Gleason's theorem

Helena Granström

August 31, 2006

Abstract

Gleason's theorem is a central result in mathematical physics. From it can be derived the standard method of calculating quantum probabilites, by taking the trace of the product between (the matrix representation of) the relevant projection operator and the so-called density matrix.
In this diploma work, a thorough presentation is first given of Gleason's original argument. We then proceed to look at a Gleason-type theorem for so-called POVMs, relaxing the assumptions of a proof published in 2003 and reaching the same result. Thereafter, a Gleason-type theorem is proved for two restricted classes of POVMs.
Kochen-Specker's theorem (KS), implied by the Gleason result, is then presented. The theorem (which is easily translated in terms of colourings) holds interesting implications for the formulation of so-called hidden variables theories.
Here, a particular (incomplete) KS colouring is explored in some depth. We investigate how effective the colouring will be in three and higher dimensions, using two different measures of this. Finally, we show that a restricted class of POVMs does not enforce the KS result.

Sammanfattning

Gleasons teorem är ett centralt resultat inom den matematiska fysiken. Från det följer det vanliga sättet att beräkna sannolikheter inom kvantmekaniken, genom att ta spåret av produkten mellan (matrisrepresentation av) en lämpligt vald projektionsoperator och den s.k. täthetsmatrisen.
Detta examensarbete inleds med en noggrann genomgång av Gleasons ursprungliga argument. Vi presenterar en sats av Gleason-typ för s.k. POVM:er som bevisades 2003, och visar att vi med svagare antaganden än de som gjordes i originalbeviset kan nå samma resultat. Därefter visar vi ett resultat analogt med Gleasons för två begränsade klasser av POVM:er.
Vi går sedan vidare med en presentation av Kochen-Speckers teorem (KS), som följer av Gleasons resultat. Denna sats (som enkelt kan översättas i termer av färgningar) har intressanta implikationer för teorier med s.k. dolda variabler.
Här utforskar vi närmare en särskild (ofullständig) KS-färgning. Vi undersöker hur effektiv denna färgning är i dimension tre och högre, med två olika mått på detta. Slutligen visar vi att en begränsad klass av POVM:er inte tvingar fram Kochen-Speckers resultat.

Acknowledgements

THANKS
Åsa, Hans, Mattias, Mathias, Andreas, Patrik, Emil for handling crises of the LaTex, integral and colouring variety. Those who should be thanked for handling all other types of crises know who they are. Thanks are also due to my supervisor Ingemar Bengtsson who, in addition to being a devoted and respectful teacher, falls into both of the above categories.

'Und es wurde fertig, das Leidenswerk. Es wurde vielleicht nicht gut, aber es wurde fertig. Und als es fertig war, siehe, da war es auch gut.'

— Thomas Mann

Contents

Chapter 1

Gleason explained

Gleason's theorem, formulated and proved by Andrew M. Gleason in 1957, is a statement about measures on Hilbert spaces of dimension at least three. The theorem states that the only possible probability measures on such spaces are measures μ of the form $\mu(a) = \mathrm{Tr}(\rho P_a)$, where ρ is a positive semi-definite self-adjoint operator of unit trace, and where P_a is a projection operator for projection onto the subspace a.
Postulating that any orthogonal basis in some Hilbert space corresponds to a measurement and that quantum systems can be represented by such spaces, we can understand the projection operators as representing yes-no observables a[1], commuting projectors corresponding to yes-no questions that can be simultaneously answered (or asked). Any (measurable) property a of the system is then uniquely associated with a subspace (which could be one-dimensional i.e. a vector) of the system's Hilbert space - within this framework, Gleason's statement is one about the probability of obtaining a given outcome when making a measurement on a quantum system. The theorem is of profound importance to modern physics due to its strong implications for how probabilities can be introduced into quantum mechanics. Put another way, it is a statement about the validity and uniqueness of the quantum probability rule.

Gleason's original proof [1] has over the years acquired a reputation of being impenetrable and hard to grasp. While the proof is rather compactly written it is not at all, however, impossible to understand. In the following section we will go through Gleason's argument step by step, trying to make explicit the implicit statements made in the original. The steps will be more or less identical to those made by Gleason himself, but some things will be expanded upon, and the structure of the proof will hopefully be made more visible.

One aspect of the theorem, which we will return to in our discussion of the Kochen-Specker theorem in chapter 4, is that it contains a non-contextuality assumption; that is, the assumption that the value[2] assigned to a vector v is independent of which basis we consider this vector to be a part of. As Gleason's theorem shows, this is a strong assump-

[1] I.e. answers to questions like 'Is the spin up in the z direction?'

[2] In effect, probability.

tion, and it has turned out to be of non-trivial importance in relation to attempts to reduce the indeterminate-probabilistic nature of quantum mechanics by means of hidden variables theories.

1.1 The proof of Gleason's theorem

We will start out with some definitions. First, let us remind ourselves of the definition of a Hilbert space.

Definition 1.1.1 *Any real or complex vector*[3] *space H with an inner product $\langle \cdot \mid \cdot \rangle$ can be ascribed a norm $\mid \cdot \mid$ according to*

$$\mid x \mid = \sqrt{\langle x \mid x \rangle}.$$

If H is complete[4] *under this norm, H is called a Hilbert space.*

We will also give some more specific definitions.

Definition 1.1.2 *A frame function of weight W for a separable*[5] *Hilbert space $\mathscr{H}$ is a real-valued function defined on the unit sphere of $\mathscr{H}$ such that for any orthonormal basis $\{x_i\}$ of $\mathscr{H}$,*

$$\sum_i f(x_i) = W \tag{1.1}$$

Definition 1.1.3 *A frame function f is regular if and only if there exists a self-adjoint operator ρ defined on $\mathscr{H}$ such that*

$$f(x) = \langle x \mid \rho \mid x \rangle$$

for all unit vectors x.

On the space of operators, itself a Hilbert space, we will use the inner product

$$\langle A \mid B \rangle = \mathrm{Tr}(A^\dagger B) \tag{1.2}$$

which will also define the norm

$$\mid A \mid = \sqrt{\langle A \mid A \rangle} \tag{1.3}$$

[3] Throughout this section the words *point* and *vector* will be used interchangeably, so that the statement that two points are orthogonal amounts to saying that the respective position vectors specifying the two points are orthogonal.

[4] A metric space M is said to be complete if every Cauchy sequence of points in M has a limit that is also in M.

[5] Separable in this context means that some countable subset of the space is *dense* in it. This means that the space has some countable subset with which all its elements can be approached, in the sense of a mathematical limit. An example of this is how any real number can be approximated to arbitrarily high accuracy by rational numbers.

On the space of vectors, the inner product will be the regular scalar product, and bracket notation will be used. No difference in notation will be made between these two operations, since it should be clear from the context which is referred to.
The crux of Gleason's proof is the realization that proving the theorem for a Hilbert space of arbitrary dimension $n \geq 3$ is accomplished by proving it for every two-dimensional subspace of a three-dimensional Hilbert space. This insight is indeed non-trivial.

Proving Gleason's theorem in three or higher dimensions is in effect equivalent to showing that any non-negative frame function defined on a real or complex Hilbert space $\mathscr{H}$ of dimension at least three is regular. Gleason arrives at this conclusion by considering subspaces of $\mathscr{H}$ of dimension two, and embedding them in a three-dimensional subspace (which can be done, because $\dim\mathscr{H} \geq 3$.) Therafter he makes use of the fact that f is regular on any such two-dimensional subspace, and that this means that f is regular on the whole of its domain of definition. (Reaching these results, however, requires some work.)
Let's start out by investigating some properties of frame functions in $\mathscr{H}^3$, and two-dimensional subspaces thereof.

Lemma 1.1.1 *In a finite-dimensional real Hilbert space a frame function is regular if and only if it is the restriction to the unit sphere of a quadratic form.*

Proof: This is clear from the definition of regularity - the restriction to the unit sphere accounts for the fact that the definition of regularity only involves unit vectors x.□

Lemma 1.1.2 *A function f on the unit circle in $\mathbb{R}^2$, $f = \cos n\theta$, where θ is defined as the angle relative the positive x-axis, is a frame function if and only if $n = 0$ or $n \equiv 2$ (mod 4).*

Proof: Assume that f has weight W. Since f is a frame function and unit vectors in the directions θ and $\theta + \frac{\pi}{2}$ constitute an orthonormal basis in two dimensions, we must have that

$$\cos n\theta + \cos n(\theta + \frac{\pi}{2}) = \cos n\theta + \cos n\theta \cos n\frac{\pi}{2} - \sin n\theta \sin n\frac{\pi}{2} =$$

$$(1 + \cos n\frac{\pi}{2}) \cos n\theta - \sin n\frac{\pi}{2} \sin n\theta = W$$

for all θ, $0 \leq \theta < 2\pi$.
This is true if and only if either $n = 0$ or

$$1 + \cos n\frac{\pi}{2} = 0 \Leftrightarrow \cos n\frac{\pi}{2} = -1 \Leftrightarrow$$

$$n\frac{\pi}{2} = \pi + k2\pi \Leftrightarrow n = 2 + 4k \Leftrightarrow n \equiv 2 (\text{mod } 4).□$$

Theorem 1.1.1 *Every continuous frame function on the unit sphere in $\mathbb{R}^3$ is regular.*

Proof: Let $\mathscr{C}$ denote the space of continuous functions on the unit sphere S in $\mathbb{R}^3$ with norm given by the standard inner product on $\mathbb{R}^3$ as defined above. The rotation group G in $\mathbb{R}^3$ is represented by a group of linear operators acting on $\mathscr{C}$ if we define

$$U_\sigma h = h \circ \sigma^{-1}, \quad \sigma \in G, \quad h \in \mathscr{C} \tag{1.4}$$

Applying the rotation σ to h is of course the same as applying the inverse rotation σ^{-1} to the argument of h before h acts on it.

Let Q_l denote the space of spherical harmonics Y_{lm} of degree l. As we know, the spherical harmonics are solutions of the Laplace equation in $\mathbb{R}^3$. These Q_l are irreducible, rotationally invariant subspaces of $\mathscr{C}$ - in fact they are the only such subspaces. Let F be the subspace of $\mathscr{C}$ consisting of continuous frame functions on S. From the definition of a frame function we can deduce that F is a closed subspace of $\mathscr{C}$, a subspace which is invariant under rotations.

We know that every continuous function on the unit sphere can be expressed as a sum of spherical harmonics Y_{lm}, and we would like to find out what harmonics contribute to the sums for the elements of F. In order to do this, we note the following properties of the spherical harmonics. The θ and ϕ, or for cartesian coordinates $(\hat{n})_z$ and $(\hat{n})_x + i(\hat{n})_y$, dependencies can be separated, according to

$$Y_{lm}(\hat{n}) = Y_{lm}(\theta, \phi) = \sqrt{\frac{(2l+1)}{4\pi} \frac{(l-m)!}{(l+m)!}} P_l^m(\cos\theta) \mathrm{e}^{im\phi} \tag{1.5}$$

From this it follows how the spherical harmonics transform under reflection, parity and conjugation.

$$Y_{lm}(\hat{n}) = (-1)^{l+m} Y_{lm}(\pi - \theta, \phi) = (-1)^m Y_{lm}(\theta, \phi + \pi) = (-1)^l Y_{lm}(-\hat{n}) = (-1)^m Y^*_{l-m}(\hat{n}) \tag{1.6}$$

Q_0 is a space of constant functions, and since every constant function is a frame function, we see that $Q_0 \subset F$, so that in the expansion of a frame function f on the sphere, Y_{00} can contribute. Q_1 consists of linear functions on $\mathbb{R}^3$ restricted to S. These functions change sign under parity as can be seen from equation (1.6), which we know is not the case for frame functions; if $\{x_i\}$ is an orthonormal basis, so of course is $\{-x_i\}$, and hence we know them to sum to the same constant, with no change of sign. Hence, $Q_1 \not\subset F$ and no Y_{1m}s occur in the parametrization of f. Q_2, on the other hand, contains the restrictions to S of quadratic forms of zero trace[6]. Quadratic forms of zero trace on S are frame functions of weight 0 on S, meaning that $Q_2 \subset F$.

[6]The reader can convince herself that this is likely, considering that the space Q_2 has five linearly independent components, namely those of different m, with $m \in [-2, 2]$, which is just equal to the number of independent components of a symmetric traceless 3×3-matrix.

To check if any Q_l with $l > 2$ is a subset of F we proceed as follows. As noted above, the spherical harmonics are solutions to the Laplace equation in $\mathbb{R}^3$. In cylindrical coordinates, this equation reads

$$\frac{1}{r}\frac{\partial}{\partial r}(r\frac{\partial \phi}{\partial r}) + \frac{1}{r^2}\frac{\partial^2 \phi}{\partial \theta^2} + \frac{\partial^2 \phi}{\partial z^2} \tag{1.7}$$

By insertion, it's easily verified that $\phi_1 = r^l \cos l\theta$ and $\phi_2 = (r^2 - 2(l-1)z^2)r^{l-2}\cos{(l-2)\theta}$ are both solutions of this equation of order l, and as such are elements in Q_l and can be written as a linear combination of Y_{lm}s.

Under the assumption that $Q_l \subset F$ these functions would be frame functions not only on S, but also on the unit circle in the xy-plane by restriction. Consider a family of orthogonal triples of vectors $\{x_1, x_2, x_3\}$ with one vector, say x_1, fixed. Let f be a frame function defined on this set such that

$$f(x_1) + f(x_2) + f(x_3) = W \tag{1.8}$$

For all orthogonal pairs of vectors $\{x_2, x_3\}$ lying in the plane orthogonal to x_1 we will have that

$$f(x_2) + f(x_3) = W - f(x_1) \tag{1.9}$$

so f is a frame function also on the great circle orthogonal to the fix vector x_1.

The assumption that $l > 2$, however, then contradicts the statement of lemma 1.1.2, namely that a function of the form $\cos n\theta$ (or, as in this case, proportional to $\cos n\theta$) will be a frame function if and only if $n = 0$ or $n \equiv 2 \pmod 4$. Because this cannot be satisfied by both l and $l-2$ simultaneously (for $l > 2$) we conclude that $Q_l \not\subset F$ for $l > 2$. By this we can conclude that F is the closed linear span[7] of Q_0 and Q_2, which in this case is equivalent with saying that $F = Q_0 + Q_2$. This means that in the parametrization of a frame function f on the unit sphere, possible under the assumption that f is continuous, only Y_{00} and Y_{2m}, $m = -2, -1, 0, 1, 2$, can contribute.

As noted in lemma 1.1.1, a frame function is regular if and only if it is the restriction to the unit sphere of a quadratic form. That this is the case with the functions of Q_2 is clear. As for the constant functions of Q_0, they too can be considered as restrictions of quadratic forms, because on S the constant $1 = x^2 + y^2 + z^2$, which no doubt is a quadratic form. So, the fact that F, the space of continuous frame functions on S, is exactly the space $Q_0 + Q_2$ and thereby consists only of restrictions to the unit sphere of quadratic forms, amounts, by lemma 1.1.1, to the result that all elements of F are regular; that is, every continuous frame function on S is regular.□

[7]The closed linear span of a subset S of some Hilbert space $\mathscr{H}$ is the smallest closed linear subspace of $\mathscr{H}$ containing S.

