BSC/HPAI

Algorithmic Learning

(with mathematics inside)

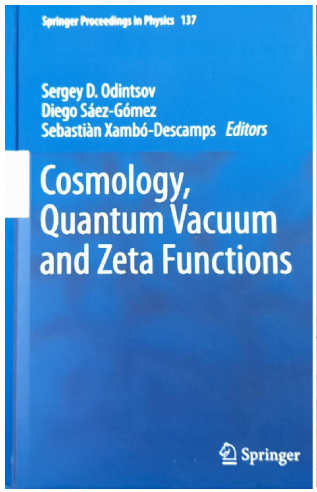
S. Xambó

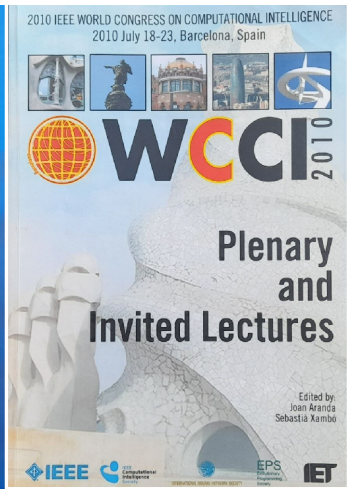
UPC & IMTech

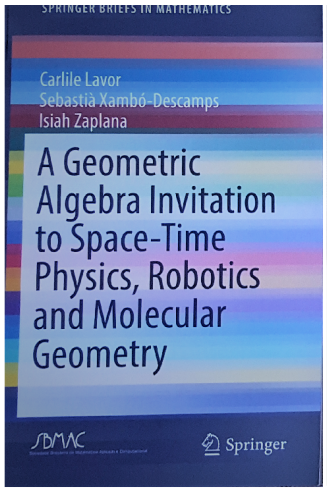
13/12/2021

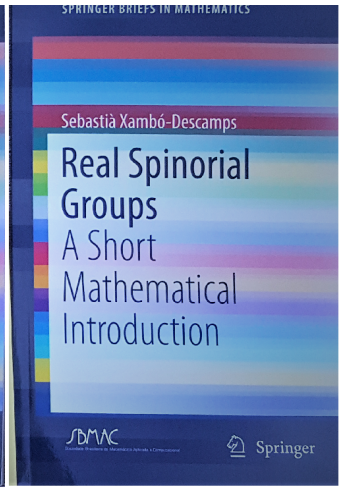
Interest is Algebraic Geometry, Geometric Algebra, Physics, Effective computations, Image Processing, Computer Vision, Biographical studies:

[1] (X-2008-Visio), [2] (X-2009-WIT), [3] (X-2009-K2),
[4] (X-2009-qc) (Lisbon talk), [5] (X-2010) WCCI, [6] (X-odintsov-saez-2011).









[7] (X-rue-2013) QC, [8] (X-2013-DM), [9] (X-roma-2013) pycones,
[10] (X-2015-jbsg) GA, Q esperanto, [11] (X-2015-review),
[12] (X-2015-GAT-A), [13] (X-2016-GAT-B), [14] (X-2016-Santalo-school),
[15] (X-2016-santalo-talks), [16] (X-2017-agacse), [17] (X-2017-icca11),
[18] (X-wainer-2018-b), [19] (X-2018-spins), [20] (lavor-X-zaplana-2018),
[21] (X-2018-Leibniz), [22] (X-greuel-narvaez-2018) AC60, [23] (X-2019-agacse),
[24] (X-2019-lightdream) Vigo, [25] (X-franch-2019) DHC AF,
[26] (X-2019-cadiz) PQ crypto, [27] (X-2019-uned), [28] (X-sayols-2020-MCS).

A pivotal shift in my interests was meeting **Eduardo U. Moya** at the AGACSE 2015 conference in Barcelona, but it was not until the ICCA11 conference in Ghent [17], in August 2017, that we started talking about *the role of GA in DL* and how to go about implementing the *computational aspects*. We have been collaborating since then:

[29] (moya-X-perez-salazar-mzortega-cortes-2020-PRL),
[30] (X-2021-iciam), [31] (X-moya-2021),
[32] (moya-X-salazar-sanchez-cortes-2021),
[33] (moya-X-sanchez-salazar-cortes-2021).

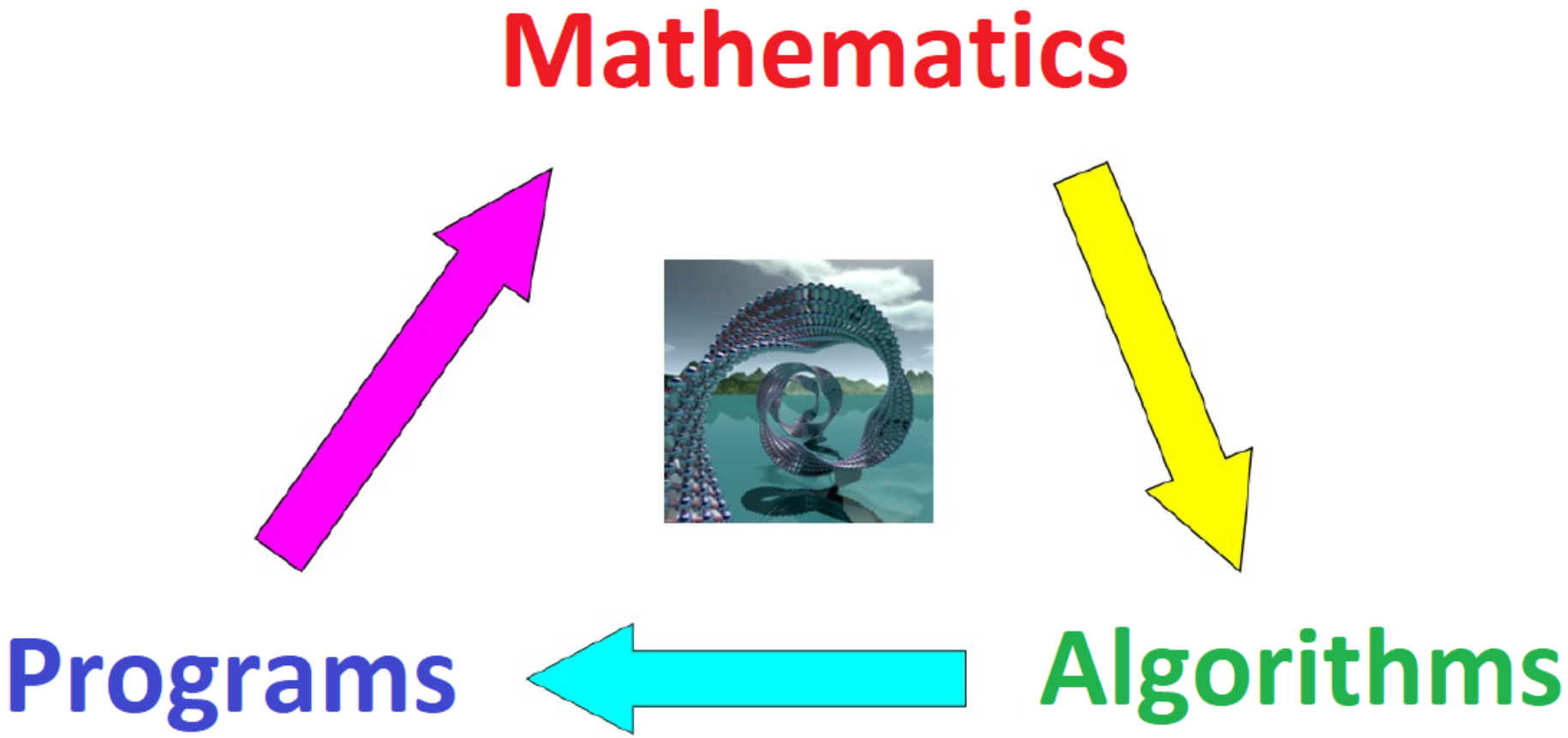
Another momentous influence has been the preparation of [34] with **Joan Bruna**, followed by our collaboration in this BGSM course:
[35] (bruna-X-2021-b) (computational aspects developed in collaboration of EUM).

October 5 to November 9: Six two-hour sessions at the BGSM/CRM course on *Mathematical Aspects of Algorithmic Learning and Deep Neural Networks*: Slides **10-05a** and **10-05b**, **10-13**, **10-14**, **10-19**, **11-02a**, **11-02b**, **11-09**. The other six two-hour sessions were delivered by **Joan Bruna**.

October 10: Chair of the round-table *New Bridges between Mathematics and Data Science: a Scientifc Debate* planned within the conference **New Bridges between Mathematics and Data Science** organized by the **REM** in Valladolid, 8-12 November 2021.

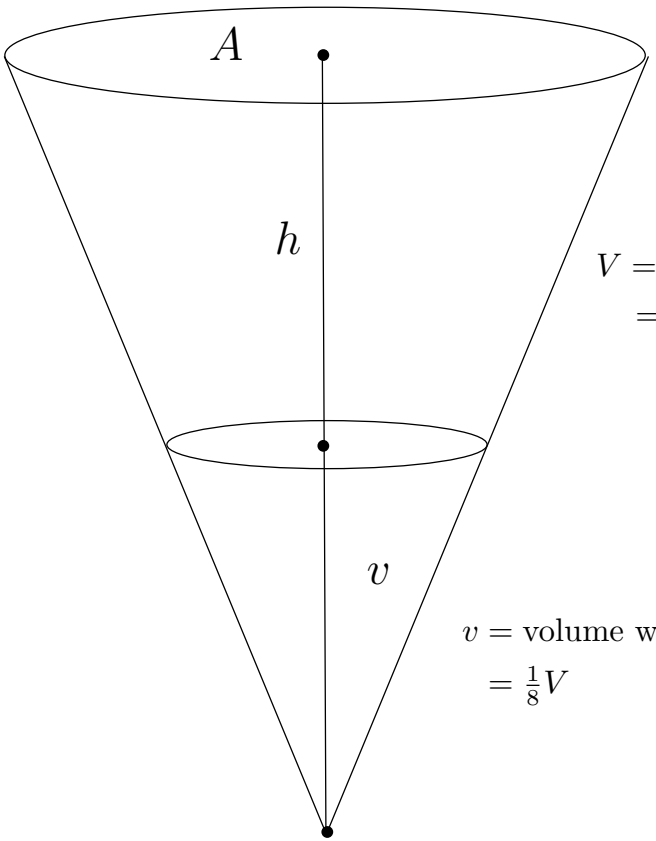
November 12: IMUVa Ateneo Lecture, *Algorithmic Learning: Analysis, Algebra and Geometry*.

Member of the Scientic Committee of the **ICACGA 2022** (*International Conference of Advanced Computational Applications of Geometric Algebra*, October 2-5, 2022, Denver, Colorado)



Slide 9 of [9] (X-roma-2013)

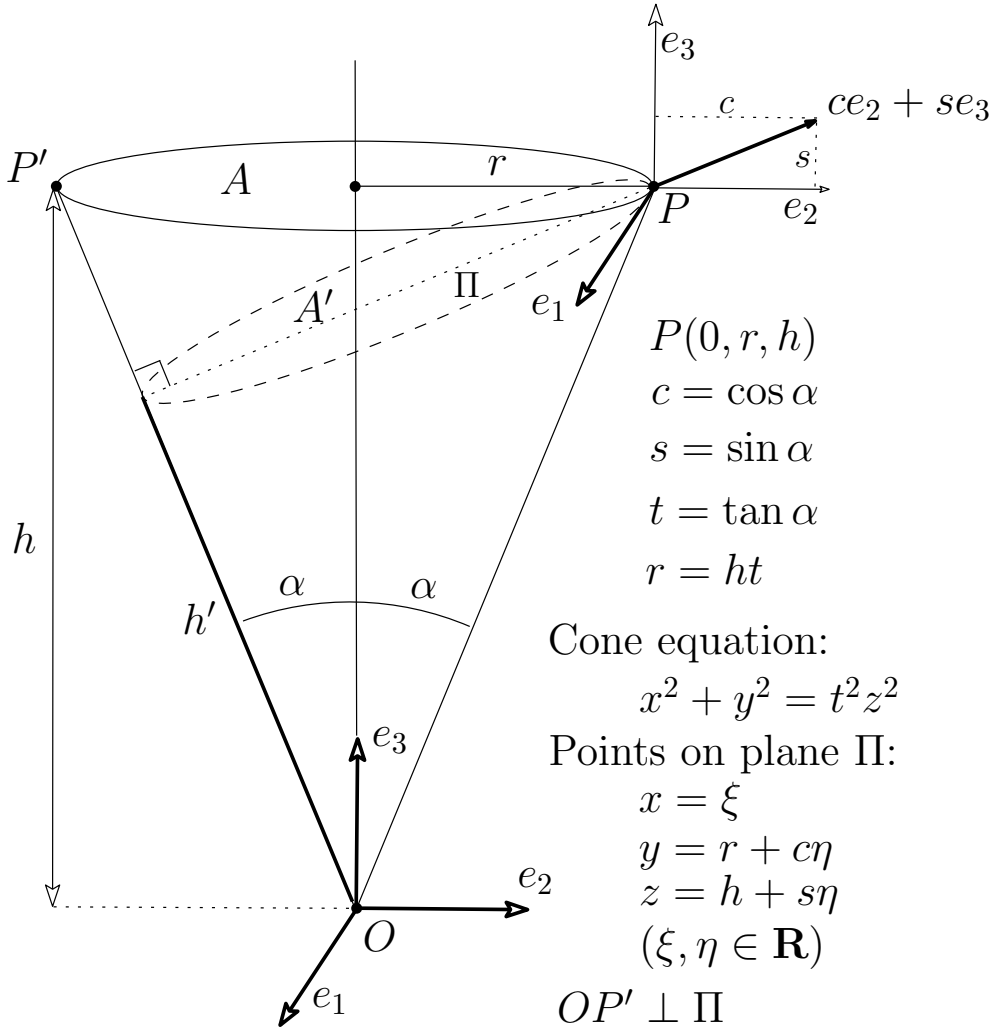
Image: Björling minimal surface



V = volume cone
$= \frac{1}{3}hA$

v = volume water
$= \frac{1}{8}V$

Rainfall: $R = v/A = \frac{1}{8}V/A = \frac{1}{8}\frac{1}{3}h = 1\text{cm} = 10\text{mm}$



$P(0, r, h)$
$c = \cos\alpha$
$s = \sin\alpha$
$t = \tan\alpha$
$r = ht$

Cone equation:
$x^2 + y^2 = t^2 z^2$

Points on plane Π:
$x = \xi$
$y = r + c\eta$
$z = h + s\eta$
$(\xi, \eta \in \mathbf{R})$

$OP' \perp \Pi$

$$R = \frac{1}{3}h\frac{\cos 2\alpha}{\cos^2\alpha}\frac{A'}{A} = \frac{1}{3}h\frac{C}{c^2}\frac{A'}{A}\ (C = \cos(2\alpha)).\ A' = \pi r^2 c\sqrt{C} = Ac\sqrt{C}.$$

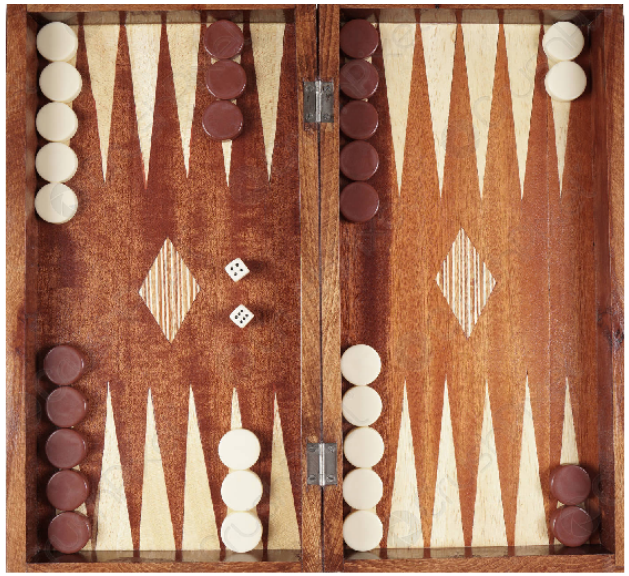
$$R = \frac{1}{3}h\frac{C}{c^2}\frac{A'}{A} = \frac{1}{3}h\frac{C}{c^2}c\sqrt{C} = \frac{1}{3}h\frac{C^{3/2}}{c}: \quad 51.5\,\text{L/m}^2.$$

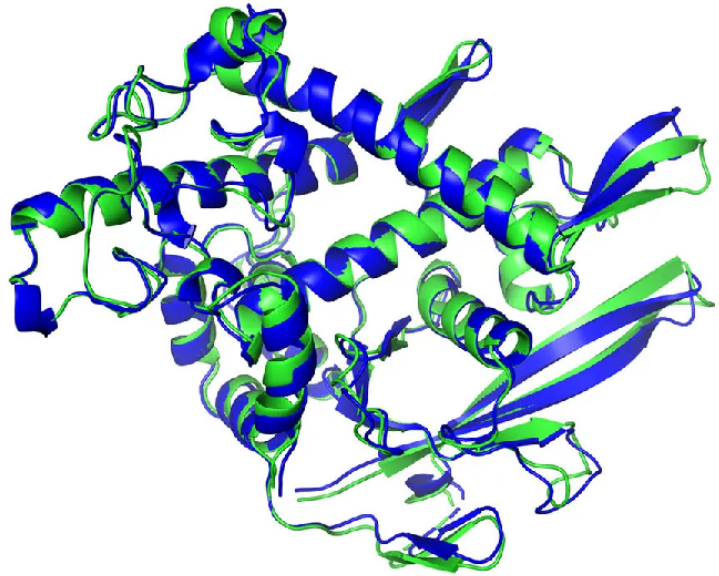
It's blowing in the wind

In the news

In textbooks and books

In arXiv et al (a sample)

Chess, Backgammon, Go, Console games, Protein folding, ...

Beethoven's Unfinished 10th Symphony Brought to Life by Artificial Intelligence (Scientific American, October 15, 2021).

Water Werzowa: I dare to say that nobody knows Beethoven as well as the AI algorithm.

I think music, when you hear it, when you feel it, when you close your eyes, it does something to your body. Close your eyes, sit back and be open for it, and I would love to hear what you felt after.

[36] *Música algorítmica: experiments i perspectives*. FME, December 1st, 2021: bayer-2021-video (55'56"-58'03").

"The new model suggests the partnership between neuroscience and AI could also move beyond our understanding of each one alone and instead find the general principles that are necessary for brains and machines to be able to learn anything at all." QM21-10-18.

NEURAL NETWORKS

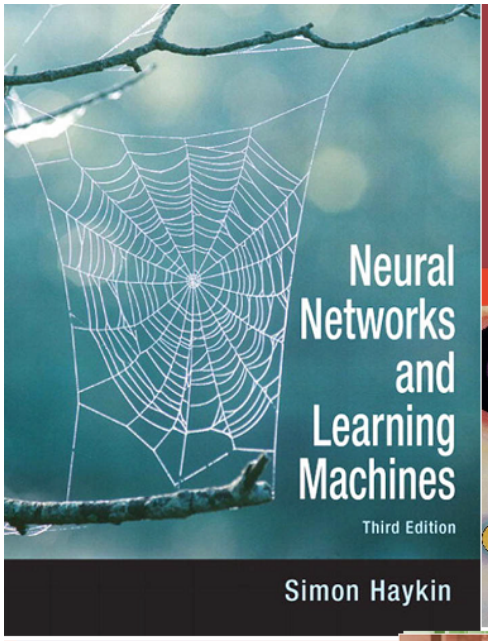
Neuron Bursts Can Mimic Famous AI Learning Strategy

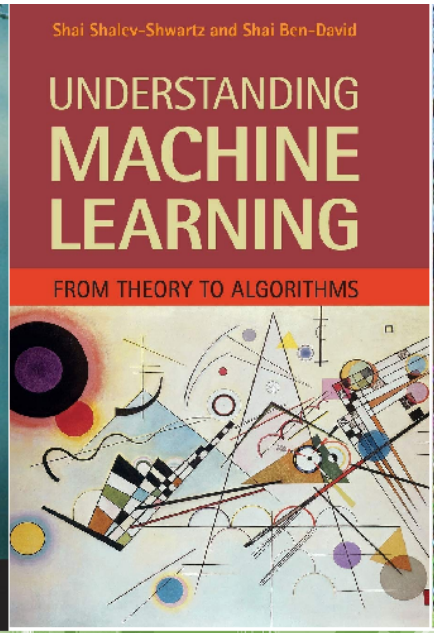
A new model of learning centers on bursts of neural activity that act as teaching signals --- approximating backpropagation, the algorithm behind learning in AI.

