Characterisation of matrix entropies

Frank Hansen and Zhihua Zhang

February 10, 2014

Abstract

The notion of matrix entropy was introduced by Tropp and Chen with the aim of measuring the fluctuations of random matrices. It is a certain entropy functional constructed from a representing function with prescribed properties, and Tropp and Chen gave some examples. We characterise the matrix entropies in terms of the second derivative of their representing function and arrive in this way at a canonical integral representation of the set of matrix entropies.

1 Introduction and main result

The search for concentration inequalities has been a flourishing field in probability theory during the past thirty years [2]. Among various inequalities the matrix concentration inequality class has applications in many fields such as random graph theory, compressed sensing et cetera [3]. Recently Chen and Tropp developed a matrix extension of the entropy method and used it to search for matrix concentration inequalities [3]. They studied in particular the matrix entropy inequalities associated with the standard entropy function $t \mapsto t \log t$ and the power functions $t \mapsto t^p$ for each $p \in [1, 2]$. Tropp and Chen [3] gave the following definition:

Definition 1.1. Let for each natural number $n$ the class $\Phi_n$ consist of the functions $\varphi: (0, \infty) \to \mathbb{R}$ that are either affine or satisfy the following three conditions.

(i) $\varphi$ is convex.

(ii) $\varphi$ is twice continuously differentiable.
Let $f = \varphi'$ be the derivative of $\varphi$. The Fréchet differential $df(x)$ of the matrix function $x \mapsto f(x)$ is an invertible linear operator on the Hilbert space $\mathcal{H}_n = B(H_n)$, and the map $x \mapsto df(x)^{-1}$ is concave.

The notation is that $H_n$ is a Hilbert space of dimension $n$, and $\mathcal{H}_n$ is the Hilbert space of bounded linear operators on $H_n$ equipped with the inner product given by the trace. The class of (representing functions for) matrix entropies $\Phi_\infty$ is then defined as the intersection

$$\Phi_\infty = \bigcap_{n=1}^\infty \Phi_n.$$  

It follows from our characterisation that $\Phi_\infty$ is a convex cone (it actually follows that each set $\Phi_n$ is a convex cone). But this is not obvious [2, 9] even for the set $\Phi_1$. The authors [2] showed that a twice differentiable strictly convex function $\varphi$ defined in the positive half-line is in $\Phi_1$ if the induced $\varphi$-entropy

$$H_\varphi(Z) = \mathbb{E}[\varphi(Z)] - \varphi(\mathbb{E}[Z])$$

is convex on the set $L_+^\infty(\Omega, \mathcal{A}, \mathbb{P})$ of bounded and non-negative random variables $Z$, where $\mathbb{E}[Z]$ denotes the expectation of $Z$.

More generally, Tropp and Chen [3] introduced to each $\varphi \in \Phi_\infty$ the following matrix $\varphi$-entropy functional

$$H_\varphi(Z) := \mathbb{E}\text{Tr}[\varphi(Z)] - \text{Tr}\varphi(\mathbb{E}[Z]),$$

where now $Z$ is a positive semi-definite random matrix. The authors established subadditivity of $H_\varphi(Z)$ and derived matrix extensions of the bounded difference inequality and the moment inequality by choosing suitable representing functions in $\Phi_\infty$. For the difference inequality they used the function $t \mapsto t \log t$, and for the moment inequality the functions $t \mapsto t^p$, where $p = q/(q - 1)$ for integers $q = 2, 3, \ldots$.

By applying and extending the techniques in [7] we are able to reformulate the defining properties of matrix entropy in a more transparent way that ultimately makes it possible to arrive at the complete characterisation given below.

**Theorem 1.2.** Let $\varphi : (0, \infty) \to \mathbb{R}$ be a twice continuously differentiable convex function, and let $f = \varphi'$ denote the derivative of $\varphi$. The following conditions are then equivalent.
(i) \( \varphi \) is the representing function of a matrix entropy.

