Quantum covariance, quantum Fisher information
and the uncertainty principle

Paolo Gibilisco\textsuperscript{1,4}, Fumio Hiai\textsuperscript{2,5} and Dénes Petz\textsuperscript{3,6}

\textsuperscript{4} Dipartimento SEFEMEQ, Università di Roma “Tor Vergata”,
Via Columbia 2, 00133-Rome, Italy.

\textsuperscript{5} Graduate School of Information Sciences, Tohoku University
Aoba-ku, Sendai 980-8579, Japan

\textsuperscript{6} Alfréd Rényi Institute of Mathematics,
H-1364 Budapest, POB 127, Hungary

Abstract

In this paper the relation between quantum covariances and quantum Fisher
informations are studied. This study is applied to generalize a recently proved
uncertainty relation based on quantum Fisher information. The proof given here
considerably simplifies the previously proposed proofs and leads to more general
inequalities.

2000 Mathematics Subject Classification. Primary 62B10, 94A17; Secondary 46L30,
46L60.

Key words and phrases. Quantum covariance, generalized variance, uncertainty
principle, operator monotone functions, quantum Fisher information.

\textsuperscript{1}E-mail: gibilisco@volterra.uniroma2.it
\textsuperscript{2}E-mail: hiai@math.is.tohoku.ac.jp. Partially supported by Grant-in-Aid for Scientific Research
(B)17340043.
\textsuperscript{3}E-mail: petz@math.bme.hu. Partially supported by the Hungarian Research Grant OTKA T068258
and 49835.
1 Introduction

Fisher information has been an important concept in mathematical statistics and it is an ingredient of the Cramér-Rao inequality. It was extended to a quantum mechanical formalism in the 1960’s by Helstrom [9] and later by Yuen and Lax [26], see [10] for the rigorous version.

The state of a finite quantum system is described by a density matrix $D$ which is positive semi-definite with $\text{Tr} \, D = 1$. If $D$ depends on a real parameter $-t < \theta < t$, then the true value of $\theta$ can be estimated by a self-adjoint matrix $A$, called observable, such that

$$\text{Tr} \, D_\theta A = \theta.$$ 

This means that expectation value of the measurement of $A$ is the true value of the parameter (unbiased measurement). When the measurement is performed (several times on different copies of the quantum system), the average outcome is a good estimate for the parameter $\theta$.

It is convenient to choose the value $\theta = 0$. Then the Cramér-Rao inequality has the form

$$\text{Tr} \, D_0 A^2 \geq \frac{1}{\text{Fisher information}},$$

where the Fisher information quantity is determined by the parametrized family $D_\theta$ and it does not depend on the observable $A$, see [10, 21].

The Fisher information depends on the tangent of the curve $D_\theta$. There are many curves through the fixed $D_0$ and the Fisher information is defined on the tangent space. The latter is the space of traceless self-adjoint matrices in case of the affine parametrization of the state space. The Fisher information is a quadratic form depending on the foot point $D_0$. If it should generate a Riemannian metric, then it should depend on $D_0$ smoothly [1].

2 From coarse-graining to Fisher information and covariance

Heuristically, coarse-graining implies loss of information, therefore Fisher information should be monotone under coarse-graining. This was proved in [3] in probability theory and a similar approach was proposed in [16] for the quantum case. The approach was completed in [19], where a class of quantum Fisher information quantities was introduced, see also [20].

Assume that $D_\theta$ is a smooth curve of density matrices with tangent $A := \dot{D}_0$ at $D_0$. The quantum Fisher information $F_D(A)$ is an information quantity associated with the pair $(D_0, A)$ and it appeared in the Cramér-Rao inequality above. Let now $\alpha$ be a
coarse-graining, that is $\alpha : M_n \to M_k$ is a completely positive trace-preserving mapping. Then $\alpha(D_0)$ is another curve in $M_k$. Due to the linearity of $\alpha$, the tangent at $\alpha(D_0)$ is $\alpha(A)$. As it is usual in statistics, information cannot be gained by coarse graining, therefore we expect that the Fisher information at the density matrix $D_0$ in the direction $A$ must be larger than the Fisher information at $\alpha(D_0)$ in the direction $\alpha(A)$. This is the monotonicity property of the Fisher information under coarse-graining:

$$F_D(A) \geq F_{\alpha(D)}(\alpha(A)) \quad (1)$$

Another requirement is that $F_D(A)$ should be quadratic in $A$, in other words there exists a (non-degenerate) real positive bilinear form $\gamma_D(A, B)$ on the self-adjoint matrices such that

$$F_D(A) = \gamma_D(A, A) \quad (2)$$

The requirements (1) and (2) are strong enough to obtain a reasonable but still wide class of possible quantum Fisher informations.

