Journal of Industrial and Management Optimization

https://www.aimspress.com/journal/jimo

JIMO, 22(5): 2301–2318.
DOI:10.3934/jimo.2026084
Received: 08 December 2025
Revised: 08 February 2026
Accepted: 20 March 2026
Published: 13 April 2026

Research article

A nonsmooth, nonconvex optimization approach over sphere constraints for Variants of regularized CCA and SVD

Amir Beck* **and Raz Sharon**

School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel

* **Correspondence:** Email: becka@tauex.tau.ac.il.

Abstract: In this paper, we study a general nonsmooth, nonconvex optimization problem over a cross product of spheres, suitable for various well-known variants of the canonical correlation analysis (CCA) and singular value decomposition (SVD) problems that promote sparsity and smoothness. We propose an alternating minimization algorithm for a smooth approximation of this problem and study its rate of convergence. Numerical experiments demonstrate the potential of the suggested method.

Keywords: sparse CCA; sparse SVD; nonsmooth nonconvex optimization; alternating minimization; sphere constraints; smooth approximation; manifold optimization
Mathematics Subject Classification: 90C26, 90C30

---

1. Introduction

We consider the following nonsmooth and nonconvex optimization problem:

$$\begin{array}{ll} \min\limits_{\mathbf{a}\in\mathbb{R}^{d_1},\mathbf{b}\in\mathbb{R}^{d_2}} & \begin{pmatrix}\mathbf{a}\\ \mathbf{b}\end{pmatrix}^T \mathbf{S}\begin{pmatrix}\mathbf{a}\\ \mathbf{b}\end{pmatrix} + \tau_1\|\mathbf{H}^{\mathbf{a}}\mathbf{a}\|_1 + \tau_2\|\mathbf{H}^{\mathbf{b}}\mathbf{b}\|_1 \\ \text{s.t.} & \|\mathbf{a}\|_2 = 1, \\ & \|\mathbf{b}\|_2 = 1, \end{array} \tag{1.1}$$

where $\tau_1, \tau_2 \geq 0$ are regularization parameters, $\mathbf{S}$ is a real symmetric matrix, $\mathbf{H}^{\mathbf{a}} \in \mathbb{R}^{h_1\times d_1}$, and $\mathbf{H}^{\mathbf{b}} \in \mathbb{R}^{h_2\times d_2}$. Problem (1.1) can model several well-known problems, such as sparse canonical correlation analysis [1–3], sparse singular value decomposition [1,4], and sparse-smooth regularized singular value decomposition [5]. Detailed explanations are provided in the following.

1.1. Model examples

1.1.1. Sparse canonical correlation analysis

Consider two paired multi-dimensional data sets, $\{(\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_n, \mathbf{y}_n)\}$, where for any i, $\mathbf{x}_i$ and $\mathbf{y}_i$ are vectors in $\mathbb{R}^{d_1}$ and $\mathbb{R}^{d_2}$, respectively. Canonical correlation analysis (CCA) seeks to find linear combinations of the variables in the two sets with the highest possible correlation. Denote by $\mathbf{X} \in \mathbb{R}^{n \times d_1}$ the matrix whose ith row is $\mathbf{x}_i^T$, and similarly $\mathbf{Y} \in \mathbb{R}^{n \times d_2}$ is the matrix whose ith row is $\mathbf{y}_i^T$. Assume that the columns of the matrices $\mathbf{X}$ and $\mathbf{Y}$ have zero mean. The sample covariance matrices are given by (ignoring a multiplicative constant)

$$\mathbf{S_X} = \mathbf{X}^T\mathbf{X}, \quad \mathbf{S_Y} = \mathbf{Y}^T\mathbf{Y}, \quad \mathbf{S_{XY}} = \mathbf{X}^T\mathbf{Y}. \tag{1.2}$$

In the one–dimensional CCA problem, we seek to find a pair of vectors $(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ that maximizes the correlation between $\mathbf{Xa}$ and $\mathbf{Yb}$, leading to the optimization problem (see [6]):

$$\text{(CCA)} \quad \max_{\mathbf{a} \in \mathbb{R}^{d_1}, \mathbf{b} \in \mathbb{R}^{d_2}} \left\{ \rho_{\mathrm{x,y}}(\mathbf{a}, \mathbf{b}) := \frac{\mathbf{a}^T\mathbf{S_{XY}}\mathbf{b}}{\sqrt{\mathbf{a}^T\mathbf{S_X}\mathbf{a}}\,\sqrt{\mathbf{b}^T\mathbf{S_Y}\mathbf{b}}} : \mathbf{a}^T\mathbf{S_X}\mathbf{a} \neq 0, \mathbf{b}^T\mathbf{S_Y}\mathbf{b} \neq 0 \right\}.$$

Notice that $\rho_{\mathrm{x,y}}(\mathbf{a}, \mathbf{b})$ is invariant to multiplication by positive scalars, and consequently, we can arbitrarily choose a scaling factor and reformulate the problem as

$$\max_{\mathbf{a} \in \mathbb{R}^{d_1}, \mathbf{b} \in \mathbb{R}^{d_2}} \left\{ \mathbf{a}^T\mathbf{S_{XY}}\mathbf{b} : \mathbf{a}^T\mathbf{S_X}\mathbf{a} = 1, \mathbf{b}^T\mathbf{S_Y}\mathbf{b} = 1 \right\},$$

Sparse CCA promotes sparse solutions of the CCA problem by adding an ℓ_1-norm regularization term. This leads to the following nonconvex problem [2]:

$$\max_{\mathbf{a} \in \mathbb{R}^{d_1}, \mathbf{b} \in \mathbb{R}^{d_2}} \left\{ \mathbf{a}^T\mathbf{S_{XY}}\mathbf{b} - \tau_1\|\mathbf{a}\|_1 - \tau_2\|\mathbf{b}\|_1 : \mathbf{a}^T\mathbf{S_X}\mathbf{a} = 1, \mathbf{b}^T\mathbf{S_Y}\mathbf{b} = 1 \right\}, \tag{1.3}$$

where $\tau_1, \tau_2 \geq 0$ are given regularization parameters. We assume that the covariance matrices $\mathbf{S_X}$ and $\mathbf{S_Y}$ are positive definite (note that positive semidefiniteness of these matrices is readily implied by their definition). Using the notation $\tilde{\mathbf{a}} = \mathbf{S}_{\mathbf{X}}^{\frac{1}{2}}\mathbf{a}$, $\quad \tilde{\mathbf{b}} = \mathbf{S}_{\mathbf{Y}}^{\frac{1}{2}}\mathbf{b}$, $\quad \tilde{\mathbf{S}}_{\mathbf{XY}} = \mathbf{S}_{\mathbf{X}}^{-\frac{T}{2}}\mathbf{S_{XY}}\mathbf{S}_{\mathbf{Y}}^{-\frac{1}{2}}$, problem (1.3) can be rewritten in a simplified form as

$$\min_{\tilde{\mathbf{a}} \in \mathbb{R}^{d_1}\ \tilde{\mathbf{b}} \in \mathbb{R}^{d_2}} \left\{ -\tilde{\mathbf{a}}^T\tilde{\mathbf{S}}_{\mathbf{XY}}\tilde{\mathbf{b}} + \tau_1\|\mathbf{S}_{\mathbf{X}}^{-\frac{1}{2}}\tilde{\mathbf{a}}\|_1 + \tau_2\|\mathbf{S}_{\mathbf{Y}}^{-\frac{1}{2}}\tilde{\mathbf{b}}\|_1 : \|\tilde{\mathbf{a}}\|_2^2 = 1, \|\tilde{\mathbf{b}}\|_2^2 = 1 \right\}. \tag{1.4}$$

Thus, problem (1.4) fits the structure of problem (1.1) with $\mathbf{S} = -\frac{1}{2}\begin{pmatrix} \mathbf{0} & \tilde{\mathbf{S}}_{\mathbf{XY}} \\ \tilde{\mathbf{S}}_{\mathbf{XY}}^T & \mathbf{0} \end{pmatrix}$, $\mathbf{H}^{\mathbf{a}} = \mathbf{S}_{\mathbf{X}}^{-\frac{1}{2}}$ and $\mathbf{H}^{\mathbf{b}} = \mathbf{S}_{\mathbf{Y}}^{-\frac{1}{2}}$.

Witten et al. [1] suggested the following alternative model for obtaining sparse canonical vectors ($\tau_1, \tau_2 > 0$ being given constants):

$$\max_{\mathbf{a} \in \mathbb{R}^{d_1}, \mathbf{b} \in \mathbb{R}^{d_2}} \left\{ \mathbf{a}^T\tilde{\mathbf{S}}_{\mathbf{XY}}\mathbf{b} : \|\mathbf{a}\|_2^2 \leq 1, \|\mathbf{b}\|_2^2 \leq 1, \|\mathbf{a}\|_1 \leq \tau_1, \|\mathbf{b}\|_1 \leq \tau_2 \right\}. \tag{1.5}$$

in which the equality constraints $||\mathbf{a}||_2^2 = 1$ and $||\mathbf{b}||_2^2 = 1$ are relaxed to the inequality constraints $||\mathbf{a}||_2^2 \leq 1$ and $||\mathbf{b}||_2^2 \leq 1$, respectively. The proposed algorithm in [1] for solving (1.5) alternates between minimizing with respect to **a** and **b**.

Mai and Zhang [3] propose an iterative algorithm, named IPLS, to handle problem (1.3), by alternating between minimizing with respect to **a** and **b**. In each step, they reformulate the problem as an unconstrained penalized least squares, followed by a projection onto the feasible set.

