Numerical estimation of the relative entropy of entanglement

Yuriy Zinchenko,[1,*] Shmuel Friedland,[2,†] and Gilad Gour[2,3,‡]

[1]*Department of Mathematics and Statistics, University of Calgary, 2500 University Drive NW, Calgary, Alberta T2N 1N4, Canada*
[2]*Department of Mathematics, Statistics and Computer Science, University of Illinois at Chicago,*
851 S. Morgan Street, Chicago, Illinois 60607-7045, USA
[3]*Institute for Quantum Information Science, University of Calgary, 2500 University Drive NW, Calgary, Alberta T2N 1N4, Canada*

We propose a practical algorithm for the calculation of the relative entropy of entanglement (REE), defined as the minimum relative entropy between a state and the set of states with positive partial transpose. Our algorithm is based on a practical semidefinite cutting plane approach. In low dimensions the implementation of the algorithm in MATLAB provides an estimation for the REE with an absolute error smaller than 10^{-3}.

I. INTRODUCTION

With the emergence of quantum information science, entanglement was recognized as a valuable resource for a variety of quantum information tasks, such as teleportation and superdense coding (for review, see, e.g., Refs. [1,2]). With this recognition, much of the research in the field focused on how to quantify this quantum resource. Among the different proposed measures quantifying entanglement, the relative entropy of entanglement (REE) is one of the most important ones. Its importance comes from the fact that its asymptotic version provides the unique rate for reversible transformations [3]; recently it was discovered that the regularized REE is the unique function that quantifies the rate of interconversion between states in a reversible theory of entanglement, where all types of nonentangling operations are allowed [4].

The REE is defined by [5]

$$E_R(\rho) = \min_{\sigma' \in \mathcal{D}} S(\rho \| \sigma') \equiv S(\rho \| \sigma), \tag{1}$$

where $\mathcal{D}$ is the set of separable states or, as will be considered in this paper, the set of positive partial transpose (PPT) states, and σ is the corresponding (possibly nonunique) minimizer. Recall that the relative entropy is defined by

$$S(\rho \| \sigma') \equiv \mathrm{Tr}(\rho \ln \rho - \rho \ln \sigma').$$

The REE quantifies to what extent a given state can be operationally distinguished from the closest PPT state. Besides being an entanglement monotone it also has useful properties such as asymptotic continuity.

The state $\sigma = \sigma(\rho)$ in Eq. (1) is called the closest PPT state (or closest separable state). Recently, the inverse problem to the long-standing problem [6] of finding the formula for the closest PPT state $\sigma(\rho)$ was solved in Ref. [7] for the case of two qubits and in Ref. [8] for all dimensions and for any number of parties. In Refs. [7,8] the authors found a closed formula for the inverse problem; that is, for a given state σ on the boundary of PPT states (or separable states) $\partial \mathcal{D}$, an explicit formula was found describing all entangled states for which σ is the closest PPT

state. However, despite the complete solution in Ref. [8] for the inverse problem, the original problem remained unsolved. Moreover, the results for the inverse problem in Refs. [7,8], suggest that the solution to the original problem may not be analytical—hence the need for numerical estimation of the REE.

In this paper, we provide an algorithm for the calculation of the REE, although other authors have also used semidefinite programming for the separability problem (see, e.g., Refs. [9–11]). At the core of our approach lies a classical geometric idea that a convex set may be, in principle, approximated by its supporting hyperplanes. Our algorithm is based on the so-called cutting planes combined with positive semidefinite optimization. Namely, we attempt to successively refine the epigraph of the relative entropy function $S(\rho \| \sigma')$ using supporting hyperplanes, building a piecewise linear lower approximation to $S(\rho \| \sigma')$, while using positive semidefinite optimization techniques to characterize our feasible set $\mathcal{D}$. In low dimensions where $\mathcal{D}$ corresponds to qubit-qubit and qubit-qutrit pairs, our numerical approach appears to recover the REE value up to a small absolute error in a relatively short time without much difficulty. For example, on a modest desktop computer we could compute the REE for 50 randomly generated qubit-qutrit problems with an average run time of under 1 s per problem with preset precision of 10^{-2} for the REE; see Sec. IV for more details. In addition, in higher dimensions, our approach readily generalizes to computing an approximate REE over PPT states rather than the more elusive separable states set.

The paper is organized as follows. In Sec. II we introduce notations and preliminary results concerning the REE. We then discuss our algorithm in Sec. III. In Sec. IV we discuss our computational results. Finally, in Sec. V we summarize our findings with a few conclusions. In addition, in Ref. [12] readers will find a link to the MATLAB driver code.

II. PRELIMINARIES

Here we will use the following notation. Let $\mathbb{C}^{m \times n}$ denote the space of $m \times n$ complex-valued matrices. $A \in \mathbb{C}^{m \times n}$ has entries $A_{i,j}$, $i = 1, \ldots, m$, $j = 1, \ldots, n$. $A^\dagger$ is the conjugate transpose of A. $\mathrm{H}_k \subset \mathbb{C}^{k \times k}$ is the space of Hermitian $k \times k$ matrices, $\mathrm{H}_{k,+} \subset \mathrm{H}_k$ is the closed convex cone of Hermitian positive semidefinite $k \times k$ matrices, $\mathrm{H}_{k,++} \subset \mathrm{H}_{k,+}$ is the open

---

*yzinchen@math.ucalgary.ca
†friedlan@uic.edu
‡gour@math.ucalgary.ca

convex cone of Hermitian positive definite $k \times k$ matrices, and $H_{k,+,1} \subset H_{k,+}$ is the set of Hermitian positive semidefinite matrices of trace 1 (i.e., density matrices).

