Portugal. Math. (N.S.)
Vol. 75, Fasc. 1, 2018, 1–10
DOI 10.4171/PM/2009

Portugaliae Mathematica

On minimal Hölder gaps and Shannon entropy balance

Immanuel M. Bomze

(Communicated by Luís Nunes Vicente)

Abstract. When estimating a bilinear form in x and y by a product of two terms depending solely on x or y, the well known Hölder inequality which uses the product of a p-norm and its dual comes easily into play. However, if one can choose p freely, one could reduce this Hölder gap accordingly. This note addresses this elementary but apparently not too popular issue by using strict log-convexity of the p-norm in $\frac{1}{p}$ (sometimes called Littlewood's inequality). The optimal p is characterized by a balance condition on the Shannon entropies of distributions related to x and y.

Mathematics Subject Classification: 26D15, 90C25, 94A17

Keywords: p-norm, duality, convexity

1. Motivation

Let $(\mathscr{L}, \|.\|, \leq)$ be a Banach lattice [18] with positive cone $\mathscr{L}_+ := \{f \in \mathscr{L} : f \geq 0\}$. We denote by $\mathscr{L}^*$ its topological dual, and

$$\|y\|_* = \sup_{\|x\| \leq 1} \langle x, y \rangle$$

the dual norm on $\mathscr{L}^*$. Here we will discuss the special case the Euclidean space $\mathscr{L} = \mathbb{R}^n$, or the infinite-dimensional sequence space $\mathscr{L} = \ell^p$, or else $\mathscr{L} = L^p(\mu)$ where μ is a finite measure on a measurable space $(\Omega, \mathscr{F})$, equipped by the p-norms given by

$$\|x\|_p^p = \sum_{i=1}^{n/\infty} |x_i|^p \qquad \text{or} \qquad \|x\|_p^p = \int_\Omega |x|^p \, d\mu.$$

Then $\mathscr{L}^* = \ell^q$ or $\mathscr{L}^* = L^q(\mu)$ with $\frac{1}{p} + \frac{1}{q} = 1$ and $\|y\|_* = \|y\|_q$, including the cases $p \in \{1, \infty\}$ for which $q \in \{\infty, 1\}$, a subscript $\square_\infty$ designating (μ-essential) sup of an element. The order $\leq$ is understood pointwise (μ-a.e.).

Now the well known Hölder inequality, more precisely, following [19], p. 135, the Rogers–Hölder–Riesz inequality [10], [16], [17], reads

$$|\langle x, y\rangle| \leq \|x\|_p \|y\|_q, \tag{1}$$

with equality if and only if $\{x, y\}$ are linearly dependent. By several authors, see e.g. [3], [9], Thm. 101(a), p. 82, [13], it was shown that essentially no extension in the form

$$|\langle x, y\rangle| \leq \Phi^{-1}(\|\Phi \circ |x|\,\|_1)\Psi^{-1}(\|\Psi \circ |y|\,\|_1)$$

beyond $\Phi(t) = t^p$ and $\Psi(t) = t^q$ is possible for any bijections Φ, Ψ on $\mathbb{R}_+$ with $\Phi(0) = \Psi(0) = 0$, if $\dim \mathscr{L} > 1$.

However, often x and y are given, and their product $\langle x, y\rangle$ is beyond our control. Rather, we want to use (1) to estimate $|\langle x, y\rangle|$ with as tight gap as possible, meaning that for given x and y, we want to minimize the right-hand side $\|x\|_p \|y\|_q$ in p which is not involved in $|\langle x, y\rangle|$. Elementary as this question is, apparently it has not received any significant attention. In this note, it is shown that the problem has a unique solution in p under very mild conditions which hold generically.

Here it should be mentioned that there are other estimates of the Hölder gap, relating it to the Minkoswki gap rather than minimizing it in p, see [11], [12].

Since all p-norms are lattice norms [18], i.e. satisfy $\|\,|x|\,\|_p = \|x\|_p$, we may and do assume in the sequel that $\{x, y\} \subset \mathscr{L}_+$. Further, we will assume that $\{x, y\}$ are linearly independent, and, to simplify discussion also that $\{x, y\} \subset \ell^1_+$ in the sequence space case, or else that $\{x, y\} \subset L^\infty_+(\mu)$ in the finite measure case. Recall that for $1 \leq p < r \leq \infty$ we have $\ell^r \subset \ell^p$ but the reverse inclusion $L^p(\mu) \subset L^r(\mu)$.

