How to Solve Gaussian Interference Channel

WPI, HKU, 2019

Fan Cheng
Shanghai Jiao Tong University

chengfan@sjtu.edu.cn

WCI 2013

- ••• Home
- Program
- Speakers
- Transport

Welcome to WCI 2013

The 2013 Workshop On Coding and Information Theory (WCI 2013) is scheduled to be held at The University of Hong Kong, December 11-13, 2013. The workshop aims at providing a platform for researchers, scientists,

WPI 2019

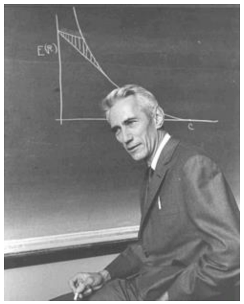
- ••• Home
- Program
- Speakers

Welcome to WPI 2019

The 2019 Workshop on Probability and Information Theory (WPI 2019) is scheduled to be held at The University of Hong Kong, August 19-22, 2019.

December 13, 2013 (Friday)			
10:00AM-11:00AM	**Estimating High-dimensional Matrices: Information-Theoretic Limits and Computational Barriers** [ABSTRACT][SLIDES]	Yihong Wu	University of Illinois, at Urbana-Champaign
11:15AM-12:15PM	**Optimal Probability Estimation with Applications to Prediction and Classification** [ABSTRACT][SLIDES]	Alon Orlitsky	University of California, San Diego
Lunch at Eliot room, 14/F, Senior Common Room (speakers invited)			
2:00PM-3:00PM	**Coding for Combined Block-Symbol Error Correction** [SLIDES]	Pascal Vontobel	Stanford University
	Informal Discussion Session		
3:20PM-3:40PM	**Some Conjectures On Entropy Power Inequality**	Fan Cheng	CUHK

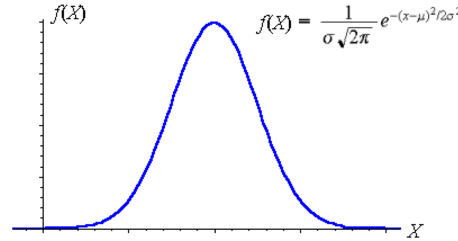
August 20, 2019 (Tuesday)			
Morning session (chaired by Venkat Anantharam)			
10:00AM-11:00AM	**Third-Order Asymptotics in Channel Coding: Old and New** [ABSTRACT][SLIDES]	Vincent Tan	National University of Singapore
11:15AM-12:15PM	**General Adversarial Channels: When Do Large Codes Exist?** [ABSTRACT][SLIDES]	Sidharth Jaggi	Chinese University of Hong Kong
12:30PM-2:20PM	Workshop lunch at Rome Cafe Graduate House HKU (speakers invited)		
Afternoon session (chaired by Pascal Vontobel)			
2:30PM-3:30PM	**Information Bottleneck Problems: An Outlook** [ABSTRACT] [SLIDES]	Shlomo Shamai	Technion
3:45PM-4:45PM	**Maximal Correlation and The Rate of Fisher Information Convergence in The Central Limit Theorem** [ABSTRACT][SLIDES]	Oliver Johnson	University of Bristol
5:00PM-6:00PM	**How to Solve Gaussian Interference Channel** [ABSTRACT][SLIDES]	Fan Cheng	Shanghai Jiao Tong University



$$\frac{\partial^2}{2\partial x^2} f(x,t) = \frac{\partial}{\partial t} f(x,t)$$

$$\boxed{h(X+\sqrt{t}Z)}$$

$$Y = X + \sqrt{t}Z$$
$$Z \sim \mathcal{N}(0,1)$$



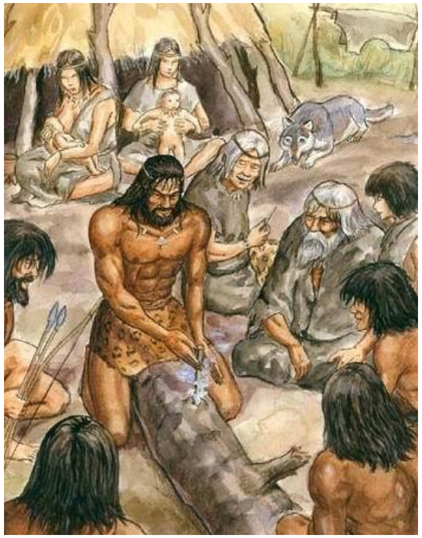
$$h(X) = -\int x\log x \, \mathrm{d}x$$

- ☐ A new mathematical theory on Gaussian distribution
- ☐ Its application on Gaussian interference channel
- ☐ History, progress, and future

