Low Rank Approximation of Entangled Bipartite Systems

Matthew M. Lin

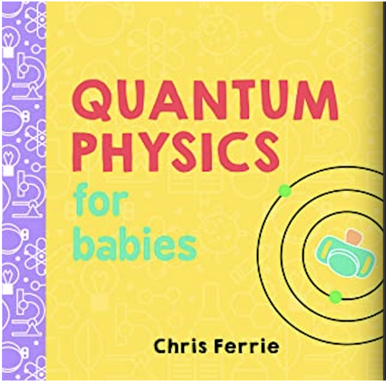
Department of Mathematics
National Cheng Kung University, Taiwan

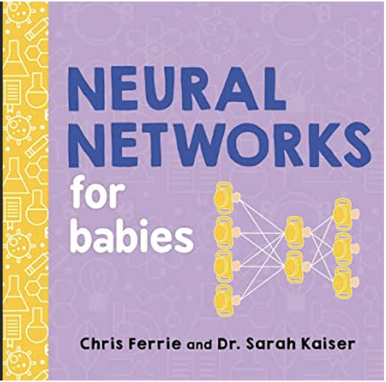
Jan. 17, 2022

Co-author

Prof. Moody T. Chu North Carolina State University

Attention is *not* all you need





Outline

Background
- Preliminaries
- Problem description

Algorithms
- Rank-1 approximation
- Quantum low-rank separability approximation
- Numerial Experiments

Conclusion

Density matrix

- Each quantum mechanical system is associated with a complex Hilbert space $\mathscr{H}$.
- Any unit vector $|\mathbf{x}\rangle \in \mathscr{H}$ is referred to as a pure state.
- Let $|\mathbf{x}\rangle \langle \mathbf{x}|\mathbf{z}\rangle$ be the orthogonal projection of any $|\mathbf{z}\rangle \in \mathscr{H}$ onto a given pure state $|\mathbf{x}\rangle$.
- A mixed state is a probabilistic mixture of finitely many pure states:

$$((|\mathbf{x}_1\rangle, \mu_1), (|\mathbf{x}_2\rangle, \mu_2), \ldots, (|\mathbf{x}_d\rangle, \mu_d))$$

- The density matrix ρ associated with such a mixed state is

$$\rho := \sum_i \mu_i |\mathbf{x}_i\rangle \langle \mathbf{x}_i|; \quad \sum_i \mu_i = 1; \quad \mu_i \geq 0,$$

  - The density matrix ρ is a positive semi-definite operator with unit trace.

Bipartite system

- Given two Hilbert spaces $\mathscr{H}_1$ and $\mathscr{H}_2$, the tensor product space is defined to be the set

$$\mathscr{H}_1 \otimes \mathscr{H}_2 := \{\sum_{s,t} \mathbf{u}_s \otimes \mathbf{v}_t | \mathbf{u}_s \in \mathscr{H}_1, \mathbf{v}_t \in \mathscr{H}_2\},$$

  - The only property required of $\otimes$ is its bi-linearity.
- An inner product can be induced via the relationship

$$\langle \mathbf{x} \otimes \mathbf{y} | \mathbf{z} \otimes \mathbf{w} \rangle := \langle \mathbf{x} | \mathbf{z} \rangle \langle \mathbf{y} | \mathbf{w} \rangle .$$

- We call $\mathscr{H}_1 \otimes \mathscr{H}_2$ the state space of a bipartite system.

