}^{\Phi}.$$

We denote the value of such a function x at ϕ by

$$\phi \mapsto x \,\square\, \phi$$

that is "yes" if x *has feature* "ϕ and "no" otherwise.

Dually, one may regard ϕ being a function (*observable* in physicist's parlance) of x, write

$$x \mapsto \phi \,\square\, x$$

instead of $\phi \mapsto x \,\square\, \phi$, and following the rules of the common language, read this as *the feature ϕ is present in/absent from x.*

Then, the reason for this will become clear below, we represent *yes* by +1 and *no* by −1 and, often, call our functions $\phi \mapsto x \,\square\, \phi$ *strings* or ±1-*strings*, despite the fact that there is no natural order in Φ. This turns $x \,\square\, \phi$ look to a kind of a numerical scalar product.[23]

The realistic number N of features in Φ may range, according to Kanerva, from 100 to 10 000 which make the cardinality $|X|$ of the space X quote large , more than 10^{30} already for $N = 100$ and truly enormous, $> 10^{100}$, for $N > 330$.

No realistic memory is large enough to encode *all these items*, but we don't need it anyway.

All we want is to be able to encode *any single item* and then to continue encoding up to, say 10^9 of these, which is quite satisfactory, at least from a human point of view:

if you record one item each second 8 hours a day every day, such a memory will suffice for more than 90 years.[24]

[22] We write "Φ" instead of "I" not to be tempted to think of this set Φ as $\{1, 2, ..., |\Phi|\}$.

[23] From a neurobiological point of view taken by Kanetva, the features ϕ correspond to "hard physical units" e.g. neurones and/or synapses in the brain, while memory items x are recorded by variable states of these units. But psychologically speaking, features are parts of the dynamical memory, very much the same as our x. In this picture, "$\square$" serves as a *coupling* rather than *function evaluation* sign, where such a coupling is a (higher order?) memory item in its own right.

[24] (Approximately) 365.25 days of the full turn of Earth around the sun make, 31 557 6 00 seconds.

And a reliably registering of this amount of information needs *only* a few billion memory locations.

KMM. Below is a (slightly mathematised) description of a memory model suggested by Kanerva , called $\mathcal{KMM}$, that, despite its shortcomings, displays certain features of the human memory (we discuss pros and cons in section ???)

Registers, Counters and their Contents. Let the "hardware" of $\mathcal{KMM}$ be represented by a set R of *registers*, say of cardinality $|R| = 10^9$, where each register r consists of a *set of counters* and where each counter corresponds to a feature $\phi \in \Phi$.

According to this correspondence, the counters from an $r \in R$ are denoted ϕ_r and the set of these by Φ_r.

The *memory content* of each counter is an integer $m(\phi_r)$. Thus, the full memory kept in all registers is an integer valued function m on the product set,

$$m : \Phi \times R \to \mathbb{Z},$$

where we agree that "no recorded memory" is represented by zeros in the corresponding counters.

These numerical functions m, unlike R and Φ_r, are modified when new items enter the memory according to the rules described below.

Vicinity Structure of $\mathcal{KMM}$. The main architectural attribute, which allows recording memory items from X in R, is a subset $D \subset X \times R$, where the inclusion

$$(x, r) \in D$$

reads as "x and r are *D-neighbours*", or "in *D-vicinity*, of each other".[25]

Memory Recording in $\mathcal{KMM}$. Whenever a new item x enters the memory, the numbers $x \,\square\, \phi$, $\phi \in \Phi$, are added to the contents of the counters $\phi_r \in \Phi_r = \Phi$ for all registers r in the D-vicinity of x.

For instance, if, originally, all registers were set on zero, then x is recorded, exactly as it is, in all D-neighbour registers r of x.

However, as we add more and more memory items, the sets of D-neighbours of different x may start overlapping and some registers will contain *sums of several* ± 1*-strings*.

Reading from Memory in $\mathcal{KMM}$. We want to decide if our memory have earlier recorded an item x, or, it contains an item similar to x.

For this we introduce a *cut-off operation* $\sigma(\phi) \mapsto \bar{\sigma}(\phi)$ on functions σ on Φ, such that the result of this cut-off is an item $x \in X$ regarded as a ± 1-function. We agree (this is negotiable) that

$\bar{\sigma}(\phi) = +1$ if $\sigma(\phi) > 0$,
$\bar{\sigma}(\phi) = -1$ if $\sigma(\phi) < 0$,
$\bar{\sigma}(\phi) = \pm 1$ randomly with probabilities $1/2$, if $\sigma(\phi) = 0$.

Granted such a cut-off, let is construct the following *(memory) search transformation* on the set X,

$$S : X \to X,$$

where $S(x)$ is defined in two steps.

[25] Think of "D-vicinity" as shorthand for "distance between x and r is less than D".

1. Add the contents $m_r = m(\phi_r), \phi_r \in \Phi_r = \Phi$, of the counters ϕ_r for those $r \in R$ which lie in the D-vicinity, say $V_x \subset R$, of x; this makes sense, since all sets Φ_r of counters in all registers r are identified with Φ.

2. Regard the resulting sum as a function on Φ,

$$\sigma = \sigma(\phi) = \sum_{r \in V_x} m_r$$

and let

$$S(x) \square \phi = \bar{\sigma}(\phi).$$

If $S(x) = x$ we conclude that x was introduced to the memory at some point. More generally, we regard all

fixed points of the iterated maps

$$S^{\circ k} = \underbrace{S \circ \ldots \circ S : X \to X}_{k}$$

for moderate k, say for $k \leq 5$, as the items recorded by the memory or at least as approximations to the actually recorded items.

The essential features of $\mathcal{KMM}$, which are motivated by properties of the human memory, are:

A. $\mathcal{KMM}$ is *distributive*: the information encoding an individual item is contained in *several* registers of $\mathcal{KMM}$.

B. $\mathcal{KMM}$ is, up to certain extent, *dynamic*: reading from $\mathcal{KMM}$ relies on iterations of transformations on the set of (possible) memory items.

These suggest further study and mathematical development of similar classes of memory models (see section???), where, in order to have a psychological plausibility, the closeness of the memory records m_1 and m_2 of x_1 and x_2 must match the closeness between x_1 and x_2 themselves

In the present $\mathcal{KMM}$ case, where in $x_1, x_2 \in X = \{-1, 1\}^\Phi$, the latter "closeness" refers to the Hamming metric and the former one is defined via the l_1-*metric* on the space of (integer valued) functions m on $\Phi \times R$. Namely if m_1 and m_2 are records of x_1 and x_2 with the same initial state $m_0 : \Phi \times R$, e.g. $m_0 = 0$, then we want to have an approximate equality

$$\sum_{\phi \in \Phi, r \in R} |m_1(\phi, r) - m_2(\phi, r)| \approx \Theta\left(dist_{Ham}(x_1, x_2)\right)$$

for a suitable (positive, bounded, monotone increasing and linear for small d) function $\Theta = \Theta(d)$, $d \geq 0$,

(The Hamming distance for binary (-1,1)-strings also admits an l_1-description:

$$dist_{Ham}(x_1, x_2) = \frac{1}{2} \sum_{\Phi} |x_1 \square \phi - x_1 \square \phi|)$$

To achieve this, we need an *adequate* vicinity structure $D \subset X \times R$, which is convenient to regard here as a *set valued map* from $X \to R$, denoted

$$\underset{\to}{D}(x) \subset R,$$

and defined by

$$r \in \underset{\rightarrow}{D}(x) \Leftrightarrow (x,r) \in D.$$

The "adequacy" of D can be now expressed in how the map $\underset{\rightarrow}{D}$ from the binary space $X = \{-1,1\}^{\Phi}$ to the, also *binary*, space of subsets in R, denoted $\mathbf{2}^R$, behaves with respects to the Hamming metrics in these two spaces. Namely, what we want of this map

$$\underset{\rightarrow}{D} : X = \{-1,1\}^{\Phi} \to \mathbf{2}^R,$$

is some approximate relation

$$dist_{Ham}(D(x_1), D(x_2)) \approx \underset{\rightarrow}{\Theta}(dist_{Ham}(x_1,x_2))$$

for some function $\underset{\rightarrow}{\Theta}(d)$ similar in its properties to the above $\Theta(d)$.

Kanerva's suggestion for such a D, hence, for $\underset{\rightarrow}{D}$, is as follows.

Implement R by a subset in X by a 1-to-1 map $R \hookrightarrow X$, which is possible since R, e.g. of cardinality $\approx 10^9$, is much smaller than $X = \{-1,1\}\Phi$ for $|\Phi| \geq 100$, and define $D \subset X \times R$, where now R is regarded as a subset $R \subset X$, by the condition

$$(x,r) \in D \Leftrightarrow dist_{Ham}(x,r) \leq \Delta$$

where $\Delta > 0$ is chosen, such that the expected numbers of R-points in the Hamming balls of radii Δ around (almost) all points $x \in X$ are close to something like 100,[26]

$$|R \cap B_x(\Delta)| \approx 100$$

for

$$B_x(\Delta) = \{y \in X\}_{dist_{Ham}(y,x) \leq \Delta} \subset X$$

Notice that the corresponding map

$$\underset{\rightarrow}{D} : X = \{-1,1\}^{\Phi} \to \mathbf{2}^R$$

for such a $D = D(\Delta)$ and $|\Delta| \leq 100$, lands in the subspace $\mathbf{2}^R_{\leq 100} \subset \mathbf{2}^R$ of subsets in R of cardinalities ≤ 100, and that the cardinality of this $\mathbf{2}^R_{\leq 100}$, say for $|R| = 10^9$, is not very far from $10^{900} = 1\,000\,000\,000^{100}$,

$$10^{900} > |\mathbf{2}^R_{\leq 100}| > 10^{800}.$$

It follows, for example, that if $\Phi = 10\,000$, then the map $\underset{\rightarrow}{D}$ *can't be* 1-to-1, since

$$|X| = 2^{10\,000} > 10^{3\,000} > 10^{900}$$

But if $|\Phi| = 1\,000$, then

$$|X| = 2^{1\,000} << 10^{800}$$

and, in principle, $\underset{\rightarrow}{D} = D(\Delta = 100)$ can (and likely to) be 1-to-1 for many embedding $R \hookrightarrow X$, but I have not check if this is indeed is so.

[26]Models of this kind always contains quite a few parameters that must be adjusted according to to what you expect of such a model.

To conclude the construction of D, we need to indicate a specific embedding $R \hookrightarrow X$ or a class of such embeddings. What Kanerva suggest is taking a *random subset* in X of cardinality $|R|$ for $R \subset X$. The essential attribute of this, guaranteed by the (Bernoulli) Hamming geometry is a relative uniform distribution if the R-points in X, e. g. concentration of the cardinality function

$$x \mapsto |R \cap B_x(\Delta)|$$

near its mean value[27]

This concentration and also connectivity properties of the Hamming spaces, bring along certain properties of $\mathcal{KMM}$ close to these of the human memory as it is explained in [Sparse Distributed Memory].

Another source of justification of "random" is a compelling biological reason – poverty of information encoded in the Human genome – to believe that, albeit the overall organisation of the human brain is genetically (pre)determined, the details (of the embryonic and post-embryonic development of the brain) are (almost) 100% random.[28] But this randomness – we shall explain in section ???– is not so straightforward as in the Kanerva memory model.

1.3 Balls, Spheres, Gaussian Measures and Maxwell Distribution

The concentration properties of N-Dimensional Euclidean balls and spheres are similar to those of cubes, where "roundness" of balls renders the pictures more transparent and the proofs easier.

$[\bullet^N_\varepsilon]$ The "law of large numbers" for balls. *The Lebesgue measure of the unit $B^N = B^N(1) \subset \mathbb{R}^N$ is concentrated near the equatorial subball*

$$B^{N-1} = \{x_1, ...x_N\}_{x_1=0} \subset B^N,$$

where "near" means in the band defined by the inequality $|x_1| \leq \varepsilon$.

More precisely the percentage of the volume of B^N in the complement to this band than consist of two $(1-\varepsilon)$-thick "spherical caps", defined by the inequalities $x_1 \leq -\varepsilon$ and $x_1 \geq \varepsilon$ and denoted $Cap^N_{\pm\varepsilon} \subset B^N$, satisfies

$$\frac{vol(Cap^N_{\pm\varepsilon})}{vol(B^N)} \underset{N\to\infty}{\to} 0 \text{ for all (fixed) } \varepsilon > 0.^{29}$$

This can be clearly seen by looking at the *normalised push-forward* [30] $\underline{\nu}_N$ of the Lebesgue measure of the ball under the projection of the ball to the first coordinate line,

$$B^N \to \mathbb{R}, \text{ for } (x_1, x_2, ..., x_N) \mapsto x_1,$$

[27] Deterministic as well as stochastic maps like $R \hookrightarrow X = \{-1,1\}^N$ and $\underset{\to}{D} : X = \{-1,1\}^\Phi \to \mathbf{2}^R$ have been extensively studied under the headings of *error correction codes*, see, e.g.
http://www.cs.yale.edu/homes/spielman/561/2009/lect11-09.pdf

[28] This also applies to *large* animal brains but not to small ones. For instance, for all we know, all *Caenorhabditis elegans* worms have identically wired nervous systems of about 300 neurones.

[29] We shall see below that this convergence is *exponentially fast.*

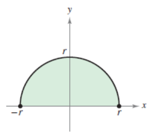
[30] The *push-forward*, sometimes called *projection* of a measure μ on X under a (continuous) map from X to Y is a measure ν on Y, such that $\nu(U) = \mu(f^{-1}(U))$ for all open subsets $U \subset Y$, where $f^{-1}(U) \subset X$ is the f-pullback of U.

The normalisation of a measure μ on a space X with finite total mass $m = \mu(X)$ is the *probability measure* $m^{-1} \cdot \mu$.

where $B^{n-1} \subset B^N$ equals the pullback of zero under this projection and

$$\underline{\nu}_N = m^{-1}\left(\sqrt{1-x_1^2}\right)^{N-1} dx_1 \text{ for } m = vol(B^N).$$

Since the relative volume[31] of the ε-band around $B^{n-1} \subset B^N$ is equal (by the



definition of $\underline{\nu}_N$) to $\underline{\nu}_N[-\varepsilon, \varepsilon]$, our concentration property says in terms of $\underline{\nu}_N$ that

$\underline{\nu}_N$ converges to the Dirac atomic δ-measure on the real line located at zero.

In fact, since the function

$$\sigma(x) = \sqrt{1-x^2},\ x \in [-1,1],$$

is continuous with a *unique* maximum at zero, almost all mass of the $(N-1)$-th power of σ, for large N, is located close to zero with *exponentially small* "tail" of what stays ε-away from zero. Namely,

$$\int_{-\varepsilon}^{-1} \sigma^{N-1}(x)dx + \int_{\varepsilon}^{1} \sigma^{N-1}(x)dx \le (1-\delta)^{N-1} \int_{-\varepsilon}^{\varepsilon} \sigma^{N-1}(x)dx$$

for all $\varepsilon > 0$ and $\delta = \delta(\varepsilon) > 0$. (We shall see later in this section that, because of the vanishing of the first derivative of our σ at zero and strict negativity of the second derivative, this δ around ε^2.)

From Balls to Spheres and Back. The above concentration property for the N-ball, implies that

the spherical measure of the boundary sphere $S^{N-1} = \partial B^N$ for large N is concentrated near the equator $S^{n-2} = S^{N-1} \cap B^{N-1}$.

In fact there is little difference between volume distributions on balls and on spheres, since

the measures of the balls are concentrate near their boundary spheres:

the points x in B^N that are ε-far from S^{N-1} sit in the concentric ball $B^N(1-\varepsilon) \subset B^N = B^N(1)$, the relative volumes of which are negligibly (exponentially) small:

$$\frac{vol\left(B^N(1-\varepsilon)\right)}{vol\left(B^N(1)\right)} = (1-\varepsilon)^N \to 0 \text{ for } \varepsilon > 0$$

Isoperimetric Exercise. Show that the Euclidean volumes (Lebesgue measures) of *all* domains in the ball of radii R, say $U \subset B^N(R)$< are concentrated at their boundaries as much as for the ball itself:

[31] "Relative volume" of a subset $A \subset B$ is $vol(A)/vol(B)$.

the proportion of the measure of a $U \subset B = B^N(R)$ *in the set of points in* U *with distance* $\geq \delta R$ *from* ∂U *is* $\leq |1-\delta|^N$, *where the equality holds only for* $U = B^N(R)$.

