Algebraic Related Structures and the Reason Behind Some Classical Constructions in Convex Geometry and Analysis

Vitali Milman

Tel-Aviv University

With deep sadness and warm memories of Victor Havin, and the fifty years of our friendship

St. Petersburg, June 2016

Instead of an Introduction:

What should we call "duality"?

Consider the class $\mathrm{Cvx}\left(\mathbb{R}^n\right)$ of all *lower-semi-continuous* convex functions $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. The Legendre transform is the map

$$\mathcal{L}\varphi(x) = \sup_{y \in \mathbb{R}^n} \left[\langle x, y \rangle - \varphi(y) \right] := \varphi^*(x)$$

(there are many "Legendre transforms": We may select 0 of the space, a scalar product and a shift for a function).

Theorem (Artstein–Milman)

1. Assume $T : \mathrm{Cvx}(\mathbb{R}^n) \to \mathrm{Cvx}(\mathbb{R}^n)$ satisfies:
   (a) $T \cdot T\varphi = \varphi$ (for any $\varphi \in \mathrm{Cvx}(\mathbb{R}^n)$);
   (b) $\varphi \le \psi$ implies $T\varphi \ge T\psi$.

   Then T is a Legendre transform.
   Note, elementary properties (a) and (b) essentially uniquely define the Legendre transform – originally a construction.
   (Side remark: It means that $\mathrm{Cvx}(\mathbb{R}^n)$ has a unique duality structure.)
   To be precise, we explain "essential uniqueness": Fix $\langle x, y\rangle$; $\exists c_0 \in \mathbb{R}$, $v_0 \in \mathbb{R}^n$, symmetric linear $B \in \mathrm{GL}_n$ s.t.

$$(T\varphi)(x) = (\mathcal{L}\varphi)(Bx + v_0) + \langle x, v_0\rangle + c_0.$$

   How far can this point of view be extended?

Consider the fundamental constructions of Convex Geometry:

- $\mathcal{K}^n$ is the class of closed convex sets [:=bodies] in $\mathbb{R}^n$.
- $\mathcal{K}_0^n := \mathcal{K}_0^n(\mathbb{R}^n) = \{K \in \mathcal{K}^n$ such that $0 \in K\}$ and fixed scalar product $\langle \cdot, \cdot \rangle$.

  Polarity $K \in \mathcal{K}_0^n \to K^\circ$:

  $$K^\circ = \left\{x \in \mathbb{R}^n \;\middle|\; \langle x, y \rangle \leq 1 \ \forall y \in K \right\} \in \mathcal{K}_0\left(\mathbb{R}^n\right).$$

  Supporting function $K \in \mathcal{K}^n \to h_K(x)$

  $$h_K(x) = \sup_{y \in K} (x, y) \in \mathcal{H}_0(\mathbb{R}^n).$$

For $K \in \mathcal{K}_0\left(\mathbb{R}^n\right)$ the gauge function M (or Minkowski functional) is the 1-homogeneous convex function $\|x\|_K$, "generalized" norm, s.t.

$$K = \left\{x \in \mathbb{R}^n : \ \|x\|_K \leq 1\right\},$$

i.e. $M(\mathbb{1}_K^\infty) = \|x\|_K$.

- Denote $\mathcal{H}_0 = \left\{\|x\|_K : \ K \in \mathcal{K}_0\left(\mathbb{R}^n\right)\right\}$

Polarity map

(i) The map $K \mapsto K^\circ$ is (essentially) a unique map $\varphi : \mathcal{K}_0 \to \mathcal{K}_0$ which

1. involution
2. reverse the order of embedding:

$$A \subset B \Rightarrow \varphi(A) \supset \varphi(B).$$

(Artstein-Milman for this class following earlier result by Böröczky-Schneider for a different but similar class.)

(This is an "analytic" characterization.)

The support and gauge maps

(ii) Also a map $\psi : \mathcal{K}_0 \to \mathcal{H}_0$ (1-1 and onto) which preserves the order (i.e. $A \subset B \Rightarrow \psi_A(x) \leq \psi_B(x)$) is (essentially, up to selecting a scalar product) the above supporting map $S(\mathbf{1}_K^\infty) \equiv S(K) := h_K(x)$.

(iii) The gauge map is (essentially) the unique order reversing map $M(\mathbb{1}_K^\infty) = \|x\|_K$.

Another characterization

Now we give an "algebraic" characterization of Polarity.

K and L are in incidence relation iff $K \not\subseteq L$ and $L \not\subseteq K$.

Theorem (Artstein-Milman)

Let $T : \mathcal{K}_0\left(\mathbb{R}^n\right) \to \mathcal{K}_0\left(\mathbb{R}^n\right)$ *be a bijection which preserves incidence relation (in both directions). Then* $\exists B \in GL_n$ *such that* T *is either*

$$TK = BK$$

or

$$TK = \left(BK\right)^{\circ}$$

To see the usefulness of these characterizations, let us extend these very geometric constructions to the setting of functions, where geometric interpretation of these constructions is impossible.

