Quantum Theory and Global Optimisation

Koenraad M.R. Audenaert

University of Wales, Bangor

February 19, 2004

This talk is not about Quantum Algorithms!

Theme I:

Additivity Problems in QIT

Algebraic Quantum Information Theory

- Quantum Information Theory (QIT) is a two-legged science:
  - Information Theory
  - Quantum Mechanics
- "My" kind of QIT = Algebraic QIT = QIT with a third leg: tools
  - Matrix Theory: matrix inequalities, eigenvalues, singular values,...
  - Convexity Theory: convex optimisation, duality theory, convex hulls,...
- Focus on problems where these tools are essential.

Open Problems

- QIT is about bi- or even multipartite states
- These are essentially higher-order tensors
- Life would be easy if a higher-order SVD existed
- It does not, so we get a lot of difficult problems in QIT
  - Separability problem
  - Proving additivity of various basic quantities
- There is hope we can tackle these problems using “ordinary” linear algebra

Entanglement Measures

- How to quantify entanglement of a state?
- Pure states: only one reasonable measure
  - $E(|\psi\rangle\langle\psi|) = S(\mathrm{Tr}_A |\psi\rangle\langle\psi|)$, S von Neumann entropy
- Mixed states: whole zoo of measures, use what you need
  - Entanglement Cost E_C
  - Entanglement of Distillation E_D
  - Entanglement of Formation (EoF, E_F)
  - Relative Entropy of Entanglement E_R
  - Squashed Entanglement E_{sq}

Entanglement Cost

- Calculating the entanglement cost E_C is one of the Big Open Problems of QIT.
- E_C defined in an operational way, nearly impossible to calculate
- Hayden, Horodecki and Terhal: E_C is equal to the *regularisation* of E_F:

$$E_C(\rho) = \lim_{n \to \infty} E_F(\rho^{\otimes n})/n.$$

- E_C would be equal to E_F if E_F were **additive**.

$$E_F(\rho_1 \otimes \rho_2) =? \; E_F(\rho_1) + E_F(\rho_2)$$

- This is Additivity Problem #1.
- One only needs to prove **superadditivity**, $E_F(\rho_1 \otimes \rho_2) \geq ? \; E_F(\rho_1) + E_F(\rho_2)$, because **subadditivity**, $E_F(\rho_1 \otimes \rho_2) \leq E_F(\rho_1) + E_F(\rho_2)$, is trivial to prove.

Strong Superadditivity of EoF

- Vollbrecht and Werner conjectured a stronger property implying superadditivity:

  Strong Superadditivity: $E_F(\rho) \geq ? \, E_F(\rho_I) + E_F(\rho_{II})$

  Here ρ is a general state over a duplicated Hilbert space and ρ_I and ρ_{II} are its reductions to the different copies of that space.

- This is Additivity Problem #2

- And now for something completely different...

Classical Capacity of Quantum Channels

- Noisy communication channels modelled as completely positive trace-preserving (CPTP) maps between operator algebras.
- One of the most fundamental questions in QIT: determination of *classical capacity of a quantum channel* i.e. the capacity of quantum channels to transmit classical information.
- Much more difficult than its purely classical counterpart due to the existence of entanglement.
- To obtain an optimal quantum channel *decoder* one has to perform entangled measurements over the channel output states.

Additivity of Holevo Capacity

- Is entanglement also necessary to obtain an optimal *encoder*?
- Widely believed not to be the case, i.e. no benefit is expected in having entanglement between the (single-letter) states sent over the channel.
- To prove this, it is necessary to show that the single-letter classical capacity (a.k.a. the Holevo capacity χ) of a quantum channel is additive.
- The Holevo capacity of a channel Φ is:

$$\chi(\Phi) = \sup_{\pi,\rho} S(\sum \pi_i \Phi(\rho_i)) - \sum \pi_i S(\Phi(\rho_i)).$$

- Is this additive: $\chi(\Phi_1 \otimes \Phi_2) = \chi(\Phi_1) + \chi(\Phi_2)$?
- This is Additivity Problem #3.

