In Wigner phase space, convolution explains why the vacuum majorizes mixtures of Fock states

Luc Vanbever
Ecole polytechnique de Bruxelles,
CP 165, Université libre de Bruxelles, 1050 Brussels, Belgium
August 2021

Abstract
I show that a nonnegative Wigner function that represents a mixture of Fock states is majorized by the Wigner function of the vacuum state. As a consequence, the integration of any concave function over the Wigner phase space has a lower value for the vacuum state than for a mixture of Fock states. The Shannon differential entropy is an example of such concave function of significant physical importance. I demonstrate that the very cause of the majorization lies in the fact that a Wigner function is the result of a convolution. My proof is based on a new majorization result dedicated to the convolution of the negative exponential distribution with a precisely constrained function. I present a geometrical interpretation of the new majorization property in a discrete setting and extend this relation to a continuous setting. Findings presented in this article might be expanded upon to explain why the Wigner function of the vacuum majorizes - beyond mixtures of Fock states - many other physical states represented by a nonnegative Wigner function.

1 Introduction
Phase space representations were conceived almost ninety years ago by Eugene Wigner [1]. Although the original application was the description of states and systems in Quantum mechanics, the Wigner function is today a successful tool in areas as diverse as the study of radiation and optical systems [2], quantum electronics, quantum chemistry or signal analysis [1]. The Wigner phase space has also proved to be an efficient framework in the analysis of both stationary and time-dependent systems.

The Wigner function keeps providing interesting insight when analysing quantum systems. This is also strengthened by the fact that the Wigner function has a series of - almost directly visible - features that nicely reflect the physics of the quantum states they represent. For instance, negative parts in the Wigner functions are viewed as the signature of a non-classical behaviour [3], whereas in the absence of negative areas, it behaves as an actual probability distribution [4]. Another illustration is the normalization of the Wigner function that simply mirrors the unit trace of the density operator.

The entropy of the Wigner function - and more specifically Shannon differential entropy - is a quantity that relates to the amount of information that can be gained on a quantum system. When entropy relations involve the two variables of the two-dimensional Wigner phase space, they can reflect the limited amount of knowledge that can be obtained simultaneously on both variables. For instance, some lower bounds on the sum of the entropies of the two phase space variables imply Heisenberg's uncertainty relations [4, 5].

1 Note however that squeezed Gaussian states have a nonnegative Wigner function but exhibit some non-classical behaviour.
It has been conjectured that the joint entropy of the Wigner function of a coherent state is the minimum joint entropy that any physical state can have if it is represented by a nonnegative Wigner function [5]. Visually, this would mean that a Wigner function can never be “more peaked” than the Wigner function of a coherent state.

In this article, I prove this conjecture for the case of mixtures of Fock states. I use the theory of majorization and I explain why the Wigner function of the vacuum state majorizes the Wigner function of mixtures of Fock states - in the cases where this Wigner function is nonnegative. Such majorization relation directly implies entropy (and seminorms) comparisons between the various distributions.

In section 2 I start by rewriting a couple of definitions and properties that I use in this article. The seven parts of my proof are presented in section 3. Section 4 details the proof of the two main theorems of this article and I conclude in section 5.

## 2 Definitions and Properties

In this article, a vector $x$ is said to be nonnegative if all its elements are nonnegative $x_i \geq 0, \forall i$. A vector is said to be normalized if the sum of its elements is equal to 1. Similarly, in the continuous case, a function $f(x)$ defined on the measure space $(H, \mathcal{F}, \nu)$ is said to be nonnegative if it takes nonnegative values on its domain $H$. A function $f(x)$ is said to be normalized if

$$
\int f(x) \, d\nu(x) = 1,
$$

where the integration is performed over the complete domain $H$.

### Some Definitions of Majorization

I first recall a couple of equivalent definitions for majorization, both in a discrete setting and in a continuous setting. I restrict the description of the equivalent definitions to the ones that will be used in this article.

In a discrete setting, if $x, y \in \mathbb{R}^N$, the following five statements are equivalent [6, 7]:

1. The vector $x$ majorizes $y$, denoted $x \succ y$.

2.

$$
\begin{align*}
\sum_{i=1}^{k} x_i^+ &\geq \sum_{i=1}^{k} y_i^+, & k = 1, \ldots, N - 1, \quad \text{and} \\
\sum_{i=1}^{N} x_i & = \sum_{i=1}^{N} y_i,
\end{align*}
$$

(1a)

(1b)

where the vectors $x^+$ and $y^+$ include the same elements, respectively as $x$ and $y$, but re-sorted in a non-increasing order.

3.

$$
\sum_{i=1}^{N} \Phi(x_i) \geq \sum_{i=1}^{N} \Phi(y_i),
$$

(2)

for all continuous convex functions $\Phi : \mathbb{R} \rightarrow \mathbb{R}$.

4. There exists a $N \times N$ doubly stochastic matrix $D$ such that

$$
y = Dx.
$$

(3)
5. The vector $y$ can be derived from $x$ by performing a limited number of Robin Hood transfers. A transfer of a positive amount $\Delta$ from $x_i$ to $x_j$, giving $y_i = x_i - \Delta$ and $y_j = x_j + \Delta$ is said to be a Robin Hood transfer if $\Delta \leq x_i - x_j$. This definition is valid for nonnegative vectors $x$ and $y$. A Robin Hood transfer is also called a Dalton transfer or sometimes a “pinch” [7].

Similarly, in a continuous setting, if $f$ and $g$ are nonnegative $\nu$-integrable functions defined on the measure space $(H, \mathcal{F}, \nu)$, the following four statements are equivalent [8, 7]:

1. The function $f$ majorizes $g$, denoted $f \succ g$.

2.
\[
\int_0^t f_\downarrow(u) \, d\nu(u) \geq \int_0^t g_\downarrow(u) \, d\nu(u), \quad \forall t \in [0, \nu(H)) \quad \text{and} \quad (4a)
\]
\[
\int f(u) \, d\nu(u) = \int g(u) \, d\nu(u), \quad (4b)
\]

where $f_\downarrow$ (and identically for $g_\downarrow$) denotes the decreasing rearrangement of $f$ defined in two steps:

\[
m_f(t) = \nu(\{x : f(x) > t\}), \quad t \geq 0, \quad \text{and},
\]
\[
f_\downarrow(u) = \sup \{t : m_f(t) > u\}, \quad 0 \leq u \leq \nu(H). \quad (5b)
\]

3. For all continuous convex functions $\Phi : \mathbb{R} \rightarrow \mathbb{R}$ with $\Phi(0) = 0$, for which the integrals exist.

4.
\[
g(y) = \int k(x, y) \, f(x) \, d\nu(x), \quad (7)
\]

where $k : H \times H \rightarrow [0, +\infty)$ is a doubly stochastic function, that is, $\int k(x, y) \, d\nu(x) = 1$ for all $y \in H$ and $\int k(x, y) \, d\nu(y) = 1$ for all $x \in H$. Note that some authors present this condition as sufficient and not as an equivalent definition [8]. In this article, Eq. (7) is only used as a sufficient condition for majorization.

