1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

A Post-Quantum Associative Memory

Ludovico Lami

Institut für Theoretische Physik und IQST, Universität Ulm, Albert-Einstein-Allee 11,
D-89069 Ulm, Germany
QuSoft, Korteweg-de Vries Institute for Mathematics, and Institute for Theoretical Physics,
University of Amsterdam, Science Park, 1098 XG Amsterdam, the Netherlands

Daniel Goldwater

School of Mathematical Sciences, University of Nottingham, University Park, Nottingham
NG7 2RD, United Kingdom

Gerardo Adesso

School of Mathematical Sciences, University of Nottingham, University Park, Nottingham
NG7 2RD, United Kingdom

E-mail: gerardo.adesso@nottingham.ac.uk

Abstract.   Associative memories are devices storing information that can be fully retrieved given partial disclosure of it. We examine a toy model of associative memory and the ultimate limitations it is subjected to within the framework of general probabilistic theories (GPTs), which represent the most general class of physical theories satisfying some basic operational axioms. We ask ourselves how large the dimension of a GPT should be so that it can accommodate 2^m states with the property that any N of them are perfectly distinguishable. Call $d(N, m)$ the minimal such dimension. Invoking an old result by Danzer and Grünbaum, we prove that $d(2, m) = m + 1$, to be compared with $O(2^m)$ when the GPT is required to be either classical or quantum. This yields an example of a task where GPTs outperform both classical and quantum theory exponentially. More generally, we resolve the case of fixed N and asymptotically large m, proving that $d(N, m) \leq m^{1+o_N(1)}$ (as $m \to \infty$) for every $N \geq 2$, which yields again an exponential improvement over classical and quantum theories. Finally, we develop a numerical approach to the general problem of finding the largest N-wise mutually distinguishable set for a given GPT, which can be seen as an instance of the maximum clique problem on N-regular hypergraphs.

A Post-Quantum Associative Memory   2

1. Introduction

A memory is a physical system which can be used to store some information which can later be retrieved. Memories can be complete, if all the information stored can be recovered at once; or incomplete, if only a part of it can be accessed. They can be perfect, if the retrieved pieces of information reproduce the original ones with probability one, or imperfect otherwise. The physical interest of designing an incomplete or imperfect memory is that in return for the loss of performance there might be an effective compression of the system size.

For example, the *quantum random access encodings* of Ambainis et al. [1] (see also [2]) allow for storing $2n$ classical bits into n qubits, in such a way that *any* given bit (but not all of them simultaneously) can be retrieved with probability $p \approx 0.79$. These memories are therefore incomplete and imperfect, but they allow for an effective compression of the physical system employed, as compared to the naïve encoding of $2n$ bits into $2n$ qubits, which is both complete and perfect.

The celebrated Hopfield network [3] is another example of an imperfect memory designed to model biological systems. An array of neurons is connected based on the desired information to be stored. The dynamics of the array result in attractors that precisely correspond to the stored states. The net effect is that upon being prepared in a certain initial configuration, the system often evolves towards the stored state that most resembles it. This mechanism amounts to an imperfect retrieval of the encoded information. The Hopfield network is in a certain sense an incomplete memory, because the recovery of a certain stored state can take place only if the initial configuration is sufficiently close to it. In other words; *some* information about the stored state has to be disclosed if we want to retrieve the rest.

In this paper we want to study and characterise the ultimate physical limitations to the performance of incomplete memories. In order to achieve this, following recent developments [4, 5, 6, 7, 8, 9] we will utilise the formalism of *general probabilistic theories* (GPTs). Within this mathematical framework, it is possible to model a vast family of physical theories, including classical probability theory, quantum mechanics; and more exotic theories such as generalised bits [10], spherical models [6], and Popescu–Rohrlich (PR) boxes [11, 12], to name a few [13, 14, 15]. We are particularly interested in finding out to what extent GPTs can exhibit an enhanced memory capacity compared to classical and quantum theories.

The paper is organised as follows. In Section 1.1 we expand upon and formalise the problem we are addressing — establishing the relationship between a physical theory and the kind of incomplete memory which could be constructed within it. Section 2 reviews the GPT formalism and introduce some well known theories for reference. Section 3 holds the first main result: we prove that a particular class of theories (those with hypercubic state spaces) are optimal for housing incomplete memories that can retrieve one lost bit. This is obtained by invoking a seminal result by Danzer and Grünbaum [16]. In Section 4 we begin to search for the optimal theory when the number of bits to be retrieved is arbitrary. There we prove our second main result, which gives the scaling of the minimal dimension of a GPT that can host very large incomplete memories capable of retrieving a fixed number of lost bits. In both of our main results, GPTs are shown to outperform classical and quantum associative memories *exponentially*. In Section 5 we recast the task of determining the largest N-wise mutually distinguishable set for a given GPT as the convex problem of finding the maximum N-clique on an N-regular hypergraph, for which we provide deterministic algorithms. Finally, we conclude in Section 6.

A Post-Quantum Associative Memory   3

1.1. The problem

We will focus on the simplest type of incomplete perfect memory, whose general working principle is as follows. We begin by storing in it an m-bit string $x \in \{0, 1\}^m$ by means of a suitable encoding. The value of x is then forgotten, with the only remaining record being stored in the memory. Later, we are given N m-bit strings $Y = \{y_1, \ldots, y_N\}$, with $y_i \in \{0, 1\}^m$, $i = 1, \ldots, N$, with the promise that one of these matches the original string, $y_i = x$. Our task is to determine x by making a suitable measurement on our device. The memory is called incomplete if the largest achievable N satisfies $N_{\max} < 2^m$, and perfect if the recovery can be achieved with unit success probability for all choices of x and Y, with the constraint that Y has cardinality N.

If we model physical systems in terms of GPTs, the problem can be equivalently seen as asking for a GPT A and an encoding function $\rho : \{0, 1\}^m \rightarrow \Omega_A$, with Ω_A being the state space on A, such that any N distinct states $\rho(x_1), \ldots, \rho(x_N) \in \Omega_A$ are perfectly distinguishable. Equivalently, we could demand the existence of 2^m states $\rho_1, \ldots, \rho_{2^m} \in \Omega_A$ that are N-wise mutually distinguishable, meaning that any N of them are (jointly) perfectly distinguishable. We will formalise our notions of *perfect distinguishability* and *mutual N-wise distinguishability* in Definitions 8 and 9. Especially the former concept has attracted considerable interest recently [17, 18].

In order to assess the capacity of a memory in system A, we will want to quantify the effective *compression* operated by the encoding ρ. This presents a problem; we cannot count the number of bits or qubits in the system A, because this will be modelled by a GPT that is in general neither classical nor quantum. However, there *is* a universal way to quantify how 'large' a GPT is: its dimension. Since a classical m-bit system can be represented by a GPT of dimension 2^m, we could employ the logarithm $\log_2 d$ of the dimension $d = \dim V$ of a certain GPT as an effective measure of its memory capacity. Along the same lines, we could employ the *compression factor*

$$\kappa = \frac{\# \text{ encoded bits}}{\log_2 \dim V} = \frac{m}{\log_2 d} \tag{1}$$

to assess the quality of the scheme. The N-wise compression factor, denoted with $\kappa(N, m)$, is the maximum such κ that is achievable with all possible GPTs. It is a universal function of the pair (N, m), as the optimisation does away with the degree of freedom represented by the choice of the underlying GPT. Clearly, it is given by $\kappa(N, m) = \frac{m}{\log_2 d(N,m)}$, where $d(N, m)$ is the minimum d such that a d-dimensional GPT hosting an N-wise mutually distinguishable set of states of cardinality 2^m can be found. This discussion allows us to precisely state our problem as follows.

> **Problem.** For all pairs of positive integers N, m, compute $\kappa(N, m)$, i.e. determine the minimum dimension of a GPT that can host an N-wise mutually distinguishable set of states of cardinality 2^m.

Remark 1 Given n strings encoded into a system A, and granted that those states are all pairwise perfectly distinguishable, it will take (at most) $n - 1$ measurements to uniquely identify the desired state via a tournament-like method, granted that the measurements are non-disturbing (see [19]). Alternatively, if the measurements *are* disturbing, we would require an equivalent number of copies of the system A. If instead those states are mutually N-wise distinguishable, these numbers reduce to $\left\lceil \frac{n-1}{N-1} \right\rceil$.

Before we get to the formalism of GPTs, by means of which we will explore more exotic theories, we can first examine the performance of the most familiar ones: quantum

A Post-Quantum Associative Memory   4

and classical mechanics. For these examples we choose $N = 2$, so that we are finding the maximum number of pairwise distinguishable states which a system can store.

- *Classical theory.* If two classical probability distributions over an alphabet $\mathcal{X}$ are pairwise perfectly distinguishable it means that they have disjoint supports inside $\mathcal{X}$. If 2^n probability distributions on $\mathcal{X}$ are pairwise perfectly distinguishable, we deduce that their supports Y_i are all disjoint, and therefore $d = |\mathcal{X}| \geq \sum_{i=1}^{2^n} |Y_i| \geq 2^n$. Expressed in words, this entails that in order to accommodate 2^n pairwise perfectly distinguishable states, a classical system must have dimension at least 2^n. This lower bound is trivially tight, so that the $N = 2$ compression factor of classical theories is precisely 1.

- *Quantum theory.* If 2^n quantum states are pairwise perfectly distinguishable, their supports must be pairwise orthogonal. This means that the total dimension *of the Hilbert space* is at least 2^n. Since the dimension of quantum mechanics as a GPT is the *square* of the Hilbert space dimension (cf. (2)), we see that a quantum system capable of accommodating 2^n pairwise perfectly distinguishable states must have dimension at least 2^{2n}. Again, this lower bound is easily seen to be tight, entailing that the $N = 2$ compression factor for quantum theory is precisely $1/2$.

Since their compression factors are at most 1, classical *as well as quantum* theory perform rather poorly at the task we are interested in here.

In [11] Popescu and Rohrlich famously showed that a hypothetical 'super-quantum' theory could outperform quantum mechanics at non-local tasks. However, results in [20, 21] indicate that such exotic theories may not beat quantum theory in terms of computational capacity. Here we will see how other theories fare at the task of implementing an associative memory and, in particular, seek out the optimal theory — that with the highest compression ratio defined above.

2. General probabilistic theories

Throughout this Section we will formally introduce and discuss general probabilistic theories. We point the interested reader to Ref. [13, 22, 15] for more details and a thorough operational justification of the construction described here.

We start by fixing some terminology. A subset $C \subseteq V$ of a finite-dimensional, real vector space V is called a **cone** if it is closed under positive scalar multiplication. It is called a **proper cone** if in addition it is (i) convex; (ii) salient, that is, $C \cap (-C) = \{0\}$; (iii) spanning, meaning that $C - C = V$; and (iv) topologically closed.‡

In what follows, we will denote the dual vector space to V, i.e. the space of linear functionals $V \to \mathbb{R}$, with V^*. If $C \subset V$ is a cone, we can construct its **dual cone** inside V^* as $C^* := \{f \in V^* : f(x) \geq 0 \ \forall x \in C\}$. If C is proper then so is C^*, and moreover $C^{**} = C$ modulo the canonical identification $V^{**} = V$. A functional $f \in C^*$ is also said to be positive; it is **strictly positive** if $f(x) > 0$ for all $x \in C$ with $x \neq 0$. It can be verified that strictly positive functionals are precisely those in the topological interior of C^*, denoted by $\mathrm{int}\,(C^*)$.

Definition 2 (General probabilistic theories) *A **general probabilistic theory** (GPT) is a triple (V, C, u) consisting of a real, finite-dimensional vector space V, a proper cone $C \subset V$, and a strictly positive functional $u \in \mathrm{int}\,(C^*)$, called the **order unit**. We call $d := \dim V$ the*

‡ Since we are in finite dimension, there is a unique Hausdorff topology on V, which we do not need to specify. For instance, it is induced by any Euclidean norm.

A Post-Quantum Associative Memory 5

dimension of the GPT, and $\Omega := C \cap u^{-1}(1) = \{x \in C : u(x) = 1\}$ its **state space**. A **pure state** is an extreme point§ of Ω. An **effect** is a functional $e \in V^*$ such that $e(\omega) \in [0, 1]$ for all $\omega \in \Omega$. We will denote the set of effects with $E = C^* \cap (u - C^*)$. A **measurement** is a finite collection $(e_i)_{i \in I}$ of effects $e_i \in E$ such that $\sum_{i \in I} e_i = u$.