By dividing by their respective $\|.\|_1$ (in the atomic case) or $\|.\|_\infty$ in the finite measure case, we can use positive bi-homogeneity of (1) to ensure that both elements x and y are bounded by unity. Further we simply can minimize the logarithm of the right-hand side of (1) and look at

$$\varphi(t) := \log\|x\|_p + \log\|y\|_q = \psi_x(t) + \psi_y(1-t), \qquad t = \frac{1}{p}, \tag{2}$$

where

$$\psi_x(t) := \log\|x\|_{1/t} = (t-1)R_{1/t}(x) \qquad \text{with } R_p(x) = \frac{1}{1-p}\log\|x\|_p^p$$

denoting Rényi entropy [15]. We won't pursue the connection to Rényi entropy here further although some of the following results may be derived also via this connection. See, e.g., [6], Lem. 2.7; see also [1], [2], [5].

For $1 < p < \infty$ and $m \in \{1, 2\}$, we have, putting $\beta = \frac{p-1}{m} > 0$,

$$x^p |\log x|^m = [x^\beta |\log x|]^m x \leq (\beta e)^{-m} x \qquad \text{for } x \in [0, 1],$$

by the above assumption, and similarly for the series case, so that integrability (or summability) of derivatives does not pose any problem whatsoever when differentiating under the integral (or summation) sign.

Next note that if x is constant, then $\psi_x(t)$ is constant in t since all norms then have the same value.

2. Analysis and discussion

To discuss φ, we note that ψ_x is convex, see, e.g. [7], p. 55 where this fact is termed Littlewood's inequality. Then obviously also φ is convex, and we will look at the first two derivatives of ψ_x which can be derived, e.g., from [8]. To this end, consider for $x \neq 0$ the probability measure $v(p) = v(p, x)$ – for brevity we will suppress the dependence on x except in Theorem 2.1 below – given by either

$$v_i(p) := \frac{|x_i|^p}{\sum_j |x_j|^p}, \qquad i \in \{1, \ldots, n\} \text{ or } i \in \mathbb{N} \tag{3}$$

as discrete weights, or in the case $\mathscr{L} = L^p(\mu)$, via the density function w.r.t. μ

$$\frac{dv(p)}{d\mu}(\omega) := \frac{1}{\|x\|_p^p} |x(\omega)|^p, \qquad \omega \in \Omega. \tag{4}$$

To simplify notation, we will use $v(p)$ also for above density function in ω, and we will write $V(p)$ instead of $v(p)$ if the expressions on the right-hand side in (3) or (4) are considered as a random variable w.r.t. $v(p)$. Under the conditions discussed above, this random variable $V(p)$, like the random variable $X = x(\omega)$ or $X = x_i$ itself, has all moments w.r.t. $v(p)$, for all $p \geq 1$. Before continuing we mention the well-known (and elementary) fact for the Euclidean case $\mathscr{L} = \mathbb{R}^n$ that

$$\lim_{p \to \infty} v_i(p) = \begin{Bmatrix} \frac{1}{m}, & \text{if } x_i = \|x\|_\infty, \\ 0, & \text{if } x_i < \|x\|_\infty. \end{Bmatrix} \tag{5}$$

where m is the cardinality of $\operatorname{Argmax}_j x_j$ (generically we have $m = 1$). Note that the space of all probability distributions (absolutely continuous w.r.t. μ) may fail to be compact w.r.t. norm topology if Ω is infinite (a classical problem in Bayesian analysis), allowing for (weak) convergence $v(p) \to 0$ as $p \to \infty$, except in case $\mathscr{L} = \mathbb{R}^n$.

