Standard Bi-Quadratic Optimization Problems and Unconstrained Polynomial Reformulations

Immanuel M. Bomze[*], Chen Ling[†], Liqun Qi[‡] and Xinzhen Zhang[§]

August 8, 2009

Abstract. A so-called Standard Bi-Quadratic Optimization Problem (StBQP) consists in minimizing a bi-quadratic form over the Cartesian product of two simplices (so this is different from a Bi-Standard QP where a quadratic function is minimized over the same set). An application example arises in portfolio selection. In this paper we present a bi-quartic formulation of StBQP, in order to get rid of the sign constraints. We study the first and second-order optimality conditions of the original StBQP and the reformulated bi-quartic problem over the product of two Euclidean spheres. Furthermore, we discuss the one-to-one correspondence between the global/local solutions of StBQP and the global/local solutions of the reformulation. We introduce a continuously differentiable penalty function. Based upon this, the original problem is converted into the problem of locating an unconstrained global minimizer of a (specially structured) polynomial of degree eight.

Key Words. Polynomial optimization, standard simplex, bi-quartic optimization, optimality conditions, penalty function.

[*]University of Vienna, Austria. E-mail: *immanuel.bomze@univie.ac.at*.

[†]School of Mathematics and Statistics, Zhejiang University of Finance and Economics, Hangzhou, 310018, China. E-mail: *linghz@163.com*. Current address: Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong. His work was supported by the National Natural Science Foundation of China (10871168), the Zhejiang Provincial National Science Foundation of China (Y606168) and a Hong Kong Polytechnic University Postdoctoral Fellowship.

[‡]Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong. E-mail: *maqilq@polyu.edu.hk*. His work is supported by the Hong Kong Research Grant Council (Projects: PolyU 5019/09P and PolyU 5018/08P).

[§]Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong. E-mail: *xzzhang@yahoo.cn*.

1 Introduction

In this paper, we consider a bi-quadratic optimization problem of the form

$$\min \left\{ p(\mathbf{x},\mathbf{y}) := \sum_{i,k=1}^{n} \sum_{j,l=1}^{m} a_{ijkl} x_i y_j x_k y_l : (\mathbf{x},\mathbf{y}) \in \Delta_n \times \Delta_m \right\}, \tag{1.1}$$

where

$$\Delta_d := \left\{ \mathbf{x} \in \mathbb{R}^d_+ : \sum_{i=1}^{d} x_i = 1 \right\}$$

is the standard simplex and $\mathbb{R}^d_+ = \{\mathbf{x} \in \mathbb{R}^d : \mathbf{x} \geq \mathbf{o}\}$ denotes the non-negative orthant in d-dimensional Euclidean space $\mathbb{R}^d$. Without loss of generality, we assume the coefficients a_{ijkl} in (1.1) satisfy the following symmetric property:

$$a_{ijkl} = a_{kjil} = a_{ilkj} \quad \text{for } \; i,k = 1,\cdots,n \; \text{ and } j,l = 1,\cdots,m\,.$$

In case that all a_{ijkl} are independent of the indices j and l, i.e., $a_{ijkl} = b_{ik}$ for every $i,k = 1,\cdots,n$, then the original problem (1.1) reduces to the following Standard Quadratic Optimization Problem (StQP)

$$\min \left\{ \sum_{i,k=1}^{n} b_{ik} x_i x_k : \mathbf{x} \in \Delta_n \right\}, \tag{1.2}$$

which is known to be NP-hard. StQPs of the form (1.2) are well studied and occur frequently as subproblems in escape procedures for general quadratic optimization, but also have manifold direct applications, e.g., in portfolio selection and in the maximum weight clique problem for undirected graphs. For details, see e.g. [3, 4, 14, 15, 16] and references therein.

On the other hand, if we fix $\mathbf{x} \in \mathbb{R}^n$ in (1.1), then we arrive at a StQP

$$\min \left\{ \mathbf{y}^\top Q(\mathbf{x}) \mathbf{y} : \mathbf{y} \in \Delta_m \right\}, \tag{1.3}$$

where $Q(\mathbf{x}) = \left[\sum_{i,k=1}^{n} a_{ijkl} x_i x_k\right]_{1\leq j,l\leq m}$ is a symmetric, possibly indefinite $m \times m$ matrix. Similarly, if we fix $\mathbf{y} \in \mathbb{R}^m$, then we have a StQP

$$\min \left\{ \mathbf{x}^\top R(\mathbf{y}) \mathbf{x} : \mathbf{x} \in \Delta_n \right\}, \tag{1.4}$$

where $R(\mathbf{y}) = \left[\sum_{j,l=1}^{m} a_{ijkl} y_j y_l\right]_{1\leq i,k\leq n}$ is a symmetric $n \times n$ matrix. Since problem (1.1) is so closely related to standard quadratic optimization, we call it a *Standard Bi-Quadratic Optimization Problem*, or a *Standard Bi-Quadratic Program (StBQP)*.

Note that the StBQP (1.1) is different from bi-quadratic optimization problems over unit spheres in [12, 22]. The latter problem arises from the strong ellipticity condition problem in solid mechanics and the entanglement problem in quantum physics; see [7, 8, 9, 11, 17, 18, 21] and the references therein. A StBQP should also be not confused with a bi-StQP, which is a special case of a multi-StQP, a problem class studied recently in [6, 19]. In bi-StQPs, the objective is a quadratic form, while the feasible set is a product of simplices, as in (1.1). Both StBQPs and bi-StQPs fall into a larger class investigated by [20]. Since the latter paper deals with general smooth objective functions, while we here make heavy use of the detailed structure of bi-quadraticity, there is no overlap of these two approaches.

Denote $\mathcal{A} := [a_{ijkl}]_{ijkl}$, then $\mathcal{A}$ is a real, partially symmetric $n \times m \times n \times m$-dimensional fourth order tensor. In terms of $\mathcal{A}$, the matrices $Q(\mathbf{x})$ and $R(\mathbf{y})$ can also be written as $\mathcal{A}\mathbf{x}\mathbf{x}^\top$ and $\mathbf{y}\mathbf{y}^\top\mathcal{A}$, respectively. So, it is clear that the objective function in (1.1) can be written briefly as

$$p(\mathbf{x}, \mathbf{y}) = (\mathcal{A}\mathbf{x}\mathbf{x}^\top) \bullet (\mathbf{y}\mathbf{y}^\top) = (\mathbf{y}\mathbf{y}^\top\mathcal{A}) \bullet (\mathbf{x}\mathbf{x}^\top),$$

where $X \bullet Y$ stands for usual Frobenius inner product for matrices, i.e., $X \bullet Y = \mathrm{tr}(X^\top Y)$. Note that the problem of finding minimizers of a non-homogeneous bi-quadratic function $(\mathcal{A}\mathbf{x}\mathbf{x}^\top) \bullet \mathbf{y}\mathbf{y}^\top + \mathbf{x}^\top H\mathbf{y}$ over $\Delta_n \times \Delta_m$ can be easily homogenized by introducing a new fourth order partially symmetric tensor $\bar{\mathcal{A}}$ with $\bar{a}_{ijkl} = a_{ijkl} + (h_{ij} + h_{kj} + h_{il} + h_{kl})/4$, where h_{ij} is the (i, j)th element in H. Indeed, since $\sum_{k=1}^n x_k = 1$ and $\sum_{l=1}^m y_l = 1$, we have

$$\begin{aligned}(\bar{\mathcal{A}}\mathbf{x}\mathbf{x}^\top) \bullet \mathbf{y}\mathbf{y}^\top &= (\mathcal{A}\mathbf{x}\mathbf{x}^\top) \bullet \mathbf{y}\mathbf{y}^\top + \frac{1}{4}\sum_{i,k=1}^{n}\sum_{j,l=1}^{m}(h_{ij} + h_{kj} + h_{il} + h_{kl})x_i y_j x_k y_l \\ &= (\mathcal{A}\mathbf{x}\mathbf{x}^\top) \bullet \mathbf{y}\mathbf{y}^\top + \sum_{i=1}^{n}\sum_{j=1}^{m} h_{ij} x_i y_j.\end{aligned}$$

Furthermore, it is easy to verify that the global/local solutions of (1.1) remain the same if $\mathcal{A}$ is replaced with $\mathcal{A} + \gamma\mathcal{E}$, where γ is an arbitrary constant and $\mathcal{E}$ is the all-ones tensor with the same structure as $\mathcal{A}$. So, without loss of generality, we assume henceforth that all entries of $\mathcal{A}$ are negative.

