A tensor product matrix approximation problem in quantum physics

Geir Dahl

Center of Mathematics for Applications, Department of Informatics,
University of Oslo P.O. Box 1053 Blindern, NO-0316 Oslo, NORWAY.
geird@math.uio.no

Jon Magne Leinaas

Department of Physics, University of Oslo
P.O. Box 1048 Blindern, NO-0316 Oslo, NORWAY.
j.m.leinaas@fys.uio.no

Jan Myrheim

Department of Physics, Norwegian University of Science and Technology
NO-7491 Trondheim, NORWAY.
jan.myrheim@ntnu.no

Eirik Ovrum

Center of Mathematics for Applications, Department of Physics,
University of Oslo P.O. Box 1048 Blindern, NO-0316 Oslo, NORWAY.
ovrum@fys.uio.no

29 August 2006

Abstract

We consider a matrix approximation problem arising in the study of *entanglement* in quantum physics. This notion represents a certain type of correlations between subsystems in a composite quantum system. The states of a system are described by a density matrix, which is a positive semidefinite matrix with trace one. The goal is to approximate such a given density matrix by a so-called separable density matrix, and the distance between these matrices gives information about the degree of entanglement in the system. Separability here is expressed in terms of tensor products. We discuss this approximation problem for a composite system with two subsystems and show that it can be written as a convex optimization problem with special structure. We investigate related convex sets, and suggest an algorithm for this approximation problem which exploits the tensor product structure in certain subproblems. Finally some computational results and experiences are presented.

AMS classification: 15A90, 15A69 , 52A41, 90C25.

Keywords: Density matrix, entanglement, tensor product, matrix approximation, positive semidefinite, convex sets.

1 Introduction

A current problem in physics is to give a precise characterization of *entanglement* in a quantum system. This describes types of correlations between subsystems of the full quantum system that go beyond the statistical correlations that can be found in a classical composite system. The interest is motivated by ideas about utilizing entanglement to perform communication and computation in qualitatively new ways.

Although a general quantitative definition of the degree of entanglement of a composite system does not exist, there is a generally accepted definition that distinguishes between quantum states with and without entanglement. The non-entangled states are referred to as *separable states*, and they are considered to only contain classical correlations between the subsystems. In addition, for some special cases a generally accepted quantitative measure of entanglement exists.

The standard mathematical formulation of a composite quantum system is in terms of *density matrices*. These describe the quantum states of either the full system or one of its subsystems, and correspond to hermitian, positive semidefinite operators that act on a complex vector space, either finite or infinite dimensional. The density matrices also satisfy a normalization condition so that they form a compact convex set in the vector space of hermitian matrices. For a composite system the separable states form a subset of the density matrices that is also compact and convex.

The physical interpretation of a density matrix is that it contains information about the statistical distribution over measurable values for any observable of the quantum system. Such an observable is represented by a hermitian matrix A, with the eigenvalues corresponding to the possible outcomes of a measurement. In particular, for a given quantum state represented by a density matrix ρ, the expectation value of the observable A is defined by the trace of the product, $\mathrm{tr}(A\rho)$. A density matrix that is a projection on a single vector is referred to as representing a *pure state*, while other density matrices are representing *mixed states*. A mixed state can be interpreted as corresponding to a statistical distribution over pure states, and in this sense includes both quantum uncertainty (through the pure states) and classical uncertainty (through the statistical distribution).

To identify whether a state, i.e., a density matrix, is entangled or not is for a general quantum state considered to be a hard problem [6]. However, some general results exist concerning the separability problem, in particular a simple sufficient condition for entanglement has been found [13], and also general schemes to test separability numerically have been suggested [4,10]. Other results refer to more special cases, for matrices of low dimensions or for restricted subsets of density matrices [9]. There have also been attempts to more general approaches to the identification of separability by use of the natural geometrical structure of the problem, where the Euclidean metric (Hilbert-Schmidt or Frobenius norm) of the hermitian matrices can be used to give a geometrical characterization of the set of separable matrices, see [16,14] and references therein.

In the present paper we focus on the geometrical description of separability and consider the problem of finding the closest separable state (in the Frobenius norm) to any given state, either entangled or non-entangled. We consider the case of a quantum system with two subsystems, where the full vector space is the tensor product of the vector spaces of the subsystems, and where the vector spaces are finite-dimensional. The distance to the closest separable state can be used as a measure of the degree of entanglement of the chosen state, although there are certain conditions a *good* measure of entanglement should satisfy, that are not obviously satisfied by the Euclidean norm, see [12]. However, beyond this question it seems that an efficient method to identify the nearest separable state should be useful in order to identify the boundary between the separable and entangled states. Obviously, if a density matrix is separable the distance to the closest separable state is zero, hence our method is a numerical test for separability.

In Section 2 we present the mathematical framework for our investigations. In Sections 4, 5 and 6 we consider a systematic, iterative method which approximates the closest separable state (the DA algorithm). This is a two step algorithm, where the first step is to optimize the convex combination of a set of tensor product matrices. The second step is to find a best new tensor product matrix to add to the existing set. These two steps are implemented iteratively so that the expansion coefficients are optimized each time a new product matrix is included and a new optimal product matrix is found after optimization of the previous set. We give some numerical examples where the algorithm is used to find the closest separable state and to efficiently identify the boundary between the separable and entangled states.

In order to be specific, and to simplify the formulation of the theory, we shall consider in this paper mostly real matrices, with the density matrices restricted to be symmetric. We want to emphasize, however, that our algorithms can be immediately applied to complex matrices, which is the relevant case for quantum theory. The only modification needed then is that transposition of real matrices is generalized to hermitian conjugation of complex matrices, and the class of real symmetric matrices is generalized to complex hermitian matrices. In Section 7 we present an example of application to quantum theory, with complex density matrices.

We may remark here that there is a qualitative difference between the separability problems for real and complex matrices. In fact, the set of separable density matrices has the same dimension as the full set of density matrices in the complex case, but has lower dimension in the real case. The last result follows from the Peres separability criterion [13], since a separable matrix must be symmetric under partial transposition if only real matrices are involved. Thus, among all real density matrices that are separable as members of the set of complex density matrices, only a subset of lower dimension are separable in terms of real matrices only.