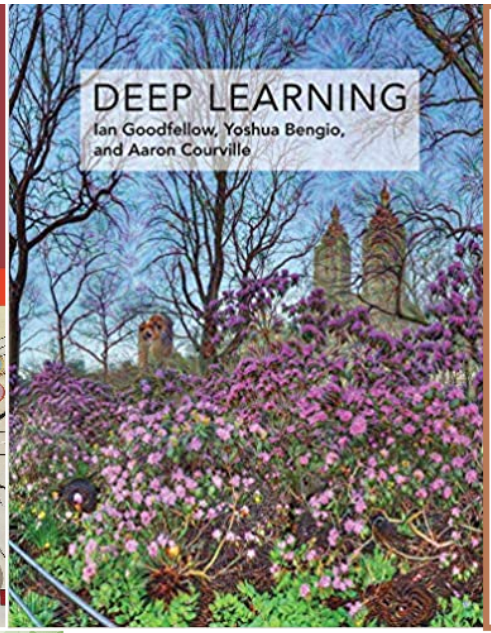
Allison Whitten

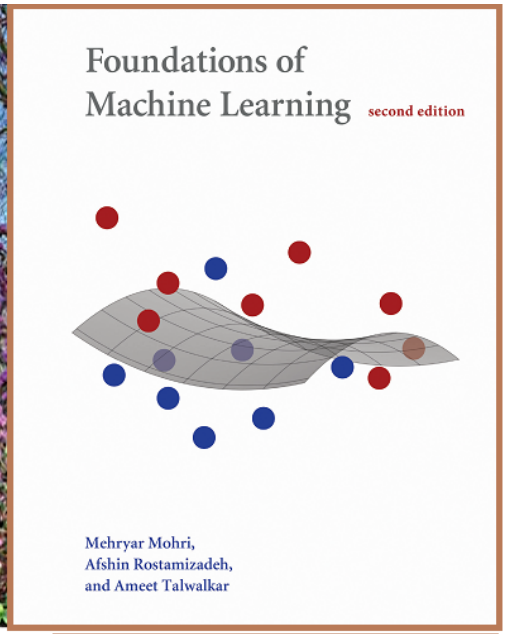
Source: Nature neuroscience

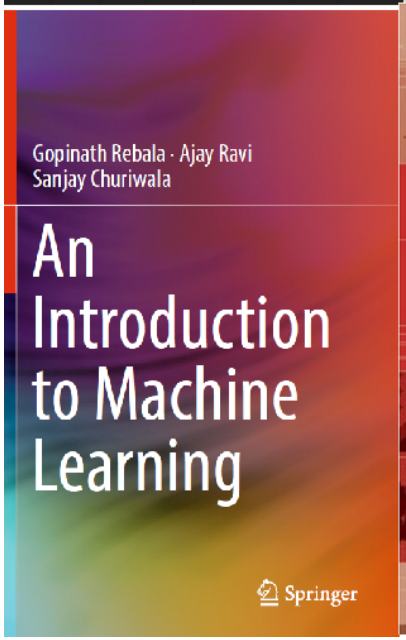
Reference: [37], *Learning representations by back-propagating errors*, Nature-1986-10-09.

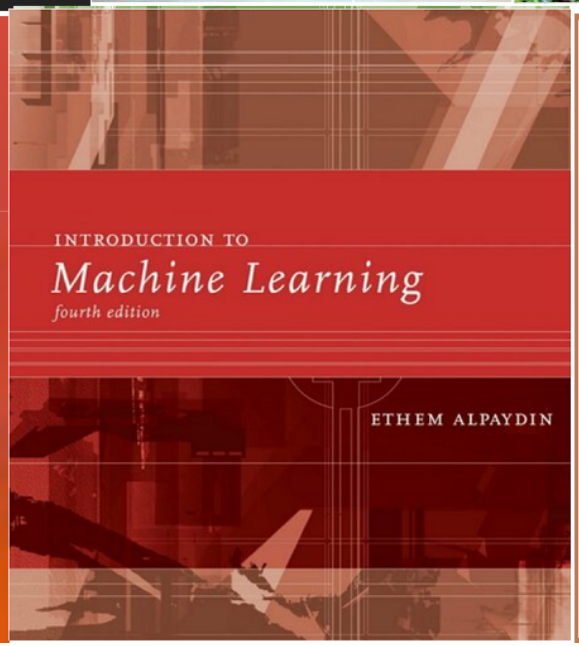

Neural Networks and Learning Machines
Third Edition
Simon Haykin

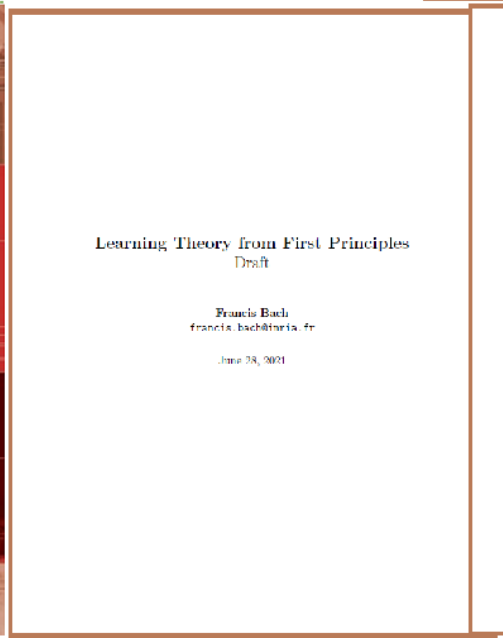

Shai Shalev-Shwartz and Shai Ben-David
UNDERSTANDING MACHINE LEARNING
FROM THEORY TO ALGORITHMS

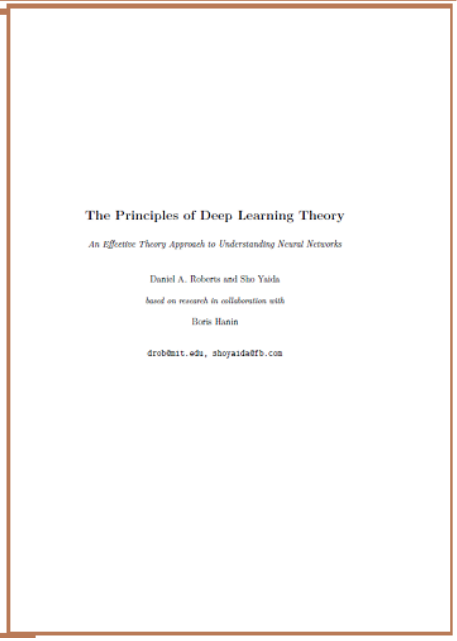

DEEP LEARNING
Ian Goodfellow, Yoshua Bengio, and Aaron Courville


Foundations of Machine Learning
second edition
Mehryar Mohri,
Afshin Rostamizadeh,
and Ameet Talwalkar


Gopinath Rebala · Ajay Ravi
Sanjay Churiwala
An Introduction to Machine Learning
Springer

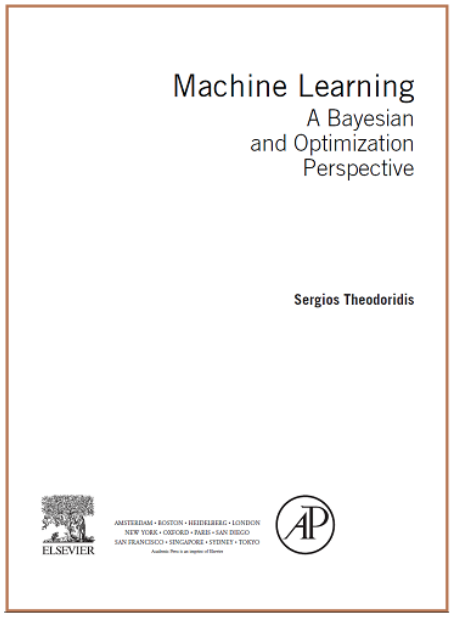

INTRODUCTION TO
Machine Learning
fourth edition
ETHEM ALPAYDIN

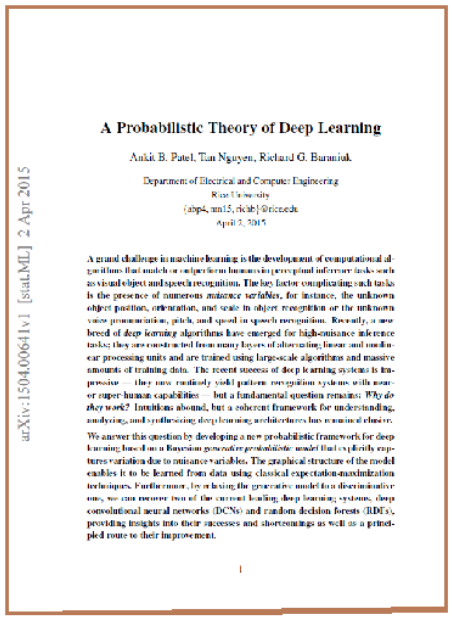

Learning Theory from First Principles
Draft
Francis Bach
francis.bach@inria.fr

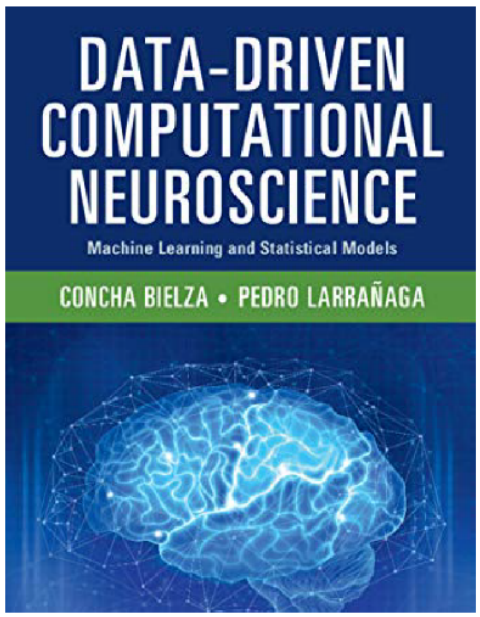

The Principles of Deep Learning Theory
An Effective Theory Approach to Understanding Neural Networks
Daniel A. Roberts and Sho Yaida
based on research in collaboration with
Boris Hanin

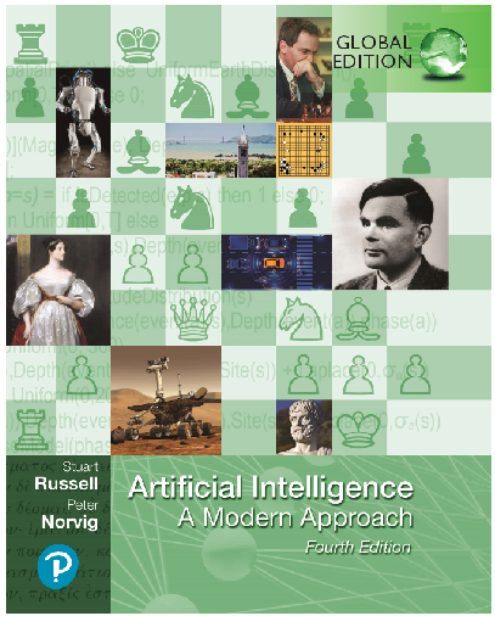
N

Bayesian approaches









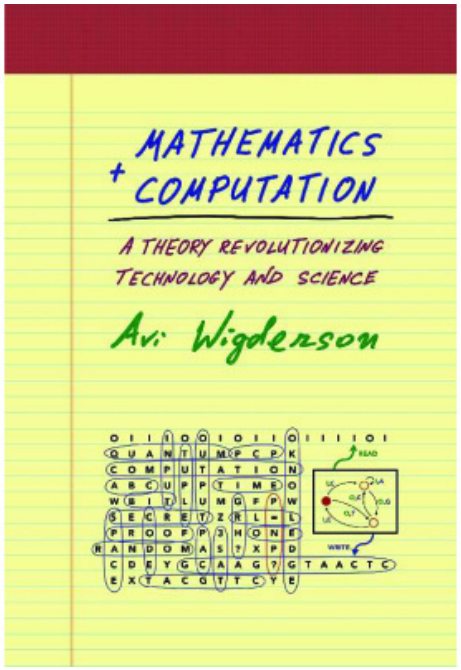
Applications

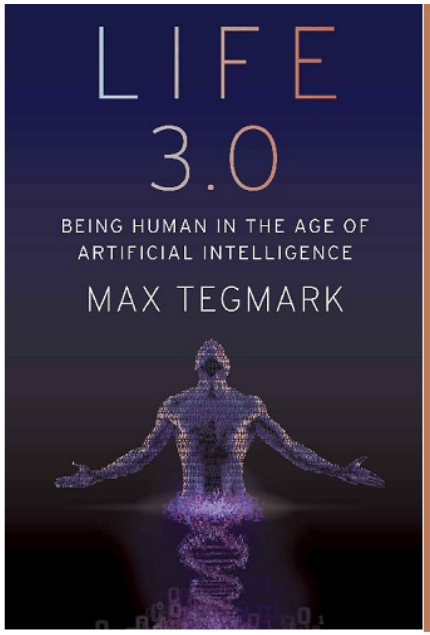
The many facets of the symbiosis Mathematics & Computation are appraised in [38] (wigderson-2019).

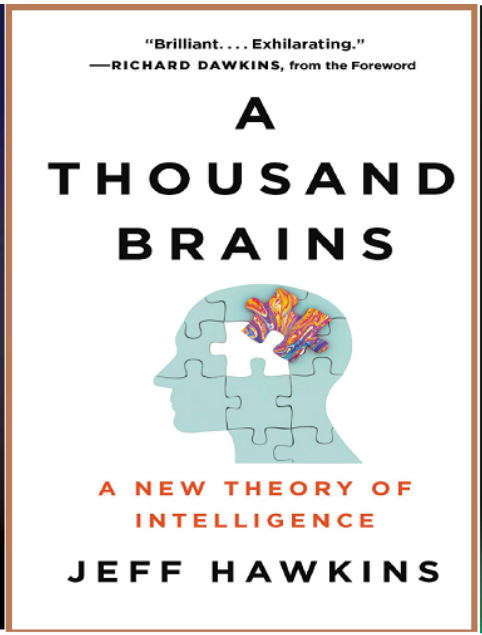
In particular, Chapter 17 is devoted to computational learning theory.

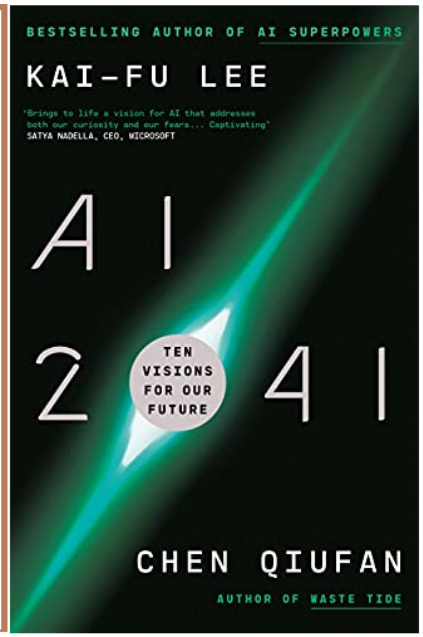
See also the extensive survey [39] and, with a more popular style, Marr's blog [40].

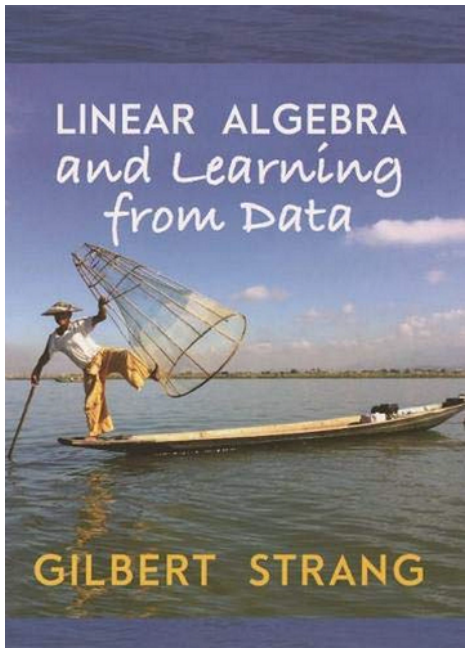
Another useful reference is [41] (strang-2019).











[42] (jin-netrapalli-ge-kakade-jordan-2019)

[43] (carleo-cirac-cranmer-daudet-schuld-tishby-vogtmaranto-zdeborova-2019)

[44] (lample-charton-2019) *symbolic integration*

[45] (du-zhai-poczos-singh-2019) GD

[46] (du-lee-li-wang-zhai-2019) GD

[47] (wang-shi-cao-2019) GA-SURF

[48] (he-kim-2019)* Learning algebraic structures

[49] (alessandretti-baronchelli-he-2019)* Birch-S-D

[50] (kusaba-liu-koyama-terakura-yoshida-2019) Periodic table

[51] (guo-wan-hu-liu-liu-bennamoun-2019) 3D point clouds

[177] (pyzerknapp-laino-2019) ML in Chemistry

[52] (garcia-cortes-napagao-cortes-2020)

[53] (charton-hayat-lample-2020)

[54] (weinan-ma-wojtowytsch-wu-2020)

[55] (zhao-birdal-lenssen-menegatti-guibas-tombari-2020)

[56] (wu-xu-wu-kong-senhadji-shu-2020)

[57] (hoffmann-schmitt-osindero-simonyan-elsen-2020)**

[58] (he-yau-2020)*, [59] (cranmer-et-5-2020)*

[60] (lamb-garcez-gori-prates-avelar-vardi-2020)*

[61] (iten-metger-wilming-delrio-renner-2020)*

[62] (townshend2-eismann-dror-2020)**, [63] (melas-2020)

[64] (scholze-2020), [65] (raghu-schmidt-2020)**,

[66] (heal-kulkarni-sertoz-2020)*, [67] (hughes-2020)*

[68] (jiang-luo-2021)**, [69] (ghojogh-ghodsi-karray-crowley-2021a)

[70] (scellier-2021), [71] (bach-2021)

[72] (bronstein-bruna-cohen-velickovic-2021)

[73] (mei-misiakiewicz-montanari-2021), [74] (cohen-2021)

[75] (beniaguev-segev-london-2021)**, [76] (wagner-2021)*

[77] (berner-grohs-kutyniok-petersen-2021), [78] (smirnov-2021)*

[79] (zaplana-2021), [80] (fernandes-2021)

[81] (franchini-vitabile-2021), [82] (raayoni-et8-2021)*,

[83] (payeur-guerguiev-zenke-richards-naud-2021)

[84] (davies-et13-2021)*, [85] (blundell-et4-2021)

[86] (davies-et3-2021), [87] (scholze-2021), [88] (he-2021)*

[89] (avigad-moura-kong-2021), [90] (yaseen-et4-2022)

On Human Communication

A REVIEW, A SURVEY, AND A CRITICISM

Third Edition

1978

Colin Cherry

Henry Mark Pease Professor of Telecommunication,
Imperial College, University of London

THE MIT PRESS

Cambridge, Massachusetts, and London, England

1957 *Preface to the First Edition*

I have written this book at the invitation of the editors of the series "Studies in Communication," to serve as an introduction to that series of volumes which will appear during the next few years. It is intended as a review, a survey, and a criticism—nothing more.

1966 *Preface to the Second Edition*

and to extend the bibliography.* The latter task has presented a great problem, since, during the past 10 years, the development of world communication techniques has proceeded at a rate commonly called "explosive." In that decade, words like *automation*, *satellites*, *space*, *computer* have come into everyday chatter and newspaper talk; they are used by the innocent as the jargon of a cult. This explosive growth of the "technology of information" has created thousands of publications in the journals and hundreds of books. Since detailed reference is impossible, I have decided to restrict the new Bibliography to useful sources of reference, a few of the more significant technical and historical works, and a few items which were omitted from the first edition.