For the above-mentioned reason, the considered problem (1.1) is NP-hard. Therefore, designing some efficient algorithms for finding approximative solutions and bounds on the optimal value of (1.1) are of interest. In order to get rid of the sign constraints $\mathbf{x} \geq \mathbf{o}$ and $\mathbf{y} \geq \mathbf{o}$, however, in this paper we focus attention on studying a bi-quartic formulation of (1.1) and some properties related to this reformulation.

Our paper is organized as follows. After motivating our study by an application example in

portfolio selection in Section 2, we first study the first and second-order optimality conditions of the original problem and the related bi-quartic optimization problem in Sections 3 and 4. In Section 5 we discuss the one-to-one correspondence between the global/local solutions of (1.1) and the global/local solutions of the reformulation. The obtained results show that the bi-quartic formulation is exactly equivalent to the original problem (1.1). Furthermore, we present in Section 6 a continuously differentiable penalty function, by which we convert the problem of locating a local/global minimizer of the constrained bi-quartic program into the problem of locating a local/global solution to an unconstrained optimization problem. This yields a method for finding second-order KKT points of the formulated bi-quartic optimization problem.

Some words about notation. The j-th component of a column vector $\mathbf{x} \in \mathbb{R}^n$ is denoted by x_j while the (i,j)-th entry of a real $m \times n$ matrix $A \in \mathbb{R}^{m\times n}$ is denoted by A_{ij}. For any matrix A and a fourth order tensor $\mathcal{A}$, respectively, $\|A\|_F$ and $\|\mathcal{A}\|_F$ denote the Frobenius norm of A and $\mathcal{A}$, respectively, i.e.,

$$\|A\|_F = \Big(\mathrm{tr}(A^\top A)\Big)^{1/2} \quad \text{and} \quad \|\mathcal{A}\|_F = \sqrt{\sum_{i,k=1}^{n}\sum_{j,l=1}^{m} a_{ijkl}^2},$$

where $\mathrm{tr}(\cdot)$ denotes the trace of a matrix. $\mathcal{S}^n$ denotes the space of real symmetric $n \times n$ matrices. For $A \in \mathcal{S}^n$, $A \succeq 0$ (resp. $A \succ 0$) means that A is positive-semidefinite (resp. positive definite). $\mathcal{S}^n_+$ denotes the cone of positive-semidefinite matrices in $\mathcal{S}^n$. I_n stands for the $n \times n$ identity matrix and $\mathbf{e}_k$ stands for its k-th column, while $\mathbf{o}$ or $\mathbf{e}$ denote generic vectors of zeroes or ones, respectively, of a size suitable to the context. Also, the sign $^\top$ denotes transpose. Finally, given the numbers $z_1, \ldots, z_n$, we denote by $\mathrm{Diag}(z_1, \cdots, z_n) \in \mathcal{S}^n$ the $n \times n$ diagonal matrix containing z_i in its diagonal.

2 Motivation: application in portfolio selection

According to Markowitz's well-known mean-variance model [14], the general single-period portfolio selection problem can be formulated as a parametric convex quadratic program. As an application example of the bi-quadratic program (1.1), we present a slightly more involved mean-variance model in portfolio selection problems, which can be converted into a bi-quadratic optimization problem.

We consider the portfolio selection problem in two groups of securities, where investment decisions have an influence on each other. Assume that the groups consist of N and M securities,

respectively. For the first group of securities, denote by $R_i^{(1)}$ the discounted return of the i-th security ($i = 1, \cdots, N$), and assume that it is independent of the relative amount x_i invested in the i-th security, but dependent on the amount y_j invested in the j-th security of the second group of security. Let $R_i^{(1)} = \xi_i^0 + \xi_{i1}y_1 + \cdots + \xi_{iM}y_M$ ($i = 1, \cdots, N$), where ξ_i^0 is an random variable with mean μ_i, and ξ_{ij} ($j = 1, \cdots, M$) are the random variables with mean zero. Here, $\mathbf{y} = [y_1, \cdots, y_M]^\top$ is the vector with y_j being the amount invested in the j-th security of the second group of securities. Then, the return of a portfolio on the first group of securities is a random variable defined by

$$R^{(1)} = \sum_{i=1}^{N} R_i^{(1)} x_i = \sum_{i=1}^{N} \xi_i^0 x_i + \sum_{i=1}^{N}\sum_{j=1}^{M} \xi_{ij} x_i y_j$$

and its expected value is $\mathbb{E}(R^{(1)}) = \boldsymbol{\mu}^\top \mathbf{x}$, where $\boldsymbol{\mu} = [\mu_1, \cdots, \mu_N]^\top$ and $\mathbf{x} = [x_1, \cdots, x_N]^\top$. By similar reasoning, we obtain the return of a portfolio on the second group of securities as

$$R^{(2)} = \sum_{j=1}^{M} R_j^{(2)} y_j = \sum_{j=1}^{M} \gamma_j^0 y_j + \sum_{j=1}^{M}\sum_{i=1}^{N} \gamma_{ji} x_i y_j,$$

where γ_j^0, γ_{ji} ($i = 1, \cdots, N, j = 1, \cdots, M$) are random variables. It is easy to see that the expected value $\mathbb{E}(R^{(2)}) = \boldsymbol{\nu}^\top \mathbf{y}$, where $\boldsymbol{\nu} = [\mathbb{E}(\gamma_1^0), \cdots, \mathbb{E}(\gamma_M^0)]^\top$. It is clear that the total return of the portfolio on the two groups of securities is $R = R^{(1)} + R^{(2)}$. We assume that ξ_i^0, ξ_{ij}, γ_j^0 and γ_{ji} are independent of each other for $i = 1, \cdots, N$ and $j = 1, \cdots, M$. Under this assumption, we know that the variance of R is $\mathbb{V}\mathrm{ar}(R) = \mathbb{V}\mathrm{ar}(R^{(1)}) + \mathbb{V}\mathrm{ar}(R^{(2)})$.

Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be the variance tensors of the random matrices $\Xi = (\xi_{ij})$ and $\Gamma = (\gamma_{ji})$ respectively, and Q_1 and Q_2 be the variance matrices of the random vectors $\boldsymbol{\xi}^0 = [\xi_1^0, \cdots, \xi_N^0]^\top$ and $\boldsymbol{\gamma}^{\mathbf{0}} = [\gamma_1^0, \cdots, \gamma_M^0]^\top$, respectively. We assume that no security may be held in negative quantities, i.e., $x_i \geq 0$ for every $i = 1, \cdots, N$ and $y_j \geq 0$ for every $j = 1, \cdots, M$. Then, given a set of values for the parameter α as well as $\mathcal{B}_1$, $\mathcal{B}_2$, Q_1, Q_2, $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$, a generalized mean-variance model can be expressed by

$$\begin{array}{ll} \min & (\mathcal{B}_1 \mathbf{x}\mathbf{x}^\top) \bullet \mathbf{y}\mathbf{y}^\top + (\mathcal{B}_2 \mathbf{x}\mathbf{x}^\top) \bullet \mathbf{y}\mathbf{y}^\top + \mathbf{x}^\top Q_1 \mathbf{x} + \mathbf{y}^\top Q_2 \mathbf{y} - \alpha \left(\boldsymbol{\mu}^\top \mathbf{x} + \boldsymbol{\nu}^\top \mathbf{y}\right) \\ \text{s.t.} & \displaystyle\sum_{i=1}^{N} x_i = a\,, \ \sum_{j=1}^{M} y_j = b\,, \ (\mathbf{x}, \mathbf{y}) \in \mathbb{R}_+^N \times \mathbb{R}_+^M\,, \end{array}$$

where a and b stand for the total amount invested in the first and the second group of securities, respectively. It is evident that the above model can be rewritten equivalently as the form of (1.1).