Proposition 2.1. *Consider an element $x \in L^\infty_+(\mu)\backslash\{0\}$ or an element $x \in \ell^1_+\backslash\{0\}$. Then w.r.t. the probability $\nu(p)$ defined in* (4) *or* (3), *we have*

$$\dot{\psi}_x(t) = \mathbb{S}\text{hanEnt}[\nu(p)] := \mathbb{E}_{\nu(p)}[-\log V(p)] \tag{6}$$

and

$$\ddot{\psi}_x(t) = p^3\, \mathbb{V}\text{ar}_{\nu(p)}[\log X] \geq 0, \tag{7}$$

the latter relation becoming an equality if and only if X is constant μ-a.e. (*or in i in the discrete case*). *In all cases, the function $x^r \log x$ is continued with zero at $x = 0$.*

Proof. As $\log V(p) = p \log X - p \log\|x\|_p$, the densities of $\nu(p)$ form an exponential family [14] in the one-dimensional parameter p with natural sufficient statistics $T(X) = \log X$. A general property in such exponential families implies $\mathbb{E}_{\nu(p)}[T(X)] = \frac{d}{dp}[\log\|x\|_p^p]$, easily obtained by differentiating the constant $\int V(p)\, d\nu(p) = 1$ under the integral sign. Using $t\dot{p} = -p$, we now arrive at

$$\begin{aligned}
\dot{\psi}_x(t) &= \log\|x\|_p^p + t\frac{d}{dt}[\log\|x\|_p^p] \\
&= \log\|x\|_p^p + t\dot{p}\frac{d}{dp}[\log\|x\|_p^p] \\
&= \log\|x\|_p^p - p\mathbb{E}_{\nu(p)}[T(X)] = \mathbb{E}_{\nu(p)}[-\log V(p)] = \mathbb{S}\text{hanEnt}[\nu(p)],
\end{aligned}$$

which is (6); after some manipulation, the same result can also be derived from [8], Lem. 1.1, see also [1], Prop. I.2, [6], Lem. 2.7. A similar calculation (related to arguments leading to the Cramér–Rao bound, cf. again [14]) shows $\frac{d}{dp}\mathbb{E}_{\nu(p)}[T(X)] = \mathbb{V}\text{ar}_{\nu(p)}[T(X)]$, implying

$$\begin{aligned}
\ddot{\psi}_x(t) &= \dot{p}\frac{d}{dp}\{\log\|x\|_p^p - p\mathbb{E}_{\nu(p)}[T(X)]\} \\
&= -\dot{p}p\frac{d}{dp}\mathbb{E}_{\nu(p)}[T(X)] = p^3\, \mathbb{V}\text{ar}_{\nu(p)}[\log X] \geq 0.
\end{aligned}$$

The argument is completely analogous in the atomic/discrete case. □

The main result seems to have gone unnoticed up to now.

Theorem 2.1. *Consider two linearly independent elements $\{x, y\} \subset \ell^1_+$, both bounded by unity. Then there is a unique p^* minimizing the Hölder gap $\|x\|_p\|y\|_q - \langle x, y\rangle$, namely either $p^* = \infty$ or $p^* = 1$, or else the unique solution in p of the fol-*

lowing transcendental equation (with $\frac{1}{p}+\frac{1}{q}=1$ for symmetric and nicer notation):

$$\log\|x\|_p^p - \frac{p}{\|x\|_p^p}\sum_i x_i^p \log x_i = \log\|y\|_q^q - \frac{q}{\|y\|_q^q}\sum_i y_i^q \log y_i. \tag{8}$$

A similar statement holds for two linearly independent elements $\{x, y\} \subset L_+^\infty(\mu)$ both bounded by unity, where μ is a finite measure. Here the equation reads

$$\log\|x\|_p^p - \frac{p}{\|x\|_p^p}\int x^p \log x \, d\mu = \log\|y\|_q^q - \frac{q}{\|y\|_q^q}\int y^q \log y \, d\mu. \tag{9}$$

