The Geometry of Information Retrieval

Quantum-inspired IR Models

Ingo Frommholz

University of Bedfordshire
Luton, UK

Herbstschule Information Retrieval
Schloss Dagstuhl

University of Bedfordshire

Table of Contents

Acknowledgements

- Some parts (in particular the quantum probability one) are based on the ECIR 2012 “Quantum Information Access and Retrieval” tutorial given by *Benjamin Piwowarski* and *Massimo Melucci*
- This tutorial is worth checking out
  - It covers some aspect more thoroughly
  - It discusses some further quantum-based IR models
- See also

  `http://www.bpiwowar.net/2012/04/`

  `slides-from-the-ecir12-quantum-information-access-and-retrieval-tutorial/`
- Direct download:

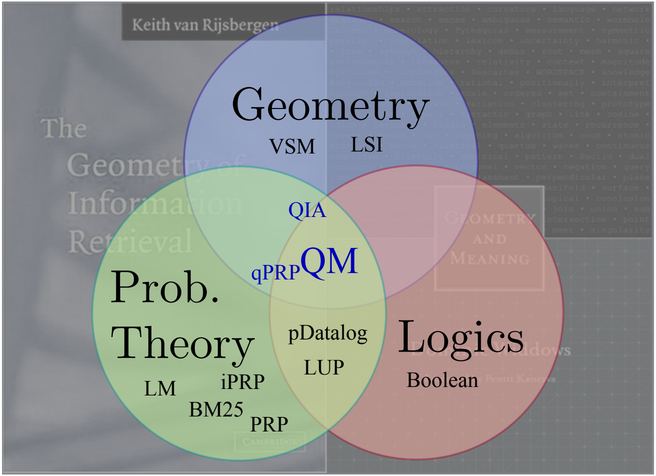
  `http://www.bpiwowar.net/wp-content/uploads/2012/04/tutorial-handout.pdf`
- I’d like to thank *Guido Zuccon* for providing material about qPRP

Introduction

- IR and Geometry

IR Models and Principles

Geometry, Probability and Logics



A Language for IR

- The geometry and mathematics behind quantum mechanics can be seen as a 'language' for expressing the different IR models [van Rijsbergen, 2004].
- Combination of geometry, probability and logics
- Leading to non-classical probability theory and logics
- Potential unified framework for IR models
- Applications in areas outside physics emerging
  - Quantum Interaction symposia (e.g. [Song et al., 2011])

IR as Quantum System?

An Analogy

Quantum System	**IR System**
Particles, physical properties in Hilbert spaces	Documents, relevance, information needs in Hilbert Spaces
System state uncertain	Information need (IN) uncertain
Observation changes system state	User interaction changes system state
Observations interfere (Heisenberg)	Document relevance interferes
Combination of systems	Combination of IN facets, polyrepresentation

Brief History of the Quantum Formalism

1890s Max Planck's hypothesis: energy not continuous but comes in *quantas*

1920s Born/Jordan: matrix reformulation of Heisenberg's work

1930s Dirac/von Neumann: mathematical formulation of quantum physics (Hilbert space), theory of quantum measurement, quantum logics

Introduction to Quantum Probabilities

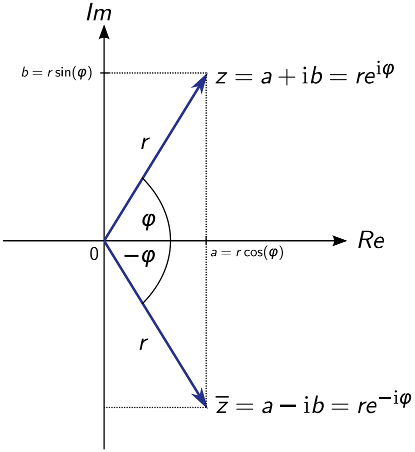
- Quantum Formalism
- Preliminaries: Hilbert Spaces and Inner Products
  - Complex Numbers
  - Hilbert Spaces
  - Operators and Projectors
  - Tensor Spaces
- Quantum Probabilities
  - Quantum and Classical Probabilities
  - Dirac Notation
  - Events
  - Quantum Logics
  - Quantum States
  - Pure States
  - Mixed States
  - Density Operators
  - Conclusion

What does this remind you of?

Quantum Formalism

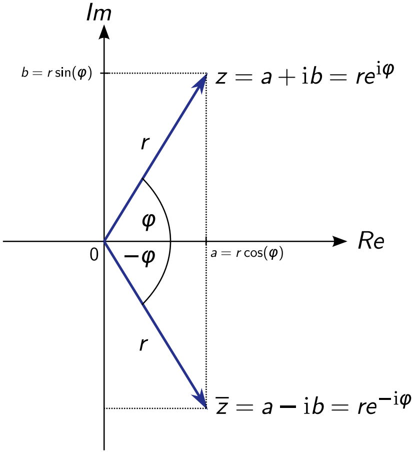
- The quantum formalism is build on top of Hilbert spaces
- Each finite-dimensional vector space with an inner product is a Hilbert space [Halmos, 1958]
  - We focus on finite-dimensional spaces here
- A vector space is defined over a field $\mathbb{K}$, e.g $\mathbb{R}$ or $\mathbb{C}$

Complex Numbers



- Complex number $z \in \mathbb{C}$
- $z = a + \mathrm{i}b,\ a, b \in \mathbb{R}, \mathrm{i}^2 = -1$
- Polar form: $z = r(\cos(\varphi) + \mathrm{i}\sin(\varphi)) = re^{\mathrm{i}\varphi}$ with $r \in \mathbb{R}^+, \varphi \in [0, 2\pi]$
- Addition/Multiplication: $z_1 = a_1 + \mathrm{i}b_1 = r_1 e^{\mathrm{i}\varphi_1}$, $z_2 = a_2 + \mathrm{i}b_2 = r_2 e^{\mathrm{i}\varphi_2}$:
  $z_1 + z_2 = (a_1 + a_2) + i(b_1 + b_2)$
  $z_1 \cdot z_2 = r_1 r_2 e^{\mathrm{i}(\varphi_1 + \varphi_2)}$

Complex Numbers



- Complex number $z \in \mathbb{C}$
- $z = a + \mathrm{i}b$, $a, b \in \mathbb{R}$, $\mathrm{i}^2 = -1$
- Polar form: $z = r(\cos(\varphi) + \mathrm{i}\sin(\varphi)) = re^{\mathrm{i}\varphi}$ with $r \in \mathbb{R}^+, \varphi \in [0, 2\pi]$
- Complex conjugate
  $\bar{z} = a - \mathrm{i}b = re^{-\mathrm{i}\varphi}$
  $b = 0 \Leftrightarrow z \in \mathbb{R} \Leftrightarrow \bar{z} = z$
- Absolute value
  $|z| = \sqrt{a^2 + b^2} = r = \sqrt{z\bar{z}}$
  $|z|^2 = z\bar{z}$

Vector Space

Vector Space

Set $\mathcal{V}$ of objects called vectors satisfying

- **Addition**: $\forall \mathbf{x}, \mathbf{y} \in \mathcal{V} : \mathbf{x} + \mathbf{y} \in \mathcal{V}$ and
  - Commutative: $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$
  - Associative: $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$
  - Origin: $\exists ! \mathbf{O} \in \mathcal{V} : \mathbf{x} + \mathbf{O} = \mathbf{x} \quad \forall \mathbf{x} \in \mathcal{V}$
  - Additive inverse: $\forall \mathbf{x} \in \mathcal{V} \quad \exists ! -\mathbf{x} \quad$ with $\quad \mathbf{x} + (-\mathbf{x}) = \phi$
- **Multiplication by scalar**: Let $\alpha \in \mathbb{K}$ be a scalar and $\mathbf{x} \in \mathcal{V}$. Then $\alpha\mathbf{x}$ is the product of α and $\mathbf{x}$ with the properties
  - Associative: $(\alpha\beta)\mathbf{x} = \alpha(\beta\mathbf{x})$
  - Distributive:
    - $\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$
    - $(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}$

Vector Space

Example

Example: n-dimensional complex vector space $\mathbb{C}^n$:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

with $x_i \in \mathbb{C}$ and

$$\mathbf{x} + \mathbf{y} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix} \text{ and } \quad \alpha\mathbf{x} = \begin{pmatrix} \alpha x_1 \\ \vdots \\ \alpha x_n \end{pmatrix}$$

Vector Space

Linear Combinations

- Linear combination: $\mathbf{y} = c_1\mathbf{x_1} + \ldots + c_n\mathbf{x_n}$
- $\{\mathbf{x_1}, \ldots, \mathbf{x_n}\}$ are linearly independent if

$$c_1\mathbf{x_1} + \ldots + c_n\mathbf{x_n} = \mathbf{0} \quad \text{iff} \quad c_1 = c_2 = \ldots = c_n = 0$$

with $\mathbf{0}$ being the zero vector

Vector Space

Basis

(Finite) Basis

A set of n linearly independent vectors $\mathcal{B} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ form a (finite) basis of a vector space $\mathcal{V}$ if every vector in $\mathcal{V}$ is a linear combination of vectors in $\mathcal{B}$:

$$\mathbf{x} = c_1\mathbf{x}_1 + \ldots + c_n\mathbf{x}_n = \sum_i c_i\mathbf{x}_i$$

- Example: Canonical basis in $\mathbb{R}^n$ (orthonormal basis)

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ . \\ 0 \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ . \\ 0 \end{pmatrix}, \ldots, \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ . \\ 1 \end{pmatrix} \text{ with } \mathbf{x} = \sum_{i=1}^{n} x_i\mathbf{e}_i$$

Vector Space

Subspace

Subspace

A non-empty subset $\mathcal{V}'$ of a vector space $\mathcal{V}$ is a subspace if along with every pair $\mathbf{x}, \mathbf{y} \in \mathcal{V}'$, every linear combination $\alpha\mathbf{x} + \beta\mathbf{y} \in \mathcal{V}'$.