3 Optimality conditions for the StBQP

In this section we recall, for ease of reference, the first and second-order necessary optimality conditions of (1.1), which are standard in constrained optimization.

Since the constraints in (1.1) are linear, constraint qualifications are met and the first-order necessary optimality conditions for a feasible point $(\bar{x}, \bar{y})$ to be a local solution to problem (1.1) require that a scalar pair $(\bar{\lambda}, \bar{\mu})$ exists such that

$$\begin{cases} \left[(\bar{\mathbf{y}}\bar{\mathbf{y}}^\top \mathcal{A})\bar{\mathbf{x}}\right]_i + \bar{\lambda} = 0, & \text{for } i \text{ with } \bar{x}_i > 0, \\ \left[(\bar{\mathbf{y}}\bar{\mathbf{y}}^\top \mathcal{A})\bar{\mathbf{x}}\right]_i + \bar{\lambda} \geq 0, & \text{for } i \text{ with } \bar{x}_i = 0, \\ \left[(\mathcal{A}\bar{\mathbf{x}}\bar{\mathbf{x}}^\top)\bar{\mathbf{y}}\right]_j + \bar{\mu} = 0, & \text{for } j \text{ with } \bar{y}_j > 0, \\ \left[(\mathcal{A}\bar{\mathbf{x}}\bar{\mathbf{x}}^\top)\bar{\mathbf{y}}\right]_j + \bar{\mu} \geq 0, & \text{for } j \text{ with } \bar{y}_j = 0. \end{cases} \tag{3.5}$$

By (3.5), it follows that $\bar{\lambda} = \bar{\mu} = -p(\bar{\mathbf{x}}, \bar{\mathbf{y}})$, since $\sum_{i=1}^n \bar{x}_i = 1$ and $\sum_{j=1}^m \bar{y}_j = 1$. In other words, the Lagrange multipliers are uniquely determined by $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$.

Further, it is well-known that the second-order necessary optimality conditions for (1.1) holds, i.e., if $(\bar{x}, \bar{y})$ is a local solution of problem (1.1), then there exists a scalar pair $(\bar{\lambda}, \bar{\mu})$ such that (3.5) holds and furthermore

$$0 \leq \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix}^\top \nabla^2 p(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} \quad \text{for all } [\mathbf{u}^\top, \mathbf{v}^\top]^\top \in \mathcal{T}(\bar{\mathbf{x}}, \bar{\mathbf{y}}), \tag{3.6}$$

where

$$\mathcal{T}(\bar{\mathbf{x}}, \bar{\mathbf{y}}) = \left\{ [\mathbf{u}^\top, \mathbf{v}^\top]^\top \in \mathbb{R}^{n+m} : \sum_{i \in I(\bar{x})} u_i = 0 \text{ and } u_i = 0 \ \forall\ i \notin I(\bar{x}) \ , \right. \\ \left. \sum_{j \in J(\bar{y})} v_j = 0 \text{ and } v_j = 0 \ \forall\ j \notin J(\bar{y}) \right\}$$

with $I(\bar{\mathbf{x}}) = \{i = 1, \cdots, n : \bar{x}_i > 0\}$ and $J(\bar{\mathbf{y}}) = \{j = 1, \cdots, m : \bar{y}_j > 0\}$.

4 Bi-quartic formulation of the StBQP

In this section, we propose a bi-quartic formulation of (1.1) and study its first and second-order necessary optimality conditions. Based upon this, we discuss the one-to-one correspondence between the global/local solutions of (1.1) and the global/local solutions of the formulated bi-quartic optimization problem. Our main technique used here is similar to that developed in [5].

To get rid of the sign constraints $\mathbf{x} \geq \mathbf{o}$ and $\mathbf{y} \geq \mathbf{o}$, we replace the variables x_i and y_j with z_i^2 and w_j^2, respectively. Then the conditions $\sum_{i=1}^n x_i = 1$ and $\sum_{j=1}^m y_j = 1$ become $\|\mathbf{z}\|^2 = 1$ and $\|\mathbf{w}\|^2 = 1$, respectively, where $\|\cdot\|$ denotes the Euclidean norm. Therefore, the original problem (1.1) can be rewritten as

$$\min \left\{ g(\mathbf{z}, \mathbf{w}) := \sum_{i,k=1}^{n} \sum_{j,l=1}^{m} a_{ijkl} z_i^2 w_j^2 z_k^2 w_l^2 : \|\mathbf{z}\|^2 = \|\mathbf{w}\|^2 = 1\,,\ (\mathbf{z}, \mathbf{w}) \in \mathbb{R}^n \times \mathbb{R}^m \right\}. \tag{4.7}$$

Since $a_{ijkl} < 0$ for all i, j, k, l, Problem (4.7) is equivalent to

$$\min \left\{ g(\mathbf{z}, \mathbf{w}) : \|\mathbf{z}\|^2 \leq 1\,,\ \|\mathbf{w}\|^2 \leq 1\,,\ (\mathbf{z}, \mathbf{w}) \in \mathbb{R}^n \times \mathbb{R}^m \right\}. \tag{4.8}$$

4.1 Optimality conditions for the bi-quartic problem

In this subsection, we study the first and second-order optimality conditions of (4.7).

Let

$$B_{jl} = [a_{ijkl}]_{1 \leq i,k \leq n}\ \ (j, l = 1, \cdots, m) \quad \text{and} \quad C_{ik} = [a_{ijkl}]_{1 \leq j,l \leq m}\ \ (i, k = 1, \cdots, n)$$

be $n \times n$ matrices and $m \times m$ matrices, respectively. Let $Z = \mathrm{Diag}(z_1, \cdots, z_n)$ and $W = \mathrm{Diag}(w_1, \cdots, w_m)$. Then the objective function $g(\mathbf{z}, \mathbf{w})$ in (4.7) can be written as

$$g(\mathbf{z}, \mathbf{w}) = \sum_{j,l=1}^{m} \left(\mathbf{z}^\top Z B_{jl} Z \mathbf{z} \right) w_j^2 w_l^2 = \sum_{i,k=1}^{n} \left(\mathbf{w}^\top W C_{ik} W \mathbf{w} \right) z_i^2 z_k^2.$$

Let $B(\mathbf{z}) = \left(\mathbf{z}^\top Z B_{jl} Z \mathbf{z} \right)_{1 \leq j,l \leq m}$ and $C(\mathbf{w}) = \left(\mathbf{w}^\top W C_{ik} W \mathbf{w} \right)_{1 \leq i,k \leq n}$. Then we further have

$$g(\mathbf{z}, \mathbf{w}) = \mathbf{w}^\top W B(\mathbf{z}) W \mathbf{w} = \mathbf{z}^\top Z C(\mathbf{w}) Z \mathbf{z}.$$

Based upon the expression for $g(\mathbf{z}, \mathbf{w})$ above, it follows, by a direct computation, that

$$\nabla_{\mathbf{z}} g(\mathbf{z}, \mathbf{w}) = 4 Z C(\mathbf{w}) Z \mathbf{z} \quad \text{and} \quad \nabla_{\mathbf{w}} g(\mathbf{z}, \mathbf{w}) = 4 W B(\mathbf{z}) W \mathbf{w}\,. \tag{4.9}$$