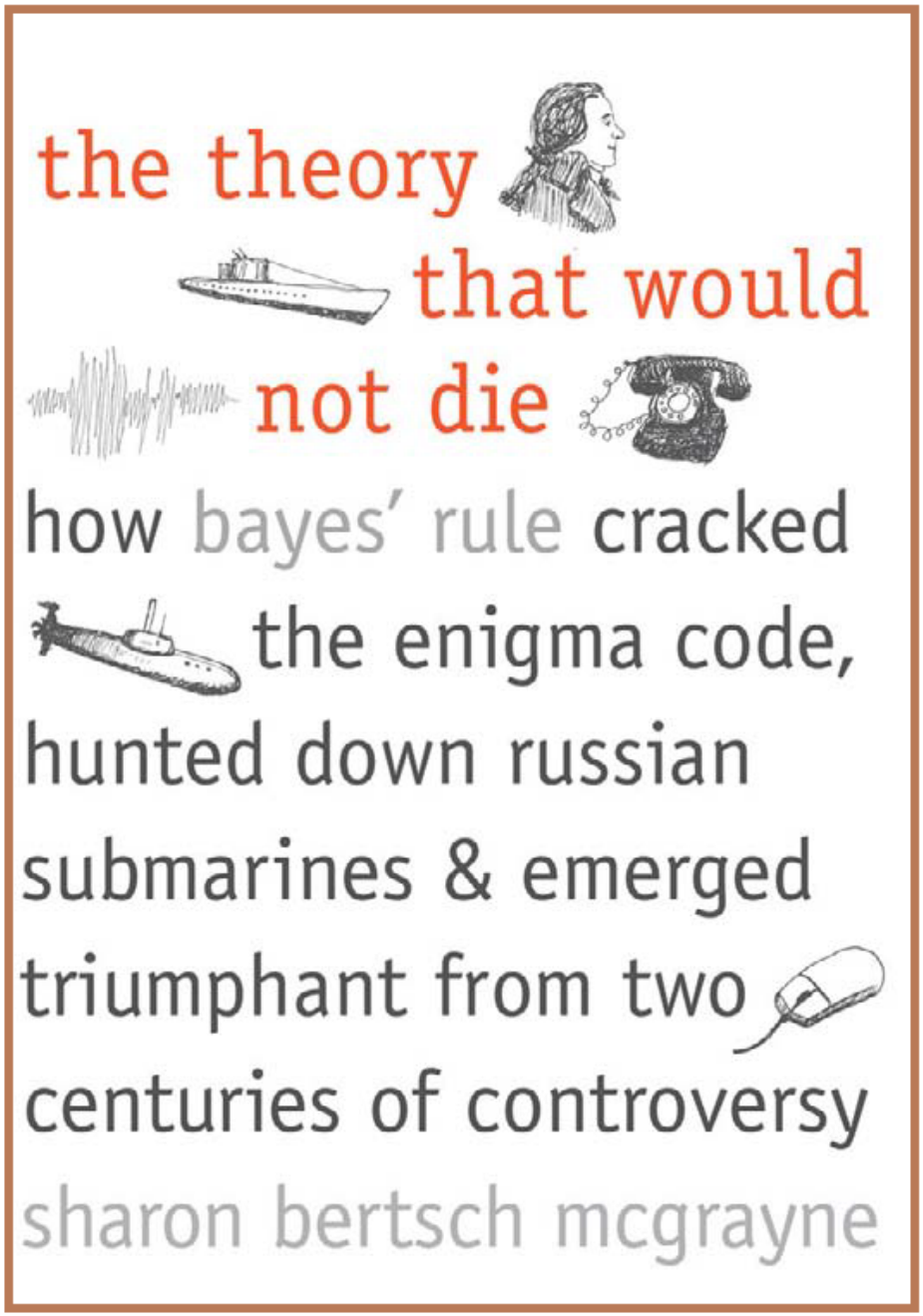
* 467 items, 57 new

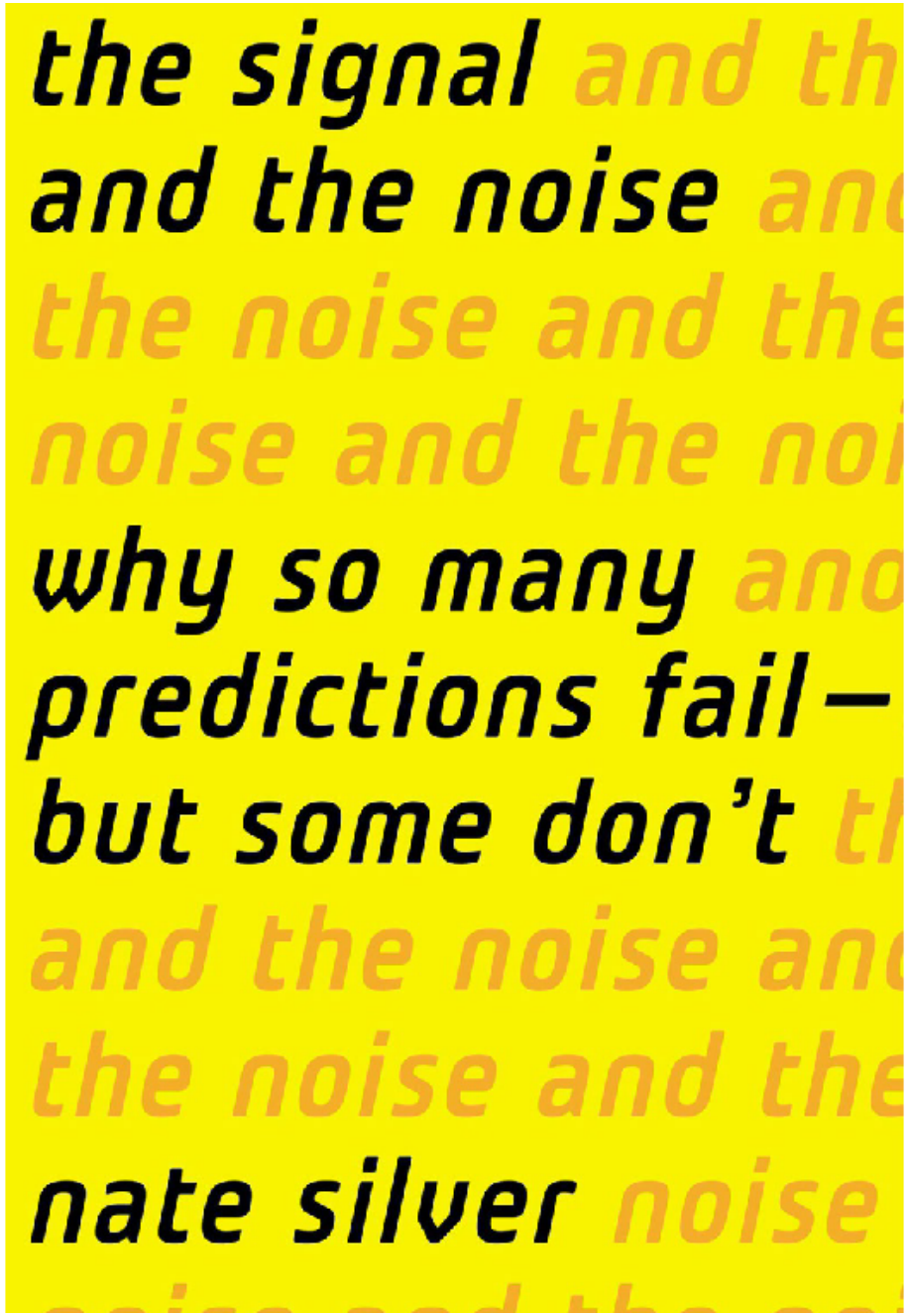
The current situation concerning AI is not unlike that of the *Human Comunication* of fifty years ago, but surely even more "explosive".

Where to begin?

Everlasting jewels

The Bayes-Laplace rule
Maximum a posteriori rule
Clustering methods (k-Means and k-NN)
Principal component analysis
Singular Value Decomposition


the theory
that would
not die
how bayes' rule cracked
the enigma code,
hunted down russian
submarines & emerged
triumphant from two
centuries of controversy
sharon bertsch mcgrayne


the signal and th
and the noise an
the noise and the
noise and the no
why so many an
predictions fail–
but some don't t
and the noise an
the noise and the
nate silver noise

Learning from experience is neatly encoded, in probabilistic terms, by the *Bayes-Laplace rule*:

$$P(Y|X) = P(Y)K(X,Y),$$

where $K(X,Y) = P(X \cap Y)/P(X)P(Y)$ is symmetric ($K(X,Y) = K(Y,X)$), which tells us how to modify our *prior* believe in Y, $P(Y)$, to the belief $P(Y|X)$ *posterior* to having observed X.

When $K > 1$ ($K < 1$), our believe is *increased* (*decreased*), and in any case it may be construed as a *learning* on Y produced by the observation of X.

The condition $K = 1$ is equivalent to say that X and Y are independent, and in this case the believe in Y is unaffected by whether or not X occurs.

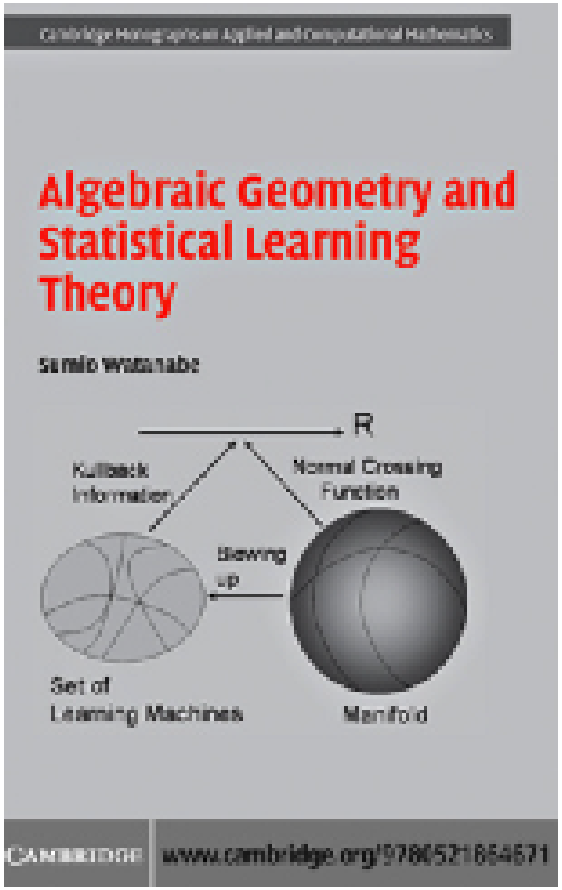
The Bayes–Laplace rule is the basis for *prediction models* of many sorts, as for example in weather forecasting.

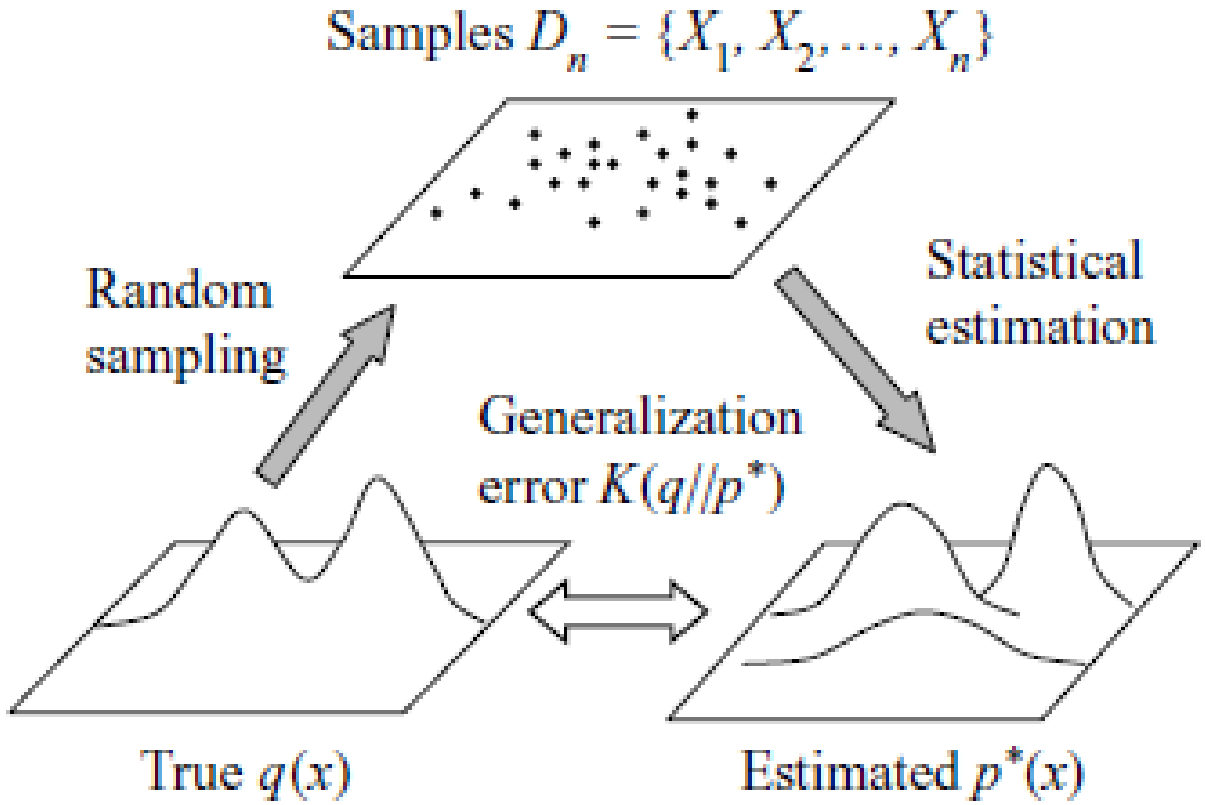
If an event X occurs, and it can be assigned to disjoint hypothesis $Y_1, \cdots, Y_r$, the MAP rule selects the hypothesis Y_j such that $P(Y_j|X)$ is maximum. The Bayes-Laplace formula tels us that this is the same as selecting the Y_j such that $P(X|Y_j)P(Y_j)$ is maximum.

In the special case in which the Y_j have the same probability, this amounts to select the Y_j such that $P(X|Y_j)$ is maximum.

[92] (silver-2012), *The signal and the noise*:

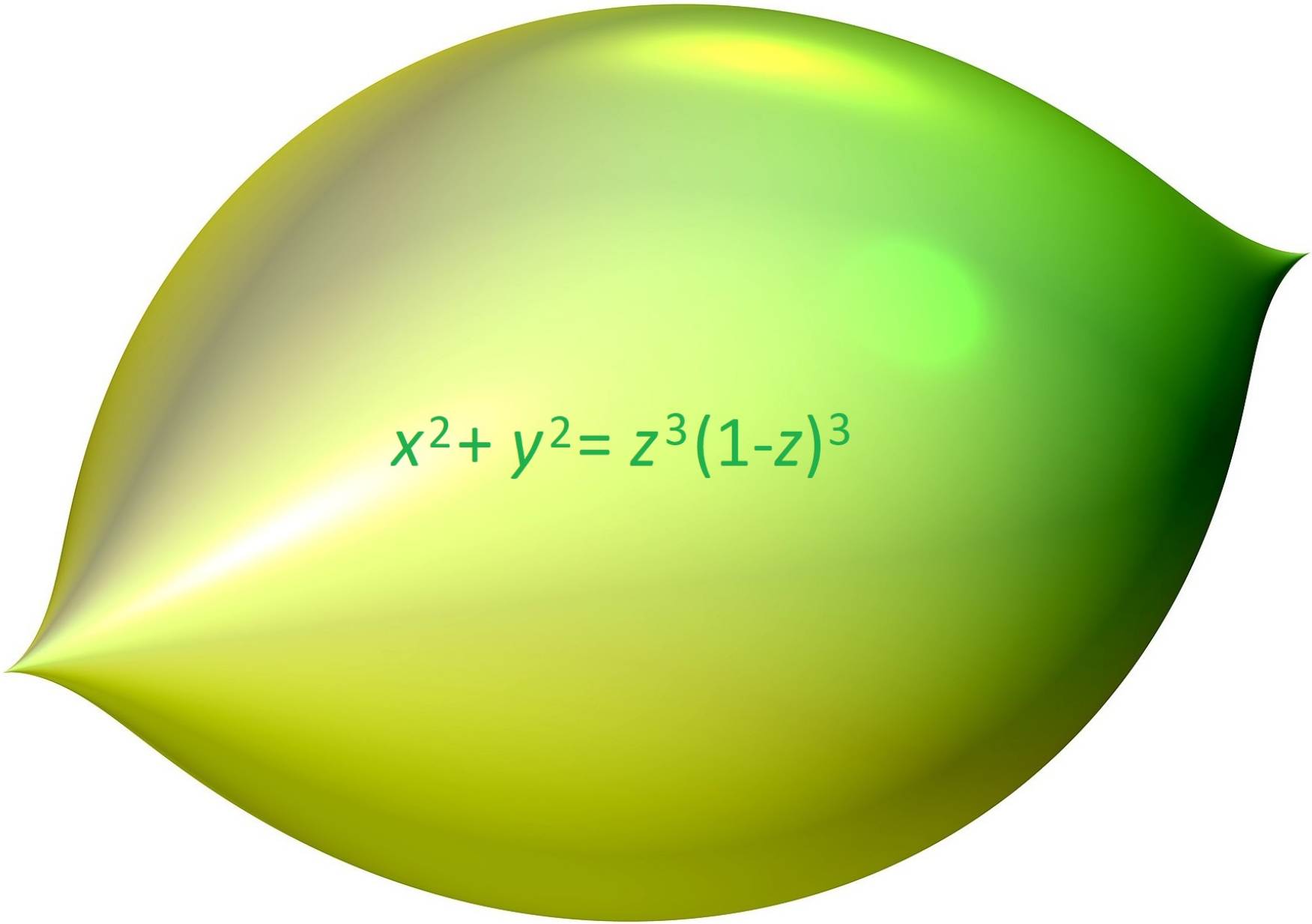
"Bayes's theorem [...] implies that we must think differently about our ideas — and how to test them. We must become more comfortable with probability and uncertainty. We must think more carefully about the assumptions and beliefs that we bring to a problem" (page 15).





In this book, we study a method which employs a parametric probability density function. A conditional probability density function $p_w(x) = p(x|w)$ of $x \in \mathbf{R}^N$ for a given parameter $w \in W$ is called a learning machine or a statistical model, where W is the set of all parameters.

From [93] (watanabe-2009).


$x^2+ y^2= z^3(1-z)^3$

Aimed at finding *hidden structure in data*, $\mathcal{D} = \{x^1, \ldots, x^m\}$.

k-Means. This algorithm groups unlabeled data $\mathcal{D}$ in k classes:

(1) Select k vectors $z^1, \cdots, z^k \in \mathcal{D}$ at random.

(2) Assign each $x^j \in \mathcal{D}$ to the first z^i nearest to x^j (initial groups).

(3) Update each z^i to the centroid (or *mean*) of the z^i group.

(4) Iterate (2) and (3) until the z^i are stable (up to a *tolerance*).

The associated *cluster predictor* assigns x to the first nearest z^i.

k-NN (nearest neighbors). Let $\mathcal{D} = \{(x^1, y^1), \cdots, (x^m, y^m)\}$ be a labeled set and k a positive integer. The label predictor of the k-NN algorithm assigns a vector x to the mode of $y^{j_1}, \cdots, y^{j_k}$, where $x^{j_1}, \cdots, x^{j_k}$ are the nearest neighbors of x from among $x^1, \cdots, x^m$.

Let X a data matrix of size $m \times n$. We regard the rows X^i of X as *observations* on n objects, $X^i = (x_1^i, \ldots, x_n^i)$, for m *features* ($i = 1, \ldots, m$).

Let $\mu^i = E(X^i)$, the *mean value*, or *expected value*, of X^i.

Let $\sigma_{ij} = \mathrm{Cov}(X^i, X^j) = E[(X^i - \mu^i)(X^j - \mu^j)] = E[X^i X^j] - \mu^i \mu^j$ and $\Sigma = (\sigma_{ij})_{1 \leqslant i,j \leqslant n}$. This is the *covariance matrix* of X, $\Sigma = \mathrm{Cov}(X)$. Notice that $\sigma_{ii} = \mathrm{Var}(X^i) = \sigma_i^2$, where $\mathrm{Var}(X^i)$ and σ_i are the *variance* and the *standard deviation* of X^i.

Given a unit m-vector u, it turns out that $\mathrm{Var}(uX) = u\Sigma u^T$, and that this is maximum precisely when *u is an eigenvector of Σ with the highest eigenvalue*. This vector is the *principal component* of X, that is, the unit eigenvector $u = u_1$ of Σ whose eigenvalue λ_1 is largest (the eigenvalues of Σ are real). It accounts for the greatest variance of the data along a direction.