- A subspace is also a vector space
- $\dim(\mathcal{V}') \leq \dim(\mathcal{V})$
- For each $\mathbf{x} \in \mathcal{V}'$, $\mathbf{x} - \mathbf{x} = \mathbf{0} \in \mathcal{V}'$ (each subspace passes through the origin)
- Example: Each 2-dimensional plane that passes through the origin is a subspace of $\mathbb{R}^3$

Hilbert Space

Inner product/1

Hilbert space $\mathcal{H}$: vector space with an inner product

(Complex) Inner Product

A function $\langle .,. \rangle \in \mathbb{C}$ with

- **Conjugate symmetry**: $\langle \mathbf{x}, \mathbf{y} \rangle = \overline{\langle \mathbf{y}, \mathbf{x} \rangle}$ (symmetric if $\langle .,. \rangle \in \mathbb{R}$, in particular $\langle \mathbf{x}, \mathbf{x} \rangle \in \mathbb{R}$!)
- **Linearity**:
  $\langle \lambda \mathbf{y}, \mathbf{x} \rangle = \lambda \langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \overline{\lambda} \mathbf{x} \rangle$
  $\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$
- **Positive definite**: $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ and $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ iff $\mathbf{x} = \mathbf{0}$

for $\lambda \in \mathbb{C}$ and $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{H}$

Hilbert Space

Inner product/2

Properties of the inner product:

- Real case (quite obvious):

$$\langle \mathbf{x}, \alpha\mathbf{y}_1 + \beta\mathbf{y}_2\rangle = \alpha\langle \mathbf{x}, \mathbf{y}_1\rangle + \beta\langle \mathbf{x}, \mathbf{y}_2\rangle$$

- Be careful in the complex case:

$$\begin{aligned} \langle \mathbf{x}, \alpha\mathbf{y}_1 + \beta\mathbf{y}_2\rangle &= \overline{\langle \alpha\mathbf{y}_1 + \beta\mathbf{y}_2, \mathbf{x}\rangle} \\ &= \overline{\alpha\langle \mathbf{y}_1, \mathbf{x}\rangle} + \overline{\beta\langle \mathbf{y}_2, \mathbf{x}\rangle} \\ &= \overline{\alpha}\langle \mathbf{x}, \mathbf{y}_1\rangle + \overline{\beta}\langle \mathbf{x}, \mathbf{y}_2\rangle \end{aligned}$$

Hilbert Space

Inner product/3

- There are many possible inner products, they need to make sense for the application
- Inner product example (standard inner product)

$$\begin{aligned} \langle \mathbf{x}, \mathbf{y} \rangle &= \mathbf{x}^\dagger \mathbf{y} = \sum_i \overline{x}_i y_i \\ &= \mathbf{x}^\mathsf{T} \mathbf{y} = \sum_i x_i y_i \text{ if real} \end{aligned}$$

with row vectors

$$\begin{aligned} \mathbf{x}^\dagger &= (\overline{x}_1, \ldots, \overline{x}_n) \qquad (\textit{adjoint}) \\ \mathbf{x}^\mathsf{T} &= (x_1, \ldots, x_n) \qquad (\textit{transpose}) \end{aligned}$$

Hilbert Space

Norm

Norm

$$||\mathbf{x}|| = \sqrt{\langle x, x \rangle}$$

is the norm of a vector $\mathbf{x}$.

- Geometric interpretation: length of the vector
- $||\mathbf{x}|| \in \mathbb{R}$
- Standard inner product:

$$||\mathbf{x}|| = \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{x_1^2 + \ldots + x_n^2}$$

- Vector $\mathbf{x}$ with $||\mathbf{x}|| = 1$ is called a unit vector , e.g. $\mathbf{x} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$

Hilbert Space

Orthogonality

Orthogonality

Two vectors **x** and **y** are orthogonal if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$

Example ($\mathbb{R}^2$):

$$\mathbf{x} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \mathbf{y} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}$$

Example ($\mathbb{C}^2$):

$$\mathbf{x} = \begin{pmatrix} i \\ i \end{pmatrix}, \mathbf{y} = \begin{pmatrix} i \\ -i \end{pmatrix}$$

Towards Projectors

- One of the most important operations for quantum probabilities are projectors
- We need to learn about linear operators and their matrix representation first (see also [van Rijsbergen, 2004, Chapter 4])

Linear Operator

Basic operations in a Hilbert space $\mathcal{H}$ are performed by linear operators (a special case of linear maps).

Linear Operator

A linear operator is a map $f : \mathcal{H} \mapsto \mathcal{H}$ such that for any scalar $\lambda \in \mathbb{C}$ we have

$$f(\mathbf{x} + \lambda \mathbf{y}) = f(\mathbf{x}) + \lambda f(\mathbf{y})$$

- Examples: rotation, projection, scaling
- Linear operators can be represented by a (square) matrix $\mathbf{A}$
- Applying $\mathbf{A}$ to vector $\mathbf{x}$: $\mathbf{Ax}$

Operators as Matrices

$$\underset{\mathbf{A}}{\begin{pmatrix} a_{11} & . & . & . & a_{1n} \\ . & . & . & . & . \\ a_{i1} & . & a_{ik} & . & a_{in} \\ . & . & . & . & . \\ a_{n1} & . & . & . & a_{nn} \end{pmatrix}} \underset{\mathbf{x}}{\begin{pmatrix} x_1 \\ . \\ . \\ . \\ x_n \end{pmatrix}} = \underset{\mathbf{y}}{\begin{pmatrix} y_1 \\ . \\ y_i \\ . \\ y_n \end{pmatrix}}$$

$$y_i = \sum_{k=1}^{n} a_{ik} x_k$$

Product of Transformations

Commutativity

- The product of two transformations **A** and **B** is defined by the effect it has on the vector **x**
- **ABx** means: Apply **B** to **x** then apply the result to **A**
- **ABx** is usually (but not always) *different* from **BAx**
- If $\mathbf{ABx} = \mathbf{BAx}$, that is $\mathbf{AB} = \mathbf{BA}$, **A** and **B** are said to be commutative (commuting)
- They are non-commutative if $\mathbf{AB} \neq \mathbf{BA}$

Product of Transformations

Matrix Multiplication

$$\underset{\mathbf{A}}{\begin{pmatrix} a_{11} & . & . & . & a_{1n} \\ . & . & . & . & . \\ a_{i1} & . & . & . & a_{in} \\ . & . & . & . & . \\ a_{n1} & . & . & . & a_{nn} \end{pmatrix}} \underset{\mathbf{B}}{\begin{pmatrix} b_{11} & . & b_{1j} & . & b_{1n} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ b_{n1} & . & b_{nj} & . & b_{nn} \end{pmatrix}} = \underset{\mathbf{C}}{\begin{pmatrix} c_{11} & . & . & . & c_{1n} \\ . & . & . & . & . \\ c_{i1} & . & c_{ij} & . & c_{in} \\ . & . & . & . & . \\ c_{n1} & . & . & . & c_{nn} \end{pmatrix}}$$

$$\mathbf{AB} = \mathbf{C}$$

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

Product of Transformations

Matrix Multiplication Example

$$
\begin{aligned}
\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} 3 & 4 \\ 5 & 6 \end{pmatrix} &= \begin{pmatrix} 13 & 16 \\ 29 & 36 \end{pmatrix} \\
&\neq \begin{pmatrix} 15 & 22 \\ 23 & 34 \end{pmatrix} \\
&= \begin{pmatrix} 3 & 4 \\ 5 & 6 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}
\end{aligned}
$$

non-communitative

Adjoints

Adjoint and Self-Adjoint (Hermitian)

The adjoint of a linear operator (or matrix) $\mathbf{A}$ is an operator $\mathbf{A}^\dagger$ so that

$$\langle \mathbf{A}^\dagger \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{A}\mathbf{y} \rangle$$

An operator/matrix is self-adjoint (Hermitian) if $\mathbf{A} = \mathbf{A}^\dagger$.

- When the scalars are complex: $\mathbf{A}^\dagger = \overline{\mathbf{A}^\top}$ (conjugate transpose)
- In the real case: $\mathbf{A}^\dagger = \mathbf{A}^\top \rightsquigarrow$ symmetric ($\mathbf{A}^\top = \mathbf{A}$) and self-adjoint matrices are the same

Adjoints

Example

- Real case:

$$\begin{pmatrix} a & x \\ u & c \end{pmatrix}^{\dagger} = \begin{pmatrix} a & u \\ x & c \end{pmatrix}$$

- Complex case:

$$\begin{pmatrix} a+\mathrm{i}b & x-\mathrm{i}y \\ u-\mathrm{i}v & c+\mathrm{i}d \end{pmatrix}^{\dagger} = \begin{pmatrix} a-\mathrm{i}b & u+\mathrm{i}v \\ x+\mathrm{i}y & c-\mathrm{i}d \end{pmatrix}$$

[van Rijsbergen, 2004, p. 55]

Orthogonal Projector

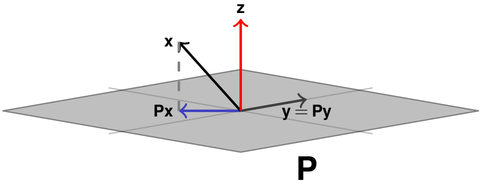
Orthogonal Projector

An orthogonal projector **P** is an idempotent, self-adjoint linear operator in $\mathcal{H}$.

- Idempotent: $\mathbf{P} = \mathbf{PP}$ (**P** leaves its image unchanged)
- There is a one-to-one correspondence between orthogonal projectors and subspaces
  - Each vector orthogonal to the subspace projects to **0**
- We use the notation **P** to denote the orthogonal projector (a matrix) that represents a subspace

Orthogonal Projector

Example



- $\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$
- $\mathbf{z} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ orthogonal to $\mathbf{P}$, hence $\mathbf{Pz} = \mathbf{0}$
- $\mathbf{y} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ contained in $\mathbf{P}$, so projection has no effect

Special Projector: Ray

Ray

If $\mathbf{x}$ is a unit vector ($||\mathbf{x}|| = 1$), then $\mathbf{x}\mathbf{x}^\dagger$ is an orthogonal projector onto the 1-dimensional subspace defined by $\mathbf{x}$. This projector is called a ray and denoted $\mathbf{P}_x$.

Example ($\mathbb{R}^2$):

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \mathbf{e}_1\mathbf{e}_1^\dagger = \begin{pmatrix} 1 \\ 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \mathbf{P}_{e_1}$$

$$\mathbf{P}_{e_1}\mathbf{e}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \mathbf{P}_{e_1}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Spectral Theorem

Spectral Theorem

To any self-adjoint matrix $\mathbf{A}$ on a finite-dimensional complex inner product space $\mathcal{V}$ there correspond real numbers $\alpha_1, \ldots, \alpha_r$ and projectors $\mathbf{E}_1, \ldots, \mathbf{E}_r$, $r \leq \dim(\mathcal{V})$, so that

1. the α_j are pairwise distinct;
2. the $\mathbf{E}_j$ are mutually orthogonal;
3. $\sum_{j=1}^{r} \mathbf{E}_j = \mathbf{I}$ ($\mathbf{I}$ is the identity matrix);
4. $\mathbf{A} = \sum_{j=1}^{r} \alpha_j \mathbf{E}_j$

- Orthogonality: $\mathbf{E}_i \perp \mathbf{E}_j$ iff $\mathbf{E}_i\mathbf{E}_j = \mathbf{E}_i\mathbf{E}_j = \mathbf{0}$
- The α_j are the distinct eigenvalues of $\mathbf{A}$
- The $\mathbf{E}_j$ are the subspaces generated by the eigenvectors

Spectral Theorem

Simple Example

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad \text{Eigenvalues} \quad \alpha_1 = 1 \quad \alpha_2 = 0$$

$$\mathbf{A} = \alpha_1 \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}}_{\mathbf{E}_1} + \alpha_2 \underbrace{\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\mathbf{E}_2}$$

$$\text{Eigenvector } \mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} : \quad \mathbf{e}_1 \mathbf{e}_1^\dagger = \mathbf{E}_1$$

Tensor Space

- A number of Hilbert spaces $\mathcal{H}^1, \ldots, \mathcal{H}^n$ can be combined to a composite tensor space $\mathcal{H} = \mathcal{H}^1 \otimes \ldots \otimes \mathcal{H}^n$
- If $\left\{\mathbf{e}^i_j\right\}$ is an orthonormal basis of $\mathcal{H}^i$, then

$$\bigotimes_{i,j} \mathbf{e}^i_j$$

(the tensor product of all combination of basis vectors) is an orthonormal basis of the tensor space

