CIMPA Research School : Data Science for Engineering and Technology
Tunis 2019

Bregman divergences
a basic tool for pseudo-metrics building for data structured by physics

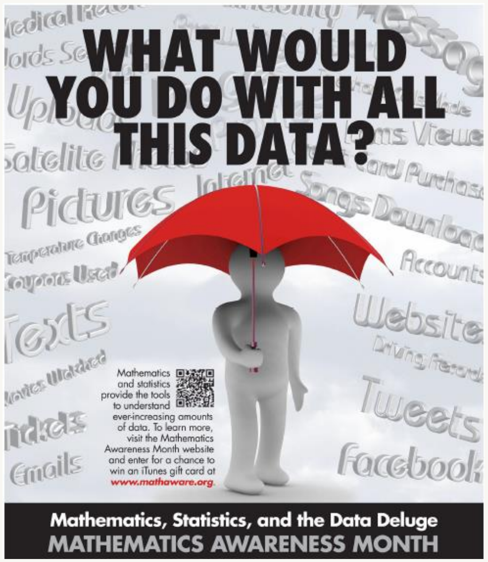
1- Introduction & Orientation

Stéphane ANDRIEUX

ONERA - France

Member of the National Academy of Technologies of France

The question : data versus knowledge



April 2012

Data Scientist:

The Sexiest Job of the 21st Century

Meet the people who can coax treasure out of messy, unstructured data.
by Thomas H. Davenport and D.J. Patil

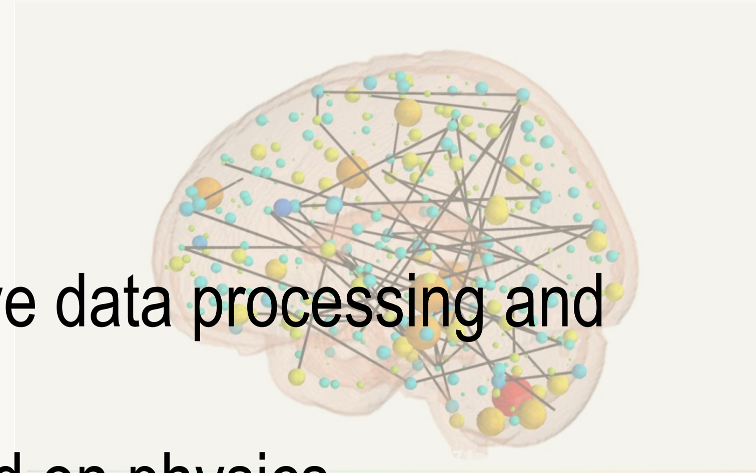
When Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren't seeking out connections with the people who were already on the site at the rate executives had expected. Something was apparently missing in the social experience. As one LinkedIn manager put it, "It was like arriving at a conference reception and realizing you don't know anyone. So you just stand in the corner sipping your drink—and you probably leave early."

70 Harvard Business Review **October 2012**

A key to the problem?

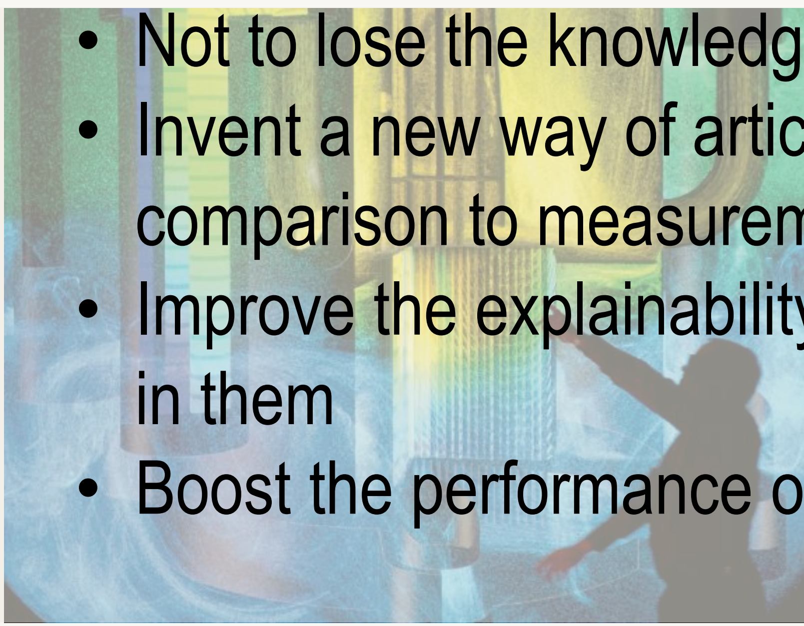
Try to deal with Data that are structured by the physics we know

