Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 10, No. 1, pp. 137–149, January 2014

AN ALTERNATIVE STEEPEST DIRECTION METHOD FOR THE OPTIMIZATION IN EVALUATING GEOMETRIC DISCORD*

Yiju Wang, Liqun Qi, Shunlong Luo and Yi Xu

Abstract: We address the evaluation of geometric discord of bipartite quantum states arising from quantum information theory. The problem corresponds to finding the best approximation of the orthogonal decomposition of a partially Hermite fourth-order tensor. By discussing the optimality condition of the problem, we reduce it to a homogenous polynomial optimization problem on the product of two unitary matrices. Based on the Riemannian manifold and Lie group theory, we propose an alternative steepest direction method for the problem. Numerical experiments show the efficiency of the method.

Key words: *geometric discord, unitary matrix constraints, geodesic, steepest direction method*

Mathematics Subject Classification: *65F99, 65K10, 15A69, 14M15, 90C53*

---

1 Introduction

In quantum information theory, one encounters the following optimization problem:

$$\begin{array}{ll} \min \|\mathcal{A} - \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n} p_{ij}\mathbf{x}_i \circ \mathbf{y}_j \circ \mathbf{x}_i^* \circ \mathbf{y}_j^*\|_F^2 & \\ \text{s.t.} \quad \mathbf{x}_i^\dagger \mathbf{x}_{i'} = \delta_{ii'}, \quad i, i' = 1, 2, \ldots, m, & \\ \qquad \mathbf{y}_j^\dagger \mathbf{y}_{j'} = \delta_{jj'}, \quad j, j' = 1, 2, \ldots, n, & \\ \qquad \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n} p_{ij} = 1,\ p_{ij} \geq 0, \quad i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n, & \end{array} \tag{1.1}$$

where $\|\cdot\|_F$ denotes the Frobenius norm of a tensor, $\mathbf{x}_i^*(\mathbf{x}_i^\dagger)$ and $\mathbf{y}_j^*(\mathbf{y}_j^\dagger)$ respectively denote the conjugate (transposed conjugate) of complex vectors $\mathbf{x}_i \in C^m$ and $\mathbf{y}_j \in C^n$, $\mathbf{x}_i \circ \mathbf{y}_j \circ \mathbf{x}_i^* \circ \mathbf{y}_j^*$ denotes the out product of vectors $\mathbf{x}_i, \mathbf{y}_j, \mathbf{x}_i^*$ and $\mathbf{y}_j^*$, which is a fourth-order rank-1 tensor [6], $\mathcal{A}$ is a partially Hermite fourth-order tensor of dimension $m \times n \times m \times n$ satisfying

$$\mathcal{A}_{iji'j'} = \mathcal{A}^*_{i'j'ij}, \quad \forall\ i, i' = 1, 2, \ldots, m; \quad j, j' = 1, 2, \ldots, n. \tag{1.2}$$

*This work was supported by the Hong Kong Research Grant Council (PolyU 501808, 501909, 502510 and 502111), the Natural Science Foundation of China (11171180) and the Specialized Research Fund for the Doctoral Program of Higher Education of China (20113705110002).

Furthermore, $\sum_{i,j} \mathcal{A}_{ijij} = 1$ and $\mathcal{A}$ is positive semi-definite in the sense that

$$\mathcal{A}\times_1\mathbf{u}\times_2\mathbf{v}\times_3\mathbf{u}^*\times_4\mathbf{v}^* := \sum_{i,i'=1}^{m}\sum_{j,j'=1}^{n} \mathcal{A}_{iji'j'}u_i v_j u_{i'}^* v_{j'}^* \geq 0, \quad \forall\, \mathbf{u} = (u_i) \in C^m,\ \mathbf{v} = (v_j) \in C^n. \tag{1.3}$$

The problem (1.1) arises from the evaluation of the (symmetrized) geometric discord, which is a natural extension of the geometric measure of quantum discord [4, 17]. The latter in turn is motivated by the quantum discord and related measures of quantum correlations in quantum information theory [9, 16, 18, 20]. The geometric discord is a significant measure for quantum correlations, and has been regarded as a key resource for certain quantum communication tasks and quantum computational models without entanglement. In quantum mechanics, a state of a system is generally described by a density matrix (non-negative matrix on a complex Hilbert space with unit trace). For an mn-th dimensional bipartite density matrix (operator) ρ^{ab} on the tensor product space $C^m \otimes C^n$, the (symmetrized) geometric discord is defined as

$$Q(\rho^{ab}) = \min_{\Pi^a,\Pi^b} \|\rho^{ab} - \Pi^{ab}(\rho^{ab})\|_F^2,$$

which can be rewritten as

$$Q(\rho^{ab}) = \min_{p_{ij},\Pi_i^a,\Pi_j^b} \|\rho^{ab} - \sum_{i,j} p_{ij}\Pi_i^a \otimes \Pi_j^b\|_F^2$$

with $p_{ij} \geq 0$ and $\sum_{i,j} p_{ij} = 1$. Here

$$\Pi^{ab}(\rho^{ab}) = \sum_{i,j} \big(\Pi_i^a \otimes \Pi_j^b\big)\rho^{ab}\big(\Pi_i^a \otimes \Pi_j^b\big)$$

represents the post-measurement state after the von Neumann measurements $\Pi^a = \{\Pi_i^a\}_1^m$, $\Pi^b = \{\Pi_j^b\}_1^n$ on C^m and C^n, respectively, i.e., $\Pi_i^a = |k_i^a\rangle\langle k_i^a|$ with $\{|k_i^a\rangle\}_1^m$ being an orthonormal basis of C^m, and Π^b is defined similarly. Now any mn-th dimensional density matrix (operator) ρ^{ab} on tensor product space $C^m \otimes C^n$ can be expressed as

