The Uncertainty Principle: A creative formula

Ilya Kavalerov*

December 26, 2017

Abstract

There is a need in machine learning to find representations of signals in physical quantities suited for particular learning tasks: feature extraction. To understand these representations, it is of interest to compare how the energy of the original signal behaves in feature space. The general framework of operators in Hilbert spaces provides uncertainty principles that could be specified to many different feature transforms. We derive these UP formulas, show the connection between operators and transforms, and examine specific cases of the UP as well as lay the groundwork to calculating UP inequalities for custom transforms and operators.

1 Introduction. With the continued rise of machine learning's popularity, so has been the weight placed on finding representations of data optimized for certain learning tasks [6]. As an example: for the task of classifying the chords played in a Beatles' song, a representation like the spectrogram is a more efficient[1] input to a machine learning classifier than the raw waveform of the song. For tasks and data found in the real world it is not always clear that a particular representation, like the signal's frequency distribution over time, is the essential physical quantity to store and send to a classifier. Therefore it is important to have a framework for general ways to expand a signal to extract a physical quantity; that quantity could then be important for some learning tasks like "what notes are in this song?" or "what animal is in this image?".

A signal can be expanded as:

$$x(t) = \int F(a)u(a,t)da$$

We call $u(a,t)$ expansion functions. The a's are values of some physical quantity (they may be continuous or discrete, but here we only consider the continuous

*UMD

[1] in terms of space vs classification accuracy, among other reasons

case). $F(a)$ is the transform of the signal, and intuitively tells us how important a particular a is for the given signal. Finding a $F(a)$ for an instance of data $x(t)$ could be called feature extraction, and it is the hope that $x(t)$ can be more simply characterized in the representation $F(a)$ and give more insight to the nature of $x(t)$, which the machine learning algorithm should exploit.

If we know how to compute some feature $F(a)$ for a signal, we are interested in characterizing the relationship between $x(t)$ and $F(a)$, for example we want to know how the localization of energy in one affects the energy spread in the other. In quantum mechanics, it is a postulate that measurements are made by applying an operator A in a Hilbert space H on a signal $x(t)$, after which the result is not known, but the average result is expressed as a linear combination of eigenfunctions $u(a,t)$ (i.e. from solving the eigenvalue problem $Au(a,t) = au(a,t)$) of that operator each weighted by some probability [8]. In this abstract setting of operators we can use basic tools to find interesting inequalities with a wide array of interpretations and consequences. Two creative formulas comparing the spread of a function in two domains, each defined by the spectrum of an operator, will provide the framework to compare two signals $x(t)$ and $F(a)$

2 Notations and essential facts. We denote the adjoint of an operator A as $A^\dagger$, which is another operator which forces the equality, $x, y \in H$ a Hilbert space:

$$\int \overline{y(t)} Ax(t)dt = \int x(t)\overline{A^\dagger y(t)}dt$$

which in the case that A is self-adjoint, i.e. $A = A^\dagger$, results in

$$\int \overline{y(t)} Ax(t)dt = \int x(t)\overline{Ay(t)}dt \tag{1}$$

Proposition 1 ([3]). *Any operator may be written as a sum of a self-adjoint operator plus i times a self-adjoint operator.*

Proof. Note that $(A + A^\dagger)^\dagger = A + A^\dagger$ and $\left(\frac{A-A^\dagger}{i}\right)^\dagger = \frac{A^\dagger - A}{-i} = \frac{A-A^\dagger}{i}$ are self-adjoint operators. Then we may write:

$$A = \frac{1}{2}(A + A^\dagger) + \frac{1}{2}i\frac{A - A^\dagger}{i} \tag{2}$$

proving the proposition. □

Repeated use of an operator n-times is denoted A^n.

Definition 1 ([3]). *A function x of an operator A is defined by first expanding the ordinary function in a Taylor series and then substituting the operator A. That is*

$$\textit{if} \quad x(a) = \sum_{n=0}^{\infty} x_n a^n \quad \textit{then} \quad x(A) = \sum_{n=0}^{\infty} x_n A^n$$

Definition 2 (Expectations of operators [3]). *The average, average square, and variance of a self-adjoint operator for a signal $x(t)$ are:*

$$\mathbb{E}_x(A) = \int \overline{x(t)} A x(t) dt$$
$$\mathbb{E}_x(A^2) = \int \overline{x(t)} A^2 x(t) dt = \int |Ax(t)|^2 dt$$
$$\sigma_x^2(A) = \mathbb{E}_x(A^2) - \mathbb{E}_x(A)^2 = \int \overline{x(t)} (A - \mathbb{E}_x(A))^2 x(t) dt = \int |(A - \mathbb{E}_x(A)) x(t)|^2 dt$$