The bilinear form $\gamma_D(A, B)$ can be canonically extended to the positive sesqui-linear form (denoted by the same $\gamma_D$) on the complex matrices, and we may assume that

$$\gamma_D(A, B) = \text{Tr} \ A^* J_D^{-1}(B)$$

for an operator $J_D$ acting on matrices. (This formula expresses the inner product $\gamma_D$ by means of the Hilbert-Schmidt inner product and the positive linear operator $J_D$.) Note that this notation transforms (1) into the relation

$$\alpha^* J_D^{-1} \alpha \leq J_D^{-1},$$

which is equivalent to

$$\alpha J_D \alpha^* \leq J_{\alpha(D)}. \quad (3)$$

Under the above assumptions, there exists a unique operator monotone function $f : \mathbb{R}^+ \to \mathbb{R}$ such that $f(t) = tf(t^{-1})$ and

$$J_D = f(L_D R_D^{-1})R_D, \quad (4)$$

where the linear transformations $L_D$ and $R_D$ acting on matrices are the left and right multiplications, that is

$$L_D(X) = DX \quad \text{and} \quad R_D(X) = XD.$$

To be adjusted to the classical case, we always assume that $f(1) = 1 \quad [19, 22]$. It seems to be convenient to call a function $f : \mathbb{R}^+ \to \mathbb{R}^+$ standard if $f$ is operator monotone, $f(1) = 1$ and $f(t) = tf(t^{-1})$. (A standard function is essential in the context of operator means [12, 19].)
If \( D = \text{Diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) \) (with \( \lambda_i > 0 \)), then
\[
\gamma_D(A, B) = \sum_{ij} \frac{1}{M_f(\lambda_i, \lambda_j)} A_{ij} B_{ij},
\]
(5)
where \( M_f \) is the mean induced by the function \( f \):
\[
M_f(a, b) := bf(a/b).
\]
When \( A \) and \( B \) are self-adjoint, the right-hand-side of (5) is real as required since \( M_f(a, b) = M_f(b, a) \).

Similarly to Fisher information, the covariance is a bilinear form as well. In probability theory, it is well-understood but the non-commutative extension is not obvious. The monotonicity under coarse-graining should hold:
\[
\text{qCov}_D(\alpha^*(A), \alpha^*(A)) \leq \text{qCov}_{\alpha(D)}(A, A),
\]
(6)
where \( \alpha^* \) is the adjoint with respect to the Hilbert-Schmidt inner product. (\( \alpha^* \) is a unital completely positive mapping.) If the covariance is expressed by the Hilbert-Schmidt inner product as
\[
\text{qCov}_D(A, B) = \text{Tr} A^* \mathbb{K}_D(B),
\]
then the monotonicity (6) has the form
\[
\alpha \mathbb{K}_D \alpha^* \leq \mathbb{K}_{\alpha(D)}.
\]
This is actually the same relation as (3). Therefore, condition (6) implies
\[
\text{qCov}_D(A, B) = \text{Tr} A^* \mathbb{J}_D(B),
\]
where \( \mathbb{J}_D \) is defined by (4). The one-to-one correspondence between Fisher information quantities and (generalized) covariances was discussed in [20]. The analogue of formula (4) is
\[
\text{qCov}_D(A, B) = \sum_{ij} M_f(\lambda_i, \lambda_j) A_{ij} B_{ij} - \left( \sum_i \lambda_i A_{ii} \right) \left( \sum_i \lambda_i B_{ii} \right).
\]
(7)