The statement of theorem 1.1.1 is an important result, which we will be able to make use of in proving the main theorem, but we will need it in a slightly stronger form. The next step is to show that all non-negative frame functions on S are in fact continuous, hence regular by theorem 1.1.1. Armed with this result we will then take the step from real to complex two-dimensional Hilbert spaces, which will then show up in our later considerations as subspaces of a larger Hilbert space. So, we now proceed to show that every non-negative frame function on S is regular. In order two show this we will make use of three intermediate results related to how much f is allowed to vary over a small open disk, one of which requires some rather subtle geometric arguments.

We will also need the following definition.

Definition 1.1.4 *: If f is a real-valued function defined on the set X we denote by $osc(f, X)$ the number $sup\{f(x) \mid x \in X\}$ - $inf\{f(x) \mid x \in X\}$.*

A great circle is defined as a circle on the sphere of maximal diameter, or equivalently, a circle on the sphere the plane of which intersects the origin. Of course, every point of the sphere lies on infinitely many great circles. Here, given a point q, we will be interested in the particular great circle for which q is the point with the smallest value of θ or equivalently, the great circle which has a tangent that coincide with that of a circle in the plane $\theta = \theta_q$ at the point q, θ_q being the value of the polar angle θ at the point q. The latter circle will, given a point q, be denoted C_q. In the following, this specific great circle will be referred to as the Great Circle through q.

Lemma 1.1.3 *Suppose $p_3 \in N$, where N is the set of points n such that $\theta_n \leq \frac{\pi}{2}$, with $p_3 \neq (0, 0, 1)$. Consider the set P_1 of all points $p_1 \in N$ such that for some point p_2*
(a) p_2 is on the Great Circle through p_1,
(b) p_3 is on the Great Circle through p_2.
Then the set P_1 has a non-empty interior[8].

Proof: The fact that the set P_1 is non-empty can be seen in figure 1.1. Consider two infinitesimally separated points p_{21} and p_{22} that are the respective highest points of two neighboring great circles in the continuum of great circles passing through a given point p_3. With some help from figure 1.1, we can realize that p_{21} will give rise to a set of points $\{p_{11}\}$ that are the highest points on great circles passing through p_{21}, a curve segment of this set is sketched in the figure. If we look at the corresponding set $\{p_{12}\}$ for p_{22} we realize that this set will be a curve lying infinitesimally close to the curve $\{p_{11}\}$. By continuously varying p_2 in this way, we can conclude that the set P_1 indeed has a non-empty interior.
It is, however, possible to find out more about this set P_1 by deriving an analytic expression for the set of points p_2 satisfying (b). Let the point p_3 have coordinates $(\sin\theta, 0, \cos\theta)$[9] in some orthonormal coordinate system (x, y, z), and let $p_2 = (a, b, c), \mid l \mid= a^2+b^2+c^2 = 1$.

[8] The interior of a set K is the union of all open sets that are subsets of K.
[9] As opposed to in Gleason's original, our θ is the usual polar angle of spherical polar coordinates defined as the angle relative the positive z-axis.

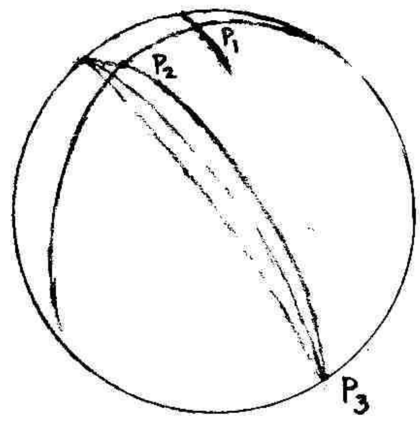

Figure 1.1: Some of the great circles that pass through p_3, the highest point p_2 of one of them and a great circle with highest point p_1 passing through p_2. In the figure is also sketched a curve segment consisting of points that are the highest points on some great circle through p_2 - p_1 of course lies on this curve.

The tangent to the great circle connecting two points p_2 and p_3 at the point p_2 is a vector in the plane in which p_2 and p_3 lie, and which is orthogonal to p_2. Such a vector, call it t_c, can be created by the following maneuver:

$$t_c = p_3 - \frac{(p_3 \cdot p_2)p_2}{p_2 \cdot p_2},$$

that is, just starting from p_3 and eliminating its component along p_2.

We want to find the points p_2 such that this tangent coincides with the tangent of the circle C_{p_2} of points on the sphere with $\theta = \theta_{p_2}$ at the point p_2, call it t_2. This tangent we know to lie in the xy-plane and to be orthogonal to p_2, which gives (normalization aside) $t_2 = (-b, a, 0)$.

The condition that t_c and t_2 be parallell at the point p_2 means that $\alpha t_2 = t_c$ for some number α, giving the three equations

$$-\alpha b = \sin\theta - a^2 \sin\theta - ac\cos\theta \qquad (1.10)$$

$$\alpha a = -ab\sin\theta - bc\cos\theta \qquad (1.11)$$

$$0 = \cos\theta - ac\sin\theta - c^2\cos\theta = \cos\theta - ac\sin\theta - (1-(a^2+b^2))\cos\theta \tag{1.12}$$

$$\Leftrightarrow 0 = (aN2 + bN2)\cos\theta - ac\sin\theta \tag{1.13}$$

However, we need only use the last of these equations in order to specify the points p_2 that satisfy (b). The reason for this is that the only way in which the tangent of a great circle through p_2 can lie entirely in the xy-plane is by also being a tangent to the surface of the sphere in the xy-plane, and thereby parallell to the tangent of C_{p_2}. So, the set P_2 of points $p_2 = (a, b, c)$ satisfying (b) are given by

$$\Psi = (a^2+b^2)\cos\theta - ac\sin\theta = 0 \tag{1.14}$$

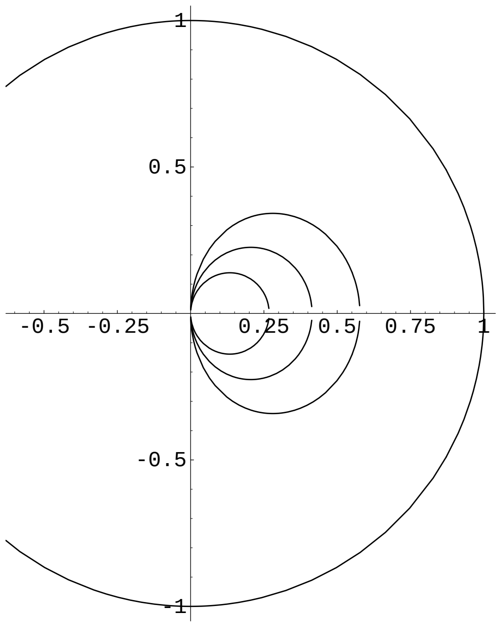

Figure 1.2: Stereographic visualization of the set P_2 for some values of p_3. Each point on the P_2 curves will in turn give rise to a new curve according to the same pattern - the union of these curves will be the set P_1.

Lemma 1.1.4 *Suppose that f is a frame function on the unit sphere S in $\mathbb{R}^3$ and that for a certain neighborhood*[10] *U of a point p on S, $osc(f, U) = \alpha$. Then every point on the great circle the plane of which is orthogonal to p has a neighborhood V for which $osc(f, V) \leq 2\alpha$.*

Proof: We choose coordinates so that $p = (0, 0, 1)$. Let us define the neighborhood U of p as all points with $\theta < \vartheta$. Let q_0 be any point on the sphere for which $\theta = \frac{\pi}{2}$ (all such points are orthogonal to p) and let r be the point with the same ϕ-coordinate as q_0, and

[10]The neighborhood of a point z_0 is defined as the set of points z satisfying $\mid z - z_0 \mid < \rho$ for some ρ, which is an open disk with radius ρ centered at z_0.

with $\theta_r = \frac{\pi}{2} + \frac{\vartheta}{2}$. Let C_0 be the unique great circle that connects r and q_0 (this circle will pass through the point p as well, because r has the same value of ϕ as q_0) and let r' and q_0' be points in $N \cap C_0$ such that $r \perp r'$ and $q_0 \perp q_0'$. For both these points to lie in U, we need to have $\frac{\vartheta}{2} \leq \vartheta$ which is satisfied for all $\vartheta \geq 0$, so both r' and q_0' lie in U. This will be true also if we substitute for q_0 a point q in some neighborhood V of q_0, holding r fixed.

So, let q_1 and q_2 be two points in V and, repeating the above procedure, let C_i be the great circle connecting r and q_i and choose points r_i' and q_i' in $N \cap C_i$ such that $r \perp r_i'$ and $q_i \perp q_i'$ $(i = 1, 2)$. Then (because all the vectors specifying points on the unit sphere have modulus one) $\{r, r_i'\}$ and $\{q_i, q_i'\}$ both form orthonormal bases for C_i $(i = 1, 2)$. Therefore, if we apply the frame function f to both these sets of vectors, using the frame function condition we get

$$f(r) + f(r_i') = f(q_i) + f(q_i'), \quad i = 1, 2 \tag{1.15}$$

Subtracting the equations obtained for $i = 1$ and $i = 2$ and taking the modulus we get

$$| f(q_1) - f(q_2) | = | f(r_1') - f(r_2') + f(q_2') - f(q_1') | \leq | f(r_1') - f(r_2') | + | f(q_2') - f(q_1') | \leq 2\alpha \tag{1.16}$$

where we in the last step made use of the fact that r_1', r_2', q_1', $q_2' \in U$. So, for any pair of points q_1 and q_2 both in V, $| f(q_1) - f(q_2) | \leq 2\alpha$, which shows that $\text{osc}(f, V) \leq 2\alpha$.□

The result of lemma 1.1.4 can be generalized quite straightforwardly - this is done in lemma 1.1.5.

Lemma 1.1.5 *Suppose that f is a frame function on the unit sphere S in $\mathbb{R}^3$ and that for a certain non-empty open set U, osc$(f, U) = \alpha$. Then every point of S has a neighborhood W such that osc$(f, W) \leq 4\alpha$.*

Proof: To realize this we need only apply lemma 1.1.4 twice - using the fact that we can, from an arbitrary point in the sphere, reach any given point in two steps of arc length $\frac{\pi}{2}$. Assuming that $\text{osc}(f, U) = \alpha$ where U is the neighborhood of some point p, there exists a point q orthogonal to p which has a neighborhood V for which $\text{osc}(f, V) \leq 2\alpha$. And, by the same reasoning there is a neighborhood W of a point w orthogonal to q for which $\text{osc}(f, W) \leq 4\alpha$. Since any point on S can be reached in this way, we are done.□

Next, we will show that every non-negative frame function on S in $\mathbb{R}^3$ is continuous, which together with theorem 1.1.1 implies that every non-negative frame function on S in $\mathbb{R}^3$ is regular. As the two preceeding lemmas indicate, we will go about this by showing that a limited oscillation in the neighborhood of some point on the sphere leads to such a limitation for all points. Showing that this oscillation is arbitrarily small amounts, of course, to showing continuity.

Theorem 1.1.2 *Every non-negative frame function on S in $\mathbb{R}^3$ is regular.*

Proof: Let f be a non-negative frame function on S with weight W. Subtracting a constant from f only changes its weight, but the result is still a frame function, as can easily be verified. Thus we can, with no loss of generality, assume that $\inf f(x) = 0$ for $x \in S$.
Now, let ϵ be some small positive number and set $\eta = \frac{\epsilon}{88}$. This choice of η is of course made for later convenience, however, η will also be a small positive number, provided that ϵ is. If $\inf f = 0$ on S we can, by the definition of infimum, always find a point $p \in S$ such that $f(p) \leq \eta$.
Let σ denote the transformation $\sigma x(\theta_x, \phi_x) = x(\theta_x, \phi_x + \frac{\pi}{2})$, and define the function

$$g(x) = f(x) + f(\sigma x).$$

If $\{x_1, x_2, x_3\}$ is an orthogonal triple, then so is $\{\sigma x_1, \sigma x_2, \sigma x_3\}$, so that

$$\sum_{i=1}^{3} g(x) = \sum_{i=1}^{3} f(x) + \sum_{i=1}^{3} f(\sigma x) = W + W = 2W$$

so g is a frame function with weight $2W$.
For q on the equator, $q \perp \sigma q$, and p, q and σq are all mutually orthogonal. This means that

$$g(q) = f(q) + f(\sigma q) = f(q) + f(\sigma q) + f(p) - f(p) = W - f(p),$$

and since this holds for any point q on the equator we can deduce that g is constant on the equator.

Now, consider a point $r \in N - \{p\}$. Let C be the Great Circle through r. Because r is the highest point on C, the point at which C intersects the equator will be orthogonal to r, call it q. This orthogonality means that $g(r) + g(q) \leq 2W$ and since we know that $g(r) + g(q) = g(r) + W - f(p)$ from above, we see that $g(x) \leq W + f(p) \leq W + \eta$ $\forall$ $x \in N - \{p\}$ where the last inequality follows from how p was chosen.

Consider also a point $s \in C \cap N$ and a point $t \in C \cap N$ orthogonal to s. Because s and t span the plane of C, as do r and q, the third orthogonal vector, call it w, is the same for these two pairs, and we have $g(r) + g(q) + g(w) = g(s) + g(t) + g(w)$, so that

$$g(r) + g(q) = g(r) + W - f(p) = g(s) + g(t) \leq g(s) + W + \eta \tag{1.17}$$

$$\Rightarrow g(r) \leq g(s) + 2\eta \tag{1.18}$$

$\forall$ $r \in N - \{p\}$ and $s \in C \cap N$, where we have again used that $f(p) \leq \eta$.

Let $\beta = \inf\{g(x) \mid x \in N - \{p\}\}$ and choose a point $z \in N - \{p\}$ for which $g(z) \leq \beta + \eta$. By definition of infimum, such a point can always be found. Now we consider a vector x satisfying the conditions of lemma 1.1.3, namely that $x \in N - \{p\}$ is such that for some y

(a) y is on the Great Circle through x
(b) z is on the Great Circle through y
with z as defined above.