Outline

- History of "Super-H" Theorem
  - Boltzmann equation, heat equation
- Shannon Entropy Power Inequality
- Complete Monotonicity Conjecture
- How to Solve Gaussian Interference Channel

Fire and Civilization



Drill

Myth: west and east

Steam engine
James Watts

The Wealth of Nations

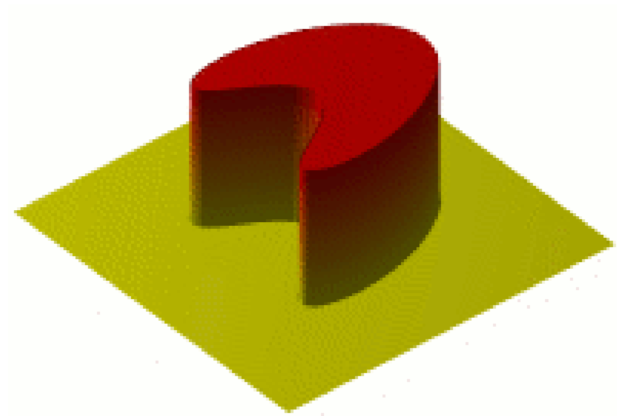
Independence of US

1776

Study of Heat

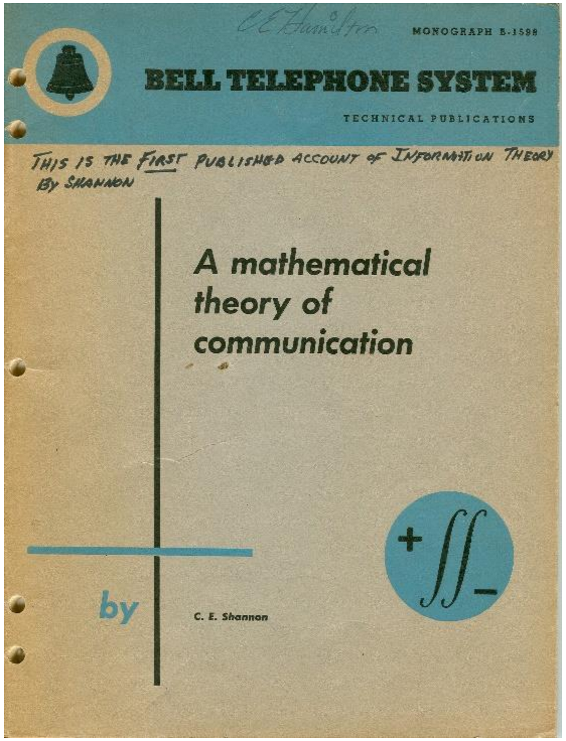
Heat transfer

$$\frac{\partial}{\partial t} f(x,t) = \frac{1}{2} \frac{\partial^2}{\partial x^2} f(x,t)$$

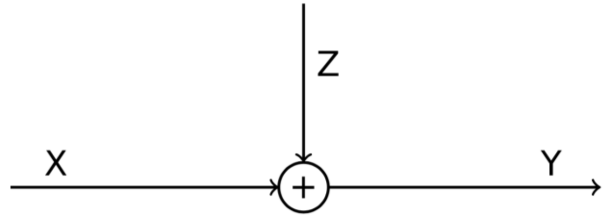


- The history begins with the work of **Joseph Fourier** around 1807
- In a remarkable memoir, Fourier invented both Heat equation and the method of Fourier analysis for its solution

Information Age



A mathematical theory of communication, Bell System Technical Journal.



