Mathematical
Inequalities
& Applications
Volume 21, Number 1 (2018), 235–249 doi:10.7153/mia-2018-21-18

ON THE LEIBNIZ RULE FOR RANDOM VARIABLES

ZOLTÁN LÉKA

(*Communicated by J.-C. Bourin*)

Abstract. We prove a Leibniz-type inequality for the spread of (real-valued) random variables in terms of their L^p-norms. The result is motivated by the Kato–Ponce inequality and Rieffel's Leibniz property.

1. Introduction

For differentiable functions on the real line the Leibniz rule and Hölder's inequality provides us with a simple way to have various estimates of the L^p norms of derivatives of products. The Leibniz inequality and its variants have recently appeared in M. Rieffel's fundamental work on the theory of quantum metric spaces, see e.g. [27, 28, 29, 30]. His study was motivated by an urgent need for a non-commutative (quantum) analogue of the Gromov–Hausdorff distance between compact metric spaces. In [18], [19] a novel approach to the long-standing problem of finding a proper metric between compact C^*-metric spaces has been offered by F. Latrémolière. For a thorough survey on these types of results, we refer the reader to [20].

We recall that a seminorm L defined on a unital normed algebra $(\mathscr{A}, \|\cdot\|, \mathbf{1}_{\mathscr{A}})$ is strongly Leibniz if

(i) $L(\mathbf{1}_{\mathscr{A}}) = 0$,

(ii) (Leibniz inequality) $L(ab) \leqslant \|a\|L(b) + \|b\|L(a)$ for all $a, b \in \mathscr{A}$,

(iii) $L(a^{-1}) \leqslant \|a^{-1}\|^2 L(a)$ whenever $a \in \mathscr{A}$ is invertible.

One of the simplest (but non-trivial) example of such seminorms is the standard deviation defined in ordinary and non-commutative probability spaces as well, see [30]. Interestingly, several examples of strongly Leibniz seminorms can be defined through derivations taking their values in Hilbert bimodules or as commutator norms, see e.g. [30, Proposition 1.1], [32, Proposition 8], [29, Example 11.5].

On the other hand we notice that the Leibniz inequality appears in the theory of symmetric Dirichlet forms on real L^2 function spaces [7], [11]. In fact, it is often

Mathematics subject classification (2010): Primary 46N30, 60E15, Secondary 26A51, 60A99.
Keywords and phrases: Kato–Ponce inequality, Leibniz inequality, Laplacian, majorization, random variable.
This study was supported by the Marie Curie IF Fellowship, Project 'Moments', Grant no. 653943 and by the Hungarian Scientific Research Fund (OTKA) grant no. K104206.

© ELEMENT, Zagreb
Paper MIA-21-18

used to establish that bounded functions in the domain form an algebra, see e.g. [7, Corollary 3.3.2]. We just recall that the standard deviation is itself a Dirichlet form over a probability space.

From the viewpoint of differential operators, Leibniz-type rules have intensively been studied in the theory of non-linear PDEs as well. Let $(-\Delta)^\alpha$ be the fractional Laplace operator defined as the Fourier multiplier

$$\widehat{(-\Delta)^\alpha f}(\xi) = |\xi|^{2\alpha}\hat{f}(\xi), \quad \xi \in \mathbb{R}^n,$$

for any f in the Schwartz space $\mathscr{S}(\mathbb{R}^n)$. Then the Kato–Ponce inequality or the fractional Leibniz rule in its simplest form asserts the following. Let $1 < r, p_1, q_1, p_2, q_2 < \infty$ such that $\frac{1}{r} = \frac{1}{p_1} + \frac{1}{q_1} = \frac{1}{p_2} + \frac{1}{q_2}$. Given $0 < \alpha \leqslant 1$, for all $f, g \in \mathscr{S}(\mathbb{R}^n)$, one has

$$\|(-\Delta)^\alpha(fg)\|_r \leqslant C(\|g\|_{p_1}\|(-\Delta)^\alpha f\|_{q_1} + \|f\|_{p_2}\|(-\Delta)^\alpha g\|_{q_2}),$$

where $C = C_{n,\alpha,p_1,q_1,p_2,q_2,r} > 0$ is a constant depending only on (n, α, p_i, q_i, r).

The result and related commutator estimates turned up in the fundamental works of Kato–Ponce [15], Christ–Weinstein [10] and Kenig–Ponce–Vega [16]. Nowadays the Kato–Ponce inequalities and their extensions have an extensive literature, see e.g. [13], [24]. For a detailed and thorough exposition of these types of results, we refer the reader to [12] and [23]. A heuristic approach to the inequality in [24] briefly says that if f oscillates more rapidly than g, then g is essentially constant with respect to f and so $(-\Delta)^\alpha(fg)$ behaves like $(-\Delta)^\alpha(f)g$. The similar statement holds if g oscillates more rapidly then f. However, the rigorous proof is based on advanced techniques of harmonic analysis.

Leibniz-type rules and related bilinear Poincaré–Sobolev inequalities have recently been proved in [3], where the interested reader may find further references.

In the spirit of the Kato–Ponce inequality, our goal is to prove a Leibniz-type rule for random variables and their dispersions around the expected values, but with a strict constant which is surprisingly independent of the Hölder exponents. Namely, the main theorem of the paper is the following.

THEOREM 1. *Let $(\Omega, \mathscr{F}, \mu)$ be a probability space. For any real $f, g \in L^\infty(\Omega, \mu)$, one has*

$$\|fg - \mathbb{E}(fg)\|_r \leqslant \|f\|_{p_1}\|g - \mathbb{E}g\|_{q_1} + \|g\|_{p_2}\|f - \mathbb{E}f\|_{q_2},$$

where $1 \leqslant r, p_1, p_2, q_1, q_2 \leqslant \infty$ and $\frac{1}{r} = \frac{1}{p_1} + \frac{1}{q_1} = \frac{1}{p_2} + \frac{1}{q_2}$.

Motivated by Rieffel's work on Leibniz seminorms, particular cases of the above theorem have already been settled in [2] and [21]. The proof here is based on discretization. In fact, we approximate the centered quantities by means of special Laplacian matrices and their products with vectors. We shall use an elementary decomposition of centered products and, instead of interpolation methods, we shall make a careful application of a Hölder-type inequality. Throughout the paper we shall use the concept of random variable but the proofs herein are built on convex analysis and majorization theory, and not on probabilistic methods.