A number of properties of discrete majorization extend to continuous majorization and we note the similarities between the equivalent definitions in the discrete and in the continuous settings (not all are reproduced here). In the continuous setting, some of the equivalent definitions are applicable to nonnegative functions; this nonnegativity restriction can be relaxed if $\nu(H)$ is finite [7]. I keep this restriction as the support of the functions studied in this article is not finite.

**Majorization Property of the Convolution of Two Probability Distributions**

The convolution vector of two probability distributions is majorized by the two original probability distributions [3]. I particularize here the proof [6, 7] to the case where the first probability vector has one more element than the second probability vector. The discrete convolution of a normalized vector $y = (y_1, y_2, \ldots, y_{N+1})^T$ and a vector $x = (x_1, x_2, \ldots, x_N)^T$ is formulated as the multiplication by a Toeplitz matrix constructed as follows:
where the symbol * denotes the convolution. The resulting probability vector \( z \) has a maximum of \( 2N \) non-zero elements. If we build a \( 2N \times 2N \) circulant matrix \( C \) by adding \( N \) columns at the right of the above Toeplitz matrix, the probability vector \( z \) can also be written as

\[
\begin{pmatrix}
  y_1 & 0 & \cdots & 0 \\
y_2 & y_1 & \ddots & \vdots \\
  \vdots & y_2 & \ddots & 0 \\
y_N & \vdots & \ddots & y_1 \\
y_{N+1} & y_N & \ddots & y_2 \\
  0 & y_{N+1} & \ddots & \vdots \\
  \vdots & \ddots & \ddots & y_N \\
  0 & \cdots & 0 & y_{N+1}
\end{pmatrix}
\begin{pmatrix}
  x_1 \\
x_2 \\
  \vdots \\
x_N \\
  0 \\
  0 \\
  \vdots \\
  0
\end{pmatrix},
\]

(8)

where \( N \) zero elements are added to the vector \( x \). With Eq. (5), the equality in Eq. (9) proves that \( x \succ z \) because the circulant matrix \( C \) is doubly stochastic.

These results have been generalized to the continuous distributions. In particular, if \( X \) and \( Y \) are independent random variables with absolutely continuous distribution functions, then the distribution of \( X + Y \) is majorized by both the distributions of \( X \) and \( Y \) [8] - the explicit proof is not provided in the referenced article.

If \( c, f : \mathbb{R}^+ \to \mathbb{R} \) are two functions whose support is the positive real axis, the convolution of \( c(x) \) and \( f(x) \) is

\[
g(x) = c * f = \int_0^x c(t) f(x - t) \, dt, \quad x \geq 0.
\]

(10)

A discrete version of this integral matches the formulation involving the Toeplitz matrix of Eq. (8).

The Wigner Function of the Fock States

The definition of the Wigner function of the quantum state with density operator \( \hat{\rho} \) is [9]

\[
W(x, p) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} (x - \frac{x'}{2} |\hat{\rho}|p + \frac{x'}{2}) e^{ix'p} \, dx',
\]

(11)
where $x$ and $p$ are canonically conjugate variables.

With this dimensionless definition, the Wigner function of the eigenstate of the Number operator with integer eigenvalue $n$ reads \[ W_n(r, \theta) = \tilde{W}_n(r^2) = \frac{(-1)^n}{\pi} L_n(2r^2) e^{-r^2}, \] where $r$ and $\theta$ are the polar coordinates in Wigner phase space and $L_n(x)$ denotes the Laguerre polynomial of the $n$-th order. The eigenstates of the Number operator are referred to as Fock states and the state corresponding to $n = 0$ as the vacuum state.

A Set of States with Nonnegative Wigner Function

Some recent results about nonnegative Wigner functions will be used in my proof. When injecting the Fock states $|m\rangle$ and $|n\rangle$ into the two inputs of a balanced beamsplitter, the states at each of the two outputs of the balanced beamsplitter can be written as the following mixture of Fock states \[ \hat{\sigma}(m,n) = \left(\frac{m!}{m+n} \right)^{-1} \sum_{z=0}^{m+n} \sum_{i=\max(0,z-n)}^{\min(z,m)} \sum_{j=\max(0,z-n)}^{\min(z,m)} (-1)^{i+j} \times \binom{m}{i} \binom{n}{z-i} \binom{m}{j} \binom{n}{z-j} z! (m+n-z)! |z\rangle\langle z|. \] It was proved that the states $\hat{\sigma}(m,n)$ have a nonnegative Wigner function \[ 10\ 11\]. For more details about the generation and properties of the states $\hat{\sigma}(m,n)$, I refer the reader to the article \[ 10\] and to the articles cited therein.

In the present article, I will use Eq. (13) as the definition of a parameterized set of states that have a nonnegative Wigner function.

3 Proof

In this section, I prove that the Wigner function of the vacuum majorizes the Wigner function of any mixture of Fock states when the latter Wigner function is nonnegative. The proof consists of seven parts.

Definition 1

In the Wigner phase space, the Wigner function $W_\alpha$ majorizes $W_\beta$, denoted as $W_\alpha \succ W_\beta$, if

\[ \int \Phi(W_\alpha(\vec{s})) \, d\vec{s} \geq \int \Phi(W_\beta(\vec{s})) \, d\vec{s}, \] for all continuous convex functions $\Phi : \mathbb{R} \rightarrow \mathbb{R}$ with $\Phi(0) = 0$, for which the integrals over the complete 2-dimensional phase space exist.

\[ 2\] A convex function of utmost interest is the opposite of the Shannon differential entropy, for which $\Phi(W(\vec{s})) = W(\vec{s}) \log(W(\vec{s}))$ with the Wigner function $W(\vec{s})$ assumed nonnegative.
For Wigner functions that do not depend on the angular coordinate but only depend on the radial coordinate \( r \), the majorization in Eq. (14) is an application of the definition in Eq. (6) with the measure \( \nu = r^2 \). I use the index \( W \) below the majorization symbol in \( W_\alpha \succ_W W_\beta \) to refer to the majorization with respect to the measure \( \nu = r^2 \).

In this article, I study convex combinations of Fock states and I only consider Wigner functions that depend on \( r^2 \), as defined in Eq. (12). A direct application of the definition in Eq. (14) to two such functions, \( W_\alpha = W_\alpha (r^2) \) and \( W_\beta = W_\beta (r^2) \) shows that

\[
W_\alpha (r^2) \succ_W W_\beta (r^2) \iff W_\alpha (z) \succ W_\beta (z),
\]

where the change of variable \( z = r^2 \) was performed. The majorization on the right-hand side of Eq. (15) is the majorization with respect to the usual measure, that is, \( \nu (z) = z \). The domain of definition of the functions \( W_\alpha (z) \) and \( W_\beta (z) \) is \( [0, +\infty) \). As an example, to compare the Wigner function of the vacuum and the Wigner function of an equally weighted mixture of the first two Fock states, we can write

\[
W_0 \succ_W \frac{1}{2} (W_0 + W_1) \iff e^{-z} \succ ze^{-z},
\]

where \( z \in \mathbb{R}^+ \).