If we want to emphasize the dependence of the Fisher information and the covariance on the function \( f \), we write \( \gamma^f_D \) and \( \text{qCov}^f_D \). The usual symmetrized covariance corresponds to the function \( f(t) = (t + 1)/2 \):
\[
\text{qCov}^f_D(A, B) = \text{Cov}_D(A, B) := \frac{1}{2} \text{Tr} (D(A^* B + BA^*)) - (\text{Tr} DA^*) (\text{Tr} DB)
\]
Of course, if \( D, A \) and \( B \) commute, then \( \text{qCov}^f_D(A, B) = \text{Cov}_D(A, B) \) for any standard function \( f \). Note that both \( \text{qCov}^f_D \) and \( \gamma^f_D \) are particular quasi-entropies [17, 18].
3 Relation to the commutator

Let $D$ be a density matrix and $A$ be self-adjoint. The commutator $i[D, A]$ appears in the discussion about Fisher information. One reason is that the tangent space $T_D := \{ B = B^* : \text{Tr} DB = 0 \}$ has a natural orthogonal decomposition:

$$\{ B = B^* : [D, B] = 0 \} \oplus \{ i[D, A] : A = A^* \}.$$ 

For self-adjoint operators $A_1, ..., A_N$, Robertson’s uncertainty principle is the inequality

$$\text{Det} \left[ \text{Cov}_D(A_i, A_j) \right]_{i,j=1}^N \geq \text{Det} \left[ -\frac{1}{2} \text{Tr} D [A_i, A_j] \right]_{i,j=1}^N,$$

see [23]. The left-hand side is known in classical probability as the generalized variance of the random vector $(A_1, ..., A_N)$. A different kind of uncertainty principle has been recently conjectured in [5] and proved in [6, 2]:

$$\text{Det} \left[ \text{Cov}_D(A_i, A_j) \right]_{i,j=1}^N \geq \text{Det} \left[ \frac{f(0)}{2} \gamma_f(i[D, A], i[D, B]) \right]_{i,j=1}^N.$$  (8)

Particular cases of inequality (8) have been proved in [4, 7, 8, 13, 14, 15, 11, 25]. Of course, we have a non-trivial inequality in the case $f(0) > 0$. The inequality can be called dynamical uncertainty principle, since the right-hand-side is the volume of a parallelepiped determined by the tangent vectors of the trajectories of the time-dependent observables $A_i(t) := D^t A_i D^{-i}$. Another remarkable property is that inequality (8) gives a non-trivial bound also in the odd case $N = 2m + 1$ and this seems to be the first result of this type in the literature.

The right-hand-side of (8) is Fisher information of commutators. If

$$\tilde{f}(x) := \frac{1}{2} \left( (x + 1) - (x - 1)^2 f(0) \right) f(x),$$  (9)

then

$$\frac{f(0)}{2} \gamma_f(i[D, A], i[D, B]) = \text{Cov}_D(A, B) - q \text{Cov}_D \tilde{f}(A, B)$$  (10)

for $A, B \in T_D$. Identity (10) is easy to check but it is not obvious that for a standard $f$ the function $\tilde{f}$ is operator monotone. It is indeed true that $\tilde{f}$ is a standard function as well, see Propositions 5.2 and 6.3 in [7]. Note that the left-hand-side of (10) was called (metric adjusted) skew information in [8].

4 Inequalities

In this section we give a simple new proof for the dynamical uncertainty principle (8). The new proof actually gives a slightly more general inequality.

5
Theorem 1 Assume that $f, g : \mathbb{R}^+ \to \mathbb{R}$ are standard functions such that
\[ g(x) \geq c \frac{(x-1)^2}{f(x)} \quad (11) \]
for some $c > 0$. Then
\[ \text{qCov}^a_D(A, A) \geq c \gamma^f_D([D, A], [D, A]). \]

Proof: We may assume that $D = \text{Diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ and $\text{Tr} \, DA = 0$. Then the left-hand-side is
\[ \sum_{ij} M_g(\lambda_i, \lambda_j)|A_{ij}|^2 \]
while the right-hand-side is
\[ c \sum_{ij} \frac{(\lambda_i - \lambda_j)^2}{M_f(\lambda_i, \lambda_j)}|A_{ij}|^2. \]
The proof is complete. □

For any standard function $f$ and its transform $\tilde{f}$ given by (9), $\tilde{f} \geq 0$ is exactly
\[ \frac{1 + x}{2} - \frac{f(0)(x-1)^2}{2f(x)} \geq 0. \]
Therefore for $g(x) = (1 + x)/2$ the assumption (11) holds for any $f$ if $c = f(0)/2$. Actually, this is the point where the operator monotonicity of $f$ is used, in Theorem 1 only inequality (11) was essential.