Remark 3 The restriction to finite-dimensional spaces is made for purely technical reasons, as it simplifies the treatment considerably. However, the GPT framework makes perfect sense in infinite dimension as well — in fact, GPTs were initially conceived to accommodate also this case [23, 24, 25] (see also [13, Chapter 1]).

The state space Ω as well as the set of effects E of a given GPT are always compact convex sets. As such, they can be equivalently described as the convex hulls of their extreme points (in the case of Ω, these are just the pure states of the theory). Two extreme points of E are always 0 and the order unit u.

Note. It is worthwhile to point out some subtleties concerning the interpretation of the above definition of a GPT that should be kept in mind:

- We implicitly assume the *no restriction hypothesis* [26]. This states that all abstract measurements as constructed in Definition 2 are actually physically implementable, and entails that defining the state space Ω of a theory is sufficient to completely determine its local structure. We deem it a fairly natural assumption, since GPTs are operationally motivated in the first place – state and effect spaces can be thought of as mutually defining – and the class of restricted GPTs can do no better than the class of unrestricted GPTs for this particular task.

- We are considering only those theories with finite-dimensional state spaces (for an exploration beyond this, see [13, Chapter 1]).

- We are only dealing with the *reliable* states and effects for a theory. Operationally, this is equivalent to having preparation and measurement procedures which always behave as desired (for example, we can produce specific states deterministically).

- We are not examining non-local correlations or entanglement-like features available in different GPTs, which are often the subject of enquiry in the GPT literature [27, 28, 29, 30, 31, 8, 9]. However, although we are only considering the geometries of single systems, it is worth emphasising that these do impact upon which non-local correlations can be attained [32, 33, 31, 8, 9].

2.1. Some Example Theories

Example 4 (Classical probability theory) States in a classical probability theory are simply probability distributions over some finite alphabet $\mathcal{X}$. The corresponding GPT will have dimension $d = |\mathcal{X}|$, where $|\mathcal{X}|$ is the size of $\mathcal{X}$. Formally, it can be defined as a triple $(\mathbb{R}^d, \mathbb{R}^d_+, u)$, where $\mathbb{R}^d_+ := \{x \in \mathbb{R}^d : x_i \geq 0 \ \forall i = 1, \ldots, d\}$ is just the positive orthant, and the unit effect is a functional acting as $u(y) = \sum_{i=1}^{d} y_i$ for all $y \in \mathbb{R}^d$. The state space is therefore formed by all non-negative vectors $x \in \mathbb{R}^d_+$ such that $u(x) = \sum_i x_i = 1$; geometrically, this set is shaped as a simplex with d vertices, which we denote by $\mathcal{S}_d$.

§ An extreme point of a convex set X is a point $x \in X$ such that $x = py + (1 - p)z$ for $p \in (0, 1)$ and $y, z \in X$ implies that $y = z = x$. The set of extreme points of X will be denoted by ext(X).

A Post-Quantum Associative Memory   6

Example 5 (Quantum mechanics) The quantum mechanical theory of a k-level system can also be phrased in the GPT language. Formally, we can define it as the triple (H_k, PSD_k, Tr), where H_k is the real vector space of $k \times k$ Hermitian matrices, PSD_k is the cone of $k \times k$ positive semidefinite matrices, and Tr is the trace functional. Observe that the *real* dimension of k-level quantum mechanics is

$$\dim H_k = k^2. \tag{2}$$

Example 6 (n-gon theories) n-gon theories (sometimes referred to as polygon theories), are those in which the state space is described by a regular n-sided polygon. These theories are well studied [34, 32, 35, 36, 37], and contain the local structure of Popescu–Rohrlich boxes as a particular case ($n = 4$). Interestingly, there is a general difference between those in which n is odd and those in which n is even: for odd n, the theories are strongly self-dual, meaning that the dual cone C^* is isomorphic to C via an isomorphism mediated by a positive definite scalar product. For even n, the theories are only weakly self-dual, meaning that C and C^* are merely linearly isomorphic.

Remark 7 One particularly nice property of n-gon theories is that they give a (restricted) version of both quantum and classical theories in limiting cases. In the limiting case of $n = 2$, the polygon collapses to the line segment; this can be taken to represent a stochastic classical bit (such as a coin). In the other extreme, at $n = \infty$, the 'polygon' describes a circle — which can be thought of representing a slice through the Bloch sphere, such as the slice of states with real-valued coefficients $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, with $\alpha, \beta \in \mathbb{R}$ and $\alpha^2 + \beta^2 = 1$.

2.2. Perfect distinguishability

Now that we have a rigorous definition of GPT in place, we can also give a precise meaning to the various notions of perfect distinguishability employed in this paper. We start with the basic definition of perfect distinguishability for a set of states in a GPT. For additional details and further motivation we refer the reader to [17, 18].

Definition 8 (Perfect distinguishability) *Let (V, C, u) be a GPT with state space Ω. We say that some finitely many states $\{\omega_i\}_{i \in I} \subseteq \Omega$ are **perfectly distinguishable** if there exists a measurement $(e_i)_{i \in I}$ such that $e_i(\omega_j) = \delta_{i,j}$ for $i, j \in I$.*

We can now give a notion of mutual distinguishability for sets of states.

Definition 9 (Mutual N-wise distinguishability) *Let (V, C, u) be a GPT with state space Ω. A set of states $\mathcal{S} \subseteq \Omega$ is said to be **mutually N-wise distinguishable** if every subset $S \in \mathcal{S}$ of cardinality $|S| = N$ is perfectly distinguishable as per Definition 8. If $N = 2$ we also say that the states in $\mathcal{S}$ are **pairwise perfectly distinguishable**.*

Remark 10 The fact that $\{\omega_1, \omega_2\}$ and $\{\omega_2, \omega_3\}$ are separately perfectly distinguishable does *not* imply, in general, that $\{\omega_1, \omega_2, \omega_3\}$ are perfectly distinguishable. More generally; the union of some sets which are N-wise distinguishable is not necessarily *mutually N-wise* distinguishable itself.

In what follows we will be interested in the minimal GPT dimension that is needed in order to achieve mutually N-wise distinguishable sets with a prescribed number of elements, or, vice versa, in the maximal number of elements that a mutually N-wise distinguishable set of states can have in GPTs of a fixed dimension. We thus formalise the following definition.

Definition 11 *For two positive integers N, m, we denote with $d(N, m)$ the minimal dimension of a GPT A such that the corresponding state space Ω_A contains a set of mutually N-wise distinguishable states of cardinality 2^m. The corresponding* **compression factor** *is defined by*

$$\kappa(N, m) := \frac{m}{\log_2 d(N, m)} \, . \tag{3}$$

If we accept the assumptions leading to the GPT framework as we have defined it above, calculating or estimating $\kappa(N, m)$ from above (equivalently, calculating or estimating $d(N, m)$ from below) amounts to establishing the ultimate physical bounds to the compression of information realised by an incomplete but perfect memory. The rest of the paper is devoted to the understanding of these quantities and to their exact computation in a few interesting cases.

We start by looking at the most extreme case, that where the memory is in fact complete, i.e. the information can be retrieved. This corresponds to setting $N = 2^m$. In this case, even GPTs do not grant any advantage over classical probability theory.

Lemma 12 *For all positive integers m, it holds that $d(2^m, m) = 2^m$ and hence $\kappa(2^m, m) = 1$. In other words, there exists a GPT (namely, classical probability theory) of dimension 2^m hosting 2^m perfectly distinguishable states, but no GPT of smaller dimension enjoying that same property.*

Proof. Since perfectly distinguishable states must be linearly independent, the dimension of the host vector space of any GPT accommodating 2^m perfectly distinguishable states must be at least 2^m. $\square$

The above Lemma 12 is slightly disappointing, as it tells us that even GPTs cannot perform better than classical probability theory at the implementation of a perfect and complete memory. However, this state of affairs changes dramatically when we consider smaller values of N, i.e. when we look instead at perfect but incomplete memories. We will see how this is possible in the next Section.

3. Pairwise distinguishability

In this Section we show that a compression factor much larger than 1, and indeed of order m up to logarithmic factors, is achievable when $N = 2$. Even more, we give an exact expression for the function $\kappa(2, m)$.

Theorem 13 *For all positive integers m, it holds that $d(2, m) = m + 1$ and hence*

$$\kappa(2, m) = \frac{m}{\log_2(m + 1)} \, . \tag{4}$$

In other words, there exists a GPT of dimension $m + 1$ hosting m pairwise distinguishable states, but no GPT of dimension m or lower enjoying this same property.

The above result is remarkable because it provides an example of a task at which GPTs outperform both classical and quantum theories dramatically. In fact, as we saw in Section 1.1 the compression factor $\kappa(2, m)$ for such theories is just a constant, while Theorem 13 tells us that in the GPT world it can be made much larger, of the order of m (up to a logarithmic factor). Another notable aspect of Theorem 13 is that it does not report an estimate but

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

A Post-Quantum Associative Memory 8

rather an exact computation of the figure of merit that is of interest here, thus establishing the ultimate physical limits to this very simple type of incomplete (perfect) memory.

The discussion and proof of Theorem 13 occupies the rest of the present Section. More in detail, in Section 3.1 we discuss the simplest non-trivial case of 3-dimensional GPTs, proving with a delightfully simple argument that $d(2,2) = 3$, or equivalently $\kappa(2,2) = 2/\log_2(3)$. Section 3.2 is devoted to the presentation of the general construction that achieves the best compression factor (4) among all GPTs. In Appendix A we revisit the proof of the Danzer–Grünbaum theorem, showing that it implies directly the optimality of the above construction.

3.1. Limits in d = 3

Before commencing, we make a disclaimer on geometric terminology. We say that a hyperplane $V \subset \mathbb{R}^n$ *supports* a set X in a point $x \in V \cap X$ if V touches X in x without 'cutting through' it, in other words, if the whole X lies in one of the two closed half-spaces determined by V, with $x \in V \cap X$. Formally:

Definition 14 *Let $X \subseteq \mathbb{R}^n$ be a subset of a Euclidean space. We say that a hyperplane $V \subset \mathbb{R}^n$ supports X in a point $x \in X$ if: (i) $x \in V$; and (ii) X is entirely contained inside one of the closed half-spaces determined by V.*

Let us consider a d-dimensional GPT with state space $\Omega \subset \mathbb{R}^{d-1}$ and the set of (distinct) states $\{\rho_i\}_{i=1,\ldots,2^m} \subset \Omega$. Assume that any pair $\{\rho_i, \rho_j\}$ with $i \neq j$ is perfectly distinguishable by means of a measurement $(e_{ij}, u - e_{ij})$, as per Definition 8. Explicitly, this means that $e_{ij}(\rho_i) = 1$ and $e_{ij}(\rho_j) = 0$. Note that the set of vectors v such that $e_{ij}(v) = 0$ and the set of vectors w such that $e_{ij}(w) = 1$ form two parallel hyperplanes V and W. Note that $\rho_j \in V$ and $\rho_i \in W$. Clearly, since $0 \le e_{ij}(\omega) \le 1$ for all states ω, the whole Ω lies between V and W. We can say that W and V support the state space Ω in ρ_i and ρ_j, respectively. Vice versa, this condition is entirely equivalent to ρ_i and ρ_j being perfectly distinguishable. To get a clear geometric intuition it is instructive to explore the special case where the state space is 2-dimensional; with our convention, this corresponds to the case where $d = 3$, because the global GPT will feature a 3-dimensional cone whose section is our 2-dimensional state space.

