# The uncertainty principle: variations on a theme

Avi Wigderson* Yuval Wigderson†

September 14, 2020

**Abstract**

We show how a number of well-known uncertainty principles for the Fourier transform, such as the Heisenberg uncertainty principle, the Donoho–Stark uncertainty principle, and Meshulam's non-abelian uncertainty principle, have little to do with the structure of the Fourier transform itself. Rather, all of these results follow from very weak properties of the Fourier transform (shared by numerous linear operators), namely that it is bounded as an operator $L^1 \to L^\infty$, and that it is unitary. Using a single, simple proof template, and only these (or weaker) properties, we obtain some new proofs and many generalizations of these basic uncertainty principles, to new operators and to new settings, in a completely unified way. Together with our general overview, this paper can also serve as a survey of the many facets of the phenomena known as uncertainty principles.

# Contents



*School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540, USA. Email: avi@ias.edu. Research supported by NSF grant CCF-1900460

†Department of Mathematics, Stanford University, Stanford, CA 94305, USA. Email: yuvalwig@stanford.edu. Research supported by NSF GRFP Grant DGE-1656518.

# 1 Introduction

## 1.1 Background

The phrase "uncertainty principle" refers to any of a wide class of theorems, all of which capture the idea that a non-zero function and its Fourier transform cannot both be "very localized". This phenomenon has been under intensive study for almost a century now, with new results published continuously to this day. So while this introduction (and the paper itself) discusses some broad aspects of it, this is not a comprehensive survey. For more information on the history of the uncertainty principle, and for many other generalizations and variations, we refer the reader to the excellent survey of Folland and Sitaram [16].

The study of uncertainty principles began with Heisenberg's seminal 1927 paper [20], with the corresponding mathematical formalism independently due to Kennard [24] and Weyl [37]. The original motivation for studying the uncertainty principle came from quantum

mechanics[1], and thus most classical uncertainty principles deal with functions on $\mathbb{R}$ or $\mathbb{R}^n$. The first, so-called Heisenberg uncertainty principle, says that the *variance* (appropriately defined, see Section 4) of a function and of its Fourier transform cannot both be small. Following Heisenberg's paper, many different notions of locality were studied. For example, it is a simple and well-known fact that if $f : \mathbb{R} \to \mathbb{C}$ has compact support, then $\hat{f}$ can be extended to a holomorphic function on $\mathbb{C}$, which in particular implies that $\hat{f}$ only vanishes on a discrete countable set, and so it is not possible for both $f$ and $\hat{f}$ to be compactly supported. This fact was generalized by Benedicks [6] (and further extended by Amrein and Berthier [1]), who showed that it is not possible for both $f$ and $\hat{f}$ to have supports of finite measure. Another sort of uncertainty principle, dealing not with the sharp localization of a function, but rather with its decay at infinity, has also been widely studied. The first such result is due to Hardy [18], who proved (roughly) that it is not possible for both $f$ and $\hat{f}$ to decay faster than $e^{-x^2}$. Yet another type of uncertainty principle is the logarithmic version conjectured by Hirschman [21] and proven by Beckner [4] and independently Białynicki-Birula and Mycielski [7], which deals with the Shannon entropies of a function and its Fourier transform, and which has connections to log-Sobolev and hypercontractive inequalities [5].

In 1989, motivated by applications to signal processing[2], Donoho and Stark [14] initiated the study of a new type of uncertainty principle, which deals not with functions defined on $\mathbb{R}$, but rather with functions defined on finite groups. Many of the concepts discussed above, such as variance and decay at infinity, do not make sense when dealing with functions on a finite group. However, other measures of "non-localization", such as the size of the support of a function, are well-defined in this context, and the Donoho–Stark uncertainty principle deals with this measure. Specifically, they proved that if $G$ is a finite abelian group[3], $\widehat{G}$ is its dual group, $f : G \to \mathbb{C}$ is a non-zero function, and $\hat{f} : \widehat{G} \to \mathbb{C}$ is its Fourier transform, then $|\mathrm{supp}(f)||\mathrm{supp}(\hat{f})| \geq |G|$. They also proved a corresponding theorem for an appropriate notion of "approximate support" (see Section 3.3 for more details). The work of Donoho and Stark led to a number of other uncertainty principles for finite groups. Three notable examples are Meshulam's extension [26] of the Donoho–Stark theorem to arbitrary finite groups, Tao's strengthening [36] of the Donoho–Stark theorem in case $G$ is a cyclic group of prime order, and the discrete entropic uncertainty principles of Dembo, Cover, and Thomas [12], which generalize the aforementioned theorems of Hirschman [21], Beckner [4] and Białynicki-Birula and Mycielski [7].

Despite the fact that all uncertainty principles are intuitively similar, their proofs use a wide variety of techniques and a large number of special properties of the Fourier transform. Here is a sample of this variety. The standard proof of Heisenberg's uncertainty principle uses integration by parts, and the fact that the Fourier transform on $\mathbb{R}$ turns differentiation

[1] As some pairs of natural physical parameters, such as the position and momentum of a particle, can be viewed as such dual functions, the phrase "uncertainty principle" is meant to indicate that it is impossible to measure both to arbitrary precision.

[2] In this context, some similar theorems can be called "certainty principles". Here, this indicates that one can use the fact that one parameter is *not* localized to measure it well by random sampling.

[3] Strictly speaking, Donoho and Stark only proved this theorem for cyclic groups, but it was quickly observed that the same result holds for all finite abelian groups.

into multiplication by $x$. Benedicks's proof that a function and its Fourier transform cannot both have finite-measure supports uses the Poisson summation formula. The logarithmic uncertainty principle follows from differentiating a deep fact of real analysis, namely the sharp Hausdorff–Young inequality of Beckner [4]. The original proof of the Donoho–Stark uncertainty principle uses the fact that the Fourier transform on a cyclic group $G$, viewed as a $|G| \times |G|$ matrix, is a Vandermonde matrix, and correspondingly certain submatrices of it can be shown to be non-singular. Tao's strengthening of this theorem also uses this Vandermonde structure, together with a result of Chebotarëv which says that in case $|G|$ is prime, *all* square submatrices of this Vandermonde matrix are non-singular. The proof of Donoho and Stark's approximate support inequality relates it to different norms of submatrices of the Fourier transform matrix. Finally, Meshulam's proof of the non-abelian uncertainty principle uses linear-algebraic considerations in the group algebra $\mathbb{C}[G]$.

## 1.2 The simple theme and its variations

In this paper, we present a unified framework for proving many (but not all) of these uncertainty principles, together with various generalizations of them. The key observation throughout is that although the proofs mentioned above use a wide variety of analytic and algebraic properties that are particular to the Fourier transform, these results can also be proved using almost none of these properties. Instead, all of our results will follow from two very basic facts about the Fourier transform, namely that it is bounded as an operator $L^1 \to L^\infty$, and that it is unitary[4]. Because the Fourier transform is by no means the only operator with these properties, we are able to extend many of these well-known uncertainty principles to many other operators.

This unified framework is, at its core, very simple. The $L^1 \to L^\infty$ boundedness of the Fourier transform gives an inequality relating $\|\hat{f}\|_\infty$ to $\|f\|_1$. Similarly, the $L^1 \to L^\infty$ boundedness of the inverse Fourier transform gives an analogous inequality relating $\|f\|_\infty$ to $\|\hat{f}\|_1$. Multiplying these two inequalities together yields our basic uncertainty principle, which has the form

$$\frac{\|f\|_1}{\|f\|_\infty} \cdot \frac{\|\hat{f}\|_1}{\|\hat{f}\|_\infty} \geq C_0,$$

for an appropriate constant $C_0$. Thus, in a sense, the "measure of localization" $H_0(g) = \frac{\|g\|_1}{\|g\|_\infty}$ is a primary one for us, and the uncertainty principle above,

$$H_0(f) \cdot H_0(\hat{f}) \geq C_0 \tag{1}$$

is the source of essentially all our uncertainty principles. Note that $H_0$ really is a yet another "measure of localization" of a function $g$, in that a function that is more "spread out" will have a larger $L^1$ norm than a more localized function, if both have the same $L^\infty$ norm. Here is how we will use this primary uncertainty principle.

[4]In fact, a far weaker condition than unitarity is needed for our results, as will become clearer in the technical sections.

Suppose we want to prove any uncertainty principle, for any potential "measure of localization" $H$ on functions, e.g. one of the form

$$H(f) \cdot H(\hat{f}) \geq C. \tag{2}$$

For example, our "measure of localization" $H$ might be the *variance* of (the square of) a function on the reals if we want to prove the Heisenberg uncertainty principle, or $H$ might be the *support size* of a finite-dimensional vector if we want to prove the Donoho–Stark uncertainty principle.

We will derive (2) by first proving a universal[5] bound, relating the measure of localization $H$ to our primary one $H_0$, that holds for *every* function $g$. This reduction will typically take the form

$$H(g) \geq C' \cdot H_0(g) = C' \cdot \frac{\|g\|_1}{\|g\|_\infty}.$$

Now this bound can be applied to both $f$ and $\hat{f}$ separately, which combined with our primary principle (1) yields (with $(C')^2 = C \cdot C_0$) the desired uncertainty principle (2)[6].

Such an approach to proving simple uncertainty principles is by no means new; it goes back at least to work of Williams [38] from 1979, who used it to prove a weak version of an approximate-support inequality for the Fourier transform on $\mathbb{R}$ (see Section 3.3 for more details). Moreover, this approach to proving the Donoho–Stark uncertainty principle has apparently been independently rediscovered several times, e.g. [9, 31]. However, we are not aware of any previous work on the wide applicability of this simple approach; indeed, we found no other applications besides the two mentioned above. Importantly, the separation of the two parts of the proof above is only implicit in these papers (after all, the whole proof in these cases is a few lines), and we believe that making this partition explicit and general, as presented above, is the source of its power. We note that though the second part of this two-part approach is often straightforward to prove, it occasionally becomes interesting and non-trivial; see for instance, Section 4.3.3.

In this paper, will see how this approach leads to a very different proof of the original Heisenberg uncertainty principle, in which the measure $H$ is the variance. We will then prove other versions of that classical principle, which yield uncertainty when $H$ captures higher moments than the variance. We will also see how it extends to uncertainty principles where $H$ captures several notions of approximate support and "non-abelian" support, as well as new uncertainty principles where $H$ captures the ratio between other pairs of norms (which in turn will be useful for other applications). Throughout the paper, we attempt to establish tightness of the bounds, at least up to constant factors.

As mentioned, the *primary* uncertainty principle uses a very basic property that far from special to the Fourier transform, but is shared by (and so applies to) many other operators in different discrete and continuous settings, which we call *$k$-Hadamard operators*.

[5]This bound is universal in the sense that no operator like the Fourier transform is involved in this inequality; it deals only with a single function $g$.

[6]Variants of this idea will come in handy as well, e.g. proving that $H(g) \geq H_0(g)^{C'}$, yielding $C = C_0^{C'}$.

Although the study of uncertainty principles is nearly a century old, it continues to be an active and vibrant field of study, with new results coming out regularly (e.g. [8, 15, 17, 27, 29, 30]—all from the past 12 months!). While many uncertainty principles are unlikely to fit into the simple framework above, we nonetheless hope that our technique will help develop this theory and find further applications.

### 1.3 Outline of the paper

The paper has two main parts, which have somewhat different natures. The first is on finite-dimensional uncertainty principles, and the second is on infinite-dimensional ones. Both parts have several sections, each with a different incarnation of the uncertainty principle. In most sections, we begin with a known result, which we then show how to reprove and generalize using our framework. We remark that most sections and subsections are independent of one another, and can be read in more or less any order. We therefore encourage the reader to focus on those sections they find most interesting.

Before delving into these, we start with the preliminary Section 2. In it, we first formally state and prove the (extremely simple) primary $L^1 \to L^\infty$ uncertainty principle, from which everything else will follow. This leads to a natural abstract definition of operators amenable to this proof, which we call $k$-Hadamard matrices; most theorems in the finite-dimensional section will be stated in these terms (and a similar notion will be developed for the infinite-dimensional section).

In Section 3, we survey, reprove, and generalize finite-dimensional uncertainty principles, including the commutative and non-commutative support uncertainty principles of (respectively) Donoho–Stark and Meshulam, several old and new approximate support uncertainty principles, including those of Williams and Donoho–Stark, as well as some new uncertainty principles on norms. In Section 4 we turn to infinite-dimensional vector spaces and the uncertainty principles one can prove there. These include support inequalities for the Fourier transform on topological groups and several variants and extensions of Heisenberg's uncertainty principle, in particular to higher moments. Moreover, these theorems apply to the general class of Linear Canonical Transforms (which vastly extend the Fourier transform). Finally, Appendix A collects the proofs of some technical theorems.

## 2 The primary uncertainty principle and $k$-Hadamard matrices

As stated in the Introduction, the primary uncertainty principle that will yield all our other results is a theorem that lower-bounds the product of two $L^1$ norms by the product of two $L^\infty$ norms. In this section, we begin by stating this primary principle, as well as giving its (extremely simple) proof. We then define $k$-Hadamard matrices, which will be our main object of study in Section 3, and whose definition is motivated from the statement of the primary uncertainty principle (roughly, the definition of $k$-Hadamard matrices is "those matrices to which the primary uncertainty principle applies"). We end this section with

several examples of $k$-Hadamard matrices, which show that such matrices arise naturally in many areas of mathematics, such as group theory, design theory, random matrix theory, coding theory, and discrete geometry.

## 2.1 The primary uncertainty principle

We begin by recalling the definition of an operator norm. Let $V, U$ be any two real or complex vector spaces, and let $\|\cdot\|_V$ and $\|\cdot\|_U$ be any norms on $V, U$, respectively. Let $A : V \to U$ a linear map. Then the *operator norm* of $A$ is

$$\|A\|_{V\to U} = \sup_{0\neq v\in V} \frac{\|Av\|_U}{\|v\|_V} = \sup_{\substack{v\in V \\ \|v\|_V=1}} \|Av\|_U.$$

For $1 \leq p, q \leq \infty$, we will denote by $\|A\|_{p\to q}$ the operator norm of $A$ when $\|\cdot\|_V$ is the $L^p$ norm on $V$ and $\|\cdot\|_U$ is the $L^q$ norm on $U$.

With this notation, we can state our main theorem, which is the primary uncertainty principle that will underlie all our other results. We remark that this theorem, as stated below, is nearly tautological—our assumptions on the operators $A$ and $B$ are tailored to give the desired result by a one-line implication. Despite this simple nature, the strength of this theorem comes from the fact that many natural operators, such as the Fourier transform, satisfy these hypotheses.

**Theorem 2.1** (Primary uncertainty principle)**.** *Let $V, U$ be real or complex vector spaces, each equipped with two norms $\|\cdot\|_1$ and $\|\cdot\|_\infty$, and let $A : V \to U$ and $B : U \to V$ be linear operators. Suppose that $\|A\|_{1\to\infty} \leq 1$ and $\|B\|_{1\to\infty} \leq 1$. Suppose too that $\|BAv\|_\infty \geq k\|v\|_\infty$ for all $v \in V$, for some parameter $k > 0$. Then for any $v \in V$,*

$$\|v\|_1\|Av\|_1 \geq k\|v\|_\infty\|Av\|_\infty.$$

*Proof.* Since $\|A\|_{1\to\infty} \leq 1$, we have that

$$\|Av\|_\infty \leq \|v\|_1.$$

Similarly, since $\|B\|_{1\to\infty} \leq 1$,

$$\|BAv\|_\infty \leq \|Av\|_1.$$

Multiplying these two inequalities together, we find that

$$\|v\|_1\|Av\|_1 \geq \|Av\|_\infty\|BAv\|_\infty \geq k\|v\|_\infty\|Av\|_\infty,$$

as claimed. □

Note that Theorem 2.1 holds regardless of the dimensions of $V$ and $U$ (so long as the $L^1$ and $L^\infty$ norms are well-defined on them), and in Section 4, we will use this primary uncertainty principle in infinite dimensions. But for the moment, let us focus on finite

dimensions, in which case we take $\|\cdot\|_1$ and $\|\cdot\|_\infty$ to be the usual $L^1$ and $L^\infty$ norms on $\mathbb{R}^n$ or $\mathbb{C}^n$. When applying Theorem 2.1, we will usually take $B = A^*$. Note that the $1 \to \infty$ norm of a matrix is simply the maximum absolute value of the entries in the matrix, so $\|A\|_{1\to\infty} \le 1$ if and only if all entries of $A$ are bounded by 1 in absolute value. Moreover, if $B = A^*$, then[7] $\|B\|_{1\to\infty} = \|A\|_{1\to\infty}$. Thus, in this case, all we need to check in order to apply Theorem 2.1 is that $\|A^*Av\|_\infty \ge k\|v\|_\infty$ for all $v \in V$. This motivates the following definition, of $k$-Hadamard matrices, which are defined essentially as "those matrices that Theorem 2.1 applies to". We note that a similar definition was made by Dembo, Cover, and Thomas in [12, Section IV.C], and they state their discrete norm and entropy uncertainty principles in similar generality.