Let us embed $\mathcal{K}(\mathbb{R}^n) = \{K \subseteq \mathbb{R}^n \mid \text{closed convex}\}$ into $\mathrm{Cvx}\left(\mathbb{R}^n\right)$ by "convex characteristic" functions:

$$K \longrightarrow \mathbf{1}_K^\infty = \begin{cases} 0 & x \in K, \\ +\infty & x \notin K. \end{cases}$$

- $\mathrm{Cvx}_0\left(\mathbb{R}^n\right) = \{f \in \mathrm{Cvx}\left(\mathbb{R}^n\right) \text{ such that } f \geq 0 \text{ and } f(0) = 0\}$ (geometric convex functions)

Obviously, after such an embedding and using inequalities for functions, $M : \mathcal{K}_0(\mathbb{R}^n) \to \mathcal{H}_0$ is an order preserving map (1-1 and onto) and $S : \mathcal{K}_0(\mathbb{R}^n) \to \mathcal{H}_0$ is an order reversing map.

The following theorem extends the notion of Support function to the $\mathrm{Cvx}(\mathbb{R}^n)$ and Polarity to the $\mathrm{Cvx}_0(\mathbb{R}^n)$ setting:

Theorem (Artstein–Milman)

1. *There is a unique order reversing extension of the support map S to $\mathrm{Cvx}(\mathbb{R}^n)$ which is the Legendre transform.*
2. *There is a **unique** order reversing extension of the polarity map $\{\mathbf{1}_K^\infty \to \mathbf{1}_{K^\circ}^\infty \mid K \in \mathcal{K}_0\}$ to $\mathrm{Cvx}_0(\mathbb{R}^n) \setminus 0$ defined by*

$$\mathcal{A}f = \sup \frac{\langle x, y\rangle - 1}{f(y)} := f^\circ(x),$$

and $\mathcal{A}(0) := \mathbf{1}_{\{0\}}^\infty$. (By extension we mean that $\mathcal{A}\mathbf{1}_K^\infty = \mathbf{1}_{K^\circ}^\infty$.)

Continuation of Theorem

3. Consider the order preserving map (involution)

$$\mathcal{J} = \mathcal{L}\mathcal{A} = \mathcal{A}\mathcal{L}$$

which connects two dualities (supporting map – Legendre transform $\mathcal{L}$, and geometric duality $\mathcal{A}$) and acts ray-wise (i.e. $(\mathcal{J}f)|_r = \mathcal{J}(f|_r)$ for any ray r).
$\mathcal{J} : Cvx_0 \to Cvx_0$ is order preserving and it is the gauge map: $\mathcal{J}$ is the only order preserving extension of the Minkowski map M onto $Cvx_0(\mathbb{R}^n)$, i.e.

$$\mathcal{J}(\mathbf{1}_K^\infty) \equiv M(\mathbf{1}_K^\infty) = \|x\|_K.$$

So, on the class of convex functions we have the notion of support function ($\mathcal{L}$ transform), Minkowski functional ($\mathcal{J}$-map) and polarity ($\mathcal{A}$-transform)!

By the way, note that there are ONLY two dualities on $\mathrm{Cvx}_0\left(\mathbb{R}^n\right)$ — $\mathcal{L}$ and $\mathcal{A}$:

Theorem (Artstein–Milman)

Let $n \geq 2$. The maps $\mathcal{L}$ and $\mathcal{A}$ are (essentially) the <u>only</u> order reversing involutions on $\mathrm{Cvx}_0(\mathbb{R}^n)$. Precisely: if $T : \mathrm{Cvx}_0 \to \mathrm{Cvx}_0$ is

1. *involution $T \cdot T = Id$.*
2. *order reversing: $\forall f, g \in Cvx_0$ we have $f \leq g \Rightarrow Tf \geq Tg$,*

then $\exists C > 0$ and $B \in GL_n$, symmetric, s.t. either

$$\forall f \in Cvx_0, \quad Tf = \mathcal{L}(f(Bx))$$

or

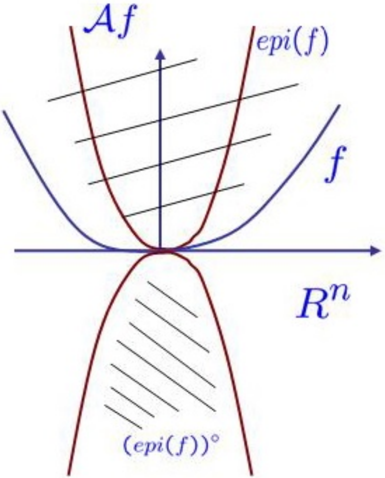
$$\forall f \in Cvx_0, \quad Tf = C\mathcal{A}(f(Bx)).$$

(When $n = 1$ there are 8 such different dualities)

- So, on the class of geometric convex functions there are exactly two dualities: one representing the support map and another geometric notion of polarity.