(ii) The non-negative function \( f' = \varphi'' \) is numerically decreasing and operator convex.

(iii) There exists a bounded measure \( \nu \) with support in \([0, \infty)\) such that
\[
f'(t) = \beta + \int_0^\infty \frac{1}{t + \lambda} d\nu(\lambda) \quad t > 0,
\]
where \( \beta \geq 0 \) is a constant.

Chen and Tropp proved that the standard entropy function \( t \mapsto t \log t \) and the power functions \( t \mapsto t^p \) for \( p \in [1, 2] \) are representing functions for matrix entropies. These statements may also be derived from earlier results by Lieb [10] and the first author [7] formulated outside the theory of matrix entropies; although it is clear that the different authors arrived at similar conclusions as the result of independent research activities.

2 Reformulating the main condition

The next result is an adaptation of a result by the first author [7] applying ideas going back to Lieb [10]. We consider, for each natural number \( n \), a Hilbert space \( H_n \) of dimension \( n \) and the Hilbert space \( H_n = B(H_n) \) of bounded linear operators on \( H_n \) with inner product given by \( (x \mid y) = \text{Tr} y^*x \).

**Theorem 2.1.** Let \( f: (0, \infty) \to \mathbb{R} \) be a strictly increasing continuously differentiable function, and let \( n \) be a natural number. The following conditions are equivalent.

(i) The map \( x \mapsto \text{Tr} h^* df(x)^{-1}h \) is, for each \( h \in H_n \), concave in positive definite operators \( x \in B(H_n) \).

(ii) The map \((x, h) \mapsto \text{Tr} h^* df(x)h \) is convex in pairs of operators in \( B(H_n) \), where \( x \) is positive definite.

**Proof.** We first assume (i) and define two quadratic forms \( \alpha \) and \( \beta \) on the direct sum \( \mathcal{H}_n \oplus \mathcal{H}_n \) by setting
\[
\alpha(X \oplus Y) = \lambda \text{Tr} X^* df(A_1)X + (1 - \lambda) \text{Tr} Y^* df(A_2)Y \\
\beta(X \oplus Y) = \text{Tr} (\lambda X^* + (1 - \lambda) Y^*) df(\lambda A)(\lambda X + (1 - \lambda) Y),
\]
where \( A_1, A_2 \) are two fixed positive definite operators in \( B(H_n) \), and \( A = \lambda A_1 + (1 - \lambda) A_2 \) for some \( \lambda \in [0, 1] \). The statement of the theorem is equivalent to the majorisation

\[
\beta(X \oplus Y) \leq \alpha(X \oplus Y)
\]

for arbitrary \( X, Y \in \mathcal{H}_n \). Let \( (e_i)^n_{i=1} \) be a basis in which \( x \) is diagonal and let \( \lambda_1, \ldots, \lambda_n \) be the corresponding eigenvalues counted with multiplicity. Expressed in this basis \( df(x)h = h \circ L_f(\lambda_1, \ldots, \lambda_n) \) is the Hadamard (entry-wise) product of \( h \) and the Löwner matrix

\[
L_f(\lambda_1, \ldots, \lambda_n) = ([\lambda_i, \lambda_j]_f)_{i,j=1}^n,
\]

where the divided difference \([t, s]_f \) is defined by setting

\[
[t, s]_f = \begin{cases} 
\frac{f(t) - f(s)}{t - s} & t \neq s \\
\frac{f'(t)}{t} & t = s.
\end{cases}
\]

The quadratic form \( h \mapsto \text{Tr} h^* df(x)h \) is positive definite since

\[
\text{Tr} h^* df(x)h = \sum_{i,j=1}^n |(he_i | e_j)|^2 [\lambda_i, \lambda_j]_f,
\]

and \([\lambda_i, \lambda_j]_f > 0 \) because \( f \) is strictly increasing. We also notice that the corresponding sesqui-linear form is given by

\[
(h, h') \mapsto \text{Tr} h' df(x)h.
\]