The second principal component of X is the eigenvector u_2 corresponding to the second eigenvector λ_2. It maximizes $\mathrm{Var}(uX) = u\Sigma u^T$ for unit vectors u orthogonal to u_1. And so on.

This frames an (unsupervised) approach to *dimension reduction* by means of the spectral decomposition $\Sigma = U\Lambda U^T$, where Λ is the diagonal matrix with the eigenvalues of Σ, ordered in non-increasing order, and U is the orthonormal matrix formed with the unit eigenvectors of Σ. In fact, the reduction of X to dimension $k \leqslant m$ is the $k \times n$ matrix $X' = U_k X$, where the rows of U_k are the first k rows of U.

Let X be an $m \times n$ data matrix as above, and let r be its rank. Then XX^T and X^TX have rank r and they have the same non-zero eigenvalues $\lambda_1^2, \ldots, \lambda_r^2$, where $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_r > 0$.

Moreover, if we let U and V denote the orthonormal matrices of eigenvectors of XX^T and X^TX, then $X = U\Lambda V^T$ where $\Lambda_{jj} = \lambda_j$ for $j = 1, \ldots, r$ are the only non-zero values of Λ.

Note that $XX^T = U(\Lambda\Lambda^T)U^T$ and $X^TX = V(\Lambda^T\Lambda)V^T$, where the first r values of the diagonals of $\Lambda\Lambda^T$ and $\Lambda^T\Lambda$ are $\lambda_1^2, \ldots, \lambda_r^2$ and all others 0 in both matrices (of sizes $m \times m$ and $n \times n$, respectively).

Since $U\Lambda = (\lambda_1 u_1, \ldots, \lambda_r u_r)$, we get the *singular value decomposition* of X:

$$X = \lambda_1 u_1^T v_1 + \cdots + \lambda_r u_r^T v_r.$$

Actually it turns out (Eckart-Young theorem) that for $k = 1, \ldots, r$ the matrix

$$M_k = \lambda_1 u_1^T v_1 + \cdots + \lambda_r u_k^T v_k$$

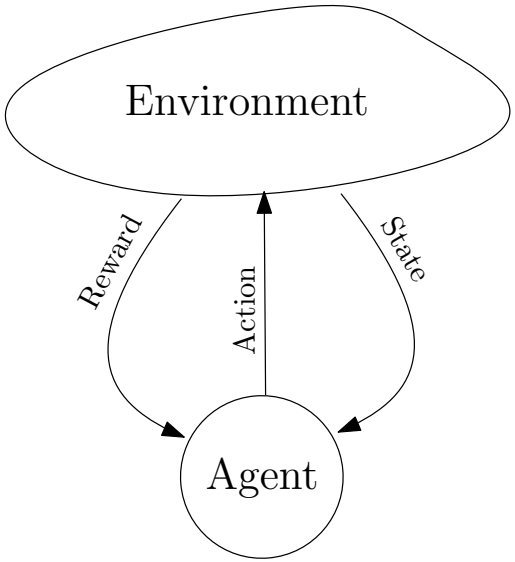
is the closest to X among the matrices of rank k.

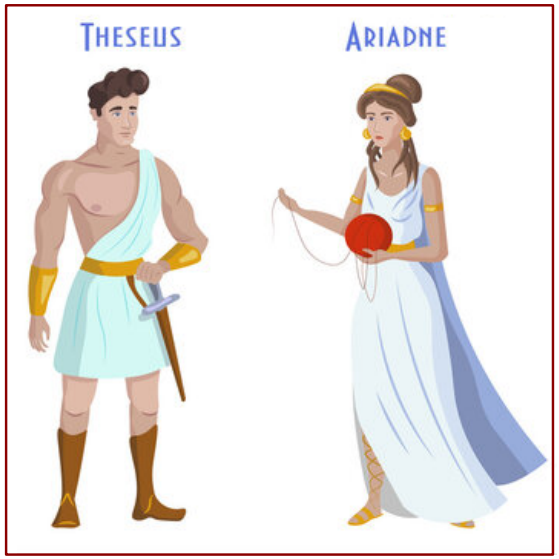
Remark. The least-squares solution to $Xa = b$ is $a = X^\dagger b$, where $X^\dagger = V\Sigma^\dagger U^T$ (Moore-Penrose pseudo-inverse of X), $\Sigma^\dagger = \mathrm{diag}(\lambda_1^{-1}, \ldots, \lambda_r^{-1}, 0, \ldots, 0)$.

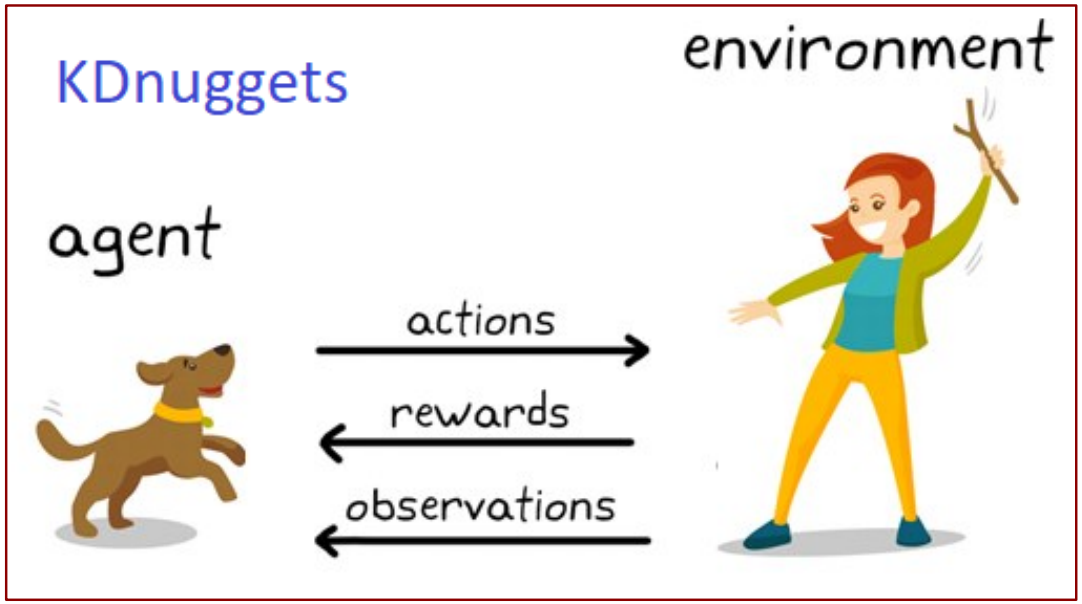
Data and learners

Reinforcement learning
Supervised learning

Algorithm learns to *react* to an *environment*







Hypothesis space[3]

$$\mathcal{H} = \{h\}$$

Data, expert tables[2]

$$\mathcal{D}$$
$$\{(x^1, y^1), \ldots, (x^m, y^m)\}$$
$$x^j \sim P, \quad y^j = f^*(y^j)$$

World – Observer[1]

Learner/LA[5]

Error measure[4]

$$\Longrightarrow$$

Prediction rule[6]

$$y = h(x)$$

[1] *Urbi et Orbi* engraving (Flammarion). **Tycho Brahe** (observer model): He experienced the solar eclipse of 21 August 1560 [he was 15], and was greatly impressed by the fact that it *had been predicted, although the prediction based on current observational data was a day off.* He realized that *more accurate observations* would be **the key to making more exact predictions**. [2] **Ephemeris**: Tables of planet and comet positions over time.
[3] *Inductive bias.* Greeks: circles around Earth. [4] *Loss*, *risk*, *regret.* How close are predictions to observations?
[5] Learner model (Kepler): Ellipses with a focus at the Sun. Today: Learning algorithm. [6] Hopefully, $h \approx f^*$.

Set of objects x^j, $j \in [m]$ (*dataset*).

Want to predict values y^j provided by a *supervisor* or *expert* in such a way that for objects x not in the dataset the value y corresponding to x is predicted with high probability (*generalization* capacity).

Hypothesis space: a space of parameterized functions,

$$\mathcal{H} = \{h_w\}_{w \in W}.$$

Problem: to find $w \in W$ such that $h_w(x^j) \approx y^j$.

Method: If the criterion for fitness depends on a function $\ell(h(x), y)$ (*local cost*), we transfer the problem to finding w that minimizes $\sum_j \ell(h_w(x^j), y^j)$ (*empirical cost* or *risk*):

$$\operatorname{argmin}_w \sum_j \ell(h_w(x^j), y^j).$$

This is the *empirical risk minimization* rule, ERM.

Supervised learning has two main modalities:

Classification: When the set $\mathcal{Y}$ is finite. In this case its elements are usually called *labels* or *classes*.

Regression: When the set $\mathcal{Y}$ is the set $\mathbf{R}$ of real numbers.

In *linear regression*, $\mathcal{H}$ is the space of functions of the form

$$h(x) = w_1 x_1 + \cdots + w_n x_n = w \cdot x$$

and the the local cost is usually $(h(x) - y)^2$.

Logistic regression is linear regression of the function $\log \frac{p}{1-p}$, $p \in [0, 1]$. If w is a solution, then $p = 1/(1 + e^{-w \cdot x})$.

Artificial neural networks

Artificial neurons
Layered architectures
$\mathcal{A}$-NNs
Examples: CxNN, QNN, GNN,...

In AL, a useful model of a *neuron* is depicted in Fig. 9.1:

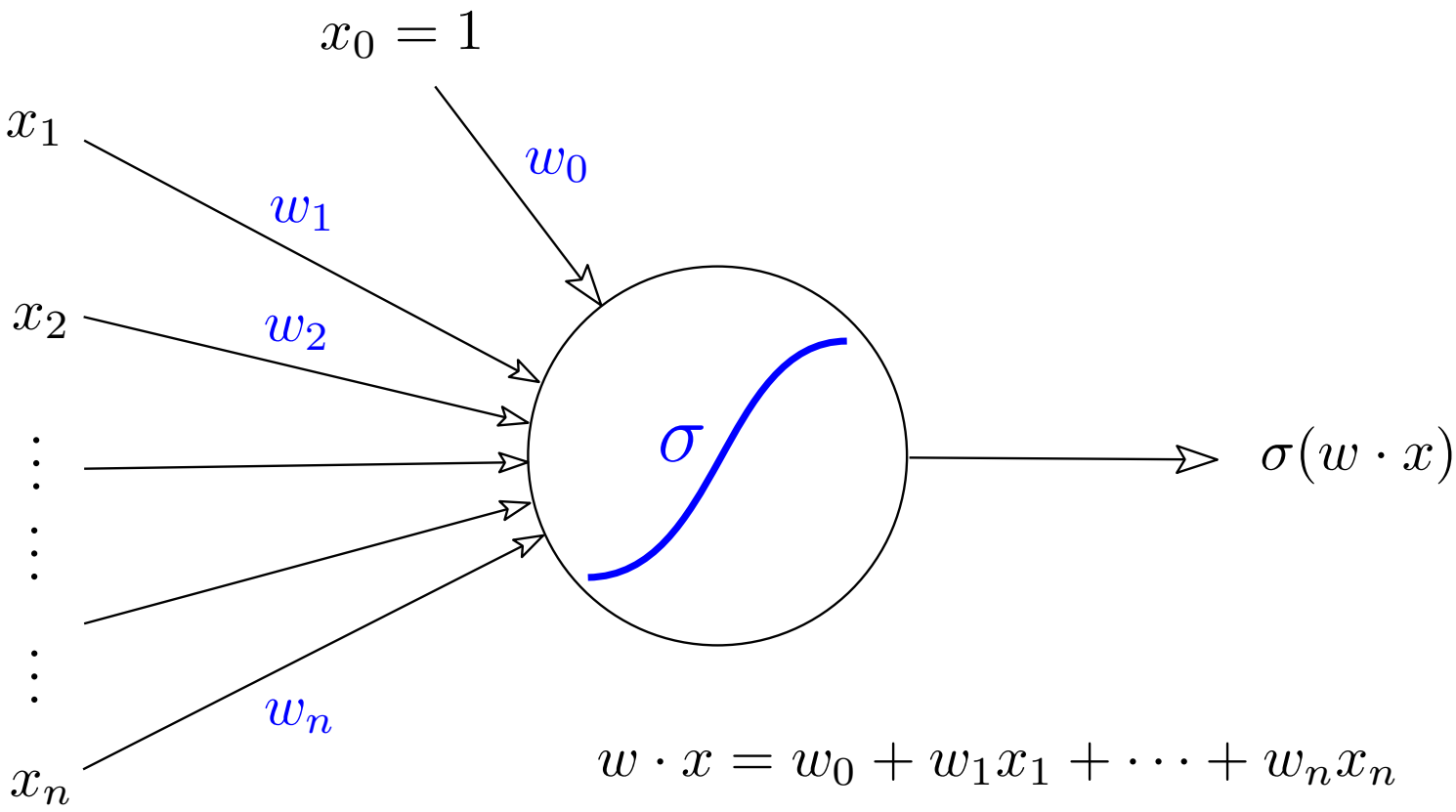


Figure 9.1: Scheme of a neuron. The neuron's output depends on the weights w and on σ (*activation function*), and this functionality is represented by the decorated circle.

In mathematical terms, a neuron is a function

$$x \mapsto f_w(x) = \sigma(x \cdot w), \tag{1}$$

where $w \in \mathbf{R}^n$ (*weights* or *parameters*) and σ is a *sigmoid* function (called *activation function*), like for instance the *logistic function* $\sigma(t) = (1 + e^{-t})^{-1}$, in which case the neuron computes a *logistic regression*.

Augmenting x with $x_0 = 1$ and providing an extra weight w_0 (called the *bias*), the neuron computes $\sigma(w_0 + w_1 x_1 + \cdots + x_n w_n)$.

To display separately the bias and the other weights, we may write f_{w,w_0} or some similar notation.

A *neural network* (NN) can be construed as a *composition of neurons* according to a graph of connections called the *architecture* of the net.

Here we will consider the case of *directed graphs* and thus leaving aside nets based on undirected graphs such as *Boltzmann machines*. Nor will we discuss networks with feedback (those having closed paths), such as the *Hopfield networks*.

The standard architecture of a NN is a directed graph structured in *layers* L_j, as illustrated in Figure 9.2, and its *functional signature* can be condensed as a chain:

$$\mathcal{N}: \quad \text{Input} \to L_0 \xrightarrow{f_1} L_1 \xrightarrow{f_2} \cdots \to L_{d-1} \xrightarrow{f_d} L_d \to \text{Output} \tag{2}$$

The integer d is the *depth* of the net. Conventionally, the net is *deep* if $d > 2$, and *shallow* otherwise. The layers $L_1, \ldots, L_{d-1}$ are considered to be *hidden*, while the input and output layers (L_0 and L_d), are *visible*.

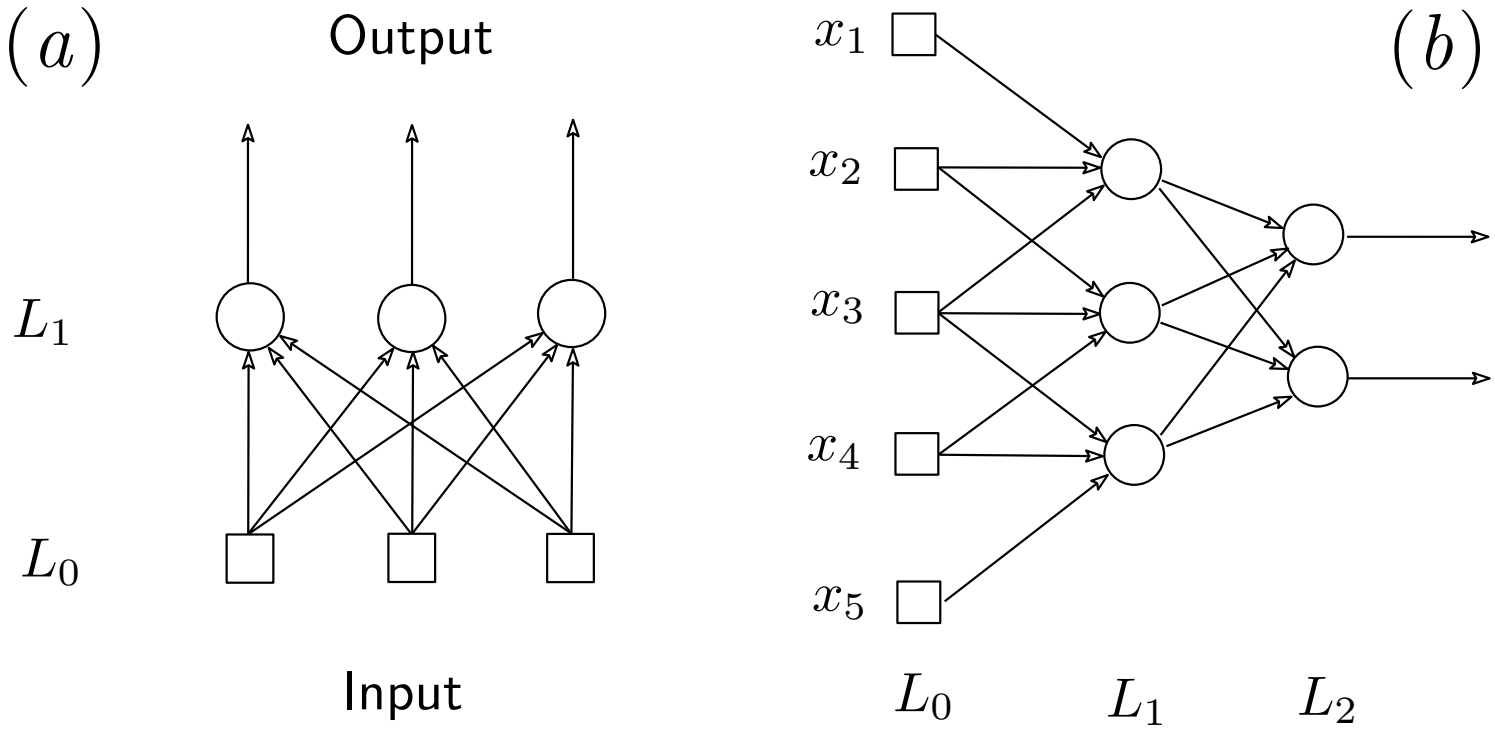


Figure 9.2: (a) Neural network with no hidden neurons and *fully connected*. (b) Network with a hidden layer L_1 of three neurons fully connected to the two output neurons of L_2. The input layer, L_0, is only partially connected to L_1.

The quantities x_j and w_j used in the neuron model introduced in equation (1) are real numbers. But we can imagine that they are entities of an algebraic structure $\mathcal{A}$ sufficient to guarantee that the expression $x \cdot w = x_1 w_1 + \cdots + x_n w_n$, and an activation function $\sigma : \mathcal{A} \to \mathcal{A}$, make sense.

For example, $\mathcal{A}$ can be a real algebra of finite dimension and σ the function of an ordinary sigmoid applied component-wise (with respect to a fixed basis of the algebra). We thus arrive at the concept of $\mathcal{A}$-*neuron* and, connecting neurons as we have done before, to the notion of $\mathcal{A}$-*neural network*, or $\mathcal{A}$-NN.

Another generalization is to replace x and w with more general data structures, such as $\mathcal{A}$-*arrays* (or *tensors*), and the product $x \cdot w$ by a suitable operation $x \star w$. Among these operations, the most commonly used are certain bilinear products, such as *cross-correlation*, as well as nonlinear ones, such as *max-pooling*.

Thus the conventional (artificial) neurons and neural networks are **R**-neurons and **R**-neuronal networks.

Beyond real numbers, among the most immediate concrete cases of algebras $\mathcal{A}$ we can mention **C** (complex numbers), **H** (quaternions), **O** (octonions), an algebra of matrices $\mathbf{R}(n)$, or a geometric algebra $\mathcal{G} = \mathcal{G}_{r,s}$ of signature (r, s). N

To simplify the terminology, we will talk about real, complex (CxNN), quaternionic (QNN), octonionic (ONN), matrix (MNN), and geometric (GNN) networks, respectively.

One advantage of $\mathcal{A}$-neurons is that the number of (real) weights they require decreases in inverse proportion to $d = \dim \mathcal{A}$. The argument is based on the simple observation that $w \in \mathcal{A}$ counts for d real weights, whereas both x and $x' = x \star w$ count for d real parameters each and hence we need d^2 real weights to connect them.