2 Formulation of the problem

We introduce first the mathematical framework needed to formulate the problem to be investigated. This includes a summary of the properties of density matrices and separable density matrices as expressed through the definitions and theorems given below. We end the section by formulating the problem to be investigated, referring to this as the density approximation problem.

We discuss here in detail only the case of real vectors and real matrices, because this simplifies our discussion, and because the generalization to the complex case is quite straightforward, based on the close correspondence between transposition of real matrices, on the one hand, and hermitian conjugation of complex matrices, on the other hand. In particular, the set of real symmetric matrices is expanded to the set of complex hermitian matrices, which is still a real vector space with the same positive definite real inner product $\mathrm{tr}(AB)$ (see below).

To explain the notation used, we consider the usual Euclidean space $\mathbb{R}^n$ of real vectors of length n equipped with the standard inner product and the Euclidean norm denoted by $\|\cdot\|$. Vectors are treated as column vectors, and the transpose of a vector x is denoted by x^T. The convex hull of a set S is the intersection of all convex sets containing S, and it is denoted by $\mathrm{conv}(S)$. Some recommended references on convexity are [17] and [1]. We let I denote the identity matrix (of order n, where n should be clear from the context). The unit ball in $\mathbb{R}^n$ is $B_n = \{x \in \mathbb{R}^n : \|x\| = 1\}$.

Let $\mathcal{S}^n$ denote the linear space consisting of all the real symmetric $n \times n$ matrices. This space has dimension $n(n+1)/2$. In $\mathcal{S}^n$ we use the standard inner product

$$\langle A, B\rangle = \mathrm{tr}(AB) = \sum_{i,j} a_{ij}b_{ij} \quad (A, B \in \mathcal{S}^n).$$

Here $\mathrm{tr}(C) = \sum_{i=1}^n c_{ii}$ denotes the trace of a matrix C. The associated matrix norm is the Frobenius norm $\|A\|_F = (\sum_{i,j} a_{ij}^2)^{1/2}$ and we use this norm for measuring distance in the matrix approximation problem of interest. A matrix $A \in \mathcal{S}^n$ is positive semidefinite provided that $x^T Ax \geq 0$ for all $x \in \mathrm{I\!R}^n$, and this will be denoted by $A \succeq 0$. We define the *positive semidefinite cone*

$$\mathcal{S}_+^n = \{A \in \mathcal{S}^n : A \succeq 0\}$$

as the set of all symmetric positive semidefinite matrices of order n. $\mathcal{S}_+^n$ is a full-dimensional closed convex cone in $\mathcal{S}^n$, so $\lambda_1 A_1 + \lambda_2 A_2 \in \mathcal{S}_+^n$ whenever $A_1, A_2 \in \mathcal{S}_+^n$ and $\lambda_1, \lambda_2 \geq 0$. For more about positive semidefinite matrices we refer to the excellent book in matrix theory [7]. A *density matrix* is a matrix in $\mathcal{S}_+^n$ with trace 1. We let $\mathcal{T}_+^n$ denote the set of all density matrices of order n, so

$$\mathcal{T}_+^n = \{A \in \mathcal{S}_+^n : tr(A) = 1\}.$$

The set $\mathcal{T}_+^n$ is convex and we can determine its extreme points. Recall that a point x in a convex set C is called an extreme point when there is no pair of distinct points $x_1, x_2 \in C$ with $x = (1/2)x_1 + (1/2)x_2$.

Theorem 2.1 *The set $\mathcal{T}_+^n$ of density matrices satisfies the following:*

(i) $\mathcal{T}_+^n$ is the intersection of the positive semidefinite cone $\mathcal{S}_+^n$ and the hyperplane $H = \{A \in \mathcal{S}^n : \langle A, I\rangle = 1\}$.

(ii) $\mathcal{T}_+^n$ is a compact convex set of dimension $n(n+1)/2 - 1$.

(iii) The extreme points of $\mathcal{T}_+^n$ are the symmetric rank one matrices $A = xx^T$ where $x \in \mathrm{I\!R}^n$ satisfies $\|x\| = 1$. Therefore

$$\mathcal{T}_+^n = conv(\{xx^T : x \in \mathrm{I\!R}^n, \|x\| = 1\}.$$

Proof. Property (i) follows from the definition of $\mathcal{T}_+^n$ as $\langle A, I\rangle = \sum_i a_{ii} = \mathrm{tr}(A)$. So H is the solution set of a single linear equation in the space $\mathcal{S}^n$ and therefore H is a hyperplane. Thus, $\mathcal{T}_+^n$ is the intersection of two closed convex sets, and this implies that $\mathcal{T}_+^n$ is also closed and convex. Moreover, $\mathcal{T}_+^n$ is bounded as one can prove that each $A \in \mathcal{T}_+^n$ satisfies $-1 \leq a_{ij} \leq 1$ for $1 \leq i, j \leq n$ (This follows from the facts that $\mathrm{tr}(A) = 1$, $e_i^T Ae_i \geq 0$

and $(e_i + e_j)^T A(e_i + e_j) \geq 0$ where e_i denotes the i'th unit vector in $\mathbb{R}^n$.) Since $\mathcal{T}_+^n$ lies in the hyperplane H, $\mathcal{T}_+^n$ has dimension at most $\dim(\mathcal{S}^n) - 1 = n(n+1)/2 - 1$. Consider the matrices $(1/2)(e_i + e_j)(e_i + e_j)^T$ $(1 \leq i < j \leq n)$ and $e_i e_i^T$ $(1 \leq i \leq n)$. One can check that these $n(n+1)/2$ matrices are affinely independent (in the space $\mathcal{S}^n$) and they all lie in $\mathcal{T}_+^n$. It follows that $\dim(\mathcal{T}_+^n) = n(n+1)/2 - 1$. This proves Property (ii).