$$\rho^{ab} = \sum_{i,j} c_{ij} X_i \otimes Y_j,$$

where $c_{ij} = \mathrm{tr}\rho^{ab}(X_i \otimes Y_j)$, $\{X_i\}_1^{m^2}$ and $\{Y_j\}_1^{n^2}$ are respectively a set of Hermitian operators on C^m and C^n, which constitute orthonormal bases for the Hilbert-Schmidt spaces $\mathcal{L}(C^m)$ and $\mathcal{L}(C^n)$ of linear operators on C^m and C^n, respectively. The density matrix ρ^{ab}, as an mn-th dimensional Hermite matrix with unit trace, can be folded into a rectangular fourth-order $m \times n \times m \times n$ dimensional partially Hermite tensor $\mathcal{A}$ by setting [6]

$$\mathcal{A}_{iji'j'} = \rho_{st}^{ab}$$

with $\sum_{i,j} \mathcal{A}_{ijij} = 1$, where $i = \lceil \frac{s}{n} \rceil$, $j = \mathrm{mod}(s, n)$, $i' = \lceil \frac{t}{n} \rceil$, $j' = \mathrm{mod}(t, n)$. Based on this convention, the evaluation of the (symmetrized) geometric discord is formulated as problem (1.1).

As a constrained optimization problem [13, 23], the problem (1.1) is quite complicated since complex variables and unitary matrix constraints [1] are involved [1, 3, 24]. It cannot be handled efficiently by the classical gradient-type optimization methods such as Lagrangian multiplier method [19], as the generated iterates always depart from the unitary constraints and it is difficult to retract them to the feasible set [1].

To attack problem (1.1), we first reduce it to a homogenous polynomial optimization problem on the product of two unitary matrices by discussing its optimality condition, and then appeal to the Riemannian manifold theory [10, 11, 7] to design a numerical method to solve it. It is well known that a set of orthogonal constraints defines a Riemannian manifold [7], and the unitary constraint defines a Lie group which is a special Riemannian manifold [8, 12, 15, 22]. Based on this, we convert the reduced problem as an unconstrained problem on an appropriate differentiable manifold, and then use the geodesic curve searching strategy to design an alternative steepest direction method to solve the problem. In this strategy, the tangent direction of the geodesic plays the same role as that in the line search method in flat spaces [15]. Compared with the classical gradient-type optimization method such as SQP method and Lagrangian multiplier method [19], the new designed method can generate a critical point of concerned problem. The given numerical simulation shows that the method can generate good solutions.

The content of this paper is organized as follows. In Section 2, we establish a reduced version of the problem by discussing the optimality condition of the problem. In Section 3, we design an alternative steepest direction method for problem (1.1) based on the Riemannian manifold and Lie group theory. The numerical performance of the method is presented in Section 4.

In the end of this section, we present some notations used in this paper. For the fourth order partially Hermite tensor $\mathcal{A}$ in form (1.2) with dimension $m \times n \times m \times n$, and vectors $\mathbf{u} = (u_i) \in C^m, \mathbf{v} = (v_j) \in C^n$, define

$$\mathcal{A} \times_1 \mathbf{u} \times_3 \mathbf{u}^* := \Big(\sum_{i,i'} \mathcal{A}_{iji'j'} u_i u_{i'}^*\Big)_{n\times n}, \quad \mathcal{A} \times_2 \mathbf{v} \times_4 \mathbf{v}^* := \Big(\sum_{j,j'} \mathcal{A}_{iji'j'} v_j v_{j'}^*\Big)_{m\times m},$$

and denote

$$\mathcal{A} \times_1 \mathbf{u} \times_2 \mathbf{v} \times_3 \mathbf{u}^* \times_4 \mathbf{v}^* := \sum_{i',i'=1}^{m} \sum_{j,j'=1}^{n} \mathcal{A}_{iji'j'} u_i v_j u_{i'}^* v_{j'}^*$$

simply as $\mathcal{A}\mathbf{u}\mathbf{v}\mathbf{u}^*\mathbf{v}^*$.

For tensors $\mathcal{A}$ and $\mathcal{B}$ in form (1.2), their inner product is defined as

$$\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i,j,i',j'} \mathcal{A}_{iji'j'} \mathcal{B}_{iji'j'}^*,$$

and the F-norm of tensor $\mathcal{A}$ is defined as

$$\|\mathcal{A}\|_F = \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle} = \Big(\sum_{i,j,i',j'} \mathcal{A}_{iji'j'} \mathcal{A}_{iji'j'}^*\Big)^{1/2} = \Big(\sum_{i,j,i',j'} |\mathcal{A}_{iji'j'}|^2\Big)^{1/2}.$$

2 Reduction of the Problem

Before deriving a simplified version of the problem (1.1), we first consider the case that the optimal value of the objective function in problem (1.1) vanishes. In this case, tensor $\mathcal{A}$ has the following decomposition

$$\mathcal{A} = \sum_{i=1}^{m} \sum_{j=1}^{n} p_{ij} \mathbf{x}_i \circ \mathbf{y}_j \circ \mathbf{x}_i^* \circ \mathbf{y}_j^*. \tag{2.1}$$

Furthermore, for any i, j, each term $\mathbf{x}_i \circ \mathbf{y}_j \circ \mathbf{x}_i^* \circ \mathbf{y}_j^*$ is a rank-1 fourth-order partially Hermite tensor, and it is orthogonal to any other terms on the right-hand side of (2.1) in the sense that

$$\begin{aligned} & \langle \mathbf{x}_i \circ \mathbf{y}_j \circ \mathbf{x}_i^* \circ \mathbf{y}_j^*,\ \mathbf{x}_{i'} \circ \mathbf{y}_{j'} \circ \mathbf{x}_{i'}^* \circ \mathbf{y}_{j'}^* \rangle \\ = & \langle \mathbf{x}_i, \mathbf{x}_{i'} \rangle \langle \mathbf{y}_j, \mathbf{y}_{j'} \rangle \langle \mathbf{x}_i^*, \mathbf{x}_{i'}^* \rangle \langle \mathbf{y}_j^*, \mathbf{y}_{j'}^* \rangle, \\ = & (\mathbf{x}_i^\dagger \mathbf{x}_{i'})(\mathbf{y}_j^\dagger \mathbf{y}_{j'})(\mathbf{x}_i^\top \mathbf{x}_{i'}^*)(\mathbf{y}_j^\top \mathbf{y}_{j'}^*) = 0 \end{aligned}$$

for any $i \neq i'$ or $j \neq j'$. Hence, these rank-1 tensors constitute a complete orthogonal decomposition of tensor $\mathcal{A}$ [14]. For the general case of the optimal value of the objective function in problem (1.1) being nonzero, we have the following conclusion.