Geometric interpretation of $\mathcal{A}$-duality, I

$\mathrm{epi}\,(\mathcal{A}f)$ is the reflection of $(\mathrm{epi}\,f)^{\circ}$:



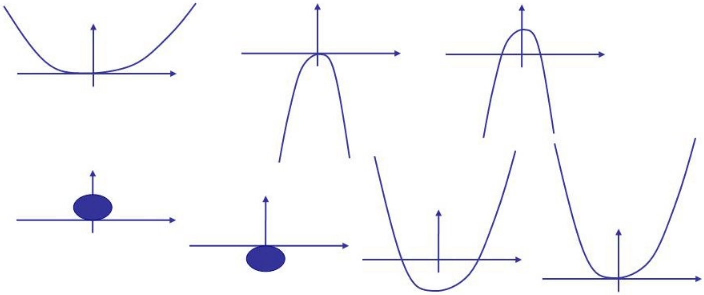
Geometric interpretation of the Legendre transform

Let e be the unit vector in $\mathbb{R}^+$. Then

$$\operatorname{epi}(\mathcal{L}f) = \Big(((\operatorname{epi} f)^\circ + e)^\circ - e\Big)^\circ + e$$

or also

$$\operatorname{epi}(\mathcal{L}f) = \Big(((\operatorname{epi} f - e)^\circ + e)^\circ - e\Big)^\circ.$$



Let me add that also the notion of mixed volumes (Minkowski) and the Minkowski Polarization theorem for the family of Convex sets may be extended for the class of log-concave functions (and also for larger classes of functions).

One should introduce an analog of Minkowski summation for this class (done with L. Rotem) which leads to the notion of mixed integrals. And many deep geometric inequalities (such as Brunn–Minkowski, isoperimetric, Urysohn, Alexandrov, Alexandrov–Fenchel) may be extended to this and other classes.

Different classes of functions

- Because convex functions are almost never integrable, we introduce different "scalings" of convex functions, and consider the classes $\{e^{-\varphi} : \ \varphi \text{ convex}\}$ (log-concavity) and $\left\{f = \left(1 + \frac{\varphi}{\beta}\right)^{-\beta} : \ \varphi \text{ is convex base of } f\right\}$ (α-concavity). The formal (structural) definitions are:

Different classes of functions

Definitions

- Fix $-\infty \le \alpha \le \infty$ (where $\beta = -1/\alpha$). A function $f : \mathbb{R}^n \to [0, \infty)$ is α-concave if for every $x, y \in \mathbb{R}^n$ such that $f(x), f(y) > 0$ and every $0 < \lambda < 1$ we have

$$f(\lambda x + (1-\lambda)y) \ge \left[\lambda f(x)^\alpha + (1-\lambda) f(y)^\alpha\right]^{\frac{1}{\alpha}}$$

- A function $f : \mathbb{R}^n \to [0, \infty)$ is called log-concave if for every $x, y \in \mathbb{R}^n$ and every $0 < \lambda < 1$ we have

$$f(\lambda x + (1-\lambda) y) \ge f(x)^\lambda f(y)^{1-\lambda}$$

  (the case $\alpha \to 0$).
- A function $f : \mathbb{R}^n \to [0, \infty)$ is quasi-concave if for every $x, y \in \mathbb{R}^n$ and every $0 < \lambda < 1$ we have

$$f(\lambda x + (1-\lambda)y) \ge \min\{f(x), f(y)\}$$

  (the case $\alpha \to -\infty$, but $\overline{\bigcup\{\alpha\text{-concave}\}} \ne$ quasi-concave)

Additions on functions

- We will always assume our α-concave functions are upper semi-continuous with $\sup f = 1$. Denote the classes of such log-concave and quasi-concave functions by $\mathrm{LC}\left(\mathbb{R}^n\right)$ and $\mathrm{QC}\left(\mathbb{R}^n\right)$ respectively.
- On convex functions, there exists a non-trivial (but by now standard) addition known as inf-convolution:

$$\left(\varphi \Box \psi\right)(x) = \inf_{y+z=x} \left[\varphi(y) + \psi(z)\right].$$

- The induced homothety is $\left(\lambda \cdot \varphi\right)(x) = \lambda \varphi\left({}^{x}/_{\lambda}\right)$.

Additions on functions

- Since there is a 1-1 correspondence between convex and α-concave functions, this also introduces operations on α-concave functions for every $-\infty < \alpha \leq 0$. For log-concave functions this is known as the sup-convolution:

$$(f \star g)(x) = \sup_{y+z=x} f(y)g(z),$$

and for α-concave functions

$$f_1 \star_\alpha f2 = \left(1 + \frac{\varphi_1 \Box \varphi_2}{\beta}\right)^{-\beta}$$

where φ_i is the base of f_i.

Minkowski's theorem

Theorem (Minkowski)

For every $K_1, K_2, \ldots, K_m \in \mathcal{K}^n$ *we have*

$$Vol\left(\sum_{i=1}^{m} \lambda_i K_i\right) = \sum_{i_1,\ldots,i_n=1}^{m} \mathrm{V}(K_{i_1}, K_{i_2}, \ldots, K_{i_n})\, \lambda_{i_1}\lambda_{i_2}\cdots\lambda_{i_n}$$

The coefficient $\mathrm{V}(K_{i_1}, K_{i_2}, \ldots, K_{i_n})$ is called the mixed volume of $K_{i_1}, K_{i_2}, \ldots, K_{i_n}$.