Compare structured data?



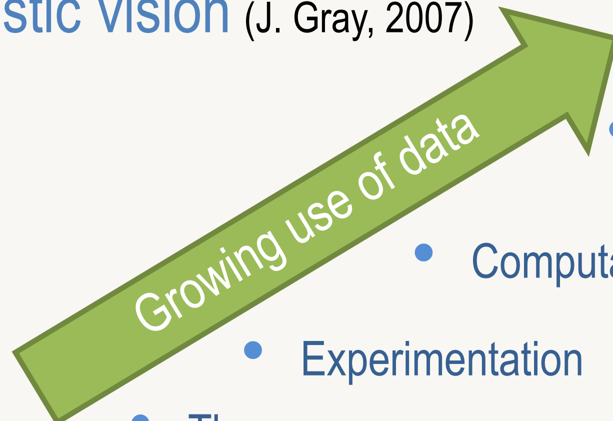
Major challenges

- Leveraging the progress of massive data processing and learning
- Not to lose the knowledge acquired on physics
- Invent a new way of articulating data and modelling, beyond comparison to measurements or validation
- Improve the explainability of results and increase confidence in them
- Boost the performance of learning algorithms



Find a way between

An enthusiastic vision (J. Gray, 2007)



- Data Driven Science
- Computational Science
- Experimentation
- Theory

The fourth paradigm

More doubts for others

Data-driven science is a failure of imagination. P. Keil

Statistics are like a bikini. What they reveal is suggestive, but what they conceal is vital. A. Levenstein

Track ? Use the geometry of solutions (mainly driven by modelling)

The most that can be expected from any model is that it can supply a useful approximation to reality: All models are wrong; some models are useful. G. Box

H. Poincaré

C'est la physique mathématique qui nous montre quels problèmes nous devons nous poser. C'est elle aussi qui nous fait prévoir la solution. (1897)