1.1.2. Regularized singular value decomposition

The ℓ_1-regularized singular value decomposition problem of a given matrix $\mathbf{X} \in \mathbb{R}^{d_1 \times d_2}$ is given by

$$\min_{\mathbf{u}\in\mathbb{R}^{d_1},\mathbf{v}\in\mathbb{R}^{d_2}} \left\{-\mathbf{u}^T\mathbf{X}\mathbf{v} + \tau_1||\mathbf{u}||_1 + \tau_2||\mathbf{v}||_1 : ||\mathbf{u}||_2 = 1, ||\mathbf{v}||_2 = 1\right\}, \tag{1.6}$$

where $\tau_1 \geq 0$ and $\tau_2 \geq 0$ are penalty parameters. Problem (1.6) fits the structure of problem (1.1) with $\mathbf{S} = -\frac{1}{2}\begin{pmatrix} \mathbf{0} & \mathbf{X} \\ \mathbf{X}^T & \mathbf{0} \end{pmatrix}$, $\mathbf{H}^{\mathbf{a}} = \mathbf{I}_{d_1}$ and $\mathbf{H}^{\mathbf{b}} = \mathbf{I}_{d_2}$.

Witten et al. [1] proposed the following alternative model for sparse SVD:

$$\max_{\mathbf{u}\in\mathbb{R}^{d_1},\mathbf{v}\in\mathbb{R}^{d_2}} \left\{\mathbf{u}^T\mathbf{X}\mathbf{v} : ||\mathbf{u}||_2 \leq 1,\ ||\mathbf{v}||_2 \leq 1,\ ||\mathbf{u}||_1 \leq \tau_1,\ ||\mathbf{v}||_1 \leq \tau_2\right\},$$

which shares the same structure as problem (1.5). The algorithm in [1] is an alternating minimization method w.r.t. **u** and **v**.

Lee et al. [4] proposed a method, referred to as SSVD, which is essentially an alternating minimization method employed on the model (1.6). Hong et al. [5] suggested the following unconstrained sparse-smooth regularized SVD model:

$$\min_{\mathbf{u}\in\mathbb{R}^{d_1},\mathbf{v}\in\mathbb{R}^{d_2}} \frac{1}{2}\left\{\left\|\mathbf{X} - \mathbf{u}\mathbf{v}^T\right\|_F^2 + \tau_1||\mathbf{u}||_1 + \lambda_2\mathbf{v}^T\Omega_2\mathbf{v}\right\}, \tag{1.7}$$

where $\mathbf{v}^T\Omega_2\mathbf{v} = v_1^2 + \sum_{j=2}^{d_2-1}\left(v_{j+1} - 2v_j + v_{j-1}\right)^2 + v_{d_2}^2$ and $\tau_1, \lambda_2 \geq 0$. The motivation behind the sparse-smooth regularization term is to analyze data matrices, particularly spatio-temporal data, such as fMRI, where only some spatial regions are activated while the signal over time is expected to be smooth. The algorithm suggested in [4] to handle model (1.7) alternates between **u** and **v**, which we refer to as SS-SVD.

Inspired by problem (1.7), we can consider the following two-sided sparse-smooth regularized singular value decomposition model with sphere constraints:

$$\begin{array}{ll} \min\limits_{\mathbf{u}\in\mathbb{R}^{d_1},\mathbf{v}\in\mathbb{R}^{d_2}} & \{-\mathbf{u}^T\mathbf{X}\mathbf{v} + \tau_1||\mathbf{u}||_1 + \tau_2||\mathbf{v}||_1 + \lambda_1\mathbf{u}^T\Omega_1\mathbf{u} + \lambda_2\mathbf{v}^T\Omega_2\mathbf{v}\} \\ \text{s.t.} & ||\mathbf{u}||_2 = 1, \\ & ||\mathbf{v}||_2 = 1, \end{array} \tag{1.8}$$

where $\tau_1, \tau_2, \lambda_1, \lambda_2 \geq 0$ and $\Omega_1 \in \mathbb{S}^{d_1}, \Omega_2 \in \mathbb{S}^{d_2}$, which fits the structure of problem (1.1) with $\mathbf{S} = \begin{pmatrix} \Omega_1 & -\frac{1}{2}\mathbf{X} \\ -\frac{1}{2}\mathbf{X}^T & \Omega_2 \end{pmatrix}$, $\mathbf{H}^{\mathbf{a}} = \mathbf{I}_{d_1}$ and $\mathbf{H}^{\mathbf{b}} = \mathbf{I}_{d_2}$.

The main contribution of this paper is to introduce and study an alternating minimization method, referred to as the AMQ method, that is invoked on a smooth approximation of (1.1), which simplifies this challenging nonsmooth optimization. We note that the alternating minimization method is a rather old and fundamental algorithm [7, 8]. For an extensive review of the method, see [9].

The rest of this paper is structured as follows. In Section 2, we introduce a smooth approximation of the nonsmooth and nonconvex problem (1.1). In Section 3, we introduce the AMQ method which is an alternating minimization algorithm for solving the smooth approximation. A rate of convergence analysis of the norms of the Riemannian gradient of the objective function is developed based on a tailored sufficient decrease property. Finally, numerical experiments presented in Section 4 demonstrate the potential of our algorithm.

1.2. Notation

Vectors are denoted by boldface lowercase letters, e.g., $\mathbf{y}$, and the ith element in $\mathbf{y}$ is denoted by y_i. Matrices are denoted by boldface uppercase letters, e.g., $\mathbf{B}$. The smallest (largest) eigenvalue of $\mathbf{S}$ is $\lambda_{\min}(\mathbf{S})$ ($\lambda_{\max}(\mathbf{S})$). The space $\mathbb{S}^d$ is the subspace of $\mathbb{R}^{d\times d}$ comprising all symmetric matrices. The space $\mathbb{S}^d_+$ ($\mathbb{S}^d_{++}$) is the subspace of $\mathbb{R}^{d\times d}$ comprising all positive semidefinite (positive definite) matrices. For any two matrices $\mathbf{A}, \mathbf{B} \in \mathbb{S}^d$, $\mathbf{A} \succeq \mathbf{B}$ ($\mathbf{A} \succ \mathbf{B}$) means that $\mathbf{A} - \mathbf{B}$ is positive semidefinite (positive definite). The subset of $\mathbb{R}^n$ consisting of all vectors in $\mathbb{R}^n$ with nonnegative components is denoted by $\mathbb{R}^n_+$. The identity matrix of size d is denoted as $\mathbf{I}_d \in \mathbb{R}^{d\times d}$. The all-zeros vector or matrix is represented as $\mathbf{0}$. We omit specifying the dimensions in case where the dimensions are clear from the context. diag($\mathbf{x}$) is a diagonal matrix with diagonal $\mathbf{x}$. We use the standard notation $[n] \equiv \{1, 2, \ldots, n\}$ for a positive integer n. We use indexing notations similar to those commonly found in numerical computing languages, for example, $\mathbf{S}_{(i,:)}$ denotes the ith row of $\mathbf{S}$, $\mathbf{x}_{(i:j)} = (x_i, x_{i+1}, \ldots, x_j)$, and for $\mathbf{X} \in \mathbb{R}^{d_1\times d_2}$, $\mathbf{X}_{(i:j,k:)} = \mathbf{X}_{(i:j,k:d_2)}$ is the submatrix of $\mathbf{X}$ consisting of rows $i, i+1, \ldots, j$ and columns $k, k+1, \ldots, d_2$. The optimal value of the optimization problem (P) is denoted by val(P). The differential of $f(\mathbf{x})$ is $Df(\mathbf{x})$. For a given $\mathbf{x} \in \mathbb{R}^d$, $\|\mathbf{x}\|$ denotes its ℓ_2-norm of $\mathbf{x}$ even when the subscript 2 does not appear.

2. The smooth approximation model

To handle problem (1.1), we first propose to consider the well-known smooth approximation in which each absolute value expression $|x|$ is replaced by $\sqrt{x^2+\eta^2}$ where $\eta > 0$ is a smoothing parameter, see [10]. Note that $|x| \le \sqrt{x^2+\eta^2} \le |x| + \eta$. We thus consider the following smooth approximation of problem (1.1):

$$\begin{aligned} \min_{\mathbf{a}\in\mathbb{R}^{d_1},\mathbf{b}\in\mathbb{R}^{d_2}} \quad & \begin{pmatrix}\mathbf{a}\\ \mathbf{b}\end{pmatrix}^T \mathbf{S} \begin{pmatrix}\mathbf{a}\\ \mathbf{b}\end{pmatrix} + \tau_1 \sum_{i=1}^{h_1} \sqrt{\left(\mathbf{H}^{\mathbf{a}}_{(i,:)}\mathbf{a}\right)^2 + \eta^2} + \tau_2 \sum_{i=1}^{h_2} \sqrt{\left(\mathbf{H}^{\mathbf{b}}_{(i,:)}\mathbf{b}\right)^2 + \eta^2} \\ \text{s.t.} \quad & \|\mathbf{a}\| = 1, \\ & \|\mathbf{b}\| = 1, \end{aligned} \tag{2.1}$$

where $\eta > 0$ is a smoothing parameter. The feasible set of (2.1) is nonempty and compact, and the objective function is continuous. Therefore, by Weierstrass's theorem, an optimal solution to problem (2.1) exists.

We follow the iteratively reweighted least squares methodology from [11] (see also [12]) and introduce two additional auxiliary vectors of variables, $\mathbf{z} \in \mathbb{R}^{h_1}$ and $\mathbf{w} \in \mathbb{R}^{h_2}$, enabling the construction of a problem equivalent to problem (2.1). Let $f : \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \mathbb{R}^{h_1} \times \mathbb{R}^{h_2} \to \mathbb{R}$ be given by

$$\begin{aligned} f(\mathbf{a},\mathbf{b},\mathbf{z},\mathbf{w}) = \begin{pmatrix}\mathbf{a}\\ \mathbf{b}\end{pmatrix}^T \mathbf{S}\begin{pmatrix}\mathbf{a}\\ \mathbf{b}\end{pmatrix} &+ \tau_1\left(\frac{1}{2}\sum_{i=1}^{h_1}\left(\frac{(\mathbf{H}^{\mathbf{a}}_{(i,:)}\mathbf{a})^2+\eta^2}{z_i}+z_i\right)\right) \\ &+ \tau_2\left(\frac{1}{2}\sum_{i=1}^{h_2}\left(\frac{(\mathbf{H}^{\mathbf{b}}_{(i,:)}\mathbf{b})^2+\eta^2}{w_i}+w_i\right)\right). \end{aligned} \tag{2.2}$$

We will analyze the following auxiliary problem:

$$\min_{\substack{\mathbf{a}\in\mathbb{R}^{d_1},\mathbf{z}\in\mathbb{R}^{h_1}\\ \mathbf{b}\in\mathbb{R}^{d_2},\mathbf{w}\in\mathbb{R}^{h_2}}} \{f(\mathbf{a},\mathbf{b},\mathbf{z},\mathbf{w}) : \|\mathbf{a}\| = \|\mathbf{b}\| = 1, \mathbf{z},\mathbf{w} > \mathbf{0}\}. \tag{2.3}$$

The equivalence between problems (2.1) and (2.3) is established in the following lemma. The proof is elementary and hence omitted.