Proposition 2.1. *Let $\{\mathbf{x}_i\}_1^m$ and $\{\mathbf{y}_j\}_1^n$ be orthonormal bases of complex spaces C^m and C^n, respectively. Put*

$$\mathcal{T}_{ij} = \mathbf{x}_i \circ \mathbf{y}_j \circ \mathbf{x}_i^* \circ \mathbf{y}_j^*, \quad \mathcal{T} = \sum_{i,j} p_{ij} \mathcal{T}_{ij},$$

where $p_{ij} \in R,\ i = 1, 2, \ldots, m, j = 1, 2, \ldots, n$. Then for any fourth-order partially Hermite tensor $\mathcal{A}$ satisfying $\sum\limits_{i,j} \mathcal{A}_{ijij} = 1$ and (1.3), $\inf_{\mathcal{T}} \|\mathcal{A} - \mathcal{T}\|$ is achieved by $\mathcal{T}$ satisfying

$$\langle \mathcal{A} - \mathcal{T}, \mathcal{T}_{ij} \rangle = 0, \quad p_{ij} = \langle \mathcal{A}, \mathcal{T}_{ij} \rangle \geq 0, \quad i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n \tag{2.2}$$

and

$$\|\mathcal{A} - \mathcal{T}\|^2 = \|\mathcal{A}\|^2 - \sum_{i,j} p_{ij}^2.$$

Proof. Clearly, for any $1 \leq i \leq m, 1 \leq j \leq n$, the rank-1 tensor $\mathbf{x}_i \circ \mathbf{y}_j \circ \mathbf{x}_i^* \circ \mathbf{y}_j^*$ is partially Hermite. Thus, from the assumption, it follows that

$$\langle \mathcal{A}, \mathcal{T}_{ij} \rangle = \langle \mathcal{A}, \mathbf{x}_i \circ \mathbf{y}_j \circ \mathbf{x}_i^* \circ \mathbf{y}_j^* \rangle = \mathcal{A} \mathbf{x}_i \mathbf{y}_j \mathbf{x}_i^* \mathbf{y}_j^* \in R.$$

To prove the first equation in (2.2), suppose on the contrary, there exist $1 \leq i_0 \leq m, 1 \leq j_0 \leq n$ such that

$$\langle \mathcal{A} - \mathcal{T}, \mathcal{T}_{i_0 j_0} \rangle = \varepsilon \neq 0.$$

Then $\varepsilon \in R$ and

$$\begin{aligned} \|\mathcal{A} - \mathcal{T} - \varepsilon \mathcal{T}_{i_0 j_0}\|^2 = & \|\mathcal{A} - \mathcal{T}\|^2 - 2\varepsilon \langle \mathcal{A} - \mathcal{T}, \mathcal{T}_{i_0 j_0} \rangle + \varepsilon^2 \|\mathcal{T}_{i_0 j_0}\|^2 \\ = & \|\mathcal{A} - \mathcal{T}\|^2 - \varepsilon^2 < \|\mathcal{A} - \mathcal{T}\|^2, \end{aligned}$$

which leads to a contradiction. Hence the first equation in (2.2) holds. Furthermore, using the orthogonality of $\{\mathcal{T}_{ij}\}$, one has that for any $1 \leq k \leq m, 1 \leq l \leq n$,

$$\begin{aligned} 0 = & \langle \mathcal{A} - \sum_{i,j} p_{ij} \mathcal{T}_{ij}, \mathcal{T}_{kl} \rangle = \langle \mathcal{A}, \mathcal{T}_{kl} \rangle - \langle \sum_{i,j} p_{ij} \mathcal{T}_{ij}, \mathcal{T}_{kl} \rangle \\ = & \langle \mathcal{A}, \mathcal{T}_{kl} \rangle - p_{kl}. \end{aligned}$$

Thus, for any $1 \leq i \leq m, 1 \leq j \leq n$,

$$p_{ij} = \langle \mathcal{A}, \mathcal{T}_{ij} \rangle.$$

From (1.3), one has

$$p_{ij} = \langle \mathcal{A}, \mathcal{T}_{ij} \rangle = \mathcal{A} \mathbf{x}_i \mathbf{y}_j \mathbf{x}_i^* \mathbf{y}_j^* \geq 0.$$

Finally,

$$\|\mathcal{A}-\mathcal{T}\|^2 = \|\mathcal{A}\|^2 - 2\langle \mathcal{A}, \mathcal{T}\rangle + \sum_{i,j} p_{ij}^2 = \|\mathcal{A}\|^2 - \sum_{i,j} p_{ij}^2. \qquad \square$$

Certainly, when condition (2.1) holds, then $\|\mathcal{A}\|^2 = \sum_{i,j} p_{ij}^2$ which reduces to the case that tensor $\mathcal{A}$ has a complete orthogonal decomposition.