Formulation as a Convolution

Thanks to Eq. (15), to obtain a majorization relation between the vacuum and the Wigner function \( W_\beta (r^2) \) of a convex combination of Fock states, we can compare \( W_\alpha (z) = e^{-z} \) with \( W_\beta (z) \). To perform this comparison, it will prove useful to formulate \( W_\beta (z) \) as a convolution.

Eq. (12) of a Fock state shows that \( W_\beta (z) \) is the product of a polynomial with \( e^{-z} \) denoted here by \( W_\beta (z) = P (z) e^{-z} \).

The expression \( P (z) e^{-z} \) can also be viewed as a linear combination of Erlang distributions of rate parameter equal to 1. The support of the Erlang distribution is \( \mathbb{R}^+ \) and its general expression is

\[
E_{k+1}(z) = \frac{z^k e^{-z}}{k!} = p_k(z) e^{-z}, \quad k = 0, 1, 2, \ldots \quad \text{and} \quad z \geq 0,
\]

which defines \( p_k(z) \). The Erlang distribution of shape parameter \( k + 1 \) is the convolution as defined by Eq. (10), of an Erlang distribution of shape parameter \( k \) with the negative exponential distribution. If I also introduce the derivative of the polynomial \( p_k(z) \), for \( k \geq 1 \),

\[
p_k(z) e^{-z} = E_{k+1}(z) = E_k(z) * e^{-z} = (p_k(z) e^{-z}) * e^{-z}, \quad z \geq 0.
\]

\( P (z) \) is a linear combination of the \( p_k(z) \) and Eq. (18) gives

\[
P (z) e^{-z} = \sum_{k=0}^{+\infty} a_k p_k(z) e^{-z} = a_0 e^{-z} + \left( \sum_{k=1}^{+\infty} a_k p_k(z) e^{-z} \right) * e^{-z}, \quad z \geq 0.
\]

The function \( W_\beta (z) \) can finally be formulated as the following convolution:
\[ W_\beta(z) = P(z) e^{-z} = (a_0 \delta(z) + P'(z) e^{-z}) * e^{-z}, \quad z \geq 0, \quad (20) \]

where the Dirac delta function takes the term \( k = 0 \) of the sum into account and \( a_0 \) was implicitly defined in Eq. (19) as the zero order coefficient of \( P(z) \).

Note that the fact that \( E_{k+1}(z) \) is obtained from the convolution of the distributions \( E_k(z) \) and \( e^{-z} \) implies that \( P(z) e^{-z} \prec e^{-z} \) in the cases where the coefficients of the polynomial \( P(z) \) are nonnegative \( [8] \). This proves for instance that \( W_0 \succ W_1 \).

Eq. (20) actually provides an indirect representation of \( W_\beta(r^2) \) as the convolution of a continuous function (except in \( z = 0 \)) with the negative exponential \( e^{-z} \) that represents the vacuum.

**Theorem 1**

Theorem 1 is a new majorization property applicable to the convolution of a discrete negative exponential vector with a vector that can have negative elements.

Let the non-normalized vector \( v_0 = (a^0, a^1, a^2, \ldots, a^{N-1})^T \) with \( 0 < a < 1 \) and let \( x \) be a normalized \( N+1 \)-dimensional vector that can be written as the sum

\[ x^T = (1, 0, \ldots, 0) + \sum_{k=1}^{N} (0, \ldots, 0, -\lambda_k, \lambda_k, 0, \ldots, 0), \quad (21) \]

where \( \lambda_k \geq 0, \ k = 1, \ldots, N, \) and the \( k \)-th and \( k+1 \)-th elements are the only non-zero elements of each vector in the sum of Eq. (21).

Let \( G \) be the \( 2N \)-dimensional vector defined by the convolution \( G = x * v_0 \). My main theorem states that

\[ \text{If} \quad G \geq 0 \quad \text{Then} \quad v_0 \succ G. \quad (22) \]

The proof of Theorem 1 is provided in section 3.

**Lemma 1**

A vector \( x \in \mathbb{R}^{N+1} \) can be decomposed into a sum of the form

\[ x^T = (\lambda_0, 0, \ldots, 0) + \sum_{k=1}^{N} (0, \ldots, 0, -\lambda_k, \lambda_k, 0, \ldots, 0), \quad (23) \]

where all \( \lambda_k \geq 0 \), if and only if

\[ \sum_{i=k}^{N} x_i \geq 0, \quad k = 0, \ldots, N. \quad (24) \]

To first prove the inverse implication, it suffices to decompose the vector \( x \) into a sum of terms of the form \( (0, \ldots, 0, -\lambda_k, \lambda_k, 0, \ldots, 0)^T \), which also proves the unicity of the decomposition. As per Eq. (24), \( x_N \) is positive and the last term of the sum must be \( (0, \ldots, 0, -x_N, x_N)^T \). Eq. (24) directly shows that the \( k \)-th term \( (0, \ldots, 0, -\sum_{i=k}^{N} x_i, \sum_{i=k}^{N} x_i, 0, \ldots, 0)^T \) of the decomposition has the form required in Eq. (23). Finally, because \( \lambda_0 = \sum_{i=0}^{N} x_i \), the parameter \( \lambda_0 \geq 0 \) and \( \lambda_0 = 1 \) if the vector \( x \) is normalized.

The direct implication derives from the fact that a single vector of the form \( (0, \ldots, 0, -\lambda_k, \lambda_k, 0, \ldots, 0)^T \) with \( \lambda_k \geq 0 \) satisfies Eq. (24) and that the sum of vectors that satisfies Eq. (24) also satisfies that same equation.
Theorem 2

Theorem 2 is the extension of Theorem 1 - combined with Lemma 1 - to the continuous case. Let \( c(z) \) be a function defined on \([0, +\infty)\), integrable and continuous on its domain, except in \( z = 0 \) where one term of the function \( c(z) \) can be proportional to a Dirac delta function. The function \( c(z) \) also fulfills the conditions

\[
\int_{x}^{+\infty} c(z) \, dz \geq 0, \quad \forall \, x > 0 \quad \text{and} \\
\int_{0}^{+\infty} c(z) \, dz = 1. \tag{25a}
\]

Let \( e^{-z} \) denote the negative exponential distribution defined on \([0, +\infty)\) and let the function \( g \) be the convolution \( g(z) = c(z) * e^{-z} \) as defined by Eq. (10). Theorem 2 states that

\[
\text{If } g(z) \geq 0 \text{ Then } e^{-z} \succ g(z). \tag{26}
\]

The proof of Theorem 2 is provided in section 4.

Lemma 2

The Wigner function of the equally weighted mixture of the first \( M + 1 \) Fock states is nonnegative, for all \( M \in \mathbb{N} \).