Tensor Space

Example

$\mathbb{C}_1^2 \otimes \mathbb{C}_2^2$ is a 4-dimensional space $\mathbb{C}^4$ with base vectors

$$\left\{\mathbf{e}_1^1 \otimes \mathbf{e}_1^2, \mathbf{e}_1^1 \otimes \mathbf{e}_2^2, \mathbf{e}_2^1 \otimes \mathbf{e}_1^2, \mathbf{e}_2^1 \otimes \mathbf{e}_2^2\right\}$$

($\left\{\mathbf{e}_i^1\right\}$ and $\left\{\mathbf{e}_j^2\right\}$ base vectors of $\mathbb{C}_1^2$ and $\mathbb{C}_2^2$, respectively)

Tensor Space

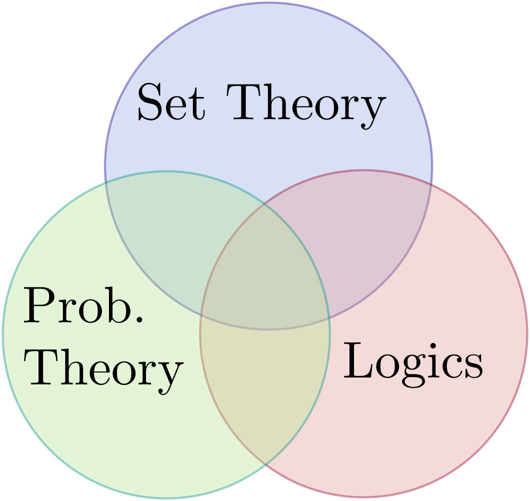
Product Operators

- If **A** is an operator in $\mathcal{H}^1$ and **B** is an operator in $\mathcal{H}^2$, then $\mathbf{A} \otimes \mathbf{B}$ is an operator in $\mathcal{H}^1 \otimes \mathcal{H}^2$ and it is

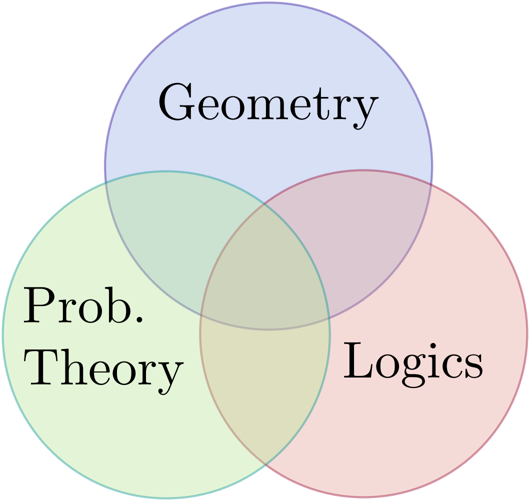
$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{a} \otimes \mathbf{b}) = \mathbf{Aa} \otimes \mathbf{Bb}$$

  for $\mathbf{a} \in \mathcal{H}^1$, $\mathbf{b} \in \mathcal{H}^2$ and $\mathbf{a} \otimes \mathbf{b} \in \mathcal{H}^1 \otimes \mathcal{H}^2$.
- A matrix representation for the tensor product is given by the *Kronecker product* (see also [Nielsen and Chuang, 2000, p. 74])

Quantum Probabilities



Quantum Probabilities



Quantum and Classical Probabilities

- Quantum probabilities are used in quantum theory to describe the behaviour of matter at atomic and subatomic scales
- Quantum probabilities are a generalisation of classical probability theory

Quantum and Classical Probabilities

Correspondance

Sample space	Set	Hilbert space
Atomic event	Element	Ray
Event	Subset	Subspace
Null element	Empty set	Empty space
Membership	Indicator function	Projector
Exclusiveness	Empty intersection	Empty intersection

Dirac Notation

In many textbooks on quantum mechanics a different (and quite handy) notation is used for vectors, the so-called Dirac notation (named after Paul Dirac).

Dirac Notation

A vector $\mathbf{y}$ in a Hilbert space $\mathcal{H}$ is represented by a $|y\rangle$, a ket. A bra $\langle x|$ denotes a *linear functional* (a map $f : \mathcal{H} \longmapsto \mathbb{K}$). Thus the bra(c)ket $\langle x|y\rangle$ denotes the *inner product* $\langle \mathbf{x}, \mathbf{y}\rangle = \mathbf{x}^\dagger \mathbf{y} = f(\mathbf{y})$.

- $\langle x| = (|x\rangle)^\dagger$ (adjoint) if $\mathbb{K} = \mathbb{C}$
- $\langle x| = (|x\rangle)^T$ (transpose) if $\mathbb{K} = \mathbb{R}$
- Bra linear: $\langle x|(\alpha|y\rangle + \beta|z\rangle) = \alpha\langle x|y\rangle + \beta\langle x|z\rangle$

Dyads

- Dyads are a special class of operators
- Outer product of a ket and a bra (a matrix!): $|x\rangle\langle y|$
- In particular useful to describe (projectors onto) rays: $\mathbf{P}_x = |x\rangle\langle x|$
- Example:

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

Notation

Vector	$\mathbf{x}$ or $\|x\rangle$
Adjoint	$\mathbf{x}^\dagger$ or $\langle x\|$ or $\mathbf{A}^\dagger$ (for matrices)
Inner product	$\langle \mathbf{x}, \mathbf{y} \rangle$ or $\mathbf{x}^\dagger \mathbf{y}$ or $\langle x\|y\rangle$
Projector	$\mathbf{S}$ projects onto subspace S (and represents it)
Ray	$\mathbf{P}_x$ ray projector determined by $\|x\rangle$
Standard probability	$\Pr$
Quantum probability	$\widehat{\Pr}$

Events

- An event S in quantum probabilities is described by the subspace $\mathbf{S}$
- Atomic events are 1-dimensional subspaces (rays)
- Combination of events (a glimpse into quantum logics):
  - Join $\vee$: spanning subspace
  - Meet $\wedge$: biggest included subspace
  - Complement $^{\perp}$: orthogonal subspace
- If $\mathbf{S}_1$ and $\mathbf{S}_2$ commute:
  - $\mathbf{S}_1 \vee \mathbf{S}_2 = \mathbf{S}_1 + \mathbf{S}_2 - \mathbf{S}_1\mathbf{S}_2$
  - $\mathbf{S}_1 \wedge \mathbf{S}_2 = \mathbf{S}_1\mathbf{S}_2$
  - $\mathbf{S}^{\perp} = \mathbf{I} - \mathbf{S}$ ($\mathbf{I}$ identity matrix)

Quantum Logic

Violation of distributive law

Let

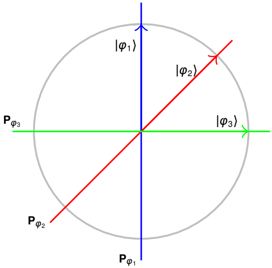
$$|\varphi_1\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad |\varphi_2\rangle = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \qquad |\varphi_3\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Then

$$(\mathbf{P}_{\varphi_1} \wedge \mathbf{P}_{\varphi_2}) \vee \mathbf{P}_{\varphi_3} = \mathbf{P}_{\varphi_3}$$

but

$$(\mathbf{P}_{\varphi_1} \vee \mathbf{P}_{\varphi_3}) \wedge (\mathbf{P}_{\varphi_2} \vee \mathbf{P}_{\varphi_3}) = \mathbf{I}$$



The Classical Case

- Quantum logics and probability reduces to classical logic and probability if all measures are compatible
- Compatibility basically means that the involved projectors are commuting (so the order does not matter)
- This happens when all event rays are orthogonal

Quantum states

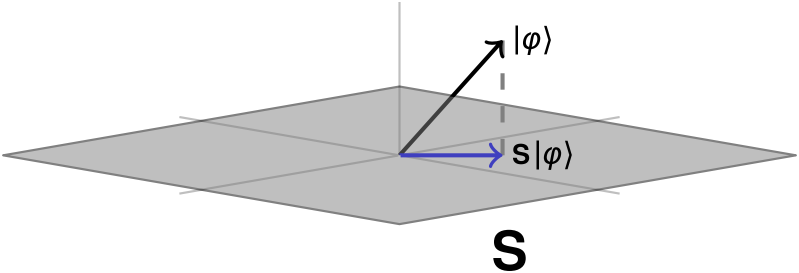
- Quantum probabilities are induced by quantum states
- A pure quantum state (or *pure state*) is represented by a *unit vector* $|\varphi\rangle$ (the state vector) in the Hilbert space $\mathcal{H}$
- A mixed (quantum) state is a probabilistic mixture of pure states
- Both kinds of states can be described by probability densities (a matrix with trace 1)

Quantum Probabilities: Pure state

Quantum probability (pure state)

The probability of the event S given the state φ is the squared length of the projection onto the corresponding subspace **S**:

$$\widehat{\Pr}(S|\varphi) = ||\mathbf{S}\,|\varphi\rangle||^2$$



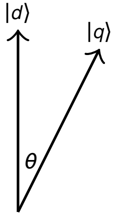
Vector Space Model

An example "quantum" system

- We are now ready to build our first "quantum" IR system
- Vector space model: document d and query q normalised vectors $|d\rangle$ and $|q\rangle$ in term space $\mathbb{R}^n$

- $|q\rangle$ state vector, $\mathbf{P}_d = |d\rangle\langle d|$

$$\widehat{\Pr}(d|q) = ||\mathbf{P}_d\,|q\rangle||^2 = |||d\rangle\,\langle d|q\rangle||^2$$
$$= ||\langle d, q\rangle||^2 = \cos^2(\theta)$$
$$= ||\langle q, d\rangle||^2 = |||q\rangle\,\langle q|d\rangle||^2$$
$$= \left|\left|\mathbf{P}_q\,|d\rangle\right|\right|^2 = \widehat{\Pr}(q|d)$$

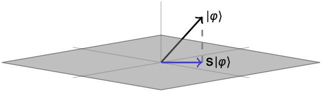


$|d\rangle$ state vector, $\mathbf{P}_q = |q\rangle\langle q|$ (dual view)

Effect of Measurement

- We have seen how we can express the probability of and event S
- What happens if we *observe* or *measure* that an event occurs (for instance the relevance of a document)?
- The state needs to be updated
- For a pure state φ this is just the *normalised projection*

$$\left|\varphi'\right\rangle = |\varphi\rangle \triangleright \mathbf{S} = \frac{\mathbf{S}\,|\varphi\rangle}{||\mathbf{S}\,|\varphi\rangle||}$$

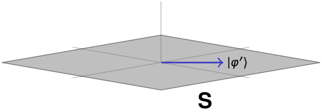


- An immediate observation of the same event would not change the state any more ($\mathbf{S}\,|\varphi'\rangle = |\varphi'\rangle$, $\mathbf{S}$ idempotent!)
- $\widehat{\Pr}(S|\varphi') = 1$

Effect of Measurement

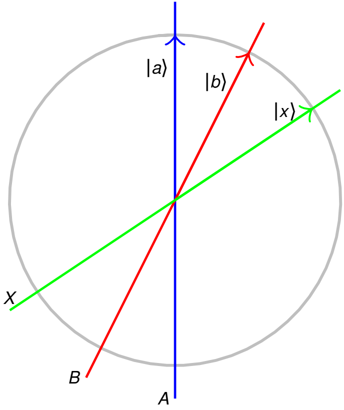
- We have seen how we can express the probability of and event S
- What happens if we *observe* or *measure* that an event occurs (for instance the relevance of a document)?
- The state needs to be updated
- For a pure state φ this is just the *normalised projection*

$$\left|\varphi'\right\rangle = |\varphi\rangle \triangleright \mathbf{S} = \frac{\mathbf{S}\,|\varphi\rangle}{\|\mathbf{S}\,|\varphi\rangle\|}$$