According to lemma 1.1.3 the set of points x satisfying these conditions is non-empty (in fact has a non-empty interior), so such an x can always be found.
Then, by equation (1.18), we have that

$$g(x) \leq g(y) + 2\eta \tag{1.19}$$

and

$$g(y) \leq g(z) + 2\eta \tag{1.20}$$

so that

$$\beta \leq g(x) \leq g(z) + 2\eta + 2\eta \leq \beta + 5\eta \tag{1.21}$$

This means that for U the non-empty interior of the set X of such points x, $\text{osc}(g, U) \leq 5\eta$, which by lemma 1.1.5 gives that there exists a neighborhood V of any point on the sphere, hence of p, with $\text{osc}(g, V) \leq 20\eta$. But because $p = \sigma p$ we have that

$$g(p) = 2f(p) \leq 2\eta$$

$$\Rightarrow \sup\{g(x) \mid x \in V\} \leq 22\eta \tag{1.22}$$

Applying lemma 1.1.5 once again, we can infer that every point u on the sphere has a neighborhood W such that

$$\text{osc}(f(u), W) \leq 88\eta = \epsilon \tag{1.23}$$

$\forall\ u \in S$. Because ϵ can be arbitrarily small this proves the continuity of f on S, and from this regularity follows, according to theorem 1.1.1.□

This concludes the second part of Gleasons proof, devoted to proving the continuity of f on S. Continuing, we will need the following

Definition 1.1.5 *We will say that a real-linear subspace $\mathscr{K}$ of a Hilbert space $\mathscr{H}$ is completely real if the inner product takes only real values on $\mathscr{K} \times \mathscr{K}$.*

Because we have proved the regularity of f on S, which can be considered as a completely real subspace of, for instance, a complex two-dimensional Hilbert space, we are interested in finding out if regularity on such a space has any implications for the regularity of f on a larger Hilbert space, of which the smaller space can be considered a subspace. The next section of the proof is devoted to this investigation.

We can start off by noting that a completely real subspace of any Hilbert space $\mathscr{H}$ is itself a Hilbert space under restriction of the inner product on $\mathscr{H}$. So, a frame function for $\mathscr{H}$ becomes a frame function when restricted to a completely real subspace of $\mathscr{H}$.

Lemma 1.1.6 *If f is a non-negative frame function of weight W on a real Hilbert space, then for any unit vectors x_1 and x_2*

$$| f(x_1) - f(x_2) | \leq 2W | x_1 - x_2 |$$

Proof: That f is regular means that there exists a symmetric self-adjoint operator ρ such that $f(x) = \langle x \mid \rho \mid x \rangle$. Because f is non-negative we have $\langle x \mid \rho \mid x \rangle \geq 0$ and because of the frame function condition, $\langle x \mid \rho \mid x \rangle \leq W$, so we have that

$$0 \leq \langle x \mid \rho \mid x \rangle \leq W \tag{1.24}$$

for all unit vectors x, and $| \rho | \leq W$.
Because we are in a real Hilbert space, $\langle x_1 \mid \rho \mid x_2 \rangle = \langle x_2 \mid \rho \mid x_1 \rangle$, so

$$\langle x_1 + x_2 \mid \rho \mid x_1 - x_2 \rangle = \langle x_1 \mid \rho \mid x_1 \rangle - \langle x_2 \mid \rho \mid x_2 \rangle = f(x_1) - f(x_2) \tag{1.25}$$

Therefore, we have that

$$| f(x_1) - f(x_2) | \leq | \rho || x_1 + x_2 || x_1 - x_2 | = | \rho | \sqrt{x_1 \cdot x_1 + 2x_1 \cdot x_2 + x_2 \cdot x_2} \, | x_1 - x_2 | \tag{1.26}$$

whence

$$| f(x_1) - f(x_2) | \leq 2W | x_1 - x_2 | \tag{1.27}$$

for unit vectors x_1 and x_2.□

We will now show that the regularity of f on every completely real subspace of a two-dimensional complex Hilbert space implies regularity on the whole space, a result which will prove very useful in our coming considerations.

Lemma 1.1.7 *Suppose that f is a non-negative frame function on a two-dimensional complex Hilbert space $\mathscr{H}^2$ which is regular on every completely real subspace. Then f is regular.*

Proof: Let W be the weight of f and let $M = \sup\{f(x) \mid x \in \mathscr{H}^2\}$. We can choose a sequence of unit vectors $\{x_n\}$ so that $\forall\, \epsilon > 0\ \exists\, N$ such that $| f(x_n) - M | \leq \epsilon\ \forall\, n \geq N$.
Because every compact metric space E, and a Hilbert space is such space, is known to have the property that every infinite sequence in E has a limit point in E, we can also assume that $\forall\, \delta > 0\ \exists\, N$ such that $| x_n - y | \leq \delta\ \forall\, n \geq N$, $y \in \mathscr{H}^2$. We now construct a vector $\lambda_n = \frac{\langle y | x_n \rangle}{| \langle y | x_n \rangle |}$, chosen to have unit norm, and to be such that $\langle \lambda_n x_n \mid y \rangle$ is real. With this choice of λ_n we see that $\lambda_n \to 1$, $\lambda_n x_n \to y$.
For any number λ with $| \lambda | = 1$ we have

$$f(\lambda x) = f(x) \tag{1.28}$$

for all unit vectors x and all frame functions f. This is because if f is a frame function for some Hilbert space $\mathscr{H}$ it is a frame function for any closed subspace S of $\mathscr{H}$ by restriction. If we consider a one-dimensional S, the frame function condition immediately gives equation (1.28). Since $\mid \lambda_n \mid = 1$, $f(\lambda_n x_n) = f(x_n)$. Also,

$$\langle \lambda_n x_n \mid y \rangle = \frac{\langle y \mid x_n \rangle \langle x_n \mid y \rangle}{\mid \langle y \mid x_n \rangle \mid} = \frac{\langle y \mid x_n \rangle \langle y \mid x_n \rangle^*}{\mid \langle y \mid x_n \rangle \mid} \in \mathbb{R} \tag{1.29}$$

so $\lambda_n x_n$ and y span a completely real subspace as defined above. Because a closed completely real subspace of a Hilbert space $\mathscr{H}$ is itself a real Hilbert space with respect to the restriction of the inner product on $\mathscr{H}$, lemma 1.1.6 applies to give (using (1.28))

$$\begin{aligned} \mid f(y) - M \mid = \mid (f(y) - f(\lambda_n x_n)) + (f(x_n) - M) \mid \leq \mid f(y) - f(\lambda_n x_n) \mid + \mid f(x_n) - M \mid \leq \\ \leq 2W \mid y - \lambda_n x_n \mid + \mid f(x_n) - M \mid \Rightarrow \\ \Rightarrow f(y) = M \end{aligned} \tag{1.30}$$

Let us now define a function F on $\mathscr{H}$ by

$$F(v) = \mid v \mid^2 f(\frac{v}{\mid v \mid}) \tag{1.31}$$

if $v \neq 0$,

$$F(0) = 0 \tag{1.32}$$

The assumption that $f(x)$ is regular, hence can be written as $\langle x \mid \rho \mid x \rangle$, on every completely real subspace, implies that F is a quadratic form under restriction to any completely real subspace. Moreover, because of (1.28) we have that $F(\lambda v) = \mid \lambda \mid^2 F(v)$ for all scalars λ and vectors v.

Let z be a unit vector orthogonal to y, and consider the completely real subspace spanned by y and z. Since F restricts to be a frame function on this subspace, we have $F(y) = f(y) = M$ and $F(z) = f(z) = W - f(y) = W - M$. On the yz-subspace, F is by the regularity assumption a quadratic form which by choice of M as $\sup f$ obtains its maximum value on the unit circle at y. Hence, the matrix for F relative to the basis (y, z) is diagonal. This can be made likely by letting $y = (1, 0)$ and $z = (0, 1)$. If F were not diagonal in the (y, z)-basis we would obtain a larger value of F than that at y at some point $\tilde{y}$ obtained by adding a component along the z-direction to y.
So,

$$F(\alpha y + \beta z) = \alpha^2 F(y) + \beta^2 F(z) = \alpha^2 M + \beta^2 (W - M), \quad \alpha, \beta \in \mathbb{R} \tag{1.33}$$

With the following maneuver we are, again using (1.28), able to take the step from real to complex Hilbert space. Let λ and μ be non-zero complex numbers, and construct the vector

$$z' = \frac{\mu}{\mid \mu \mid} \frac{\mid \lambda \mid}{\lambda} z$$

It's easy to check that z' will also be a unit vector orthogonal to y. Thus, using (1.28) and (1.33)

$$F(\lambda y + \mu z) = F((\frac{\mid \lambda \mid}{\lambda})(\lambda y + \mu z)) = F(\mid \lambda \mid y + \mid \mu \mid z') = M \mid \lambda \mid^2 + (W - M) \mid \mu \mid^2 \tag{1.34}$$

Because every vector in $\mathscr{H}$ can be expressed as a complex linear combination of y and z we have shown that

$$F(x) = \langle x \mid \rho \mid x \rangle \tag{1.35}$$

for all $x \in \mathscr{H}$, where ρ is the self-adjoint operator with the matrix representation

$$\begin{pmatrix} M & 0 \\ 0 & W - M \end{pmatrix}$$

relative the basis (y, z). This completes the proof that f is regular on all of $\mathscr{H}$.□

Above, the regularity was extended from every completely real two-dimensional subspace to all two-dimensional Hilbert spaces. The next step serves to generalize regularity on every two-dimensional subspace to regularity on any Hilbert space $\mathscr{H}$ for which regularity on any two-dimensional subspace holds.

Lemma 1.1.8 *Suppose that f is a non-negative frame function for a Hilbert space $\mathscr{H}$ (that can be either real or complex) and suppose that f is regular when restricted to any two-dimensional subspace of $\mathscr{H}$. Then f is regular.*

Proof: We repeat the definition of F from the previous lemma:

$$F(v) = \mid v \mid^2 f(\frac{v}{\mid v \mid}) \tag{1.36}$$

if $v \neq 0$,

$$F(0) = 0 \tag{1.37}$$

Because f is regular on any two-dimensional subspace S of $\mathscr{H}$ by assumption, there exists a bilinear or Hermitian[11] form A_S such that

$$F(x) = A_S(x, x) \propto \langle x \mid \rho \mid x \rangle \tag{1.38}$$

for all $x \in S$. Now, we can define a form A on all of $\mathscr{H} \times \mathscr{H}$ by

$$A(x, y) = A_S(x, y) \tag{1.39}$$

if $x \in S$, $y \in S$. A is defined on all of $\mathscr{H} \times \mathscr{H}$ because any pair of vectors x and y can of course be considered as elements in the subspace that they span. This subspace is two-dimensional in all cases except for when x and y are parallell, so that $y = \lambda x$ for some λ. Then, on the other hand, the bilinearity of A_S gives that

$$A_S(x, y) = A_S(x, \lambda x) = \lambda F(x) \tag{1.40}$$

irrespective of how S is chosen.
Using the polarization identity we define $A_S(x, y)$ by

$$A_S(x, y) = \frac{1}{4}(A_S(x + y, x + y) - A_S(x - y, x - y)) \tag{1.41}$$

[11]Corresponding to the real and complex case, respectively.

when $\mathscr{H}$ is real and

$$A_S(x,y) = \frac{1}{4}(A_S(x+y,x+y) - A_S(x-y,x-y) + i(A_S(x+iy,x+iy) - A_S(x-iy,x-iy)) \tag{1.42}$$

when $\mathscr{H}$ is complex. Using that A_S is linear in its first argument and conjugate linear in its second (as was used in deriving (1.40)) it's easily verified that these expressions give the desired result in the respective cases.
From (1.39), (1.41) and (1.42) we can deduce that

$$\begin{aligned} A(\alpha x, y) &= \alpha^* A(x,y) \\ A(x,y) &= A(y,x)^* \\ 4ReA(x,y) &= F(x+y) - F(x-y) \\ 2F(x) + 2F(y) &= F(x+y) + F(x-y) \end{aligned}$$

Using these relations, we have that

$$8ReA(x,z) + 8ReA(y,z) = 2F(x+z) - 2F(x-z) + 2F(y+z) - 2F(y-z) =$$

$$= F(x+z+y+z) - F(x-z+y-z) + F(x-y) - F(x-y) = F(x+y+2z) - F(x+y-2z) =$$

$$4ReA(x+y,2z) = 8ReA(x+y,z)$$

so that

$$ReA(x,z) + ReA(y,z) = ReA(x+y,z) \tag{1.43}$$

Letting $x \to ix$ and $y \to iy$, we find

$$ImA(x,z) + ImA(y,z) = Im(A, x+y, z) \tag{1.44}$$

$$(1.43) + (1.44) \Rightarrow A(x,z) + A(y,z) = A(x+y,z) \tag{1.45}$$

which with the first two properties of A shows that A is bilinear or Hermitian on all of $\mathscr{H} \times \mathscr{H}$.

The next step towards the proof of full regularity is showing that A is bounded, which is done as follows.

Let x and y be vectors such that $| \; x \; | \leq 1$, $| \; y \; | \leq 1$. We can choose $\omega \in \mathbb{C}$, with $| \; \omega \; | = 1$, so that $A(\omega x, y)$ is real, which gives

$$4 \mid A(x,y) \mid = 4A(\omega x, y) = 4ReA(\omega x, y) = F(\omega x + y) - F(\omega x - y) \leq$$

$$\leq M(\mid \omega x + y \mid^2 + \mid \omega x - y \mid^2) = 2M(\mid \omega x \mid^2 + \mid y \mid^2) \leq 4M$$

so

$$\mid A(x,y) \mid \leq M \tag{1.46}$$

and A is a bounded sesquilinear form. Thus, we can apply the Riesz representation theorem, which states the following.

Theorem 1.1.3 *Let $\mathscr{H}_1$, $\mathscr{H}_2$ be Hilbert spaces and let $h : \mathscr{H}_1 \times \mathscr{H}_2 \rightarrow k$ be a bounded sesquilinear form.Then h has a representation $h(x, y) = \langle x \mid S \mid y \rangle$ where $S : \mathscr{H}_1 \rightarrow \mathscr{H}_2$ is a bounded linear operator. S is uniquely determined by h and has a norm $\mid S \mid=\mid h \mid$.*

By this theorem, there exists a bounded self adjoint operator, let us call it ρ such that

$$A(x, y) = \langle x \mid \rho \mid y \rangle \tag{1.47}$$

for all $x, y \in H$. Since

$$f(x) = F(x) = A(x, x) = \langle x \mid \rho \mid x \rangle \tag{1.48}$$

for all unit vectors $x \in \mathscr{H}$, this concludes the proof.□

At this point, the results of theorem 1.1.2 and lemma 1.1.7 will be tied together in the strong statement that every non-negative frame function on a Hilbert space $\mathscr{H}^N$ with $N \geq 3$ is regular. The method that has been used throughout the proof, namely that of transferring properties of subspaces to the Hilbert space of which they are a part, will prove useful in this step as well.