Following the parallel between differential operators on products and centered random variables, in the last section of the paper we shall prove a simple 'chain rule' for the L^p-norms of compositions with monotone Lipschitz functions.

2. Symmetric norms on $\mathbb{R}^n$

First, we collect a few properties of symmetric norms. Let x be an n-dimensional real vector. We write $x^{\downarrow}$ for the non-increasing rearrangement of x. We recall that a norm $\|\cdot\|$ on $\mathbb{R}^n$ is symmetric if it is invariant under the permutation of the components and their sign changes. For instance, the ℓ_p norms given by $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$ $(1 \leqslant p < \infty)$ and the vector k-norms $\|x\|_{(k)} = \sum_{i=1}^k |x_i|^{\downarrow}$ $(1 \leqslant k \leqslant n)$ are symmetric. Basic properties of symmetric norms are the absolute property, $\|x\| = \||x|\|$, and the monotonicity, i.e. $\|x\| \leqslant \|y\|$ if $|x| \leqslant |y|$ (see [4, Section 2]).

We should point out that von Neumann [25] proved that there is a one-to-one correspondence between the symmetric norms on $\mathbb{R}^n$ and unitarily invariant norms on the space of $n \times n$ matrices. It is also appropriate to mention here that the vector k-norms are extremal in the following sense: the well–known Ky Fan Dominance Theorem (see [4] or [31, Chapter 15]) says that for all symmetric norms $\|\cdot\|$ the inequality $\|x\| \leqslant \|y\|$ holds if and only if $\|x\|_{(k)} \leqslant \|y\|_{(k)}$ is satisfied for all vector k-norms. Note that we may obtain an infinite family of inequalities from a finite one.

Let us now introduce the concept of weak majorization or submajorization relation denoted by $\prec_w$. If x, y are real n-dimensional vectors then

$$x \prec_w y \quad \text{if and only if} \quad \sum_{i=1}^k x_i^{\downarrow} \leqslant \sum_{i=1}^k y_i^{\downarrow} \text{ for all } k = 1, \dots, n.$$

Following the previous notations, $|x| \prec_w |y|$ precisely holds if $\|x\|_{(k)} \leqslant \|y\|_{(k)}$ for all $k = 1, \dots, n$. In order to prove our main theorem, we shall preliminarily need the concept of weighted vector k-norms as well. We say that a vector $w = (w_1, w_2, \dots, w_n)^T \in \mathbb{R}^n$ is a weight vector if its coordinates are positive and decreasingly ordered $w_1 \geqslant w_2 \geqslant \dots \geqslant w_n > 0$. Then the weighted vector k-norm is given by

$$\|x\|_{(k)}^w = \sum_{i=1}^k w_i |x_i|^{\downarrow}.$$

Recently, these norms have appeared in matrix optimization problems related to the Ky Fan k-norms [34] and robust linear optimization with special weights $w_1 = \dots = w_{n-1} \geqslant w_n > 0$, see [6].

We recall that the dual norm of any norm $\|\cdot\|$ on $\mathbb{R}^n$ can be given by $\|x\|_* = \max\{\langle x, y\rangle \colon \|y\| \leqslant 1\}$, where $\langle \cdot, \cdot \rangle$ stands for the usual inner product on $\mathbb{R}^n$. It is simple to see that if $\|\cdot\|$ is symmetric then $\|\cdot\|_*$ has the same property.

Our first lemma gives an expression for the dual norm of $\|\cdot\|_{(k)}^w$. The proof here closely follows the proof of [6, Proposition 2].

LEMMA 1. *The dual norm of the weighted vector k-norm is given by*

$$\|x\|^w_{(k)^*} = \max\left\{\frac{\|x\|_{(1)}}{w_1}, \frac{\|x\|_{(2)}}{w_1+w_2}, \ldots, \frac{\|x\|_{(k-1)}}{w_1+\ldots+w_{k-1}}, \frac{\|x\|_{(n)}}{w_1+\ldots+w_k}\right\}.$$

Proof. The rearrangement inequality (e.g. see [31, p. 208]) and the absolute property of the symmetric norm $\|\cdot\|^w_{(k)}$ imply that

$$\begin{aligned}\|x\|^w_{(k)^*} &= \max\{\langle |x|, y\rangle \colon \|y\|^w_{(k)} \leqslant 1\}\\ &= \max\{\langle |x|^{\downarrow}, y\rangle \colon \|y\|^w_{(k)} \leqslant 1 \text{ and } 0 \leqslant y_1, \ldots, y_n\}.\end{aligned}$$

On the other hand, it is simple to see that $\|y\|^w_{(k)}$ $(0 \leqslant y \in \mathbb{R}^n)$ can be written as the optimal solution of the linear program:

$$\begin{aligned}\max \sum_{i=1}^n u_i y_i & \\ \text{s.t. } \sum_{i=1}^j u_i \leqslant \sum_{i=1}^{\min(j,k)} w_i & \quad j=1,\ldots,n,\\ u_j \geqslant 0, & \quad j=1,\ldots,n.\end{aligned}$$

From the strong duality of LP problems we may find that

$$\begin{aligned}\|y\|^w_{(k)} = \min \sum_{i=1}^n \sum_{j=1}^{\min(i,k)} s_i w_j & \\ \text{s.t. } \sum_{i=j}^n s_i \geqslant y_j & \quad j=1,\ldots,n,\\ s_j \geqslant 0, & \quad j=1,\ldots,n.\end{aligned}$$

Thus $\|y\|^w_{(k)} \leqslant 1$ $(y \in \mathbb{R}^n_+)$ if and only if

$$\sum_{i=1}^n \sum_{j=1}^{\min(i,k)} s_i w_j \leqslant 1 \text{ and } \sum_{i=j}^n s_i \geqslant y_j, s_j \geqslant 0,\ j=1,\ldots,n$$

is feasible. Hence we get

$$\begin{aligned}\|x\|^w_{(k)^*} = \max \sum_{i=1}^n |x_i|^{\downarrow} y_i & \\ \text{s.t. } \sum_{i=1}^n \sum_{j=1}^{\min(i,k)} s_i w_j \leqslant 1, & \\ \sum_{i=j}^n s_i \geqslant y_j & \quad j=1,\ldots,n\\ s_j, y_j \geqslant 0, & \quad j=1,\ldots,n.\end{aligned}$$