Gaussian Channel:

$$Z_t \sim \mathcal{N}(0, t)$$

X and Z are mutually independent. The p.d.f of X is g(x)

Y_t is the convolution of X and Z_t.

$$Y_t := X + Z_t$$

The probability density function (p.d.f.) of Y_t

$$f(y;t) = \int g(x) \frac{1}{\sqrt{2\pi t}} e^{\frac{(y-x)^2}{2t}}$$

$$\frac{\partial}{\partial t} f(y;t) = \frac{1}{2} \frac{\partial^2}{\partial y^2} f(y;t)$$

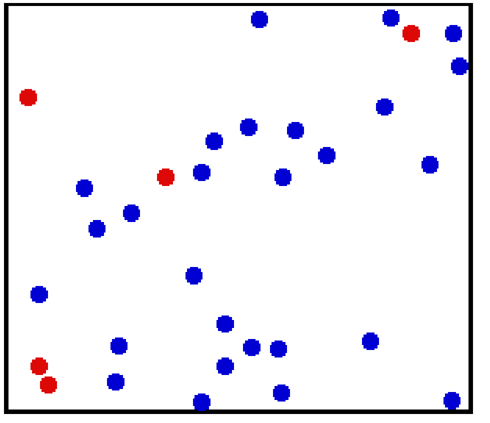
The p.d.f. of Y is the solution to the heat equation, and vice versa.
Gaussian channel and heat equation are identical in mathematics.

Ludwig Boltzmann

Ludwig Eduard Boltzmann
1844-1906
Vienna, Austrian Empire

Boltzmann formula: $S = -k_B \ln W$

Gibbs formula: $S = -k_b \sum_i p_i \ln p_i$



Boltzmann equation:

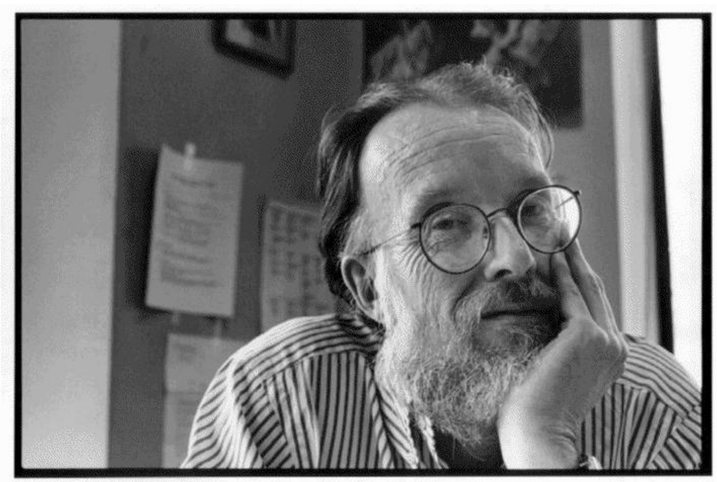
$$\frac{df}{dt} = (\frac{\partial f}{\partial t})_{force} + (\frac{\partial f}{\partial t})_{diff} + (\frac{\partial f}{\partial t})_{coll}$$

H-theorem:

$$H(f(t))\, is\ non{-}decreasing$$

“Super H-theorem” for Boltzmann Equation

A function is **completely monotone** (CM) iff all the signs of its derivatives are alternating in $+/$-: +, -, +, -,…… (e.g., $1/t, e^{-t}$)



H. P. McKean, NYU.
National Academy of Sciences

- McKean’s Problem on Boltzmann equation (1966):
  - $H(f(t))$ is CM in t, when $f(t)$ satisfies Boltzmann equation
  - False, disproved by E. Lieb in 1970s
  - the particular Bobylev-Krook-Wu explicit solutions, this “theorem” holds true for $n \leq 101$ and breaks downs afterwards

“Super H-theorem” for Heat Equation

- Heat equation: Is $H(f(t))$ CM in t, if $f(t)$ satisfies heat equation
  - Equivalently, is $H(X+\sqrt{t}Z)$ CM in t?
  - The signs of the first two order derivatives were obtained
  - Failed to obtain the 3rd and 4th. (It is easy to compute the derivatives, it is hard to obtain their signs)