Theorem 1.1.4 *Every non-negative frame function on a (real or complex) Hilbert space $\mathscr{H}$ of dimension at least three is regular.*

Proof: As has been noted earlier, a frame function for a Hilbert space $\mathscr{H}$ by restriction becomes a frame function (in general, of course, of different weight) for any completely real subspace of $\mathscr{H}$. Because we have assumed that the dim$\mathscr{H} \geq 3$, any such two-dimensional subspace can be embedded in a completely real three-dimensional subspace of $\mathscr{H}$, making possible the application of theorem 1.1.2 because this three-dimensional space will be isomorphic to $\mathbb{R}^3$. This shows that any non-negative frame function f is regular on any completely real two-dimensional subspace of $\mathscr{H}$. By lemma 1.1.7, if a frame function is regular on every completely real two-dimensional subspace it is in fact real on all two-dimensional subspaces; and by the statement of the last lemma, f is regular on all of $\mathscr{H}$.□

We are almost there - the crux of Gleason's theorem is contained in theorem 1.1.4. The main result, however, is the following.

Theorem 1.1.5 *Let p be a measure on the closed subspaces of a separable (real or complex) Hilbert space $\mathscr{H}$ with dim$\mathscr{H} \geq 3$. There exists a positive semi-definite self-adjoint operator ρ of the trace class*[12] *such that for all closed subspaces A of $\mathscr{H}$*

$$p(A) = Tr(\rho P_A) \tag{1.49}$$

[12]A trace class operator is a compact operator for which a trace may be defined, such that the trace is finite and independent of the choice of basis. When the dimension is finite any operator will be of the trace class, because we can always choose a matrix representation for it and define its trace as the sum of its diagonal elements - the concept of trace class becomes meaningful only when dealing with infinite dimensional spaces.

where P_A is the orthogonal projection of $\mathscr{H}$ onto A.
In particular, any assignment of probabilities to the vectors in $\mathscr{H}$ has to be of this form.

Proof: If B_x is the one-dimensional subspace spanned by the unit vector x, $f(x) = p(B_x)$ defines a non-negative frame function f. Since f is regular on all Hilbert spaces of dimension at least three, there is by definiton of regularity a self-adjoint operator ρ such that

$$f(x) = p(B_x) = \langle x \mid \rho \mid x \rangle$$

for all unit vectors x. The fact that $\langle x \mid \rho \mid x \rangle \geq 0$ for all unit vectors x shows that ρ has to be positive semi-definite. Given an orthonormal basis $\{x_i\}$ of $\mathscr{H}$, we have that

$$p(\mathscr{H}) = \sum_i p(B_{x_i}) = \sum_i \langle x_i \mid \rho \mid x_i \rangle \tag{1.50}$$

The sum on the far right converges; so we see that ρ is in the trace class with $\mathrm{Tr}\rho = p(\mathscr{H})$.

For any closed subspace A of $\mathscr{H}$ we can always expand an orthonormal basis $\{y_i\}$ for A to an orthonormal basis for $\mathscr{H}$ by adjoining to $\{y_i\}$ vectors $\{z_i\}$, so that $\{y_i, z_i\}$ is an orthonormal basis for $\mathscr{H}$.
Then, with P_A the projection operator for orthogonal projection onto the subspace A, $P_A y_i = y_i$ for all i, and of course $P_A z_j = 0$ for all j. Consequently,

$$p(A) = \sum_i p(B_{y_i}) = \sum_i \langle y_i \mid \rho \mid y_i \rangle = \sum_i \langle P_A y_i \mid \rho \mid y_i \rangle + \sum_i \langle P_A z_i \mid \rho \mid z_i \rangle = \mathrm{Tr}(\rho P_A)$$

and we have shown that

$$p(A) = \mathrm{Tr}(\rho P_A) \tag{1.51}$$

thereby deriving the standard quantum rule, and proving Gleason's theorem. □

It is of some importance to note that the last step is valid only for Hilbert spaces of dimension three and higher. Nowhere in this proof we have shown that a frame function has to be regular on any two-dimensional space - the statement, based primarily on theorem 1.1.2, is that every frame function is regular on the real unit sphere S, considered as a *sub*space of $\mathbb{R}^3$. From Gleason's result follows the uniqueness of the density matrix as the means of assigning probabilities to vectors in Hilbert spaces in a consistent way.

Chapter 2

A Gleason-type theorem for POVMs

The statement of Gleason's theorem, that any quantum state is given by a density operator, was originally formulated using frame functions defined on sets of orthogonal projective operators summing to one. The probability of outcome A when performing a measurement is then given by an inner product between the projector P_A corresponding to outcome A and the density matrix of the system, which fully specifies (our knowledge of) its state. The assumption made was

$$\sum_{P_j \in X} f(P_j) = 1 \tag{2.1}$$

$$X = \{P_j \in D_d \mid \sum_j P_j = 1, P_i P_j = \delta_{ij}\} \tag{2.2}$$

with D_d the set of projection operators in d dimensions.

By allowing ourselves to make somewhat stronger assumptions about the frame function than those originally made by Gleason, we can prove a Gleason-type theorem by much simpler means than those available to Gleason. The frame function assumption, originally made only for sets of orthogonal projectors, will here be made for the more general POVMs. A POVM (Positive Operator Valued Measure) is a resolution of the identity operator into positive operators, called effects, which just as the set considered by Gleason sum to one but with the orthogonality constraint relaxed. Formally, a POVM is a set of n positive operators E_i that act on an N dimensional Hilbert space $\mathscr{H}^N$, and that satisfy

$$\sum_{i=1}^{n} E_i = 1, \quad E_i^\dagger = E_i, \quad E_i \geq 0, \quad i = 0, 1, ..., n \tag{2.3}$$

Note in particular that the number of elements of a POVM need not equal the dimension of the Hilbert space. Both PVMs (Projective Valued Measures - in effect, ON bases; orthogonal resolutions of the identity) and POVMs can be said to represent measurements of some quantity, but the POVM is a realization of a more general notion of measurement. For a PVM, the results obtained are mutually exclusive (for instance, $m = j$ means $m \neq j - 1, ..., -j$). For a POVM, however, this is not the case. Also, while the projectors of a PVM always commute, no such assumption is made for the effects of a general POVM. Any POVM can (according to what is known as Naimark's theorem), however, always be described in terms of a projective measurement, if the latter is performed in a higher dimension [2]. Because of this, the concept of a POVM is naturally inherent in the formalism.

A Gleason-type derivation of the standard quantum probability rule has indeed proved possible using frame functions defined on POVMs, both general and restricted. For some configurations, the fact that the quantum rule is not valid, has also been shown. In a 2003 paper, Caves et al [3] prove a Gleason-type theorem, originally formulated and proved by Busch [4], and also investigate some specific types of POVMs, using the assumption

$$\sum_{E_j \in X} f(E_j) = 1 \tag{2.4}$$

$$X = \{E_j \in K_d \mid \sum_j E_j = 1\} \tag{2.5}$$

K_d being the set of effects in d dimensions.

It turns out that making the frame function assumption only for POVMs with two or three elements, rather than for all POVMs, still enforces the quantum probability rule. In the following, we will stay close to what is done in [3], noting specifically when extra considerations have to be made related to the number of elements in the POVM. The proof is divided into several steps; first linearity with respect to non-negative rationals is proved in section 2.1, thereafter continuity in section 2.2. These two results add up to linearity, which is related to an inner product in section 2.3.

The assumption we make is the following.

$$\sum_{j=1}^{n} f(E_j) = 1 \tag{2.6}$$

for

$$E_j \in K_d, \quad \sum_{j=1}^{n} E_j = 1, \quad n = 2, 3 \tag{2.7}$$

Note that we here have chosen to normalize the frame function f so that the weight is 1, rather than ascribing to it the weight W. This, however, is only a question of normalization and means no loss of generality.

Apart from the domain of definition of the frame function f, and the fact that this result is valid also for qubits (two-dimensional systems) the theorem is fully equivalent with Gleason's original; stating that the density matrix description of a system is the only one possible. The $N = 2$ case is of great importance in Gleasons original proof as well, although the two-dimensional spaces he considers are always assumed to be subspaces of some larger $\mathscr{H}$; in the end, his result is not valid for the qubit case. In using POVMs with three elements, corresponding to projective measurements in higher dimensions, the same assumption is in some sense implicit, which is arguably one of the reasons for the great simplification of the proof in the POVM case.

The result we set out to prove is the following.

Theorem 2.0.6 *For every frame function $f : K_d \to [0, 1]$, there is a unique unit-trace positive operator ρ such that $f(E) = (\rho, E) = Tr(\rho E)$, where K_d is the set of effects in d dimensions.*

The proof follows.

2.1 Linearity with respect to the non-negative rationals

The first step is to prove additivity. This step necessitates making the frame function assumption for both two- and three-element POVMs, since we consider the two POVMs $\{E_1, E_2, E_3\}$ and $\{E_1 + E_2, E_3\}$.

Using the frame function property we obtain $f(E_1) + f(E_2) + f(E_3) = f(E_1 + E_2) + f(E_3)$ and hence

$$f(E_1) + f(E_2) = f(E_1 + E_2).$$

From additivity homogenity linearity with respect to non-negative rationals follows. Consider the effect $\frac{n}{m}E$. Applying the frame function and making use of the additivity yields

$$mf(\frac{n}{m}E) = \underbrace{f(\frac{n}{m}E) + f(\frac{n}{m}E) + ... + f(\frac{n}{m}E)}_{\text{m times}} = f(m\frac{n}{m}E) = f(nE) = f(\underbrace{E + E + ... + E}_{\text{n times}})$$
$$= \underbrace{f(E) + f(E) + ... + f(E)}_{\text{n times}} = nf(E)$$
$$\Rightarrow f(\frac{n}{m}E) = \frac{n}{m}f(E) \tag{2.8}$$

This linearity with respect to non-negative rationals in combination with continuity of course implies full linearity, since the set of rationals is dense in the space of real numbers, meaning that any real number s can be given to arbitrary accuracy by two rational numbers, one smaller and one larger than s. Hence, the next step is to prove that the frame function is continuous.

2.2 Continuity

We will show that discontinuity of the frame function would lead to a contradiction with how f is defined. Continuity in metric spaces is defined as follows

Definition 2.2.1 *: f is continuous at x_0 if $\forall\ \epsilon > 0\ \exists\ \delta > 0$ such that $\mid f(x) - f(x_0) \mid < \epsilon$, $\forall\ x$ satisfying $\mid x - x_0 \mid < \delta$. The norm on the space of operators $\mid A \mid = \sqrt{(A, A)}$ is defined through the inner product $(A, B) = Tr(A^\dagger B)$.*

We begin by showing continuity at the zero operator. Additivity implies that $f(0) = 0$; $f(E) = f(E+0) = f(E) + f(0)$. Assume that f is discontinuous at the zero operator.This means that $\exists\ \epsilon > 0$ such that $\forall\ \ \delta > 0\ \exists$ some effect E such that $\mid E \mid < \delta$ and $f(E) \geq \epsilon$. Choose $\delta = \frac{1}{N} < \epsilon$, with $N \in \mathbb{Z}^+$, and let E be an effect satisfying $\mid E \mid < \frac{1}{N}$ and $f(E) \geq \epsilon$. Multiplying E by N gives $F = NE$, which is also an effect since $\mid F \mid = N \mid E \mid < 1$, implying that the sum of the squares of its eigenvalues is less than 1. However, the additive property (or the linearity with respect to non-negative rationals) of the frame function gives $f(F) = Nf(E) \geq N\epsilon > 1$. f, however, is a function to the closed interval $[0, 1]$ from the set of effects that are members of POVMs with two or three elements, and since for any arbitrary effect E_1 we can always find E_2 such that $E_1 + E_2 = 1$, this is a contradiction. From this we can conclude that f is indeed continuous at the zero operator.

To generalize this result to include any arbitrary effect E_0 we proceed as follows. Let E be an effect in the neighborhood of E_0, and consider the difference $\mid E - E_0 \mid$. This difference can be diagonalized and divided into non-negative and negative eigenvalue parts, so that $E - E_0 = A - B$, where A is the non-negative part, and -B is the part with negative eigenvalues.

We have that $\mid A \mid, \mid B \mid \quad \leq \quad \mid A - B \mid \quad = \quad \mid E - E_0 \mid$. This follows from

$$\mid A - B \mid^2 = \mathrm{Tr}((A - B)^\dagger (A - B))$$

$$= \mathrm{Tr}(A^\dagger A - A^\dagger B - B^\dagger A + B^\dagger B) = \mathrm{Tr}(A^\dagger A + B^\dagger B) \geq \mathrm{Tr}(A^\dagger A), \mathrm{Tr}(B^\dagger B) \tag{2.9}$$

From this, and from the fact that A and B are positive operators, we can conclude, provided $\mid E - E_0 \mid \leq 1$, that A and B are effects.

The frame function can be applied to the equation $E + B = E_0 + A$ (f is defined on these operators for the same reason as given above - they are positive, and can always be considered as being part of a two or three element POVM). Using additivity, this yields $f(E) - f(E_0) = f(A) - f(B)$.

Because of continuity at the zero operator as shown, we know that $\forall\ \epsilon = \frac{\epsilon'}{2} > 0\ \exists\ \delta > 0$ such that $\mid A \mid, \mid B \mid < \delta \Rightarrow f(A), f(B) < \epsilon'$.

This means that if $\mid E - E_0 \mid = \mid A - B \mid < \delta$ we have $\mid A \mid, \mid B \mid < \delta$ and $\mid f(E) - f(E_0) \mid =$ $= \mid f(A) - f(B) \mid \leq \mid f(A) \mid + \mid f(B) \mid < 2\epsilon' = \epsilon$.

This completes the proof that f is continuous on all of K_d, and taken together with the results of section 2.2 this establishes that f is a linear function on K_d.

2.3 Linearity and the inner product

Having proved that f is linear, we will make use of the fact that any linear function on a vector space can be expressed by means of an inner product on this space. In order to do so, we need to extend f to the entire vector space of operators, which is done as follows.
Let H be an arbitrary Hermitian operator. Such an operator can always be expressed as a difference between two positive operators, call them G_1 and G_2. The most obvious way to accomplish this is to diagonalize H and let G_1 and $-G_2$ be the positive- and negative-eigenvalue parts, respectively.
Any positive operator G can be expressed as $G = \alpha E$ for some positive real number α and some effect E.
Now, define $f(H) = f(G_1) - f(G_2) = \alpha_1 f(E_1) - \alpha_2 f(E_2)$, using the additivity and linearity of f. Although the decomposition of H is not unique the extension is, which can be proved as follows. Assume that $H = \alpha_1 E_1 - \alpha_2 E_2 = \alpha_3 E_3 - \alpha_4 E_4$. Choose β such that $\beta \geq max\{\alpha_i\}$ and divide both sides of the above equation by β to give

$$\frac{\alpha_1}{\beta} E_1 + \frac{\alpha_4}{\beta} E_4 = \frac{\alpha_2}{\beta} E_2 + \frac{\alpha_3}{\beta} E_3 \tag{2.10}$$

It is clear that these operators are all now in the original domain of f, meaning that the frame function can be applied. This gives

$$\alpha_1 f(E_1) + \alpha_4 f(E_4) = \alpha_2 f(E_2) + \alpha_3 f(E_3) \tag{2.11}$$

By this we have a linear function f on the whole space of Hermitian operators. To make the extension to the space of all operators we note that any operator C can be written (uniquely) as $C = A + iB$ for Hermitian operators A and B.