It remains to determine the extreme points of $\mathcal{T}_+^n$. Let $A \in \mathcal{T}_+^n$. Since A is real and symmetric it has a spectral decomposition

$$A = VDV^T$$

where V is a real orthogonal $n \times n$ matrix and D is a diagonal matrix with the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of A on the diagonal. By partitioning V by its columns $v_1, v_2, \ldots, v_n$, where v_i is the eigenvector corresponding to λ_i, we get

$$A = \sum_{i=1}^{n} \lambda_i v_i v_i^T. \tag{1}$$

Since A is positive semidefinite, all the eigenvalues are nonnegative. Moreover, $\sum_i \lambda_i = \mathrm{tr}(A) = 1$. Therefore, the decomposition (1) actually represents A as a convex combination of rank one matrices $v_i v_i^T$ where $\|v_i\| = 1$ (as V is orthogonal). From this it follows that all extreme points of $\mathcal{T}^n$ are rank one matrices $xx^T$where $\|x\| = 1$. One can also verify that *all* matrices of this kind are indeed extreme points, but we omit the details here. Finally, from convexity theory (see [17]) the Krein-Milman theorem says that a compact convex set is the convex hull of its extreme points which gives the final property in the theorem. □

The theorem shows that the extreme points of the set of density matrices $\mathcal{T}_+^n$ are the symmetric rank one matrices $A_x = xx^T$ where $\|x\| = 1$. Such a matrix A_x is an orthogonal projector (so $A_x^2 = A_x$, and A_x is symmetric) and $A_x y$ is the orthogonal projection of a vector y onto the line spanned by x.

Remark. The spectral decomposition (1) is interesting in this context. First, it shows that any matrix $A \in \mathcal{T}_+^n$ may be written as a convex combination of at most n extreme points. This improves upon a direct application of Carathéodory's theorem (see [17]) which says that A may be represented using at most $\dim(\mathcal{T}^n)+1 = n(n+1)/2$ extreme points. Secondly, (1) shows how we can decompose A as a convex combination of its extreme points by calculation eigenvectors and corresponding eigenvalues of A. Finally, we remark that the argument above based on the spectral decomposition also shows the well-known fact that the extreme rays of the positive semidefinite cone correspond to the symmetric rank one matrices.

We now proceed and introduce a certain subset of $\mathcal{T}_+^n$ which will be of main interest below. Recall that if $A \in \mathbb{R}^{p\times p}$ and $B \in \mathbb{R}^{q\times q}$ then the tensor product $A \otimes B$ is the square matrix of order pq given by its (i,j)'th block $a_{ij}B$ $(1 \leq i,j \leq p)$. A general reference on tensor products is [8].

For the rest of this section we fix two positive numbers p and q and let $n = pq$. We call a matrix $A \in \mathbb{R}^{n\times n}$ *separable* if A can be written as a convex combination

$$A = \sum_{j=1}^{N} \lambda_j \, B_j \otimes C_j$$

for some positive integer N, matrices $B_j \in \mathcal{T}_+^p$, $C_j \in \mathcal{T}_+^q$ $(j \leq N)$ and nonnegative numbers λ_j $(j \leq N)$ with $\sum_{j=1}^N \lambda_j = 1$. Let $\mathcal{T}_+^{n,\otimes}$ denote the set of all separable matrices of order n. Note that $n = pq$, but p and q are suppressed in our notation. For sets U and W of matrices we let $U \otimes W$ denote the set of matrices that can be written as the tensor product of a matrix in U and a matrix in W. The following theorem summarizes important properties of $\mathcal{T}_+^{n,\otimes}$.

Theorem 2.2 *The set $\mathcal{T}_+^{n,\otimes}$ of separable matrices satisfies the following:*

(i) $\mathcal{T}_+^{n,\otimes} \subseteq \mathcal{T}_+^n$.

(ii) $\mathcal{T}_+^{n,\otimes}$ is a compact convex set and $\mathcal{T}_+^{n,\otimes} = conv(\mathcal{T}_+^p \otimes \mathcal{T}_+^q)$.

(iii) The extreme points of $\mathcal{T}_+^{n,\otimes}$ are the symmetric rank one matrices

$$A = (x \otimes y)(x \otimes y)^T$$

where $x \in \mathbb{R}^p$ and $y \in \mathbb{R}^q$ both have Euclidean length one. So

$$\mathcal{T}_+^{n,\otimes} = conv(\{(x \otimes y)(x \otimes y)^T : x \in B_p, y \in B_q\}.$$

Proof. (i) Let $B \in \mathcal{T}_+^p$ and $C \in \mathcal{T}_+^q$, and let $A = B \otimes C$. Then, $A^T = (B \otimes C)^T = B^T \otimes C^T = B \otimes C = A$, so A is symmetric. Moreover, A is positive semidefinite as both B and C are positive semidefinite (actually, the eigenvalues of A are the products of eigenvalues of B and eigenvalues of C, see e.g. [8]). Finally, $\mathrm{tr}(A) = \mathrm{tr}(B)\mathrm{tr}(C) = 1$, so $A \in \mathcal{T}_+^n$. Since a separable matrix is a convex combination of such matrices, and $\mathcal{T}_+^n$ is convex, it follows that every separable matrix lies in $\mathcal{T}_+^n$.

(ii) By definition the set $\mathcal{T}_+^{n,\otimes}$ is the set of all convex combinations of matrices in $\mathcal{T}_+^p \otimes \mathcal{T}_+^q$. From a basic result in convexity (see e.g. [17]) this means that

$\mathcal{T}_+^{n,\otimes}$ coincides with the convex hull of $\mathcal{T}_+^p \otimes \mathcal{T}_+^q$. Consider now the function $g : \mathcal{T}^p \times \mathcal{T}^q \to \mathcal{S}^n$ given by $g(B, C) = B \otimes C$ where $B \in \mathcal{S}^p$ and $C \in \mathcal{S}^q$. Then $g(\mathcal{T}_+^p \times \mathcal{T}_+^q) = \mathcal{T}_+^p \otimes \mathcal{T}_+^q$ and the function g is continuous. Therefore $\mathcal{T}_+^{n,\otimes}$ is compact as $\mathcal{T}_+^p \times \mathcal{T}_+^q$ is compact (and the convex hull of a compact set is again compact, [17]).