We write $A \geqslant 0, A \gneqq 0, A > 0$ if A is positive semidefinite, nonzero positive semidefinite, and positive definite, respectively. More generally for $A, B \in H_k$, we let

$$
\begin{aligned}
A \geqslant B &\iff A - B \geqslant 0, \\
A \gneqq B &\iff A - B \gneqq 0, \\
A > B &\iff A - B > 0.
\end{aligned}
$$

We denote by I_k the identity matrix of order k, which is also denoted by I when no ambiguity arises. A standard inner product on $H_k \times H_k$ is the trace inner product defined as $\langle A, B \rangle = A \circ B = \text{trace}(A^\dagger B)$; for concreteness, we may fix our norm choice $\|A\| = \sqrt{\langle A, A \rangle}$.

Remark: Unlike our discussion in Sec. I, for the remainder of the paper we adapt more math-and-optimization-friendly notations, because in detailing our approach we generously borrow on optimization techniques. For an excellent introduction to the area of so-called *convex optimization,* see Ref. [13].

Here $\text{eig}(X) = \{\xi_i, \mathbf{x}^{(i)}\}_{i=1,k}$ denotes the set of eigenvalue-eigenvector pairs of $X \in H_k$. We assume that $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k)}$ is an orthonormal basis in $\mathbb{C}^k$, called an eigenbasis of X. Then X is a diagonal matrix $\text{diag}(\xi_1, \ldots, \xi_k)$ in its eigenbasis, which is also denoted as $\text{diag}(\boldsymbol{\xi})$, where $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_k)$ is the eigenvalue multiset.

For $X \in H_{k,++}$ denote by $\ln X$ the logarithm function of X. In the eigenbasis of X, $\ln X$ is the diagonal matrix $\text{diag}(\ln \boldsymbol{\xi}) = \text{diag}(\ln \xi_1, \ldots, \ln \xi_k)$. It is well known that $\ln X$ is order preserving, i.e., $\ln X \geqslant \ln Y$ for $X \geqslant Y > 0$, and matrix-concave, i.e., $\ln[tX + (1-t)Y] \geqslant t \ln X + (1-t) \ln Y$ for $X, Y > 0, t \in [0,1]$ on $H_{k,++}$; see, e.g., Ref. [14]. Moreover, $\ln X$ is also strongly order preserving [8]: if $X \gneqq Y > 0$, then $\ln X \gneqq \ln Y$ (more results about strict concavity of $\ln X$ can be found in Ref. [8]).

Assume we are given fixed integers $m, n > 1$, and a fixed nonzero $A \in H_{mn,+,1}$. We are interested in finding the REE of A, that is, in finding

$$
\begin{aligned}
E_R(A) &= \inf_{X \in \mathcal{D}_{m,n}} S(A \| X) \\
&= \inf_X \left\{ A \circ \ln A - A \circ \ln X : X \in \mathcal{D}_{m,n} \right\},
\end{aligned}
$$

where the set of *separable states* $\mathcal{D}_{m,n}$ is defined as

$$
\mathcal{D}_{m,n} = \text{conv}(H_{m,+,1} \otimes H_{n,+,1}) \subset H_{mn,+,1}
$$

with $\text{conv}(\cdot)$ denoting a convex hull and $\otimes$ denoting the standard Kronecker product. Recall that the Kronecker product $A \otimes B$ of two matrices A, B may be defined using a block-matrix notation as a matrix with (i, j) blocks $(A \otimes B)_{i,j} = a_{i,j} B, \forall i, j$. The following basic fact justifies the inclusion $\mathcal{D}_{m,n} \subset H_{mn,+,1}$ and is readily established from the definition: If $\text{eig}(A) = \{\alpha_i, \mathbf{a}^{(i)}\}_{i=1,m}$ denotes the set of eigenvalue-eigenvector pairs of $A \in H_m$, and, likewise, $\text{eig}(B) = \{\beta_j, \mathbf{b}^{(j)}\}_{j=1,n}, B \in H_n$, then $\text{eig}(A \otimes B) = \{\alpha_i \beta_j, \mathbf{a}^{(i)} \otimes \mathbf{b}^{(j)}\}_{i=1,m, j=1,n}$.

Clearly, finding $E_R(A)$ in this definition is equivalent to solving

$$
\inf_X \{-A \circ \ln X : X \in \mathcal{D}_{m,n}\}. \tag{2}
$$

Since $\mathcal{D}_{m,n}$ is compact and the objective $-A \circ \ln X$ is continuous, the infimum is achieved and may as well be replaced by the minimum.

We turn our attention to the objective function in (2). If we assume that X is a diagonal matrix $\text{diag}(\boldsymbol{\xi}) \geqslant 0$ and $A \gneqq 0$, then $-A \circ \ln X = \sum_{i=1}^{k} A_{i,i} \ln \xi_i$. Therefore, in the limiting sense, $-A \circ \ln X = \infty$ if and only if $A_{i,i} > 0$ and $\xi_i = 0$ for some i. Likewise, if $A > 0$ and $X \geqslant 0$ is singular, then $A \circ \ln X = -\infty$. In the same limiting sense we may write $0 \ln \xi = 0$ for any $\xi \geqslant 0$. So our objective function $-A \circ \ln X$ may approach infinity or stay bounded as X approaches the boundary of $\mathcal{D}_{m,n}$. We may state our first elementary proposition.

Proposition II.1. The objective function of (2), $-A \circ \ln X$, is convex with respect to X on $H_{mn,++}$. Moreover, the minimum of $-A \circ \ln X$ over $H_{mn,++}$ is $-A \circ \ln A$ and is achieved at $X = A$. Consequently, if $A \notin \text{int } \mathcal{D}_{m,n}$, the interior of $\mathcal{D}_{m,n}$, then the solution to (2) lies on the boundary of $\mathcal{D}_{m,n}$.

