END OF MATHEMATICS - liberation of mathematics from physical science to data science

August Lau

Abstract. Data is abundant but has few answers. Massive data challenges mathematics when no equation exists. Data processing has no universal operator but uses a combination of operators. Physical science has done well with formal rules and equations, typically low order equation and low dimensional space. Data science lives in high/infinite dimensional space where an object has many attributes. Finally, "point set topology" liberated us from the rigidity of differential calculus and extended our mind to perceive infinite/infinitesimal objects. Our response should be to embrace more data and more operators. We should be inventors of new mathematical tools to process and interpret data.

Introduction: Mathematics has been making the transition from "physical science" to "data science." We make physical measurements and we depend on mathematics based on "physical mathematics" like wave equation. Mathematics subsequently extends its application to "data science" where the mathematical form is similar but the application is non-physical like artificial neural network. Data science is not experimentally verifiable like physical science. At first glance, physical science appears to have a firmer foundation than data science.

NON-UNIQUENESS AS "THE LAW" IN PHYSICAL SCIENCE AND DATA SCIENCE:

Classical mathematics: "The inclusion of thermodynamic irreversibility through a non-unitary transformation theory leads to a deep alteration of the structure of dynamics. We are led from groups to semigroups, from trajectories to processes." ILYA PRIGOGINE (1977 Nobel Lecture)

Modern mathematics: The initial state has different topology compared to final state, hence likely to be irreversible. Qualitative approach like topology liberated our thinking from quantitative approach.

We will tease out some of the meaning of "**from groups to semigroups, from trajectories to processes**."

Group: A simple example of a group is the set of rotation matrices. The opposite of the rotation will undo the original rotation. So it will go back to the original position as though nothing has been changed. Group is a generalization of such reversible system similar to a group of rotations that every operator A has an inverse B where AB=BA=I and I is the identity operator. Unitary transformation is a special kind of invertible operator.

Semigroup: A semigroup is composed of two things: a set and a binary operation obeying associative law of a(bc)=(ab)c. A good example of a semigroup is the set of ALL nxn matrices with matrix multiplication. Not all matrices have inverse. All invertible matrices form a group and all nxn matrices form a semigroup. Semigroup is a generalization of a group. Semigroup might not have inverses but preserves associative law A(BC)=(AB)C.

Trajectories: A trajectory can be viewed as a metaphor for a distinct path that can be calculated in a deterministic way. A trajectory is symbolically written as A(x) = y where x is a vector and A is the operator and y is the output.

Processes: Classically, a process is an extension from deterministic equations to probability theory. Even though probability could be useful, it does not solve the basic problem of non-uniqueness.

Semigroups and processes were good attempts to liberate the mathematics for physical science. But we need other kinds of mathematics if we want to deal with both physical science and data science.

MATHEMATICAL FORMALISM:

"Static and dynamic" are prominent in physical science:
Classical view of physical science could be divided into static view and dynamic view. Static view is the description of an object at a fixed time. Objects could be planets, molecules, DNA, etc. Dynamic is the study of the object changing in time. In planetary motion, we could write a differential equation to describe the change in time. Group theory is generalized from differential equation as a group of transformations. With irreversibility and non-uniqueness, semigroup is needed as a generalization of group.

"Data and operator" are unifying concepts for both physical science and data science:
Static description is now replaced by data. Data could be analog or digital. Instead of a physical object, all objects are measured and recorded as data. A movie is a collection of data with each frame as analog/digital data. The dynamic is a change from frame to frame. A movie has a plot which is the dynamic changing from one frame to another. But a movie is written by a script and not a semigroup/differential equation. Dynamic description is replaced by an operator. Instead of differential equation or transformation semigroup, we use operator as a general term of transforming dataset to dataset.

Numerical blurring of physical science and data science: Here are some examples of transition from "physical science" to "data science" in numerical method and we will just mention them in passing.
<u>Wave equation</u>: Wave equation mathematically is second order in time and second order in space. It originated with wave propagation but the form of the wave equation can be used in graph theory in data science and is detached from its physical origin.
<u>Heat equation</u>: Heat equation is first order in time and second order in space. It can be used in diffusion of data without any physical basis. The form of the heat equation is used in clustering in data science and is detached from its physical origin.

DATA AND OPERATOR:

What is data?