The two quadratic forms \( \alpha \) and \( \beta \) are in particular positive definite. Therefore, there exists an operator \( \Gamma \) on \( \mathcal{H}_n \oplus \mathcal{H}_n \) which is positive definite in the Hilbert space structure given by \( \beta \) such that

\[
\alpha(X \oplus Y, X' \oplus Y') = \beta(\Gamma(X \oplus Y), X' \oplus Y') \quad X, X', Y, Y' \in \mathcal{H}_n,
\]

where we retain the notation \( \alpha \) and \( \beta \) also for the corresponding sesqui-linear forms. Let \( \gamma \) be an eigenvalue of \( \Gamma \) corresponding to an eigenvector \( X \oplus Y \). Then

\[
\alpha(X \oplus Y, X' \oplus Y') = \beta(\gamma(X \oplus Y), X' \oplus Y') \quad \text{for} \quad X', Y' \in \mathcal{H}_n
\]
or equivalently

\[ \lambda \text{Tr} X' df(A_1)X + (1 - \lambda) \text{Tr} Y' df(A_2)Y = \gamma \text{Tr} (\lambda X' + (1 - \lambda)Y') df(A)(\lambda X + ((1 - \lambda)Y) \]

for arbitrary \( X', Y' \in H_n \). From this we may derive the identities

\[ df(A_1)X = \gamma df(A)(\lambda X + (1 - \lambda)Y) = df(A_2)Y \]

and thus by setting \( M = df(A)(\lambda X + (1 - \lambda)Y) \), we obtain

\[ df(A)^{-1}(M) = \lambda X + (1 - \lambda)Y = \lambda df(A_1)^{-1}(\gamma M) + (1 - \lambda) df(A_2)^{-1}(\gamma M). \]

By multiplying from the left with \( M^* \) and taking the trace we obtain

\[ \gamma (\lambda \text{Tr} M^* df(A_1)^{-1}M + (1 - \lambda) \text{Tr} M^* df(A_2)^{-1}M) = \text{Tr} M^* df(A)^{-1}M \]

\[ \geq \lambda \text{Tr} M^* df(A_1)^{-1}M + (1 - \lambda) \text{Tr} M^* df(A_2)^{-1}M, \]

where the last inequality is implied by the concavity of \( x \mapsto \text{Tr} h^* df(x)^{-1}h \). This shows that the positive definite operator \( \Gamma \geq 1 \) from which (1) and thus statement (ii) of the theorem follows.

If we instead assume statement (ii) in the theorem and consider the same construction as above, then the eigenvalue \( \gamma \geq 1 \) and the last inequality therefore implies that

\[ \text{Tr} M^* df(A)^{-1}M \geq \lambda \text{Tr} M^* df(A_1)^{-1}M + (1 - \lambda) \text{Tr} M^* df(A_2)^{-1}M \]

for each \( M \in H_n \) on the form \( M = df(A)(\lambda X + (1 - \lambda)Y) \). Since the Fréchet differential \( df(A) \) is bijective, any vector \( M \in H_n \) may be written in this form. We conclude that the map \( x \mapsto df(x)^{-1} \) is concave which is statement (i) in the theorem. \( \text{QED} \)

Condition (ii) in the above theorem is obviously preserved under convex combinations of functions. We therefore realise that each set \( \Phi_n \) is convex. The set of matrix entropies is therefore a convex cone.
3 Proof of the main theorem

