UCMERCED 北京交通大學 BEIJING JIAOTONG UNIVERSITY MESA LAB

Why big data and machine learning must meet fractional calculus?

YangQuan Chen, Ph.D., Director,

MESA (Mechatronics, Embedded Systems and Automation) LAB

ME/EECS/SNRI/HSRI/CITRIS, School of Engineering,

University of California, Merced

E: yqchen@ieee.org; *or*, yangquan.chen@ucmerced.edu

T: (209)228-4672; **O**: SRE-327; **Lab**: Castle #22 (**T**: 228-4398)

July 6th, 2020. Monday 2-3PM

Seminar over internet hosted by Beijing Jiaotong University

Acknowledgements

- **Professor Yongguang Yu and BJTU**
- **You all, for coming !**
- **Am I nervous? - *Fractionally***

Skip Ad in a *fractional* hour

University of California, Merced

- The Research University of the Central Valley
- Centrally Located
  - Sacramento – 2 hrs
  - San Fran. – 2 hrs
  - Yosemite – 1.5 hrs
  - LA – 4 hrs
- Surrounded by farmlands and sparsely populated areas

MESA LAB

UC Merced

http://www.ucmerced.edu/fast-facts

- Established 2005
- 1st research university in 21st century in USA.
- 6815/**7,375** Undergraduates
- 521/**592** Grads (most Ph.Ds)
- 60% (**70%**) 1st generation; 60% Pell
- Strong Undergraduate Research Presence (HSI, MSI)
- **2020**: 9K undergrads, 1K grads
- **$1.3B expansion: now ~ 2020**

2017 (2018) U.S. News and World Report Rankings

Campus	Public	National
UC Berkeley	1	20
UCLA	2 / 1	24
UC Santa Barbara	8	37
UC Irvine	9	39
UC Davis	10	44
UC San Diego	10	44
UC Santa Cruz	30	79
UC Riverside	56	118
UC Merced	**78 / 67 / 44**	**152 / 136 / 107**

https://www.universityofcalifornia.edu/news/6-uc-campuses-named-among-nation-s-top-10-public-universities
https://news.ucmerced.edu/news/2018/uc-merced-rises-nearly-30-spots-us-news-rankings

2020 U.S. News and World Report Rankings

Graduate / Engineering

Campus	National
UC Berkeley	3
UC San Diego	11
UCLA	16
UC Santa Barbara	24
UC Davis	31
UC Irvine	36
UC Riverside	75
UC Santa Cruz	87
UC Merced	**134/107**

https://www.universityofcalifornia.edu/news/thinking-about-graduate-school-uc-programs-top-us-news-2020-rankings

UCMERCED w: mechatronics.ucmerced.edu MESA LAB

Dr. YangQuan Chen, yqchen@ieee.org

Mechatronics, Embedded Systems and Automation Lab

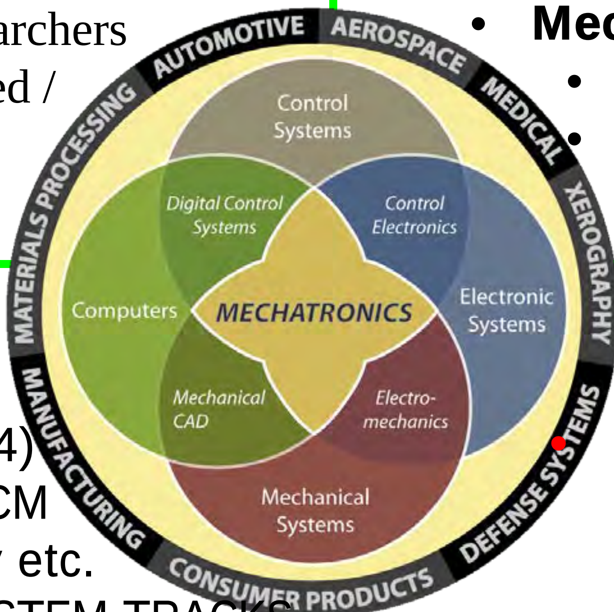
Real solutions for sustainability!

Established Aug. 2012 @ Castle, 4,500+ sq ft
6 Ph.D/10+ undergraduate researchers
10+ visiting scholars || sponsored / mentored many capstone teams

Research Areas of Excellence:

(ISI H-index=61, Google H=83; i10=463, HCR-2018,19)

- Unmanned Aerial Systems & UAV-based Personal Remote Sensing (PRS)
- **Cyber-Physical Systems (CPS)**
  - **Mechatronics**
    - **Applied Fractional Calculus**
    - **Modeling and Control of Renewable Energy Systems**



Education and Outreach Activities:

- Eng Service Learning(Sp14)
- AIAA Student Branch @UCM
- Preview Days, Bobcat Day etc.
- Robots-n-Ribs|MESABox! STEM-TRACKS TEAM-E; UAS4STEM. USDA HSI: 2016-20
- **ME142 Mechatronics (**take-home labs**)**
- **ME280 Fractional Order Mechanics**
- **ME211 Nonlinear Control**
- **ME143 Unmanned Aerial Systems**
- **ME212 Robustness and Optimality**

Projects Related to San Joaquin Valley:

- **Energy** [Solar/wind energy, Building efficiency (HVAC lighting), smart grids integration, NG pipelines]
- **Water** (Water/soil salinity management, water sampling UAVs)
- **Precision Ag/Environment** (Crop dynamics, optimal harvesting, pest, methane sniffing/mapping, DH ...)

UC Multi-campus Synergy on CIDERS

California Institute of Data-drone Engineering and Services

UCM, UCSC,UCB, UCSD, LLNL

CIDERS in Scientific data-drones: platforms, operation, and certification

UCM
UCD
UCSD

CIDERS in precision agriculture

UCM
UCD
LBL
SNL
UCSD

CIDERS in environmental monitoring: water, **fire**, soil, dust, AQ …

Outline

- **Fractional Calculus, Complexity, and Fractional Order Thinking**
- **Big Data, Variability, and Fractional Calculus**
- **Machine Learning, Optimal Randomness and Fractional Calculus**
- **Looking Into Future: Fractional Calculus is Physics Informed**

What is "Fractional Calculus"?

- **Calculus**: integration and differentiation.
- **"Fractional Calculus":** integration and differentiation of non-integer orders.
  – Orders can be real numbers (and even complex numbers!)
  – Orders are not constrained to be "integers" or even "fractionals"

How this is possible?

Why should I care?

Any (good) consequences (to me)?

... from integer to non-integer ...

$$x^n = \underbrace{x\cdot x\cdot\ldots\cdot x}_{n}$$

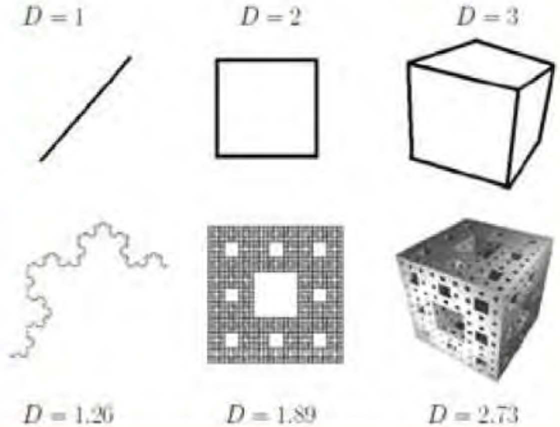
$$x^n = e^{n\ln x}$$

$$n! = 1\cdot 2\cdot 3\cdot\ldots\cdot(n-1)\cdot n,$$

$$\Gamma(x) = \int_0^\infty e^{-t}t^{x-1}dt, \qquad x>0,$$

$$\Gamma(n+1) = 1\cdot 2\cdot 3\cdot\ldots\cdot n = n!$$

... from integer to non-integer ...



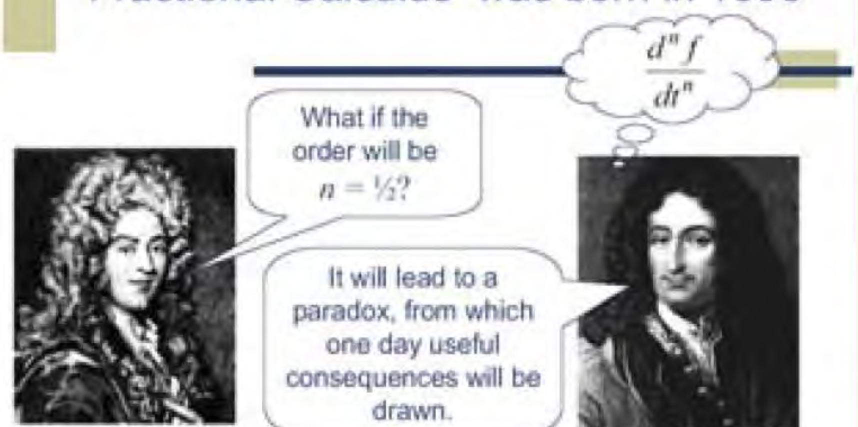
Interpolation of operations

$$f,\ \frac{df}{dt},\ \frac{d^2f}{dt^2},\ \frac{d^3f}{dt^3},\ \ldots$$

$$f,\ \int f(t)dt,\ \int dt\int f(t)dt,\ \int dt\int dt\int f(t)dt,\ \ldots$$

$$\ldots,\ \frac{d^{-2}f}{dt^{-2}},\ \frac{d^{-1}f}{dt^{-1}},\ f,\ \frac{df}{dt},\ \frac{d^2f}{dt^2},\ \ldots$$

Fractional Calculus was born in 1695



G.F.A. de L'Hôpital (1661–1704)

G.W. Leibniz (1646–1716)

The beginning of a new stage

static models	dynamical models	fractional order modeling
geometry, algebra	differential and integral calculus	***Do better than*** **fractional calculus** ***the best doable before!***

07/06/2020

Integer-Order Calculus

Fractional-Order Calculus

Fractional Order Mechanics!

Hooke's law: $F = kx$

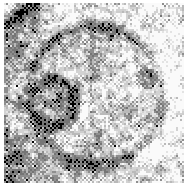
Newton's fluid: $F = kx'$

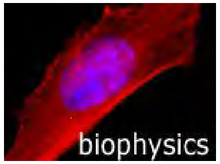
Newton's 2nd law: $F = kx''$

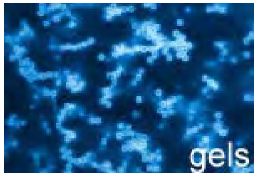
$\rightarrow F(t) = kx^{(\alpha)}(t)$

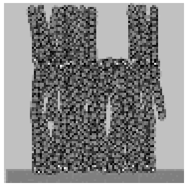
Going in-between: interpolation of operators:

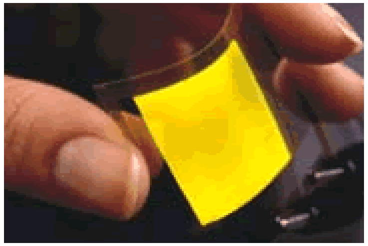
$$\ldots, \quad \frac{d^{-2}f}{dt^{-2}}, \quad \frac{d^{-1}f}{dt^{-1}}, \quad f, \quad \frac{df}{dt}, \quad \frac{d^2 f}{dt^2}, \quad \ldots$$

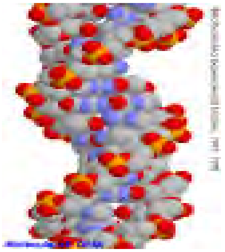


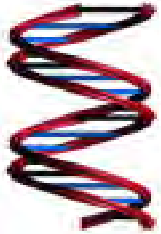


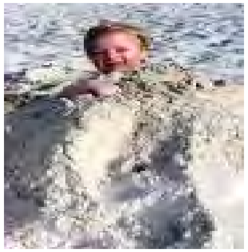


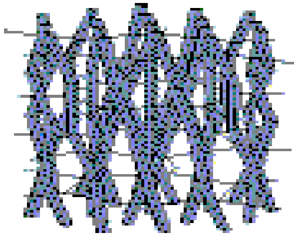


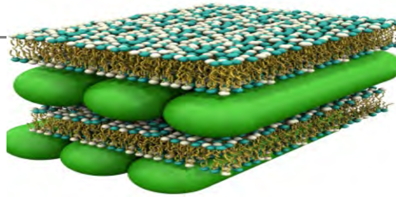












Soft matters, **complex** fluids

"Fractional Order Thinking" or, "In Between Thinking"

- For example
  - Between integers there are non-integers;
  - Between logic 0 and logic 1, there is the "**fuzzy logic**";
  - Between integer order splines, there are "**fractional order splines**"
  - Between integer high order moments, there are **noninteger order moments (e.g. FLOS)**
  - Between "integer dimensions", there are **fractal dimensions**
  - Fractional Fourier transform (FrFT) – in-between time-n-freq.
  - Non-Integer order calculus (fractional order calculus – abuse of terminology.) (FOC)

Rule of thumb for "Fractional Order Thinking"

- Self-similar
- Scale-free/Scale-invariant
- Power law
- Long range dependence (LRD)
- $1/f^{a}$ noise
- Porous media
- Particulate
- Granular
- Lossy
- Anomaly
- Disorder
- Soil, tissue, electrodes, bio, nano, network, transport, diffusion, soft matters (**bio**x) …

What is considered as complex?

（Inverse） Power Law （IPL）

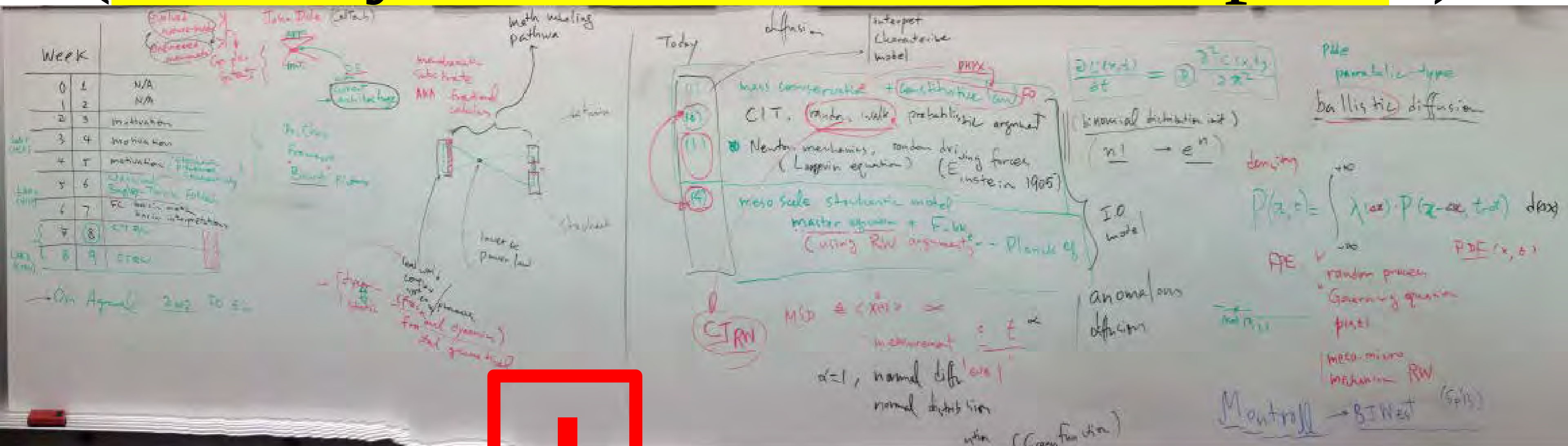
（逆）幂律现象

https://www.zhihu.com/question/20313934

为什么我国的概率与统计学教科书里不怎么讲幂律分布?

Complexity “bow tie”

(When you start to call it complex?)



Complex systems. phenomena, behaviors, …

I P L

Scale-Free, Heavy-Tailedness, Long Range Dependence, Long Memory …

Empirical Power Laws

Discipline	Law's name	Form of law
Anthropology		
1913 [4]	Auerbach	Pr(city size rank r) $\propto 1/r$
1998 [65]	War	Pr(intensity $> I$) $\propto 1/I^{\alpha}$
1978 [86]	$1/f$ Music	Spectrum(f) $\propto 1/f$
Biology		
1992 [87]	DNA sequence	Symbol spectrum(frequency f) $\propto 1/f^{\alpha}$
2000 [49]	Ecological web	Pr(k species connections) $\propto 1/k^{1.1}$
2001 [35]	Protein	Pr(k connections) $\propto 1/k^{2.4}$
2000 [34]	Metabolism	Pr(k connections) $\propto 1/k^{2.2}$
2001 [40]	Sexual relations	Pr(k relations) $\propto 1/k^{\alpha}$
Botany		
1883 [64]	da Vinci	Branching; $d_0^{\alpha} = d_1^{\alpha} + d_2^{\alpha}$
1922 [101]	Willis	No. of genera(No. of species N) $\propto 1/N^{\alpha}$
1927 [51]	Murray	$d_0^{2.5} = d_1^{2.5} + d_2^{2.5}$
Economics		
1897 [56]	Pareto	Pr(income x) $\propto 1/x^{1.5}$
1998 [24]	Price variations	Pr(stock price variations x) $\propto 1/x^3$
Geophysics		
1894 [55]	Omori	Pr(aftershocks in time t) $\propto 1/t$
1933 [67]	Rosen–Rammler	Pr(No. of ore fragments $<$ size r) $\propto r^{\alpha}$
1938 [44]	Korčak	Pr(island area $A > a$) $\propto 1/a^{\alpha}$
1945 [31]	Horton	No. of segments at n/No. of segments at $n+1$ constant
1954 [26]	Gutenberg–Richter	Pr(earthquake magnitude $< x$) $\propto 1/x^{\alpha}$
1957 [27]	Hack	River length $\propto$ (basin area)$^{\alpha}$
1977 [44]	Richardson	Length of coastline $\propto$ 1/(ruler size)$^{\alpha}$
2004 [84]	Forest fires	Frequency density(burned area A) $\propto 1/A^{1.38}$
Information theory		
1999 [32]	World Wide Web	Pr(k connections) $\propto 1/k^{1.94}$
1999 [19]	Internet	Pr(k connections) $\propto 1/k^{\alpha}$

Discipline	Law's name	Form of law
Physics		
1918 [70]	$1/f$ noise	Spectrum(f) $\propto 1/f$
2002 [25]	Solar flares	Pr(time between flares t) $\propto 1/t^{2.14}$
2003 [69]	Temperature anomalies	Pr(time between events t) $\propto 1/t^{2.14}$
Physiology		
1959 [61]	Rall	Neurons; $d_0^{1.5} = d_1^{1.5} + d_2^{1.5}$
1963 [76]	Mammalian vascular network	Veins and arteries; $d_0^{2.7} = d_1^{2.7} + d_2^{2.7}$
1963 [90]	Bronchial tree	$d_0^3 = d_1^3 + d_2^3$
1973 [48]	McMahon	Metabolic rate(body mass M) $\propto M^{0.75}$
1976 [103]	Radioactive clearance	Pr(isotope expelled in time t) $\propto 1/t^{\alpha}$
1987 [93]	West–Goldberger	Airway diameter(generation n) $\propto 1/n^{1.25}$
1991 [30]	Mammalian brain	Surface area $\propto$ volume$^{0.90}$
1992 [77]	Interbreath variability	No. of breaths(interbreath time t) $\propto 1/t^{2.16}$
1993 [58]	Heartbeat variability	Power spectrum(frequency f) $\propto f$
2007 [23]	EEG	Pr(time between EEG events) $\propto 1/t^{1.61}$
2007 [13]	Motivation and addiction	Pr(k behavior connections) $\propto 1/k^{\alpha}$
Psychology		
1957 [75]	Psychophysics	Perceived response(stimulus intensity x) $\propto x^{\alpha}$
1963 [71]	Trial and error	Reaction time(trial N) $\propto 1/N^{0.91}$
1961 [29]	Decision making	utility(delay time t) $\propto 1/t^{\alpha}$
1991 [3]	Forgetting	Percentage correct recall(time t) $\propto 1/t^{\alpha}$
2001 [20]	Cognition	Response spectrum(frequency f) $\propto 1/f^{\alpha}$
2009 [37]	Neurophysiology	Pr(phase-locked interval $< \tau$) $\propto 1/\tau^{\alpha}$
Sociology		
1926 [41]	Lotka	Pr(No. of papers published rank r) $\propto 1/r^2$
1949 [104]	Zipf	Pr(word has rank r) $\propto 1/r$
1963 [16]	Price	Pr(citation rank r) $\propto 1/r^3$
1994 [8]	Urban growth	Population density(radius R) $\propto 1/R^{\alpha}$
1998 [88]	Actors	Pr(k connections) $\propto 1/k^{2.3}$

Complex Webs: Anticipating the Improbable, B.J. West and P. Grigolini, Cambridge (2011).

IPL in Different Contexts

- Scale-free networks (degree distributions)
- Pink noise (power spectrum)
- Probability density function (PDF)
- Autocorelation function (ACF)
- Allometry ($Y=a\ X^b$)
- Anomalous relaxation (evolving over time)
- Anomalous diffusion (MSD versus time)
- Self-similar

Other connectedness to FC? (hidden)

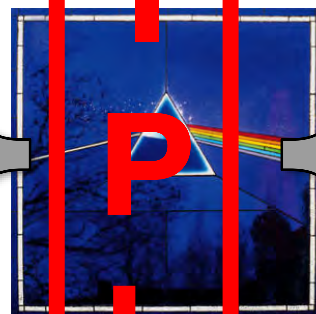
- Fractal, irregular, anomalous, rough, Hurst
  - Multifractal, multi-scale, scale-rich
- Renormalization (?), Universality
- Extreme events– spikiness, bursty, intermittence
- Fluctuation in fluctuations; Variability,
- Emergence, Surprise, **Black swan**
- Nonlocality, Long term memory
- Complex (behavior, processes, network, fluid, dynamics, systems …)
- When the forest is big, there are all types of birds ("It takes all kinds" 林子大了什么鸟都有), 20/80 rule(二八定律)

My submission:

Fractional dynamics
point of view
of
complex systems for complexity characterization and regulation

MESA LAB

I



P

L

Complex systems. phenomena, behaviors, …

Scale-Free,
Heavy-Tailedness,
Long Range Dependence,
Long Memory …

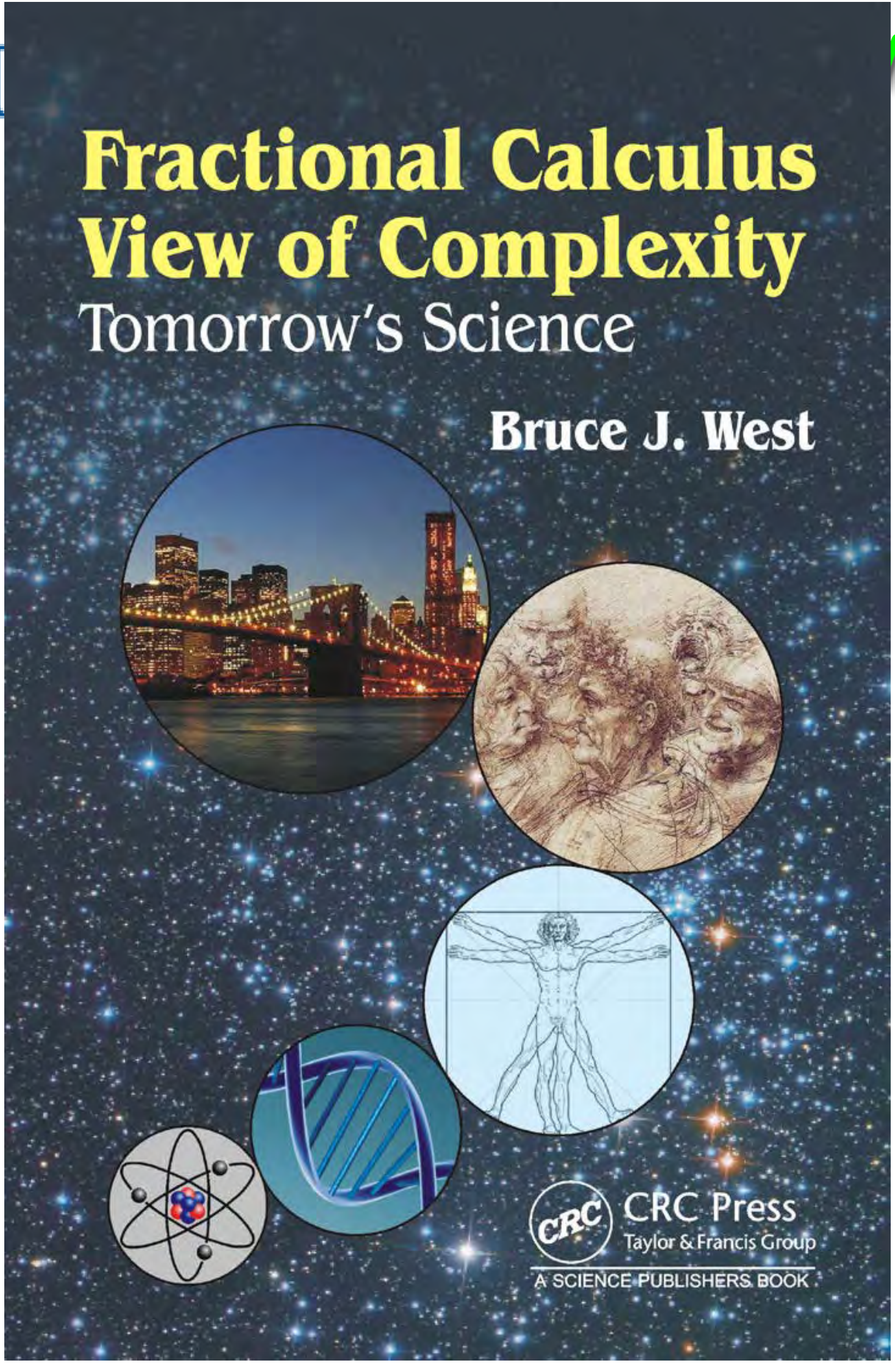
Mittag-Leffler Law

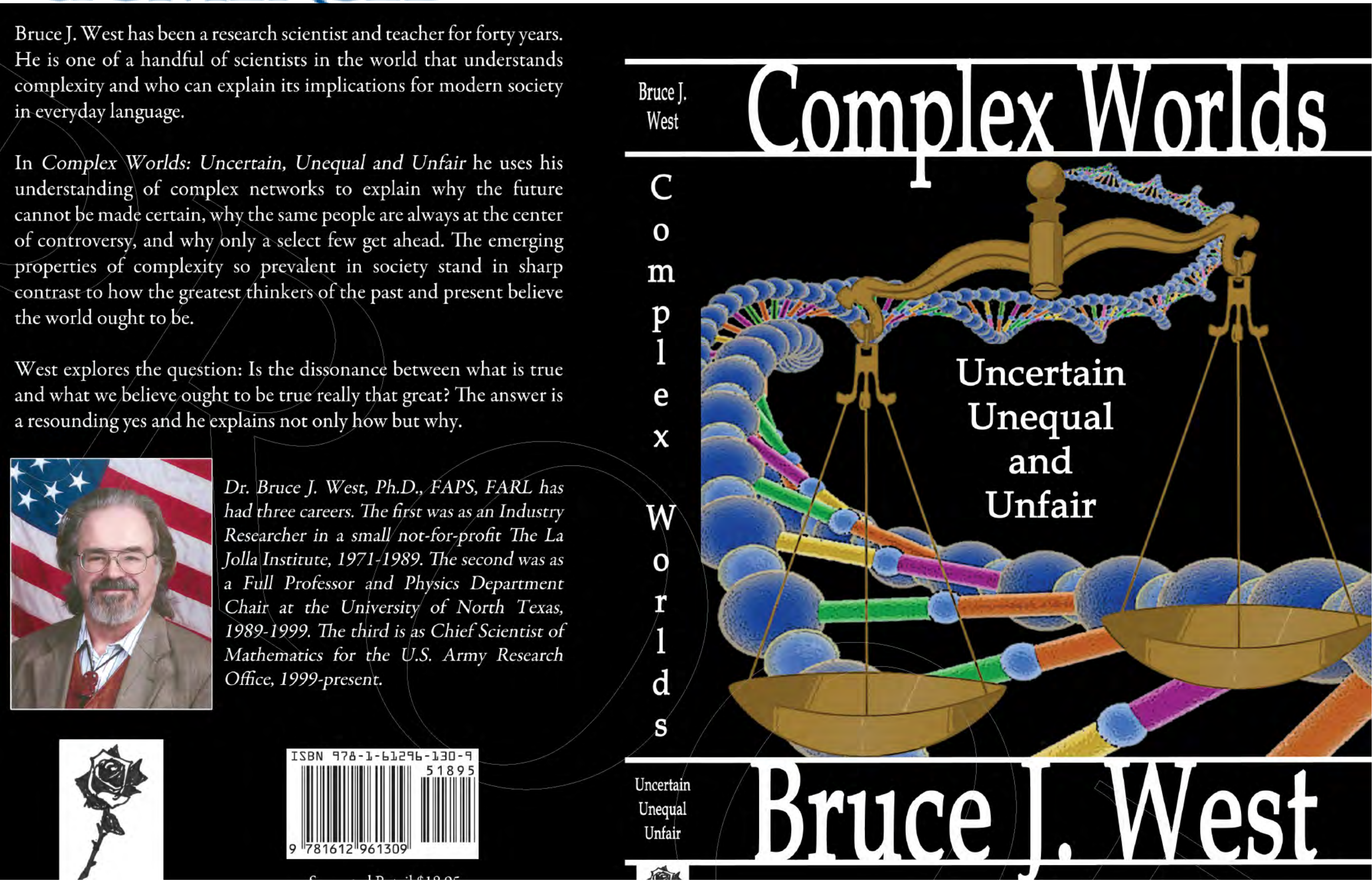
Fractional dynamic

Fractional calculus based models

UCMERCED

MESA LAB




Bruce J. West has been a research scientist and teacher for forty years. He is one of a handful of scientists in the world that understands complexity and who can explain its implications for modern society in everyday language.
In *Complex Worlds: Uncertain, Unequal and Unfair* he uses his understanding of complex networks to explain why the future cannot be made certain, why the same people are always at the center of controversy, and why only a select few get ahead. The emerging properties of complexity so prevalent in society stand in sharp contrast to how the greatest thinkers of the past and present believe the world ought to be.
West explores the question: Is the dissonance between what is true and what we believe ought to be true really that great? The answer is a resounding yes and he explains not only how but why.
Dr. Bruce J. West, Ph.D., FAPS, FARL has had three careers. The first was as an Industry Researcher in a small not-for-profit The La Jolla Institute, 1971-1989. The second was as a Full Professor and Physics Department Chair at the University of North Texas, 1989-1999. The third is as Chief Scientist of Mathematics for the U.S. Army Research Office, 1999-present.
ISBN 978-1-61296-130-9
51895
9 781612 961309
Bruce J. West
Complex Worlds
Complex Worlds
Uncertain
Unequal
and
Unfair
Uncertain
Unequal
Unfair
Bruce J. West