One way to do measurement of a physical object is by a ruler. A ruler is just a metaphor for a known object. A measurement can be viewed mathematically as a function from a known object to an unknown object. The unknown object is our object of interest which is under our investigation. A measurement is a function (sometimes called mapping in mathematics) from known to unknown. What about a function from unknown to known? It can be thought of as a measurement also. If we use X-ray on a body projected to a plate, it is a measurement from the unknown object to a known object of a plate. A function from known to unknown (or unknown to known) can be used in physical science or data science.

A direct function is from a known object to an unknown object. An indirect function is from an unknown object to a known object. A spreadsheet is an example of recording direct data of measuring an object by length, width, etc. Length and width are using known measuring sticks. If we use X-ray as indirect data, the data on the plate is a collection of images and not just numbers. Images are data also.

Data require **processing** and **interpretation** to find the description of the unknown object. In fact, most measurements are indirect, e.g., cameras with videos and sound. Indirect measurements are now overwhelming our digital storage.

What is operator?

Mathematically, an operator defines the relationship from one dataset to another dataset. An operator could be explicit or implicit. An explicit operator could be $f(x) = 2x$ where we take a measurement x and multiply it by 2. An implicit operator could be a differential equation like $dx/dt = x$ where x and t are related with derivative on the left hand side and non-derivative on the right hand side.

In physical science, we collect data and the trick is to find the operator connecting different data. After we discover the operator, we can verify the operator by experiments or more data observations. We will fix the operator and assume that the (dynamical) system behaves as the operator prescribes, e.g., planetary motion, chemical equations, combination of DNA. In data science, it does not have the support of physical experiments and depends on mathematical prediction/statistics. One such example is neural network which is a composition of many simple operators. There is no explicit equation connecting inputs to outputs.

Fundamental questions:
How to find operators if data is given?
If operator is only approximate, how to interpret data with imperfect operator?
Is there a difference between data and operator?

In physical science, we have folklore of how great the physical insight is… and how inspirational it is to come up with the right equation. We do not want to dispute the romance but it was inspiration and perspiration … and a lot of trial and error. Let us ramble on “artificial intelligence” of data science. We use tools for automatic theorem proving, syntactic tools for pattern recognition, graph tools, etc. We appreciate the honesty of data science that it is ad hoc and not systematic. It might be more important to improve the **power of the tool** to handle big data than the **intelligence of the tool**. An empirical tool like neural network could be more **powerful** than differential equations or semigroup theory. The tools might not be elegant but they are **powerful**. They might be brute force but **powerful** to control massive amount of data.

NON-UNIQUENESS LEADS TO INTERPRETATION:

Data science can be roughly divided into processing and interpretation. Processing uses a “processing flow” which is like a flow chart with operators mapping one dataset to another. Processing tends to use operators that are “physical mathematics” like wave equation. But non-uniqueness forces us to use interpretation (human intelligence and machine learning). Human intelligence is needed to make decisions when we could not invert the operator to a unique solution. Machine learning could be viewed as an interpretation tool since the machine learning solution by itself is non-unique. Machine learning requires human interpretation (“tweaking”) when we change from one application to another.

CONCLUSION:

Mathematics is transitioning from physical science (experimental) to data science (experiential). In data science, we want to experience the data and “manipulate“ the data like physical objects. One approach is to generate tools for **PROCESSING** and **INTERPRETATION** of data. It is a collection of tools rather than a well-defined procedure. It is **empirical** and might not seem systematic to some. Physical science uses “hard” mathematics and data science uses “soft” mathematics. Hardness means that the equation does not change regardless of data, e.g., wave equation. Softness means that there is no well-defined equation. Sometimes we do not even know the equation until we see the data, e.g., neural network. Non-uniqueness leads us to interpretation to make decisions. Finally, we need to be aware of the **power of AI** rather than the intelligence.

End of mathematics could mean mathematics at its historical end. It is difficult to see that mathematics would end because of its infinite/infinitesimal process.

End of mathematics could also mean the goal or purpose of mathematics. If the end of mathematics is to understand data, with or without physical constraint, we will be more liberated and willing to invent mathematics based on our interpretation of data. This could be a rebirth of mathematics if we allow data to speak to us.