For brevity we sketch only the proof; using a limiting argument one has to carefully work out the boundary case where $-A \circ \ln X$ approaches infinity. Convexity follows from $\ln X$ being matrix-concave on $H_{mn,++}$. The first differential of $-A \circ \ln X$ vanishes at affine-feasible $X = A + \Delta$, which is a sufficient condition for the minimum of a smooth convex function. Namely, borrowing from expressions (4) and (3) in Sec. III A, we know that for small $\Delta \in H_{mn}$ with trace $\Delta = 0$, up to first order $-A \circ \ln(A + \Delta) \approx -A \circ \ln A + \Delta \circ I = -A \circ \ln A$. Finally, if $A \notin \text{int } \mathcal{D}_{m,n}$, the last part follows from both the t-level sets of the objective function $\{X \in H_{mn,+,1} : -A \circ \ln X \leqslant t\}$ as well as $\mathcal{D}_{m,n}$, being convex.

In fact, one may state an even stronger statement about $\ln X$ and the objective as in Ref. [8, App. A].

Theorem II.2. The function $\ln X$ is strongly matrix-concave on $H_{k,++}$. Consequently, the function $-A \circ \ln X$ is strictly convex on $H_{k,++}$ for each fixed $A \in H_{k,++}$.

Therefore, if $A > 0$, then the solution to (2) is given by a unique boundary point [8]. If A is singular, then $X(A)$ does not have to be unique [7]. In general, the set of minimizing separable states form a compact convex subset of $\mathcal{D}_{m,n}$ that intersects the boundary of $\mathcal{D}_{m,n}$.

The problem (2) may look deceptively simple. In general, even deciding if a certain X is separable is NP-hard [15]; see Ref. [16] for a similar problem pertaining to the graph isomorphism. Thus, despite the fact that (2) is a convex optimization problem—the objective function is convex—we do not attempt to solve (2) simply relying on, for example, the so-called (polynomial) ellipsoid method, as finding a separating oracle for $\mathcal{D}_{m,n}$ might already be quite difficult. To emphasize the distinction between $\mathcal{D}_{m,n}$ and its containing superset $H_{mn,+,1}$ we state our next simple proposition.

Proposition II.3. $\mathcal{D}_{m,n}$ is a proper subset of $H_{mn,+,1}$ and, in particular, has flat faces besides those that coincide with faces of $H_{mn,+,1}$.

Proof. Consider the rank-1, a.k.a. a *pure state*, matrix $X = \mathbf{x}\mathbf{x}^\top$, where $\mathbf{x} = \frac{1}{\sqrt{2}}(1, 0, \ldots, 0, 1)$. Clearly $X \in H_{mn,+,1}$ but not in $\mathcal{D}_{m,n}$, so indeed $\mathcal{D}_{m,n} \neq H_{mn,+,1}$. (Here and in

the subsequent discussion, to see that $X \notin \mathcal{D}_{m,n}$ we rely on Fact II.4.)

Likewise, let $Y = ((1,0) \cdot (1,0)^\top) \otimes ((1,0) \cdot (1,0)^\top) \in \mathcal{D}_{2,2}$ and $Z = ((0,1) \cdot (0,1)^\top) \otimes ((0,1) \cdot (0,1)^\top) \in \mathcal{D}_{2,2}$, and consider $\frac{Y+Z}{2} \in \mathcal{D}_{2,2}$: clearly $\frac{Y+Z}{2}$ cannot be written as $A \otimes B$ for any $A, B \in H_{2,+,1}$. The last construction may be easily generalized to arbitrary $m, n > 1$. ∎

Surprisingly, in small dimensions $m = 2$, $n = 2, 3$, the set $\mathcal{D}_{m,n}$ admits very easy alternative characterization using the so-called *partial transpose* [17]. The partial transpose (with respect to the second factor) is the linear map $PT_2 : H_{mn} \to H_{mn}$, which is given by partial exchange of matrix entry indices and may be conveniently stated as $PT_2(A \otimes B) = A \otimes B^\top$.

Fact II.4. For $m = 2$, $n = 2, 3$, $X \in \mathcal{D}_{m,n}$ if and only if $X \in H_{mn,+,1}$ and the corresponding partial transpose $PT_2(X) \in H_{mn,+}$. Moreover, for higher dimensions m, n the partial transpose condition is necessary.

(The necessity easily follows from the mentioned property of the eigenvalues for the Kronecker product.)

The characterization of $\mathcal{D}_{m,n}$, $m = 2, n = 2, 3$, is equivalent to viewing the set as an affine slice of $H_{k,+,1}$ for some $k > 0$, which in turn allows us to capture $\mathcal{D}_{m,n}$ via *positive semidefinite optimization* [18,19] techniques. Thus, in the latter case of $m = 2$, $n = 2, 3$ where the feasible set is relatively simple and the derivative of the objective function $-A \circ \ln X$ is relatively easy to compute (see Sec. III), we do expect to find an efficient approach to solving (2), as in, for example, Ref. [13].

III. THE APPROACH

Our main goal is solving (2), and the proposed technique is fairly simple. Since the objective function is convex, we rely on constructing a successively refined sequence of approximations to the epigraph of the objective restricted to int $\mathcal{D}_{m,n} \subset H_{mn,+,1}$, defined as

$$\mathrm{epi}(-A \circ \ln X)|_{\mathrm{int}\,\mathcal{D}_{m,n}}$$
$$= \{(X,t) \in \mathrm{int}\,\mathcal{D}_{m,n} \times \mathbb{R} : -A \circ \ln X \leqslant t\},$$

which in turn may be viewed as a convex set. The resulting relaxations to (2) corresponding to the positive semidefinite optimization may be efficiently solved with off-the-shelf (and even free) numerical software.

We do not attempt to prove the convergence of the proposed approach. Instead, we provide numerical evidence of its efficiency. We want to remark that our approach is consistent with the classical cutting-planes approach in nonlinear optimization, with the latter being already well established in many practical settings.