Finite Dimensional Quantum Mechanical Systems

- Suppose $\mathscr{H}_1$ and $\mathscr{H}_2$ are finite dimensional with orthonormal basis states $\{\mathbf{e}_i\}_{i=1}^m$ and $\{\mathbf{f}_j\}_{j=1}^n$, respectively. Then
  1. $\{\mathbf{e}_i \otimes \mathbf{f}_j\}$ is a natural orthonormal basis for $\mathscr{H}_1 \otimes \mathscr{H}_2$.
  2. Elements in $\mathscr{H}_1$ and $\mathscr{H}_2$ can be interpreted as column vectors $\mathbf{x} \in \mathbb{C}^m$ and $\mathbf{y} \in \mathbb{C}^n$, respectively.
  3. The action $\mathbf{x} \otimes \mathbf{y}$ is equivalent to $\mathbf{x}\mathbf{y}^\top$ (i.e, outer product or tensor product $\mathbf{x} \circ \mathbf{y}$).
  4. An element in $\mathscr{H}_1 \otimes \mathscr{H}_2$ can be represented by a matrix in $\mathbb{C}^{m\times n}$, or, simply, a column vector in $\mathbb{C}^{mn}$.
- $|C\rangle$ in $\mathscr{H}_1 \otimes \mathscr{H}_2$ is a pure state if its matrix representation $C \in \mathbb{C}^{m\times n}$ has unit Frobenius norm.
- A density matrix ρ over $\mathscr{H}_1 \otimes \mathscr{H}_2$ should be of the form

$$\rho = \sum_i \mu_i \left|C_i\right\rangle \left\langle C_i\right|; \quad \sum_i \mu_i = 1; \quad \mu_i \geq 0,$$

where each $|C_i\rangle$ represents a pure state in $\mathscr{H}_1 \otimes \mathscr{H}_2$.

Example

- Consider $\mathscr{H}_i = \mathbb{C}^2$, $i = 1, 2$, with the standard basis denoted by $|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $|1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.
- In the quantum formalism, a tensor product $|\uparrow\rangle \otimes |\downarrow\rangle$ is often abbreviated as $|\uparrow\downarrow\rangle$.
- A natural basis for the tensor product space $\mathbb{C}^2 \otimes \mathbb{C}^2$ is:

$$\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$$

whose corresponding matrix representations are:

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

respectively.

Bell States

- In quantum information science, however, a more commonly used basis is the Bell states

$$\begin{cases} |\Phi^+\rangle & := & \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), \\ |\Phi^-\rangle & := & \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle), \\ |\Psi^+\rangle & := & \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle), \\ |\Psi^-\rangle & := & \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle). \end{cases}$$

- The Bell states form an orthonormal basis with the matrix representations given by

$$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},$$

Density matrices of Bell States

- The corresponding density matrices $\rho_{|\Phi^+\rangle} = |\Phi^+\rangle\langle\Phi^+|$ and so on should be expressed respectively as

$$\frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \quad \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix},$$

$$\frac{1}{2}\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \frac{1}{2}\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Entanglement

- If a pure state $|\psi\rangle \in \mathscr{H}_1 \otimes \mathscr{H}_2$ can be expressed as

$$|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle\,, \tag{1}$$

  where $|\psi_i\rangle \in \mathscr{H}_i$, $i = 1, 2$, are pure states, respectively, then we say that the pure state $|\psi\rangle$ is separable; otherwise, it is said to be entangled.
- The Bell states are entangled.

Schmidt decomposition

Lemma (Schmidt decomposition)

Any pure state $|\psi\rangle \in \mathscr{H}_1 \otimes \mathscr{H}_2$ *can be written in the form*

$$|\psi\rangle = \sum_j \sigma_j \left|\mathbf{u}_j\right\rangle \otimes \left|\mathbf{v}_j\right\rangle$$

where $\left|\mathbf{u}_j\right\rangle \in \mathscr{H}_1$ *and* $\left|\mathbf{v}_j\right\rangle \in \mathscr{H}_2$ *are orthonormal vectors,* $\sigma_j \geq 0$ *and* $\sum_j \sigma_j^2 = 1$.