The function $V : (\mathcal{K}^n)^n \to [0, \infty]$ is the unique symmetric multi-linear function such that

$$\mathrm{V}(K, K, \ldots, K) = \mathrm{Vol}\,(K)$$

for all $K \in \mathcal{K}^n$. This means that volume is polarized on the class of convex bodies with respect to the Minkowski addition.

Minkowski sum for functions

- The class $\mathrm{QC}\left(\mathbb{R}^n\right)$ is NOT in natural correspondence with $\mathrm{Cvx}\left(\mathbb{R}^n\right)$. However, formally sending $\alpha \to -\infty$ in the α-convolution $f \star_\alpha g$ we arrive at the following definitions:

$$(f \oplus g)(x) = \sup_{y \in \mathbb{R}^n} \min\left\{f(y), g(x-y)\right\}$$

$$(\lambda \odot f)(x) = f\left(\frac{x}{\lambda}\right)$$

- For $f \in \mathrm{QC}\left(\mathbb{R}^n\right)$ and $0 \le t \le 1$ define

$$K_t(f) = \{x \in \mathbb{R}^n : \ f(x) \ge t\} \in \mathcal{K}^n.$$

- For every $f, g \in \mathrm{QC}\left(\mathbb{R}^n\right)$, $\lambda > 0$ and $0 \le t \le 1$:

$$K_t\left(f \oplus (\lambda \odot g)\right) = K_t(f) + \lambda K_t(g).$$

- From here we may check that $\mathrm{LC}\left(\mathbb{R}^n\right)$ is closed under $\oplus$, so we may apply this “Minkowski” summation for the log-concave (and also α-concave) classes.

Minkowski theorem for functions

This new summation leads to polynomiality of Lebesgue integral and allows to polarize Lebesgue integral on log-concave (or quasi-concave) functions.

Theorem (Milman–Rotem)

For every $f_1, f_2, \dots, f_m \in \mathrm{QC}\left(\mathbb{R}^n\right)$ *we have*

$$\int \bigoplus_{i=1}^{m} (\lambda_i \odot f_i) = \sum_{i_1,\dots,i_n=1}^{m} \mathrm{V}(f_{i_1}, f_{i_2}, \dots, f_{i_n})\, \lambda_{i_1}\lambda_{i_2}\cdots\lambda_{i_n}$$

The coefficient $\mathrm{V}(f_{i_1}, f_{i_2}, \dots, f_{i_n})$ will be called the mixed integral of $f_{i_1}, f_{i_2}, \dots, f_{i_n}$. The function $V : \mathrm{QC}\left(\mathbb{R}^n\right)^n \to [0, \infty]$ is the unique symmetric multi-linear function such that

$$\mathrm{V}(f, f, \dots, f) = \int f$$

for all $f \in \mathrm{QC}\left(\mathbb{R}^n\right)$.

Minkowski theorem for functions

- Notice that this theorem is applicable for the category of pairs (K, μ) where K is convex and compact and μ is a log-concave measure supported on K (A very natural family!).
- The proof of the theorem is not long, using Minkowski's theorem and Fubini's theorem.
- From the proof one also obtains the following formula:

$$V(f_1, f_2, \dots, f_n) = \int_0^1 V\left(K_t(f_1), K_t(f_2), \dots, K_t(f_n)\right) dt.$$

Inequalities

Many well-known geometric inequalities (like Brunn–Minkowski, Alexandrov, Alexandrov–Fenchel) are extended to mixed integrals.

Definition

- For $K \in \mathcal{K}^n$ let K^* be the ball with the same volume as K:

$$K^* = \left(\frac{\mathrm{Vol}K}{\mathrm{Vol}D}\right)^{\frac{1}{n}} \cdot D.$$

- For $f \in \mathrm{QC}\left(\mathbb{R}^n\right)$ the symmetric decreasing rearrangement f^* is defined by $K_t\left(f^*\right) = K_t(f)^*$

The Brunn–Minkowski inequality is equivalent to the inclusion $A^* + B^* \subseteq (A+B)^*$.

Similarly, we have the functional Brunn–Minkowski $f^* \oplus g^* \leq (f \oplus g)^*$.

Inequalities

A standard corollary of the Alexandrov–Fenchel inequality is

$$V(K_1, K_2 \dots, K_n) \geq \left(\prod_{i=1}^{n} \mathrm{Vol}(K_i) \right)^{1/n} = V(K_1^*, K_2^*, \dots, K_n^*).$$

This extends for example the isoperimetric inequality:

$$S(K) = \frac{1}{n} V(K, K, \dots, K, D) \geq \frac{1}{n} \mathrm{Vol}(K)^{\frac{n-1}{n}} \mathrm{Vol}(D)^{\frac{1}{n}}.$$

For quasi-concave functions we have the similar $V(f_1, f_2 \dots, f_n) \geq V(f_1^*, f_2^*, \dots, f_n^*)$.

Define a particular case of mixed integrals:

$$W_i(f) = V(\underbrace{f, f, \dots, f}_{n-i \text{ times}}, \underbrace{\mathbf{1}_D, \mathbf{1}_D, \dots, \mathbf{1}_D}_{i \text{ times}}),$$

the i-th quermassintegral of f. This notion was independently studied by Bobkov, Colesanti and Fragalà.