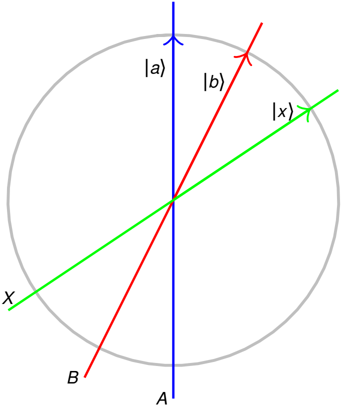
- An immediate observation of the same event would not change the state any more ($\mathbf{S}\,|\varphi'\rangle = |\varphi'\rangle$, $\mathbf{S}$ idempotent!)
- $\widehat{\Pr}(S|\varphi') = 1$

Order Effects



- Quantum probabilities provide a theory for explaining *order effects*
- Such effects appear when incompatible measures are involved
- We sketch this with a simple example

Order Effects



- 3 Events A,B,X
  $\mathbf{A} = |a\rangle\langle a|$
  $\mathbf{B} = |b\rangle\langle b|$
  $\mathbf{X} = |x\rangle\langle x|$
  All non-commutative!
- $\widehat{\Pr}(X|AB) \neq \widehat{\Pr}(X|BA)$
- The probability that we observe X is different if we observed A then B or if we observed B then A
  - [van Rijsbergen, 2004]: Determining relevance then aboutness is not the same as determining aboutness then relevance

Interference

The Double Slit Experiment

(Taken from [Feynman, 1951])

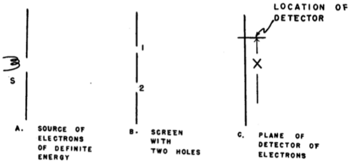


FIGURE 1
An experiment to determine the probability that electrons arrive at a detector at X

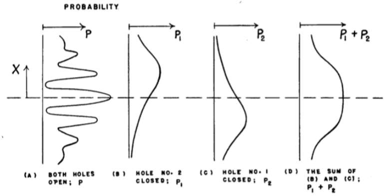


FIGURE 2
Results of the experiment. Probability of arrival of electrons at X plotted against the position X of the detector.

- Physical experiment that motivates interference
  - Some works (e.g. [Zuccon et al., 2009, Melucci, 2010b]) use this analogy for IR
- Particle either passes slit 1 or slit 2 before it appears somewhere on the screen
- Probability $\Pr(x)$ that it appears at position x?

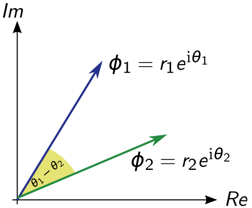
Classical Kolmogorovian probabilities:

Interference

Interference Term

$I_{12} = 2 \cdot \sqrt{\widehat{\Pr}_1(x)}\sqrt{\widehat{\Pr}_2(x)} \cdot \cos(\theta_1 - \theta_2)$ is called the interference term

It also depends on the phase $\theta_1 - \theta_2$ of the two complex numbers involved



$I_{12} = 0 \Leftrightarrow \cos(\theta_1 - \theta_2) = 0 \Leftrightarrow \theta_1 - \theta_2 = \frac{\pi}{2} + k\pi$
(both numbers are perpendicular in the complex plane)

Composite Systems

Tensor Spaces

- Quantum systems (Hilbert spaces) can be combined using the tensor product (see also [Griffiths, 2002])
- If $|\varphi_i\rangle \in \mathcal{H}^i$ is the state of system i then

$$\bigotimes_i |\varphi_i\rangle$$

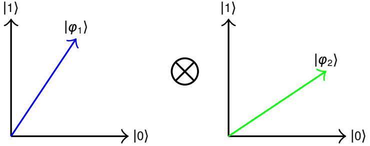
is the state in the composite system $\bigotimes_i \mathcal{H}^i$ (product state)

- Let S_i be a subspace (event) in $\mathcal{H}^i$. Then

$$\widehat{\Pr}\left(\bigotimes_i S_i | \bigotimes_i \varphi_i\right) = \prod_i \widehat{\Pr}(S_i|\varphi_i) \tag{1}$$

Composite Systems

2 Qubit Example, Separable State



Combining two qubits with

$$|\varphi_1\rangle = a_1 |0\rangle + a_2 |1\rangle$$
$$|\varphi_2\rangle = b_1 |0\rangle + b_2 |1\rangle$$

- State of composite system is the product state

  $$|\varphi_1\rangle \otimes |\varphi_2\rangle = a_1 b_1 |00\rangle + a_1 b_2 |01\rangle + a_2 b_1 |10\rangle + a_2 b_2 |11\rangle$$

  (with, e.g., $|01\rangle = |0\rangle \otimes |1\rangle$)
- If composite state is a product state, it is said to be separable
- Both systems are independent – if we measure, say, $|1\rangle$in the first qubit[1], we can still measure either $|0\rangle$ or $|1\rangle$ in the second one!
- Bivariate distribution with a_i, b_i as marginals [Busemeyer, 2012]

[1] Expressed by the subspace $|10\rangle\langle 10| + |11\rangle\langle 11|$

Composite Systems

Entanglement

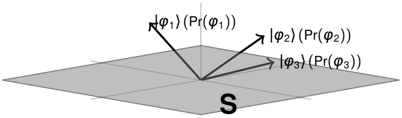
- There are states in a composite system that cannot be expressed as product states
- For example in the 2 qubit system,

$$|\varphi\rangle = \frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle$$

  is such a non-separable state
- The systems are not independent any more – if for instance we measure $|0\rangle$ in the first qubit, this means the second qubit will be in state $|0\rangle$!
- The composite system is in an entangled state
- Equation 1 does not hold any more

Mixed State

- In general we don't know the state of the system (or in IR we don't know what the user really wants)
- We assume the system to be in a certain state with a certain (classical!) probability



Mixed (Quantum) State

We assume the system is in a state φ_i with probability $\Pr(\varphi_i)$ so that $\sum_i \Pr(\varphi_i) = 1$. Then the probability of an event S is

$$\widehat{\Pr}(S) = \sum_i \Pr(\varphi_i)\widehat{\Pr}(S|\varphi_i) = \sum_i \Pr(\varphi_i) \, ||\mathbf{S}\, |\varphi_i\rangle||^2$$

The Effect of Measurement

Mixed State

- When observing/measuring S all state vectors are projected and renormalised, resulting in a new state set $\{\psi_i\}$
- The probability of each new state ψ_i is computed as follows:

$$\Pr(\psi_i|S) = \sum_{\varphi|\psi_i=|\varphi\rangle \triangleright \mathbf{S}} \frac{\widehat{\Pr}(S|\varphi)\Pr(\varphi)}{\widehat{\Pr}(S)}$$

- $\varphi|\psi_i = |\varphi\rangle \triangleright \mathbf{S}$ means all vectors that have the same normalised projection $|\psi_i\rangle$
- Conditional quantum probabilities:

$$\widehat{\Pr}(S_2|S_1) = \sum_{\psi} \widehat{\Pr}(S_2|\psi)\Pr(\psi|S_1)$$

Preliminaries: Trace

Trace

$$\mathrm{tr}(\mathbf{T}) = \sum_{i=1}^{n} \langle e_i | \mathbf{T} | e_i \rangle$$

is known as the trace of **T** with $\{|e_i\rangle\}$ as an orthonormal basis. It is equal to the sum of the diagonal elements of **T**.

Some important properties (see [van Rijsbergen, 2004, p. 79]):

- Linearity: $\mathrm{tr}(\alpha\mathbf{T_1} + \beta\mathbf{T_2}) = \alpha\mathrm{tr}(\mathbf{T_1}) + \beta\mathrm{tr}(\mathbf{T_2})$
- Cyclic permutation: e.g. $\mathrm{tr}(\mathbf{T_1}\mathbf{T_2}) = \mathrm{tr}(\mathbf{T_2}\mathbf{T_1})$
- $\mathrm{tr}\mathbf{T}^\dagger = \overline{\mathrm{tr}(\mathbf{T})}$
- $\mathrm{tr}(\mathbf{T}) \geq 0$ if **T** is positive definite
- An operator **T** is of trace class if **T** is positive and its trace is finite

Trace and Probability

Density Operator

$$\widehat{\Pr}(S) = \sum_i \Pr(\varphi_i)\,||\mathbf{S}\,|\varphi_i\rangle||^2$$

$$= \sum_i \Pr(\varphi_i)\langle \mathbf{S}\varphi_i|\mathbf{S}\varphi_i\rangle \qquad \text{Def. norm}$$

$$= \sum_i \Pr(\varphi_i)\,\langle \varphi_i|\,\mathbf{S}\,|\varphi_i\rangle \qquad \mathbf{S} \text{ self-adjoint, idempotent}$$

$$= \sum_i \Pr(\varphi_i)\,\mathrm{tr}(\mathbf{S}\,|\varphi_i\rangle\langle\varphi_i|) \qquad \text{[Nielsen and Chuang, 2000, p. 76]}$$

$$= \mathrm{tr}\left(\mathbf{S}\underbrace{\sum_i \Pr(\varphi_i)\,|\varphi_i\rangle\langle\varphi_i|}_{=\rho}\right) \qquad \text{Trace linearity}$$

Density Operator

- We saw that $\widehat{\Pr}(S) = \mathrm{tr}(\mathbf{S}\rho)$
- ρ is a density operator (usually a *density matrix*)
- ρ encodes a quantum probability distribution (either mixed or pure)

Density Operator

A density operator ρ is a trace-class operator with $\mathrm{tr}(\rho) = 1$.

Simple example (2 events with probability 1/2):

$$\rho = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$$

Gleason's Theorem

Gleason's Theorem

Let μ be any measure on the closed subspaces of a separable (real or complex) Hilbert space $\mathcal{H}$ of dimension of at least 3. There exists a positive self-adjoint operator $\mathbf{T}$ of trace class such that, for all closed subspaces S of $\mathcal{H}$,

$$\mu(S) = \mathrm{tr}(\mathbf{TS})$$

- If μ is a probability measure, then $\mathrm{tr}(\mathbf{T}) = 1$, so $\mathbf{T}$ is a density operator.

Gleason's Theorem

In other words...

Following the Piwowarski/Melucci tutorial:

Distribution over a Hilbert Space

A distribution over a Hilbert space $\mathcal{H}$ is any function $\widehat{\Pr} : S \subseteq \mathcal{H} \mapsto [0, 1]$ such that:

- $\widehat{\Pr}(\emptyset) = 0$ and $\widehat{\Pr}(\mathbf{P}_\varphi) \geq 0 \quad \forall \varphi \in \mathcal{H}$
- $\sum_i \widehat{\Pr}(\mathbf{P}_{e_i}) = 1$ for any basis $\{e_i\}$

Gleason's Theorem

To every probability distribution over a Hilbert space $\mathcal{H}$ (dimension ≥ 3), there exists a unique density matrix ρ such that for any $S \subseteq \mathcal{H}$

$$\widehat{\Pr}(S) = \mathrm{tr}(\rho \mathbf{S})$$

What does this mean?

- Gleason's theorem provides a 1-to-1 relationship between quantum probability distributions and density operators
- We can approximate density operators using spectral techniques, decompositions etc.