The commutator of two operators is $[A, B] = AB - BA$, the anti-commutator is $[A, B]_+ = AB + BA$

For random variables X, Y, the covariance may be expressed as

$$\operatorname{cov}(X, Y) = E[XY] - E[X]E[Y] = \frac{1}{2} E[XY + YX] - E[X]E[Y]$$
$$= E[(X - E[X])(Y - E[Y])] = \frac{1}{2} E[(X - E[X])(Y - E[Y]) + (Y - E[Y])(X - E[X])]$$

This may act as inspiration for a similar quantity for general operators:

Definition 3 (The covariance of two operators [4]).

$$\begin{aligned} \operatorname{Cov}_x(A, B) &= \frac{1}{2} \mathbb{E}_x(AB + BA) - \mathbb{E}_x(A)\mathbb{E}_x(B) \\ &= \frac{1}{2} \mathbb{E}_x([A - \mathbb{E}_x(A), B - \mathbb{E}_x(B)]_+) \end{aligned}$$

3 The Uncertainty Principle for arbitrary variables.

Theorem 1 (First Uncertainty Principle [2, 7, 5]). *Let A, B be self-adjoint operators on a complex Hilbert space H. If $x \in D(A^2) \cap D(B^2) \cap D(i[A, B])$ and $\|x\| \leq 1$ then*

$$4\sigma_x^2(A)\sigma_x^2(B) \geq (\mathbb{E}_x(i[A, B]))^2$$

Proof. First of all, since A and B are self adjoint we have:

$$\mathbb{E}_x(A) = \langle Ax, x \rangle = \langle x, Ax \rangle = \overline{\langle Ax, x \rangle} = \overline{\mathbb{E}_x(A)} \in \mathbb{R}$$
$$\mathbb{E}_x(rA) = \langle rAx, x \rangle = r \langle Ax, x \rangle = \overline{r} \langle x, Ax \rangle = \langle x, rAx \rangle = \overline{\mathbb{E}_x(rA)} \in \mathbb{R}$$

and the same for B (where the second line shows that multiplication by $r \in \mathbb{R}$ preserves self-adjointness). We note

$$\mathbb{E}_x(x) i[A, B] = 2\operatorname{Im}(\mathbb{E}_x(BA)) \tag{3}$$

since the inner product is linear:

$$\begin{aligned}\mathbb{E}_x(i[A,B]) &= i\mathbb{E}_x(AB - BA) = i(\langle ABx, x\rangle - \langle BAx, x\rangle)\\ &= i(\langle Bx, Ax\rangle - \langle Ax, Bx\rangle) = i(\overline{\langle Ax, Bx\rangle} - \langle Ax, Bx\rangle)\\ &= 2\mathrm{Im}(\langle Ax, Bx\rangle) = 2\mathrm{Im}(\mathbb{E}_x(BA))\end{aligned}$$

Next:

$$|\mathbb{E}_x(B+iA)|^2 = |\mathbb{E}_x(B) + i\mathbb{E}_x(A)|^2 = \mathbb{E}_x(A)^2 + \mathbb{E}_x(B)^2 \tag{4}$$

And by Cauchy Schwartz:

$$|\mathbb{E}_x(B+iA)|^2 = |\langle (B+iA)x, x\rangle|^2 \leq \|x\|^2 \|(B+iA)x\|^2 \leq \|(B+iA)x\|^2 = \mathbb{E}_x(|B+iA|^2) \tag{5}$$

Also

$$\mathbb{E}_x(|B+iA|^2) = \mathbb{E}_x(B^2) + \mathbb{E}_x(A^2) - 2\mathrm{Im}(\mathbb{E}_x(BA)) \tag{6}$$

since:

$$\begin{aligned}\mathbb{E}_x(|B+iA|^2) &= \|(B+iA)x\|^2\\ &= \langle (B+iA)x, (B+iA)x\rangle\\ &= \langle Bx, Bx\rangle - i\langle Bx, Ax\rangle + i\langle Ax, Bx\rangle + \langle Ax, Ax\rangle\\ &= \mathbb{E}_x(B^2) + \mathbb{E}_x(A^2) - i(\langle Bx, Ax\rangle - \langle Ax, Bx\rangle)\\ &= \mathbb{E}_x(B^2) + \mathbb{E}_x(A^2) - 2\mathrm{Im}(\mathbb{E}_x(BA))\end{aligned}$$