Hint. Compare the volume of U to that of the cone from the center of B over the boundary ∂U and show that $vol(U) \leq \frac{R}{N} vol(\partial U)$. Then apply this to the subdomains $U_{-\varepsilon} \subset U \cap B^N(R-\varepsilon)$ defined by $dist(u, \partial U) \geq \varepsilon$ for all $\varepsilon \in [0, \delta]$.

Isoperimetric Remark. The above inequality is valid for *all* domains $U \subset \mathbb{R}^N$, with $vol(U) \leq vol\left(B^N(R)\right)$:

$$\frac{vol\{x \in U\}_{dist(x,\partial U) \geq \delta R}}{vol(U)} \leq |1-\delta|^N,$$

but you can't (at least nobody has succeeded in it) get it by brute force of straightforward integration as in the case of $U \subset B(R)$: all known proofs are more imaginative than that.[32]

Equatorial concentration for the sphere S^{N-1} (and hence for the ball $B^N \supset S^{N-1}$) can be also derived from the relation

$$\mathbf{E}(x_1^2)_{S^{N-1}} = \frac{1}{N} \to 0,$$

where $\mathbf{E}(x_1^2)_{S^{N-1}}$ denotes the expectation (average over S^{N-1}) of the squared coordinate x_1^2, and where the identity $\mathbf{E}(x_1^2)_{S^{N-1}} = \frac{1}{N}$ follows from the equalities

$$\mathbf{E}(x_i^2)_{S^{N-1}} = \mathbf{E}(x_1^2)_{S^{N-1}}, \text{ for all } i = 1, 2, \ldots N,$$

and

$$1 = \mathbf{E}\left(\sum_{i=1}^{N} x_i^2\right)_{S^{N-1}} = \sum_{i}^{N} \mathbf{E}(x_i^2)_{S^{N-1}} = N\mathbf{E}(x_1^2)_{S^{N-1}}$$

for $\sum_{i=1}^N x_i^2 = 1$ on S^{N-1}. (To pass from $\mathbf{E}(x_1^2)_{S^{N-1}} \to 0$ to concentration one uses the Markov inequality as in 1.1, which is obvious anyway.)

Yet another way to visualise the concentration of the measures of spheres at their equators is by observing that the spherical distance function $x \mapsto dist_{S^{N-1}}(x, x_0)$ pushes forward the spherical measure to the segment $[0, \pi]$, where it's density function is equal, up to a scaling constant, to $(\sin x)^{N-2}$, which sharply concentrates at $x = \pi/2$ for large N for the same reason the function $\left(\sqrt{1-x^2}\right)^{N-1}$ does at $x = 0$.

$[\perp]$ Random and Orthogonal. The concentration properties of the unit balls B^N and spheres $S^{N-1} = \partial B^N$ have the following corollary.

Randomly independently chosen vectors x_j, $j = 1, 2, \ldots, k$, *in the unit ball* B^N *are, with high probability, nearly unitary and nearly mutually orthogonal, provided* N *is sufficiently large.*

[32] This is called the ***isoperimetric inequality*** and it must be on the list of ten (may be five) greatest theorems in geometry (in all of mathematics?), in the company of Pythagorean $a^2 + b^2 = c^2$ – the numero uno theorem in mathematics – and Bernoulli's law of the large numbers together with Leibniz' $1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \ldots = \frac{\pi}{4}$ which compete for being the second greatest ones.

This may be not especially exciting for $k < N$, but remarkably, (and counterintuitively) the almost orthogonality holds true for k *much larger* than N.

Namely,

there exists a strictly positive function $\epsilon(\varepsilon) > 0$, $\varepsilon > 0$, such that if, for a given $\varepsilon > 0$,

$$\log M \leq \epsilon(\varepsilon) N$$

then

$$||x_j|| > 1 - \varepsilon \text{ and } \langle x_{j_1}, x_{j_2} \rangle \leq \varepsilon$$

with probability $> 1 - \frac{1}{(1+\epsilon(\varepsilon))^N}$.

This trivially follows from the Exponential Tail Bound for balls, which is proven below with a Gaussian approximation of the measure $(\sqrt{1-x^2})^{N-1}dx$.

$\sqrt{N}$-SCALING OF BALLS AND GAUSSIAN MEASURES. The concentration of measures of unit balls at their equators can be seen in a finer detail under $\sqrt{N}$*-magnification*, that is by looking at the *balls of radii* $R = \sqrt{N}$, instead of $R = 1$ and at the projections of their (normalised) measures to lines.

The densities of the normalised push-forward measures now visibly

converge to the density functions of normalised Gaussian measures $G = \exp - \frac{x^2}{2} dx$*, for*

$$\left(\sqrt{1 - \left(\frac{x}{\sqrt{N}}\right)^2}\right)^{N-1} = \left(\left(1 - \frac{1}{x^{-2}N}\right)^{x^{-2}N}\right)^{\frac{N-1}{N}\frac{x^2}{2}} = (e+\epsilon)^{-\frac{x^2}{2}}.$$

where $\epsilon = \epsilon(x, N)$ (as well as the function $\sqrt{1 - x^2/N}$ itself) is defined for $|x| \leq \sqrt{N}$ and where $\epsilon \to 0$ for $\frac{x}{\sqrt{N}} \to 0$.

This implies the following quantitative version of the above $[\bullet_{\varepsilon^N}]$concentration.[33]

$[\bullet_{\lambda/\sqrt{N}}]$ *Almost all volume of the unit* N*-ball* $B^N = B^N(1)$ *is contained in the* $\frac{2\lambda}{\sqrt{N}}$*-thick band*

$$\{x_1, x_2, ..., x_N\}_{|x_1| \leq \frac{\lambda}{\sqrt{N}}} \subset B^N$$

for $\lambda >> 1$ *and* $N \to \infty$.

In fact, the relative volumes of the complementary δ-thick spherical caps for $\delta = 1 - \lambda/\sqrt{N}$, denoted $Cap_\delta^N \subset B^N = B^N(1)$, are bounded by

$$\frac{vol\left(Cap_\delta^N\right)}{vol\left(B^N\right)} \leq \frac{\int_\lambda^\infty (e-\epsilon)^{-\frac{x^2}{2}} dx}{\int_{-\infty}^\infty (e+\epsilon)^{-\frac{x^2}{2}} dx}, \ \lambda = \sqrt{N}(1-\delta)$$

For instance, if $\lambda \leq 0.1\sqrt{N}$, then a rough estimate shows that $\epsilon < 1/2$ and the above inequality implies that

$$\frac{vol\left(Cap_\delta^N\right)}{vol\left(B^N\right)} \leq 2^{-\frac{\lambda^2}{2}} = 2^{-\frac{N}{4}(1-\delta)^2}.$$

[33] A sleeker version of is presented in section 2.2.

$[\bigcirc_\rho]$ *Spherical Geodesic Balls.* Let us even rougher evaluate the spherical volumes of geodesic balls $B_\circ^{N-1}(\rho) \subset S^{N-1}$ of radii $\rho \leq \pi/2$,

$$\frac{vol\left(B_\circ^{N-1}(\rho)\right)}{vol(S^{N-1})} = \frac{\int_0^\rho \sin^{N-2} t dt}{\int_0^\pi \sin^{N-2} t dt} \leq \frac{\rho \sin^{N-2} \rho}{2\int_0^{\pi/2} t^{N-2} dt} = \frac{2^{N-2}(N-1)\rho \sin^{N-2} \rho}{\pi^{N-1}}.$$

For example, the geodesic balls of radii $\pi/6$, $\pi/4$ and $\pi/3$ satisfy

$$vol(B_\circ^{N-1}(\pi/6)) < \frac{(N-1)}{6\pi^{N-2}} vol(S^{N-1}) \lessapprox \frac{1}{3^N} vol(S^{N-1}),$$

$$vol(B_\circ^{N-1}(\pi/4)) < \frac{(N-1)(\sqrt{2})^{N-2}}{6\pi^{N-2}} vol(S^{N-1}) \lessapprox \frac{1}{2^N} vol(S^{N-1})$$

and

$$vol(B_\circ^{N-1}(\pi/3)) < \frac{(N-1)(\sqrt{3})^{N-2}}{6\pi^{N-2}} vol(S^{N-1}) \lessapprox \frac{1}{1.8^N} vol(S^{N-1}),$$

where "$\lessapprox$" turns to "$<$" for $N > 10$.

1.4 Gaussian Cubes and Maxwell Distribution.

It is satisfying to see that the normalised Gaussian measure,

$$G(x)dx = (2\pi)^{-\frac{1}{2}} e^{-\frac{x^2}{2}} dx,$$

albeit in analytic garments, arrives from geometry as the limit for $N \to \infty$ of the coordinate projections of the normalised measures of N-balls (of N-spheres if you wish) of radii $\sqrt{N}$.

But the full beauty of "Gaussian" comes to life in the Cartesian powers $G^{\times N}(\overline{x})d\overline{x}^N$ of G, kind of "*Gaussian cubes*", that are the measures on $\mathbb{R}^N$, which are also called *Gaussian*, defined as follows

$$G^{\times N}(\overline{x})d\overline{x}^N = \prod_{i=1}^N G(x_i)dx_i = (2\pi)^{-\frac{N}{2}} e^{-\frac{1}{2}(x_1^2+x_2^2+...+x_n^2)} dx_1 dx_2 ... dx_n.$$

The amazing – obvious once being said – property of these "cubes" is that they are

fully rotationally symmetric, i.e. invariant under the orthogonal group $O(N)$.

Equally obviously,

the Gaussian measure $G^{\times N}(\overline{x})d\overline{x}^N$ on $\mathbb{R}^N$ is the only *rotationally symmetric measure the projection of which to the x_1-line is equal to $G(x_1)dx_1$.*

Exercises. (a) Recall the standard proof of the identity $\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$ by computing the integral $\int\int e^{-x_1^2 - x_2^2}$ in the polar coordinates.

(b) Let μ be a Borel measure on $\mathbb{R}$ and let $\mu \times \mu$ on $\mathbb{R}^2$ be the Cartesian square of μ , that is a measure on $\mathbb{R}^2$, such that

$$\mu \times \mu(U \times V) = \mu(U) \cdot \mu(V)$$

for all open subsets $U, V \subset \mathbb{R}$.

Show that if $\mu \times \mu$ is rotationally symmetric, then μ is Gaussian $ae^{-bx^2}dx$, or it is a (weak) limit of Gaussian, namely Lebesgue's cdx or Dirac's $c\delta(x)dx$. [34]

(c) Show that Gaussian function $G(x) = (2\pi)^{-1/2}e^{-x^2/2}$ is distinguished among all ae^{-bx^2} by the two normalisation conditions

$$\int_{-\infty}^{\infty} ae^{-bx^2}dx = 1 \text{ and } \int_{-\infty}^{\infty} x^2 ae^{-bx^2}dx = 1.$$

(d) Show that the push-forwards $\nu_{N,k}$ of the normalised measures of the balls B^N under the normal projection $B^N \to \mathbb{R}^k$, $k \leq N$, for the subspace $\mathbb{R}^k \subset \mathbb{R}^N$ spanned by the first k-coordinate vectors converge, for a fixed k and $N \to \infty$, to the Gaussian measure,

$$\nu_{N,k}(U) \to \int_U G^{\times k}(\overline{x})d\overline{x}^k = (2\pi)^{-\frac{k}{2}} \int_U e^{-\frac{1}{2}(x_1^2+x_2^2+...+x_k^2)}dx_1dx_2...dx_k$$

for all open subsets $U \subset \mathbb{R}^k$.

(e) Show that the Gaussian N-power measure on $\mathbb{R}^N$

$$G^{\times N}(\overline{x})d\overline{x}^N = (2\pi)^{-\frac{N}{2}} \int_U e^{-\frac{1}{2}(x_1^2+x_2^2+...+x_n^2)}dx_1dx_2...dx_N$$

concentrates near the sphere $S = S^{N-1}(\sqrt{N} \subset \mathbb{R}^N$: the value of this measure on the subset

$$U_\rho = \{\overline{x} \in \mathbb{R}^N\}_{dist(\overline{x},S)>\rho}$$

satisfies

$$\int_{U_\rho} G^{\times N}(\overline{x})d\overline{x}^N \to 0 \text{ for } \frac{\rho}{\sqrt{N}} \to 0.$$

Two Words about Gas Kinetics. All of the can be traced to Maxwell's papers on the kinetic theory of gases,[35] where, instead of our $\sqrt{N}$-ball, one deals with the sphere of radius $\sqrt{M}$ in $\mathbb{R}^M$ for $M = 3N$, where this $\mathbb{R}^M = \mathbb{R}^{3N}$ comes as $(\mathbb{R}^3)^N$ and where the points in this sphere

$$\sum_{i=1}^{n} ||x_i||^2 = M, \text{ equally defined by } \frac{1}{N}\sum_{i=1}^{n}\frac{1}{2}||x_i||^2 = 3/2, \ x_i \in \mathbb{R}^3$$

represent the velocity vectors of N particles in the 3-space with average kinetic energy per particle equal $3/2$.[36]

Then the velocity distribution of an individual particle averaged over this sphere, which equals the projection of the normalised spherical measure on S^{3N-1} to $\mathbb{R}^3$ by evaluating the x_1-coordinate on this sphere, converges, according to the Maxwell law, to the Gaussian distribution on $\mathbb{R}^3$, that is

$$(2\pi)^{-3/2}e^{-\frac{1}{2}||x_1||^2}dx_1,$$

[34] See chapter 4 in *Normal Distribution characterisations with applications* by Wlodzimierz Bryc for a stronger result,
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.64.1799&rep=rep1&type=pdf

[35] Possibly the mathematical aspects of it which we are concerned with here, discussed here were known, in different terms, prior to Maxwell/s work.

[36] Following physicists' convention we use the $\frac{1}{2}$-coefficient for the kinetic energy, which, however, makes little sense unless a particular system of units is chosen.

where dx_1 stands for the Euclidean (Lebesgue's) volume element in $\mathbb{R}^3$.

Finally, the concentration property of the spherical measure shows, according to Maxwell, (this needs a little thought) that the *average* Gaussian limit behaviour is also *typical*, which is seen in the particle picture as follows.

Let $\delta_{\bar{x},i}$, $\bar{x} = (x_1, ...x_i, ...x_N) \in \mathbb{R}^{3N}$ be the Dirac atomic δ-measures in $\mathbb{R}^3$ located at the points $x = x_i \in \mathbb{R}^3$, $i = 1, 2, ...N$. Then

the average of these measures (weakly) converges to the Gaussian one for an overwhelming majority of points $\bar{x} \in S^{3N}(\sqrt{2N})$,[37]

$$\frac{1}{N}\sum_i^N \delta_{\bar{x},i}(x) \underset{N\to\infty}{\to} (2\pi)^{-3/2} e^{-\frac{1}{2}||x||^2} dx.$$

Typical Versus Average. Recall – this was mentioned earlier – that averages over configuration spaces, such as S^{N-1}, are (with a luck) are analytically computable. But they remain pure mathematical abstractions having no physical meaning and being non-accessible to experimental verification, while *typical* behaviours are what physicists can observe.

For instance, in the classical *Stern-Zartman verification of the Maxwell distribution law*, one sees, mathematically speaking, that the averages of the Dirac measures

$$\frac{1}{n}\sum_{i\in S_n} \delta_{\bar{x},i}(x)$$

for random samples S_n of n particles, where n is large but much smaller than N, are indeed close to the Gaussian distribution predicted by Maxwell.