The question of causality and correlation

We are drowning in information, but starving for knowledge J. Naisbett, 1996

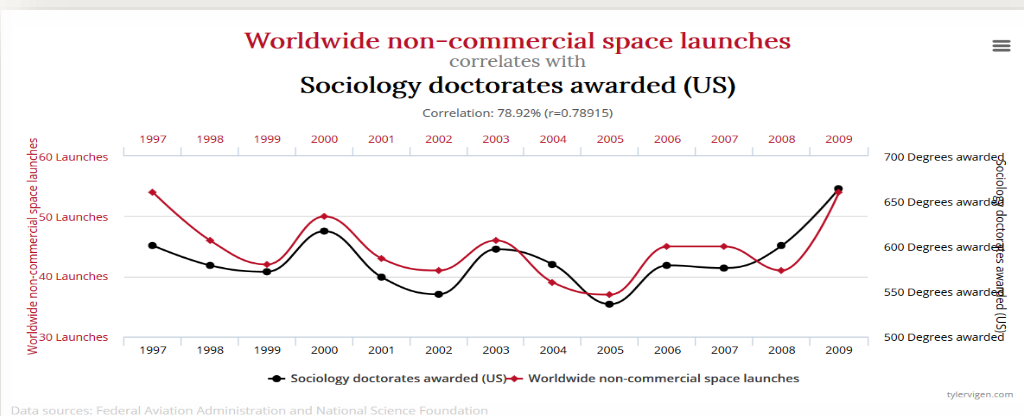
Gross correlation can lead to disasters

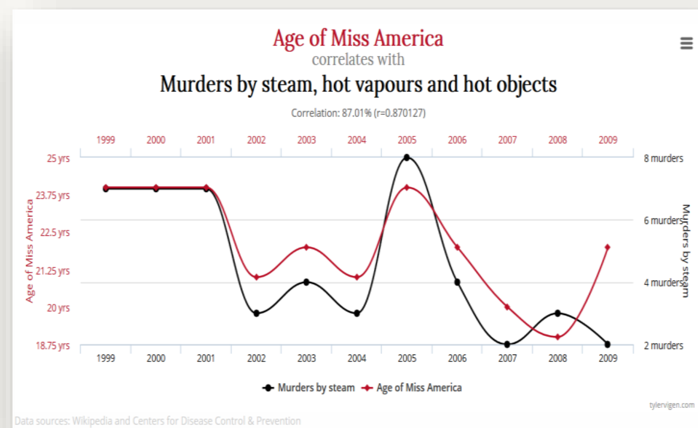
Truism: All linear dependence pairs can always be correlated

Semi-truism: even for highly non-linear dependencies there are fortuitous correlations that can be extracted with sufficient data and computation

For applications with reliability issues, precautions are mandatory

validation | extrapolation capabilities | aberrant or unpredictable behavior | fictitious causalities





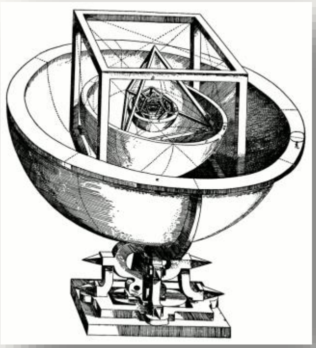
Tyler Vigen, Spurious Correlations

A brief historical review

Tycho BRAHE
1546 -1601

The very first separation between **collecting** the data and **mining** the data

Jonathan KEPLER
1571 -1630



Improves observations, collects a massive amount of data,
Observation overcomes tradition

Model fitting (musical harmony) with wrong model
Then correlation => 3 laws

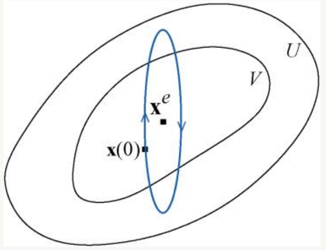
A first approach : when geometry matters (I)

Stability of trajectory for autonomous systems

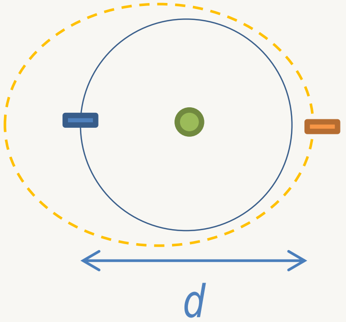
$$\boxed{\begin{cases} \dot{x} = f(x) \\ x(0) = x_0 \end{cases}}$$

Definition of Lyapunov stability of an equilibrium $f(x^e)=0$

x^e is a stable equilibrium if for every neighborhood U of x^e there is a neighborhood $V \subseteq U$ of x^e such that every solution $x(t)$ starting in V ($x_0 \in V$) remains in U for all $t \geq 0$



Stability of orbits ?