Getting back to the task of writing this linear function as an inner product on the vector space of operators, we choose an orthonormal basis of operators $\{\tau_j\}$, enabling us to expand any arbitrary operator A as the sum $A = \sum_j^d \tau_j(\tau_j, A)$, d being the dimension of the

operator space.
Applying the frame function, we get $f(A) = \sum_j^d f(\tau_j)(\tau_j, A)$.
We can now define the operator ρ as being the solution of the equations $f(\tau_j) = (\rho, \tau_j)$. This is d^2 equations for the d^2 components of ρ, so the solution is unique.
The requirements that the frame function is non-negative and normalized (that is, $\sum_i f(x_i) = 1$) guarantee that ρ has the density matrix properties positivity and unit trace. Positivity follows from the fact that for any normalized vector $\mid \psi\rangle$ we have, due to non-negativity of the frame function, $0 \leq f(\mid \psi\rangle\langle\psi \mid) = \langle\psi \mid \rho \mid \psi\rangle$. That ρ has unit trace is seen by expanding the unit matrix using the normalization of the frame function:

$$\mathrm{Tr}\rho = (\rho, 1) = (\rho, \sum_j E_j) = \sum_j (\rho, E_j) = \sum_j f(E_j) = 1.$$

This completes the proof. □

It should be noted that this last consideration is not affected by the weaker frame function assumption, and also that the dimension of the Hilbert space of operators does not enter into these calculations.

As noted above, the POVM version of Gleason's theorem is valid for two-dimensional Hilbert spaces, in contrast to the original statement. One may say that one reason that the proof simplifies to such a great extent for POVMs is that we in this case get a frame function that is regular on two-dimensional Hilbert spaces, so that no embedding in higher dimensional spaces is needed. However, the two-dimensional case is highly present in both proofs.

Chapter 3

Gleason's theorem for informationally complete POVMs

As noted above, the POVM analogue of Gleason's theorem is valid also for two-state systems, so-called qubits. The purpose of this chapter is to investigate whether a Gleason-type theorem can be proved for two restricted classes of POVMs, namely two 'semi-symmetrical' families of asymmetrical POVMs with four elements, in the qubit case. As we know, every POVM can be seen as representing a measurement, the outcomes of which are in general not mutually exclusive. In this particular case, we will be looking at the situation when we have two possible states for our system, but four possible outcomes of a single measurement.

The four element POVMs discussed below are particularly interesting because they are informationally complete. In dimension N the density matrix (a Hermitian operator of trace one) is represented by a $N \times N$ matrix, and is fully specified by N^2 real numbers, of which $N^2 - 1$ are independent. A POVM with N^2 elements will give exactly N^2 probabilities, of which $N^2 - 1$ are independent. Therefore, when the number of POVM elements is the square of the dimension of the system (two in the qubit case) it is informationally complete, in the sense that its statistics give enough information to construct the density matrix.[1] The symmetric variety of such POVMs, the symmetric informationally complete POVM, or sic-POVM for short, has been shown by Caves et al [2] not to necessitate the quantum probability rule. In some sense, the failure of the proof in the sic case is due to the great degree of symmetry present[2] - the argument appears to go through for all other types of four element POVMs.
In the following, in contrast to in chapter 1 and 2, the continuity of the frame function will *not* be proved, but assumed.

Using the fact that (two-dimensional) Hermitian operators can be expressed in terms of

[1] A PVM will always have its number of elements equal to the dimension of the system, and will give $N - 1$ independent numbers upon measurement. Because of this, no PVM can be complete - we need $N + 1$ PVMs to determine the density matrix; $(N + 1)(N - 1) = N^2 - 1$.

[2] This was suggested by C. Fuchs, private communication.

the Pauli matrices, we can write the general two-dimensional effect as

$$E = r1 + s \cdot \sigma = r1 + s\hat{n} \cdot \sigma \tag{3.1}$$

where 1 is the unit matrix and $\hat{n}$ is a unit vector.

Using this expression, we will end up working on the Bloch-sphere, which is a two-sphere quite distinct from the S^2 of the Gleason proof. The points on this sphere are non-zero vectors in a two-dimensional complex Hilbert space taken modulo a complex number, and orthogonal vectors (states) as usually defined, correspond to antipodal points on the sphere (on the sphere considered by Gleason, antipodal points were identified).

The restricted sets of POVMs considered in this section all consist of effects that are multiples of one-dimensional projectors so that $r = s \leq \frac{1}{2}$ and $E = r(1 + \hat{n} \cdot \sigma)$, and all effects also have the same weight r, namely $\frac{1}{N}$ for a N outcome POVM. We see that a POVM is then fully specified by the $\hat{n}$ vectors of its elements, that sum to zero in order for the effects to sum to one, which is clear from the equality

$$\sum_{i=1}^{N} E_i = N\frac{1}{N} + \frac{1}{N}\sum_{i=1}^{N} \hat{n}_i \cdot \sigma \tag{3.2}$$

So,

$$\sum_{j=1}^{N} \hat{n}_j = 0 \tag{3.3}$$

for any POVM with N elements.
Moreover, we will assume rotational invariance, so that all POVMs that are the same up to a three-dimensional rotation are considered equivalent.

We will now be interested in frame functions $\tilde{f}(\frac{1}{N}(1 + \hat{n} \cdot \sigma)) \equiv f(\hat{n})$ defined on this set. From the right hand side of this equation it is clear that f is a function on the unit sphere in three dimensions. The frame function condition will here as in chapter 2 be normalized according to

$$\sum_{i=1}^{N} f(\hat{n}_i) = 1 \tag{3.4}$$

Proving a Gleason-type theorem is equivalent to showing that a frame function satisfying (3.4) has to be of the form

$$f(\hat{n}) = \mathrm{Tr}(\rho E) = \frac{1}{N}(1 + \hat{n} \cdot P) \tag{3.5}$$

for some ρ, where $P = \mathrm{Tr}(\rho\sigma)$ is a three component vector satisfying $\mid P \mid \leq 1$. Because the frame function is evidently a function on the unit sphere (continuous by assumption) it can be written as a sum of spherical harmonics Y_{lm};

$$f(\hat{n}) = \sum_{lm} c_{lm} Y_{lm} \tag{3.6}$$

The quantum rule for the frame function evidently contains only harmonics with $l = 0$ and $l = 1$; explicitly

$$\frac{1}{N}(1+\hat{n}\cdot P) = \sqrt{\frac{4\pi}{N}}Y_{00}+\sqrt{\frac{2\pi}{3N}}P_x(Y_{1,-1}-Y_{1,1})+i\sqrt{\frac{2\pi}{3}}P_y(Y_{1,-1}+Y_{1,1})+\sqrt{\frac{4\pi}{3N}}P_z Y_{10} \tag{3.7}$$

Note here the difference between this and the case considered by Gleason, where the quantum rule allowed only $l = 0$ and $l = 2$. Again, this is because we are working with a different sphere than did Gleason.

So, what we will want to do is to expand the frame function over the particular POVM we are looking at in terms of spherical harmonics, and check what values of l contribute to the sum. In fact, it is possible to derive a property that has to hold if the lth harmonic is to be allowed in the expansion of a frame function $f(\hat{n})$, namely $c_{lm}\sum_{j=1}^{N} Y_{lr}(\hat{n}_j) = \delta_{l0}$ for all l, m and r.

For $l = 0$ this condition is trivial, and is accomplished simply by normalization.

For $l \geq 1$ it is equivalent to either

$$c_{lm} = 0, \quad m = -l, ..., l \tag{3.8}$$

or

$$\sum_{j=1}^{N} Y_{lr}(\hat{n}_j) = 0, \quad r = -l, ..., l \tag{3.9}$$

This means that if the lth harmonic contributes to a frame function $f(\hat{n})$ that is, if not all of the c_{lm} are equal to zero, then the $\hat{n}$ vectors of the POVM must satisfy equation (3.9).

If we can show that for a certain POVM no l-values other than $l = 0$ and $l = 1$ can satisfy this for all m, we will have derived the quantum rule, thereby proving a Gleason-type theorem. Due to the property of the POVM $\hat{n}$-vectors (3.3), equation (3.9) is automatically satisfied for $l = 1$.

In investigating whether higher values of l can contribute we will make use of some properties of the spherical harmonics (restated here just as in equation (1.5) and equation (1.6) for convenience), namely the way that the θ and ϕ, or for cartesian coordinates $(\hat{n})_z$ and

$(\hat{n})_x + i(\hat{n})_y$, dependencies can be separated, according to

$$Y_{lm}(\hat{n}) = Y_{lm}(\theta,\phi) = \sqrt{\frac{(2l+1)}{4\pi}\frac{(l-m)!}{(l+m)!}}P_l^m(\cos\theta)\mathrm{e}^{im\phi} = \tag{3.10}$$

$$= h_{lm}((\hat{n})_z)((\hat{n})_x + i(\hat{n})_y)^m \tag{3.11}$$

where

$$h_{lm}((\hat{n})_z) = \sqrt{\frac{(2l+1)}{4\pi}\frac{(l-m)!}{(l+m)!}}\frac{P_l^m((\hat{n})_z)}{(\sqrt{1-(\hat{n})_z N2})^m} \tag{3.12}$$

and also, how the spherical harmonics transform under reflection, parity and conjugation

$$Y_{lm}(\hat{n}) = (-1)^{l+m}Y_{lm}(\pi-\theta,\phi) = (-1)^m Y_{lm}(\theta,\phi+\pi) = (-1)^l Y_{lm}(-\hat{n}) = (-1)^m Y^*_{l-m}(\hat{n}) \tag{3.13}$$

It should be noted that continuity of the frame function, necessary for a spherical harmonics expansion to be possible, is here assumed, as opposed to in chapters 1 and 2.
A particularly useful form of equation (3.11) is that for $m = l$, in which case

$$Y_{lm}(\hat{n}) = Y_{ll}(\hat{n}) \propto ((\hat{n})_x + i(\hat{n})_y)^l \tag{3.14}$$

The function $h_{lm}((\hat{n})_z)$ will in this case be independent of $(\hat{n})_z$ because the factor $(\sqrt{1-(\hat{n})_z^2})^l$ in $P_l^l((\hat{n})_z)$ will cancel the explicit dependence of equation (3.12), so that

$$\sum_{j=1}^{4} Y_{ll}(\hat{n}) \propto \sum_{j=1}^{4}((\hat{n}_j)_x + i(\hat{n}_j)_y)^l \tag{3.15}$$

Another result that will be of great importance in the following is

Theorem 3.0.1 *Any two non-zero associated Legendre functions $P_n^m(z)$ and $P_n^s(z)$, where n is an integer such that $n \geq 1$ and $m \neq \pm s$, have on the open interval $(-1,1)$ either no common zero or exactly one common zero. The latter occurs if and only if $n - \mid m \mid$ and $n - \mid s \mid$ are both odd and positive.*

This 1984 result is due to N.H.J. Lacroix [5] and its proof, which is purely analytic and quite elementary[3], uses the fact that the associated Legendre functions are solutions of Legendre's associated equation, which is a second order differential equation.

For a POVM with four elements, the vectors $\{\hat{n}_i\}$ specify the vertices of a tetrahedron. What we will do in the following is to express this tetrahedron in terms of $Y_{lm}(\hat{n}_i)$ and make use of the fact that the sum $\sum_{i=1}^4 Y_{lm}(\hat{n}_i)$ has to be zero for all m in order for the l:th harmonic to contribute to the expression for $f(\hat{n})$. Having concluded that only harmonics with $l = 0, 1$ can contribute to the sum for a given POVM a Gleason result immediately follows, because the fact that the frame function is real requires that the harmonics appear in exactly the combinations of equation (3.7).

[3] It does, however, involve properties of the so-called Prüfer polar coordinates, see [6] for more on those.

3.1 The quantum probability rule for a first class of POVMs

As mentioned above, the unit vectors $\hat{n}$ of the effects of a POVM determine the POVM completely. The unit vectors of a sic-POVM form a regular tetrahedron. If this tetrahedron is stretched, using a parameter θ, a class of POVMs is obtained. In this first case, we will choose the θ dependence so that $\cos\theta = 0$ and $\cos\theta = \pm 1$, correspond to the square and the line, respectively. All values in between correspond to a specific tetrahedron (i.e. a four element POVM).

The four unit vectors specifying a POVM in this class of deformations of the sic-POVM can be expressed as

$$\begin{array}{rcll} \hat{n}_1 & = & (\sin\theta\cos\frac{3\pi}{2}, \sin\theta\sin\frac{3\pi}{2}, \cos\theta) & = (0, -\sin\theta, \cos\theta) \\ \hat{n}_2 & = & (\sin\theta\cos\frac{\pi}{2}, \sin\theta\sin\frac{\pi}{2}, \cos\theta) & = (0, \sin\theta, \cos\theta) \\ \hat{n}_3 & = & (\sin\theta - \pi\cos 0, \sin\theta - \pi\sin 0, \cos\pi - \theta) & = (\sin\theta, 0, -\cos\theta) \\ \hat{n}_4 & = & (\sin\theta - \pi\cos\pi, \sin\theta - \pi\sin\pi, \cos\pi - \theta) & = (-\sin(\theta), 0, -\cos\theta) \end{array} \tag{3.16}$$

It's easily verified that the vectors sum to zero, just as they are supposed to if they are to represent a POVM. Using the symmetrical condition that the inner product between the vectors spanning the POVM is the same for any pair in the set, one sees that the regular tetrahedron corresponds to

$$\theta_{sym} = \arccos\frac{1}{\sqrt{3}} \tag{3.17}$$

To investigate what harmonics can contribute, we start by looking at the sum

$$\sum_{j=1}^{4} Y_{ll}(\hat{n}_j)$$

According to equation (3.15), this sum is proportional to

$$\sum_{j=1}^{4}((\hat{n}_j)_x + i(\hat{n}_j)_y).$$

Using (3.16),

$$\sum_{j=1}^{4} Y_{ll}(\hat{n}_j) \propto 1^l + (-1)^l + (i)^l + (-i)^l \tag{3.18}$$

We see that this sum is zero for $l = 0, 1, 2$ and odds. This means that harmonics with these values of l can contribute to the sum. To find out if they actually do, we have to check whether the sum $\sum_{j=1}^{4} Y_{lm}(\hat{n}_j)$ is zero for all m, not only for $m = l$. For $l = 2$, the sum will be zero for all m and all θ except for $m = 0$. However, for the values of θ that are solutions of $P_2^0(\cos\theta) = 0$ the $l = 2$ harmonic will contribute to the sum, and the quantum rule will not hold.