More on quermassintegrals

If $f \in \mathrm{LC}\left(\mathbb{R}^n\right)$ and $K \in \mathcal{K}^n$, then for every $\lambda > 0$ we have

$$\left[f \oplus \left(\lambda \odot \mathbf{1}_K\right)\right](x) = \left[f \star \left(\lambda \cdot \mathbf{1}_K\right)\right](x) = \sup_{y \in \lambda K} f(x-y)$$

and the same is true for the α-sum on α-concave functions.

In particular this implies:

Theorem

The function $\lambda \mapsto \int f \star \left(\lambda \cdot \mathbf{1}_D\right)$ is polynomial with non-negative coefficients.

This was observed by Bobkov, Colesanti and Fragalà. In the case $f = \mathbf{1}_K$ this polynomial is the Steiner polynomial of K, so we call it the Steiner polynomial of f. Its coefficients are (up to normalization) the quermassintegrals. It reflects the polynomiality of volume of ε-neighborhoods of K, and in our generalization of f.

More on quermassintegrals

Theorem (Bobkov–Colesanti–Fragalà)

For $f, g \in \mathrm{LC}(\mathbb{R}^n)$, $0 < \lambda < 1$ *and* $i = 0, 1, \ldots, n-1$ *we have*

$$W_i\left(\lambda \cdot f \star (1-\lambda) \cdot g\right) \geq W_i(f)^{\lambda} W_i(g)^{1-\lambda}$$

This generalizes the (very non-trivial) Brunn–Minkowski type inequality for mixed volumes:

$$W_i(\lambda K + (1-\lambda)\,T) \geq W_i(K)^{\lambda} W_i(T)^{1-\lambda}$$

($i = 0$ is the standard Brunn–Minkowski inequality)

The case $i = 0$ in the theorem is the Prékopa–Leindler inequality. The full theorem they proved deals with α-concave functions as well, and generalizes the Borell–Brascamp–Lieb inequalities (which again corresponds to $i = 0$).

More on quermassintegrals

Another very nice inequality proved by Bobkov, Colesanti and Fragalà is the following:

Theorem

For $f \in \mathrm{QC}(\mathbb{R}^n)$, *and* $0 \leq i \leq k \leq n-1$ *we have*

$$W_k(f) \geq c \cdot W_i(f^p)^{1/p},$$

where $p = \frac{n-i}{n-k}$ *and* $c = |D_n|^{1-\frac{1}{p}}$.

For $f = \mathbf{1}_K$ one recovers the Alexandrov inequalities

$$\left(\frac{W_k(K)}{W_k(D_n)}\right)^{\frac{1}{n-k}} \geq \left(\frac{W_i(K)}{W_i(D_n)}\right)^{\frac{1}{n-i}}$$

Alexandrov inequalities for log-concave functions

Theorem (Milman–Rotem)

Define $g(x) = e^{-|x|}$*. For every log-concave function* f *with* $\max f = 1$ *and every integers* $0 \leq k < m < n$ *we have*

$$\left(\frac{W_k(f)}{W_k(g)}\right)^{\frac{1}{n-k}} \leq \left(\frac{W_m(f)}{W_m(g)}\right)^{\frac{1}{n-m}},$$

with equality if and only if $f(x) = e^{-c|x|}$ *for some* $c > 0$*.*

Both the assumption that f is log-concave and the assumption that $\max f = 1$ are essential.

(And again, some fundamental and non-trivial constructions are consequences of some very elementary, basic properties.)

Introduction

Let the classical **Fourier transform** $\mathbb{F}$ on $\mathbb{R}^n$ be

$$\mathbb{F}f = \int e^{-2\pi i\langle x,y\rangle} f(y)dy.$$

Let S be the Schwartz class of "rapidly" decreasing (infinitely smooth) functions on $\mathbb{R}^n$.

Theorem (Artstein, Faifman, Milman)

Assume we are given a bijective transform $\mathcal{F} : S \to S$, s.t. $\forall f, g \in S$ we have

$$\mathcal{F}(f \cdot g) = \mathcal{F}f * \mathcal{F}g.$$

Then $\exists$ diffeomorphism $\omega : \mathbb{R}^n \to \mathbb{R}^n$ s.t. either $\forall f \in S$, $\mathcal{F}f = \mathbb{F}(f \circ \omega)$ or $\forall f \in S$, $\mathcal{F}f = \overline{\mathbb{F}(f \circ \omega)}$. Real linearity and continuity of $\mathcal{F}$ is the automatic consequence.

Previous versions contained more conditions and were proved jointly with S. Alesker. Joining these results with the previous theorem we may state that if $\mathcal{F} : S \to S$ s.t. $\forall f, g \in S$,

$$\mathcal{F}(f \cdot g) = \mathcal{F}f * \mathcal{F}g,$$

$$\mathcal{F}(f * g) = \mathcal{F}f \cdot \mathcal{F}g,$$

then $\exists$ linear $A \in GL_n$, $|\det(A)| = 1$, s.t. either

$$\forall f \in S, \quad \mathcal{F}f = \mathbb{F}(f \circ A)$$

or

$$\forall f \in S, \quad \mathcal{F}f = \overline{\mathbb{F}(f \circ A)}.$$

There was one "weak" point in all the previous results on recovering some constructions in an essentially unique way by very elementary properties.