**Definition 2.2.** Let $A \in \mathbb{C}^{m\times n}$ be a matrix and $k > 0$. We say that $A$ is *$k$-Hadamard* if every entry of $A$ has absolute value at most 1 and $\|A^*Av\|_\infty \ge k\|v\|_\infty$ for all $v \in \mathbb{C}^n$. Equivalently, $A$ is $k$-Hadamard if all its entries are bounded by 1 in absolute value, $A^*A$ is invertible, and $\|(A^*A)^{-1}\|_{\infty\to\infty} \le 1/k$.

The next subsection consists of a large number of examples of $k$-Hadamard matrices which arise naturally in many areas of mathematics. Before proceeding, we end this subsection with three general observations. The first is the observation that the simplest way to ensure that[8] $\|A^*Av\|_\infty \ge k\|v\|_\infty$ is to assume that $A^*A = kI$. As we will see in the next section, many natural examples of $k$-Hadamard matrices have this stronger unitarity property.

Our second general observation is a rephrasing of Theorem 2.1, using the terminology of $k$-Hadamard matrices. Note that it is really identical to Theorem 2.1, but we state it separately for convenience.

**Theorem 2.3** (Primary uncertainty principle, rephrased)**.** *Let $A \in \mathbb{C}^{m\times n}$ be $k$-Hadamard. Then for any $v \in \mathbb{C}^n$, we have that*

$$\|v\|_1\|Av\|_1 \ge k\|v\|_\infty\|Av\|_\infty.$$

Thirdly, one can observe that the proof of Theorem 2.1 never actually used any properties whatsoever of the (usual) $L^1$ and $L^\infty$ norms, and the same result holds (with appropriately modified assumptions) for any choice of norms. We chose above to state the primary uncertainty principle specifically for the $L^1$ and $L^\infty$ norms simply because it is that statement that will be used to prove all subsequent theorems. However, for completeness, we now state the most general version which holds for all norms. The proof is identical to that of Theorem 2.1, so we omit it.

**Theorem 2.4** (Primary uncertainty principle, general version)**.** *Let $V, U$ be real or complex vector spaces, and let $A : V \to U$ and $B : U \to V$ be linear operators. Let $\|\cdot\|_{V^{(1)}}, \|\cdot\|_{V^{(2)}}$ be two norms on $V$, and $\|\cdot\|_{U^{(1)}}, \|\cdot\|_{U^{(2)}}$ two norms on $U$. Suppose that $\|A\|_{V^{(1)}\to U^{(2)}} \le 1$ and $\|B\|_{U^{(1)}\to V^{(2)}} \le 1$, and suppose that $\|BAv\|_{V^{(2)}} \ge k\|v\|_{V^{(2)}}$ for all $v \in V$ and some $k > 0$. Then for any $v \in V$,*

$$\|v\|_{V^{(1)}}\|Av\|_{U^{(1)}} \ge k\|v\|_{V^{(2)}}\|Av\|_{U^{(2)}}.$$

[7] In fact, this holds in greater generality: as long as $\|\cdot\|_1$ and $\|\cdot\|_\infty$ are dual norms on any inner product spaces, we will have that $\|A\|_{1\to\infty} = \|A^*\|_{1\to\infty}$.

[8] And indeed, to have that $\|A^*Av\| = k\|v\|$ for any norm whatsoever.

## 2.2 Examples of $k$-Hadamard matrices

We end this section by collecting several classes of examples of $k$-Hadamard matrices. While this subsection may be skipped at first reading, its main point is to demonstrate the richness of operators for which uncertainty principles hold. As remarked above, many of these matrices actually satisfy the stronger property that $A^*A = kI$.

**Hadamard matrices** Observe that if $A$ is a $k$-Hadamard $n \times n$ matrix, then $k \leq n$, and thus $n$-Hadamard matrices are best possible. One important class of $n$-Hadamard matrices are the ordinary[9] Hadamard matrices, which are $n \times n$ matrices $A$ with all entries in $\{-1, 1\}$ and with $A^*A = nI$. There are many constructions of Hadamard matrices, notably Paley's constructions [28] coming from quadratic residues in finite fields. Moreover, one can always take the tensor product of two Hadamard matrices and produce a new Hadamard matrix, which allows one to generate an infinite family of Hadamard matrices from a single example, such as the $2 \times 2$ matrix $\left(\begin{smallmatrix} 1 & 1 \\ -1 & 1 \end{smallmatrix}\right)$. Of course, these examples are nothing but the Fourier transform matrices over Hamming cubes. We remark that there are still many open questions about Hadamard matrices, most notably the so-called Hadamard conjecture, which asserts that $n \times n$ Hadamard matrices should exist for all $n$ divisible by 4.

**Complex Hadamard matrices** However, $n$-Hadamard $n \times n$ matrices are more general than Hadamard matrices, because we do not insist that the entries be real. In fact, one can show that $n$-Hadamard matrices are precisely complex Hadamard matrices, namely matrices with entries on the unit circle $\{z \in \mathbb{C} : |z| = 1\}$ whose rows are orthogonal. There is a rich theory to these matrices, with connections to operator algebras, quantum information theory, and other areas of mathematics; for more, we refer to the survey [2].

**The Fourier transform** A very important class of complex Hadamard matrices consists of Fourier transform matrices: if $G$ is a finite abelian group, then we may normalize its Fourier transform matrix so that all entries have norm 1, and the Fourier inversion formula precisely says that this matrix multiplied by its adjoint is $|G|I$. Thus, Fourier transform matrices are $n$-Hadamard $n \times n$ matrices, where $n = |G|$. More generally, quantum analogues of the Fourier transform can also be seen as $k$-Hadamard matrices; for more information, see [22].

**Other explicit square matrices** We do not insist that $k = n$ in our $n \times n$ Hadamard matrices. For $k < n$, special cases of $k$-Hadamard matrices with $A^*A = kI$ have been studied in the literature. For instance, such matrices with $k = n - 1$, real entries, and zeros on the diagonal are called *conference matrices*, and *weighing matrices* are such matrices with $k < n$ and all entries in $\{-1, 0, 1\}$; both of these have been studied in connection with design theory. See [11] for more details.

[9] Whence our name for such matrices.

**Random matrices** For less structured examples of $k$-Hadamard matrices, let $M$ be a random $n \times n$ unitary (or orthogonal) matrix, i.e. a matrix sampled from the Haar measure on $\mathrm{U}(n)$ (or $\mathrm{O}(n)$). It is well-known that with high probability as $n \to \infty$, every entry of $M$ will have norm $O(\sqrt{\log n/n})$; see [13, Theorem 8.1] for a simple proof, and [23] for far more precise results, including the determination of the correct constant hidden in the big-$O$. Thus, if we multiply $M$ by $c\sqrt{n/\log n}$ for an appropriate constant $c > 0$, we will obtain with high probability a $k$-Hadamard matrix with $k = \Omega(n/\log n)$. This shows that an appropriately chosen random matrix will be $\Omega(n/\log n)$-Hadamard with high probability, which is best possible up to the logarithmic factor.

**Rectangular matrices and codes** Recall that we do not require our $k$-Hadamard matrices to be square, which corresponds to not insisting that $V$ and $U$ have the same dimension in Theorem 2.4. If $A$ is an $m \times n$ $k$-Hadamard matrix, then $k \leq m$. One example of a non-square matrix attaining this bound is the $2^n \times n$ matrix $S$ whose rows consist of all vectors in $\{-1, 1\}^n$. Then distinct columns of $S$ are orthogonal because they disagree in exactly $2^{n-1}$ coordinates, which implies that $S^*S = 2^n I$, and so $S$ is $2^n$-Hadamard.

Note that the columns of $S$ are simply the codewords of the Hadamard code, viewed as vectors in $\{-1, 1\}^{2^n}$ rather than in $\{0, 1\}^{2^n}$. A similar construction works for all binary codes with an appropriate minimum distance. If we view the codewords as vectors with $\pm 1$ entries and form a matrix $S$ whose columns are these codewords, then the minimum distance condition will imply that the columns are nearly orthogonal, and thus that the diagonal entries of $S^*S$ will be much larger than the off-diagonal entries. Thus, $S$ will be $k$-Hadamard for a value of $k$ depending on the minimum distance of the code.

**Incidence matrices of finite geometries** If $q$ is a prime power, let $\mathrm{PG}(2, q)$ be the projective plane over the field $\mathbb{F}_q$. Then setting $n = q^2 + q + 1$, we can let $A$ be the $n \times n$ incidence matrix of points and lines in $\mathrm{PG}(2, q)$, namely the matrix whose rows and columns are indexed by the points and lines, respectively, of $\mathrm{PG}(2, q)$, and whose $(p, \ell)$ entry is 1 if $p \in \ell$, and 0 otherwise. Then $A$ certainly has all its entries bounded by 1. Moreover, each column has exactly $q + 1$ ones, and distinct columns have inner product 1, since any two lines intersect at exactly one point. This implies that $A^*A = qI + J$, where $J$ is the all-ones matrix. It is not too hard to see from this that $\|A^*Av\|_\infty \geq \frac{q}{2}\|v\|_\infty$ for any $v \in \mathbb{C}^n$, which implies that $A$ is a $k$-Hadamard $n \times n$ matrix, where $k \geq q/2 = \Theta(\sqrt{n})$.

Similarly, one can consider the $d$-dimensional projective space $\mathrm{PG}(d, q)$ over $\mathbb{F}_q$, and form the incidence matrix of $a$-flats and $b$-flats, for any $0 \leq a < b < d$. It will be a (not necessarily square) $k$-Hadamard matrix, for some value of $k$ depending on $a, b$, and $d$.

# 3 Finite-dimensional uncertainty principles

In this section, we show how to use our general uncertainty principle for $k$-Hadamard matrices, Theorem 2.3, to prove a number of uncertainty principles in finite dimensions. We start with the basic support-size uncertainty principle of Donoho and Stark, and then move on to Meshulam's generalization of it to arbitrary finite groups. We next proceed to prove several uncertainty principles for various notions of approximate support, and conclude with a collection of uncertainty principles for ratios of other norms, which will be useful for us later when we prove the Heisenberg uncertainty principle.

Most of our results in this section generalize known theorems about the Fourier transform on finite groups. However, as we demonstrate below, they do not actually need any of the algebraic structure of the Fourier transform (or even of an underlying group), and instead all follow from the fact that Fourier transform matrices are $k$-Hadamard.

## 3.1 The Donoho–Stark support-size uncertainty principle

For a vector $v \in \mathbb{C}^n$, let $\mathrm{supp}(v)$ be its *support*, namely the set of coordinates $i$ where $v_i \neq 0$. Similarly, if $f : G \to \mathbb{C}$ is a function on a finite group, we denote by $\mathrm{supp}(f)$ the set of $x \in G$ for which $f(x) \neq 0$. Recall that if $G$ is a finite abelian group, we denote by $\widehat{G}$ the *dual group*, which consists of all homomorphisms from $G$ to the circle group $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$. $\widehat{G}$ forms an abelian group under pointwise multiplication, and it is in fact (non-canonically) isomorphic to $G$. We define the Fourier transform $\hat{f} : \widehat{G} \to \mathbb{C}$ of a function $f : G \to \mathbb{C}$ by $\hat{f}(\chi) = \sum_{x \in G} f(x)\overline{\chi(x)}$. The basic uncertainty principle for the Fourier transform on finite abelian groups is the following theorem of Donoho and Stark.

**Theorem 3.1** (Donoho–Stark [14])**.** *Let $G$ be a finite abelian group. If $f : G \to \mathbb{C}$ is a non-zero function and $\hat{f} : \widehat{G} \to \mathbb{C}$ denotes its Fourier transform, then*

$$|\mathrm{supp}(f)||\mathrm{supp}(\hat{f})| \geq |G|.$$

Our first finite-dimensional result is an extension of Theorem 3.1 to arbitrary $k$-Hadamard matrices.

**Theorem 3.2** (Support-size uncertainty principle)**.** *Let $A \in \mathbb{C}^{m \times n}$ be a $k$-Hadamard matrix. Then for any non-zero $v \in \mathbb{C}^n$,*

$$|\mathrm{supp}(v)||\mathrm{supp}(Av)| \geq k.$$

*Proof.* This is the first demonstration of the principle articulated in the Introduction. We already have, from Theorem 2.3, that for any non-zero $v$, $\|v\|_1\|Av\|_1 \geq k\|v\|_\infty\|Av\|_\infty$. Thus, all we need is to bound the support-size of a function by the ratio of its norms, which is obvious: for any vector $u$,

$$\|u\|_1 = \sum_{i=1}^{n} |u_i| = \sum_{i \in \mathrm{supp}(u)} |u_i| \leq |\mathrm{supp}(u)|\|u\|_\infty.$$

Applying this bound to both $v$ and $Av$, we obtain the result. ∎

In this proof, the measure of localization we wished to study was the support of a vector. The uncertainty principle for this measure follows from the primary one (on the ratio of norms) via an inequality that holds for all vectors (bounding it by that ratio). This is an instance of the basic framework discussed in the Introduction, which will recur throughout.

**Remark.** We remark that in general, the bound in Theorem 3.1 (and thus also in Theorem 3.2) is tight. For instance, if $f$ is the indicator function of some subgroup $H \subseteq G$, then $\hat{f}$ will be a constant multiple of the indicator function of the dual subgroup $H^\perp \subseteq \widehat{G}$, and we have that $|H||H^\perp| = |G|$. Thus, $|\mathrm{supp}(f)||\mathrm{supp}(\hat{f})| = |G|$.

## 3.2 Support-size uncertainty principles for general finite groups

In this section, we show how our general framework can be used to extend the Donoho–Stark support-size uncertainty principle to the Fourier transform over arbitrary finite groups, abelian or non-abelian. Such an extension was already proved by Meshulam [26], for a *linear-algebraic* notion of support-size. Here we propose a natural *combinatorial* notion of support[10], and prove an uncertainty principle for it within our framework. Further, we prove that these two uncertainty principles are almost equivalent: they are identical for a certain class of functions, and are always equivalent up to a factor of 4. We note that both notions of support-size are natural and both extend the abelian case. Finally, at the end of the section, we provide another, new uncertainty principle of norms for general groups, proved by Greg Kuperberg, which provides a different proof Meshulam's theorem using our framework.

To facilitate a natural combinatorial definition of support, we embed both the "time domain" (namely, functions on the group), and the "Fourier domain" (namely, their image under the Fourier transform) as sub-algebras of the matrix ring $\mathbb{C}^{n\times n}$, where $n = |G|$. Then the notion of support becomes the standard one, namely the set of non-zero entries of these matrices. This embedding does much more. It gives as well a natural definition of norms (treating these matrices as vectors), and accommodates a description of the Fourier transform as a $k$-Hadamard operator. These yield a proof of the our support-size uncertainty inequality that is almost identical to the one in the abelian case.

### 3.2.1 Preliminaries: the Fourier transform in general finite groups

We now recall the basic notions of the Fourier transform of general finite groups (aka their representation theory) using the embedding above, which also affords a definition of the inverse Fourier transform which looks nearly identical to the abelian case. We refer the reader to the comprehensive text [33] on the representation theory of finite groups for a standard exposition of these concepts.

Let $G$ be an arbitrary finite group of order $n$, and let $\mathbb{C}[G]$ denote its group algebra. We embed $\mathbb{C}[G]$ as a sub-algebra of $\mathbb{C}^{n\times n}$ as follows. Given an element $f = \sum_{x\in G} f(x)x$, let $T_f$ denote left-multiplication by $f$ in $\mathbb{C}[G]$. Then $T_f$ is a linear map $\mathbb{C}[G] \to \mathbb{C}[G]$. Moreover,

[10] In his paper, Meshulam also defines a certain combinatorial measure of support size, which (as he points out) is much weaker than his linear-algebraic one.

since $\mathbb{C}[G]$ is equipped with a standard basis, namely the basis of delta functions on $G$, we can represent $T_f$ as an $n \times n$ matrix, and it is straightforward to see that both addition and multiplication of matrices corresponds to addition and multiplication in $\mathbb{C}[G]$. So we henceforth think of $\mathbb{C}[G]$ as the subspace of $\mathbb{C}^{n \times n}$ consisting of all matrices $T_f$.