The zeros of $P_2^0(\cos\theta)$ are $\cos\theta = \pm 1, 0, \pm\frac{1}{\sqrt{3}}$; that is, the line, the square and the regular tetrahedron. For $l = 3$ the only sum which does not give zero for all values of θ is $\sum_{j=1}^{4} P_3^2(\cos\theta)$, the zeros of which are $\cos\theta = 0$ and ± 1.

Also for $l = 5$, the sum for $l = 2$ differs from zero for all θ but those satisfying $P_5^2(\cos\theta) = 0$. The zeros are the same as those for $P_3^2(\cos\theta)$. This is consistent with the results of Caves et al.[2].

For any odd l, $l + 2$ will be odd, so that $P_l^m(\cos(\pi - \theta)) = -P_l^m(\cos\theta)$ by equation (3.13). Also, the ϕ dependence is periodic with period 4π. This means that $\sum_{j=1}^{4} P_l^m(\cos\theta)$ will be non-zero for all odd l and $m \equiv 2 \pmod 4$. The condition for the l:th harmonic to contribute is that the sum $\sum_{j=1}^{4} Y_{lm}(\hat{n}_j)$ is zero for all m.

For odd $l \geq 7$ (even values of l have already been excluded) only those values of θ can contribute that are zeros of both $P_l^2(\cos\theta)$, $P_l^6(\cos\theta)$ and so on all the way up to $m = l-1$.

Hence, in order for harmonics with $l \geq 7$, for which at least $\sum_{j=1}^{4} Y_{l2}(\hat{n}_j)$ and $\sum_{j=1}^{4} Y_{l6}(\hat{n}_j)$ are non-zero, to contribute for specific values θ_0 of θ we need these θ_0 to be common zeros of the l:th associated Legendre functions $P_l^m(\cos\theta)$ for different m. By theorem 3.0.1 no such common zeros of the associated Legendre functions exist, other than $\cos\theta = \pm 1$ and 0. That ± 1 will always be a root of the associated Legendre functions is clear from the first factor in the expression

$$P_l^m(x) = (1 - x^2)^{\frac{m}{2}} \frac{d^m}{dx^m}(P_l(x)) \tag{3.19}$$

So, we have found that for this family of tetrahedrons, we have a Gleason-type theorem (meaning that the quantum rule for calculating probabilites is valid and unique) for all configurations except for the (lower-dimensional) extreme points and the regular tetrahedron representing the sic-POVM.

3.2 The quantum probability rule for a second class of POVMs

The same reasoning can be applied to another family of POVMs, with similar semi-regular properties. This time, we choose the deformation parameter θ so that the unit vectors of this class of POVMs span tetrahedrons whose base is always a regular triangle (in the extremal point, one of the vectors is the zero vector, and the configuration is just a trine).

This subset of tetrahedrons can be parametrized as

$$\begin{array}{rcll} \hat{n}_1 & = & (0,0,-3\cos\theta\cos 0) & = (0,0,-3\cos\theta) \\ \hat{n}_2 & = & (\sin\theta cos\frac{2\pi}{3}, \sin\theta\sin\frac{2\pi}{3}, \cos\theta) & = (-\frac{1}{2}\sin\theta, \frac{\sqrt{3}}{2}\sin\theta, \cos\theta) \\ \hat{n}_3 & = & (\sin\theta cos\frac{4\pi}{3}, \sin\theta\sin\frac{4\pi}{3}, \cos\theta) & = (-\frac{1}{2}\sin\theta, -\frac{\sqrt{3}}{2}\sin\theta, \cos\theta) \\ \hat{n}_4 & = & (\sin\theta\cos 0, \sin\theta\sin 0, \cos\theta) & = (\sin\theta, 0, \cos\theta) \end{array} \tag{3.20}$$

The symmetrical case corresponds to $\cos\theta = \frac{1}{3}$, which is seen as follows.
The symmetrical condition of the pairwise inner product between the POVM vectors being the same for all pairs in the set, and specifically

$$\hat{n}_1 \cdot \hat{n}_2 = \hat{n}_2 \cdot \hat{n}_3 \tag{3.21}$$

gives

$$-3\cos\theta^2 = \frac{1}{4}\sin\theta^2 - \frac{3}{4}\sin\theta^2 + \cos\theta^2 \Rightarrow$$

$$\Rightarrow 4cos\theta^2 = \frac{1}{2}(1-\cos\theta^2) \Rightarrow \frac{9}{2}\cos\theta^2 = \frac{1}{2} \Rightarrow$$

$$\Rightarrow \cos\theta = \pm\frac{1}{3} \tag{3.22}$$

The case $\theta = \frac{\pi}{2}$ is just the two-dimensional regular trine, while $\theta = \pi$ and $\theta = 0$ both give the straight line.

Proceeding in analogue with the previous case, we consider the sum

$$\begin{array}{l} \sum_{j=1}^{4} Y_{ll}(\hat{n}_j) \propto (-\frac{1}{2} + i\frac{\sqrt{3}}{2})^l + (-\frac{1}{2} - i\frac{\sqrt{3}}{2})^l + 1^l = \\ = (\mathrm{e}^{il\alpha} + \mathrm{e}^{il\beta} + 1) \propto \mathrm{e}^{il\alpha} + \mathrm{e}^{il\beta} + 1 \end{array}$$

with

$$\begin{array}{l} \alpha \equiv \arctan -\sqrt{3} = \frac{\pi}{3} \\ \beta \equiv \arctan \sqrt{3} = \frac{4\pi}{3} \end{array}$$

This expression being equal to zero of course means that its real and imaginary parts are zero separately. This leads to the condition

$$\cos l\alpha + \cos l\beta + 1 = 0$$

and

$$\sin l\alpha + \sin l\beta = 0$$

which gives

$$\cos l\alpha = \cos l\beta$$

$$\Rightarrow 2\cos l\alpha + 1 = 0$$

So, the condition for the Y_{ll}'s to sum to zero is

$$\cos l\alpha = \cos l\frac{\pi}{3} = -\frac{1}{2} \quad \Rightarrow \quad l = 0, \quad l = 2, \quad l \equiv 1 \pmod 2 \quad \wedge \quad l \not\equiv 0 \pmod 3 \tag{3.23}$$

Hence, the harmonics that can possibly contribute for this type of tetrahedrons have l equal to zero, two or to an odd number that is not a multiple of three.

To find out if these values of l really do contribute, resulting, for any $l \neq 0, 1$ contributing, in the lack of a Gleason theorem, we proceed as in the previous case. Due to the ϕ dependence, the cases in which the sum

$$\sum_{j=1}^{4} Y_{lm}(\hat{n}_j)$$

will be non-zero occur for $m = 3, 6, 9$...and so on. For $l = 5$, $m = 3$ is the only allowed multiple of three, and for the values of θ that give $P_5^3(\cos\theta) = 0$ the $l = 5$ harmonic contributes, and we do not have a Gleason theorem. The zeros of P_5^3 are, apart from 0 and ± 1, $\pm\frac{1}{3}$, which as noted above, is exactly the regular tetrahedron. Higher values of l either will be even or will allow at least two m-values that are multiples of three. Due to the lack of common zeroes of the associated Legendre polynomials for different m as stated in theorem 3.0.1, we can deduce that only $l = 0, 2$ can contribute, so that we do get the standard quantum rule, except for the cases which give $P_5^3(\cos\theta) = 0$.

3.3 Summary

To summarize, we have found (*assuming* continuity, not proving it, *n.b.*) that a Gleason-type theorem can be proved for all POVMs in these two families, apart from the regular one. These two ways of deforming a regular tetrahedron are arguably the two most symmetric ways of creating an irregular tetrahedron, and seeing as how it appears to be the high degree of symmetry that causes the proof to fail in the sic case, it is not likely that any other tetrahedrons would exhibit a behaviour like that of the regular tetrahedron.

Chapter 4

Kochen-Specker's theorem

4.1 Kochen-Specker's theorem

The theorem known as Kochen and Specker's theorem (KS) was formulated by Simon Kochen and Ernst Specker [7] in 1967. The effective statement of the theorem, sometimes referred to as the Bell-Kochen-Specker theorem [9], is that it in a Hilbert space of dimension $N \geq 3$ is impossible to assign definite values from $\{0, 1\}$ to all projection operators, i.e. vectors[1], in such a way that in each set of N orthogonal vectors exactly one vector is assigned the value 1. This is, in effect, a corollary to Gleason's theorem, since it can be shown that no density matrices give rise to such probabilities, but it can be and was proved independently of Gleason's result (albeit 10 years later).

The physical implications of KS in relation to theories of so-called hidden variables have been much discussed, and the most common interpretation is that the theorem places severe restrictions on any such theory; in effect, that the theorem implies that any well-defined properties possessed by particles would necessarily have to be contextual - the original authors themselves thought their result to establish the 'nonexistence of hidden variables'. Recently, however, arguments have been made for an understanding of KS rather as an epistemological statement about the limitation of the knowledge possible to obtain through measurement. [8]

A problem that over the years has evolved into a downright contest, is that of finding a finite set of KS uncolourable vectors. Kochen and Specker in their proof used 117 vectors arranged in an ingenious way; the current record (in three dimensions) is due to Conway and Kochen and lies at 31 - a number that can be further reduced in higher dimensions.[2] With a slight change of the rules, however, all of these records are easily broken. In two dimensions, consider four points equally spaced on a circle, and first treat them as a four

[1] We will in the following look at unit vectors but are in fact interested in rays rather than vectors, because in establishing orthogonality relations only directions are relevant; consequently, each unit vector will represent all vectors with the same direction.

[2] See for example Peres, [10]

element POVM. This necessitates a colouring that renders one of the points black and the remaining three white. However, the two pairs of anti-parallell vectors also form two two-dimensional PVMs, which requires exactly one vector in each pair to be coloured black, resulting in two of the four points to be black - and we have a contradiction.
In section 4.4 we will consider another finite set of vectors in an attempt to prove a KS result for sic-POVMs - a set that is shown not to suffice for this purpose.

4.2 A KS colouring in arbitrary dimensions

Let us first consider the three-dimensional case. We are interested in assigning value 0 or 1 to vectors in $\mathscr{H}^3$ in such a way that no set of three mutually orthogonal vectors are all assigned the value 0, and no pair of orthogonal vectors both have the value 1. These conditions can be expressed as

$$g : S^2 \to \{0, 1\} \tag{4.1}$$

$$g(P_1) + g(P_2) + g(P_3) = 1 \tag{4.2}$$

for all sets of orthogonal vectors $\{P_1, P_2, P_3\}$, S^2 being the unit two-sphere. Letting white represent the value 0 and black the value 1, this problem can be translated into the problem of colouring S^2, in a way that satisfies the conditions just stated.

Any such assignment of truth values (probabilites from $\{0, 1\}$) to all vectors in the Hilbert space of some system would correspond to (the possibility of) the system having well-defined properties, existing independent of measurement. That is, for any possible observable the outcome of the corresponding measurement would be fully determined in advance. However, by the Kochen-Specker theorem, a complete such assigment of truth values is impossible. Hence, what we will try to do in the following is to assign probabilites from $\{0, 1\}$ according to (4.2) to some of the vectors in $\mathscr{H}$ - some vectors will necessarily remain uncoloured, by KS.

One way to go about this, suggested by Appleby [8], is to start out by colouring the two polar caps defined by $| \tan\theta | < 1$ black, and the region around the equator bounded by $| \tan\theta | = \sqrt{2}$ white, where θ is the usual polar angle. This type of colouring is sketched in figure 4.1.

These limits are derived as follows. The two polar caps are made small enough so that no two vectors in an orthogonal triple can simultaneously lie in the black region, which means that they will extend down to $\theta = \frac{\pi}{4}$. The white section around the equator is just wide enough so that not all three vectors can lie in it at the same time.

An explicit expression for the limit of the latter section is derived as follows, see figure 4.2. The case we primarily have to guard against is when all three vectors lie at the same θ

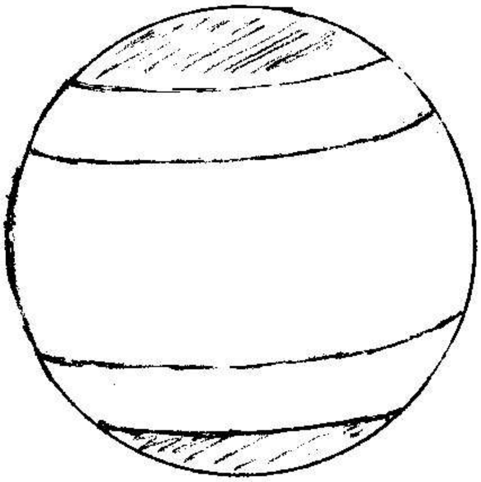

Figure 4.1: A possible (incomplete) KS colouring of the unit two-sphere.

coordinate. The three points on the sphere specified by these vectors are then the corners of a regular triangle, with side length $\sqrt{2}$, because we are on the unit sphere. The radius of a circle in which such a triangle can be inscribed is

$$R_2 = \sqrt{\frac{2}{(2+1)}} = \sqrt{\frac{2}{3}} \tag{4.3}$$

which as can be seen from the figures is exactly $\arcsin \theta_w$, where θ_w is the desired limiting angle. This gives the white area to be the region around the equator with θ values between $\arctan \sqrt{2}$ and $\arctan -\sqrt{2}$.

This colouring of S^2 satisfies the Kochen-Specker criteria for $1 - \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} = 87\%$ of all vectors; that is, all orthogonal triples consisting of vectors from the so coloured areas will satisfy equation (4.2).
An analoguous colouring can be done for S^n, yielding the percentage results 79% for $n = 3$, 74% for $n = 4$ and 71% for $n = 5$.

The above numbers are obtained using the fact that equation (4.3) generalizes to

$$R_n = \sqrt{\frac{n}{n+1}} = \sqrt{\frac{N-1}{N}} \tag{4.4}$$

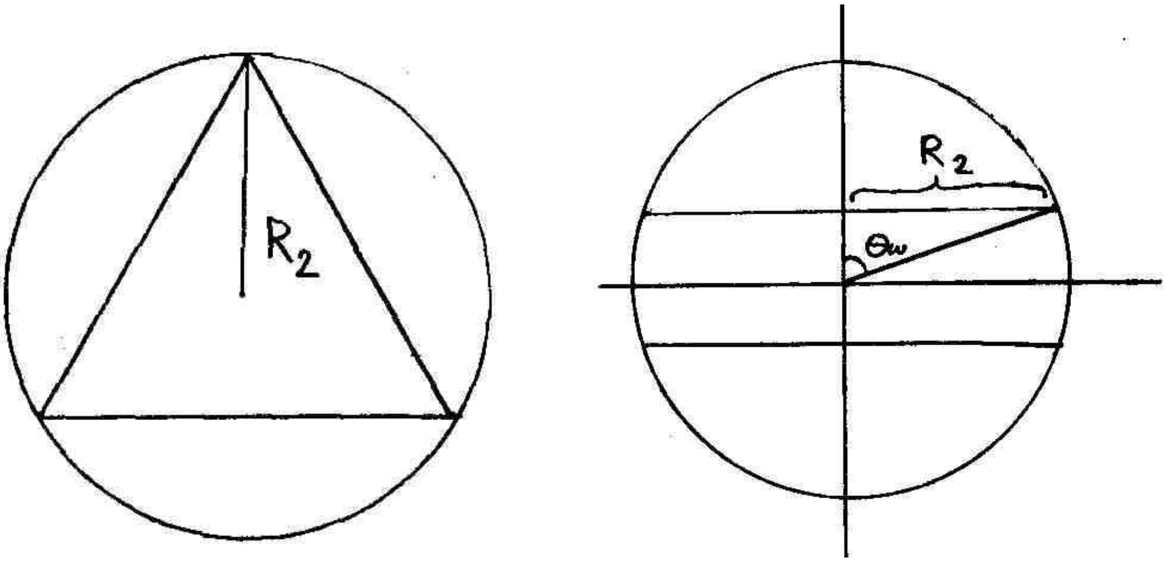

Figure 4.2: Derivation of the limiting angles for the colouring in figure 4.1.

for the radius of the circumsphere of a regular n-simplex, where $n = N - 1$ is the dimension of the sphere in N dimensions.