1.5 Cubes, Diamonds, Simplices, and Balls

To gain some high dimensional intuition let us compare geometric invariants of the four most symmetric N-dimensional convex Euclidean bodies, which are:

0-Centered Cube :

$$\blacksquare = \blacksquare^N = [-1, 1]^N = \{x_i\}_{|x_i|\le 1} \subset \mathbb{R}^N,$$

Diamond:

$$\blacklozenge = \blacklozenge^N = \{x_i\}_{\sum_i |x_i|\le 1} \subset \mathbb{R}^N,$$

N-Simplex:

$$\blacktriangle = \blacktriangle^N = \{x_i\}_{\sum_i x_i = 1} \subset \mathbb{R}_+^{N+1},$$

Unit ball:

$$\bullet = \bullet^N = \{x_i\}_{\sum_i x_i^2 \le 1} \subset \mathbb{R}^N,$$

The invariants we want to include in our list are as follows:

Dimensionless Volume:

$$\sqrt[N]{|X|} = |X|^{1/N} = \sqrt[N]{vol(X)},$$

Dimensionless Boundary Volume:

$$\sqrt[N-1]{|\partial X|} = |\partial X|^{1/N-1} = \sqrt[N]{vol_{N-1}(\partial X)},$$

[37] "Overwhelming" means, that the convergence takes place for $\bar{x}$ away from "bad" subsets $\Sigma_N \subset S^{3N}(\sqrt{2N})$ the (probability) measures of which exponentially fast converge to zero.

Diameter

$$diam(X) = \sup_{x,y\in X} dist(x,y),$$

Inradius:

$$inrad(X) = \sup_{x\in X} dist(x,\partial X)$$

Square Root of the Average/Expectation of the Distance Squared:

$$\sqrt{\mathbf{E}}dist^2(X) = \sqrt{\frac{1}{|X|^2}\int_{X\times X} dist(x,y)^2 dxdy},$$

These invariants "*inv*" were tailored to satisfy the following two conditions.

$\bullet_1$ *Monotonicity under inclusions.*

$$X \subset Y \Rightarrow inv(X) \le inv(Y).$$

$\bullet_2$ *Linearity under scaling.*

$$inv(\lambda X) = \lambda inv(X) \text{ for all } \lambda > 0.$$

Besides, the most relevant at the present moment $inv = \sqrt{\mathbf{E}}dist^2$ enjoys the following two properties. which help its evaluation in specific examples.

$\star_1$ if $X \subset \mathbb{R}^N$ is *centred*, i.e. if the center of mass of X is at 0, then, by the Pythagorean theorem,

$\mathbf{E}dist^2(X)$ *equals* $2N$*-times the the "double average" of the squared* L_2*-norms of unit linear functions* $l:\mathbb{R}^N\to\mathbb{R}$*:*

$$\sqrt{\mathbf{E}}dist^2(X) = \sqrt{\frac{2N}{|S^{N-1}|\cdot|X|}\int_{S^{N-1}} dl \int_X l(x)^2 dx},$$

where this S^{N-1} denotes the unit sphere in the (dual to $\mathbb{R}^N \supset X$)) space of linear functions $l(x)$ on $\mathbb{R}^N \supset X$ of norm 1;[38]

$\star_2$ moreover, if $X \subset \mathbb{R}^N$ is *centred* and *irreducibly (orthogonally) symmetric*, i.e. the isometry group of X , as it acts on $\mathbb{R}^N$, is *irreducible,*[39] Then

X is *isotropic*:

the integral $\int_X l(x)^2dx$ *is constant in* l *for* $||l|| = 1$.

On Isotropic Bodies and Measures. A Borel measure μ on $\mathbb{R}^N$, e.g the Lebesgue measure restricted an $X\subset\mathbb{R}^N$ is called *isotropic* if the integral

$$\int_{\mathbb{R}^N} l(x)^2 d\mu$$

is constant in l for $||l|| = 1$.

---

[38]Recall that $||l|| = sup_{||x||=1} l(x)$, $x\in\mathbb{R}^N$; conversely, $||x|| = \sup_{||l||=1} l(x)$.

[39]"Irreducible" means that there is no G-invariant linear subspace in $\mathbb{R}^N$ except for $\{0\}$ and $\mathbb{R}^N$ itself.

An equivalent way to put is by saying that the norm the scalar product on the space of linear functions $l:\mathbb{R}^N\to\mathbb{R}$ is equal, up to a scalar multiple, to that on the Hilbert space of linear functions on X,

$$||l|| = C\sqrt{\int_{\mathbb{R}^N} l(x)^2 d\mu}.$$

$$\langle l_1, l_2\rangle_{\mathbb{R}^N} = C\int_{\mathbb{R}^N} l_1(x) l_2(x) l d\mu.$$

for a constant $C>0$ which depends on μ.

It follows - this is classical – (Bernoulli? Laplace? Legendre? Binet?) that if

μ is central,

μ has finite mass, $\mu(\mathbb{R}^N)<\infty$,

the support of μ linearly spans $\mathbb{R}^N$,

then *there is a Euclidean/Hilbertian norm* $||...||_{new}$ *on* $\mathbb{R}^N$ *with respect to which* μ *is isotropic.*

Indeed, define the scalar product $\langle ...\rangle_{new}$ corresponding to $||...||_{new}$ by prescribing

$$\langle l_i, l_j\rangle_{new} = \int_{\mathbb{R}^N} l_1(x) l_2(x) d\mu$$

for some linear basis $\{l_i\}$ in the (dual to our $\mathbb{R}^N$) space of linear functions on $\mathbb{R}^N$. [40]

Also observe here (this is obvious) that the scaler product that makes μ isotropic is unique up to scaling. Thus, in particular,

irreducibly symmetric (as in $\star_2$) finite Borel measures on $\mathbb{R}^N$ *are isotropic.*

Let us look at the values of the above invariants for our bodies for large $N\to\infty$ where the most interesting point is comparative values of dimensionless volume $\sqrt[N]{|...|}=\sqrt[N]{Vol}$ and $\sqrt{\mathbf{E}dist^2}$.

$$\sqrt[N]{|\blacksquare|}=2,\quad \sqrt[N-1]{|\partial\blacksquare|}\sim 2,\quad diam\blacksquare = 2\sqrt{N},\quad inrad\blacksquare = 1,\quad \sqrt{\mathbf{E}dist^2\blacksquare} = 2\sqrt{\frac{N}{3}}$$

$$\sqrt[N]{|\blacklozenge|}\sim\frac{2e}{N},\quad \sqrt[N-1]{|\partial\blacklozenge|}\sim\frac{2e}{N},\quad diam\blacklozenge = 2,\quad inrad\blacklozenge = \frac{1}{\sqrt{N}},\quad \sqrt{\mathbf{E}dist^2\blacklozenge}\sim\sqrt{\frac{2}{N}}$$

$$\sqrt[N]{|\blacktriangle|}\sim\frac{e}{N},\quad \sqrt[N-1]{|\partial\blacktriangle|}\sim\frac{e}{N},\quad diam\blacktriangle = \sqrt{2},\quad inrad\blacktriangle\sim\frac{1}{N},\quad \sqrt{\mathbf{E}dist^2\blacktriangle}\sim\sqrt{\frac{2}{N}}$$

$$\sqrt[N]{|\bullet|}\sim\frac{\sqrt{2e\pi}}{\sqrt{N}},\quad \sqrt[N-1]{|\partial\bullet|}\sim\frac{\sqrt{2e\pi}}{\sqrt{N}},\quad diam\bullet = 2,\quad inrad\bullet = 1,\quad \sqrt{\mathbf{E}dist^2\bullet}\sim\sqrt{2}.$$

Comments. (I) Among the relations in this table the only ones that need explanation are those for $\sqrt{\mathbf{E}dist^2}=\sqrt{\frac{1}{vol^2}\int\int dist(\circ,\bullet)^2}$.

This is computed for the N-cube $\blacksquare=\blacksquare^N$ by integrating a single coordinate, say x_1 over $\blacksquare$ and by applying the above $\star_{1,2}$.

Similarly $\mathbf{E}dist^2$ is evaluated for the N-diamond $\blacklozenge$ and for the regular N-simplex $\blacktriangle$ (hopefully, there is no mistake with the constants here) where (and everywhere) $A\sim B$ signifies that $A/B\to 1$ for $N\to\infty$.

[40] The Euclidean/Hilbertian norm $||...||_{new}$ on $\mathbb{R}^N\supset X$ is often defined via what is called *Binet (Legendre?) ellipsoid* that is the unit ball $B^N_{new}=\{x_i\}_{||l|_{new}\le 1}\subset\mathbb{R}^N$.

Finally, $\sqrt{\mathbf{E}}dist^2$ for $\bullet$ is evaluated by observing that, since the measure of he N-ball is concentrated near its boundary sphere $\partial\bullet^N = \bigcirc^{N-1}$,

$$\sqrt{\mathbf{E}}dist^2(\bullet) \sim \sqrt{\mathbf{E}}dist^2(\bigcirc) = \sqrt{2}.$$

(II) It is instructive to compare the volume of the *cube* $\blacksquare = \frac{1}{\sqrt{N}}\blacksquare$ *inscribed in the unit sphere*, that is the unit cube $\blacksquare$ scaled down by the factor $\frac{1}{\sqrt{N}}$, to the volume of the unit ball,

$$vol\,(\blacksquare) \sim \left(\sqrt{\frac{2}{e\pi}}\right)^N \cdot vol(\bullet),$$

which implies, at least for $N > 10$, that

$$\frac{1}{3^N}vol(\bullet) < vol\,(\blacksquare) < \frac{1}{2^N}vol(\bullet).$$

(III) The relation

$$\int_{\blacksquare^N} x_1^2 dx_1, ..., dx_n = \int_{-1}^{1} x_1^2 dx_1 = 2/3,$$

which we used to prove $\sqrt{\mathbf{E}}dist^2\blacksquare = 2\sqrt{\frac{N}{3}}$, implies, by the *irreducible symmetry* of the cube the same equality for all unit liner functions l on $\mathbb{R}^N$, which for l being the projection to the principal diagonal of the cube, yields, as it was explained in section 1.2, Bernoulli' *law of large numbers* in agreement with the general principle:

SYMMETRY BEGETS STATISTICS.

(IV) In all four cases, our X satisfy:

$$\sqrt{\mathbf{E}}dist^2(X) = const\sqrt{N}\,\sqrt[N]{vol(X)} \text{ for } 0.01 \le const \le 100.$$

Conjecturally

this relation holds for all *centred isotropic* convex *bodies X in $\mathbb{R}^N$.*

(This is a long standing conjecture related to *Bourgain's slicing problem*, as was pointed out to me by Vitali Milman.[41])

(V) The (unit size) diamond and the cube $\frac{1}{N}\blacksquare^N$ with the edge size $2/N$ inscribed into this diamond have comparable volumes:

$$\sqrt[N]{vol\left(\frac{1}{N}\blacksquare^N\right)} = \frac{2}{\sqrt{N}} \approx \frac{2e}{N} \sim \sqrt[N]{vol(\blacklozenge^N)}.$$

Furthermore, the (obvious) exponential concentration of the volume of $\blacklozenge^N$ near $\blacklozenge^{N-1} \subset \blacklozenge^N$, implies that cubes scaled by $\lambda >> 1$ contain almost all of $\blacklozenge^N$.

Namely, the part of $\blacklozenge^N$ with the coordinates $x_i \ge \lambda/N$ satisfies

$$\frac{vol\left(\blacklozenge^N \setminus \frac{\lambda}{N}\blacksquare^N\right)}{vol\left(\blacklozenge^N\right)} \le \exp -c\lambda \text{ for } c > 0.1$$

[41]See *Notes on isotropic convex bodies* by A Giannopoulos, `users.uoa.gr/~apgiannop/isotropic-bodies.pdf` and `http://users.uoa.gr/~apgiannop/isotropic-sections.pdf` `https://arxiv.org/pdf/1511.05525`

(VI) It follows, the same is true for intersections of diamonds with balls $r\bullet^N = \bullet^N(r)$ of radii $r = \lambda/\sqrt{N}$, since these balls contain the cubes $\frac{r}{\sqrt{N}}\blacksquare^N$.

$$\frac{vol\left(\blacklozenge^N \smallsetminus \bullet^N\left(\lambda/\sqrt{N}\right)\right)}{vol\left(\blacklozenge^N\right)} \leq \exp -c\lambda \text{ for } c > 0.1.$$

Thus,

most of the volume of the N-diamond is located close to its center.

Similarly, one shows that

most of the volume of the regular N-simplex is is located close to its center of mass.

Exercise. Evaluate the percentage of the volume of the unit ball $\bullet^N$ contained in the intersection of this ball with the cube with the edge length $\lambda/\sqrt{N}$. for $1 < \lambda < \sqrt{N}$.

1.6 High Dimensional Convex Polyhedra

Let us briefly describe some remarkable polyhedra besides simplices, cubes and diamonds and also say a few words about about volumes, of general convex polyhedra. often called *polytopes* in the Euclidean $\mathbb{R}^N$.

1. Besides simplices, cubes and diamonds there are other remarkable convex polyhedra.

Symmetric and Bisymmetric Polytopes. A polytope $P \subset \mathbb{R}^N$ is *symmetric* if the isometry group of P is *transitive* on the set of vertices and it is *bisymmetric* if this group is transitive on the sets of the top dimensional faces of P as well.

For instance, the Cartesian products of simplices: $P = \blacktriangle^{N_1} \times \ldots \times \blacktriangle^{N_i} \times \ldots \times \blacktriangle^{N_j}$ are symmetric and if $N_1 = N_2 = \ldots = N_j$ these P are *bi-symmetric.*

Another power-like operation, which applies to all N-dimensional convex polyhedra P and which preserves (bi)symmetry, is as follows.

Let G be a finite group of orthogonal transformations of $\mathbb{R}^M$ and $A^N \subset \mathbb{R}^M$ be an N-dimensional affine subspace, such that $g(A)$ is normal to A for all non-identity elements $g \in G$.

Imbed P into A and let P^{*G} be the *convex hull of the G-orbit of P.*

Clearly

if P is symmetric then P^{*G} is also symmetric and if P is bisymmetric then P^{*G} is bisymmetric as well.

For instance the $(N-1)$-simplex $\blacktriangle^{N-1}$ and the N-diamond $\blacklozenge^N$ are thus obtained from the segment $[-1,1]$ by suitably embedding it to $\mathbb{R}^N$ with the coordinate permutation group Π_N action on it.

Besides $[-1,1]$, there are two other "prime" bisymmetric polyhedra: the regular icosahedron and dodecahedron in the 3-space, which generate higher dimensional ones by successive application of "$\times j$" and "$*k$" to them.

Bistochastisity versus Bisymmetry I am not certain –this may be known – if there are other sources of bisymmetric polyhedra, but there is a family of symmetric polyhedra which are as beautiful as the bisymmetric ones.

These are the spaces $BS^{NN} \subset \mathbb{R}^{N^2}$ of *bistochastic matrices* that are *convex hulls of the sets of coordinate permutations* in the space $\mathbb{R}^{\mathbb{N}^2}$ of linear transformations of $\mathbb{R}^N$.

Remarkably, these BS^{NN}, which are symmetric of dimension $N^2 - 2N + 1$ with $N!$ vertices, have *only* $N^2 + 2N$ *top dimensional faces*, where the action of the automorphism group of BS^{NN} on the set of these faces has *only two orbits*, the one of which consists of the faces defined by N^2 inequalities $x_{ij} \geq 0$ and the faces from the second orbit are defined by $2N$ equations $\sum_i x_{ij} = 1$ and $\sum_j x_{ij} = 1$.

(Traditionally, the polyhedra BS^{NN} are *defined* by these inequalities and equations, while identification of the *extremal points* with *permutation matrices* makes the content of *the Birkhoff-von Neumann theorem*, while the frequently used combinatorial interpretation/corollary of this goes under the name of *König's matching theorem.* The proof of *the Birkhoff-von Neumann theorem* is nontrivial but not difficult either – we invite you to find it by yourself.)

Cyclic Polyhedra and the Upper Bound Theorem. *The moment curve* $\mathsf{momc} \subset \mathbb{R}^{\mathsf{N}}$ is the image of the map $\eta : \mathbb{R} \to \mathbb{R}^N$ for

$$\eta : t \mapsto (x_1 = t, x_2 = t^2, ..., x_N = t^N).$$

A cyclic polyhedron $\mathsf{CyP_k} \subset \mathbb{R}^{\mathsf{N}}$ is the convex hulls of k points from this curve momc.

An amusing feature of such a $\mathsf{CyP_k}$ is that it has lots faces of all dimensions.[42] For instance, if $M \leq N/2$, then the number of the M-faces is

$$\frac{k(k-1)...(k-M)}{M!},$$

that is

every M-tuple of vertices of $\mathsf{CyP_k}$ spans a face.