Separable density matrix

- A more intriguing question is to determine whether a given density matrix ρ over $\mathscr{H}_1 \otimes \mathscr{H}_2$ can be decomposed as

$$\rho = \sum_k \eta_k \mathcal{D}_k^{(1)} \otimes \mathcal{D}_k^{(2)}, \quad \sum_k \eta_k = 1, \quad \eta_k \geq 0.$$

  - $\{\mathcal{D}_k^{(1)}\}$ and $\{\mathcal{D}_k^{(2)}\}$ are density matrices in $\mathscr{H}_1$ and $\mathscr{H}_2$.
- We call a density matrix ρ over the bipartite space is separable if and only if

$$\rho = \sum_\ell \theta_\ell (|\mathbf{x}_\ell\rangle \langle \mathbf{x}_\ell|) \otimes (|\mathbf{y}_\ell\rangle \langle \mathbf{y}_\ell|).$$

  - $\mathbf{x}_\ell \in \mathscr{H}_1$ and $\mathbf{y}_\ell \in \mathscr{H}_2$ are unit vectors.
  - $\theta_\ell \geq 0$ and $\sum_\ell \theta_\ell = 1$.

Lemma (Chen, Wu 2003)

Given a density matrix $\rho \in \mathbb{C}^{mn \times mn}$, let $\mathscr{R}(\rho) \in \mathbb{C}^{m^2 \times n^2}$ denote the $\mathscr{R}$-folding[1] of ρ. If ρ is separable, then necessarily the Ky Fan norm , i.e., the sum of all singular values of $\mathscr{R}(\rho)$, is less than 1.

- The Bell state Φ^+ is entangled and even more its density matrix $\rho_{|\Phi^+\rangle}$ is entangled since the $\mathscr{R}$-folding of the density matrix $\rho_{|\Phi^+\rangle}$ is $\frac{1}{2}I_4$ whose Ky Fan norm is 1.
- Similar arguments can be applied to show that none of $\rho_{|\Phi^-\rangle}$, $\rho_{|\Psi^+\rangle}$, and $\rho_{|\Psi^-\rangle}$.

[1] Also defined in our later discussion.

Approximation

- If ρ is not separable, then seeking its nearest separable approximation is a problem of practical importance.
- Different operational paradigms have been proposed:
  1. The trace metric

$$D_T(\rho,\sigma) := \frac{1}{2}\mathrm{Tr}\sqrt{(\rho-\sigma)^2},$$

  2. The Bures distance

$$D_B(\rho,\sigma) := \sqrt{2 - 2\mathrm{Tr}\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}},$$

  3. The Frobenius norm

$$D_F(\rho,\sigma) = \frac{1}{2}\|\rho-\sigma\|_F = \frac{1}{2}\sqrt{\mathrm{Tr}(\rho-\sigma)^2}.$$

Entangled Bipartite Quantum Systems

Problem

Given a positive definite (PD) matrix $\rho \in \mathbb{C}^{mn \times mn}$ with unit trace, find its approximation in the form

$$\min_{\substack{\lambda_r \geq 0, \sum_{r=1}^{R} \lambda_r = 1, \mathbf{a}_r \in \mathbb{C}^m, \mathbf{b}_r \in \mathbb{C}^n \\ \|\mathbf{a}_r\|=1, \|\mathbf{b}_r\|=1}} \|\rho - \sum_{r=1}^{R} \lambda_r (\mathbf{a}_r \mathbf{a}_r^*) \otimes (\mathbf{b}_r \mathbf{b}_r^*)\|_F^2, \qquad (2)$$

where * denotes the conjugate transpose.

Difficulities

- Deciding whether a density matrix is entangled or not is an NP hard problem.
- In our case, our formulation is not for the task of "deciding" whether a given mixed state is entangled or not.
- Instead, per given density matrix ρ and a fixed rank R, we look for a local separable approximation.
  1. The Cauchy–Riemann equations do not hold.
  2. Approximation over real field is not realistic:

$$\mathbf{x} \otimes \mathbf{y} = (\mathbf{u} \otimes \mathbf{p} - \mathbf{v} \otimes \mathbf{q}) + \imath(\mathbf{v} \otimes \mathbf{p} + \mathbf{u} \otimes \mathbf{q})$$

  if $\mathbf{x} = \mathbf{u} + \imath\mathbf{v}$ and $\mathbf{y} = \mathbf{p} + \imath\mathbf{q}$.