Hence $\nabla^2_{\mathbf{zz}} g(\mathbf{z}, \mathbf{w}) = 8 Z C(\mathbf{w}) Z + 4\mathrm{Diag}[C(\mathbf{w}) Z \mathbf{z}]$, $\nabla^2_{\mathbf{ww}} g(\mathbf{z}, \mathbf{w}) = 8 W B(\mathbf{z}) W + 4\mathrm{Diag}[B(\mathbf{z}) W \mathbf{w}]$ and

$$\nabla^2_{\mathbf{zw}} g(\mathbf{z}, \mathbf{w}) = 16 \begin{bmatrix} z_1 \sum_{k=1}^n z_k^2 \mathbf{w}^\top W C_{1k} W \\ z_2 \sum_{k=1}^n z_k^2 \mathbf{w}^\top W C_{2k} W \\ \vdots \\ z_n \sum_{k=1}^n z_k^2 \mathbf{w}^\top W C_{nk} W \end{bmatrix}, \tag{4.10}$$

which together form

$$\nabla^2 g(\mathbf{z},\mathbf{w}) = \begin{bmatrix} \nabla^2_{\mathbf{zz}} g(\mathbf{z},\mathbf{w}) & \nabla^2_{\mathbf{zw}} g(\mathbf{z},\mathbf{w}) \\ \left[\nabla^2_{\mathbf{zw}} g(\mathbf{z},\mathbf{w})\right]^\top & \nabla^2_{\mathbf{ww}} g(\mathbf{z},\mathbf{w}) \end{bmatrix}. \tag{4.11}$$

Let $(\bar{\mathbf{z}},\bar{\mathbf{w}})$ be an optimal solution to (4.7). Since constraint qualifications are met, the first-order optimality conditions are necessary, so we know that there exist $\bar\alpha$, $\bar\beta \in \mathbb{R}$ such that

$$\begin{cases} \nabla_{\mathbf{z}} g(\bar{\mathbf{z}},\bar{\mathbf{w}}) + 2\bar\alpha\bar{\mathbf{z}} = \mathbf{o}\,, \\ \nabla_{\mathbf{w}} g(\bar{\mathbf{z}},\bar{\mathbf{w}}) + 2\bar\beta\bar{\mathbf{w}} = \mathbf{o}\,, \end{cases} \tag{4.12}$$

which implies, together with (4.9), that the KKT conditions are equivalent to

$$\begin{cases} 2\bar{Z}C(\bar{\mathbf{w}})\bar{Z}\bar{\mathbf{z}} + \bar\alpha\bar{\mathbf{z}} = \mathbf{o}\,, \\ 2\bar{W}B(\bar{\mathbf{z}})\bar{W}\bar{\mathbf{w}} + \bar\beta\bar{\mathbf{w}} = \mathbf{o}\,. \end{cases} \tag{4.13}$$

From (4.13), it holds that $\bar\alpha = \bar\beta = -2g(\bar{\mathbf{z}},\bar{\mathbf{w}})$. In other words, the Lagrange multipliers $\bar\alpha$ and $\bar\beta$ of (4.7) are uniquely determined by $(\bar{\mathbf{z}},\bar{\mathbf{w}})$.

Now (4.10) implies that $\nabla^2_{\mathbf{zw}} g(\bar{\mathbf{z}},\bar{\mathbf{w}})\bar{\mathbf{w}} = 16\bar{Z}C(\bar{\mathbf{w}})\bar{Z}\bar{\mathbf{z}}$ and $[\nabla^2_{\mathbf{zw}} g(\bar{\mathbf{z}},\bar{\mathbf{w}})]^\top\bar{\mathbf{z}} = 16\bar{W}B(\bar{\mathbf{z}})\bar{W}\bar{\mathbf{w}}$. Hence, by (4.11), we have

$$\nabla^2 g(\bar{\mathbf{z}},\bar{\mathbf{w}}) \begin{bmatrix} \bar{\mathbf{z}} \\ \bar{\mathbf{w}} \end{bmatrix} = 28 \begin{bmatrix} \bar{Z}C(\bar{\mathbf{w}})\bar{Z}\bar{\mathbf{z}} \\ \bar{W}B(\bar{\mathbf{z}})\bar{W}\bar{\mathbf{w}} \end{bmatrix}. \tag{4.14}$$

By this, we know that the first-order optimality condition (4.13) can be rewritten as

$$\left(\nabla^2 g(\bar{\mathbf{z}},\bar{\mathbf{w}}) + 14\bar\alpha I_{n+m}\right) \begin{bmatrix} \bar{\mathbf{z}} \\ \bar{\mathbf{w}} \end{bmatrix} = 0\,. \tag{4.15}$$

It is well-known that the second-order necessary optimality conditions for problem (4.7) involve the Hessian of the Lagrangian (recall that $\bar\alpha = -2g(\bar{\mathbf{z}},\bar{\mathbf{w}}) = \bar\beta$),

$$H_{\bar\alpha,\bar\beta}(\bar{\mathbf{z}},\bar{\mathbf{w}}) = \nabla^2 g(\bar{\mathbf{z}},\bar{\mathbf{w}}) + 2\bar\alpha I_{n+m} =: H_{\bar\alpha}(\bar{\mathbf{z}},\bar{\mathbf{w}})$$

and require in addition to (4.15) that

$$\begin{bmatrix} \mathbf{z} \\ \mathbf{w} \end{bmatrix}^\top H_{\bar\alpha}(\bar{\mathbf{z}},\bar{\mathbf{w}}) \begin{bmatrix} \mathbf{z} \\ \mathbf{w} \end{bmatrix} \geq 0 \quad \text{for all } [\mathbf{z}^\top,\mathbf{w}^\top]^\top \in \mathbb{R}^{n+m} \text{ with } \bar{\mathbf{z}}^\top\mathbf{z} = \bar{\mathbf{w}}^\top\mathbf{w} = 0. \tag{4.16}$$

Based upon the obtained first and second-order necessary optimality conditions of (4.7), we may further study their properties. To this end, in the next subsection we will discuss the case of general bi-homogeneous optimization over the two balls and the two spheres.

4.2 General bi-homogeneous optimization

In this subsection, we consider a general objective function $g(\mathbf{z}, \mathbf{w})$ which is homogeneous of degrees $r_{\mathbf{z}} \geq 2$ and $r_{\mathbf{w}} \geq 2$ with respect to the variables $\mathbf{z}$ and $\mathbf{w}$, respectively, and the problem $\min\{g(\mathbf{z}, \mathbf{w}) : \|\mathbf{z}\|^2 = 1, \|\mathbf{w}\|^2 = 1\}$ (later, we shall specialize to our case $r_{\mathbf{z}} = r_{\mathbf{w}} = 4$). In this case, the Hessian matrix of the Lagrangian becomes

$$H_{\bar{\alpha},\bar{\beta}}(\bar{\mathbf{z}}, \bar{\mathbf{w}}) = \nabla^2 g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) + 2 \begin{bmatrix} \bar{\alpha} I_n & 0 \\ 0 & \bar{\beta} I_m \end{bmatrix}, \tag{4.17}$$

and the second-order necessary optimality condition is

$$0 \leq \begin{bmatrix} \mathbf{z} \\ \mathbf{w} \end{bmatrix}^\top H_{\bar{\alpha},\bar{\beta}}(\bar{\mathbf{z}}, \bar{\mathbf{w}}) \begin{bmatrix} \mathbf{z} \\ \mathbf{w} \end{bmatrix} \quad \text{for all } [\mathbf{z}^\top, \mathbf{w}^\top]^\top \in \mathbb{R}^{n+m} \text{ with } \bar{\mathbf{z}}^\top \mathbf{z} = \bar{\mathbf{w}}^\top \mathbf{w} = 0. \tag{4.18}$$