Power Law

$$f(x) = ax^k$$

When k is negative: Inverse power law

Scale-free

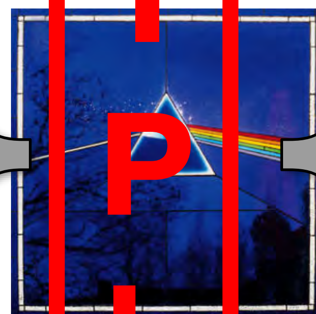
Scale invariance

$$f(cx) = a(cx)^k = c^k f(x) \propto f(x).$$

- "Scaling laws in cognitive sciences" by CT Kello, GDA Brown, R Ferrer-i-Cancho, JG Holden, K Linkenkaer-Hansen, T. Trends in Cognitive Sciences 14 (5), 223-232, 2010

Complex systems. phenomena, behaviors, …

I



P

L

Scale-Free,
Heavy-Tailedness,
Long Range Dependence,
Long Memory …

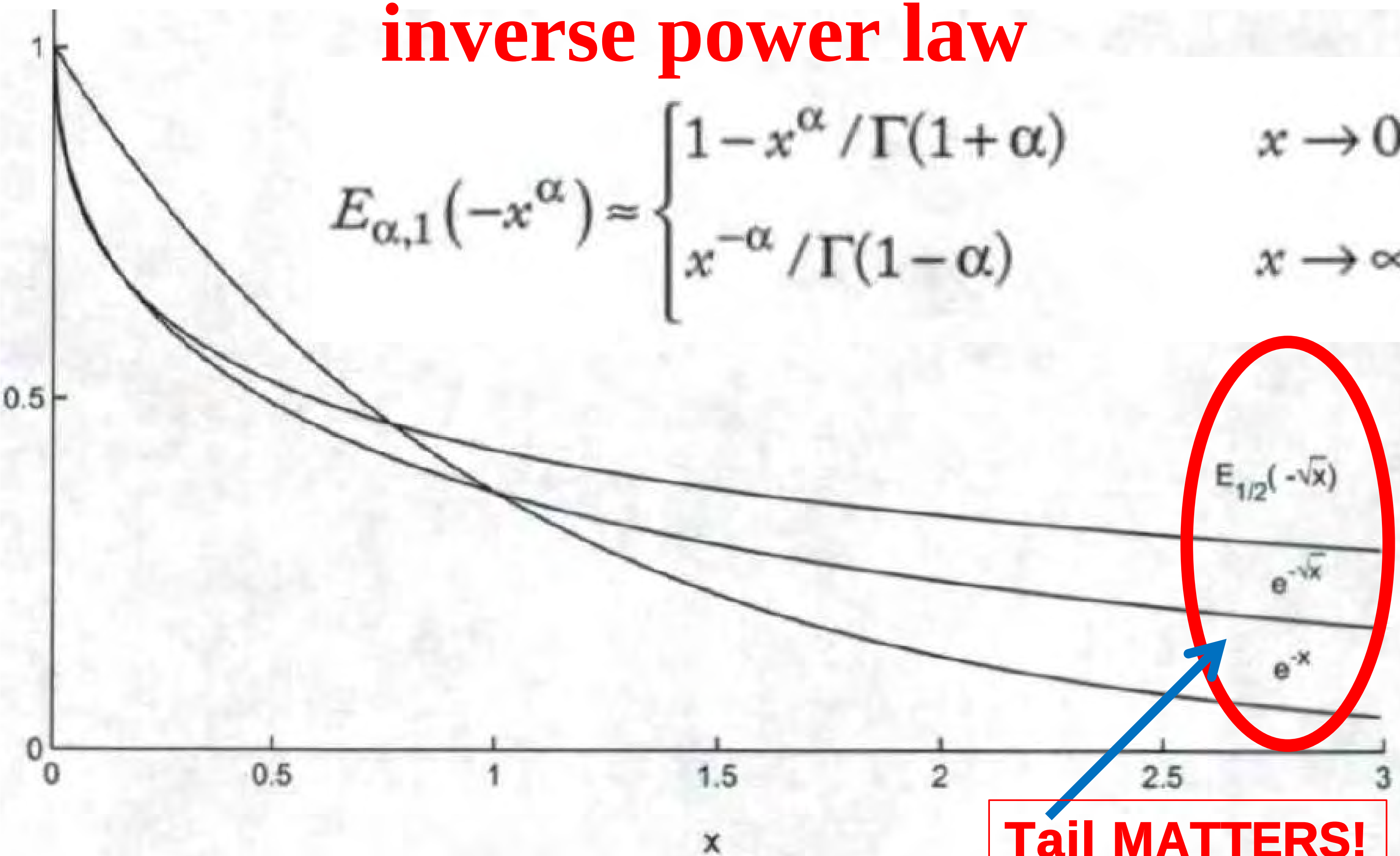
Mittag-Leffler Law

Fractional dynamic

Fractional calculus based models

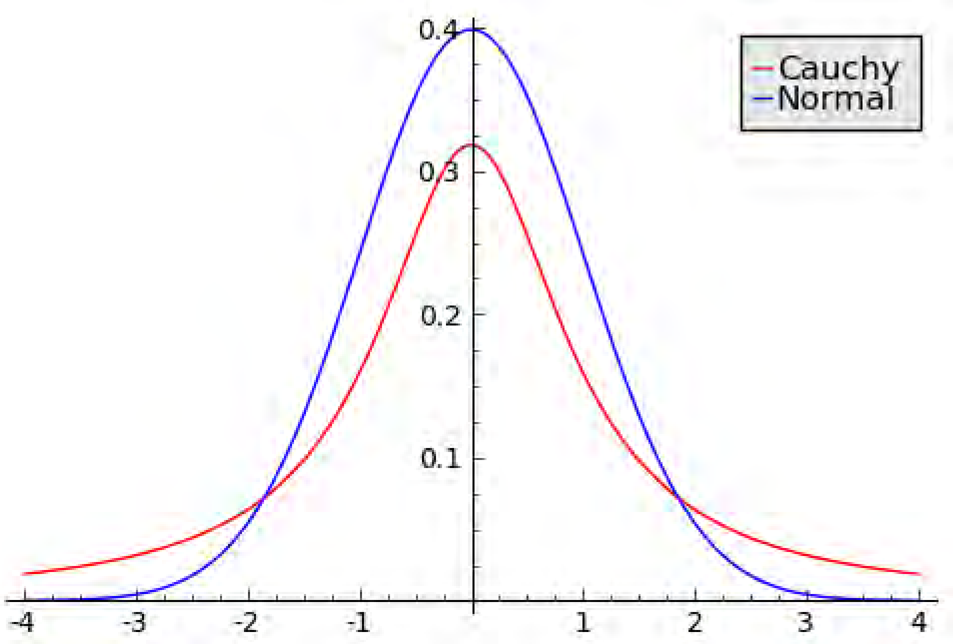
Root of long (algebraic) tail, or inverse power law

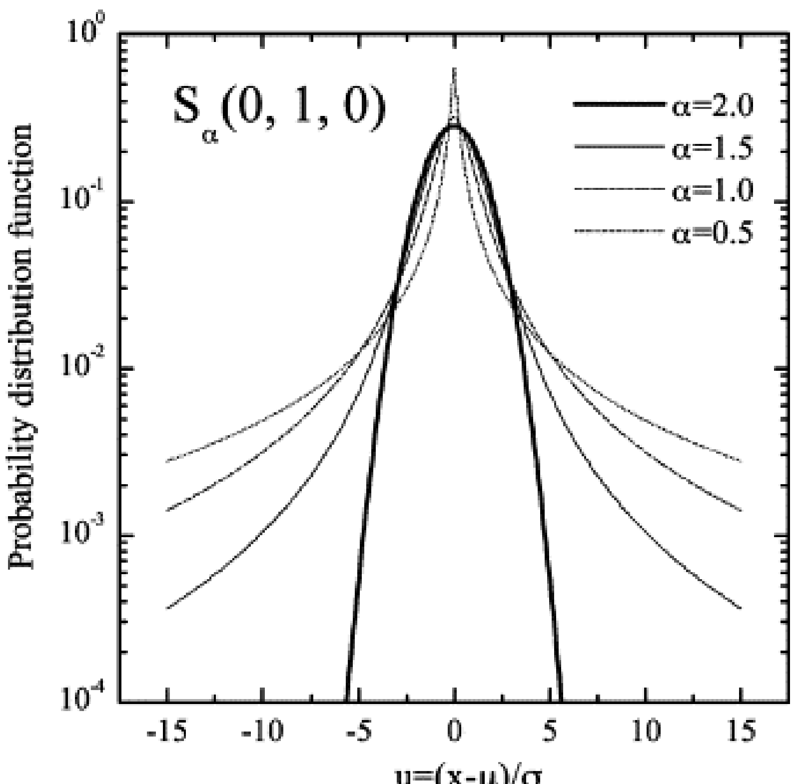
$$E_{\alpha,1}(-x^{\alpha}) \approx \begin{cases} 1 - x^{\alpha}/\Gamma(1+\alpha) & x \to 0^{+} \\ x^{-\alpha}/\Gamma(1-\alpha) & x \to \infty \end{cases}$$

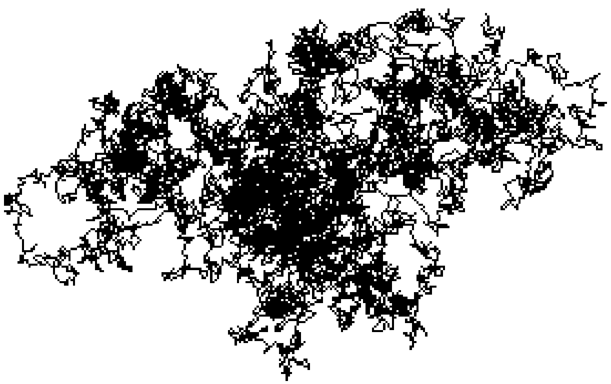


Heavy tail, fat tail

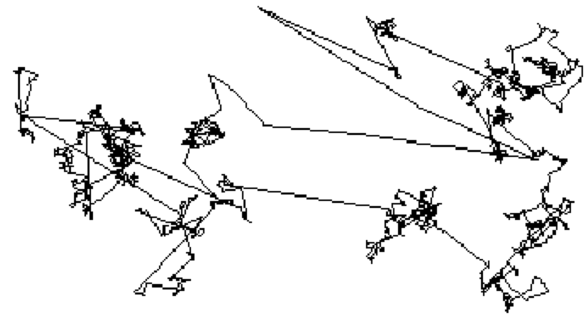
$$P[X > x] \sim x^{-\alpha}$$





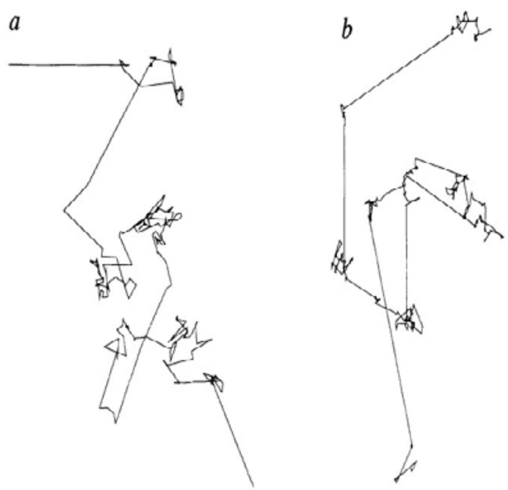


Brownian motion



Lévy flights

Wandering albatrosses



flight search patterns

G.M. Viswanathan, et al. ***Nature*** 381 (1996) 413–415.

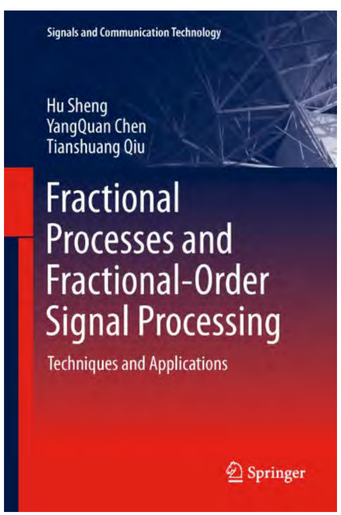
UCMERCED

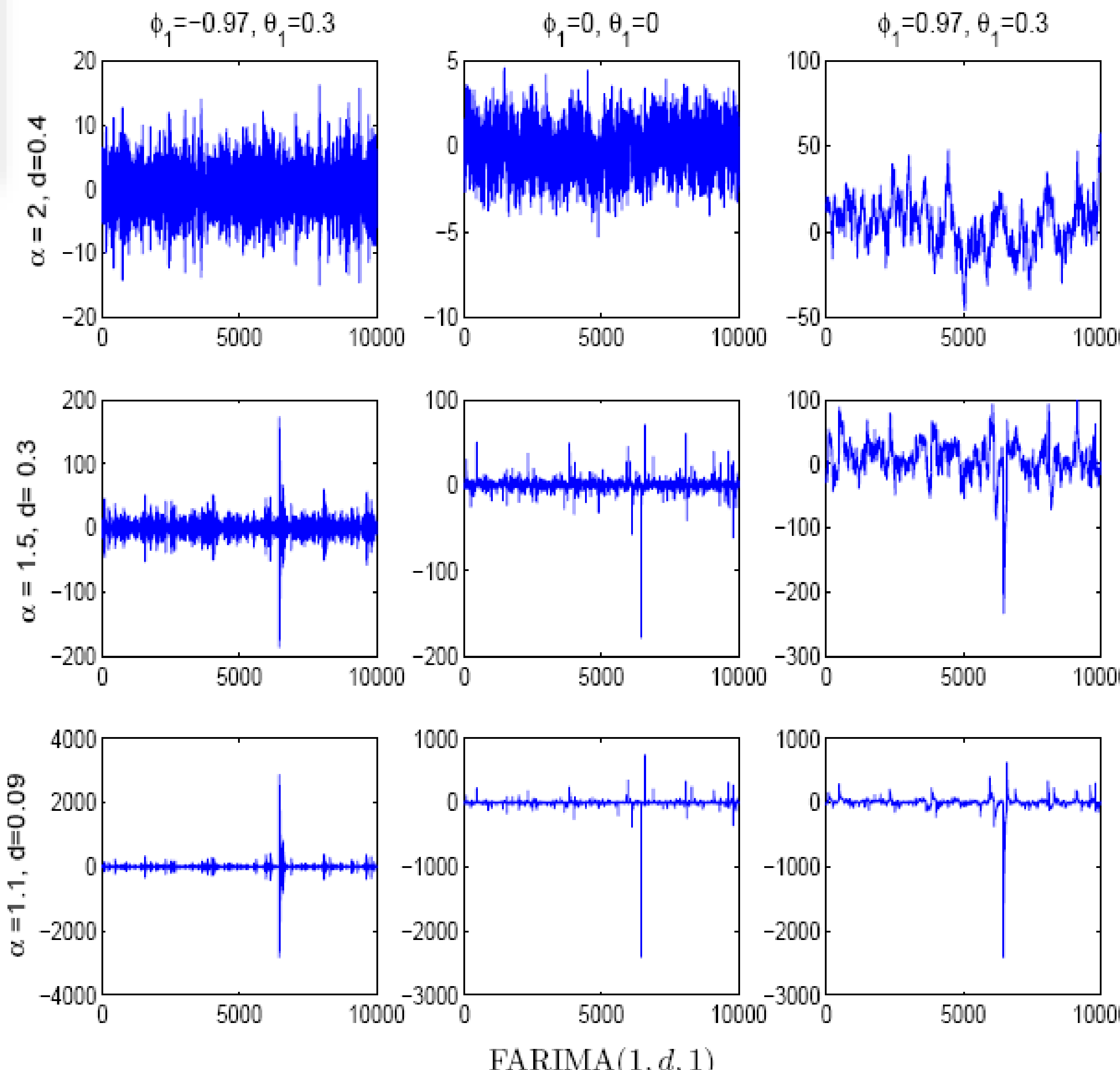
MESA LAB

Spikiness/Burstiness

[7] S. Stoev, and M.S. Taqqu, "Simulation Methods for Linear Fractional Stable Motion and FARIMA Using the Fast Fourier Transform". *Fractals*, 2004.

Kai Liu, YangQuan Chen,* and Xi Zhang. **An Evaluation of ARFIMA (Autoregressive Fractional Integral Moving Average) Programs.** *Axioms* **2017**, *6*(2), 16; doi:10.3390/axioms6020016





UCMERCED

MESA LAB

Long-range dependence

Self-similar

Hurst parameter

ARFIMA

α-stable distributions

Complex signals

Network Traffic, Smart Grids,
HRV, Outliers in time series,
biox signals ...

Fractional Gaussian noise (FGn)

“Spikiness” intermittence

“Heavy tails”

Fractional Brownian motion (FBm)

Connection to FC via PDF

- "*Fractional Calculus and Stable Probability Distributions*" (1998) by Rudolf Gorenflo , Francesco Mainardi http://arxiv.org/pdf/0704.0320.pdf

$$\frac{\partial u}{\partial t} = D(\alpha)\,\frac{\partial^\alpha u}{\partial |x|^\alpha}\,, \quad -\infty < x < +\infty\,, \quad t \geq 0\,,$$

$$\text{with} \quad u(x,0) = \delta(x) \quad 0 < \alpha \leq 2$$

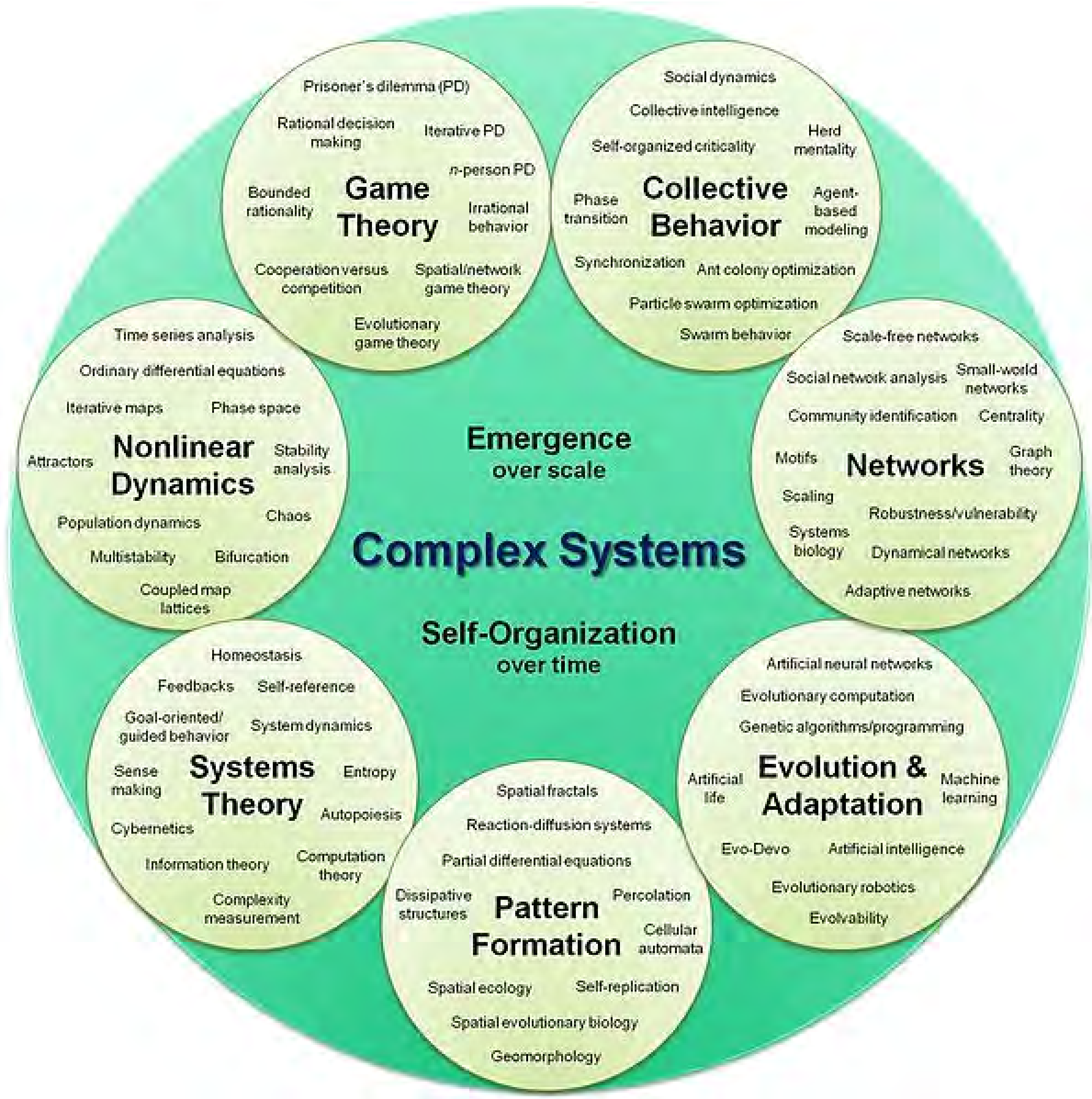
$$\frac{\partial^{2\beta} u}{\partial t^{2\beta}} = D(\beta)\,\frac{\partial^2 u}{\partial x^2}\,, \quad x \geq 0\,, \quad t \geq 0\,,$$

$$\text{with} \quad u(0,t) = \delta(t) \quad 0 < \beta < 1$$

UCMERCED MESA LAB

Credit: Bruce West

- Can these be synthesized?
- Is the fractional calculus entailed by complexity?



Take home messages

- Triangle:
  - Complexity
  - Power Law
  - Fractional Calculus

- Stochasticity with rich forms (heavytailedness)
  - Fractional order master equations

Outline

- **Fractional Calculus, Complexity, and Fractional Order Thinking**
- **Big Data, Variability, and Fractional Calculus**
- **Machine Learning, Optimal Randomness and Fractional Calculus**
- **Looking Into Future: Fractional Calculus is Physics Informed**

10 V's of Big Data

- **#1: Volume**
- **#2: Velocity**
- **#3: Variety**
- **#4: Variability**
- **#5: Veracity**
- **#6: Validity**
- **#7: Vulnerability**
- **#8: Volatility**
- **#9: Visualization**
- **#10: Value**

https://tdwi.org/articles/2017/02/08/10-vs-of-big-data.aspx

When talking about big data, we have to talk about "10V"s.

#1: Volume
#2: Velocity
#3: Variety
#4: Variability
#5: Veracity
#6: Validity
#7: Vulnerability
#8: Volatility
#9: Visualization
#10: Value

- Variability in big data's context refers to a few different things. One is the number of inconsistencies in the data. These need to be found by *anomaly and outlier detection* methods in order for any meaningful analytics to occur.
- Variability can also refer to diversity. In practice, the data can be classified into several different types, for example, healthy or unhealthy.

https://tdwi.org/articles/2017/02/08/10-vs-of-big-data.aspx

#1: Volume

- -- 300 hours of video are uploaded to YouTube every minute.
- -- An estimated 1.1 trillion photos were taken in 2016, and that number is projected to rise by 9 percent in 2017. As the same photo usually has multiple instances stored across different devices, photo or document sharing services as well as social media services, the total number of photos stored is also expected to grow from 3.9 trillion in 2016 to 4.7 trillion in 2017.
- -- In 2016 estimated global mobile traffic amounted for 6.2 exabytes per month. That's 6.2 billion gigabytes.