To prove this result, I consider the \( M + 1 \) states \( \hat{\sigma}(m, M - m), \ m = 0, \ldots, M \). The definition of \( \hat{\sigma}(m, n) \) was reproduced in Eq. (13) and I keep a constant value for \( M = m + n \),

\[
\hat{\sigma}(m, M - m) = \sum_{z=0}^{M} a_{m,z}|z\rangle\langle z|, \quad m = 0, \ldots, M, \tag{27}
\]

where

\[
a_{m,z} = 2^{-M} \sum_{i=\max(0,z+m-M)}^{\min(z,m)} \sum_{j=\max(0,z+m-M)}^{\min(z,m)} (-1)^{i+j} \binom{m}{i} \binom{M-m}{z-i} z!(M-z)! \binom{M}{j} \binom{M-m}{z-j} / (m!(M-m)!). \tag{28}
\]

We can notice that each term of the double sum in Eq. (28) has a symmetry relatively to the variables \( m \) and \( z \):

\[
\binom{m}{i} \binom{M-m}{z-i} z!(M-z)! = \binom{z}{i} \binom{M-z}{m-i} m!(M-m)!, \tag{29}
\]

and that the indices \( i \) and \( j \) also run through the same values when \( m \) and \( z \) are swapped. This indicates that \( a_{m,z} = a_{z,m} \) and the normalization \( \sum_{z=0}^{M} a_{m,z} = 1 \) for \( m = 0, \ldots, M \), implies that summing on the index \( m \) also gives the value \( \sum_{m=0}^{M} a_{m,z} = 1 \) for \( z = 0, \ldots, M \).
Hence, the mixture with equal weights of the states $\hat{\sigma} (m,M - m)$ provides

\[
\frac{1}{M+1} \sum_{m=0}^{M} \hat{\sigma} (m,M - m) = \frac{1}{M+1} \sum_{z=0}^{M} |z\rangle\langle z|,
\] (30)

which has a nonnegative Wigner function. This completes the proof of Lemma 2.

**Final Result**

I have established in Eq. (20) that a function of the form $W_{\beta} (z) = P(z) e^{-z}$ can be expressed as the convolution of $a_0 \delta(z) + P'(z) e^{-z}$ with the negative exponential $e^{-z}$ distribution, where $a_0$ is the zero order coefficient of the polynomial $P(z)$. The variable $z \in \mathbb{R}^+$ and the convolution is defined by Eq. (10).

To analyse mixtures of Fock states, I choose $P(z) = (-1)^n L_n (2z)$, where $L_n (x)$ is the Laguerre polynomial of the $n$-th order. The first entry condition of Theorem 2, Eq. (25a), becomes

\[
\int_{x}^{+\infty} (-1)^n L'_n (2z) e^{-z} dz \geq 0, \quad \forall x > 0,
\] (31)

where $L'_n (2z)$ denotes the first derivative of the polynomial $L_n (2z)$ with respect to the variable $z$. The integral over the complete positive real axis shows that the second entry condition of Theorem 2, Eq. (25b), is satisfied:

\[
\int_{0}^{+\infty} (-1)^n (\delta(z) + L'_n (2z) e^{-z}) dz = 1.
\] (32)

The nonnegativity of the integral in Eq. (31) can be verified by using the recursion relation obtained as follows:

\[
\int_{x}^{+\infty} L'_n (2z) e^{-z} dz = \int_{x}^{+\infty} L'_{n-1} (2z) e^{-z} dz - 2 \int_{x}^{+\infty} L_{n-1} (2z) e^{-z} dz,
\] (33a)

\[
= - \int_{x}^{+\infty} L'_{n-1} (2z) e^{-z} dz + 2 \int_{x}^{+\infty} (L'_{n-1} (2z) e^{-z} - L_{n-1} (2z) e^{-z}) dz,
\]

\[
= - \int_{x}^{+\infty} L'_{n-1} (2z) e^{-z} dz - 2 L_{n-1} (2x) e^{-x}.
\] (33b)

where I have used the relation $L'_n (x) = L'_{n-1} (x) - L_{n-1} (x)$ applicable to a sequence of Laguerre polynomials in Eq. (33a) and integrated by parts in Eq. (33b). By making use of this recursion, we eventually get

\footnote{The physical reason for this result is not commented upon in this article.}
\[
\int_x^{+\infty} (-1)^n L_n'(2z) e^{-z} dz = 2 \sum_{i=0}^{n-1} (-1)^i L_i(2x) e^{-x}.
\] (34)

The integral in Eq. (31) is proportional to the radial profile of the Wigner function of an equally weighted mixture of the first \(n-1\) Fock states with the radial coordinate \(r = x^{1/2}\). By Lemma 2, this quantity is nonnegative \(\forall x \geq 0\). This means that each function \(\tilde{W}_n(z) = \frac{(-1)^n}{z} L_n(2z) e^{-z}\) can be retrieved as the convolution of \(e^{-z}\) with a generalized function that meets the entry conditions of Theorem 2, Eqs. (25a) and (25b).

We note also - as I did already in Lemma 1 - that if various functions fulfill Eqs. (25a) and (25b), then a convex combination of those functions also fulfills the entry conditions of Theorem 2 - including in the case where the convex combination comprises an infinite number of such functions.

My final result is then a direct corollary of Eqs. (12), (15) and Theorem 2: if the Wigner function of a mixture of Fock states is nonnegative, then it is majorized by the Wigner function of the vacuum.

4 Proofs of Theorem 1 and Theorem 2

Proof of Theorem 1

Let \(v_0 = (a^0, a^1, a^2, \ldots, a^{N-1})^\top\) with \(0 < a < 1\) and let \(v_k\) and \(u_k, k = 1, \ldots, N\) be the \(N\)-dimensional vectors defined by the elementary convolution

\[
\begin{pmatrix}
0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \vdots & \ddots & \vdots & \vdots \\
-1 & 0 & \vdots & \vdots & \vdots \\
1 & -1 & \ddots & \vdots & \vdots \\
0 & 1 & \ddots & 0 & \vdots \\
\vdots & \vdots & \ddots & -1 & 0 \\
\vdots & \vdots & \ddots & 1 & -1 \\
0 & 0 & \cdots & 0 & 0
\end{pmatrix}
\begin{pmatrix}
a^0 \\
a^1 \\
\vdots \\
a^{N-1}
\end{pmatrix} =
\begin{pmatrix}
0 \\
\vdots \\
0 \\
-a^0 \\
a^0 - a^1 \\
\vdots \\
a^{N-2} - a^{N-1} \\
a^{N-1} \\
0 \\
\vdots \\
0
\end{pmatrix}
\]

where \(-a^0\) and \(a^{N-1}\) are respectively the \(k\)-th and the \(k+N\)-th elements of the vector \(\begin{pmatrix} v_k \\ u_k \end{pmatrix}\). The convolution \(G\) is defined as

\[
G = \begin{pmatrix} V \\ U \end{pmatrix} = \begin{pmatrix} v_0 \\ 0 \end{pmatrix} + \sum_{k=1}^{N} \lambda_k \begin{pmatrix} v_k \\ u_k \end{pmatrix}, \quad \lambda_k \geq 0,
\]

(35)
where \( \bar{0} \) denotes the vector of \( N \) zero values. Theorem 1 states that

\[
\text{If } \begin{pmatrix} V \\ U \end{pmatrix} \succeq 0 \text{ Then } \begin{pmatrix} v_0 \\ 0 \end{pmatrix} \succ \begin{pmatrix} V \\ U \end{pmatrix}.
\]

(37)

For the sake of clarity and to provide a geometric interpretation of the theorem, I first present the proof for \( N = 3 \).