Lemma 2.1. *Let* $g : \mathbb{R}^d \to \mathbb{R}$ *be a smooth function,* $\eta > 0$, $C \subseteq \mathbb{R}^d$ *and* $\mathbf{H} \in \mathbb{R}^{h\times d}$. *Consider the following two optimization problems:*

$$\begin{aligned} \min_{\mathbf{a}\in\mathbb{R}^d} \quad & g_1(\mathbf{a}) \equiv g(\mathbf{a}) + \textstyle\sum_{i=1}^h \sqrt{(\mathbf{H}_{(i,:)}\mathbf{a})^2+\eta^2} \\ s.t. \quad & \mathbf{a} \in C, \end{aligned} \tag{2.4}$$

and

$$\begin{aligned} \min_{\mathbf{a}\in\mathbb{R}^d,\mathbf{z}\in\mathbb{R}^h} \quad & g_2(\mathbf{a},\mathbf{z}) \equiv g(\mathbf{a}) + \textstyle\sum_{i=1}^h \frac{1}{2}\left(\frac{(\mathbf{H}_{(i,:)}\mathbf{a})^2+\eta^2}{z_i}+z_i\right) \\ s.t. \quad & \mathbf{a} \in C, \\ & \mathbf{z} > \mathbf{0}. \end{aligned} \tag{2.5}$$

Then,

(a) For any $\mathbf{a} \in C$*, the unique optimal solution of* $\min_{\mathbf{z}>\mathbf{0}} g_2(\mathbf{a},\mathbf{z})$ *is*

$$z_i = \sqrt{(\mathbf{H}_{(i,:)}\mathbf{a})^2+\eta^2}, \; i = 1\ldots,h. \tag{2.6}$$

(b) If $(\mathbf{a},\mathbf{z})$ *is an optimal solution of* (2.5)*, then* $\mathbf{a}$ *is an optimal solution of* (2.4).

(c) If $\mathbf{a}$ *is an optimal solution of* (2.4)*, then* $(\mathbf{a},\mathbf{z})$ *is an optimal solution of* (2.5)*, where* $\mathbf{z}$ *is given by* (2.6).

(d) $g_1(\mathbf{a}) = g_2(\mathbf{a},\mathbf{z})$ *where* $\mathbf{z}$ *is given by* (2.6).

(e) $\nabla g_1(\mathbf{a}) = \nabla_{\mathbf{a}} g_2(\mathbf{a},\mathbf{z})$*, where* $\nabla_{\mathbf{a}}$ *is the gradient with respect to* $\mathbf{a}$ *and* $\mathbf{z}$ *is given by* (2.6).

3. The alternating minimization scheme

3.1. The AMQ method

We suggest employing an alternating minimization scheme for problem (2.3), involving sequential minimization with respect to $(\mathbf{z}, \mathbf{w})$ and $(\mathbf{a}, \mathbf{b})$ alternately, repeatedly employing the following two steps:

$$\text{(A)}\ \mathbf{z}^{(k)},\ \mathbf{w}^{(k)}\ \leftarrow\ \underset{\mathbf{z},\mathbf{w}}{\operatorname{argmin}}\Big\{f(\mathbf{a}^{(k)}, \mathbf{b}^{(k)}, \mathbf{z}, \mathbf{w})\ :\ \mathbf{z},\ \mathbf{w} > 0\Big\};$$

$$\text{(B)}\ \mathbf{a}^{(k+1)},\ \mathbf{b}^{(k+1)}\ \in\ \underset{\mathbf{a},\mathbf{b}}{\operatorname{argmin}}\Big\{f(\mathbf{a}, \mathbf{b}, \mathbf{z}^{(k)}, \mathbf{w}^{(k)}) : \|\mathbf{a}\|_2 = 1,\ \|\mathbf{b}\|_2 = 1\Big\}.$$

Step A is straightforward, with the optimal solution stated in Lemma 2.1(a), that is, $z_j^{(k)} = \sqrt{(\mathbf{H}^{\mathbf{a}}_{(j,:)}\mathbf{a}^{(k)})^2 + \eta^2}$ and $w_l^{(k)} = \sqrt{(\mathbf{H}^{\mathbf{b}}_{(l,:)}\mathbf{b}^{(k)})^2 + \eta^2}$ for $j \in [h_1]$ and $l \in [h_2]$. We will rewrite the problem in step B as a quadratically constrained quadratic programming (QCQP) problem. Denote by $\mathbf{D}_{\mathbf{z}} \in \mathbb{R}^{h_1 \times h_1}$ and $\mathbf{E}_{\mathbf{w}} \in \mathbb{R}^{h_2 \times h_2}$ the diagonal matrices:

$$\mathbf{D}_{\mathbf{z}} = \operatorname{diag}\left(\frac{\tau_1}{2}\left(z_1^{-1}, \ldots, z_{h_1}^{-1}\right)\right), \qquad \mathbf{E}_{\mathbf{w}} = \operatorname{diag}\left(\frac{\tau_2}{2}\left(w_1^{-1}, \ldots, w_{h_2}^{-1}\right)\right). \tag{3.1}$$

Step B can now be rewritten as finding an optimal solution of the following QCQP problem:

$$\min_{\mathbf{a},\mathbf{b}}\left\{\begin{pmatrix}\mathbf{a}^T & \mathbf{b}^T\end{pmatrix}\mathbf{Q}^{(k)}\begin{pmatrix}\mathbf{a}\\ \mathbf{b}\end{pmatrix} : \|\mathbf{a}\|_2 = 1, \|\mathbf{b}\|_2 = 1\right\}, \tag{3.2}$$

where

$$\mathbf{Q}^{(k)} = \mathbf{S} + \begin{pmatrix}(\mathbf{H}^{\mathbf{a}})^T\mathbf{D}^{(k)}\mathbf{H}^{\mathbf{a}} & \mathbf{0}\\ \mathbf{0} & (\mathbf{H}^{\mathbf{b}})^T\mathbf{E}^{(k)}\mathbf{H}^{\mathbf{b}}\end{pmatrix}, \tag{3.3}$$

with $\mathbf{D}^{(k)} = \mathbf{D}_{\mathbf{z}^{(k)}}$ and $\mathbf{E}^{(k)} = \mathbf{E}_{\mathbf{w}^{(k)}}$. Note that $\mathbf{D}_{\mathbf{z}^{(k)}}$ and $\mathbf{E}_{\mathbf{z}^{(k)}}$ are well defined since by Lemma 2.1(a) the components of $\mathbf{z}^{(k)}$ and $\mathbf{w}^{(k)}$ are positive. Problem (3.2) is a nonconvex homogeneous QCQP with two quadratic constraints. For this class of problems, it is known that the semidefinite relaxation is tight, in the sense that there always exists an optimal rank-one solution of the relaxed problem. Moreover, such a solution can be recovered from an SDP optimum via a rank-reduction procedure. Consequently, solving the semidefinite relaxation followed by rank reduction yields a globally optimal solution of the original QCQP; see [13–15].

We will refer to the resulting method as the AMQ method as it requires the solution of a QCQP at each iteration of an alternating minimization scheme.

Algorithm 3.1. ***AMQ method for solving*** (2.3)

```
Initialize: $\mathbf{a}^{(0)} \in \mathbb{R}^{d_1}$, $\mathbf{b}^{(0)} \in \mathbb{R}^{d_2}$ with $\|\mathbf{a}^{(0)}\|_2 = \|\mathbf{b}^{(0)}\|_2 = 1$.
for $k = 0, 1, 2, \ldots$ do
    Set $z_j^{(k)} = \sqrt{(\mathbf{H}^{\mathbf{a}}_{(j,:)}\mathbf{a}^{(k)})^2 + \eta^2}$ and $w_l^{(k)} = \sqrt{(\mathbf{H}^{\mathbf{b}}_{(l,:)}\mathbf{b}^{(k)})^2 + \eta^2}$ for $j \in [h_1]$ and $l \in [h_2]$.
    Set $\mathbf{D}^{(k)} = \frac{\tau_1}{2}\operatorname{diag}(\mathbf{z}^{(k)})^{-1}$,   $\mathbf{E}^{(k)} = \frac{\tau_2}{2}\operatorname{diag}(\mathbf{w}^{(k)})^{-1}$.
    Form the matrix
        $$\mathbf{Q}^{(k)} = \mathbf{S} + \begin{pmatrix}(\mathbf{H}^{\mathbf{a}})^T\mathbf{D}^{(k)}\mathbf{H}^{\mathbf{a}} & \mathbf{0}\\ \mathbf{0} & (\mathbf{H}^{\mathbf{b}})^T\mathbf{E}^{(k)}\mathbf{H}^{\mathbf{b}}\end{pmatrix}.  \tag{3.4}$$
```

```
    Solve the QCQP problem
        $\min_{\mathbf{x}} \left\{ \mathbf{x}^T \mathbf{Q}^{(k)} \mathbf{x} : \|\mathbf{x}_{(1:d_1)}\|^2 = 1, \|\mathbf{x}_{(d_1+1:d_1+d_2)}\|^2 = 1 \right\}.$    (3.5)
    Update $\mathbf{a}^{(k+1)} \leftarrow \mathbf{x}_{(1:d_1)}, \mathbf{b}^{(k+1)} \leftarrow \mathbf{x}_{(d_1+1:d_1+d_2)}$.
end for
```

3.2. Convergence

We adopt standard concepts from smooth manifold optimization and refer the reader to [16, 17] for a detailed exposition. Here, we outline only the essentials.