Consider a function $g: D \to \mathbb{R}$ of two variables defined in a convex domain $D \subseteq \mathbb{R}^2$. Let $x$ and $y$ be commuting self-adjoint operators on a Hilbert space of finite dimension $n$ with spectra $\sigma(x)$ and $\sigma(y)$ such that $\sigma(x) \times \sigma(y) \subseteq D$. We say that $(x, y)$ is in the domain of $g$. There exists a common resolution $P_1, \ldots, P_n$ of the identity in one-dimensional projections such that

$$x = \sum_{i=1}^{n} \lambda_i P_i \quad \text{and} \quad y = \sum_{i=1}^{n} \mu_i P_i,$$

where $\lambda_1, \ldots, \lambda_n$ and $\mu_1, \ldots, \mu_n$ respectively are the eigenvalues of $x$ and $y$ counted with multiplicity. The functional calculus is defined by setting

$$g(x, y) = \sum_{i=1}^{n} g(\lambda_i, \mu_i) P_i.$$

We recall that $g$ is said to be matrix convex of order $n$ if

$$g(\lambda x_1 + (1 - \lambda)x_2, \lambda y_1 + (1 - \lambda) y_2) \leq \lambda g(x_1, y_1) + (1 - \lambda) g(x_2, y_2)$$

for sets of commuting operators $(x_1, y_1)$ and $(x_2, y_2)$, and $\lambda \in [0, 1]$ such that the commutator

$$[\lambda x_1 + (1 - \lambda)y_1, \lambda x_2 + (1 - \lambda)y_2] = 0.$$

If the commutativity condition (3) is not satisfied for a $\lambda \in [0, 1]$ then the condition in (2) is void.

We say that $g$ is operator convex if $g$ is matrix convex of all orders.

**Proof of Theorem 1.2** (i) $\Rightarrow$ (ii). Let $\varphi$ be the representing function of a matrix entropy and set $f = \varphi'$. We consider operators on $H_n$ for a fixed but arbitrary $n$. The map of two variables

$$(x, h) \mapsto \text{Tr} h^* df(x) h$$

is then convex by Theorem 2.1. Fix an operator $h$ in $B(H_n)$ and consider self-adjoint operators $y \in B(H_n)$. By composing with the linear map $y \mapsto yh$, we obtain that the map

$$(x, y) \mapsto \text{Tr} h^* y df(x)(yh)$$

is then convex by Theorem 2.1.
is convex in pairs of operators \((x, y)\) such that \(x\) is positive definite and \(y\) is self-adjoint. Since the Fréchet differential \(df(x)\) acts as a linear operator on \(H_n\) we observe that

\[
df(x)(yh) = df(x)(Ly h) = (df(x)Ly)(h),
\]
where \(Ly\) denotes left multiplication with \(y\). The map

\[
(x, y) \mapsto \text{Tr} h^* \left( L_y df(x) Ly \right) (h)
\]

is therefore convex and since \(h\) is arbitrary, we obtain that the map

\[
(x, y) \mapsto Ly df(x) Ly \in B(H_n)
\]
is convex in pairs of operators \((x, y)\) such that \(x\) is positive definite and \(y\) is self-adjoint. To each \(t \in [0, 1]\) we set \(x_t = (1 - t)x + ty\). By composing with the linear map \((x, y) \mapsto (x_t, y - x)\) we obtain that the map

\[
(x, y) \mapsto L_{y-x} df(x_t) L_{y-x} \in B(H_n)
\]
is convex in pairs of positive definite operators. We then define an operator \(T(x, y) \in B(H_n)\) by setting

\[
(4) \quad T(x, y) = L_{y-x} \int_0^1 df(x_t) L_{y-x} dt
\]

for positive definite operators on \(H_n\). It follows that the map \((x, y) \mapsto T(x, y)\) is convex in positive definite operators. We now consider the real function,

\[
(5) \quad g(t, s) = (t - s)(f(t) - f(s)) \quad t, s > 0,
\]

and intend to show that it is operator convex. Take commuting positive definite operators \(x\) and \(y\). Since \(x_t\) and \(y - x\) commute we obtain

\[
df(x_t)(y - x) = f'(x_t)(y - x),
\]
hence

\[
(6) \quad T(x, y)h = (y - x) \int_0^1 f'(x_t)(y - x)h dt
\]

\[
= (y - x)(f(y) - f(x))h
\]

\[
= g(x, y)h
\]
for each $h \in B(H_n)$. For commuting positive definite operators $x$ and $y$, the left multiplication by $g(x,y)$ thus coincides with the action of the convex operator map $(x,y) \mapsto T(x,y) \in B(H_n)$. This implies that $g$ is operator convex.