Proposition 2.2. *For any fourth-order partially Hermite tensor $\mathcal{A}$ such that $\sum_{i,j} \mathcal{A}_{ijij} = 1$ and any orthonormal basis $\{\mathbf{x}_i\}_1^m$ of C^m and $\{\mathbf{y}_j\}_1^n$ of C^n, it holds that*

$$\sum_{i,j} \mathcal{A}\mathbf{x}_i\mathbf{y}_j\mathbf{x}_i^*\mathbf{y}_j^* = \sum_{i,j} \mathcal{A}_{ijij} = \sum_{i,j} p_{ij} = 1.$$

Proof. For any orthonormal base $\{\mathbf{x}_i\}_1^m$ of C^m,

$$A(\mathbf{x}) := \sum_{i=1}^{m} \mathcal{A} \times_1 \mathbf{x}_i \times_3 \mathbf{x}_i^*$$

is an n-dimensional Hermite matrix, and for any orthonormal base $\{\mathbf{y}_j\}_1^n$ of C^n,

$$\sum_{i,j} \mathcal{A}\mathbf{x}_i\mathbf{y}_j\mathbf{x}_i^*\mathbf{y}_j^* = \sum_{j=1}^{n} \mathbf{y}_j^\dagger \mathcal{A}(\mathbf{x})\mathbf{y}_j = \mathrm{tr}(A(\mathbf{x})).$$

Similarly,

$$B(\mathbf{y}) := \sum_{j=1}^{n} \mathcal{A} \times_2 \mathbf{y}_j \times_4 \mathbf{y}_j^*$$

is an m-dimensional Hermite matrix, and for any orthonormal base $\{\mathbf{x}_i\}_1^m$ of C^m,

$$\sum_{i,j} \mathcal{A}\mathbf{x}_i\mathbf{y}_j\mathbf{x}_i^*\mathbf{y}_j^* = \sum_{i=1}^{m} \mathbf{x}_i^\dagger B(\mathbf{y})\mathbf{x}_i = \mathrm{tr}(B(\mathbf{y})).$$

This means that $\sum_{i,j} \mathcal{A}\mathbf{x}_i\mathbf{y}_j\mathbf{x}_i^*\mathbf{y}_j^*$ is independent of the choice of orthonormal bases $\{\mathbf{x}_i\}_1^m$ and $\{\mathbf{y}_j\}_1^n$. Hence, we may take a set of special base $\{\mathbf{x}_i\}_1^m$ such that the i-th element of $\mathbf{x}_i$ is one and other elements are all zero, and similarly for $\{\mathbf{y}_j\}_1^n$, to obtain

$$\sum_{i,j} \mathcal{A}\mathbf{x}_i\mathbf{y}_j\mathbf{x}_i^*\mathbf{y}_j^* = \mathrm{tr}A(\mathbf{x}) = \mathrm{tr}B(\mathbf{y}) = \sum_{i,j} \mathcal{A}_{ijij} = \sum_{i,j} p_{ij} = 1.$$

The desired result follows. $\square$

From Propositions 2.1 and 2.2, the problem (1.1) is reduced to

$$\begin{aligned} &\min \|\mathcal{A} - \sum_{i,j} p_{ij}\mathbf{x}_i \circ \mathbf{y}_j \circ \mathbf{x}_i^* \circ \mathbf{y}_j^*\|_F^2 \\ &\text{s.t.} \quad \mathbf{x}_i^\dagger \mathbf{x}_{i'} = \delta_{ii'}, \quad i, i' = 1, 2, \ldots, m, \\ &\qquad\ \ \mathbf{y}_j^\dagger \mathbf{y}_{j'} = \delta_{jj'}, \quad j, j' = 1, 2, \ldots, n, \\ &\qquad\ \ p_{ij} \in R \end{aligned}$$

which is also equivalent to

$$\begin{array}{ll} \max F(X,Y) = \sum_{i,j} \left(\mathcal{A}\mathbf{x}_i\mathbf{y}_j\mathbf{x}_i^*\mathbf{y}_j^*\right)^2 & \\ \text{s.t.} \quad \mathbf{x}_i^\dagger\mathbf{x}_{i'} = \delta_{ii'}, \quad i,i' = 1,2,\ldots,m, & \\ \qquad \mathbf{y}_j^\dagger\mathbf{y}_{j'} = \delta_{jj'}, \quad j,j' = 1,2,\ldots,n. & \end{array} \tag{2.3}$$

where $X = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m) \in C^{m\times m}$ and $Y = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n) \in C^{n\times n}$.

Problem (2.3) is a continuously differentiable optimization problem on the product of two complex Riemannian manifolds, which is also the product of two Lie groups [11]. Note that here each Riemannian manifold is not connected, or more precisely, it has two components [21]. However, we can maximize the objective function over one component, as for any unit vectors $\mathbf{x}_i \in C^m$ and $\mathbf{y}_j \in C^n$, it holds that

$$\begin{aligned} \mathcal{A}\mathbf{x}_i\mathbf{y}_j\mathbf{x}_i^*\mathbf{y}_j^* &= \mathcal{A}(-\mathbf{x}_i)\mathbf{y}_j(-\mathbf{x}_i^*)\mathbf{y}_j^* \\ &= \mathcal{A}\mathbf{x}_i(-\mathbf{y}_j)\mathbf{x}_i^*(-\mathbf{y}_j^*) \\ &= \mathcal{A}(-\mathbf{x}_i)(-\mathbf{y}_j)(-\mathbf{x}_i^*)(-\mathbf{y}_j^*). \end{aligned}$$

It should be noted that the condition described in Proposition 2.1 is not sufficient since for any orthonormal bases $\{\mathbf{x}_i\}_1^m$ and $\{\mathbf{y}_j\}_1^n$, it always holds that

$$\langle \mathcal{A} - \mathcal{T}, \mathcal{T}_{ij} \rangle = 0, \quad i = 1,2,\ldots,m, \; j = 1,2,\ldots,n$$

and

$$\|\mathcal{A} - \mathcal{T}\|^2 = \|\mathcal{A}\|^2 - \sum_{i,j} p_{ij}^2.$$

3 Alternative Steepest Direction Algorithm

In this section, we design a numerical method for solving problem (2.3). As complex variables are involved in this problem, we need to introduce the gradient of a function w.r.t. complex variables. In general, for a real-valued function $G(\mathbf{z})$ in complex variables $\mathbf{z} \in C^d$, if it is analytic with respect to variables $\mathbf{z}$ and $\mathbf{z}^*$ independently, then its gradient can be described by a pair of complex-valued operators in terms of real differentials w.r.t. the real part and imaginary part [3], i.e.,

$$G_{\mathbf{z}} := \frac{1}{2}\Big(\frac{\partial G}{\partial \mathbf{z}_R} - i\frac{\partial G}{\partial \mathbf{z}_I}\Big), \qquad G_{\mathbf{z}^*} := \frac{1}{2}\Big(\frac{\partial G}{\partial \mathbf{z}_R} + i\frac{\partial G}{\partial \mathbf{z}_I}\Big),$$

where $\mathbf{z}_R$ and $\mathbf{z}_I$ respectively denote the real part and the imaginary part of $\mathbf{z}$. From Theorem 2 in [3], $\mathbf{z}$ is a critical point of function $G(\mathbf{z})$ if and only if $G_{\mathbf{z}} = 0$ and/or $G_{\mathbf{z}^*} = 0$