In fact, if $M \leq N/2$, then, given M points on the moment curve, say,

$$\eta(t_j) \in \mathsf{momc},\ \mathsf{j} = 1, 2, ..., \mathsf{M},$$

there is a supporting hyperplane for momc *defined by a non-constant linear (better to say affine) function $f(x)$ on $\mathbb{R}^N$, which is non negative on* momc *and which vanishes at these M points.*

Indeed, let $p(t)$ be a non-zero polynomial of degree N, which vanishes at our M points – such a p exists for $M \leq N/2$ – and write

$$p(t)^2 = \sum_{i=0}^{N} c_i t^i.$$

. Then the function $f(x_1, ..., x_N) = \sum_{i=0}^{N} c_i x_i$ is the one we want.

It follows that every M-tuple of vertices in $\mathsf{CyP_k}$ is contained in a *supporting hyperplane for* $\mathsf{CyP_k}$, hence, it spans a face of $\mathsf{CyP_k}$.

What is less obvious is that

the number of M-faces of a polytope with k vertices in $\mathbb{R}^N$, is bounded by that number for $\mathsf{CyP_k}$.

[42] `https://www.cs.mcgill.ca/~fukuda/soft/polyfaq/node12.html`

This is called the *Upper Bound Theorem* conjectured by Motzkin in 1957, proven by MucMullen in 1970, generalised by Stanley in 1975 to all triangulations of the sphere S^{N-1} and extended to *Minkowski sums* of polyhedra by Adiprasito and Sanyal in 2014. [43]

2. Now let us turn to volumes and start with the following pretty observation[44] concerning the *volumes of the convex hulls* of k-tuples of points x_i[45] in the Euclidean space, denoted

$$conv\{x_1, ..., x_k\} \subset \mathbb{R}^N.$$

Lemma. The subset $conv\{x_1, ..., x_k\} \subset \mathbb{R}^N$ is contained in the union of the balls

$$B_i = B_{\frac{1}{2}x_i}\left(\frac{1}{2}||x_i||\right), i = 1, ..., k,$$

which have radii $\frac{1}{2}||x_i||$ and centers at $\frac{1}{2}x_i \in B^N$.

Indeed, since the segments $[0, x_i] \subset B_i$ serves as diameters in these balls B_i, a point $x \notin B_i$ if and only if the angle of the triangle $[0, x, x_i]$ at x is *acute*,

$$\angle_x([0, x], [x, x_i]) < \frac{\pi}{2},$$

where these inequalities for all i imply that

some hyperplane normal to the segment $[0, x]$ *and crossing it next to* x *separates* x *from* $\{x_1, ..., x_n\}$.

Hence, *every* $x \in conv\{x_1, ..., x_n\}$ *is contained* in some of B_i. QED.

Corollary. The volumes of the convex hulls of all k-tuples of points in the unit ball $B^N \subset \mathbb{R}^N$ are bounded by

$$[\bullet/2^N] \qquad \frac{vol(conv\{x_1, ..., x_k\})}{vol(B^N)} \le \frac{k}{2^N}.$$

This inequality gives a fair idea of what happens for $k = c^N$ if $1_\varepsilon \le c \le 2 - \varepsilon$ but dismally fails for subexponential k and for $c \ge 2$.

3. Let us prove a simple volume bound applicable to all $k = c^N$.

The volumes of the convex hulls of all k-tuples of points in the unit ball B^N with $k \le c^N$ satisfy

$$[\bullet_{\beta(c)}] \qquad \frac{vol(conv\{x_1, ..., x_k\})}{vol(B^N)} \le \beta^N + \frac{1}{2^N}(1 - \beta^N)$$

for some $\beta = \beta(c) < 1$, where a specific (rough) bound on β is as follows:

$$\beta \le \sqrt{1 - \frac{1}{16c^{\frac{2N}{N-1}}}}.$$

[43]See https://en.wikipedia.org/wiki/Upper_bound_theorem and https://arxiv.org/abs/1405.7368

[44][G. Elekes 1986] *A geometric inequality and the complexity of computing* https://www.math.cmu.edu/~af1p/Teaching/MCC17/Papers/elevol.pdf, also see section I.2 in *Combinatorial complexity* by Pach and Agarwal.

[45]Depending on context, x_i may denote either points in $\mathbb{R}^N$ or coordinates of points.

This is shown similarly to [$\bullet/2^N$] with a cruder but more general

Covering Lemma. If positive numbers, r, ρ and $h = 1 - r$ satisfy

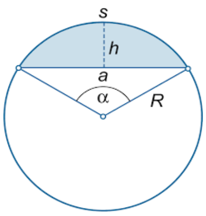
$$\rho^2 + r^2 \geq 1 + h^2,$$

then the balls

$$B_{x_i}(\rho), i = 1, ...k, \text{ together with } B_0(r)$$

cover all of the hull $conv\{x_1, ..., x_k\} \subset B^N = B_0(1)$.

Proof If the points x_i lie on the unit sphere $S^{N-1} = \partial B^N$ (which is the only case we need) the covering property is clear from the picture below[46] and then the general case follows.



Using this, a simple evaluation of the volume of the part of the h-band in B^N around $S^{N-1} = \partial B^N$ covered by the ρ-balls shows that

$$\frac{vol(conv\{x_1, ..., x_k\})}{vol(B^N)} \leq r^N + (1 - r^N)c^N(2\rho)^{N-1}$$

and if we set

$$\rho = \frac{1}{4c^{\frac{N}{N-1}}} \text{ and } r = \sqrt{1 - \rho^2},$$

then

$$r^N + (1 - r^N)c^N(2\rho)^{N-1} = \left(1 - \frac{1}{16c^{\frac{2N}{N-1}}}\right)^{\frac{N}{2}} + \frac{1}{2^N}\left(1 - \left(1 - \frac{1}{16c^{\frac{2N}{N-1}}}\right)^{\frac{N}{2}}\right).$$

QED.

4. Convex polyhedra *circumscribed* around the unit ball satisfy lower bounds on their volumes dual to the above [$\bullet_{\beta(c)}$].

For instance, *if a convex polyhedron Q contains the unit ball $B^N = B_0(1)$ and if the number k of the $(N-1)$-dimensional faces F_i of P is bounded by $k \leq (10)^N$, then the volume of P is bounded from below by*

[CIRC$_{10}$]
$$\frac{vol(Q)}{vol(B^N)} \geq 1.0008^N.$$

Sketch of the Proof. Let $y_i \in S^{N-1} = \partial B^N$ be the nearest points to the faces F_i an let $U \supset \mathbb{R}^N$ be obtained by removing the balls $B_{y_i}(\rho)$ from the external h-band around S^{N-1} in $\mathbb{R}^N$,

$$U = B_0(1 + h) \setminus B_0(1) \setminus \bigcup_i B_{y_i}(\rho).$$

[46] The only notation in this picture which agrees with ours is h, while s corresponds to some of x_i and R must be 1.

If $\rho \gtrapprox 2\sqrt{h}$, then the set U *is contained* in Q, and the proof follows with a suitable choice h and ρ.

5. Let the above $Q \supset B^N$ be actually *circumscribed* around the unit ball B^N, which means that the faces F_i of Q meet B^N at the points y_i.

Santalo Inequality. If Q is symmetric with respect to the origin,

$$-Q = Q,$$

then the volume of the convex hull $P = conv\{y_i\}$ satisfies

$$\frac{vol(Q)}{vol(B^N)} \cdot \frac{vol(P)}{vol(B^N)} \leq 1.$$

Consequently,

the maximal volume of an arbitrary, not necessarily symmetric, inscribed convex polyhedron with k vertices in B^N is bounded in terms of the minimal volume of the circumscribed ones with 2k faces as follows,

$$\frac{max_k vol_{iscr}}{vol(B^N)} \leq \frac{vol(B^N)}{\min_{2k} vol_{circ}}.$$

6. *Inverse Santalo Inequality. According to* a Bourgain-Milman theorem, *circumscribed symmetric Q satisfy*

$$\frac{vol(Q)}{vol(B^N)} \cdot \frac{vol(P)}{vol(B^N)} \geq const^N.$$

Therefore,

$$\frac{min_k vol_{circ}}{vol(B^N)} \geq const^N \frac{vol(B^N)}{\max_{2k} vol_{inscr}}.$$

(According to the *Mahler conjecture* the extremal Q is the cube $\blacksquare^N = [-1,1]^N \supset B^N$ and P dual to $\blacksquare^N$ is the diamond $\blacklozenge^N \subset B^N$, thus, the (still conjectural) optimal constant comes from the relation

$$vol(\blacksquare^N) \cdot vol(\blacklozenge^N) = \tfrac{4}{n!} = const_{opt} \cdot (vol(B^N))^2.)$$

7. *Questions.* The above upper and lower bounds on the volumes are sharp up to a $const^N$ factor. This is more or less satisfactory for $c >> 2$, where, moreover, one can get a better estimate for this universal *const* with a little extra work.

Also, I guess there is an inequality interpolating between $[\bullet/2^N]$ and $[\bullet_{\beta(c)}]$ buried somewhere in textbooks on convexity.

It is also known[47] what happens for subexponential k, but I am not certain if the following questions have been answered.[48]

7.A. What is an asymptotically sharp bound on of the volumes of convex polyhedra P with $k = c^N$ vertices in the unit ball B^N for $N, c \to \infty$ and, more interestingly, when $N, c^{-1} \to \infty$?

[47] I Bárány,, Z. Füredi. *Computing the Volume is Difficult.* `https://pdfs.semanticscholar.org/d443/b44c8037c27426445c2694411447fa5f729b.pdf`

[48] There are hundreds of papers on volumes of polyhedra inscribed in and circumscribed around balls and other convex sets as well as estimates on the volume efficiency of approximation of convex sets by polyhedra. (Some papers listed at the end of this section.) But I mainly failed to extract relevant information. Maybe you'll have better luck.

(Nothing of the above tells you, for instance, that the volume of $conv\{x_1, ..., x_k\}$, where $x_i \in B^N$, $k \leq 0.000001^N$, and $N \geq 1\,000\,000\,000$, is bounded by $vol(B^N)/10^N$.)

7.B. How close is the volume of the cube inscribed in S^{N-1} to the maximum of the volumes of polyhedra in the ball with 2^N vertices? [49]

7.C. What are upper bounds on the volumes of polyhedra with $k = c^N$ vertices in the N-cube for $c < 2$?

7.D. How do other *(bi)symmetric* polyhedra fare in this regard?

7.E. .What are lower bounds on the volumes of convex polyhedra $P \subset \mathbb{R}^N$, with k vertices, which intersect the unit ball across a "significantly large" set, i.e. where

$$\frac{vol(P \cap B^N)}{vol(B^N)} \geq c^N$$

for a given positive $c < 1$, e.g. for $k = 1, 1^N$ and $c = 0.9$?

7.F *Veronese Polyhedra.* Besides the moment curve there other distinguished algebraic subvarieties in $\mathbb{R}^N$, where the most beautiful one is the *Veronese variety* $\mathsf{VER}^\mathsf{n} \subset \mathbb{R}^\mathsf{N}$ for $N = \frac{(n+1)(n+2)}{2} - 1$ that is the image of the map from the unit sphere $S^n \subset \mathbb{R}^{n+1}$.

What is the geometry of convex hulls VERP_k of k-tuples of points from VER^n?

(Every VERP_k defines a convex cone in the space of positive definite quadratic forms on $\mathbb{R}^{n+1}$ which also can be seen as a curve-linear polyhedron in the symmetric space $SL(n+1)/SO(n+1)$.)

1.7 Random Convex Hulls and Error Correcting Programs

Let μ, be a probability measure in $\mathbb{R}^N$, and let $\{x_1, ..., x_k\}_\mu$ denote a *μ-random k-tuple of points* $x_i \in \mathbb{R}^N$, which is a shorthand for *the power measure* $\mu^{\otimes k}$ *on* $(\mathbb{R}^N)^k$.

Questions. What are expectations, i.e. $\mu^{\otimes k}$-mean values, of geometric invariants of the convex hulls $conv\{x_1, ..., x_k\}_\mu \subset \mathbb{R}^N$?

What is the overall geometry of a typical $conv\{x_1, ..., x_k\}_\mu$?

Our basic (but not the only) example is where k-tuples of points x_i are taken on random from the unit ball $B^N \subset \mathbb{R}^N$, where their convex hull is denoted

$$conv\{x_1, ..., x_k\}_\bullet \subset B^N,$$

with "$\bullet$" referring to the Lebesgue measure "dx" in the ball.

A significant feature of these $conv\{x_1, ..., x_k\}_\bullet$, is as follows.

If $k \leq c^N$, $1 \leq c \leq \sqrt{2}$, then, with probability $\geq 1 - \left(\frac{c^2}{2}\right)^N$, the convex hull $conv\{x_1, ..., x_k\}_\bullet \subset B^N$, regarded as a convex polyhedron, has exactly k vertices, i.e. all k points $x_i \in conv\{x_1, ..., x_k\}_\bullet$ are extremal – none of x_i is contained in the convex hull of the remaining $(k-1)$ points; in other words, each x_i can be separated by a hyperplane from the subset $\{x_1, ..., x_{i-1}, x_{i+1}, ..., x_k\} \subset B^N$.

[49] János Pach pointed out to me that the maximum volume inscribed polyhedra must be simplicial.

In fact, the probability of *non-separability* of an individual x_{i+1}, from $\{x_1, ..., x_{i-1}\}$, that is the relative volume of $conv\{x_1, ..., x_{i-1}\}_\circ \subset B^N$, is bounded by

$$\frac{vol(conv\{x_1, ..., x_{i-1}\}_\bullet)}{vol(B^N)} \le \frac{i-1}{2^N}$$

by Elekes inequality (the above $[\bullet/2^N]$) and non-separability of at least one of x_i from the rest is at most

$$\sum_{i=1}^{k} \frac{i-1}{2^N} \le \frac{k^{2N}}{2^N}.$$

QED. separated

This is already refreshingly counterintuitive and then you may be not sufficiently impressed to learn – we shall explain this below – that, typically, not only individual x_i can't be separated but even i-tuples of several of them. But high dimensions are, I believe, full of much greater surprises the time of which is yet to come.

m-Faces Estimate. Denote by $\mathsf{P}[\triangle^m, N, k]$ the probability that all $(m+1)$-sub-tuples in the random k-tuples of points $x_i \in B^N$ span m-faces in the convex hull $conv\{x_1, ...x_k\}_\bullet \subset B^N$. Then,

If $m \le \frac{N}{100}$ and $k \le 1.01^{\frac{N}{m^2}}$,then, for all $N = 1, 2, ...,$ this probability satisfies:

$$\mathsf{P}[\triangle^m, N, k] \ge 1 - 0.99^{\frac{N}{m^2}}.$$

Proof. Start by observing that the distance $d_N(m)$ between an m face and the opposite $(N-m-2)$-face in the $(N-1)$-simplex spanned by orthonormal unit vectors in $\mathbb{R}^N$ that is,

$$d_N(m) = \sqrt{\frac{1}{m+1} + \frac{1}{N-m-1}},$$

is bounded from below for $1 \le m \le N/10$ by

$$d_m \ge \frac{1}{2\sqrt{m}}.$$

Then we recall (see$[\perp]$ and $[\bullet_{\lambda/\sqrt{N}}]$ in section 1.3) that a random $x \in B^n$ has $||x|| \ge 1-\epsilon$ with probability $1-(1-\epsilon)^N$ and, for a given $x_0 \in \mathbb{R}^N$, the scalar product $\langle x, x_0\rangle$ is bounded by ϵ with probability

$$\mathbf{P}(\epsilon) \ge 1 - 2^{-\frac{N\epsilon^2}{4}}.$$

(This is written in $[\bullet_{\lambda/\sqrt{N}}]$ with $1-\delta$ instead of ϵ.)

It follows by a simple computation that

all edges in the random polyhedron $conv\{x_1, ..., x_k\}_\bullet \subset B^N$ for $k \le 1.01^{\frac{N}{m^2}}$ have the lengths in the range

$$\sqrt{2} \pm \frac{1}{10\sqrt{m}}$$

with probability $p \geq 1 - 0.99^{\frac{N}{m^2}}$.