(5), (6) give us

$$|\mathbb{E}_x(B+iA)|^2 \leq \mathbb{E}_x(B^2) + \mathbb{E}_x(A^2) - 2\mathrm{Im}(\mathbb{E}_x(BA))$$

and after substituting (4) and shuffling we have:

$$\mathbb{E}_x(B^2) - \mathbb{E}_x(B)^2 + \mathbb{E}_x(A^2) - \mathbb{E}_x(A)^2 \geq 2\mathrm{Im}(\mathbb{E}_x(BA))$$

we can rewrite the above setting $A \to rA$ and $B \to sB$:

$$s^2(\mathbb{E}_x(B^2) - \mathbb{E}_x(B)^2) + r^2(\mathbb{E}_x(A^2) - \mathbb{E}_x(A)^2) \geq 2rs\mathrm{Im}(\mathbb{E}_x(BA))$$

finally plugging in $r^2 = \mathbb{E}_x(B^2) - \mathbb{E}_x(B)^2$ and $s^2 = \mathbb{E}_x(A^2) - \mathbb{E}_x(A)^2$ and squaring both sides yields:

$$(\mathbb{E}_x(A^2) - \mathbb{E}_x(A)^2)(\mathbb{E}_x(B^2) - \mathbb{E}_x(B)^2) \geq (\mathrm{Im}(\mathbb{E}_x(BA)))^2 = \frac{1}{4}(\mathbb{E}_x(i[A,B]))^2$$

which proves the theorem. □

We may use the Cauchy Schwartz inequality in a different way to arrive at a slightly stronger version of the uncertainty principle.

Theorem 2 (Second Uncertainty Principle [4, 3]). *Let A, B be self-adjoint operators on a complex Hilbert space H. If $x \in D(A^2) \cap D(B^2) \cap D([A, B])$. Then*

$$4\sigma_x^2(A)\sigma_x^2(B) \geq |\mathbb{E}_x([A, B])|^2 + 4\text{Cov}_x(A, B)^2$$

Proof. Let $A_0 = A - \mathbb{E}_x(A)$ and $B_0 = B - \mathbb{E}_x(B)$. We may see

$$\sigma_x^2(A)\sigma_x^2(B) \geq |\mathbb{E}_x(A_0B_0)|^2 \tag{7}$$

by expanding the integrals:

$$\begin{aligned}
\sigma_x^2(A)\sigma_x^2(B) &= \int |A_0x(t)|^2 dt \int |B_0x(t)|^2 dt \\
&\geq \left| \int \overline{A_0x(t)} B_0x(t) dt \right|^2 \\
&= \left| \int \overline{x(t)} A_0B_0x(t) dt \right|^2 \\
&= |\mathbb{E}_x(A_0B_0)|^2
\end{aligned}$$

We note that $\mathbb{E}_x(A_0B_0)$ may not be real since A_0B_0 may not be self-adjoint. Now we plug $A = A_0B_0$ (and therefore $A^\dagger = B_0^\dagger A_0^\dagger$) into proposition 1 obtaining, since A_0 and B_0 are self-adjoint:

$$A_0B_0 = \frac{1}{2}(A_0B_0 + B_0A_0) + \frac{1}{2}i\frac{A_0B_0 - B_0A_0}{i} = \frac{1}{2}([A_0, B_0]_+) + \frac{1}{2}i\frac{[A_0, B_0]}{i}$$

where $([A_0, B_0]_+)$ and $\frac{[A_0, B_0]}{i}$ are self adjoint by the proposition. Therefore using linearity of the inner product:

$$\begin{aligned}
|\mathbb{E}_x(A_0B_0)|^2 &= \left| \mathbb{E}_x(\frac{1}{2}([A_0, B_0]_+) + \frac{1}{2}i\frac{[A_0, B_0]}{i}) \right|^2 \\
&= \frac{1}{4}|\mathbb{E}_x([A_0, B_0]_+)|^2 + \frac{1}{4}|\mathbb{E}_x([A_0, B_0])| \\
&= \frac{1}{4}(|\mathbb{E}_x([A_0, B_0])|^2 + |\mathbb{E}_x([A_0, B_0]_+)|^2)
\end{aligned}$$

showing:

$$4\sigma_x^2(A)\sigma_x^2(B) \geq |\mathbb{E}_x([A_0, B_0])|^2 + |\mathbb{E}_x([A_0, B_0]_+)|^2$$