If $x$ and $y$ commute, then so does $x$ and $x + y$. We may therefore apply the transformation $(t,s) \mapsto (t,t + s)$ to obtain that also the function

$$(t,s) \mapsto t(f(t + s) - f(s))$$

is operator convex; in particular it is operator convex in the one variable $s$. After division by $t^2$ we obtain that the function

$$s \mapsto \frac{f(t + s) - f(s)}{t}$$

is operator convex. Finally, by letting $t \to 0$, we obtain $f'$ as the point-wise limit of a sequence of operator convex functions. Hence $f'$ is operator convex.

We also need to prove that $f'$ is numerically decreasing. However, since by assumption $x \to \text{Tr} h^* df(x)^{-1} h$ is concave for fixed $h$, we obtain by restricting to numbers that the function

$$t \to \frac{s^2}{f'(t)}$$

is concave for fixed $s$. Since $f'$ is positive it follows that

$$t \to \frac{1}{f'(t)}$$

is increasing. Thus $f'$ is numerically decreasing. QED

Proof of Theorem 1.2 (ii) $\Rightarrow$ (i). We retain the notation as above and define the positive function

$$k(t,s) = \frac{f(t) - f(s)}{t - s} = \int_0^1 f'((\lambda t + (1 - \lambda)s) \, d\lambda \quad t, s > 0,$$

where we used Hermite’s formula. Let now $x$ be a positive definite operator in $B(H_n)$ and take an orthonormal basis $(e_1, \ldots, e_n)$ in which $x$ is diagonal with eigenvalues given by

$$xe_i = \lambda_i e_i \quad i = 1, \ldots, n.$$
By calculation we obtain that the expectation of the Fréchet differential is given by

$$\text{Tr } h^* df(x)h = \sum_{i,j=1}^{n} |(he_i | e_j)|^2 \frac{f(\lambda_i) - f(\lambda_j)}{\lambda_i - \lambda_j} = \text{Tr } h^* k(L_x, R_x)h,$$

where $L_x$ and $R_x$ denote left and right multiplication with $x$, respectively.

Putting the formulas together we obtain the identity

$$\text{Tr } h^* df(x)h = \int_0^1 \text{Tr } h^* f'(\lambda L_x + (1-\lambda)R_x)h d\lambda.$$

The first author proved [6, Page 100] that a mapping of the type

$$(x, \xi) \mapsto (f(x)\xi | \xi)$$

is convex in pairs consisting of a positive definite operator $x$ on a Hilbert space $H$ and a vector $\xi \in H$, if $f$ is operator convex and decreasing. Subsequently, Ando and Hiai [11, Theorem 3.1] proved that the condition is not only sufficient but also necessary.

Since the transformation $x \mapsto \lambda L_x + (1-\lambda)R_x$ is affine and $f'$ is operator convex and numerically decreasing, we realise that the mapping of two variables in (7) is convex. By Theorem 2.1 it now follows that $f$ is the derivative of the representing function of a matrix entropy. QED

Notice that a function $f : (0, \infty) \to \mathbb{R}$ with operator convex and numerically decreasing derivative $f'$ necessarily is operator monotone.

We still need to establish the integral formula in Theorem 1.2. But it follows directly from the characterisation of operator convex and numerically decreasing functions, see for example the analysis in [5, Page 9 and 10].

### 4 Appendix

A corollary to the above analysis gives a result of independent interest in the theory of operator monotone functions.