Let $\mathcal{M}$ be a Riemannian submanifold of a linear space $\mathcal{E}$. For all $\mathbf{x} \in \mathcal{M}$, the tangent space [17, Definition 3.7] of $\mathcal{M}$ at $\mathbf{x}$ is given by $\mathrm{T}_{\mathbf{x}}\mathcal{M} = \{c'(0) \mid c : I_{\mathbf{x}} \to \mathcal{M}$ is smooth around 0, $c(0) = \mathbf{x}$ and $I_{\mathbf{x}}$ is an open interval containing $0\}$. Let $f : \mathcal{M} \to \mathbb{R}$ be a smooth function. For any $\mathbf{x} \in \mathcal{M}$, the Riemannian gradient [17, Proposition 3.53] of f on $\mathcal{M}$ is given by $\nabla_{\mathcal{M}} f(\mathbf{x}) \equiv \mathcal{P}_{\mathrm{T}_{\mathbf{x}}\mathcal{M}}(\nabla \bar{f}(\mathbf{x}))$, where $\mathcal{P}_{\mathrm{T}_{\mathbf{x}}\mathcal{M}} : \mathbb{R}^d \to \mathrm{T}_{\mathbf{x}}\mathcal{M}$ is the orthogonal projection onto the tangent space $\mathrm{T}_{\mathbf{x}}\mathcal{M}$, and $\bar{f}$ is any smooth extension of f to a neighborhood of $\mathcal{M}$ in $\mathcal{E}$. A point $\mathbf{x} \in \mathcal{M}$ is called a stationary point of the problem

$$\min_{\mathbf{x} \in \mathcal{M}} f(\mathbf{x})$$

if $\nabla_{\mathcal{M}} f(\mathbf{x}) = \mathbf{0}$. By [17, Proposition 4.4], if $\mathbf{x}$ is a local minimizer of $\min_{\mathbf{x} \in \mathcal{M}} f(\mathbf{x})$ then $\nabla_{\mathcal{M}} f(\mathbf{x}) = \mathbf{0}$.

For embedded submanifolds $\mathcal{M}_1 \subseteq \mathbb{R}^{d_1}$ and $\mathcal{M}_2 \subseteq \mathbb{R}^{d_2}$, the product $\mathcal{M}_1 \times \mathcal{M}_2$ is also an embedded submanifold of $\mathbb{R}^{d_1+d_2}$. Moreover, its tangent space at any point $(\mathbf{x}_1, \mathbf{x}_2)$ satisfies $\mathrm{T}_{(\mathbf{x}_1,\mathbf{x}_2)}(\mathcal{M}_1 \times \mathcal{M}_2) = \mathrm{T}_{\mathbf{x}_1}\mathcal{M}_1 \times \mathrm{T}_{\mathbf{x}_2}\mathcal{M}_2$ [17, Proposition 3.14].

The tangent space of the sphere manifold $\mathcal{M} = \{\mathbf{x} : \|\mathbf{x}\|_2 = 1\}$ is $\mathrm{T}_{\mathbf{x}}\mathcal{M} = \{\mathbf{v} : \mathbf{x}^T\mathbf{v} = 0\}$, and the projection onto the tangent space is given by $\mathcal{P}_{\mathrm{T}_{\mathbf{x}}\mathcal{M}}(\mathbf{u}) = \mathbf{u} - \langle \mathbf{x}, \mathbf{u} \rangle \mathbf{x}$ ([17, Section 7.2]).

3.2.1. A sufficient decrease property

Let $\mathcal{M}^S$ be the feasible set of problem (2.3), meaning the Cartesian product $\mathcal{M}^S \equiv \mathcal{M}_1^S \times \mathcal{M}_2^S$ where $\mathcal{M}_1^S \equiv \{\mathbf{x}_1 : \|\mathbf{x}_1\| = 1, \mathbf{x}_1 \in \mathbb{R}^{d_1}\}$ and $\mathcal{M}_2^S \equiv \{\mathbf{x}_2 : \|\mathbf{x}_2\| = 1, \mathbf{x}_2 \in \mathbb{R}^{d_2}\}$. In this section, the division of a vector $\mathbf{x} \in \mathbb{R}^{d_1+d_2}$ into its sub-elements is denoted by $\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}$, where $\mathbf{x}_1 \in \mathbb{R}^{d_1}$ and $\mathbf{x}_2 \in \mathbb{R}^{d_2}$. The tangent space of $\mathbf{x} \in \mathcal{M}^S$ (meaning $\mathbf{x}_i \in \mathcal{M}_i^S$ for $i = 1, 2$) is

$$\mathrm{T}_{\mathbf{x}}\mathcal{M}^S = \left\{ \mathbf{v} \in \mathbb{R}^{d_1+d_2} : \langle \mathbf{x}_1, \mathbf{v}_1 \rangle = 0,\ \langle \mathbf{x}_2, \mathbf{v}_2 \rangle = 0 \right\},$$

and the orthogonal projection onto the tangent space is thus given by

$$\mathcal{P}_{\mathrm{T}_{\mathbf{x}}\mathcal{M}^S}(\mathbf{u}) = \begin{pmatrix} \mathbf{u}_1 - \langle \mathbf{x}_1, \mathbf{u}_1 \rangle \mathbf{x}_1 \\ \mathbf{u}_2 - \langle \mathbf{x}_2, \mathbf{u}_2 \rangle \mathbf{x}_2 \end{pmatrix}. \tag{3.6}$$

We denote by $\nabla_{\mathcal{M}_i^S} f$ $(i = 1, 2)$ the Riemannian gradient with respect to the ith manifold given by $\nabla_{\mathcal{M}_i^S} f(\mathbf{x}) = P_{\mathrm{T}_{\mathbf{x}_i}\mathcal{M}_i^S}((\nabla f(\mathbf{x}))_i) = (P_{\mathrm{T}_{\mathbf{x}}\mathcal{M}^S}(\nabla f(\mathbf{x})))_i$. We define the curve $c_{\mathbf{x},\mathbf{v}}(t)$ by

$$c_{\mathbf{x},\mathbf{v}}(t) = \frac{\mathbf{x} + t\mathbf{v}}{\sqrt{1 + t^2}}, \tag{3.7}$$

for $\mathbf{x} \in \mathcal{M}^S$ and $\mathbf{v} \in V_{\mathbf{x}}^{eq}$, with

$$V_{\mathbf{x}}^{eq} \equiv \{\mathbf{v} \in \mathrm{T}_{\mathbf{x}}\mathcal{M}^S : \|\mathbf{v}_i\| = 1 \text{ for } i = 1, 2\}. \tag{3.8}$$

Our objective in this section is to establish a sufficient decrease property of the following problem:

$$\min_{\mathbf{x}\in\mathcal{M}^S} \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x}, \tag{3.9}$$

where $\mathbf{A} \in \mathbb{S}^{d_1+d_2}$. Note that a problem of the form (3.9) is solved at each iteration of the AMQ method, see problem (3.2).

The next lemma provides a descent property for problem (3.9).

Lemma 3.1. *Let $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x}$ for $\mathbf{A} \in \mathbb{S}^{d_1+d_2}$. Given the curve $c_{\mathbf{x},\mathbf{v}}$ from* (3.7) *and $V_{\mathbf{x}}^{eq}$ as in* (3.8)*, it holds that for any $t \in \mathbb{R}$, $\mathbf{x} \in \mathcal{M}^S$, and $\mathbf{v} \in V_{\mathbf{x}}^{eq}$, the following inequality holds,*

$$f\left(c_{\mathbf{x},\mathbf{v}}(t)\right) \le f(\mathbf{x}) + \langle\nabla_{\mathcal{M}^S} f(\mathbf{x}), \mathbf{v}\rangle t + Lt^2, \tag{3.10}$$

where $L = \lambda_{\max}(\mathbf{A}) - \min\{\lambda_{\min}(\mathbf{A}), 0\}$.

Proof. Note that $c_{\mathbf{x},\mathbf{v}}(t) \in \mathcal{M}^S$, as $\frac{(\mathbf{x}_i+t\mathbf{v}_i)^T}{\sqrt{1+t^2}}\frac{(\mathbf{x}_i+t\mathbf{v}_i)}{\sqrt{1+t^2}} = 1$ for $i = 1, 2$. We will show that for any $\mathbf{u} \in \mathbb{R}^{d_1+d_2}$ and $\mathbf{s} \in \mathrm{T}_{\mathbf{x}}\mathcal{M}^S$, it holds that $\langle\mathcal{P}_{\mathrm{T}_{\mathbf{x}}\mathcal{M}^S}(\mathbf{u}), \mathbf{s}\rangle = \langle\mathbf{u}, \mathbf{s}\rangle$. Indeed,

$$\begin{aligned}
\langle\mathcal{P}_{\mathrm{T}_{\mathbf{x}}\mathcal{M}^S}(\mathbf{u}), \mathbf{s}\rangle &= \textstyle\sum_{i=1}^2\langle\mathcal{P}_{\mathrm{T}_{\mathbf{x}_i}\mathcal{M}_i^S}(\mathbf{u}_i), \mathbf{s}_i\rangle \\
&= \textstyle\sum_{i=1}^2\langle\mathbf{u}_i - \langle\mathbf{x}_i, \mathbf{u}_i\rangle\,\mathbf{x}_i, \mathbf{s}_i\rangle \\
&= \textstyle\sum_{i=1}^2\langle\mathbf{u}_i, \mathbf{s}_i\rangle \\
&= \langle\mathbf{u}, \mathbf{s}\rangle,
\end{aligned}$$

where the second equality is by (3.6) and the third by the fact that $\langle\mathbf{x}_i, \mathbf{s}_i\rangle = 0$ as $\mathbf{s} \in \mathrm{T}_{\mathbf{x}}\mathcal{M}^S$. We can thus conclude that

$$\langle\nabla_{\mathcal{M}^S} f(\mathbf{x}), \mathbf{v}\rangle = \langle\mathcal{P}_{\mathrm{T}_{\mathbf{x}}\mathcal{M}^S}\left(\nabla f(\mathbf{x})\right), \mathbf{v}\rangle = \langle\nabla f(\mathbf{x}), \mathbf{v}\rangle = \mathbf{x}^T\mathbf{A}\mathbf{v}. \tag{3.11}$$