#2: Velocity

- Velocity refers to the speed at which data is being generated, produced, created, or refreshed.
- Sure, it sounds impressive that Facebook's data warehouse stores upwards of 300 petabytes of data, but the velocity at which new data is created should be taken into account. Facebook claims 600 terabytes of incoming data per day.
- Google alone processes on average more than "40,000 search queries every second," which roughly translates to more than 3.5 billion searches per day.

#3: Variety 多样化

- When it comes to big data, we don't only have to handle structured data but also semistructured and mostly unstructured data as well. As you can deduce from the above examples, most big data seems to be unstructured, but besides audio, image, video files, social media updates, and other text formats there are also log files, click data, machine and sensor data, etc.

#4: Variability 变化性

- Variability in big data's context refers to a few different things. One is the number of inconsistencies in the data. These need to be found by *anomaly and outlier detection* methods in order for any meaningful analytics to occur.
- Big data is also variable because of the multitude of data dimensions resulting from multiple disparate data types and sources. Variability can also refer to the inconsistent speed at which big data is loaded into your database.

#5: Veracity 真实性

- This is one of the unfortunate characteristics of big data. As any or all of the above properties increase, the veracity (confidence or trust in the data) drops. This is similar to, but not the same as, validity or volatility (see below). Veracity refers more to the provenance or reliability of the data source, its context, and how meaningful it is to the analysis based on it.

#6: Validity 有效性

- Similar to veracity, validity refers to how accurate and correct the data is for its intended use. According to Forbes, an estimated 60 percent of a data scientist's time is spent cleansing their data before being able to do any analysis. The benefit from big data analytics is only as good as its underlying data, so you need to adopt good data governance practices to ensure consistent data quality, common definitions, and metadata.

#7: Vulnerability 脆弱性

- Big data brings new security concerns. After all, a data breach with big data is a big breach. Does anyone remember the infamous AshleyMadison hack in 2015?
- Unfortunately there have been many big data breaches. Another example, as reported by CRN: in May 2016 "a hacker called Peace posted data on the dark web to sell, which allegedly included information on 167 million LinkedIn accounts and ... 360 million emails and passwords for MySpace users."
- Information on many others can be found at Information is Beautiful.

#8: Volatility 挥发性

- How old does your data need to be before it is considered irrelevant, historic, or not useful any longer? How long does data need to be kept for?
- Before big data, organizations tended to store data indefinitely -- a few terabytes of data might not create high storage expenses; it could even be kept in the live database without causing performance issues. In a classical data setting, there not might even be data archival policies in place.
- Due to the velocity and volume of big data, however, its volatility needs to be carefully considered. You now need to establish rules for data currency and availability as well as ensure rapid retrieval of information when required. Make sure these are clearly tied to your business needs and processes -- with big data the costs and complexity of a storage and retrieval process are magnified.

#9: Visualization

- Current big data visualization tools face technical challenges due to limitations of in-memory technology and poor scalability, functionality, and response time. You can't rely on traditional graphs when trying to plot a billion data points, so you need different ways of representing data such as data clustering or using tree maps, sunbursts, parallel coordinates, circular network diagrams, or cone trees.
- Combine this with the multitude of variables resulting from big data's variety and velocity and the complex relationships between them, and you can see that developing a meaningful visualization is not easy.

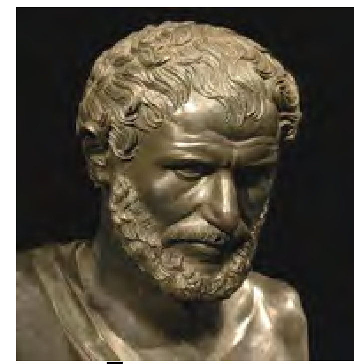
#10: Value

- Last, but arguably the most important of all, is value. The other characteristics of big data are meaningless if you don't derive business value from the data.
- Substantial value can be found in big data, including understanding your customers better, targeting them accordingly, optimizing processes, and improving machine or business performance. You need to understand the potential, along with the more challenging characteristics, before embarking on a big data strategy.

万变不离其宗

- 荀子·儒效. 《荀子》是战国时期(475-221BC)荀子和弟子们整理或记录他人言行的哲学著作。

- What is “宗” for Big Data?
  - **Variability**



- **“The Only Thing That Is Constant Is Change -” — Heraclitus 赫拉克利特(纪元前五世纪的希腊哲学家)**

(Fractional Order) Dynamics: per pre-Socratic Greek philosopher, Heraclitus

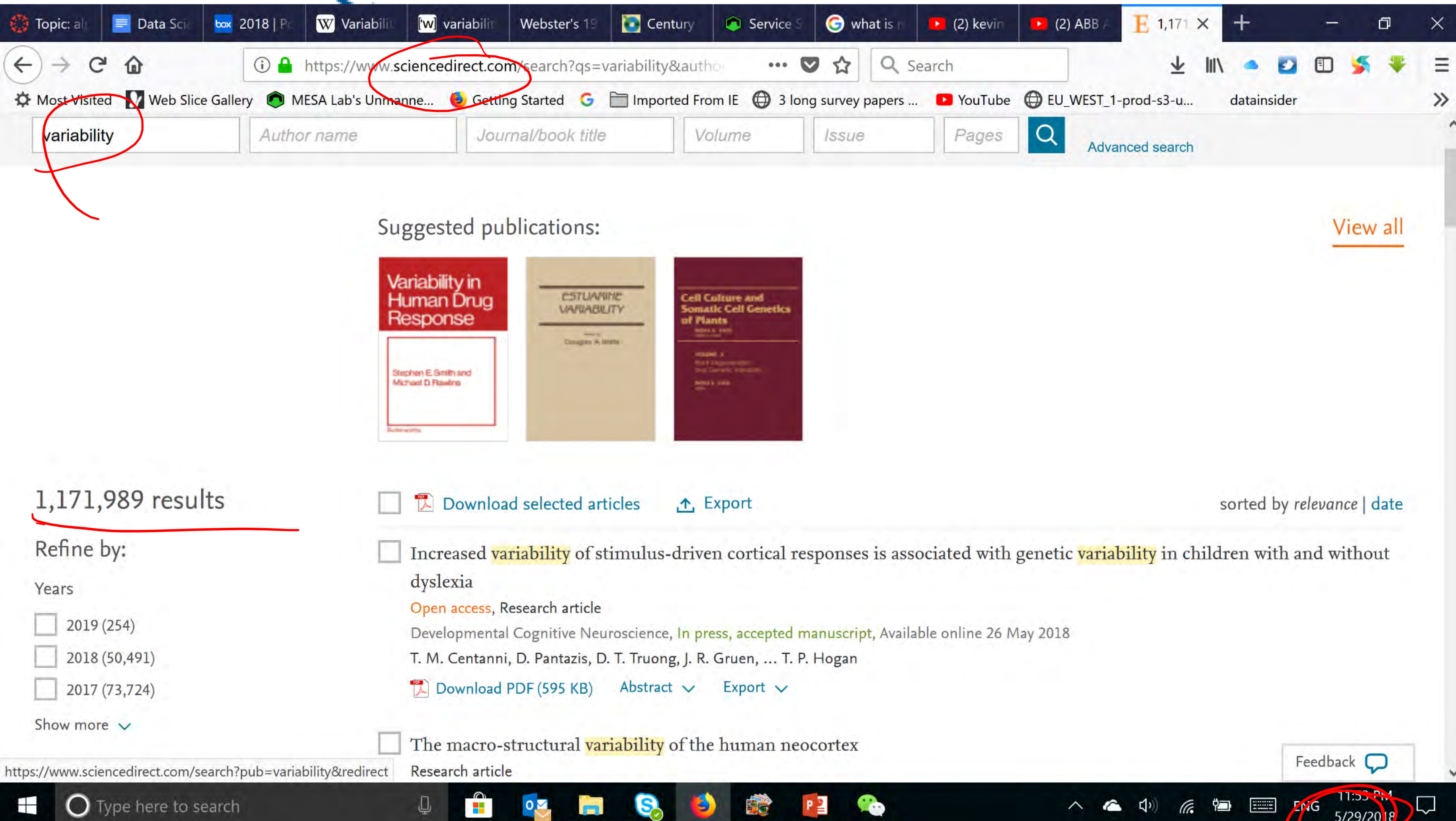
- "It is in changing that we find purpose."
  — Heraclitus
- "Nothing endures but change."
  — Heraclitus
- "**No man ever steps in the same river twice, for it's not the same river and he's not the same man.**"
  — Heraclitus

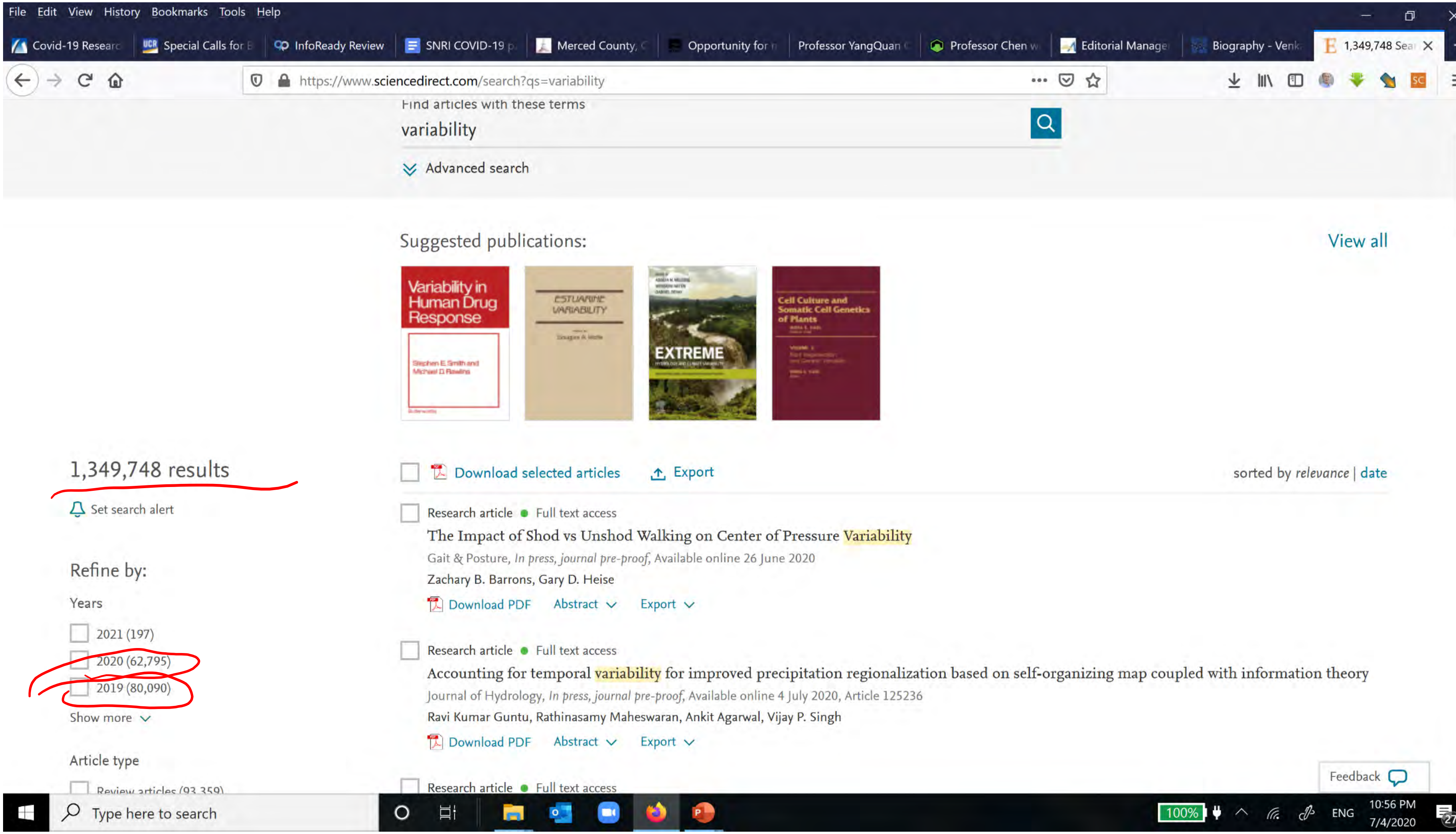
2000+ years later
Integer Order Calculus

- Invented late 17th century by Isaac Newton and Gottfried Wilhelm Leibniz
- Note: Integer Oder Dynamic view of Changes is only for our own "convenience" (Scott Blair)
- Denying fractional calculus is as saying there is no non-integers in between integers!

"We may express our concepts in Newtonian terms if we find this convenient but, if we do so, we must realize that we have made a translation into a language which is foreign to the system which we are studying." (1950)

G W Scott Blair


https://www.sciencedirect.com/search?qs=variability&autho
variability
Author name
Journal/book title
Volume
Issue
Pages
Advanced search
Suggested publications:
View all
Variability in Human Drug Response
Stephen E Smith and Michael D Rawlins
ESTUARINE VARIABILITY
Cell Culture and Somatic Cell Genetics of Plants
1,171,989 results
Refine by:
Years
2019 (254)
2018 (50,491)
2017 (73,724)
Show more
Download selected articles
Export
sorted by relevance | date
Increased variability of stimulus-driven cortical responses is associated with genetic variability in children with and without dyslexia
Open access, Research article
Developmental Cognitive Neuroscience, In press, accepted manuscript, Available online 26 May 2018
T. M. Centanni, D. Pantazis, D. T. Truong, J. R. Gruen, ... T. P. Hogan
Download PDF (595 KB)
Abstract
Export
The macro-structural variability of the human neocortex
Research article
Feedback
https://www.sciencedirect.com/search?pub=variability&redirect
Type here to search
5/29/2018

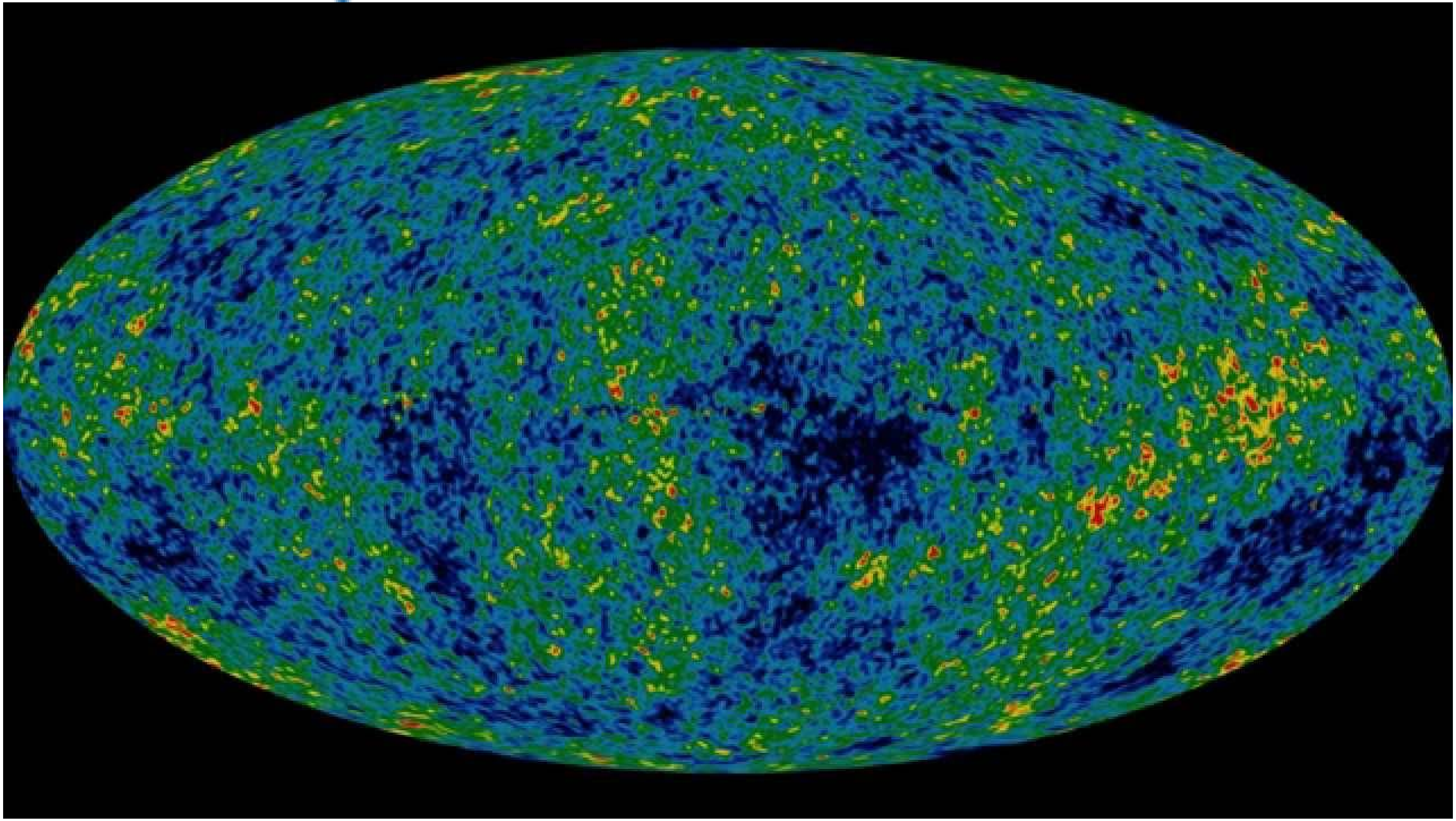

File Edit View History Bookmarks Tools Help
https://www.sciencedirect.com/search?qs=variability
Find articles with these terms
variability
Advanced search
Suggested publications:
View all
Variability in Human Drug Response
ESTUARINE VARIABILITY
EXTREME
Cell Culture and Somatic Cell Genetics of Plants
1,349,748 results
Set search alert
Refine by:
Years
2021 (197)
2020 (62,795)
2019 (80,090)
Show more
Article type
Download selected articles
Export
sorted by relevance | date
Research article Full text access
The Impact of Shod vs Unshod Walking on Center of Pressure Variability
Gait & Posture, In press, journal pre-proof, Available online 26 June 2020
Zachary B. Barrons, Gary D. Heise
Download PDF Abstract Export
Research article Full text access
Accounting for temporal variability for improved precipitation regionalization based on self-organizing map coupled with information theory
Journal of Hydrology, In press, journal pre-proof, Available online 4 July 2020, Article 125236
Ravi Kumar Guntu, Rathinasamy Maheswaran, Ankit Agarwal, Vijay P. Singh
Download PDF Abstract Export
Research article Full text access
Feedback
Type here to search
10:56 PM
7/4/2020

(anything's) variability

- Climate variability, changes in the components of Earth's climate system and their interactions
- Genetic variability, a measure of the tendency of individual genotypes in a population to vary from one another
- Heart rate variability, a physiological phenomenon where the time interval between heart beats varies
- Human variability, the range of possible values for any measurable characteristic, physical or mental, of human beings
- Spatial variability, when a quantity that is measured at different spatial locations exhibits values that differ across the locations
- Statistical variability, a measure of dispersion in statistics

- Gait, breath, cognitive, temperature, soil, crop, ……

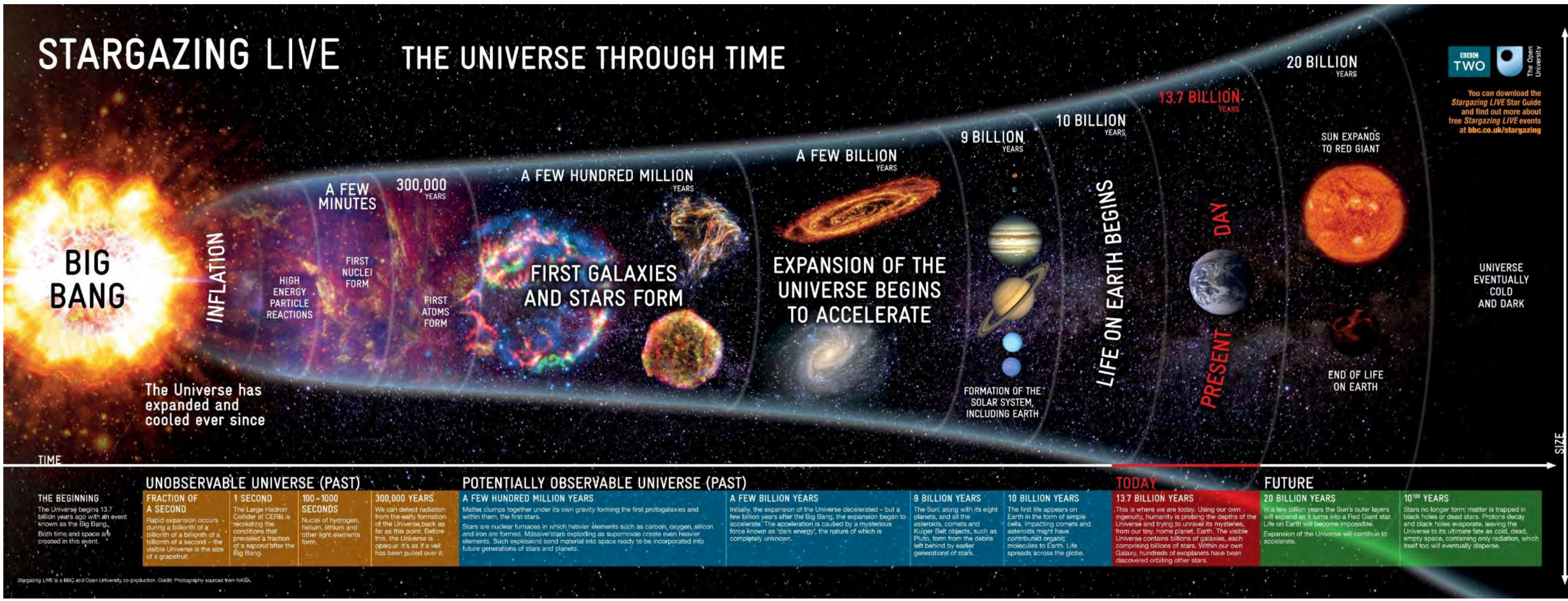
https://en.wikipedia.org/wiki/Variability



Cosmic Background Radiation

https://i.ytimg.com/vi/WB5jmdJvQeU/maxresdefault.jpg

Variability in universe? True



http://www.americaspace.com/wp-content/uploads/2015/08/Page11.jpg

Variability in "big data"?

- Sure!
- But how MAD about it?
  - We need Fractional Calculus!

So, to be complex to have big data??

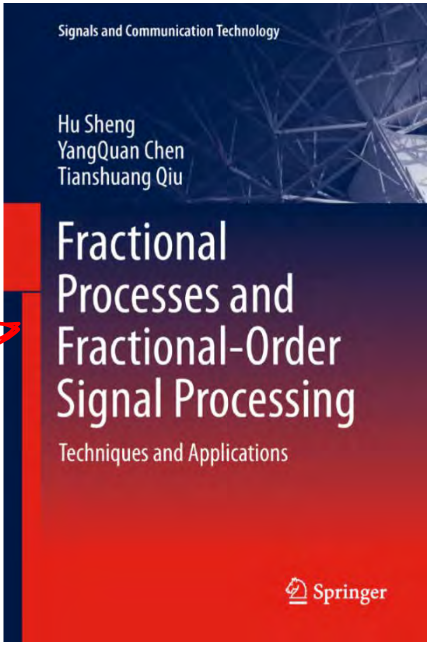
- Sure!
- But how MAD about it?
  - We need Fractional Calculus!

Yes, BD should meet FC!

HOW?

FODA: Fractional Order Data Analytics

- First proposed by Prof. YangQuan Chen 2015.
- *Metrics based on using fractional order signal processing techniques for quantifying the generating dynamics of observed or perceived variabilities.*
  – Hurst parameter, fGn, fBm, …
  – Fractional order integral, differentiation
  – FLOM/FLOS (fractional order lower order moments/statistics)
  – Alpha stable processes, Levy flights
  – ARFIMA, GARMA (Gegenbauer), CTRW .