> **“This suggests that……, etc., but I could not prove it”**
>
> **-- H. P. McKean**

C. Villani, 2010 Fields Medalist

A review of mathematical topics in collisional kinetic theory

Cédric Villani

completed: October 4, 2001

revised for publication: May 9, 2002

most recent corrections: June 7, 2006

Claude E. Shannon and EPI

- Central limit theorem
- Capacity region of Gaussian broadcast channel
- Capacity region of Gaussian Multiple-Input Multiple-Output broadcast channel
- Uncertainty principle

All of them can be proved by Entropy Power Inequality (EPI)

- Entropy power inequality (Shannon 1948): For any two independent continuous random variables X and Y,

$$e^{2h(\boldsymbol{X}+\boldsymbol{Y})} \geq e^{2h(\boldsymbol{X})} + e^{2h(\boldsymbol{Y})}$$

  Equality holds iff X and Y are Gaussian
  - Motivation: Gaussian noise is the worst noise
  - Impact: A new characterization of Gaussian distribution in information theory
  - Comments: most profound! (Kolmogorov)

Entropy Power Inequality

- Shannon himself didn't give a proof but an explanation, which turned out to be wrong
- The first proof is given by A. J. Stam (1959), N. M. Blachman (1966)
- Research on EPI

  Generalization, new proof, new connection. E.g., Gaussian interference channel is open, some stronger "EPI" should exist.

  - Stanford Information Theory School: Thomas Cover and his students: A. El Gamel, M. H. Costa, A. Dembo, A. Barron (1980-1990)
  - After 2000, Princeton && UC Berkeley

Heart of Shannon theory

Ramification of EPI

Shannon EPI

Gaussian perturbation: $h(X+\sqrt{t}Z)$

Fisher Information: $I\left(X+\sqrt{t}Z\right)=\frac{\partial}{\partial t}h(X+\sqrt{t}Z)/2$

Fisher Information is decreasing in t

$e^{2h(X+\sqrt{t}Z)}$ is concave in t

Fisher information inequality (FII):

$$\frac{1}{I(X+Y)}\geq\frac{1}{I(X)}+\frac{1}{I(Y)}$$

Tight Young's inequality

$|X+Y|_r\geq c|X|_p|Y|_q$

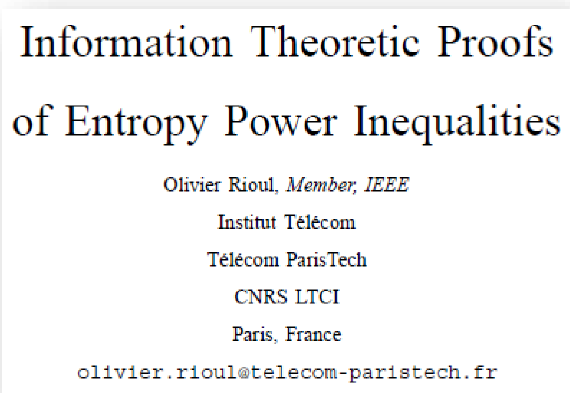
Status Quo: FII can imply EPI and all its generalizations. Many network information problems remain open even the noise is Gaussian.

--Only EPI is not sufficient

Where our journey begins

- Shannon Entropy power inequality
- Fisher information inequality
- $h(X + \sqrt{t}Z)$

- ~~$h\big(f(t)\big)$ is CM~~
- ~~When $f(t)$ satisfied Boltzmann equation, disproved~~
- ~~When $f(t)$ satisfied heat equation, unknown~~ Mathematician ignored it
- ~~We even don't know what CM is!~~


Information Theoretic Proofs of Entropy Power Inequalities

Olivier Rioul, *Member, IEEE*
Institut Télécom
Télécom ParisTech
CNRS LTCI
Paris, France
olivier.rioul@telecom-paristech.fr

- **Raymond introduced this paper to me in 2008**
- **I made some progress with Chandra Nair in 2011 (MGL)**
- **Complete monotonicity (CM) was discovered in 2012**
- **The third derivative in 2013 (Key breakthrough)**
- **The fourth order in 2014**
- **Recently, CM → GIC**

Motivation

Motivation: to find some inequalities to obtain a better rate region; e.g., the convexity of $\boldsymbol{h(X+\sqrt{e^{-t}}Z)}$, the concavity of $\boldsymbol{\frac{I(X+\sqrt{t}Z)}{t}}$, etc.