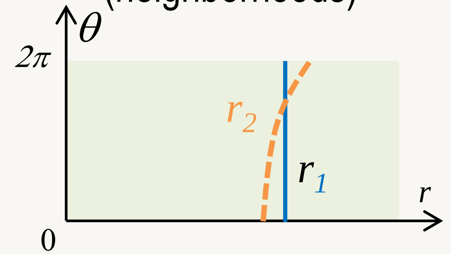


Not stable with this definition the because the periods differ

Adapt the definition via a parametrization of time

The orbit x_1 is stable if there exists a smooth monotonic function $T(t)$ such that $x_1(t)$ and $x_2(T(t))$ remain close for close initial conditions

Adapt the geometry (neighborhoods)



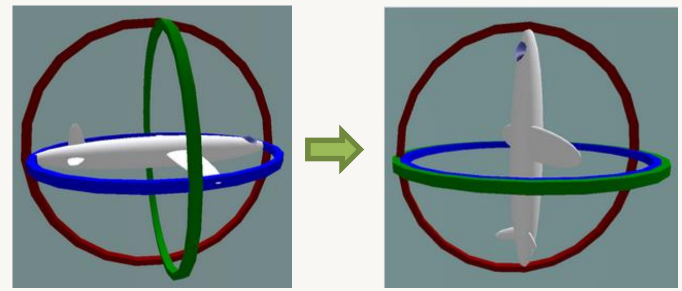
$$\left\|x_1 - x_2\right\|^2 = \int_0^{2\pi} \left(r\left[x_1(\theta)\right] - r\left[x_1(\theta)\right]\right)^2 d\theta$$

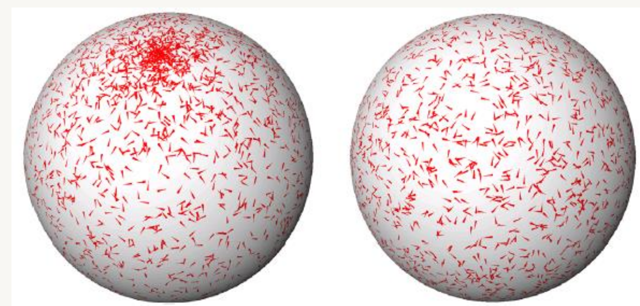
A first approach : when geometry matters (II)

Dealing with rigid body rotations

Describing the three rotations by the Euler angles can be troublesome or inefficient

- Gimpel block (loss of a dof -the parallels and meridians degenerates at poles)
- Bad (non smooth) rotation interpolation
- Bad sampling
- Non stability of orthogonal matrices with respect to noise





(a) Naïve sampling (b) Uniform sampling

J. Kuffner (ICRA 2004)

- Representation of the rotations $SO(3)$ by the unit sphere in IR^4
- Quaternions algebra
- Use of the geodesic distance on the sphere (largest circles : natural and easy to compute)

Another approach : Compare structured data?

Illustration in thermomechanics

$$div\,\underline{q} = s$$

$$div\,\underline{\underline{\sigma}} = 0$$

Conservation Laws

$$\underline{q} = -\underline{\underline{K}}\,\underline{\nabla}T$$

$$\underline{\underline{\sigma}} = \underline{\underline{\underline{\underline{A}}}}\left(\underline{\underline{\varepsilon}} - T\,\underline{\underline{\alpha}}\right)$$

Constitutive equations

$$\int_{\partial\Omega} \underline{\underline{\sigma}}_1 : \underline{\underline{\varepsilon}}_2 = \int_{\partial\Omega} \underline{\underline{\sigma}}_2 : \underline{\underline{\varepsilon}}_1$$

Global properties

What for ?

Mechanics analysis

Error estimation
Construction of cost functions for optimization
Model reduction Projections
Stability analyses

Inverse Problems

Minimization of a gap between solutions of well posed problems
Least squares alternatives in multiphysical data problems