We thus consider a 3-dimensional GPT with states confined to a set $\Omega \subset \mathbb{R}^2$. The situation is as depicted in Figure 1. The two states $\rho_i, \rho_j \in \Omega$ in Figure 1 are indeed perfectly distinguishable, because the entire set Ω is enclosed between two parallel lines supporting it in ρ_i and ρ_j, respectively. However, one can see that not all pairs among the 6 states marked with black dots can be perfectly distinguishable.‖

Let us make this discussion a bit more rigorous. Assume that we are given k states $\rho_1, \ldots, \rho_k \in \Omega$, with the promise that they are pairwise perfectly distinguishable. We can ask ourselves: *how large can k be?* The convex hull of $\rho_1, \ldots, \rho_k$ will naturally form a polygon $P \subseteq \Omega$. In fact, we have that every ρ_i must correspond to a vertex of P in order for the perfect distinguishability condition to be obeyed. Consider now two neighbouring vertices ρ_i, ρ_{i+1} of P, as well as the edge connecting them. Call α_i, α_{i+1} the internal angles of P at vertices ρ_i, ρ_{i+1}. It can be shown that, in order for ρ_i, ρ_{i+1} to be perfectly distinguishable, it has to hold that $\alpha_i + \alpha_{i+1} \le \pi$ (cf. Figure 2). Summing over $i = 1, \ldots, k$, with the convention that $k+1 \equiv 1$, we obtain that

$$k\pi \ge \sum_{i=1,\ldots,k} (\alpha_i + \alpha_{i+1}) = 2\sum_i \alpha_i. \tag{5}$$

‖ Indeed, for example the two dots at the bottom of the grey figure are not. This makes sense, because we see from Theorem 13 that in dimension 3 there can be at most $2^{3-1} = 4$ states with such property.
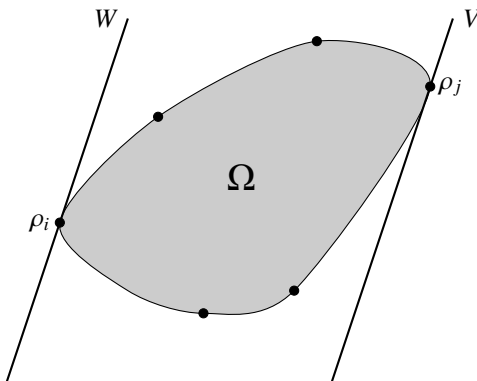
Figure 1: A set of states in a two-dimensional GPT with state space Ω — here we are viewing the state space top-down. If ρ_i and ρ_j are perfectly distinguishable then the two lines W and V *support* the state space in ρ_i and ρ_j, respectively.

The sum on the right-hand side is just the sum of all internal angles of a convex polygon with k vertices. From elementary geometry, this is well known to be $(k-2)\pi$. Therefore, we obtain the inequality

$$k\pi \geq 2(k-2)\pi\,, \tag{6}$$

which yields immediately $k \leq 4 = 2^2$, in line with Theorem 13. This bound is tight, because the four vertices of a square state space correspond to pairwise perfectly distinguishable states — a more general version of this latter statement will be proved in the next Section.
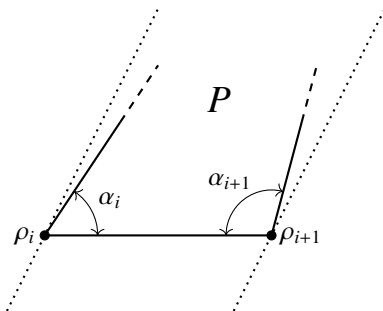

Figure 2: A geometric sketch of a possible proof for the simplest non-trivial case $N = 2$, $d = 3$.

3.2. Generalisation to arbitrary dimension and optimality

We now set out to generalise the analysis to any dimension.¶ In light of the geometric construction discussed in Section 3.1 (the situation is entirely analogous to that depicted in Figure 1 for $d = 3$), we can reformulate our problem as follows:

¶ In a related spirit, theories using hyperspheres of generalised dimension, so-called D-balls, are discussed in [38, 39]; the authors aim to isolate the 3-sphere as a the necessary state space for quantum theory based on physical requirements. See also [6] for a different use of spherical theories.

> **Problem (reformulation).** Determine the minimum $d_m = d(2, m)$ such that there exists a set $X = \{\rho_1, \ldots, \rho_{2^m}\} \subset \mathbb{R}^{d_m-1}$ with the following property: for any two distinct $\rho_i, \rho_j \in X$, there are two parallel hyperplanes $W, V \subset \mathbb{R}^{d_m-1}$ of which one supports X in ρ_i and the other supports X in ρ_j.

We now explain how to achieve a construction with the above properties in dimension $d = m + 1$. The argument is quite simple, and it is worthwhile explaining it in words before delving into the mathematical formalism. The state space of the GPT we pick to achieve the bound is shaped as a hypercube of dimension m. Since the whole theory includes also multiples of normalised spaces, its dimension is in fact $m + 1$. The 2^m states we choose correspond to the vertices of the hypercube. The crucial point now is that any two distinct vertices will be sitting each on one of two parallel hyperplanes that enclose the whole state space. Those hyperplanes, that are spanned by two opposite faces of the hypercube, will define the binary measurement needed to discriminate the states in question. This bit of reasoning already shows that any two vertices of the hypercube indeed represent perfectly distinguishable states.

We now make this argument rigorous. Construct the GPT $\left(\mathbb{R}^{m+1}, C_{g,m}, u \right)$, where

$$C_{g,n} := \left\{ (x_0, x_1, \ldots, x_n)^\intercal \in \mathbb{R}^{m+1} : \; x_0 \geq \max_{1 \leq i \leq n} |x_i| \right\} \tag{7}$$

and moreover $u\left((x_0, x_1, \ldots, x_m)^\intercal \right) := x_0$. The state space of this GPT is clearly a hypercube of dimension m. Now, for $\epsilon \in \{\pm 1\}^n$, define $\rho_\epsilon \in \mathbb{R}^{m+1}$ by

$$(\rho_\epsilon)_i := \begin{cases} 1 & \text{if } i = 0, \\ \epsilon_i & \text{if } i \geq 1. \end{cases} \tag{8}$$

Note that there are exactly 2^m distinct choices for ϵ. We deduce that the 2^m states ρ_ϵ are pairwise perfectly distinguishable. To see why, consider $\epsilon, \epsilon' \in \{\pm 1\}^n$ that are distinct. Then, they will differ at some position $i \in \{1, \ldots, m\}$. Without loss of generality, we can assume that $\epsilon_i = +1$ and $\epsilon'_i = -1$. Now, consider the two-element collection $(e_i, u - e_i)$, where the functional e_i is defined by

$$e_i \left((x_0, x_1, \ldots, x_n)^\intercal \right) := \frac{x_0 + x_i}{2} \, . \tag{9}$$

Note that for all $x \in C_{g,m}$ we have that $0 \leq e_i(x) \leq u(x)$; hence, the collection $(e_i, u - e_i)$ defines a binary measurement. It is now elementary to verify that

$$e_i(\rho_\epsilon) = 1 \, , \quad (u - e_i)(\rho_\epsilon) = 0 \, , \quad e_i(\rho_{\epsilon'}) = 0 \, , \quad (u - e_i)(\rho_{\epsilon'}) = 1 \, . \tag{10}$$

These are precisely the conditions needed to ensure that ρ_ϵ and $\rho_{\epsilon'}$ are perfectly distinguishable.[+]

We have therefore constructed a GPT of dimension $m + 1$ which is capable of accommodating 2^m pairwise perfectly distinguishable states; hence,

$$d(2, m) \leq m + 1 \, , \qquad \text{or} \qquad \kappa(2, m) \geq \frac{m}{\log_2(m + 1)} \, . \tag{11}$$

[+] These 'hypercubic' theories have been employed in a similar spirit by Ver Steer and Wehner [2, Claim 6.2] to construct superior random access codes.

What is remarkable here is that those limits far exceed the capabilities of both classical and quantum theory — each of these have an exponential scaling in the number of dimensions required to store m bits, whereas hypercubic theories scale only linearly. Equivalently, the compression factor for the case $N = 2$ of pairwise perfect distinguishability is at most 1 for classical and quantum theory, but scales almost linearly in m (up to logarithmic factors) for the best conceivable GPT. This demonstrates a sort of *exponential advantage* of general GPTs over classical and quantum theories.

It remains to show that the above construction is optimal. From the mathematical standpoint, this is highly non-trivial. To overcome this hurdle, we exploit the reformulation of the problem presented in Section 3.2: in that form, the problem was posed for the first time by Klee [40] and was solved not long after by Danzer and Grünbaum [16]. Their solution shows that $d_m = m + 1$ is a minimum for any m. We restate their result for our convenience below.*

Theorem 15 (Danzer–Grünbaum [16]) *For a positive integer n, the maximum cardinality of a set $X \subset \mathbb{R}^n$ such that for any two distinct $x_1, x_2 \in X$ there are two parallel hyperplanes $V_1, V_2 \subset \mathbb{R}^n$ with the property that V_i supports X in x_i ($i = 1, 2$) is precisely 2^n. This cardinality is achieved by the set of vertices of a hypercube.*

Note that the above result is entirely equivalent, via the problem reformulation in Section 3.2, to Theorem 13. For the interested reader, in Appendix A we present a brief but self-contained account of, and homage to, the beautiful proof by Danzer and Grünbaum [16]; see also [41, Chapter 17].

4. Perfect distinguishability beyond pairwise: Asymptotic results

In the previous Section we have established the maximum number of pairwise perfectly distinguishable states which can be housed in a GPT, and hence the limits to the capacity of an associative memory of the type described in our introduction, in the case where $N = 2$. The situation is much less clear for $N \geq 3$, in which case we cannot exhibit an explicit expression for $\kappa(N, m)$, nor a tight general estimate. However, in Theorem 17 below we determine the exact asymptotics in m for every fixed N. Before we do so, it is instructive to see how a naïve generalisation of the hypercube construction actually fails to yield an exact computation of $\kappa(N, m)$.

4.1. A naïve generalisation and its fall

At first, we could hope that a simple generalisation of the hypercube construction may work. To explain how to obtain such a generalisation, we start by observing that from the geometric standpoint a hypercube can equivalently be seen as a Cartesian product of segments. Indeed, the extreme points of a simple line segment can be thought of as having coordinates (± 1). A square, of having all four combinations of $(\pm 1, \pm 1)$, and so on for cubes, and hypercubes in any dimension. This operation of combining vertices by concatenating their coordinates corresponds precisely to the geometric construction of the Cartesian product. Such a construction can be translated into the world of GPTs in a fully general fashion, giving rise to the notion of *prism theories*, which we explore in more detail in Appendix B.

Noticing this, we could be tempted to conjecture that Theorem 13 could be extended to any N in the naïve way, i.e. that the extreme states of a GPT with state space $\mathcal{S}_N^{\times l}$, the

---

* Our poor knowledge of German meant that we employed a translation of the original paper, realised by Rolf Schneider.

A Post-Quantum Associative Memory 12

l-fold Cartesian product of the N-vertex simplex, form a mutually N-wise distinguishable set. However, we can quickly see that this is not the case, and that the relationship between simplex structure and the size of mutual distinguishability does not extend beyond $N = 2$. We show this with an example:

Example 16 Call $\rho_1, \ldots, \rho_q$ the vectors of the canonical basis of $\mathbb{R}^q$, thought of as states in the classical GPT $(\mathbb{R}^q, \mathbb{R}^q_+, u)$ described in Example 4. Explicitly, we will have $\rho_i = (0, \ldots, 1, \ldots 0)^\intercal$, where the single entry equal to 1 is in the ith position. Then the extremal (pure) states of the l-fold product $\mathcal{S}_q^{\times l}$ are of the form $(\rho_{i_1}, \ldots, \rho_{i_l})$, where $i_1, \ldots, i_l \in \{1, \ldots, q\}$. Consider the 3 states

$$\omega_1 := (\rho_1, \ldots, \rho_1), \qquad \omega_2 := (\rho_1, \rho_2, \rho_1, \ldots, \rho_1), \qquad \omega_3 := (\rho_2, \rho_1, \ldots, \rho_1). \qquad (12)$$

Then $\omega_1, \omega_2, \omega_3$ are not jointly distinguishable. In fact, note that

$$\omega_2 + \omega_3 = (\rho_1 + \rho_2, \rho_2 + \rho_1, \rho_1, \ldots, \rho_1) \geq \omega_1. \qquad (13)$$

Thus, if $e_1 \cdot \omega_2 = e_1 \cdot \omega_3 = 0$, then also $e_1 \cdot \omega_1 = 0$. In other words, there cannot be a measurement singling out ω_1 from this triple of states.