Rank-1 Approximation of Entangled Bipartite Systems

Example

Given a fixed positive semi-definite matrix A in $\mathbb{C}^{mn \times mn}$, consider

$$\min_{\substack{\lambda \in \mathbb{R}_+, \mathbf{x} \in \mathbb{C}^m, \mathbf{y} \in \mathbb{C}^n \\ \|\mathbf{x}\|=1, \|\mathbf{y}\|=1}} \|A - \lambda(\mathbf{x}\mathbf{x}^*) \otimes (\mathbf{y}\mathbf{y}^*)\|_F^2. \tag{3}$$

We can think of (3) as a special case of (2) with $R = 1$

- The minimization above is equivalent to maximizing the absolute value of

$$\lambda(\mathbf{x}, \mathbf{y}) := \langle A, (\mathbf{x} \otimes \mathbf{y})(\mathbf{x} \otimes \mathbf{y})^* \rangle$$

subject to the constraints that **x** and **y** are of unit lengths.

Related rank-1 tensor approximation

- This approximation can be recast as a special type of rank-1 approximation with “shared” factors:

$$\min_{\substack{\lambda\in\mathbb{R}_+,\mathbf{x}\in\mathbb{R}^m,\mathbf{y}\in\mathbb{R}^n\\ \|\mathbf{x}\|=1,\|\mathbf{y}\|=1}} \|\mathfrak{A} - \lambda\,\mathbf{x}\circ\mathbf{x}\circ\mathbf{y}\circ\mathbf{y}\|_F^2,$$

  where $\circ$ denotes the outer product and $\mathfrak{A} \in \mathbb{R}^{m\times m\times n\times n}$ is a special refolding of the original $A \in \mathbb{R}^{mn\times mn}$ into an order-4 tensor.
    - This specially structured problem can be handled by some conventional techniques, say, the Tensorlab toolbox.
- To this, we propose two new rank-1 approximation methods which are easily constructed and have higher efficiency when comparing with some state-of-the-art optimization techniques.
- These methods could be served as a first step toward a more general problem.

Wirtinger calculus

- Let $f : \mathbb{C} \to \mathbb{R}$ be a given real-valued function over a complex variable $z = x + \imath y$ such that $f(z) = u(x, y)$.
  - The Wirtinger derivatives are defined by

$$\begin{cases} \frac{\partial f}{\partial z} & := & \frac{1}{2}\big(\frac{\partial u}{\partial x} - \imath \frac{\partial u}{\partial y}\big), \\ \frac{\partial f}{\partial \overline{z}} & := & \frac{1}{2}\big(\frac{\partial u}{\partial x} + \imath \frac{\partial u}{\partial y}\big), \end{cases}$$

  - In other words, the two symbols z and $\overline{z}$ are formally regarded as independent with respect to each other.

Gradient information

Lemma

If $f : \mathbb{C}^n \to \mathbb{R}$ *is regarded as* $f(\mathbf{z}) = f(\mathbf{u}, \mathbf{v})$ *for* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$*, where* $\mathbf{z} = \mathbf{u} + \imath \mathbf{v} \in \mathbb{C}^n$*. Then the "true" gradient of* f *is given by*

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial \mathbf{u}} \\ \frac{\partial f}{\partial \mathbf{v}} \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial \mathbf{z}} + \frac{\partial f}{\partial \overline{\mathbf{z}}} \\ \imath\left(\frac{\partial f}{\partial \mathbf{z}} - \frac{\partial f}{\partial \overline{\mathbf{z}}}\right) \end{bmatrix}.$$

Block matrix A

- Consider an $m \times m$ block matrix A with blocks $A_{ij} \in \mathbb{R}^{n \times n}$,

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1,m} \\ A_{21} & A_{22} & \cdots & A_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m,1} & A_{m,2} & \cdots & A_{m,m} \end{bmatrix} \in \mathbb{C}^{mn \times mn}.$$

- Associated with A, we define the so called $\mathscr{R}$-folding:

$$\mathscr{R}(A) := \begin{bmatrix} \mathbf{vec}(A_{1,1})^\top \\ \mathbf{vec}(A_{2,1})^\top \\ \vdots \\ \mathbf{vec}(A_{m,m})^\top \end{bmatrix} \in \mathbb{C}^{m^2 \times n^2},$$

where **vec** denotes the conventional vectorization of a matrix by its columns.