Some Observations

- If the system is in a pure state φ, the density operator is $\rho = |\varphi\rangle\langle\varphi| \rightsquigarrow$ pure distributions are represented by projectors
- Density matrices are Hermitian
- Applying the spectral theorem, we can decompose ρ:

$$\rho = \sum_i p_i \mathbf{E_i}$$

  The $p_i \geq 0$ (with $\sum_i p_i = 1$ and $p_i \in \mathbb{R}$) are the eigenvalues and probabilities associated to the events represented by the projectors E_i
- Example

$$\rho = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

Measurement and Conditional Probabilities

- Update of density matrix after measuring/observing S_1:

$$\rho' = \frac{\mathbf{S_1}\rho\mathbf{S_1}}{\mathrm{tr}(\mathbf{S_1}\rho\mathbf{S_1})} = \frac{\mathbf{S_1}\rho\mathbf{S_1}}{\mathrm{tr}(\rho\mathbf{S_1})}$$

- Lüders' Rule for conditional probabilities:

$$\widehat{\Pr}(S_2|S_1) = \frac{\mathrm{tr}(\mathbf{S_1}\rho\mathbf{S_1}\mathbf{S_2})}{\mathrm{tr}(\rho\mathbf{S_1})}$$

If $\mathbf{S_1}$ and $\mathbf{S_2}$ are compatible, this reduces to classical conditionalisation (see [Hughes, 1992, p. 224])

Tensor product

- Let ρ_i be the state of the system represented by $\mathcal{H}^i$. Then

$$\rho_i \otimes \ldots \otimes \rho_n$$

  is the state of the composite system $\mathcal{H}^1 \otimes \ldots \otimes \mathcal{H}^n$
- $\widehat{\Pr}(\bigotimes_i S_i | \bigotimes_i \rho_i) = \prod_i \widehat{\Pr}(S_i|\rho_i)$
- Note that there can be a state ρ in the composite system that is not separable any more

Quantum Probabilities

Conclusion

- Introduced quantum probabilities based on Hilbert Spaces
- Some salient features: pure and mixed states, densities, measurement/observation, order effects, interference, composite systems, entanglement
- Quantum probabilities as a generalisation of classical probabilities

Now: How can we use this for information retrieval?

Again: What does this remind you of?

Quantum-inspired IR Models

- The QIA Model
  - Quantum Information Access
  - Polyprepresentation
- Quantum Probability Ranking Principle

Quantum Information Access (QIA)

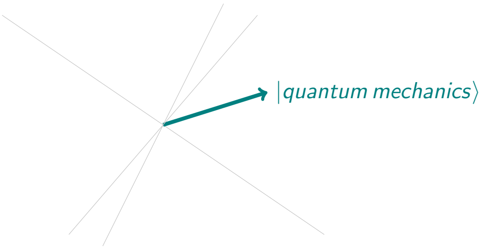
Notation Wrap-Up

- Hilbert space: vector space with an inner product
- Dirac Notation:
  $|\varphi\rangle$ is a ket (a vector φ)
  $\langle\varphi|$ is a bra (a transposed vector φ^{T})
  $\langle\varphi|\psi\rangle \in \mathbb{C}$ is a bra(c)ket (inner product)
  $|\varphi\rangle\langle\psi|$ is a ketbra (a matrix)
- A subspace S is represented by a projector (another matrix)
- $|\varphi\rangle\langle\varphi|$ projector onto 1-dimensional subspace
- Orthogonal projection of vector $|\varphi\rangle$ onto S: $S|\varphi\rangle$

Assumptions underlying QIA

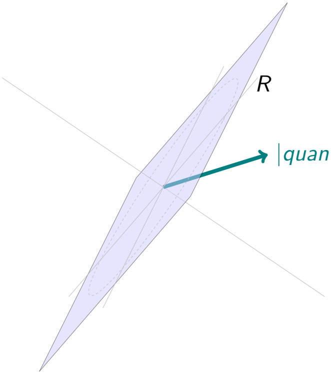
- IR system uncertain about user's information need (IN)
  - System view of the user's IN becomes more and more specific through interaction
- The IN may change from the user's point of view
- There is an IN Space, a Hilbert space

Information Need Space



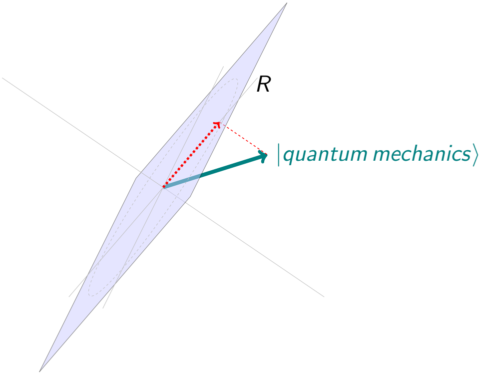
- INs as vectors: IN vector $|\varphi\rangle$
- Event "document d is relevant" represented by subspace R
- Probability of relevance: squared length of projection $\Pr(R|d,\varphi) = ||R|\varphi\rangle||^2$
- Unit vector imposes relevance distribution on subspaces (events)

Information Need Space



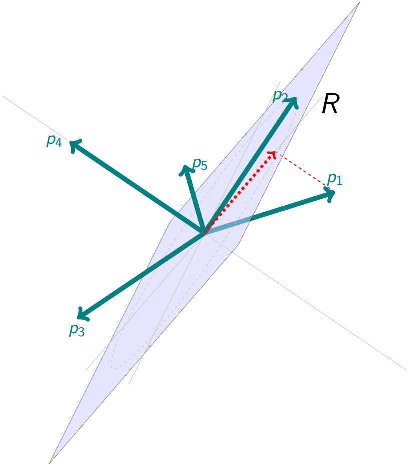
- INs as vectors: IN vector $|\varphi\rangle$
- Event "document *d* is relevant" represented by subspace *R*
- Probability of relevance: squared length of projection $\Pr(R|d,\varphi) = ||R|\varphi\rangle||^2$
- Unit vector imposes relevance distribution on subspaces (events)

Information Need Space



- INs as vectors: IN vector $|\varphi\rangle$
- Event "document *d* is relevant" represented by subspace *R*
- Probability of relevance: squared length of projection $\Pr(R|d,\varphi) = ||R|\varphi\rangle||^2$
- Unit vector imposes relevance distribution on subspaces (events)

System's Uncertainty about User's Intentions



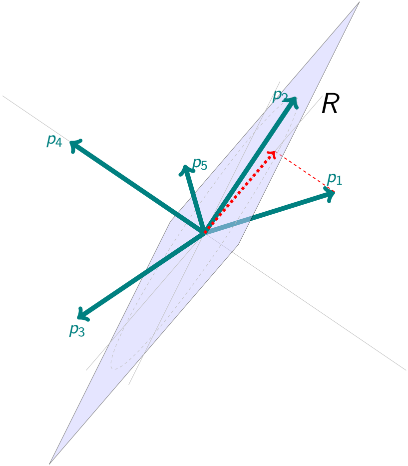
- System uncertain about user's IN
- Expressed by an ensemble S of possible IN vectors (density ρ):

$$S = \{(p_1, |\varphi_1\rangle), \ldots, (p_n, |\varphi_n\rangle)\}$$

- Probability of relevance:

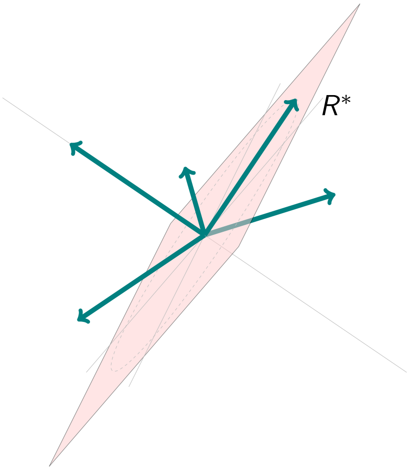
$$\Pr(R|d,S) = \sum_i p_i \cdot \underbrace{\Pr(R|d,\varphi_i)}_{=||R|\varphi\rangle||^2}$$

System's Uncertainty about User's Intentions



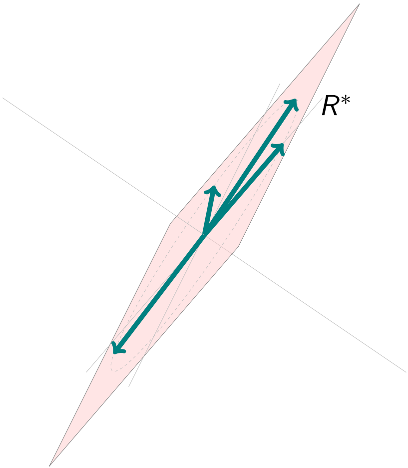
- Dual representation using density operator and trace function
- $\rho = \sum_i p_i \cdot |\varphi_i\rangle\langle\varphi_i|$
- $\Pr(R|d, S) = \mathrm{tr}(\rho R)$

User Interaction and Feedback



- Outcome of feedback: Query and query reformulation, (click on) relevant document, ...
- Expressed as subspace
- Project IN vectors onto document subspace
- Document now gets probability 1
- System's uncertainty decreases
- Also reflects changes in information needs

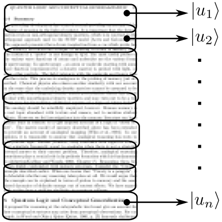
User Interaction and Feedback

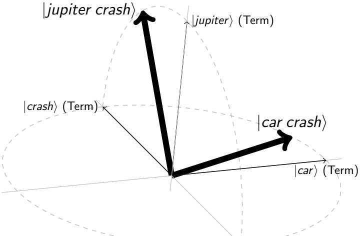


- Outcome of feedback: Query and query reformulation, (click on) relevant document, ...
- Expressed as subspace
- Project IN vectors onto document subspace
- Document now gets probability 1
- System's uncertainty decreases
- Also reflects changes in information needs

Textual Representation

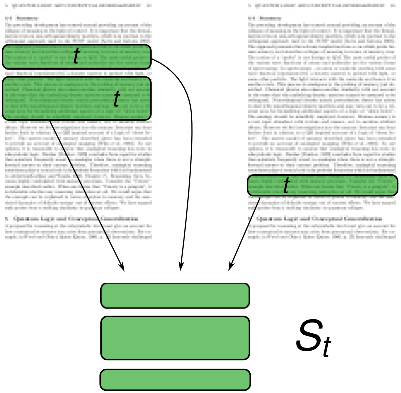
IN Space / Documents





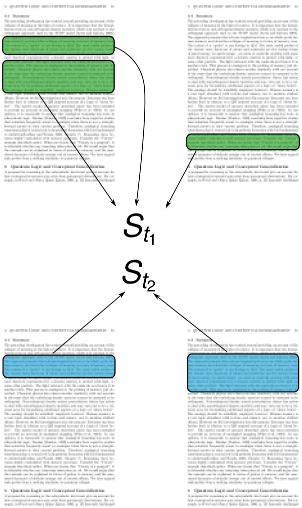
- IN space based on term space
- IN vectors made of document fragments
- Weighting scheme (e.g., tf, tf-idf,...)
- Document is relevant to all INs found in its fragments
- Document subspace *R* spanned by IN vectors
- No length normalisation necessary