Based on gradients $G_{\mathbf{z}^*}$ and $G_{\mathbf{z}}$, we obtain the first-order Taylor expression of $G(\mathbf{z})$:

$$G(\mathbf{z}_0 + \triangle\mathbf{z}) = G(\mathbf{z}_0) + G_{\mathbf{z}}^\top(\mathbf{z}_0)\triangle\mathbf{z} + G_{\mathbf{z}^*}^\top(\mathbf{z}_0)\triangle\mathbf{z}^* + o(\|\triangle\mathbf{z}\|).$$

Since $G_{\mathbf{z}^*}(\mathbf{z}_0) = G_{\mathbf{z}}^*(\mathbf{z}_0)$, we have

$$G_{\mathbf{z}}^\top(\mathbf{z}_0)\triangle\mathbf{z} + G_{\mathbf{z}^*}^\top(\mathbf{z}_0)\triangle\mathbf{z}^* = 2R(G_{\mathbf{z}}^\top(\mathbf{z}_0)\triangle\mathbf{z}) = 2R(G_{\mathbf{z}^*}^\dagger(\mathbf{z}_0)\triangle\mathbf{z}),$$

where $R(\cdot)$ denotes the real part of a quantity. This means that $G_{\mathbf{z}^*}(\mathbf{z}_0)$ can be taken as the steepest direction of the real-valued function $G(\mathbf{z})$ at $\mathbf{z}_0$, and based on $G_{\mathbf{z}^*}$, we may design a steepest direction method to optimize a real-valued function in complex spaces.

As the objective function in (2.3) is maximized on the product of two special Riemannian manifolds, i.e., two Lie groups, we recall some related definitions on the Riemannian manifold $M = \{X \in C^{n\times n} \mid X^\dagger X = I\}$ which will be used subsequently. The tangent space at point $X \in M$ is defined as [5]

$$T_X(M) = \{\Delta \in C^{n\times n} \mid X\Delta^\dagger + \Delta X^\dagger = 0\}.$$

Any tangent vector $\Delta \in T_X(M)$ can be written as $\Delta = SX$ for some $S \in C^{n\times n}$ such that $S + S^\dagger = 0$ by setting $S = \Delta X^\dagger$. For function $G(X)$ defined on $C^{n\times n}$, denote the $n \times n$ matrix of partial derivatives of G with respect to the elements of X by G_X, i.e.,

$$(G_X)_{ij} = (\partial G/\partial X_{ij}).$$

Then the gradient of $G(X)$ on the manifold M, denoted by ∇G, is defined via [7]

$$R\Big(\mathrm{tr}(G_X^\dagger \Delta)\Big) = \frac{1}{2} R\Big(\mathrm{tr}(\nabla G^\dagger \Delta)\Big), \quad \forall\ \Delta \in T_X(M),$$

and can be evaluated as [1, 2]:

$$\nabla G = G_X - X G_X^\dagger X. \tag{3.1}$$

It can be taken as the projected gradient of function $G(X)$ on the tangent space $T_X(M)$. Another useful notion is the geodesic which is a curve with minimal length on the manifold and is uniquely determined by the tangent vector at the initial point. For manifold M defined above, the Geodesics along tangent direction $\Delta = SX$ from X has the following expression [11]

$$X(t) = \exp(tS)X.$$

Based on the above analysis, we now design a steepest direction method for problem (2.3). Since two unitary matrix constraints are involved, i.e., two independent manifolds are involved, we need to handle them independently and alternatively in the iteration.

First, consider the optimization problem in $X = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m) \in C^{m\times m}$:

$$\begin{array}{ll} \max & \sum\limits_i \left(\mathcal{A}\mathbf{x}_i\mathbf{y}_j\mathbf{x}_i^*\mathbf{y}_j^*\right)^2 \\ \text{s.t.} & \mathbf{x}_i^\dagger \mathbf{x}_{i'} = \delta_{ii'}, \quad i, i' = 1, 2, \ldots, m \end{array}$$

with $Y = Y^k = (\mathbf{y}_1^k, \mathbf{y}_2^k, \ldots, \mathbf{y}_n^k) \in C^{n\times n}$ fixed. After $X^{k+1} = (\mathbf{x}_1^{k+1}, \mathbf{x}_2^{k+1}, \ldots, \mathbf{x}_m^{k+1}) \in C^{m\times m}$, the solution of the optimization problem above, is obtained, then consider the optimization problem in Y

$$\begin{array}{ll} \max & \sum\limits_i \left(\mathcal{A}\mathbf{x}_i\mathbf{y}_j\mathbf{x}_i^*\mathbf{y}_j^*\right)^2 \\ \text{s.t.} & \mathbf{y}_j^\dagger \mathbf{y}_{j'} = \delta_{jj'}, \quad j, j' = 1, 2, \ldots, n \end{array}$$

with $X = X^{k+1}$ fixed. Repeat this process until certain optimal conditions are satisfied. The following is the proposed steepest direction method for problem (2.3).

Algorithm 3.1.

Step 0. Choose parameters $\eta > 0$, $\varepsilon \geq 0$ and $\sigma, \gamma \in (0, 1)$. Take initial point $X^0 = I_m, Y^0 = I_n$, set $k = 0$, and $flag_1 = flag_2 = 1$.

Step 1. Compute the steepest direction of function $F(X, Y^k)$ at X^k on the Riemannian manifold $M_X = \{X \in C^{m\times m} \mid X^\dagger X = I_m\}$, i.e., compute $\nabla_{X^*} F(X^k, Y^k)$ via formula (3.1) and denote it as d_k^X. If $\|d_k^X\|_{\rm F} \le \varepsilon$, set $X^{k+1} = X^k$ and $flag_1 = 0$, goto Step 3; otherwise, goto the next step.