Applying again the strong duality of LP problems, one has

$$\begin{aligned}\|x\|^w_{(k)^*} = \min\theta \\ \text{s.t.} \quad \sum_{i=1}^{\min(j,k)} \theta w_i - \sum_{i=1}^{j} \alpha_i \geqslant 0, \qquad & j=1,\dots,n,\\ \alpha_j \geqslant |x_j|^{\downarrow} \qquad & j=1,\dots,n,\\ \theta \geqslant 0, \alpha_j \geqslant 0, \qquad & j=1,\dots,n.\end{aligned}$$

Therefore, the equality

$$\|x\|^w_{(k)^*} = \max\left\{\frac{\|x\|_{(1)}}{w_1}, \frac{\|x\|_{(2)}}{w_1+w_2}, \dots, \frac{\|x\|_{(k-1)}}{w_1+\dots+w_{k-1}}, \frac{\|x\|_{(n)}}{w_1+\dots+w_k}\right\}$$

immediately follows. □

Throughout the paper let e_i stand for the ith standard basis element of $\mathbb{R}^n$. Now we may get some description of the extreme points of the unit ball $\mathfrak{B}^w_{(k)} := \{x \in \mathbb{R}^n \colon \|x\|^w_{(k)} \leqslant 1\}$. These points will be denoted by $\operatorname{ext}\mathfrak{B}^w_{(k)}$.

LEMMA 2. *For the extreme points of* $\mathfrak{B}^w_{(k)}$, *one has*

$$\operatorname{ext}\mathfrak{B}^w_{(k)} \subseteq \left\{\sum_{i\in S} \frac{\pm e_i}{\sum_{j=1}^{\min(k,|S|)} w_j} \colon S \subseteq [n] \text{ and } 1 \leqslant |S| \leqslant k-1 \text{ or } |S| = n\right\}.$$

Proof. Let $\mathscr{C}$ stand for the points of the right-hand side in the above inclusion, and let us consider its convex hull $\mathfrak{B}_{\mathscr{C}} = \operatorname{conv}\mathscr{C}$. Then it is immediate that $\operatorname{ext}\mathfrak{B}_{\mathscr{C}} \subseteq \mathscr{C}$. Furthermore, we get

$$\mathfrak{B}^w_{(k)^*} = \{x \in \mathbb{R}^n \colon |\langle x,y\rangle| \leqslant 1 \text{ for } \|y\|^w_{(k)} \leqslant 1\} = \{x \in \mathbb{R}^n \colon |\langle x,y\rangle| \leqslant 1 \text{ for } y \in \mathscr{C}\}.$$

Indeed, the last inclusion $\supseteq$ is clear. On the other hand, for any $1 \leqslant j \leqslant n$, one has that $\|x\|_{(j)} = \max\{|\langle x,y\rangle| \colon y_i \in \{0,\pm 1\} \text{ and } \|y\|_1 = j\}$, hence Lemma 1 implies the inclusion $\subseteq$ in the last equality. Thus the polar of $\mathfrak{B}_{\mathscr{C}}$ equals the unit ball $\mathfrak{B}^w_{(k)^*}$. The Bipolar Theorem [31, Theorem 5.5] now gives that $\mathfrak{B}_{\mathscr{C}} = \mathfrak{B}^w_{(k)}$, so $\operatorname{ext}\mathfrak{B}^w_{(k)} \subseteq \mathscr{C}$. □

The expression in Lemma 2 might be a bit misleading if $k=1$. Then the set of the right-hand side is given by the case $|S| = n$. We also remark that it might happen that the right-hand side of the above inclusion is strictly larger than $\operatorname{ext}\mathfrak{B}^w_{(k)}$. Indeed, if w is the constant 1 vector, then it is simple to show that

$$\operatorname{ext}\mathfrak{B}^1_{(k)} = \frac{1}{k}\{-1,1\}^n \cup \{\pm e_i\} \qquad (1 < k < n). \tag{2.1}$$

Actually, this now follows from the fact that the dual norm of the vector k-norm is $\max\left(\|\cdot\|_\infty, \frac{\|\cdot\|_1}{k}\right)$.

Applying a completely different method, one can precisely describe the extreme points of the unit ball in $\mathbb{R}^n$ with respect to the weighted vector k-norms [9, Theorem 3.1] (called c-norms in [9]). However, for our purposes the inclusion of Lemma 2 is enough.

3. Proof of Theorem 1

Let x be an n-dimensional vector. We introduce an $n \times n$ symmetric matrix Θ_x associated to x with zero row and column sums defined by

$$(\Theta_x)_{ij} = \begin{cases} \frac{1}{2n}(x_i + x_j) & \text{if } i \neq j \\ -\sum_{k:k\neq i}(\Theta_x)_{ik} & \text{if } i = j. \end{cases}$$

For any vectors x and y, let $x \otimes y$ denote the rank-1 matrix xy^T. We will need the following simple but useful lemma.

LEMMA 3. *Let us consider two vectors x and y in $\mathbb{R}^n$. For any symmetric norm $\|\cdot\|$ on $\mathbb{R}^n$, we have*

$$\|(\Theta_x - n^{-1}\mathbf{1} \otimes x)y\| \leqslant \||x|^{\downarrow}|y|^{\downarrow}\|.$$

Proof. Relying upon the Ky Fan dominance theorem, it is sufficient to prove that

$$|(\Theta_x - n^{-1}\mathbf{1} \otimes x)y| \prec_w |x|^{\downarrow}|y|^{\downarrow};$$

that is, the statement of the lemma holds with the vector k-norms.

Now fix k. By a continuity argument, we may assume that the components of x are nonzero. Since we are maximizing a convex function over a compact convex polytope, we have

$$\|(\Theta_x - n^{-1}\mathbf{1} \otimes x)v\|_{(k)} \leqslant \||x|^{\downarrow}|v|^{\downarrow}\|_{(k)} = \|v\|_{(k)}^{|x|^{\downarrow}} \quad \text{for all } v \in \mathbb{R}^n$$

if and only if

$$\|(\Theta_x - n^{-1}\mathbf{1} \otimes x)v\|_{(k)} \leqslant 1 \quad \text{for all } v \in \operatorname{ext} \mathfrak{B}_{(k)}^{|x|^{\downarrow}}.$$

In view of Lemma 2, this is equivalent to show for all $l \in \{1, \ldots, k-1\} \cup \{n\}$ that

$$\|(\Theta_x - n^{-1}\mathbf{1} \otimes x)v\|_{(k)} \leqslant \|x\|_{(\min(l,k))}$$

holds for $v \in \{0, \pm 1\}^n, \|v\|_1 = l$. Fix l. By homogeneity, we might have $\|x\|_{(\min(l,k))} = 1$. If we consider the previous inequality as a convex optimization problem, we need to prove that the objective value of

$$\max_v \|(\Theta_x - n^{-1}\mathbf{1} \otimes x)v\|_{(k)} \quad \text{s.t. } v \in \{0, \pm 1\}^n, \|v\|_1 = l, \tag{3.1}$$

is less than or equal to 1. (Of course, if $k = 1$ then $l = n$ must hold.)