Maximal Output Purity of Channels

- Holevo capacity looks quite complicated. It has resisted any attempt so far at proving its additivity.
- For that reason, a simpler channel property, called the minimal output entropy (MOE), has been introduced.
- In general, if an operation acts on a pure state, the output will be a mixed state.
- The *maximal output purity* (MOP), ν_q, of an operation or channel quantifies how close to purity one can get by choosing the input state. As a measure of purity, one can use the Schatten q-norm $||X||_q = (\mathrm{Tr}\, X^q)^{1/q}$:

$$\nu_q(\Omega) = \max_\psi \{||\Omega(|\psi\rangle\langle\psi|)||_q : ||\psi|| = 1\}.$$

- The *minimal output entropy* (MOE), ν_S, is the entropic version of the MOP, where the von Neumann entropy S is used as a measure of purity.

$$\nu_S(\Omega) = \min_\psi \{S(\Omega(|\psi\rangle\langle\psi|)) : ||\psi|| = 1\}.$$

Additivity Problem #4

- Is MOE additive?
- Could shed some light on the additivity problem for the Holevo capacity.
- Additivity of MOE would follow from multiplicativity of MOP for "small" q (Amosov, Holevo and Werner): $\nu_q(\Phi \otimes \Omega) = \nu_q(\Phi)\nu_q(\Omega)$.
- Multiplicativity of MOP proven by Chris King for
  - entanglement breaking channels
  - unital qubit maps
  - depolarising channels
- Refuted for values of $q > 4.79$ (Holevo and Werner) (Baka!)
- Nevertheless, it could still hold for $q \downarrow 1$ (Hope springs eternal...)

Theme II:

Optimisation Theory

Local and Global Optimisation

- There are lots of techniques for finding the maxima of a general function
- Most methods can get stuck in local maxima
- Global optimisation is about finding the biggest local maximum
- No method exists that guaranteedly finds that
- So we are very happy when we have a problem for which there is only one local optimum
- Convex problems are in that league.

Convex Sets and Functions

- A set S is **convex** if and only if

$$\forall p, q \in S : \overline{pq} \in S$$

where $\overline{pq}$ is the line segment joining p and q

- Whatever one can say about sets, one can say about functions
- The **Graph** of a function $y = f(x)$ is the set of points $\{(x, y) : y = f(x)\}$
- The **Epigraph** of a function $y = f(x)$ is the set of points

$$\mathrm{Epi}(f) := \{(x, y) : y \geq f(x)\}$$

- A function is convex if and only if its epigraph is a convex set
- A function f is **concave** if $-f$ is convex

Duality Theory

- A dual view of convexity: a convex set S equals the intersection of all halfplanes containing S
- A function f is convex if and only if

$$\forall\{(a_i, x_i)\} : f(\sum_i a_i x_i) \leq \sum a_i f(x_i)$$

- Dually, f is convex if and only if
  f equals the pointwise supremum of all affine functions majorised by f

$$\forall x : f(x) = \sup_{a,b}\{a^T x + b : (\forall y : a^T y + b \leq f(y))\}$$

- For every a, one can calculate $\hat{b}(a)$, the largest b satisfying the condition.

Duality in Optimisation

- Let's tackle the following constrained optimisation:
- Minimise a convex function $f(x)$ over the interval $x \leq x_u$.

$$\hat{f} = \min_x \{f(x) : x \leq x_u\}.$$

- We can easily prove

$$\min_x \{f(x) : x \leq x_u\} \geq \max_a \{a x_u + \hat{b}(a) : a \leq 0\}.$$

(In fact, equality holds!)

- The maximisation over a is a convex problem, called the dual problem.
- The minimisation over x is called the primal problem.

Certificates of Convergence

- By picking specific $x \leq x_u$ (so-called feasible x), you get upper bounds on $\hat{f}$.
- By picking specific tangents $a \leq 0$, you get lower bounds on $\hat{f}$.
- Thus if you solve the primal problem together with the dual one, you can bracket the solution within an upper and a lower bound.
- The difference between the bounds is an upper bound on how far you are from the real solution.
- You, therefore, get a certificate of convergence.
- This works for general convex optimisation problems.
- No other optimisation problem has this feature.