A. First differential of ln X and approximate epi($-A \circ \ln X$)|$_{\mathrm{int}\,\mathcal{D}_{m,n}}$

Recall that for diagonal matrix $\mathrm{diag}(\boldsymbol{\lambda}) \in H_{k,++,1}, \boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_k)$, the first-order expansion of $\ln[\mathrm{diag}(\boldsymbol{\lambda}) + \Delta X]$ for small $\Delta \in H_k$ may be written as [20], (6.6.31)]

$$\ln[\mathrm{diag}(\boldsymbol{\lambda}) + \Delta X] = \ln \mathrm{diag}(\boldsymbol{\lambda}) + D(\boldsymbol{\lambda}) \cdot \Delta + O(\|\Delta\|^2),$$

where $D(\boldsymbol{\lambda})$ is the $k \times k$ Hermitian matrix

$$D(\boldsymbol{\lambda})_{i,j} = \begin{cases} \frac{\ln \lambda_i - \ln \lambda_j}{\lambda_i - \lambda_j}, & \lambda_i \neq \lambda_j, \\ \frac{1}{\lambda_i}, \text{ i.e.,} & \text{the limiting value}, \quad \lambda_i = \lambda_j, \end{cases}$$

$A \cdot B$ denotes the entrywise (Hadamard) product of two matrices, and $O(\|\Delta\|^2)$ is a $k \times k$ matrix with $O(\|\Delta\|^2)$ entries; the regular matrix product takes precedence over the entry-wise product.

Recall that $\ln(UXU^\dagger) = U \ln(X) U^\dagger$ for unitary U and any X. Thus, for arbitrary $X \in H_{k,++}$, the first-order expansion of $\ln(X + \Delta)$ may be written by applying spectral decomposition of $X = U_X \mathrm{diag}(\boldsymbol{\lambda}_X) U_X^\dagger$ with the unitary matrix of eigenvectors U_X and the matrix $\mathrm{diag}(\boldsymbol{\lambda}_X)$ of the corresponding eigenvalues $\boldsymbol{\lambda}_X$ of X, resulting in:

$$\ln(X + \Delta)$$
$$= \ln[U_X \mathrm{diag}(\boldsymbol{\lambda}_X) U_X^\dagger + U_X U_X^\dagger \Delta U_X U_X^\dagger]$$
$$= U_X[\ln \mathrm{diag}(\boldsymbol{\lambda}_X) + D(\boldsymbol{\lambda}_X) \cdot U_X^\dagger \Delta U_X + O(\|\Delta\|^2)]U_X^\dagger$$
$$= \ln X + U_X[D(\boldsymbol{\lambda}_X) \cdot U_X^\dagger \Delta U_X]U_X^\dagger + O(\|\Delta\|^2).$$

Using this and a sequence of fixed points $X^{(i)} \in$ int $\mathcal{D}_{m,n}$, $i = 0, \ldots, N$, we may proceed with constructing an approximation to epi$(-A \circ \ln X)|_{\mathrm{int}\,\mathcal{D}_{m,n}}$. Namely, since $-A \circ \ln X$ is convex on int $\mathcal{D}_{m,n}$, its epigraph is supported by tangent hyperplanes at every $X^{(i)}$, and so epi$(-A \circ \ln X)|_{\mathrm{int}\,\mathcal{D}_{m,n}}$ is a subset of all $(X,t) \in$ int $\mathcal{D}_{m,n} \times \mathbb{R}$ satisfying

$$\max_i -A \circ \{\ln X^{(i)}$$
$$+ U_{(i)}[D(\boldsymbol{\lambda}_{(i)}) \cdot U_{(i)}^\dagger (X - X^{(i)}) U_{(i)}]U_{(i)}^\dagger\} \leqslant t,$$

where for brevity of notation we write $U_{(i)}, \boldsymbol{\lambda}_{(i)}$ for $U_{X^{(i)}}, \boldsymbol{\lambda}_{X^{(i)}}$, or equivalently, for all i,

$$A \circ [U_{(i)}(D(\boldsymbol{\lambda}_{(i)}) \cdot U_{(i)}^\dagger X U_{(i)})U_{(i)}^\dagger] + t - s_i$$
$$= -A \circ \ln X^{(i)} + A \circ [U_{(i)}(D(\boldsymbol{\lambda}_{(i)}) \cdot U_{(i)}^\dagger X^{(i)} U_{(i)})U_{(i)}^\dagger]$$

with $s_i \geqslant 0$, which is the same as

$$E^{(i)} \circ X + t - s_i = -A \circ \ln X^{(i)} + E^{(i)} \circ X^{(i)}, \forall i$$

with $s_i \geqslant 0$ and

$$E^{(i)} = [U_{(i)}(D(\boldsymbol{\lambda}_{(i)}) \cdot U_{(i)}^\dagger A U_{(i)})U_{(i)}^\dagger], \qquad (3)$$

since $A \cdot B = B \cdot A$. Note that trace$[(A \cdot B)C] = $ trace$[A(B^\top \cdot C)]$ and $D(\boldsymbol{\lambda}_{(i)}) = D(\boldsymbol{\lambda}_{(i)})^\top$. Finally, if we recall that from Proposition II.1 we have $t \geqslant -A \circ \ln A$, the approximation to epi$(-A \circ \ln X)|_{\mathrm{int}\,\mathcal{D}_{m,n}}$ may be refined to

$$E^{(i)} \circ X + t - s_i = -A \circ \ln X^{(i)} + E^{(i)} \circ X^{(i)}, \qquad (4)$$
$$t \geqslant -A \circ \ln A, \ s_i \geqslant 0, \ \forall i = 0, \ldots, N.$$

Note that (4) contains only linear inequalities. Intuitively, if N is large and the sequence $X^{(i)}$ forms a nearly dense cover of $\mathcal{D}_{m,n}$, then the approximation (4) to the epigraph of the objective would be fairly accurate. Consequently, in order to find an approximate solution to our problem (2), we might as well work with the approximation to epi$(-A \circ \ln X)|_{\mathrm{int}\,\mathcal{D}_{m,n}}$, minimizing t. That is, instead of (2) we want to find the minimum of t satisfying (4) over all $X \in \mathcal{D}_{m,n}$. It turns out that the latter problem may be handled efficiently numerically

as described in Sec. III C, where we also propose a scheme for generating $X^{(i)}$ incrementally with the end goal of keeping N small.