Step 2. Determine stepsize s_k via Armijo rule, i.e., take the smallest nonnegative integer m_k such that

$$F(X^k(s_k), Y^k) \ge F(X^k, Y^k) + \sigma s_k R\big(\langle d_k^X, \nabla_X F(X^k, Y^k)\rangle\big)$$

where $s_k = \eta\gamma^{m_k}$ and $X^k(s) = \exp^{s d_k^X X^{k\dagger}} X^k$. Set $X^{k+1} = X^k(s_k)$ and goto the next step.

Step 3. Compute the steepest direction of function $F(X^{k+1}, Y)$ at Y^k on the Riemannian manifold $M_Y = \{Y \in C^{n\times n} \mid Y^\dagger Y = I_n\}$, i.e., compute $\nabla_{Y^*} F(X^{k+1}, Y^k)$ via (3.1) and denote it as d_k^Y. If $\|d_k^Y\|_{\rm F} \le \varepsilon$, set $Y^{k+1} = Y^k$, $flag_2 = 0$ and goto Step 5; otherwise, goto the next step.

Step 4. Determine stepsize t_k via Armijo rule, i.e., take the smallest nonnegative integer m_k such that

$$F(X^{k+1}, Y^k(t_k)) \ge F(X^{k+1}, Y^k) + \sigma t_k R\big(\langle d_k^Y, \nabla_Y F(X^{k+1}, Y^k)\rangle\big)$$

where $t_k = \eta\gamma^{m_k}$ and $Y^k(t) = \exp^{t d_k^Y Y^{k\dagger}} Y^k$. Set $Y^{k+1} = Y^k(t_k)$ and goto the next step.

Step 5. If $flag_1 = flag_2 = 0$, stop; otherwise, set $k = k + 1$ and goto step 1.

In this method, the new iteration is obtained by searching along the projected curve of the steepest direction. Hence, the generated sequence $\{F(X^k, Y^k)\}$ of the algorithm is monotonically increasing and each cluster point of the generated sequence $\{X^k, Y^k\}$ is a critical point of the problem.

4 Numerical Simulation

To test the efficiency of the proposed method, we perform some numerical experiments. In our numerical computing, the parameters used in the algorithm are set as $\eta = 1, \rho = 0.1, \gamma = 0.5$. We take $\varepsilon = 10^{-2}$ as the stop criterion. All codes are written in MATLAB 7.0 and run on a PIV 2.0 GHz personal computer.

It should be noted that Algorithm 3.1 is a gradient-type method, and the generated sequence only converges to a critical point of the problem [1] and the global minimizer can hardly be obtained. Here, we test two partially Hermite tensors with $\max\{m, n\} > 2$. The following numerical experiments show that the algorithm can find a good solution of the problem.

Example 4.1. Take $m = 2, n = 3$, and take $\mathcal{A}$ with randomly generated elements

$$A(:,:,1,1) = \begin{pmatrix} 0.2732 + 0.0000i & -0.0816 - 0.0347i & 0.1448 - 0.0977i \\ 0.0706 - 0.0365i & 0.0338 - 0.0320i & 0.0530 - 0.0288i \end{pmatrix}$$

$$A(:,:,2,1) = \begin{pmatrix} 0.0706 + 0.0365i & 0.0473 + 0.0052i & 0.0562 + 0.0263i \\ 0.1458 + 0.0000i & -0.0404 - 0.0187i & 0.0782 - 0.0509i \end{pmatrix}$$

$$A(:,:,1,2) = \begin{pmatrix} -0.0816 + 0.0347i & 0.1772 + 0.0000i & 0.0171 + 0.1154i \\ 0.0473 - 0.0052i & -0.0349 + 0.0149i & 0.0091 - 0.0167i \end{pmatrix}$$

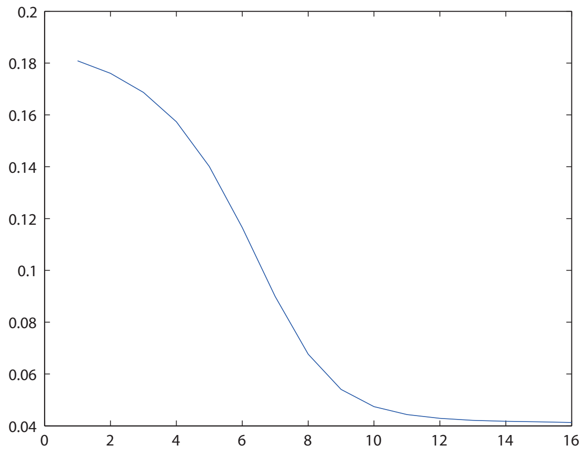

Figure 4.1: Numerical result of Example 4.1.

$$A(:,:,2,2) = \begin{pmatrix} 0.0338 + 0.0320i & -0.0349 - 0.0149i & 0.0167 - 0.0040i \\ -0.0404 + 0.0187i & 0.0905 - 0.0000i & 0.0096 + 0.0596i \end{pmatrix}$$

$$A(:,:,1,3) = \begin{pmatrix} 0.1448 + 0.0977i & 0.0171 - 0.1154i & 0.2042 + 0.0000i \\ 0.0562 - 0.0263i & 0.0167 + 0.0040i & 0.0547 - 0.0282i \end{pmatrix}$$

$$A(:,:,2,3) = \begin{pmatrix} 0.0530 + 0.0288i & 0.0091 + 0.0167i & 0.0547 + 0.0282i \\ 0.0782 + 0.0509i & 0.0096 - 0.0596i & 0.1091 - 0.0000i \end{pmatrix}$$

This tensor has a highly approximated complete orthogonal decomposition. For this tensor, Algorithm 3.1 terminates after 16 iteration with the objective function value 0.0413. The computer running time is 0.25 seconds. The varying tendency of the objective function of problem (1.1) is shown in Figure 4.1.