4.2. Asymptotics in m for fixed N

Although no simple generalisations of the exact computation in Theorem 13 are available, we can obtain a general result that guarantees that for fixed N and very large m, the scaling of the compression factor $\kappa(N, m)$ in m is *exactly the same* as that given by Theorem 13. In other words, the scaling of $\kappa(N, m)$ in m for a fixed N does not depend on N. To prove this somewhat surprising result, we will make use of a probabilistic argument, while we leave open the task of finding a constructive proof of the result below.

Theorem 17 *For all fixed integers $N \geq 2$, it holds that*

$$\lim_{m \to \infty} \frac{\kappa(N, m)}{\kappa(2, m)} = \lim_{m \to \infty} \frac{\kappa(N, m)}{m / \log_2(m)} = 1. \qquad (14)$$

In particular, for every fixed $N \geq 2$ we have that

$$d(N, m) \leq m^{1 + o_N(1)} \qquad (m \to \infty). \qquad (15)$$

Proof. Clearly, if a set of GPT states is mutually N-wise distinguishable for some $N \geq 2$, it is also 2-wise distinguishable (i.e. we have pairwise perfect distinguishability). Hence, $\kappa(N, m) \leq \kappa(2, m) = \frac{m}{\log_2(m+1)}$, entailing that $\limsup_{m \to \infty} \frac{\kappa(N, m)}{\kappa(2, m)} \leq 1$.

To establish the lower bound, it suffices to prove (15), because that would imply that

$$\frac{\kappa(N, m)}{m / \log_2(m)} = \frac{\log_2(m)}{\log_2 d(N, m)} \geq 1 - o_N(1) \xrightarrow[m \to \infty]{} 1. \qquad (16)$$

Now, to establish (15) we need to find, for fixed N and very large m, an example of a GPT of dimension $m^{1 + o_N(1)}$ that can accommodate a mutually N-wise distinguishable set of states of cardinality approximately 2^m. Consider the GPT with state space $\mathcal{S}_q^{\times l}$ of Example 16. For now, q and l are free parameters, to be fixed later. Let us now draw states at random in a i.i.d. fashion from $\mathcal{S}_q^{\times l}$. Every state, of the form $\omega = (\rho_{i_1}, \ldots, \rho_{i_l})$, is in turn constructed by drawing

A Post-Quantum Associative Memory   13

$i_1, \ldots, i_l \in \{1, \ldots, q\}$ uniformly at random, again in an i.i.d. manner. We now ask ourselves: given N random states $\omega^{(1)}, \ldots, \omega^{(N)} \in \mathcal{S}_q^{\times l}$, when are they distinguishable by looking at the first component, which we will also express by saying that the first component *discriminates* them? (Remember that each $\omega^{(j)}$ has l components, because it belongs to $\mathcal{S}_q^{\times l}$.) The answer is clear: whenever the first components of $\omega^{(1)}, \ldots, \omega^{(N)}$ are all different. This happens with probability

$$\Pr\left\{1^{\text{st}} \text{ component discriminates } \omega^{(1)}, \ldots, \omega^{(N)}\right\} = \prod_{k=1}^{N-1}\left(1 - \frac{k}{q}\right), \qquad (17)$$

because $\prod_{k=1}^{N}\left(1 - \frac{k}{q}\right)$ is the probability that N random numbers between 1 and q (representing the first components of $\omega^{(1)}, \ldots, \omega^{(N)}$) are all different. Hence,

$$\Pr\left\{1^{\text{st}} \text{ component does not discriminate } \omega^{(1)}, \ldots, \omega^{(N)}\right\} = 1 - \prod_{k=1}^{N-1}\left(1 - \frac{k}{q}\right). \qquad (18)$$

Since we can look at any component of our choice, there are l of them, and these are all independent,

$$\Pr\left\{\text{no component discriminates } \omega^{(1)}, \ldots, \omega^{(N)}\right\}$$
$$= \left(\Pr\left\{1^{\text{st}} \text{ component does not discriminate } \omega^{(1)}, \ldots, \omega^{(N)}\right\}\right)^l \qquad (19)$$
$$= \left(1 - \prod_{k=1}^{N-1}\left(1 - \frac{k}{q}\right)\right)^l.$$

So far we have only considered one N-tuple of states. If we draw a subset $\mathcal{M} \subset \mathcal{S}_q^{\times l}$ of $M = |\mathcal{M}| = 2^m$ states in total, there are $\binom{M}{N}$ distinct such N-tuples (up to re-ordering). Therefore, the probability that at least one of them is such that no component discriminates it is

$$\Pr\left(\bigcup_{\omega^{(1)}, \ldots, \omega^{(N)} \in \mathcal{M} \text{ distinct}} \left\{\text{no component discriminates } \omega^{(1)}, \ldots, \omega^{(N)}\right\}\right)$$
$$\leq \binom{M}{N}\Pr\left\{\text{no component discriminates } \omega^{(1)}, \ldots, \omega^{(N)}\right\} \qquad (20)$$
$$= \binom{M}{N}\left(1 - \prod_{k=1}^{N-1}\left(1 - \frac{k}{q}\right)\right)^l.$$

As long as we can guarantee that the rightmost side of (20) stays below, say, $1/2$, we will know that there exists a choice of $\mathcal{M}$ such that for every distinct $\omega^{(1)}, \ldots, \omega^{(N)} \in \mathcal{M}$, some component will discriminate them. Hence, we will have implicitly constructed a mutually N-wise distinguishable set $\mathcal{M}$ — this is, of course, an instance of the celebrated probabilistic method [42, 43].

If the above condition is met, then, noting that $l(q - 1) + 1$ is the dimension of the GPT $\mathcal{S}_q^{\times l}$, we will have that

$$\frac{\kappa(N, m)}{m/\log_2(m)} \geq \frac{\frac{m}{\log_2(l(m)(q(m)-1)+1)}}{m/\log_2(m)} = \frac{\log_2(m)}{\log_2\left(l(m)(q(m) - 1) + 1\right)}. \qquad (21)$$

A Post-Quantum Associative Memory 14

We now proceed to identify a scaling of the parameters $l = l(m)$ and $q = q(m)$ such that not only the rightmost side of (20) converges to 0, but also the rightmost side of (21) converges to 1, i.e.

$$\frac{\log_2(m)}{\log_2\left(l(m)(q(m)-1)+1\right)} \xrightarrow[m\to\infty]{} 1 \,. \tag{22}$$

Doing so would conclude the proof, because we would have implicitly constructed a sequence of mutually N-wise distinguishable sets of cardinality 2^m in a GPT of dimension $D(m)$, such that $m/\log_2 D(m) \approx m/\log_2(m)$ for large m.

To this end, set

$$q(m) \sim \left(\log_2(m)\right)^2, \qquad l(m) \sim \frac{2Nm}{\log_2\left(\frac{2q(m)}{N(N-1)}\right)} \,, \tag{23}$$

where $\sim$ simply indicates that the ratio tends to 1 as $m \to \infty$. The precise values of $q(m)$ and $l(m)$ are not important, what we really care about is the scaling in m. Note that (22) is clearly satisfied, because $\log_2\left(l(m)q(m)\right) = \log_2(m) + O\!\left(\log_2\left(\log_2 m\right)\right)$. As for the rightmost side of (20), note that

$$\log_2\left(\binom{M}{N}\left(1 - \prod_{k=1}^{N-1}\left(1 - \frac{k}{q(m)}\right)\right)^{l(m)}\right) \le Nm\left\{1 + \frac{l(m)}{Nm}\log_2\left(1 - \prod_{k=1}^{N-1}\left(1 - \frac{k}{q(m)}\right)\right)\right\}, \tag{24}$$

where we used the crude approximation $\binom{M}{N} \le M^N$. We will now prove that

$$\frac{l(m)}{Nm}\log_2\left(1 - \prod_{k=1}^{N-1}\left(1 - \frac{k}{q(m)}\right)\right) \xrightarrow[m\to\infty]{} -2 \,, \tag{25}$$

implying that the right-hand side of (24) tends to $-\infty$, which would in turn entail that the rightmost side of (20) converges to 0, and hence would conclude the proof. Now, observing that

$$1 - \prod_{k=1}^{N-1}\left(1 - \frac{k}{q(m)}\right) = (1 + o(1))\sum_{k=1}^{N-1}\frac{k}{q(m)} = (1 + o(1))\frac{N(N-1)}{2q(m)} \tag{26}$$

and moreover $l(m)/m \xrightarrow[m\to\infty]{} 0$, we see that

$$\frac{l(m)}{Nm}\log_2\left(1 - \prod_{k=1}^{N-1}\left(1 - \frac{k}{q(m)}\right)\right) = \frac{l(m)}{Nm}\log_2\left(\frac{N(N-1)}{2q(m)}\right) + \frac{l(m)}{Nm}\log_2(1 + o(1))$$
$$= -2 + o(1)\,, \tag{27}$$

as claimed. □

The above Theorem 17 is the most general result that we could prove analytically (albeit not constructively) in this Section. At this point, it is wise to pause for a moment our search for mutual N-distinguishable sets and ask ourselves a basic question: how do we decide whether a given a set of states is jointly perfectly distinguishable?

A Post-Quantum Associative Memory　　　　　　　　　　　　　　　　　　15

5. Perfect distinguishability beyond pairwise: Numerical methods

5.1. Perfect distinguishability as a convex program

We record here the simple observation that not only the question of perfect distinguishability, but actually the calculation of the minimal error probability in joint discrimination of a set of states in a given GPT is in fact a convex program [44]. This is particularly interesting and useful, as in many situations arising naturally in applications the underlying cone admits an efficient description in terms of linear inequalities, or else in terms of inequalities in the Löwner partial order, i.e. the one determined by positive semi-definiteness. The former is the case, for instance, for classical theories (Example 4). A description in terms of positive semi-definite constraints, instead, can be formulated not only for quantum theory itself (Example 5), but also for several GPTs that are of great interest in entanglement theory [45]. Notable examples in this context include the theory of NPT entanglement [46, 47, 48, 49, 50] and that of extendibility [51, 52, 53, 54, 55].

Lemma 18 *Let (V, C, u) be a d-dimensional GPT with state space Ω. Given states $\{\omega_i\}_{i=1}^{N} \subset \Omega$ and a priori probabilities $\{p_i\}_{i=1}^{N}$, the maximal success probability in the associated task of state discrimination is given by the convex program*

$$
\begin{aligned}
P_s^{\max}\left(\{p_i, \omega_i\}_{i=1}^{N}\right) \quad = \quad &\max \quad \sum_{i=1}^{N} p_i\, e_i \cdot \omega_i \\
&\text{s.t.} \quad e_1, \dots, e_N \in C^*, \ \sum_{i=1}^{N} e_i = u\,.
\end{aligned}
\tag{28}
$$

If C is polyhedral with M extremal rays, i.e. if there exist finitely many $v_1, \dots, v_M \in V$ such that $C = \left\{ \sum_{j=1}^{M} a_j v_j : a_j \geq 0 \ \forall\, j \right\}$, then (28) can be rephrased as a linear program, namely,

$$
\begin{aligned}
P_s^{\max}\left(\{p_i, \omega_i\}_{i=1}^{N}\right) = &\max \ \sum_{i=1}^{N} p_i\, e_i \cdot \omega_i \\
&\text{s.t.} \ \ e_i \cdot v_j \geq 0 \quad \forall\, i = 1, \dots, N, \quad \forall\, j = 1, \dots, M, \ \sum_{i=1}^{N} e_i = u\,.
\end{aligned}
\tag{29}
$$

The above program can be solved efficiently, in time $O\left(d(d + M)^{3/2} N^{5/2}\right)$.

Proof. The most general state discrimination procedure consists of making a measurement $(e_i)_{i=1,\dots,N}$, and guessing the unknown state to be ω_i upon having obtained outcome i. The average probability of success of this strategy is precisely $\sum_i p_i\, e_i \cdot \omega_i$. The constraints in (28) are those required to make sure that $(e_i)_{i=1,\dots,N}$ is in fact a valid measurement in the GPT (V, C, u).