If $G$ were abelian, then conjugating by the Fourier transform matrix would simultaneously diagonalize all $T_f$, with the diagonal entries precisely being the values of $\hat{f}$. If $G$ is non-abelian, then such a complete simultaneous diagonalization is impossible, but we can get maximal one possible; namely, conjugating by an appropriate matrix, which we also call the Fourier transform, turns each $T_f$ into a block-diagonal matrix with specified block sizes, uniformly for all $f$, as follows.

Let $\rho_1, \ldots, \rho_t$ be the irreducible representations of $G$ over $\mathbb{C}$, i.e. each $\rho_i$ is a homomorphism $G \to GL(W_i)$, where $W_i$ is a vector space over $\mathbb{C}$ of dimension $d_i$. We may assume that $\rho_1, \ldots, \rho_t$ are unitary representations, meaning that $\rho_i(x)$ is a unitary transformation on $W_i$ for all $x \in G$. Recall that $n = d_1^2 + \cdots + d_t^2$. Then we define the Fourier transform as follows.

**Definition 3.3** (The Fourier transform)**.** Given a function $f : G \to \mathbb{C}$, its *Fourier transform* is defined by $\hat{f}(\rho_i) = \sum_{x \in G} f(x)\rho_i(x)$, so that $\hat{f}(\rho_i)$ is a linear transformation $W_i \to W_i$.

We also henceforth fix an orthonormal basis $E_i$ of each $W_i$, and everything that follows will implicitly depend on these choices of bases. In particular, we may now think of the linear maps $\rho_i(x)$ and $\hat{f}(\rho_i)$ as a $d_i \times d_i$ matrices, represented in the basis $E_i$. We remark that, as above, one can define the Fourier transform without reference to any bases, but that everything we do from now on, such as defining the support and its size, will need these bases.

To define the Fourier transform matrix, we describe its columns (first, up to scaling): these are the so-called *matrix entry* vectors. For indices $i \in [t]$ and $j, k \in [d_i]$, we define the matrix entry vector $c(i; j, k) \in \mathbb{C}^n$ as follows. It is a vector whose coordinates are indexed by elements of $G$, and whose $x$ coordinate is the $(j, k)$ entry of the matrix $\rho_i(x)$; observe that in the abelian case these vectors are simply the $n$ characters of $G$. A simple consequence of Schur's lemma is that these vectors are orthogonal; see [33, Corollaries 2–3] for a proof.

**Proposition 3.4** (Orthogonality of matrix entries)**.** *We have*

$$\langle c(i; j, k), c(i'; j', k') \rangle = \begin{cases} n/d_i & \text{if } i = i', j = j', k = k' \\ 0 & \text{otherwise.} \end{cases}$$

We can now formally define the Fourier transform matrix and establish its basic properties.

**Definition 3.5** (Fourier transform matrix)**.** Let $F$ be the $n \times n$ matrix whose rows are indexed by $G$ and whose columns are indexed by tuples $(i; j, k)$ in lexicographic order, and whose $(i; j, k)$ column is the vector $\sqrt{d_i}c(i; j, k)$. We call $F$ the *Fourier transform matrix*.

Observe that Proposition 3.4 implies that $F^*F = FF^* = nI$. Moreover, the key diagonalization property of $F$ mentioned above is that for any function $f : G \to \mathbb{C}$, we have that $FT_fF^*$ is a block-diagonal matrix, whose blocks are just $d_i$ copies of the matrices[11] $n\hat{f}(\rho_i)$.

Thus, we think of the Fourier transform as simply a change of basis (and a dilation) on the matrix space $\mathbb{C}^{n\times n}$. Recall that we had already embedded $\mathbb{C}[G]$ in this space by mapping $f \in \mathbb{C}[G]$ to the matrix $T_f$. We think of its Fourier transform as the block-diagonal matrix $\widehat{T_f} \coloneqq FT_fF^*$. Moreover, we think of the subspace of $\mathbb{C}^{n\times n}$ consisting of all block-diagonal matrices with $d_i$ identical blocks of size $d_i \times d_i$ as the "Fourier subspace". Then the change of basis given by $F$ precisely maps the subspace corresponding to $\mathbb{C}[G]$ to this Fourier subspace. Note that if $G$ is abelian, then each $W_i$ is one-dimensional, and we find that $\widehat{T_f}$ is simply a diagonal matrix whose diagonal entries are the values of $\hat{f}$.

### 3.2.2 Notions of support for $\hat{f}$

With this setup, there is a clear candidate for the support of $\hat{f}$. Namely, we define $\mathrm{supp}(\hat{f})$ to simply be the set of non-zero entries of the matrix $\widehat{T_f}$. Note that since the block $\hat{f}(\rho_i)$ appears $d_i$ times in $\widehat{T_f}$, we have that

$$|\mathrm{supp}(\hat{f})| = \sum_{i=1}^{t} d_i |\mathrm{supp}(\hat{f}(\rho_i))|,$$

where $\mathrm{supp}(\hat{f}(\rho_i))$ denotes the set of non-zero entries of the matrix $\hat{f}(\rho_i)$. Recall that this matrix depended on the choice of the bases $E_i$, so we also make the following definition.

**Definition 3.6.** The *minimum support-size* of $\hat{f}$ is

$$|\text{min-supp}(\hat{f})| = \min_{E_1,\ldots,E_t} |\mathrm{supp}(\hat{f})|,$$

where the minimum is over all choices of orthonormal bases $E_1, \ldots, E_t$ for $W_1, \ldots, W_t$.

Thus, the minimum support-size of $\hat{f}$ is simply its support size in its most efficient representation. We note that if $G$ is abelian, then all $W_i$ are one-dimensional, and in particular the choice of basis affects nothing. So if $G$ is abelian, then both $|\mathrm{supp}(\hat{f})|$ and $|\text{min-supp}(\hat{f})|$ simply recover our earlier notion of the support-size of $\hat{f}$.

Meshulam proposed an alternative notion for the support-size of $\hat{f}$, which we call the rank-support, and which is defined as follows.

**Definition 3.7** (Meshulam [26])**.** Given $f : G \to \mathbb{C}$, the *rank-support* of $\hat{f}$ is $\text{rk-supp}(\hat{f}) = \operatorname{rank} T_f$.

[11] It is a little strange to have the blocks of $\widehat{T_f}$ be $n\hat{f}(\rho_i)$, rather than simply $\hat{f}(\rho_i)$. Of course, we could have normalized $F$ differently, so as to avoid this factor of $n$. However, we chose not to do this to be consistent with our earlier normalization of $k$-Hadamard matrices.

We note that, with this definition, it is not at all clear how rk-supp($\hat{f}$) even depends on $\hat{f}$, let alone in what sense it can be thought of as a support-size. However, since similar matrices have the same rank, we see that $\operatorname{rank} T_f = \operatorname{rank} \widehat{T_f}$, and since $\widehat{T_f}$ is a block-diagonal matrix, we see that

$$\operatorname{rank} \widehat{T_f} = \sum_{i=1}^{t} d_i \operatorname{rank}(\hat{f}(\rho_i)).$$

In particular, this connection shows that if $G$ is abelian, then we also get that rk-supp($\hat{f}$) = |supp($\hat{f}$)|. Meshulam's definition of rank-support has some advantages over our definition of minimum support-size; importantly, the rank-support does not depend on the choices of bases $E_1, \dots, E_t$. However, it is not obviously related to any notion of *support* of the Fourier transform—instead, it jumps directly to a notion of its *size*. In contrast, we offer, for each basis, a notion of support of $\hat{f}$, namely the set of non-zero entries in $\widehat{T_f}$, and then we pick the smallest possible (i.e. in the most "efficient" basis) to define its size. As mentioned, both definitions agree with |supp($\hat{f}$)| for abelian groups $G$, so both can be considered reasonable notions of support-size. The two notions can be related as follows, where we say that a function $f$ is *Hermitian* if $f(x) = \overline{f(x^{-1})}$ for all $x \in G$.

**Lemma 3.8.** *For any function $f : G \to \mathbb{C}$, we have that* rk-supp($\hat{f}$) $\leq$ |min-supp($\hat{f}$)|. *Moreover, if $f$ is a Hermitian function, then* rk-supp($\hat{f}$) = |min-supp($\hat{f}$)|.

*Proof.* If $\widehat{T_f}$ has $s$ non-zero entries, then in particular it has at most $s$ non-zero columns, which implies that $\operatorname{rank} \widehat{T_f} \leq |\operatorname{supp}(\hat{f})|$ for any bases $E_1, \dots, E_t$. This implies the first inequality by minimizing over these bases.

For the second, suppose that $f$ is Hermitian. This implies that each $\hat{f}(\rho_i)$ is Hermitian, since

$$\left(\hat{f}(\rho_i)\right)^* = \sum_{x \in G} \overline{f(x)} \left(\rho_i(x)\right)^* = \sum_{x \in G} \overline{f(x)} \rho_i(x^{-1}) = \sum_{x \in G} f(x^{-1}) \rho_i(x^{-1}) = \hat{f}(\rho_i).$$

Thus, there exists an orthonormal basis $E_i$ for $W_i$ in which $\hat{f}(\rho_i)$ is a diagonal matrix. Using such a basis for each $W_i$, we see that $\widehat{T_f}$ is diagonal, at which point its rank precisely equals the number of non-zero diagonal entries. This proves the reverse inequality for Hermitian $f$. □

### 3.2.3 Uncertainty principles for the min-support and the rank-support

One other connection between the rank-support and the minimum support-size is that both of them satisfy an uncertainty principle like that of Donoho and Stark. For the rank-support, this was proven by Meshulam.

**Theorem 3.9** (Meshulam [26])**.** *For any finite group $G$ and any $f : G \to \mathbb{C}$,*

$$|\operatorname{supp}(f)| \operatorname{rk-supp}(\hat{f}) \geq |G|.$$

For the minimum support-size, this is main result of this section.

**Theorem 3.10.** *For any finite group $G$ and any $f : G \to \mathbb{C}$,*

$$|\mathrm{supp}(f)||\text{min-supp}(\hat{f})| \geq |G|.$$

From Lemma 3.8, we see that Meshulam's Theorem 3.9 implies our Theorem 3.10. Moreover, for Hermitian functions $f$, the two theorems are precisely equivalent. Finally, we can prove a reverse implication, up to a factor of 4.

**Lemma 3.11.** *Theorem 3.10 implies that for any function $f : G \to \mathbb{C}$,*

$$|\mathrm{supp}(f)|\,\text{rk-supp}(\hat{f}) \geq \frac{|G|}{4}.$$

*Proof.* Consider the function $g : G \to \mathbb{C}$ defined by $g(x) = f(x) + \overline{f(x^{-1})}$. Then $g$ is Hermitian, so by Lemma 3.8 we have that $\text{rk-supp}(\hat{g}) = |\text{min-supp}(\hat{g})|$. As both the rank of matrices and the support of vectors are subadditive, we have that

$$|\mathrm{supp}(g)| \leq 2|\mathrm{supp}(f)| \qquad \text{and} \qquad \text{rk-supp}(\hat{g}) \leq 2\,\text{rk-supp}(\hat{f}).$$

Putting this all together, we find that

$$|\mathrm{supp}(f)|\,\text{rk-supp}(\hat{f}) \geq \frac{1}{4}|\mathrm{supp}(g)|\,\text{rk-supp}(\hat{g}) = \frac{1}{4}|\mathrm{supp}(g)||\text{min-supp}(\hat{g})| \geq \frac{|G|}{4}. \qquad \square$$

**Remark.** The bound in Meshulam's Theorem 3.9 is tight if $f$ is the indicator function of some subgroup of $G$, as observed in [26]. As such an indicator function is Hermitian, this also shows that the bound in Theorem 3.10 is in general best possible.

Our proof of Theorem 3.10 more or less follows the abelian case, namely it's a direct application of our general framework. We define linear operators $A, B : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$ by[12] $A \circ M = FMF^*$ and $B \circ M = F^*MF$, for any $M \in \mathbb{C}^{n\times n}$. Note that $A \circ T_f = \widehat{T_f}$. The main properties of these operators are captured in the following lemma, which simply says that $A$ acts as an $n^2$-Hadamard operator on the subspace $\mathbb{C}[G] \subset \mathbb{C}^{n\times n}$.

**Lemma 3.12.** *Let $\mathbb{C}[G] \cong V \subset \mathbb{C}^{n\times n}$ be the subspace consisting of the matrices $T_f$, and let $U \subset \mathbb{C}^{n\times n}$ be the Fourier subspace consisting of all block-diagonal matrices with $d_i$ identical blocks of size $d_i \times d_i$. Then the following hold.*

*(i) For any $M \in V$, we have that $B \circ (A \circ M) = n^2 M$.*

*(ii) For any $M \in V$, we have that $\|A \circ M\|_\infty \leq \|M\|_1$*

*(iii) For any $N \in U$, we have that $\|B \circ N\|_\infty \leq \|N\|_1$.*

[12] We use the notation $\circ$ to denote the action of $A$ and $B$ to avoid confusion with the notation for matrix multiplication.

The proof is straightforward, and we defer it to Appendix A. However, with this lemma in hand, we can prove Theorem 3.10.

*Proof of Theorem 3.10.* We begin by fixing bases $E_1, \dots, E_t$ that are minimizers in the definition of $|\text{min-supp}(\hat{f})|$. By applying the support-size uncertainty principle for $k$-Hadamard matrices, Theorem 2.1, to the operators $A, B$ as above, we see that

$$|\text{supp}(T_f)||\text{supp}(A \circ T_f)| \geq n^2.$$

Recall that $A \circ T_f$ is simply $\widehat{T_f}$, so the second term is simply $|\text{min-supp}(\hat{f})|$. For the first term, observe that each column of $T_f$ is simply a permutation of the values that $f$ takes. This implies that $|\text{supp}(T_f)| = n|\text{supp}(f)|$, as every non-zero value of $f$ is repeated exactly $n$ times in $T_f$. Thus, dividing by $n$ gives the claimed bound. □

### 3.2.4 Kuperberg's proof of Meshulam's uncertainty principle

After reading a draft of this paper, Greg Kuperberg (personal communication) discovered a new norm uncertainty principle for the Fourier transform over non-abelian groups, which can be proved using our framework and which implies Meshulam's Theorem 3.9 as a simple corollary. To state Kuperberg's theorem, we first need to define the Schatten norms of a matrix.

**Definition 3.13** (Schatten norms)**.** Let $M \in \mathbb{C}^{n \times n}$ be a matrix and $p \in [1, \infty]$ be a parameter. The *Schatten $p$-norm* of $M$, denoted $\|M\|_p^{(S)}$, is defined by

$$\|M\|_p^{(S)} = \text{Tr}\left((M^* M)^{p/2}\right)^{1/p}.$$

Equivalently, if $\sigma = (\sigma_1, \dots, \sigma_n)$ is the vector of singular values of $M$, then the Schatten $p$-norm of $M$ is simply the ordinary $L^p$ norm of $\sigma$, i.e. $\|M\|_p^{(S)} = \|\sigma\|_p$.

The Schatten norms are invariant under left- or right-multiplication by unitary matrices, so $\|T_f\|_p^{(S)} = \frac{1}{n}\|\widehat{T_f}\|_p^{(S)}$, with the factor of $n$ coming from our normalization of $F$ so that $\frac{1}{\sqrt{n}}F$ is unitary. Moreover, since $\widehat{T_f}$ is a block-diagonal matrix with $d_i$ blocks of $n\hat{f}\rho_i$, we have that

$$\|T_f\|_1^{(S)} = \frac{1}{n}\|\widehat{T_f}\|_1^{(S)} = \sum_{i=1}^{t} d_i \|\hat{f}(\rho_i)\|_1^{(S)} \quad \text{and} \quad \|T_f\|_\infty^{(S)} = \frac{1}{n}\|\widehat{T_f}\|_\infty^{(S)} = \max_{i \in [t]} \|\hat{f}(\rho_i)\|_\infty^{(S)}. \quad (3)$$

With this definition, we can state Kuperberg's norm uncertainty principle for the Fourier transform over finite groups. We state it for the Schatten norms of $\widehat{T_f}$, but by (3), we could just as well replace $\widehat{T_f}$ by $T_f$ in the following theorem.

**Theorem 3.14** (Kuperberg)**.** *Let $G$ be a group of order $n$ and $f : G \to \mathbb{C}$ a non-zero function. We have that*

$$\frac{\|f\|_1}{\|f\|_\infty} \cdot \frac{\|\widehat{T_f}\|_1^{(S)}}{\|\widehat{T_f}\|_\infty^{(S)}} \geq n.$$

Meshulam's Theorem 3.9 is a simple corollary of this theorem.