Both conditions are equivalent to the Shannon entropy balance formula

$$\mathbb{S}\text{hanEnt}[v(p, x)] = \mathbb{S}\text{hanEnt}[v(q, y)].$$

Proof. The first two derivatives of φ w.r.t. $t \in]0, 1[$ are $\dot{\varphi}(t) = \dot{\psi}_x(t) - \dot{\psi}_y(1-t)$ and $\ddot{\varphi}(t) = \ddot{\psi}_x(t) + \ddot{\psi}_y(1-t)$. So (7) yields strict convexity of φ. If $\dot{\varphi}(0) \geq 0$, then $p^* = \infty$. If $\dot{\varphi}(1) \leq 0$, then $p^* = 1$. Else $\dot{\varphi}(0) < 0 < \dot{\varphi}(1)$ (this includes the possibility of $\dot{\varphi}(t) \searrow -\infty$ as $t \searrow 0$ or $\dot{\varphi}(t) \nearrow +\infty$ as $t \nearrow 1$ or both), and strict convexity yields uniqueness of the minimizer $p^* \in]1, \infty[$ satisfying (8) or (9), using (6). □

Let us discuss conditions for an interior solution $p^* \in]1, \infty[$. For ease of presentation only, we focus on the finite-dimensional case $\mathscr{L} = \mathbb{R}^n$:

Theorem 2.2. *Consider two vectors $\{x, y\} \subset \mathbb{R}^n$. Suppose that both x and y have a unique coordinate with maximum modulus less than unity, i.e.*

$$|x_j| < |x_i| = \|x\|_\infty < 1 = \|x\|_1 \quad \textit{for all } j \neq i$$

and likewise

$$|y_j| < |y_k| = \|y\|_\infty < 1 = \|y\|_1 \quad \textit{for all } j \neq k.$$

Then the minimizer p^* of the Hölder gap solves (8) *and is thus interior.*

Proof. By assumption $m = 1$, so (5) and (6) yield

$$\dot{\varphi}(t) \to 0 - \log\|y\|_1 + \sum_i |y_i| \log|y_i| < 0 \qquad \text{as } t \searrow 0,$$

because $\|y\|_1 = 1$ and $|y_j| \log|y_j| < 0$ as by the above assumptions we have $0 < |y_j| < 1$ for at least one j. By the same arguments we have $\lim_{t\nearrow 1} \dot{\varphi}(t) > 0$, and the result follows. □

3. An application

As an application we improve upon Cauchy–Schwarz ($p_0 = 2$, i.e. $t_0 = \frac{1}{2}$):

Theorem 3.1. *For $x \neq 0$ define $\rho(x) = \frac{\langle x, x \log|x| \rangle}{\langle x, x \rangle}$ and $\psi(x) = \log\langle x, x\rangle - 2\rho(x)$.*

(a) *If $\psi(x) = \psi(y)$, then Cauchy–Schwarz gives the tightest Hölder gap.*

(b) *However, if $\psi(x) \neq \psi(y)$, then there is a tighter Hölder gap for $p = \frac{1}{t}$ with t a (suitable) convex combination of t_0 and $t_1 = proj_{[0,1]}(t_n)$ where*

$$8t_n = 4 - \frac{\psi(x) - \psi(y)}{\gamma(x) + \gamma(y)} \qquad \text{with } \gamma(x) = \left[\frac{\| x \log|x| \|_2}{\|x\|_2}\right]^2 - \rho^2(x). \tag{10}$$

Here $proj_{[0,1]}(s) = \min\{1, \max\{0, s\}\}$ is the point in $[0,1]$ closest to s.

Proof. Consider a full Newton step for t, starting with $t_0 = \frac{1}{2}$. The projection ensures feasibility, and the sign of $\dot{\varphi}(t_0)$ decides on moving left or right. □

Remark 3.1. Suppose X and Y are two random variables with joint distribution μ on Ω which may be stochastically dependent under μ, but have the same marginal distribution. For the vast field of dependence models in this situation via Sklar's copula refer, e.g. to [4]. Then obviously the Hölder gap equals

$$[\mathbb{E}_\mu(|X|^p)]^{1/p}[\mathbb{E}_\mu(|Y|^q)]^{1/q} - \mathbb{E}_\mu(XY) = [\mathbb{E}_\mu(|X|^p)]^{1/p}[\mathbb{E}_\mu(|X|^q)]^{1/q} - \mathbb{E}_\mu(XY)$$

which is symmetric in t around t_0. So t_0 minimizes the convex function φ in this situation, and the Cauchy–Schwarz gap – which in the uncorrelated (in particular in the independent) case equals $\mathbb{V}\mathrm{ar}_\mu(X)$ – is the smallest Hölder gap. The above theorem generalizes this observation to other copulas which may be of some help in case of non-identical marginals.