In the sequel of this subsection, we will study some properties with respect to the first and second-order optimality conditions. From the homogeneity assumption on g, it holds, by Euler's identity, that

$$\nabla_{\mathbf{z}} g(\mathbf{z}, \mathbf{w})^\top \mathbf{z} = r_{\mathbf{z}} g(\mathbf{z}, \mathbf{w}), \quad \nabla_{\mathbf{w}} g(\mathbf{z}, \mathbf{w})^\top \mathbf{w} = r_{\mathbf{w}} g(\mathbf{z}, \mathbf{w}) \tag{4.19}$$

and

$$\left.\begin{aligned} \nabla^2_{\mathbf{zz}} g(\mathbf{z}, \mathbf{w})\mathbf{z} &= (r_{\mathbf{z}} - 1)\nabla_{\mathbf{z}} g(\mathbf{z}, \mathbf{w}), \\ \nabla^2_{\mathbf{ww}} g(\mathbf{z}, \mathbf{w})\mathbf{w} &= (r_{\mathbf{w}} - 1)\nabla_{\mathbf{w}} g(\mathbf{z}, \mathbf{w}). \end{aligned}\right\}$$

On the other hand, cross-differentiating (4.19), it holds that

$$\left.\begin{aligned} \nabla^2_{\mathbf{zw}} g(\mathbf{z}, \mathbf{w})\mathbf{w} &= r_{\mathbf{w}} \nabla_{\mathbf{z}} g(\mathbf{z}, \mathbf{w}), \\ \left[\nabla^2_{\mathbf{zw}} g(\mathbf{z}, \mathbf{w})\right]^\top \mathbf{z} &= r_{\mathbf{z}} \nabla_{\mathbf{w}} g(\mathbf{z}, \mathbf{w}), \end{aligned}\right\} \tag{4.20}$$

which implies

$$\left.\begin{aligned} \nabla^2_{\mathbf{zz}} g(\mathbf{z}, \mathbf{w})\mathbf{z} + \nabla^2_{\mathbf{zw}} g(\mathbf{z}, \mathbf{w})\mathbf{w} &= (r_{\mathbf{z}} + r_{\mathbf{w}} - 1)\nabla_{\mathbf{z}} g(\mathbf{z}, \mathbf{w}), \\ \left[\nabla^2_{\mathbf{zw}} g(\mathbf{z}, \mathbf{w})\right]^\top \mathbf{z} + \nabla^2_{\mathbf{ww}} g(\mathbf{z}, \mathbf{w})\mathbf{w} &= (r_{\mathbf{z}} + r_{\mathbf{w}} - 1)\nabla_{\mathbf{w}} g(\mathbf{z}, \mathbf{w}), \end{aligned}\right\} \tag{4.21}$$

and, together with (4.11), that

$$\nabla^2 g(\mathbf{z}, \mathbf{w}) \begin{bmatrix} \mathbf{z} \\ \mathbf{w} \end{bmatrix} = (r_{\mathbf{z}} + r_{\mathbf{w}} - 1)\nabla g(\mathbf{z}, \mathbf{w}). \tag{4.22}$$

Let $(\bar{\mathbf{z}}, \bar{\mathbf{w}})$ be a local solution to (4.8) with a general bi-homogeneous objective function g. It is easy to see that still constraint qualifications are met, so the KKT condition (4.12) for the

considered problem holds, i.e.,

$$\left.\begin{aligned} \nabla_{\mathbf{z}} g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) + 2\bar{\alpha}\bar{\mathbf{z}} &= \mathbf{o}\,, \\ \nabla_{\mathbf{w}} g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) + 2\bar{\beta}\bar{\mathbf{w}} &= \mathbf{o}\,. \end{aligned}\right\} \tag{4.23}$$

where we in addition know that $\bar{\alpha} \geq 0$ and $\bar{\beta} \geq 0$ as the multipliers of inequality constraints. We first establish a uniqueness result for these multipliers under the stated problem assumptions.

Theorem 4.1 *For any local solution $(\bar{\mathbf{z}}, \bar{\mathbf{w}})$ to (4.8) with a general bi-homogeneous objective function g, the Lagrange multipliers satisfy $\bar{\alpha} = -\frac{r_{\mathbf{z}}}{2} g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) \geq 0$ and $\bar{\beta} = -\frac{r_{\mathbf{w}}}{2} g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) \geq 0$. Hence necessarily $g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) \leq 0$. More precisely, we have either $g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) = 0$ or $g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) < 0$, in which case $\|\mathbf{z}\| = \|\mathbf{w}\| = 1$, i.e., $(\bar{\mathbf{z}}, \bar{\mathbf{w}})$ is feasible for (4.7).*

Proof. We distinguish cases. If $\bar{\mathbf{z}} = \mathbf{o}$ and $\bar{\mathbf{w}} = \mathbf{o}$, then both constraints are not binding and $\bar{\alpha} = \bar{\beta} = 0 = -\frac{r}{2} g(\mathbf{o}, \mathbf{o})$ for any $r > 0$. If $\bar{\mathbf{z}} \neq \mathbf{o}$ but $\bar{\mathbf{w}} = \mathbf{o}$, we infer from (4.23) and (4.19) that $\bar{\alpha} = -\frac{r_{\mathbf{z}}}{2\bar{\mathbf{z}}^\top \bar{\mathbf{z}}} g(\bar{\mathbf{z}}, \mathbf{o})$ holds. However, by homogeneity in $\mathbf{w}$ we also have $g(\bar{\mathbf{z}}, \mathbf{o}) = 0$, so that again $\bar{\alpha} = \bar{\beta} = 0 = -\frac{r}{2} g(\bar{\mathbf{z}}, \mathbf{o})$ for any $r > 0$. The case $\bar{\mathbf{w}} \neq \mathbf{o}$ but $\bar{\mathbf{z}} = \mathbf{o}$ is completely symmetric. So finally we have to deal with $\bar{\mathbf{z}} \neq \mathbf{o}$ and $\bar{\mathbf{w}} \neq \mathbf{o}$. As above, (4.23) and (4.19) imply that the multipliers are uniquely determined and given by $\bar{\alpha} = -\frac{r_{\mathbf{z}}}{2\bar{\mathbf{z}}^\top \bar{\mathbf{z}}} g(\bar{\mathbf{z}}, \bar{\mathbf{w}})$ as well as $\bar{\beta} = -\frac{r_{\mathbf{w}}}{2\bar{\mathbf{w}}^\top \bar{\mathbf{w}}} g(\bar{\mathbf{z}}, \bar{\mathbf{w}})$. Hence we are done if $g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) = 0$, and we only have to prove that $g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) < 0$ implies $\|\bar{\mathbf{z}}\| = \|\bar{\mathbf{w}}\| = 1$. But this is clear again from the homogeneity assumptions on g, studying the behaviour of $g(t\bar{\mathbf{z}}, \bar{\mathbf{w}}) = t^{r_{\mathbf{z}}} g(\bar{\mathbf{z}}, \bar{\mathbf{w}})$ as $t \in \mathbb{R}$ varies around $t = 1$. ∎