If C is polyhedral with M extremal rays spanned by vectors $v_1, \dots, v_M$, then naturally $e \in V^*$ satisfies that $e \in C^*$ if and only if $e \cdot v_j \geq 0$ for all $j = 1, \dots, M$. In this way one derives (29) from (28). Finally, the estimates on the efficiency of the linear program solution are taken from the work by Vaidya [56]. To make the comparison precise, note that in our case $n = d(N - 1) \sim dN$ is the number of real variables♯ and $m = NM$ is the number of constraints. □

Based on the above result, we can state its implications for the problem of perfect discrimination, which is of interest here:

♯ We have d variables for each of the vectors $e_1, \dots, e_{N-1}$ living in a d-dimensional space V^*. Note that e_N is uniquely determined by the normalisation condition $\sum_i e_i = u$.

Corollary 19 *Let (V, C, u) be a d-dimensional GPT with state space Ω. Deciding whether the states $\{\omega_i\}_{i=1}^{N} \subset \Omega$ are perfectly distinguishable is a convex feasibility problem [44]:*

$$
\begin{array}{ll}
\text{find} & e_1, \ldots, e_N \in V^* \\
\text{s.t.} & e_1, \ldots, e_N \in C^*, \quad \sum_{i=1}^{N} e_i = u, \quad e_i \cdot \omega_i = 1 \ \forall \ i = 1, \ldots, N.
\end{array} \tag{30}
$$

If C is polyhedral with M extremal rays then the above program becomes linear, and can be solved in time at most $O\left(d(d + M)^{3/2} N^{5/2}\right)$.

5.2. Restricting the search

Consider for simplicity a GPT (V, C, u) whose cone is polyhedral. Thanks to Lemma 18, we know that whether a given set of states $\{\omega_i\}_i$ can be discriminated perfectly can be decided efficiently. But how do we start searching for a maximal mutually N-wise distinguishable set of states? Before we proceed to answer this, we first show that the search can be restricted to pure states, i.e. to extremal points of the state space.

Lemma 20 *Let (V, C, u) be a d-dimensional GPT with state space Ω. Given some $N \geq 2$, a mutually N-wise distinguishable set can be searched among pure states, i.e. extremal points of Ω.*

Proof. Let us assume that a mutually N-wise distinguishable set $\mathcal{S} \subseteq \Omega$ has been found. Every $\omega \in \mathcal{S}$ will admit a (not necessarily unique) decomposition of the form $\omega = \sum_{i=1}^{d} p_i^{\omega} \varphi_i^{\omega}$, where $\varphi_i^{\omega} \in \Omega$ are pure states. Let us pick i such that $p_i^{\omega} > 0$, and consider the associated pure state φ_i^{ω}. Repeating this procedure for every $\omega \in \mathcal{S}$, we can form a set of pure states $\mathcal{S}' = \{\varphi_i^{\omega} : \omega \in \mathcal{S}, \ p_i^{\omega} > 0\}$.

We claim that also $\mathcal{S}'$ is mutually N-wise distinguishable. To see why, pick some pure states $\varphi_{i_1}^{\omega_1}, \ldots, \varphi_{i_N}^{\omega_N} \in \mathcal{S}'$, and consider the corresponding states $\omega_1, \ldots, \omega_N \in \mathcal{S}$. Let $(e_j)_{j=1,\ldots,N}$ be the measurement that achieves perfect discrimination of the set $\{\omega_j\}_j$, i.e. such that $1 = e_j \cdot \omega_j = \sum_i p_i^{\omega_j} e_j \cdot \varphi_i^{\omega_j}$ for all j. We immediately deduce that $e_j \cdot \varphi_i^{\omega_j} = 1$ for all i and j such that $p_i^{\omega_j} > 0$, and in particular $e_j \cdot \varphi_{i_j}^{\omega_j} = 1$ for all j. This implies that the states $\varphi_{i_1}^{\omega_1}, \ldots, \varphi_{i_N}^{\omega_N}$ are perfectly distinguishable by means of the measurement $(e_j)_{j=1,\ldots,N}$. $\qquad\square$

We could wonder whether a similar restriction applies to the measurements as well, i.e. whether it suffices to restrict the search to extremal effects. After all, if $e \cdot \omega = 1$ and $e = \sum_i p_i^e f_i$ with f_i extremal effects, it follows that $f_i \cdot \omega = 1$ whenever $p_i > 0$; we could therefore imagine to replace e with any f_i such that $p_i > 0$. The reason why this does not work, however, is that doing so in general alters the sum of all the effects, which needs to be equal to the order unit. This means that in general restricting to extremal effects is not guaranteed to yield all possible feasible measurements. We construct an example to demonstrate this in Appendix C.

Nonetheless, the fact that we can restrict ourselves to the finite set of pure states makes our search for the largest N-wise mutually distinguishable set of states much easier to approach. Restricting ourselves to extremal effects would have been useful in that it would have enabled us to simplify the search for distinguishing measurements, but the convex approach described in Section 5.1 serves perfectly well to that purpose. The restriction to pure states, on the other hand, means that the next component of our search can take place on the terrain of a finite, rather than infinite, set.

A Post-Quantum Associative Memory   17

5.3. Finding the largest N-wise mutually distinguishable set of states

Our search can be split into two distinct steps:

- **Joint distinguishability:** For a given GPT with a state space Ω, discover all subsets of $\text{ext}(\Omega)$ which are N-wise distinguishable. For example, if we had $N = 3$, we would be finding all triples of pure states which were distinguishable by a tripartite measurement. We call such sets θ^i, and the set of such sets $\Theta = \{\theta^i\}$.

- **Mutual distinguishability:** Find the largest set $\Phi \subseteq \text{ext}(\Omega)$ such that every subset $\phi \subset \Phi$ of cardinality $|\phi| = N$ is also an element of Θ, $\phi = \theta^j$. This means that any N size subset of Φ is N-wise distinguishable — or, equivalently, that Φ is *N-wise mutually distinguishable*.

The first of these steps can be achieved using the methods described in Section 5.1 using the following straightforward routine:

Algorithm 1: Finding the set Θ of N-wise distinguishable subsets

---
$\Theta = \emptyset$
$\Gamma = \texttt{Combinations}(\text{ext}(\Omega), N)$
for $\theta \in \Gamma$ **do**
   if *Conv(θ)* **then**
      | $\Theta := \Theta \cup \theta$
   end
end

In which $\texttt{Combinations}(A, N)$ is a function which yields all distinct subsets of A which have cardinality N; and $\texttt{Conv}(\theta)$ refers to the program outlined in Section 5.1, returning $\texttt{TRUE}$ if its input is jointly distinguishable, and $\texttt{FALSE}$ otherwise.

Now, with our set of jointly distinguishable sets Θ in hand, we can proceed to finding the largest N-wise mutually distinguishable set Φ. Given that we know the elements of Θ, we know all groups of states which are N-wise distinguishable. We can think of this relationship between states — that of being N-wise jointly distinguishable — as a connection between them. In fact, we can take this logic literally; we can construct a hypergraph overlaying our state space. Formally, recall that an (undirected) hypergraph is a pair (V, E), where V is a (finite) set of so-called nodes, and E is a subset of the power set of V, i.e. a collection of subsets of V. We refer to the elements of E as hyperedges, and to E itself as the hyperedge set. A hypergraph is called N-regular if each hyperedge has cardinality precisely N.

In the hypergraph we construct, the vertices correspond to the states we are considering (typically the pure states of the theory), and the hyperedges are all the subsets of N states that are perfectly distinguishable. Note that if $N = 2$ then every edge connects two vertices, yielding an ordinary graph.

Definition 21 (Distinguishablity hypergraphs) *Given some integer $N \geq 2$ and a GPT with state space Ω and finitely many pure states, i.e. such that $|\text{ext}(\Omega)| < \infty$, the N-distinguishability hypergraph of Ω, denoted $\mathscr{G}(\Omega; N)$, is the N-regular hypergraph with node set $\text{ext}(\Omega)$ and hyperedges given by all subsets of $\text{ext}(\Omega)$ of cardinality N which are perfectly distinguishable in Ω according to Definition 8.*

Above, we described the task of finding Φ as follows:

Find the largest set $\Phi \subseteq \text{ext}(\Omega)$ such that every subset $\phi \subset \Phi$ of cardinality $|\phi| = N$ is also an element of Θ, $\phi = \theta^j$.

A Post-Quantum Associative Memory   18

With the graph-theoretic view of our problem in mind, we can re-formulate this problem as follows:

> What is the largest sub-graph $\mathscr{G}'$ of $\mathscr{G}(\Omega; N)$ which is N-complete, in the sense that every subset of nodes of $\mathscr{G}'$ of cardinality N is a hyperedge?

This being a particular phrasing of the well-known *maximum clique problem*. Or, to be more precise; in our case we seek the maximum N-clique on an N-regular hypergraph. On ordinary graphs (ordinary in the sense that they are not hypergraphs), the problem is well studied [57, 58, 59, 60, 61], and algorithms are known both for exact solutions, and for faster, inexact solutions — a review appears in [62]. This problem is known to be **NP**-complete [63].

In the case of hypergraphs, however, less is known. We have developed an exact algorithm, based on an existing one called `hClique` [64].

Algorithm 2: Finding the largest N-wise mutually distinguishable set $\mathbf{\Phi}$

$\mathbf{\Phi} = \emptyset$
for $\theta \in \mathbf{\Theta}$ **do**
$\quad Q := \theta$
$\quad$ **for** $\omega \in \text{ext}(\Omega) \setminus Q$ **do**
$\quad\quad T = \text{True}$
$\quad\quad$ **for** $a \in Q$ **do**
$\quad\quad\quad Q' = (Q \setminus a) \cup \omega$
$\quad\quad\quad$ **if** $Q' \notin Q$ **then**
$\quad\quad\quad\quad T = \text{False}$
$\quad\quad\quad$ **end**
$\quad\quad$ **end**
$\quad\quad$ **if** T **then**
$\quad\quad\quad Q := Q \cup \omega$
$\quad\quad$ **end**
$\quad$ **end**
$\quad$ **if** $|Q| > |\mathbf{\Phi}|$ **then**
$\quad\quad \mathbf{\Phi} := Q$
$\quad$ **end**
end

This algorithm works by examining each edge in turn, and finding the largest clique Q branching out from the nodes on that edge. In order to do this, we begin with an edge θ^i, and set the initial clique to the nodes connected by that edge θ^i. We then examine the set of nodes not included in θ^i, $\Omega' = \text{ext}(\Omega) \setminus \theta^i$. Then, for each $\omega \in \Omega'$, we check if ω is fully connected to Q. If it is, then it can be added to Q, and the clique can grow.

Definition 22 (Fully connected cliques) *Let Ω be the state space for a GPT (V, C, u), and assume that Ω has finitely many pure states, i.e. that $|\text{ext}(\Omega)| < \infty$. Let $\mathscr{G}(\Omega; N)$ be the N-distinguishability hypergraph of Ω, as per Definition 21. Let Q be a sub-graph of $\mathscr{G}$. A node ω is fully N-connected to Q iff: Given $Q' = (Q \setminus a) \cup \{\omega\}$, $\exists \theta^j \in \Theta : \theta^j = q_l \; \forall q_l \in Q'$, where $|q_l| = N$; and this is true $\forall a \in Q$.*

In words: we remove an element a from our clique and replace it with our candidate node ω (giving the set Q'). We then create every sub-group of size N from Q' (these are $\{q_l\}$). If every one of these sub-groups matches an edge $\theta \in \Theta$, then we can say that ω is fully connected to $Q \setminus a$. If this is true for all $a \in Q$, then ω is fully connected to Q. Now, if this is the case, we can add ω to the clique, $Q := Q \cup \{\omega\}$.