Calculation of λ: Way 1

- Observe that

$$\begin{aligned}\lambda(\mathbf{x},\mathbf{y}) &= \langle A, (\bar{\mathbf{x}}\otimes\bar{\mathbf{y}})(\mathbf{x}\otimes\mathbf{y})^\top\rangle_{\mathbb{R}} \\ &= \langle \mathscr{A}(\mathbf{y},\bar{\mathbf{y}})\mathbf{x},\bar{\mathbf{x}}\rangle_{\mathbb{R}} = \langle \mathscr{B}(\mathbf{x},\bar{\mathbf{x}})\mathbf{y},\bar{\mathbf{y}}\rangle_{\mathbb{R}},\end{aligned}$$

- where

$$\begin{aligned}\mathscr{A}(\mathbf{y},\bar{\mathbf{y}}) &:= \mathbf{reshape}(\mathscr{R}(A)(\mathbf{y}\otimes\bar{\mathbf{y}}),[m,m]), \\ \mathscr{B}(\mathbf{x},\bar{\mathbf{x}}) &:= \mathbf{reshape}(\mathscr{R}(A)^\top(\mathbf{x}\otimes\bar{\mathbf{x}}),[n,n]).\end{aligned}$$

First Order Optimality Condition

Lemma (FOC)

The first order optimality condition for maximizing $\lambda(\mathbf{x}, \mathbf{y})$ *is that*

$$\begin{cases} \mathscr{A}(\mathbf{y}, \bar{\mathbf{y}})\mathbf{x} & = & \lambda(\mathbf{x}, \mathbf{y})\mathbf{x}, \\ \mathscr{B}(\mathbf{x}, \bar{\mathbf{x}})\mathbf{y} & = & \lambda(\mathbf{x}, \mathbf{y})\mathbf{y}. \end{cases}$$

Power-like iterative scheme

- To obtain the (local) maximizer of $\lambda(\mathbf{x}, \mathbf{y})$, we start from an initial value $(\mathbf{x}^{[0]}, \mathbf{y}^{[0]})$ and repeat the following process:

$$\begin{cases} \mathbf{x}^{[p+1]} & := \dfrac{\mathscr{A}(\mathbf{y}^{[p]},\overline{\mathbf{y}^{[p]}})\mathbf{x}^{[p]}}{\|\mathscr{A}(\mathbf{y}^{[p]},\overline{\mathbf{y}^{[p]}})\mathbf{x}^{[p]}\|_2} \\ \mathbf{y}^{[p+1]} & := \dfrac{\mathscr{B}(\mathbf{x}^{[p+1]},\overline{\mathbf{x}^{[p]}})\mathbf{y}^{[p]}}{\|\mathscr{B}(\mathbf{x}^{[p+1]},\overline{\mathbf{x}^{[p]}})\mathbf{y}^{[p]}\|_2}, \end{cases} \quad p = 0, 1, 2, \ldots.$$

  - If the iteration ever converges, the fixed-point of this iteration satisfies precisely the first order optimality condition

First Order Optimality Condition

Let $\mathcal{C}(\mathbf{x}, \mathbf{y}) := \text{reshape}(A(\mathbf{x} \otimes \mathbf{y}), [n, m]) \in \mathbb{C}^{n \times m}$.