(iii) Let A be an extreme point of $\mathcal{T}_+^{n,\otimes}$. Then $A = B \otimes C$ for some $B \in \mathcal{T}_+^p$ and $C \in \mathcal{T}_+^q$ (for a convex combination of more than one such matrix is clearly not an extreme point). From Theorem 2.1 we have that $B = \sum_{j=1}^m \lambda_j x_j x_j^T$ and $C = \sum_{k=1}^r \mu_j y_j y_j^T$ for suitable vectors $x_j \in \mathbb{R}^p$ and $y_k \in \mathbb{R}^q$ with $\|x_j\| = \|y_k\| = 1$, and nonnegative numbers λ_j $(j \leq m)$ and μ_k $(k \leq r)$ with $\sum_j \lambda_j = \sum_k \mu_k = 1$. Using basic algebraic rule for tensor products we then calculate

$$\begin{aligned} A = B \otimes C &= (\textstyle\sum_{j=1}^m \lambda_j x_j x_j^T) \otimes \sum_{k=1}^r \mu_k y_k y_k^T \\ &= \textstyle\sum_{j,k} \lambda_j \mu_k (x_j x_j^T) \otimes (y_k y_k^T) \\ &= \textstyle\sum_{j,k} \lambda_j \mu_k (x_j \otimes y_k)(x_j^T \otimes y_k^T) \\ &= \textstyle\sum_{j,k} \lambda_j \mu_k (x_j \otimes y_k)(x_j \otimes y_k)^T. \end{aligned}$$

Since $\sum_{j,k} \lambda_j \mu_k = 1$ this shows that A can be written as a convex combination of matrices of the form $(x \otimes y)(x \otimes y)^T$ where $\|x\| = \|y\| = 1$. Since A is an extreme point we have $m = r = 1$ and we have shown the desired form of the extreme points of $\mathcal{T}_+^{n,\otimes}$. Finally, one can verify that all these matrices are really extreme points, but we omit these details. □

We now formulate the following *density approximation problem* (DA), which we shall examine further in subsequent sections of the paper,

> (DA) *Given a density matrix* $A \in \mathcal{T}_+^n$ *find a separable density matrix* $X \in \mathcal{T}_+^{n,\otimes}$ *which minimizes the distance* $\|X - A\|_F$.

Separability here refers to the tensor product decomposition $n = pq$, as discussed above.

3 An approach based on projection

In this section we study the DA problem and show that it may be viewed as a projection problem associated with the convex set $\mathcal{T}_+^{n,\otimes}$ introduced in Section 2. This leads to a projection algorithm for solving DA.

The DA problem is to find the best approximation (in Frobenius norm) to a given density matrix $A \in \mathcal{T}_+^n$ in the convex set $\mathcal{T}_+^{n,\otimes}$ consisting of all separable density matrices. This corresponds to the optimization problem

$$\inf\{\|X - A\|_F : X \in \mathcal{T}_+^{n,\otimes}\}. \tag{2}$$

In this problem the function to be minimized is $f : \mathcal{S}_+^n \to \mathrm{I\!R}$ given by $f(X) = \|X - A\|_F$. Now, f is strictly convex (on the underlying space $\mathcal{S}^n$) and therefore continuous. Moreover, as shown in Theorem 2.2, the set $\mathcal{T}_+^{n,\otimes}$ is compact and convex, so the infimum is attained. These properties imply the following fact, see also [1], [2], [17].

Theorem 3.1 *For given $A \in \mathcal{T}_+^n$ the approximation problem (2) has a unique optimal solution X^*.*

The unique solution X^* will be called the *projection* of A onto $\mathcal{T}_+^{n,\otimes}$, and we denote this by $X^* = \mathrm{Proj}_{\otimes}(A)$.

The next theorem gives a variational inequality characterization of the projection $X^* = \mathrm{Proj}_{\otimes}(A)$. Let $\mathrm{Ext}(\mathcal{T}_+^{n,\otimes})$ denote the set of all extreme points of $\mathcal{T}_+^{n,\otimes}$; these extreme points were found in Theorem 2.2. The theorem is essentially the *projection theorem* in convexity, see e.g. [1]. We give a proof following the lines of [1] (except that we consider a different inner product space). The ideas in the proof will be used in our algorithm for solving DA below.

Theorem 3.2 *Let $A \in \mathcal{T}_+^n$ and $X \in \mathcal{T}_+^{n,\otimes}$. Then the following three statements are equivalent:*

$$\begin{aligned}
&\textit{(i)}\quad X = \mathit{Proj}_{\otimes}(A).\\
&\textit{(ii)}\ \ \langle A - X, Y - X\rangle \le 0 \quad \textit{for all } Y \in \mathcal{T}_+^{n,\otimes}.\\
&\textit{(iii)}\ \langle A - X, Y - X\rangle \le 0 \quad \textit{for all } Y \in \mathit{Ext}(\mathcal{T}_+^{n,\otimes}).
\end{aligned} \tag{3}$$

Proof. Assume first that (ii) holds and consider $Y \in \mathcal{T}_+^{n,\otimes}$. Then we have

$$\begin{aligned}
\|A - Y\|_F^2 &= \|(A - X) - (Y - X)\|_F^2\\
&= \|A - X\|_F^2 + \|Y - X\|_F^2 - 2\langle A - X, Y - X\rangle\\
&\ge \|A - X\|_F^2 + \|Y - X\|_F^2\\
&\ge \|A - X\|_F^2.
\end{aligned}$$

So, $\|A - X\|_F \le \|A - Y\|_F$ for all $Y \in \mathcal{T}_+^{n,\otimes}$ and therefore $X = \mathrm{Proj}_{\otimes}(A)$ and (i) holds.