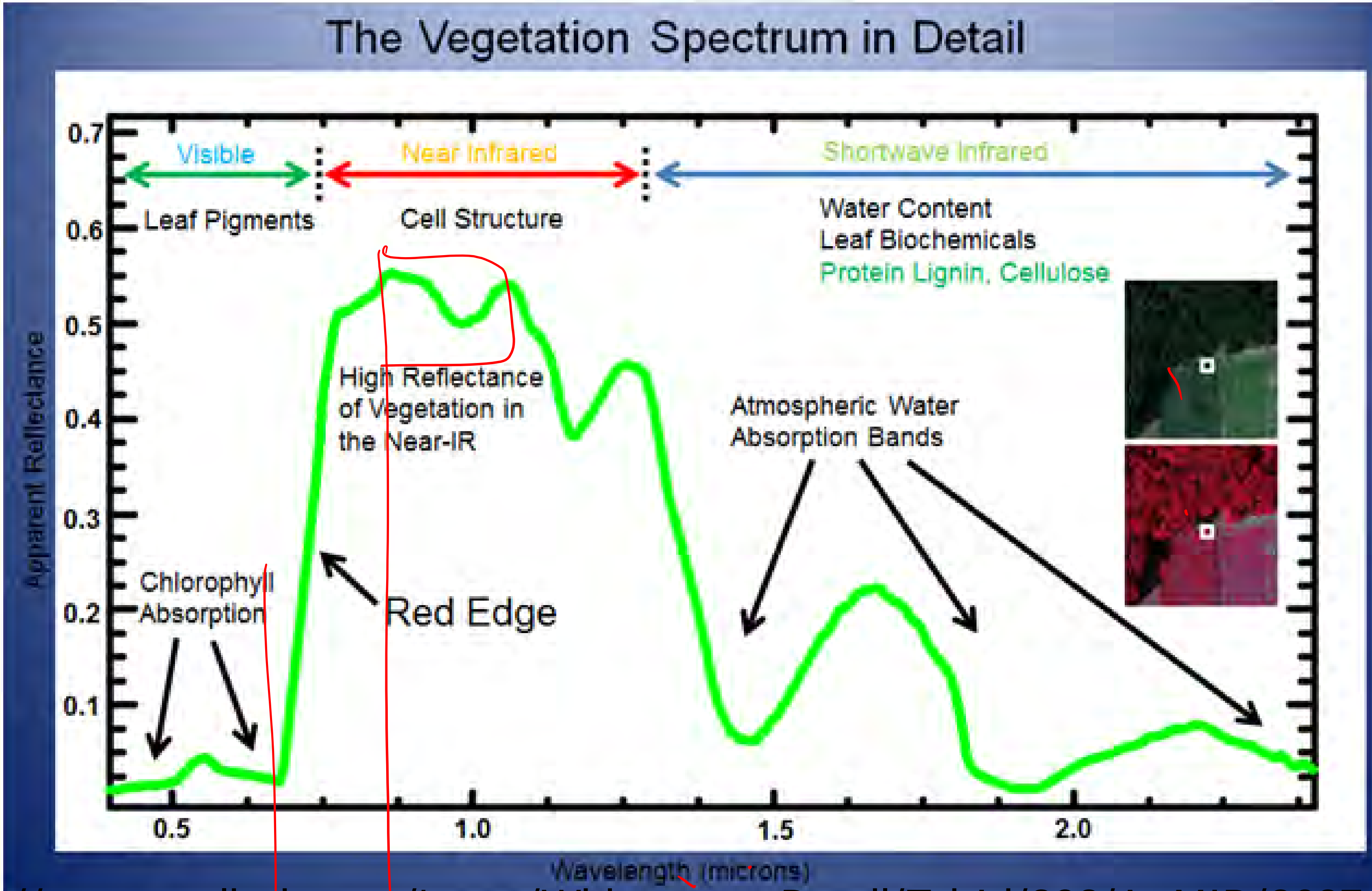
UCMERCED MESA LAB

Fractional Order Data Analytics:

connecting dots of Drones, Big Data, and Fractional Calculus

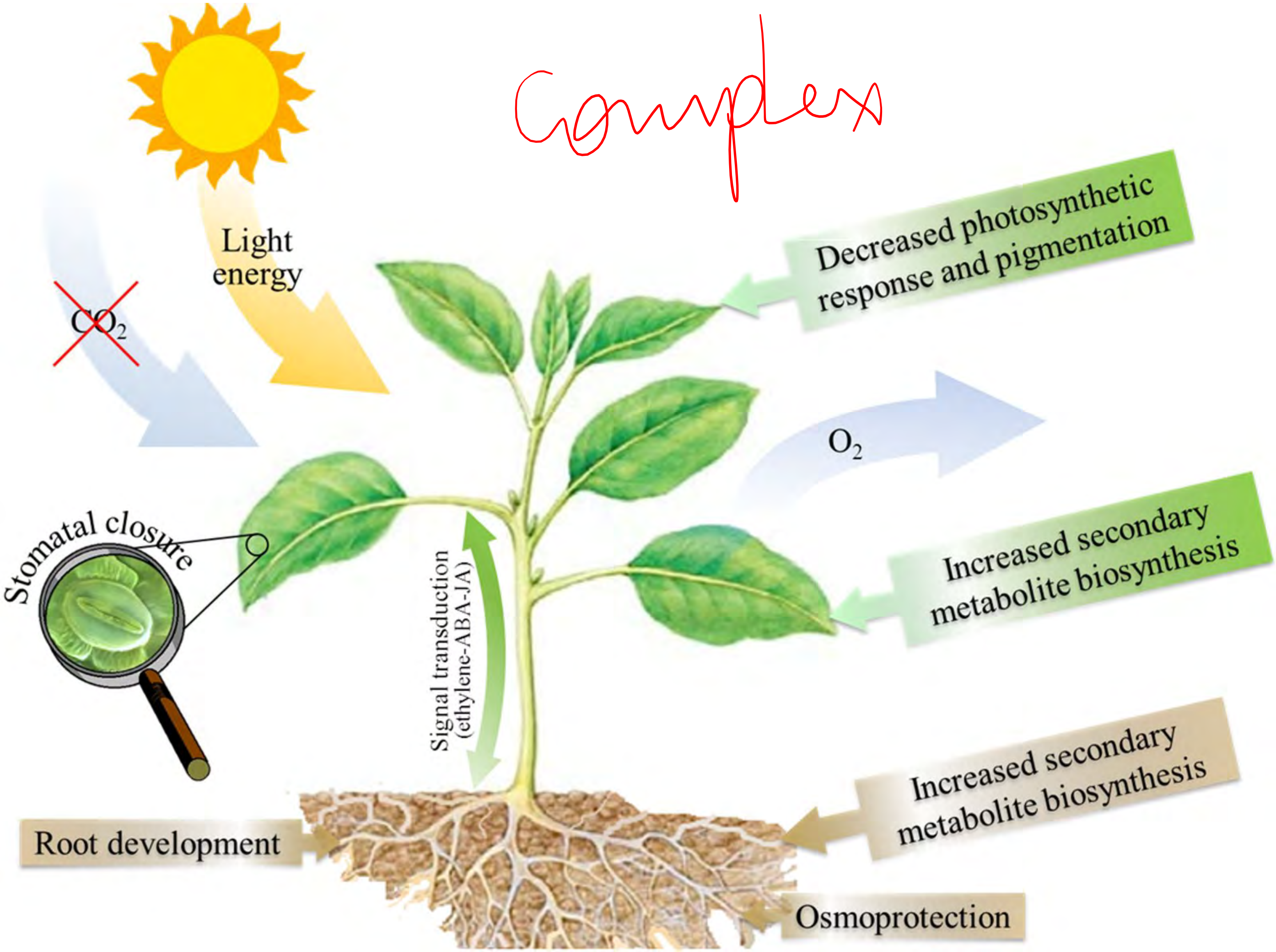
YangQuan Chen, Ph.D., Director,
MESA(Mechatronics, Embedded Systems and Automation)LAB
ME/EECS/SNRI/HSRI, School of Engineering,
University of California, Merced
E: yqchen@ieee.org; *or*, yangquan.chen@ucmerced.edu
T: (209)228-4672; **O**: SE1-254; **Lab**: CAS Eng 820 (**T**: 228-4398)

March 21, 2015. Saturday 2:00-2:15 PM
Robots & Ribs Day @ MESA LAB Symposium @ UC Merced



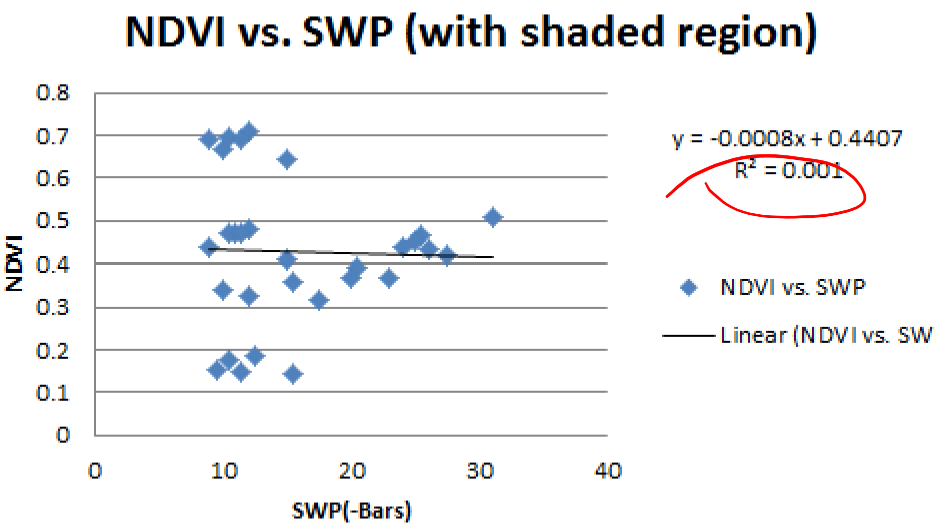
https://www.exelisvis.com/Learn/WhitepapersDetail/TabId/802/ArtMID/2627/ArticleID/13742/Vegetation-Analysis-Using-Vegetation-Indices-in-ENVI.aspx

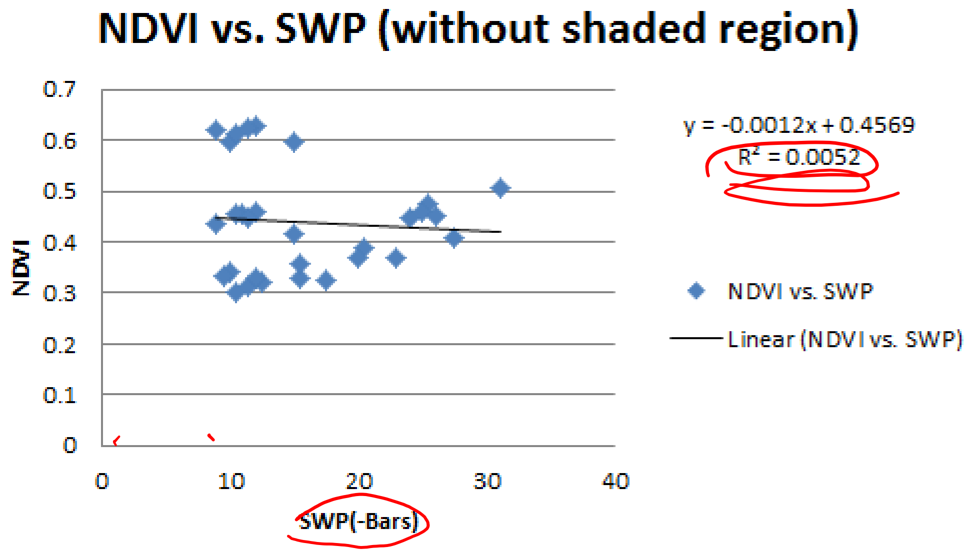
Complex

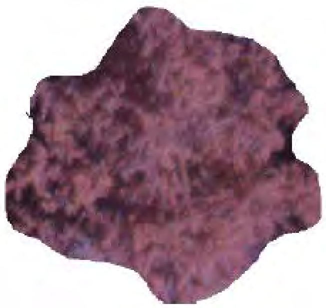


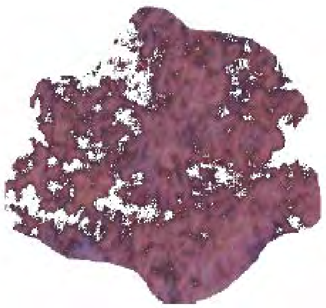
http://www.intechopen.com/books/responses-of-organisms-to-water-stress/water-stress-and-agriculture

NDVI vs. water stress??







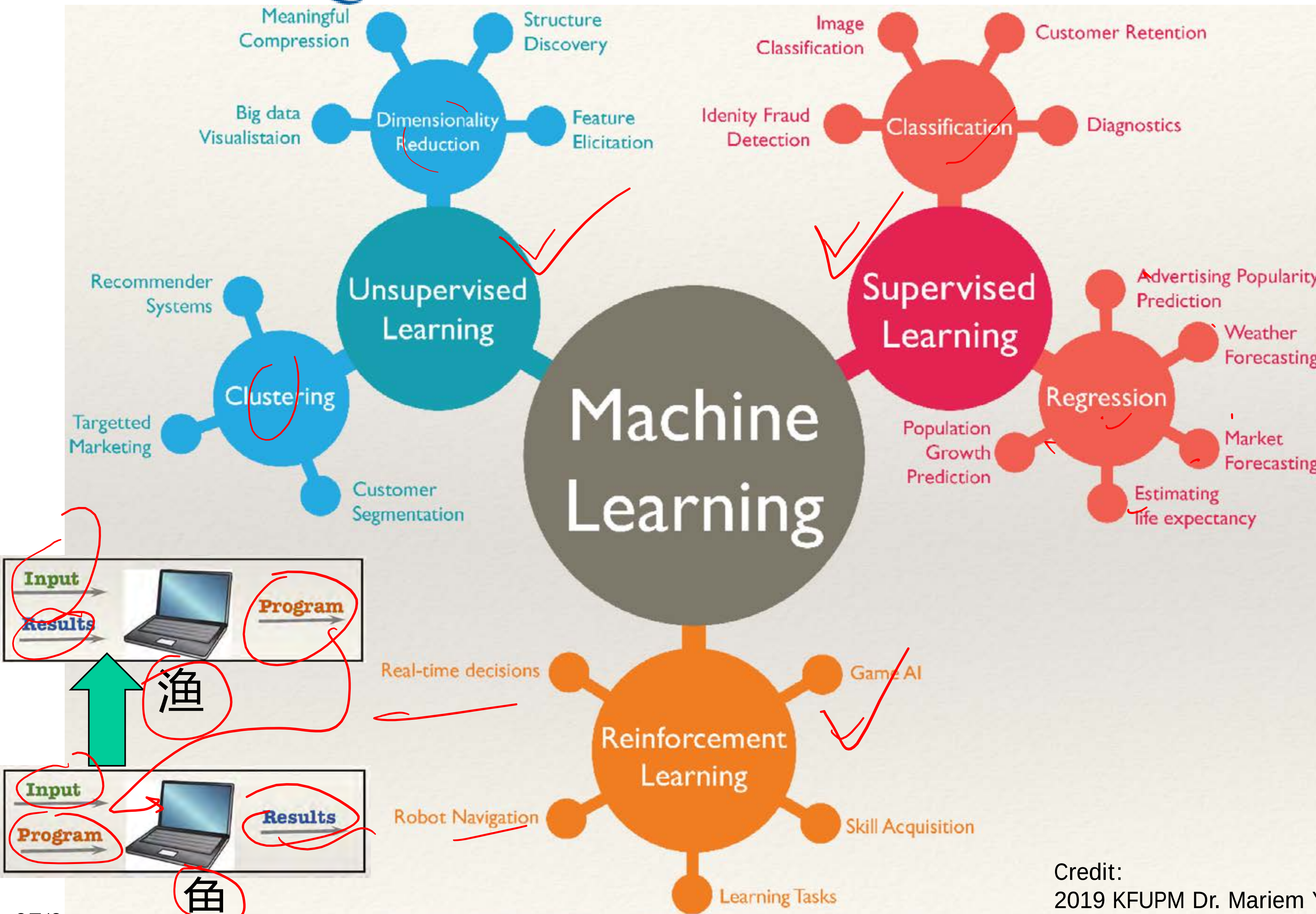


Drones as “Tractor 2.0” for Farmers

- RRR or SSM of water, fertilizers, pesticides etc.
- Fractional Calculus may save the world one day.
- Drones create big data and demand FODA due to “complexity” thus variability, inherent in life process.

Outline

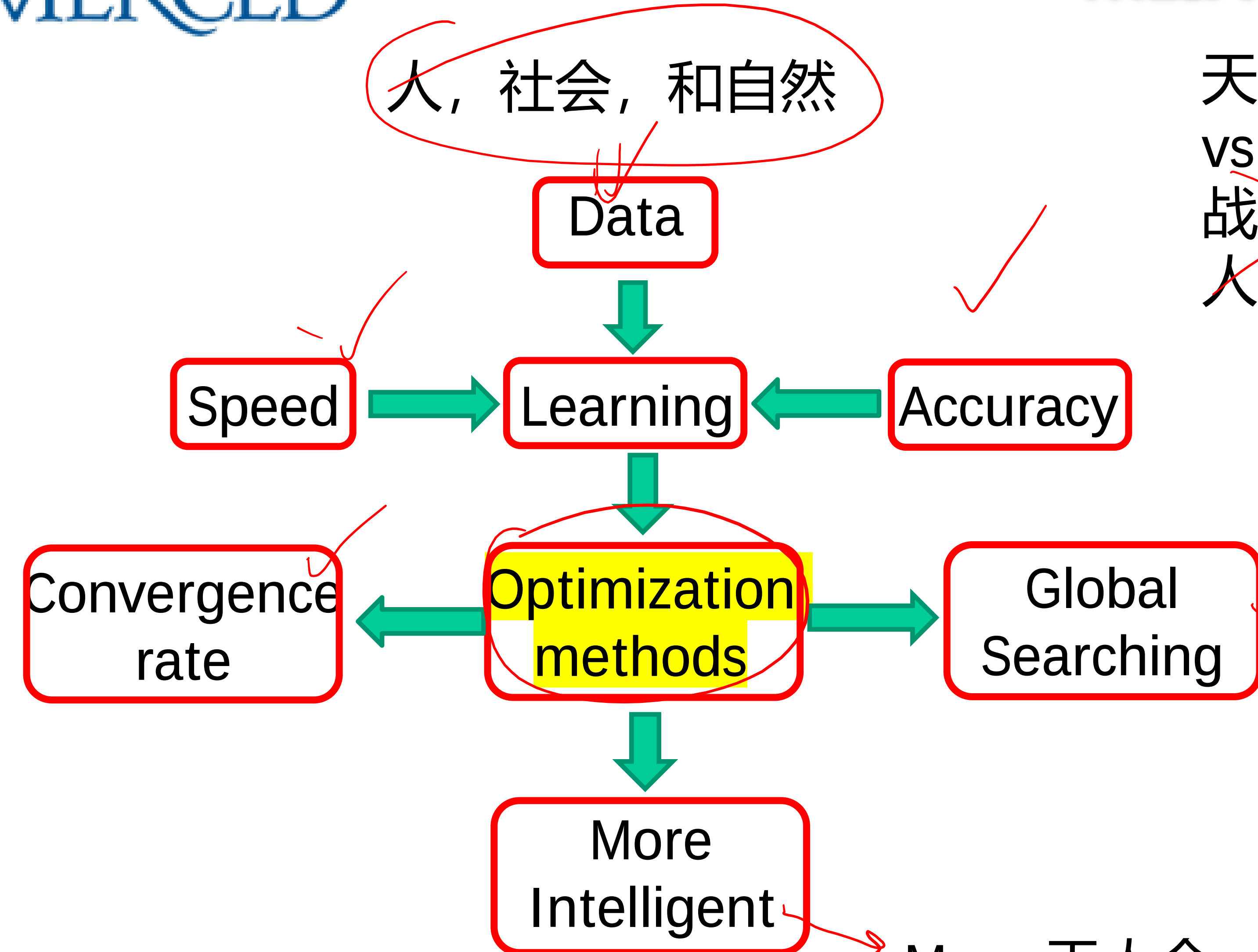
- **Fractional Calculus, Complexity, and Fractional Order Thinking**
- **Big Data, Variability, and Fractional Calculus**
- **Machine Learning, Optimal Randomness and Fractional Calculus**
- **Looking Into Future: Fractional Calculus is Physics Informed**



Credit:
2019 KFUPM Dr. Mariem Y. Zouch

UCMERCED

MESA LAB



天人合一

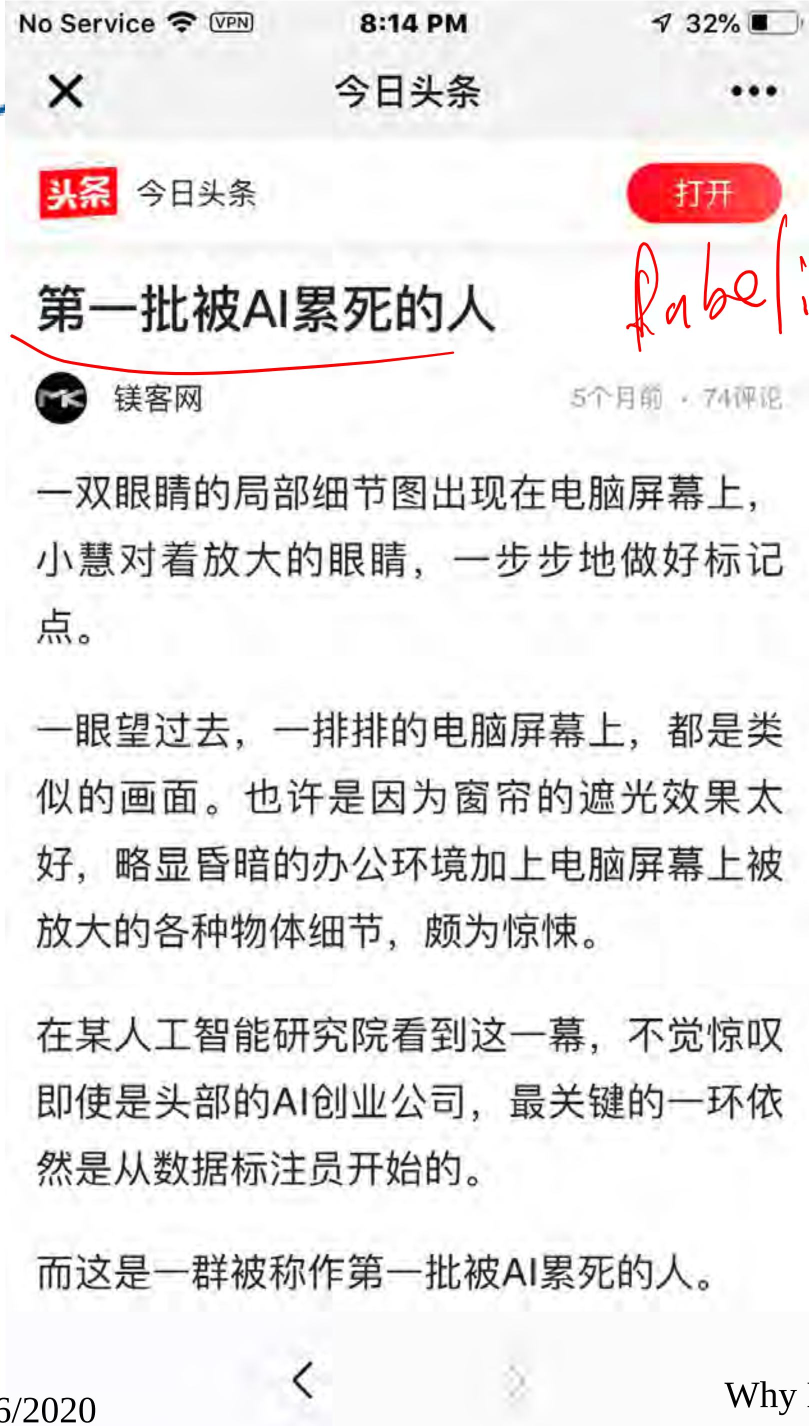
vs.

~~战天斗地~~

~~人定胜天?~~

More 天人合一

MESA LAB


No Service 8:14 PM 32%
今日头条
头条 今日头条 打开
第一批被AI累死的人
镁客网 5个月前 · 74评论
一双眼睛的局部细节图出现在电脑屏幕上，小慧对着放大的眼睛，一步步地做好标记点。
一眼望过去，一排排的电脑屏幕上，都是类似的画面。也许是因为窗帘的遮光效果太好，略显昏暗的办公环境加上电脑屏幕上被放大的各种物体细节，颇为惊悚。
在某人工智能研究院看到这一幕，不觉惊叹即使是头部的AI创业公司，最关键的一环依然是从数据标注员开始的。
而这是一群被称作第一批被AI累死的人。

Labeling

人工智能：

人工然后智能

Artificial then intelligent

学习（算法）很重要

不然会累死人的

Learning algorithm is important, or we will tire to death

UCMERCED

MESA LAB

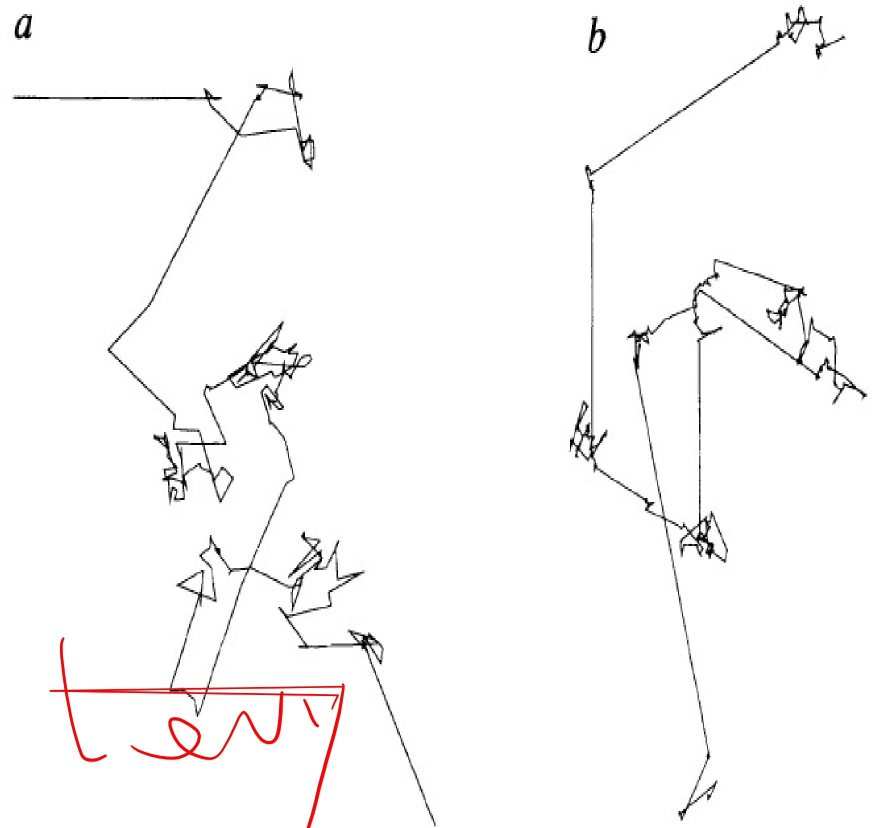
Reflection 沉思:

- (Machine) Learning is now a hot research topic;
- How to learn efficiently (optimally) is always important;
- The key for learning is the optimization method;
- Thus designing an efficient optimization method is the most important topic now
  - What is the optimal way to optimize?
  - What is the *more* optimal way to optimize?

ML core is optimization, can we demand "More Optimal Machine Learning" (i.e., DL with minimum/smallest labelled data) ?

Levy flight is optimized randomness for albatross via millions of year evolution or slow optimization

Wandering albatrosses

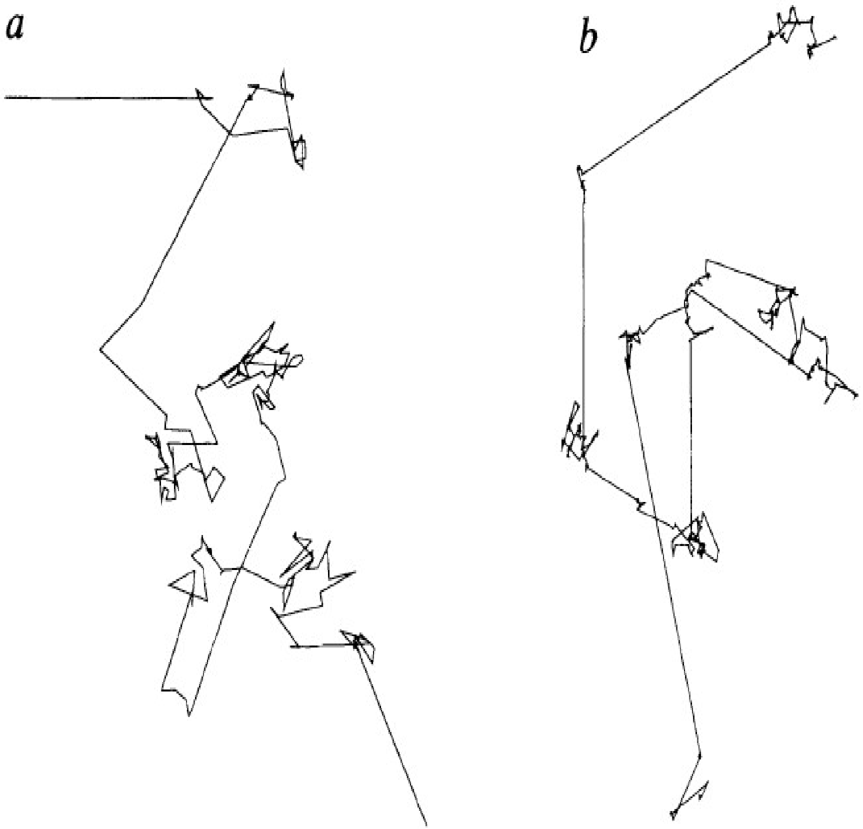


Levy

flight search patterns

G.M. Viswanathan, et al. ***Nature*** 381 (1996) 413–415.