“The essence of mathematics lies in its freedom.”   Georg Cantor

Appendix I (General mathematics)

Mathematics is a collection of formal rules (quantitative and qualitative).
MATHEMATICS => REDUCTIONISM + EXPANSIONISM
Reductionism decomposes operator/data into simpler forms. Applicable to physical science.
Expansionism combines operator/data into complex forms. Applicable to data science.

We will illustrate the transitions in mathematics for ALGEBRA + GEOMETRY:

ALGEBRA (quantitative) => POLYNOMIAL, SEMIGROUP, FUNCTIONS, NEURAL NETWORK **(NN)**

Fundamental theorem of arithmetic
Any positive integer can be decomposed as product of prime numbers. It illustrates the power of reductionism. It reduces an integer problem to prime integers.

Fundamental theorem of algebra
Any polynomial with complex coefficients can be factored into products of monomial (power=1). It illustrates the power of reductionism. It reduces a polynomials to monomials.

Fundamental theorem of linear algebra
The goal is to reduce a matrix into simpler matrices like unitary matrices and diagonal matrices. Another triumph for reductionism.

Semigroup of functions
A matrix is a special kind of function from vector space to vector space. A set of functions from a set to another set is a more general semigroup with binary associative algebra. Semigroup has the algebraic flexibility of being non-invertible.

Cascading semigroups (Composition of functions)
We can compose two different semigroups of functions and expand it to a bigger and more complex semigroups. The obvious example of composition of semigroups is composition of functions, e.g., composition of F and G is defined by F(G(x)). In linear algebra, composition is similar to matrices and multiplication of matrices.

Neural network (NN) = composition of semigroups/algebras
This brings us to modern application of mathematics. NN historically has one hidden layer but it is easily generalized to multiple layers. A set of functions connecting one layer to the next layer is a semigroup. One example of NN is a composition of semigroups connecting input data to output data.

GEOMETRY (qualitative) => SHAPE, TOPOLOGY, TOPOLOGICAL DATA ANALYSIS **(TDA)**

Fundamental theorem of calculus
The fundamental theorem of calculus is a theorem that links the concept of differentiating a function with the concept of integrating a function.

Stokes' theorem
It connects a body (e.g. volume V) to the boundary of a body (e.g. surface dV), e.g., Integral (f over dV) = Integral (df over V). It connects the boundary integral with interior integral.

De Rham cohomology
This is differential-geometric thinking which subsequently leads to topology. With differentials, we could describe geometric complexity of an object like number of holes in an object. Cohomology maps an object to an algebraic group and it motivates further research in topology.

Topology
We can finally liberate mathematics from derivatives. This is "point set topology" which can be defined as a set of points with a topological structure. Topology yields cohomology without having derivatives, e.g., Alexander cohomology. This leads to even more abstraction like "category theory" which gives algebraic topology a "functorial" interpretation.

Topological Data Analysis (TDA) = mapping topology to algebra
This brings us to modern application of mathematics. One popular application of TDA is mapping data to topological space. One common application uses algebraic topology to map topological space to homology groups. We can interpret the topological space with or without algebraic topology, i.e., we can directly perceive the topological space.

NN and TDA are modern applications after a long history of algebra and geometry.
Bottomline: Physical reality is normally seen as THE reality. Is physical science too RESTRICTIVE? Does data science have more freedom? Data science might inspire us to come back to the classical view of pure mathematics. Pure mathematics was investigated for its own sake and might be separated from physical reality. Mathematics could liberate us from physicalism to dataism.

Study of pure mathematics should be encouraged. It might surprise us to find that pure mathematics like "Chinese Remainder Theorem" could be the mathematical motivation for crypto-currency and crypto-economy. Or the study of point set topology and algebraic topology leads to topological data analysis of time series of seismic data or financial data.

Automata and semigroup had common interest in the past but have developed into separate disciplines. Maybe it is time to unite automata and semigroup as abstract machine, hopefully with a topological spin.

Appendix II (Seismic Processing)

Seismic processing converts real data recorded in the field to interpretable images.

REDUCTIONISM + EXPANSIONISM:
REDUCTIONISM (structured, bounded, convergent, compact, “manageable”)
EXPANSIONISM (less structured, unbounded, divergent, self-replicating, “reproductive”)
INTERPRETATION (small data versus big data, restrictions versus freedom)

Reductionism compresses large datasets into small datasets so it is “easier” to interpret. It reduces the paths from input to output to only one path or small number of paths. It limits the number of outputs for interpretation.