Lemma (FOC2)

A critical point must satisfies the relationship

$$\begin{cases} \mathcal{C}(\mathbf{x}, \mathbf{y})^\top \overline{\mathbf{y}} & = & (\mathbf{y}^\top \mathcal{C}(\mathbf{x}, \mathbf{y})\mathbf{x})\mathbf{x}, \\ \mathcal{C}(\mathbf{x}, \mathbf{y})\overline{\mathbf{x}} & = & (\mathbf{y}^\top \mathcal{C}(\mathbf{x}, \mathbf{y})\mathbf{x})\mathbf{y}. \end{cases}$$

That is, with respect to $\mathcal{C}(\mathbf{x}, \mathbf{y})$,

- $(\lambda, \mathbf{y}, \overline{\mathbf{x}})$ is the dominant singular triplets of $\mathcal{C}(\mathbf{x}, \mathbf{y})$.
- $\mathbf{y}$ is the dominant left singular vector.
- $\overline{\mathbf{x}}$ is the dominant right singular vector of $\mathcal{C}(\mathbf{x}, \mathbf{y})$.
- An SVD-like iteration can be seen in [Chu & Lin, 2021].

Gradient flow for quantum low-rank approximation

- For convenience, introduce the abbreviations

$$\begin{aligned} \Theta &= \Theta(\lambda_1, \ldots, \lambda_R, \mathbf{x}_1, \ldots, \mathbf{x}_R, \mathbf{y}_1, \ldots, \mathbf{y}_R) \\ &:= \rho - \sum_{r=1}^{R} \lambda_r (\mathbf{x}_r \otimes \mathbf{y}_r)(\mathbf{x}_r \otimes \mathbf{y}_r)^* \in \mathbb{C}^{mn \times mn}, \end{aligned}$$

and, for each $r \in [\![R]\!]$,

$$\begin{aligned} \omega_r &= \omega_r(\lambda_1, \ldots, \lambda_R, \mathbf{x}_1, \ldots, \mathbf{x}_R, \mathbf{y}_1, \ldots, \mathbf{y}_R) \\ &:= \langle \mathbf{x}_r \otimes \mathbf{y}_r, \Theta(\mathbf{x}_r \otimes \mathbf{y}_r) \rangle \in \mathbb{R}, \\ \mathscr{C}_r &= \mathscr{C}_r(\lambda_1, \ldots, \lambda_R, \mathbf{x}_1, \ldots, \mathbf{x}_R, \mathbf{y}_1, \ldots, \mathbf{y}_R) \\ &:= \mathbf{reshape}(\Theta(\mathbf{x}_r \otimes \mathbf{y}_r), n, m) \in \mathbb{C}^{n \times m}. \end{aligned}$$

Calculation of the gradient

Lemma

Suppose $\mathbf{x}_r = \mathbf{u}_r + \imath \mathbf{v}_r$ *and* $\mathbf{y}_r = \mathbf{p}_r + \imath \mathbf{q}_r$ *with* $\mathbf{u}_r, \mathbf{v}_r \in \mathbb{R}^m$ *and* $\mathbf{p}_r, \mathbf{q}_r \in \mathbb{R}^n$*. Let* $g := \langle \Theta, \Theta \rangle$ *be a function of the real variables* λ_r, $\mathbf{u}_r$, $\mathbf{v}_r$, $\mathbf{p}_r$, *and* $\mathbf{q}_r$, $r \in [\![R]\!]$*. Then the portions of the gradient* ∇g *with respect to the respective real variables are given by*

$$\begin{cases} \frac{\partial g}{\partial \lambda_r} = -2\omega_r, \\ \frac{\partial g}{\partial (\mathbf{u}_r, \mathbf{v}_r)} = -4\lambda_r \begin{bmatrix} \mathrm{Re}(\mathscr{C}_r^\top \bar{\mathbf{y}}_r) \\ \mathrm{Im}(\mathscr{C}_r^\top \bar{\mathbf{y}}_r) \end{bmatrix}, \qquad r \in [\![R]\!]. \\ \frac{\partial g}{\partial (\mathbf{p}_r, \mathbf{q}_r)} = -4\lambda_r \begin{bmatrix} \mathrm{Re}(\mathscr{C}_r \bar{\mathbf{x}}_r) \\ \mathrm{Im}(\mathscr{C}_r \bar{\mathbf{x}}_r) \end{bmatrix}. \end{cases}$$

Projected gradient

- Since our problem is constrained to the pure states, we need the projected gradient.
- The projection can be obtained by projecting the blocks of ∇g onto the corresponding unit spheres, S^{2m-1} and S^{2n-1}, respectively.