Single Query Term



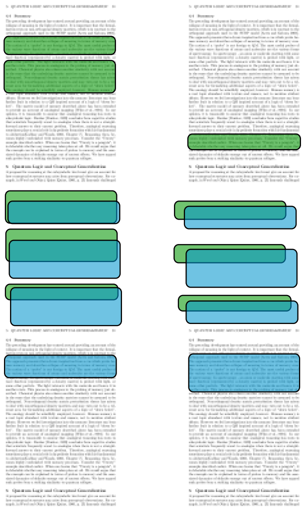
- Take all fragments vectors (IN vectors) containing term t
- This makes up ensemble S_t

Mixture



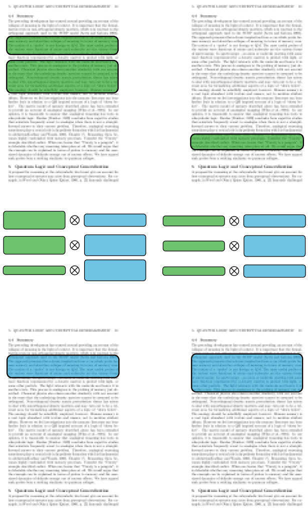
- Mixture of all combinations of term fragments
- Document must at least satisfy one term fragment
- The more term fragments are contained, the more relevant a document is
- $S^{(M)} = \sum_{i=1}^{n} w_i S_{t_i}$
- w_i is term weight

Mixture of Superposition



- Superpose all combinations (e.g. $\frac{1}{\sqrt{2}}(|\varphi\rangle + |\psi\rangle)$)
- At least one query term fragment superposition must be contained
- The more fragment superpositions are contained, the more relevant a document is
- Indication that it works well with multi-term concepts (e.g. “digital libraries”)

Tensor product



- Assumption: each term covers an IN aspect
- Tensor product of all fragment vectors $\rightsquigarrow$ combination of IN aspects
- Document must satisfy all IN aspects
- The more tensor products are satisfied, the more relevant is the document

What can it bring to IR?

- Evaluation with several TREC collections [Piwowarski et al., 2010]
- Tensor representation of query could compete with BM25
- We don't lose retrieval effectiveness in an ad hoc scenario
- Framework is open for possible extension:
    - Different forms of interactions (query reformulations, relevance judgements) $\rightsquigarrow$ sessions
    - Diversity and novelty
    - Structured queries (Boolean; based on mixture, superposition and tensor)
    - Polyrepresentation [Frommholz et al., 2010]

Book Store Example

Books › **"quantum mechanics good introduction"**

Showing 10 Results — Sort by: Avg. Customer Review

1.

Game of Life Cellular Automata by Andrew Adamatzky (**Hardcover** - 5 Jul 2010)

Buy new: ~~£108.00~~ **£99.41**

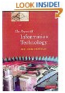
5 new from £99.41 4 used from £117.56

Get it by **Tuesday, Aug 17** if you order in the next **47 hours** and choose express delivery.

Eligible for **FREE** Super Saver Delivery.

Excerpt - page 465: " **...** Rolf Laundauer [21] 23.1 Introductory Concepts of **Quantum** Mechanics A **good introduction** to all aspects of **quantum** computation is provided by a number of recent books"

2.



The Physics of Information Technology (Cambridge Series on Information and the Natural Sciences) by Neil Gershenfeld (**Hardcover** - 16 Oct 2000)

Buy new: ~~£56.00~~ **£47.04**

10 new from £47.03 6 used from £30.90

Get it by **Tuesday, Aug 17** if you order in the next **50 hours** and choose express delivery.

Eligible for **FREE** Super Saver Delivery.

Excerpt - page 284: " **...** 1973). Lectures on **Quantum** Mechanics. Reading: W.A. Benjamin. A **good** intuitive **introduction** to **quantum** mechanics. [Peres, 1993] Peres, Asher. (1993). **Quantum** Theory **...** "

3.

An Introduction to Quantum Computing Algorithms (Progress in Computer Science and Applied Logic (PCS)) by Arthur O. Pittenger (**Hardcover** - 12 Nov 1999)

Buy new: ~~£53.99~~ **£51.29**

12 new from £36.89 7 used from £21.17

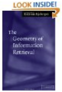
Usually dispatched within 1 to 3 weeks

Eligible for **FREE** Super Saver Delivery.

★★★★★ (2)

Excerpt - page 2: " **...** development, the reader is also referred to Shankar 1631 and Sakurai [601 for a **good introduction** to the mathematics of **quantum** mechanics"

4.



The Geometry of Information Retrieval by C. J. van Rijsbergen (**Hardcover** - 12 Aug 2004)

Buy new: ~~£39.00~~ **£37.05**

12 new from £29.95 3 used from £30.95

Get it by **Tuesday, Aug 17** if you order in the next **50 hours** and choose express delivery.

Eligible for **FREE** Super Saver Delivery.

Excerpt - page 13: "philosophically minded, Barrett (1999) is worth reading. There are several **good** popular introductions to **quantum** mechanics, for example Penrose (1989, 1994), Polkinghorne (1986 **...** "

Book Store Example

Customer Reviews

An Introduction to Quantum Computing Algorithms (Progress in Computer Science and Applied Logic (PCS))

2 Reviews

5 star:		(2)
4 star:		(0)
3 star:		(0)
2 star:		(0)
1 star:		(0)

Average Customer Review
★★★★★ (2 customer reviews)

Share your thoughts with other customers

Create your own review

Search Customer Reviews

GO!

Only search this product's reviews

Most Helpful First | Newest First

★★★★★ **An invitation.**, 10 April 2003

By **Palle E T Jorgensen "Palle Jorgensen"** (Iowa City, Iowa United States) - See all my reviews

TOP 500 REVIEWER REAL NAME

This review is from: An Introduction to Quantum Computing Algorithms (Progress in Computer Science and Applied Logic (PCS)) (Hardcover)

A handful of good introductions to ideas in quantum computing have appeared in the past two years. The present one stands out in being both friendly and brief. There is no way into the subject, getting around the fundamentals in quantum physics and in math. Through this little book, an uninitiated reader can get some insight into the ideas of Deutsch-Jozsa, and the algorithms of Peter Shor and Lov Grover. The author does his job, as well as any, and the book is pleasant reading.

Help other customers find the most helpful reviews

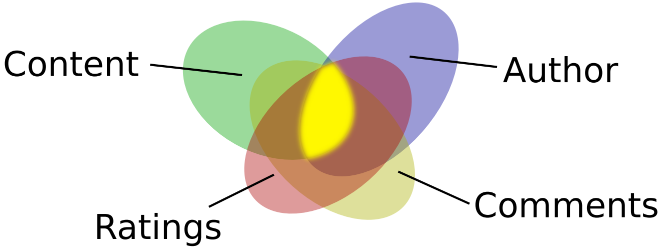
Was this review helpful to you? Yes No

Report this | Permalink

Comment

The Principle of Polyrepresentation

Book Store Scenario



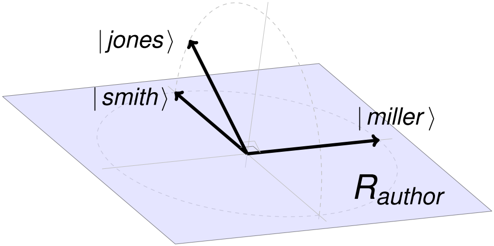
1. Get ranking for different representations
2. Find the cognitive overlap

- Based on different document representations, but also different representations of user's information need
- Hypothesis: cognitive overlap contains highly relevant documents (experiments support this)

How can we apply this in QIA?

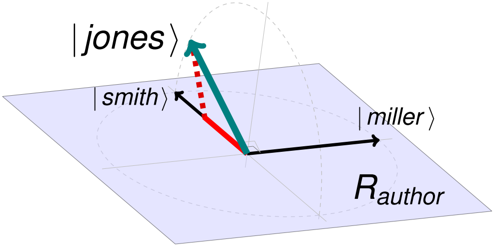
- Model single representations in a vector space (by example)
  - Authors
  - Ratings
- Combine the representations

Example: Author Space



- Each author is a dimension
- *Non-orthogonal* vectors: *dependencies*
- Angle between vectors reflects the degree of dependency (90° = orthogonal = upright = independent)
- Example: Jones and Smith (somehow) related, Smith and Miller not

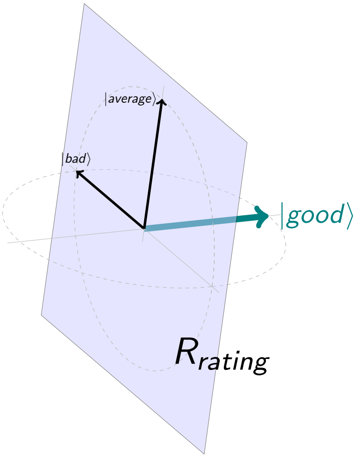
Example: Author Space



- Document by Smith and Miller
- User seeks for documents by Jones
- Document retrieved due to relationship between Jones and Smith

Rating Space

- Example: rating scale good/bad/average – each is a dimension
- "Average" rated book represented by 2-dimensional subspace
- User wants books which are rated good
  $\Rightarrow$ not relevant ($|good\rangle$ orthogonal)



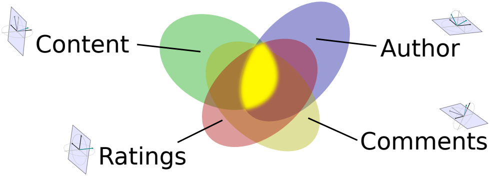
Combining the Evidence

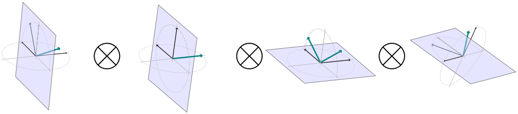
Total Cognitive Overlap and Tensors

- Modelled different representations in vector space
- Probabilities w.r.t. single representations
- How do we express user's IN w.r.t. all representations?
- How do we get a cognitive overlap?