This in combination with the above lower bound $d_m \geq \frac{1}{2\sqrt{m}}$, shows that, with the same probability p, the convex hulls of disjoint $l+1$-sub-tuples and of $N-l$ sub-tuples for $l \leq m$ don't intersect, and the proof follows since, obviously,

if an l-dimensional simplex $\triangle^l \subset \mathbb{R}^N$ intersect the interior of a compact convex subset $C \subset \mathbb{R}^N$, then the intersection of an n-face $\triangle^n \subset \triangle^l$, $n \leq l$, with the interior of C non-trivially intersects some $(N-n)$-simplex $\triangle^{N-n} \subset C$ spanned by $N-n$ *extremal points* of C,

and since

every point in a regular $(N-n)$-simplex with edges $\sqrt{2}$ lies within distance $\leq \frac{1}{\sqrt{N-n}}$ from some $(N-n-1)$-face of this simplex,

where the proof is concluded by actually performing all computations, which is not difficult with our generous choice of constants.

Remarks (a) The above argument applies to the convex hulls $conv\{x_1, ...x_k\}_\mu$ of μ-random k-tuples of points in $\mathbb{R}^N$, whenever the distance function $||x-y||$ is $\mu^{\otimes}2$-concentrated with exponential tail bound. This includes the uniform measure in the N-cube $\blacksquare^N$, as well as the Bernoulli measure supported on the vertices of $\blacksquare^N$ and, probably, (this seem easy is, probably, written somewhere) the natural measures supported on the n-skeleta of $\blacksquare^N$ for all n.

(b) In their 2006 paper[50] Donoho and Tanner establish, – this takes about 30 pages of calculation in their 80 page paper – a much stronger lower bound on m which allows $m = const \cdot N$ for $const \geq \frac{1}{6}$ in some cases. Strictly speaking, their result (corollary 7.1 on p.38), as it is stated in the paper, applies only to subexponential $k = k(N)$, but, probably, their techniques do say something about $k = c^N$.

However, it seems to remain unclear what the maximal m could be, say for $k = 2N$. Is it $\frac{N-\varepsilon}{2}$?

CUTTING AWAY ERRONEOUS VERTICES.

In a 2017 paper, Gorban, Burton, Romanenko and Tyukin[51] prove sharp bounds for linear separability of members of random k-tuples $\{x_1, ...x_k\}_\mu$ for several classes of probability measures μ in $\mathbb{R}^N$ and propose effective and relatively simple algorithms A for correcting errors of a certain class of existing fairly complicated (heuristic) algorithms $\mathcal{A}$, where x_i represent the input, output and internal parameters of $\mathcal{A}$, where the errors are represented by certain extremal points of $conv\{x_i\}_\mu$ and where the correcting algorithms A are designed to linearly separate these errors.

1.8 Archimedean and Other Measure Preserving Maps.

Start with a question.

[50] *Counting faces of randomly-projected polytopes when the projection radically lowers dimension.*
https://arxiv.org/abs/math/0607364

[51] A. Gorban, R Burton, I. Romanenko, I. Tyukin, *One-Trial Correction of Legacy AI Systems and Stochastic Separation Theorems.* `https://arxiv.org/abs/1610.00494` . Also see [Blessing of dimensionality] in the reference list in section 1.10.

What are measure preserving maps, preferably simple and natural ones, between simple spaces with simple measures on them?

Here is a beautiful instance of this.

$\star_1$ *the (Archimedean) map from the complex plane $\mathbb{C}$ to the positive ray $\mathbb{R}^+$, for*

$$A = A_1 : z \mapsto |z|^2$$

sends the Euclidean (Lebesgue) measure in the plane to π times the usual (Lebesgue) measure on $\mathbb{R}^+$.

In fact, the A-pullback of the segment $[x, x+\delta] \subset \mathbb{R}_+$ is an annulus in the plane between the circles of radii $r = \sqrt{x}$ and $r + \Delta$ for

$$\Delta = \left|\frac{dr^2}{dr}\right|^{-1} \cdot \delta + o(\delta) = \left(\frac{1}{2}r\right)^{-1} \delta + o(\delta),$$

which makes

$$area(A^{-1}[x,\delta]) = 2\pi r\Delta + o(\Delta) = \pi\delta + o(\delta)$$

for infinitesimally small δ; hence,

$$area(A^{-1}[x,\delta]) = \pi\delta$$

for all $x \in \mathbb{R}_+$ and $\delta > 0$. QED.

However simple, this is amazing. If you disagree, try to find another polynomial map from the unit disc B^2 to the segment $[0,1]$ that would push forward the normalised Lebesgue measure on B^2 to that on $[0,1]$.)

But this is not the original Archimedean map, which, in fact,

sends the unit sphere $S^2 \subset \mathbb{R}^3$ to the segment $[-1,1]$ and pushes forward the spherical measure to 2π times the Lebesgue measure.

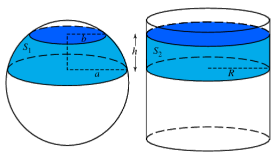


Figure 1: Archimedes' Theorem

The map itself is plain and simple: it is the normal projection of the sphere to the vertical coordinate line, as in the picture. What is remarkable – Archimedes believed that was his main accomplishment – is that

this projection sends the spherical measure to
2π times the linear Lebesgue measure.

Archimedes' Proof. By the Pythagorean theorem, this map – think of it as the hight function h on the sphere for $-1 \le h \le 1$ – slices the sphere into circles, which are, on the level h, have length $2\pi\sqrt{1-h^2}$, while the the differential of this map at the hight h is $\sqrt{1-h^2}$, also by the Pythagorean theorem.

Hence, the areas of the spherical annuli between the circles on the levels h and $h+\delta$, for $-1 \le h, h+\delta < 1$ are, *independently of* h, equal to $2\pi\delta$, for

$$\left(2\pi\sqrt{1-h^2}\right)\left(\sqrt{1-h^2}\right)^{-1}\delta = 2\pi\delta.$$

Let us give an alternative proof of Archimedes' theorem by deriving it from the corresponding measure preservation property of the map A_1 in above $\star_1$.

Let $A_N = A_1^{\times(N)}$ be the $(N+1)$th Cartesian power of A_1, that is the map from the complex $(N+1)$-space to the positive "quadrant" $\mathbb{R}_+^N$,

$$A_N : \mathbb{C}^N \to \mathbb{R}_+^N,$$

defined by

$$A_N : (z_1, ..., z_N) \mapsto (|z_1|^2, |z_2|^2, ..., |z_N|^2).$$

Since A_1 "multiplies" measure by π, its power $A_1^{\times(N)}$ multiplies measure by π^N, that is

$\star_N$ *the map A_N pushes forward the Euclidean (Lebesgue) measure on $\mathbb{C}^N$ to π^N times the Euclidean (Lebesgue) measure on $\mathbb{R}_+^N$.*

Next, observe that A_N sends the unit ball $B^{2N} \subset \mathbb{C}^N = \mathbb{R}^{2N}$ to the "rectangular" simplex

$$\{x_0 \geq 0, x_1 \geq 0, ...x_N \geq 0, \sum_i |x_i| \leq 1\} \subset \mathbb{R}_+^N \subset \mathbb{R}^N$$

which equals the positive part of the $(N+1)$-diamond from the previous section, and, accordingly denoted by

$$\blacklozenge_+^N = \blacklozenge^N \cap \mathbb{R}_+^N \subset \mathbb{R}^N.$$

Thus, for instance, the relation $\star_N$ yields the usual formula for the volumes of even dimensional balls: $vol(B^{2N}) = \pi^N vol(\blacklozenge_+^N) = \pi^N / N!$.

This, in turn, gives the formula for odd dimensional spheres, for $vol(S^{n-1}) = n \cdot vol(B^n)$ for all n, odd or even. But $\star_N$ does not cover the Archimedean case of (even!) $n = 2$.

To this end, restrict A_{N+1} the unit sphere in $\mathbb{C}^{N+1}$ and observe that this sphere S^{2N-1} goes to the regular N-simplex $\blacktriangle^N \subset \mathbb{R}_+^{N+1}$, defined by $\sum_i x_i = 1$ in $\mathbb{R}_+^{N+1}$ and that, because of $\star_N$, the map A_N pushes forward the spherical measure on S^{2N-1} to π^N times Lebesgue measure on the N-simplex.

It is worth observing at this point that the map $A_{N+1} : S^{2N-1} \to \blacktriangle^N$ can be seen as the quotient map $S^{2N-1} \to S^{2N-1}/\mathbb{T}^{N+1} = \blacktriangle^N$ for the $(N+1)$-torus which acts on $\mathbb{C}^{N+1}$, hence on $S^{2N-1} \subset \mathbb{C}^{N+1}$, by multiplication of the coordinates by complex numbers τ_i with norm one,

$$(z_0, z_1, ...x_N) \mapsto (\tau_0 z_0, \tau_1 z_1, ...\tau_n z_N).$$

Thus, A_{N+1} factors via the Hopf quotient map

$$S^{2N-1} \to S^{2N-1}/\mathbb{T}_\circ^1 = \mathbb{C}P^N \to \blacktriangle^N = \mathbb{C}P^N/\mathbb{T}^N$$

for $\mathbb{T}_\circ^1 \subset \mathbb{T}^{N+1}$ being the diagonal circle in the torus which acts on $\mathbb{C}^{N+1}$ by:

$$(z_0, z_1, ...x_N) \mapsto (\tau z_0, \tau z_1, ...\tau z_N).$$

If $n = 2$, then (the complex projective line) $\mathbb{C}P^1 = S^3/\mathbb{T}_\circ^1$ naturally identifies with the two sphere S^2 where the circle acts by rotation around an axes and where the Archimedean appears as the quotient map $S^2 \to S^2/\mathbb{T}^1$.

The naturality/symmetry of the above maps shows that, indeed, the spherical measure on S^2 goes to a multiple of the Lebesgue measure on $[-1,1] = S^2/\mathbb{T}^1$, where evaluation of this "multiple" needs that for the Hopf map between the *unit* spheres $S^3 \to S^2$.

To do this notice that the diameter of S^2 with the quotient metric is equal to $\pi/2 = \frac{1}{2}diam(S^3) = \frac{1}{2}diam(S^2)$, since the pullbacks of opposite points in S^2 are orthogonal as vectors in $\mathbb{C}^2 \supset S^3$.

Hence, $area(S^3/\mathbb{T}^1_\circ) = \frac{1}{4}area(S^2)$, and since $vol(S^3) = 2\pi \cdot area(S^3/\mathbb{T}^1_\circ)$, we come up with the Archimedean value for the area of the unit 2-sphere:

$$area(S^2) = \frac{4}{2\pi}vol(S^3) = \frac{4}{2\pi}\cdot\frac{4\pi^2}{2} = 4\pi.$$

Exercises.(i) Show that the pullbacks $A^{-1}(x) \in \mathbb{C}^N$, $x \in \mathbb{R}^{N+1}_+$, are tori of dimensions $\le N+1$, where, actually, $dim(A^{-1}(x))$ is equal to the number of non-zero components in the vector $x = (x_0, x_1, ...x_N)$.

(ii) Show that the N-dimensional volume of $A^{-1}(x)$ for $x \in \blacktriangle^N \subset \mathbb{R}^{N+1}_+$ – such a torus lies in $S^{2N-1} \subset \mathbb{C}^N$ – is bounded by $\left(\frac{2\pi}{N}\right)^N$.

(iii) Show that the tori $A^{-1}(x)$ are ortogonal to $\mathbb{R}^{N+1} \subset \mathbb{C}^{N+1} = \mathbb{R}^{N+1} \oplus \sqrt{-1}\mathbb{R}^{N+1}$

(iv) Show that the interior of the positive cone $\mathbb{R}^{N+1}_+ \subset \mathbb{R}^{N+1}$, call it $\mathbb{R}^{N+1}_{++} \subset \mathbb{R}^{N+1}_+$, meets each torus ar a single point; thus, the region $\mathcal{F} \subset \mathbb{C}^N$, where the action of $\mathbb{T}^{N+1}$ is free, naturally decomposes into the product , $\mathcal{F} = \mathbb{R}^{N+1}_{++} \times \mathbb{T}^{N+1}$.

Two Words about Moment Maps. Apparently –this is suggested by how his theorem is usually depicted[52] Archimedes himself had not visualised his theorem by mentally focusing at the image of the normal projection $S^2 \to [-1,1]$, but he saw it more geometrically and more informatively in the light of the radial, with respect to the vertical axes, projection from the sphere to the circumscribed cylinder $S^1 \times [-1,1]$ for S^1 being the (unit) equatorial circle in the sphere.

The radial projection map $S^2 \to S^1 \times [-1,1]$ is area preserving.

Yet another way to think of the Archimedes theorem – this was explained to me by Michael Atiyah many years ago – is as of the *moment map* for the (area preserving!) action of $\mathbb{T}^1$ on S^2, where the circle $\mathbb{T}^1$ acts on the sphere $S^2 \subset \mathbb{R}^3$ by rotation around the vertical axes.

We shall explain this in ???; now let us return to our main topic and look more closely at the geometry of the map (which is also an instance of a moment map) $A_{N+1} : S^{N-1} \to \blacktriangle^N$ for large N.

$|\bullet|$ Predominant Distance Contraction by the Archimedes Map. *The relative volume of the subset $\Delta = \Delta_N \subset \blacktriangle^N$, such that the the map $A = A_{N+1} : S^{N-1} \to \blacktriangle^N$, for $(z_i) \mapsto (|z_i|^2)$, is $\frac{\log_2 N}{\sqrt{N}}$ distance decreasing over Δ tends to one for $N \to \infty$.*

Namely,

for every $\varepsilon > 0$ and all sufficiently large N, the inequality

$$dist(A(s_1), A(s_2)) \le \frac{\log_2 N}{\sqrt{N}} dist(s_1, s_2)$$

[52] According to Cicero, a sphere with the circumscribed cylinder was surmounted on Archimedes' tomb in Syracuse.

is satisfied for all s_1, s_2 *in the pullback* $A^{-1}(\Delta) \subset S^{2N-1}$ *of a subset* $\Delta \subset \blacktriangle^N$, *such that*

$$\frac{vol(\Delta)}{vol(\blacktriangle^N)} \geq 1 - \varepsilon,$$

where, recall, $vol(\Delta)/vol(\blacktriangle^N) = vol(A^{-1}(V))/vol(S^{2N-1})$ *by the Archimedes' theorem.*

Proof. Take the intersection of $\blacktriangle^N$ with the $(N+1)$-cube $[-\delta^2, \delta^2]^{N+1} \subset \mathbb{R}^{N+1}$ for $\Delta = \Delta_\delta \subset \blacktriangle^N$ and observe the following.

$\bullet_1$ *The norm of the differential* dA *over* Δ_δ *is* $\leq 2\delta$.

In fact, the A-pullback of Δ_δ is defined by the inequalities $|z_i| \leq \delta$ and since dA is given by the diagonal matrix with the entries $2z_0, 2z_1, ... 2z_N$.

$\bullet_1'$ *The bound on* $\|dA(s)\|$ *on* $\tilde{\Delta} = A^{-1}(\Delta)$ *implies the same bound on the dilation (Lipschitz constant) of* A:

$$dist(A(s_1), A(s_2)) \leq \left(\sup_{s \in \tilde{A}} \|dA(s)\|\right) \cdot dist(s_1, s_2) \text{ for all } s_1, s_2 \in \tilde{\Delta}.$$

This would be automatic if the set $\tilde{\Delta} \subset S^{2N-1} \subset \mathbb{C}^N$ were convex but it is not even geodesically convex in the sphere S^{2N-1}.

However, albeit straight segments $[s_1, s_2] \subset \mathbb{C}^N \supset \tilde{\Delta}$ are not necessarily contained in $\tilde{\Delta}$, the norm of dA on such a segment is bounded by

$$\max(\|dA(s_1)\|, \|dA(s_2)\|) \leq \sup_{s \in \tilde{A}} \|dA(s)\|$$

by the convexity of the function $max_i \|z_1\|$. Hence, the length of the curve $\gamma = A[s_1, s_2] \subset \mathbb{R}_+^{N+1}$, which joins $x_1 = A(s_1)$ with $x_2 = A(s_2)$ is bounded by

$$\sup \|dA(s)\| \cdot dist(s_1, s_2)$$

and since, obviously,

$$dist(x_1, x_2) \leq length(\gamma),$$

this distance is bounded by $\sup \|dA(s)\| \cdot dist(s_1, s_2)$.