"Any progress?"
"Nope…"

It is widely believed that there should be no new EPI except Shannon EPI and FII.

Observation: $\boldsymbol{I(X+\sqrt{t}Z)}$ is convex in $\boldsymbol{t}$

$$I\left(X+\sqrt{t}Z\right)=\frac{\partial}{2\partial t}h\left(X+\sqrt{t}Z\right)\geq 0$$ (de Bruijn, *1958*)

$$I^{(1)}=\frac{\partial}{\partial t}I\left(X+\sqrt{t}Z\right)\leq 0$$ (McKean1966, Costa 1985)

Could the third one be determined?

Discovery

Observation: $\boldsymbol{I(X+\sqrt{t}Z)}$ **is convex in** $\boldsymbol{t}$

- ☐ $h\left(X+\sqrt{t}Z\right)=\frac{1}{2}\ln 2\pi e t, I\left(X+\sqrt{t}Z\right)=\frac{1}{t}$. I is CM: +, -, +, -…
- ☐ If the observation is true, the first three derivatives are: +, -, +
- ☐ Q: Is the 4th order derivative -? Because Z is Gaussian! If so, then…
- ☐ The signs of derivatives of $h(X+\sqrt{t}Z)$ are independent of X. **Invariant!**
- ~~☐ Exactly the same problem in McKean's 1966 paper~~

My own opinion:

- A new fundamental result on Gaussian distribution
- Invariant is very important in mathematics
- In mathematics, the more beautiful, the more powerful
- Very hard to make any progress

To convince people, must prove its convexity

Challenge

Let $X \sim g(x)$

$$Y_t := X + \sqrt{t}Z$$

$$Y_t \sim f(y,t) = \int g(x) \frac{1}{\sqrt{2\pi e t}} e^{-\frac{(y-x)^2}{2t}} \mathrm{d}x$$

$$f_n = f^{(n)}(y,t) := \frac{\partial^n}{\partial y^n} f(y,t)$$

- $h(Y_t) = -\int f(y,t) \ln f(y,t)\, dy$: no closed-form expression except for some special $g(x)$.
- $f(y,t)$ satisfies heat equation.
- $I(Y_t) = \int \frac{f_1^2}{f} dy$
- $I^{(1)}(Y_t) = -\int \left(\frac{f_2}{f} - \frac{f_1^2}{f^2}\right)^2 dy$
- So what is $I^{(2)}$? (Heat equation, integration by parts)

Challenge (cont'd)

$$\frac{\partial}{\partial t}h(Y_t) = \frac{1}{2}\boldsymbol{I}(Y_t);$$

$$\frac{\partial^2}{\partial t^2}h(Y_t) = -\frac{1}{2}\int f\left(\frac{f_{yy}}{f} - \frac{f_y^2}{f^2}\right)^2 \mathrm{d}y.$$

$$\frac{\partial^3 h(Y_t)}{\partial t^3} = \frac{1}{2}\int \frac{2f_y^3 f_{yyy}}{3f^3} - \frac{f_y^4 f_{yy}}{2f^4} - \frac{f_{yy}f_{yyyy}}{f} + \frac{f_{yy}^3}{2f^2}\mathrm{d}y$$

It is trivial to calculate derivatives.
It is not generally obvious to prove their signs.