Combining the Evidence

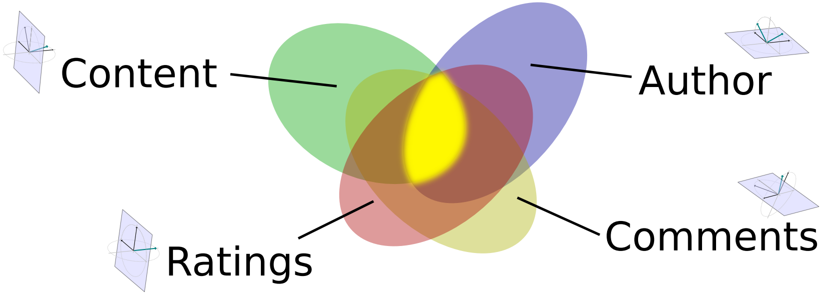
Total Cognitive Overlap and Tensors





- Polyrepresentation space as tensor product (“⊗”) of single spaces
- Probability that document is in total cognitive overlap: $\Pr_{polyrep} = \Pr_{content} \cdot \Pr_{ratings} \cdot \Pr_{author} \cdot \Pr_{comments}$

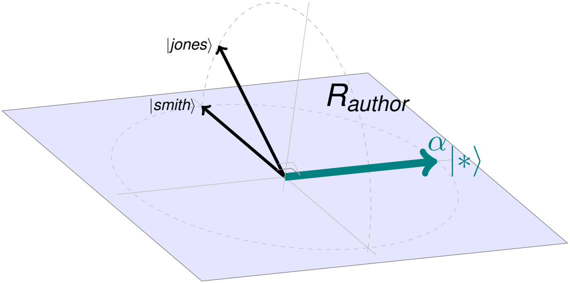
Wishlist



- Documents not relevant in one representation should not get a value of 0
- Ignore selected representations
- Relative importance of representations to user (mixing and weighting)

"Don't care" dimension

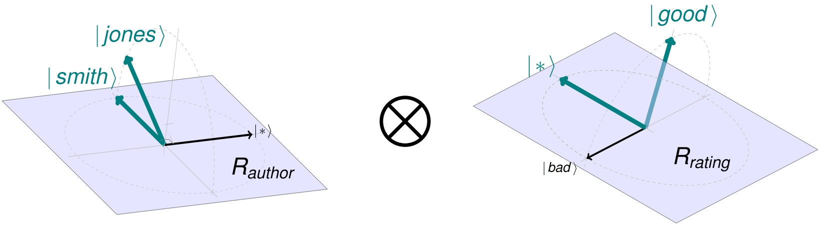
- Introduction of a "don't care" dimension
- Part of each document subspace
  ⇝ each document "satisfies" the don't care "need"
- Example: Document by Smith, user doesn't care about authors with probability α



- $\alpha = 1$ means representation is ignored at all

Example

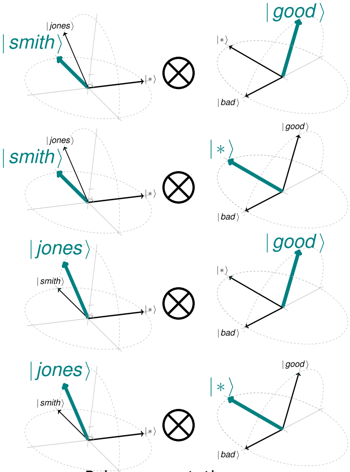
- What the system assumes about the user's IN:
  - Seeks books either by Jones or by Smith
  - Looks either for good books or doesn't care about ratings
- Assume a document *d* by Smith which is rated "bad"



- Polyrepresentation space: 9-dimensional (3×3)

IN Vectors in Polyrepresentation Space

How do they look like and what do they mean?

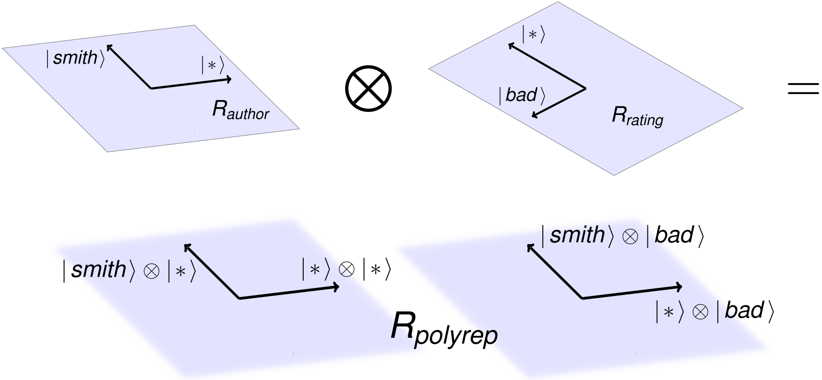


Polyrepresentation space

Reflects all 4 possible combinations of INs w.r.t. single representations:

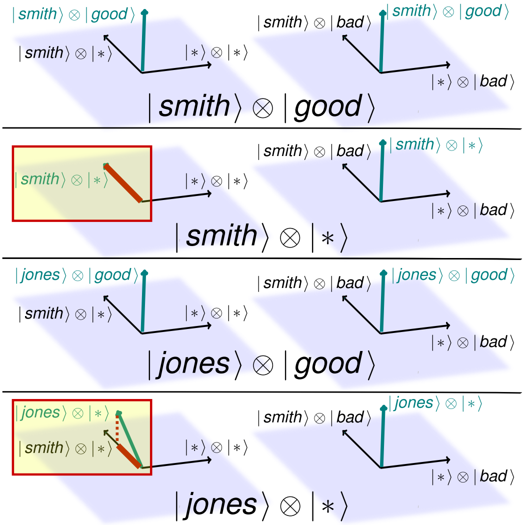
Smith/good: $|smith\rangle \otimes |good\rangle$
Smith/dont' care: $|smith\rangle \otimes |*\rangle$
Jones/good: $|jones\rangle \otimes |good\rangle$
Jones/dont' care: $|jones\rangle \otimes |*\rangle$

Documents in Polyrepresentation Space



- Represented as tensor product of single document subspaces
- Here: 4-dimensional subspace (2×2)

Determining the Retrieval Weight



Why the system retrieves the bad book by Smith

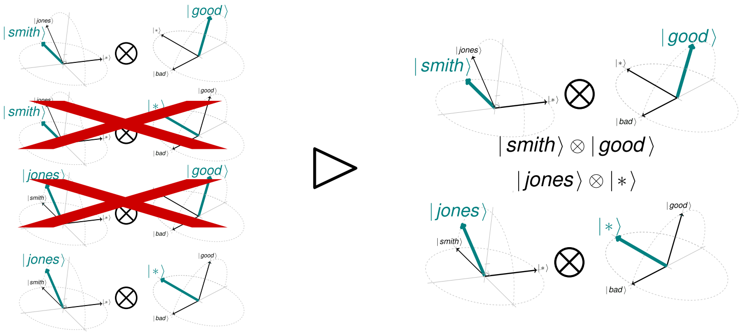
Something left on the Wishlist

Relationships between Representations

System observes (interaction/feedback) user preferences:

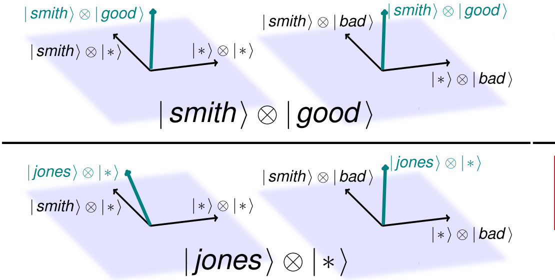
- If book is by Smith, it has to be rated good
- If book is by Jones, don't care about the ratings

System evolves to new state in polyrepresentation space (2 combinations not allowed any more)



What does that mean?

- Only two assumed IN cases left:
  1. Smith/good
  2. Jones/don't care
- Cannot be expressed as combination of single representations
- Bad book by Smith only retrieved due to relationship to Jones!



QIA Summary

We showed how we can express polyrepresentation in a mathematical framework inspired by quantum mechanics:

- Presented IN space
- Modelled different example representations
- Combined representations in polyrepresentation space

QIA Conclusion

- QIA framework
  - User’s IN as ensemble of vectors
  - Documents as subspaces
  - User interaction and feedback
  - Term space, query construction
  - Can compete in an ad hoc scenario
- Polyrepresentation
  - Different non-topical representations as subspaces
  - Polyrepresentation space as tensor space to calculate cognitive overlap
  - “Don’t care” dimension for weighting of representations
  - Non-separate states reflect interdependencies

QIA Extensions

- Queries in sessions [Frommholz et al., 2011]
  - Use geometry and projections to determine type of and handle follow-up query (generalisation, information need drift, specialisation)
- Summarisation [Piwowarski et al., 2012]
  - QIA interpretation of LSA-based methods
- Query algebra for the QIA framework [Caputo et al., 2011]

The Quantum Probability Ranking Principle (qPRP) [Zuccon, 2012]

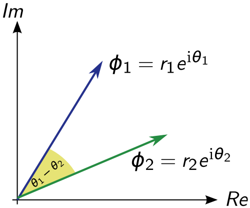
Quantum Interference (again)

- Recall the double slit experiment
- Probability amplitudes instead of probabilities to model interference
- Derived the interference term

$$I = \phi_1 \overline{\phi_2} + \phi_2 \overline{\phi_1} = 2 \cdot \sqrt{\widehat{\Pr}_1(x)} \sqrt{\widehat{\Pr}_2(x)} \cdot \cos(\theta_1 - \theta_2)$$

to compute

$$\widehat{\Pr}_{12}(x) = \widehat{\Pr}_1(x) + \widehat{\Pr}_2(x) + I$$



Probability Ranking Principle

- Probability Ranking Principle (PRP): Rank according to decreasing $\Pr(R|d, q)$
- Which document to present next?

$$\underset{d \in \mathcal{B}}{\text{argmax}} \left(\Pr\left(R|d, q\right)\right)$$

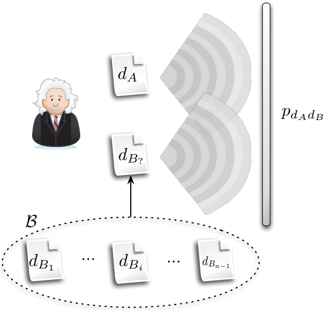
  $\mathcal{B}$: candidate documents not presented yet
- Assumes relevance judgements are independent, no dependencies between documents
- Potentially does not suit tasks considering novelty and diversity well!

Quantum Probability Ranking Principle

Double Slit Analogy

- IR analogy to double slit experiment
  - User is "particle"
  - Each slit corresponds to a document
  - Event that particle passes through slit: user analyses document
  - Screen measures user satisfaction (proportional to $\widehat{\Pr}(R|d,q)$)
- Which document to present next (after user saw d_A)?

$$\underset{d_B \in \mathcal{B}}{\operatorname{argmax}} \left(\widehat{\Pr}(R|d_B, q) + I_{d_A d_B} \right)$$



(taken from [Zuccon, 2012])

Quantum Probability Ranking Principle

Assumptions

1. Ranking is performed sequentially
2. Empirical data is best described by quantum probabilities
3. Relevance of documents not assessed in isolation; documents that have been ranked before influence future relevance assessments

Quantum Probability Ranking Principle (qPRP)

Let $\mathcal{A}$ (where $\mathcal{A} = \varnothing$ is also considered) be the set containing the documents that have been already retrieved until rank $i-1$ and let $\mathcal{B}$ be the set of candidate documents for being retrieved at rank i. In order to maximise its effectiveness, an Information Retrieval system has to rank document $d_B \in \mathcal{B}$ at rank i if and only if

$$\widehat{\Pr}(R|d_B, q) + I_{d_A d_B} \geq \widehat{\Pr}(R|d_C, q) + I_{d_A d_C},$$

for any $d_C \in \mathcal{B}$. $I_{d_A d_B}$ is the sum of the quantum interference terms produced considering all the pairs composed by d_B and each already retrieved document $d_A \in \mathcal{A}$ (similarly for $I_{d_A d_C}$).