Depending on the context, I use the words vector and point interchangeably and similarly for the words elements and coordinates (of a point).

Because \( \lambda_k \geq 0 \) in Eq. (36), the vector \( U \) is always nonnegative and the nonnegativity of \( \begin{pmatrix} V \\ U \end{pmatrix} \) is determined by the elements of the vector \( V \).

In \( \mathbb{R}^3 \), the point \( V \) is in a convex cone \( C \) whose vertex is in \( v_0 \) and that is defined by the three linearly independent vectors \( v_1, v_2 \) and \( v_3 \). This convex cone and its position with respect to the axes \( X, Y \) and \( Z \) are schematized in Figure 1.

A given point \( V \) in \( \mathbb{R}^3 \) uniquely defines a point \( \begin{pmatrix} V \\ U \end{pmatrix} \) in \( \mathbb{R}^6 \) via Eqs. (35) and (36). The subset of points \( V \) of the convex cone \( C \) such that \( \begin{pmatrix} v_0 \\ 0 \end{pmatrix} \succ \begin{pmatrix} V \\ U \end{pmatrix} \) is convex; this is a direct consequence of the convexity of majorization [7].

Figure 1: 3-dimensional convex cone spanned by the three vectors \( v_1, v_2 \) and \( v_3 \) and cut by the planes \( X = 0, Y = 0 \) and \( Z = 0 \). Any point \( V \) with positive coordinates in this convex cone corresponds to a 6-dimensional vector \( \begin{pmatrix} V \\ U \end{pmatrix} \) majorized by \( \begin{pmatrix} v_0 \\ 0 \end{pmatrix} \).
The proof proceeds in two steps. In a first step, I decompose the convex cone \( C \) into six convex cones. In each of those six convex cones, one of the plane \( X = 0, Y = 0 \) or \( Z = 0 \) separates the points with nonnegative coordinates from the points that have at least one negative coordinate. In the second step, I complete the proof by showing that \( v_0 \) majorizes each vertices of the six convex cones cut by one of the plane \( X = 0, Y = 0 \) or \( Z = 0 \).

Step 1. To explain the first step, I replace the vectors \( v_1, v_2 \) and \( v_3 \) by the proportional vectors \( v_{0\lambda\lambda}, v_{0\lambda\lambda} \) and \( v_{0\lambda\lambda} \) respectively. The example of the vector \( v_{0\lambda\lambda} \) can be used to describe how the indices are chosen: the 0 as second index in \( v_{0\lambda\lambda} \) indicates that the second coordinate of \( v_0 + v_{0\lambda\lambda} \) is zero whereas the \( \lambda \) as 1\textsuperscript{st} and 3\textsuperscript{rd} indices of \( v_{0\lambda\lambda} \) indicate that for this vector, \( \lambda_1 = \lambda_3 = 0 \) in Eq. (36). This means that the 1\textsuperscript{st} and 3\textsuperscript{rd} coordinates of \( v_0 + v_{0\lambda\lambda} \) are nonnegative as a direct consequence of Eq. (35). The points \( v_0 + v_{i\lambda} \) with \( i = 0 \) or \( i = \lambda \), are illustrated in Figure 1.

Let \( V \) be a given point of the convex cone \( C \) in \( \mathbb{R}^3 \),

\[
V = v_0 + \lambda'_1 v_{0\lambda\lambda} + \lambda'_2 v_{0\lambda\lambda} + \lambda'_3 v_{0\lambda\lambda}, \quad \lambda'_k \geq 0.
\]  

(38)

As can be visualized in Figure 1 and checked in Table 1 the vector \( \bar{0} = v_0 + v_{000} \) is within the convex cone \( C \) and can be used to decompose it into three convex cones. I assume without loss of generality that \( V \) is in the convex cone \( C_3 \) defined by the three vectors \( v_{0\lambda\lambda}, v_{0\lambda\lambda} \) and \( v_{000} \). \( C_3 \) can be retrieved in Eq. (39), so that \( V \) is in the convex cone \( C_{31} \) defined as

\[
V = v_0 + \lambda''_1 v_{0\lambda\lambda} + \lambda''_2 v_{0\lambda\lambda} + \lambda''_3 v_{000}, \quad \lambda''_k \geq 0.
\]  

(39)

I further decompose the convex cone \( C_3 \) by introducing the linear combination of the first two vectors that span the cone, \( v_{0\lambda\lambda} = \mu_1 v_{0\lambda\lambda} + \mu_2 v_{0\lambda\lambda} \), where the two first coordinates of \( v_0 + v_{0\lambda\lambda} \) equal to zero and, for this vector, \( \lambda_3 = 0 \) in Eq. (36). It will be clear in Table 1 that the vector \( v_0 + v_{0\lambda\lambda} \) is within the convex cone \( C_3 \) (both \( \mu_1 \) and \( \mu_2 \) are nonnegative) and can be used to decompose \( C_3 \) into two convex cones. I assume without loss of generality that the vector \( v_{0\lambda\lambda} \) can be replaced in Eq. (39) so that \( V \) is in the convex cone \( C_{31} \) defined as

\[
V = v_0 + \lambda'''_1 v_{0\lambda\lambda} + \lambda'''_2 v_{0\lambda\lambda} + \lambda'''_3 v_{000}, \quad \lambda'''_k \geq 0.
\]  

(40)

In summary, the initial convex cone \( C \) can be decomposed into six convex cones. The initial convex cone \( C \) is the union of the six convex cones and any point of the convex cone \( C \) can be retrieved in one of the six convex cones. Each of the six convex cones has properties similar to \( C_{31} \):

- The convex cone \( C_{31} \) is defined by the three vectors \( v_{0\lambda\lambda}, v_{0\lambda\lambda} \) and \( v_{000} \), denoted below by \( v_{0i} \) where \( i = 0 \) or \( i = \lambda \);
- by construction, each of the three vectors \( v_0 + v_{0i} \) has the first coordinate \( X = 0 \) and a nonnegative value for the two other coordinates;
- as a consequence, within the convex cone \( C_{31} \), the plane \( X = 0 \) separates the points that have all nonnegative coordinates from the points that have at least the first coordinate \( X < 0 \). The hatched area in Figure 1 illustrates this separation for the convex cone \( C_{31} \).