The next lemma is standard but the proof is given for completeness.

Lemma 2 Let $\mathcal{K}$ be a finite dimensional real Hilbert space with inner product $\langle \cdot, \cdot \rangle$. Let $\langle \cdot, \cdot \rangle$ be a real (not necessarily strictly) positive bilinear form on $\mathcal{K}$. If
\[ \langle f, f \rangle \leq \langle f, f \rangle \]
for every vector $f \in \mathcal{K}$, then
\[ \text{Det} \left( [ \langle f_i, f_j \rangle ]_{i,j=1}^m \right) \leq \text{Det} \left( [ \langle f_i, f_j \rangle ]_{i,j=1}^m \right) \]
holds for every $f_1, f_2, \ldots, f_m \in \mathcal{K}$. Moreover, if $\langle \cdot, \cdot \rangle - \langle \cdot, \cdot \rangle$ is strictly positive, then inequality (12) is strict whenever $f_1, \ldots, f_m$ are linearly independent.

Proof: Consider the Gram matrices $G := [ \langle f_i, f_j \rangle ]_{i,j=1}^m$ and $H := [ \langle f_i, f_j \rangle ]_{i,j=1}^m$, which are symmetric and positive semidefinite. For every $a_1, \ldots, a_m \in \mathbb{R}$ we get
\[ \sum_{i,j=1}^m (\langle f_i, f_j \rangle - \langle f_i, f_i \rangle) a_ia_j = \sum_{i=1}^m a_i f_i \sum_{i=1}^m a_i f_i - \sum_{i=1}^m a_i f_i \sum_{i=1}^m a_i f_i \geq 0 \]
by assumption. This says that \( G - H \) is positive semidefinite, hence it is clear that \( \text{Det} (G) \geq \text{Det} (H) \).

Moreover, assume that \( \langle \cdot , \cdot \rangle - \langle \cdot , \cdot \rangle \) is strictly positive and \( f_1, \ldots, f_m \) are linearly independent. Then \( G - H \) is positive definite and hence \( \text{Det} (G) > \text{Det} (H) \). \( \square \)

The previous general result is used now to have a determinant inequality, an extension of the dynamical uncertainty relation.

**Theorem 3** Assume that \( f, g : \mathbb{R}^+ \to \mathbb{R} \) are standard functions such that

\[
g(x) \geq c \frac{(x - 1)^2}{f(x)}
\]

for some \( c > 0 \). Then for self-adjoint matrices \( A_1, A_2, \ldots, A_m \) the determinant inequality

\[
\text{Det} \left( [\text{qCov}_D^g (A_i, A_j)]^m_{i,j=1} \right) \geq \text{Det} \left( [c \gamma_D^f ( [D, A_i], [D, A_j])]^m_{i,j=1} \right)
\]

(13)

holds. Moreover, equality holds in (13) if and only if \( A_i - (\text{Tr} DA_i)I, 1 \leq i \leq m \), are linearly dependent, and both sides of (13) are zero in this case.

**Proof:** Let \( K \) be the real vector space \( T_D = \{ B = B^* : \text{Tr} DB = 0 \} \). We have \( \text{qCov}_D^g (A, A) = 0 \) if and only if \( A = \lambda I \), therefore

\[
\langle A, B \rangle := \text{qCov}_D^g (A, B)
\]

is an inner product on \( K \). From formulas (5), (7) and from the hypothesis, we have

\[
c \gamma_D^f ([D, A], [D, A]) = \sum_{ij} c \frac{(\lambda_i - \lambda_j)^2}{M_f(\lambda_i, \lambda_j)} |A_{ij}|^2
\]

\[
\leq \sum_{ij} M_g(\lambda_i, \lambda_j) |A_{ij}| = q \text{Cov}_D^g (A, A) = \langle A, A \rangle.
\]

If

\[
\langle A, B \rangle := c \gamma_D^f ([D, A], [D, B]) ,
\]

then \( \langle A, A \rangle \leq \langle A, A \rangle \) holds and (12) gives the statement when \( \text{Tr} DA_1 = \text{Tr} DA_2 = \ldots = \text{Tr} DA_m = 0 \). The general case follows by writing \( A_i - (\text{Tr} DA_i)I \) in place of \( A_i \), \( 1 \leq i \leq m \).