B. Equivalent reparametrization of $\mathcal{D}_{m,n}, m = 2, n = 2,3$

Hereafter $\mathcal{D}_{m,n}$ represents the set of separable states only for the cases of $m = 2$ and $n = 2$ or 3. In higher dimensions, our approach will work if we take $\mathcal{D}_{m,n}$ to be the set of PPT states. We adopt a somewhat unusual reparametrization of $\mathcal{D}_{m,n}$ relying on Fact II.4, which will prove to be useful in Sec. III C.

Recall that $X \in \mathcal{D}_{m,n}$ if and only if $X \in H_{mn,+,1}$ and $PT_2(X) \in H_{mn,+}$. That is, $X \in H_{mn,+}$, trace $X = 1$, and $PT_2(X) \in H_{mn_+}$. These conditions may be rewritten in the following way. Denote

$$F^{(i,j,o,p)} = \left(\mathbf{e}^{(i)}\mathbf{e}^{(j)\top}\right) \otimes \left(\underline{\mathbf{e}}^{(o)}\underline{\mathbf{e}}^{(p)\top}\right),$$
$$G^{(i,j,o,p)} = \left(\mathbf{e}^{(i)}\mathbf{e}^{(j)\top}\right) \otimes \left(\underline{\mathbf{e}}^{(p)}\underline{\mathbf{e}}^{(o)\top}\right). \tag{5}$$

Here $\mathbf{e}^{(i)}, \mathbf{e}^{(j)}$ and $\underline{\mathbf{e}}^{(o)}, \underline{\mathbf{e}}^{(p)}$ are the standard unit vectors in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively; i.e., $\mathbf{e}_i^{(i)} = 1$ and $\mathbf{e}_j^{(i)} = 0, j \neq i$. Then the conditions that $X \in \mathcal{D}_{m,n}$ can be written as follows:

(a) Unit trace of X:

$$I \circ X = 1. \tag{6}$$

(b) Diagonal blocks of $PT_2(X)$:

$$F^{(i,i,o,p)} \circ X - G^{(i,i,o,p)} \circ Y = 0,$$
$$\forall i = 1, \ldots, m,$$
$$\forall o = 1, \ldots, n, \ p = o, \ldots, n. \tag{7}$$

(c) Strictly upper-triangular blocks of $PT_2(X)$:

$$F^{(i,j,o,p)} \circ X - G^{(i,j,o,p)} \circ Y = 0,$$
$$\forall i = 1, \ldots, m, \ j = i + 1, \ldots, m,$$
$$\forall o = 1, \ldots, n, \ p = 1, \ldots, n \tag{8}$$

(d) Positive semidefiniteness of

$$PT_2(X) \ X, Y \in H_{mn,+}. \tag{9}$$

Here the matrix Y corresponds to $PT_2(X)$. Note that due to X, Y being Hermitian, it suffices to specify the reordering of the entries of X into $Y = PT_2(X)$ only for the upper-triangular part of X.

C. Sequential positive semidefinite optimization approach

A problem corresponds to positive semidefinite optimization [18] if it may be written as

$$\inf_X \{C \circ X : A^{(i)} \circ X = b_i, i = 1, \ldots \ell, X \in H_{k,+}\},$$

where $k \times k$ matrices $C, A^{(i)}$ are fixed. Note that this includes block-diagonally structured matrices by choosing appropriate $A^{(i)}, b_i$ to zero-out off-diagonal blocks, nonnegativity constraints $\mathbf{x} \geqslant \mathbf{0}$ by considering diagonal 1×1 positive semidefinite blocks, and the so-called free variables by taking $\mathbf{x} = \mathbf{y} - \mathbf{z}$, where $\mathbf{y}, \mathbf{z} \geqslant \mathbf{0}$. Without loss of generality we may assume $C, A^{(i)} \in H_k$; indeed, for any $k \times k$ matrix C and

$X \in H_k$ we have

$$C \circ X = \text{trace}\left(\frac{CX + CX}{2}\right)$$
$$= \frac{\text{trace}(CX)}{2} + \frac{\text{trace}(X^\dagger C^\dagger)}{2} = \left(\frac{C + C^\dagger}{2}\right) \circ X.$$

So, even if $C, A^{(i)}$ are non-Hermitian initially, we may easily symmetrize those by following this procedure.

Efficient numerical algorithms exist for solving this type of problem. In particular, the so-called *interior-point methods* [18] provide both the theoretical polynomial bound on the number of iterations needed to find an ϵ-approximate solution and good practical performance. Numerous freely available interior-point method solvers, e.g., SeDuMi, SDPT3, CSDP, and DSDP, are suitable for solving this type of problem. For our purposes we rely on the MATLAB-based SeDuMi solver.

The approximation of $\text{epi}(-A \circ \ln X)|_{\text{int } \mathcal{D}_{m,n}}$ given by *** combined with the parametrization of $\mathcal{D}_{m,n}$ given by (6), (7), (8), and (9) forms a basis for our numerical scheme to compute an approximate solution to (2). Based on refining our approximation of $\text{epi}(-A \circ \ln X)|_{\text{int } \mathcal{D}_{m,n}}$, we build a sequence of positive semidefinite optimization problems, with each problem resulting in a finer and finer approximation to a solution of (2).