Actually, this was the study of identity, the rigidity of identity.

Indeed, because the Fourier transform $\mathcal{F}$ is already known to us, we just apply it to *a priori* unknown transform T (which has the property $T(f \cdot g) = Tf * Tg$) and then

$$(\mathcal{F}T)(f \cdot g) = (\mathcal{F}Tf) \cdot (\mathcal{F}Tg)\,.$$

So we need to show that the map preserving product is, essentially, *identity*!!

We want a different example. And here a new series of results of purely analytic nature starts.

The chain rule operator equation and derivation construction

Chain rule for $f, g \in C^1(\mathbb{R}) : D(f \circ g) = (Df) \circ g \cdot Dg$.
Let $T : C^1(\mathbb{R}) \to C(\mathbb{R})$ be an operator satisfying the functional equation

$$T(f \circ g)(x) = (Tf)(g(x)) \cdot (Tg)(x); \quad f, g \in C^1(\mathbb{R}),\ x \in \mathbb{R}. \tag{1}$$

Which operators T satisfy (1)?
Examples:

a) $p > 0, \quad (Tf)(x) = |f'(x)|^p$ and
$(Tf)(x) = \operatorname{sgn} f'(x)\ |f'(x)|^p$ both satisfy (1).

b) Let $H \in C(\mathbb{R}),\ H > 0$. Define $(Tf)(x) = H(f(x))/H(x)$.
Then T satisfies (1).

c) Consider $T : C^1(\mathbb{R}) \to C(\mathbb{R}), \quad (Tf)(x) =$
$\left\{ \begin{array}{cc} f'(x) & f \in C^1(\mathbb{R}) \text{ bijective} \\ 0 & \textit{else} \end{array} \right\}$. Then T satisfies (1).

Solutions of the chain rule operator equation

Multiplying two solutions of the chain rule yields again a solution. $T : C^k(\mathbb{R}) \to C(\mathbb{R})$ is C^k-non-degenerate if $T|_{C_b^k(\mathbb{R})} \neq 0$ where $C_b^k(\mathbb{R})$ are the (half-) bounded functions in $C^k(\mathbb{R})$. Here $k \in \mathbb{N}_0$.

Theorem [Artstein-König-Milman]

Assume $T : C^k(\mathbb{R}) \to C(\mathbb{R})$ satisfies the chain rule

$$T(f \circ g) = (Tf) \circ g \cdot Tg; \quad f, g \in C^k(\mathbb{R})$$

for $k \in \mathbb{N}_0$ and that T is C^k-non-degenerate. Then there is $p \geq 0$ and $H \in C_{>0}(\mathbb{R})$ such that for any $f \in C^k(\mathbb{R})$

$$(Tf)(x) = \frac{H(f(x))}{H(x)} \, |f'(x)|^p \{\text{sgn } f'(x)\},$$

and this also holds for $k = \infty$, i.e. $f \in C^\infty(\mathbb{R})$.

Remarks

1. By this formula T automatically extends to $C^1(\mathbb{R})$. So the natural Dom T is C^1.

2. Let us add the following normalization condition:

$$T(-2 \cdot \mathrm{Id}) = -2\,.$$

Then the unique solution of the chain rule operator equation is the derivative $T(f) = f'$.

(Now we don't see it as a study of the identity map.)

3. The result is also valued for operators acting on the family of polynomes $P(\mathbb{R}^n)$, however, ONLY if we add some (although weak) pointwise continuity assumption (actually, at one point 0). Under such assumptions the statement is true also for the class of entire functions. Note, we don't have **any** continuity assumptions in the theorem.

However, on so small a family as Polynomials, the statement without such an assumption is not true (example: $T(P) = \operatorname{degr} P$, and many others. An interesting question is to describe all $T : P \to P$ s.t. $T(P \circ Q) = TP \cdot TQ$).

The chain rule is a very natural (and very elementary) way to define a non-trivial construction – derivation.

Steps in the Proof

a) Show "localization": There is $F : \mathbb{R}^{k+2} \to \mathbb{R}$ such that $Tf(x) = F(x, f(x), ..., f^{(k)}(x))$ for all $f \in C^k(\mathbb{R})$ and $x \in \mathbb{R}$.

b) Analyze the structure of the representing function F : $F(x, \alpha_0, ..., \alpha_k) = \frac{H(\alpha_0)}{H(x)} K(\alpha_1)$, K multiplicative, F independent of $\alpha_2, ..., \alpha_k$ if $k \geq 2$.

c) Show the measurability and then the continuity of the coefficient functions occurring in F.

Stability of the Chain Rule

$T : C^1(\mathbb{R}) \to C(\mathbb{R})$ is locally non-degenerate if $\forall$ open interval $J \subset \mathbb{R}$, $\forall x \in J$, $\exists g \in C^1(\mathbb{R})$, $y \in \mathbb{R}$, s.t. $g(y) = x$, $\mathrm{Im}(g) \subset J$ and $Tg(y) \neq 0$.