Levy flight is optimized randomness for albatross via millions of year evolution or slow optimization



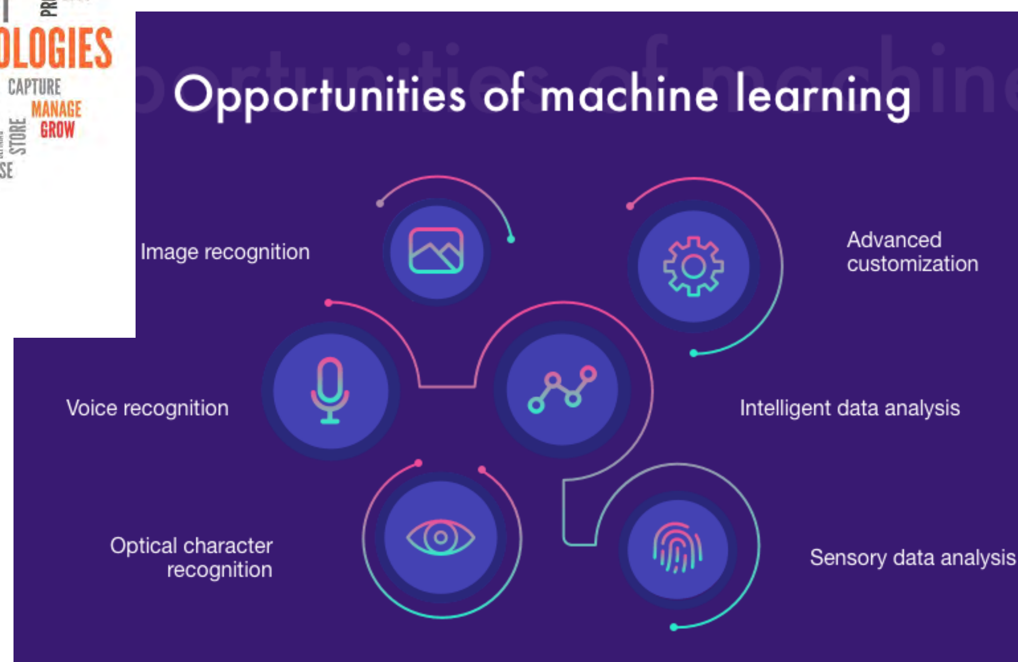
Can it be <u>more optimal?</u>

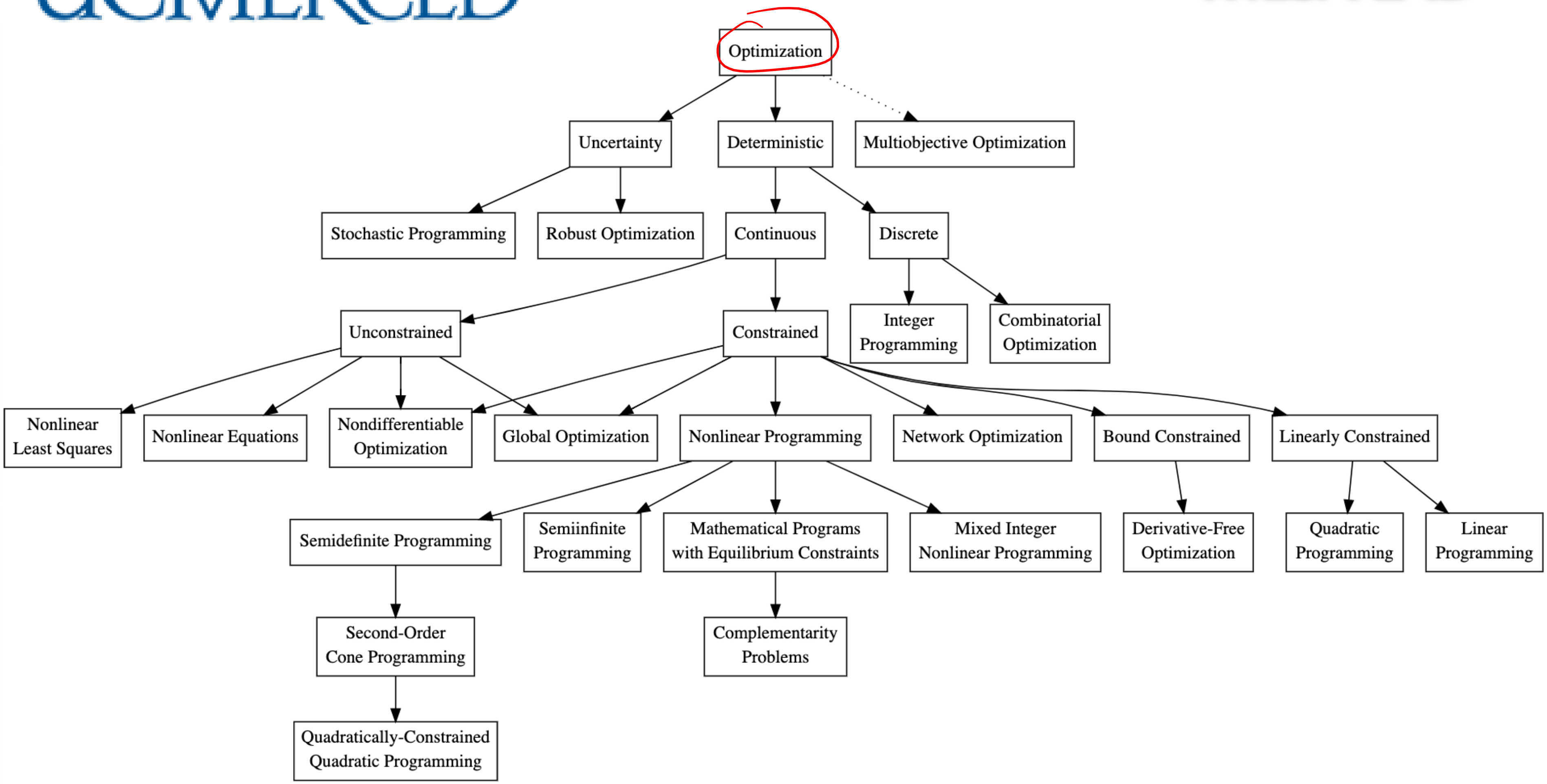
Wandering albatrosses

flight search patterns

G.M. Viswanathan, et al. ***Nature*** 381 (1996) 413–415.

Nowadays, Big Data and Machine Learning are two hottest topics and they are closely related to each other. To better understand them, we also need F.C. （分数阶微积分） and H.T. （重尾）





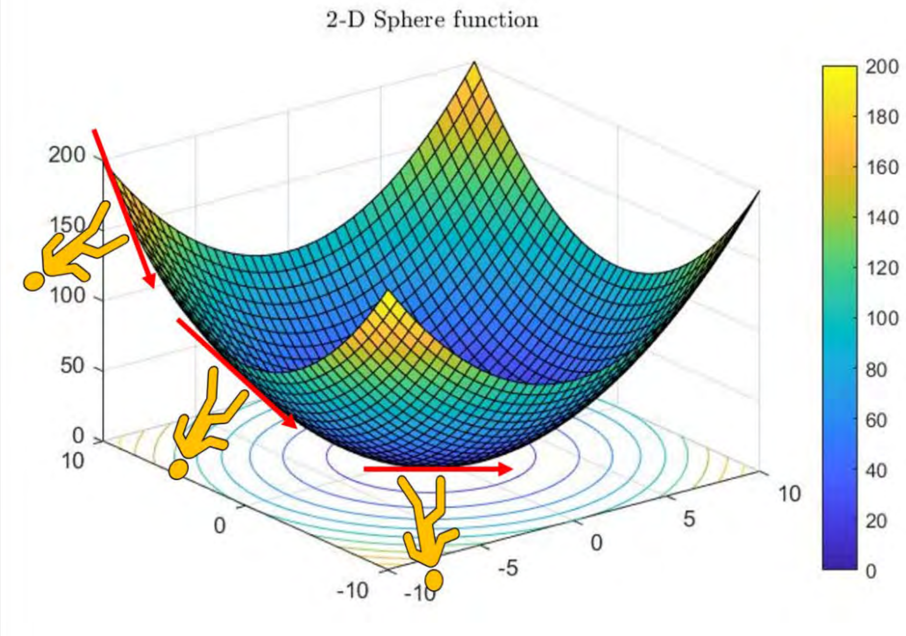
https://math.stackexchange.com/questions/3664716/taxonomy-overview-of-optimization-methods

Two broad categories

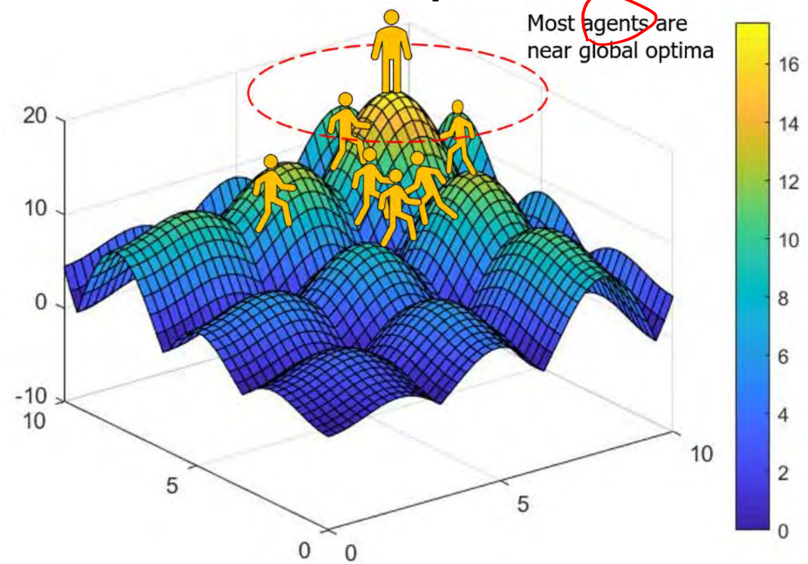
- Derivative-free
  - Direct Search, NM, PSO GA
- Gradient-based
  - GD and its variants

Derivative-free

Single agent search vs. swarm-based search



2-D Alpine function



$$f(\mathbf{x}) = f(x_1, \ldots, x_n) = \sum_{i=1}^{n} |x_i \sin(x_i) + 0.1 x_i|$$

Exploration 探索 is often achieved by randomization or random numbers in terms of some predefined probability distributions.

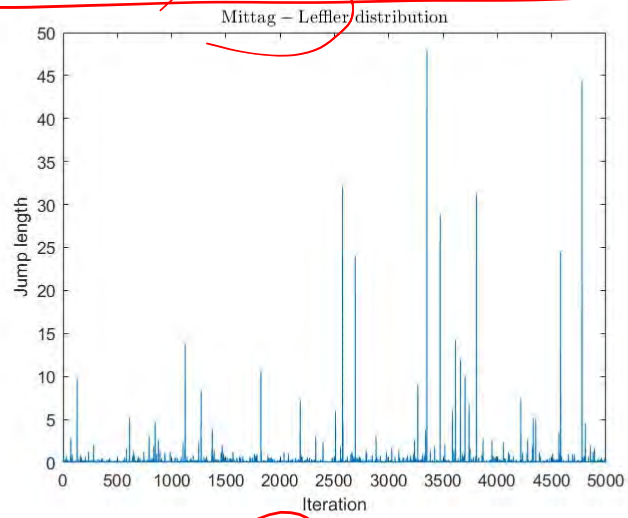
Exploitation **开拓利用** uses local information such as gradients to search local regions more intensively, and such intensification can enhance the rate of convergence.

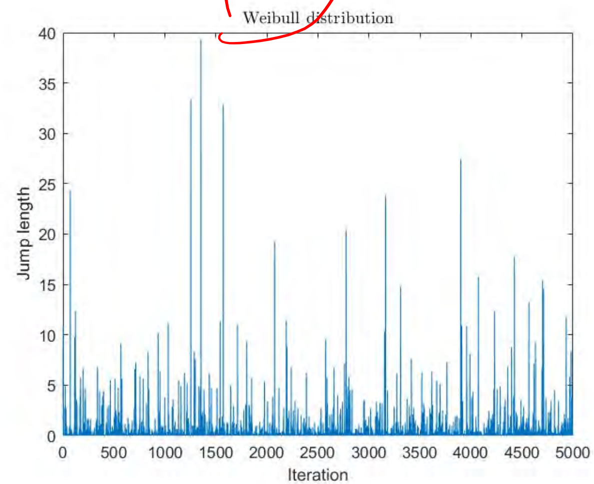
What is the optimal randomness?

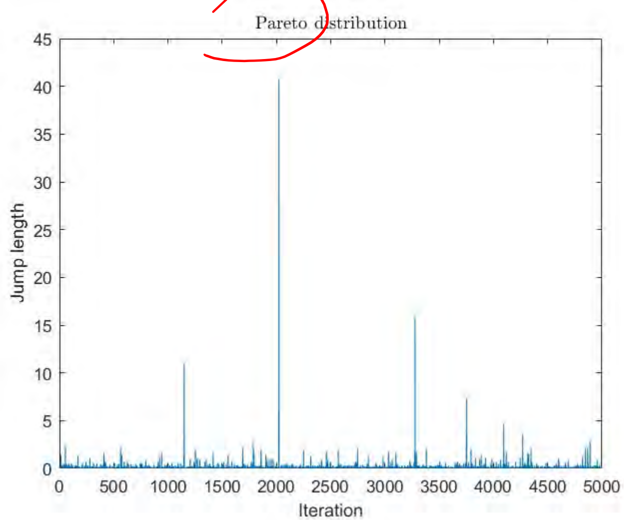
?

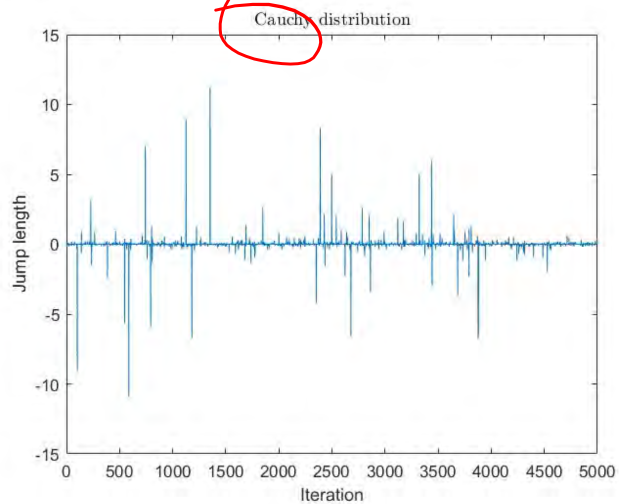
J Wei, YQ Chen, Y Yu, Y Chen (2019). *Optimal Randomness in Swarm-Based Search*. **Mathematics** 7 (9), 828 [PDF]

4 HT 重尾 distributions – sample paths









2.3 Experimental results

No.	Test function
1	F_{sph}: Sphere's Function
2	F_{ros}: Rosenbrock's Function
3	F_{ack}: Ackley's Function
4	F_{grw}: Griewank's Function
5	F_{ras}: Rastrigin's Function
6	F_{sch}: Generalized Schwefel's Problem 2.26
7	F_{sal}: Salomon's Function
8	F_{wht}: Whitely's Function
9	F_{pn1}: Generalized Penalized Function 1
10	F_{pn2}: Generalized Penalized Function 2
11	F_1: Shifted Sphere Function
12	F_2: Shifted Schwefel's Problem 1.2
13	F_3: Shifd Rotated High Conditioned Elliptic Function
14	F_4: Shifted Schwefel's Problem 1.2 with Noise in Fitness
15	F_5: Schwfel's Problem 2.6 with global Optimum on Bounds
16	F_6: Shifted Rosenbrock's Function
17	F_7: Shifted Rotated Griewank's Function without Bounds
18	F_8: Shifted Rotated Ackley's Function with Global Optimum on Bounds
19	F_9: Shifted Rastrigin's Function
20	F_{10}: Shifted Rotated Rastrigin's Function

Parameters configuration
$NP = 20$ (population size),
$P_a = 0.25$ (discovery probability),
Max_FEs=10,000*D (termination criterion),
$runs = 50$ (Running times)

Table 2. Comparisons between CS and four randomness-enhanced CS algorithms at $D = 30$.

Fun	CS	CSML	CSP	CSC	CSW
F_{sph}	9.58E-31	4.90E-54‡	**4.74E-59**‡	1.17E-57‡	4.40E-51‡
F_{ros}	1.20E+01	5.22E+00‡	3.10E+00‡	**2.74E+00**‡	8.62E+00‡
F_{ack}	7.70E-13	1.06E-14‡	1.07E-14‡	9.56E-15‡	**8.28E-15**‡
F_{grw}	7.11E-17	**0.00E+00**‡	**0.00E+00**‡	**0.00E+00**‡	**0.00E+00**‡
F_{ras}	2.32E+01	1.38E+01‡	1.88E+01‡	1.49E+01‡	**8.34E+00**‡
F_{sch}	1.57E+03	5.37E+02‡	1.32E+03‡	4.80E+02‡	**3.56E+01**‡
F_{sal}	3.76E-01	2.96E-01‡	3.00E-01‡	2.84E-01‡	**2.20E-01**‡
F_{wht}	3.73E+02	2.00E+02‡	2.49E+02‡	2.27E+02‡	**1.93E+02**‡
F_{pn1}	2.07E-03	**1.57E-32**‡	**1.57E-32**‡	2.07E-03$^≈$	**1.57E-32**‡
F_{pn2}	4.82E-28	**1.35E-32**‡	**1.35E-32**‡	**1.35E-32**‡	**1.35E-32**‡
F_1	6.48E-30	**0.00E+00**‡	**0.00E+00**‡	**0.00E+00**‡	**0.00E+00**‡
F_2	1.05E-02	1.10E-03‡	**2.77E-04**‡	1.40E-03‡	1.23E-02†
F_3	**2.17E+06**	3.04E+06†	2.99E+06†	3.25E+06†	3.61E+06†
F_4	1.79E+03	4.98E+02‡	**3.58E+02**‡	4.02E+02‡	5.51E+02‡
F_5	3.17E+03	2.44E+03‡	1.98E+03‡	2.11E+03‡	**1.94E+03**‡
F_6	2.78E+01	1.57E+01‡	**9.91E+00**‡	1.23E+01‡	1.59E+01‡
F_7	**1.34E-03**	2.22E-03†	5.79E-03†	3.73E-03†	2.49E-03†
F_8	**2.09E+01**	**2.09E+01**$^≈$	**2.09E+01**$^≈$	**2.09E+01**$^≈$	**2.09E+01**$^≈$
F_9	2.84E+01	1.30E+01‡	2.74E+01‡	1.28E+01‡	**6.81E+00**‡
F_{10}	1.69E+02	1.21E+02‡	1.31E+02‡	1.18E+02‡	**1.03E+02**‡
‡/≈ /†	-	17/1/2	17/1/2	16/2/2	16/1/3
p-value	-	8.97E-03	1.00E-02	1.00E-02	1.87E-02
Avg. rank	4.35	2.78	2.88	2.58	**2.43**

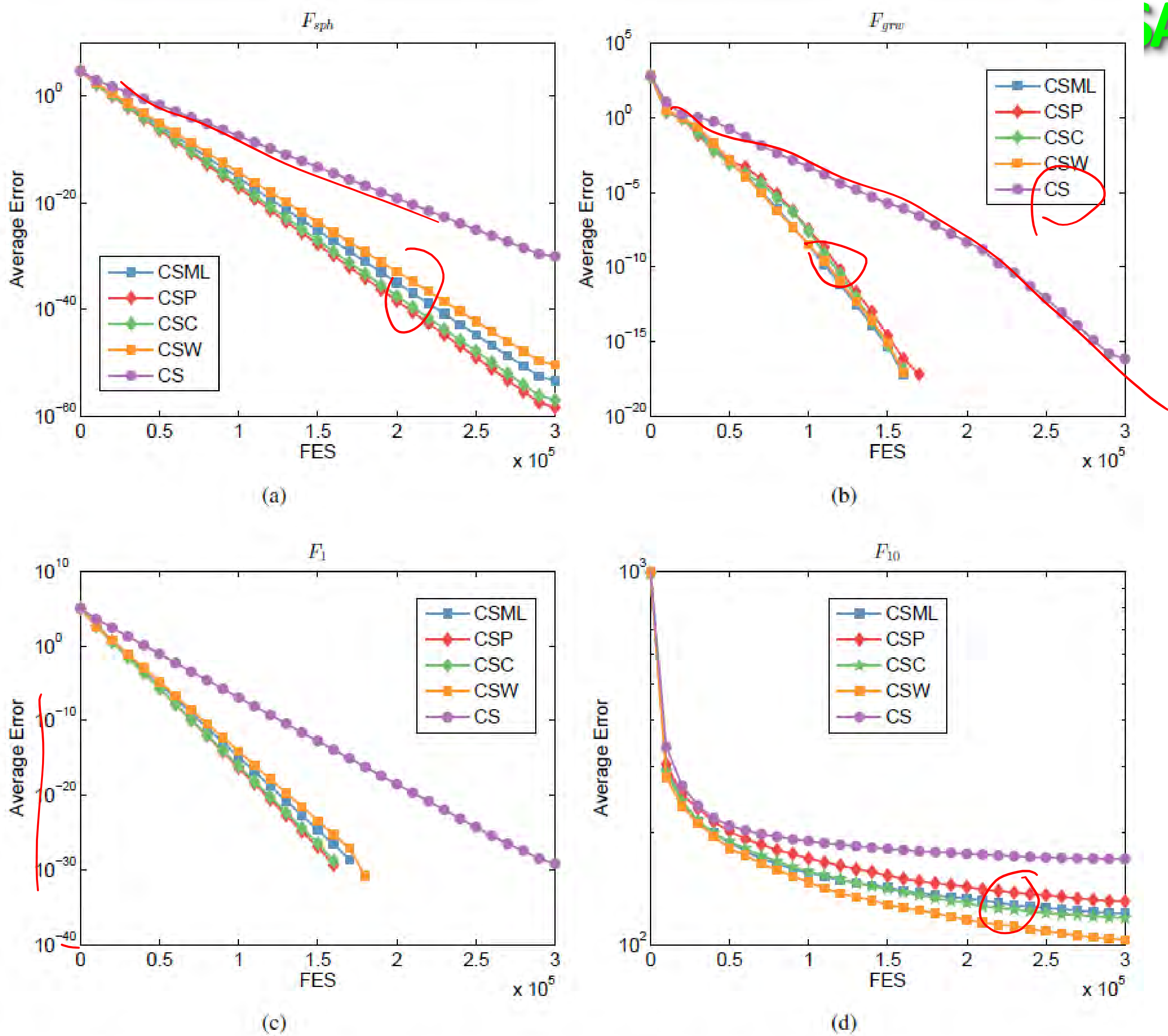


Figure 2. Convergence curves of CS and different improved CS algorithms for selected functions at $D = 30$.

Why BD and ML must meet FC?

Connection to FC via PDF

- "*Fractional Calculus and Stable Probability Distributions*" (1998) by Rudolf Gorenflo , Francesco Mainardi http://arxiv.org/pdf/0704.0320.pdf

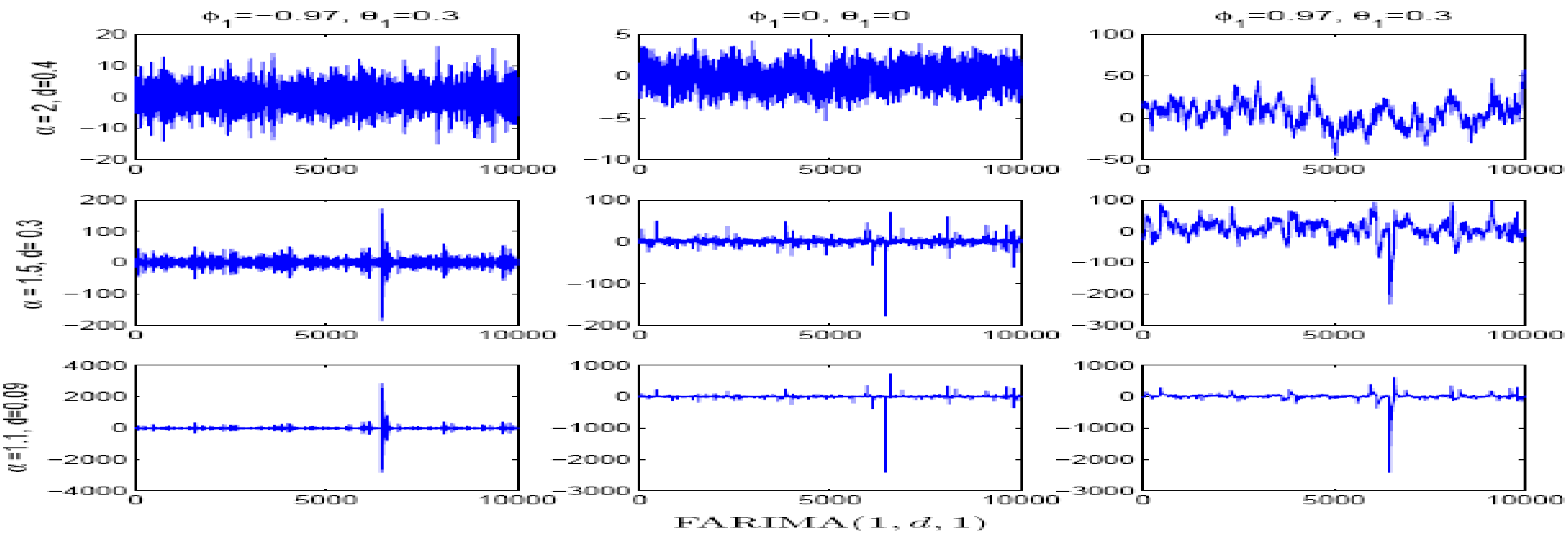
$$\frac{\partial u}{\partial t} = D(\alpha)\,\frac{\partial^\alpha u}{\partial |x|^\alpha}\,, \quad -\infty < x < +\infty\,, \quad t \geq 0\,,$$

$$\text{with} \quad u(x,0) = \delta(x) \quad 0 < \alpha \leq 2$$

$$\frac{\partial^{2\beta} u}{\partial t^{2\beta}} = D(\beta)\,\frac{\partial^2 u}{\partial x^2}\,, \quad x \geq 0\,, \quad t \geq 0\,,$$

$$\text{with} \quad u(0,t) = \delta(t) \quad 0 < \beta < 1$$

Optimal randomness means fractional calculus!