Finally, note:

$$\begin{aligned}[A_0, B_0] &= A_0B_0 - B_0A_0 \\ &= (A - \mathbb{E}_x(A))(B - \mathbb{E}_x(B)) - (B - \mathbb{E}_x(B))(A - \mathbb{E}_x(A)) \\ &= AB - BA = [A, B]\end{aligned}$$

and that $\mathbb{E}_x([A_0, B_0]_+) = 2\text{Cov}_x(A, B)$, showing that:

$$4\sigma_x^2(A)\sigma_x^2(B) \geq |\mathbb{E}_x([A, B])|^2 + 4\text{Cov}_x(A, B)^2$$

Proving the theorem. □

As stated in the proof, $[A_0, B_0]_+$ is self-adjoint, making $\mathbb{E}_x([A_0, B_0]_+) = 2\text{Cov}_x(A, B) \in \mathbb{R}$ so that $4\text{Cov}_x(A, B)^2 \geq 0$. Therefore $4\sigma_x^2(A)\sigma_x^2(B) \geq |\mathbb{E}_x([A, B])|^2 + 4\text{Cov}_x(A, B)^2 \geq |\mathbb{E}_x([A, B])|^2$ which is compatible with the first UP.

4 Representation of Signals.

We are interested in using these two abstract UPs to understand signals in terms of some physical quantity represented by an operator. The UPs can help us understand, for example, a signal's density, average value, and spread in a physical quantity. To do this, we first expand the signal as a linear combination of basis functions obtained by solving the eigenvalue for some operator of interest.

For the discrete eigenvalue case we have the Hilbert-Schmidt theorem, also known as the Spectral Theorem:

Theorem 3 (Hilbert-Schmidt[9]). *Let A be a self-adjoint compact operator on a Hilbert space H. Then, there is a complete orthonormal basis $\{u_n\}$ for H so that*

$$Au_n(t) = a_n u_u(t) \tag{8}$$

and $a_n \to 0$ as $n \to \infty$.

We will primarily be concerned with the continuous eigenvalue case:

Theorem 4 ([1]). *Let A be a self-adjoint operator on a Hilbert space H. Then:*

$$Au(a, t) = au(a, t)$$

where

i. The eigenvalues a of A are real.

ii. The eigenfunctions $u(a, t)$ of A are orthogonal.

iii. If A is a linear, second order differential operator then the eigenfunctions $u(a, t)$ of A form a complete set.

Part i is important from a physical standpoint because in nature measurable quantities are often real. Cohen alludes that Part iii remains true for all self-adjoint operators [4, 3].

If these three properties are in place we can easily find the coefficients in a synthesis equation for x

$$x(t) = \int F(a)u(a,t)da \tag{9}$$

in the following way. First note:

$$\int \overline{u(a',t)}u(a,t)dt = \delta(a-a') \tag{10}$$

so that after multiplying (9) by $\overline{u(a',t)}$ and integrating with respect to time:

$$\begin{aligned}\int x(t)\overline{u(a',t)}dt &= \iint F(a)u(a,t)\overline{u(a',t)}dadt \\ &= \int F(a)\delta(a-a')da \\ &= F(a')\end{aligned} \tag{11}$$

leaving us with an analysis equation:

$$F(a) = \int x(t)\overline{u(a,t)}dt \tag{12}$$

We refer to $F(a)$ as the A-transform of $x(t)$.

Proposition 2 ([3]). *If $u(a,t)$ is an eigenfunction of the operator A then:*

$$f(A)u(a,t) = f(a)u(a,t)$$

Proof. Since $u(a,t)$ is an eigenfunction of A we have that $Au(a,t) = au(a,t)$, moreover $A^n u(a,t) = a^n u(a,t)$. Plugging this into the definition of a function of an operator and recognizing the ordinary Taylor series, we have:

$$f(A)u(a,t) = \sum_{n=0}^{\infty} f_n A^n u(a,t) = \sum_{n=0}^{\infty} f_n a^n u(a,t) = f(a)u(a,t)$$

proving the proposition. □

5 Computing Averages. Now we are ready to prove an interesting correspondence rule. If the density of some physical quantity a is taken to be $|F(a)|^2$ (which may not integrate to 1), then the average value of a is

$$E_F(a) = \int a|F(a)|^2 da$$

and more generally the average value of some function $g(a)$ is:

$$E_F(g(a)) = \int g(a)|F(a)|^2 da$$

It turns out that we can calculate these averages without calculating the transform $F(a)$.