*Proof of Theorem 3.9.* We already know that $|\mathrm{supp}(f)| \geq \|f\|_1/\|f\|_\infty$. Additionally, it is well-known that for any matrix $M$,

$$\mathrm{rank}\, M \geq \frac{\|M\|_1^{(S)}}{\|M\|_\infty^{(S)}}.$$

This can be seen from the definition of $\|M\|_p^{(S)}$ as the $L^p$ norm of the vector $\sigma$ of singular values of $M$. Indeed, the rank of $M$ is simply the number of non-zero singular values, i.e. $\mathrm{rank}\, M = |\mathrm{supp}(\sigma)|$, from which the above inequality follows. This shows, using Theorem 3.14, that

$$|\mathrm{supp}(f)|\, \text{rk-supp}(\hat{f}) = |\mathrm{supp}(f)|\, \mathrm{rank}\, \widehat{T_f} \geq \frac{\|f\|_1}{\|f\|_\infty} \cdot \frac{\|\widehat{T_f}\|_1^{(S)}}{\|\widehat{T_f}\|_\infty^{(S)}} \geq n. \qquad \square$$

So to finish the proof, we need to prove Kuperberg's Theorem 3.14, whose proof is another application of our general framework.

*Proof of Theorem 3.14.* If we think of $f \in \mathbb{C}[G]$ as a row vector, then we can think of $F$ as a linear operator $\mathbb{C}[G] \to U$, which sends $f$ to $\frac{1}{n}\widehat{T_f}$; note the additional factor of $\frac{1}{n}$, coming from our earlier normalization of $\widehat{T_f} = FT_fF^*$. We first claim that $\|\frac{1}{n}\widehat{T_f}\|_\infty^{(S)} \leq \|f\|_1$, i.e. that $F$ has norm 1 as an operator from $\|\cdot\|_1$ to $\|\cdot\|_\infty^{(S)}$. By convexity of the Schatten-$\infty$ norm, it suffices to check this on the extreme points of the $L^1$ unit ball, i.e. for a delta function $f = \delta_x$, the function that takes value 1 at $x \in G$ and 0 on all other elements of $G$. But $T_{\delta_x}$ is simply a permutation matrix, so all of its singular values are 1, implying that $\|T_{\delta_x}\|_1^{(S)} = 1$. This in turn implies that $\|\widehat{T_{\delta_x}}\|_1^{(S)} = n$, by (3).

Recall that the Schatten-1 and Schatten-$\infty$ norms are dual on the matrix space $\mathbb{C}^{d_i \times d_i}$, which implies that they are dual on $U$ by the formulas in (3). Since the $L^1$ and $L^\infty$ norms on $\mathbb{C}[G]$ are also dual, the above also implies that $F^*$ has norm 1 as an operator from $\|\cdot\|_1^{(S)}$ to $\|\cdot\|_\infty$. Finally, we already know that $F^*F = nI$, so we conclude by the primary uncertainty principle, Theorem 2.1, that

$$\|f\|_1 \|\widehat{T_f}\|_1^{(S)} \geq n\|f\|_\infty \|\widehat{T_f}\|_\infty^{(S)}. \qquad \square$$

## 3.3 Uncertainty principles for notions of approximate support

The support-size uncertainty principle of Theorem 3.2 is rather weak, in the sense that the support size of a vector is a very fragile measure: coordinates with arbitrarily small non-zero values contribute to it. Stronger versions of this theorem, in which one considers instead the "essential support", namely the support of a vector after deleting such tiny entries[13], are much more robust. Such versions were sought first by Williams [38] in the continuous

[13]More precisely, deleting a small fraction of the total mass in some norm.

setting, and by Donoho and Stark [14] in the discrete setting. It turns out that using our approach it is easy to extend Theorem 3.2 to such a robust form for the $L^1$ norm, but not for $L^2$ (although Donoho and Stark's original $L^2$ proof does generalize to $k$-Hadamard matrices with $A^*A = kI$). We will describe both and compare them.

We start with some notation. If $v \in \mathbb{C}^n$ is a vector and $T \subseteq [n]$ is a set of coordinates, we denote by $v[T]$ the vector in $\mathbb{C}^{|T|}$ obtained by restricting $v$ to the coordinates in $T$. We use $T^{\mathsf{c}}$ to denote the complement of $T$ in $[n]$.

**Definition 3.15.** Let $\varepsilon \in [0,1]$ and $p \in [1,\infty]$. For a vector $v \in \mathbb{C}^n$ and a set $T \subseteq [n]$, we say that $v$ is $(p,\varepsilon)$-*supported* on $T$ if $\|v[T^{\mathsf{c}}]\|_p \leq \varepsilon \|v\|_p$.

We also define the $(p,\varepsilon)$-*support size* of $v$ to be

$$|\mathrm{supp}_\varepsilon^p(v)| = \min\{|T| : T \subseteq [n], v \text{ is } (p,\varepsilon)\text{-supported on } T\}.$$

**Remark.** In general, there may not exist a unique minimum-sized set $T$ in the definition of $|\mathrm{supp}_\varepsilon^p(v)|$, so the set "$\mathrm{supp}_\varepsilon^p(v)$" is not well-defined. However, we will often abuse notation and nevertheless write $\mathrm{supp}_\varepsilon^p(v)$ to denote an arbitrary set $T$ achieving the minimum in the definition of $|\mathrm{supp}_\varepsilon^p(v)|$.

The basic uncertainty principle concerning approximate supports was also given by Donoho and Stark, who proved the following.

**Theorem 3.16** (Donoho–Stark [14])**.** *Let $G$ be a finite abelian group and $f : G \to \mathbb{C}$ a non-zero function. For any $\varepsilon, \eta \in [0,1]$, we have that*

$$|\mathrm{supp}_\varepsilon^2(f)||\mathrm{supp}_\eta^2(\hat{f})| \geq |G|(1-\varepsilon-\eta)^2.$$

If one attempts to apply our basic framework to prove an uncertainty principle for approximate supports, one is naturally led to the following result. The analogous theorem for the Fourier transform on $\mathbb{R}$ was proven by Williams [38], in what we believe is the earliest application of this paper's approach to uncertainty principles.

**Theorem 3.17** ($L^1$-approximate support uncertainty principle)**.** *Let $A \in \mathbb{C}^{m\times n}$ be a $k$-Hadamard matrix and let $v \in \mathbb{C}^n$ be a non-zero vector. For any $\varepsilon, \eta \in [0,1]$, we have that*

$$|\mathrm{supp}_\varepsilon^1(v)||\mathrm{supp}_\eta^1(Av)| \geq k(1-\varepsilon)(1-\eta) \geq k(1-\varepsilon-\eta).$$

*Proof.* The second inequality follows from the first, so it suffices to prove the first. We may assume that $\varepsilon, \eta < 1$. The primary uncertainty principle, Theorem 2.3, says that

$$\frac{\|v\|_1}{\|v\|_\infty} \cdot \frac{\|Av\|_1}{\|Av\|_\infty} \geq k.$$

Following our framework, what we need to prove is the following claim: for every $\delta \in [0,1)$ and for every vector $u$,

$$\frac{|\mathrm{supp}_\delta^1(u)|}{1-\delta} \geq \frac{\|u\|_1}{\|u\|_\infty}$$

Applying this inequality to $v$ with $\varepsilon$ and to $Av$ with $\eta$ yields the desired result, so it suffices to prove this claim.

Let $T = \text{supp}^1_\delta(u)$. Note that since $\|u\|_1 = \|u[T]\|_1 + \|u[T^\complement]\|_1$, the definition of $T$ implies that $\|u[T]\|_1 \geq (1-\delta)\|u\|_1$. Observe too that $\|u[T]\|_\infty = \|u\|_\infty$, since $T$ just consists of the coordinates of $u$ of maximal absolute value (with ties broken arbitrarily). Since $u[T]$ has length $|T|$, this implies that $\|u[T]\|_1 \leq |T|\|u[T]\|_\infty = |T|\|u\|_\infty$. Combining our inequalities, we find that

$$(1-\delta)\|u\|_1 \leq \|u[T]\|_1 \leq |T|\|u\|_\infty = |\text{supp}^1_\delta(u)|\|u\|_\infty,$$

as claimed. □

Again, we obtain as a special case an uncertainty principle for the $L^1$ approximate support of a function and its Fourier transform.

**Corollary 3.18.** *Let $G$ be a finite abelian group and $f : G \to \mathbb{C}$ a non-zero function. For any $\varepsilon, \eta \in [0,1]$, we have that*

$$|\text{supp}^1_\varepsilon(f)||\text{supp}^1_\eta(\hat{f})| \geq |G|(1-\varepsilon)(1-\eta) \geq |G|(1-\varepsilon-\eta).$$

At first glance, Corollary 3.18 looks quite similar to Theorem 3.16. However, it is easy to construct vectors whose $(2,\varepsilon)$-support is much smaller than their $(1,\varepsilon)$-support. For instance, we can take $v_H \in \mathbb{C}^n$ to be the "harmonic vector" $(1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n})$. Then for any fixed $\varepsilon \in (0,1)$ and large $n$, we have that

$$\text{supp}^1_\varepsilon(v_H) = \Theta_\varepsilon(n^{1-\varepsilon}) \qquad \text{and} \qquad \text{supp}^2_\varepsilon(v_H) = \Theta(\varepsilon^{-2}).$$

In particular, if we keep $\varepsilon$ fixed and let $n \to \infty$, we see that $\text{supp}^1_\varepsilon(v_H)$ will be much larger than $\text{supp}^2_\varepsilon(v_H)$, which suggests that Theorem 3.16 will be stronger than Corollary 3.18 for such "long-tailed" vectors. In fact, this is not a coincidence or a special case: for constant $\varepsilon$, the 1-support will be at least as large than the 2-support of any vector. More precisely, we have the following.

**Proposition 3.19.** *For any vector $v \in \mathbb{C}^n$ and any $\varepsilon \in (0,1)$,*

$$|\text{supp}^1_{\varepsilon^2}(v)| \geq |\text{supp}^2_\varepsilon(v)|.$$

This proposition demonstrates that in general, Theorem 3.16 is stronger than our Corollary 3.18. Because the proof is somewhat technical, we defer it to Appendix A.

To conclude this section, we state a generalization of Theorem 3.16 that applies to all unitary $k$-Hadamard matrices.

**Theorem 3.20** ($L^2$-approximate support uncertainty principle)**.** *Let $A \in \mathbb{C}^{n\times n}$ be a $k$-Hadamard matrix with $A^*A = kI$, and let $\varepsilon, \eta \in [0,1]$. Let $v \in \mathbb{C}^n$ be a non-zero vector. Then*

$$|\text{supp}^2_\varepsilon(v)||\text{supp}^2_\eta(Av)| \geq k(1-\varepsilon-\eta)^2.$$

This proof is essentially identical to the original proof from [14], so we defer it to Appendix A. We stress that this proof we need to assume the stronger condition that $A^*A = kI$: unlike our other proofs, this proof uses crucially and repeatedly the fact that $\frac{1}{\sqrt{k}}A$ preserves the $L^2$ norm under this assumption. Similarly, it is important for this proof that the matrix $A$ be square, since we will need the same property for $A^*$.

We also note that it is impossible to prove this result in the same way we proved Theorem 3.17. Indeed, such a proof would necessarily need a $2 \to \infty$ norm uncertainty principle of the form $\|v\|_2\|Av\|_2 \geq C\|v\|_\infty\|Av\|_\infty$. In the next subsection, we will show (in Theorem 3.23) that such an inequality cannot hold unless $C$ is a constant independent of $k$. Therefore, one necessarily has to use an alternative approach (and stronger properties) to prove Theorem 3.20.

## 3.4 Uncertainty principles for other norms: possibility and impossibility

Generalizing the primary uncertainty principle, it is natural and useful to try to prove other uncertainty principles on norms for the finite Fourier transform, or more generally for other $k$-Hadamard operators. Indeed, it seems that one can use Theorem 2.4 directly to derive inequalities of the form $\|v\|_p\|Av\|_p \geq c(k)\|v\|_q\|Av\|_q$ for other norms $p$ and $q$, and for some constant $c(k)$.

However, the situation for other norms is trickier than for $p = 1$ and $q = \infty$. It is instructive to try this for two prototypical cases: first, $p = 1$ and $q = 2$, and next, $p = 2$ and $q = \infty$. In both cases, the constant $c(k)$ we obtain from a direct use of the general theorem is 1. As we shall see, this happens for different reasons in these two cases. In the first (and generally for $p = 1$ and any $q$), one can obtain a much better inequality, indeed a tight one, *indirectly* from the case $p = 1$ and $q = \infty$. In the second (and in general for $p \geq 2$ and any $q$), the constant $c(k) = 1$ happens to be essentially optimal, and one can only obtain a trivial result. We turn now to formulate and prove each of these statements. We note that, besides natural mathematical curiosity, there is a good reason to consider uncertainty principles for other norms: indeed, (a version of) the one we prove here for $p = 1$ and $q = 2$ will be key to our new proof of Heisenberg's uncertainty principle in Section 4.3.

### 3.4.1 Optimal norm uncertainty inequalities for $p = 1$

Consider first the case $p = 1$ and $q = 2$, and suppose we wish to prove such a norm uncertainty principle for the Fourier transform on a finite abelian group $G$. To apply Theorem 2.4, we would need to scale the Fourier transform matrix into a matrix $A$ with $\|A\|_{1\to 2}, \|A^*\|_{1\to 2} \leq 1$. The way to do so is to rescale so all the entries of $A$ have absolute value $1/\sqrt{|G|}$. But in that case $A^*A = I$, so one would simply obtain the inequality $\|v\|_1\|Av\|_1 \geq \|v\|_2\|Av\|_2$, which is not sharp. In fact, this inequality is trivial (and has nothing to do with "uncertainty"), since any vector $u$ satisfies $\|u\|_1 \geq \|u\|_2$. To obtain a stronger inequality, which is sharp, we instead use a simple *reduction* to the $1 \to \infty$ result of Theorem 2.3, which is a variant of the two-step framework articulated in the Introduction.

**Theorem 3.21** (Norm uncertainty principle, $p = 1$)**.** *Let $A \in \mathbb{C}^{m\times n}$ be $k$-Hadamard, and let $v \in \mathbb{C}^n$ be non-zero. Then for any $1 \le q \le \infty$, we have*

$$\|v\|_1\|Av\|_1 \ge k^{1-1/q}\|v\|_q\|Av\|_q.$$

*Proof.* The case $q = \infty$ is precisely Theorem 2.3, so we may assume that $q < \infty$. So, following our approach, all we need to prove is the bound

$$\frac{\|u\|_1}{\|u\|_q} \ge \left(\frac{\|u\|_1}{\|u\|_\infty}\right)^{(q-1)/q} \tag{4}$$

for any non-zero vector $u$, and plug it in our primary inequality $\|v\|_1\|Av\|_1 \ge k\|v\|_\infty\|Av\|_\infty$ for both $v$ and $Av$. This is simple: we compute

$$\|u\|_q^q = \sum_{i=1}^n |u_i|^q \le \|u\|_\infty^{q-1}\sum_{i=1}^n |u_i| = \|u\|_\infty^{q-1}\|u\|_1,$$

which implies that $\|u\|_1^{q-1}\|u\|_q^q \le \|u\|_1^q\|u\|_\infty^{q-1}$, which yields the bound (4). □

We get as a special case an uncertainty principle for the discrete Fourier transform, which we believe has not been previously observed.

**Corollary 3.22.** *For any $1 \le q \le \infty$, any finite abelian group $G$, and any non-zero function $f : G \to \mathbb{C}$, we have*

$$\|f\|_1\|\hat{f}\|_1 \ge |G|^{1-1/q}\|f\|_q\|\hat{f}\|_q.$$

**Remark.** This is tight if $f$ is the indicator function of a subgroup $H \subseteq G$. In that case, $\hat{f}$ is a constant multiple of the indicator function of the dual subgroup $H^\perp \subseteq \widehat{G}$, and we have that $|H||H^\perp| = |G|$. This shows that the result is tight, since the $q$-norm of an indicator function is exactly the $1/q$ power of its support size.