More optimal than the “optimized PSO” CS (Cuckoo Search)

Two broad categories

- Derivative-free (fractional calculus helps)
  - Direct Search, NM, PSO
- Gradient-based
  - GD and its variants

GD and SGD

$$x_{t+1} = x_t - \eta_t g_t, \text{ where } \mathbb{E}[g_t] = \nabla f(x_t)$$

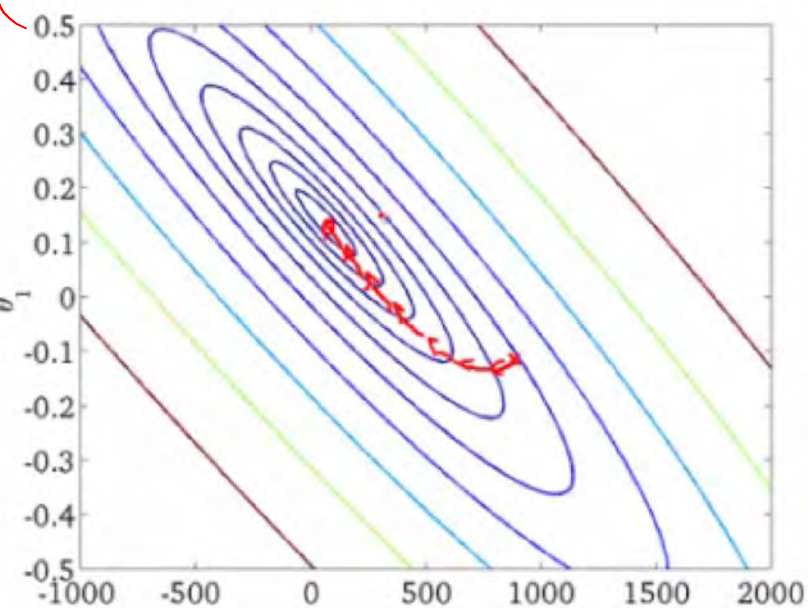
$$g_t = \nabla f(x_t) + \xi_t$$

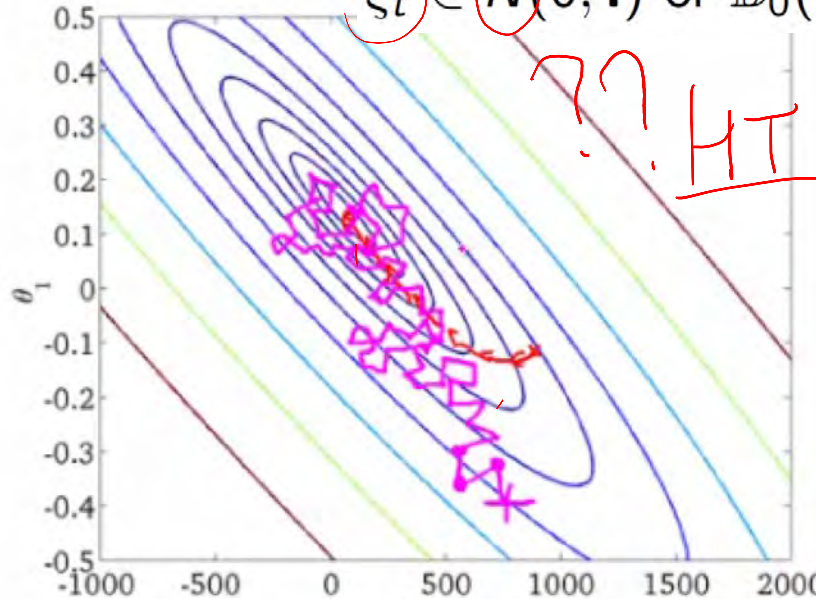
$$\xi_t \in N(0, \mathbf{I}) \text{ or } \mathbb{B}_0(r)$$

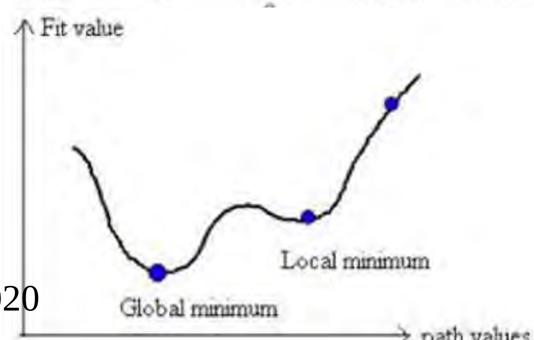
??. HT

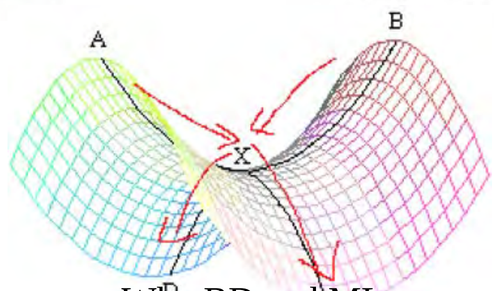
Task: $\min_x f(x)$

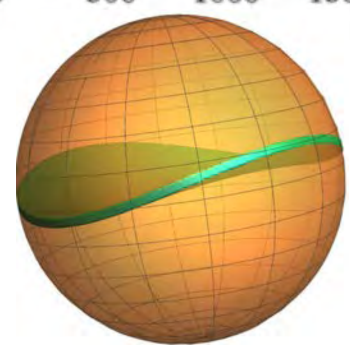
GD does iterative updates $x_{t+1} = x_t - \eta_t \nabla f(x_t)$

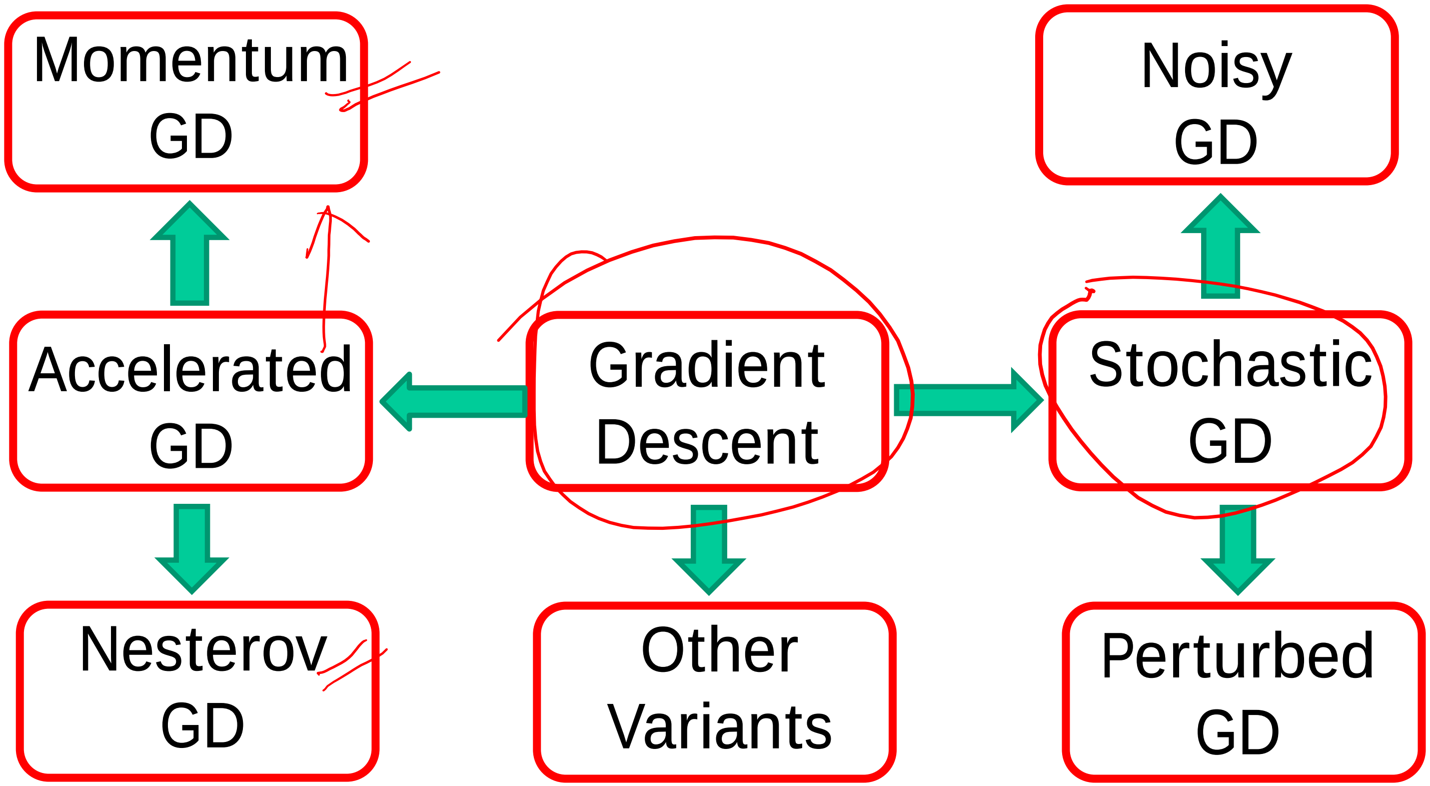




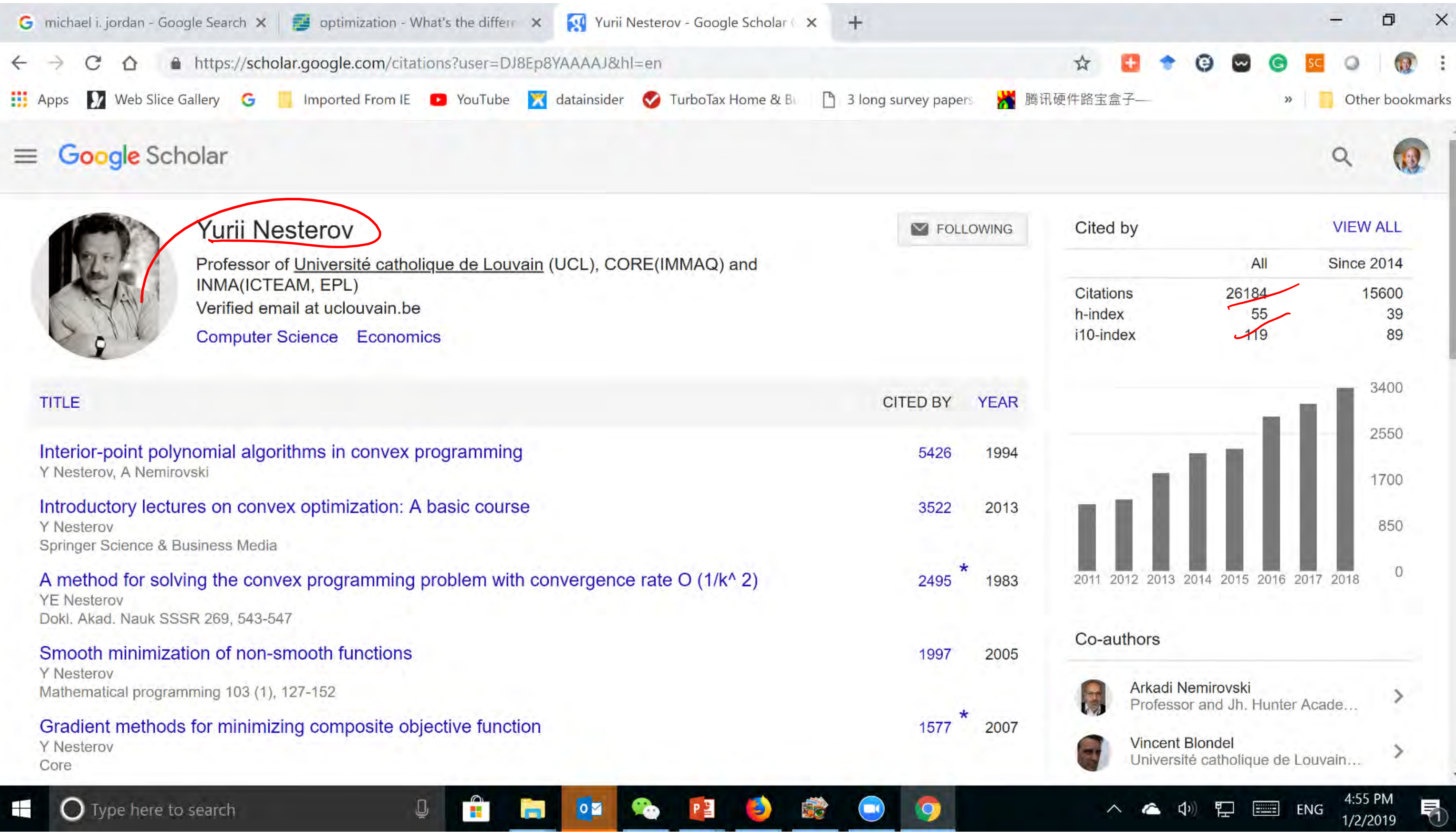







Momentum GD
Noisy GD
Accelerated GD
Gradient Descent
Stochastic GD
Nesterov GD
Other Variants
Perturbed GD

NAGD


michael i. jordan - Google Search
optimization - What's the differ
Yurii Nesterov - Google Scholar
https://scholar.google.com/citations?user=DJ8Ep8YAAAAJ&hl=en
Apps
Web Slice Gallery
Imported From IE
YouTube
datainsider
TurboTax Home & B
3 long survey papers
腾讯硬件路宝盒子—
Other bookmarks
Google Scholar
Yurii Nesterov
FOLLOWING
Professor of Université catholique de Louvain (UCL), CORE(IMMAQ) and INMA(ICTEAM, EPL)
Verified email at uclouvain.be
Computer Science
Economics
TITLE
CITED BY
YEAR
Interior-point polynomial algorithms in convex programming
Y Nesterov, A Nemirovski
5426
1994
Introductory lectures on convex optimization: A basic course
Y Nesterov
Springer Science & Business Media
3522
2013
A method for solving the convex programming problem with convergence rate O (1/k^ 2)
YE Nesterov
Dokl. Akad. Nauk SSSR 269, 543-547
2495 *
1983
Smooth minimization of non-smooth functions
Y Nesterov
Mathematical programming 103 (1), 127-152
1997
2005
Gradient methods for minimizing composite objective function
Y Nesterov
Core
1577 *
2007
Cited by
VIEW ALL
All
Since 2014
Citations
26184
15600
h-index
55
39
i10-index
119
89
3400
2550
1700
850
0
2011
2012
2013
2014
2015
2016
2017
2018
Co-authors
Arkadi Nemirovski
Professor and Jh. Hunter Acade…
Vincent Blondel
Université catholique de Louvain…
Type here to search
ENG
4:55 PM
1/2/2019

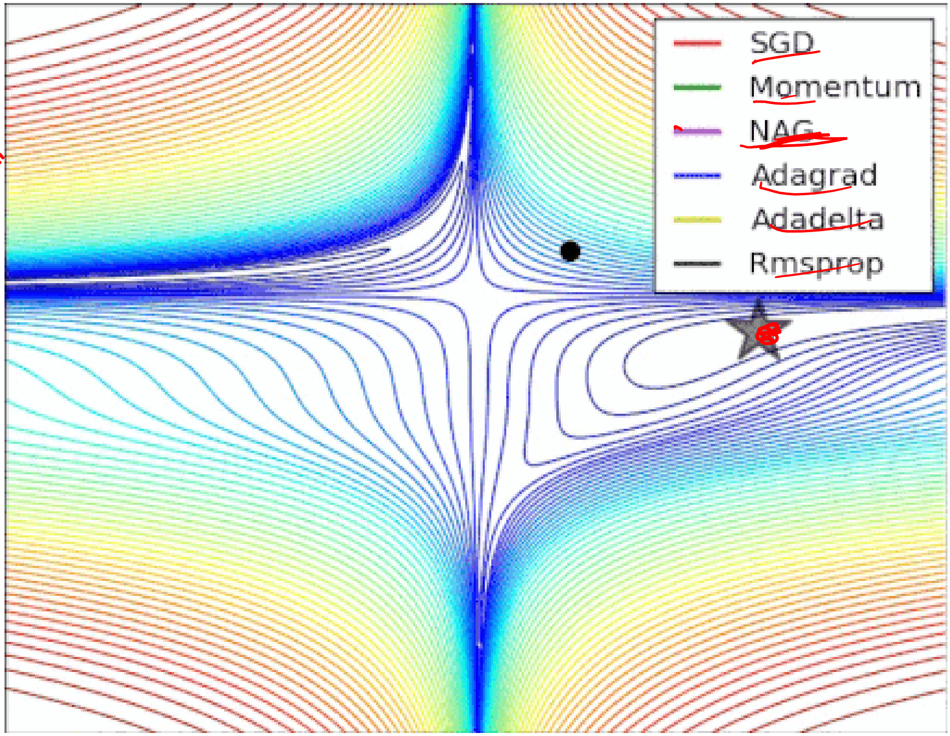
NAGD: Nesterov Accelerated GD

$$\begin{cases} y_{k+1} = a y_k - \mu \nabla f(x_k) \\ x_{k+1} = x_k + y_{k+1} + b y_k \end{cases}$$

x_k x_{k-1}

- Set $b = \frac{-a}{1+a}$ and one can derive the NAGD.

- Set b=0 and one can derive GDM.

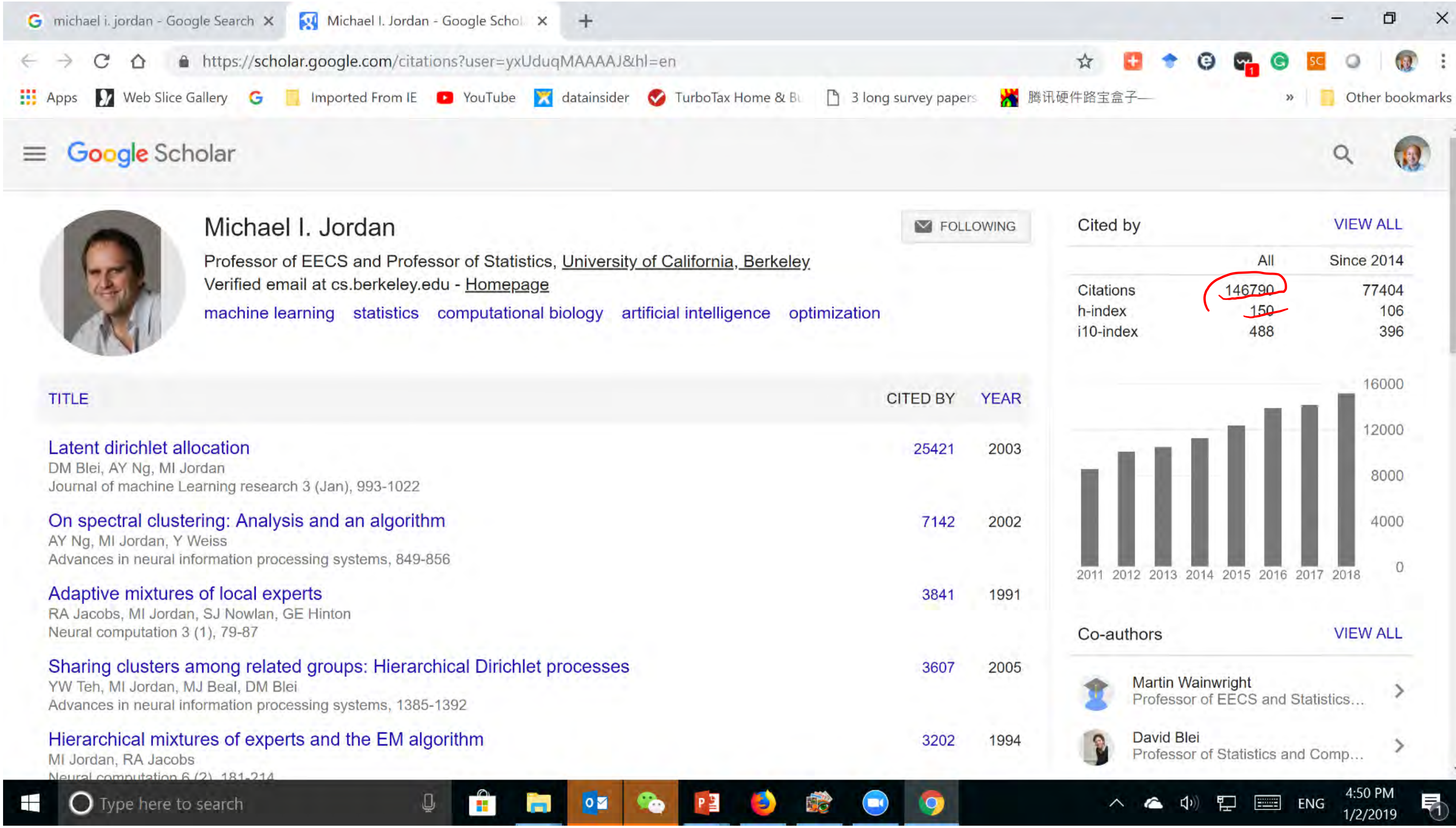


https://stats.stackexchange.com/questions/179915/whats-the-difference-between-momentum-based-gradient-descent-and-nesterovs-acc

Michael I. Jordan:

Is there an optimal way to optimize?

- ICM2018 1 hour report” **Dynamical, symplectic and stochastic perspectives on optimization”** https://youtu.be/wXNWVhE2Dl4
- SIAM Mathematics of Data Science (MDS20) Distinguished Lecture Series. **Machine Learning: Dynamical, Statistical, and Economic Perspectives.** https://youtu.be/Z4u3EA2k8vg
- IAS 6/11/20. **On Langevin Dynamics in Machine Learning - Michael I. Jordan.** https://youtu.be/QTnjqdxG99c
- ACM 4/1/2020. **"The Decision-Making Side of Machine Learning"** https://youtu.be/K3lFvai4viY


michael i. jordan - Google Search
Michael I. Jordan - Google Schol
https://scholar.google.com/citations?user=yxUduqMAAAAJ&hl=en
Google Scholar
Michael I. Jordan
FOLLOWING
Professor of EECS and Professor of Statistics, University of California, Berkeley
Verified email at cs.berkeley.edu - Homepage
machine learning
statistics
computational biology
artificial intelligence
optimization
TITLE
CITED BY
YEAR
Latent dirichlet allocation
DM Blei, AY Ng, MI Jordan
Journal of machine Learning research 3 (Jan), 993-1022
25421
2003
On spectral clustering: Analysis and an algorithm
AY Ng, MI Jordan, Y Weiss
Advances in neural information processing systems, 849-856
7142
2002
Adaptive mixtures of local experts
RA Jacobs, MI Jordan, SJ Nowlan, GE Hinton
Neural computation 3 (1), 79-87
3841
1991
Sharing clusters among related groups: Hierarchical Dirichlet processes
YW Teh, MI Jordan, MJ Beal, DM Blei
Advances in neural information processing systems, 1385-1392
3607
2005
Hierarchical mixtures of experts and the EM algorithm
MI Jordan, RA Jacobs
3202
1994
Cited by
VIEW ALL
All
Since 2014
Citations
146790
77404
h-index
150
106
i10-index
488
396
16000
12000
8000
4000
0
2011
2012
2013
2014
2015
2016
2017
2018
Co-authors
VIEW ALL
Martin Wainwright
Professor of EECS and Statistics...
David Blei
Professor of Statistics and Comp...
Type here to search
ENG
4:50 PM
1/2/2019

Review of Michael Jordan's work:

~~*Nesterov accelerated GD*~~ *(NAGD)* can be formulated as

$$\begin{cases} x_k = y_{k-1} - \mu \nabla f\left(y_{k-1}\right) \\ y_k = x_k + \frac{k-1}{k+2}\left(x_k - x_{k-1}\right) \end{cases}$$

$y_k(y_{k-1}, y_{k-2})$

Set $t = k\sqrt{\mu}$ and one can derive its corresponding differential equation

$$\ddot{X} + \frac{3}{t}\dot{X} + \nabla f\left(X\right) = 0$$

MSD

- The main idea of Michael Jordan's work is to analyze the iteration algorithm in *the continuous-time domain*.
- For differential equation, one can use *Lyapunov method or variational method* to analyze its properties.

UCMERCED MESA LAB

Review of Michael Jordan's work

Take Lyapunov functional as

$$V(t) = t^2 \left(f(X(t)) - f^*\right) + 2\left\|X + \frac{t}{2}\dot{X} - x^*\right\|$$

whose time derivative is

$$\dot{V}(t) = 2t\left(f(X(t)) - f^*\right) - 2t\left\langle X - x^*, \nabla f(X)\right\rangle \le 0$$

due to the convexity of $f(X)$.

Then one has

$$f(X(t)) - f^* \le \frac{V(t)}{t^2} \le \frac{V(0)}{t^2} = \frac{2\|x_0 - x^*\|^2}{t^2}$$

which indicates a $O(\frac{1}{t^2})$ convergence rate.

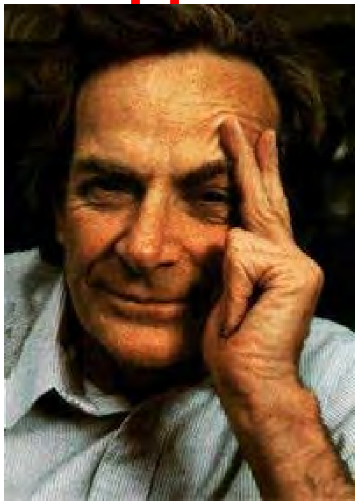
One can also use the variational method to derive the master differential equation for an optimization method.

Variational principle

- **Maupertuis: Least Action Principle**
- **Hamilton: Hamilton's Variational Principle**
- **Feynman: Quantum-Mechanical Path Integral Approach**

Pierre-Louis Moreau de Maupertuis (1698 – 1759)

Sir William Rowan Hamilton (1805 – 1865)



Richard Phillips Feynman (1918 – 1988)

UCMERCED

MESA LAB

Review of Michael Jordan's work

Nesterov accelerated GD (NAGD) can be formulated as

$$\begin{cases} x_k = y_{k-1} - \mu \nabla f(y_{k-1}) \\ y_k = x_k + \frac{k-1}{k+2}(x_k - x_{k-1}) \end{cases}$$

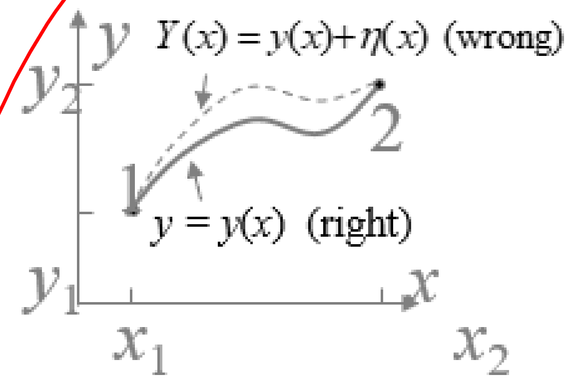
Set $t = k\sqrt{\mu}$ and one can derive its corresponding differential equation

$$\ddot{X} + \frac{3}{t}\dot{X} + \nabla f(X) = 0$$

Define the Lagrangian

$$\boxed{\mathcal{L} = \mathcal{L}[q_1, q_2, \cdots, q_N, \dot{q}_1, \dot{q}_2, \cdots, \dot{q}_N, t].}$$

where $(q_1, q_2, \cdots, q_N)$ are generalized coordinates.



We will then make the **action integral** stationary

$$\boxed{S = \int \mathcal{L}[q_1, q_2, \cdots, q_N, \dot{q}_1, \dot{q}_2, \cdots, \dot{q}_N, t]\, dt,}$$

which requires that $\mathcal{L}$ satisfy

$$\boxed{\frac{\partial \mathcal{L}}{\partial q_1} = \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{q}_1}; \frac{\partial \mathcal{L}}{\partial q_2} = \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{q}_2}; \cdots; \frac{\partial \mathcal{L}}{\partial q_N} = \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{q}_N}.}$$

E-L.