Conversely, assume that (i) holds and let $Y \in \mathcal{T}_+^{n,\otimes}$. Let $0 \leq \lambda \leq 1$ and consider the matrix $X(\lambda) = (1-\lambda)X + \lambda Y$. Then $X(\lambda) \in \mathcal{T}_+^{n,\otimes}$ as $\mathcal{T}_+^{n,\otimes}$ is convex. Consider the function

$$\begin{aligned} g(\lambda) &= \|A - X(\lambda)\|_F^2 = \|(1-\lambda)(A-X) + \lambda(A-Y)\|_F^2 \\ &= (1-\lambda)^2\|A-X\|_F^2 + \lambda^2\|A-Y\|_F^2 + 2\lambda(1-\lambda)\langle A-X, A-Y\rangle. \end{aligned}$$

So g is a quadratic function of λ and its (right-sided) derivative in $\lambda = 0$ is

$$g'_+(0) = -2\|A-X\|_F^2 + 2\langle A-X, A-Y\rangle = -2\langle A-X, Y-X\rangle$$

and this derivative must be nonnegative as $X(0) = X = \mathrm{Proj}_\otimes(A)$. But this gives the inequality $\langle A-X, Y-X\rangle \leq 0$ so (ii) holds.

To see the equivalence of (ii) and (iii) recall from Theorem 2.2 that each $Y \in \mathcal{T}_+^{n,\otimes}$ may be represented as a convex combination $Y = \sum_{j=1}^t \lambda_j Y_j$ where $\lambda_j \geq 0$ $(j \leq t)$ and $\sum_j \lambda_j = 1$ and each Y_j is a rank one matrix of the form described in the theorem. Therefore

$$\begin{aligned} \langle A-X, Y-X\rangle &= \langle A-X, \textstyle\sum_{j=1}^t \lambda_j Y_j - X\rangle = \langle A-X, \textstyle\sum_{j=1}^t \lambda_j (Y_j - X)\rangle \\ &= \textstyle\sum_{j=1}^t \lambda_j \langle A-X, Y_j - X\rangle. \end{aligned}$$

from which the desired equivalence easily follows. □

We remark that the *variational inequality* characterization given in (3) is the same as the optimality condition one obtains when formulating DA as the convex minimization problem

$$\min\{(1/2)\|A-X\|_F^2 : X \in \mathcal{T}_+^{n,\otimes}\}.$$

In fact, the gradient of $f(X) = (1/2)\|X-A\|_F^2$ is $\nabla f(X) = X - A$ and the optimality characterization here is $\langle \nabla f(X), Y-X\rangle \geq 0$ for all $Y \in \mathcal{T}_+^{n,\otimes}$, and this translates into (3). In the next section we consider an algorithm for DA that is based on the optimality conditions we have presented above.

4 The Frank-Wolfe method

We discuss how Theorem 3.2 may be the basis for an algorithm for solving the DA problem. The algorithm is an adaption of a general algorithm in convex programming called the *Frank-Wolfe method* (or the conditional gradient algorithm), see [2]. This is an iterative algorithm where a decent direction is found in each iteration by linearizing the objective function.

The main idea is as follows. Let $X \in \mathcal{T}_+^{n,\otimes}$ be a candidate for being the projection $\text{Proj}_\otimes(A)$. We check if $X = \text{Proj}_\otimes(A)$ by solving the optimization problem

$$\gamma(X) := \max\{\langle A - X, Y - X\rangle : Y \in \text{Ext}(\mathcal{T}_+^{n,\otimes})\}. \tag{4}$$

We discuss *how* this problem can be solved in the next section.

The algorithm for solving DA may be described in the following way.

The DA algorithm.

1. Choose an initial candidate $X \in \mathcal{T}_+^{n,\otimes}$.
2. Optimality test: solve the corresponding problem (4).
3. If $\gamma(X) \leq 0$, stop; the current solution X is optimal. Otherwise, let Y^* be an optimal solution of (4). Determine the matrix X' which is nearest to A on the line segment between X and Y^*.
4. Replace X by X' and repeat Steps 1-3 until an optimal solution has been found.

We now discuss this algorithm in some detail. Consider a current solution $X \in \mathcal{T}_+^{n,\otimes}$ and solve (4) as in Step 2. There are two possibilities. First, if $\gamma(X) \leq 0$, then, due to Theorem 3.2, we must have that $X = \text{Proj}_\otimes(A)$. Thus, the DA problem has been solved. Alternatively, $\gamma(X) > 0$ and we have found $Y^* \in \text{Ext}(\mathcal{T}_+^{n,\otimes})$ such that $\langle A-X, Y^*-X\rangle > 0$. This means that $g'_+(0) < 0$ for the function g introduced in the proof of Theorem 3.2: $g(\lambda) = \|A - X(\lambda)\|_F^2$ where $X(\lambda) = (1-\lambda)X + \lambda Y^*$. Let λ^* be an optimal solution in the (line search) problem $\min\{g(\lambda) : 0 \leq \lambda \leq 1\}$. Since g is a quadratic function in one variable, this minimum is easy to find analytically. Since $g'_+(0) < 0$, we have that $\lambda^* > 0$. The corresponding matrix $X' = X(\lambda^*)$ is the projection of A onto the line segment between X and Y^*, and (by convexity) this matrix lies in $\mathcal{T}_+^n$. In the final step we replace our candidate matrix X by X' and repeat the whole procedure for this new candidate.

The convergence of this Frank-Wolfe method for solving DA is assured by the following theorem (which follows from the general convergence theorem in Chapter 2 of [2]).

Theorem 4.1 *The DA algorithm produces a sequence of matrices $\{X^{(k)}\}$ that converges to $\mathit{Proj}_\otimes(A)$.*

Note here, however, that the method is based on solving the subproblem (4) in each iteration. We discuss this subproblem in the next section.

5 The projection subproblem

The DA algorithm is based on solving the subproblem (4). This is the optimality test of the algorithm. We now discuss an approach to solving this problem.