Let r^* denote the optimal value of (2),

$$r^* = \inf_X \{-A \circ \ln X : X \in \mathcal{D}_{m,n}\}.$$

As we explained in Sec. II, r^* is attainable, and we may replace inf by min. Let $\underline{r} \leqslant r^* \leqslant \bar{r}$ denote lower and upper bounds on r^*, respectively. For a given $\varepsilon > 0$, we say that $X^* \in \mathcal{D}_{m,n}$ is an ε-approximate solution to (2) if

$$-A \circ \ln X^* \leqslant r^* + \varepsilon. \tag{10}$$

Recalling that $\frac{1}{mn}I \in \mathcal{D}_{m,n}$, the basic approach to find X^* and approximate the true value r^* may be summarized as follows:

(1) **If** $A \in \mathcal{D}_{m,n}$ **then** return $X^* = A, \underline{r} = \bar{r} = 0$ (recall Fact II.4) and **stop**, else

(2) Set the desired accuracy $\varepsilon > 0$, initialize $N = 0, X^* = X^{(0)} = \frac{1}{m+n}I, \underline{r} = -A \circ \ln A, \bar{r} = -A \circ \ln X^*$,

(3) Solve the positive semidefinite optimization problem corresponding to:

$$\min_{t,X}\{t : (t, X) \text{ satisfy } (4),(6),(7),(8),(9) \text{ and } t \geqslant \underline{r}\}, \tag{11}$$

and store its optimal solution pair $(\underline{t}, \underline{X})$,

(4) Update $\underline{r} = \underline{t}$,

(5) **If** $\bar{r} - \underline{r} \leqslant \varepsilon$ **then** return $X^*, \underline{r}, \bar{r}$ and **stop**, **else** set $X^{(N+1)}$ to be the next $\underline{X}$-*refinement point* for $\text{epi}(-A \circ \ln X)|_{\text{int } \mathcal{D}_{m,n}}$ and increment $N = N + 1$, and **if** $-A \circ \ln X^{(N)} \leqslant -A \circ \ln X^*$ **then** update $X^* = X^{(N)}, \bar{r} = -A \circ \ln X^*$, **go to (2)**.

To justify our termination criteria in step 4, observe that since the approximation to (convex) $\text{epi}(-A \circ \ln X)|_{\text{int } \mathcal{D}_{m,n}}$ is based on supporting hyperplanes, the solution to (11) always corresponds to a lower bound on the optimal value of $-A \circ \ln X$ in (2). Note that $\underline{r}$ and $\bar{r}$ are nondecreasing and nonincreasing, respectively, at each iteration of our scheme.

It is left to explain how we construct $\underline{X}$-refinement points.

D. X-refinement point for epi$(-A \circ \ln X)|_{\text{int } \mathcal{D}_{m,n}}$

Given a sequence of points $X^{(0)}, \ldots, X^{(N)}$ in int $\mathcal{D}_{m,n}$, a point $\underline{X}$ solving (11), and some fixed reference point $Z \in$ int $\mathcal{D}_{m,n}$, we generate $X^{(N+1)}$ simply by minimizing $-A \circ \ln X$ on the linear segment $[Z, \underline{X}]$; that is,

$$X^{(N+1)} = \arg\min_X \{-A \circ \ln X :$$
$$X = \alpha Z + (1-\alpha)\underline{X}, \alpha \in [0,1]\}. \quad (12)$$

In fact, we are interested in finding the approximate minimizer in (12) with a given precision, so this equality should be interpreted in an approximate sense. Given the tolerance $\delta > 0$, the one-dimensional minimization may be efficiently performed using, say, the standard derivative-based bisection scheme as follows:

(i) Set the end points $X_{\text{start}} = Z, X_{\text{end}} = \underline{X}$ of the interval containing the approximate minimizer of (12) $X_{\min} \ni [X_{\text{start}}, X_{\text{end}}]$, set the initial value of $X_{\min} = \frac{X_{\text{start}} + X_{\text{end}}}{2}$,

(ii) While $\|X_{\text{end}} - X_{\text{start}}\| \geqslant \delta$ repeat
    if the directional derivative along $X_{\text{end}} - X_{\text{start}}$ of $-A \circ \ln X$ at $X_{\min}$ is negative set $X_{\text{start}} = X_{\min}$ else set $X_{\text{end}} = X_{\min}$, and set $X_{\min} = \frac{X_{\text{start}} + X_{\text{end}}}{2}$.

This scheme quickly terminates with an interval $[X_{\text{start}}, X_{\text{end}}] \ni X_{\min}$ such that $\|X_{\text{end}} - X_{\text{start}}\| < \delta$; note that the directional derivative of $-A \circ \ln X$ may be easily evaluated using (3). We set $X^{(N+1)} = X_{\min}$.

Particularly, we experiment with two choices of the reference point Z:

(A) (Conservative strategy) $Z = \frac{1}{mn}I$,

(B) (Aggressive strategy) $Z = X^*$, where X^* corresponds to the best upper bound $\bar{r}$ on r^* obtained so far.

Recall that generally in the absence of degeneracies the solution to (11) lies in the boundary of a semidefinite cone. Therefore, intuitively, choice A may encourage the next refinement point $X^{(N+1)}$ to be more centered, while choice B would be a more aggressive choice that in a sense targets "minimizing the discrepancy" between the lower and upper bounds $\underline{r}, \bar{r}$, and so we would generally expect choice B to result in a smaller N being needed, i.e., a fewer number of cutting planes added. Indeed, our intuition seems to be confirmed by numerics; see Sec. IV.

Despite its seeming simplicity, our choice of $X^{(N+1)}$ proves to be very effective numerically in refining the objective epigraph epi$(-A \circ \ln X)|_{\text{int } \mathcal{D}_{m,n}}$ near the region of interest, i.e., close to an ε-approximate solution to (2); see Sec. V for some illustrations.

Next we comment on a few possible reasons for this surprising efficiency.

Remark III.1. We solve (11) only approximately since, in general, at present this is the only option for handling positive semidefinite optimization problems efficiently, i.e., in polynomial time. However, this does not seem to negatively affect our approach to (2) because practically the optimization problem (11) is solved to a very high (near-machine) precision.

Remark III.2. It is only for notational convenience that beginning in Sec. III A we restrict $X^{(i)}$ to belong to the interior of $\mathcal{D}_{m,n}$ rather than the whole set itself. Observe that in general $-A \circ \ln X$ may approach infinity when X approaches the boundary of $\mathcal{D}_{m,n}$. Since we want to use (4) in forming

(11) and subsequently solve the problem numerically, such a situation is clearly undesirable. In practice, to overcome such a tendency for $X^{(i)}$ we implement a simple numerical safeguard *recentering* procedure, briefly outlined at the end of this section.