$\bullet_2$ *If* $\delta = 2\lambda/\sqrt{N+1}$, *then the relative volume of the complement* $\blacktriangle^N \setminus \Delta$ *is bounded by* $4(N+1)2^{-\lambda^2}$.

Indeed, the A-pullback of this complement in the $(2N-1)$-sphere[53] equals the union of $N+1$ regions in the sphere, which are defined by $|z_i| \geq \delta$ and where each of them is contained in the union of four spherical caps, defined by $|x_i| \geq \delta/\sqrt{2}$ and $|y_i| \geq \delta/\sqrt{2}$ for $x_i + y_i\sqrt{-1} = z_i$.

According to the "spherical remark" after $[\bullet_{\lambda/\sqrt{N}}]$ in section 1.3, the volumes of these caps are bounded by $2^{-\lambda^2}$, which yields $\bullet_2$.

Now $|\bullet|$ trivially follows from the above for $\Delta = \Delta_\delta$ with $\delta = \frac{\log_2 N}{2\sqrt{N}}$.

Remark/Questions. The above estimate on the volume of the complement to $\Delta_\delta \subset \blacktriangle^N$ can be improved a little, but it remains unclear if our $A : S^{2N-1} \to \blacktriangle$

[53]One doesn't have to go to the sphere: the relative volume estimates one needs is as easy (slightly easier) to make in $\blacktriangle$ as in the sphere.

is the *optimal map*, in the sense that it has *maximal* distance contraction (i.e. minimal Lipschitz constant) among all *measure preserving* maps $S^{2N+1} \to \blacktriangle$.

In general, it may be interesting to look for measure preserving maps, similar to the Archimedean A, between standard high dimensional Euclidean (and not only Euclidean) domains and/or natural measures, where these maps would be as contracting as possible away from small parts of these domains.

Exercises. (a) Construct a (non-strictly) distance decreasing map from the unit N-sphere to a real segment, $f : S^N \to [0, \delta]$, for

$$\delta = \frac{vol(S^N)}{vol(S^{N-1})} \sim \frac{\pi}{\sqrt{N}},$$

such that f pushes forward the normalised spherical measure to the normalised Lebesgue measure on the segment $[0, \delta]$.

Hint compose the normal projection $S^N \to [-1, 1]$ with a suitable distance decreasing map $[-1, 1][0, \delta]$.

(b) Given $n \leq N$, construct a (non-strictly) distance decreasing map from the unit N-sphere to the n-ball B, $n \leq N$, such that

$$vol(B) = \frac{vol(S^N)}{vol(S^{N-n})},$$

and such that this map pushes forward the normalised spherical measure to the normalised Lebesgue measure on B.

Remark. One can show[54] that this is sharp: there is no such maps from S^N to the ball B^n with $vol(B^n) > \frac{vol(S^N)}{vol(S^{N-n})}$.

Symmetry and Entropy: Archimedes → Boltzmann → Fisher.

By the above (iv), the quotient space $\mathbb{C}^{N+1}/\mathbb{T}^{N+1}$ can be identified with the positive cone $\mathbb{R}_+^{N+1}$, and the complex Archimedean map $A : \mathbb{C}^{N+1} : \mathbb{C}^{N+1} \to \mathbb{R}_+^{N+1}$ factors through the real one,

$$\mathbb{R}A : \mathbb{R}_+^{N+1} \to \mathbb{R}_+^{N+1} \text{ for } (x_0, x_1, ..., x_N) \mapsto (x_1^2, x_2^2, ..., x_N^2).$$

Since the quotient metric in $\mathbb{C}^{N+1}/\mathbb{T}^{N+1}$ is equal to the standard Euclidean metric in $\mathbb{R}_+^{N+1} = \mathbb{C}^{N+1}/\mathbb{T}^{N+1}$, the Riemannian metric in the receiving positive cone $\mathbb{R}_+^{N+1}$ which is induced from $\mathbb{C}^{N+1}$ equals the transport of the Euclidean (regarded as Riemannian) metric $\sum_i dx_i^2$ from $\mathbb{R}_+^{N+1} = \mathbb{C}^{N+1}/\mathbb{T}^{N+1}$ to $\mathbb{R}_+^{N+1} = \mathbb{R}A(\mathbb{R}_+^{N+1})$ by the map $\mathbb{R}A$, where this metric is

$$\mathbb{R}A_*(g_{Eu}) = \sum_{i=0}^{N} \frac{1}{2x_i} dx_i^2,$$

since the differential of the map $\mathbb{R}A$ is given by the diagonal matrix with the entries $2x_i = \frac{dx_i^2}{dx_i}$.

Remarkably – only exceptional quadratic forms have this property –

[54]This follows from a lower bound on the *waist* of the sphere, see `https://pdfs.semanticscholar.org/9e73/ae93261043a0c721f77e631a130a78daf379.pdf` and references therein.

there exists a smooth function, say Φ *on* $\mathbb{R}_+^{N+1}$, *the Hessian of which is equal to the (differential) quadratic form* $\mathbb{R}A_*(g_{Eu})$,

$$\frac{\partial_{ij}\Phi}{dx_i dx_j} dx_i dx_j = \mathbb{R}A_*(g_{Eu}).$$

In fact, the function $\frac{1}{2}\sum_i x_i \log x_i$can be taken for Φ, since $\frac{d^2}{dx^2}(x\log x) = \frac{1}{x}$.
It follows that
the entropy function on the simplex $\blacktriangle^N \subset \mathbb{R}_+^{N+1}$,

$$ent(x_0, x_1, ...x_N) = -(x_0 \log x_0 + x_1 \log x_1 + ... + x_N \log x_N)$$

is concave, where, moreover - this, apparently, goes back to Ronald. Fisher–

$$Hess(ent)) = -2\mathbb{R}A_*(g_{Eu}).$$

In particular, the Riemannin metric defined by $-Hess(ent)$ *has constant positive curvature.*[55]

It is also clear that the entropy is *the only* function which satisfies this equality and vanishes at the vertices $(1, 0, ...0)$, $(0, 1, ...0)$,,,, $(0, 0, ...1)$ of the simplex $\blacktriangle^N$. Therefore one can take the equality $Hess(ent)) = -2\mathbb{R}A_*(g_{Eu})$ for the definition of entropy[56] entropy by Boltzmann-Shannon *computational*(!) formula $-\sum_i p_i \log p_i$ is as inappropriate as *defining* (rather than *computing*) the area of the disk as the limit of the areas of the inscribed regular n-gons. (The original Boltzmann's definition, translated to the 21st century mathematical language, is described in "*search for a structure-entropy*" on my page https://cims.nyu.edu/ gromov/).

Conversely, by taking the Hessian of $\sum_i x_i \log x_i$, one arrives from the clumsy simplex to the beautifully round sphere with the huge symmetry group.

(This, possibly, may explain the "unreasonable effectiveness" of entropy in mathematical physics and in math generated by physics, and which points toward "quantum nature" of entropy. But exactly this beautiful hidden symmetry makes one wary of transplanting the idea of entropy from physics to mathematical models of Life.)

1.9 Selected References to Chapter 1

[Blessing of dimensionality] A.N. Gorban, I.Y. Tyukin, Blessing of dimensionality: mathematical foundations of the statistical physics of data.
https://arxiv.org/abs/1801.03421

[Concentration of Measure] Michel Ledoux, The Concentration of Measure Phenomenon. American Mathematical Society (2001).

[Concentration Property] A. A. Giannopoulos and V. Milman, Concentration property on probability spaces, Advances in Mathematics 156 (2000), 77-106.
https://core.ac.uk/download/pdf/82426777.pdf

[Elementary Introduction] Keith Ball, *An Elementary Introduction to Modern Convex Geometry.* library.msri.org/books/Book31/files/ball.pdf

[55] A direct computation of the curvature tensor of $Hess(ent))$ seems cumbersome and I don't know whether Fisher proved that $curv(-\frac{1}{2}Hess(ent)) = +1$ this way.

[56] *Defining*

[High Dimensional Probability] Roman Vershynin, *High Dimensional Probability: An Introduction with Applications in Data Science.*
www-personal.umich.edu/~romanv/papers/HDP-book/HDP-book.pdf

[Metric Structures] M. Gromov, *Chapter $3\frac{1}{2}$ in Metric structures for Riemannian and non-Riemannian spaces.* Birkhauser, Boston, 1999.

[Probability in High Dimension] Ramon van Handel, *Probability in High Dimension.*
APC 550 Lecture Notes Princeton University.
https://web.math.princeton.edu/~rvan/APC550.pdf

[Sparse Distributed Memory] Pentti Kanerva, *Sparse Distributed Memory and Related Models*
https://pdfs.semanticscholar.org/5dee/60f4618b6e24771380b2004bc4f9db5dfbc7.pdf

2 Compression versus Concentration.

2.1 Comparison between Concentrations.

A. *Definitions.* Given two probability measures on metric spaces, μ_1 on X_1 and μ_2 on X_2, say that that

μ_1 is more concentrated at a point $x_1 \in X$ than μ_2 at $x_2 \in X_2$,

if the measures of the r-balls in these spaces with the centers at the points x_1 and x_2 satisfy

$$\mu(B_{x_1}(r)) \geq \nu(B_{x_2}(r))$$

for all $r \geq 0$.

In general, when we speak in this context of concentrations of *non-probability* measures μ of finite mass, we automatically normalise the measures. Thus, our comparison inequality becomes

$$\frac{\mu_1(B_{x_1}(r))}{\mu_1(X_1)} \geq \frac{\mu_2(B_{x_2}(r))}{\mu_2(X_2)}.$$

Concentration of Functions and Maps. Concentration of a map $f : X \to Y$, at a particular value $y = f(x)$, where $X = (X, \mu)$ is a measure space and Y is a metric space, e.g. $Y = \mathbb{R}$, means, by definition, concentration of the push-forward measure $f_*(\mu)$ at y in Y.

This comparison relation between μ_1 and μ_2 entirely depends on the push-forwards of these measures to the the half line $\mathbb{R}_+ = [0, \infty)$ by the maps

$$d_1 : X_1 \to \mathbb{R}_+ \text{ and } d_2 : X_2 \to \mathbb{R}_+$$

for $d_1 : x \mapsto dist_{X_1}(x, x_1)$ and $d_2 : x \mapsto dist_{X_2}(x, x_2)$, denoted $\mu_{1*} = (d_1)_*(\mu_1)$ and $\mu_{2*} = (d_2)_*(\mu_2)$. Namely

μ_1 is is more concentrated at a point x_1 in X_1 than μ_2 at x_2 in X_2 if and only if the function d_1 is more concentrated at the value 0 than d_2, that is the measure μ_{1*} is more concentrated at 0 in $\mathbb{R}_+$ than μ_{2_*}.

MEASURES ON THE HALF LINE.

B. *Concentration comparison* between probability measures on half lines, call them now μ_1 on $[x_1, \infty)$ and μ_2 on $[x_2, \infty)$ can be implemented by maps $f : [x_1, \infty) \to [x_2, \infty)$. Namely,

$(*)$*if μ_1 is more concentrated at x_1 than μ_2 at x_2, then, provided μ_2 has no atoms, there exists a monotone increasing map $f : [x_2, \infty) \to [x_1, \infty)$, which pushes forward the measure μ_2 to μ_1,*

$$f_*(\mu_2) = \mu_1$$

and which shrinks the segments $[x_2, x]$, that is

$$f(x) - x_1 \leq x - x_2$$

for all $x \geq x_2$.

Proof. Let the values of the function $f : [x_2, \infty) \to [x_1, \infty)$ at all $x \in [x_2, \infty)$ be defined as the the infima of $x' \geq x_1$, such that

$$\mu_1[x_1, x'] \geq \mu_2[x_2, x]$$

and observe that if μ_2 is atomless μ_1 is more concentrated than μ_2 than, indeed, $f_*(\mu_1) = \mu_2$ and $f(x) - x_1 \leq x - x_2$.

Exercises. (a) Let μ and ν be probability measures without atoms and having connected supports. Then there exists a unique monotone increasing homeomorphism f from the support of μ to the support of ν, such that $f_*(\mu) = \nu$ (with an obvious convention in the case where μ has finite support and *nu* an infinite one. some of the supports is infinite).

(Thus, to a horror of a statistician, an arbitrary probability distribution on the line with strictly positive measurable density can be made normal by a mere change of variable.[57])

(b) Formulate and prove a version of $(*)$ for purely atomic measures μ_1 and μ_2.

(C) Let $\mu_1 = \phi_1(x)dx$ and $\mu_2 = \phi_2(x)dx)$ be probability measures on $[0.\infty$ with continuous density functions $\phi_1(x)$ and $\phi_2(x)$. Notice that since

$$\int_0^\infty \phi_1(x)dx = \int_0^\infty \phi_2(x)dx \ (= 1),$$

the the difference $\phi_1(x) - \phi_2(x)$, unless it is identically zero, must somewhere change sign and let it *change sign at most once.* This means that subset where the functions are strictly positive and mutually equal is connected. (Typically there would be a single such point.)

In this case – it is (almost) 100% obvious –

one of the two measures is more concentrated at 0 *than the other.*

For instance,

if $\phi_1(0) > \phi_2(0)$ than μ_1 is more concentrated than μ_2.

In general

if $\phi_1(x) = \phi_2(x)$ up to a point x_0, and $\phi_1(x) > \phi_2(x)$ for $x = x_0 + \varepsilon$, then the same conclusion holds: μ_1 is more concentrated at 0 *in $[0, \infty)$ than μ_2.*

[57]This is not a laughing matter, at least not in biology and psychology, where numerically expressed features may have no *natural* parameters attached to them.

$\left[\text{从}\right]$ *Corollary.* If the functions ϕ_1 and ϕ_2 are differentiable and

$$\frac{d\phi_1(x)}{dx} < \frac{d\phi_2(x)}{dx},$$

whenever $\phi_1(x) = \phi_2(x)$, then the measure $\mu_1 = \phi_1(x)dx$ is more concentrated at 0 than $\mu_2 = \phi_2(x)dx$.

2.2 Convexity and Concentration

The bulk of published results on *concentration of measure*[58] explicitly or implicitly relies on the concept of convexity, with the motto

the more concave you are the more concentrated you are.

Below are simple examples of this.

A. Let $\mu_1 = \phi_1(x)dx$ and $\mu_2 = \phi_2(x)dx)$ be probability measures on $[0,\infty)$ with continuous density functions $\phi_1(x)$ and $\phi_2(x)$ which are C^2-differentiable on the (finite or infinite) segments $[0, a_1)$ and $[0, a_2)$ correspondingly and vanish outside these segments, i.e. $\phi_i(x) = 0$ for $x \geq a_i$, $i = 1, 2,$. Also let

$$\frac{d\phi_1(0)}{dx} \leq 0 \text{ and } \frac{d\phi_2(0)}{dx} \geq 0.$$

Let one the following ???? inequalities on the second derivatives be satisfied for all $x \in [0, \min(x_1, x-2)]$.

$$[\mathsf{A}_1] \qquad \frac{d^2\phi_1(x)}{dx^2} \leq \frac{d^2\phi_2(x)}{dx^2},$$

$$[\mathsf{A}_2] \qquad \frac{d^2\log\phi_1(x)}{dx^2} \leq \frac{d^2\log\phi_2(x)}{dx^2},$$

$$[\mathsf{A}_3] \qquad d\phi_2(x)dx \leq 0\frac{d}{dx}\log - \frac{d\phi_1(x)}{dx} \leq \frac{d}{dx}\log - \frac{d\phi_2(x)}{dx},$$

Then *the measure* $\mu_1 = \phi_1(x)dx$ *is more concentrated at* 0 *than* $\mu_2 = \phi_2 dx$.