By the expression (4.17) for $H_{\bar{\alpha},\bar{\beta}}(\bar{\mathbf{z}}, \bar{\mathbf{w}})$, one can now obtain

$$\begin{aligned} &\left(H_{\bar{\alpha},\bar{\beta}}(\bar{\mathbf{z}}, \bar{\mathbf{w}}) + 2(r_{\mathbf{z}} + r_{\mathbf{w}} - 2) \begin{bmatrix} \bar{\alpha} I_n & 0 \\ 0 & \bar{\beta} I_m \end{bmatrix} \right) \begin{bmatrix} \bar{\mathbf{z}} \\ \bar{\mathbf{w}} \end{bmatrix} \\ &= \left(\nabla^2 g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) + 2(r_{\mathbf{z}} + r_{\mathbf{w}} - 1) \begin{bmatrix} \bar{\alpha} I_n & 0 \\ 0 & \bar{\beta} I_m \end{bmatrix} \right) \begin{bmatrix} \bar{\mathbf{z}} \\ \bar{\mathbf{w}} \end{bmatrix} \\ &= (r_{\mathbf{z}} + r_{\mathbf{w}} - 1) \left(\nabla g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) + 2 \begin{bmatrix} \bar{\alpha}\bar{\mathbf{z}} \\ \bar{\beta}\bar{\mathbf{w}} \end{bmatrix} \right) \\ &= \mathbf{o}\,, \end{aligned}$$

where the last equality is due to (4.23). Hence, unless $[\bar{\mathbf{z}}^\top, \bar{\mathbf{w}}^\top] = \mathbf{o}^\top$, the matrix

$$H_{\bar{\alpha},\bar{\beta}}(\bar{\mathbf{z}}, \bar{\mathbf{w}}) + 2(r_{\mathbf{z}} + r_{\mathbf{w}} - 2) \begin{bmatrix} \bar{\alpha} I_n & 0 \\ 0 & \bar{\beta} I_m \end{bmatrix}$$

is singular.

For the general homogeneous problem with a single ball constraint, a second-order condition has been proven in [5, Theorem 1]. The obtained conclusion establishes positive semidefiniteness of the corresponding matrix; see also [2]. The following theorem extends [5, Theorem 1] to the case of bi-homogeneous optimization over the product of two balls.

Theorem 4.2 *Let $(\bar{\mathbf{z}}, \bar{\mathbf{w}})$ be a local solution to the problem*

$$\min \left\{ g(\mathbf{z}, \mathbf{w}) : \|\mathbf{z}\|^2 \leq 1 \,,\, \|\mathbf{w}\|^2 \leq 1 \right\} \,,$$

where g is homogeneous of degrees $r_{\mathbf{z}} \geq 2$ and $r_{\mathbf{w}} \geq 2$ with respect to the variables $\mathbf{z}$ and $\mathbf{w}$. Let $\bar{\alpha} = -\frac{r_{\mathbf{z}}}{2} g(\bar{\mathbf{z}}, \bar{\mathbf{w}})$ and $\bar{\beta} = -\frac{r_{\mathbf{w}}}{2} g(\bar{\mathbf{z}}, \bar{\mathbf{w}})$. Then

$$\overline{\Theta} := H_{\bar{\alpha},\bar{\beta}}(\bar{\mathbf{z}}, \bar{\mathbf{w}}) + \begin{bmatrix} \left((2r_{\mathbf{z}} + r_{\mathbf{w}} - 4)\bar{\alpha} + r_{\mathbf{z}}\bar{\beta}\right) I_n & 0 \\ 0 & \left(r_{\mathbf{w}}\bar{\alpha} + (r_{\mathbf{z}} + 2r_{\mathbf{w}} - 4)\bar{\beta}\right) I_m \end{bmatrix} \succeq 0. \quad (4.24)$$

Proof. Let $(\mathbf{z}, \mathbf{w}) \in \mathbb{R}^n \times \mathbb{R}^m$ be arbitrary and put

$$\delta = \left\{ \begin{array}{ll} \frac{\bar{\mathbf{z}}^\top \mathbf{z}}{\bar{\mathbf{z}}^\top \bar{\mathbf{z}}} \,, & \text{if } \bar{\mathbf{z}} \neq \mathbf{o} \\ 0 \,, & \text{if } \bar{\mathbf{z}} = \mathbf{o} \end{array} \right\} \quad \text{as well as} \quad \gamma = \left\{ \begin{array}{ll} \frac{\bar{\mathbf{w}}^\top \mathbf{w}}{\bar{\mathbf{w}}^\top \bar{\mathbf{w}}} \,, & \text{if } \bar{\mathbf{w}} \neq \mathbf{o} \\ 0, & \text{if } \bar{\mathbf{w}} = \mathbf{o} \end{array} \right\} .$$

Then $[\mathbf{z}^\top - \delta\bar{\mathbf{z}}^\top, \mathbf{w}^\top - \gamma\bar{\mathbf{w}}^\top]^\top$ is the orthoprojection of $[\mathbf{z}^\top, \mathbf{w}^\top]^\top$ onto $\bar{\mathbf{z}}^\perp \times \bar{\mathbf{w}}^\perp$ and satisfies $\bar{\mathbf{z}}^\top(\mathbf{z} - \delta\bar{\mathbf{z}}) = \bar{\mathbf{w}}^\top(\mathbf{w} - \gamma\bar{\mathbf{w}}) = 0$. We have

$$\begin{aligned} H_{\bar{\alpha},\bar{\beta}}(\bar{\mathbf{z}}, \bar{\mathbf{w}}) \begin{bmatrix} \delta\bar{\mathbf{z}} \\ \gamma\bar{\mathbf{w}} \end{bmatrix} &= \left(\begin{bmatrix} \nabla^2_{\mathbf{zz}} g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) & \nabla^2_{\mathbf{zw}} g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) \\ \left[\nabla^2_{\mathbf{zw}} g(\bar{\mathbf{z}}, \bar{\mathbf{w}})\right]^\top & \nabla^2_{\mathbf{ww}} g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) \end{bmatrix} + 2 \begin{bmatrix} \bar{\alpha} I_n & 0 \\ 0 & \bar{\beta} I_m \end{bmatrix} \right) \begin{bmatrix} \delta\bar{\mathbf{z}} \\ \gamma\bar{\mathbf{w}} \end{bmatrix} \\ &= \begin{bmatrix} \delta\nabla^2_{\mathbf{zz}} g(\bar{\mathbf{z}}, \bar{\mathbf{w}})\bar{\mathbf{z}} + \gamma\nabla^2_{\mathbf{zw}} g(\bar{\mathbf{z}}, \bar{\mathbf{w}})\bar{\mathbf{w}} + 2\bar{\alpha}\delta\bar{\mathbf{z}} \\ \delta\left[\nabla^2_{\mathbf{zw}} g(\bar{\mathbf{z}}, \bar{\mathbf{w}})\right]^\top \bar{\mathbf{z}} + \gamma\nabla^2_{\mathbf{ww}} g(\bar{\mathbf{z}}, \bar{\mathbf{w}})\bar{\mathbf{w}} + 2\bar{\beta}\gamma\bar{\mathbf{w}} \end{bmatrix} \\ &= \begin{bmatrix} \delta(r_{\mathbf{z}} - 1)\nabla_{\mathbf{z}} g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) + \gamma r_{\mathbf{w}} \nabla_{\mathbf{z}} g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) + 2\bar{\alpha}\delta\bar{\mathbf{z}} \\ \delta r_{\mathbf{z}} \nabla_{\mathbf{w}} g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) + \gamma(r_{\mathbf{w}} - 1)\nabla_{\mathbf{w}} g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) + 2\bar{\beta}\gamma\bar{\mathbf{w}} \end{bmatrix} \\ &= \begin{bmatrix} -2\bar{\alpha}(\delta r_{\mathbf{z}} + \gamma r_{\mathbf{w}} - 2\delta)\bar{\mathbf{z}} \\ -2\bar{\beta}(\delta r_{\mathbf{z}} + \gamma r_{\mathbf{w}} - 2\gamma)\bar{\mathbf{w}} \end{bmatrix} , \end{aligned} \quad (4.25)$$

where the last equality follows from (4.12). Now obviously every local solution $(\bar{\mathbf{z}}, \bar{\mathbf{w}})$ to the considered problem also is a local solution to the equality-constrained problem

$$\min \left\{ g(\mathbf{z}, \mathbf{w}) : \|\mathbf{z}\| = \|\bar{\mathbf{z}}\| \,,\, \|\mathbf{w}\| = \|\bar{\mathbf{w}}\| \right\} .$$