Next, note that the objective value of (3.1) is convex in x if x varies on any face of the unit ball $\mathfrak{B}^1_{\min(l,k)}$. The extreme points of any face are readily extreme points of the unit ball. Hence, in view of (2.1), it turns out that it is sufficient to solve (3.1) in the next cases.

Case (a) $1 \leqslant l \leqslant k-1$ and $x = \pm e_{i_0}$. Let us assume that $v_{i_0} = 1$ (if $v_{i_0} = -1$ the proof is similar). Then

$$\begin{aligned}
\|(\Theta_x - n^{-1}\mathbf{1}\otimes x)v\|_{(k)} &\leqslant \|(\Theta_x - n^{-1}\mathbf{1}\otimes x)v\|_{(n)} \\
&= \frac{1}{2n}\sum_{i=1}^{n}\Big|\sum_{j=1}^{n}((x_i - x_j)v_j - (x_i + x_j)v_i)\Big| \\
&= \frac{1}{2n}\left((n+1) - \sum_{j\neq i_0} v_j\right) + \frac{1}{2n}\sum_{i\neq i_0}(1+v_i) \\
&= 1.
\end{aligned}$$

If $v_{i_0} = 0$, we get

$$\begin{aligned}
\|(\Theta_x - n^{-1}\mathbf{1}\otimes x)v\|_{(k)} &\leqslant \frac{1}{2n}\sum_{i=1}^{n}\Big|\sum_{j=1}^{n}((x_i - x_j)v_j - (x_i + x_j)v_i)\Big| \\
&= \frac{1}{2n}\Big|\sum_{j\neq i_0} v_j\Big| + \frac{1}{2n}\sum_{j\neq i_0}|v_j| \\
&\leqslant 1.
\end{aligned}$$

Case (b) $1 \leqslant l \leqslant k-1$ and $x \in \frac{1}{l}\{+1,-1\}^n$. Clearly, we do not decrease the objective value of (3.1) if we allow v to be any vector such that $\|v\|_1 = l$. Hence it is enough to show that

$$\max\{\|(\Theta_x - n^{-1}\mathbf{1}\otimes x)v\|_{(k)} \colon x \in \{-1,1\}^n, v \in \{\pm e_1, \ldots, \pm e_n\}\} = 1.$$

Notice that the above expression now is independent of l. Let i_0 denote the index of the only non-zero component of v, and let $S_{x,-1}, S_{x,+1}$ be the support sets of x taking values -1 and $+1$, respectively. We can assume that $x_{i_0} = 1$. Then

$$\begin{aligned}
\|(\Theta_x - n^{-1}\mathbf{1}\otimes x)v\|_{(k)} &\leqslant \|(\Theta_x - n^{-1}\mathbf{1}\otimes x)v\|_{(n)} \\
&= \frac{1}{2n}\sum_{i=1}^{n}\Big|\sum_{j=1}^{n}((x_i - x_j)v_j - (x_i + x_j)v_i)\Big| \\
&= \frac{1}{2n}\sum_{j\in S_{x,+1}}(x_{i_0} + x_j) + \frac{1}{2n}\sum_{j\in S_{x,-1}}(x_{i_0} - x_j) \\
&= 1.
\end{aligned}$$

Case (c) $l = n$. We need to check that

$$\max\{\|(\Theta_x - n^{-1}\mathbf{1}\otimes x)v\|_{(k)} \colon v \in \{\pm 1\}^n\} = \|x\|_{(k)}. \tag{3.2}$$

Looking again at the extreme points of $\mathfrak{B}^1_{(k)}$, we may assume that $x = e_{i_0}$ or $x \in \frac{1}{k}\{\pm 1\}^n$. In the latter case, with $\tilde{x} = kx$,

$$\begin{aligned}\frac{1}{k}\|(\Theta_{\tilde{x}} - n^{-1}\mathbf{1}\otimes\tilde{x})v\|_{(k)} &\leqslant \|(\Theta_{\tilde{x}} - n^{-1}\mathbf{1}\otimes\tilde{x})v\|_\infty \\ &= \max_{1\leqslant i\leqslant n}\frac{1}{2n}\Big|\sum_{j=1}^n((\tilde{x}_i - \tilde{x}_j)v_j - (\tilde{x}_i + \tilde{x}_j)v_i)\Big| \\ &\leqslant 1,\end{aligned}$$

because $(\tilde{x}_i - \tilde{x}_j)(\tilde{x}_i + \tilde{x}_j) = 0$ and $|\tilde{x}_i| = |\tilde{x}_j| = 1$.

Lastly, if $x = e_{i_0}$ one can apply an argument similar to Case (b). Let $S_{v_{i_0},-1}, S_{v_{i_0},+1}$ be the support sets of v taking values $-v_{i_0}$ and v_{i_0}, respectively. Then

$$\begin{aligned}\|(\Theta_x - n^{-1}\mathbf{1}\otimes x)v\|_{(k)} &\leqslant \|(\Theta_x - n^{-1}\mathbf{1}\otimes x)v\|_{(n)} \\ &= \frac{1}{2n}\sum_{i=1}^n\Big|\sum_{j=1}^n((v_j - v_i)x_i - (v_i + v_j)x_j)\Big| \\ &= \frac{1}{2n}\left|\left(\sum_{j=1}^n(v_j - v_{i_0})\right) - 2v_{i_0}\right| + \frac{1}{2n}\sum_{i\neq i_0}|v_{i_0} + v_i| \\ &\leqslant \frac{1}{2n}\sum_{j=1}^n|v_j - v_{i_0}| + \frac{2}{2n} + \frac{1}{2n}\sum_{i\neq i_0}|v_{i_0} + v_i| \\ &= \frac{1}{n}(|S_{v_{i_0},-1}| + 1 + |S_{v_{i_0},+1}| - 1) = 1.\end{aligned}$$

Therefore we may conclude that the optimum value of (3.1) is at most 1 as desired, and the proof is complete. □

Proof of Theorem 1. First, assume that μ is the uniform probability measure on $[n] = \{1,\dots,n\}$. It is simple to see that

$$fg - \mathbb{E}(fg) = \frac{1}{n}\left[\sum_{j=1}^n(f_ig_i - f_jg_j)\right]_{i=1}^n.$$

Next, we observe that

$$f_ig_i - f_jg_j = \frac{1}{2}(f_i + f_j)(g_i - g_j) + \frac{1}{2}(g_i + g_j)(f_i - f_j).$$

Hence

$$fg - \mathbb{E}(fg) = -\Theta_f g - \Theta_g f = -\Theta_f(g - \mathbb{E}g) - \Theta_g(f - \mathbb{E}f), \tag{3.3}$$

where the last equality follows from the fact $\Theta_f 1 = \Theta_g 1 = 0$.