Proposition 3 ([4]). *If A is an operator, and $F(a)$ is the A-transform of $x(t)$, then:*

$$\mathbb{E}_x(g(A)) = \int \overline{x(t)} g(A) x(t) dt = \int g(a)|F(a)|^2 da = E_F(g(a))$$

Proof. From the definition, we plug in (9), use proposition 2, and (10):

$$\begin{aligned}
\mathbb{E}_x(g(A)) &= \int \overline{x(t)} g(A) x(t) dt \\
&= \int \left(\overline{\int F(a')u(a',t)da'} \right) g(A) \left(\int F(a)u(a,t)dadt \right) \\
&= \int \int \int \overline{F(a')u(a',t)} g(A) u(a,t) F(a) da' dadt \\
&= \int \int \int \overline{F(a')u(a',t)} g(a) u(a,t) F(a) da' dadt \\
&= \int \int \int \overline{F(a')} F(a) g(a) \overline{u(a',t)} u(a,t) dt da' da \\
&= \int \int \overline{F(a')} F(a) g(a) \delta(a-a') da' da \\
&= \int \overline{F(a)} F(a) g(a) da \\
&= \int g(a)|F(a)|^2 da \\
&= E_F(g(a))
\end{aligned}$$

proving the proposition. □

Proposition 3 also means that we can compute expectations of operators with expectations of ordinary functions.

6 Instances of the UP. An important operator is the frequency operator defined by $D = \frac{1}{i}\frac{d}{dt}$.[2] The eigenvalue problem is $Du(\omega, t) = \omega u(\omega, t)$ and the solutions are $u(\omega, t) = ce^{j\omega t}, \omega \in \mathbb{R}$ [4]. To find the normalization constant c we consider:

$$\int u^*(\omega', t)u(\omega, t)dt = c^2 \int e^{-j\omega t}e^{j\omega' t}dt = 2\pi c^2\delta(\omega - \omega')$$

implying $c = \frac{1}{\sqrt{2\pi}}$ so that $u(\omega, t) = \frac{1}{\sqrt{2\pi}}e^{j\omega t}$. This leaves us with a familiar looking version of (12):

$$F(\omega) = \frac{1}{\sqrt{2\pi}} \int x(t)e^{-j\omega t}dt = \hat{x}(\omega)$$

the Fourier transform.

From Proposition 3 we have a rule of associating a function of ordinary variables with an operator. In the case of D:

$$\int g(\omega)|\hat{x}(\omega)|^2 d\omega = \int \overline{x(t)}g(D)x(t)dt \tag{13}$$

$$\text{This correspondence means } D = \begin{cases} \frac{1}{i}\frac{d}{dt} & \text{in the } t \text{ representation} \\ \omega & \text{in the Fourier representation} \end{cases} \tag{14}$$

If we have a signal, and want to calculate average frequencies without calculating the Fourier transform, we can use (13). Similarly if we have a Fourier transform and want to calculate time averages without calculating the signal, we can use the time operator defined by: $X = i\frac{d}{d\omega}$. From Proposition 3 we have

$$\int g(t)|x(t)|^2 dt = \int \overline{\hat{x}(\omega)}g(X)\hat{x}(\omega)d\omega \tag{15}$$

$$\text{This correspondence means } X = \begin{cases} t & \text{in the } t \text{ representation} \\ i\frac{d}{d\omega} & \text{in the Fourier representation} \end{cases} \tag{16}$$

In this case the eigenvalue problem is $tu(t, t') = t'u(t, t')$, and we could conclude $u(t, t') = \delta(t - t')$.

Now we specify the abstract uncertainty principles with A, B to D, X and recognize the well known Heisenberg uncertainty principle. For the left hand side we use the two correspondence rules (13), (15) to write down:

$$\begin{aligned} \sigma_x^2(D) &= \int \overline{x(t)}(D - \mathbb{E}_x(D))^2 x(t)dt = \int (\omega - E_{\hat{x}}(\omega))^2 |\hat{x}(\omega)|^2 d\omega \\ \sigma_x^2(X) &= \int \overline{x(t)}(X - \mathbb{E}_x(X))^2 x(t)dt = \int (t - E_x(t))^2 |x(t)|^2 dt \end{aligned} \tag{17}$$

[2] To prove it is self-adjoint do integration by parts and make an argument similar to this one.

where we have moved from calculations with non-commuting operators to calculations with ordinary commuting variables.