Review of Michael Jordan's work

Consider the convex function $f(x)$ and define the *Bregman divergence*

$$D_h(y, x) = h(y) - h(x) - \langle \nabla h(x), y - x \rangle$$

where, $h(x)$ is also a convex function.

Define the *second Bregman Lagrangian* as

$$L(x, v, t) = e^{\alpha_t + \beta_t + \gamma_t} \left(\mu D_h \left(x, x + e^{-\alpha_t} v \right) - f(x) \right)$$

with ideal scaling conditions

$$\dot{\gamma}_t = e^{\alpha_t}$$
$$\dot{\beta}_t \le e^{\alpha_t}$$

UCMERCED

MESA LAB

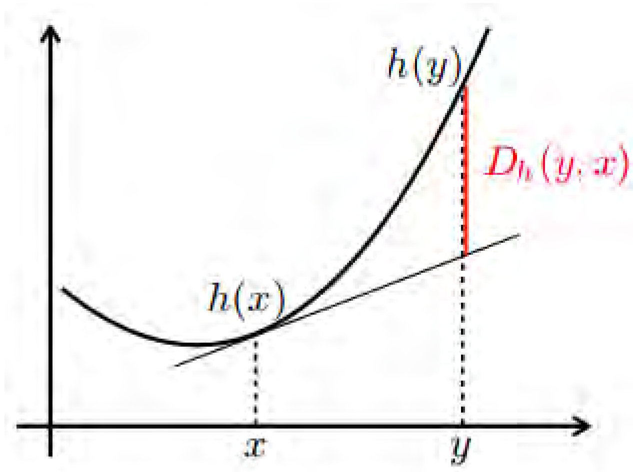
Review of Michael Jordan's work

The *second Bregman Lagrangian*

$$L(x, v, t) = e^{\alpha_t + \beta_t + \gamma_t} \left(\mu D_h \left(x, x + e^{-\alpha_t} v \right) - f(x) \right)$$

- $D_h(x, x + e^{-\alpha_t} v)$ can be viewed as the kinetic energy and $-f(x)$ is the potential energy.

- $\alpha_t, \beta_t, \gamma_t$ are arbitrary smooth functions to help analyzing the convergence rate.



Review of Michael Jordan's work

Define a functional on curves via integration of the Lagrangian:

$$J = \int_R L(x, v, t)dt$$

and the Euler-Lagrange equation is

$$\frac{\partial}{\partial x} L(x, v, t) = \frac{d}{dt}\frac{\partial}{\partial v} L(x, v, t)$$

The Euler-Lagrange equation for the second Bregman Lagrangian reduces to

$$\frac{d}{dt}\nabla h\left(X_t + e^{\alpha_t}\dot{X}_t\right) = \dot{\beta}_t \nabla h(X_t) - \dot{\beta}_t \nabla h\left(X_t + e^{\alpha_t}\dot{X}_t\right) - \frac{e^{\alpha_t}}{\mu}\nabla f(X_t)$$

Review of Michael Jordan's work

Define $h = \frac{1}{2}\|x\|^2$ and $\beta_t = \gamma t$, and one has

$$\ddot{X}_t + 2\gamma \dot{X}_t + \frac{\gamma^2}{\mu}\nabla f\left(X_t\right) = 0$$

NAGD

which in the same form as the master differential equation of NAGD.

Assume f is μ-uniformly convex with respect to h (strictly convex) and the scaling condtion $\dot{\beta}_t = e^{\alpha_t}$ holds. One can conclude that

$$V(t) = e^{\beta_t}\left(\mu D_h\left(x, X_t + e^{-\alpha_t}\dot{X}_t\right) + f\left(X_t\right) - f\left(x\right)\right)$$

is a Lyapunov functional ($\dot{V} \leq 0$), which indicates a $O(e^{-\beta_t})$ convergence rate.

UCMERCED MESA LAB

Review of Michael Jordan's work

Discretize the previous Lagrangian system, and we can get following two discrete algorithms:

$$\begin{cases} x_{k+1} = \tau_k z_k + (1 - \tau_k)\, y_k \\ \nabla h\,(z_{k+1}) = \nabla h\,(z_k) - \alpha_k \nabla f\,(x_{k+1}) \\ y_{k+1} = \Theta\,(x) \end{cases}$$

and

$$\begin{cases} x_{k+1} = \tau_k z_k + (1 - \tau_k)\, y_k \\ y_{k+1} = \Theta\,(x) \\ \nabla h\,(z_{k+1}) = \nabla h\,(z_k) - \alpha_k \nabla f\,(y_{k+1}) \end{cases}$$

where, $\tau_k = \frac{A_{k+1} - A_k}{A_k} := \frac{\alpha_k}{A_k}$, and $\frac{d}{dt} e^{\beta_t} = (A_{k+1} - A_k)/\delta$ can be well approximated by a discrete-time sequence A_i. Moreover, Θ is an arbitrary map whose domain is the previous state $x = (x_{k+1}, z_{k+1}/z_k, y_k)$

[1] Wilson, A. C., Recht, B., & Jordan, M. I. (2016). A Lyapunov analysis of momentum methods in optimization. *arXiv preprint arXiv:1611.02635*.

Review of Michael Jordan's work

- One can transform an iterative (optimization) algorithm to its continuous-time limit case, which can simplify the analyses (*Lyapunov methods*).
- One can directly design a differential equation (of motion) and then discretize it to derive an iterative algorithm (*Variational method*).
- The key is to *find a suitable Lyapunov functional* to analyze the stability and convergent rate.

Exciting **new** Fact: Optimization algorithms can be systematically synthesized using Lagrangian mechanics (E-L) EOM

Obviously, why not fractional order?

Define $h = \frac{1}{2}\|x\|^2$ and $\beta_t = \gamma t$, and one has

$$\ddot{X}_t + 2\gamma \dot{X}_t + \frac{\gamma^2}{\mu} \nabla f(X_t) = 0$$

which in the same form as the master differential equation of NAGD.

$$\dot{x}(t) \Rightarrow x^{(\alpha)}(t)$$

Could to be "More Optimal"? (Better than the best)

Go "fractional" to be more optimal !

Fractional calculus of variation (FOCV), FO EL equation

GD and SGD

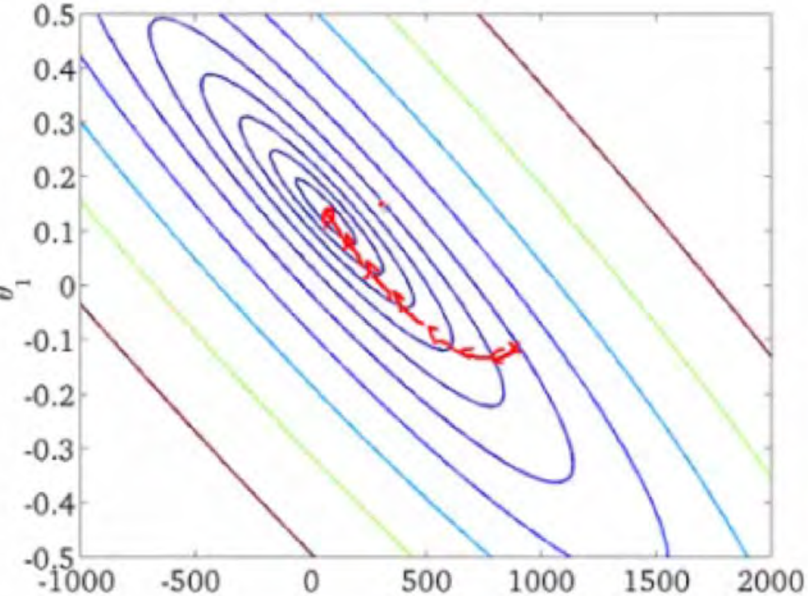
$$x_{t+1} = x_t - \eta_t g_t, \text{ where } \mathbb{E}[g_t] = \nabla f(x_t)$$

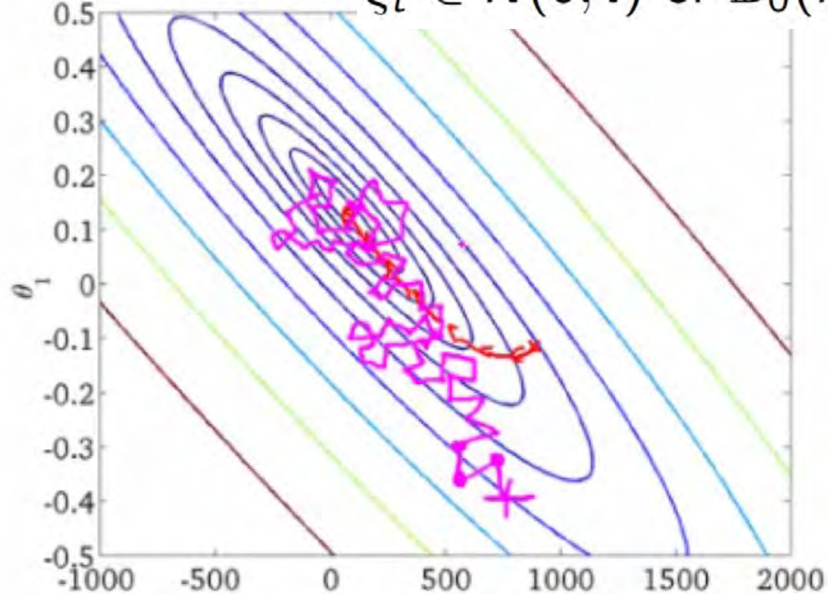
Task: $\min_x f(x)$

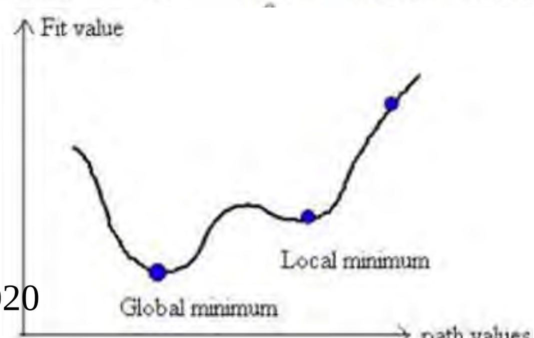
$$g_t = \nabla f(x_t) + \xi_t$$

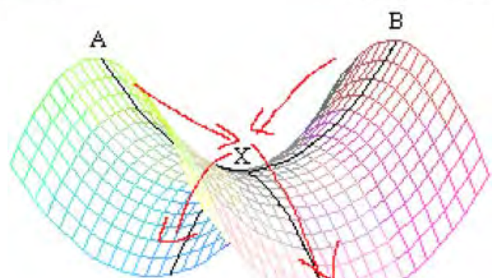
GD does iterative updates $x_{t+1} = x_t - \eta_t \nabla f(x_t)$

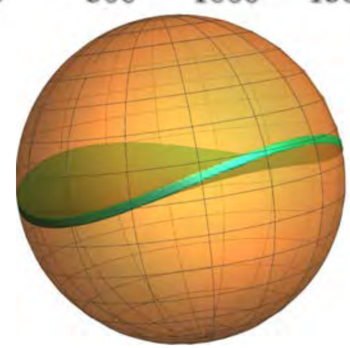
$$\xi_t \in N(0, \mathbf{I}) \text{ or } \mathbb{B}_0(r)$$











Why BD and ML must meet FC?

SGD

- Optimal randomness using fractional order noises can offer better than the best performance, similarly shown in

J Wei, YQ Chen, Y Yu, Y Chen (2019). *Optimal Randomness in Swarm-Based Search*. **Mathematics** 7 (9), 828 [PDF]

https://www.mathworks.com/matlabcentral/fileexchange/71758-optimal-randomness-in-swarm-based-search

UC MERCED

MESA LAB

What control community can/should offer to CS/ML community?

- ***"The Three Musketeers"***
  - Internal model principle (IMP)
  - Nu-Gap metric
  - Model discrimination

https://www.ifac2020.org/program/workshops/machine-learning-meets-model-based-control.html Eric Kerrigan, Imperial College London, UK

Analyses & Design with System Theory

- In [1], the authors transfer the convergence problem of numerical algorithms into a stability problem of a discrete-time system;
- In [2], the authors explained that the commonly used SGD-Momentum algorithm in Machine Learning is a PI controller and designed a PID algorithm.
- Motivated by [2] and different from Michael Jordan's work, we will directly design and analyze the algorithms *in S or Z domain*.

[1] Kashima, K., & Yamamoto, Y. (2007). System theory for numerical analysis. *Automatica*, 43(7), 1156-1164.

[2] An, W., Wang, H., Sun, Q., Xu, J., Dai, Q., & Zhang, L. (2018). A PID Controller Approach for Stochastic Optimization of Deep Networks. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 8522-8531).

Analyses & Design with System Theory

Gradient Descent (GD) is a first-order algorithm:

$$x_{k+1} = x_k - \mu \nabla f(x_k)$$

where $\mu > 0$ is the step size.

Using the Z-transform, we have that

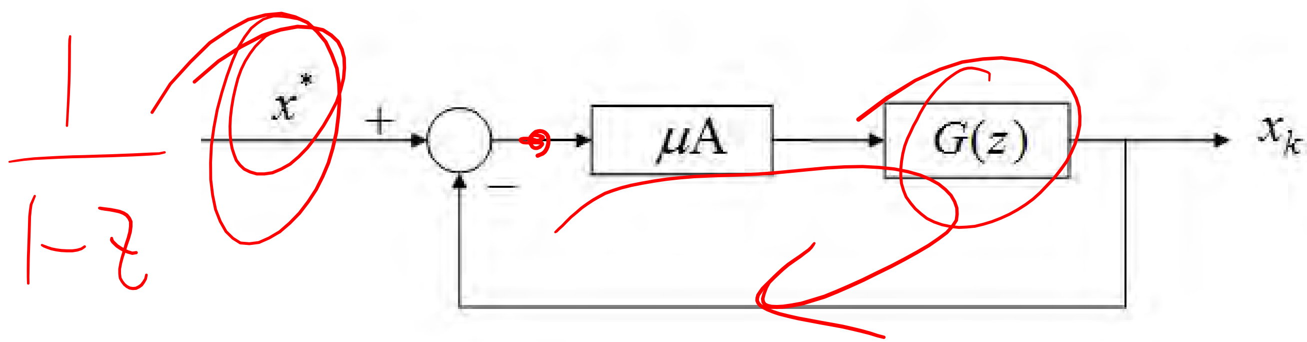
$$X(z) = \frac{\mu}{z-1}\left[-\nabla f(x_k)\right]_z$$

UC MERCED

MESA LAB

Analyses & Design with System Theory

Approximate the gradient around the extreme point x^*, and one has $\nabla f(x_k) \approx A(x_k - x^*)$ with $A = \nabla^2 f(x^*)$.



For GD, we have that $G(z) = \frac{1}{z-1}$, which is an integrator.

- Integrator in the forward loop is to eliminate the tracking error for a constant reference signal (**Internal Model Principle**).

UCMERCED MESA LAB

Analyses & Design with System Theory

GD-Momentum (GDM) is then designed to accelerate the conventioanl GD, which is polpularly used in Machine Learning.

$$\begin{cases} y_{k+1} = \alpha y_k - \mu \nabla f(x_k) \\ x_{k+1} = x_k + y_{k+1} \end{cases}$$

where y_k is the accumulation of history gradient and $\alpha \in (0,1)$ is the rate of moving average decay.

Do Z-transform for the update rule and derive

$$\begin{cases} zY(z) = \alpha Y(z) - \mu [\nabla f(x_k)]_z \\ zX(z) = X(z) + zY(z) \end{cases}$$

UCMERCED MESA LAB

Analyses & Design with System Theory

Then one has

$$X(z) = \frac{\mu z}{(z-1)(z-\alpha)}\left[-\nabla f(x_k)\right]_z$$

For GD-Momentum, we have that $G(z) = \frac{z}{(z-1)(z-\alpha)}$, with an integrator in the forward loop.

- GD-Momentum is a second-order (*G(z)*) algorithm with an additional pole α and zero *0*.
- The "second-order" means the order of *G(z)*, which is different from the algorithm using the Hessian matrix information.

Analyses & Design with System Theory

NAGD can be simplified as

$$\begin{cases} y_{k+1} = x_k - \mu \nabla f(x_k) \\ x_{k+1} = (1-\lambda) y_{k+1} + \lambda y_k \end{cases}$$

where μ is the step size and λ is a weighting coffecient.
Do Z-transform for the update rule and derive

$$\begin{cases} zY(z) = X(z) - \mu [\nabla f(x_k)]_z \\ zX(z) = (1-\lambda) zY(z) + \lambda Y(z) \end{cases}$$

UCMERCED

MESA LAB

Analyses & Design with System Theory

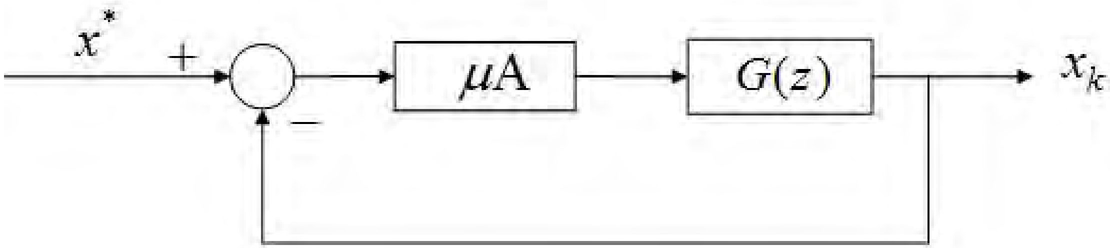
Then one has

$$X(z) = \frac{-(1-\lambda)z-\lambda}{(z-1)(z+\lambda)}\mu[\nabla f(x_k)]_z$$
$$= \frac{z+\frac{\lambda}{1-\lambda}}{(z-1)(z+\lambda)}\mu(1-\lambda)[-\nabla f(x_k)]_z$$

For NAGD, we have that $G(z) = \frac{z+\frac{\lambda}{1-\lambda}}{(z-1)(z+\lambda)}$, with an integrator in the forward loop.

- NAGD is a second-order algorithm with an additional pole $-\lambda$ and a zero $\frac{-\lambda}{1-\lambda}$.

Analyses & Design with System Theory

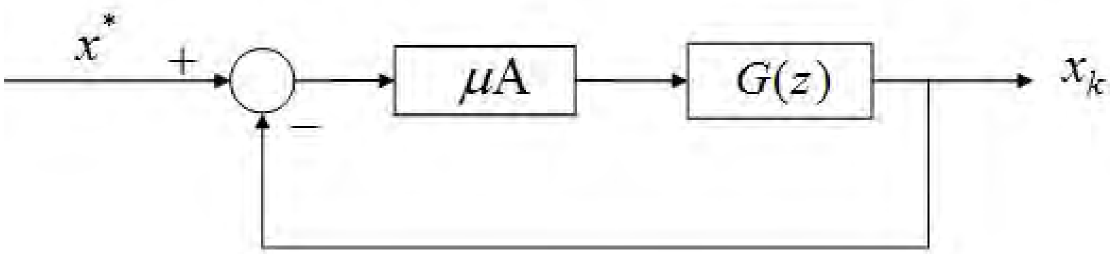


- A necessary condition for the stability of the algorithm is that all the poles of closed-loop system are within a unit disc.

- If the Lipschitz continuous gradient constant *L* is given, one can replace *A* with *L* and then the condition is sufficient.

UCMERCED MESA LAB

Analyses & Design with System Theory



- For each $G(z)$, it has a corresponding iterative algorithm.
- $G(z)$ can be third or higher order systems.
- **$G(z)$ can also be a fractional order system.**

Analyses & Design with System Theory

General second-order algorithm design

Consider a general second-order discrete system

$$G(z) = \frac{z + b}{(z-1)(z-a)}$$

whose corresponding iterative algorithm is

$$\begin{cases} y_{k+1} = a y_k - \mu \nabla f(x_k) \\ x_{k+1} = x_k + y_{k+1} + b y_k \end{cases}$$

- Set $b = \frac{-a}{1+a}$ and one can derive the NAGD.
- Set b=0 and one can derive GDM.

Analyses & Design with System Theory

General second-order algorithm design

- The iterative algorithm can be viewed as the ***state-space realization*** of the corresponding system. Thus, it has many different realizations (all are equivalent).

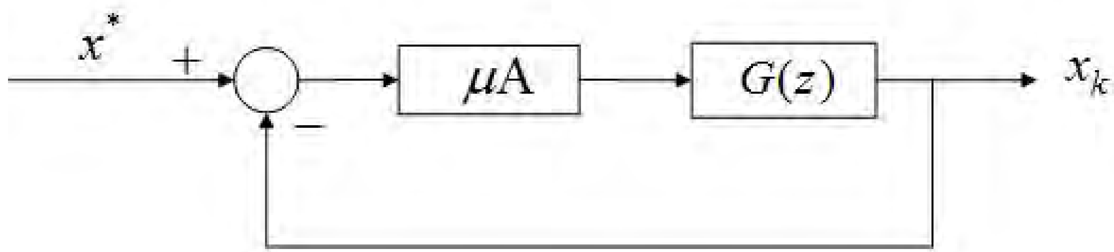
$$\begin{cases} Y(z) = \frac{1}{z-a}\left[-\mu\nabla f(x_k)\right]_z \\ X(z) = \frac{z-b}{z-1}Y(z) \end{cases} \sim \begin{cases} y_{k+1} = a y_k - \mu\nabla f(x_k) \\ x_{k+1} = x_k + y_{k+1} + b y_k \end{cases}$$

$$\begin{cases} Y(z) = \frac{1}{z-1}\left[-\mu\nabla f(x_k)\right]_z \\ X(z) = \frac{z-b}{z-a}Y(z) \end{cases} \sim \begin{cases} y_{k+1} = y_k - \mu\nabla f(x_k) \\ x_{k+1} = a x_k + y_{k+1} + b y_k \end{cases}$$

Analyses & Design with System Theory

General second-order algorithm design

We have introduced two parameters a and b, but how to optimize them?



We can use *Integral Square Error (ISE)* as the criterion to optimize the parameters. Since for different target function $f(x)$, its Lipschitz continuous gradient constant is different. Thus, define $\rho := \mu A$ as the loop forward gain.

Analyses & Design with System Theory

General second-order algorithm design

ρ	0.4	0.8	1.2	1.6	2.0	2.4
a	-0.6	-0.2	0.2	0.6	1	1.4
b	1.5	0.25	-0.1667	-0.3750	-0.5	-0.5833

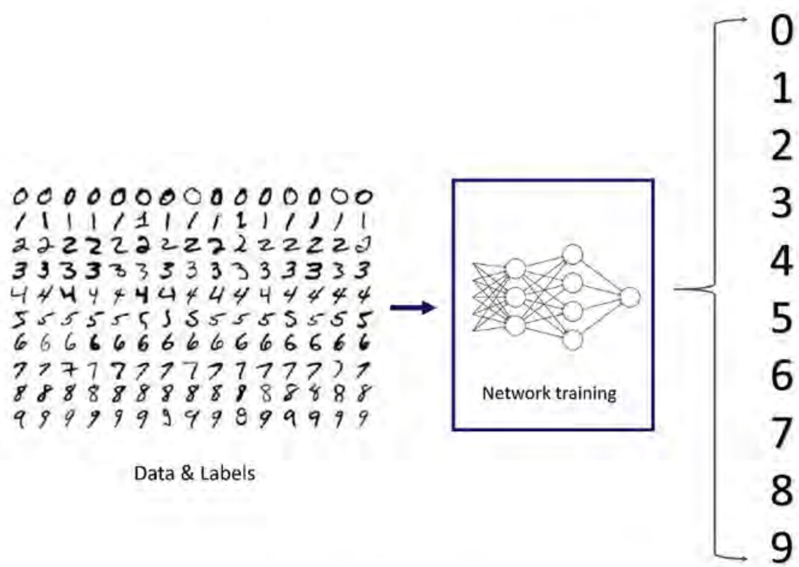
- It is found that the *optimal a and b satisfies* $b = \frac{-a}{1+a}$, which is the same design as NAGD.
- We have used other criteria such as IAE, ITASE to find other optimal parameters, but the results are *the same as ISE*.
- Different from NAGD, we derive the parameters by optimization rather than mathematically design, which can be extended to more general cases.

MESA LAB

Analyses & Design with System Theory

How does the zero influence the convergence performance?

Image Classification in 10 Minutes with MNIST Dataset



MNIST Dataset and Number Classification [1]

The MNIST database is a collection of hand-written digits, which contains 60,000 training images and 10,000 testing images. It is widely used as a Benchmark for Machine Learning algorithms.

- ❑ In the following, x-axis is always the epoch number.

Analyses & Design with System Theory

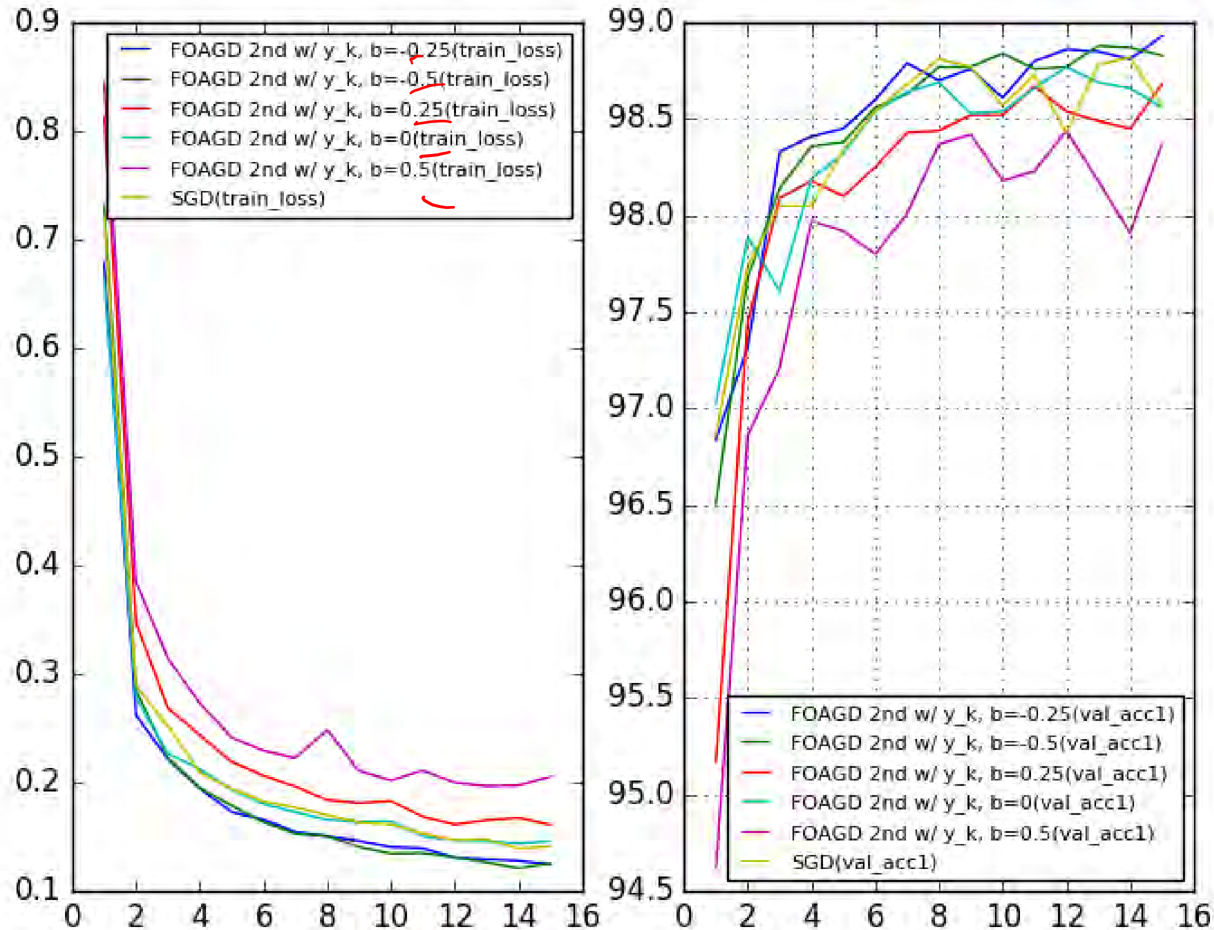


Figure: Training loss (Left), Test accuracy (Right)

Analyses & Design with System Theory

How does the zero influence the convergence performance?