Let us consider two vectors x and y in $\mathbb{R}^n$. We claim that the following version of Hölder's inequality holds

$$\|\Theta_x(y - \mathbb{E}y)\|_r \leqslant \|x\|_p\|y - \mathbb{E}y\|_q, \tag{3.4}$$

where $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$ and $r, p, q \geqslant 1$. One then applies the decomposition (3.3) and the proof is straightforward.

To prove (3.4), let $\mathfrak{X}_0$ denote the subspace of n-dimensional vectors of zero mean value:

$$\mathfrak{X}_0 = \{x \in \mathbb{R}^n \colon \sum_{i=1}^{n} x_i = 0\}.$$

We denote by q^* (and r^*) the conjugate exponent of q (and r, resp.), as usual. We now verify that the norm of the operator $\Theta_x \colon (\mathfrak{X}_0, \|\cdot\|_q) \to (\mathbb{R}^n, \|\cdot\|_r)$ is at most $\|x\|_p$. By a duality argument, the adjoint

$$\begin{aligned} \Theta_x^* \colon (\mathbb{R}^n, \|\cdot\|_{r^*}) &\to (\mathbb{R}^n/\mathbb{R}, \|\cdot\|_{q^*}) \\ y &\mapsto \Theta_x y + \mathbb{R} \end{aligned}$$

has the same norm as Θ_x (see e.g. [26, Proposition 2.3.10]), and $1/p + 1/r^* = 1/q^*$ holds. To get an estimate in the quotient space $(\mathbb{R}^n/\mathbb{R}, \|\cdot\|_{q^*})$, define the constant

$$\lambda_y = \frac{1}{n}\langle x, y\rangle$$

for any $y \in \mathbb{R}^n$. Clearly, $\Theta_x y - \lambda_y \mathbf{1} = (\Theta_x - n^{-1}\mathbf{1} \otimes x)y$. Now let us apply Lemma 3 to the symmetric norm $\|\cdot\| = \|\cdot\|_{q^*}$ and Hölder's inequality to conclude

$$\begin{aligned} \inf_{\lambda \in \mathbb{R}} \|\Theta_x y - \lambda \mathbf{1}\|_{q^*} \leqslant \|\Theta_x y - \lambda_y \mathbf{1}\|_{q^*} = \|(\Theta_x - n^{-1}\mathbf{1} \otimes x)y\|_{q^*} &\leqslant \||x|^{\downarrow}|y|^{\downarrow}\|_{q^*} \\ &\leqslant \|x\|_p \|y\|_{r^*}. \end{aligned}$$

Hence the claim readily follows and the uniform case is proved.

If we have a general probability space $(\Omega, \mathscr{F}, \mu)$, applying a uniform approximation by simple functions, we can reduce the problem to finite state spaces. Let us now use the approximation method in [2, Proposition 2.1], and the finite discrete case reduces to the uniform one. We briefly recall the trick. Let $\mu_{\mathbb{Q}}$ be any probability measure on an n-points state space S_n such that every atom carries a rational probability. Let us write $(\mu_{\mathbb{Q}})_i = r_i/m$, where $r_i, m \in \mathbb{N}$, for all $1 \leqslant i \leqslant n$. The map

$$\Phi \colon (x_1, \ldots, x_n) \mapsto (\underbrace{x_1, \ldots, x_1}_{r_1}, \ldots, \underbrace{x_n, \ldots, x_n}_{r_n})$$

is an injective algebra homomorphism from $\mathbb{R}^n$ into $\mathbb{R}^m$. Furthermore, if λ denotes the uniform probability measure on the m-points space, we have $\|x\|_{\ell^p(\mu_{\mathbb{Q}})} = \|\Phi(x)\|_{\ell^p(\lambda)}$ and $\|xy - \mathbb{E}_{\mu_{\mathbb{Q}}}(xy)\|_{\ell^p(\mu_{\mathbb{Q}})} = \|\Phi(x)\Phi(y) - \mathbb{E}_{\lambda}(\Phi(x)\Phi(y))\|_{\ell^p(\lambda)}$ in the weighted ℓ^p spaces for any $x, y \in \mathbb{R}^n$ and $1 \leqslant p \leqslant \infty$. Since any probability measure on the n-points space S_n can be approximated by rational probability measures, the proof is complete. $\square$

REMARK 1. Let L be the $n \times n$ Laplacian matrix defined by $L = n^{-1}\mathbf{1} \otimes \mathbf{1} - I$. Then $-Lf = f - \mathbb{E}_\lambda f$, where λ is the uniform probability measure on $[n]$. Let us consider the derivation $\partial \colon \ell_n^2(\lambda) \to \ell_n^2(\lambda) \otimes \ell_n^2(\lambda) = \ell_{n\times n}^2(\lambda \otimes \lambda)$ given by

$$(\partial f)_{ij} = \frac{f_i - f_j}{\sqrt{2}},$$

and let the left and right actions be $f(a_{ij}) = (f_i a_{ij})$ and $(a_{ij})f = (f_j a_{ij})$ on $\ell_{n\times n}^2(\lambda \otimes \lambda)$. It is easy to see that

$$-L = \partial^* \partial$$

and from the Leibniz rule that

$$-L(fg) = \partial^*(f\partial g) + \partial^*((\partial f)g).$$

Moreover, a little computation gives that $\partial^*(f\partial g) = -\Theta_f g$ and $\partial^*((\partial f)g) = -\Theta_g f$. Hence we get back decomposition (3.3) through the derivation ∂.

We also note that

$$\partial^*(f\partial g) = -\frac{1}{2}(L(fg) - gLf + fLg)$$

holds.