A Post-Quantum Associative Memory   19
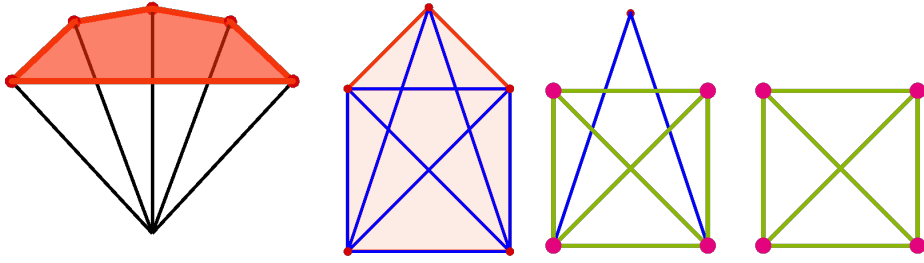

Figure 3: Illustration of how we would isolate the maximum *N*-clique for a simple state space. From left to right: The first image the state space of a hypothetical GPT. In the next image, we discard the third dimension and connect the states which are pairwise distinguishable using blue lines. The third image then isolates those states which form a clique: all four of the larger, pink vertices are connected to one another by blue edges, meaning they are *mutually pairwise distinguishable*. The top state is only connected to two of these, so is excluded from the clique. Since no node can be added to this clique, it is maximal. In this case this clique is also the maximum clique. The rightmost and final image isolates this maximum clique as a graph of four states.

     This process will discover the largest clique which can be built out from each edge; this is the set of maximal cliques. The largest of these will be the *maximum clique*, our object of interest.

     If we take $N = 2$, we have the simpler problem of finding the maximum clique on a (non-hyper) graph. Note that this is the case even for high-dimensional state spaces, because the dimensionality of the hypergraph described in Definition 21 depends only upon the number of states connected by each edge (perfectly distinguishable through a single measurement), not upon the dimension of *V* itself. The case depicted in Figure 3 is two dimensional in two senses: the original state space occupies a two-dimensional surface embedded in $\mathbb{R}^3$, and the graph formed from it can be represented on the plane, since each edge is a line.

     As stated above, the problem of discovering the largest set of *N*-wise mutually distinguishable states for a given GPT — equivalent to finding the maximum *N*-clique for an undirected hypergraph — may not be amenable to a closed analytical solution in general. Even though we do not yet know which theories would be optimal for associative memories in the case that $N > 2$, our results in this Section establish an exact method for probing candidate theories, in any dimension and for any *N*.

6. Discussion

In this paper we discussed a simple model of associative memory, in the form of a GPT system capable of being in any one of 2^m states in such a way that any *N* of them are perfectly distinguishable. When $N = 2$, we could characterise precisely the GPTs performing optimally at this task: they are theories whose state space is shaped as an *m*-dimensional hypercube. We proved in Theorem 13 that such theories outperform classical and quantum

A Post-Quantum Associative Memory 20

theories exponentially, in the sense that they have dimension $d(2, m) = m + 1$, while any classical or quantum system with the same properties needs to have dimension $O(2^m)$. We extended our analysis to the asymptotic case of arbitrary fixed N and very large m, proving in Theorem 17 that there exist GPTs with dimension still scaling effectively linearly with m, $d(N, m) \leq m^{1+o_N(1)}$ (as $m \to \infty$), for every $N \geq 2$. This means that, in such a "big data" scenario, the exponential improvement enabled by GPTs over classical and quantum theories is independent of N; in other words, *there's plenty of room in the post-quantum world.*

Though we were not able to generalise our optimality construction – we do not know, for any given value of $N > 2$, what the optimal GPT would be – we have shown that there exists a reliable and computationally tractable method for discovering the memory capacity of theories for any N. To recap the method, we first showed that we can restrict the search to N-sized subsets of *pure* jointly distinguishable states, which can be generated as described in Algorithm 1. The set of such subsets of jointly distinguishable states can be thought of as a connection hypergraph overlaying the set of pure states of the GPT. Our search for the largest N-wise mutually distinguishable set thus becomes equivalent to the search for the maximum clique on this hypergraph, which can be performed with deterministic success as per Algorithm 2.

We hope this work could inspire further research into the cognitive abilities of intelligent agents in generalised probabilistic theories as well as their interplay with classical and quantum learning models [65, 66].

Acknowledgements

References

[1] Ambainis A, Nayak A, Ta-Shma A and Vazirani U 1999 *Proc. 31st ACM Symp. on Theory of Computing* STOC '99 (New York, NY, USA) pp 376–383 ISBN 1581130678
[2] Ver Steeg G and Wehner S 2009 *Quantum Info. Comput.* **9** 801–832 ISSN 1533-7146
[3] Hopfield J J 1982 *Proc. Natl. Acad. Sci. U.S.A.* **79** 2554–2558 ISSN 0027-8424
[4] van Dam W 2013 *Nat. Comput.* **12** 9–12
[5] Linden N, Popescu S, Short A J and Winter A 2007 *Phys. Rev. Lett.* **99**(18) 180502
[6] Lami L, Palazuelos C and Winter A 2018 *Commun. Math. Phys.* **361** 661–708
[7] Aubrun G, Lami L, Palazuelos C, Szarek S J and Winter A 2020 *Commun. Math. Phys.* **375** 679–724
[8] Aubrun G, Lami L, Palazuelos C and Plávala M 2021 *Geom. Funct. Anal.* **31** 181–205
[9] Aubrun G, Lami L, Palazuelos C and Plávala M 2022 *Phys. Rev. Lett.* **128**(16) 160402 editors' Suggestion
[10] Barrett J 2007 *Phys. Rev. A* **75** 032304
[11] Popescu S and Rohrlich D 1994 *Found. Phys.* **24** 379–385
[12] Popescu S and Rorlich D 1997 *Phys. Rev. A* **56** R3319
[13] Lami L 2017 *Non-classical correlations in quantum mechanics and beyond* Ph.D. thesis Universitat Autònoma de Barcelona preprint arXiv:1803.02902
[14] Müller M 2021 *SciPost Phys. Lect. Notes*
[15] Plávala M 2021 *Preprint arXiv:2103.07469*
[16] Danzer L and Grünbaum B 1962 *Math. Zeitschrift* **79** 95–99
[17] Arai H, Yoshida Y and Hayashi M 2019 *J. Phys. A* **52** 465304

A Post-Quantum Associative Memory   21

[18] Yoshida Y, Arai H and Hayashi M 2020 *Phys. Rev. Lett.* **125**(15) 150402
[19] Chiribella G, D'Ariano G M and Perinotti P 2010 *Phys. Rev. A* **81**(6) 062348
[20] Lee S Y, Lee C W, Nha H and Kaszlikowski D 2015 *J. Opt. Soc. Am. B* **32** 1186–1192
[21] Lee M 2016 *Bounds on computation from physical principles* Ph.D. thesis Oxford University
[22] Mueller M P 2020 *Preprint arXiv:2011.01286*
[23] Ludwig G 1985 *An Axiomatic Basis for Quantum Mechanics: Derivation of Hilbert space structure* vol 1 (Springer-Verlag)
[24] Hartkämper A and Neumann H 1974 *Foundations of Quantum Mechanics and Ordered Linear Spaces: Advanced Study Institute held in Marburg 1973* (Springer Berlin Heidelberg)
[25] Davies E B and Lewis J T 1970 *Commun. Math. Phys.* **17** 239–260
[26] Janotta P and Lal R 2013 *Phys. Rev. A* **87**(5) 052131
[27] Barnum H, Barrett J, Leifer M and Wilce A 2012 *Proc. Sympos. Appl. Math.* vol 71 pp 25–48
[28] Shahandeh F 2021 *PRX Quantum* **2** 1 ISSN 2691-3399 (*Preprint* 1911.11059)
[29] Schmid D, Selby J H, Wolfe E, Kunjwal R and Spekkens R W 2021 *PRX Quantum* **2**(1) 010331
[30] D'Ariano G M, Erba M and Perinotti P 2020 *Phys. Rev. A* **101**(4) 042118
[31] Aubrun G, Lami L and Palazuelos C 2019 *Preprint arXiv:1910.04745*
[32] Janotta P, Gogolin C, Barrett J and Brunner N 2011 *New J. Phys.* **13** ISSN 13672630 (*Preprint* 1012.1215)
[33] Short A J and Wehner S 2010 *New J. Phys.* **12** ISSN 13672630 (*Preprint* 0909.4801)
[34] Massar S and Patra M K 2014 *Physical Review A - Atomic, Molecular, and Optical Physics* **89** 1–8 ISSN 10941622 (*Preprint* 1403.2509)
[35] Heinosaari T, Leppäjärvi L and Plávala M 2019 *Quantum* **3** 1–39 ISSN 2521327X (*Preprint* 1808.07376)
[36] Kobayshi M 2017 *J. Math. Phys.* **58** ISSN 00222488
[37] Pfister C and Wehner S 2013 *Nat. Commun.* **4** 1–9 ISSN 20411723
[38] Masanes L, Müller M P, Pérez-García D and Augusiak R 2014 *J. Math. Phys.* **55** ISSN 00222488 (*Preprint* 1111.4060)
[39] Krumm M and Müller M P 2019 *NPJ Quantum Information* **5** ISSN 20566387 (*Preprint* 1804.05736)
[40] Klee V 1960 *Unsolved problems in intuitive geometry* (Hektographiert, Seattle)
[41] Aigner M and Ziegler G M 2010 *Proofs from The Book* 4th ed (Springer-Verlag, Berlin) ISBN 978-3-642-00855-9
[42] Erdos P and Spencer J 1974 *AMC* **10** 12
[43] Alon N and Spencer J H 2016 *The probabilistic method* (John Wiley & Sons)
[44] Boyd S P and Vandenberghe L 2004 *Convex Optimization* Berichte über verteilte messyteme (Cambridge University Press)
[45] Horodecki R, Horodecki P, Horodecki M and Horodecki K 2009 *Rev. Mod. Phys.* **81**(2) 865–942
[46] Peres A 1996 *Phys. Rev. Lett.* **77**(8) 1413–1415
[47] Audenaert K, Plenio M B and Eisert J 2003 *Phys. Rev. Lett.* **90**(2) 027901
[48] Wang X and Duan R 2017 *Phys. Rev. Lett.* **119**(18) 180506
[49] Wang X and Wilde M M 2020 *Phys. Rev. Lett.* **125**(4) 040502
[50] Huber M, Lami L, Lancien C and Müller-Hermes A 2018 *Phys. Rev. Lett.* **121**(20) 200503
[51] Doherty A C, Parrilo P A and Spedalieri F M 2004 *Phys. Rev. A* **69**(2) 022308
[52] Rajarama Bhat B V, Parthasarathy K R and Sengupta R 2017 *Rev. Math. Phys.* **29** 1750012
[53] Kaur E, Das S, Wilde M M and Winter A 2019 *Phys. Rev. Lett.* **123**(7) 070502
[54] Kaur E, Das S, Wilde M M and Winter A 2021 *Phys. Rev. A* **104**(2) 022401
[55] Lami L, Khatri S, Adesso G and Wilde M M 2019 *Phys. Rev. Lett.* **123**(5) 050501
[56] Vaidya P M 1989 *Proc. 30th Annual Symp. Found. Computer Science* SFCS '89 (USA: IEEE Computer Society) pp 332–337 ISBN 0818619821
[57] Vassilevska V 2009 *Information Processing Letters* **109** 254–257 ISSN 00200190
[58] Cazals F and Karande C 2008 *Theoretical Computer Science* **407** 564–568 ISSN 03043975
[59] Sun B, Danisch M, Chan T H and Sozio M 2020 *Proceedings of the VLDB Endowment* **13** 1628–1640 ISSN 21508097
[60] Carraghan R and Pardalos P M 1990 *Operations Research Letters* **9** 375–382 ISSN 01676377
[61] Segundo P S, Rodríguez-Losada D and Jiménez A 2011 *Computers and Operations Research* **38** 571–581 ISSN 03050548
[62] Wu Q and Hao J K 2015 *European Journal of Operational Research* **242** 693–709 ISSN 03772217
[63] Karp R M 1972 *Complexity of Computer Computations* 85–103
[64] Torres-Jimenez J, Perez-Torres J C and Maldonado-Martinez G 2017 *Discrete Mathematics, Algorithms and Applications* **9** ISSN 17938317
[65] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł and Polosukhin I 2017 *Advances in neural information processing systems* pp 6000–6010
[66] Adesso G 2022 *Authorea Preprints* https://doi.org/10.22541/au.167052124.41804127/v2

A Post-Quantum Associative Memory 22

[67] Rockafellar R T 1970 *Convex analysis* (Princeton University Press, Princeton, N.J.)
[68] Lang R 1986 *Arch. Math.* **47** 90–92
[69] Tabia G N M and Appleby D M 2013 *Phys. Rev. A* **88** 1–8 ISSN 10502947 (*Preprint* 1304.8075)

Appendix A. Proof of the Danzer–Grünbaum Theorem

Preliminaries: Minkowski addition

Before delving into the proof, we need to fix some terminology. A *convex body* in the Euclidean space $\mathbb{R}^n$ is a compact convex subset $A \subset \mathbb{R}^n$ with non-empty interior, in formula $\mathrm{int}(A) \neq \emptyset$. We say that two convex bodies $A, B \subset \mathbb{R}^n$ *touch each other* if $A \cap B \neq \emptyset$ but $\mathrm{int}(A) \cap \mathrm{int}(B) = \emptyset$, which corresponds to the intuitive notion of two solids touching only at their surfaces.