### 3.4.2 No non-trivial norm uncertainty inequalities for $p \ge 2$

Two special cases of Corollary 3.22, one of which is just Theorem 2.3, are that if $A$ is $k$-Hadamard, then

$$\|v\|_1\|Av\|_1 \ge k\|v\|_\infty\|Av\|_\infty \qquad \text{and} \qquad \|v\|_1\|Av\|_1 \ge \sqrt{k}\|v\|_2\|Av\|_2.$$

Looking at these two bounds, it is natural to conjecture that

$$\|v\|_2\|Av\|_2 \ge \sqrt{k}\|v\|_\infty\|Av\|_\infty, \tag{5}$$

which would of course be best possible if true. If we again attempt to prove this for the Fourier transform directly from Theorem 2.4, we need to scale the Fourier transform to a matrix $A$ with $2 \to \infty$ norm at most 1, which again requires taking $A$ to have all entries of absolute value $1/\sqrt{|G|}$. Then we again get that $A^*A = I$, and only obtain the trivial inequality $\|v\|_2\|Av\|_2 \ge \|v\|_\infty\|Av\|_\infty$.

In contrast to the previous subsection, this trivial bound is essentially tight, as shown by the following theorem.

**Theorem 3.23.** *Let $G$ be a finite abelian group of order $n$, and let $A$ be the Fourier transform matrix of $G$. Let $p \in [2, \infty], q \in [1, \infty]$ be arbitrary. There exists a vector $v \in \mathbb{C}^n$ with*

$$\|v\|_p \|Av\|_p \leq 2\|v\|_q \|Av\|_q.$$

*In particular, (5) is false in general.*

*Proof.* We normalize $A$ so that all its entries have absolute value 1, and assume without loss of generality that the first row and column of $A$ consist of all ones[14]. We define the vector $v = (1 + \sqrt{n}, 1, 1, \ldots, 1) \in \mathbb{C}^n$. Then $Av = \sqrt{n}v$, i.e. $v$ is an eigenvector of $A$ with eigenvalue $\sqrt{n}$; this can be seen by observing that $v$ is the sum of $(\sqrt{n}, 0, 0, \ldots 0)$ and $(1, 1, \ldots, 1)$, and the action of $A$ on these vectors is to swap them and multiply each by $\sqrt{n}$.

Moreover, we can compute that

$$\|v\|_p = \left[\left(\sqrt{n}+1\right)^p + n - 1\right]^{1/p} \qquad \text{and} \qquad \|v\|_q = \left[\left(\sqrt{n}+1\right)^q + n - 1\right]^{1/q}.$$

We claim that for any $a \geq 1, b \geq 0$, the function $h(x) = (a^x + b)^{1/x}$ is monotonically non-increasing for $x \geq 1$. Indeed, its derivative is

$$h'(x) = \frac{h(x)}{x^2(a^x + b)} \left(a^x \log(a^x) - (a^x + b)\log(a^x + b)\right).$$

The term $h(x)/(x^2(a^x + b))$ is positive, and the function $t \mapsto t \log t$ is increasing for $t \geq 1$, which implies that the parenthesized term is non-positive, so $h'(x) \leq 0$. This implies that $\|v\|_x$ is a non-increasing function of $x$, so we have that

$$\|v\|_p \leq \|v\|_2 = \sqrt{2n + 2\sqrt{n}} \qquad \text{and} \qquad \|v\|_q \geq \|v\|_\infty = \sqrt{n} + 1.$$

In particular, we find that $\sqrt{2}\|v\|_q \geq \sqrt{2}\|v\|_\infty \geq \|v\|_2 \geq \|v\|_p$. This shows us that

$$\frac{\|v\|_p}{\|v\|_q} \cdot \frac{\|Av\|_p}{\|Av\|_q} = \frac{\|v\|_p^2}{\|v\|_q^2} \leq 2. \qquad \square$$

### 3.4.3 The Hausdorff–Young inequality and the regime $1 < p < 2$

For the remaining range of $1 < p < 2$, we can obtain norm inequalities like those for $p = 1$. However, we need two additional hypotheses. First, we need to assume that our $k$-Hadamard matrix satisfies the stronger unitarity property that $A^*A = kI$, while such an assumption was unnecessary in the $p = 1$ case. Second, we will need to assume that the second norm index, $q$, is at most $p' = p/(p-1)$; this assumption was immaterial in the $p = 1$ case, since the dual index of 1 is $\infty$. We remark that we include this subsection only for completeness; the results here are known and use standard techniques, namely the Riesz–Thorin interpolation theorem and the log-convexity of the $L^p$ norms.

[14] This simply corresponds to indexing the rows and columns of $A$ so that the identity element of $G$ and $\widehat{G}$ come first.

To do this, we first prove a discrete analogue of the Hausdorff–Young inequality. This inequality was already observed[15] by Dembo, Cover, and Thomas [12, Equation (52)], who also stated it in the same general setting of unitary $k$-Hadamard matrices, as we now do.

**Proposition 3.24** (Discrete Hausdorff–Young inequality)**.** *Let $A \in \mathbb{C}^{n\times n}$ be a $k$-Hadamard matrix with $A^*A = kI$. Fix $1 < p < 2$, and let $p' = p/(p-1) \in (2,\infty)$. Then $\|A\|_{p\to p'} \leq k^{(p-1)/p}$.*

*Proof.* We already know that $\|A\|_{1\to\infty} \leq 1$, and our assumption that $A^*A = kI$ implies that $\|A\|_{2\to 2} = \sqrt{k}$. We may apply the Riesz–Thorin interpolation theorem [32, Theorem IX.17] to these bounds, which implies that $\|A\|_{p\to p'} \leq k^{(p-1)/p}$, as claimed. □

As a corollary, we obtain the following norm uncertainty principle for $1 < p < 2$.

**Theorem 3.25** (Norm uncertainty principle, $1 < p < 2$)**.** *Let $A \in \mathbb{C}^{n\times n}$ be a $k$-Hadamard matrix with $A^*A = kI$. Let $p \in (1,2)$ and $q \in [p,p']$ be norm indices. Then for any $v \in \mathbb{C}^n$,*

$$\|v\|_p \|Av\|_p \geq k^{\frac{q-p}{pq}} \|v\|_q \|Av\|_q$$

*Proof.* First, suppose that $q = p'$. In that case, we may multiply the conclusion of Proposition 3.24 for $A$ and $A^*$ and conclude that

$$k^{\frac{2(p-1)}{p}} \|v\|_p \|Av\|_p \geq \|Av\|_{p'} \|A^*Av\|_{p'} = k\|v\|_{p'} \|Av\|_{p'},$$

which implies the desired bound $\|v\|_p \|Av\|_p \geq k^{\frac{2-p}{p}} \|v\|_{p'} \|Av\|_{p'}$.

For smaller values of $q$, we use the above as a primary uncertainty principle, and derive the result by showing that

$$\frac{\|u\|_p}{\|u\|_q} \geq \left(\frac{\|u\|_p}{\|u\|_{p'}}\right)^{\frac{p-q}{pq-2q}} \tag{6}$$

for any non-zero vector $u$ and any $p \leq q \leq p'$. Let $\theta \in [0,1]$ be the unique number such that

$$\frac{1}{q} = \frac{1-\theta}{p} + \frac{\theta}{p'},$$

namely $\theta = \frac{p-q}{pq-2q}$. Then the log-convexity of the $L^p$ norms (also known as the generalized Hölder inequality) says that $\|u\|_q \leq \|u\|_p^{1-\theta} \|u\|_{p'}^{\theta}$, and rearranging this yields (6).

We now apply (6) to $u = v$ and $u = Av$, and conclude that

$$\frac{\|v\|_p}{\|v\|_q} \cdot \frac{\|Av\|_p}{\|Av\|_q} \geq \left(\frac{\|v\|_p}{\|v\|_{p'}} \cdot \frac{\|Av\|_p}{\|Av\|_{p'}}\right)^{\frac{p-q}{pq-2q}} \geq \left(k^{\frac{2-p}{p}}\right)^{\frac{p-q}{pq-2q}} = k^{\frac{q-p}{pq}}.$$ □

[15] They used this discrete Hausdorff–Young inequality to prove a discrete entropic uncertainty principle, analogous to that of Hirschman [21].

**Remark.** We remark that, as with the $1 \to q$ norm uncertainty principles above, Theorem 3.25 is tight for the Fourier transform. Indeed, if $v$ is the indicator vector of a subgroup of $G$, then $Av$ will be a constant multiple of the indicator vector of the dual subgroup, and the inequality in Theorem 3.25 will be an equality in this case.

What these subsections demonstrate is that the $1 \to \infty$ result of Theorem 2.3 is the strongest result of its form, in two senses. First, it implies the optimal $1 \to q$ inequalities for any $1 \leq q \leq \infty$, which cannot be obtained by a direct application of Theorem 2.4. Second, such $p \to q$ uncertainty principles for $p > 1$ are false in general, as shown by the fact that one cannot obtain a super-constant uncertainty even for the Fourier transform when $p \geq 2$. In the regime $1 < p < 2$, one can obtain tight inequalities whenever $p \leq q \leq p'$, at least for $k$-Hadamard matrices that satisfy the unitarity property $A^*A = kI$.

# 4 Uncertainty principles in infinite dimensions

In this section, we will state and prove various uncertainty principles that hold in infinite-dimensional vector spaces, primarily the Heisenberg uncertainty principle and its generalizations. We begin in Section 4.1 with general results that hold for the Fourier transform on arbitrary locally compact abelian groups. We then restrict to $\mathbb{R}$, and discuss in Section 4.2 a large class of operators for which our results hold, namely infinite-dimensional analogues of the $k$-Hadamard matrices we focused on in Section 3. These include the so-called Linear Canonical Transforms (LCT), a family of integral transforms generalizing of the Fourier and other transforms, which arise primarily in applications to optics. Finally, we move to prove the Heisenberg uncertainty principle and its variations for such operators in Section 4.3. In addition to obtaining a new proof which avoids using the analytic tools common in existing proofs, we also prove a number of generalizations. Most notably, we establish uncertainty principles for higher moments than the variance[16]. We also give new inequalities which are similar to Heisenberg's but are provably incomparable. We remark that in some of our proofs of existing inequalities, the constants obtained are worse than in the classical proofs.

## 4.1 The Fourier transform on locally compact abelian groups

We begin by recalling the basic definitions of the Fourier transform on locally compact abelian[17] groups, and proving some generalizations of earlier results in this context.

Let $G$ be a locally compact abelian group. Then $G$ can be equipped with a left-invariant Borel measure $\mu$, called the *Haar measure*, which is unique up to scaling. If we let $\widehat{G}$ denote the set of continuous group homomorphisms $G \to \mathbb{T}$, then $\widehat{G}$ is a group under pointwise multiplication. Moreover, if we topologize $\widehat{G}$ with the compact-open topology, then $\widehat{G}$ becomes an-

[16]Such results were already obtained by Cowling and Price [10], but again, our proof avoids their heavy analytic machinery.

[17]In fact, we believe that, as in Section 3.2, many of our results can be extended to infinite non-abelian groups, at least as long as all their irreducible representations are finite-dimensional.

other locally compact abelian group, which is called the *Pontryagin dual* of $G$. Given a function $f \in L^1(G)$, we can define its Fourier transform $\hat{f} : \widehat{G} \to \mathbb{C}$ by $\hat{f}(\chi) = \int f(x)\overline{\chi(x)}\,\mathrm{d}\mu(x)$, and it is easy to see that $\hat{f}$ is a well-defined element of $L^\infty(\widehat{G})$. Moreover, having chosen $\mu$, there exists a unique Haar measure $\nu$ on $\widehat{G}$ so that the *Fourier inversion formula* holds, namely so that $f(x) = \int \hat{f}(\chi)\chi(x)\,\mathrm{d}\nu(\chi)$ for $\mu$-a.e. $x$, as long as $\hat{f} \in L^1(\widehat{G})$. With this choice of $\nu$, we also have the *Plancherel formula*, that $\int |f|^2\,\mathrm{d}\mu = \int |\hat{f}|^2\,\mathrm{d}\nu$, as long as one side is well-defined. From now on, we will fix these measures $\mu$ and $\nu$, and all $L^p$ norms of functions will be defined by integration against these measures. Observe that from the definition of $\hat{f}$ and from the Fourier inversion formula, we have that the Fourier transform and inverse Fourier transform have norm at most 1 as operators $L^1 \to L^\infty$. Using this, we can prove an infinitary version of our primary uncertainty principle, Theorem 2.1.

**Theorem 4.1** (Primary uncertainty principle, infinitary version)**.** *Let $G$ be a locally compact abelian group with a Haar measure $\mu$, and let $\widehat{G}, \nu$ be the dual group and measure. Fix $1 \le q \le \infty$ and let $f \in L^1(G)$ be such that $\hat{f} \in L^1(\widehat{G})$. Then*

$$\|f\|_1\|\hat{f}\|_1 \ge \|f\|_\infty\|\hat{f}\|_\infty.$$

**Remark.** Throughout this section, we will frequently need the assumption that $f \in L^1(G)$ and $\hat{f} \in L^1(\widehat{G})$. To avoid having to write this every time, we make the following definition.

**Definition 4.2** (Doubly $L^1$ function)**.** We call function $f : G \to \mathbb{C}$ *doubly* $L^1$ if $f \in L^1(G)$ and $\hat{f} \in L^1(\widehat{G})$.

Note that $f$ being doubly $L^1$ implies that $f, \hat{f} \in L^\infty$, and thus that $f, \hat{f} \in L^p$ for all $p \in [1, \infty]$ by Hölder's inequality.

*Proof of Theorem 4.1.* For any $\chi \in \widehat{G}$, we have that

$$|\hat{f}(\chi)| = \left|\int f(x)\overline{\chi(x)}\,\mathrm{d}\mu(x)\right| \le \int |f(x)|\,\mathrm{d}\mu(x) = \|f\|_1,$$

since $|\chi(x)| = 1$. This implies that $\|\hat{f}\|_\infty \le \|f\|_1$. For the same reason, we see that $\|f\|_\infty \le \|\hat{f}\|_1$. Multiplying these inequalities gives the desired result. □

**Remark.** If we take $G$ to be a finite abelian group, this result appears to be a factor of $|G|$ worse than Theorem 2.3. However, this discrepancy is due to the fact that previously, we were equipping both $G$ and $\widehat{G}$ with the counting measure, which are not dual Haar measures. If we instead equip them with dual Haar measures (e.g. equipping $G$ with the counting measure and then equipping $\widehat{G}$ with the uniform probability measure), then this "extra" factor of $|G|$ would disappear, and we would get the statement of Theorem 4.1.

Using this theorem, we can obtain an analogue of the Donoho–Stark uncertainty principle, which holds for every locally compact abelian group. This result was first proved by Matolcsi and Szűcs [25], using the theory of spectral integrals.

**Theorem 4.3** (Support-size uncertainty principle for general abelian groups)**.** *Let $G, \mu, \widehat{G}, \nu$ be as above. Let $f : G \to \mathbb{C}$ be non-zero and doubly $L^1$. Then $\mu(\mathrm{supp}(f))\nu(\mathrm{supp}(\hat{f})) \geq 1$.*

*Proof.* The proof follows that of Theorem 3.2. Following our general approach, we claim that for any non-zero integrable function $g$ on any measure space $(X, \lambda)$, we have that

$$\lambda(\mathrm{supp}(g)) \geq \frac{\|g\|_1}{\|g\|_\infty}.$$

Applying this to $f$ and $\hat{f}$ and combining it with the primary uncertainty principle, Theorem 4.1, yields the desired result. To prove the claim, we simply compute

$$\|g\|_1 = \int_X |g(x)| \, \mathrm{d}\lambda(x) = \int_{\mathrm{supp}(g)} |g(x)| \, \mathrm{d}\lambda(x) \leq \|g\|_\infty \int_{\mathrm{supp}(g)} \mathrm{d}\lambda(x) = \lambda(\mathrm{supp}(g))\|g\|_\infty. \quad \square$$

In general, Theorem 4.3 is tight. This can be seen, for instance, by recalling that it is equivalent to Theorem 3.1 when $G$ is finite, and we already know that theorem to be tight when $f$ is the indicator function of a subgroup. However, Theorem 4.3 is tight even for some infinite groups. For instance, let $G$ be any compact abelian group, and let $\mu$ be the Haar probability measure on $G$. Then $\widehat{G}$ is a discrete group, and $\nu$ is the counting measure on $\widehat{G}$. If we let $f : G \to \mathbb{C}$ be the constant 1 function, then $\mu(\mathrm{supp}(f)) = 1$. Moreover, $\hat{f}$ will be the indicator function of the identity in $\widehat{G}$, so $\nu(\mathrm{supp}(\hat{f})) = 1$ as well.