Proof. All three inequalities imply that the first derivative of ϕ_1 decay faster than that of ϕ_2. Therefore, when the two functions meet at a point $x_{\text{从}} \geq 0$, i.e. where

$$\phi_1(x_{\text{从}}) = \phi_1(x_{\text{从}}),$$

the derivative of ϕ_1 will be smaller than that of ϕ_2,

$$\frac{d\phi_1(x_{\text{从}})}{dx} \leq \frac{d\phi_2(x_{\text{从}})}{dx},$$

and the proof follows by the above $\left[\text{从}\right]$. (One needs a minor additional effort to handle the case where $\frac{d\phi_1(x_{\text{从}})}{dx} \leq \frac{d\phi_2(x_{\text{从}})}{dx}$, which is not formally covered by the strict "$<$" in $\left[\text{从}\right]$.)

[58]See `https://en.wikipedia.org/wiki/Concentration_of_measure` and references therein

*Regularity Remark.*The above 3 conditions which, a priori, need both functions to be twice differentiable, in fact, makes sense if ψ_1 is continuous and ψ_2 is C^2-smooth. But, to simplify, we always assume our function C^2-differentiable.

In Praise for log-Convexity. A positive function ϕ on an affine, e.g. Euclidean, space is log-convex/log-concave if $\log f$ is convex/concave. Thus, log-convexity means that

$$\phi\left(\frac{x_1+x_2}{2}\right) \le \sqrt{\phi(x_1)\phi(x_2)}$$

and log-concavity says that

$$\phi\left(\frac{x_1+x_2}{2}\right) \ge \sqrt{\phi(x_1)\phi(x_2)}$$

Then the second derivative

$$\frac{d^2 \log phi(x)}{dx^2}$$

serves as a measure of log-convexity/concavity:

the greater this derivative is the more $\phi(x)$ is log-convex and less log-concave.

This is well adapted to the study of concentration of densities ϕ of measures $\mu = f(x)dx$, since both: concentration and $\frac{d^2 \log phi(x)}{dx^2}$ is invariant under multiplications of ϕ by a constant.

B. *Concentration of* $(\sqrt{1-x^2})^{N-1}dx$ *Revisited.* (Compare section 1.3.) Since the second derivative of the logarithm of the function $\phi_N = (\sqrt{1-x^2})^{N-1}$ satisfies

$$\frac{d^2 \log \phi_N(x)}{dx^2} = -\frac{N-1}{2}\left(\frac{2}{1-x^2}+\frac{4x^2}{(1-x^2)^2}\right) \le -(N-1) = \frac{d^2 \log \exp - \frac{(N-1)x^2}{2}}{dx^2},$$

the measure $\phi_N dx$ is more concentrated at 0 than the Gaussian measure $\exp - \frac{(N-1)x^2}{2} dx$ by the above $[\mathsf{A}_2]$.

Therefore, the coordinate function $(x_1, ..., x_i, ...x_N) \mapsto x_1$ on the unit ball

$$B^N = \{x_1, ..., x_i, ..., x_N\}_{\sum_i x_i^2 \le 1} \subset \mathbb{R}^N$$

with the usual (Lebesgue's) measure $dx_1 dx_2 ... dx_N$ is more concentrated than $\exp - \frac{(N-1)x^2}{2} dx$.

2.3 Convexity and Levy-Milman Concentration of Lipschitz Functions.

The Levy mean of a function $f : X \to \mathbb{R}$ on a measure space X with finite mass $M = \mu(X)$ is the value $y \in \mathbb{R}$, such that, essentially, the y-level $f^(y) \subset X$, divides X into *equal* halves. More precisely, both, the sub-level and the sup-level of f, have measures $\ge \frac{M}{2}$,

$$\mu(f^{-1}(-\infty, y]) \ge \frac{M}{2} \text{ and } \mu(f^{-1}[y, \infty)) \ge \frac{M}{2}.$$

(This is equivalent to $\mu(f^{-1}(-\infty,y]) = \mu(f^{-1}[y,\infty))$ in the typical case of $\mu(f^(y) = 0.)$

This concept was introduced in 1951 book by Paul Levy who has made the following key observation.

The isoperimetric inequality in the sphere $S^{N-1}\subset\mathbb{R}^N$ implies that

$\ast$ *all 1-Lipshitz function $f: S^{N-1}\to\mathbb{R}$ are more concentrated at their Levy means, than the signed geodesic distance function to the equator $S^{N-2}\subset S^{N-1}$.*[59]

To make it clear, recall, that

(i) *the signed distance* is the function $x\mapsto\pm dist(x, S^{N-2})$, $x\in S^{N-1}$, with the "+"sign in the upper hemisphere and with "−" in the lower hemisphere.

(ii) *The isoperimetric inequality in S^{N-1}* says that

$\star$ *among all domains $V\subset S^{n-1}$ with a given $(n-1)$-volume the geodesic balls have the minimal $(n-2)$-volumes of their boundaries.*

There several interpretation of this for domains with non-smooth boundaries ∂V depending on how the $(n-2)$-volume of ∂V is defined.

What we need here is the following rendition of $\star$ which entirely relies on the spherical measure in S^{n-1} with no reference to any $(n-2)$-measure.)

$\star_\varepsilon$ *among all domains $V\subset S^{n-1}$ with a given $(n-1)$-volume the geodesic balls have the minimal $(n-1)$-volumes of the ε-neighbourhoods of V for all $\varepsilon\geq 0$.*

Notice, that $\star_\varepsilon$ for $\varepsilon\to 0$ implies $\star$ with $(n-2)$-volume understood as *Minkowski content* of the boundary.

Conversely, $\star$, interpreted as $\star_\varepsilon$ with infinitesimal ε, (almost) obviously integrates to $\star_\varepsilon$ for all ε.

Granted this, we see that the distance function to an arbitrary subset $V\subset S^{N-1}$ of half measure is more concentrated than the distance to a hemisphere, and then $\ast$ easily follows.

2.4 Concentration for maps to Metric Spaces of Dimension>1

Such concentration is known for maps from S^{N-1}, but not, for instance. from hypersurfaces with all principal curvatures > 1

A. Convexity and concentration,
Levy-Milman-Talagrand
Topology and measure
Questionable Meaning of Probability And Measure in High Dimensions.

[59]Apparently, this was the first instance of the relation between isoperimetric and functional inequalities which was explicitly stated and proved. This idea became widely used in geometry and analysis since the contributions by Maz′ya (1960), Federer-Fleming (1960) and by Cheeger (1970).

3 Polypeptide and Protein Spaces.

Polypeptides are molecular chains, think of them as "words" in an alphabet of 20 "letters" – molecules of *amino acids*.[60]

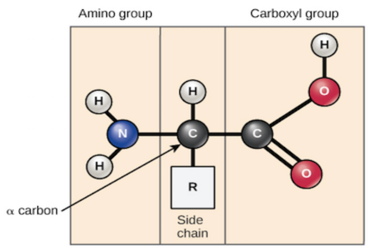
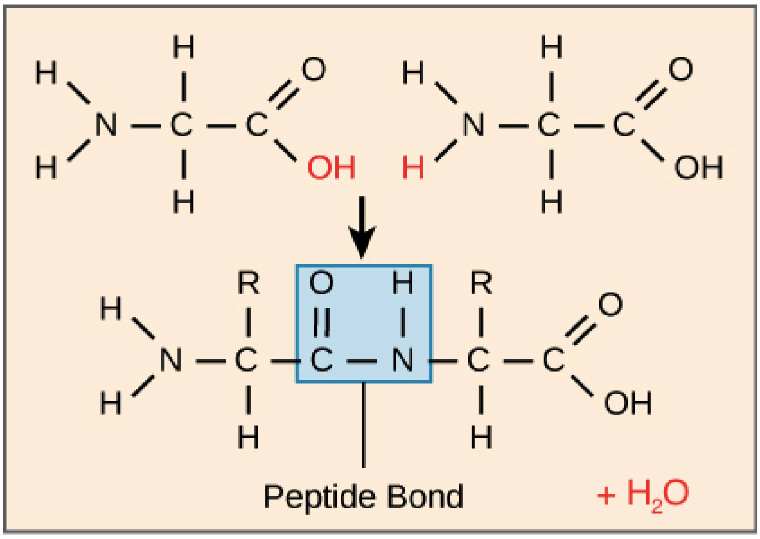
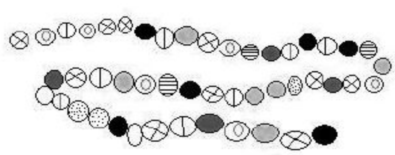


Figure 2: generic amino acid





A chain of amino acids

In a friendly (water + something) environment, polypeptides turn into proteins by acquiring a particular spacial structure by the process of *protein folding*.

[60]No statement in biology comes without an exception. For instance, there are two rare genetically-encoded amino acids *selenocysteine* (discovered in 1950s) and *pyrrolysine* (discovered in 2002 in archaea and bacteria). See `https://en.wikipedia.org/wiki/Selenocysteine`
`http://www.pitt.edu/~koide/group/Selenocyst-AG.pdf`
`https://en.wikipedia.org/wiki/Pyrrolysine`

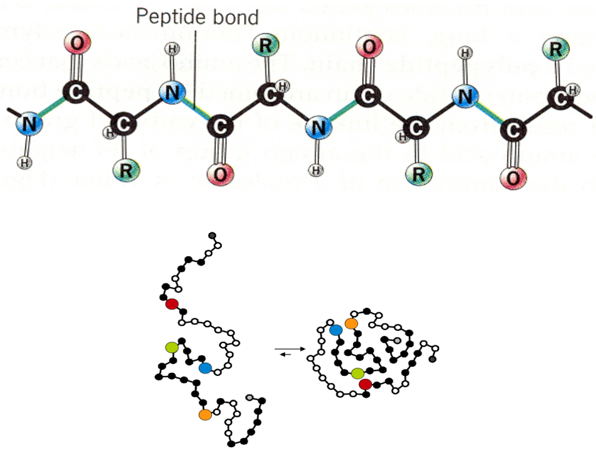


Figure 3: protein folding and unfolding

The length of proteins found in cells ranges between 40 and 40 000 "laters",[61] while most "easily foldable" *globular* (potato shaped) proteins are composed of 100-300 *amino acid residues.*[62]

At this point, don't get depressed by your non-understanding of the physical/chemical meaning of the words "protein", "amino acid" etc,[63] – the time for worry is ahead of you – but attempt to *formulate* the Protein Problem(s)" in (quasi)mathematical terms being content (at this stage only!) with a minimal input from biology.

Direct your thinking of proteins toward a subset $\mathcal{P}$ in the set of all sequences of length 300 in 20 symbols,

$$\mathcal{P} \subset \mathbf{20}^{300},$$

try to define what this $\mathcal{P}$ could conceivably be and what kind of properties it makes sense to ask about it. [64]

Biologists think of $\mathcal{P}$ as of "Protein Universe" – the set of all proteins of all organisms and ask the following questions.

[61]The longest known proteins, – *titin* or *connectin*, which is the third most abundant in human muscle, is composed of ≈ 30 000 amino acids. `https://en.wikipedia.org/wiki/Titin`

The actual "physical" length of titin, if a molecule is stretched, is more than a micron, and its full chemical name, if written in English, would go in more than 180 000 letters.

No "natural" English word, however, exceeds 30 letters in length, but *agglutinative languages*, such, for instance as *Afrikaans, German, Tagalog, Turkish*, may, in principle, contain arbitrary long words, where the longest recorded one is a 136 letter word in Afrikaans which can be translated as

issuable media conference's announcement at a press release regarding the convener's speech at a secondhand car dealership union's strike meeting.

See `https://en.wikipedia.org/wiki/Longest_words`
`https://www.quora.com/Which-language-uses-the-longest-words-in-their-daily-speech-vocabulary`

[62]Polypeptide synthesis is accompanied by *dehydration* – loss of a few atoms from amino acids molecules in the form of water; what remain of these molecules in the polypeptide are called *residues*.

[63]What is really depressing is the difficulty of specifically articulating what you don't understand. Saying I know that I know nothing is senseless if you don't know what stands behind this "nothing".

[64]Limiting to 300 doesn't seem to change the overall picture

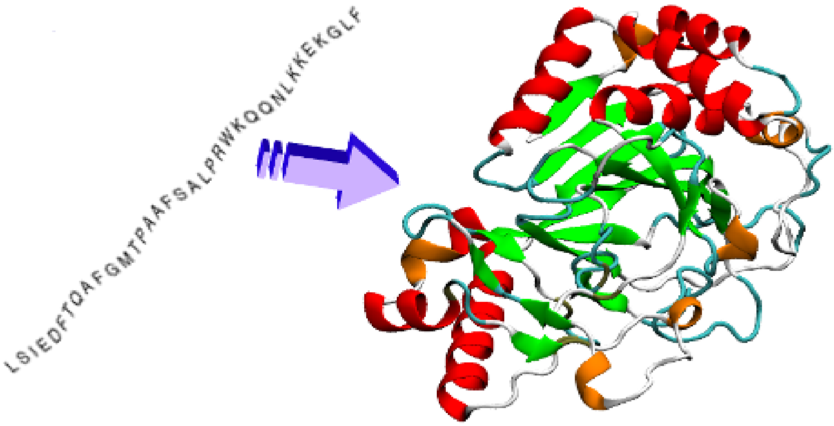


Figure 4: Protein Folding and Unfolding

How many protein sequences are there? How many sequences are novel vs. repetitious? How many sequences are characterised at structural and functional levels? Are sequences of prokaryotes, eukaryotes, and viruses different? Is the number of sequence families saturating or is it still expanding rapidly?"[65]

The first question we, mathematicians, understand; a preliminary answer is $\approx 5 \cdot 10^7$ - the number of protein sequences in the data banks.[66]

From another angle: one estimates the number of *non-bacterial* species of organisms on Earth by 10^7, while the number of different bacteria may be in the range 10^9-10^{12} with, possibly, a comparable number of different kinds of viruses. [67]

Since bacteria have 1-2 thousand of protein coding genes, there may be 10^{12}-10^{15} different proteins on in living organisms on Earth, and, probably, in the range below $10^{20} << 20^{100} > 10^{130}$ even if you count the, by now extinct, organisms that have ever lived on Earth.

All in all, we are faced with the problem:

describe in *mathematical terms* a given subset $\mathcal{P} \subset \mathbf{20}^{300}$
of the size between 10^{10} and 10^{20}.

To start thinking of this, we need first to clarify the meaning of "a given" in "a subset" and then specify "mathematics" we are allowed to use.

The problem with "a" is that $\mathcal{P}$ – the protein universe – is not, strictly

[65] *Nature of the protein universe* by M. Levitte, PNAS, 2009,
`https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2698892/`

[66] https://www.gqlifesciences.com/the-largest-public-or-private-biological-sequence-database-on-earth/

[67] Here is what I found about it on the web:
https://www.livescience.com/54660-1-trillion-species-on-earth.html
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC539005/
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3160642/
https://www.sciencedaily.com/releases/2011/08/110823180459.htm
https://mbio.asm.org/content/7/4/e00999-16
http://www.virology.ws/2013/09/06/how-many-viruses-on-earth/

speaking "a set", it is not described in terms easily translatable to mathematical language and it is by no means "given" to us. The only thing we truly (modulo errors) know about $\mathcal{P}$ is what is contained in data banks, say a sample $\mathcal{S} \subset \mathcal{P}$, of size 10^8 of $\mathcal{P}$. Apart from this we have no and will never have knowledge of when a given sequence p is contained in $\mathcal{P}$.

In fact, we do not expect from a mathematical description, better to say model $\mathcal{MP}$, of $\mathcal{S} \subset \mathcal{P}$ to deliver such knowledge, since the actual $\mathcal{P}$ is the result of billions historical *accidents*,. This renders a mathematical description of $\mathcal{P}$, even is an approximate one, unrealistic, since the construction/description of a desired model $\mathcal{MP}$ must (mostly) rely on *general principles* (whatever they can be).

To better appreciate illogic of $\mathcal{P}$, see how it is with languages:

"define" the set $\mathcal{R}$ of grammatically correct and meaningful sentences in twenty words in English that have been and will have been written, say in the four hundred year interval between July 5, 1687[68] and July 5, 2087.

A simple counting shows that the set $\mathcal{R}$ can't contain more than 10^{15} entries[69] while the number of all such *conceivable* sentences can be estimated above 10^{20}. (The actual number depends on where you put the border line between meaning and balderdash. [70])

This number juggling + rules of logic inescapably tell you that *there exists* a bona fide English sentence which is not in $\mathcal{R}$. But where is it? Can you show it to me? No, you can't.