Duality and Convex Hulls

- The **Convex Hull** $\mathrm{Conv}(S)$ of a set S is the *union* of all line segments $\overline{pq}$ where $p, q \in S$
- Dually, the convex hull of S is the *intersection* of all halfplanes containing S.
- The convex closure $\hat{f}$ of a function f is defined by

$$\mathrm{Epi}(\hat{f}) = \mathrm{Conv}(\mathrm{Epi}(f))$$

- Functionally, the convex closure of f can be calculated by taking convex combinations

$$\hat{f}(x) = \min_{\{(a_i,x_i)\}} \{\sum_i a_i f(x_i) : \sum_i a_i x_i = x\}$$

- And, dually, the convex closure of f is
  the pointwise supremum of all affine functions majorised by f

$$\hat{f}(x) = \sup_{a,b} \{a^T x + b : (\forall y : a^T y + b \leq f(y))\}$$

Allegro:

Convex Optimisation and Additivity

Entanglement of Formation Defined

- Any state ρ can be realised by an **ensemble** of pure states
- An ensemble is specified by a set of pairs $\{(p_i, \psi_i)\}_{i=1}^N$
  - of N state vectors ψ_i and statistical weights p_i
  - with $p_i \geq 0$ and $\sum_i p_i = 1$
- The entanglement of formation (EoF) of a bipartite state ρ (over the bi-partite Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$), is

$$E_F(\rho) = \min_{\{(p_i,\psi_i)\}} \{ \sum_i p_i S(\mathrm{Tr}_A \Psi_i) : \sum_i p_i \Psi_i = \rho \}.$$

EoF is a Convex Closure

- **Observation 1**: The definition of the EoF really means that the EoF is the **convex closure** of the pure state entanglement function

$$E(\Psi) = S(\mathrm{Tr}_A \Psi)$$

- Consider bounded functions f whose domain is the set of states $\mathcal{S}(\mathcal{H})$
- We can apply real convex analysis because $\mathcal{B}(\mathcal{H})$ with $\langle A, B \rangle = \mathrm{Tr}[AB]$ is a real vector space
- **Definition 1**: The convex closure of f is

$$\hat{f}(\rho) = \min_{\{(p_i,\rho_i)\}} \{\sum_i p_i f(\rho_i) : \sum_i p_i \rho_i = \rho\}$$

EoF is a Convex Closure

- **Observation 2**: The second, dual, formulation of the convex closure looks easier
- **Definition 2**: The convex closure of a function f is
  the pointwise supremum of all affine functions on $\mathcal{S}(\mathcal{H})$ majorised by f

$$\hat{f}(\rho) = \sup_{X \in \mathcal{B}(\mathcal{H})} \{\mathrm{Tr}[\rho X] : (\forall \sigma \in \mathcal{S}(\mathcal{H}) : \mathrm{Tr}[\sigma X] \leq f(\sigma))\}$$

- Moreover, this definition can be pulled apart into two identical parts, based on the concept of the **conjugate function**

Conjugate function

- Define the **conjugate function** f^*:

$$f^*(X) = \max_{\rho \in \mathcal{S}(\mathcal{H})} \mathrm{Tr}[\rho X] - f(\rho)$$

- If f is continuous this is called the **Legendre transform** of f.
- The conjugate is convex in X: pointwise maximum of affine functions

- The conjugate and convex closure determine each other completely

$$f \quad \xrightarrow{*} \quad f^* \quad \xleftrightarrow{*} \quad \hat{f}$$

  - The convex closure of f is the conjugate of the conjugate of f: $\hat{f} = f^{**}$
  - The conjugate of the convex closure of f is the conjugate of f: $\hat{f}^* = f^*$

Additivity of EoF

Theorem 1 (Audenaert and Braunstein) *With g defined by*

$$g(A) = \max_{\psi \in \mathcal{H}} \mathrm{Tr}[\Psi \log(A)] - E(\Psi)$$

strong superadditivity of the EoF,

$$E_F(\rho) \geq ?\, E_F(\rho_I) + E_F(\rho_{II}),$$

is equivalent to subadditivity of g,

$$g(A_1 \otimes A_2) \leq ?\, g(A_1) + g(A_2).$$

Relation to MOP

- The above Theorem reduces the additivity problem for the EoF, defined as a minimisation over *ensembles*, to an equivalent problem for the conjugate function, defined as a maximisation over *pure states*.
- Using the Lie-Trotter relation, entropic quantities can be converted to power-law quantities.
- Using some involved mathematics, Audenaert and Braunstein also proved:

  Theorem 2 *If ν_q is multiplicative for $q \downarrow 1$ and for all competely positive maps, then the entanglement of formation is strongly superadditive.*

- So two additivity problems are closely related!