Another advantage is that the algebraic structure of $\mathcal{A}$ can be regarded as a resource for describing and implementing AL algorithms, a point that is particularly relevant when $\mathcal{A}$ is a geometric algebra on account of its intimate connection with the geometry of its geometric space.

Remark. This is not unlike the use of finite fields as alphabets for coding information, for being able to sum, multiply and divide alphabet symbols turns out to represent a great bonus with respect to a set with no structure.

Besides the further references provided henceforth, we find that the text [94] is a remarkably inspiring early reference for most of the topics discussed in this section. In particular, it studies complex NNs in chapter 2 and QNNs in chapter 5. It also features interesting applications of these algebras to predict chaotic time series (chapter 6) and to robotics (chapter 7).

Perhaps the most important idea of these networks is that they can exploit the phase properties of complex numbers.

At the beginning of the study of these networks, the contributions of Hirose and his school stand out. They focus on signal processing, with collections such as [95] (2003) and treated as [96] (2009), or [97] (2012; a second edition of a book of the same title and author published in 2006), and the collection of ten articles collected in [98] (2013), of which the first stands out, by Hirose himself (the editor of the volume), with the title *Application fields and fundamental merits of complex-valued neural networks*.

The text [99] belongs to the same circle, which illustrates with very convincing graphic experiments the value of considering the phase.

More recently we have [100] (2016), on complex convolutional networks; [101] (2017), focused on image classification; [102] (2017), where the emphasis is on deep networks; and [103] (2018), which provides an assessment of complex networks in real signal classification tasks.

Finally we mention [104], which reveals the significance of complex networks from other points of view, particularly that of deep AL, an "umbrella term for emerging techniques that attempt to generalize deep (structured) neural models to non-Euclidean domains such as graphs and varieties."

The interest of these networks comes from the relation that the quaternions keep with the group of rotations of the ordinary Euclidean space, a relation especially transparent in terms of $\mathbf{H} = \mathcal{G}_3^+$, for the expression $\underline{h}(x) = hx\bar{h}$, $h \in \mathbf{H}$ non-zero, is a vector and $\underline{h}$ is a similarity of ratio $|h|^2$ (a rotation if h is unitary).

Another reason is that quaternions have three phases and that these phases can be used to extract valuable information from the signals to be processed.

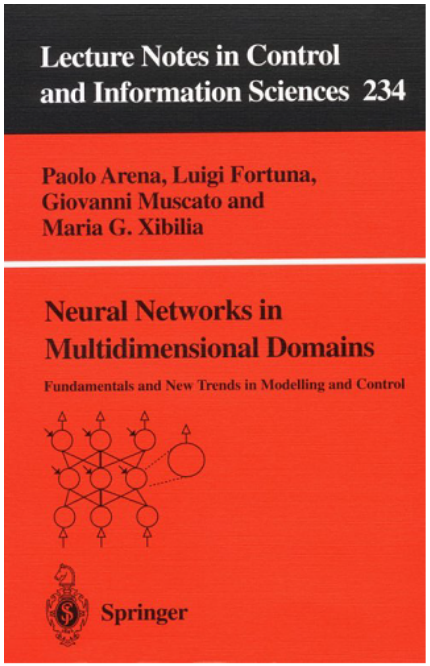
Research in QNNs also comes from long ago, even before that of complex networks. We refer to [105] and [106] for relevant historical information regarding what is called *Clifford's analysis*, especially in relation to Fourier and wavelet transforms in a quaternionic context and their generalization to the geometric context.

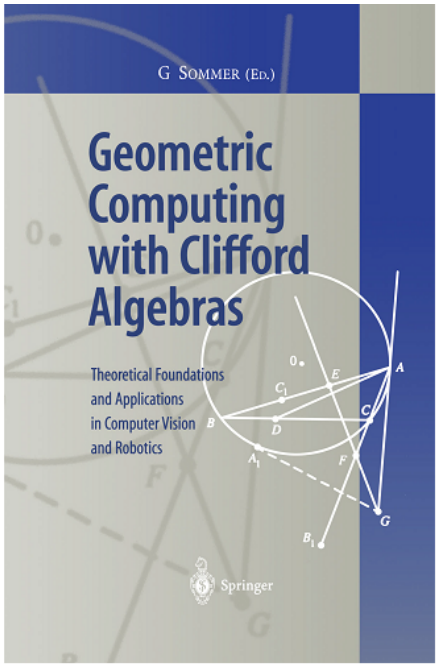
In the origin of the more specific topic we are considering, we find Gerald Sommer and his collaborators: [107] (generalization of Gabor filters) and [108] (generalization of the real multilayer perceptron, cf. [109]).

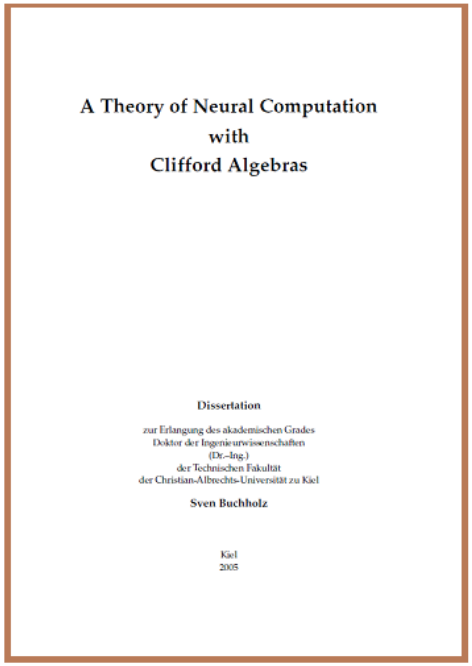
The report [110] presents a quaternion wavelet theory "for image analysis and processing" and [111] an overview of the properties and applications of quaternion networks up to that point.

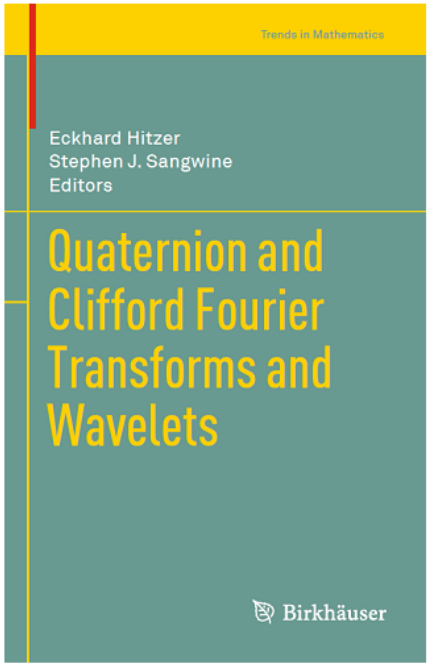
In the last decade, research on QNNs has continued both on the applied and theoretical fronts.

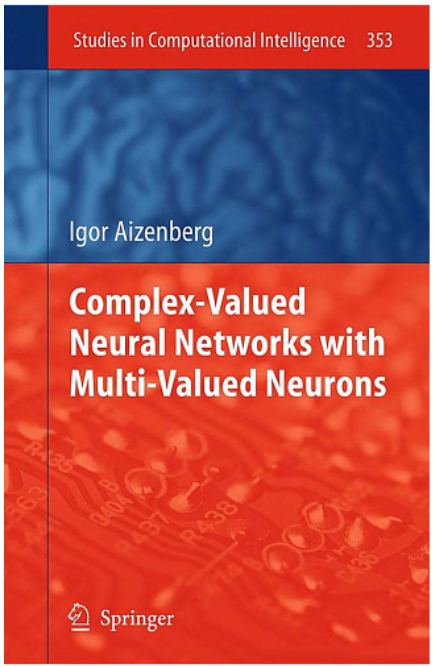
The article [112] deals with the quaternionic multilayer perceptron. Hopfield QNNs and their rotation invariance are investigated in [113]. In the works [114] and [115], the QNNs are applied to the comprehension of the spoken language. Deep QNNs are studied in [116] and convolutional ones in [117]. Finally [55] presents a quaternionic version of capsule networks aimed at processing point clouds in Euclidean space and in [29] a new QNN deterministic layer is introduced that provides contrast invariance and sensitivity to rotation angles using quaternionic Gabor functions and Hilbert transforms, while in [32] the authors use the Riesz transform in the quatermion monogenic representation to propose a novel deterministic convolution layer in the Fourier domain robust to contrast and haze changes in image classification.

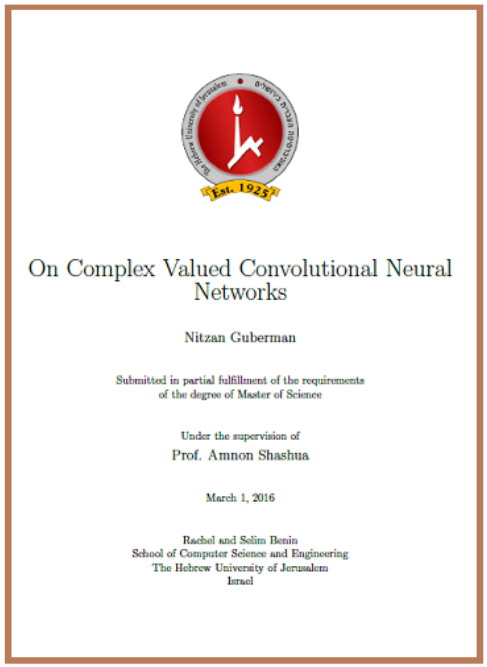

Lecture Notes in Control and Information Sciences 234
Paolo Arena, Luigi Fortuna, Giovanni Muscato and Maria G. Xibilia
Neural Networks in Multidimensional Domains
Fundamentals and New Trends in Modelling and Control
Springer

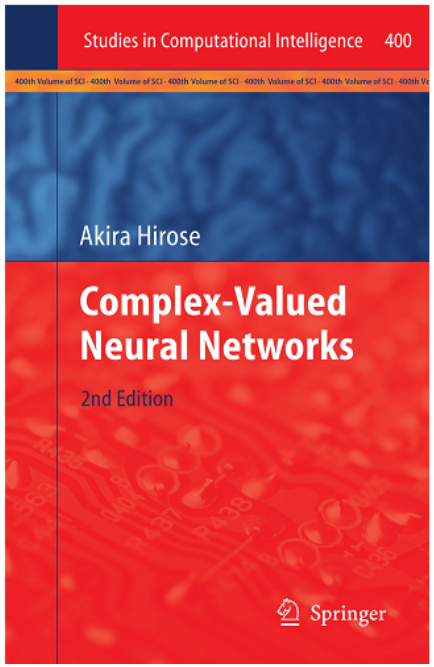

G Sommer (Ed.)
Geometric Computing with Clifford Algebras
Theoretical Foundations and Applications in Computer Vision and Robotics
Springer

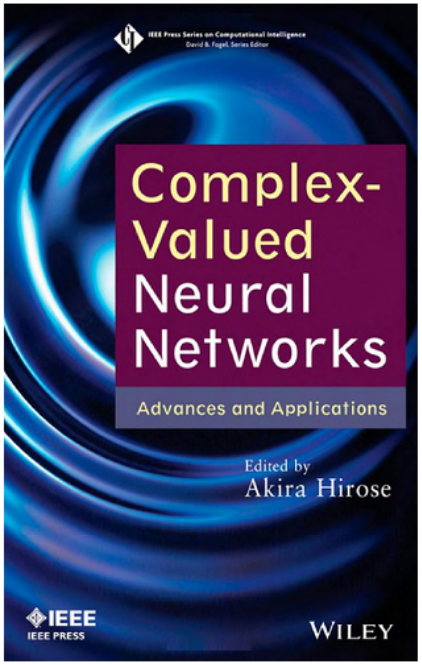

A Theory of Neural Computation with Clifford Algebras
Dissertation
zur Erlangung des akademischen Grades Doktor der Ingenieurwissenschaften (Dr.-Ing.) der Technischen Fakultät der Christian-Albrechts-Universität zu Kiel
Sven Buchholz
Kiel
2005


Trends in Mathematics
Eckhard Hitzer
Stephen J. Sangwine
Editors
Quaternion and Clifford Fourier Transforms and Wavelets
Birkhäuser


Studies in Computational Intelligence 353
Igor Aizenberg
Complex-Valued Neural Networks with Multi-Valued Neurons
Springer


On Complex Valued Convolutional Neural Networks
Nitzan Guberman
Submitted in partial fulfillment of the requirements of the degree of Master of Science
Under the supervision of
Prof. Amnon Shashua
March 1, 2016
Rachel and Selim Benin
School of Computer Science and Engineering
The Hebrew University of Jerusalem
Israel


Studies in Computational Intelligence 400
Akira Hirose
Complex-Valued Neural Networks
2nd Edition
Springer


Complex-Valued Neural Networks
Advances and Applications
Edited by
Akira Hirose
IEEE PRESS
WILEY

Interestingly, the study of GNNs began even before that of QNNs, as in [118], and G. Sommer was a strong proponent of this inquiry at the beginning of the millennium with works such as [119], in which he developed the theoretical foundations that served him well for problems such as artificial vision and robotics; [120], dedicated to a $\mathcal{G}$-version of the multilayer perceptron; [121] and [122], which develop the notion of *monogenic signal*. A culmination of these efforts was Sven Buchholz's thesis, [123], which should be considered, as its title indicates, a theory of neuronal computation with geometric algebras. As a sample of applications, we cite [124] (image segmentation), [125] (support vectors in the geometric context), the volumes [126] (geometric computing for wavelet transforms, artificial vision, learning, control and action) and [127] (geometric computing in engineering and computer science), [128] (use of geometric algebra for edge detection in color images), [129] (clustering methods based on the conformal geometric algebra $\mathcal{G}_{4,1}$), and [47] (treatment of multispectral images with geometric algebra). We end with [130] and [131], a 2-volume set of what should be a systematic treatment on these developments, but see also [132] and [133].

In the recent article [56], convolutional octonion networks are constructed and applied to CIFAR-10 and CIFAR-100 image classification. According to the authors, they have better convergence and accuracy than other networks applied to the same tasks.

Octonions have also been successfully applied to *dictionary learning*, as for instance in [134], an approach that can in fact be formulated for more general algebras, including geometric ones, as in [135].

Another recent example is the case when $\mathcal{A}$ is the algebra of *commutative quaternions*, $\mathbf{H}^c = \langle 1, \boldsymbol{i}_1, \boldsymbol{i}_2, \boldsymbol{i}_2 \rangle$. They were introduced by C. Segre in 1892 (see [136]) and can be defined by the relations $\boldsymbol{i}_1^2 = \boldsymbol{i}_3^2 = -1$, $\boldsymbol{i}_2^2 = 1$, $\boldsymbol{i}_1\boldsymbol{i}_2\boldsymbol{i}_3 = -1$. These imply that $\boldsymbol{i}_1\boldsymbol{i}_2 = \boldsymbol{i}_2\boldsymbol{i}_1 = \boldsymbol{i}_3$, $\boldsymbol{i}_2\boldsymbol{i}_3 = \boldsymbol{i}_3\boldsymbol{i}_2 = \boldsymbol{i}_1$, $\boldsymbol{i}_3\boldsymbol{i}_1 = \boldsymbol{i}_1\boldsymbol{i}_3 = -\boldsymbol{i}_2$, and hence $\mathbf{H}^c$ is commutative. This algebra has been revived in [137] at the level of what, in our notations, would be called $\mathbf{H}^c$-neurons.

Finally let us have a look to the recent paper [57]. In its Abstract we read:

> Our work considers a richer set of objects for activations and weights, and undertakes a comprehensive study of alternative algebras as number representations by studying their performance on two challenging problems: large-scale image classification using the ImageNet dataset and language modeling using the enwiki8 and WikiText-103 datasets. We denote this broader class of models as AlgebraNets. Our findings indicate that the conclusions of prior work, which explored neural networks constructed from **C** (complex numbers) and **H** (quaternions) on smaller datasets, do not always transfer to these challenging settings. However, our results demonstrate that there are alternative algebras which deliver better parameter and computational efficiency compared with **R**. We consider **C**, **H**, $M_2(\mathbf{R})$ (the set of 2×2 real-valued matrices), $M_2(\mathbf{C})$, $M_3(\mathbf{R})$, $M_4(\mathbf{R})$, dual numbers and the $\mathbf{R}^3$ cross product. Additionally, we note that multiplication in these algebras has higher compute density than real multiplication, a useful property in situations with inherently limited parameter reuse such as auto-regressive inference and sparse neural networks.

These are all $\mathcal{A}$-NNs. Are they GNNs?

By our comments on the isomorphism class of $\mathcal{G}_{r,s}$, this is certainly the case for $2\mathbf{R} \simeq \mathcal{G}_{1,0}$,[1] $\mathbf{C} \simeq \mathcal{G}_{0,1}$, $\mathbf{R}(2) \simeq \mathcal{G}_{2,0}$, $\mathbf{H} \simeq \mathcal{G}_{0,2} = \mathcal{G}^{+}_{3,0}$, $\mathbf{C}(2) \simeq \mathcal{G}_{1,2}$, and $\mathbf{R}(4) \simeq \mathcal{G}_{2,2}$.

The exceptions are $\mathbf{R}(3)$ and $(\mathbf{R}^3, \times)$, as their dimensions are not powers of 2. Note, however, that the nature of the latter is also geometric, as the cross product is the Hodge dual of their wedge product, which lives in $\mathcal{G}_{3,0}$. See also [94, Ch. 3] (on Vectorial NNs).

To remark also that although the octonions are not a geometric algebra, they are nevertheless a subalgebra of $\mathcal{G}_{0,7}$ (see [138, §7.4]).

[1] $2\mathbf{R} = \mathbf{R} \oplus \mathbf{R}$ is the algebra of dual numbers, and in general $2\mathcal{A} = \mathcal{A} \oplus \mathcal{A}$. By $\mathcal{A}(n)$, or $M_n(\mathcal{A})$, we denote tha algebra of $n \times n$ matrices with entries in $\mathcal{A}$.

In this section we try to establish some connections between what has been said or hinted before and possible lines of inquiry in the area of AL by means of what can be described, in a broad sense, as *geometric calculi*. Our comments will refer to the following topics:

- AL of mathematical structures.
- Other faces of geometric AL.
- Robotics.
- Computational resources and techniques.
- Recent advances on k-NN (*nearest neighbors*).
- Other liaisons.

A compelling illustration of this theme is reported in [44].

In our view, it represents a line of research that may be promising for AL with geometric calculi as well. Quote from the Abstract:

> Neural networks have a reputation for being better at solving statistical or approximate problems than at performing calculations or working with symbolic data. In this paper, we show that they can be surprisingly good at more elaborated tasks in mathematics, such as *symbolic integration* and solving *differential equations*. We propose a syntax for representing mathematical problems, and methods for generating large datasets that can be used to train sequence-to-sequence models. We achieve results that *outperform* commercial Computer Algebra Systems such as Matlab or Mathematica.

For other works of a similar potential, see [48] (on learning *algebraic structures*), [49] (the bearing of AL on current research in *number theory*), [43] (a kindred report in the realm of *physical sciences*, with many useful insights in various aspects of AL), [58] ("a foray into discrete analogues of Riemannian manifolds, providing a *rich interplay between combinatorics, geometry and theoretical physics*"), [59] (on finding *symbolic equations* that match a given dataset, with the surprising illustration of an "overdensity *equation for dark matter*"), [53] (showing that "neural networks can learn *advanced theorems and complex computations* without built-in mathematical knowledge"), [139] (a version of AL that learns *mappings between function spaces*, with impressive applications to partial differential equations).

Altogether, these works point out to novel avenues for inquiries in *AL* that are *transforming the understanding of science in general, and of mathematics in particular, in ways never seen hitherto*.

For people working in geometric algebra/calculus, it is natural to term AL as "geometric" if based on those formalisms. But AL researchers came up with a different use for this qualification, as in [140]:

> [...] we consider the general question of how to construct deep architectures with small learning complexity on general non-Euclidean domains, which are typically unknown and need to be estimated from the data.

Even more explicit in these appraisals is [141]:

> Geometric deep learning is an umbrella term for emerging techniques attempting to generalize (structured) deep neural models to non-Euclidean domains such as *graphs* and *manifolds*. The purpose of this paper is to overview different examples of geometric deep learning problems and present available solutions, key difficulties, applications, and future research directions in this *nascent field*.

Further evidence for the great potential of this paradigm can be gleaned in the survey [60], whose main thrust lies in *linking graph neural networks* and (neural) *symbolic computing*:

> The need for improved *explainability*, interpretability and trust of AI systems in general demands *principled methodologies*, as suggested by neural-symbolic computing. In this paper, we review the state-of-the-art on the use of GNNs as a model of neural-symbolic computing.