Even if the full Newton step is feasible, it is not necessarily improving which requires the above mentioned convex combination. In our experiments below, we simply took the arithmetic mean of t_0 and t_1, reducing the full step length by 50%, and obtained satisfactory results by this heuristic approach, for the case $t_n \in [0, 1]$:

$$t_h = \frac{1}{2}(t_1 + t_0) = t_0 - \frac{\psi(x) - \psi(y)}{16[\gamma(x) + \gamma(y)]}.$$

This procedure basically replaced the Taylor expansion of φ around t_0 with a quadratic model q_h having double curvature of the Taylor polynomial but same tangent and value at t_0, so q_h is overestimating φ locally around t_0, and most of the time on the whole interval between t_0 and t_h, which means that $\varphi(t_h) \le q_h(t_h) < q_h(t_0) = \varphi(t_0)$. According to the empirical results reported in the next section, this seems to be often the case.

Examples. For $x = [0.5, 0.45, 0.05]$ and $y = [0.3, 0.3, 0.4]$ the minimal value of φ is attained at around $t^* = 0.82$ (so $p^* = 1.22$); here $t_n > 1$, and $t_1 = 1$ neither improves upon Cauchy–Schwarz nor on the heuristic value at $t_h = \frac{3}{4}$, as the (rounded) φ-values at (t^*, t_1, t_h, t_0) are, respectively $(-1.01, -0.92, -1.00, -0.93)$. A slight variation of x to $x' = [0.45, 0.45, 0.1]$ – which violates the assumptions in Theorem 2.2 but using the arguments of its proof we still get $\dot{\varphi}(0) = \log 2 + \sum_i y_i \log y_i \approx -0.396 < 0$ – and above y give φ-values at $(t^*, t_1 = t_n, t_h, t_0) = (0.78, 0.96, 0.73, 0.5)$ now equal to $(-1.024, -0.954, -1.022, -0.979)$. Here $p^* = 1.28$. Both phenomena reported here occur however rarely, and the fact that t_h is almost optimal seems to be typical in view of the empirical evidence reported in the following section.

4. A small simulation study

In this study, we consider dimensions $n \in \{3, 5, 10\}$ and look at the statistics for the gap ratios $\frac{\gamma(x,y,t)}{\gamma(x,y,t')}$ across 10^4 randomly drawn samples (x, y) from the unit interval $[0, 1]$, renormalized such that $\|x\|_1 = \|y\|_1 = 1$. Recall the gap $\gamma(x, y, t) = \|x\|_p \|y\|_q - \langle x, y\rangle$ with $t = \frac{1}{p}$. The gap ratios were taken for different combinations $(t, t') \in \{(t^*, t_1), (t^*, t_h), (t^*, t_0), (t_1, t_0), (t_h, t_0)\}$. We report the median value as well as the lower and upper quartile for these ratios. Of course, the conditions of Theorem 2.2 were satisfied so that always $0 < t^* < 1$ holds. In all these 30,000 samples, it never occurred that $\varphi(t_h) > \varphi(t_0)$ and in fact $\frac{\gamma(x,y,t^*)}{\gamma(x,y,t_h)}$ dominates $\frac{\gamma(x,y,t^*)}{\gamma(x,y,t_0)}$ by a significant amount at all quantiles and across all considered dimensions.

In case of $n = 3$, among 10^4 samples we observed 213 cases where the (truncated) Newton step t_1 did not improve the Cauchy–Schwarz gap, 1571 cases where $\varphi(t_1) < \varphi(t_h)$, and 1212 cases where $t_n \notin [0, 1]$. The statistics are detailed in Table 1 below. In case of $n = 5$, among 10^4 samples we observed 103 cases where the (truncated) Newton step t_1 did not improve the Cauchy–Schwarz gap, 846 cases where $\varphi(t_1) < \varphi(t_h)$, and 303 cases where $t_n \notin [0, 1]$. The statistics are detailed in Table 2 below.