Lemma

The projected gradients of objective function g can be condensed into the expressions

$$\begin{cases} \mathrm{Proj}_{S^{2m-1}} \frac{\partial g}{\partial(\mathbf{u}_r, \mathbf{v}_r)} & = & -4\lambda_r(\mathscr{C}_r^\top \bar{\mathbf{y}}_r - \omega_r \mathbf{x}_r), \\ \mathrm{Proj}_{S^{2n-1}} \frac{\partial g}{\partial(\mathbf{u}_p, \mathbf{v}_q)} & = & -4\lambda_r(\mathscr{C}_r \bar{\mathbf{x}}_r - \omega_r \mathbf{y}_r), \end{cases} \qquad r \in [\![R]\!].$$

Projected gradient flow

- we now define the complex-valued differential system

$$\begin{cases} \frac{d\lambda_r}{dt} & = & 2\omega_r, \\ \frac{d\mathbf{x}_r}{dt} & = & 4\lambda_r(\mathscr{C}_r^\top \overline{\mathbf{y}}_r - \omega_r \mathbf{x}_r), \qquad r \in [\![R]\!], \\ \frac{d\mathbf{y}_r}{dt} & = & 4\lambda_r(\mathscr{C}_r \overline{\mathbf{x}}_r - \omega_r \mathbf{y}_r), \end{cases}$$

  where t stands for a dimensionless parameter of time.
- The gradient flow therefore converge globally to a singleton as its limit point.

Maintaining nonnegativity and rand reduction

1. **Event detection:** Use an event function to detect when any $\lambda_r(t)$, $r \in [\![R]\!]$ becomes zero during the integration.
2. **Rank deduction:** When the event $\lambda_r(\widehat{t}) = 0$ is detected for one particular value r and time $\widehat{t}$, the term

$$\lambda_r(\mathbf{x}_r \otimes \mathbf{y}_r)(\mathbf{x}_r \otimes \mathbf{y}_r)^*$$

contributes nothing to the objective value g at that instant.

- We drop this term entirely.
- The low rank R is decreased by 1.
- We build an algorithm that can dynamically lower the rank R when a certain component is not needed.

Maintain sum-to-one

- To satisfy the constraint $\sum_{r=1}^{R} \lambda_r(t) = 1$ for all $t \geq 0$, it is necessary to impose the consistency condition

$$\sum_{r=1}^{R} \frac{d\lambda_r(t)}{dt} = 0, \quad \text{for all } t \geq 0,$$

- We propose to remedy the situation by modifying the flow for $\lambda_r(t)$ to

$$\frac{d\lambda_r}{dt} = 2(\omega_r - \widetilde{\omega}), \quad r \in [\![R]\!],$$

  where $\widetilde{\omega} := \frac{\sum_{r=1}^{R} \omega_r}{R}$, while the original governing equations for $\frac{d\mathbf{x}_r}{dt}$ and $\frac{d\mathbf{y}_r}{dt}$, $r \in [\![R]\!]$ are kept invariant.
- The resulting system is no longer in the steepest descent direction. We have to show that a descent flow is kept.

Descent flow

Lemma

Let $Z(t)$ denote the newly defined flow

$$Z(t) := (\lambda_1(t), \ldots, \lambda_R(t), \mathbf{x}_1(t), \ldots, \mathbf{x}_R(t), \mathbf{y}_1(t), \ldots, \mathbf{y}_R(t)).$$

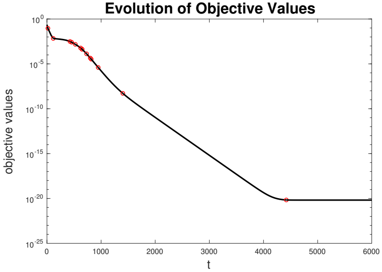
Then the objection value of g is descending along the trajectory $Z(t)$.