We do not regard the two views of "geometric" that we are considering as antagonistic in any way, as in fact we sense that each can benefit from the other.

We have already mentioned the application of quaternions to robotics presented in [94, Ch. 7].

Among later texts, let us refer to the pioneer book [132], particularly chapters 2 and 7; Selig's treatise *Geometric fundamentals of robotics*, [142]; the collection [133], especially the papers in Part VIII (Geometry and Robotics), and the extensive compilation [126], especially Part IV (Geometric computing of *robot kinematics and dynamics*) and Part VI (Applications II: Robotics and *medical robotics*).

For recent summaries of robotics analyzed with CGA, see [20, Ch. 4] and [79].

Concerning AL in robotics, it has proceeded largely in parallel to the geometric developments, as witnessed by [143] (how machine learning has been applied to robotic path-planning and path-planning related concepts), the survey [144] (reinforcement learning in robotics; see also [145]), Lenz' PhD thesis [146], and the surveys [147] (DL techniques for mobile robot applications), [148] (*DL* methods for robot vision), [149] (learning *control in robotics*).

It appears ever more clearly that advanced AL is playing a major role in robotics aimed at providing all sorts of *assisting services to humans*, as epitomized by the memoir [150].

In all these cases, *the opportunities for applying geometric methods to gain theoretical and applied advantages seem clearly plentiful, if only because of the many engineering aspects that concur in any such system*.

Currently, there is a wealth of software (frameworks) for deep learning (see `Comparison_of_deep-learning_software` in Wikipedia).

For example, *Tensorflow* (see [151]) provides

> ... an interface for expressing machine learning algorithms, and an implementation for executing such algorithms.

Most of them offer a Python interface and increasingly also a *Julia* interface, as for instance *Tensorflow*. An interesting case is Flux (2017), which is pure Julia (framework and interface).

But as far as we know, none of these frameworks can deal with GNNs beyond complex NNs.

On the other hand, there is a rich variety of systems that perform computations with geometric algebras (see, for example, the Software section in the Wikipedia `Geometric_algebra` article).

But again, and as far as we know, none offers a deep learning framework. By its design, the Julia system described in [152] has perhaps the highest potentiality to serve as a basis for developing such a framework.

A first step in this direction would be a framework supporting QNNs. Another useful resource is provided by *template libraries*, as for instance [80].

The authors of [153] also express the view (end of §2.1), that

> the theory [...] for many branches of *unsupervised learning* is still in its infancy.

For our inquiry, there are two main directions to look at.

One concerns recent advances in conventional (non-geometric) unsupervised learning, as for example [154], which orchestrates a powerful scenario for an automatic physicist with no supervision.

In our appraisal, there is much that can conceivably be transferred to other domains, like the strategies that it advocates and the algorithmic ways by which they are marshaled.

For other instances of a similar kind, see [50] (reconstruction of the *periodic table*), [155] (proposing "a family of biologically plausible artificial neural networks (NNs) for unsupervised learning") and [61] (steps "towards the long-term goal of *machine-assisted scientific discovery from experimental data without making prior assumptions about the system*").

The other direction is linking unsupervised learning with GNNs.

Aside from contributions such as the innovative paper [129], which develops a clustering method based on CGA, it appears to be a largely uncharted terrain.

Many of the ideas in the preceding paragraph may be relevant for these explorations.

In this, it may bear further fruits the *unsupervised learning of Lie group transformations* studied in [156] on account of its generality and the geometric character of Lie groups (cf. [19, §6.5]).

Earlier we have met layered $\mathcal{A}$-NNs, but now it is convenient for us to allow more flexible architectures.

By adapting the conventional notions about graph NNs (cf. [157], [158]), we find that a suitable class, among many other possible generalizations, is formed by directed acyclic graphs (N, E) with no isolated nodes and endowed with (trainable) *weights* $w_e \in \mathcal{A}$ $(e \in E)$ and, for each non-initial node n, (trainable) *biases* $b_n \in \mathcal{A}$ and activation functions $\sigma_n : \mathcal{A} \to \mathcal{A}$.

The states of a node are in one-to-one correspondence with elements of $a \in \mathcal{A}$. The initial nodes are input nodes. For a non-initial node n, its state a_n is determined by the formula $a_n = \sigma_n(b_n + \sum_{e:e_0=n} w_e a_{e_1})$, where e_0 and e_1 are the nodes connected by the edge e.

The output of the net is given by the states of the terminal nodes produced by these rules.

In the layered $\mathcal{A}$-NN, the initial (final) nodes are those of L_0 (L_d).

Let us also suggest that it may be productive, particularly in the case of GNNs, to allow that weights w be operators acting on states a according to suitable law $w \star a$ (let us dub $\star$NNs these structures).

These notions draw some inspiration from [20, Ch. 5] and [159] (on oriented CGA and its application to *molecular distance geometry*), and actually it looks puzzling to see whether they could help in porting AL to bear on the problems tackled by molecular distance geometry (see the more specific comments on AL in Chemistry at the end of this section). In doing so, it is important to bear in mind early trailblazers on Clifford neurons such as [160], [161] and [162].

Other areas where the scheme may provide analytic and geometric advantages is in the treatment of *3D point clouds* (see the survey [51], and papers like [163], [81], [164]), as well as in devising more powerful *capsule nets*: see [165], [166], [167], [168].

Of these, only the latter operates with complex numbers. Since CapsNets process elementary patterns, they should benefit from drawing ideas about *pattern theories*, say in the sense of the monograph [169], and also to enhance *explainability* along the lines of [170].

A few hints on the aptness of $\star$NNs to properly deal with *invariance* and *covariance* properties are in order.

These concepts always refer to the action of some group. If a group $G = \{g\}$ acts on a set $\mathcal{X}$, a function $f(x)$ is *invariant* under this action if $f(g \cdot x) = f(x)$ for all $x \in \mathcal{X}$ and $g \in G$.

Similarly, if G also acts on a set $\mathcal{Y}$, a map $f : \mathcal{X} \to \mathcal{Y}$ is *covariant* (or *equivariant*) with respect to the actions of G in $\mathcal{X}$ and $\mathcal{Y}$ if $f(g \cdot x) = g \cdot f(x)$ for all $x \in \mathcal{X}$ and $g \in G$.

Note that an invariant function f is covariant if we make G act trivially on the range of f, so that $g \cdot f(x) = f(x)$ for all $x \in \mathcal{X}$.

The main reason in the context of AL to care about G-covariance is that no *data augmentation* is required to recognize features in arbitrary G-poses, as in [171] for discrete groups of rigid motions.

Let us go back to AL for Chemistry. We note a perceptible $\star$ character of the networks studied in works such as [172], [173], [174], [175], [176], which motivates a careful study of their contributions from the $\star$NN point of view.

See also the collection [177] and especially the paper [178], in which the relevant group is $SE(3)$, the group of distance-preserving transformations of the ordinary Euclidean space. The main claim is that the authors "directly verify that the performance gains are connected with the unique $SE(3)$-*equivariant convolution architecture* of the new model".

Even closer to the spirit of our disquisition is [62], as for us geometric algebras are optimally suited for the treatment of tensors ($\mathcal{A}$-arrays), and many other geometric entities and formalisms as well.

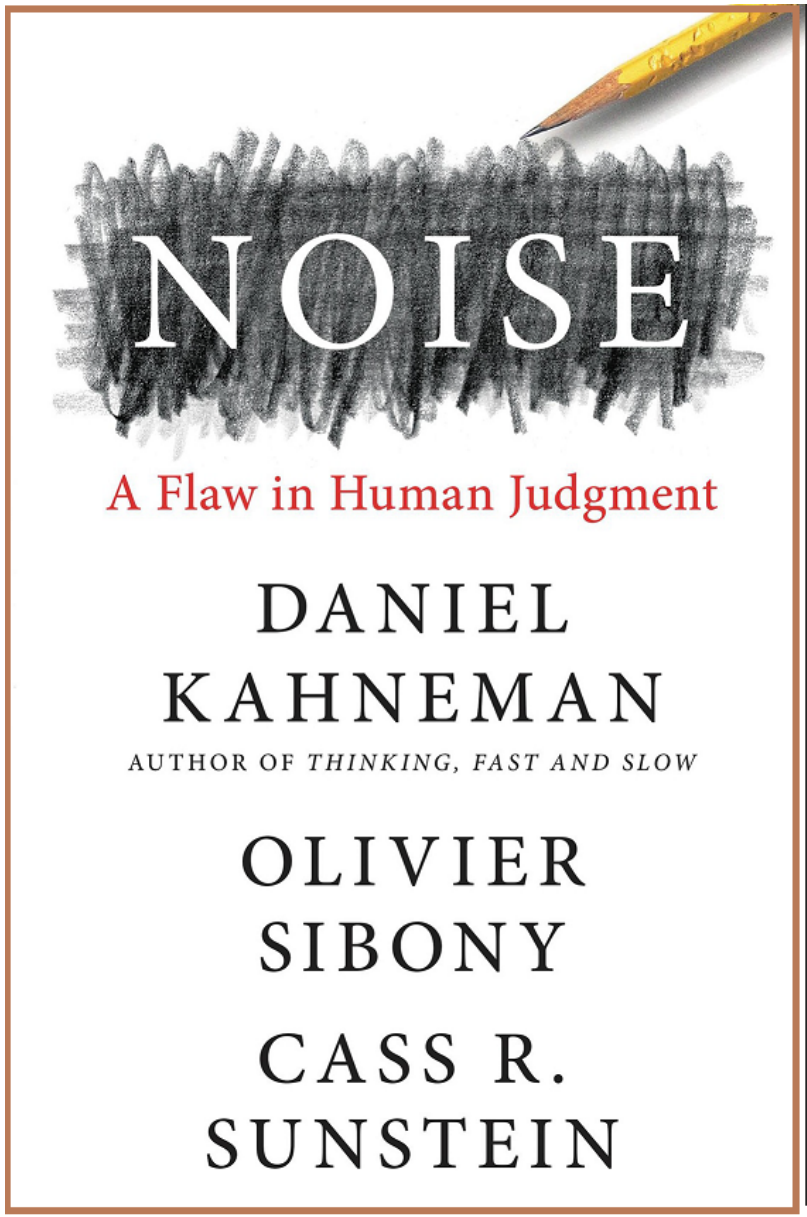
By the way, we note that the AlgebraNets that we have seen before are special cases of $\star$NNs, and that they have been mainly applied to classification problems, but more research could uncover properties and applications based on their geometric character.

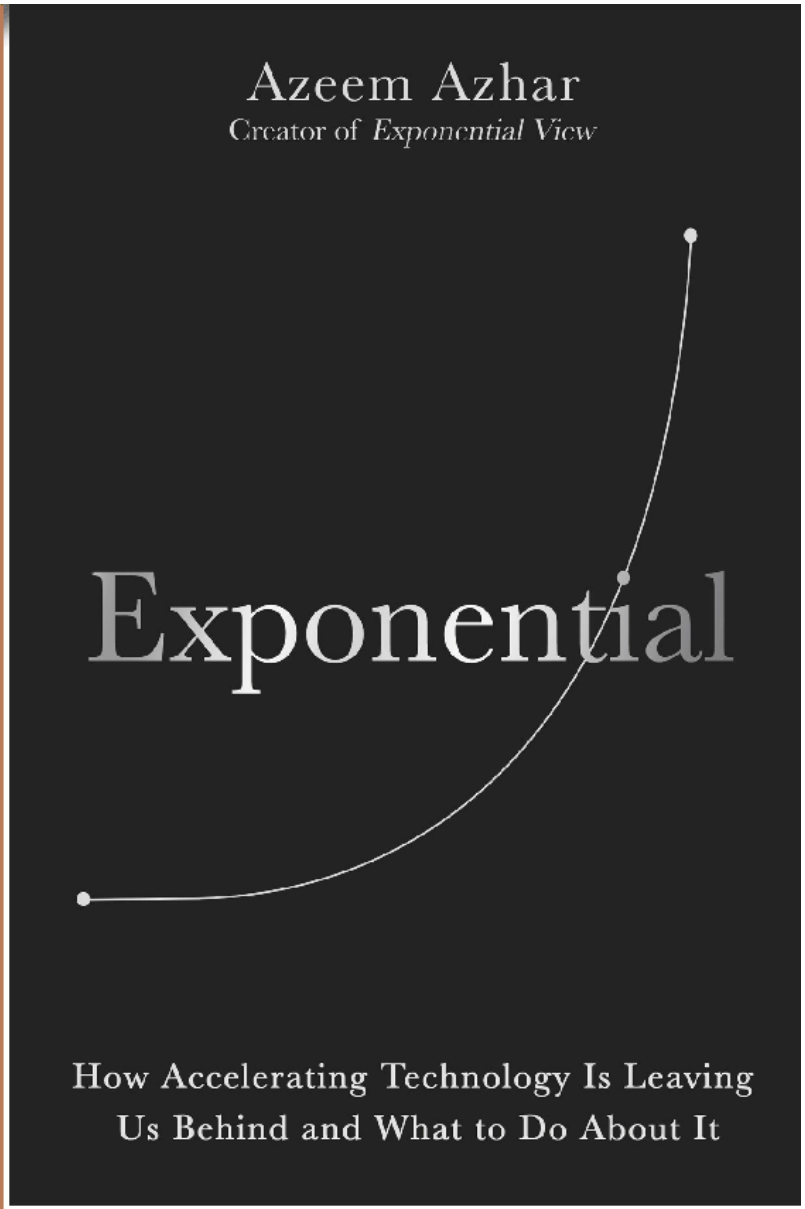
We end with a few remarks on the scattering transforms (a special kind of CNN) introduced in [179] and further studied in [180] (for graph networks), [181] and [182]. The computational side of this transform has produced the system [183].

Altogether, it would be worthwhile to define and study a geometric scattering transform based on the geometric algebra wavelet theory first introduced in [184] and further exploited in [110] (for quaternions), the collection [185] (particularly the paper by P. Cerejeiras, M. Ferreira, and U. Kähler), and [106].

It would also be gainful to devise a scattering transform network that could be trained, both in the conventional sense and in the geometric realm just mentioned, and a computational platform that could deal with both.

Readings for the season's break:





References I

[1] S. Xambó-Descamps, G. Tarragó, G. Perarnau, R. Luna, and F. Lisa, "Algorisme LPBP per a la visió estereoscòpica," 2008.

Final report of the VISIO Project. 115 p.

[2] S. Xambó-Descamps, "A Structured and Comprehensive Package for Computations in Intersection Theory," 2009.

Lecture at the Special Session on Computational and Analytic Geometry of the 2009 AMS/JMM in Washington DC (5-8 January, 2009). Slides.

References II

[3] S. Xambó-Descamps, "A Clifford perspective on Klein's geometry," 2009.

Lecture at the Didactics of Mathematics as a Mathematical Discipline. Universidade da Madeira, Funchal, Portugal (1-4 October, 2009). Slides.

[4] S. Xambó-Descamps, "A mathematical view of quantum computation," 2009.

Lecture at the "Quantum Computation and Information Seminar", Security and Quantum Information Group, Technical University of Lisbon (9 October 2009). Slides.

References III

[5] S. Xambó-Descamps and J. e. Aranda, "Plenary and Invited Lectures," 2010.

2010 IEEE WORLD CONGRESS ON COMPUTATIONAL INTELLIGENCE (WCCI 2010, July 18-23, Barcelona, Spain). Published by the Institute of Electrical and Electronics Engineering, Inc., 2010, x + 261 pp.

[6] S. Xambó-Descamps, S. D. Odintsov, and D. Sáez, *Cosmology, Quantum Vacuum and Zeta Functions*.

Proceedings of Physics, Springer, 2011.

Proceedings of the Conference to celebrate Emili Elizalde's 60th birthday (Mars 8-10, 2010). xii + 382 p.

References IV

[7] S. Xambó-Descamps, "Mathematical Essentials of Quantum Computing," 2013.

pdf (50 pp). Summarized in poster. See also the Ateneo Colloquium at the IMUVA (24 November 2013), Slides.

[8] S. Xambó-Descamps, "David Mumford: diálogo y semblanza," La Gaceta *de la RSME*, vol. 16, no. 4, pp. 643–668, 2013.

pdf.

[9] S. Xambó-Descamps and G. Roma, "Procesamiento de Patrones en Python," 2013.

Lecture at the Conference PyConEs-2013, Universidad Politécnica de Madrid (23-24 November 2013). Slides.

References V

[10] S. Xambó-Descamps, "Escondidas sendas de la geometría proyectiva a los formalismos cuánticos," in *El legado matemático de Juan Bautista Sancho Guimerá*, pp. 233–274, Real Sociedad Matemática Española y Ediciones Universidad de Salamanca, 2015.

pdf.

[11] S. Xambó-Descamps, "Review of Hestenes' Space-Time Algebra *(second edition)*," *EMS Newsletter*, vol. 98, pp. 63–66, 2015.

pdf.

[12] S. Xambó-Descamps, "Geometric Algebra Techniques in Mathematics, Physics and Engineering, A," 2015.

Intensive course at the IMUVA: 1 Firsts steps, 2 Grassmann and Clifford algebras, 3 Geometric Algebra, and 4 Space-time algebra.

References VI

[13] S. Xambó-Descamps, "Geometric Algebra Techniques in Mathematics, Physics and Engineering, B," 2016.

Intensive course at the IMUVA: 5 Classical Mechanics, 6 Quantum Esperanto and Pauli spinors, 7 Conformal Geometric Algebra 8. Classifications.

[14] S. Xambó-Descamps, "*17th Lluís Santaló Research Summer School:* Geometric Algebra and Geometric Calculus, with Applications in Mathematics, Physics and Engineering (director," 2016.

The speakers, other than the Director, were D. Hestenes, J. Lasenby, L. Dorst, and A. Lasenby. For more details, see LLuís Santaló 2016.

References VII

[15] S. Xambó-Descamps, "*Lectures at the* 17th "Lluís Santaló" Research Summer School," 2016.

1. Algebra and geometry in current curricula, 2. On axiom systems for GA, 3. A view of F. KLEIN Erlangen Program through GA and 4. Enriching Abstract Algebra with GA.

[16] S. Xambó-Descamps, "Proceedings of the AGACSE 2015 Conference, *Barcelona, July 27-31, 2015 (Editor in Chief)*," 2017.

Special issue of *Advances of Applied Clifford Algebras* **27**/1 (2017), 550 pp. Springer link.

[17] S. Xambó-Descamps, "Waldyr Rodrigues Jr.: Sketches on his Life and Work," 2017.

Invited plenary lecture at ICCA11 (Ghent, 9 August 2017). pdf.

References VIII

[18] S. Xambó-Descamps and S. Wainer, "Waldyr Alves Rodrigues Jr.: Fruits of a unifying philosophy," 2018.

Plenary lecture at AGACSE 2018 (July 23, 2018). Slides.

[19] S. Xambó-Descamps, *Real spinorial groups—a short mathematical introduction*.

SBMA/Springerbrief, Springer, 2018.

[20] C. Lavor, S. Xambó-Descamps, and I. Zaplana, *A Geometric Algebra Invitation to Space-Time Physics, Robotics and Molecular Geometry*.

SBMA/Springerbrief, Springer, 2018.

References IX

[21] S. Xambó-Descamps, "From Leibniz' *characteristica geometrica* to contemporary geometric algebra," *Quaderns d'Història de l'Enginyeria*, vol. 16 , no. Special issue dedicated to commemorate Leibniz (1646-1716), pp. 103–134, 2018.

pdf.

[22] S. Xambó-Descamps, G.-M. Greuel, and L. Narváez (eds), "Singularities, Algebraic Geometry, Commutative Algebra, and Related Topics. Festschrift for Antonio Campillo on the Occasion of his 65th Birthday," 2018.

Springer, 2018 (27 papers, xvi + 604 pages). Summary and Springer link.

References X

[23] S. Xambó-Descamps, "Proceedings of the AGACSE 2018 Conference, *Campinas, July 27-31, 2018 (Editor in Chief)*," 2019.

Published as a Topical Collection in *Advances in Applied Clifford Algebras*.

[24] S. Xambó-Descamps, "A Light Dream," in *Proceedings of the 2019 Interdisciplinary Colloquium in Topology and its Applications* (M. Bruguera, M. J. Chasco, and X. E. Domínguez, eds.), pp. 40–45, Servizo de Publicacións da Universidade de Vigo, 2019.

pdf.