Theorem (König-Milman)

Fix $T : C^1(\mathbb{R}) \to C(\mathbb{R})$ *and* $B : \mathbb{R}^3 \to \mathbb{R}$ *such that* $\forall f, g \in C^1$ *and* $\forall x \in \mathbb{R}$

$$T(f \circ g)(x) = Tf \circ g(x) \cdot Tg(x) + B(x, f \circ g(x), g(x)).$$

Assume that T *is locally non-degenerate and* Tf *depends non-trivially on* f'.

Then $B = 0$ *(and* T *satisfies the chain rule).*

Even more rigidity

Consider the "chain rule inequality"

$$T\left(f\circ g\right)\leq\left(Tf\right)\circ g\cdot Tg \qquad (*)$$

for $T: C^1(\mathbb{R})\to C(\mathbb{R})$, $\mathrm{Dom}(T)=C^1(\mathbb{R})$.

Assume that T satisfies the following:

- non-degeneration: $\forall$ open interval $I\subset\mathbb{R}$, $\forall x\in I$, $\exists g\in C^1(\mathbb{R})$ s.t. $g(x)=x$, $\mathrm{Im}(g)\subset I$ and $Tg(x)>1$.
- T is pointwise continuous: $\forall f, f_n\in C^1(\mathbb{R})$ s.t $f_n\to f$, $f_n'\to f'$ uniformly on compact subsets we have $\left(Tf_n\right)(x)\to\left(Tf\right)(x)$ pointwise for all $x\in\mathbb{R}$.

Theorem (König-Milman)

For T as above assume also $\exists x \in \mathbb{R}$ s.t $T(-Id)(x) < 0$. Then $\exists H \in C(\mathbb{R})$, $H > 0$, $\exists p > 0$ and $A \geq 1$ s.t

$$Tf = \begin{cases} \frac{H \circ f}{H} \left| f' \right|^p & \text{for } f' \geq 0 \\ -A \frac{H \circ f}{H} \left| f' \right|^p & \text{for } f' < 0. \end{cases}$$

Note:

- For $A = 1$, T satisfies the chain rule equation: we have equality in $(*)$.
- For both f and g non-decreasing we automatically have equality in $(*)$.
- Actually, the same is true if for some $C > 0$

$$T\left(f \circ g\right) \leq C \cdot \left(Tf\right) \circ g \cdot Tg$$

  and even much more generally (the answer is slightly modified).

Theorem (König-Milman)

Assume that $T : C^1(\mathbb{R}) \to C(\mathbb{R})$ *is pointwise continuous and non-degenerate. Suppose further that there is a function* $S : \mathbb{R}^3 \to \mathbb{R}$ *such that the perturbed chain operator inequality*

$$|T(f \circ g)(x) - (Tf)(g(x)) \cdot (Tg)(x)| \leq S(x, (f \circ g)(x), g(x))$$

holds for all $f, g \in C^1(\mathbb{R})$ *and all* $x \in \mathbb{R}$. *Then there are* $p > 0$ *and a positive continuous function* $H : \mathbb{R} \to \mathbb{R}_{>0}$ *such that for all* $f \in C^1(\mathbb{R})$ *and all* $x \in \mathbb{R}$

$$Tf(x) = \frac{H(f(x))}{H(x)} sgn\ f'(x)|f'(x)|^p.$$

This implies that we may choose $S = 0$, *i.e. that we have equality*

$$T(f \circ g)(x) = (Tf)(g(x)) \cdot (Tg)(x), \qquad f, g \in C^1(\mathbb{R}), x \in \mathbb{R}.$$

The following classically sound functional statement is used:

We say that $K : \mathbb{R} \to \mathbb{R}$ is submultiplicative if

$$K(\alpha\beta) \leq K(\alpha) K(\beta), \qquad \forall \alpha, \beta \in \mathbb{R}.$$

Theorem (König-Milman)

Let K be submultiplicative, measurable and continuous at 0 and at 1. Assume $K(-1) < 0 < K(1)$. Then $\exists p > 0$ s.t.

$$K(\alpha) = \begin{cases} \alpha^p & \text{for } \alpha \geq 0 \\ -A\,|\alpha|^p & \text{for } \alpha < 0 \end{cases}$$

(and $K(-1) = -A \leq -1$).

[every assumption in the theorem is needed]

As a corollary, K must be multiplicative on $\mathbb{R}^+$.

Second order chain rule formulas

(We will see a new phenomenon here)

For $f, g \in C^2(\mathbb{R})$ one has

$$D^2(f \circ g) = D^2 f \circ g \cdot g'^2 + f' \circ g \cdot D^2 g. \qquad (*)$$

We study the solutions of a generalized operator functional equation:

Let $k \geq 2$ and $T : C^k(\mathbb{R}) \to C(\mathbb{R}),\ A_1, A_2 : C^{k-1}(\mathbb{R}) \to C(\mathbb{R})$ be such that

$$T(f \circ g) = Tf \circ g \cdot A_1 g + A_2 f \circ g \cdot Tg; \quad f, g \in C^k(\mathbb{R}). \qquad (2)$$

A is *isotropic* if it commutes with all shift operators.