In principle, one can prove the convergence of a minimization algorithm if it is possible to show that the objective function gets reduced by some nontrivial amount at each iteration of the algorithm. If we have a smooth objective function with no constraints, one natural choice is to follow the steepest descent direction—the direction of the negative gradient—at each iteration, until a significant decrease in the objective is achieved. However, when constraints are present, such as in the case of (2), following the steepest descent direction might not work well because the direction might send us back to the boundary of the feasible region over and over again, thus hindering the progress of our algorithm. Therefore, intuitively, we want to make long enough steps that guarantee a sufficient decrease of the objective at each iteration, and, at the same time, stay away from the boundary of $\mathcal{D}_{m,n}$ whenever possible.

The first goal is accomplished by considering (11) instead of, say, merely minimizing $-A \circ \ln X$ along the steepest descent direction. Potentially, this allows us to traverse the whole feasible region $\mathcal{D}_{m,n}$ in search of a better next iterate. In other words, solving (11) rather than minimizing the objective function along the gradient hopefully prevents the algorithm from being stuck near the boundary of $\mathcal{D}_{m,n}$ and allows for long enough steps. Recall that to guarantee that we actually decrease the objective at each iteration rather than overshoot toward a point with an even larger objective value as compared to the one we started with, we combine solving (11) with a backtracking line search (12) along $[Z, \underline{X}]$.

The second goal is in part accomplished by relying on the maximal-rank property of the path-following interior-point methods used to solve (11). Here, interestingly, a seemingly trivial refinement in (4) that replaces unrestricted t with $t \geqslant -A \circ \ln A$ makes a dramatic difference in the numerics. That is, although $t \geqslant -A \circ \ln A$ probably does not have such a huge impact on refining the epigraph itself, it certainly helps to produce a well-centered solution X^* to (11) by finding a full-rank solution X^*—hopefully far away from the boundary of $\mathcal{D}_{m,n}$—if such exists at least during the very first iterations of our scheme.

Furthermore, if the addition of the next point $X^{(N+1)}$ to the approximate description of the epi$(-A \circ \ln X)|_{\text{int } \mathcal{D}_{m,n}}$ according to (4) causes numerical problems for solving (11), we augment $X^{(N+1)}$ (as many times as necessary) by re-centering it toward $\frac{1}{mn}I$, that is, by setting $X^{(N+1)} = \beta X^{(N+1)} + (1-\beta)\frac{1}{mn}I$, where $\beta \in (0,1)$ is some preset constant. In our experiments we use the default value of $\beta = 0.5$. The rationale for the latter is that $\frac{1}{mn}I$ is a well-centered point in $\mathcal{D}_{m,n}$ and hence does not cause any numerical problems in (4).

IV. COMPUTATIONAL RESULTS

Our numerical approach was implemented in MATLAB, using SeDuMi solver version 1.1R3 and tested with MATLAB version R2007a. We tested our algorithm on an Intel Xeon 2.4 GHz machine with 2 Gb RAM.

We ran our numerical scheme on a set of 100 randomly generated matrices A: 50 instances corresponding to $m = 2, n = 2$ with targeted default solution precision of $\varepsilon = 10^{-3}$, and 50 instances corresponding to $m = 2, n = 3$ with targeted default solution precision of $\varepsilon = 10^{-2}$, setting $\delta = 10^{-10}$. All the runs were successful.

Comparing strategies A and B, the latter aggressive approach appears to produce superior numerical results. For example, on the 50 problems corresponding to $m = 2, n = 3$, the average solution times per problem instance were decreased almost 5-fold, with an average of 65.14 outer iterations per problem in case A resulting in slightly over 5-s run time per instance, compared with only average 13.8 outer iterations and under 1-s run time per problem while using case B.

For the randomly generated problems, our choice of the termination precision level ε was stipulated by the following performance metric that may be used to judge the quality of the approximation. For a given problem instance, one may introduce *relative gap,* defined as

$$g_r = \frac{\bar{r} - \underline{r}}{r^*}.$$

Note that since $A \circ \ln A$ is negative because of the eigenvalues of A being nonnegative and adding up to 1, and so $\underline{r} > 0$, we can bound

$$g_r \leqslant \frac{\bar{r} - \underline{r}}{\underline{r}}.$$

So, for example, despite a seemingly high value of $\varepsilon = 10^{-2}$ in the case of $m = 2, n = 3$, most of the problems were terminated with $g_r < 10\%$.

For a more detailed illustration, we report on the following three examples, where, in addition, one may compute the relative entropy

$$A \circ \ln A - \inf_X \{-A \circ \ln X : X \in \mathcal{D}_{m,n}\}$$

analytically:

(i) 4×4 simple rank-1 (pure) state, here $m = n = 2$ and

$$A = \begin{pmatrix} 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 \end{pmatrix}.$$

An ε-approximate solution is found in two iterations with the reported approximate relative entropy value $r = 0.6931 \cdots$ and $|\ln 2 - r| \approx 2.8227 \times 10^{-11}$ computed with MATLAB (where $\ln 2$ is the true analytic value) and

$$X^* \approx \begin{pmatrix} 1/3 & 0 & 0 & 1/6 \\ 0 & 1/6 & 0 & 0 \\ 0 & 0 & 1/6 & 0 \\ 1/6 & 0 & 0 & 1/3 \end{pmatrix},$$

with $\approx$ meaning that the actual computed values in X^* differ from the values by no more than 10^{-7}, relying on strategy A; in case B the results are similar with a solution being found in two iterations.