- One can find that both $b = -0.25$ and $b = -0.5$ cases perform better than the SGD-Momentum. For $b = 0.25$ and $b = 0.5$, they perform worse.

- One can find the additional zero can improve the performance if we carefully adjust it.

- Both our method and Nesterov method give an optimal/good choice of the zero which is closely related to the pole ($b = \frac{-a}{1+a}$).

Analyses & Design with System Theory

General third-order algorithm design

Consider a general second-order discrete system

$$G(z) = \frac{z^2 + cz + d}{(z-1)(z^2 + az + b)}$$

- Set $b = d = 0$, it will reduce to the second-order algorithm.
- Compared with the second-order case, poles can now be complex number.
- More generally, a higher order system can contain *more internal models*.

Analyses & Design with System Theory

General third-order algorithm design

If all the poles are real, then one has that

$$G(z) = \frac{1}{z-1}\left(\frac{z-c}{z-a}\right)\left(\frac{z-d}{z-b}\right)$$

whose corresponding iterative algorithm is

$$\begin{cases} y_{k+1} = y_k - \mu \nabla f(x_k) \\ z_{k+1} = a z_k + y_{k+1} - c y_k \\ x_{k+1} = b x_k + z_{k+1} - d z_k \end{cases}$$

Analyses & Design with System Theory

General third-order algorithm design (ISE)

ρ	0.4	0.8	1.2	1.6	2.0	2.4
a	0.6439	0.5247	-0.4097	-0.5955	-1.0364	-1.4629
b	0.0263	0.0649	0.0419	-0.0398	0.0364	0.0880
c	1.5439	0.5747	-0.3763	-0.3705	-0.5364	-0.6462
d	0.0658	0.0812	0.0350	-0.0408	0.0182	0.0367

ρ	0.4	0.8	1.2	1.6	2.0	2.4
Roots of numerator	-1.5000	-0.3250	0.2082	0.1624	1.0000	1.4000
	-0.0439	-0.2500	0.1681	0.6000	0.0364	0.0629
Roots of denominator	-0.6000	-0.3250	0.2128	0.1624	0.5000	0.5833
	-0.0439	-0.1997	0.1969	0.3750	0.0364	0.0629

Analyses & Design with System Theory

General third-order algorithm design (ISE)

- Since we use the ISE for tracking a step signal (it is quite simple), the optimal poles and zeros are the same as the second-order case *with a pole-zero cancellation*.

- In this optimization results, all the poles and zeros are real.

- Compared with the second-order case, the only difference is that we can have *complex poles*.

How to derive complex poles in the design? FC helps.

Analyses & Design with System Theory

General fractional-order algorithm design

Borrowing the idea from Micheal Jordan, we directly design a continuous time fractional order system

$$G(s) = \frac{1}{s(s^{\alpha} + \beta)}, \alpha \in (0, 2), \beta \in (0, 20]$$

at first. Then, find the optimal parameters using the ISE criterion.

ρ	0.3	0.5	0.7	0.9
α	1.8494	1.6899	1.5319	1.2284
β	20	20	20	20

Analyses & Design with System Theory

General fractional-order algorithm design

It is the continuous-time design. One can use the *Numerical Inverse Laplace Transform (NILT)* and Matlab command `stmcb()` to derive its discrete form.

$$\rho = 0.3 : G_3(z) = \frac{0.0010z^2+0.0054z-0.0032}{z^3-2.851z^2+2.767z-0.9163}, G_2(z) = \frac{0.04689z-0.04471}{z^2-1.958z+0.9582}$$
$$\rho = 0.5 : G_3(z) = \frac{0.0025z^2+0.0056z-0.0042}{z^3-2.774z^2+2.625z-0.8509}, G_2(z) = \frac{0.0426z-0.03876}{z^2-1.927z+0.9266}$$
$$\rho = 0.7 : G_3(z) = \frac{0.0061z^2+0.0037z-0.005}{z^3-2.652z^2+2.398z-0.7462}, G_2(z) = \frac{0.01962z-0.00637}{z^2-1.748z+0.7481}$$
$$\rho = 0.9 : G_3(z) = \frac{0.0187z^2-0.0082z-0.0044}{z^3-2.357z^2+1.836z-0.4788}, G_2(z) = \frac{0.01895z-0.0098}{z^2-1.434z+0.4345}$$

where $G_2(z)$ is the second order approximation and $G_3(z)$ is the third order approximation

https://www.mathworks.com/matlabcentral/fileexchange/39035-numerical-inverse-laplace-transform

Analyses & Design with System Theory

Poles analyses:

$$\begin{cases} \rho = 0.3 : 1, \cancel{0.9256 \pm 0.2444j} \\ \rho = 0.5 : 1, 0.8870 \pm 0.2534j \\ \rho = 0.7 : 1, \cancel{0.8260 \pm 0.2529j} \\ \rho = 0.9 : 1, 0.6786 \pm 0.1354j \end{cases}$$

Zeros analyses:

$$\begin{cases} \rho{=}0.3 : -5.9388, 0.5388 \\ \rho{=}0.5 : -2.8330, 0.5930 \\ \rho{=}0.7 : -1.2581, 0.6515 \\ \rho{=}0.9 : -0.3131, 0.7516 \end{cases}$$

- If we direct design the algorithm in the discrete domain, all the poles are real.
- The fractional order design contributes to the arise of complex poles.

Analyses & Design with System Theory

General fractional-order algorithm design

If we have complex poles, then one has that

$$G(z) = \frac{z+c}{z-1}\left(\frac{1}{z-a+jb} + \frac{1}{z-a-jb}\right)$$

whose corresponding iterative algorithm is

$$\begin{cases} y_{k+1} = a y_k - b z_k - \mu \nabla f(x_k) \\ z_{k+1} = a z_k + b y_k \\ x_{k+1} = x_k + y_{k+1} + c y_k \end{cases}$$

Analyses & Design with System Theory

General fractional-order algorithm design

We still use the MINST data set to compare the designed algorithm with the commonly used SGD algorithm and 50 epoches are used to train the Network. For the fractional order design, take the $\rho = 0.9$ case, where $a = 0.6786, b = 0.1354$, and different zero c are designed. When $c = 0$, it is similar to the second-order SGD, while when $c \neq 0$, it is similar to second-order NAGD. For the SGD case, we set $\alpha = 0.9$. When simulation, set learning rate $\mu = 0.1$.

UCMERCED

MESA LAB

Analyses & Design with System Theory

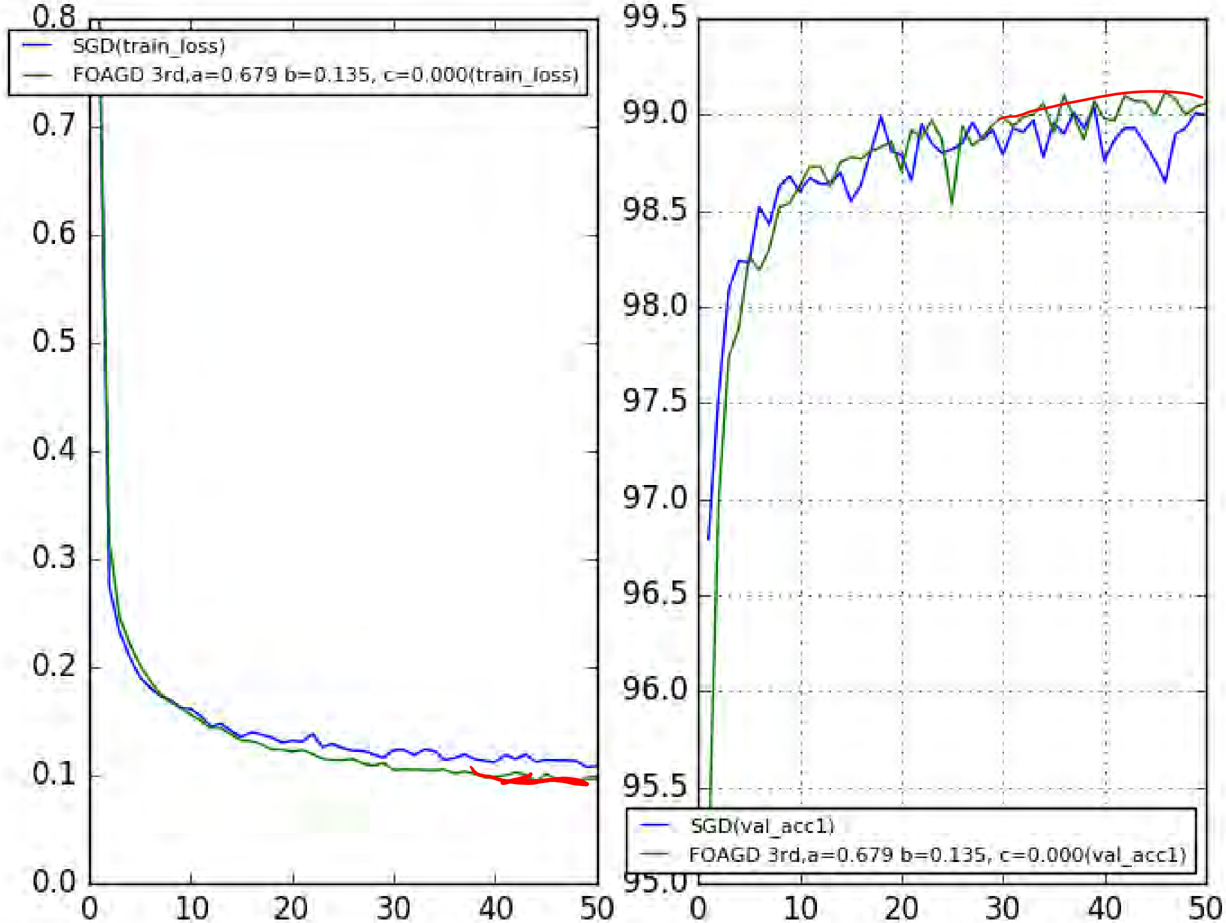


Figure: $c = 0$: Training loss (Left), Test accuracy (Right)

Analyses & Design with System Theory

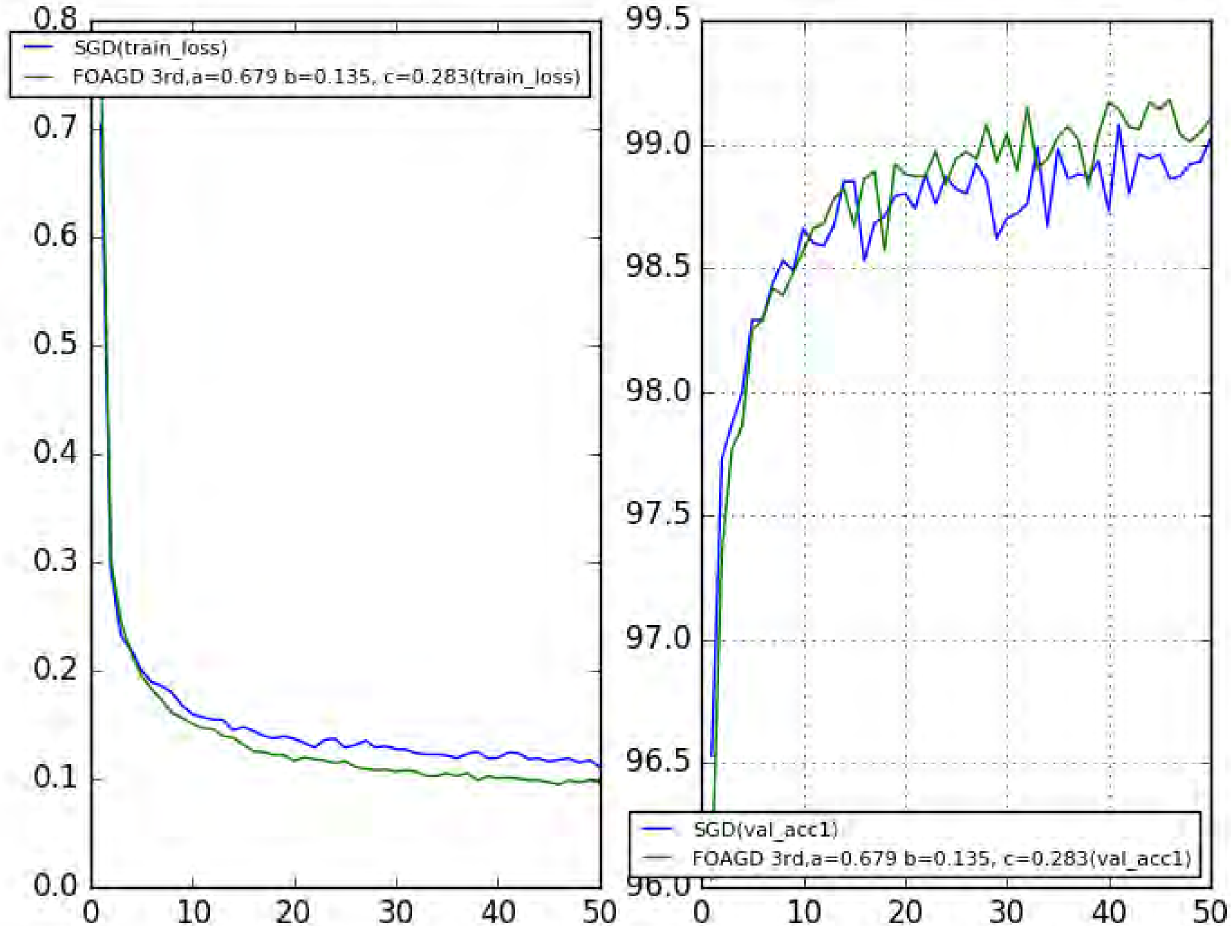


Figure: $c = 0.283$ Training loss (Left), Test accuracy (Right)

Analyses & Design with System Theory

- Both $c = 0$ and $c = 0.283$ cases perform better than commonly used SGD-Momentum.

- Generally, with c carefully designed, better performance can be achieved as the second-order case.

- The simulation results can only prove that fractional calculus (complex poles) can potentially improve the performance, which is closely related to learning rate.

Observation and take home messages

Michael I. Jordan: *Is there an optimal way to optimize*? Yes, via limiting dynamics analysis and discretization, via SGD with other randomness like Langevin motion.

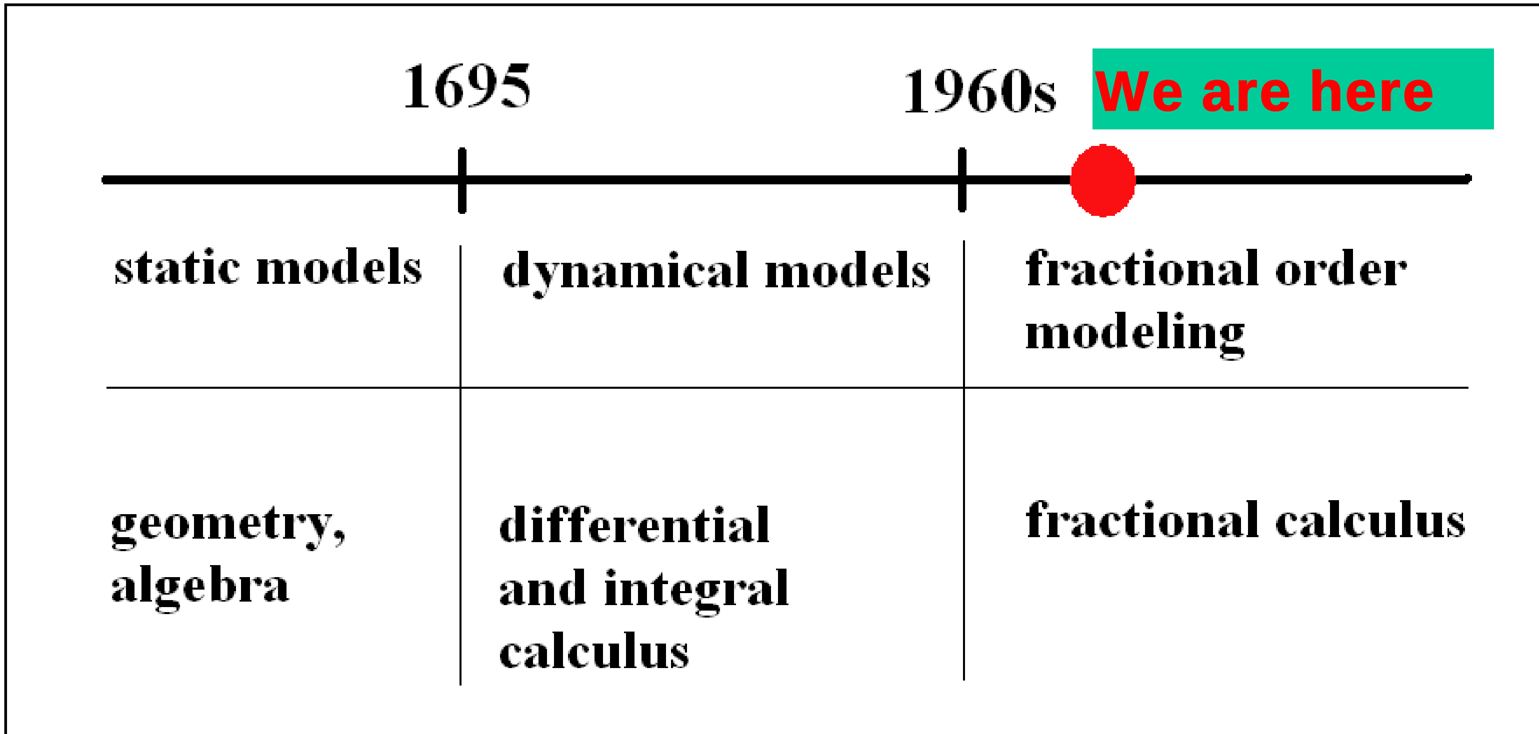
YangQuan Chen: *Is there a more optimal way to optimize*? Yes if Fractional Calculus is used:

- Optimal randomness in SGD, random search
- AGD limit dynamics is fractional order designed via FO E-L
- IMP
- … to be discovered

Outline

- **Fractional Calculus, Complexity, and Fractional Order Thinking**
- **Big Data, Variability, and Fractional Calculus**
- **Machine Learning, Optimal Randomness and Fractional Calculus**
- **Looking Into Future: Fractional Calculus is Physics Informed**

Fractional Calculus: a response to more advanced characterization of our more complex world at too small or two large scale



Slide credit: Igor Podlubny

Take home messages

Want to do better than the best?

Want to be more optimal?

Go Fractional!

Decision and Control in the Era of Big Data ?

- Yes, we must use fractional calculus!
  - Fractional order signals, systems, controls.
  - **Fractional order data analytics**

Future of Machine Learning

- Physics-informed ML
- Scientific ML – (cause-effect embedded or cause-effect discovery)

- Involving fractional calculus, we are closer to the nature, i.e., “道”

http://220.178.124.24:8080/wbbbs/archiver/?tid-16226.html

New wisdom equipped with FC

- 道可道，非常道。------世间万物的运行规律是可以被描述的，但它们并非一成不变的。

way

by non-normal

be explained

could

The nature’s rule (of complexity)

Non-normal way:

Fractional Calculus!

Heavytailedness

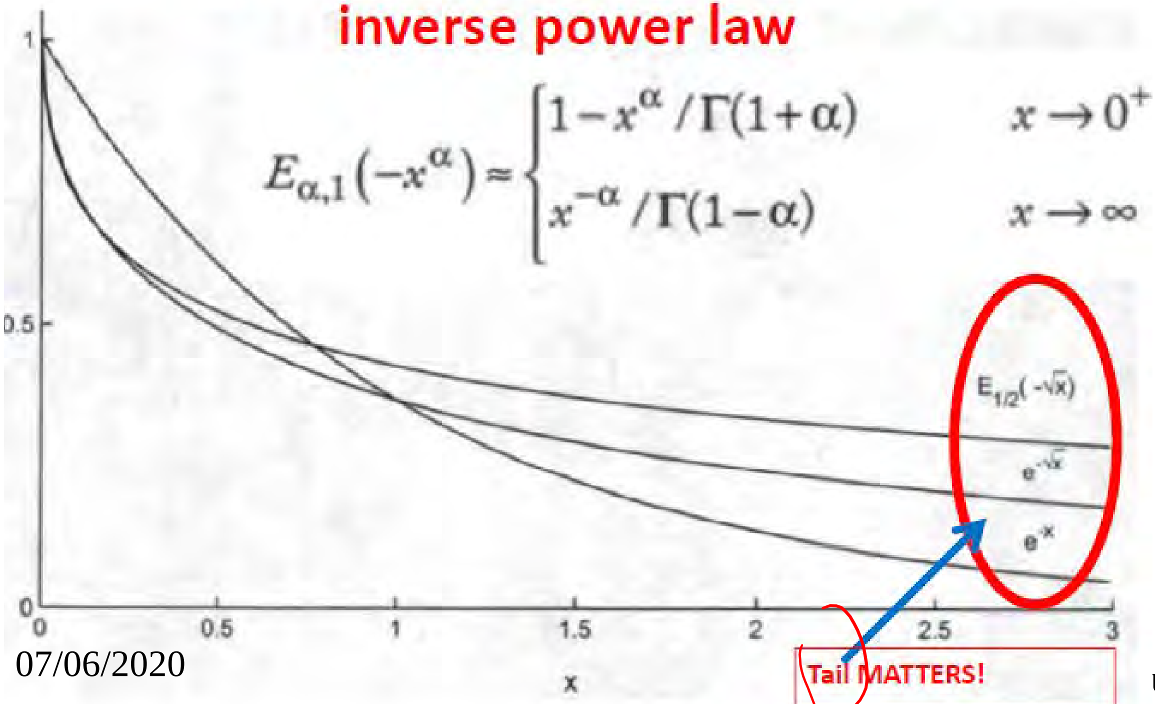
http://220.178.124.24:8080/wbbbs/archiver/?tid-16226.html

New wisdom equipped with FC

- 玄之又玄，众妙之门。-------了解这类对立统一体相互转变的规律，就是通向对世间万物理解的大门。

Root of long (algebraic) tail, or inverse power law

Non-normal way:



Fractional Calculus!

Heavytailedness

?

ust meet FC?

To have a better life, learn FC

- 老子说：“人法地，地法天，天法道，道法自然。”
  - “道法自然” – prompts the use fractional calculus
- Better understanding complexity using fractional calculus leads to 积极入世的态度（王阳明）then “天人合一”
  - 天人不合一例子：逆水行舟，冒雨走夜路，冬天穿背心。。。[1]
  - 天人合一例子：夫妻恩爱，团队精神，人养狗、狗护主。。。

中庸之道 + 道法自然 + 分数阶微积分 – 傲慢 = ~~天人合一美好人生~~

Q/A session

Thank you for your attendance and patience! Your comments and critiques are welcome!

- Thanks go to
  - Dr. Yuquan Chen, Dr. Jiamin Wei, Lihong Guo, Dr. Zhenlong Wu, Dr. Yanan Wang, Panpan Gu, Jairo Viola, Haoyu Niu, Dr. Jie Yuan etc. for walks, chats and tea/coffee breaks at Castle, Atwater, CA before COVID-19 era.

Backup slides

- **Socrates**,
  - https://en.wikipedia.org/wiki/The_unexamined_life_is_not_worth_living
- Platos, Aristotle

c. 570 – c. 495 BC
Pythagoras

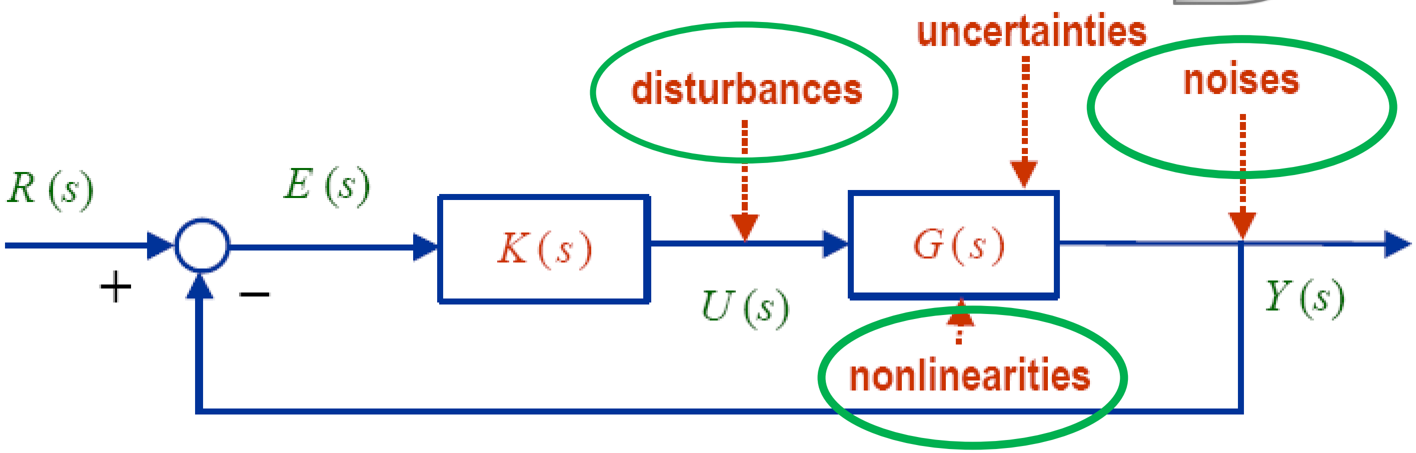
around 475 BC
Parmenides

c. 535 – c. 475 BC
Heraclitus

Fractional Order Controls

- **IO Controller + IO Plant**
- **FO Controller + IO Plant**
- **FO Controller + FO Plant**
- **IO Controller + FO Plant**

4 + 4 + ...



D. Xue and **Y. Chen***, "A Comparative Introduction of Four Fractional Order Controllers". Proc. of The 4th IEEE World Congress on Intelligent Control and Automation (WCICA02), June 10-14, 2002, Shanghai, China. pp. 3228-3235.