Consider problem (4) for a given X (and A, of course). Letting $B = A - X$ and separating out the constant term $\langle B, X\rangle$ we are led to the problem

$$\eta(B) := \max\{\langle B, Y\rangle : Y \in \mathrm{Ext}(\mathcal{T}_+^{n,\otimes})\}. \tag{5}$$

Based on Theorem 2.2 we know that the extreme points of $\mathcal{T}_+^n$ are the rank one separable density matrices $(x \otimes y)(x \otimes y)^T$ where $x \in \mathrm{I\!R}^p$ and $y \in \mathrm{I\!R}^q$ satisfy $\|x\| = \|y\| = 1$. So problem (5) becomes

$$\max\{g(x, y) : \|x\| = \|y\| = 1\}$$

where we define

$$g(x, y) = \langle B, (x \otimes y)(x \otimes y)^T\rangle.$$

This function g is a multivariate polynomial of degree 4, i.e. it is a sum of terms of the form $c_{ijkl}x_i x_j y_k y_l$. One can verify that g may not be concave (or convex). Therefore it seems difficult to find a global maximum of g subject to the two given equality constraints. However, the function g has a useful decomposable structure in the two variables x and y which leads to a practical and fast algorithm for finding a local maximum of g.

The idea is to use a *block coordinate ascent* approach (also called the *nonlinear Gauss-Seidel method*, see [2]) to the maximization of g. This iterative method consists in alternately fixing x and y and maximize with respect to the other variable. We now show that the corresponding subproblems (when either x or y is fixed) can be solved by eigenvalue methods.

First note that, by the mixed-product rule for tensor products,

$$Y = (x \otimes y)(x \otimes y)^T = (xx^T) \otimes (yy^T) = \begin{bmatrix} Y_{11} & \cdots & Y_{1p} \\ \vdots & & \vdots \\ Y_{p1} & \cdots & Y_{pp} \end{bmatrix}$$

where $Y_{ij} = x_i x_j (yy^T) \in \mathrm{I\!R}^{q\times q}$ $(i, j \le p)$. Partition the fixed matrix B conformly as

$$B = \begin{bmatrix} B_{11} & \cdots & B_{1p} \\ \vdots & & \vdots \\ B_{p1} & \cdots & B_{pp} \end{bmatrix}$$

where each B_{ij} is a $q \times q$ matrix. Note here that $B_{ij} = B_{ji}^T$ $(i, j \leq p)$ as B is symmetric. With this block partitioning we calculate

$$\begin{aligned} g(x,y) &= \langle B, Y \rangle = \textstyle\sum_{i,j \leq p} \langle B_{ij}, Y_{ij} \rangle \\ &= \textstyle\sum_{i,j \leq p} \langle B_{ij}, x_i x_j (yy^T) \rangle \\ &= \textstyle\sum_{i,j \leq p} x_i \langle B_{ij}, yy^T \rangle x_j \\ &= x^T \tilde{B}(y) x \end{aligned} \tag{6}$$

where $\tilde{B}(y) = [\tilde{b}_{ij}(y)]$ is a $p \times p$ matrix with entries $\tilde{b}_{ij}(y) = \langle B_{ij}, yy^T \rangle = y^T B_{ij} y$. The matrix $\tilde{B}(y)$ is symmetric.

Next we find another useful expression for $g(x, y)$.

$$\begin{aligned} g(x,y) &= x^T \tilde{B}(y) x \\ &= \textstyle\sum_{i,j \leq p} \tilde{b}_{ij}(y) x_i x_j \\ &= \textstyle\sum_{i,j \leq p} y^T B_{ij} y \cdot x_i x_j \\ &= y^T (\textstyle\sum_{i,j \leq p} x_i x_j B_{ij}) y = y^T \hat{B}(x) y \end{aligned} \tag{7}$$

where we define the matrix $\hat{B}(x)$ by $\hat{B}(x) = \sum_{i,j \leq p} x_i x_j B_{ij}$. Note that this matrix is symmetric as $B_{ij} = B_{ji}^T$.

We now use these calculations to solve the mentioned subproblems where x respectively y is fixed.

Theorem 5.1 *The following equations hold*

$$\begin{aligned} \eta(B) &= \max\nolimits_{x,y} g(x,y) \\ &= \max\{\lambda_{max}(\tilde{B}(y)) : \|y\| = 1\} \\ &= \max\{\lambda_{max}(\hat{B}(x)) : \|x\| = 1\}. \end{aligned}$$

Moreover, for given x, $\max_y g(x,y)$ is attained by a normalized eigenvector of

$\hat{B}(x)$*, and for fixed* y*,* $\max_x g(x,y)$ *is attained by a normalized eigenvector of* $\tilde{B}(y)$*.*

Proof. From equations (6) and (7) we get

$$\begin{aligned}\eta(B) &= \max\nolimits_{x,y} g(x,y)\\ &= \max\nolimits_{\|y\|=1} \max\nolimits_{\|x\|=1} x^T \tilde{B}(y) x\\ &= \max\nolimits_{\|x\|=1} \max\nolimits_{\|y\|=1} y^T \hat{B}(x) y.\end{aligned}$$

We now obtain the theorem from the following general fact: for every real symmetric matrix C we have that $\max_{\|z\|=1} z^T C z = \lambda_{max}(C)$ and that a maximizing z is a normalized eigenvector of C corresponding to $\lambda_{max}(C)$. □

Due to this theorem the block coordinate ascent method applied to the projection subproblem (4) gives the following scheme.

Algorithm: Eigenvalue maximization.

```
1. Choose an initial vector $y$ of length one.
2. Repeat the following two steps until convergence (or $g$ no
   longer increases).
      2a. Let $x$ be a normalized eigenvector corresponding to the largest
          eigenvalue of the matrix $\tilde{B}(y)$.
      2b. Let $y$ be a normalized eigenvector corresponding to the largest
          eigenvalue of the matrix $\hat{B}(x)$.
```

We now comment on the convergence issues for this algorithm. The constructed sequence of vectors $\{(x^{(k)}, y^{(k)})\}$ must have a convergent subsequence. Moreover, the sequence $\{g(x^{(k)}, y^{(k)})\}$ is convergent. These facts follow from standard compactness/continuity arguments since the direct product of the unit balls is compact, g is continuous, and the sequence $\{g(x^{(k)}, y^{(k)})\}$ is nondecreasing. If we assume that each of the coordinate maxima found by the algorithm is unique (which seems hard to verify theoretically), then it is known that every limit point of $\{(x^{(k)}, y^{(k)})\}$ will be a local maximum of g (see Proposition 2.7.1 in [2]).