(ii) 4×4 simple rank-2 state, here $m = n = 2$ and

$$A = \begin{pmatrix} 1/2 & 0 & 0 & 1/4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1/4 & 0 & 0 & 1/2 \end{pmatrix}.$$

An ε-approximate solution is found in 62 iterations with the reported approximate relative entropy value $r = 0.1308 \cdots$ and $|\ln 2 - (1/4 \ln 4 + 3/4 \ln 4/3) - r| \approx -2.2268 \times 10^{-8}$ computed with MATLAB [where $\ln 2 - (1/4 \ln 4 + 3/4 \ln 4/3)$ is the true analytic value) and

$$X^* \approx \begin{pmatrix} 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 \end{pmatrix},$$

with $\approx$ meaning that the actual computed values in X^* differ from the values by no more than 10^{-5}, relying on strategy A; in case B the results are similar with a solution being found in five iterations.

(iii) 4×4 full-rank (entangled) Werner state, here $m = n = 2$ and

$$A = \frac{1}{4} \begin{pmatrix} 1 - f & 0 & 0 & 0 \\ 0 & f + 1 & -2f & 0 \\ 0 & -2f & f + 1 & 0 \\ 0 & 0 & 0 & 1 - f \end{pmatrix}, f = 2/3.$$

An ε-approximate solution is found in 93 iterations with the reported approximate relative entropy value $r = 0.1308 \cdots$ and $|3/4 \ln 3/2 + 1/4 \ln 1/2 - r| \approx 2.4489 \times 10^{-8}$ computed with MATLAB (where $3/4 \ln 3/2 + 1/4 \ln 1/2$ is the true analytic value) and

$$X^* \approx \begin{pmatrix} 1/6 & 0 & 0 & 0 \\ 0 & 1/3 & -1/6 & 0 \\ 0 & -1/6 & 1/3 & 0 \\ 0 & 0 & 0 & 1/6 \end{pmatrix},$$

with $\approx$ meaning that the actual computed values in X^* differ from these values by no more than 6×10^{-5}, relying on strategy A; in case B the results are similar with a solution being found in 25 iterations.

For these examples, in an attempt to make an even more convincing case, we set the targeted precision to $\varepsilon = 10^{-7}$ (with the corresponding $\delta = 10^{-12}$).

Reference [12] contains the MATLAB driver routine that calls SeDuMi solver as its internal subroutine; note that functions vec and mat are internal to SeDuMi but are trivial to reimplement as stand-alone routines. For more details on the implementation of SeDuMi see Ref. [21] and the references therein. If needed, one may replace SeDuMi by some other appropriate solver such as SDPT3 or CSDP. The driver routine is written with the primary goal of having the code transparent and readable, rather than achieving maximal computational efficiency. For example, although the use of MATLAB's cells allows for greater transparency, it is definitely not the most efficient and fastest approach from a computational point of view.

V. CONCLUSION

We propose a numerical algorithm for computing the REE $E_R(\sigma)$ in small dimensions for general σ. Our numerical experiments support the viability of our approach.

The efficiency of our numerical scheme may potentially be improved by considering the second-order-based approximation to $\mathrm{epi}(-A \circ \ln X)|_{\mathrm{int}\ \mathcal{D}_{m,n}}$ and consequently relying on the so-called convex quadratic positive semidefinite optimization problems instead of (11), which are still amenable to the interior-point methods. We do not pursue this route since we are quite satisfied with the performance of our numerical approach so far.

Optimistically, we hope that one may actually establish a (polynomial) complexity bound on finding an ε-approximate solution to (2) by following a variant of the proposed scheme.

Also, we hypothesize that in turn ε gives rise to a bound on the proximity to the true solution of (2). Presently, both are under further investigation.

Note that our approach is also applicable in a straightforward fashion to computing the lower bound on (2) where $\mathcal{D}_{m,n}$ is replaced by positive partial transpose states. We experimented with larger matrix sizes going up to $m = 5$, $n = 10$ and still were able to get the 10^{-2} approximate solutions, although the actual computations took much longer. If desired, in general in low dimensions we could achieve far greater numerical precision than 10^{-2} as illustrated in Sec. IV.

[1] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Rev. Mod. Phys. **81**, 865 (2009).

[2] M. B. Plenio and S. Virmani, Quantum Inf. Comput. **7**, 1 (2007).

[3] M. Horodecki, J. Oppenheim, and R. Horodecki, Phys. Rev. Lett. **89**, 240403 (2002).

[4] F. G. S. L. Brandao and M. B. Plenio, Nature Phys. **4**, 873 (2008).

[5] V. Vedral and M. B. Plenio, Phys. Rev. A **57**, 1619 (1998).

[6] J. Eisert, e-print arXiv:quant-ph/0504166v1.

[7] A. Miranowicz and S. Ishizaka, Phys. Rev. A **78**, 032310 (2008).

[8] S. Friedland and G. Gour, e-print arXiv:1007.4544 [quant-ph].

[9] A. C. Doherty, P. A. Parrilo, and F. M. Spedalieri, Phys. Rev. Lett. **88**, 187904 (2002)

[10] F. G. S. L. Brandao and R. O. Vianna, Phys. Rev. Lett. **93**, 220503 (2004).

[11] J. Eisert, P. Hyllus, O. Gühne, and M. Curty, Phys. Rev. A **70**, 062317 (2004).

[12] [http://math.ucalgary.ca/~yzinchen/relEntropy.m].

[13] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, UK, 2004).

[14] R. Bhatia, *Matrix Analysis* (Springer, New York, 1997).

[15] L. Gurvits, in *Proceedings of the 35th ACM Symposium on Theory of Computing* (ACM Press, New York, 2003).

[16] S. Friedland, On the graph isomorphism problem, e-print arXiv:0801.0398.

[17] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Lett. A **223**, 1 (1996).

[18] J. Renegar, *A Mathematical View of Interior-Point Methods in Convex Optimization*, MPS-SIAM Series on Optimization (SIAM, Philadelphia, 2001).

[19] H. Wolkowicz, R. Saigal, and L. Vandenberghe, *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications* (Springer, New York, 2000).

[20] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis* (Cambridge University Press, Cambridge, UK, 1999).

[21] J. Sturm, Opt. Methods Soft. **11**, 625 (1999).