Then, by (4.18) with obvious modifications if $\|\bar{\mathbf{z}}\|\,\|\bar{\mathbf{w}}\| = 0$, it results from (4.25) that

$$\begin{aligned} 0 \;\leq\; & \begin{bmatrix} \mathbf{z}-\delta\bar{\mathbf{z}} \\ \mathbf{w}-\gamma\bar{\mathbf{w}} \end{bmatrix}^{\top} H_{\bar\alpha,\bar\beta}(\bar{\mathbf{z}},\bar{\mathbf{w}}) \begin{bmatrix} \mathbf{z}-\delta\bar{\mathbf{z}} \\ \mathbf{w}-\gamma\bar{\mathbf{w}} \end{bmatrix} \\ = \; & \begin{bmatrix} \mathbf{z} \\ \mathbf{w} \end{bmatrix}^{\top} H_{\bar\alpha,\bar\beta}(\bar{\mathbf{z}},\bar{\mathbf{w}}) \begin{bmatrix} \mathbf{z} \\ \mathbf{w} \end{bmatrix} + \begin{bmatrix} \delta\bar{\mathbf{z}}-2\mathbf{z} \\ \gamma\bar{\mathbf{w}}-2\mathbf{w} \end{bmatrix}^{\top} H_{\bar\alpha,\bar\beta}(\bar{\mathbf{z}},\bar{\mathbf{w}}) \begin{bmatrix} \delta\bar{\mathbf{z}} \\ \gamma\bar{\mathbf{w}} \end{bmatrix}. \\ = \; & \begin{bmatrix} \mathbf{z} \\ \mathbf{w} \end{bmatrix}^{\top} H_{\bar\alpha,\bar\beta}(\bar{\mathbf{z}},\bar{\mathbf{w}}) \begin{bmatrix} \mathbf{z} \\ \mathbf{w} \end{bmatrix} + 2 \begin{bmatrix} 2\mathbf{z}-\delta\bar{\mathbf{z}} \\ 2\mathbf{w}-\gamma\bar{\mathbf{w}} \end{bmatrix}^{\top} \begin{bmatrix} \bar\alpha(\delta r_{\mathbf{z}}+\gamma r_{\mathbf{w}}-2\delta)\bar{\mathbf{z}} \\ \bar\beta(\delta r_{\mathbf{z}}+\gamma r_{\mathbf{w}}-2\gamma)\bar{\mathbf{w}} \end{bmatrix}. \end{aligned}$$

Next we use $(2\mathbf{z}-\delta\bar{\mathbf{z}})^{\top}\bar{\mathbf{z}} = \mathbf{z}^{\top}\bar{\mathbf{z}}$ and $(2\mathbf{w}-\gamma\bar{\mathbf{w}})^{\top}\bar{\mathbf{w}} = \mathbf{w}^{\top}\bar{\mathbf{w}}$ to arrive at

$$0 \leq \begin{bmatrix} \mathbf{z} \\ \mathbf{w} \end{bmatrix}^{\top} H_{\bar\alpha,\bar\beta}(\bar{\mathbf{z}},\bar{\mathbf{w}}) \begin{bmatrix} \mathbf{z} \\ \mathbf{w} \end{bmatrix} + 2\left[\bar\alpha\mathbf{z}^{\top}\bar{\mathbf{z}}(\delta r_{\mathbf{z}}+\gamma r_{\mathbf{w}}-2\delta) + \bar\beta\mathbf{w}^{\top}\bar{\mathbf{w}}(\delta r_{\mathbf{z}}+\gamma r_{\mathbf{w}}-2\gamma)\right]. \quad (4.26)$$

Now let us again distinguish cases: if $g(\bar{\mathbf{z}},\bar{\mathbf{w}}) = 0$, then by Theorem 4.1 $\bar\alpha = \bar\beta = 0$ and $\widetilde{\Theta} = H_{0,0}$ is positive-semidefinite by (4.26), since $\mathbf{z}$ and $\mathbf{w}$ were arbitrary. If, however, $g(\bar{\mathbf{z}},\bar{\mathbf{w}}) < 0$, then by Theorem 4.1 we know $\|\bar{\mathbf{z}}\| = \|\bar{\mathbf{w}}\| = 1$ so that $\delta = \mathbf{z}^{\top}\bar{\mathbf{z}}$ and $\gamma = \mathbf{w}^{\top}\bar{\mathbf{w}}$. We continue (4.26) to obtain

$$0 \leq \begin{bmatrix} \mathbf{z} \\ \mathbf{w} \end{bmatrix}^{\top} H_{\bar\alpha,\bar\beta}(\bar{\mathbf{z}},\bar{\mathbf{w}}) \begin{bmatrix} \mathbf{z} \\ \mathbf{w} \end{bmatrix} + 2(r_{\mathbf{z}}-2)\bar\alpha\delta^2 + 2(r_{\mathbf{w}}-2)\bar\beta\gamma^2 + 2\left(\bar\alpha r_{\mathbf{w}}+\bar\beta r_{\mathbf{z}}\right)\delta\gamma\,. \quad (4.27)$$

Moreover, from the fact that $\delta^2 \leq \|\mathbf{z}\|^2$, $\gamma^2 \leq \|\mathbf{w}\|^2$ and $2\delta\gamma \leq \delta^2+\gamma^2 \leq \|\mathbf{z}\|^2+\|\mathbf{w}\|^2$, it follows

$$\begin{aligned} & 2(r_{\mathbf{z}}-2)\bar\alpha\delta^2 + 2(r_{\mathbf{w}}-2)\bar\beta\gamma^2 + 2\left(\bar\alpha r_{\mathbf{w}}+\bar\beta r_{\mathbf{z}}\right)\delta\gamma \\ \leq \; & 2(r_{\mathbf{z}}-2)\bar\alpha\|\mathbf{z}\|^2 + 2(r_{\mathbf{w}}-2)\bar\beta\|\mathbf{w}\|^2 + \left(\bar\alpha r_{\mathbf{w}}+\bar\beta r_{\mathbf{z}}\right)\left(\|\mathbf{z}\|^2+\|\mathbf{w}\|^2\right), \end{aligned}$$

so that we derive from (4.27)

$$0 \leq \begin{bmatrix} \mathbf{z} \\ \mathbf{w} \end{bmatrix}^{\top} \left(H_{\bar\alpha,\bar\beta}(\bar{\mathbf{z}},\bar{\mathbf{w}}) + \begin{bmatrix} \bar{c}I_n & 0 \\ 0 & \bar{d}I_m \end{bmatrix} \right) \begin{bmatrix} \mathbf{z} \\ \mathbf{w} \end{bmatrix} \quad \text{for all } (\mathbf{z},\mathbf{w}) \in \mathbb{R}^n \times \mathbb{R}^m,$$

where $\bar{c} = (2r_{\mathbf{z}}+r_{\mathbf{w}}-4)\bar\alpha + r_{\mathbf{z}}\bar\beta$ and $\bar{d} = r_{\mathbf{w}}\bar\alpha + (r_{\mathbf{z}}+2r_{\mathbf{w}}-4)\bar\beta$, and the theorem is proved. ∎