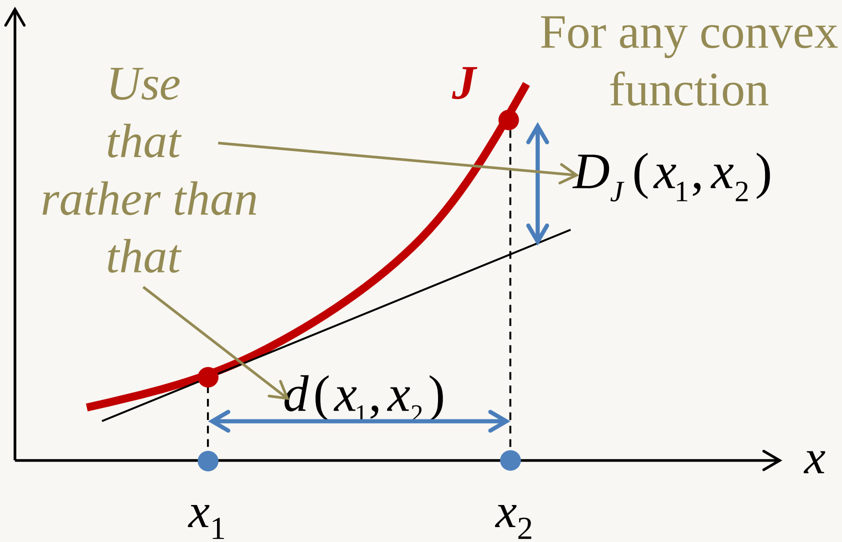
Exploitation of massive data

Learning (constitutive equations, wall law in boundary layers, closing equations,...)
Pattern detection for post-processing simulations or massive experimental data (Classification, Manifold Learning)

Today → Tomorrow

Finally the lecture's menu

Introduce and study
the **Bregman divergence**



Then

Basic computational geometry with the Bregman divergence

Extension of Proper Orthogonal Decomposition (POD) :
Using Bregman divergences to Product Spaces for multiphysics

Clustering with Bregman divergences

A about data, metrics and the triangle inequality

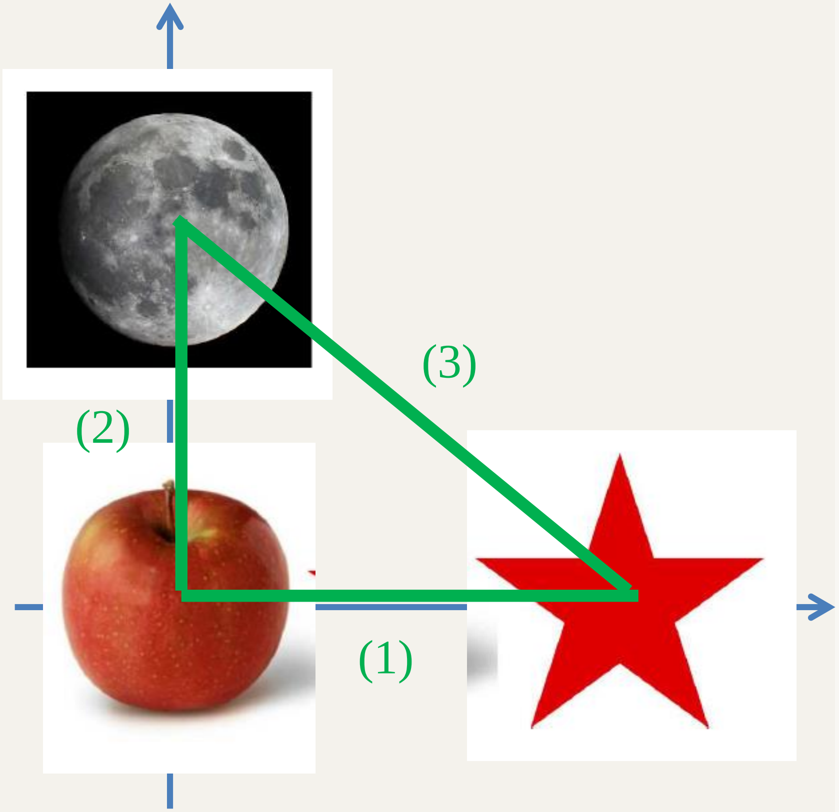
The Bregman divergence and derived measures of dissimilarity **do not** satisfy the triangle inequality

Leading the to a lot of challenges for

- computational geometry
- scalar-product based concepts in Hilbert spaces

For different dimensions and types of dissimilarities the triangle inequality is **not** pertinent

(3) > **(1)** + **(2)**



Thanks for your attention