Example 4.2. Take $m = 5, n = 2$, and take $\mathcal{A}$ with randomly generated elements

$$A(:,:,1,1) = \begin{pmatrix} 0.1399 + 0.0000i & 0.0523 + 0.0411i \\ -0.0107 - 0.0703i & 0.0274 - 0.0208i \\ 0.0378 - 0.0582i & 0.0315 - 0.0170i \\ -0.0298 - 0.0041i & -0.0098 + 0.0084i \\ 0.0111 - 0.0228i & 0.0131 - 0.0075i \end{pmatrix}$$

$$A(:,:,2,1) = \begin{pmatrix} -0.0107 + 0.0703i & -0.0220 + 0.0367i \\ 0.1545 - 0.0000i & 0.0550 + 0.0409i \\ 0.0697 - 0.0232i & 0.0347 + 0.0094i \\ 0.0565 + 0.0505i & -0.0014 + 0.0277i \\ 0.0005 + 0.0178i & -0.0019 + 0.0046i \end{pmatrix}$$

$$A(:,:,3,1) = \begin{pmatrix} 0.0378 + 0.0582i & -0.0088 + 0.0301i \\ 0.0697 + 0.0232i & 0.0162 + 0.0296i \\ 0.1558 + 0.0000i & 0.0509 + 0.0336i \\ -0.0248 - 0.0091i & -0.0172 - 0.0139i \\ 0.0825 - 0.0159i & 0.0256 + 0.0109i \end{pmatrix}$$

$$A(:,:,4,1) = \begin{pmatrix} -0.0298+0.0041i & 0.0041+0.0016i \\ 0.0565-0.0505i & 0.0326-0.0127i \\ -0.0248+0.0091i & -0.0099-0.0148i \\ 0.0979+0.0000i & 0.0354+0.0268i \\ -0.0006+0.0552i & -0.0070+0.0056i \end{pmatrix}$$

$$A(:,:,5,1) = \begin{pmatrix} 0.0111+0.0228i & -0.0056+0.0127i \\ 0.0005-0.0178i & 0.0020-0.0048i \\ 0.0825+0.0159i & 0.0242+0.0180i \\ -0.0006-0.0552i & -0.0016-0.0193i \\ 0.1254-0.0000i & 0.0388+0.0235i \end{pmatrix}$$

$$A(:,:,1,2) = \begin{pmatrix} 0.0523-0.0411i & 0.0817+0.0000i \\ -0.0220-0.0367i & 0.0170-0.0496i \\ -0.0088-0.0301i & 0.0125-0.0404i \\ 0.0041-0.0016i & 0.0113+0.0147i \\ -0.0056-0.0127i & 0.0051-0.0188i \end{pmatrix}$$

$$A(:,:,2,2) = \begin{pmatrix} 0.0274+0.0208i & 0.0170+0.0496i \\ 0.0550-0.0409i & 0.0809+0.0000i \\ 0.0162-0.0296i & 0.0387-0.0186i \\ 0.0326+0.0127i & 0.0139+0.0341i \\ 0.0020+0.0048i & -0.0002+0.0036i \end{pmatrix}$$

$$A(:,:,3,2) = \begin{pmatrix} 0.0315+0.0170i & 0.0125+0.0404i \\ 0.0347-0.0094i & 0.0387+0.0186i \\ 0.0509-0.0336i & 0.0656-0.0000i \\ -0.0099+0.0148i & -0.0293+0.0079i \\ 0.0242-0.0180i & 0.0277-0.0010i \end{pmatrix}$$

$$A(:,:,4,2) = \begin{pmatrix} -0.0098-0.0084i & 0.0113-0.0147i \\ -0.0014-0.0277i & 0.0139-0.0341i \\ -0.0172+0.0139i & -0.0293-0.0079i \\ 0.0354-0.0268i & 0.0531+0.0000i \\ -0.0016+0.0193i & -0.0144+0.0039i \end{pmatrix}$$

$$A(:,:,5,2) = \begin{pmatrix} 0.0131+0.0075i & 0.0051+0.0188i \\ -0.0019-0.0046i & -0.0002-0.0036i \\ 0.0256-0.0109i & 0.0277+0.0010i \\ -0.0070-0.0056i & -0.0144-0.0039i \\ 0.0388-0.0235i & 0.0453+0.0000i \end{pmatrix}$$

This tensor also has a highly approximated complete orthogonal decomposition. For this tensor, Algorithm 3.1 terminates after 37 iteration with the objective function value 0.0216. The computer running time is 0.64 seconds. The varying tendency of the objective function of problem (1.1) is shown in Figure 4.2.

For the case $m = n = 5$, the numerical result on one randomly generated tensor can be seen from the following figure with running time being 8.859 seconds. The varying tendency of the objective function of problem (1.1) is shown in Figure 4.3.

From our large amount of numerical experiments, we see that the objective function value has a large decrease during the iteration in most cases and its is efficient for solving the problem. It should be noted that the algorithm encounters difficulties for larger dimensions m, n. Hence, how to make the algorithm to be efficient for large scale problem is an important topic for future research.

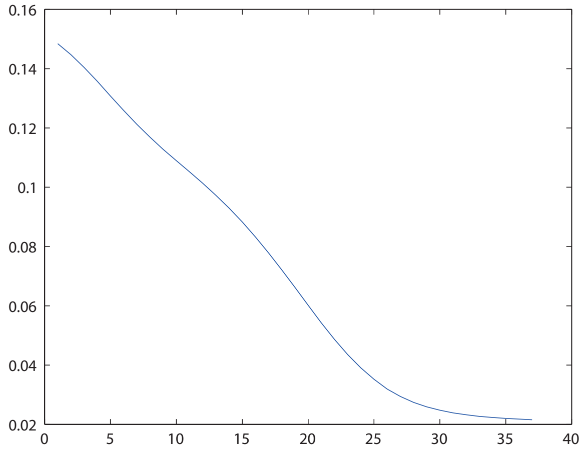

Figure 4.2: Numerical results of Example 4.2

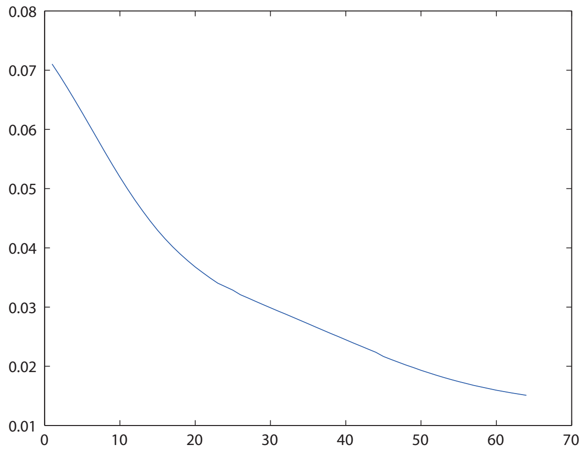

Figure 4.3: Decreasing procedure of the objective function (1.1) with $m = n = 5$