To prove the statement on equality case, we show that \( g(x) > c(x - 1)^2/f(x) \) or \( f(x)g(x) > c(x - 1)^2 \) for all \( x > 0 \). Since \( f(x)g(x) \) is increasing while \( c(x - 1)^2 \) is decreasing for \( 0 < x \leq 1 \), it is clear that \( f(x)g(x) > c(x - 1)^2 \) for \( 0 < x \leq 1 \). Since \( f(x) \) and \( g(x) \) are (operator) concave, it follows that \( f(x)g(x)/x^2 = (f(x)/x)(g(x)/x) \) is decreasing for \( x > 0 \). But \( c(x - 1)^2/x^2 \) is increasing for \( x \geq 1 \), so that we have \( f(x)g(x) > c(x - 1)^2 \) for \( x \geq 1 \) as well. The inequality shown above implies that

\[
M_g(\lambda_i, \lambda_j) > c \frac{(\lambda_i - \lambda_j)^2}{M_f(\lambda_i, \lambda_j)}
\]
for all $1 \leq i, j \leq m$. Hence $\langle \cdot, \cdot \rangle - \langle \cdot, \cdot \rangle$ is strictly positive on $K$, and the latter statement follows from Lemma 2.

Recall that (8) is obtained by the choice $g(x) = (1 + x)/2$ and $c = f(0)/2$. Assume we put $c = f(0)/2$. Then (13) holds for a standard $f$ if

$$g(x) \geq \frac{f(0)(x - 1)^2}{2f(x)}.$$ 

In particular, $g(0) \geq 1/2$. The only standard $g$ satisfying this inequality is $g(t) = (t + 1)/2$. This corresponds to the case where the left-hand-side is the usual covariance.

Motivated by [13, 24], Kosaki [11] studied the case when $f(x)$ equals to

$$h_\beta(x) = \frac{\beta(1 - \beta)(x - 1)^2}{(x^\beta - 1)(x^{1-\beta} - 1)}.$$ 

In this case $g(x) = h_\beta(x)$ is possible for every $0 < \beta < 1$ if the constant $c$ is chosen properly. More generally, inequality (13) holds for any standard $f$ and $g$ when the constant $c$ is appropriate. It follows from the lemma below that $c = f(0)g(0)$ is good, see (14).

**Lemma 4** For every standard function $f$,

$$f(x) \geq f(0) \left| x - 1 \right|.$$ 

**Proof:** The inequality is not trivial only if $f(0) > 0$ and $x > 1$, so assume these conditions. Let $q(x_0)$ be the constant such that the tangent line to the graph of $f$ at the point $x_0 > 1$ has the equation

$$y = f'(x_0)x + q(x_0).$$ 

Since $f$ is (operator) concave one has $q(x_0) \geq f(0)$. Using again (operator) concavity and symmetry one has

$$f'(x_0) \geq \lim_{x \to +\infty} f'(x) = \lim_{x \to +\infty} \frac{f(x)}{x} = \lim_{x \to +\infty} f(x^{-1}) = f(0) > 0.$$ 

This implies

$$f(x_0) = f'(x_0) \cdot x_0 + q(x_0) \geq f(0) \cdot x_0 + f(0) \geq f(0) \cdot x_0 - f(0) = f(0) \cdot (x_0 - 1)$$

and the proof is complete.

The lemma gives the inequality

$$f(x)g(x) \geq f(0)g(0)(x - 1)^2$$

for standard functions. If $f(0) > 0$ and $g(0) > 0$, then Theorem 3 applies.

Similarly to the proof of Theorem 3, one can prove that the right-hand-side of (13) is a monotone function of the variable $f$. 