4. On the chain rule for random variables

In this section, we will prove a 'chain rule' for the L^p norm of compositions of bounded random variables with Lipschitz functions.

Let us introduce the fundamental concept of discrete Laplacians. We recall that an $n \times n$ symmetric matrix L is *Laplacian* if it has zero row and column sums and all of its off-diagonals are non-negative. A straightforward corollary of the definition is that $-L$ is positive semi-definite.

Our first proposition gives a conditional norm estimate of L. The proof is an application of the Calderón–Mityagin interpolation (see [8], [22] or [31, Theorem 15.17]).

THEOREM 2. *Let L be an $n \times n$ Laplacian. For any symmetric norm $\|\cdot\|$ on $\mathbb{R}^n$ and $x \in \mathfrak{X}_0$, we have*

$$\|Lx\| \leqslant n(\max_{i\neq j} L_{ij})\|x\|.$$

Proof. Let us consider L as a linear operator which maps the normed space $(\mathfrak{X}_0, \|\cdot\|)$ into $(\mathbb{R}^n, \|\cdot\|)$. It is simple to see that the dual space is $(\mathfrak{X}_0, \|\cdot\|)^* = (\mathbb{R}^n, \|\cdot\|_*)/\mathbb{R}$, where $\|\cdot\|_*$ denotes the dual norm. Since $L1 = 0$, the adjoint of L is

$$\begin{aligned} L^* \colon (\mathbb{R}^n, \|\cdot\|_*) &\to (\mathbb{R}^n, \|\cdot\|_*)/\mathbb{R}, \\ v &\mapsto Lv + \mathbb{R}. \end{aligned}$$

Let us associate a non-negative vector $x_\infty \in \mathbb{R}^n$ to L defined by

$$x_\infty(i) = \max_{1\leqslant j, j\neq i\leqslant n} L_{ij}, \qquad 1 \leqslant i \leqslant n.$$

Set the matrix

$$\hat{L} = L - x_\infty \otimes 1.$$

We claim that the estimates hold:

$$\|\hat{L}\|_{1\to 1} \leqslant n \max_{i\neq j} L_{ij} \quad \text{and} \quad \|\hat{L}\|_{\infty\to\infty} \leqslant n \max_{i\neq j} L_{ij}.$$

In fact, L has zero row and column sums hence for all $1 \leqslant i \leqslant n$ we have

$$\begin{aligned}\sum_{j=1}^n |\hat{L}_{ij}| &= \left(x_\infty(i) + \sum_{1\leqslant j, j\neq i\leqslant n} L_{ij}\right) + \sum_{1\leqslant j, j\neq i\leqslant n} (x_\infty(i) - L_{ij}) \\ &= n x_\infty(i) \leqslant n(\max_{i\neq j} L_{ij}).\end{aligned}$$

Similarly, for any $1 \leqslant j \leqslant n$,

$$\begin{aligned}\sum_{i=1}^n |\hat{L}_{ij}| &= \left(x_\infty(j) + \sum_{1\leqslant i, i\neq j\leqslant n} L_{ij}\right) + \sum_{1\leqslant i, i\neq j\leqslant n} (x_\infty(i) - L_{ij}) \\ &= \sum_{i=1}^n x_\infty(i) \leqslant n(\max_{i\neq j} L_{ij}).\end{aligned}$$

This implies the claim.

Since the dual norm $\|\cdot\|_*$ is symmetric, the Calderón-Mityagin interpolation readily gives that

$$\|\hat{L}^T v\|_* \leqslant n \max_{i\neq j} L_{ij} \|v\|_*, \qquad v \in \mathbb{R}^n. \tag{4.1}$$

Pick a $v \in \mathbb{R}^n$. Set

$$\lambda_v = \langle x_\infty, v\rangle.$$

Then

$$\begin{aligned}\inf_{\lambda\in\mathbb{R}} \|Lv - \lambda 1\|_* \leqslant \|Lv - \lambda_v 1\|_* &= \|Lv - \langle x_\infty, v\rangle 1\|_* \\ &= \|\hat{L}^T v\|_*.\end{aligned}$$

But from (4.1) it now follows that

$$\|L^*\| \leqslant \|\hat{L}^T\|_{\|\ \|_*\to\|\ \|_*} \leqslant n \max_{i\neq j} L_{ij}.$$

Since $\|L|\mathfrak{X}_0\| = \|L^*\|$, the proof is complete. $\square$

REMARK 2. Conditional properties of Hermitians have turned out to be particularly useful concepts. For instance, a theorem proved by Bhatia and Sano [5] says that if φ is an operator convex function then the related Loewner matrix

$$L_\varphi = \left[\frac{\varphi(x_i)-\varphi(x_j)}{x_i - x_j}\right]_{i,j},$$

where $x_1,\dots,x_n$ are distinct points in $(0,\infty)$, must be conditionally negative definite, i.e. it is negative definite on the subspace $\mathfrak{X}_0$. We refer the reader to [1], [14] for further interesting examples.

Let L be an $n\times n$ Laplacian matrix. If $\varphi\colon \mathbb{R}\to\mathbb{R}$ is a concave function, then Jensen's inequality implies that $L\varphi(f)\leqslant \varphi'(f)Lf$ holds for any $f\in\mathbb{R}^n$. On the other hand, if we consider the Dirichlet form given by $\mathscr{E}(f)=-\langle f, Lf\rangle$, one has with any Lipschitz function φ that

$$\mathscr{E}(\varphi\circ f)\leqslant \mathrm{Lip}(\varphi)^2\mathscr{E}(f),$$

which is usually referred to as the Markovian property of $\mathscr{E}$, see e.g. [11, p. 5],[17, p. 42]. From now on $\mathrm{Lip}(\varphi)$ denotes the Lipschitz constant of the function φ. Interestingly, if the above inequality holds only for monotone Lipschitz functions then it is valid for any Lipschitzian, see e.g. [17, Proposition 2.1.3]. Specifically, the variance $\mathrm{Var}_\mu(x)=\mathbb{E}_\mu(|x-\mathbb{E}_\mu x|^2)$ of any vector x (with respect to the probability measure μ) satisfies the inequality

$$\mathrm{Var}_\mu(\varphi(x))\leqslant \mathrm{Lip}(\varphi)^2\mathrm{Var}_\mu(x).$$

DEFINITION 1. If $\varphi\colon\mathbb{R}\to\mathbb{R}$ is a real function, and $x_1,\dots,x_n$ are distinct points in $\mathbb{R}$, we define an $n\times n$ symmetric matrix $\Theta[\underline{x};\varphi]$ with zero row and column sums by

$$\Theta_{ij}[x_1,\dots,x_n;\varphi]=\begin{cases}\frac{\varphi(x_i)-\varphi(x_j)}{x_i-x_j} & \text{if } i\neq j\\ -\sum_{k:k\neq i}\Theta_{ik}[x_1,\dots,x_n;\varphi] & \text{if } i=j.\end{cases}$$

Given the uniform probability measure on the state space $\{1,\dots,n\}$, the next lemma easily links the vector $\varphi(x)-\mathbb{E}\varphi(x)$ and the matrix $\Theta[\underline{x};\varphi]$.