Breakthrough

Integration by parts: $\int u dv = uv - \int v du$

$$\int \frac{f_1^4 f_2}{f^4} \mathrm{d}y = \int \frac{4 f_1^6}{5 f^5} \mathrm{d}y$$

$$\int \frac{f_1^3 f_3}{f^3} \mathrm{d}y = \int -\frac{3 f_1^2 f_2^2}{f^3} + \frac{12 f_1^6}{5 f^5} \mathrm{d}y$$

$$\int \frac{f_1 f_2 f_3}{f^2} \mathrm{d}y = \int -\frac{f_2^3}{2 f^2} + \frac{f_1^2 f_2^2}{f^3} \mathrm{d}y$$

$$\int \frac{f_2 f_4}{f} \mathrm{d}y = \int -\frac{f_3^2}{f} - \frac{f_2^3}{2 f^2} + \frac{f_1^2 f_2^2}{f^3} \mathrm{d}y$$

Higher Order Derivatives in Costa's Entropy Power Inequality 7 2015
F Cheng, Y Geng
IEEE Transactions on Information theory 61 (11), 5892-5905

First breakthrough since McKean 1966

$$\frac{\partial^3}{\partial t^3} h(Y_t) = \frac{1}{2} \frac{\partial^2}{\partial t^2} I(Y_t)$$

$$= \frac{1}{2} \int f \left(\frac{f_3}{f} - \frac{f_1 f_2}{f^2} + \frac{1}{3} \frac{f_1^3}{f^3} \right)^2 + \frac{f_1^6}{45 f^5} dy$$

$$
\begin{aligned}
&\frac{\partial^4}{\partial t^4} h(Y_t) \\
&= -\frac{1}{2} \int f \left(\frac{f_4}{f} - \frac{6}{5} \frac{f_1 f_3}{f^2} - \frac{7}{10} \frac{f_2^2}{f^2} + \frac{8}{5} \frac{f_1^2 f_2}{f^3} - \frac{1}{2} \frac{f_1^4}{f^4} \right)^2 \\
&+ f \left(\frac{2}{5} \frac{f_1 f_3}{f^2} - \frac{1}{3} \frac{f_1^2 f_2}{f^3} + \frac{9}{100} \frac{f_1^4}{f^4} \right)^2 \\
&+ f \left(-\frac{4}{100} \frac{f_1^2 f_2}{f^3} + \frac{4}{100} \frac{f_1^4}{f^4} \right)^2 \\
&+ \frac{1}{300} \frac{f_2^4}{f^3} + \frac{56}{90000} \frac{f_1^4 f_2^2}{f^5} + \frac{13}{70000} \frac{f_1^8}{f^7} dy
\end{aligned}
$$

GCMC

Gaussian complete monotonicity conjecture (GCMC):

$I(X+\sqrt{t}Z)$ **is CM in** t

Conjecture 2: $\log I(X+\sqrt{t}Z)$ **is convex in** t

A general form: number partition. Hard to determine the coefficients.

$$\mathcal{P}_j : n = \sum n_i \Longrightarrow M_j = \prod_{\mathcal{P}_j} \frac{f_{n_i}}{f}$$

$$\frac{\partial^n}{\partial t^n} h(Y_t) = \operatorname{sgn}(n) \int f \sum_k \left(\sum_j \beta_{k,j} M_j \right)^2 \mathrm{d}y$$

Hard to find $\beta_{k,j}$!

Moreover

C. Villani showed the work of H. P. McKean to us.
G. Toscani cited our work within two weeks:

- the consequences of the evolution of the entropy and of its subsequent derivatives along the solution to the heat equation have important consequences.
- Indeed the argument of McKean about the signs of the first two derivatives are equivalent to the proof of the logarithmic Sobolev inequality.

Gaussian optimality for derivatives of differential entropy using linear matrix inequalities
X. Zhang, V. Anantharam, Y. Geng - Entropy, 2018 - mdpi.com

- A new method to prove signs by LMI
- Verified the first four derivatives
- For the fifth order derivative, current methods cannot find a solution

Complete monotone function

Theorem (Hausdorff-Bernstein-Widder theorem)

If $f(t)$ is completely monotone, then there is a non-negative finite Borel measure on $[0, \infty)$, with cumulative distribution function μ such that

$$f(t) = \int_0^\infty e^{-tx}\, d\mu(x)$$

How to construct $\mu(x)$?