Set $\rho = \min\{\lambda_{\min}(\mathbf{A}), 0\}$ and define $\hat{\mathbf{A}} \equiv \mathbf{A} - \rho\mathbf{I}$. Then, $\hat{\mathbf{A}}$ is positive semidefinite. For the function $\hat{f}(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\hat{\mathbf{A}}\mathbf{x}$, the following holds,

$$\begin{aligned}
\hat{f}\left(c_{\mathbf{x},\mathbf{v}}(t)\right) &= \frac{1}{2(1+t^2)}(\mathbf{x}+t\mathbf{v})^T\hat{\mathbf{A}}(\mathbf{x}+t\mathbf{v}) \\
&\le \frac{1}{2}(\mathbf{x}+t\mathbf{v})^T\hat{\mathbf{A}}(\mathbf{x}+t\mathbf{v}) = \frac{1}{2}\mathbf{x}^T\hat{\mathbf{A}}\mathbf{x} + t\mathbf{v}^T\hat{\mathbf{A}}\mathbf{x} + \frac{t^2}{2}\mathbf{v}^T\hat{\mathbf{A}}\mathbf{v} \\
&= \hat{f}(\mathbf{x}) + t\langle\nabla_{\mathcal{M}^S}\hat{f}(\mathbf{x}), \mathbf{v}\rangle + \frac{t^2}{2}\mathbf{v}^T\hat{\mathbf{A}}\mathbf{v},
\end{aligned} \tag{3.12}$$

where the first inequality follows from the fact that $\hat{\mathbf{A}}$ is positive semidefinite and $t^2 \ge 0$, and the equality in the third row follows by (3.11) (with $\mathbf{A}$ replaced by $\hat{\mathbf{A}}$). The gradient of $\hat{f}$ is given by

$$\nabla\hat{f}(\mathbf{x}) = \nabla f(\mathbf{x}) - \rho\mathbf{x}.$$

Since $\mathcal{P}_{\mathrm{T}_{\mathbf{x}}\mathcal{M}^S}$ is a linear operator, $\mathcal{P}_{\mathrm{T}_{\mathbf{x}}\mathcal{M}^S}(\nabla\hat{f}(\mathbf{x})) = \mathcal{P}_{\mathrm{T}_{\mathbf{x}}\mathcal{M}^S}(\nabla f(\mathbf{x})) + \mathcal{P}_{\mathrm{T}_{\mathbf{x}}\mathcal{M}^S}(-\rho\mathbf{x})$. It holds that $\mathcal{P}_{\mathrm{T}_{\mathbf{x}}\mathcal{M}^S}(-\rho\mathbf{x}) = \mathbf{0}$ as the following is satisfied for $i = 1, 2$:

$$\mathcal{P}_{\mathrm{T}_{\mathbf{x}_i}\mathcal{M}_i^S}(-\rho\mathbf{x}_i) \overset{(3.6)}{=} -\rho\mathbf{x}_i + \rho\|\mathbf{x}_i\|^2\mathbf{x}_i \overset{\|\mathbf{x}_i\|=1}{=} \mathbf{0}.$$

Therefore,

$$\nabla_{\mathcal{M}^S}\hat{f}(\mathbf{x}) = \mathcal{P}_{\mathrm{T}_{\mathbf{x}}\mathcal{M}^S}(\nabla\hat{f}(\mathbf{x})) = \mathcal{P}_{\mathrm{T}_{\mathbf{x}}\mathcal{M}^S}(\nabla f(\mathbf{x})) = \nabla_{\mathcal{M}^S} f(\mathbf{x}). \tag{3.13}$$

For all $\mathbf{x} \in \mathcal{M}^S$,

$$\hat{f}(\mathbf{x}) = f(\mathbf{x}) - \rho. \tag{3.14}$$

Plugging the equalities (3.13) and (3.14) into (3.12), we finally obtain that

$$\begin{aligned}
f(c_{\mathbf{x},\mathbf{v}}(t)) - \rho &= \hat{f}(c_{\mathbf{x},\mathbf{v}}(t)) \\
&\le \hat{f}(\mathbf{x}) + t\,\langle\nabla_{\mathcal{M}^S}\hat{f}(\mathbf{x}), \mathbf{v}\rangle + \tfrac{t^2}{2}\,\mathbf{v}^T\hat{\mathbf{A}}\,\mathbf{v} \\
&= f(\mathbf{x}) - \rho + t\,\langle\nabla_{\mathcal{M}^S} f(\mathbf{x}), \mathbf{v}\rangle + \tfrac{t^2}{2}\,(\mathbf{v}^T\mathbf{A}\mathbf{v} - \rho\,\|\mathbf{v}\|^2) \\
&\le f(\mathbf{x}) - \rho + t\,\langle\nabla_{\mathcal{M}^S} f(\mathbf{x}), \mathbf{v}\rangle + \tfrac{t^2}{2}\left(\lambda_{\max}(\mathbf{A})\,\|\mathbf{v}\|^2 - \rho\,\|\mathbf{v}\|^2\right) \\
&= f(\mathbf{x}) - \rho + t\,\langle\nabla_{\mathcal{M}^S} f(\mathbf{x}), \mathbf{v}\rangle + t^2\,(\lambda_{\max}(\mathbf{A}) - \rho),
\end{aligned}$$

where in the last equality we used the fact that $\|\mathbf{v}\|^2 = \|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2 = 1 + 1 = 2$. This completes the proof. □

We now introduce a Riemannian gradient step for problem (3.9). To do so, we utilize the following functions:

- The function $g : \mathcal{M}^S \to \mathbb{R}$ is defined as

$$g(\mathbf{x}) = \sum_{i=1}^{2} \|\nabla_{\mathcal{M}_i^S} f\,(\mathbf{x})\,\|. \tag{3.15}$$

- For each $i = 1, 2$, the function $u_i : \mathcal{M}^S \to \mathrm{T}_{\mathbf{x}_i}\mathcal{M}_i^S$ is defined as

$$u_i(\mathbf{x}) = \begin{cases} \frac{\mathbf{t}}{\|\mathbf{t}\|_2} \text{ for some arbitrary } \mathbf{0} \neq \mathbf{t} \in \mathrm{T}_{\mathbf{x}_i}\mathcal{M}_i^S, & \text{if } \nabla_{\mathcal{M}_i^S} f(\mathbf{x}) = \mathbf{0}, \\ -\frac{\nabla_{\mathcal{M}_i^S} f(\mathbf{x})}{\|\nabla_{\mathcal{M}_i^S} f(\mathbf{x})\|}, & \text{otherwise.} \end{cases} \tag{3.16}$$

Consider problem (3.9). Let $\mathbf{x} \in \mathcal{M}^S$ and $c_{\mathbf{x},\mathbf{v}}$ be the curve defined in (3.7), g as in (3.15), u_i as in (3.16) for $i = 1, 2$, and L as in Lemma 3.1. We denote $\mathbf{v} = (u_1(\mathbf{x}), u_2(\mathbf{x})) \in \mathrm{T}_{\mathbf{x}}\mathcal{M}^S$. Then, we consider the Riemannian gradient step $\mathbf{x}^+ = c_{\mathbf{x},\mathbf{v}}\left(\frac{1}{2L}g(\mathbf{x})\right)$.

A sufficient decrease lemma for the above Riemannian gradient step is given below.

Lemma 3.2 (sufficient decrease)**.** *Let* $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x}$ *for* $\mathbf{A} \in \mathbb{S}^{d_1+d_2}$ *and* $L = \lambda_{\max}(\mathbf{A}) - \min\{\lambda_{\min}(\mathbf{A}), 0\}$. *Given* $\mathbf{x} \in \mathcal{M}^S$*, let*

$$\mathbf{x}^+ = c_{\mathbf{x},\mathbf{v}}\left(\frac{1}{2L}g(\mathbf{x})\right),$$

where $c_{\mathbf{x},\mathbf{v}}$ *is defined in* (3.7)*,* g *in* (3.15)*,* $\mathbf{v}$ *is defined by* $\mathbf{v}_i = u_i(\mathbf{x})$ *for* $i = 1, 2$ *and* u_i *is defined in (3.16). Then,*

$$f(\mathbf{x}) - f(\mathbf{x}^+) \geq \frac{1}{4L} \|\nabla_{\mathcal{M}^S} f(\mathbf{x})\|^2 .$$

Proof. For any $l > 0$, by Lemma 3.1 and the definition of $\mathbf{v}$, it holds that

$$f(\mathbf{x}) - f\left(c_{\mathbf{x},\mathbf{v}}(l)\right) \geq -\langle \nabla_{\mathcal{M}^S} f(\mathbf{x}), \mathbf{v}\rangle l - Ll^2 = l\left(\sum_{i=1}^{2} \|\nabla_{\mathcal{M}_i^S} f\left(\mathbf{x}\right)\|\right) - Ll^2. \tag{3.17}$$

Plugging $l = \frac{1}{2L} g(\mathbf{x})$ into (3.17), we obtain that

$$f(\mathbf{x}) - f\left(c_{\mathbf{x},\mathbf{v}}\left(\frac{g(\mathbf{x})}{2L}\right)\right) \geq \frac{1}{4L} g(\mathbf{x})^2 = \frac{1}{4L}\left(\sum_{i=1}^{2} \|\nabla_{\mathcal{M}_i^S} f\left(\mathbf{x}\right)\|\right)^2 \geq \frac{1}{4L} \|\nabla_{\mathcal{M}^S} f\left(\mathbf{x}\right)\|^2,$$

which is the desired result. □

3.2.2. Convergence of the AMQ algorithm

Utilizing the sufficient decrease result of Lemma 3.2, we now establish a rate of convergence result of the norm of the Riemannian gradient of the sequence generated by the AMQ method.