Example 1: Evolution of Objective Values

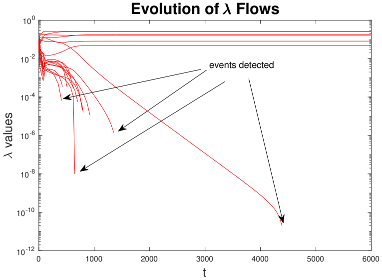
- Generate a test matrix

$$\rho = \sum_{r=1}^{6} \lambda_r (\mathbf{x}_r \mathbf{x}_r^*) \otimes (\mathbf{y}_r \mathbf{y}_r^*)$$

  - $\mathbf{x}_r, \mathbf{y}_r \in \mathbb{C}^5$: with randomly generated unit vectors
  - $\lambda_r > 0$, $r \in [\![6]\!]$, satisfying $\sum_{r=1}^{6} \lambda_r = 1$, as the target.
  - $\rho \in \mathbb{C}^{25 \times 25}$ is already separable in itself with rank 6.
  - Starting an experiment with $R = 20$ initially, we are interested in finding out whether ρ can be completely recovered by our method.



(a) Descending behavior of the objective value

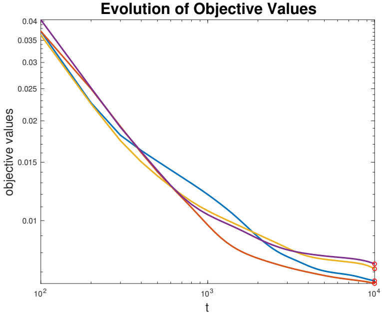


(b) Dynamics of $\lambda_r(t)$, $r \in [\![20]\!]$

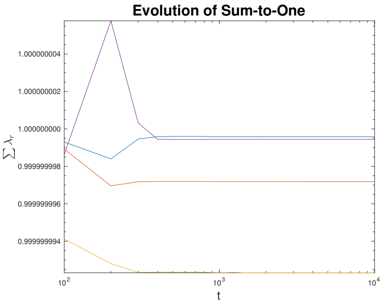
- Each circle indicates an event occurs.
- At the end of integration, the rank is indeed reduced to $R = 6$ and the objective value is nearly zero in this particular example.

Example 2: Sum-to-one

- $\rho \in \mathbb{C}^{40\times 40}$: a randomly generated symmetric and positive definite matrix.
- Search for unit vectors $\mathbf{x}_r \in \mathbb{C}^8$ and $\mathbf{y}_r \in \mathbb{C}^5$ with initial $R = 10$ and four sets of randomly generated starting values.
- This is a hard problem in that at $t = 10^4$ the flows have not reached convergence yet, but their descent property is clear. It is also likely they will converge to different optimal values.
- The property $\sum_{r=1}^{10} \lambda_r = 1$ is reasonably preserved within a fairly narrow window of approximately 10^{-8}. This confirms that our strategy for maintaining both sum-to-one and descent achieves its goal.



(c) Distinct trajectories lead to distinct objective values.



(d) Preservation of $\sum_{r=1}^{10} \lambda_r(t) = 1$.

1. We interpret the study of the the rank-1 approximation to entangled bipartite systems as a nonlinear eigenvalue problem as well as a nonlinear singular value problem.
2. Low rank approximation for entangled bipartite quantum systems is interesting because of its potential application as a way to certify the quality of an entanglement.
3. We describes a complex-valued gradient dynamics for the low rank approximation problem using the Wirtinger calculus.
4. Advantages:
   - Easy-to-program numerical schemes
   - Global convergence

Thank you very much !