For the right hand sides we compute the commutator $[X, D] = i$ via:

$$
\begin{aligned}
(XD - DX)x(t) &= (t\frac{1}{i}\frac{d}{dt} - \frac{1}{i}\frac{d}{dt}t)x(t) \\
&= \frac{1}{i}(t\frac{dx(t)}{dt} - \frac{d(tx(t))}{dt}) \\
&= \frac{1}{i}(t\frac{dx(t)}{dt} - x(t) - t\frac{dx(t)}{dt}) \\
&= -\frac{1}{i}x(t) = ix(t)
\end{aligned}
$$

so that theorem 1 becomes:

$$
\begin{aligned}
4\sigma_x^2(X)\sigma_x^2(D) &\geq (\mathbb{E}_x(i[X, D]))^2 \\
4\left(\int (t - E_x(t))^2|x(t)|^2 dt\right)\left(\int (\omega - E_{\hat{x}}(\omega))^2|\hat{x}(\omega)|^2 d\omega\right) &\geq 1
\end{aligned}
\tag{18}
$$

Note in this instantiation of the UP that the left hand side is the variance of the signal in time (σ_t^2) times the variance of the signal in frequency (σ_p^2) (with the magnitude of the signal squared and magnitude of the Fourier transform squared taken as densities). This is consistent with the famed Heisenberg uncertainty principle in quantum mechanics: $\sigma_t\sigma_p \geq \frac{\hbar}{2}$ where operators X and D are multiplied by the Planck constant $\hbar$ and called the "position" and "momentum" operators.

6.1 An instance of the second UP. To instantiate theorem 2 it is meaningful to consider an example signal. Let $x(t)$ be a quadratic phase signal with Gaussian envelope:

$$x(t) = (\alpha/\pi)^{1/4}\exp(-\alpha t^2/2 + i\beta t^2/2 + i\gamma t) \tag{19}$$

First note $|x(t)|^2 = x(t)\overline{x(t)} = \sqrt{\frac{\alpha}{\pi}}\exp(-\alpha t^2)$. Then:

$$\mathbb{E}_x(X) = \int \overline{x(t)}Xx(t)dt = \int tx(t)\overline{x(t)}dt = \sqrt{\frac{\alpha}{\pi}}\int t\exp(-\alpha t^2)dt = 0$$

$$\mathbb{E}_x(X^2) = \int \overline{x(t)}X^2x(t)dt = \int tx(t)\overline{x(t)}dt = \sqrt{\frac{\alpha}{\pi}}\int t^2\exp(-\alpha t^2)dt = \frac{1}{2\alpha}$$

where we recognized the mean and variance of a mean zero normally distributed random variable.

In fact $\int \sqrt{\frac{\alpha}{\pi}}\exp(-\alpha t^2) = \int |x(t)|^2 = \|x\|^2 = 1$. Next we can notice

$$Dx(t) = (i\alpha t + \beta t + \gamma)x(t)$$

so that:

$$\mathbb{E}_x(D) = \int \overline{x(t)} D x(t) dt = \sqrt{\frac{\alpha}{\pi}} \int (i\alpha t + \beta t + \gamma) \exp(-\alpha t^2) = \gamma$$

similarly:

$$\begin{aligned}\mathbb{E}_x(D^2) = \int \overline{x(t)} D^2 x(t) dt &= \sqrt{\frac{\alpha}{\pi}} \int |i\alpha t + \beta t + \gamma|^2 \exp(-\alpha t^2) \\ &= \int (\alpha^2 t^2 + \beta^2 t^2 + 2\beta t\gamma + \gamma^2) \exp(-\alpha t^2) \\ &= \frac{\alpha^2 + \beta^2}{2\alpha} + \gamma^2\end{aligned}$$

Therefore:

$$\begin{aligned}\sigma_x^2(D) = \mathbb{E}_x(D^2) - \mathbb{E}_x(D)^2 &= \frac{\alpha^2 + \beta^2}{2\alpha} \\ \sigma_x^2(X) = \mathbb{E}_x(X^2) - \mathbb{E}_x(X)^2 &= \frac{1}{2\alpha}\end{aligned}$$

And for the covariance $\text{Cov}_x(X, D) = \frac{1}{2}\mathbb{E}_x(XD + DX) - \mathbb{E}_x(X)\mathbb{E}_x(D)$ we first note:

$$\mathbb{E}_x(XD+DX) = \mathbb{E}_x(XD+DX-DX+XD) = 2\mathbb{E}_x(XD)-E[X,D] = 2\mathbb{E}_x(XD)-i$$

and calculate:

$$\begin{aligned}\mathbb{E}_x(XD + DX) &= \int \overline{x(t)}(2XD - i)x(t)dt \\ &= \int \overline{x(t)}(2tDx(t) - ix(t))dt \\ &= \int \overline{x(t)}(2t(i\alpha t + \beta t + \gamma)x(t) - ix(t))dt \\ &= \sqrt{\frac{\alpha}{\pi}} \int (2i\alpha t^2 + 2\beta t^2 + 2\gamma t - i) \exp(-\alpha t^2) dt \\ &= \frac{2i\alpha + 2\beta}{2\alpha} - i \\ &= \frac{\beta}{\alpha}\end{aligned}$$