In case $n = 10$, due to reasons detailed below, the interquantile range was extended from $\left[\frac{1}{4}, \frac{3}{4}\right]$ to $[0.1, 0.9]$, to include also more extreme observations where only the lowest and the highest decile are treated as outliers. All other specifications were kept. Among 10^4 samples we observed 7 cases where the (truncated) Newton step t_1 did not improve the Cauchy–Schwarz gap, 138 cases where $\varphi(t_1) < \varphi(t_h)$, and 8 cases where $t_n \notin [0, 1]$. The statistics are detailed in Table 3 below.

Table 1. Statistics on gap quotients $\frac{\gamma(x,y,t)}{\gamma(x,y,t')}$ across 10,000 random instances for $\{x, y\} \subset \mathbb{R}^3$; rounded percentages reported.

$(t, t') =$	(t^*, t_1)	(t^*, t_h)	(t^*, t_0)	(t_1, t_0)	(t_h, t_0)
lower quartile	93.60	80.57	46.20	55.58	58.55
median	99.78	96.08	81.56	85.40	85.58
upper quartile	99.99	99.16	96.18	96.95	97.12

Table 2. Statistics on gap quotients $\frac{\gamma(x,y,t)}{\gamma(x,y,t')}$ across 10,000 random instances for $\{x, y\} \subset \mathbb{R}^5$; rounded percentages reported.

$(t, t') =$	(t^*, t_1)	(t^*, t_h)	(t^*, t_0)	(t_1, t_0)	(t_h, t_0)
lower quartile	99.57	94.82	78.48	81.58	83.24
median	99.98	98.32	92.69	93.36	94.44
upper quartile	100.00	99.61	98.37	98.54	98.78

Table 3. Statistics on gap quotients $\frac{\gamma(x,y,t)}{\gamma(x,y,t')}$ across 10,000 random instances for $\{x, y\} \subset \mathbb{R}^{10}$; rounded percentages reported.

$(t, t') =$	(t^*, t_1)	(t^*, t_h)	(t^*, t_0)	(t_1, t_0)	(t_h, t_0)
lower decile	99.76	95.48	82.66	83.44	86.66
median	100.00	99.26	96.98	97.01	97.74
upper decile	100.00	99.97	99.90	99.90	99.93

Pursuing the same simulation strategy for larger dimensions, the differences become less pronounced but this is an artefact of the approach: indeed, let $\gamma_n(t)$ denote the Hölder gap based on a sample $(\mathsf{x}_n, \mathsf{y}_n) = (x_i, y_i)_{i=1}^n \in \mathbb{R}^{2n}$ of $2n$ i.i.d. $U[0, 1]$-distributed random variables. Then the law of large numbers implies that for all $t = \frac{1}{p}$ and $t' = \frac{1}{p'}$ we have almost surely

$$\frac{\gamma_n(t)}{\gamma_n(t')} = \frac{\frac{1}{n}(\|\mathsf{x}_n\|_p\|\mathsf{y}_n\|_q - \langle \mathsf{x}_n, \mathsf{y}_n\rangle)}{\frac{1}{n}(\|\mathsf{x}_n\|_{p'}\|\mathsf{y}_n\|_{q'} - \langle \mathsf{x}_n, \mathsf{y}_n\rangle)} \to \exp[\bar{\varphi}(t) - \bar{\varphi}(t')] \quad \text{as } n \to \infty, \tag{11}$$

with $\bar{\varphi}(t) = \log(\mathbb{E}_U\|X\|_p\mathbb{E}_U\|X\|_q - [\mathbb{E}_U X]^2)$ where X is a $U[0, 1]$-distributed random variable. Remark 3.1 ensures that $\bar{\varphi}(t) \geq \bar{\varphi}(t_0) = \log \mathbb{V}\mathrm{ar}_U X = -\log 12$ for all $t \in [0, 1]$, so that the gap ratios (11) approach 100% for all (t, t') combinations considered in above tables, as convergence when $n \to \infty$ is uniform in $t \in [t_0 - \delta, t_0 + \delta]$ for small enough $\delta > 0$. Again, this does not preclude considerable Hölder gap improvements for a particular instance in high dimensions; but this cannot be detected in our simulation environment.