References XI

[25] S. Xambó-Descamps and J. Franch, "Alessio Figalli, Fields Medal 2018," 2019.

Poster exhibit to celebrate the DHC of Alessio FIGALLI by UPC (pdf).

[26] S. Xambó-Descamps, "Post-quantum Cryptography," 2019.

Talk at the "V Jornadas Doctorales del Programa de Doctorado en Matemáticas (UCA, November 2019)". Slides.

[27] S. Xambó-Descamps, "Geometric Algebra: Mathematical Structures and Applications," 2019.

A 2-hour seminar at the UNED (3-4 October, 2019). Slides.

References XII

[28] S. Xambó-Descamps and N. Sayols, "Computer algebra tales on Goppa codes and McEliece cryptography," *Mathematics in Computer Science*, vol. 14, no. 2, pp. 457–469, 2020.

Springer link. pdf.

[29] E. U. Moya-Sánchez, S. Xambó-Descamps, A. S. Pérez, S. Salazar-Colores, J. Martínez-Ortega, and U. Cortés, "A bio-inspired quaternion local phase CNN layer with contrast invariance and linear sensitivity to rotation angles," *Pattern Recognition Letters*, vol. 131, pp. 56–62, 2020.

References XIII

[30] S. Xambó-Descamps (ed), *Systems, Patterns and Data Engineering with Geometric Calculi*.

ICIAM2019 SEMA SIMAI Springer Series, Springer, 2021.

Springer link.

[31] S. Xambó-Descamps and E. U. Moya, "Geometric calculi and automatic learning–An outline," in *Systems, patterns and data engineering with geometric calculi* (S. Xambó-Decamps, ed.), vol. 13 of *ICIAM2019 SEMA SIMAI*, pp. 153–178, Springer, 2021.

Springer link.

References XIV

[32] E. U. Moya-Sánchez, S. Xambó-Descamps, S. Salazar Colores, A. Sánchez Pérez, and U. Cortés, "A Quaternion Deterministic Monogenic CNN Layer for Contrast Invariance," in *Systems, Patterns and Data Engineering with Geometric Calculi (see [30])*, pp. 133–152, Springer, 2021.

Springer link.

[33] E. U. Moya-Sánchez, S. Xambo-Descamps, A. Sánchez, S. Salazar-Colores, and U. Cortés, "A trainable monogenic ConvNet layer robust in front of large contrast changes in image classification," *IEEEAccess*, 2021.

Accepted Oct 15, 2021.
`https://arxiv.org/pdf/2109.06926.pdf`.

References XV

[34] J. Bruna and S. Xambó-Descamps, "Aprenentatge algorísmic i xarxes neuronals profundes," Butlletí de la SCM, vol. 36, no. 1, pp. 5–67, 2021.

[35] J. Bruna and S. Xambó-Descamps, "Mathematical Aspects of Algorithmic Learning and Deep Neural Networks," 2021.

BGSMath/CRM course.

[36] P. Bayer, "Música algorítmica: experiments i perspectives," 2021.

Imaginary, una simfonia matemàtica: Jornada música i matemàtiques a l'FME. Lecture video.

References XVI

[37] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[38] A. Wigderson, *Mathematics and computation*.

Princeton University Press, 2019.

https://www.math.ias.edu/files/mathandcomp.pdf, 372 p.

[39] G. Nguyen, S. Dlugolinsky, M. Bobák, V. Tran, Á. L. García, I. Heredia, P. Malík, and L. Hluchỳ, "Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey," *Artificial Intelligence Review*, vol. 52, no. 1, pp. 77–124, 2019.

References XVII

[40] B. Marr, "The Key Definitions Of Artificial Intelligence (AI) That Explain Its Importance," Consulted 2021.

`https://bernardmarr.com/the-key-definitions-of-artificial-intelligence-ai-that-explai`

[41] G. Strang, *Linear algebra and learning from data*.

Wellesley-Cambridge Press, 2019.

[42] C. Jin, P. Netrapalli, R. Ge, S. M. Kakade, and M. I. Jordan, "On nonconvex optimization for machine learning: Gradients, stochasticity, and saddle points," 2019.

`https://arxiv.org/pdf/1902.04811.pdf`.

References XVIII

[43] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, "Machine learning and the physical sciences," *Reviews of Modern Physics*, vol. 91, no. 4, p. 045002, 2019.

`https://arxiv.org/pdf/1903.10563.pdf`.

[44] G. Lample and F. Charton, "Deep learning for symbolic mathematics," 2019.

`https://arxiv.org/pdf/1912.01412.pdf`.

[45] S. S. Du, X. Zhai, B. Poczos, and A. Singh, "Gradient descent provably optimizes over-parameterized neural networks," 2019.

Published as a conference paper at ICLR 2019:
`https://arxiv.org/pdf/1810.02054.pdf`.

References XIX

[46] S. Du, J. Lee, H. Li, L. Wang, and X. Zhai, "Gradient descent finds global minima of deep neural networks," in *International Conference on Machine Learning*, pp. 1675–1685, PMLR, 2019.

http://proceedings.mlr.press/v97/du19c/du19c.pdf.

[47] R. Wang, Y. Shi, and W. Cao, "GA-SURF: A new speeded-up robust feature extraction algorithm for multispectral images based on geometric algebra," *Pattern Recognition Letters*, vol. 127, pp. 11–17, 2019.

[48] Y.-H. He and M. Kim, "Learning algebraic structures: Preliminary investigations," 2019.

https://arxiv.org/pdf/1905.02263.pdf.

[49] L. Alessandretti, A. Baronchelli, and Y.-H. He, "Machine learning meets number theory: The data science of Birch-Swinnerton-Dyer," 2019.

https://arxiv.org/pdf/1911.02008.pdf.

References XX

[50] M. Kusaba, C. Liu, Y. Koyama, K. Terakura, and R. Yoshida, "Recreation of the periodic table with an unsupervised machine learning algorithm," 2019.

https://arxiv.org/ftp/arxiv/papers/1912/1912.10708.pdf.

[51] Y. Guo, H. Wang, Q. Hu, H. Liu, L. Liu, and M. Bennamoun, "Deep Learning for 3D Point Clouds: A Survey," 2019.

https://arxiv.org/pdf/1912.12033.pdf.

[52] D. Garcia-Gasulla, A. Cortés, S. Alvarez-Napagao, and U. Cortés, "Signs for ethical AI: A route towards transparency," 2020.

https://arxiv.org/pdf/2009.13871.pdf.

References XXI

[53] F. Charton, A. Hayat, and G. Lample, "Deep differential system stability–learning advanced computations from examples," 2020.

https://arxiv.org/pdf/2006.06462.pdf.

[54] E. Weinan, C. Ma, S. Wojtowytsch, and L. Wu, "Towards a mathematical understanding of neural network-based machine learning: what we know and what we don't," *CSIAM Transactions on Applied Mathematics*, vol. 1, no. 4, 2020.

https://arxiv.org/pdf/2009.10713.pdf.

[55] Y. Zhao, T. Birdal, J. E. Lenssen, E. Menegatti, L. Guibas, and F. Tombari, "Quaternion Equivariant Capsule Networks for 3D Point Clouds," 2019.

http://arxiv.org/abs/1912.12098, v2, 2020.

References XXII

[56] J. Wu, L. Xu, F. Wu, Y. Kong, L. Senhadji, and H. Shu, "Deep octonion networks," *Neurocomputing*, vol. 397, pp. 179–181, July 2020.

[57] J. Hoffmann, S. Schmitt, S. Osindero, K. Simonyan, and E. Elsen, "Algebranets," 2020.

`https://arxiv.org/pdf/2006.07360.pdf`.

[58] Y.-H. He and S.-T. Yau, "Graph laplacians, riemannian manifolds and their machine-learning," 2020.

`https://arxiv.org/pdf/2006.16619.pdf`.

[59] M. Cranmer, A. Sanchez-Gonzalez, P. Battaglia, R. Xu, K. Cranmer, D. Spergel, and S. Ho, "Discovering symbolic models from deep learning with inductive biases," 2020.

`https://arxiv.org/pdf/2006.11287v1.pdf`.

References XXIII

[60] L. Lamb, A. Garcez, M. Gori, M. Prates, P. Avelar, and M. Vardi, "Graph neural networks meet neural-symbolic computing: A survey and perspective," 2020.

https://arxiv.org/pdf/2003.00330.pdf.

[61] R. Iten, T. Metger, H. Wilming, L. Del Rio, and R. Renner, "Discovering physical concepts with neural networks," *Physical Review Letters*, vol. 124, no. 1, p. 010508, 2020.

[62] R. J. Townshend, B. Townshend, S. Eismann, and R. O. Dror, "Geometric prediction: Moving beyond scalars," 2020.

https://arxiv.org/pdf/2006.14163.pdf.

[63] L. Melas-Kyriazi, "The Mathematical Foundations of Manifold Learning," 2020.

https://arxiv.org/pdf/2011.01307.pdf.

References XXIV

[64] P. Scholze, "Liquid tensor experiment," 2020.

Xena guest post: *Xena* guest post, and pdf version.

[65] M. Raghu and E. Schmidt, "A survey of deep learning for scientific discovery," 2020.

`https://arxiv.org/pdf/2003.11755`.

[66] K. Heal, A. Kulkarni, and E. C. Sertöz, "Deep learning gauss-manin connections," 2020.

`https://arxiv.org/pdf/2007.13786.pdf`.

[67] M. C. Hughes, "A neural network approach to predicting and computing knot invariants," *Journal of Knot Theory and Its Ramifications*, vol. 29, no. 03, p. 2050005, 2020.

`https://arxiv.org/pdf/1610.05744.pdf`.

References XXV

[68] W. Jiang and J. Luo, "Graph neural network for traffic forecasting: A survey," 2021.

https://arxiv.org/pdf/2101.11174.pdf.

[69] B. Ghojogh, A. Ghodsi, F. Karray, and M. Crowley, "Laplacian-based dimensionality reduction including spectral clustering, laplacian eigenmap, locality preserving projection, graph embedding, and diffusion map: Tutorial and survey," 2021.

https://arxiv.org/pdf/2106.02154.pdf.

[70] B. Scellier, *A deep learning theory for neural networks grounded in physics*.

PhD thesis, Département d'informatique et de recherche opérationnelle, Faculté des arts et des sciences, Université de Montréal, 2021.

https://arxiv.org/pdf/2103.09985.pdf.

References XXVI

[71] F. Bach, "Learning theory from first principles, draft," 2021.

https://www.di.ens.fr/~fbach/ltfp_book.pdf. xii+260 p.

[72] M. M. Bronstein, J. Bruna, T. Cohen, and P. Veličković, "Geometric deep learning: Grids, groups, graphs, geodesics, and gauges," 2021.

https://arxiv.org/pdf/2104.13478.pdf.

[73] S. Mei, T. Misiakiewicz, and A. Montanari, "Learning with invariances in random features and kernel models," 2021.

https://arxiv.org/pdf/2102.13219.pdf.

[74] T. Cohen, *Equivariant Convolutional Networks*.

PhD thesis, University of Amsterdam, 2021.

https://pure.uva.nl/ws/files/60770359/Thesis.pdf.

References XXVII

[75] D. Beniaguev, I. Segev, and M. London, "Single cortical neurons as deep artificial neural networks," *Neuron*, vol. 109, no. 17, pp. 2727–2739, 2021.

[76] A. Z. Wagner, "Constructions in combinatorics via neural networks," 2021.

https://arxiv.org/pdf/2104.14516.pdf.

[77] J. Berner, P. Grohs, G. Kutyniok, and P. Petersen, "The modern mathematics of deep learning," 2021.

https://arxiv.org/pdf/2105.04026.pdf.

[78] O. Smirnov, "TensorFlow RiemOpt: a library for optimization on Riemannian manifolds," 2021.

https://arxiv.org/pdf/2105.13921.

References XXVIII

[79] I. Zaplana, "New Perspectives on Robotics with Geometric Calculus," in *Systems, Patterns and Data Engineering with Geometric Calculi* (A. D. et al., ed.), ICIAM2019 SEMA SIMAI Springer Series, pp. 1–17, Springer, 2021.

[80] L. A. F. Fernandes, "Exploring Lazy Evaluation and Compile-Time Simplifications for Efficient Geometric Algebra Computations," in *Systems, Patterns and Data Engineering with Geometric Calculi* (A. D. et al., ed.), ICIAM2019 SEMA SIMAI Springer Series, Springer, 2021.

111-131.

[81] S. Franchini and S. Vitabile, "Geometric Calculus Applications to Medical Imaging: Status and Perspectives," in *Systems, Patterns and Data Engineering with Geometric Calculi* (A. D. et al., ed.), ICIAM2019 SEMA SIMAI Springer Series, pp. 31–46, Springer, 2021.

References XXIX

[82] G. Raayoni, S. Gottlieb, Y. Manor, G. Pisha, Y. Harris, U. Mendlovic, D. Haviv, Y. Hadad, and I. Kaminer, "Generating conjectures on fundamental constants with the ramanujan machine," *Nature*, vol. 590, no. 7844, pp. 67–73, 2021.

[83] A. Payeur, J. Guerguiev, F. Zenke, B. A. Richards, and R. Naud, "Burst-dependent synaptic plasticity can coordinate learning in hierarchical circuits," *Nature neuroscience*, pp. 1–10, 2021.

[84] A. Davies, P. Velickovic, L. Buesing, S. Blackwell, D. Zheng, N. Tomasev, R. Tanburn, P. Battaglia, C. Blundell, A. Juhasz, *et al.*, "Advancing mathematics by guiding human intuition with ai," *Nature*, 2021.

`https://www.nature.com/articles/s41586-021-04086-x`.

References XXX

[85] C. Blundell, L. Buesing, A. Davis, P. Velickovic, and G. Williamson, "Towards combinatorial invariance for Kazhdan-Lusztig polynomials," 2021.

`https://arxiv.org/pdf/2111.15161.pdf`.

[86] A. Davies, M. Lackenby, and N. Tomasev, "The signature and cusp geometry of hyperbolic knots," 2021.

`https://arxiv.org/pdf/2111.15323.pdf`.

[87] P. Scholze, "Half a year of the Liquid Tensor Experiment: amazing developments *Xena*," 2021.

Half-a-year history of the LTE.

[88] Y.-H. He, "Machine-learning mathematical structures," 2021.

`https://arxiv.org/pdf/2101.06317.pdf`.

References XXXI

[89] J. Avigad, L. de Moura, and S. Kong, "Theorem proving in Lean *(Release 3.23.0)*," 2021.

Tutorial.

[90] M. U. Yaseen, A. Anjum, G. Fortino, A. Liotta, and A. Hussain, "Cloud based scalable object recognition from video streams using orientation fusion and convolutional neural networks," *Pattern Recognition*, vol. 121, p. 108207, 2022.

`https://arxiv.org/pdf/2106.15329.pdf`.

[91] S. B. McGrayne, *The theory that would not die*.

Yale University Press, 2011.

References XXXII

[92] N. Silver, *The signal and the noise: why so many predictions fail–but some don't*.

Penguin, 2012.

[93] S. Watanabe, *Algebraic geometry and statistical learning theory*.

No. 25 in Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, 2009.

[94] P. Arena, L. Fortuna, G. Muscato, and M. G. Xibilia, *Neural networks in multidimensional domains: fundamentals and new trends in modelling and control*, vol. 234 of *LNCIS*.

Springer, 1998.

References XXXIII

[95] A. Hirose, *Complex-valued neural networks: theories and applications*, vol. 5.

World Scientific, 2003.

[96] T. Nitta, *Complex-valued neural networks: Utilizing high-dimensional parameters*.

IGI Global, 2009.

[97] A. Hirose, *Complex-valued neural networks (second edition)*.

Springer, 2012.

Japanese edition 2004, first English edition 2006.

[98] A. e. Hirose, *Complex-valued neural networks: Advances and applications*.

IEEE Computational Intelligence, John Wiley & Sons, 2013.

References XXXIV

[99] I. Aizenberg, *Complex-valued neural networks with multi-valued neurons*, vol. 353.

Springer, 2011.

[100] N. Guberman, "On complex valued convolutional neural networks," 2016.

`https://arxiv.org/pdf/1602.09046.pdf`.

[101] C.-A. Popa, "Complex-valued convolutional neural networks for real-valued image classification," in *2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 816–822, IEEE, 2017.

[102] C. Trabelsi, O. Bilaniuk, Y. Zhang, D. Serdyuk, S. Subramanian, J. F. Santos, S. Mehri, N. Rostamzadeh, Y. Bengio, and C. J. Pal, "Deep complex networks," *arXiv: 1705. 09792*, 2017.

References XXXV

[103] N. Mönning and S. Manandhar, "Evaluation of complex-valued neural networks on real-valued classification tasks," *arXiv: 1811. 12351*, 2018.

[104] J. Bruna, S. Chintala, Y. LeCun, S. Piantino, A. Szlam, and M. Tygert, "A mathematical motivation for complex-valued convolutional networks," *arXiv: 1503. 03438*, 2015.

[105] F. Brackx, E. Hitzer, and S. J. Sangwine, "History of quaternion and Clifford-Fourier transforms and wavelets," *Quaternion and Clifford Fourier transforms and wavelets*, vol. 27, pp. XI–XXVII, 2013.

[106] E. Hitzer and S. J. Sangwine, *Quaternion and Clifford Fourier transforms and wavelets*.

Springer, 2013.

References XXXVI

[107] T. Bulow and G. Sommer, "Quaternionic Gabor filters for local structure classification," in *Proceedings. Fourteenth International Conference on Pattern Recognition (Cat. No. 98EX170)*, vol. 1, pp. 808–810, IEEE, 1998.

[108] S. Buchholz and G. Sommer, "Quaternionic spinor MLP," in *ESANN'2000 Proceedings*, pp. 377–382, D-Facto, 2000.

European Symposium on Artificial Neural Networks, Bruges (Belgium), 26-28 April 2000.

[109] S. Haykin, "Neural networks and learning machines," 2009.

xxx + 906 pp.

[110] W. L. Chan, H. Choi, and R. Baraniuk, "Quaternion wavelets for image analysis and processing," in *IEEE International Conference on Image Processing*, vol. 5, pp. 3057–3060, 2004.

References XXXVII

[111] T. Isokawa, N. Matsui, and H. Nishimura, "Quaternionic neural networks: Fundamental properties and applications," in *Complex-valued neural networks: utilizing high-dimensional parameters*, pp. 411–439, IGI Global, 2009.

[112] T. Isokawa, H. Nishimura, and N. Matsui, "Quaternionic multilayer perceptron with local analyticity," *Information*, vol. 3, no. 4, pp. 756–770, 2012.

`https://arxiv.org/pdf/1901.09342`.

[113] M. Kobayashi, "Rotational invariance of quaternionic Hopfield neural networks," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 11, no. 4, pp. 516–520, 2016.

References XXXVIII

[114] T. Parcollet, M. Morchid, P.-M. Bousquet, R. Dufour, G. Linarès, and R. De Mori, "Quaternion neural networks for spoken language understanding," in *2016 IEEE Spoken Language Technology Workshop (SLT)*, pp. 362–368, IEEE, 2016.

[115] T. Parcollet, Y. Zhang, M. Morchid, C. Trabelsi, G. Linarès, R. De Mori, and Y. Bengio, "Quaternion convolutional neural networks for end-to-end automatic speech recognition," *arXiv: 1806. 07789*, 2018.

[116] C. J. Gaudet and A. S. Maida, "Deep quaternion networks," in *2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, IEEE, 2018.

[117] X. Zhu, Y. Xu, H. Xu, and C. Chen, "Quaternion convolutional neural networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 631–647, 2018.

References XXXIX

[118] J. K. Pearson and D. L. Bisset, "Neural networks in the Clifford domain," in *Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94)*, vol. 3, pp. 1465–1469, IEEE, 1994.

[119] G. Sommer (ed.), *Geometric computing with Clifford algebras: theoretical foundations and applications in computer vision and robotics*.

Springer, 2001.

[120] S. Buchholz and G. Sommer, "Clifford algebra multilayer perceptrons," in *Geometric computing with Clifford algebras*, pp. 315–334, Springer, 2001.

[121] M. Felsberg and G. Sommer, "The monogenic signal," *IEEE Transactions on Signal Processing*, vol. 49, no. 12, pp. 3136–3144, 2001.

References XL

[122] T. Bülow and G. Sommer, "Hypercomplex signals – a novel extension of the analytic signal to the multidimensional case," *IEEE Transactions on signal processing*, vol. 49, no. 11, pp. 2844–2852, 2001.