However, when we restrict to $G = \mathbb{R}$ and $\mu$ the Lebesgue measure, we find that Theorem 4.3 is far from tight. Instead, the correct inequality is $\mu(\mathrm{supp}(f))\nu(\mathrm{supp}(\hat{f})) = \infty$, as proven by Benedicks [6] and strengthened by Amrein and Berthier [1]. The proofs of these results use the specific structure of $\mathbb{R}$, and we are not able to reprove them with our framework, presumably because our approach should work for any $G$, and Benedicks's result is simply false in general. There has been a long line of work on how much Theorem 4.3 can be strengthened for other locally compact abelian groups $G$; see [16, Section 7] for more.

## 4.2 $k$-Hadamard operators in infinite dimensions

Continuing to restrict to functions on $\mathbb{R}$, one can ask for other transforms which satisfy an uncertainty principle, just as previously we investigated all $k$-Hadamard matrices, and not just the Fourier transform matrices. From the proof of Theorem 4.1, and from the definition of $k$-Hadamard matrices, the following definition is natural.

**Definition 4.4.** We say that a linear operator $A : L^1(\mathbb{R}) \to L^\infty(\mathbb{R})$ is *$k$-Hadamard* if $\|A\|_{1\to\infty} \leq 1$ and if $\|A^*Af\|_\infty \geq k\|f\|_\infty$ for all functions $f$ with $f, Af \in L^1(\mathbb{R})$.

**Remark.** Extending our earlier use of the word, we will say that $f$ is *doubly $L^1$ for $A$* if $f, Af \in L^1(\mathbb{R})$. We will usually just say *doubly $L^1$* and omit "for $A$" when $A$ is clear from context.

The primary uncertainty principle for $k$-Hadamard operators, extending Theorem 4.1, is the following, whose proof is identical to that of Theorem 4.1.

**Theorem 4.5** (Primary uncertainty principle for $k$-Hadamard operators)**.** *Suppose $A$ is a $k$-Hadamard operator and $f$ is doubly $L^1$. Then*

$$\|f\|_1\|Af\|_1 \geq k\|f\|_\infty\|Af\|_\infty.$$

We can also extend the uncertainty principles for other norms seen in Theorem 3.21 to this infinite-dimensional setting, as follows.

**Theorem 4.6** (Norm uncertainty principle, infinitary version)**.** *Suppose $A$ is a $k$-Hadamard operator and $f$ is doubly $L^1$. Then for any $1 \leq q \leq \infty$,*

$$\|f\|_1\|Af\|_1 \geq k^{1-1/q}\|f\|_q\|Af\|_q.$$

*Proof.* The proof follows that of Theorem 3.21. We may assume that $q < \infty$, since the case of $q = \infty$ is precisely Theorem 4.5. It suffices to prove that for any non-zero function $g \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$,

$$\frac{\|g\|_1}{\|g\|_q} \geq \left(\frac{\|g\|_1}{\|g\|_\infty}\right)^{(q-1)/q}, \tag{7}$$

since we may then apply this bound to $f$ and $\hat{f}$ and use the primary uncertainty principle, Theorem 4.5. To prove (7), we simply compute

$$\|g\|_q^q = \int |g(x)|^q \,\mathrm{d}x \leq \|g\|_\infty^{q-1} \int |g(x)| \,\mathrm{d}x = \|g\|_\infty^{q-1}\|g\|_1,$$

which implies (7) after multiplying both sides by $\|g\|_1^{q-1}$ and rearranging. □

We already saw in the previous section that the Fourier transform on $\mathbb{R}$ is 1-Hadamard. As it turns out, the Fourier transform is one instance of a large class of $k$-Hadamard operators (with arbitrary values of $k$) known as *linear canonical transformations (LCT)*, which we define below. These transformations arise in the study of optics, and generalize many other integral transforms on $\mathbb{R}$, such as the fractional Fourier and Gauss–Weierstrass transformations. Although their analytic properties are somewhat more complicated than those of the Fourier transform, our framework treats them equally, since the only property we will need of them is that they are $k$-Hadamard. For more information on LCT, see [39, Chapters 9–10] or [19].

We now define the LCT, following [3]. This is a family of integral transforms, indexed by the elements of $\mathrm{SL}_2(\mathbb{R})$. Specifically, given a matrix $M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R})$ with $b \neq 0$, we can define the LCT $L_M$ associated to $M$ to be

$$(L_M f)(\xi) = \frac{e^{-i\pi \operatorname{sgn}(b)/4}}{\sqrt{|b|}} \int f(x) e^{i\pi(d\xi^2 - 2x\xi + ax^2)/b} \,\mathrm{d}x.$$

One can also take the limit $b \to 0$ and obtain a consistent definition of $L_M$ for all $M \in \mathrm{SL}_2(\mathbb{R})$. It turns out that this definition yields an infinite-dimensional representation of $\mathrm{SL}_2(\mathbb{R})$; in

particular, one sees that the inverse transform $L_M^{-1}$ is given by $L_{M^{-1}} = (L_M)^*$. From the definition, we see that if $b \neq 0$,

$$|(L_M f)(\xi)| = \frac{1}{\sqrt{|b|}} \left| \int f(x) e^{i\pi(d\xi^2 - 2x\xi + ax^2)/b} \, \mathrm{d}x \right| \leq \frac{1}{\sqrt{|b|}} \int |f(x)| \, \mathrm{d}x = \frac{\|f\|_1}{\sqrt{|b|}},$$

so $\|L_M\|_{1\to\infty} \leq 1/\sqrt{|b|}$. This implies the following result.

**Theorem 4.7.** *Let $M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R})$ be a matrix with $b \neq 0$. Let $A = \sqrt{|b|} L_M$ be a rescaling of the LCT $L_M$. Then $A$ is $|b|$-Hadamard.*

*Proof.* By the above, we see that $\|A\|_{1\to\infty} = \sqrt{|b|}\|L_M\|_{1\to\infty} \leq 1$. Similarly, if we set $B = A^* = \sqrt{|b|} L_{M^{-1}}$, then $\|B\|_{1\to\infty} \leq 1$ and $BAf = |b| L_{M^{-1}} L_M f = |b| f$ for any doubly $L^1$ function $f$. □

By combining the primary uncertainty principle for $k$-Hadamard operators with the argument of Theorem 4.3, we obtain the following generalization of the Matolcsi–Szűcs (or Donoho–Stark) uncertainty principle for the LCT, or indeed for any $k$-Hadamard operator.

**Corollary 4.8.** *If $M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R})$ and $f : \mathbb{R} \to \mathbb{C}$ is doubly $L^1$ and non-zero, then*

$$\lambda(\mathrm{supp}(f))\lambda(\mathrm{supp}(L_M f)) \geq |b|,$$

*where $\lambda$ denotes Lebesgue measure.*

*Proof.* From the primary uncertainty principle, Theorem 4.5, we find that

$$\frac{\|f\|_1}{\|f\|_\infty} \cdot \frac{\|L_M f\|_1}{\|L_M f\|_\infty} \geq |b|.$$

In proving Theorem 4.3, we saw that $\frac{\|g\|_1}{\|g\|_\infty} \leq \lambda(\mathrm{supp}(g))$ for all $g$, which yields the claim. □

We believe that this fact was not previously observed for the LCT. Of course, one expects that in general a much stronger result should hold, namely that $\lambda(\mathrm{supp}(f))\lambda(\mathrm{supp}(L_M f)) = \infty$ whenever $b \neq 0$; this would generalize the result of Benedicks [6] from the Fourier transform to all LCT. However, we are not able to obtain such a result with our approach, for the same reason that we cannot reprove Benedicks's theorem.

## 4.3 The Heisenberg uncertainty principle

In this section, we prove (with a somewhat worse constant) the well-known Heisenberg uncertainty principle, as well as some extensions of it. Again, as in all previous proofs we have seen, we use the elementary two-step process explained in the Introduction. Our proof differs drastically from the classical ones, which use analytic techniques (integration by parts) and special properties of the Fourier transform (that it turns differentiation into

multiplication by $x$). Indeed it is not clear if these classical techniques can be used to prove our generalizations.

For a doubly $L^1$ function $f$, we define the *variance* of $f$ to be

$$V(f) = \int x^2 |f(x)|^2 \, \mathrm{d}x.$$

If $\|f\|_2 = 1$, then we may think of $|f|^2$ as a probability distribution, in which case $V$ really does measure the variance of this distribution (assuming, without loss of generality[18], that its mean is 0). This interpretation is natural from the perspective of quantum mechanics (whence the original motivation for studying uncertainty principles): in quantum mechanics, we would think of $f$ as a wave function, and then $|f|^2$ would define the probability distribution for measuring some quantity associated to the wave function, such as a particle's[19] position or momentum. Heisenberg's[20] uncertainty principle asserts that $V(f)$ and $V(\hat{f})$ cannot both be small.

**Theorem 4.9** (Heisenberg's uncertainty principle [20, 24, 37])**.** *There exists a constant $C > 0$ such that for any doubly $L^1$ function $f \neq 0$,*

$$V(f)V(\hat{f}) \geq C\|f\|_2^2 \|\hat{f}\|_2^2.$$

**Remark.** It is in fact known that the optimal constant is $C = 1/(16\pi^2)$, with equality attained for Gaussians.

Additionally, versions of the Heisenberg uncertainty principle have been established for the LCT, see [35] for a survey. The most basic such extension is the following, stated without proof as [39, Exercise 9.10] and first proven in print by Stern [34].

**Theorem 4.10** (LCT uncertainty principle [34, 39])**.** *There exists a constant $C > 0$ such that the following holds for all doubly $L^1$ functions $f$. If $M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R})$ and $L_M$ is the associated LCT, then*

$$V(f)V(L_M f) \geq Cb^2 \|f\|_2^2 \|L_M f\|_2^2.$$

### 4.3.1 A Heisenberg uncertainty principle for other norms

We begin by proving the following generalization of Heisenberg's uncertainty principle. It lets us bound $V(f)V(Af)$ by any $L^q$ norm of $f$ and $Af$, for any $k$-Hadamard operator $A$ (recovering, for $q = 2$, the classical results of the previous subsection[21]). As far as we know,

[18] If its mean is at some point $a$, we can simply replace $f(x)$ by $f(x-a)$ to make $V$ be the actual variance of the distribution.

[19] Because of this interpretation, it is natural to have $f$ be a function defined on $\mathbb{R}^n$, to model a particle moving in $n$-dimensional space. For the moment we focus on the case $n = 1$, though we discuss the multidimensional analogue in Section 4.3.3.

[20] Though the physical justification for the uncertainty principle is due to Heisenberg [20], the proof of the mathematical fact is due to Kennard [24] and Weyl [37].

[21] Note e.g. that one recovers the correct dependence on $b$ when deducing the LCT uncertainty principle above from this one.

this result is new for $q \neq 2$, even for the Fourier transform. As we show below, the statements for different $q$ are in general of incomparable strength.

**Theorem 4.11** (Heisenberg uncertainty principle for arbitrary norms)**.** *For any $k$-Hadamard operator $A$, any doubly $L^1$ function $f$, and any $1 < q \leq \infty$,*

$$V(f)V(Af) \geq C_q k^{3-2/q} \|f\|_q^2 \|\hat{f}\|_q^2,$$

*where $C_q = 2^{-\frac{10q-8}{q-1}}$ depends only on $q$. In particular, $V(f)V(\hat{f}) \geq C_q \|f\|_q^2 \|\hat{f}\|_q^2$.*

**Remark.** No attempt was made to optimize the constant $C_q$. However, our proof is unlikely to give the optimal constant even after optimization; for instance, in the case $q = 2$, it is known that the optimal constant for the Fourier transform is $C_2 = 1/(16\pi^2) \approx 6.3 \times 10^{-3}$, whereas our proof gives the somewhat worse constant $2^{-12} \approx 2.4 \times 10^{-4}$.

As with our other proofs, the proof of this result proceeds in two stages. The first, already done in Theorem 4.6, is establishing the "norm uncertainty principle" $\|f\|_1 \|Af\|_1 \geq \|f\|_q \|Af\|_q$. After this, all that remains is to lower-bound $V(g)$ as a function of $\|g\|_1$ and $\|g\|_q$ for an arbitrary function $g$. Combining these two bounds will yield the result.

However, an important new ingredient which we did not use in the finite-dimensional setting is a different way to upper-bound $\|g\|_1$. The idea is to choose a constant $T$, depending on $g$ and the target norm $q$, so that most of the the $L^1$-mass of $g$ is outside the interval $[-T, T]$. This will allow us to lower bound the variance through a simple use of Hölder's inequality. Note that in the proof and subsequently, we use the usual conventions of manipulating $q$ as though it is finite, though everything works identically for $q = \infty$ by taking a limit, or by treating expressions like $\infty/(\infty - 1)$ as equal to 1.

*Proof of Theorem 4.11.* We may assume that $f \neq 0$. Following our general framework, we claim that the bound

$$\frac{\|g\|_1}{\|g\|_q} \leq \left(2^{\frac{5q-4}{q-1}} \frac{V(g)}{\|g\|_q^2}\right)^{\frac{q-1}{3q-2}} \tag{8}$$

holds for any non-zero function $g \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$. Observe that this bound is homogeneous, in that it is unchanged if we replace $g$ by $cg$ for some constant $c$. Once we have this bound, we can apply it to the norm uncertainty principle, Theorem 4.6, which says that

$$\frac{\|f\|_1}{\|f\|_q} \cdot \frac{\|Af\|_1}{\|Af\|_q} \geq k^{1-1/q}.$$

Plugging in (8) for $g = f$ and $g = Af$, we find that

$$\left(2^{\frac{10q-8}{q-1}} \frac{V(f)}{\|f\|_q^2} \frac{V(Af)}{\|Af\|_q^2}\right)^{\frac{q-1}{3q-2}} \geq k^{\frac{q-1}{q}},$$

and rearranging gives the desired conclusion. So it suffices to prove (8).

Let $T = \frac{1}{2}(\|g\|_1/(2\|g\|_q))^{q/(q-1)}$, so that $(2T)^{1-1/q}\|g\|_q = \frac{1}{2}\|g\|_1$. By Hölder's inequality, we have that

$$\int_{-T}^{T} |g(x)|\,\mathrm{d}x = \int \mathbf{1}_{[-T,T]}(x)|g(x)|\,\mathrm{d}x \leq \|\mathbf{1}_{[-T,T]}\|_{q/(q-1)}\|g\|_q = (2T)^{1-1/q}\|g\|_q = \frac{1}{2}\|g\|_1,$$

where the last step follows from the definition of $T$. This implies that the interval $[-T, T]$ contains at most half of the $L^1$ mass of $g$, so $\frac{1}{2}\|g\|_1 \leq \int_{|x|>T}|g(x)|\,\mathrm{d}x$. Applying the Cauchy–Schwarz inequality to this bound, we find that

$$\begin{aligned}
\frac{1}{2}\|g\|_1 &\leq \int_{|x|>T} |g(x)|\,\mathrm{d}x \\
&= \int_{|x|>T} \frac{1}{x}(x|g(x)|)\,\mathrm{d}x \\
&\leq \left(\int_{|x|>T} \frac{1}{x^2}\,\mathrm{d}x\right)^{1/2} \left(\int_{|x|>T} x^2|g(x)|^2\,\mathrm{d}x\right)^{1/2} \\
&\leq \left(\frac{2}{T}\right)^{1/2} V(g)^{1/2} \\
&= 2\left(\frac{2\|g\|_q}{\|g\|_1}\right)^{q/(2(q-1))} V(g)^{1/2}.
\end{aligned}$$

Rearranging this inequality yields (8). □

Theorem 4.11 contains within it infinitely many "Heisenberg-like" uncertainty principles, one for each $q \in (1, \infty]$ (and one for each $k$-Hadamard operator $A$). It is natural to wonder whether these are really all different, or whether one of them implies all the other ones. As a first step towards answering this question in the case of the Fourier transform, we can show that the $q = 2$ and $q = \infty$ cases are incomparable, in the sense that there exist functions for which one is arbitrarily stronger than the other. More precisely, we have the following result. We use $\mathcal{S}(\mathbb{R})$ to denote the Schwartz class of rapidly decaying smooth functions.

**Theorem 4.12.** *Define a function $F : \mathcal{S}(\mathbb{R}) \setminus \{0\} \to \mathbb{R}_{>0}$ by*

$$F(f) = \frac{\|f\|_\infty \|\hat{f}\|_\infty}{\|f\|_2\|\hat{f}\|_2} = \frac{\|f\|_\infty \|\hat{f}\|_\infty}{\|f\|_2^2}.$$

*Then the image of $F$ is all of $\mathbb{R}_{>0}$.*

We defer the proof of Theorem 4.12 to Appendix A.