Two sets $A, B \subseteq \mathbb{R}^n$ can be added together via the *Minkowski addition*, defined by

$$A + B := \{a + b : a \in A, b \in B\}. \tag{A.1}$$

In what follows, for some $x \in \mathbb{R}^n$ we will often write $x + A$ instead of $\{x\} + A$. We can also multiply a given set by any real number $\lambda \in \mathbb{R}$, by setting

$$\lambda A := \{\lambda a : a \in A\}. \tag{A.2}$$

Naturally, the Minkowski difference between two sets $A, B \subseteq \mathbb{R}^n$ can now be constructed as $A - B := A + (-B)$. If A and B are convex then also $A + B$ and λA are such. If they are convex bodies and $\lambda \neq 0$, then also $A + B$ and λA are convex bodies. A special type of Minkowski addition is the *Minkowski symmetrisation*. For $A \subseteq \mathbb{R}^n$, this is defined by

$$\widetilde{A} := \frac{A - A}{2}. \tag{A.3}$$

Clearly, if A is a convex body then so is $\widetilde{A}$. In what follows we will need the following standard lemma, whose proof is included only for the sake of completeness (it follows e.g. from [67, Corollary 6.6.2]).

Lemma 23 *Let $A \subset \mathbb{R}^n$ be a convex body. Then*

$$\widetilde{\mathrm{int}\,A} = \mathrm{int}\,\widetilde{A}. \tag{A.4}$$

Proof. We start by showing that $\widetilde{\mathrm{int}\,A} \subseteq \mathrm{int}\,\widetilde{A}$. First, note that $\widetilde{\mathrm{int}\,A} \subseteq \widetilde{A}$, simply because $\mathrm{int}(A) \subseteq A$. Second, observe that $\widetilde{\mathrm{int}\,A}$ is open. To show this, pick some $a \in \widetilde{\mathrm{int}\,A}$, so that $a = \frac{b-c}{2}$ with $b, c \in \mathrm{int}\,A$. Let $\epsilon > 0$ be such that $\|\delta\| < \epsilon$ implies that $b + \delta, c + \delta \in \mathrm{int}\,A$, where $\|\cdot\|$ denotes the Euclidean norm. Then as long as $\|\delta\| \leq \epsilon$ we also have that $a + \delta = \frac{(b+\delta)-(c-\delta)}{2} \in \widetilde{\mathrm{int}\,A}$. This confirms that $\widetilde{\mathrm{int}\,A}$ is indeed open. Since the interior of a set is nothing but its largest open subset, from this and the inclusion $\widetilde{\mathrm{int}\,A} \subseteq \widetilde{A}$ we deduce that $\widetilde{\mathrm{int}\,A} \subseteq \mathrm{int}\,\widetilde{A}$.

For the other inclusion, take $a \in \mathrm{int}\,\widetilde{A}$, and some sufficiently small $\epsilon > 0$ such that $\frac{a}{1-\epsilon} = \frac{b-c}{2} \in \widetilde{A}$, where $b, c \in A$ (note that the left-hand side converges to a as $\epsilon \to 0^+$ and is thus eventually in $\widetilde{A}$). Consider a point $p \in \mathrm{int}\,A$; we now claim that $(1-\epsilon)b+\epsilon p, (1-\epsilon)c+\epsilon p \in \mathrm{int}\,A$ for all $0 < \epsilon < 1$. To see this geometrically intuitive fact, fix $\epsilon > 0$ and pick $\eta > 0$ such that $\|\delta\| \leq \eta$ implies that $p + \delta \in A$. Then as soon as $\|\delta'\| \leq \epsilon\eta$ we have that for

A Post-Quantum Associative Memory 23

example $(1 - \epsilon)b + \epsilon p + \delta' = (1 - \epsilon)b + \epsilon(p + \delta) \in A$, where $\delta := \delta'/\epsilon$. This proves that $(1 - \epsilon)b + \epsilon p, (1 - \epsilon)c + \epsilon p \in \mathrm{int}\,A$, as claimed. Now,

$$a = (1 - \epsilon)\frac{b - c}{2} = \frac{((1 - \epsilon)b + \epsilon p) - ((1 - \epsilon)c + \epsilon p)}{2} \in \frac{1}{2}\,(\mathrm{int}\,A - \mathrm{int}\,A) = \widetilde{\mathrm{int}\,A},$$

concluding the proof.　　　　　　　　　　　　　　　　　　　　　　　　　　　　□

The proof

We are now ready to present Danzer and Grünbaum's argument [16], in a slightly simplified form.

Proof of Theorem 15 and therefore of Theorem 13. For a positive integer n, some finite subset $X \subset \mathbb{R}^n$, and a convex body $A \subset \mathbb{R}^n$, we define the following properties:

- $P(n, X)$: X is not contained in any hyperplane of $\mathbb{R}^n$ (in other words, its affine hull has dimension n) and for all distinct $x_1, x_2 \in X$ there are parallel hyperplanes $V_1, V_2 \subset \mathbb{R}^n$ such that V_i supports X in x_i, for $i = 1, 2$.
- $Q(n, A, X)$: For all $x_1, x_2 \in X$, the convex bodies $x_1 + A$ and $x_2 + A$ touch each other.
- $Q^*(n, A, X)$: Same as $Q(n, A, X)$, but we additionally require that $A = -A$ (i.e. that A be centrally symmetric).

Furthermore, let us set

$$p_n := \sup\{|X| : \exists X \subset \mathbb{R}^n \text{ finite}: P(n, X)\}, \tag{A.5}$$

$$q_n := \sup\{|X| : \exists X \subset \mathbb{R}^n \text{ finite}, A \subset \mathbb{R}^n \text{ convex body}: Q(n, X, A)\}, \tag{A.6}$$

$$q_n^* := \sup\{|X| : \exists X \subset \mathbb{R}^n \text{ finite}, A = -A \subset \mathbb{R}^n \text{ convex body}: Q^*(n, X, A)\}, \tag{A.7}$$

where $|\cdot|$ denotes the cardinality of a finite set, i.e. the number of elements it contains. The geometrically intuitive fact that the 2^n vertices of the hypercube satisfy $P(n, X)$ — and hence $p_n \geq 2^n$ — has been discussed in Section 3.2, so we will not dwell on it further. The problem is to show that $p_n \leq 2^n$. The proof can be broken down into the following chain of inequalities:

$$2^n \leq p_n \overset{\text{(i)}}{\leq} q_n \overset{\text{(ii)}}{=} q_n^* \overset{\text{(iii)}}{\leq} 2^n. \tag{A.8}$$

We now justify one by one the three crucial steps (i)–(iii):

i In fact, for all n and for all sets $X \subseteq \mathbb{R}^n$ we have that $P(n, X) \implies Q(n, -\mathrm{conv}(X), X)$, where conv denotes the convex hull. To see this, assume that $P(n, X)$ holds. Then, for $x_1, x_2 \in X$ with $x_1 \neq x_2$ there exists a hyperplane $V \subset \mathbb{R}^n$ such that the set X, and hence also the convex body $\mathrm{conv}(X)$, is entirely contained between $x_1 + V$ and $x_2 + V$. Multiplying by -1 and translating, we see that $x_1 - \mathrm{conv}(X)$ is entirely contained between V and $x_1 - x_2 + V$, and analogously $x_2 - \mathrm{conv}(X)$ is entirely contained between V and $x_2 - x_1 + V$. Since $x_1 \neq x_2$, we see that the convex bodies $x_1 - \mathrm{conv}(X)$ and $x_2 - \mathrm{conv}(X)$ are entirely contained into each of the two closed half-spaces determined by V. This implies that their interiors, which are instead contained into the corresponding *open* half-spaces, are disjoint. Remembering that $0 \in (x_1 - \mathrm{conv}(X)) \cap (x_2 - \mathrm{conv}(X))$, we see that in fact $x_1 - \mathrm{conv}(X)$ and $x_2 - \mathrm{conv}(X)$ touch each other.

A Post-Quantum Associative Memory   24

ii We show that for all n, for all finite $X \subseteq \mathbb{R}^n$, and for all $A \subseteq \mathbb{R}^n$,

$$Q(n, A, X) \iff Q\left(n, \widetilde{A}, X\right), \tag{A.9}$$

so that naturally $q_n = q_n^*$. Start by noting the following: for a set $A \subseteq \mathbb{R}^n$ and two points $x, y \in \mathbb{R}^n$,

$$(x + A) \cap (y + A) \neq \emptyset \iff \frac{x - y}{2} \in \widetilde{A}, \tag{A.10}$$

where $\widetilde{A}$ is the Minkowski symmetrisation of A. Therefore, for fixed $x_1, x_2 \in X$, we have that $(x_1 + A) \cap (x_2 + A) \neq \emptyset$ if and only if $\frac{x_1 - x_2}{2} \in \widetilde{A}$. Since A and $\widetilde{A}$ have the same Minkowski symmetrisation, this is also equivalent to $\left(x_1 + \widetilde{A}\right) \cap \left(x_2 + \widetilde{A}\right) \neq \emptyset$. In other words,

$$(x_1 + A) \cap (x_2 + A) \neq \emptyset \iff \left(x_1 + \widetilde{A}\right) \cap \left(x_2 + \widetilde{A}\right) \neq \emptyset. \tag{A.11}$$

Applying this to $\text{int } A$ instead of A, we get that

$$\begin{aligned}
&\text{int}(x_1 + A) \cap \text{int}(x_2 + A) = (x_1 + \text{int}(A)) \cap (x_2 + \text{int}(A)) = \emptyset \\
&\iff \left(x_1 + \widetilde{\text{int } A}\right) \cap \left(x_2 + \widetilde{\text{int } A}\right) = \text{int}\left(x_1 + \widetilde{A}\right) \cap \text{int}\left(x_2 + \widetilde{A}\right) = \emptyset,
\end{aligned} \tag{A.12}$$

where the identity $x_i + \widetilde{\text{int } A} = \text{int}\left(x_i + \widetilde{A}\right)$ follows from Lemma 23. We have therefore proved that the convex bodies $x_1 + A$ and $x_2 + A$: (a) intersect if and only if so do $x_1 + \widetilde{A}$ and $x_2 + \widetilde{A}$; and (b) have disjoint interiors if and only if so do $x_1 + \widetilde{A}$ and $x_2 + \widetilde{A}$. In other words, $x_1 + A$ and $x_2 + A$ touch each other if and only if also $x_1 + \widetilde{A}$ and $x_2 + \widetilde{A}$ touch each other.

iii We now show that $q_n^* \leq 2^n$. To this end, pick a convex body $A = -A \subset \mathbb{R}^n$ and some set $X \subseteq \mathbb{R}^n$ such that $Q^*(n, A, X)$ holds. Set $B := \text{conv}(X)$. By (A.10), for all $x_1, x_2 \in X$ it must hold that $\frac{x_1 - x_2}{2} \in \widetilde{A} = A$. Then we claim that for all $x \in X$,

$$\frac{x + B}{2} \subseteq x + A. \tag{A.13}$$

To see this, up to taking the convex hull it suffices to show that $\frac{x+y}{2} \subseteq x + A$ for all $y \in X$. And indeed, thanks to the above observation $\frac{x+y}{2} = x + \frac{y-x}{2} \in x + A$. This proves (A.13). As an immediate consequence of this together with $Q^*(n, A, X)$, observe that the interiors of the convex bodies $\frac{x+B}{2}$, indexed by $x \in X$, are all disjoint and moreover contained in B, because this is convex. Therefore, their volumes†† satisfy that

$$\text{vol}(B) \geq \sum_{x \in X} \text{vol}\left(\frac{x + B}{2}\right) = \sum_{x \in X} \frac{1}{2^n} \text{vol}(x + B) = \sum_{x \in X} \frac{1}{2^n} \text{vol}(B) = \frac{|X|}{2^n} \text{vol}(B). \tag{A.14}$$

Since $\text{vol}(B) > 0$ because B is a convex body, we obtain that $|X| \leq 2^n$, as claimed.