[123] S. Buchholz, *A theory of neural computation with Clifford algebras*.

PhD thesis, Christian-Albrechts Universität Kiel, 2005.

[124] J. Rivera-Rovelo and E. Bayro-Corrochano, "Medical image segmentation using a self-organizing neural network and Clifford geometric algebra," in *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, pp. 3538–3545, IEEE, 2006.

[125] E. J. Bayro-Corrochano and N. Arana-Daniel, "Clifford support vector machines for classification, regression, and recurrence," *IEEE Transactions on Neural Networks*, vol. 21, no. 11, pp. 1731–1746, 2010.

References XLI

[126] E. Bayro-Corrochano, *Geometric computing: for wavelet transforms, robot vision, learning, control and action*.

Springer, 2010.

[127] E. Bayro-Corrochano and G. Scheuermann (editors), *Geometric algebra computing: In engineering and computer science*.

Springer Science & Business Media, 2010.

[128] S. Franchini, A. Gentile, F. Sorbello, G. Vassallo, and S. Vitabile, "Clifford algebra based edge detector for color images," in *2012 Sixth International Conference on Complex, Intelligent, and Software Intensive Systems*, pp. 84–91, IEEE, 2012.

[129] M. T. Pham and K. Tachibana, "A conformal geometric algebra based clustering method and its applications," *Advances in Applied Clifford Algebras*, vol. 26, no. 3, pp. 1013–1032, 2016.

References XLII

[130] E. Bayro-Corrochano, *Geometric Algebra Applications Vol. I: Computer Vision, Graphics and Neurocomputing*.

Springer, 2018.

[131] E. Bayro-Corrochano, *Geometric Algebra Applications Vol. II: Robot Modelling and Control*.

Springer, 2020.

[132] E. B. Corrochano, *Geometric Computing for Perception Action Systems: Concepts, Algorithms, and Scientific Applications*.

Springer, 2001.

[133] E. Bayro-Corrochano, *Handbook of Geometric Computing*.

Springer, 2005.

Paperback edition 2010.

References XLIII

[134] S. Lazendic, H. De Bie, and A. Pizurica, "Octonion sparse representation for color and multispectral image processing," in *2018 26th European Signal Processing Conference (EUSIPCO)*, pp. 608–612, IEEE, 2018.

[135] S. Lazendic, A. Pizurica, and H. De Bie, "Hypercomplex algebras for dictionary learning," in *Early Proceedings of the AGACSE 2018 Conference*, pp. 57–64, Unicamp/IMECC, 2018.

`https://biblio.ugent.be/publication/8570237/file/8570249.pdf`.

[136] C. Segre, "The real representations of complex elements and extension to bicomplex systems," *Mathematishcen Annalen*, vol. 40, pp. 413–467, 1892.

References XLIV

[137] T. Nitta and H. H. Gan, "Fundamental structure of orthogonal variable commutative quaternion neurons," in *Proceedings of the Joint 11th International Conference on Soft Computing and Intelligent Systems and 21st International Symposium on Advanced Intelligent Systems, SCIS & ISIS2020 (Online conference, December 5-7, 2020)*, pp. 434–436, 2020.

[138] P. Lounesto, *Clifford Algebras and Spinors (2nd edition)*, vol. 286 of *LMS Lecture Notes Series*.

Cambridge University Press, 2001.

[139] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar, "Fourier neural operator for parametric partial differential equations," 2020.

`https://arxiv.org/pdf/2010.08895.pdf`.

References XLV

[140] M. Henaff, J. Bruna, and Y. LeCun, "Deep convolutional networks on graph-structured data," 2015.

arXiv:1506.05163.

[141] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric deep learning: going beyond Euclidean data," *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 18–42, 2017.

[142] J. M. Selig, *Geometric fundamentals of robotics*.

Springer, 2004.

[143] M. Otte, "A survey of machine learning approaches to robotic path-planning," *International Journal of Robotics Research*, vol. 5, no. 1, pp. 90–98, 2008.

References XLVI

[144] J. Kober, J. A. Bagnell, and J. Peters, "Reinforcement learning in robotics: A survey," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1238–1274, 2013.

[145] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 26–38, 2017.

`https://discovery.ucl.ac.uk/id/eprint/10083557/1/1708.05866v2.pdf`.

[146] I. Lenz, *Deep learning for robotics*.

PhD thesis, Cornell University, 2016.

[147] J. Shabbir and T. Anwer, "A survey of deep learning techniques for mobile robot applications," *arXiv: 1803. 07608*, 2018.

References XLVII

[148] J. Ruiz-del Solar, P. Loncomilla, and N. Soto, "A survey on deep learning methods for robot vision," *arXiv: 1803. 10862*, 2018.

[149] L. Tai, J. Zhang, M. Liu, J. Boedecker, and W. Burgard, "A survey of deep network solutions for learning control in robotics: From reinforcement to imitation," *arXiv: 1612. 07139 , v4*, 2018.

[150] A. Colomé and C. Torras, *Reinforcement Learning of Bimanual Robot Skills*, vol. 134 of *Springer Tracts in Advanced Robotics*.

Springer, 2020.

xx + 182 p.

[151] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.

References XLVIII

[152] M. Reed, "Differential geometric algebra with Leibniz and Grassmann," *Proceedings JuliaCon*, vol. 1, no. 1, 2019.

See `https://github.com/chakravala/Grassmann.jl`.

[153] U. Von Luxburg and B. Schölkopf, "Statistical learning theory: Models, concepts, and results," in *Handbook of the History of Logic*, vol. 10, Inductive logic, pp. 651–706, Elsevier, 2011.

[154] T. Wu and M. Tegmark, "Toward an AI Physicist for Unsupervised Learning," 2019.

`https://arxiv.org/pdf/1810.10525.pdf`, v4.

[155] C. Pehlevan and D. B. Chklovskii, "Neuroscience-inspired online unsupervised learning algorithms: Artificial neural networks," *IEEE Signal Processing Magazine*, vol. 36, no. 6, pp. 88–96, 2019.

`https://arxiv.org/pdf/1908.01867.pdf`.

References XLIX

[156] J. Sohl-Dickstein, C. M. Wang, and B. A. Olshausen, "An unsupervised algorithm for learning Lie group transformations," 2010.

`https://arxiv.org/pdf/1001.1027.pdf`.

[157] Y. Zhou, H. Zheng, and X. Huang, "Graph neural networks: Taxonomy, advances and trends," *https://arxiv.org/pdf/2012.08752.pdf*, 2020.

[158] R. DeVore, B. Hanin, and G. Petrova, "Neural network approximation," 2020.

[159] C. Lavor and R. Alves, "Recent advances on oriented conformal geometric algebra applied to molecular distance geometry," in *Systems, Patterns and Data Engineering with Geometric Calculi* (A. D. et al., ed.), ICIAM2019 SEMA SIMAI Springer Series, pp. 19–30, Springer, 2021.

References L

[160] S. Buchholz, K. Tachibana, and E. Hitzer, "Optimal learning rates for Clifford neurons," 2007.

International conference on artificial neural networks (Springer), 864-873.

[161] S. Buchholz, E. Hitzer, and K. Tachibana, "Coordinate independent update formulas for versor Clifford neurons," 2008.

SCIS & ISIS 2008, Japan Society for Fuzzy Theory and Intelligent Informatics, 814-819.

[162] E. Hitzer, "Geometric operations implemented by conformal geometric algebra neural nodes," 2013.

`https://arxiv.org/pdf/1306.1358.pdf`.

[163] D. Hildenbrand and E. Hitzer, "Analysis of point clouds using conformal geometric algebra," 2008.

References I

[164] H. Wang, Q. Liu, X. Yue, J. Lasenby, and M. J. Kusner, "Pre-training by completing point clouds," 2020.

`https://arxiv.org/pdf/2010.01089.pdf`.

[165] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Advances in neural information processing systems*, pp. 3856–3866, Springer, 2017.

[166] C. Xiang, L. Zhang, Y. Tang, W. Zou, and C. Xu, "MS-CapsNet: A Novel Multi-Scale Capsule Network," *IEEE Signal Processing Letters*, 2018.

[167] G. E. Hinton, S. Sabour, and N. Frosst, "Matrix capsules with EM routing," in *ICLR 2018*, Springer, 2018.

15 pages.

References LII

[168] E. Xi, S. Bing, and Y. Jin, "Capsule Network Performance on Complex Data," 2017.

http://arxiv.org/abs/1712.03480.

[169] D. Mumford and A. Desolneux, *Pattern theory: the stochastic analysis of real-world signals*.

A. K. Peters, 2010.

[170] A. Shahroudnejad, P. Afshar, K. N. Plataniotis, and A. Mohammadi, "Improved explainability of capsule networks: Relevance path by agreement," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 549–553, 2018.

https://arxiv.org/pdf/1802.10204.pdf.

References LIII

[171] T. Cohen and M. Welling, "Group equivariant convolutional networks," in *International conference on machine learning*, pp. 2990–2999, 2016.

http://proceedings.mlr.press/v48/cohenc16.pdf.

[172] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," 2017.

https://arxiv.org/pdf/1704.01212.pdf.

[173] N. Thomas, T. Smidt, S. Kearnes, L. Yang, L. Li, K. Kohlhoff, and P. Riley, "Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds," 2018.

https://arxiv.org/pdf/1802.08219.pdf.

References LIV

[174] J. Klicpera, J. Gross, and S. Günnemann, "Directional message passing for molecular graphs," 2020.

https://arxiv.org/pdf/2003.03123.pdf.

[175] C. W. Park, M. Kornbluth, J. Vandermause, C. Wolverton, B. Kozinsky, and J. P. Mailoa, "Accurate and scalable multi-element graph neural network force field and molecular dynamics with direct force architecture," 2020.

https://arxiv.org/ftp/arxiv/papers/2007/2007.14444.pdf.

[176] T. E. Smidt, M. Geiger, and B. K. Miller, "Finding symmetry breaking order parameters with Euclidean neural networks," *Physical Review Research*, vol. 3, no. 1, p. L012002, 2021.

References LV

[177] E. O. Pyzer-Knapp and T. Laino, *Machine Learning in Chemistry: Data-Driven Algorithms, Learning Systems, and Predictions*.

ACS Publications, 2019.

[178] S. Batzner, T. E. Smidt, L. Sun, J. P. Mailoa, M. Kornbluth, N. Molinari, and B. Kozinsky, "SE(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials," 2021.

`https://arxiv.org/pdf/2101.03164.pdf`.

[179] J. Bruna and S. Mallat, "Invariant scattering convolution networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1872–1886, 2013.

[180] F. Gama, A. Ribeiro, and J. Bruna, "Diffusion scattering transforms on graphs," 2018.

`https://arxiv.org/pdf/1806.08829.pdf`.

References LVI

[181] J. Bruna, "The scattering representation," 2019.

[182] C. Pan, S. Chen, and A. Ortega, "Spatio-temporal graph scattering transform," 2020.

[183] M. Andreux, T. Angles, G. Exarchakis, R. Leonarduzzi, G. Rochette, L. Thiry, J. Zarka, S. Mallat, J. Andén, E. Belilovsky, *et al.*, "Kymatio: Scattering Transforms in Python," *Journal of Machine Learning Research*, vol. 21, no. 60, pp. 1–6, 2020.

[184] M. Mitrea, *Clifford wavelets, singular integrals, and Hardy spaces*. Springer, 1994.

References LVII

[185] T. Qian, M. I. Vai, and Y. Xu (editors), *Wavelet analysis and applications*.

Applied and numerical harmonic analysis, Birkhäuser, 2007.

Includes: *Clifford analysis and the continuous spherical wavelet transform* (P. Cerejeiras, M. Ferreira, U. Kähler).

[186] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*.

Cambridge university press, 2014.

440 pp.

[187] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*.

MIT press, 2016.

`http://www.deeplearningbook.org`, 781 p.

References LVIII

[188] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of machine learning*.

MIT press, 2018.

xvi + 486 pp.

[189] G. Rebala, A. Ravi, and S. Churiwala, *An Introduction to Machine Learning*.

Springer, 2019.

xxii + 263 p.

[190] E. Alpaydin, *Introduction to machine learning (fourth edition)*.

Adaptive computation and machine learning, MIT press, 2020.

xxiv + 682 pp. 1st edition: 2004; 2nd, 2010; 3rd, 2014.

References LIX

[191] D. A. Roberts, S. Yaida, and B. Hanin, *The Principles of Deep Learning Theory: An Effective Theory Approach to Understanding Neural Networks*.

Cambridge University Press, 2021.

Forthcoming. Draft: `https://arxiv.org/pdf/2106.10165.pdf`. viii+441 p.

[192] S. Theodoridis, *Machine learning: a Bayesian and optimization perspective*.

Academic Press, 2015.

xxii+1050 p.

[193] A. B. Patel, T. Nguyen, and R. G. Baraniuk, "A probabilistic theory of deep learning," *arXiv preprint arXiv:1504.00641*, 2015.

References LX

[194] S. Russell and P. Norvig, *Artificial intelligence: A modern approach (Fourth edition)*.

Pearson, 2020.

1166 p.

[195] C. Bielza and P. Larrañaga, *Data-Driven Computational Neuroscience: Machine Learning and Statistical Models*.

Cambridge University Press, 2020.

xiv+747 p.

[196] J. Guttag, *Introduction to computation and programming using Python: With application to understanding data*.

MIT Press, 2016.

xvii + 447 p.

References LXI

[197] A. Nandy and M. Biswas, *Reinforcement Learning: With Open AI, TensorFlow and Keras Using Python*.

Apress, 2018.

xiii + 167 p.

[198] A. Said and V. Torra, *Data Science in Practice*.

Springer, 2019.

viii + 195 p.

[199] V. E. Balas, S. S. Roy, D. Sharma, and P. Samui, *Handbook of deep learning applications*, vol. 136.

Springer, 2019.

vi + 383 p.

References LXII

[200] F. Chollet, *Deep learning with Python (2nd edition, V4)*.
Manning Early Access Program, 2020.
210 p.

[201] H. Li, *Invariant algebras and geometric reasoning*.
World Scientific, 2008.

For me, the most surprising aspect of these winning algorithms is that its authors are not experts in the games, nor on the chemistry of proteins for that matter.

In some sense they are not unlike the infinite variety of tools invented by humanity that extend the human capacities beyond the biological nature. Any tool can serve to illustrate this: the fire and the wheel; hammers and anvils; screws and screwdrivers; bicycles, cars, planes and submarines; photography, movies and digital cameras; and so on. Among the tools, there is the outstanding class of scientific instruments, like lenses, telescopes, microscopes, or the detectors of gravitational waves.

P

General

[109] (haykin-2009)
[186] (shalevshwartz-bendavid-2014)
[187] (goodfellow-bengio-courville-2016)
[188] (mohri-rostamizadeh-talwalkar-2018)
[189] (rebala-ravi-churiwala-2019)
[190] (alpaydin-2020)
[71] (bach-2021)
[191] (roberts-yaida-hanin-2021)

Bayesian approaches

[192] (theodoridis-2015)
[193] (patel-nguyen-baraniuk-2015)
[194] (russell-norvig-2020)
[195] (bielza-larranaga-2020)

P

Applications

[196] (guttag-2016)
[197] (nandy-biswas-2018)
[198] (said-torra-2019)
[199] (balas-roy-sharma-samui-2019)
[200] (chollet-2020)
[150] (colome-torras-2020)

P

[94] (arena-fortuna-muscato-xibilia-1998)

[119] (sommer-2001)

[123] (buchholz-2005)

[99] (aizenberg-2011)

[97] (hirose-2012)

[98] (hirose-2013)

[106] (hitzer-sangwine-2013)

[100] (guberman-2016)

P

Sketch of GA

The reason for using geometric algebras is that their formalism is optimally adapted to express the geometric facts of any *linear geometric space*, that is, of a real vector space $E = E_{r,s}$ endowed with a metric (a bilinear symmetric real-valued product $x \cdot x'$, $x, x' \in E$) of signature (r, s). The most direct way to introduce the geometric algebra $\mathcal{G}_{r,s}$ of this space, one that is arguably the closest to the ideas on which W. K. Clifford (1845-1879) based his creation, is that Grassmann's exterior algebra of E, ΛE, has *a unique bilinear* **associative** *product* with unit 1 (called *geometric product* by Clifford himself) such that

$$xa = x \cdot a + x \wedge a \quad (x \in E, a \in \Lambda E), \tag{3}$$

where $x \cdot a = i_x(a)$ (the contraction of x with a).

Since i_x is the unique skew derivation of ΛE such that $i_x(x') = x \cdot x'$ for any $x' \in E$, the formula (3) shows how to multiply any multivector a by any vector x on the left.

In fact, *the formula suffices for the calculation of any product of multivectors* because of the following reasoning.

By bilinearity, it is enough to know how to multiply a non-zero exterior product $b = x_1 \wedge \cdots \wedge x_r$ ($r \geqslant 2$) of vectors $x_1, \ldots, x_r$ (such products are called r-*blades*) by an arbitrary multivector a.

We can further assume that $x_1, \ldots, x_r$ are *pair-wise orthogonal*, for the space $\langle x_1, \ldots, x_r \rangle$ has orthogonal bases and the exterior product of any such basis is equal, up to a multiplicative constant, to b. Finally we have that $x_1 \cdots x_r = b$ (by induction on r we may assume that $x_2 \cdots x_r = x_2 \wedge \cdots \wedge x_r$, and then $x_1 x_2 \cdots x_r = x_1(x_2 \wedge \cdots \wedge x_r)$ $= x_1 \wedge x_2 \wedge \cdots \wedge x_r$ because $x_1 \cdot (x_2 \wedge \cdots \wedge x_r) = 0$). So $ba = x_1 \cdots x_r a$, which can be determined by r applications of (3).

Now the *geometric algebra* $\mathcal{G}_{r,s}$ is the exterior algebra $\Lambda E_{r,s}$ enriched with the geometric product (this structure is also known as *Clifford's algebra*). It is clear then that it has dimension 2^n, where $n = r + s = \dim E$.

Note that the equation (3) shows that the linear grading of $\mathcal{G}_{r,s}$, which is in fact a grading with respect to the exterior product, *is not a grading with respect to the geometric product*.

But the decomposition $\mathcal{G} = \mathcal{G}^+ \oplus \mathcal{G}^-$ into *even* ($\mathcal{G}^+$) and *odd* ($\mathcal{G}^-$) grade components is a grading mod 2 *also with respect to the geometric product* (ultimately this is derived from the equation (3), by which the product of two vectors is resolved as the sum of a scalar, which has grade 0, and a bivector, which has grade 2).

In particular, $\mathcal{G}^+$ is a subalgebra.

↘

The isomorphisms $\mathcal{G}_{1,0} \simeq \mathbb{R} \oplus \mathbb{R}$, $\mathcal{G}_{0,1} \simeq \mathcal{G}^+_{2,0} \simeq \mathbb{C}$, $\mathcal{G}_{2,0} \simeq \mathbb{R}(2)$, or $\mathcal{G}_{0,2} \simeq \mathcal{G}^+_{3,0} \simeq \mathbb{H}$, easy to derive directly, are in fact examples of a general trend (cf. [19]):

$\mathcal{G}_{r,s}$ is isomorphic to a matrix algebra $F_\nu(m)$, where $\nu = s - r \mod 8$, $F_\nu = \mathbb{R}, \mathbb{C}, \mathbb{H}, 2\mathbb{H}, \mathbb{H}, \mathbb{C}, \mathbb{R}, 2\mathbb{R}$ for $\nu = 0, 1, 2, 3, 4, 5, 6, 7$, and $\dim(F_\nu)m^2 = 2^n$. For example, $\mathcal{G}_{1,3} = F_2(m) = \mathbb{H}(2)$.

Of these isomorphisms, those that most closely connect algebra with geometry are $\mathbf{C} = \mathcal{G}^+_{2,0} \simeq \mathbb{C}$ and $\mathbf{H} = \mathcal{G}^+_{3,0} \simeq \mathbb{H}$ in the case of the Euclidean plane and space, respectively (of **C** and **H** we say that they are the *geometric* complex numbers and quaternions, respectively, since they emerge directly from the geometry and not from ad hoc definitions as the usual ones for $\mathbb{C}$ and $\mathbb{H}$).

For samples of various applications of geometric algebra, see [19, 20] and their bibliographies. For a discussion of a broader perspective of geometric algebra and its applications, see [201] (especially Ch. 1).

P