Quantum Probability Ranking Principle

- Each document would be selected according to

$$\underset{d_B \in \mathcal{B}}{\operatorname{argmax}} \left(\widehat{\Pr}(R|d_B, q) + \sum_{d_A \in \mathcal{A}} I_{d_A d_B} \right)$$

- Documents may interfere (at relevance level) with already presented ones
- Encodes dependencies between documents

qPRP: Interference Term

Estimation

- Recall

$$I_{d_A d_B} = \phi_{d_A}\overline{\phi_{d_B}} + \phi_{d_B}\overline{\phi_{d_A}}$$
$$= 2 \cdot \sqrt{\widehat{\Pr}(R|d_A, q)}\sqrt{\widehat{\Pr}(R|d_B, q)} \cdot \cos(\theta_{AB})$$

(with $\cos(\theta_{AB}) = \cos(\theta_{d_A} - \theta_{d_B})$)

- Interference governed by the phase difference θ_{AB} between $\phi_{d_A} \in \mathbb{C}$ and $\phi_{d_B} \in \mathbb{C}$
  - Destructive: $\cos(\theta_{AB}) < 0$
  - Constructive: $\cos(\theta_{AB}) > 0$
- Estimate $I_{d_A d_B}$ with *similarity function* f_{sim}

$$\cos(\theta_{AB}) \approx \beta f_{sim}(d_A, d_B)$$

$\beta \in \mathbb{R}$: normalisation, sign of interference

qPRP: Evaluation

- Different similarity functions applied in diversity task (TREC 6,7,8 and ClueWeb B)
- qPRP outperformed PRP and (sometimes) a Portfolio Theory setting (but no parameter tuning required)
- Kullback-Leibler divergence performed best
- Pearson's correlation coefficient seems most robust

qPRP: Evaluation

Further Experiments

- Comparison with PRP, Maximal Marginal relevance (MMR), Portfolio Theory (PT) and interactive PRP (iPRP)
- Ad hoc task
    - PRP best performing ranking approach when independence assumption hold
- Diversity task
    - PRP often outperformed by PT, iPRP and qPRP

qPRP: Discussion

- Kolmogorovian probabilities (in PRP) adequate when using independence assumption (ad hoc task)
- Quantum probabilities seem good choice if documents are not independent and interfere (diversity task)
- qPRP reduces to PRP if phases are perpendicular ($\cos(\theta_{AB}) = 0$)
- Integration into QIA framework possible

Further Models and Conclusion

- Further Work and Software
- Discussion and Conclusion

Further Selected Works

- Quantum probability in context [Melucci, 2008]
- Effective query expansion with quantum interference [Melucci, 2010b]
- Semantics and meaning [Widdows, 2004]
- Entanglement and word meaning [Bruza et al., 2009b, Bruza et al., 2009a]
- Lattice structures and documents [Huertas-Rosero et al., 2009]
- Quantum theory and search [Arafat, 2007]
- Query expansion and query drift [Zhang et al., 2011]
- Document re-ranking [Zhao et al., 2011]
- Complex numbers in IR [Zuccon et al., 2011]
- DB+IR: Commutative Quantum Query Language [Schmitt, 2008]
- Further overview [Song et al., 2010]

Software

Kernel Quantum Probability API by Benjamin Piwowarski

`http://kqp.bpiwowar.net/`

Other Resources

- http://www.quantuminteraction.org/home
  A collection of useful links and resources in the context of the Quantum Interaction series
- http://www.mendeley.com/groups/496611/quantum-and-geometry-ir/
  Aiming at providing an updated list of quantum and geometry IR research papers
- You may also follow me on Twitter: @iFromm

Discussion

Quantum Theory and IR

The quantum formalism is a powerful 'language' for IR – isn't it?

- We've seen some examples of quantum-inspired models (QIA, qPRP)
- Quantum probabilities may give us a hint of what is wrong with existing approaches (but not always!) [Piwowarski et al., 2012]
- But there is criticism: "Ornamental but not useful" (Kantor, who hopes to be proven wrong) [Kantor, 2007]

Conclusion

- Quantum probabilities
- 2 quantum-inspired approaches: QIA and qPRP
- Further approaches

Questions?

Bibliography I

Aerts, D. and Gabora, L. (2005).
A theory of concepts and their combinations II: A Hilbert space representation.
Kybernetes, 34(1/2):192–221.

Arafat, S. (2007).
Foundations Research in Information Retrieval Inspired by Quantum Theory.
Phd thesis, University of Glasgow.

Bibliography II

Bruza, P., Kitto, K., Nelson, D., and McEvoy, C. (2009a).
Extracting Spooky-Activation-at-a-Distance from Considerations of Entanglement.
In *Proc. of the Third Quantum Interaction Symposium (QI 2009)*, volume 5494 of *Lecture Notes in Computer Science*, pages 71–83, Berlin, Heidelberg. Springer Berlin Heidelberg.

Bruza, P., Kitto, K., Nelson, D., and McEvoy, C. (2009b).
Is there something quantum-like about the human mental lexicon?
Journal of mathematical psychology, 53(5):363–377.

Bibliography III

Busemeyer, J. R. (2012).
Introduction to Quantum Probability for Social and Behavioral Scientists.
In Rudolph, L. and Valsiner, J., editors, *Qualitative Mathematics for Social Sciences - Mathematical Models for Research on Cultural Dynamics*. Routledge.

Caputo, A., Piwowarski, B., and Lalmas, M. (2011).
A Query Algebra for Quantum Information Retrieval.
In *Proceedings of the 2nd Italian Information Retrieval Workshop 2011*.

Bibliography IV

Di Buccio, E., Melucci, M., and Song, D. (2011).
Towards Predicting Relevance Using a Quantum-Like Framework.
In Clough, P., Foley, C., Gurrin, C., Jones, G. J. F., Kraaij, W., Lee, H., and Mudoch, V., editors, *Proceedings ECIR 2011*, volume 6611 of *Lecture Notes in Computer Science*, pages 755–758. Springer.

Feynman, R. P. (1951).
The Concept of Probability in Quantum Mechanics.
In *Proc. Second Berkeley Symp. on Math. Statist. and Prob. (Univ. of Calif. Press, 1951)*, pages 533–541.

Bibliography V

- Frommholz, I., Larsen, B., Piwowarski, B., Lalmas, M., Ingwersen, P., and van Rijsbergen, K. (2010).
  Supporting Polyrepresentation in a Quantum-inspired Geometrical Retrieval Framework.
  In *Proceedings of the 2010 Information Interaction in Context Symposium*, pages 115–124, New Brunswick. ACM.

- Frommholz, I., Piwowarski, B., Lalmas, M., and van Rijsbergen, K. (2011).
  Processing Queries in Session in a Quantum-Inspired IR Framework.
  In Clough, P., Foley, C., Gurrin, C., Jones, G. J. F., Kraaij, W., Lee, H., and Mudoch, V., editors, *Proceedings ECIR 2011*, volume

Bibliography VI

6611 of *Lecture Notes in Computer Science*, pages 751–754. Springer.

Griffiths, R. B. (2002).
Consistent quantum theory.
Cambridge University Press, Cambridge; New York.

Halmos, P. R. (1958).
Finite-Dimensional Vector Spaces.
Springer-Verlag New York Heidelberg Berlin.

Huertas-Rosero, A., Azzopardi, L., and van Rijsbergen, C. J. (2009).
Eraser Lattices and Semantic Contents.
In *Proceedings Quantum Interaction 2009*, pages 266–275.

Bibliography VII

- Hughes, R. (1992).
  The Structure and Interpretation of Quantum Mechanics.
  Harvard University Press.
- Kantor, P. B. (2007).
  Keith van Rijsbergen, The Geometry of Information Retrieval.
  Information Retrieval, 10(4-5):485–489.
- Melucci, M. (2008).
  A basis for information retrieval in context.
  ACM Transactions on Information Systems (TOIS), 26(3).

Bibliography VIII

Melucci, M. (2010a).
An Investigation of Quantum Interference in Information Retrieval.
In Cunningham, H., Hanbury, A., and Rüger, S., editors, *Proceedings of the First Information Retrieval Facility Conference (IRFC 2010)*, pages 136–151. Springer.

Melucci, M. (2010b).
An Investigation of Quantum Interference in Information Retrieval.
In Cunningham, H., Hanbury, A., and Rüger, S., editors, *Proceedings of the First Information Retrieval Facility Conference (IRFC 2010)*, pages 136–151. Springer.

Bibliography IX

Melucci, M. and Rijsbergen, K. V. (2011).
Quantum Mechanics and Information Retrieval.
In Melucci, M. and Baeza-Yates, R., editors, *Advanced Topics in Information Retrieval*, volume 33 of *The Information Retrieval Series*, chapter 6, pages 125–154. Springer Berlin Heidelberg, Berlin, Heidelberg.

Nielsen, M. A. and Chuang, I. L. (2000).
Quantum Computation and Quantum Interaction.
Cambridge University Press, Cambridge, UK.

Bibliography X

Piwowarski, B., Amini, M.-R., and Lalmas, M. (2012).
On using a Quantum Physics formalism for Multi-document Summarisation.
Journal of the American Society for Information Science and Technology (JASIST).

Piwowarski, B., Frommholz, I., Lalmas, M., and Van Rijsbergen, K. (2010).
What can Quantum Theory Bring to Information Retrieval?
In *Proc. 19th International Conference on Information and Knowledge Management*, pages 59–68. Springer.

Bibliography XI

Piwowarski, B. and Lalmas, M. (2009).
A Quantum-based Model for Interactive Information Retrieval.
In *Proceedings of the 2nd International Conference on the Theory of Information Retrieval (ICTIR 2009)*, Cambridge, UK.

Schmitt, I. (2008).
QQL : A DB&IR Query Language.
The VLDB Journal, 17(1):39–56.

Song, D., Lalmas, M., van Rijsbergen, K., Frommholz, I., Piwowarski, B., Wang, J., Zhang, P., Zuccon, G., Bruza, P., Arafat, S., Azzopardi, L., Huertas-Rosero, A., Hou, Y., Melucci, M., and Rüger, S. (2010).
How Quantum Theory is Developing the Field of Information Retrieval.

Bibliography XII

In *Proceedings Quantum Interaction 2010*.

Song, D., Melucci, M., Frommholz, I., Zhang, P., Wang, L., and Arafat, S., editors (2011).
Quantum Interaction: 5th International Symposium, QI 2011, Aberdeen, UK, Revised Selected Papers.
Springer Berlin Heidelberg, lecture no edition.

van Rijsbergen, C. J. (2004).
The Geometry of Information Retrieval.
Cambridge University Press, New York, NY, USA.

Widdows, D. (2004).
Geometry and Meaning.
Center for the Study of Language and Information, Stanford University.

Bibliography XIII

Zhang, P., Song, D., Zhao, X., and Hou, Y. (2011).
Investigating Query-Drift Problem from a Novel Perspective of Photon Polarization.
In Amati, G. and Crestani, F., editors, *Advances in Information Retrieval Theory - Proc. of ICTIR 2011*, pages 332–336. Springer.

Zhao, X., Zhang, P., Song, D., and Hou, Y. (2011).
A Novel Re-ranking Approach Inspired by Quantum Measurement.
In Clough, P., Foley, C., Gurrin, C., Jones, G. J. F., Kraaij, W., Lee, H., and Mudoch, V., editors, *Proceedings ECIR 2011*, volume 6611 of *Lecture Notes in Computer Science*, pages 721–724. Springer.

Bibliography XIV

Zuccon, G. (2012).
Document Ranking with Quantum Probabilities.
Phd thesis, University of Glasgow.

Zuccon, G., Azzopardi, L., and van Rijsbergen, K. (2009).
The Quantum Probability Ranking Principle for Information Retrieval.
In *Advances in Information Retrieval Theory - Proc. of ICTIR 2009*, volume 5766 of *Lecture Notes in Computer Science*, pages 232–240, Berlin, Heidelberg. Springer Berlin Heidelberg.

Bibliography XV

Zuccon, G., Piwowarski, B., and Azzopardi, L. (2011).
On the use of Complex Numbers in Quantum Models for Information Retrieval.
In Amati, G. and Crestani, F., editors, *Advances in Information Retrieval Theory - Proc. of ICTIR 2011*, pages 346–350. Springer.