Theorem 3.1. *Let* $\{(\mathbf{a}^{(k)}, \mathbf{b}^{(k)}, \mathbf{z}^{(k)}, \mathbf{w}^{(k)})\}_{k\geq 0}$ *be the sequence generated by the AMQ method and let* $\mathbf{x}^{(k)} = (\mathbf{a}^{(k)}, \mathbf{b}^{(k)})$*. Let* F *be the objective function of problem* (2.1)*. Then, the following inequality holds:*

$$\min_{k=0,1,\ldots,N} \|\nabla_{\mathcal{M}^S} F(\mathbf{x}^{(k)})\| \leq \frac{\sqrt{4\tilde{L}(F(\mathbf{x}^{(0)}) - F_{\text{opt}})}}{\sqrt{N+1}}, \tag{3.18}$$

where F_{opt} *is the optimal value of problem* (2.1) *and*

$$\tilde{L} \equiv 2\lambda_{\max}(\mathbf{S}) - 2\min\{\lambda_{\min}(\mathbf{S}), 0\} + \frac{1}{\eta} \max\{\tau_1 \lambda_{\max}((\mathbf{H}^{\mathbf{a}})^T \mathbf{H}^{\mathbf{a}}), \tau_2 \lambda_{\max}((\mathbf{H}^{\mathbf{b}})^T \mathbf{H}^{\mathbf{b}})\}.$$

Proof. First, by Weierstrass's theorem [18, Theorem 2.12], an optimal solution of problem (2.1) exists, and by Lemma 2.1 $F_{\text{opt}} = \text{val}(2.3) = \text{val}(2.1)$. Denote the objective function of problem (3.5) by h_k. Applying Lemma 3.2 with $\mathbf{A} = 2\mathbf{Q}^{(k)}$ (defined in (3.4)) and $\mathbf{x} = \mathbf{x}^{(k)} \in \mathcal{M}^S$, it holds that there exists with $\mathbf{x}^+ \in \mathcal{M}^S$ for which

$$h_k(\mathbf{x}^{(k)}) - h_k(\mathbf{x}^+) \geq \frac{1}{4L^{(k)}} \|\nabla_{\mathcal{M}^S} h_k(\mathbf{x}^{(k)})\|^2,$$

where $L^{(k)} = 2\lambda_{\max}(\mathbf{Q}^{(k)}) - 2\min\{\lambda_{\min}(\mathbf{Q}^{(k)}), 0\}$. Since $\mathbf{x}^{(k+1)}$ is the minimizer of h_k over $\mathcal{M}^S$, we can conclude that

$$h_k(\mathbf{x}^{(k)}) - h_k(\mathbf{x}^{(k+1)}) \geq \frac{1}{4L^{(k)}} \|\nabla_{\mathcal{M}^S} h_k(\mathbf{x}^{(k)})\|^2. \tag{3.19}$$

By the definition of $\mathbf{Q}^{(k)}$ given in (3.4) and the fact that $\mathbf{D}^{(k)} \leq \frac{\tau_1}{\eta}\mathbf{I}, \mathbf{E}^{(k)} \leq \frac{\tau_2}{\eta}\mathbf{I}$, it follows that

$$\lambda_{\max}(\mathbf{Q}^{(k)}) \leq \lambda_{\max}(\mathbf{S}) + \frac{1}{2\eta} \max\{\tau_1 \lambda_{\max}((\mathbf{H}^{\mathbf{a}})^T \mathbf{H}^{\mathbf{a}}), \tau_2 \lambda_{\max}((\mathbf{H}^{\mathbf{b}})^T \mathbf{H}^{\mathbf{b}})\}.$$

In addition, since $(\mathbf{H^a})^T\mathbf{D}^{(k)}\mathbf{H^a}, (\mathbf{H^b})^T\mathbf{E}^{(k)}\mathbf{H^b} \geq \mathbf{0}$, it follows that $\lambda_{\min}(\mathbf{Q}^{(k)}) \geq \lambda_{\min}(\mathbf{S})$. Consequently,

$$L^{(k)} \leq \tilde{L} \equiv 2\lambda_{\max}(\mathbf{S}) - 2\min\{\lambda_{\min}(\mathbf{S}), 0\} + \frac{1}{\eta}\max\{\tau_1\lambda_{\max}((\mathbf{H^a})^T\mathbf{H^a}), \tau_2\lambda_{\max}((\mathbf{H^b})^T\mathbf{H^b})\},$$

which, combined with (3.19), implies that

$$h_k(\mathbf{x}^{(k)}) - h_k(\mathbf{x}^{(k+1)}) \geq \frac{1}{4\tilde{L}}\|\nabla_{\mathcal{M}^S} h_k(\mathbf{x}^{(k)})\|^2. \tag{3.20}$$

Denote $\hat{\mathbf{x}}^{(k)} = (\mathbf{x}^{(k)}, \mathbf{z}^{(k)}, \mathbf{w}^{(k)})$, and let $f : \mathcal{M}_1 \times \mathcal{M}_2 \to \mathbb{R}$ be the objective function of problem (2.3) where $\mathcal{M}_1 = \mathcal{M}^S$ and $\mathcal{M}_2 = \mathbb{R}^{h_1} \times \mathbb{R}^{h_2}$. Then, the result (3.20) can be rewritten as ($\nabla_{\mathcal{M}_1}$ and $\nabla_{\mathcal{M}_2}$ denote the Riemannian gradient w.r.t. to the first and second manifolds in the cross product, respectively)

$$f(\mathbf{x}^{(k)}, \mathbf{z}^{(k)}, \mathbf{w}^{(k)}) - f(\mathbf{x}^{(k+1)}, \mathbf{z}^{(k)}, \mathbf{w}^{(k)}) \geq \frac{1}{4\tilde{L}}\|\nabla_{\mathcal{M}_1} f(\mathbf{x}^{(k)}, \mathbf{z}^{(k)}, \mathbf{w}^{(k)})\|^2. \tag{3.21}$$

By the definition of the AMQ method, $(\mathbf{z}^{(k)}, \mathbf{w}^{(k)}) \in \underset{\mathbf{z}\in\mathcal{M}_2}{\operatorname{argmin}} f(\mathbf{x}^{(k)}, \mathbf{z}, \mathbf{w})$. Consequently,

$$\|\nabla_{\mathcal{M}_2} f(\mathbf{x}^{(k)}, \mathbf{z}^{(k)}, \mathbf{w}^{(k)})\| = 0 \tag{3.22}$$

and

$$f(\mathbf{x}^{(k+1)}, \mathbf{z}^{(k)}, \mathbf{w}^{(k)}) - f(\mathbf{x}^{(k+1)}, \mathbf{z}^{(k+1)}, \mathbf{w}^{(k+1)}) \geq 0. \tag{3.23}$$

Thus, for all $k \geq 0$,

$$\begin{aligned}
&f(\mathbf{x}^{(k)}, \mathbf{z}^{(k)}, \mathbf{w}^{(k)}) - f(\mathbf{x}^{(k+1)}, \mathbf{z}^{(k+1)}, \mathbf{w}^{(k+1)})\\
&= f(\mathbf{x}^{(k)}, \mathbf{z}^{(k)}, \mathbf{w}^{(k)}) - f(\mathbf{x}^{(k+1)}, \mathbf{z}^{(k)}, \mathbf{w}^{(k)})\\
&\quad + f(\mathbf{x}^{(k+1)}, \mathbf{z}^{(k)}, \mathbf{w}^{(k)}) - f(\mathbf{x}^{(k+1)}, \mathbf{z}^{(k+1)}, \mathbf{w}^{(k+1)})\\
&\overset{(3.21),(3.23)}{\geq} \frac{1}{4\tilde{L}}\left\|\nabla_{\mathcal{M}_1} f(\mathbf{x}^{(k)}, \mathbf{z}^{(k)}, \mathbf{w}^{(k)})\right\|^2 + 0\\
&\overset{(3.22)}{=} \frac{1}{4\tilde{L}}\left\|\nabla_{\mathcal{M}_1\times\mathcal{M}_2} f(\mathbf{x}^{(k)}, \mathbf{z}^{(k)}, \mathbf{w}^{(k)})\right\|^2.
\end{aligned}$$

Summing the above inequality over $k = 0, 1, \dots, N$, we obtain that

$$f(\mathbf{x}^{(0)}, \mathbf{z}^{(0)}, \mathbf{w}^{(0)}) - f(\mathbf{x}^{(N+1)}, \mathbf{z}^{(N+1)}, \mathbf{w}^{(N+1)}) \geq \frac{1}{4\tilde{L}}\sum_{k=0}^{N}\left\|\nabla_{\mathcal{M}_1\times\mathcal{M}_2} f(\mathbf{x}^{(k)}, \mathbf{z}^{(k)}, \mathbf{w}^{(k)})\right\|^2.$$

Using the facts that $f(\mathbf{x}^{(N+1)}, \mathbf{z}^{(N+1)}, \mathbf{w}^{(N+1)}) \geq F_{\text{opt}}$, $F(\mathbf{x}^{(0)}) = f(\mathbf{x}^{(0)}, \mathbf{z}^{(0)}, \mathbf{w}^{(0)})$ (by Lemma 2.1(d)) and that $\sum_{k=0}^{N} a_k \geq (N+1)\min_{k=0,\dots,N} a_k$ for any $a_0, a_1, \dots, a_N \in \mathbb{R}$, it follows that the following inequality holds:

$$\min_{k=0,1,\dots,N}\|\nabla_{\mathcal{M}_1\times\mathcal{M}_2} f(\mathbf{x}^{(k)}, \mathbf{z}^{(k)}, \mathbf{w}^{(k)})\| \leq \frac{\sqrt{4\tilde{L}(F(\mathbf{x}^{(0)}) - F_{\text{opt}})}}{\sqrt{N+1}} \tag{3.24}$$

holds. Since $\|\nabla_{\mathcal{M}_2} f(\mathbf{x}^{(k)}, \mathbf{z}^{(k)}, \mathbf{w}^{(k)})\| = 0$, it follows that

$$\|\nabla_{\mathcal{M}_1\times\mathcal{M}_2} f(\mathbf{x}^{(k)}, \mathbf{z}^{(k)}, \mathbf{w}^{(k)})\| = \|\nabla_{\mathcal{M}_1} f(\mathbf{x}^{(k)}, \mathbf{z}^{(k)}, \mathbf{w}^{(k)})\| \tag{3.25}$$