The definition of $\mathcal{R}$ doesn't allow one to write down such a sentence until July 5, 2087.[71]

But don't get despondent. Instead, take the lesson from languages: temporally forget about the real $\mathcal{P}$ and try to define

$\mathcal{CP}$ – the "set" of all *conceivable* grammatical and meaningful *protein sequences.*

The word grammar must be understood in a broad sense, specifically here, as

a set of rules restricting possible structure of *individual* proteins as well as of *relations between different* proteins.

The structure of an individual protein often refers to its *spacial structure*, which is customary presented in *graphical language* as in the figure below.

This structure is somehow encoded by the protein sequence, but a universal decoder:

$$\text{sequence} \rightsquigarrow \text{structure}$$

remains unavailable at the present day.

[68]This is publication date of Newton's Mathematical Principles of Natural Philosophy , originally, in Latin and translated to English in 1728.

[69]It may be more with future computer programs generating billions sentences per second for the sake of other computers, but we leave it as it is.

[70]Write down a sentence in 20 words (better, counting twenty only for nouns, verbs adjectives, and adverbs) and try to evaluate how many variations of individual words the sentence can sustain and remain meaningful.

[71]Do you see a loophole? – You may present this sentence orally. But instead of resolving the difficulty this adds a new layer of perverted logic to the problem.

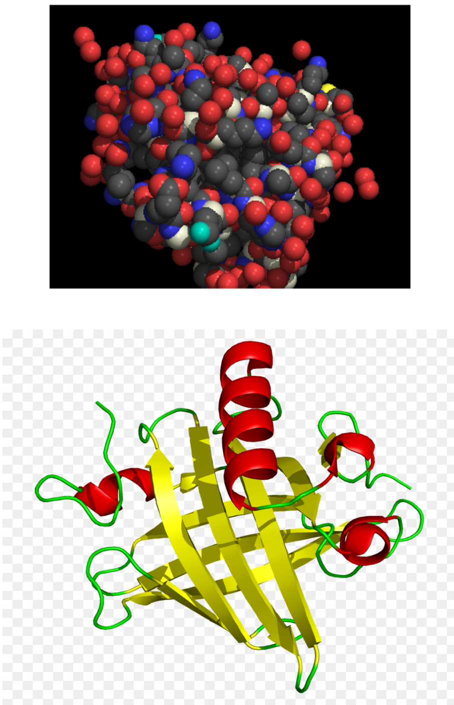


Figure 5: protein structure

3.1 Secondary Structure of Proteins, Folding and Unfolding.

Against logic and reason: knots in proteins.

3.2 Relations Between Proteins

1. Similarities between sequences and between patterns within sequences.

$1_\circ$ Similarities/dissimilarities between amino acids of two kinds: given a priori by their structure and implied by their positions in protein sequences.

Compare this to similarities/dissimilarities between phonemes/morpheme/letters/words in languages

2. Evolution and alignment.
3. Protein networks.
4. Alternative splicing.
5. Similarities of 3d structures, protein domains.
6. Similarities in functions.

7 Stability under mutations.

8. Presence in the same organism(s).

4 Lessons from Protein Folding.

Proteins seen on different spacial scales sequence, secondary, super-secondary tertiary quaternary structure protein interactions, protein, functions https://www.khanacademy.org/science/biology/macromolecules/proteins-and-amino-acids/a/orders-of-protein-structure

Secondary Structures: Symmetries, Fast Evolution and Fast Folding.

Protein Knots.

4.1 Structure of Functions on Large Spaces.

4.2 Symbols, Information, Structure, Function, Meaning.

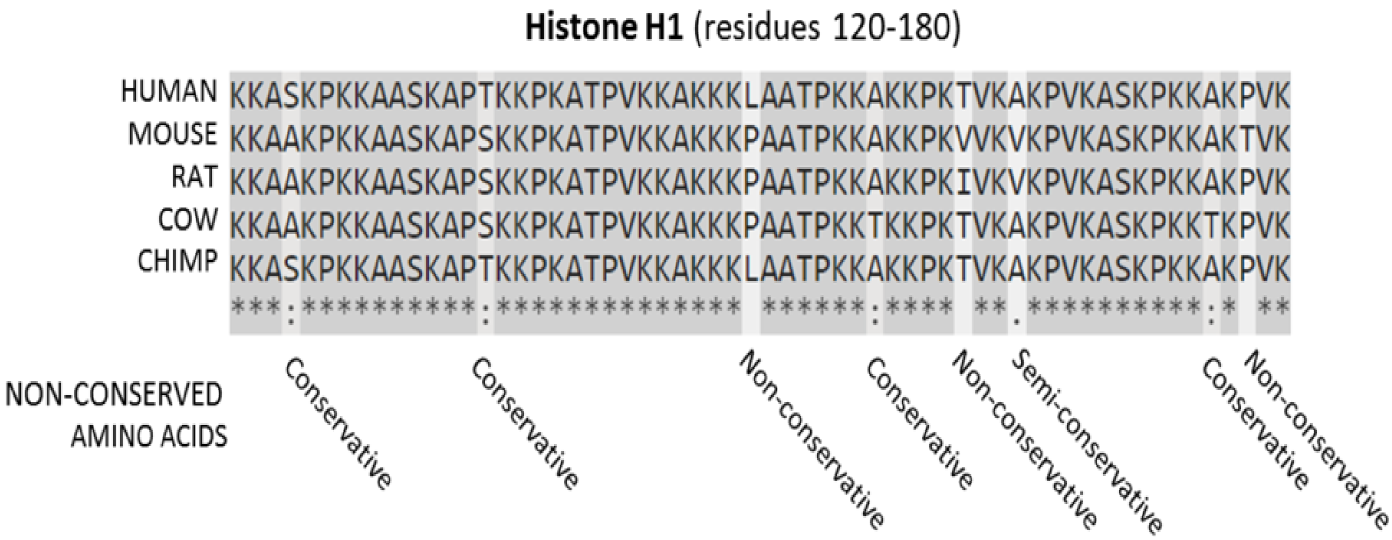


Figure 6: Alignment

5 Mathematics of Life.

5.1 Sources id Symmetry, Doubling, Conservation, Repetition, Complementarity

5.2 Trees Everywhere

(more to come)

/Users/misha/Desktop/large dimensions +applications /landscape.jpg

6 Structures on the Chess Board

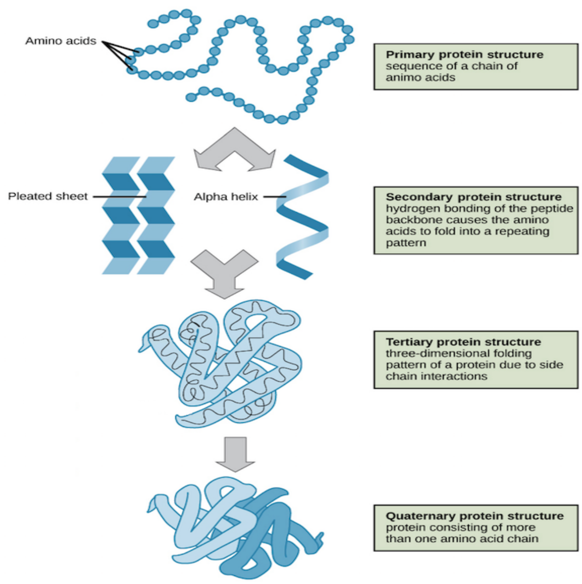


Figure 7: Protein Structures

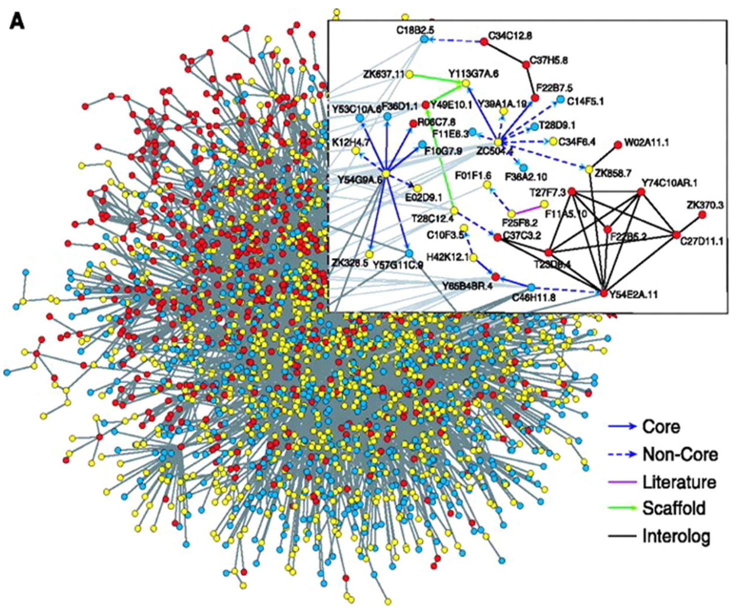
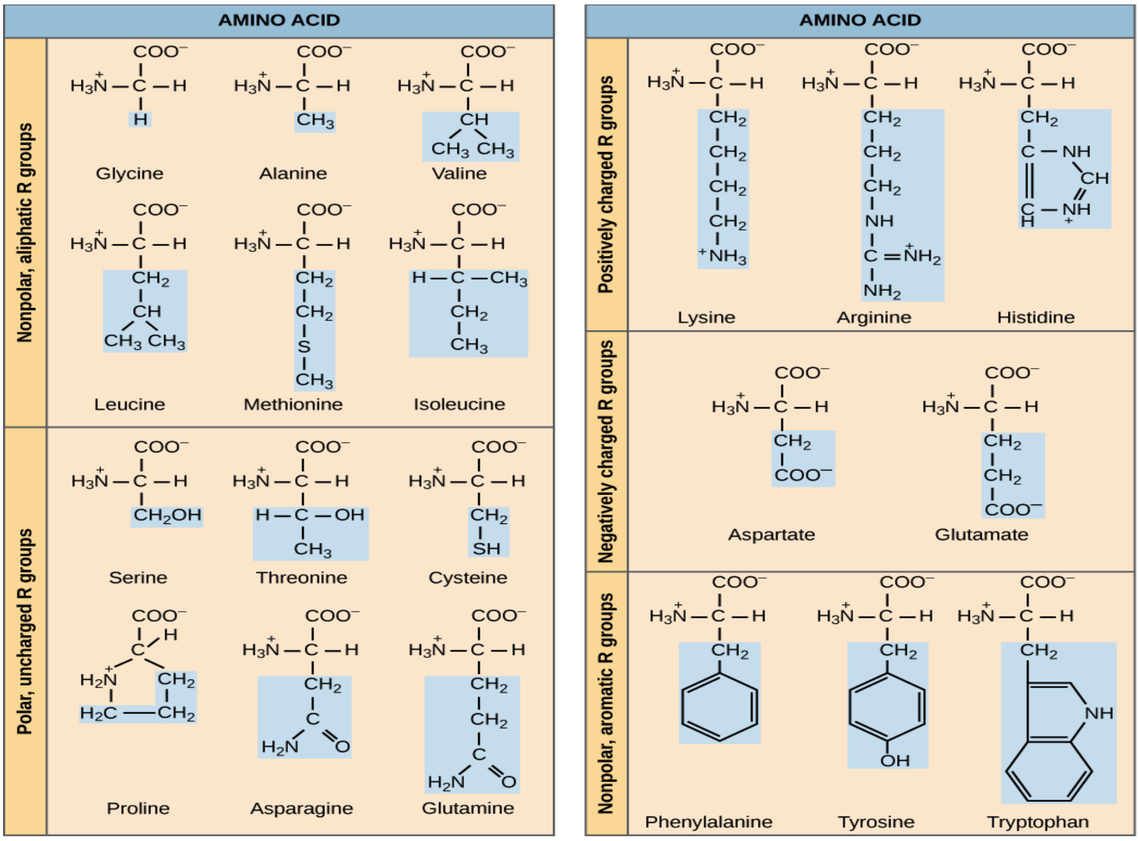


Figure 8: Protein network



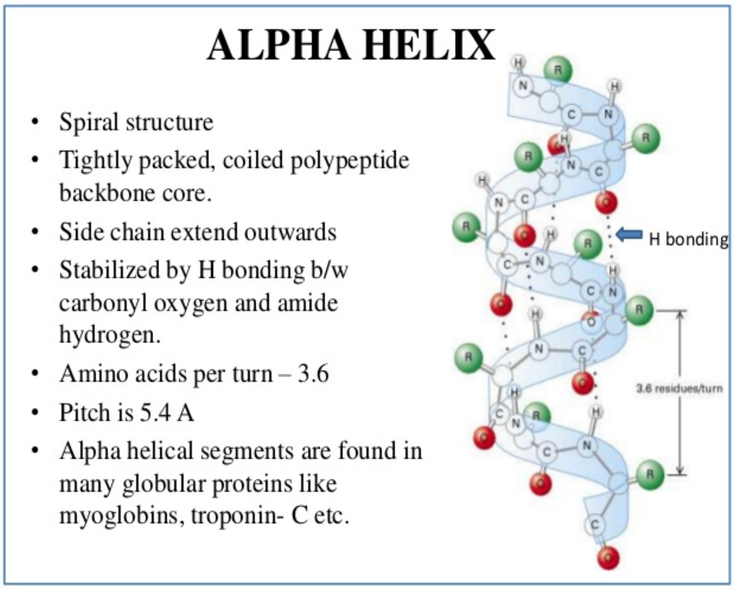


Figure 9: alpha helix

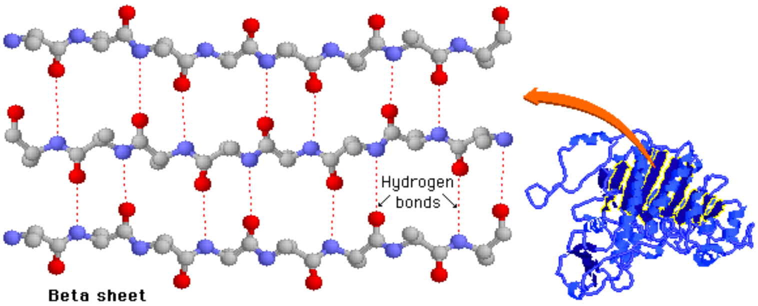


Figure 10: beta sheet

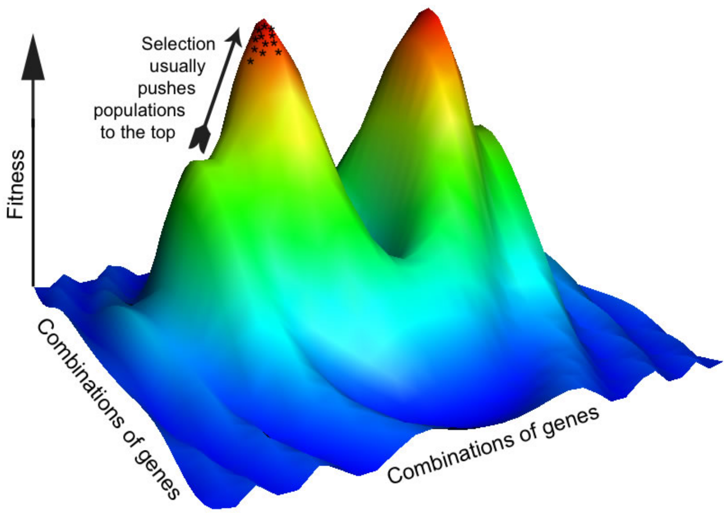


Figure 11: fitness landscape

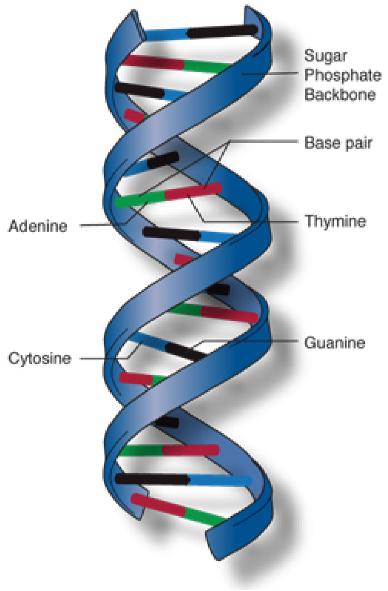


Figure 12: DNA

Figure 13: sefsimilarity