Step 2. The second step of the proof consists in proving that the vertices of the six convex cones limited by one of the plane \( X = 0, Y = 0 \) or \( Z = 0 \) correspond to vectors \( \begin{pmatrix} V \\ U \end{pmatrix} \) that are all majorized by \( v_0 \).

Table 1 shows the coordinates of each of the points \( v_0 + v_{i\lambda} \) in \( \mathbb{R}^3 \), the corresponding values \( \lambda_1, \lambda_2 \) and \( \lambda_3 \) as well as the corresponding vectors \( \begin{pmatrix} V \\ U \end{pmatrix} \) in \( \mathbb{R}^6 \).
Table 1: Values of $V\top$ and corresponding $(V\top,U\top)$ for the vertices of the polytope defined by the convex cone $C$ cut by the planes $X = 0$, $Y = 0$ and $Z = 0$.

| $v_0 + v_{i_1i_2...i_N}$ | $V\top$ | $(V\top,U\top)$ | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ |
|--------------------------|---------|------------------|-------------|-------------|-------------|
| $v_0$                    | $(a^0,a^1,a^2)$ | $(a^0,a^1,a^2,0,0,0)$ | 0           | 0           | 0           |
| $v_0 + v_{000}$          | $(0,0,0)$   | $(0,0,0,a^0,a^1,a^2)$ | 1           | 1           | 1           |
| $v_0 + v_{0\lambda\lambda}$ | $(0,a^0,a^1)$ | $(0,a^0,a^1,a^2,0,0)$ | 1           | 0           | 0           |
| $v_0 + v_{\lambda\lambda\lambda}$ | $(a^0,0,a^1)$ | $(a^0,0,a^1,a^2,a^3,0)$ | 0           | $a^1$       | 0           |
| $v_0 + v_{\lambda\lambda0}$ | $(a^0,a^1,0)$ | $(a^0,a^1,0,a^2-a^3,a^3-a^4,a^4)$ | 0           | 0           | $a^2$       |
| $v_0 + v_{000}$          | $(a^0,0,0)$   | $(a^0,0,0,a^1-a^2,a^3)$ | 1           | 0           | $a^1$       |
| $v_0 + v_{0\lambda\lambda}$ | $(0,a^0,0)$ | $(0,a^0,0,a^1-a^3,a^3)$ | 1           | 1           | 0           |

It is directly apparent in Table 1 that the eight vectors $\left(\begin{array}{c} V \\ U \end{array}\right)$ are majorized by the first vector $\left(\begin{array}{c} v_0 \\ 0 \end{array}\right)$. The first vector $\left(\begin{array}{c} v_0 \\ 0 \end{array}\right)$ indeed majorizes itself and each other vector $\left(\begin{array}{c} V \\ U \end{array}\right)$ may be obtained from $\left(\begin{array}{c} v_0 \\ 0 \end{array}\right)$ by performing Robin Hood transfers.

The convexity property mentioned earlier completes the proof of the theorem in the case $N = 3$.

The following three observations remain valid for any value $N$.

- a majorized vector $\left(\begin{array}{c} V \\ U \end{array}\right)$ can be the convolution of $v_0$ with a vector that has negative elements.

As an example, the point $v_0 + v_{0\lambda\lambda}$ corresponds to the convolution of $(a^0,-a^1,a^1)^\top$ with $v_0$;

- a majorized vector $\left(\begin{array}{c} V \\ U \end{array}\right)$ can correspond to the convex combination of vectors of the convex cone $C$ that have negative elements. As an example, the point $v_0 + v_{0\lambda\lambda}$ is the convex combination with equal weights of $v_0 + 2v_1$ and $v_0 + 2v_2$ that both contain negative elements;

- it is not necessary for the convolution to have the form specified by Eq. (36) to get a vector majorized by $v_0$. As an example, the convolution with $\lambda_1 = a$, $\lambda_2 = -\epsilon$ and $\lambda_3 = a^2$ with $0 < \epsilon \leq (1-a)^2$ defines a point outside of the convex cone $C$ that corresponds to a vector $\left(\begin{array}{c} V \\ U \end{array}\right)$ majorized by $v_0$.

The same two steps of the proof for $N = 3$ directly generalize to other values for $N$.

Step 1. The step by step decomposition of the convex cone generalizes in the $N$-dimensional space into a decomposition that will end up with $N!$ convex subcones.

A point of the initial convex cone $C$ is defined as

$$
V = v_0 + \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_{N-1} v_{N-1} + \lambda_N v_N, \quad \lambda_k \geq 0, \quad (41a)
$$

$$
v_0 + \lambda_1 v_{0\lambda\lambda} + \lambda_2 v_{0\lambda0} + \cdots + \lambda_N v_{\lambda\lambda\lambda} + \lambda'_{N-1} v_{\lambda\lambda0} + \lambda'_N v_{\lambda0\lambda}, \quad \lambda'_k \geq 0, \quad (41b)
$$

where, in Eq. (41b), the $N$ indices uniquely define each vector $v_{i_1i_2...i_N}$ and are chosen as in the case $N = 3$: an index $i_k = 0$ indicates that the $k$-th coordinate of $v_0 + v_{i_1i_2...i_N}$ is zero and an index $i_k = \lambda$ indicates that $\lambda_k = 0$ when this vector is expressed in the initial convex cone defined by Eq. (36).

It will be shown in Step 2 that all vectors $v_0 + v_{i_1i_2...i_N}$, with any combination of the indices, are inside the convex cone $C$. This property, added to the fact that the $N$ vectors $v_k$ are linearly independent,
allows us to express any vector of the initial convex cone as a vector of one of the \(N\) subcones:  

\[
V = v_0 + \lambda'_1 v_0 \alpha_1 + \lambda'_2 v_0 \alpha_2 + \cdots + \lambda'_N v_0 \alpha_N + \lambda'' v_0 \ldots 0, \quad \lambda'' \geq 0, \tag{42}
\]

where the first vector has one index \(i_1 = 0\), the second vector has two indices \(i_1 = i_2 = 0\), up to the last vector with \(N\) zero indices \(v_0..0\).

Similarly to the \(N = 3\) case, the following statements apply to the \(N\) convex subcones (the indices and coordinates must be permuted to describe properly each subcone):

- The \(N\)-dimensional convex subcone is defined by its vertex \(v_0\) and by the \(N\) vectors \(v_0 \alpha_1, \ldots, v_0 \ldots 0\),
- the \(N\) vectors \(v_0 + v_0 \alpha_1, \ldots, v_0 + v_0 \ldots 0\) and \(v_0 + v_0 \ldots 0\) have a first coordinate \(X_1 = 0\) and a nonnegative value for the other coordinates;
- as a consequence and by linearity, within the convex subcone defined by the \(N\) vectors \(v_0 \alpha_1, \ldots, v_0 \ldots 0\), the hyperplane \(X_1 = 0\) separate the points that have all nonnegative coordinates from the points that have at least the first coordinate \(X_1 < 0\).