**Theorem 4.1.** Let $f : (0, \infty) \to \mathbb{R}$ be an increasing differentiable function such that the derivative $f'$ is operator convex and numerically decreasing. The function

$$(x, y) \mapsto \text{Tr } (y - x)(f(y) - f(x))$$

## 9
of two variables is then convex in positive definite operators acting on a finite dimensional Hilbert space.

Proof. Consider again the operator $T(x, y)$ defined in equation (4). By taking the expectation of $T(x, y)$ in the unit operator we obtain the identity

\[
\text{Tr} T(x, y) = \text{Tr} L_{y-x} \int_0^1 df(x_t)(y-x) \, dt = \text{Tr} (y-x)(f(y) - f(x)),
\]

cf. for example [4, Theorem 2.1]. The result then follows by noticing that the map $(x, y) \mapsto T(x, y)$ is convex under the given assumptions. QED

Consider an arbitrary function $f : (0, \infty) \to \mathbb{R}$. One may ask what can be inferred about $f$ if the function defined in (8) is convex? Since

\[
(y - x)(f(y) - f(x)) = yf(y) - yf(x) - xf(y) + xf(x),
\]

we realise by setting $y = h^* h$ for a fixed $h$ that the function

\[
x \mapsto -\text{Tr} h^* f(x) h + \text{Tr} xf(x)
\]

is convex. We may replace $h$ with $th$ for any constant $t \neq 0$ and obtain after division by $t^2$ that

\[
x \mapsto -\text{Tr} h^* f(x) h + \frac{1}{t^2} \text{Tr} xf(x)
\]

is convex for arbitrary $h$. By letting $t \to \infty$ we realise that $f$ is operator concave. If $f$ is positive then it is also operator monotone, cf. for example [8, Theorem 2.3]. However, if we choose an extreme operator monotone function which is of the form

\[
f(t) = \frac{t}{t + \lambda} \quad t > 0
\]

for some $\lambda > 0$, then the corresponding function in (8) is not convex.
References

[1] T. Ando and F. Hiai. Operator log-convex functions and operator means. Mathematische Annalen, 350(3):611–630, 2011.

[2] S. Boucheron, O. Bousquet, G. Lugosi and P. Massart. Moment inequalities for functions of independent random variables. The Annals of Probability, 33(2): 514-560, 2005.

[3] R.Y. Chen and J.A. Tropp, Subadditivity of matrix $\varphi$-entropy and concentration of random matrices. arXiv:1308.2952v1, 13 Aug., 2013.

[4] F. Hansen and G.K. Pedersen. Perturbation formulas for traces on $C^*$-algebras. Publ. RIMS, Kyoto Univ., 31:169–178, 1995.

[5] F. Hansen. Trace functions as Laplace transforms. Journal of Mathematical Physics, 47:043504, 2006.

[6] F. Hansen. Extensions of Lieb’s concavity theorem. Journal of Statistical Physics, 124:87-101, 2006.

[7] F. Hansen. Trace functions with applications in quantum physics. Journal of Statistical Physics, DOI 10.1007/s10955-013-0890-x. November, 2013.

[8] F. Hansen. The fast track to Löwner’s theorem. Linear Algebra Appl., 438: 4557-4571, 2013.

[9] R. Latała and C. Oleszkiewicz. Between Sobolev and Poincaré. Geometric Aspects of Functional Analysis, Israel Semiar (GAFA) 1996-2000. Lecture Notes in Math. 1745: 147-168, 2000. Springer, Berlin. MR1796718.

[10] E. Lieb. Convex trace functions and the-Wigner-Yanase-Dyson conjecture. Advances in Math., 11:267–288, 1973.

Frank Hansen: Institute for International Education, Tohoku University, Japan. Email: frank.hansen@m.tohoku.ac.jp.

Zhihua Zhang: School of Mathematical Sciences, University of Electronic Science and Technology of China, P. R. China, and Department of Mathematics, Graduate School of Science, Tohoku University, Japan.