Expansionism creates a lot more datasets from real data. It generates many datasets to give flexibility of interpretation, e.g., thousands of datasets created from real data. Hence, interpretation needs more tools.

REDUCTIONISM
Typical examples are Fourier transform, eigenfunction decomposition, functional analysis, singular value decomposition, matrix decomposition. They attempt to convert large dataset to small dataset, from high dimension to low dimension, e.g., seismic inversion using inverse problems to reduce dimension. It is viewed as more “manageable” dataset for interpretation.

EXPANSIONISM
The goal of expansionism is to generate more datasets to give more **freedom to interpret**.

If A is mxn matrix, we decompose A = BC where B is mxp and C is pxn.
For reductionism, we want p to be close to m or n.
For expansionism, we might have p >> m or n.
For further expansionism, we might want A = A1*A2*…*Aq where Aj could be dimensionally larger SIZE than A and q is much larger than 2 to increase the DEPTH of the decomposition. It seems “abnormal” to blow up a matrix into larger dimensional matrices, but that is what we do in generating more complex systems like neural network.

Neural network (NN) started with only one hidden layer but expanded to many layers.
We can expand a network to many depths and we can expand each layer to have long vector.
We can also expand a network to have many branches, i.e., increasing the topology of network.
Expansionism increases complexity of the network topology by length and/or depth.

Historically, expansionism has existed in seismic imaging for a long time. Let us look at the classical idea of imaging as datuming. We record seismic data at the surface of the earth. We estimate the velocity field V1 which uses wave equation to extrapolate datum from D0 to D1 where D0 is the surface and D1 is some surface below. Then V2 is estimated to datum from D1 to D2, V3 from D2 to D3, etc.

Combination of V1, V2, V3 could generate N1xN2xN3 output datasets where Nj is the number of perturbations of Vj. If Vj has 10 perturbations, then 1000 = 10x10x10 seismic images could be produced. In the past, we shy away from generate 1000 image cubes for interpretation. It was viewed as "unmanageable" to interpret 1000 seismic images.

Bottomline: We suppressed expansionism which created large number of datasets even in classical seismic imaging. However, expansionism gives greater freedom to interpret many seismic cubes, e.g., 1000 image cubes. It needs new research to obtain interpretation tools for thousand or even million cubes. Topological data analysis, neural network, and diffusion semigroup are some examples of interpretation tool.

REFERENCES:

References in Mathematics:

Small Semilattices, Semigroup Forum 1972

Concerning costability of compact semigroups, Duke Math Journal 1972

Coverable Semigroups, Proceedings of the American Math Society 1973

The boundary of a semilattice of an n-cell, Pacific Journal of Math 1975

Finite abelian semigroups represented into the power of finite groups, Czchoslovak Math Journal 1979

Images of compact 0-dimensional semigroups, Colloquium Mathematicum, 1979

Certain local homeomorphisms of continua are homeomorphisms, II, Bulletin Polish Acad. Science 1979

Plane continua and transformation groups, Proceedings of the American Math Society 1980

References in viXra and arXiv:

Transformation Semigroup and Complex Topology: a study of inversion with increasing complexity, arXiv 2010

Solvability by semigroup : Application to seismic imaging with complex decomposition of wave equations and migration operators with idempotents, arXiv 2011

Geometric theory of inversion and seismic imaging II: INVERSION + DATUMING + STATIC + ENHANCEMENT,   viXra 2017

Quantitative and Qualitative Seismic Imaging and Seismic Inversion,   arXiv 2017

The Unreasonable Expectation of Mathematics,   viXra 2018

Nullspace and Mitigation in seismic imaging and seismic inversion,   viXra 2019

References in SEG (Society of Exploration Geophysicists):

Common-Offset Processing with DMO-Driven Statics**,** SEG 1991

Prediction of resistivity from seismic : an integrated study of reprocessing, well logs, neuralnet and prospecting, SEG 2001

Imaging with complex decomposition: Numerical applications to seismic processing in difficult areas,   SEG 2008

Diffusion semigroups: A diffusion-map approach to nonlinear decomposition of seismic data without predetermined basis,   SEG 2009

Geometric simplicity as a migration criterion: an application of computational topology to seismic imaging,   SEG 2009