8
Theorem 5 Assume that \( f, g : \mathbb{R}^+ \rightarrow \mathbb{R} \) are standard functions. If

\[
\frac{c}{f(t)} \geq \frac{d}{g(t)}
\] (15)

for some positive constants \( c, d \) and \( A_1, A_2, \ldots, A_m \) are self-adjoint matrices, then

\[
\text{Det} \left( \left[ c \gamma_D^f ([D, A_i], [D, A_j]) \right]_{i,j=1}^m \right) \leq \text{Det} \left( \left[ d \gamma_D^g ([D, A_i], [D, A_j]) \right]_{i,j=1}^m \right)
\] (16)

holds.

References

[1] S. Amari and H. Nagaoka, Methods of information geometry, Translations of Mathematical Monographs, Vol. 191, AMS and Oxford University Press, Oxford, 2000.

[2] A. Andai, Uncertainty principle with quantum Fisher information, arXiv:0707.1147

[3] N. N. Cencov, Statistical decision rules and optimal inferences, Translation of Math. Monog. 53, Amer. Math. Society, Providence, 1982.

[4] P. Gibilisco and T. Isola, Uncertainty principle and quantum Fisher information, Ann. Inst. Stat. Math, 59 (2007), 147–159.

[5] P. Gibilisco, D. Imparato and T. Isola, A volume inequality for quantum Fisher information and the uncertainty principle, to be published in J. Statist. Phys.

[6] P. Gibilisco, D. Imparato and T. Isola, A Robertson-type uncertainty principle and quantum Fisher information, to be published in Lin. Alg. Appl.

[7] P. Gibilisco, D. Imparato and T. Isola, Uncertainty principle and quantum Fisher information II, J. Math. Phys. 48 (2007), 072109, arXiv:math-ph/0701062v3.

[8] F. Hansen, Metric adjusted skew information, arXiv:math-ph/0607049v3, 2006.

[9] C. W. Helstrom, Quantum detection and estimation theory, Academic Press, New York, 1976.

[10] A. S. Holevo, Probabilistic and statistical aspects of quantum theory. North-Holland, Amsterdam, 1982.

[11] H. Kosaki, Matrix trace inequality related to uncertainty principle, Internat. J. Math. 16 (2005), 629–645.

[12] F. Kubo and T. Ando, Means of positive linear operators, Math. Ann. 246 (1979/80), 205–224.
[13] S. Luo and Z. Zhang, An informational characterization of Schrödinger’s uncertainty relations, J. Stat. Phys. 114, 1557–1576 (2004).

[14] S. Luo and Q. Zhang, On skew information, IEEE Trans. Inform. Theory, 50(2004), 1778–1782.

[15] S. Luo and Q. Zhang, Correction to On skew information. IEEE Trans. Inform. Theory 51, 4432 (2005).

[16] E.A. Morozova and N.N. Chentsov, Markov invariant geometry on state manifolds (in Russian), Itogi Nauki i Tehniki 36(1990), 69–102.

[17] M. Ohya and D. Petz, Quantum Entropy and Its Use, Springer-Verlag, Heidelberg, 1993. Second edition 2004.

[18] D. Petz, Quasi-entropies for finite quantum systems, Rep. Math. Phys., 23(1986), 57-65.

[19] D. Petz, Monotone metrics on matrix spaces. Linear Algebra Appl. 244(1996), 81–96.

[20] D. Petz, Covariance and Fisher information in quantum mechanics. J. Phys. A: Math. Gen. 35(2003), 79–91.

[21] D. Petz, Quantum Information Theory and Quantum Statistics, Springer-Verlag, Berlin and Heidelberg, 2007.

[22] D. Petz and Cs. Sudár, Geometries of quantum states, J. Math. Phys. 37(1996), 2662–2673.

[23] H. P. Robertson, An indeterminacy relation for several observables and its classical interpretation. Phys. Rev. 46, 794–801, 1934.

[24] E. P. Wigner and M. M. Yanase, Information content of distributions. Proc. Nat. Acad. Sci. USA 49(1963), 910–918.

[25] K. Yanagi, S. Furuichi and K. Kuriyama, A generalized skew information and uncertainty relation, IEEE Trans. Inform. Theory, 51(2005), 4401–4404 (2005).

[26] H. P. Yuen and M. Lax, Multiple-parameter quantum estimation and measurement of nonselfadjoint observables, IEEE Trans. Information Theory IT-19(1973), 740–750.