Step 2. Because the initial convex cone \(C\) is the union of \(N\) convex subcones similar to the subcone defined by Eq. \(42\), it must be proved that the vectors \(v_0 + v_{1,2}\ldots i_N\) with all possible combinations of the indices \(i_k = 0\) or \(i_k = \lambda\) correspond to 2\(N\)-dimensional vectors \(\begin{pmatrix} V \\ U \end{pmatrix}\) that are majorized by \(v_0\).

I use a step by step procedure to construct the points \(v_0 + v_{1,2}\ldots i_N\) of the convex cone \(C\) such that either \(i_k = 0\) or \(i_k = \lambda\), starting from the lowest index.

Let us assume that the \(j\)-th element is the first (lowest index) zero of a given vector \(v_0 + v_{1,2}\ldots i_N\): \(i_j = 0\) and \(i_1 = \cdots = i_{j-1} = \lambda\). We can observe in Eq. \(35\) that any vector \(v_k\) has only one negative element, its \(k\)-th element, whose value is \(-a^0 = -1\).

The initial value of the \(j\)-th element of \(v_0\) is \(a^j - 1\) and - within the convex cone \(C\) - a zero can only be obtained by adding a vector \(a^j \cdot \bar{v}_j\) to \(v_0\). Using Eq. \(35\), the effects of adding \(a^j \cdot \begin{pmatrix} v_j \\ u_j \end{pmatrix}\) to \(\begin{pmatrix} v_0 \\ 0 \end{pmatrix}\) can be described as follows:

- The first \(j - 1\) elements of \(V\) are unchanged and keep their values \(a^0, a^1, \ldots, a^{j-2}\);  
- the \(j\)-th element of \(V\) decreases from the value \(a^{j-1}\) to 0; we can consider that this element decreases by slices, that is, first from \(a^{j-1}\) to \(a^j\), then from \(a^j\) to \(a^{j+1}\), and so on, until the \(N\)-th slice, from \(a^{N+j-2}\) to 0;
- the decrease in value of the \(j\)-th element of \(V\) is compensated by increases of the values of the \(N\)-th subsequent elements of \(\begin{pmatrix} V \\ U \end{pmatrix}\): the \(j + 1\)-th element is increased by the first slice, that is, from \(a^j\) to \(a^{j-1}\), the next element is increased by the second slice from \(a^{j+1}\) to \(a^j\) and so on. Those transfers within the \(V\) part of \(\begin{pmatrix} V \\ U \end{pmatrix}\) are all Robin Hood transfers;
- after the transfers to the \(V\) part of \(\begin{pmatrix} V \\ U \end{pmatrix}\) are completed, the subsequent slices are transferred to the \(U\) part of \(\begin{pmatrix} V \\ U \end{pmatrix}\). Before these transfers, all the elements of \(U\) had a zero value, which means that all the transfers from the \(j\)-th element of \(V\) to the \(U\) part of \(\begin{pmatrix} V \\ U \end{pmatrix}\) are also Robin Hood transfers;
- after this first step, the \(N\) elements of \(\begin{pmatrix} V \\ U \end{pmatrix}\) that have been increased (\(j + 1\)-th to \(j + N\)-th) have the form \(a^j\) or \(a^j - a^{j+p}\) where \(p\) is a positive integer and where the exponent \(i\) increases continuously from element to element, by steps equal to 1.
The effect of adding an additional vector $a^{j-2}\binom{v_j}{u_j}$ to the obtained vector can be sliced into Robin Hood transfers in the same way as for the first vector. After each addition, the vector $\binom{V}{U}$ keeps the same structure and in particular, the elements that follow the element that has just been put to zero keep the same structure $a^i$ or $a^i - a^{i+p}$ where $p$ is a positive integer and where the exponent $i$ increases from element to element, by steps equal to 1.

Two steps of this procedure are illustrated in Table 2 to construct the vector $v_0 + v_{\lambda_0\lambda_0\lambda}$ in the case $N = 7$, first for $j = 3$ and then for $j' = 6$.

\[
\begin{align*}
\begin{pmatrix} v_0 \\ \lambda \end{pmatrix} & \quad \rightarrow \quad \begin{pmatrix} v_3 \\ \lambda \end{pmatrix} + a^2 \begin{pmatrix} u_3 \\ \lambda \end{pmatrix} \quad \rightarrow \quad \begin{pmatrix} v_6 \\ \lambda \end{pmatrix} + a^4 \begin{pmatrix} u_6 \\ \lambda \end{pmatrix} \\
\begin{pmatrix} a^0 \\ a^1 \\ a^2 \\ a^3 \\ a^4 \\ a^5 \\ a^6 \\ \lambda \end{pmatrix} & \quad \rightarrow \quad \begin{pmatrix} a^0 \\ a^1 \\ a^2 \\ a^3 \\ a^4 \\ a^5 \\ a^6 - a^7 \\ \lambda \end{pmatrix} \quad \rightarrow \quad \begin{pmatrix} a^0 \\ a^1 \\ a^2 \\ a^3 \\ a^4 \\ a^5 \\ a^6 - a^7 \\ a^8 - a^9 \end{pmatrix} \end{align*}
\]

Table 2: Illustration of the two successive steps to construct the vector $v_0 + v_{\lambda_0\lambda_0\lambda}$ in the case $N = 7$. It is clear that $v_0$ majorizes the vector $\binom{V}{U}$ corresponding to $v_0 + v_{\lambda_0\lambda_0\lambda}$, which in turn majorizes the vector $\binom{V}{U}$ corresponding to $v_0 + v_{\lambda_0\lambda_0\lambda}$. The construction described above demonstrates how a vector $v_0 + v_{i_1 i_2 \ldots i_N}$ with any combination of the indices $i_k = 0$ or $i_k = \lambda$, can be obtained from the initial vector $\binom{v_0}{\lambda}$ by successively adding vectors $\lambda_k \binom{v_k}{u_k}$ for each position $k$ where the index $i_k = 0$ in $v_0 + v_{i_1 i_2 \ldots i_N}$. Each addition of vector $\lambda_k \binom{v_k}{u_k}$ - with $\lambda_k$ properly chosen - can be decomposed into Robin Hood transfers. Consequently, the vector $\binom{v_0}{\lambda}$ majorizes all vectors $\binom{V}{U}$ constructed in this manner. This completes the proof of Theorem 1.

**Proof of Theorem 2**

I outline (only) the main steps of the extension of Theorem 1 to the continuous case. My approach consists in showing that the continuous convolution of Theorem 2 can be obtained as the limit of a sequence of vectors that match the entry criteria of Theorem 1.
Let $c(z)$ be the function defined in the statement of Theorem 2 in section 3.

I define an interval $[z_0, z_{\text{end}}]$ on the positive real axis with $z_0 = 0$ and I partition that interval into $2N$ equal sub-intervals such that $z_i - z_{i-1} = \Delta z$, $i = 1, \ldots, 2N$. Within the first $N$ sub-intervals $[z_{i-1}, z_i]$, $i = 1, \ldots, N$, I choose $z_i'$ so that $c(z_i')$ is the supremum of the function $c(z)$ on that sub-interval.