LEMMA 4. *Let $x=(x_1,\dots,x_n)\in\mathbb{R}^n$. Then*

$$-\frac{1}{n}\Theta[\underline{x};\varphi]\left(x-\frac{1}{n}\sum_{j=1}^n x_j\mathbf{1}\right)=\varphi(x)-\frac{1}{n}\sum_{j=1}^n\varphi(x_j)\mathbf{1}.$$

Proof. Fix an index $1\leqslant k\leqslant n$. We get

$$\begin{aligned}\left[-\frac{1}{n}\Theta[\underline{x};\varphi]x\right]_k &= -\frac{1}{n}\sum_{i=1}^n\frac{\varphi(x_k)-\varphi(x_i)}{x_k-x_i}(x_i-x_k)\\ &=\varphi(x_k)-\frac{1}{n}\sum_{j=1}^n\varphi(x_j)\mathbf{1}.\end{aligned}$$

Since $\Theta[\underline{x};\varphi]\mathbf{1} = 0$, the proof is complete. $\square$

It is clear that if $\varphi\colon \mathbb{R} \to \mathbb{R}$ is a monotone increasing function then $\Theta[\underline{x};\varphi]$ is a Laplacian. Now a straightforward corollary of Theorem 2 is the following.

COROLLARY 1. *Let $\|\cdot\|$ denote any symmetric norm on $\mathbb{R}^n$. For any monotone Lipschitz function $\varphi\colon \mathbb{R} \to \mathbb{R}$, distinct points $x_1,\dots,x_n \in \mathbb{R}$ and $u \in \mathfrak{X}_0$, we have*

$$\|\Theta[x;\varphi]u\| \leqslant n\,\mathrm{Lip}(\varphi)\|u\|.$$

From Lemma 4 we arrive at the next result.

COROLLARY 2. *Let $\|\cdot\|$ be a symmetric norm on $\mathbb{R}^n$. For any monotone Lipschitz function $\varphi\colon \mathbb{R} \to \mathbb{R}$ and $f = (f_1,\dots,f_n) \in \mathbb{R}^n$, we have*

$$\Big\|\varphi(f) - \frac{1}{n}\sum_{i=1}^{n}\varphi(f_i)\mathbf{1}\Big\| \leqslant \mathrm{Lip}(\varphi)\Big\|f - \frac{1}{n}\sum_{i=1}^{n} f_i\mathbf{1}\Big\|.$$

Now by means of the approximation described in the proof of Theorem 1, we get the following.

THEOREM 3. *Let $(\Omega,\mathscr{F},\mu)$ be a probability space and fix $1 \leqslant p \leqslant \infty$. For any monotone Lipschitz function $\varphi\colon \mathbb{R} \to \mathbb{R}$ and real $f \in L^\infty(\Omega,\mu)$, we have*

$$\|\varphi(f) - \mathbb{E}\varphi(f)\|_p \leqslant \mathrm{Lip}(\varphi)\|f - \mathbb{E}f\|_p.$$

REMARK 3. In general, there is no distinction if we have a real Lipschitz function given on the real line or a subset of $\mathbb{R}$. Applying Zorn's lemma we may find an extension from the smaller set to $\mathbb{R}$ with the same Lipschitz number, see [33, Theorem 1.5.6].

Unfortunately, the next examples show that one cannot remove the restriction to monotone functions in the previous theorem.

EXAMPLE 1. Let us consider the probability vector $\mu = (1/36, 3/4, 2/9)$, and let $f = (-0.3, 0.28, 0.38)$. The function φ, defined on sufficiently small open neighborhoods of the components of f, is the map $x \mapsto x^{-1}$, and take any of its extension to $\mathbb{R}$ with the same Lipshitz number. Specifically, let ℓ^1_μ be the ℓ^1-space of functions on a 3-points space with the weight function μ. Then we get

$$\|\varphi(f) - \mathbb{E}_\mu\varphi(f)\|_{\ell^1(\mu)} = \|f^{-1} - \mathbb{E}_\mu f^{-1}\|_{\ell^1(\mu)} = 0.57783,$$

whilst

$$\|f^{-1}\|_\infty^2\|f - \mathbb{E}_\mu f\|_{\ell^1(\mu)} = 0.5417.$$

Another but piecewise linear example shows the failure of the Markov property in weighted ℓ^1 spaces.

EXAMPLE 2. Let us consider the probability vector $\mu = (1/6, 9/12, 1/12)$. Define φ on the real line by

$$\varphi(x) = \begin{cases} -\left(x + \frac{11}{15}\right) & \text{if } x \leqslant \frac{1}{15} \\ \frac{3}{5}x - \frac{21}{25} & \text{if } x \geqslant \frac{1}{15}. \end{cases}$$

Clearly, $\mathrm{Lip}(\varphi) = 1$. Set $f = (-11/15, 1/15, 13/15)$. A straightforward calculation gives that

$$\|f - \mathbb{E}_\mu f\|_{\ell^1(\mu)} = 0.244\ldots$$

and

$$\|\varphi(f) - \mathbb{E}_\mu \varphi(f)\|_{\ell^1(\mu)} = 0.26.$$

On the other hand, we stress that if φ is the square function (restricted to the range of f) then

$$\|f^2 - \mathbb{E} f^2\|_p \leqslant 2\|f\|_\infty \|f - \mathbb{E} f\|_p,$$

for all $1 \leqslant p \leqslant \infty$.

Our numerical experiments lead us to suspect that if $2 \leqslant p \leqslant \infty$ then one may remove the monotonicity assumption in Theorem 3, but we shall leave open this question.