Recall that the $q = 2$ case of Theorem 4.11 (which is just the classical Heisenberg uncertainty principle) says that $V(f)V(\hat{f}) \geq C\|f\|_2^2\|\hat{f}\|_2^2$, whereas the $q = \infty$ case of Theorem 4.11 says that $V(f)V(\hat{f}) \geq C'\|f\|_\infty^2\|\hat{f}\|_\infty^2$, for appropriate constants $C, C' > 0$. Thus, Theorem 4.12 says that these two results are in general incomparable: there exist functions for

which the $q = 2$ gives an arbitrarily stronger lower bound on $V(f)V(\hat{f})$, while there exist other functions for which the $q = \infty$ case gives an arbitrarily stronger bound. In fact, we expect that in general, the uncertainty principles for any $q \neq 2$ should be incomparable to the Heisenberg uncertainty principle, namely the case where $q = 2$. We are unable to prove this, and therefore leave it as a conjecture.

**Conjecture 4.13.** *Let $2 \neq q \in (1, \infty]$, and define a function $F_q : \mathcal{S}(\mathbb{R}) \setminus \{0\} \to \mathbb{R}_{>0}$ by*

$$F_q(f) = \frac{\|f\|_q \|\hat{f}\|_q}{\|f\|_2 \|\hat{f}\|_2} = \frac{\|f\|_q \|\hat{f}\|_q}{\|f\|_2^2}.$$

*Then the image of $F_q$ is all of $\mathbb{R}_{>0}$.*

### 4.3.2 An uncertainty principle for higher moments

Theorem 4.11 is itself a special case of a much more general uncertainty principle, which we now state. Rather than proving uncertainty for the variance functional $V(f)$, it proves it for any moments greater than 1 of the distributions $|f|^2$ and $|Af|^2$. Namely, for any $1 < r < \infty$, let us define

$$M_r(f) = \int |x|^r |f(x)|^2 \, \mathrm{d}x,$$

which is precisely the $r$th moment of the distribution $|f|^2$ if $\|f\|_2 = 1$. Even when $\|f\|_2 \neq 1$, $M_r(f)$ is still a good measure of how "spread" $f$ is, in that it computes how much $L^2$ mass of $f$ is far from the origin, weighted according to $|x|^r$. Uncertainty principles for such functionals were studied by Cowling and Price [10], who established the $q = 2$ case of the following result for the Fourier transform, as well as many more general results of this flavor for the Fourier transform. As with the basic Heisenberg uncertainty principle, we believe that the $q \neq 2$ case is new, as is the extension to arbitrary $k$-Hadamard operators.

**Theorem 4.14** (Heisenberg uncertainty principle for higher moments)**.** *Let $1 < r, s < \infty$ and $1 < q \leq \infty$. Then for any $k$-Hadamard operator $A$ and any doubly $L^1$ function $f$,*

$$M_r(f)^{\frac{q-1}{qr+q-2}} M_s(Af)^{\frac{q-1}{qs+q-2}} \geq C_{r,q} C_{s,q} k^{1-\frac{1}{q}} \|f\|_q^{\frac{2q-2}{qr+q-2}} \|Af\|_q^{\frac{2q-2}{qs+q-2}},$$

*for some constants $C_{r,q}, C_{s,q} > 0$ depending only on $r, q$, and $s, q$, respectively. In particular, if $s = r$, we have*

$$M_r(f) M_r(Af) \geq C'_{r,q} k^{\frac{qr+q-2}{q}} \|f\|_q^2 \|\hat{f}\|_q^2$$

*for another constant $C'_{r,q} = C_{r,q}^{2(qr+q-2)/(q-1)}$.*

We defer the proof of this theorem to Appendix A, but the basic idea is the same as the proof of Theorem 4.11: by the general framework, it suffices to upper-bound $\|f\|_1 / \|f\|_q$ as a function of $M_r(f)$. To do so, we again choose an appropriate $T$ so that most of the $L^1$ mass of $f$ is outside of $[-T, T]$, and then proceed by simple applications of the Hölder and Cauchy–Schwarz inequalities.

### 4.3.3 Further extensions and open questions

Cowling and Price [10] actually study a much more general question than what is in Theorem 4.14 (though restricting to $q = 2$ and the Fourier transform). They investigate integrals of the form $\int w(x)|f(x)|^p$ for values of $p$ other than 2 and for quite general functions $w$, and finding necessary and sufficient conditions for an uncertainty principle to hold for such functionals of $f$ and $\hat{f}$. Using the same techniques, we can also obtain such results in the case $w(x) = |x|^r$, as above; the proof is identical to that of Theorem 4.14, except that instead of using Cauchy–Schwarz to write

$$\int_{|x|>T} |f(x)|\,\mathrm{d}x = \int_{|x|>T} \frac{1}{|x|^{r/2}}(|x|^{r/2}|f(x)|)\,\mathrm{d}x \leq \left(\int_{|x|>T} \frac{\mathrm{d}x}{|x|^r}\right)^{1/2} \left(\int |x|^r |f(x)|^2\,\mathrm{d}x\right)^{1/2},$$

we would instead use Hölder's inequality to say

$$\int_{|x|>T} |f(x)|\,\mathrm{d}x = \int_{|x|>T} \frac{1}{|x|^{r/p}}(|x|^{r/p}|f(x)|)\,\mathrm{d}x \leq \left(\int_{|x|>T} \frac{\mathrm{d}x}{|x|^{\frac{r}{p-1}}}\right)^{\frac{p-1}{p}} \left(\int |x|^r |f(x)|^p\right)^{\frac{1}{p}}.$$

Then as long as $r > p - 1$, this first integral will converge, and the argument would go through as before. We omit the details, since they are very similar to (but messier than) the computations in the proof of Theorem 4.14.

However, an interesting point is raised by this argument, which is the fact that it only works for $r > p-1$. Cowling and Price's theorem works for all $r > (p-2)/2$, which is a larger range, and they in fact prove a converse which says that no such result is true if $r$ is any smaller. It is not at present clear to us whether there is a genuine obstruction that prevents our technique from working for all possible $r$, or if there is some variant manipulation that would yield the full strength of Cowling and Price's theorem.

A similar convergence issue arises when attempting to prove the multidimensional version of Heisenberg's uncertainty principle with our framework. The multidimensional version says the following.

**Theorem 4.15.** *Let $n \geq 1$ be an integer and $f \in L^2(\mathbb{R}^n)$. Then*

$$\left(\int \|x\|_2^2 |f(x)|^2\,\mathrm{d}x\right)\left(\int \|\xi\|_2^2 |\hat{f}(\xi)|^2\,\mathrm{d}\xi\right) \geq Cn^2 \|f\|_2^2 \|\hat{f}\|_2^2, \tag{9}$$

*where the constant $C > 0$ does not depend on the dimension $n$.*

If one attempts to prove this by using the argument from the proof of Theorem 4.11, the natural thing to try is to pick $T$ appropriately and then to write

$$\frac{1}{2}\|f\|_1 \leq \int_{\|x\|_2>T} |f(x)|\,\mathrm{d}x \leq \left(\int_{\|x\|_2>T} \frac{1}{\|x\|_2^2}\,\mathrm{d}x\right)^{1/2} \left(\int_{\|x\|_2>T} \|x\|_2^2 |f(x)|^2\,\mathrm{d}x\right)^{1/2}.$$

However, once $n > 1$, the first integral is infinite for any $T$, causing this proof to break down. The issue is again a convergence issue; in fact, one can make this proof go through

by integrating $1/\|x\|_2^r$ for any $r > n$, which means that one can prove a multidimensional uncertainty principle for $M_r(f)$ for any $r > n$. However, we still do not know how to prove the ordinary Heisenberg uncertainty principle in any dimension greater than 1 using our framework, and it would be very interesting to understand if these convergence issues represent a real limitation of our approach, or if there is a way around them. We leave this tantalizing question as an open problem.

**Open problem.** Can one prove the multidimensional Heisenberg uncertainty principle, Theorem 4.15, using our framework? For instance, can one prove that if $g \in L^1(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$, then

$$\frac{\|g\|_1}{\|g\|_\infty} \leq \left( Cn \frac{V(g)}{\|g\|_\infty^2} \right)^c,$$

where $V(g) = \int_{\mathbb{R}^n} \|x\|_2^2 |g(x)|^2 \, \mathrm{d}x$ is the $n$-dimensional variance of $g$, and $C, c > 0$ are constants independent of $n$? Alternately, can one prove such an inequality with $\|g\|_\infty$ replaced by $\|g\|_q$? In all these questions, the main interest is to obtain the correct dependence on the dimension $n$.

**Acknowledgments** We would like to thank Zhengwei Liu and Assaf Naor for helpful discussions, and Tomasz Kosciuszko for correcting an error in an earlier draft of this paper. We would also like to thank Greg Kuperberg for meticulous reading and many insightful comments on an earlier draft, as well as for permission to include his Theorem 3.14 in this paper.

# References

[1] W. O. Amrein and A. M. Berthier, On support properties of $L^p$-functions and their Fourier transforms, *J. Functional Analysis* **24** (1977), 258–267.

[2] T. Banica, *Complex Hadamard matrices and applications*, 2019. `https://banica.u-cergy.fr/pdf/ham.pdf`.

[3] M. J. Bastiaans and T. Alieva, The linear canonical transformation: definition and properties, in J. J. Healy, M. A. Kutay, H. M. Ozaktas, and J. T. Sheridan (eds.), *Linear canonical transforms*, *Springer Ser. Optical Sci.*, vol. 198, Springer, New York, 2016, pp. 29–80.

[4] W. Beckner, Inequalities in Fourier analysis, *Ann. of Math. (2)* **102** (1975), 159–182.

[5] W. Beckner, Pitt's inequality and the uncertainty principle, *Proc. Amer. Math. Soc.* **123** (1995), 1897–1905.

[6] M. Benedicks, On Fourier transforms of functions supported on sets of finite Lebesgue measure, *J. Math. Anal. Appl.* **106** (1985), 180–183.

[7] I. Białynicki-Birula and J. Mycielski, Uncertainty relations for information entropy in wave mechanics, *Comm. Math. Phys.* **44** (1975), 129–132.

[8] T. Carroll, X. Massaneda, and J. Ortega-Cerda, An enhanced uncertainty principle for the Vaserstein distance, 2020. Preprint available at arXiv:2003.03165.

[9] K. Conrad, Characters of finite abelian groups, accessed 2020. `https://kconrad.math.uconn.edu/blurbs/grouptheory/charthy.pdf`.

[10] M. G. Cowling and J. F. Price, Bandwidth versus time concentration: the Heisenberg-Pauli-Weyl inequality, *SIAM J. Math. Anal.* **15** (1984), 151–165.

[11] R. Craigen and H. Kharaghani, Orthogonal designs, in C. J. Colbourn and J. H. Dinitz (eds.), *Handbook of combinatorial designs*, Discrete Mathematics and its Applications (Boca Raton), second ed., Chapman & Hall/CRC, Boca Raton, FL, 2007, pp. 280–295.

[12] A. Dembo, T. M. Cover, and J. A. Thomas, Information-theoretic inequalities, *IEEE Trans. Inform. Theory* **37** (1991), 1501–1518.

[13] D. L. Donoho and X. Huo, Uncertainty principles and ideal atomic decomposition, *IEEE Trans. Inform. Theory* **47** (2001), 2845–2862.

[14] D. L. Donoho and P. B. Stark, Uncertainty principles and signal recovery, *SIAM J. Appl. Math.* **49** (1989), 906–931.

[15] W. Erb, Shapes of uncertainty in spectral graph theory, 2019. Preprint available at arXiv:1909.10865.

[16] G. B. Folland and A. Sitaram, The uncertainty principle: a mathematical survey, *J. Fourier Anal. Appl.* **3** (1997), 207–238.

[17] F. Gonçalves, D. Oliveira e Silva, and J. P. G. Ramos, New sign uncertainty principles, 2020. Preprint available at arXiv:2003.10771.

[18] G. H. Hardy, A theorem concerning Fourier transforms, *J. London Math. Soc.* **8** (1933), 227–231.

[19] J. J. Healy, M. A. Kutay, H. M. Ozaktas, and J. T. Sheridan (eds.), *Linear canonical transforms: theory and applications*, *Springer Series in Optical Sciences*, vol. 198, Springer, New York, 2016.

[20] W. Heisenberg, Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik, *Z. Physik* **43** (1927), 172–198.

[21] I. I. Hirschman, Jr., A note on entropy, *Amer. J. Math.* **79** (1957), 152–156.

[22] A. Jaffe, C. Jiang, Z. Liu, Y. Ren, and J. Wu, Quantum fourier analysis, *Proceedings of the National Academy of Sciences* **117** (2020), 10715–10720.

[23] T. Jiang, Maxima of entries of Haar distributed matrices, *Probab. Theory Related Fields* **131** (2005), 121–144.

[24] E. H. Kennard, Zur Quantenmechanik einfacher bewegungstypen, *Z. Physik* **44** (1927), 326–352.

[25] T. Matolcsi and J. Szűcs, Intersection des mesures spectrales conjuguées, *C. R. Acad. Sci. Paris Sér. A-B* **277** (1973), A841–A843.

[26] R. Meshulam, An uncertainty inequality for groups of order $pq$, *European J. Combin.* **13** (1992), 401–407.

[27] M. Northington V, Uncertainty principles for Fourier multipliers, 2019. Preprint available at arXiv:1907.08812.

[28] R. E. A. C. Paley, On orthogonal matrices, *J. Math. and Phys.* **12** (1933), 311–320.

[29] A. Poria, Uncertainty principles for the Fourier and the short-time Fourier transforms, 2020. Preprint available at arXiv:2004.04184.

[30] S. Quader, A. Russell, and R. Sundaram, Small-support uncertainty principles on $\mathbb{Z}/p$ over finite fields, 2019. Preprint available at arXiv:1906.05179.

[31] M. Ram Murty, Some remarks on the discrete uncertainty principle, in *Highly composite: papers in number theory*, *Ramanujan Math. Soc. Lect. Notes Ser.*, vol. 23, Ramanujan Math. Soc., Mysore, 2016, pp. 77–85.

[32] M. Reed and B. Simon, *Methods of modern mathematical physics. II. Fourier analysis, self-adjointness*, Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1975.

[33] J.-P. Serre, *Linear representations of finite groups*, Graduate Texts in Mathematics, Vol. 42, Springer-Verlag, New York-Heidelberg, 1977. Trans. L. L. Scott.

[34] A. Stern, Uncertainty principles in linear canonical transform domains and some of their implications in optics, *J. Opt. Soc. Amer. A* **25** (2008), 647–652.

[35] R. Tao and J. Zhao, Uncertainty principles and the linear canonical transform, in J. J. Healy, M. A. Kutay, H. M. Ozaktas, and J. T. Sheridan (eds.), *Linear canonical transforms*, *Springer Ser. Optical Sci.*, vol. 198, Springer, New York, 2016, pp. 97–111.

[36] T. Tao, An uncertainty principle for cyclic groups of prime order, *Math. Res. Lett.* **12** (2005), 121–127.

[37] H. Weyl, *Gruppentheorie und Quantenmechanik*, S. Hirzel, Leipzig, 1928.

[38] D. N. Williams, New mathematical proof of the uncertainty relation, *Amer. J. Phys.* **47** (1979), 606–607.

[39] K. B. Wolf, *Integral transforms in science and engineering*, *Mathematical Concepts and Methods in Science and Engineering*, vol. 11, Plenum Press, New York-London, 1979.

# A Proofs of some technical results

In this section, we present the proofs of Lemma 3.12, Proposition 3.19, Theorem 3.20, Theorem 4.12, and Theorem 4.14, which were omitted from the main text.

## A.1 Proof of Lemma 3.12

*Proof of Lemma 3.12.*

(i) Recall that the operators $A, B : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$ are defined by $A \circ M = FMF^*$ and $B \circ N = F^*NF$, and that we showed from Proposition 3.4 that $F^*F = FF^* = nI$. This implies that

$$B \circ (A \circ M) = F^*(FMF^*)F = (nI)M(nI) = n^2M.$$

(ii) Recall that $V$ consists of all matrices $T_f$ for $f \in \mathbb{C}[G]$. To prove this norm bound, it suffices to prove it for the extreme points of the $L^1$ ball. So we may assume that $f = \delta_x$ is the function that takes value 1 at some $x \in G$ and value 0 elsewhere. In that case, $T_f$ is a permutation matrix, and therefore every entry of $FT_{\delta_x}F^*$ is the inner product of two columns of $F$. By Proposition 3.4, all these inner products are either 0 or $n$, which implies that $\|A \circ T_{\delta_x}\|_\infty \leq n = \|T_{\delta_x}\|_1$.