A new expression for entropy involved special functions in mathematical physics

Bessis-Moussa-Villani, 1975

Let A and B be two $n \times n$ Hermitian matrices and let B be positive semidefinite.

$$f(t) := \mathrm{Tr} e^{A-tB}, t \geq 0,$$

is completely monotone.

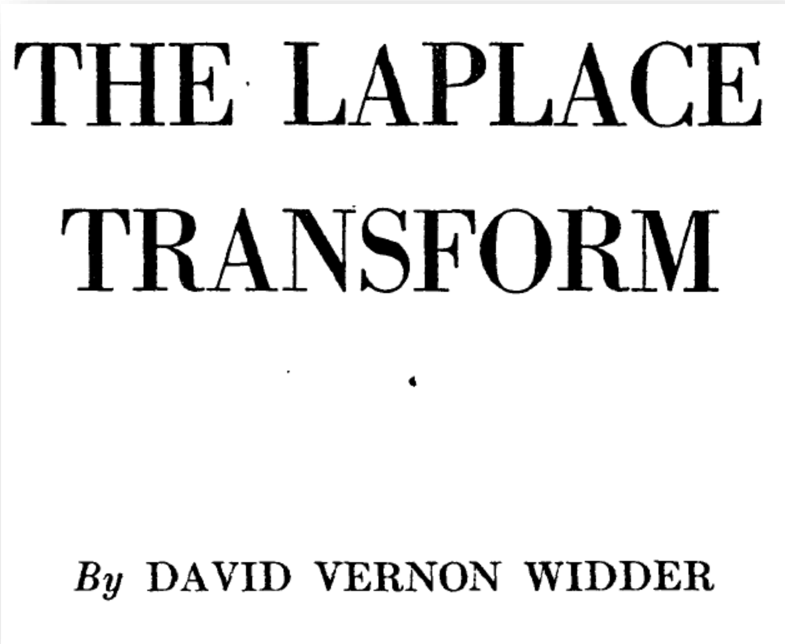
Herbert R. Stahl, 2013

Complete monotone function

Theorem: A function $f(t)$ is CM in t, then $\log f(t)$ is also convex in t

- $I(Y_t)$ is CM in t, then $\log I(Y_t)$ is convex in t (Conjecture 1 implies Conjecture 2)

- A function f(t) is CM, a Schur-convex function can be obtained by f(t)
- Schur-convex → Majority theory


THE LAPLACE TRANSFORM

By DAVID VERNON WIDDER

1946

Remarks: The current tools in information theory don't work. More sophisticated tools should be built to attack this problem.

A new mathematical foundation of information theory

True Vs. False

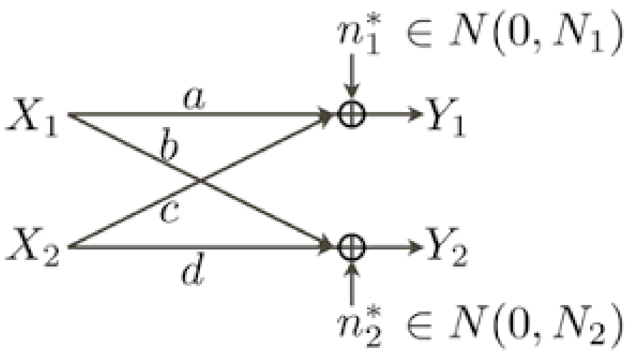
- If GCMC is true
  - A fundamental breakthrough in mathematical physics, information theory and any disciplines related to Gaussian distribution
  - A new expression for Fisher information
  - Derivatives are an invariant
  - Though $h(X+\sqrt{t}Z)$ looks very messy, certain regularity exists
  - Application: Gaussian interference channel?



- If GCMC is false
  - No Failure, as heat equation is a physical phenomenon
  - A Gauss constant (e.g. 2019), where Gaussian distribution fails. Painful!