REMARK 4. In [30] the strong Leibniz inequality related to the standard deviation was studied in the commutative and non-commutative context as well. We recall that a seminorm L defined on a unital normed algebra $(\mathscr{A}, \|\cdot\|, \mathbf{1}_{\mathscr{A}})$ is strongly Leibniz if it satisfies the Leibniz inequality and $L(a^{-1}) \leqslant \|a^{-1}\|^2 L(a)$ whenever $a \in \mathscr{A}$ is invertible.

Now if we choose the algebra $L^\infty(\Omega, \mu)$ over a probability space $(\Omega, \mathscr{F}, \mu)$ and define the seminorm $L(f) = \|f - \mathbb{E} f\|_p$, $1 \leqslant p \leqslant \infty$, the Leibniz inequality is satisfied with real functions. This was completely established in [21] and the result is included in Theorem 1 in the end-point case $p_1 = p_2 = \infty$. Unfortunately, the strong Leibniz property, in general, fails in probability spaces, see Example 1 if $p = 1$. However, our numerical simulations suggest that if $2 \leqslant p \leqslant \infty$ the strong property may hold as well.

REFERENCES

[1] R. B. BAPAT AND T. E. S. RAGHAVAN, *Nonnegative Matrices and Applications*, Cambridge University Press, Cambridge, 1997.

[2] Á. BESENYEI AND Z. LÉKA, *Leibniz seminorms in probability spaces*, J. Math. Anal. Appl., **429** (2015), 1178–1189.

[3] F. BERNICOT, D. MALDONADO, K. MOEN AND V. NAIBO, *Bilinear Sobolev–Poincaré inequalities and Leibniz-type rules*, J. Geom. Anal., **24** (2014), 1144–1180.

[4] R. BHATIA, *Matrix analysis*, Springer–Verlag New York, 1997.

[5] R. BHATIA AND T. SANO, *Loewner matrices and operator convexity*, Math. Ann., **344** (2009), 703–716.

[6] D. BERTSIMAS, D. PACHAMANOVA AND M. SIM, *Robust linear optimization under general norms*, Operations Research Letters, **32** (2004), 510–516.

[7] N. BOULEAU AND F. HIRSCH, *Dirichlet Forms and Analysis on Wiener space*, de Gruyter Studies in Mathematics, De Gruyter, Berlin, 1991.

[8] A.-P. CALDERÓN, *Spaces between L^1 and L^∞ and the theorem of Marcinkiewicz*, Stud. Math., **26** (1966), 273–299.
[9] S. CHANG AND C. K. LI, *Certain isometries on $\mathbb{R}^n$*, Lin. Alg. Appl., **165** (1992), 251–265.
[10] F. CHRIST AND M. WEINSTEIN, *Dispersion of small-amplitude solutions of the generalized Korteweg-de Vries equation*, J. Funct. Anal., **100** (1991), 87–109.
[11] M. FUKUSHIMA, *Dirichlet Forms and Markov Processes*, North Holland Mathematical Library, 1980.
[12] L. GRAFAKOS, *Modern Fourier Analysis*, Graduate Texts in Mathematics, Springer, New York, 2014.
[13] L. GRAFAKOS AND S. OH, *The Kato–Ponce inequality*, Comm. Partial Differential Equations, **39** (2014), 1128–1157.
[14] R. A. HORN, C. R. JOHNSON, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.
[15] T. KATO AND G. PONCE, *Commutator estimates and the Euler and Navier-Stokes equations*, Comm. Pure Appl. Math., **41** (1988), 891–907.
[16] C. KENIG, G. PONCE, AND L. VEGA, *Well-posedness and scattering results for the generalized Korteweg-de Vries equation via the contraction principle*, Comm. Pure Appl. Math., **46** (1993), 527–620.
[17] J. KIGAMI, *Analysis on Fractals*, Cambridge University Press, Cambridge, 2001.
[18] F. LATRÉMOLIÈRE, *The dual Gromov–Hausdorff propinquity*, J. Math. Pures Appl., **103** (2015), 303–351.
[19] F. LATRÉMOLIÈRE, *The quantum Gromov–Hausdorff propinquity*, Trans. Amer. Math. Soc., **368** (2016), 365–411.
[20] F. LATRÉMOLIÈRE, *Quantum metric spaces and the Gromov–Hausdorff propinquity*, preprint, `http://arxiv.org/pdf/1506.04341`.
[21] Z. LÉKA, *Symmetric seminorms and the Leibniz property*, J. Math. Anal. Appl., **452** (2017), 708–725.
[22] B. S. MITYAGIN, *An interpolation theorem for modular spaces* (Russian), Mat. Sb. (N. S.), **66** (1965), 473–482.
[23] C. MUSCALU AND W. SCHLAG, *Classical and Multilinear Harmonic Analysis*, vol. 2., Cambridge University Press, 2013.
[24] C. MUSCALU, J. PIPHER, T. TAO AND C. THIELE, *Bi-parameter paraproducts*, Acta Math., **193** (2004), 269–296.
[25] J. VON NEUMANN, *Some matrix inequalities and metrization of matrix-space*, Rev. Tomsk Univ., **1** (1937), 286–300.
[26] G. K. PEDERSEN, *Analysis Now*, Springer–Verlag, 1989.
[27] M. A. RIEFFEL, *Metrics on state spaces*, Doc. Math., **4** (1999), 559–600.
[28] M. A. RIEFFEL, *Leibniz seminorms for "matrix algebras converge to the sphere"*, Quanta of Maths **11**, Amer. Math. Soc., Providence, RI, 2010, 543–578.
[29] M. A. RIEFFEL, *Non-commutative resistance networks*, *SIGMA Symmetry Integrability Geom. Methods Appl.*, **10** (2014), 2259–2274.
[30] M. A. RIEFFEL, *Standard deviation is a strongly Leibniz seminorm*, New York J. Math., **20** (2014), 35–56.
[31] B. SIMON, *Convexity: An Analytic Viewpoint*, Cambridge University Press, 2011.
[32] N. WEAVER, *Lipschitz algebras and derivations of von Neumann algebras*, J. Funct. Anal., **139** (1996), 261–300.
[33] N. WEAVER, *Lipschitz algebras*, World Scientific, 1999.
[34] B. WU, C. DING, D. SUN AND K.-C. TOH, *On the Moreau–Yosida regularization of the vector k-norm related functions*, SIAM Journal on Optimization, **24** (2014), 766–794.

(Received November 23, 2016)

Zoltán Léka
Royal Holloway
University of London
Egham Hill, Egham, Surrey, TW20 0EX, United Kingdom
e-mail: `zoltan.leka@rhul.ac.uk`