Equivalence of Additivity Problems

- One month later...
- Peter Shor proved equivalence of all four additivity problems!
  - additivity of EoF,
  - strong superadditivity of EoF,
  - additivity of classical capacity of a quantum channel,
  - additivity of the MOE: $\nu_S(\Phi \otimes \Omega) = \nu_S(\Phi) + \nu_S(\Omega)$.
- The stakes for proving multiplicativity of MOP have raised.
- So what's next?

Theme III:

The Quantum de Finetti Theorem

Unknown Probabilities

- Probability Estimation is a procedure to estimate an unknown probability from the results of repeated trials in identical circumstances.
- In Bayesian view of probability theory, a probability is a "measure of credible belief, reflecting one's state of knowledge".
- In this view, "unknown probability" is an oxymoron.
- Bruno de Finetti (early 1930's) tried to eliminate this offending concept.
- He focused on the equivalence of repeated trials:
  - indistinguishability of the different trials w.r.t. predictions
  - a probability assignment for multiple trials should be symmetric under permutation of the trials

Exchangeable Probabilities

- Principle of Equivalence of Repeated Trials:
  - If an experimenter judges a collection of N trials to have a probability $P^{(N)}$, he will judge any permutation of the trials to have that same probability.
  - This will be true for every N
  - Consistency condition: $P^{(N)}$ must be derivable from $P^{(N+1)}$
- Sequence of probabilities $(P^{(N)})_N$ obeying this condition are called exchangeable.

de Finetti's Theorem

- A sequence of probabilities $(P^{(N)})_N$ is exchangeable if and only if

$$P^{(N)} = \int d\mu(P) P^{\times N},$$

where $d\mu(P)$ is a positive measure over (single-trial) probabilities.

- "Probability distribution over probabilities" replaces "unknown probability"
- Experimenter can act as if...
  - there is an objective (single-trial) probability assignment, P,
  - yielding $P^{\times N}$ for N repeated trials,
  - his uncertainty about P is expressed by $\int d\mu(P)$.

Exchangeable Quantum States

- In Quantum Theory, the analogon of a probability is the *quantum state*
- In quantum tomography, one tries to measure an unknown quantum state using repeated trials on identically prepared particles.
- In the information-based interpretation of quantum mechanics, a state represents the state of knowledge of an observer.
- Again, in that interpretation, “unknown quantum state” is an oxymoron.
- Sequence of exchangeable quantum states $(\rho^{(N)})_N$:
  - State $\rho^{(N)}$ defined over N-fold copy of Hilbert space $\mathcal{H}^{\otimes N}$;
  - Every $\rho^{(N)}$ is symmetric under permutation of copies of $\mathcal{H}$
  - Consistency: $\rho^{(N)} = \mathrm{Tr}_1\, \rho^{(N+1)}$

The Quantum de Finetti Theorem

- Theorem (Hudson and Moody 1976; Størmer 1969):

  A sequence of states $(\rho^{(N)})_N$ is exchangeable if and only if

  $$\rho^{(N)} = \int d\mu(\rho)\rho^{\otimes N},$$

  where $d\mu(\rho)$ is a positive measure over the state space of $\mathcal{H}$.

- Concept of “unknown state” replaced by “probability distribution over states”
- Tomographer can act as if...
  - there is an “observer-in-the-box” preparing systems in the same state ρ,
  - yielding $\rho^{\otimes N}$ for N repeated trials,
  - his uncertainty about ρ is expressed by $\int d\mu(\rho)$.