Complete Monotonicity:

How to Solve Gaussian Interference Channel



- Two fundamental channel coding problem: BC and GIC
- $h(aX_1 + cX_2 + N_1), h(bX_1 + dX_2 + N_2)$ exceed the capability of EPI
- Han-Kobayashi inner bound
- Many researchers have contributed to this model
- Foundation of wireless communication

The Thick Shell over $h(X+\sqrt{t}Z)$

$h(X+\sqrt{t}Z)$ is hard to estimate:

- ☐ The p.d.f of $X+\sqrt{t}Z$ is messy
- ☐ $f(x)\log f(x)$
- ☐ $\int f(x)\log f(x)$

No generally useful lower or upper bounds

--The thick shell over $X+\sqrt{t}Z$

Analysis: alternating is the worst

- If the CM property of $h(X + \sqrt{t}Z)$ is not true
  - Take 5 for example: if CM breaks down after n=5
  - If we just take the 5th derivative, there may be nothing special. (So GIC won't be so hard)
  - CM affected the rate region of GIC

- Prof. Siu, Yum-Tong: "Alternating is the worst thing in analysis as the integral is hard to converge, though CM is very beautiful"
  - It is not strange that Gaussian distribution is the worst in information theory

- Common viewpoint: information theory is about information inequality: EPI, MGL, etc.
- CM is a class of inequalities. We should regard it as a whole in application. We should pivot our viewpoint from inequalities.

Information Decomposition

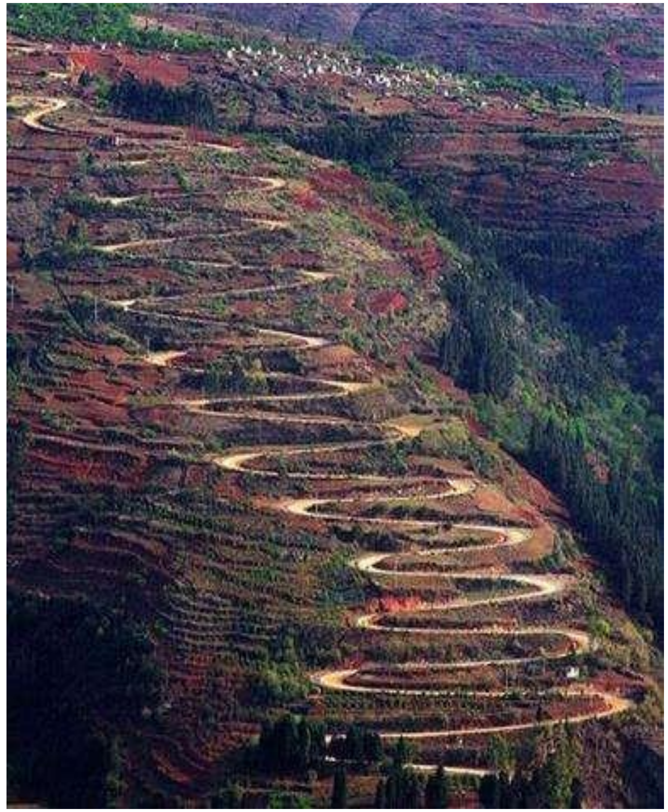
- The lesson learned from complete monotonicity

$$I(X+\sqrt{t}Z)=\int_0^{\infty} e^{-tx}\,d\mu(x)$$

- Two independent components:
  - e^{-tx} stands for complete monotonicity
  - $d\mu(x)$ serves as the **identity** of $I(X+\sqrt{t}Z)$
- Information decomposition:

  Fisher Information = Complete Monotonicity + Borel Measure
- CM is the thick shell. It can be used to estimate in majority theory
  - Very useful in analysis and geometry
- $d\mu(x)$ involves only x, and t is removed
  - The thick shell is removed from Fisher information
  - $d\mu(x)$ is relatively easier to study than Fisher information
  - WE know very little about $d\mu(x)$
- Only CM is useless for (network) information theory
  - The current constraints on $d\mu(x)$ are too loose
  - Only the "special one" is useful, otherwise every CM function should have the same meaning in information theory

CM && GIC



A fundamental problem should have a nice and clean solution.
To understand complete monotonicity is not an easy job (10 years).
Top players are ready, but the football is missing…

Thanks!

Guangyue

Raymond, Chandra, Venkat, Vincent...