By the definition of the Riemannian gradient, we have that

$$\begin{aligned}\nabla_{\mathcal{M}_1} f(\mathbf{x}^{(k)}, \mathbf{z}^{(k)}, \mathbf{w}^{(k)}) &= \mathcal{P}_{\mathcal{M}_1}(\nabla f(\mathbf{x}^{(k)}, \mathbf{z}^{(k)}, \mathbf{w}^{(k)})) \\ &= \mathcal{P}_{\mathcal{M}_1}(\nabla F(\mathbf{x}^{(k)})) \\ &= \nabla_{\mathcal{M}_1} F(\mathbf{x}^{(k)}) \\ &= \nabla_{\mathcal{M}^S} F(\mathbf{x}^{(k)}), \end{aligned} \tag{3.26}$$

where the second equality holds by Lemma 2.1(e). Plugging (3.25) and (3.26) into (3.24), the result (3.18) follows. □

The bound in (3.18) implies that

$$\lim_{N\to\infty} \min_{k=0,\dots,N} \|\nabla_{\mathcal{M}_S} F(x^{(k)})\| = 0.$$

Since $\mathcal{M}_S$ is compact, there exists a subsequence $\{x^{(k_j)}\}$ converging to a point $x^\star \in \mathcal{M}_S$. Moreover, along this subsequence, $\|\nabla_{\mathcal{M}_S} F(x^{(k_j)})\| \to 0$, and by continuity of the Riemannian gradient, it follows that $\nabla_{\mathcal{M}_S} F(x^\star) = 0$. Thus, $x^\star$ is a stationary point of problem (2.1) [17, Definition 4.5]. The vanishing of the Riemannian gradient is equivalent to the satisfaction of the first order Karush–Kuhn–Tucker conditions of the constrained problem; see [17, Exercise 7.8]. In particular, all local minimizers satisfy these conditions.

4. Numerical experiments

We study the empirical performance of the AMQ method by comparing it to a general-purpose Riemannian gradient descent (RGD) method [17, Algorithm 4.1]. Both methods are applied to synthetic instances of the smoothed problem (2.1), which is a differentiable approximation of the original nonsmooth, nonconvex formulation (1.1). Since (2.1) is smooth and constrained to a product of spheres, RGD provides a natural baseline. The RGD method used in our experiments has the following update step:

$$(\mathbf{a}^+, \mathbf{b}^+) = R_{(\mathbf{a},\mathbf{b})}\left(-t\nabla_{\mathcal{M}^S} f^s(\mathbf{a}, \mathbf{b})\right),$$

where $f^s(\mathbf{a}, \mathbf{b})$ is the objective function of problem (2.1) and R is the following retraction onto the spheres manifold $\mathcal{M}_s$:

$$R_{(\mathbf{a},\mathbf{b})}(\mathbf{v}) = \left(\frac{\mathbf{a}+\mathbf{v}_1}{\|\mathbf{a}+\mathbf{v}_1\|}, \quad \frac{\mathbf{b}+\mathbf{v}_2}{\|\mathbf{b}+\mathbf{v}_2\|}\right),$$

with $\mathbf{v} = (\mathbf{v}_1, \mathbf{v}_2)$ being in the tangent space $\mathrm{T}_{(\mathbf{a},\mathbf{b})}\mathcal{M}^S$. The step size t is selected via standard backtracking [17, Section 4.5], initialized with $t = s$ for some $s > 0$ and parameters $\alpha, \beta \in (0, 1)$:

$$\text{while } f^s(\mathbf{a}, \mathbf{b}) - f^s\left(R_{(\mathbf{a},\mathbf{b})}\left(-t\nabla_{\mathcal{M}^S} f^s(\mathbf{a}, \mathbf{b})\right)\right) < \alpha t \left\|\nabla_{\mathcal{M}^S} f^s(\mathbf{a}, \mathbf{b})\right\|^2, \quad \text{set } t \leftarrow \beta t.$$

We compared the performance of AMQ with that of RGD with backtracking parameters $s = 1$, $\alpha = 10^{-4}$, and $\beta = 0.8$. Each QCQP subproblem arising in the AMQ method was solved via a semidefinite relaxation, which was handled using an interior-point SDP method [19] implemented as part of our numerical pipeline. A feasible rank-one solution was then extracted from the SDP solution using a rank-reduction procedure, following standard approaches described in [13–15].

The AMQ method was allowed at most 500 outer iterations, while RGD was allowed up to 10^6 iterations. Both methods terminate earlier if the norm of the difference between two consecutive iterates drops below 10^{-12}. All runs were initialized at the same deterministic point: a vector of ones projected onto the feasible set.

Synthetic instance generation For each pair of structural dimensions $d \in \{10, 20, 30, 40, 80\}$ and $h \in \{10, 30, 50\}$, we set $d_1 = d_2 = d$ and $h_1 = h_2 = h$ in (2.1). We construct a symmetric random matrix $\mathbf{S} \in \mathbb{R}^{2d\times 2d}$ by first generating a matrix $\tilde{\mathbf{S}}$ with entries drawn uniformly from $[-1, 1]$ and then symmetrizing it via $\mathbf{S} = \frac{1}{2}(\tilde{\mathbf{S}} + \tilde{\mathbf{S}}^\top)$. The matrices $\mathbf{H}^\mathbf{a} \in \mathbb{R}^{h\times d}$ and $\mathbf{H}^\mathbf{b} \in \mathbb{R}^{h\times d}$ are also generated with entries sampled uniformly from $[-1, 1]$. For each configuration (d, h), we generate 50 independent random instances.

The regularization parameters in (2.1) are fixed to $\tau_1 = \tau_2 = 1$, and the smoothing parameter is set to $\eta^2 = 10^{-6}$.

All experiments were performed on a workstation equipped with an Intel Xeon Silver 4314 CPU @ 2.40 GHz, with 16 cores and 16 logical processors, using `Python`. All methods were implemented with standard scientific computing packages, including `NumPy` and `SciPy`.

Evaluation methodology For each of the 50 random seeds per (d, h) setting, we record the objective value of problem (2.1) and the cumulative runtime at regular intervals: every $0.1s$ during the first half-second and every $0.5s$ thereafter. These traces are used to compute success rates, time-to-accuracy statistics, and budgeted performance profiles.

We define the *best value* $f^\star$ of a given problem instance as the lowest objective value attained by either AMQ or RGD.

4.1. *Success rates with and without time constraints*

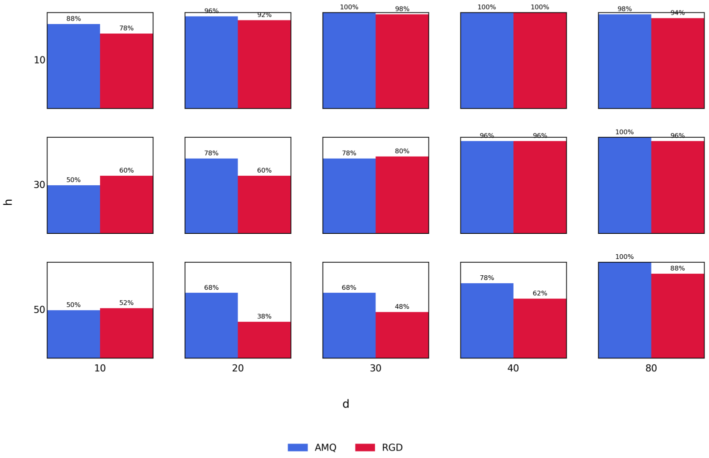


Figure 1. Success rate (no time limit).

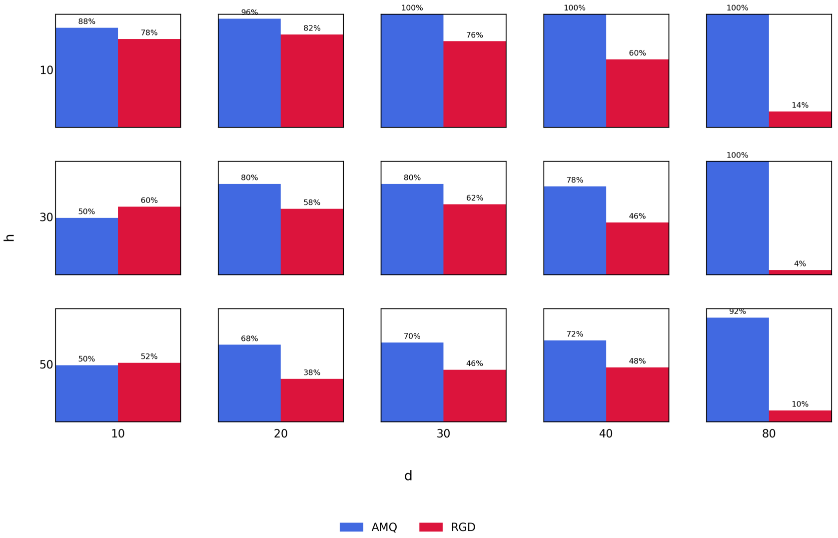


Figure 2. Success rate (30s budget).

Figures 1 and 2 summarize the success rates over 50 random seeds for each (d, h) configuration, under two settings: without any time constraint, and with a 30-second budget. In both cases, a run is counted as successful if its lowest achieved objective value is within 10^{-6} of the *best value* $f^{\star}$ achieved by either method in the corresponding time limit.

In the unlimited-time scenario, in most settings, AMQ (blue) achieves a higher proportion of successful runs than RGD (red), often by a considerable margin. Under the 30-second time constraint, AMQ attains higher or comparable success rates in the majority of configurations.