In the above percentages, the area of the black cap is included, but already in four dimensions the contribution to the total area is close to negligible. As we will see below it will reduce further with increasing dimension, which is why we in the following will primarily be interested in looking at the area taken up by the white section.

The fraction of the sphere in N dimensions that can be coloured white with the given restriction is, using equation (4.4),

$$F = \frac{\int_{\arcsin\sqrt{\frac{N-1}{N}}}^{\frac{\pi}{2}} \sin^{N-2}\theta d\theta}{\int_0^{\frac{\pi}{2}} \sin^{N-2}\theta d\theta} = 2\frac{\mathrm{vol}(S^{N-2})}{\mathrm{vol}(S^{N-1})}\int_{\arcsin\sqrt{\frac{N-1}{N}}}^{\frac{\pi}{2}} \sin^{N-2}\theta d\theta \tag{4.5}$$

where $\mathrm{vol}(S^d)$ denotes the surface area of the d-dimensional sphere.
As for the black area, B_N, it will in analogy with the $N = 3$ case be located around the poles of the sphere, with limiting angle $\frac{\pi}{4}$;

$$B_N = \mathrm{vol}(S^{N-2})\int_0^{\frac{\pi}{4}} \sin^{N-2}\theta d\theta \tag{4.6}$$

What, one may ask, is the fraction of the sphere in N dimensions that can be coloured using this method in the limit $N \to \infty$? As can be seen from the expression

$$\mathrm{vol}(S^d) = \mathrm{vol}(S^{d-1}) \int_0^\pi \sin^{d-1}\theta d\theta \tag{4.7}$$

for high dimensions, the fraction of the area of the sphere that will lie around the poles is negligible, due to the increasingly sharp peak around $\theta = \frac{\pi}{2}$ of the sine function when raised to a large number. Thus the fraction of the surface area taken up by the black section will be very small.

To determine the fraction of the sphere taken up by the white section requires a bit more careful analysis. We will need to evaluate the expression

$$\lim_{N\to\infty} 2\frac{\mathrm{vol}(S^{N-2})}{\mathrm{vol}(S^{N-1})} \int_{\arcsin\sqrt{\frac{N-1}{N}}}^{\frac{\pi}{2}} \sin^{N-2}\theta d\theta \tag{4.8}$$

The volume of the sphere in d dimensions is

$$\mathrm{vol}(S^d) = \frac{2\pi^{\frac{d+1}{2}}}{\Gamma(\frac{d+1}{2})} \tag{4.9}$$

so that

$$\frac{\mathrm{vol}(S^{N-2})}{\mathrm{vol}(S^{N-1})} = \frac{1}{\sqrt{\pi}}\frac{\Gamma(\frac{N}{2})}{\Gamma(\frac{N-1}{2})} \tag{4.10}$$

Equation (4.8) can then be written as

$$\lim_{N\to\infty} 2\frac{1}{\sqrt{\pi}}\frac{\Gamma(\frac{N}{2})}{\Gamma(\frac{N-1}{2})} \int_{\arcsin\sqrt{\frac{N-1}{N}}}^{\frac{\pi}{2}} \sin^{N-2}\theta d\theta$$

We will treat the Gamma function part and the integral part of the expression separately, starting out by looking at the fraction

$$\frac{\Gamma(\frac{N}{2})}{\Gamma(\frac{N-1}{2})}$$

In the limit of large N we can apply Stirling's approximation to the Gamma function

$$\Gamma(z) = \sqrt{2\pi}\mathrm{e}^{-z} z^{z-\frac{1}{2}}(1 + \frac{1}{12z} + O(\frac{1}{z^2})), \quad |\ \mathrm{arg}z\ | < \pi, \quad |\ z\ | \to \infty \tag{4.12}$$

and consider the expression

$$\frac{\sqrt{2\pi}\mathrm{e}^{-\frac{N}{2}} \frac{N}{2}^{\frac{N}{2}-\frac{1}{2}}(1 + \frac{2}{12N} + O(\frac{1}{N^2}))}{\sqrt{2\pi}\mathrm{e}^{-\frac{N-1}{2}} \frac{N-1}{2}^{\frac{N-1}{2}-\frac{1}{2}}(1 + \frac{2}{12(N-1)} + O(\frac{1}{(N-1)^2}))} \to \frac{\sqrt{\mathrm{e}}}{\sqrt{\mathrm{e}}}\frac{\sqrt{N}}{\sqrt{2}}(1 - \frac{1}{N}) \to \frac{\sqrt{N}}{\sqrt{2}}, \quad N \to \infty \tag{4.13}$$

where we have used that

$$(\frac{N-1}{2})^{\frac{N}{2}-1} = (\frac{N}{2})^{\frac{N}{2}-1}(1-\frac{1}{N})^{\frac{N}{2}-1} =$$

$$= (\frac{N}{2})^{\frac{N}{2}-1}((1-\frac{1}{N})^{-N})^{-\frac{1}{2}}\frac{1}{(1-\frac{1}{N})} \to (\frac{N}{2})^{\frac{N}{2}-1}\frac{1}{\sqrt{e}}\frac{1}{(1-\frac{1}{N})}, \quad N \to \infty \tag{4.14}$$

and

$$\frac{(1+\frac{2}{12N}+O(\frac{1}{N^2}))}{(1+\frac{2}{12(N-1)}+O(\frac{1}{(N-1)^2}))} \to 1, N \to \infty \tag{4.15}$$

Thus, we can conclude that

$$\lim_{N\to\infty}\frac{\Gamma(\frac{N}{2})}{\Gamma(\frac{N-1}{2})} = \frac{\sqrt{N}}{\sqrt{2}}$$

Next, let us take a look at the behaviour of the integral

$$\int_{\arcsin\sqrt{\frac{N-1}{N}}}^{\frac{\pi}{2}} \sin^{N-2}\theta d\theta$$

in the limit of large N.

It's not hard to convince oneself that this is equivalent to looking at the integral

$$\int_0^{\arccos\sqrt{\frac{N-1}{N}}} \cos^{N-2}\theta d\theta$$

which simplifies the calculations because all expansions can be done around zero.
In the limit of large N we can use

$$\arccos\sqrt{\frac{N-1}{N}} = \frac{1}{\sqrt{N}} + O(\frac{1}{N^{\frac{3}{2}}}) \tag{4.17}$$

Expanding $\cos t$ around $t = 0$, we get

$$\cos t|_{t=0} = 1 - \frac{t^2}{2!} + O(t^4) \tag{4.18}$$

so that, using the regular binomial expansion and the fact that when N is large $N-2$ can be approximated with N,

$$\lim_{N\to\infty}\cos^{N-2}\theta = (1-\frac{\theta^2}{2})^N + h(\theta,N) = h(\theta,N) + 1 - N\frac{\theta^2}{2} + \frac{N^2}{2!}\frac{\theta^4}{4} - \frac{N^3}{3!}\frac{\theta^6}{8} + ... \tag{4.19}$$

where $h(\theta, N)$ is a function such that

$$\lim_{N\to\infty}\sqrt{N}\int_0^{\arccos\sqrt{\frac{N-1}{N}}} h(\theta,N) = 0$$

Integrating term by term and using equations (4.18) and (4.17), we get

$$\lim_{N\to\infty}\int_0^{\arccos\sqrt{\frac{N-1}{N}}}\cos^{N-2}\theta d\theta = [\theta - \frac{N}{6}\theta^3 + \frac{N^2}{40}\theta^5 - \frac{N^3}{336}\theta^7 + ...]_{\theta=0}^{\theta=\arccos\sqrt{\frac{N-1}{N}}}$$

$$= \frac{1}{\sqrt{N}}\sum_{k=0}^{\infty}\frac{1}{2^k}\frac{1}{k!}\frac{(-1)^k}{(2k+1)} \tag{4.21}$$

The sum in (4.21) is in fact equal to

$$\sqrt{\frac{\pi}{2}}\text{erf}(\frac{1}{\sqrt{2}})$$

with erf the statistic-probabilistic error function;

$$\text{erf}(z) \equiv \frac{2}{\sqrt{\pi}}\int_0^z \mathrm{e}^{-t^2}dt \tag{4.22}$$

Putting all of this together, we have the result

$$\lim_{N\to\infty} 2\frac{\text{vol}(S^{n-2})}{\text{vol}(S^{n-1})}\int_{\arcsin\sqrt{\frac{N-1}{N}}}^{\frac{\pi}{2}}\sin^{N-2}\theta d\theta = \text{erf}(\frac{1}{\sqrt{2}}) \approx 0.68 \tag{4.23}$$

So, approaching the limit of an infinite number of dimensions of the Hilbert space $\mathscr{H}$ in which our projective measurements are conducted, binary probabilities (corresponding to well-defined, non-contextual properties of the system with available states in $\mathscr{H}$) can be assigned to approximately 68% of the vectors in $\mathscr{H}$. The behaviour of the percentage as a function of dimension[3] is given in figure 4.3.

The minimum occurs around 12.465, and the integer giving the least percentage is $N = 12$; about 66.76%.

What has been derived above is a lower limit for the area of the sphere that is KS colourable in arbitrary dimensions. The possibility remains, however, that a maximally effective colouring could cover a much larger area - possibly, in fact, as much as 99%[4] of the sphere in $\mathbb{R}^3$ [8].

[3]Treated in the diagram as a continuous variable.

[4]There are some measure theoretical subtleties that can be used to colour even more of the sphere, in fact 'almost' all of it, in a very special sense of the word. For more on this, see for example [11].

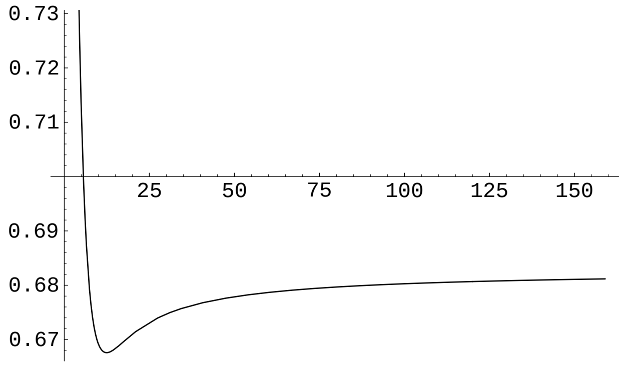

Figure 4.3: Percentage of the sphere in N dimensions that is colourable using the above method, as a function of N.

4.3 KS coloured bases

The physically relevant question, however, is arguably not how large a fraction of all states can be assigned probabilities 1 or 0, but rather what percentage of all complete orthogonal bases, corresponding to measurements, can have all their basis vectors assigned binary probabilities in a consistent way. A geometrical consideration (that can likely be further generalized, albeit not without some effort) allows us to answer this question in three and four dimensions.

Let us first consider the colouring of the two-sphere proposed above - a black cap and a white equatorial belt covering in total 87% of the sphere - and make use of the regular measure on $\mathbb{R}^3$ to compare the number of properly coloured bases consisting of vectors from these sections with the total number of ordered orthonormal triples in $\mathbb{R}^3$.

In a properly coloured base exactly one vector has to be black, so one of the three vectors in an orthogonal triple has to be chosen to lie on one of the black caps. The remaining two orthogonal vectors can then be chosen from a great circle orthogonal to the first vector - the question is how large a fraction of this great circle will lie within the white section and also how the second vector (which, of course, completely determines the third basis vector up to a sign) can be chosen so that the third vector will also be contained within the white section.

Figure 4.4 depicts the plane of the great circle orthogonal to the first (black) vector on which the remaining two vectors in the orthogonal triple will have to lie. The circle segment bounding the striped area is the cut between the white belt and the orthogonal great circle. For any choice of second vector from this section, the third vector will be fully determined (up to a sign). Hence, we cannot choose our second vector in a satisfactorily

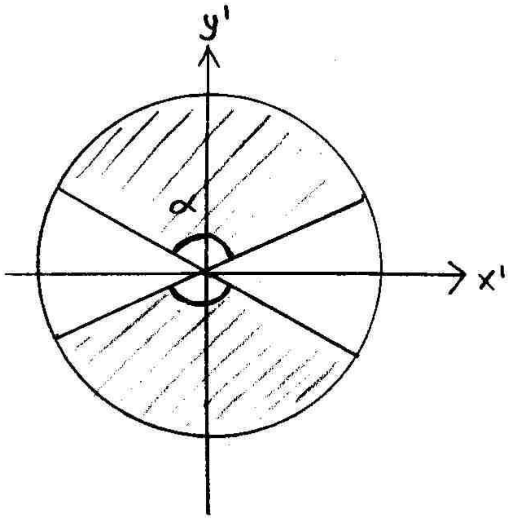

Figure 4.4: A cut through the plane of the great circle orthogonal to the vector chosen to lie on the black cap.

coloured triple from any part of the circle-belt overlap in figure 4.4, but only from the sectors that will result in the third vector lying in the white belt as well. Given a second vector, the third is obtained by rotation in the great circle plane by an angle of $\frac{\pi}{2}$. The allowed choices for second vector are then the points such that the points corresponding to a $\frac{\pi}{2}$ rotation of these points are also white. This set of points is just the overlap between the white (striped) sector in figure 4.4 and the same sector rotated by $\frac{\pi}{2}$, as illustrated in figure 4.5, an overlap that can be shown to always be non-empty. Hence, what we will need to find is the total angle taken up by the striped section in figure 4.5 - this will be denoted by β.