Allegretto:

There, and Back Again

The Plan

- Optimisation theory can help us in solving questions in quantum theory.
- It has done so before.
- It can't help us very much with the MOP:
- Not a convex problem: it is a *maximisation* of a convex function over a convex set
- All we get from convexity theory is that the maximum will be obtained in an extreme point
- Now here is where Quantum de Finetti comes in!
- So, in return, quantum theory can offer solutions to questions in optimisation theory!

Maximal Output Purity (again!)

- Consider the maximal output purity ν_q for integer q; then

$$\nu_q^q(\Phi) = \max_{\rho \in \mathcal{S}(\mathcal{H})} \mathrm{Tr}[(\Phi(\rho))^q],$$

- Note that $\mathrm{Tr}[(\Phi(\rho))^q] = \mathrm{Tr}[\Phi(\rho)\Phi(\rho)\ldots\Phi(\rho)]$, with q factors.
- We can write $\mathrm{Tr}[(\Phi(\rho))^q] = \mathrm{Tr}[A\rho^{\otimes q}]$, with

$$A_{(i),(j)} = \mathrm{Tr}[\boldsymbol{\Phi}_{i_1,j_1}\ldots\boldsymbol{\Phi}_{i_q,j_q}],$$

where $\boldsymbol{\Phi}$ is the Choi matrix of the map Φ.

Symmetry

- Consider permutations of n copies of $\mathcal{H}$, $\pi \in S_n$.
- If, for every permutation $\pi \in S_n$, a matrix A over $\mathcal{H}^{\otimes n}$ obeys

$$A_{(i_1,\ldots,i_q),(j_1,\ldots,j_q)} = A_{(i_{\pi(1)},\ldots,i_{\pi(q)}),(j_{\pi(1)},\ldots,j_{\pi(q)})},$$

  then the matrix A is *symmetric*.
- Let P_π permute the indices according to π, i.e. $(P_\pi x)_{(i)} = x_{\pi(i)}$.
- Thus A is symmetric if and only if $\forall \pi \in S_n$, $P_\pi^\dagger A P_\pi = A$.
- The linear map P_n that projects all operators to the symmetric subspace is

$$\mathsf{P}_n(A) = \frac{1}{n!} \sum_{\pi \in S_n} P_\pi^\dagger A P_\pi.$$

  We call $\mathsf{P}_n(A)$ the *symmetric part* of A.

Yet Another Main Theorem

The symmetric part of A is:

$$\mathsf{P}_n(A) = \frac{1}{n!} \sum_{\pi \in S_n} P_\pi^\dagger A P_\pi.$$

Theorem 3 *For any* $q, n \in \mathbb{N}$*, and for any operator* A *over* $\mathcal{H}^{\otimes q}$ *with Hermitian symmetric part, the sequence* $(\mu_n(A))_n$*, with*

$$\mu_n(A) := \lambda_{\max}(\mathsf{P}_{q+n}(A \otimes \mathbf{I}^{\otimes n})),$$

is non-increasing and converges to

$$\lim_{n\to\infty} \mu_n(A) = \max_\rho \mathrm{Tr}[A\rho^{\otimes q}].$$

Proof of Theorem (I)

- We want to maximise $\mathrm{Tr}[A\rho^{\otimes q}]$ over all states ρ.
- First turn this into the more general optimisation problem

$$\max_{d\mu(\rho)\geq 0} \{\mathrm{Tr}[A\int \rho^{\otimes q} d\mu(\rho)] : \int d\mu(\rho) = 1\}$$

- We can replace this by a maximisation of $\mathrm{Tr}[A\rho^{(q)}]$ over all $\rho^{(q)}$ that are q-th element in some exchangeable sequence of states:
  - Any state of the form $\rho^{\otimes q}$ obviously forms part of an exchangeable sequence.
  - Conversely, by the QdF theorem, the q-th element in any exchangeable sequence must be of the form $\int \rho^{\otimes q} d\mu(\rho)$ for some positive measure $d\mu(\rho)$.