(T, A) is C^k*-non-degenerate* if for all open sets $J \subset \mathbb{R}$ and $x \in J$ there are $y_1, y_2 \in \mathbb{R}$ and $g_1, g_2 \in C^k(\mathbb{R})$ such that $Im(g_i) \subset J$, $g_1(y_1) = x = g_2(y_2)$ and $z_i = (Tg_i(y_i), Ag_i(y_i)) \in \mathbb{R}^2$ are linearly independent for $i = 1, 2$.

Note: For $A_1 = A_2 = \frac{1}{2} T$, (2) is just the chain rule operator equation (1). Non-degeneration excludes this.

There are very few operators $A_1; A_2$ which lead to (2) with non-trivial solutions. Let us list all of them by Dom T:

(0) Dom $T = C$: $A_1 f = A_2 f \equiv \mathbb{1}$ and "tuning" is not needed, and $Tf = H \circ f - H$ for some $H(x) \in C(\mathbb{R})$.

(1) Dom $T = C^1$: let $p > 0$, $q > 0$, $\exists$ three families

- 1a. $A_1 g = A_2 g = |g'|^p \{\operatorname{sgn} g'\}$
- 1b. $A_1 g = |g'|^q \{\operatorname{sgn} g'\}$; $A_2 g = |g'|^p [\operatorname{sgn} g']$
- 1c. $A_1 g = A_2 g = |g'|^p \cos(d(x) \ln |g'|) \{\operatorname{sgn} g'\}$ for some $d(x) \in C(\mathbb{R})$.

(2) Dom $T = C^2$: $\quad p \geq 1$
$A_2 g = |g'|^p \{\operatorname{sgn} g'\}$; $A_1 g = g' \cdot A_2 g$.

(3) Dom $T = C^3$: $\quad p \geq 2$
$A_2 g = |g'|^p \{\operatorname{sgn} g'\}$; $A_1 g = g'^2 A_2 g$.

That is all. No other combination of A_1 and A_2 may lead to any (non-trivial) solutions. We know the formulas for solutions in any of these cases. Of course, solutions in the cases (0) Dom $T = C$ and (1) Dom $T = C'$ are also solutions in the remaining cases. Let me describe the additional solutions for the cases (2) and (3).

Case Dom $T = C^2$. $\exists\, p \geq 1$, $c(x) \not\equiv 0$ and $H(x) \in C(\mathbb{R})$ s.t.

$$Tf = \left(cf'' + [H \circ f \cdot f' - H] \cdot f' \right) \cdot |f'|^{p-1} \{\operatorname{sgn} f'\}$$

(So, for $H = 0$; $p = 1$, the answer is $Tf = cf''$.)

To describe the case Dom $T = C^3$, we need to introduce the Schwarzian derivative S of a $C^3(\mathbb{R})$-function f:

$$Sf = \left(\frac{f''}{f'} \right)' - \frac{1}{2} \left(\frac{f''}{f'} \right)^2 = \frac{f'''}{f'} - \frac{3}{2} \left(\frac{f''}{f'} \right)^2 .$$

Note $f'^2 Sf = f''' f' - \frac{3}{2} f''^2$ is also defined if $f' = 0$. S satisfies the composition rule

$$S(f \circ g) = Sf \circ g \cdot g'^2 + Sg.$$

The kernel of S consists of the fractional linear transformations $f(x) = \frac{ax+b}{cx+d}$.
Hence for such f, S is invariant: $S(f \circ g) = Sg$. Although the Schwarzian derivative is mainly important in complex analysis, e.g. for conformal mappings, univalent functions and complex dynamics, we only study real versions of S.

Now, for Dom $T = C^3(\mathbb{R})$ (in addition to the previous solutions) we have (for $c(x)$ and $H(x) \subset C(\mathbb{R})$, $p \geq 2$)

$$Tf = (c \cdot Sf + H(f)f'^2 - H)|f'|^p\{\operatorname{sgn} f'\}$$

(So both f'' and f''' are coming through S.)

Note that there is no solution of (2) depending on the fourth or higher derivative of f (!)

So the natural domains of solutions of (2) are $C^k(\mathbb{R})$ for $k \in \{0, 1, 2, 3\}$.

Two initial conditions may determine the form of T, e.g.

$T(x^2) = 2,\ T(x^3) = 6x$ yields

$$Tf = f'', \qquad\qquad A_1 f = f'^2,\ A_2 f = f'$$

and $T(x^2) = -6,\ T(x^3) = -36x^2$ implies the solution

$$Tf = f'^2 Sf, \qquad\qquad A_1 f = f'^4,\ A_2 f = f'^2.$$

(4) The solutions for A_1, A_2 depend only on f'. Therefore the natural domain for A_1 and A_2 is $C^1(\mathbb{R})$.

Extra information

The solutions in the case C^1

1a. $(Tg)(x) =$
$(c \ln |g'(x)| + H(g(x)) - H(x))|g'(x)|^p\{\operatorname{sgn} g'(x)\}$;

1b. $(Tg)(x) =$
$H(g(x))|g'(x)|^q[\operatorname{sgn} g'(x)] - H(x)|g'(x)|^p\{\operatorname{sgn} g'(x)\}$;

1c. $(Tg)(x) = c|g'(x)|^p \sin(d \ln |g'(x)|)\{\operatorname{sgn} g'(x)\}$.