(iii) Note that $B = A^*$, and recall that the $L^1$ and $L^\infty$ norms are dual to one another. This implies that $\|B\|_{1\to\infty} = \|A\|_{1\to\infty}$, and thus this is a direct consequence of (ii). □

## A.2 Proof of Proposition 3.19

*Proof of Proposition 3.19.* Assume this were not the case, and let $n$ be the smallest dimension in which a counterexample exists. Consider the set of counterexamples $v$ with $\|v\|_1 = 1$. Since the $L^1$ and $L^2$ norms of a vector are unchanged if we replace each entry by its absolute value and are unchanged if we permute the coordinates, we may restrict ourselves to counterexamples $v$ with non-negative real entries such that $v_1 \geq v_2 \geq \cdots \geq v_n$. Finally, observe that for any $s$, the set of vectors with $|\mathrm{supp}^1_{\varepsilon^2}(v)| = s$ is a closed set, and similarly for $|\mathrm{supp}^2_\varepsilon(v)|$. So the set of all counterexamples $v$ with $v_1 \geq v_2 \geq \cdots \geq v_n \geq 0$ and $\|v\|_1 = 1$ is a compact subset of $\mathbb{R}^n$, and we may therefore pick a counterexample $v$ of minimal $L^2$ norm. So from now on, let $v$ be a counterexample with $v_1 \geq \cdots \geq v_n \geq 0$, $\|v\|_1 = 1$, $n$ chosen to be minimal, and $\|v\|_2$ minimal among all such counterexamples. Let $s_1 = |\mathrm{supp}^1_{\varepsilon^2}(v)|$ and $s_2 = |\mathrm{supp}^2_\varepsilon(v)|$, so we assume for contradiction that $s_1 < s_2$. We split $v$ into three sub-vectors,

$$v_L = (v_1, v_2, \ldots, v_{s_1}) \qquad v_M = (v_{s_1+1}, \ldots, v_{s_2-1}) \qquad v_R = (v_{s_2}, \ldots, v_n),$$

with the subscripts indicating *Left, Middle,* and *Right*. Note that $v_M$ may be the empty vector if $s_2 = s_1 + 1$. Let the lengths of these vectors be $\ell = s_1, m = s_2 - s_1 - 1$, and $r = n - s_2 + 1$. We know that $v_L$ contains at least $1 - \varepsilon^2$ of the $L^1$ mass of $v$, meaning that

$$\|v_L\|_1 \geq (1-\varepsilon^2)\|v\|_1 = (1-\varepsilon^2)(\|v_L\|_1 + (\|v_M\|_1 + \|v_R\|_1)),$$

which implies that

$$\|v_L\|_1 \geq \frac{1-\varepsilon^2}{\varepsilon^2}(\|v_M\|_1 + \|v_R\|_1). \tag{10}$$

Similarly, since $v_L$ and $v_M$ together contain $s_2 - 1$ coordinates of $v$, they must collectively have *less than* $1 - \varepsilon$ of the $L^2$ mass of $v$, meaning that

$$\|v_L\|_2^2 + \|v_M\|_2^2 < (1-\varepsilon^2)\|v\|_2^2 = (1-\varepsilon^2)((\|v_L\|_2^2 + \|v_M\|_2^2) + \|v_R\|_2^2),$$

which implies

$$\|v_L\|_2^2 + \|v_M\|_2^2 < \frac{1-\varepsilon^2}{\varepsilon^2}\|v_R\|_2^2. \tag{11}$$

Now, we claim that $v_L$ and $v_M$ are both constant vectors. Indeed, suppose not, and let $w$ be the vector gotten by replacing the first $\ell$ entries of $v$ by the average value of $v_1, \ldots, v_\ell$, and replacing the next $m$ entries by the average value of $v_{s_1+1}, \ldots, v_{s_2-1}$. Then $\|w\|_2 < \|v\|_2$, since the $L^2$-norm is strictly convex, but $\|w\|_1 = \|v\|_1$. Therefore, inequalities (10) and (11) both hold for $w$, meaning that $w$ is a new counterexample with smaller $L^2$ norm, which we assumed did not exist. Thus, $v_L$ and $v_M$ are both constant vectors. In other words, we've found that there exist constants $a \geq b \geq 0$ such that $v_L = (a, a, \ldots, a), v_M = (b, b \ldots, b)$, and every entry of $v_R$ is at most $b$. In that case, inequalities (10) and (11) become

$$\ell a \geq \frac{1-\varepsilon^2}{\varepsilon^2}(mb + \|v_R\|_1) \tag{12}$$

$$\ell a^2 + mb^2 < \frac{1-\varepsilon^2}{\varepsilon^2}\|v_R\|_2^2. \tag{13}$$

Multiplying (12) by $a$ and using the fact that $m, b \geq 0$, we find that

$$\ell a^2 \geq \frac{1-\varepsilon^2}{\varepsilon^2}a\|v_R\|_1 \geq \frac{1-\varepsilon^2}{\varepsilon^2}\|v_R\|_2^2,$$

where the last step uses the fact that every entry of $v_R$ is at most $a$. However, (13) implies that

$$\ell a^2 < \frac{1-\varepsilon^2}{\varepsilon^2}\|v_R\|_2^2,$$

a contradiction. □

## A.3 Proof of Theorem 3.20

*Proof of Theorem 3.20.* Let $U = \frac{1}{\sqrt{k}}A$ be a rescaling of $A$, chosen so that $U$ is unitary, meaning that $\|Uw\|_2 = \|w\|_2$ for all $w \in \mathbb{C}^n$. Note that since the definition of $\mathrm{supp}^2$ is invariant under rescaling, we have that $|\mathrm{supp}^2_\eta(Av)| = |\mathrm{supp}^2_\eta(Uv)|$.

Let $S = \mathrm{supp}^2_\varepsilon(v)$ and $T = \mathrm{supp}^2_\eta(Uv)$. Let $P_S, P_T$ denote the orthogonal projections onto the coordinates indexed by $S, T$, respectively, and let $M = P_S U^* P_T$ be the submatrix of $U^*$ with rows indexed by $S$ and columns by $T$. Our goal is to obtain upper and lower bounds on $\|M\|_{2\to 2}$, which we will combine to conclude the desired result. To begin with the upper bound, we observe that for any vector $w$,

$$\|Mw\|_2^2 = \sum_{i=1}^n |(Mw)_i|^2 = \sum_{i\in S} |\langle M_i, w\rangle|^2 \le \sum_{i\in S} \|M_i\|_2^2 \|w\|_2^2 \le \frac{|S||T|}{k}\|w\|_2^2,$$

where $M_i$ denotes the $i$th row of $M$. The first inequality is Cauchy–Schwarz, while the second uses the fact that every entry of $M$ has absolute value at most $1/\sqrt{k}$, and that there are $|S||T|$ entries in $M$. This shows that $\|M\|_{2\to 2} \le \sqrt{|S||T|/k}$.

For the lower bound, we first observe that the unitarity of $U$ implies that

$$\|v - U^*P_TUv\|_2 = \|U^*Uv - U^*P_TUv\|_2 = \|U^*(Uv - P_TUv)\|_2 = \|Uv - P_TUv\|_2 \le \eta\|v\|_2,$$

where the last step follows from the definition of $T$. Since $P_S$ is a projection, it is a contraction in $L^2$, so

$$\|P_Sv - MUv\|_2 = \|P_S(v - U^*P_TUv)\|_2 \le \|v - U^*P_TUv\|_2 \le \eta\|v\|_2.$$

Moreover, by the definition of $S$, we know that $\|v - P_Sv\|_2 \le \varepsilon\|v\|_2$. Therefore,

$$\|v - MUv\|_2 \le \|P_Sv - MUv\|_2 + \|v - P_Sv\|_2 \le (\eta + \varepsilon)\|v\|_2.$$

Using the inequality $\|a - b\|_2 \ge \|a\|_2 - \|b\|_2$, we conclude that

$$\|MUv\|_2 \ge (1 - \varepsilon - \eta)\|v\|_2 = (1 - \varepsilon - \eta)\|Uv\|_2.$$

Combining this with our bound $\|M\|_{2\to 2} \le \sqrt{|S||T|/k}$ gives the desired result. □

## A.4 Proof of Theorem 4.12

*Proof of Theorem 4.12.* Let $a > b > 0$ be real numbers. Define

$$f_{a,b}(x) = e^{-\pi((a+bi)x)^2} = e^{-\pi(a^2-b^2)x^2}e^{-2\pi iabx^2}.$$

From the definition, we see that $|f_{a,b}(x)| = e^{-\pi(a^2-b^2)x^2}$. Since we assumed that $a > b > 0$, we have that $a^2 - b^2 > 0$, and therefore $|f_{a,b}|$ decays superexponentially at infinity, and we see that $f_{a,b} \in \mathcal{S}(\mathbb{R})$. We can compute

$$\|f_{a,b}\|_2^2 = \int |f_{a,b}(x)|^2 \,\mathrm{d}x = \int e^{-2\pi(a^2-b^2)x^2} = \frac{1}{\sqrt{2(a^2-b^2)}}.$$

Additionally, for any fixed $a > b > 0$, we see that $\pi(a^2 - b^2)x^2$ is minimized at $x = 0$, implying that $|f_{a,b}(x)|$ is maximized at $x = 0$, and therefore

$$\|f_{a,b}\|_\infty = |f_{a,b}(0)| = 1.$$

We can also compute the Fourier transform of $f_{a,b}$ explicitly, by recalling that the Fourier transform of $e^{-\pi(cx)^2}$ is $\frac{1}{c}e^{-\pi(\xi/c)^2}$, and that this holds for all $c \in \mathbb{C}$ for which $e^{-\pi(cx)^2} \in L^2$. Setting $c = a + bi$, we find that

$$\widehat{f_{a,b}}(\xi) = \frac{1}{a+bi}e^{-\pi\left(\frac{\xi}{a+bi}\right)^2} = \frac{1}{a+bi}e^{-\pi\xi^2(a^2-b^2)/(a^2+b^2)^2}e^{-2\pi i\xi^2 ab/(a^2+b^2)^2}.$$

In particular,

$$\left|\widehat{f_{a,b}}(\xi)\right| = \frac{1}{\sqrt{a^2+b^2}}e^{-\pi\xi^2(a^2-b^2)/(a^2+b^2)^2}.$$

Again for fixed $a > b > 0$, we have that $|\widehat{f_{a,b}}(\xi)|$ is maximized at $\xi = 0$, and we conclude that

$$\|\widehat{f_{a,b}}\|_\infty = \frac{1}{\sqrt{a^2+b^2}}.$$

This implies that

$$F(f_{a,b}) = \frac{\|f_{a,b}\|_\infty \|\widehat{f_{a,b}}\|_\infty}{\|f_{a,b}\|_2^2} = \sqrt{\frac{2(a^2-b^2)}{a^2+b^2}}.$$

Finally, let us set $b = \sqrt{a^2-1}$, and insist that $a > 1$ so that $b > 0$. Then we get that

$$F(f_{a,\sqrt{a^2-1}}) = \sqrt{\frac{2}{2a^2-1}},$$

and as $a$ ranges over $(1, \infty)$, the function $\sqrt{2/(2a^2-1)}$ ranges over $(0, \sqrt{2})$. So we conclude that $(0, \sqrt{2})$ is in the image of $F$.

Next, we wish to construct a family of functions whose images under $F$ cover the remaining interval $[\sqrt{2}, \infty)$. For a real number $c > 0$, we define

$$g_c(x) = \frac{1}{\sqrt{c}}e^{-\pi(x/c)^2} + \sqrt{c}e^{-\pi(cx)^2}.$$

From the property mentioned above about the Fourier transform of a Gaussian, we see that $g_c$ is its own Fourier transform for all $c$. We can compute

$$\|\widehat{g_c}\|_\infty = \|g_c\|_\infty = |g_c(0)| = \sqrt{c} + \frac{1}{\sqrt{c}}.$$

Moreover, we can also compute

$$\begin{aligned}
\|g_c\|_2^2 &= \int_{-\infty}^{\infty} \left( \frac{1}{\sqrt{c}} e^{-\pi(x/c)^2} + \sqrt{c} e^{-\pi(cx)^2} \right)^2 \mathrm{d}x \\
&= \frac{1}{c} \int_{-\infty}^{\infty} e^{-2\pi(x/c)^2} \,\mathrm{d}x + 2 \int_{-\infty}^{\infty} e^{-\pi x^2 (c^2 + 1/c^2)} \,\mathrm{d}x + c \int_{-\infty}^{\infty} e^{-2\pi(cx)^2} \,\mathrm{d}x \\
&= \frac{1}{c} \left( \frac{c}{\sqrt{2}} \right) + 2 \left( \frac{1}{\sqrt{c^2 + \frac{1}{c^2}}} \right) + c \left( \frac{1}{c\sqrt{2}} \right) \\
&= \sqrt{2} + \frac{2c}{\sqrt{c^4 + 1}}.
\end{aligned}$$

Thus, we find that

$$F(g_c) = \frac{\|g_c\|_\infty \|\widehat{g_c}\|_\infty}{\|g_c\|_2^2} = \frac{(\sqrt{c} + \frac{1}{\sqrt{c}})^2}{\sqrt{2} + \frac{2c}{\sqrt{c^4+1}}} = \frac{c + 2 + \frac{1}{c}}{\sqrt{2} + \frac{2c}{\sqrt{c^4+1}}}.$$

This function is minimized at $c = 1$, where its value is $\sqrt{2}$. Moreover, as $c \to \infty$, the denominator converges to $\sqrt{2}$, whereas the numerator grows like $c$. Thus, we see that $\lim_{c\to\infty} F(g_c) = \infty$. Since this is a continuous function of $c$, we conclude that $[\sqrt{2}, \infty)$ is in the image of $F$. Combining this with our result that $(0, \sqrt{2})$ is in the image of $F$, we conclude that the image of $F$ is all of $\mathbb{R}_{>0}$. □

## A.5 Proof of Theorem 4.14

*Proof of Theorem 4.14.* We may assume that $f \neq 0$. We mimic the proof of Theorem 4.11. We first claim that for any non-zero $g \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$,

$$\frac{\|g\|_1}{\|g\|_q} \leq \frac{1}{C_{r,q}} \left( \frac{M_r(g)}{\|g\|_q^2} \right)^{\frac{q-1}{qr+q-2}}, \tag{14}$$

where

$$C_{r,q} = (r-1)^{\frac{q-1}{qr+q-2}} 2^{-\frac{2qr+q-r-2}{qr+q-2}}$$

Once we have this, we can apply it with the norm uncertainty inequality, Theorem 4.6, to obtain that

$$\frac{1}{C_{r,q}C_{s,q}} \left( \frac{M_r(f)}{\|f\|_q^2} \right)^{\frac{q-1}{qr+q-2}} \left( \frac{M_s(Af)}{\|Af\|_q^2} \right)^{\frac{q-1}{qs+q-2}} \geq \frac{\|f\|_1}{\|f\|_q} \cdot \frac{\|Af\|_1}{\|Af\|_1} \geq k^{\frac{q-1}{q}},$$

which is the claimed result. So it suffices to prove (14).

As before, we set $T = \frac{1}{2}(\|g\|_1/(2\|g\|_q))^{q/(q-1)}$, so that $(2T)^{1-1/q}\|g\|_q = \frac{1}{2}\|g\|_1$. Hölder's inequality again gives that

$$\int_{-T}^{T} |g(x)| \,\mathrm{d}x \leq \frac{1}{2}\|g\|_1, \qquad \text{implying} \qquad \int_{|x|>T} |g(x)| \,\mathrm{d}x \geq \frac{1}{2}\|g\|_1.$$

Therefore, applying the Cauchy–Schwarz inequality, we can compute

$$\begin{aligned}\frac{1}{2}\|g\|_1 &\leq \int_{|x|>T} |g(x)|\,\mathrm{d}x \\ &= \int_{|x|>T} \frac{1}{|x|^{r/2}} \left(|x|^{r/2}|g(x)|\right)\,\mathrm{d}x \\ &\leq \left(\int_{|x|>T} \frac{1}{|x|^r}\,\mathrm{d}x\right)^{1/2} \left(\int_{|x|>T} |x|^r|g(x)|^2\,\mathrm{d}x\right)^{1/2} \\ &\leq \frac{\sqrt{2}}{\sqrt{(r-1)T^{r-1}}} M_r(g)^{1/2},\end{aligned}$$

where we use the assumption that $r > 1$ to evaluate the (convergent) integral. Rearranging, and using the definition of $T$, we obtain (14). □