Proof of Theorem (II)

- From the definition of exchangeable sequence, we infer that $\rho^{(q)}$ must be the partial trace of a symmetric state $\rho^{(\infty)}$ over $\mathcal{H}^{\otimes\infty}$, where all but q copies of $\mathcal{H}$ have been traced out.
- As a result, we can express the states $\int \rho^{\otimes q} d\mu(\rho)$ in

$$\max_{\rho} \mathrm{Tr}[A\rho^{\otimes q}] = \max_{d\mu(\rho)\geq 0} \{\mathrm{Tr}[A \int \rho^{\otimes q} d\mu(\rho)] : \int d\mu(\rho) = 1\}$$

as a partial trace of symmetric states over $\mathcal{H}^{\otimes\infty}$:

$$\begin{aligned} \max_{\rho} \mathrm{Tr}[A\rho^{\otimes q}] &= \lim_{n\to\infty} \max_{\rho} \{\mathrm{Tr}[A \ \mathrm{Tr}_n[\rho]] : \rho \text{ symmetric over } \mathcal{H}^{\otimes(q+n)}\} \\ &= \lim_{n\to\infty} \max_{\rho} \{\mathrm{Tr}[(A \otimes \mathbf{I}^{\otimes n}) \, \rho] : \rho \text{ symmetric over } \mathcal{H}^{\otimes(q+n)}\}. \end{aligned}$$

Proof of Theorem (III)

- So we have

$$\max_\rho \mathrm{Tr}[A\rho^{\otimes q}] \;=\; \lim_{n\to\infty} \max_\rho \{\mathrm{Tr}[(A \otimes \mathbf{I}^{\otimes n})\,\rho] : \rho \text{ symmetric}\}.$$

- This maximisation over *symmetric* states can be replaced by a maximisation over *all* states, provided they are projected first onto the symmetric subspace:

$$\max_\rho \mathrm{Tr}[A\rho^{\otimes q}] = \lim_{n\to\infty} \max_\rho \mathrm{Tr}[(A \otimes \mathbf{I}^{\otimes n})\,\mathsf{P}_{q+n}(\rho)\,].$$

- The projection can equally well be applied to the factor $A \otimes \mathbf{I}^{\otimes n}$:

$$\begin{aligned}\max_\rho \mathrm{Tr}[A\rho^{\otimes q}] \;&=\; \lim_{n\to\infty} \max_\rho \mathrm{Tr}[\mathsf{P}_{q+n}(A \otimes \mathbf{I}^{\otimes n})\,\rho\,] \\ &=\; \lim_{n\to\infty} \lambda_{\max}(\mathsf{P}_{q+n}(A \otimes \mathbf{I}^{\otimes n})).\end{aligned}$$

Proof of Theorem (IV)

- We only have to show that $\mu_n(A)$ is non-increasing with n.
- To do so, we use the convexity of $\lambda_{\max}$:

$$\begin{aligned}
\mu_{n+1}(A) &= \lambda_{\max}(\mathsf{P}_{q+n+1}(A \otimes \mathbf{I}^{\otimes(n+1)})) \\
&= \lambda_{\max}(\mathsf{P}_{q+n+1}(\mathsf{P}_{q+n}(A \otimes \mathbf{I}^{\otimes n}) \otimes \mathbf{I}) \\
&\leq \frac{1}{(q+n+1)!} \sum_{\pi \in S_{q+n+1}} \lambda_{\max}(P_\pi^\dagger(\mathsf{P}_{q+n}(A \otimes \mathbf{I}^{\otimes n}) \otimes \mathbf{I})P_\pi) \\
&= \lambda_{\max}(\mathsf{P}_{q+n}(A \otimes \mathbf{I}^{\otimes n}) \otimes \mathbf{I}) \\
&= \lambda_{\max}(\mathsf{P}_{q+n}(A \otimes \mathbf{I}^{\otimes n})) \\
&= \mu_n(A).
\end{aligned}$$

- This finishes the Proof.

So what is this?

- Algorithmic issues: with some more symmetry theory, we get a modestly efficient algorithm that calculates guaranteed upper bounds and has no problems at all with local maxima!
- See quant-ph/0402076 for more details than you'd care for
- There might be more in store here!
- Theoretical Issues: reduces a difficult optimisation problem to an eigenvalue problem
- This might just be the simplification needed for tackling additivity
- (Watch out for forthcoming stuff on ArXiv)