It should be remarked here that there are some remaining open questions concerning convergence of our method. However, in view of the hardness of the DA problem (shown to be NP-hard in [6]) one can expect that solving the projection subproblem (4) is also hard. We should therefore not expect anything more than local maxima in general, although we may be lucky to find a global maximum of g in certain cases. We refer to Section 7 for some preliminary computational results for our methods.

A final remark is that it may also be of interest to consider other numerical approaches to the problem of maximizing g than the one proposed here. We have not tried this since the described eigenvalue approach seems to work quite well.

6 Improvement of the DA algorithm

The DA algorithm, as described in Section 4, turns out to show very slow convergence. In this section we discuss a modification of the method which improves the convergence speed dramatically.

The mentioned slow convergence of the DA algorithm may be explained geometrically as follows. Assume that the given matrix A is non-separable, and that the current separable matrix X is not on the boundary of the set $\mathcal{T}_+^{n,\otimes}$ of separable matrices. To find a separable matrix closer to A, in this case, a good strategy would be to move in the direction $A - X$ until the boundary is reached. The algorithm moves instead in a direction $Y - X$ which is typically almost orthogonal to $A - X$, because the product matrix Y (an extreme point) will typically be far away from X.

The basic weakness of the algorithm is that from iteration k to $k+1$ it retains only the current best estimate X, throwing away all other information about X. An alternative approach, which turns out to allow a much faster convergence, is to retain all information, writing X explicitly as a convex combination of the previously generated product matrices Y_k,

$$X = \sum_{r=1}^{k} \lambda_r Y_r \ ,$$

where $\lambda_r \geq 0$ and $\sum_r \lambda_r = 1$. After generating the next product matrix Y_{k+1} as a solution of the optimization problem (5) we find a new best convex combination varying all the coefficients λ_r, $r = 1, \ldots, k+1$.

An obvious modification of this scheme is to throw away in each iteration every Y_r getting a coefficient $\lambda_r = 0$. This means in practice that the number of product matrices retained does not grow too fast.

Thus, we are faced with the quadratic programming problem to minimize the squared distance $\|A - X\|^2$ as a quadratic polynomial in the coefficients λ_k. We have implemented a version of the conjugate gradient method (see [5]) for this problem. Theoretically, in the absence of rounding errors and inequality constraints, this method converges in a finite number of steps, and it also works well if the problem is degenerate, as is likely to happen here. The algorithm was adapted so it could handle the linear constraints $\lambda_i \geq 0$ for each i and

$\sum_i \lambda_i = 1$, but we omit describing the implementation details here. (Several fast algorithms for quadratic optimization with linear inequality constraints are available.)

7 Computational results

In this section we present preliminary results for the modified DA algorithm as described in Section 6. Moreover we present some results and experiences with the eigenvalue maximization algorithm (see Section 5), and, finally, we discuss an application which may serve as a test of our methods.

Since we want to apply the DA algorithm to the problem of separabilty in quantum physics, we need to work with complex matrices. Therefore we have implemented and tested the complex version of the algorithm, rather than the real version as described above. As already remarked, the generalization is quite straightforward and is not expected to change in any essential way the performance of the algorithm. The main change is that matrix transposition has to be replaced by hermitian conjugation, and hence real symmetric matrices will in general become complex and hermitian. Thus, for example, the matrices $\tilde{B}(y)$ and $\hat{B}(x)$ both become hermitian, which means that their eigenvalues remain real, and the problem of finding the largest eigenvalue becomes no more difficult.

7.1 Eigenvalue maximization

We have tested the performance of the eigenvalue maximization algorithm on a number of randomly generated matrices, and on some special matrices used in other calculations. We ran the algorithm for each input matrix a number of times with random starting vectors x and y, comparing the maximum values and maximum points found.

For completely random matrices we often find only one maximum. Sometimes we find two maximum points with different maximum values. In certain symmetric cases it may happen that two or more maximum points have the same maximum value. Thus, we are not in general guaranteed to find a global maximum, but we always find a local maximum.

One possible measure of the speed of convergence of the algorithm is the absolute value of the scalar product $\langle x \otimes y, x' \otimes y' \rangle$, where x, y are input vectors and x', y' output vectors in one iteration. It takes typically about 5 iterations before this overlap is about 10^{-3} from unity (when $p = q = 3$), which means that the convergence to a local maximum is fast. The algorithm involves the diagonalization of one $p \times p$ matrix and one $q \times q$ matrix for each iteration, and in addition it may happen that more iterations are needed when the dimensions increase.

7.2 Results with the present program version

In Table 1 we show the performance of the modified DA algorithm for different dimensions of the problem. We take $p = q$, and we use the maximally entangled matrices in the given dimensions, as we know the distance to the closest separable matrix in these special cases. A maximally entangled matrix is a pure state $A = uu^T$ (or $A = uu^\dagger$ if u is a complex vector), where

$$u = \frac{1}{\sqrt{p}} \sum_{i=1}^{p} e_i \otimes f_i$$

and where $\{e_i\}$ and $\{f_i\}$ are two sets of orthonormal basis vectors in $\mathbb{R}^p$ (or in $\mathbb{C}^p$). The closest separable state is $A' = \lambda A + (1-\lambda)(1/p^2)I$ with $\lambda = 1/(p+1)$, and the distance to it from A is $\sqrt{(p-1)/(p+1)}$. The density matrix $(1/n)I$, where I is the $n \times n$ unit matrix, is called the maximally mixed state.