It's clear that this β can be expressed in terms of the α of figure 4.4 as

$$\beta = 4\alpha - 2\pi \tag{4.24}$$

The angle α, in turn, can be expressed in terms of the regular polar angle θ that specifies our choice of black vector using the following procedure.
First, consider the plane spanned by the vector chosen to lie in the black section, call it z', and a vector y' in the plane orthogonal to z'; $\{x, y, z\}$ is a reference coordinate system as shown. The vector x' orthogonal to y' and z' is chosen so that its z component equals zero. From figure 4.6 it is clear that

$$z = 0x' + \sin\theta y' + \cos\theta z' \tag{4.25}$$

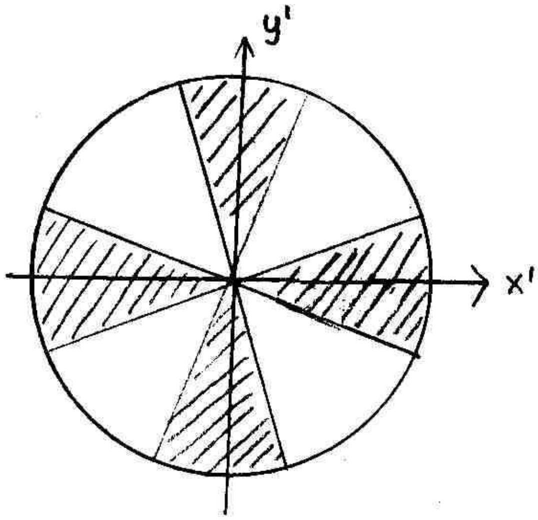

Figure 4.5: Overlap between the white section of the great circle, and its rotation by $\frac{\pi}{2}$.

Meanwhile, as can (hopefully) be seen from figure 4.7, a vector v lying just on the boundary of the white belt can be expressed in terms of y' and x' as

$$v = \cos\alpha' x' + \sin\alpha' y' \tag{4.26}$$

with $\alpha' = \frac{\alpha}{2}$, its z component being equal to zero. We also know that the z component of our vector v is just h, with $h = \frac{1}{\sqrt{3}}$ according to our earlier deliberations. Taken together, this gives

$$v \cdot z = h = (\cos\alpha' x' + \sin\alpha' y') \cdot z = \sin\alpha' y' \cdot z = \sin\alpha' \sin\theta \tag{4.27}$$

so that

$$\alpha = 2 \arcsin \frac{h}{\sin\theta} \tag{4.28}$$

and

$$\beta = 8 \arcsin \frac{h}{\sin\theta} - 2\pi \tag{4.29}$$

When $\theta < \arcsin \frac{1}{\sqrt{3}}$ expression (4.28) for α will not be defined; for those angles all of the vectors orthogonal to the black section vector defined by the angle θ will lie within the white section.
This enables us to express the fraction of the orthogonal great circle corresponding to every choice of vector z' in terms of the angle θ, making possible integration over all values of θ and thereby the comparison we have in mind.

So, the integrals we will want to evaluate are

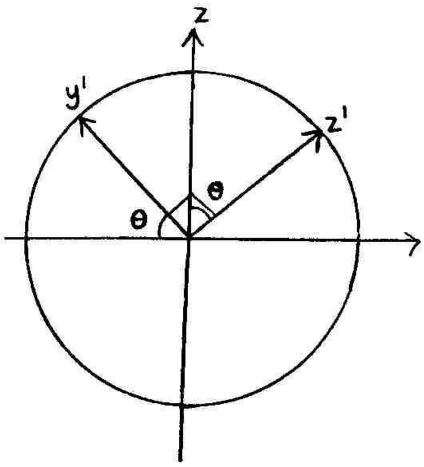

Figure 4.6: The vector y' will make an angle $\pi - \theta$ with the z axis.

$$I = 2\pi \int_0^{\arcsin\frac{1}{\sqrt{3}}} \sin\theta d\theta + \int_{\arcsin\frac{1}{\sqrt{3}}}^{\frac{\pi}{4}} (8 \arcsin \frac{h}{\sin\theta} - 2\pi) \sin\theta d\theta \tag{4.30}$$

the value of which turns out to be 1.4572.
This, multiplied by a combinatorial factor of three because what we considered the first vector could as well have been the second or third, should be compared to the value of the integral

$$2\pi \int_0^{\frac{\pi}{2}} \sin\theta d\theta = 2\pi \tag{4.31}$$

- the result is that approximately 69% of all possible ordered bases in $\mathbb{R}^3$ can be satisfactorily KS-coloured using the given construction.

The above considerations for the three-dimensional case can with some modifications be applied also in four dimensions. In this case, the white equatorial belt will be a three-dimensional object, and the orthogonal great circle will have turned into a two-sphere. Introducing spherical coordinates $\{\phi, \theta_1, \theta_2\}$ on the three-sphere, we will start out by finding the intersectional area of the orthgonal two-sphere and the white 'belt'. Let z' denote the black vector, let z be a reference coordinate, and let y' be a vector on the orthogonal two-sphere as in figure 4.8.

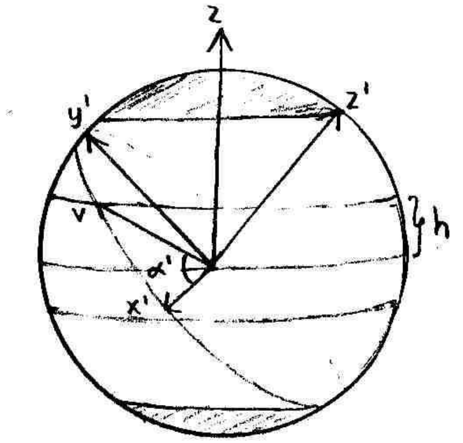

Figure 4.7: Coordinates x' and y' are introduced in the plane of the great circle orthogonal to the vector z'.

The white section is the set of vectors

$$\{u : \quad | u \cdot z | \leq A\}, \quad A = \frac{1}{2} \tag{4.32}$$

For any vector u in this set we have that

$$u \cdot z' = 0 \tag{4.33}$$

Now, let's make the ansatz

$$y' = az + bz' \tag{4.34}$$

Normalization together with condition (4.33) then gives

$$a^2 + b^2 + 2abz' \cdot z = 1 \quad \text{and} \quad az' \cdot z + b = 0 \tag{4.35}$$

which combines to

$$a = \frac{1}{\sin\theta_2}, \quad b = -\frac{\cos\theta_2}{\sin\theta_2} \tag{4.36}$$

Also,

$$u \cdot z' = 0 \Rightarrow u \cdot y' = u \cdot (az + bz') = au \cdot z \tag{4.37}$$

So, using (4.32), the belt on the orthogonal two-sphere will be the set of vectors

$$\{v : \quad | v | \leq B\}, \quad B = aA = \frac{1}{2\sin\theta_2} \tag{4.38}$$

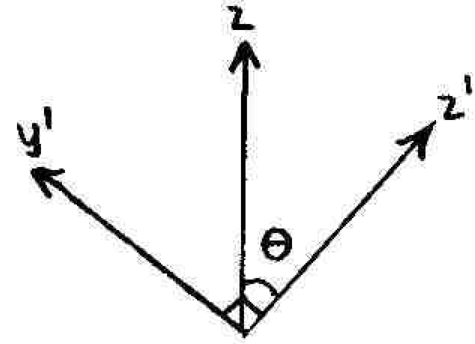

Figure 4.8: y' lies on the two-sphere orthogonal to the vector z'.

For $0 \leq \theta_2 \leq \arcsin\frac{1}{2}$ the entire orthogonal two-sphere will lie within the white section.

This intersection between the orthogonal two-sphere and the white section on the three-sphere can now be treated in analogue with the previuos case. Given a black first vector, when placing the second vector in the white section, the segment of the great circle orthogonal to this second vector on which we can choose the third in order for the fourth to lie in the white section is given by

$$\gamma = 8\arcsin\frac{B}{\sin\theta_1} - 2\pi \tag{4.39}$$

Also in analogy with the previous case, all of the orthogonal great circle will be white for $\arccos B \leq \theta_1 \leq \arcsin B$. To summarize, we have integration over the angle θ_2 which runs between 0 and $\frac{\pi}{2}$, covering the black cap, and the possibilities available for choosing the remaining three vectors are governed by a function of θ_2, obtained from an integration over the angle θ_1 between $\arccos B$ and $\frac{\pi}{2}$, that is, over the white section of the two-sphere orthogonal to the first vector specified by θ_2, B being a function of θ_2.

To make all of this explicit, we have the following integrals

$$I = 2\pi\int_{\arccos B}^{\arcsin B}\sin\theta_1 d\theta_1 + \int_{\arcsin B}^{\frac{\pi}{2}}(8\arcsin\frac{B}{\sin\theta_1} - 2\pi)\sin\theta_1 d\theta_1 \tag{4.40}$$

and, finally

$$4\pi\int_0^{\arcsin\frac{1}{2}}\sin^2\theta_2 d\theta_2 + \int_{\arcsin\frac{1}{2}}^{\frac{\pi}{4}} I\sin^2\theta_2 d\theta_2 \tag{4.41}$$

The result when comparing this, multiplied by an overall combinatorial factor of four, to the value of the expression

$$4\pi \int_0^{\frac{\pi}{2}} \sin^2 \theta d\theta \tag{4.42}$$

is that 32% of the ordered orthogonal triples in $\mathbb{R}^4$ are properly coloured using the chosen method.

4.4 KS for a restricted class of sic-POVMs

As mentioned in the introduction to this chapter, a lot of people have been working on the problem of proving the KS theorem by finite means. This quest has focused mainly on PVMs, that is orthogonal resolutions of the identity, but some work has also been done for POVMs of different types. Cabello, for instance, has shown the KS theorem for a single qubit using eight element POVMs [12] by inscribing cubes in dodecahedrons. Masahiro Nakamura has done the same using POVMs with four elements. As noted earlier, the KS result follows trivially if one simultaneously allows POVMs with different numbers of elements.

It should be noted that the lack of a Gleason type theorem for sic-POVMs in the qubit case opens up for the possibility that we don't have the restriction of the KS result either. It gives no further information, though, if this is actually the case. It is therefore worthwhile to investigate some more obvious sets of four-element POVMs to see if they can be KS coloured.

Here, we will make use of the dodecahedron method proposed by Cabello. For the restricted class of regular tetrahedrons that can be inscribed in a regular dodecahedron so that the vertices of the two configurations coincide, it is possible to by explicit construction prove that the KS result is not applicable. The reader can convince herself that an inscribed regular tetrahedron will have corners lying on four mutually next-to-next-to-adjacent vertices. For a fixed dodecahedron, there are 10 possible ways to inscribe a regular tetrahedron.

A satisfactory Kochen-Specker colouring is then given, as in figure 4.9, by

A I M Q **F** C P S
A H P R **J** C L T
E G N Q **J** B R N
E H L U **K** D G T
F D M U **K** B I S

where boldface characters imply that the corresponding vertex is coloured black.

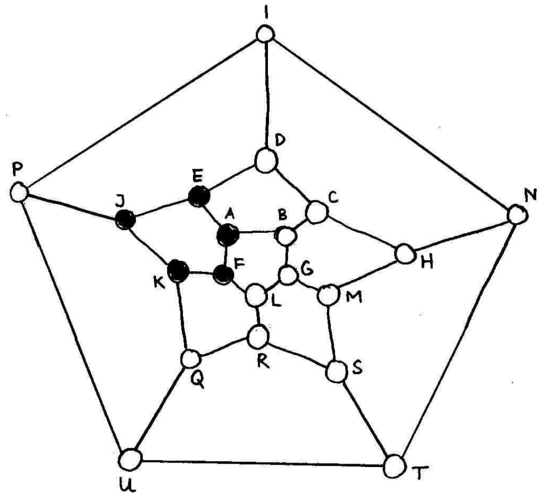

Figure 4.9: Colouring of the vertices of a regular dodecahedron, corresponding to a negation of the KS-theorem for the class of regular tetrahedrons that can be inscribed.

Consequently, the KS result is not valid for this restricted class of symmetric informationally complete POVMs, and whether or not we have the KS result for all sic-POVMs is still an open question.

Chapter 5

Conclusions and open questions

We have seen that it suffices to make the frame function assumption for POVMs with two and three elements in order to prove a Gleason type theorem for general POVMs. In the case of two-dimensional quantum systems, qubits, the sic-POVM does not yield the Gleason result, as has been shown by Caves et al. [3]. The considerations of this diploma work, however, have made likely that for all other four element POVMs the quantum rule holds.

Generalizing a method of colouring proposed by Appleby in three dimensions [8], we have also found a lower limit for the area of the n-sphere that can be KS coloured, but we are still ignorant as to a sharp upper limit.

In three and four dimensions, we have calculated how many of all possible bases the coloured area corresponds to. In order to make the lower limit mentioned above more physically interesting, one would need to answer this question in arbitrary dimensions - something that appears to be non-trivial.

The Appleby colouring has the advantage that it generalizes easily to higher dimensions. As for a maximally effective colouring, there are no arguments to support that this would be the case. In particular, it is in no way obvious that the same method of colouring would be maximal in different dimensions.

Another question left unanswered is that about the existence of a finite set of vectors that can be used to prove the KS theorem for sic-POVMs - as of yet, no such set has been presented.

Bibliography

[1] Andrew M. Gleason, *Measures on the closed subspaces of a Hilbert space*, Journal of Mathematics and Mechanics, Vol. 6, No. 6, 885-893 (1957)

[2] N.I. Akhiezer, I.M. Glazman, *Theory of Linear Operators in Hilbert Space*, (Ungar, New York, 1963)

[3] C.M. Caves, C.A. Fuchs, K. Manne and J.M. Renes,'Gleason-type Derivations of the Quantum Probability Rule for Generalized Measurements', `quant-ph/0306179`; Found. Phys. 34, 193 (2004)

[4] Paul Busch, 'Quantum states and generalized observables: a simple proof of Gleason's theorem', `quant-ph/9909073`; Phys. Rev. Lett. 91, 120403 (2003)

[5] Norbert H.J. Lacroix, *On Common Zeros of Legendre's Associated Functions*, Mathematics of Computation, Vol. 43, No. 167, 243-245 (1984)

[6] Hans Sagan, *Boundary and Eigenvalue Problems in Mathematical Physics*, (Chelsea, New York, 1955)

[7] S. Kochen and E.P. Specker, *The Problem of Hidden Variables in Quantum Mechanics*, J. Math. Mech., Vol. 17, No. 1, 59-87 (1967)

[8] D.M. Appleby, 'The Bell-Kochen-Specker theorem', `quant-ph/0308114`; Stud. Hist. Philos. Mod. Phys. 36 (2005)

[9] J.S. Bell, *On the Problem of Hidden Variables in Quantum Mechanics*, Rev. Mod. Phys., Vol. 38, No. 3, 447-452 (1966)

[10] A. Peres, *Quantum theory: Concepts and Methods*,(Kluwer Academic Publishers, Dordrecht, 1995)

[11] I. Pitowsky, *Quantum Mechanics and Value definiteness*, Philos. Sci., Vol. 52, No. 1, 154-156 (1985)

[12] Adán Cabello, 'Kochen-Specker Theorem for a Single Qubit Using Positive Operator-Valued Measures', `quant-ph/0210082`; Phys. Rev. Lett. 90 (2003)