5. Conclusion

If you want to reduce the Hölder gap, don't stick with Cauchy–Schwarz inequality. It may pay off by a significant improvement (around 20% or even more).

Acknowledgments. The author wants to thank Alfredo Iusem, Mokshay Madiman and Werner Schachinger for inspiring discussions, Luis Nunes Vicente for his cooperative and diligent editorial work, and two anonymous referees for their valuable suggestions.

References

[1] Sergey Bobkov and Mokshay Madiman. The entropy per coordinate of a random vector is highly constrained under convexity conditions. *IEEE Trans. Inform. Theory*, 57(8):4940–4954, 2011.

[2] Sergey Bobkov, Mokshay Madiman, and Liyao Wang. Fractional generalizations of Young and Brunn-Minkowski inequalities. *Contemporary Mathematics*, 545:38–53, 2011.

[3] Ralph Cooper. Note on the Cauchy-Hölder inequality. *Notes of the London Math. Soc.*, 26(2):8–9, 1928.

[4] Giorgio dall'Aglio, Samuel Kotz, and Gabriella Salinetti (eds.), *Advances in Probability Distributions with Given Marginals*, volume 67 of *Mathematics and its Applications*. Springer Netherlands, 1991.

[5] Amir Dembo, Thomas M. Cover, and Joy A. Thomas. Information Theoretic Inequalities. *IEEE Trans. Inform. Theory*, 37(6):1501–1518, 1991.

[6] Matthieu Fradelizi, Mokshay Madiman, and Liyao Wang. Optimal concentration of information content for log-concave densities. In C. Houdré, D. Mason, P. Reynaud-Bouret, and J. Rosinski (eds.), *High Dimensional Probability VII: The Cargèse Volume*, Progress in Probability. Birkhäuser, Basel, 2016.

[7] David John Haldane Garling. *Inequalities: a Journey into Linear Analysis*. Cambridge Univ. Press, Cambridge UK, 2007.

[8] Leonard Gross. Logarithmic Sobolev inequalities. *Am. J. Math.*, 97(4):1061–1083, 1975.

[9] Godfrey Harold Hardy, John Edensor Littlewood, and George Pólya. *Inequalities*. Cambridge University Press, Cambridge, 1952.

[10] Otto Hölder. Über einen Mittelwerthssatz (On a mean value theorem, in German). *Nachr. Ges. Wiss. Göttingen*, 5:38–45, 1889.

[11] Alfredo Noel Iusem, Carlos Augusto Sholl Isnard, and Dan Butnariu. A mixed Hölder and Minkowski inequality. *Proc. Am. Math. Soc.*, 127:2405–2415, 1999.

[12] Alfredo Noel Iusem and Carlos Augusto Sholl Isnard. On mixed Hölder-Minkowski inequalities and total convexity of certain functions in $\mathscr{L}^p(\Omega)$. *Math. Inequalities Appl.*, 3:519–537, 2000.

[13] Janusz Matkowski. The converse of the Hölder inequality and its generalizations. *Studia Mathematica*, 109(2):171–182, 1994.

[14] Erich L. Lehmann and George Casella. *Theory of Point Estimation*, 2nd ed. Springer, New York, 1998.

[15] Alfréd Rényi. On measures of entropy and information. *Proc. Fourth Berkeley Symp. on Math. Statist. and Prob.* 1:547–561, 1961.

[16] Frigyes Riesz. Untersuchungen über Systeme integrierbarer Funktionen (Investigations on systems of integrable functions, in German). *Math. Annal*, 69:449–497, 1910.

[17] Leonard James Rogers. An extension of a certain theorem in inequalities. *Messenger of Mathematics*, 17(10):145–150, 1888.

[18] Helmut H. Schaefer. *Banach Lattices and Positive Operators*, volume 215 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, 1974.

[19] J. Michael Steele. *The Cauchy-Schwarz Master Class. An Introduction to the Art of Mathematical Inequalities*. Cambridge Univ. Press, New York, 2004.

Received September 9, 2017; revision received April 13, 2018

I. M. Bomze, ISOR/VCOR and ds:UniVie, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria
E-mail: immanuel.bomze@univie.ac.at