I construct a $N + 1$-dimensional vector $C_{N+1}$ that fulfills (for $\Delta z$ sufficiently small) the condition specified by Eq. (44):

$$C_{N+1} = \hat{K}_{N+1} \begin{pmatrix} c_{\delta 0} \\ c(z_1') (1 - e^{-\Delta z}) \\ \vdots \\ c(z_N') (1 - e^{-\Delta z}) \\ c_{\text{last}} \frac{(1 - e^{-\Delta z})}{\Delta z} \end{pmatrix},$$

(43)

where $c_{\delta 0}$ is the factor multiplying the Dirac delta function of $c(z)$ in $z = 0$, $\hat{K}_{N+1}$ ensures that the vector $C_{N+1}$ is normalized and

$$c_{\text{last}} = \int_{z_N}^{+\infty} c(z) \, dz.$$  

(44)

I also define the $N$-dimensional vector $e_0 = (e^{-0\Delta z}, e^{-1\Delta z}, e^{-2\Delta z}, \ldots, e^{-(N-1)\Delta z})^T$.

It can be seen that the discrete convolution $G_{2N} = C_{N+1} * e_0$ is nonnegative by decomposing the continuous convolution $g(z) = c(z) * e^{-z}$ into a discrete sum. It is calculated at one of the points that define the first $N - 1$ sub-intervals, $z = z_j$ as follows

$$g(z_j) = \sum_{i=1}^{j} \int_{z_{i-1}}^{z_i} c(x) e^{-(z_j - x)} \, dx$$

(45a)

$$= c_{\delta 0} e^{-z_j} + \sum_{i=1}^{j} c(z_i') \int_{z_{i-1}}^{z_i} e^{-(z_j - x)} \, dx$$

(45b)

$$= c_{\delta 0} e^{-z_j} + (1 - e^{-\Delta z}) \sum_{i=1}^{j} c(z_i'') e^{-(z_j - z_i)},$$

(45c)

where the points $z_i''$ within each sub-intervals are determined by the mean value theorem.

The points $z_i'$ were chosen so that $c(z_i') \geq c(z_i'')$, $i = 1, \ldots, N$. The sum in Eq. (45c) has the same form as the discrete convolution $G_{2N}$ and it shows that if the continuous convolution $g(z)$ is nonnegative, then the $j + 1$-th element of the discrete convolution $G_{2N}$ is nonnegative - also at the points where the continuous convolution $g(z)$ might have a zero value.

A second variable name $y$ is associated with the index $j$ and the $j + 1$-th element of the discrete convolution $G_{2N} = C_{N+1} * e_0$ is denoted $G_{2N}(y_j)$, $j = 0, \ldots, 2N - 1$.

The three vectors $C_{N+1}$, $e_0$ and $G_{2N}$ are within the conditions of Theorem 1 for any sufficiently small value of $\Delta z$ as well as for any choice of $z_{\text{end}}$ that defines the initial interval.

Because in this case $e_0 \succ G_{2N}$, Eq. (3) can be used to write the $2N$ equations
\[ G_{2N} (y_j) = \sum_{i=0}^{N-1} D_{2N} (z_i, y_j) e^{-z_i}, \quad j = 0, \ldots, 2N - 1, \quad (46) \]

where \(0 \leq D_{2N} (z_i, y_j) \leq 1\) are the elements of a \(2N \times 2N\) doubly stochastic matrix.

As \(\Delta z\) tends to zero, \(N\) to infinity and \(z_{\text{end}}\) to infinity,

- the vector \(G_{2N}\) tends to the continuous convolution \(g(z)\) of the distribution \(e^{-z}\) with the function \(c(z)\) (both are integrable functions and improper integrals are absolutely convergent);
- the sum in Eq. (46) takes the form of a convergent Riemann–Stieltjes integral whose integrand is \(e^{-z}\) and whose integrator function is monotone and of bounded variation (this integrator function could be a generalized function).

At the limit, the sum in Eq. (46) identifies with the condition for continuous majorization set out in Eq. (7). This completes the proof of Theorem 2.

5 Conclusion

I have proved that the nonnegative Wigner function of any mixture of Fock states is majorized by the Wigner function of the vacuum.

The fact that the Wigner function results from a convolution is revealed to be a decisive factor in this majorization relation. I have described the underlying new majorization property dedicated to convolutions with the negative exponential distribution.

As a corollary to the majorization relation, the Shannon differential entropy of the Wigner function of the vacuum is lower than the entropy of the Wigner function of any mixture of Fock states. This same order between the Wigner functions of the vacuum and any mixture of Fock states is also respected for any other entropy, provided that the latter is defined by a concave function.

I conjecture that properties similar to the ones described in this article may explain further majorization relations between the Wigner function of the vacuum and other states than mixtures of Fock states - after all, any Wigner function can be obtained as a convolution of the Sudarshan-Glauber P representation.

References

[1] J. Weinbub and D. K. Ferry. Recent advances in wigner function approaches. *Applied Physics Reviews*, 5(4):041104, 2018.

[2] Ivan V. Bazarov. Synchrotron radiation representation in phase space. *Phys. Rev. ST Accel. Beams*, 15:050703, May 2012.

[3] Anatole Kenfack and Karol yczkowski. Negativity of the wigner function as an indicator of non-classicality. *Journal of Optics B: Quantum and Semiclassical Optics*, 6(10):396–404, aug 2004.

[4] Iwo Białynicki-Birula and Jerzy Mycielski. Uncertainty relations for information entropy in wave mechanics. *Communications in Mathematical Physics*, 44(2):129–132, 1975.

[5] Anaëlle Hertz, Michael G Jabbour, and Nicolas J Cerf. Entropy-power uncertainty relations: towards a tight inequality for all gaussian pure states. *Journal of Physics A: Mathematical and Theoretical*, 50(38):385301, 2017.
[6] Raymond J. Hickey. Majorisation, randomness and some discrete distributions. *Journal of Applied Probability*, 20(4):897–902, 1983.

[7] Barry C Marshall Albert W; Olkin Ingram; Arnold. *Inequalities : Theory of Majorization and Its Applications*. Springer, New York, 2010.

[8] Raymond J. Hickey. Continuous majorisation and randomness. *Journal of Applied Probability*, 21(4):924–929, 1984.

[9] Ulf Leonhardt. *Essential Quantum Optics: From Quantum Measurements to Black Holes*. Cambridge University Press, 2010.

[10] Zacharie Van Herstraeten and Nicolas J. Cerf. Quantum wigner entropy, 2021.

[11] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information (Cambridge Series on Information and the Natural Sciences)*. Cambridge University Press, 1 edition, January 2004.

[12] Peter Day. Decreasing rearrangements and doubly stochastic operators. *Transactions of The American Mathematical Society*, 178:383–383, 04 1973.

[13] Harry Joe. An ordering of dependence for distribution of k-tuples, with applications to lotto games. *Canadian Journal of Statistics*, 15(3):227–238, 1987.