This concludes the proof.   □

††Convex bodies are well known to be Lebesgue measurable, see e.g. [68] and references therein.

A Post-Quantum Associative Memory   25

Appendix B. Prism Theories

Expanding a GPT to higher dimensions is a way to explore systems with a variable number of degrees of freedom, but which are governed by a consistent set of relationships. For example, the state space of an n-sided ordinary, classical die is represented in GPT form by a simplex with n vertices (representing deterministic preparations of a particular outcome); this allows us to accommodate systems having many degrees of freedom by generalising the same basic geometric pattern to higher dimensions. Though this works in a straightforward way for classical theory, the situation in quantum theory is more nuanced. The state space of the qubit is represented by the Bloch sphere in three dimensions, but the state space of a qutrit possesses a complicated geometry [69] which is not simply given by a sphere in higher dimensions. Theories using hyperspheres of higher dimension, so-called D-balls, are discussed in [38, 39]; in these the authors aim to isolate the 3-sphere as a the necessary state space for quantum theory based on physical requirements.

Here, we introduce a method for expanding given geometries to higher dimensions in a generic way. We do this by taking the Cartesian product of shapes in lower dimensional spaces. In Figure B1 we visualise some state spaces shaped as simplices and their corresponding effects, as well as a simplex prism $\mathcal{S}_3 \times \mathcal{S}_3$.

However, this way of incorporating new degrees of freedom, although mathematically consistent, does not have a direct operational interpretation: new variables do not have to be independent of the old ones.

It is a feature of the Cartesian product that the product of any two convex sets will produce a new convex set. We can use this feature as the basis to construct new, higher dimensional GPTs by taking the Cartesian product (denoted $\times$) of lower dimensional state spaces. The resulting GPT is called a *prism theory*. We give a formal definition below:

Definition 24 (Prism theories) *Let $A = (V_A, C_A, u_A)$ and $B = (V_B, C_B, u_B)$ be two GPTs. The* **prism theory** *$A \oplus B = (V_{A \oplus B}, C_{A \oplus B}, u_{A \oplus B})$ is defined as follows:*

   i $V_{A \oplus B} := \ker((u_A, 0) - (0, u_B)) \subset V_A \oplus V_B$ *is the subspace of $V_A \oplus V_B$ given by the kernel of the functional $(u_A, 0) - (0, u_B)$ whose action is defined by $((u_A, 0) - (0, u_B))(x, y) := u_A(x) - u_B(y)$;*

   ii $C_{A \oplus B} := (C_A \oplus C_B) \cap V_{A \oplus B}$;

   iii $u_{A \oplus B}$ *is the restriction of $(u_A, 0)$ (equivalently, of $(0, u_B)$) to $V_{A \oplus B}$.*

To unpack the above somewhat complicated definition, it is useful to look at the state spaces. Since the host vector space $V_{A \oplus B}$ is a subspace of the simple direct sum $V_A \oplus V_B$, any state of $A \oplus B$ can also be seen as a vector of the form $\omega_{A \oplus B} = (x, y) \in V_A \oplus V_B$. We observe that item (ii) implies that in fact $x \in C_A$ and $y \in C_B$, so that $x = \lambda \omega_A$ and $y = \mu \omega_B$, for $\lambda, \mu \geq 0$ and $\omega_A \in \Omega_A := C_A \cap u_A^{-1}(1)$, $\omega_B \in \Omega_B := C_B \cap u_B^{-1}(1)$. Now, since $\omega_{A \oplus B}$ must belong to the kernel of $(u_A, 0) - (0, u_B)$, we also see that $\lambda = \mu$; if it is a normalised state, then by (iii) we have that $1 = u_{A \oplus B}(\omega_{A \oplus B}) = (u_A, 0)(\lambda \omega_A, \lambda \omega_B) = \lambda$. Therefore, $\omega_{A \oplus B}$ can be simply identified with the pair of states (ω_A, ω_B), and vice versa any such pair constitutes a state of $A \oplus B$. We have thus proved the following, which amounts to an intuitive description of the rather cumbersome Definition 24:

Lemma 25 *For any two GPTs A, B with state spaces Ω_A, Ω_B, the state space of the prism theory $A \oplus B$ is simply the Cartesian product of Ω_A and Ω_B. In formula,*

$$\Omega_{A \oplus B} = \Omega_A \times \Omega_B. \tag{B.1}$$

(a) States and effects for simplex with 2 pure states

(b) States and effects for simplex with 3 pure states

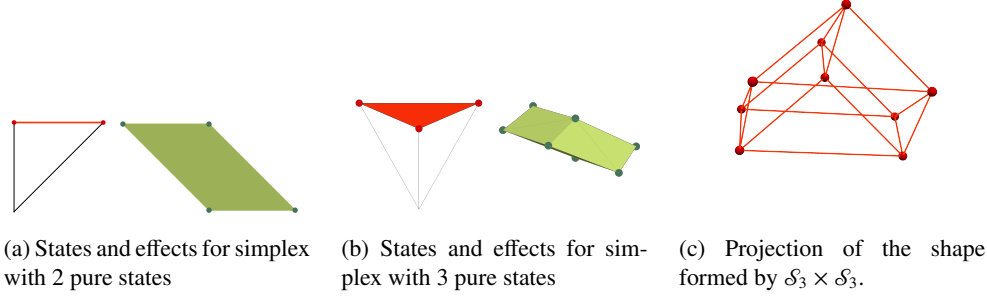(c) Projection of the shape formed by $\mathcal{S}_3 \times \mathcal{S}_3$.

Figure B1: Depiction of GPTs based on simplices. Image (a) shows the states in a theory based on $\mathcal{S}_2$. The two pink points correspond to the extremal states of the theory, and the thin black lines connect these to the origin. The green square shows the space of possible effects, with the darker points signifying the extremal effects. In (b) we see the theory corresponding to $\mathcal{S}_3$, in which we've added a dimension. On the right, in (c) we see a representation of the product of two simplices, $\mathcal{S}_3 \times \mathcal{S}_3$. Since this yields a shape in four dimensions, we have used a projection to visualise it in three dimensions. Here the entire volume of the polytope should be understood as representing the space of mixed states, with the pink points again corresponding to pure states.

Remark 26 If A, B are two GPTs with dimensions $\dim A = d_A$ and $\dim B = d_B$, thanks to Lemma 25 we have that

$$\dim (A \oplus B) = (d_A - 1)(d_B - 1) + 1 \,. \tag{B.2}$$

Appendix C. On measurement normalisation

Here we construct an example of a GPT with state space Ω and effect space E in which one can find three states $(\omega_i)_{i=1,2,3} \subset \Omega$ and three extremal effects $(e_j)_{j=1,2,3} \subset E$ satisfying $e_j \cdot \omega_i = \delta_{ij}$, but such that $(\omega_i)_{i=1,2,3}$ are not perfectly distinguishable, i.e. there does not exist a *measurement* $(f_k)_{k=1,2,3}$ such that $f_k \cdot \omega_i = \delta_{ik}$. The reason why this is possible, naturally, is that only collections of effects $(f_k)_k$ satisfying $\sum_k f_k = u$, with u being the order unit, can represent physical measurements.

The state space of the GPT we have in mind is — once again! — shaped as a 3-dimensional cube. More precisely, we consider the $n = 3$ case of the GPT constructed in Section 3.2 (see in particular (7) there). Its state space is depicted in Figure C1. We identify there three states $\omega_1, \omega_2, \omega_3$, with coordinates

$$\omega_1 \coloneqq (1, 1, 1, 1)^\mathsf{T}, \quad \omega_2 \coloneqq (1, -1, 1, -1)^\mathsf{T}, \quad \omega_3 \coloneqq (1, -1, -1, 1)^\mathsf{T}, \tag{C.1}$$

and five auxiliary states ρ_0, ρ_1 and $\sigma_1, \sigma_2, \sigma_3$, defined by

$$\rho_0 \coloneqq (1, -1, 1, 1)^\mathsf{T}, \qquad \rho_1 \coloneqq (1, 1, -1, -1)^\mathsf{T}, \tag{C.2}$$

$$\sigma_1 \coloneqq (1, -1, -1, -1)^\mathsf{T}, \quad \sigma_2 \coloneqq (1, 1, -1, 1)^\mathsf{T}, \quad \sigma_3 \coloneqq (1, 1, 1, -1)^\mathsf{T}. \tag{C.3}$$

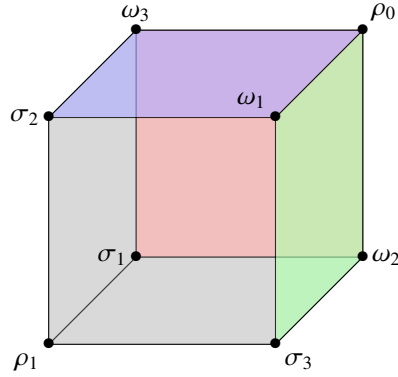*A Post-Quantum Associative Memory*   27

Figure C1: A pictorial representation of the construction in Appendix C. The three coloured faces represent the set of states for which $e_1 = 0$ (red), $e_2 = 0$ (blue), and $e_3 = 0$ (green).

Note that the first coordinate represents the normalisation, in accordance with the notation of (7), and the last three identify the position of the state in the 3-dimensional 'section' space depicted in Figure C1.

We now construct the three extremal effects $(e_j)_{j=1,2,3} \subset E$ satisfying $e_j \cdot \omega_i = \delta_{ij}$. In the dual space set

$$e_1 := \frac{1}{2}(1,1,0,0)\,, \quad e_2 := \frac{1}{2}(1,0,0,-1)\,, \quad e_3 := \frac{1}{2}(1,0,-1,0)\,. \tag{C.4}$$

(Note that the states were represented by column vectors, so the effects are represented by row vectors.) Note that indeed $e_j \cdot \omega_i = \delta_{ij}$. Moreover, since a generic effect is of the form (c, y_1, y_2, y_3), with $\min\{c, 1-c\} \geq \sum_i |y_i|$, it follows that each e_j is an extremal effect. The faces of the state space on which $e_1 = 0$, $e_2 = 0$, and $e_3 = 0$ are depicted in Figure C1 as coloured in red, blue, and green, respectively.

We now show that the states in (C.1) are not perfectly distinguishable. A first clue that this may be the case can be obtained by noting that the three effects in (C.4) satisfy $\sum_i e_i = \frac{1}{2}(3,1,-1,-1) \not\leq (1,0,0,0) = u$, where $\not\leq$ signifies that the inequality $\leq$ can be violated if both sides are evaluated on certain states in Ω. This means that the collection (e_1, e_2, e_3) does not constitute a measurement. ITo turn this observation into a fully-fledged proof, one observes that the three effects in (C.4) are the *only* ones that can satisfy $e_j \cdot \omega_i = \delta_{ij}$: since they do not form a measurement, the states in (C.1) cannot be perfectly distinguishable.

We will however follow a different reasoning, which has the advantage of providing some quantitative insights. To this end, we will employ the auxiliary states in (C.2) and (C.3). We start by noticing that for all $k = 1, 2, 3$ it holds that $\rho_1 = 2(\sigma_k + \omega_k) - \sum_i \omega_i$. Now, assume by contradiction that we have found a measurement $(f_k)_k$ satisfying both $\sum_k f_k = u$ and $f_k \cdot \omega_i = \delta_{ik}$. Then

$$1 = u(\rho_1) = \sum_k f_k(\rho_1) = \sum_k f_k \left(2(\sigma_k + \omega_k) - \sum_i \omega_i\right)$$
$$\geq \sum_k \left(2 - \sum_i \delta_{ik}\right) = \sum_k (2-1) = 3\,, \tag{C.5}$$

and we have reached a contradiction.