Recall $[X, D] = i$ so $|\mathbb{E}_x([X, D])|^2 = 1$. Since $\mathbb{E}_x(X) = 0$ we have $2\text{Cov}_x(X, D) = \mathbb{E}_x(XD + DA) = \frac{\beta}{\alpha}$ so that the right hand side of theorem 2 is:

$$|\mathbb{E}_x([X, D])|^2 + 4\text{Cov}_x(X, D)^2 = 1 + \frac{\beta^2}{\alpha^2}$$

this meets the left hand side of theorem 2:

$$4\sigma_x^2(X)\sigma_x^2(D) = 4\frac{1}{2\alpha}\frac{\alpha^2+\beta^2}{2\alpha} = 1 + \frac{\beta^2}{\alpha^2}$$

which is in turn greater than the lower bound of theorem 1, $(\mathbb{E}_x(i[A,B]))^2 = 1$.

7 The Weyl Correspondence.

In section 6 we were able to associate a function of one ordinary variable with a function of an operator in (13) and (15), and compute expectations. But if we have a function of two variables such as $a(t,\omega) = t\omega^2$ does this correspond to $A(X,D) = XD^2$, DXD, D^2X or some other concoction? All of these choices are different since X and D do not commute. It is a relevant question because for example, for the lower bounds in the UP we in general will need to compute expectations of operators that are a function of two other operators. In the $[X,D]$ case we were lucky that it reduced to a constant i, but we may not be so lucky with future $[A,B]$.

To associate a function of two ordinary variables with a function of an operator what Hermann Weyl chose to do in 1928 is the following.

Definition 4 (Weyl Operator, Weyl Rule/Correspondence [3])**.** *For a function $a(t,\omega)$ define its Fourier transform by:*

$$\hat{a}(\theta,\tau) = \frac{1}{4\pi^2}\iint a(t,\omega)\exp(-i\theta t - i\tau\omega)dtd\omega$$

in which case

$$a(t,\omega) = \iint \hat{a}(\theta,\tau)\exp(i\theta t + i\tau\omega)d\theta d\tau \tag{20}$$

The Weyl operator $A(X,D)$ corresponding to $a(t,\omega)$ is defined by substituting $t \to X$ and $\omega \to D$ in (20)*:*

$$A(X,D) = \iint \hat{a}(\theta,\tau)\exp(i\theta X + i\tau D)d\theta d\tau \tag{21}$$

To understand the operator $\exp(i\tau D)$ recall the Taylor series definition of a function of an operator:

$$\exp(i\tau D)x(t) = \sum_{n=0}^{\infty}\frac{(i\tau)^n D^n}{n!}x(t) = \sum_{n=0}^{\infty}\frac{\tau^n}{n!}\frac{d^n}{dt^n}x(t) = x(t+\tau) \tag{22}$$

This is called the translation operator. We are interested in disentangling $\exp(i\theta X + i\tau D)$ in (21), since operators are not commuting there are methods such as:

Proposition 4 (The Zassenhaus formula [3]). *Let A and B be two linear operators, then:*

$$e^{A+B} = e^A e^B e^{-\frac{1}{2}[A,B]} e^{\frac{1}{3}[B,[A,B]]+\frac{1}{6}[A,[A,B]]} e^{-\frac{1}{24}([[[A,B],A]A]+3[[[A,B]A],B]+3[[[A,B],B],B])\cdots}$$

As a special case consider $A = i\theta X$ and $B = i\tau D$, in which case $[A,B] = -i\theta\tau$ so that this series terminates ($[A,[A,B]] = [B,[A,B]] = 0$). Then

$$\exp(i\theta X + i\tau D) = \exp(i\theta\tau/2)\exp(i\theta X)\exp(i\tau D) = \exp(-i\theta\tau/2)\exp(i\tau D)\exp(i\theta X) \tag{23}$$

Using (23) we rewrite (21) and apply it on the Fourier transform $\hat{x}(\omega)$ of a signal:

$$A(X,D)\hat{x}(\omega) = \iint \hat{a}(\theta,\tau)\exp(-i\theta\tau/2)\exp(i\tau D)\exp(i\theta X)\hat{x}(\omega)d\theta d\tau$$

After plugging in (4) and some laborious substitutions this can be shown to be equivalent to:

$$A(X,D)\hat{x}(\omega) = \frac{1}{2\pi}\iint a(t,\frac{1}{2}(\omega+\theta))\exp(i(\theta-\omega)t)\hat{x}(\theta)dtd\theta \tag{24}$$