Acknowledgement

The authors would like to express their sincerely gratitude to Profs. T.E. Abrudan and L.H. Lim for their assistance on this research, and wish to give their sincere thanks to the anonymous referees for their valuable suggestions and helpful comments which help improve the quality of the correspondence significantly.

References

[1] T.E. Abrudan, J. Eriksson and V. Koivunen, Steepest descent algorithms for optimization under unitary matrix constraint, *IEEE Transactions on Signal Processing* 56 (2008) 1134–1147.

[2] T. Abrudan, J. Eriksson and V. Koivunen, Conjugate gradient algorithm for optimization under unitary matrix constraint, *Signal Processing* 89 (2009) 1704–1714.

[3] D.H. Brandwood, A complex gradient operator and its applications in adptive array theory, Microwaves, *Optics and Antennas, IEE Proceedings H*, 130 (1983) 11–16.

[4] B. Dakić, V. Vedral and C. Brukner, Necessary and sufficient condition for nonzero quantum discord, *Physical Review Letters* 105 (2010) 190502.

[5] M.P. Do Carmo, *Riemannian Geometry. Mathematics: Theory and Applications*, Birkhauser, Basel, 1992.

[6] P. Comon, G. Golub and L.H. Lim, B. Mourrain, Symmetric tensors and symmetric tensor rank, *SIAM Journal on Matrix Analysis and Applications* 30 (2008) 1254–1279

[7] A. Edelman, T.A. Arias and S.T. Smith, The geometry of algorithms with orthogonality constraints, *SIAM Journal on Matrix Analysis and Applications* 20 (1998) 303–353.

[8] L. Eldén and B. Savas, A Newton-Grassmann method for computing the best multilinear rank-(r_1, r_2, r_3) approximation of a tensor, *SIAM Journal on Matrix Analysis Applications* 31 (2009) 248–271.

[9] A. Ferraro, L. Aolita, D. Cavalcanti, F.M. Cucchietti, and A. Acín, Almost all quantum states have nonclassical correlations, *Physical Review A* 81 (2010) 052318.

[10] D. Gabay, Minimizing a differentiable function over a differential manifold, *Journal on Optimization Theory and Applications* 37 (1982) 177–219.

[11] S. Helgason, Differential Geometry, *Lie Groups and Symmetric Spaces*, Academic Press, New York, 1978.

[12] M. Ishteva, P.A. Absil, S.V. Huffel, and L. De Lathauwer, Best low multilinear rank approximation of higher-order tensors, based on the Riemannian trust-region scheme, *SIAM Journal on Matrix Analysis and Applications* 32 (2011) 115–135.

[13] C. Ling, X.Z. Zhang, and L.Q. Qi, Semidefinite relaxation approximation for multivariate bi-quadratic optimization with quadratic constraints, *Numer Linear Algebra and Applications* 19 (2012) 113–131.

[14] T.G. Kolda, Orthogonal tensor decompositions, *SIAM Journal on Matrix Analysis and Applications* 23 (2001) 243–255.

[15] D.G. Luenberger, The gradient projection method along geodesics, *Management Science* 18 (1972) 620–631.

[16] S. Luo, Quantum discord for two-qubit systems, *Physical Review A* 77 (2008) 042303.

[17] S. Luo and S. Fu, Geometric measure of quantum discord, *Physical Review A* 82 (2010) 034302.

[18] S. Luo, Using measurement-induced disturbance to characterize correlations as classical or quantum, *Physical Review A* 77 (2008) 022301.

[19] J. Nocedal and S. J. Wright, *Numerical Optimization*, New York, Springer, 1999.

[20] H. Ollivier and W.H. Zurek, Quantum discord: a measure of the quantumness of correlations, *Physical Review Letters* 88 (2001) 017901.

[21] T. Rapcsák, On minimization on Stiefel manifolds, European Journal of *Operations Research* 143 (2002) 365–376.

[22] B. Savas and L.H. Lim, Quasi-Newton methods on Grassmannians and multilinear approximations of tensors, *SIAM Journal on Scientific Computing* 32 (2010) 3352–3393.

[23] S.T. Smith, Geometric optimization methods for adaptive filtering, Ph.D. Thesis, Harvard University, Cambridge, MA, 1993.

[24] G.L. Zhou, L Caccetta, K.L. Teo and S.Y. WU, Nonnegative polynomial optimization over unit spheres and convex programming relaxations, *SIAM Journal on Optimization* 22 (2012) 987–1008.

Manuscript received 3 March 2013
revised 4 June 2013
accepted for publication 13 October 2013

Yiju Wang
School of Management Science, Qufu Normal University
Rizhao, Shandong, 276800, China
E-mail address: `wyiju@hotmail.com`

Liqun Qi
Department of Applied Mathematics, The Hong Kong Polytechnic University
Hong Kong
E-mail address: `maqilq@polyu.edu.hk`

Shunlong Luo
Academy of Mathematics and Systems Science, Chinese Academy of Sciences
Beijing, 100190, China
E-mail address: `luosl@amt.ac.cn`

Yi Xu
Academy of Department of Applied Mathematics, The Hong Kong Polytechnic University
Hong Kong
E-mail address: `yi.xu1983@gmail.com`