4.2. Average time to best value

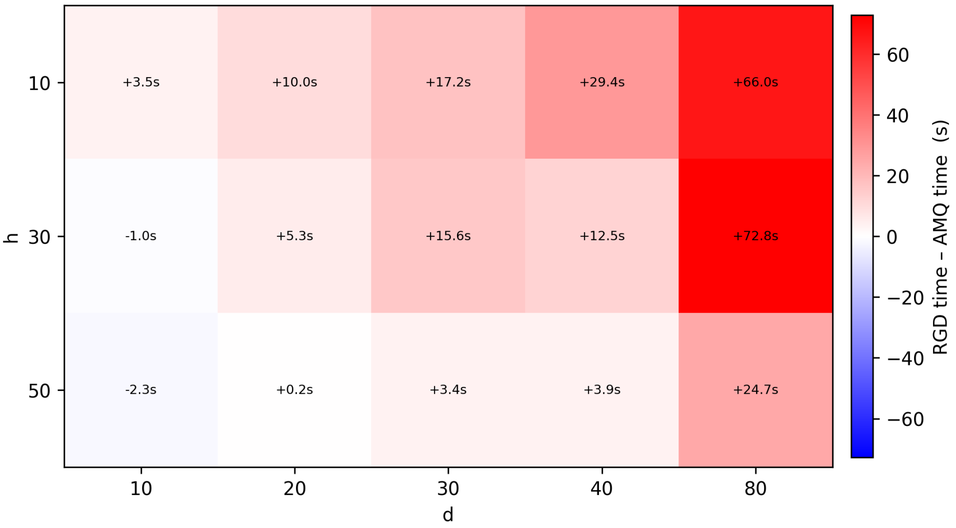
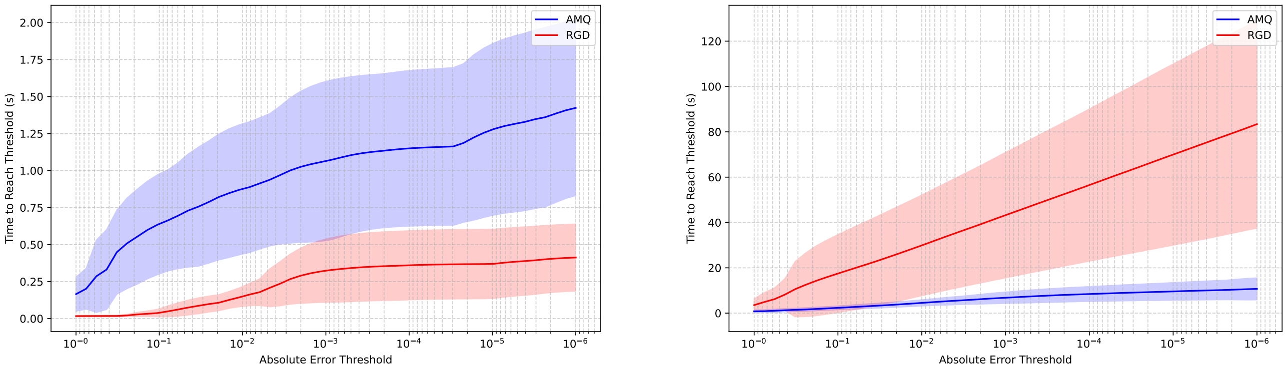


Figure 3. Time difference (RGD minus AMQ). *Mean runtime difference (in seconds) to reach within* 10^{-6} *of the unlimited-time* best value $f^{\star}$.

Figure 3 shows the average time difference (RGD minus AMQ) required to reach the unlimited-budget *best value* $f^\star$ within a tolerance of 10^{-6}, based on seeds where both methods succeeded. Positive (red) values indicate AMQ was faster; negative (blue) values indicate RGD was faster.

AMQ is generally faster across most configurations. Only in the smallest-scale instances (e.g., $[d = 10, h = 50]$ and $[d = 10, h = 30]$) is RGD marginally faster. Overall, AMQ often achieves a substantial increase in speed over RGD when both succeed.

4.3. *Error–time trajectories*



(a) $d = 10,\ h = 30$ **(b)** $d = 80,\ h = 30$

Figure 4. Runtime vs. accuracy. *Mean time (in seconds) to reach target error thresholds* $|f(t) - f^\star|$. *Shaded areas denote one standard deviation.*

Figure 4 illustrates the average runtime required to reach varying levels of target accuracy, measured relative to the unlimited-budget *best value* $f^\star$. The horizontal axis shows the absolute error threshold $|f(t) - f^\star|$, and the vertical axis indicates the average runtime (in seconds) needed to reach that level of accuracy.

Each curve represents the mean over all seeds for which both methods reached within 10^{-6} of $f^\star$; the shaded region indicates one standard deviation.

In the lower-dimensional case $[d = 10, h = 30]$, both methods perform comparably across the full range of thresholds. In contrast, for the case where $[d = 80, h = 30]$, AMQ converges significantly faster.

4.4. *Summary*

The experiments in this section compare AMQ and RGD on synthetic instances of the smoothed problem (2.1), constructed to reflect the general model (1.1). The goal is to assess the behavior of the two methods under controlled synthetic conditions that match the structure of our formulation.

Across most medium and large configurations, AMQ attains the best objective value $f^\star$ more frequently and within a shorter wall-clock time, while RGD is competitive on the smallest instances.

We also observe that AMQ tends to reach a given target accuracy in less time when both methods succeed, especially as the problem dimensions grow. These results indicate that on problems matching the structure of (1.1), solving the alternating subproblems via quadratic relaxations can yield a practical speed advantage over a generic Riemannian gradient scheme.

5. Discussion

This work introduced a unified nonsmooth, nonconvex optimization framework over a product of spheres that encompasses several variants of regularized CCA and SVD. Building on a smooth approximation of this framework, we developed the AMQ algorithm, which alternates between quadratic subproblems, and established a convergence-rate bound for the norms of the Riemannian gradients of the smoothed objective.

The numerical experiments on synthetic instances of the smoothed model support the view that exploiting structured quadratic subproblems can accelerate practical convergence compared with a baseline Riemannian gradient method. At the same time, the experiments were deliberately aligned with the abstract formulation rather than with specific sparse CCA or SVD benchmarks, so the findings should be interpreted as a proof of concept rather than as a comprehensive performance study.

We believe that our general formulation captures structural elements that recur in practical problems over products of spheres, including, but not limited to, the sparse CCA and SVD variants discussed here, and can therefore serve as a useful framework for formulating and studying other models in future applications.

Acknowledgments

This research is supported by the Israel Science Foundation (Grant No. 926/21).

Author contributions

The authors contributed equally to the work.

Use of Generative-AI tools declaration

During the preparation of this manuscript, the authors used ChatGPT (OpenAI) and GitHub Copilot (powered by Claude) for language polishing, expression improvement, and assistance with LaTeX formatting. All AI-assisted content was critically reviewed and revised by the authors, who take full responsibility for the final published version.

Conflict of interest

The authors declare no conflict of interest.

References

1. D. M. Witten, R. Tibshirani, T. Hastie, A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis, *Biostatistics*, **10** (2009), 515–534. https://doi.org/10.1093/biostatistics/kxp008
2. X. Suo, V. Minden, B. Nelson, R. Tibshirani, M. Saunders, Sparse canonical correlation analysis, *arXiv preprint arXiv:1705.10865*, (2017). https://doi.org/10.48550/arXiv.1705.10865
3. Q. Mai, X. Zhang, An iterative penalized least squares approach to sparse canonical correlation analysis, *Biometrics*, **75** (2019), 734–744. https://doi.org/10.1111/biom.13043
4. M. Lee, H. Shen, J. Z. Huang, J. S. Marron, Biclustering via sparse singular value decomposition, *Biometrics*, **66** (2010), 1087–1095. https://doi.org/10.1111/j.1541-0420.2010.01392.x
5. Z. Hong, H. Lian, Sparse-smooth regularized singular value decomposition, *J. Multivariate Anal.*, **117** (2013), 163–174. https://doi.org/10.1016/j.jmva.2013.02.011
6. D. R. Hardoon, S. Szedmak, J. Shawe-Taylor, Canonical correlation analysis: An overview with application to learning methods, *Neural Comput.*, **16** (2004), 2639–2664. https://doi.org/10.1162/0899766042321814
7. D. P. Bertsekas, J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Prentice Hall, 1989.
8. J. M. Ortega, W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.
9. S. J. Wright, Coordinate descent algorithms, *Math. Program.*, **151** (2015), 3–34. https://doi.org/10.1007/s10107-015-0892-3
10. A. Beck, M. Teboulle, Smoothing and first order methods: a unified framework, *SIAM J. Optim.*, **22** (2012), 557–580. https://doi.org/10.1137/100818327
11. I. Daubechies, R. DeVore, M. Fornasier, C. S. Güntürk, Iteratively reweighted least squares minimization for sparse recovery, *Comm. Pure Appl. Math.*, **63** (2010), 1–38. https://doi.org/10.1002/cpa.20303
12. A. Beck, On the convergence of alternating minimization for convex programming with applications to iteratively reweighted least squares and decomposition schemes, *SIAM J. Optim.*, **25** (2015), 185–209. https://doi.org/10.1137/13094829X
13. A. Beck, Quadratic matrix programming, *SIAM J. Optim.*, **17** (2007), 1224–1238. https://doi.org/10.1137/05064816X
14. G. Pataki, The geometry of semidefinite programming, *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*, Kluwer Academic Publishers, 2000, 29–65. https://doi.org/10.1007/978-1-4615-4381-7_3
15. Y. Ye, S. Zhang, New results on quadratic minimization, *SIAM J. Optim.*, **14** (2003), 245–267. https://doi.org/10.1137/S105262340139001X

16. P. A. Absil, R. Mahony, R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, 2008. https://doi.org/10.1515/9781400830244

17. N. Boumal, *An Introduction to Optimization on Smooth Manifolds*, Cambridge University Press, 2023. https://doi.org/10.1017/9781009166164

18. A. Beck, *First-Order Methods in Optimization*, SIAM, 2017. https://doi.org/10.1137/1.9781611974997

19. C. Helmberg, F. Rendl, R. J. Vanderbei, H. Wolkowicz, An interior-point method for semidefinite programming, *SIAM J. Optim.*, **6** (1996), 342–361. https://doi.org/10.1137/0806020

AIMS Press