However, from (24), by multiplying by $\overline{\hat{x}(\omega)}$, integrating with respect to ω, and making the substitution $\omega \to \omega + \frac{1}{2}, \theta \to \omega - \frac{1}{2}\theta$, we obtain:

$$\begin{aligned} \int \overline{\hat{x}(\omega)} A(X,D)\hat{x}(\omega)d\omega &= \frac{1}{2\pi}\iiint \overline{\hat{x}(\omega+\frac{1}{2}\theta)} a(t,\omega)\exp(-i\theta t)\hat{x}(\omega-\frac{1}{2}\theta)dtd\omega \\ \mathbb{E}_{\hat{x}}(A(X,D)) &= \iint a(t,\omega)W(t,\omega)dtd\omega = E_W(a(t,\omega)) \end{aligned} \tag{25}$$

letting

$$W(t,\omega) = \frac{1}{2\pi}\int \overline{\hat{x}(\omega+\frac{1}{2}\theta)}\exp(-i\theta t)\hat{x}(\omega-\frac{1}{2}\theta)d\theta$$

which is called the Wigner distribution. It can also be shown that $\mathbb{E}_{\hat{x}}(A(X,D)) = \mathbb{E}_x(A(X,D))$ [3].

The left hand side of (25) deals with operators, and the right hand side deals with ordinary functions. The right hand side is an average of $a(t,\omega)$ with respect to the density $W(t,\omega)$ (which is not always positive).

8 Conclusion. We have started in the most general setting and written down two UP inequalities in terms of expectations of operators. With the foundational operators of time/position and frequency/momentum, X and D, we have looked at a function (19) that minimizes the left hand side of the second UP: $4\sigma_x^2(X)\sigma_x^2(D) = |\mathbb{E}_x([X,D])|^2 + 4\text{Cov}_x(X,D)^2$. One further direction of interest taken in [2] is to look at Gabor systems $\varphi_{m,n}$ and wavelet systems $\psi_{m,n}$, and examine how far $\sigma^2_{\varphi_{m,n}}(X)\sigma^2_{\varphi_{m,n}}(D)$ and $\sigma^2_{\psi_{m,n}}(X)\sigma^2_{\psi_{m,n}}(D)$ are from the UP lower bound, and how constraints like being a frame or an orthonormal basis affect the size of $\sigma_x^2(X)\sigma_x^2(D)$.

Another direction is to form UPs for other feature spaces. We have seen one instance of the UPs with X and D. But we also know how to compute the expectations for the UPs for any operator from proposition 3 and (25). X and D can act as generators for other operators, like the translation operator from (22). Another such example is $C = \frac{1}{2}(XD + DX)$ the scale operator; the operation of $e^{i\theta C}$ on $x(t)$ is given by $e^{i\theta C}x(t) = e^{\theta/2}x(e^{\theta}t)$, a time dilation. It is a big point of Cohen's in [3, 4] that equations of regular commuting variables may be associated with equations of noncommuting operators through their expectations. It is these features that make the two general UPs here usable in many contexts as a creative formula.

Acknowledgements. The author would like to thank Professor John Benedetto and Professor Wojtek Czaja at UMD for their guidance and helpful discussions.

References

[1] George Arfken. "Mathematical Methods for Physicists". In: *Mathematical Methods for Physicists.* Academic Press, 1985, pp. 510–511.

[2] John J. Benedetto. "Frame decompositions, sampling, and uncertainty principle inequalities". In: *Wavelets: mathematics and applications.* Ed. by John J. Benedetto and MichaelEditors Frazier. CRC Press, 1994.

[3] Leon Cohen. *The Weyl operator and its generalization.* Birkhäuser/Springer, 2013.

[4] Leon Cohen. *Time frequency analysis.* Prentice Hall, 1995.

[5] Gerald B. Folland and Alladi Sitaram. "The Uncertainty Principle: A Mathematica Survey". In: *The Journal of Fourier Analysis and Applications* 3.3 (1997).

[6] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning.* `http://www.deeplearningbook.org`. MIT Press, 2016.

[7] Begue Matthew. *The Uncertainty Principle: A Brief Survey.* `https://www.math.umd.edu/~begue/Expository/Uncertainty.pdf`. Dec. 2012.

[8] Ravinder R Puri. "Basic Quantum Mechanics". In: *Mathematical Methods of Quantum Optics.* Springer, 2001, pp. 1–36.

[9] Michael C. Reed and Barry Simon. "Methods of modern mathematical physics". In: *Methods of modern mathematical physics.* Academic Press, 1980, pp. 203–203.