The number of iterations of the main algorithm is set to 1000, and the fixed number of iterations used in the eigenvalue maximization algorithm (see Section 5) is set to 20. The tabulated time t is the total execution time on one computer, and the tabulated error is the difference between the calculated distance and the true distance.

p, q	$n = pq$	t (s)	error
2	4	312	$3 \cdot 10^{-13}$
3	9	155	$3 \cdot 10^{-12}$
4	16	127	$3 \cdot 10^{-8}$
5	25	773	$1 \cdot 10^{-6}$
6	36	2122	$5 \cdot 10^{-6}$
7	49	3640	$1.0 \cdot 10^{-5}$
8	64	4677	$1.5 \cdot 10^{-5}$
9	81	5238	$2.2 \cdot 10^{-5}$
10	100	6566	$3.5 \cdot 10^{-5}$

Table 1
Performance of the modified DA algorithm

The main conclusion we may draw is that the accuracy obtained for a fixed number of iterations decreases with increasing dimension. It should be noted that the rank one matrices used here are somewhat special, and that higher rank matrices give less good results. We want to emphasize that this is work in progress, and some fine tuning remains to be done. Nevertheless, we conclude at this stage that the method can be used for quite large matrices, giving useful results in affordable time.

7.3 An application

In the special cases $p = q = 2$ and $p = 2, q = 3$ there exists a simple necessary and sufficient criterion for separability of complex matrices (see [13], [9]). We have checked our method against this criterion for $p = q = 2$.

Figure 1 shows a two-dimensional cross section of the 15 dimensional space of complex 4×4 density matrices. The section is defined by three matrices: a maximally entangled matrix as described above, called a Bell matrix when $p = q = 2$; the maximally mixed state $(1/4)I$; and a rank one product matrix. The plot shows correct distances according to the Frobenius norm.

The algorithm finds the minimal distance from a given matrix A to a separable matrix. We choose an outside matrix A and gradually mix it with $(1/4)I$, until we find an $A' = \lambda A + (1 - \lambda)(1/4)I$, with $0 \leq \lambda \leq 1$, for which the distance is less than $5 \cdot 10^{-5}$. Then we plot A' as a boundary point.

When we start with A entangled, the closest separable matrix does not in general lie in the plane plotted here. Hence, we may certainly move in the plotting plane as much as the computed distance without crossing the boundary we are looking for. In this way we get very close to the boundary, approaching it from the outside, in very few steps.

In Figure 1, the curved line is generated from the necessary, and in this case sufficient, condition for separability. All matrices below this line are separable, while the others are not. The 6 plotted boundary points are computed by our algorithm. The matrices to the right of the vertical straight line and below the skew straight line are positive definite, and the Bell matrix is located where the two lines cross. The maximally mixed state $(1/4)I$ is the origin of the plot.

Finally, we refer to the recent paper [11] for a further study of geometrical aspects of entanglement and applications of our algorithm in a study of bound entanglement in a composite system of two three-level systems.

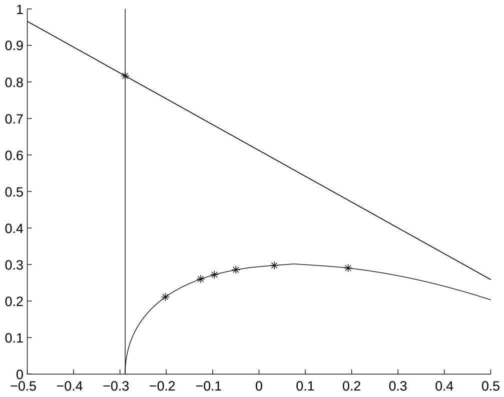

Fig. 1. The boundary of the set of separable matrices. The axes show distances by the Frobenius norm.

Acknowledgment. The authors thank a referee for several useful comments and suggestions.

References

[1] D.P. Bertsekas. *Convex Analysis and Optimization.* Athena Scientific, 2003.

[2] D.P. Bertsekas. *Nonlinear Programming.* Athena Scientific, 1999.

[3] S. Boyd and L. Vandenberge. *Convex Optimization.* Cambridge University Press, Cambridge, 2004.

[4] A.C. Doherty, P.A. Parillo and F.M. Spedalieri. *Distinguuishing separable and entangled states*, Phys. Rev. Lett. **88**, 187904 (2002).

[5] G.H. Golub and C.F. Van Loan. *Matrix Computations.* The John Hopkins University Press, Baltimore, 1993.

[6] L. Gurvits. *Classical deterministic complexity of Edmonds' problem and quantum entanglement.* In *Proceedings of the Thirty-Fifth ACM Symposium on Theory of Computing* (ACM, New York, 2003), pp. 10-19.

[7] R.A. Horn and C.R. Johnson. *Matrix Analysis.* Cambridge University Press, 1991.

[8] R.A. Horn and C.R. Johnson. *Topics in Matrix Analysis.* Cambridge University Press, 1995.

[9] M. Horodecki, P. Horodecki and R. Horodecki. *Separability of mixed states: necessary and sufficient conditions*, Phys. Lett. A **223**, 1 (1996).

[10] L.M. Ioannou, B.C. Travaglione, D. Cheung and A.K. Ekert. *Improved algorithm for quantum separability and entanglement detection*, Phys. Rev. A **70**, 060303 (2004).

[11] J.M. Leinaas, J. Myrheim and E. Ovrum. *Geometrical aspects of entanglement*, Phys. Rev. A 74, 12313 (2006)

[12] M. Ozawa. *Entanglement measures and the Hilbert–Schmidt distance*, Phys. Lett. A **268**, 158 (2000).

[13] A. Peres. *Separability criterion for density matrices*, Phys. Rev. Lett. **77**, 1413 (1996).

[14] A.O. Pittenger and M.H. Rubin. *Geometry of entanglement witnesses and local detection of entanglement*, Phys. Rev. A **67**, 012327 (2003).

[15] C.F. Van Loan and N. Pitsianis. *Approximation with Kronecker products.* Moonen, Marc S. (ed.) et al., Linear algebra for large scale and real- time applications. Proceedings of the NATO Advanced Study Institute, Leuven, Belgium, August 3 - 14, 1992. Dordrecht: Kluwer Academic Publishers. NATO ASI Ser., Ser. E, Appl. Sci. 232, 293-314 (1993).

[16] F. Verstraete, J. Dehaene and B. De Moor. *On the geometry of entangled states* Jour. Mod. Opt. **49**, 1277 (2002).

[17] R. Webster. *Convexity.* Oxford University Press, Oxford, 1994.