From the above theorem we immediately conclude

Corollary 4.1 *Let $(\bar{\mathbf{z}},\bar{\mathbf{w}})$ be a local solution to the problem*

$$\min\left\{g(\mathbf{z},\mathbf{w}) : \|\mathbf{z}\|^2 \leq 1, \|\mathbf{w}\|^2 \leq 1\right\},$$

where g is homogeneous of degree r with respect to both the variables $\mathbf{z}$ and $\mathbf{w}$. Then necessarily $g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) \leq 0$, and for $\bar{\alpha} = \bar{\beta} = -\frac{r}{2} g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) \geq 0$, we have

$$H_{\bar{\alpha},\bar{\alpha}}(\bar{\mathbf{z}}, \bar{\mathbf{w}}) + 4\bar{\alpha}(r-1) I_{n+m} \succeq 0\,. \tag{4.28}$$

In case of $r_{\mathbf{z}} = r_{\mathbf{w}} = 4$, for a local minimizer $(\bar{\mathbf{z}}, \bar{\mathbf{w}})$ of (4.8) and $\bar{\alpha} = \bar{\beta} = -2g(\bar{\mathbf{z}}, \bar{\mathbf{w}})$, we get via (4.11), (4.10) and preceding relations, and dividing (4.28) by 4,

$$2 \begin{bmatrix} \bar{Z} C(\bar{\mathbf{w}}) \bar{Z} & 2\bar{G} \\ 2G^\top & \bar{W} B(\bar{\mathbf{z}}) \bar{W} \end{bmatrix} + \mathrm{Diag} \begin{bmatrix} C(\bar{\mathbf{w}}) \bar{Z} \bar{\mathbf{z}} \\ B(\bar{\mathbf{w}}) \bar{W} \bar{\mathbf{w}} \end{bmatrix} + \frac{7}{2} \bar{\alpha} I_{n+m} \succeq 0, \tag{4.29}$$

where $\bar{G} = \left(\bar{W} \bar{C}_1 \bar{W} \bar{\mathbf{w}}, \cdots, \bar{W} \bar{C}_n \bar{W} \bar{\mathbf{w}} \right)^\top$ and $\bar{C}_i = \bar{z}_i \sum_{k=1}^n \bar{z}_k^2 C_{ik}$.

As mentioned in [5] for the single ball constraint case, in our proof of Theorem 4.2, the fact that $\bar{\alpha} \geq 0$ and $\bar{\beta} \geq 0$ is essential. For the general bi-homogeneous optimization over the product of two spheres, we have the following result.

Theorem 4.3 *Let $(\bar{\mathbf{z}}, \bar{\mathbf{w}})$ be a local solution to the problem*

$$\min \left\{ g(\mathbf{z}, \mathbf{w}) : \|\mathbf{z}\|^2 = 1, \|\mathbf{w}\|^2 = 1 \right\},$$

where g is homogeneous of degrees $r_{\mathbf{z}}$ and $r_{\mathbf{w}}$ with respect to the variables $\mathbf{z}$ and $\mathbf{w}$, respectively. Then for $\bar{\alpha} = -\frac{r_{\mathbf{z}}}{2} g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) \in \mathbb{R}$ and $\bar{\beta} = -\frac{r_{\mathbf{w}}}{2} g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) \in \mathbb{R}$, we have that (4.23) holds and

$$\widetilde{\Theta} := H_{\bar{\alpha},\bar{\beta}}(\bar{\mathbf{z}}, \bar{\mathbf{w}}) + \begin{bmatrix} 2(r_{\mathbf{z}} - 2)\bar{\alpha} \bar{\mathbf{z}} \bar{\mathbf{z}}^\top & (\bar{\beta} r_{\mathbf{z}} + \bar{\alpha} r_{\mathbf{w}}) \bar{\mathbf{z}} \bar{\mathbf{w}}^\top \\ (\bar{\beta} r_{\mathbf{z}} + \bar{\alpha} r_{\mathbf{w}}) \bar{\mathbf{w}} \bar{\mathbf{z}}^\top & 2(r_{\mathbf{w}} - 2)\bar{\beta} \bar{\mathbf{w}} \bar{\mathbf{w}}^\top \end{bmatrix} \succeq 0. \tag{4.30}$$

Proof. The assertion (4.23) is obviously true. Now we prove (4.30). By the same arguments that lead to the proof of Theorem 4.2, we arrive at (4.26). Since $\delta = \mathbf{z}^\top \bar{\mathbf{z}}$ and $\gamma = \mathbf{w}^\top \bar{\mathbf{w}}$ here, this inequality can be rewritten as

$$0 \leq \begin{bmatrix} \mathbf{z} \\ \mathbf{w} \end{bmatrix}^\top \left(H_{\bar{\alpha},\bar{\beta}}(\bar{\mathbf{z}}, \bar{\mathbf{w}}) + \begin{bmatrix} 2(r_{\mathbf{z}} - 2)\bar{\alpha} \bar{\mathbf{z}} \bar{\mathbf{z}}^\top & (\bar{\beta} r_{\mathbf{z}} + \bar{\alpha} r_{\mathbf{w}}) \bar{\mathbf{z}} \bar{\mathbf{w}}^\top \\ (\bar{\beta} r_{\mathbf{z}} + \bar{\alpha} r_{\mathbf{w}}) \bar{\mathbf{w}} \bar{\mathbf{z}}^\top & 2(r_{\mathbf{w}} - 2)\bar{\beta} \bar{\mathbf{w}} \bar{\mathbf{w}}^\top \end{bmatrix} \right) \begin{bmatrix} \mathbf{z} \\ \mathbf{w} \end{bmatrix},$$

which implies that (4.30) holds. We complete the proof of the theorem. ∎

If $g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) \leq 0$ and $\min\{r_{\mathbf{z}}, r_{\mathbf{w}}\} \geq 2$, then (4.30) implies (4.24). Indeed, since $\|\bar{\mathbf{z}}\| \leq 1$ and $\|\bar{\mathbf{w}}\| \leq 1$, we know that $I_n - \bar{\mathbf{z}} \bar{\mathbf{z}}^\top \succeq 0$ and $I_m - \bar{\mathbf{w}} \bar{\mathbf{w}}^\top \succeq 0$, and also

$$\begin{bmatrix} I_n & -\bar{\mathbf{z}} \bar{\mathbf{w}}^\top \\ -\bar{\mathbf{w}} \bar{\mathbf{z}}^\top & I_m \end{bmatrix} \succeq 0\,.$$

Consequently, it follows that

$$\begin{aligned}
&\overline{\Theta} - \widetilde{\Theta} \\
&= -g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) \left\{ r_{\mathbf{z}} r_{\mathbf{w}} \begin{bmatrix} I_n & -\bar{\mathbf{z}}\bar{\mathbf{w}}^\top \\ -\bar{\mathbf{w}}\bar{\mathbf{z}}^\top & I_m \end{bmatrix} + \begin{bmatrix} r_{\mathbf{z}}(r_{\mathbf{z}} - 2)(I_n - \bar{\mathbf{z}}\bar{\mathbf{z}}^\top) & 0 \\ 0 & r_{\mathbf{w}}(r_{\mathbf{w}} - 2)(I_m - \bar{\mathbf{w}}\bar{\mathbf{w}}^\top) \end{bmatrix} \right\} \\
&\succeq 0\,,
\end{aligned}$$

since $\bar{\alpha} = -\frac{r_{\mathbf{z}} g(\bar{\mathbf{z}}, \bar{\mathbf{w}})}{2} \geq 0$ and $\bar{\beta} = -\frac{r_{\mathbf{w}} g(\bar{\mathbf{z}}, \bar{\mathbf{w}})}{2} \geq 0$.

The following corollary comes immediately from Theorem 4.3.

Corollary 4.2 *Let $(\bar{\mathbf{z}}, \bar{\mathbf{w}})$ be a pair of local solution for the problem of form*

$$\min \left\{ g(\mathbf{z}, \mathbf{w}) : \|\mathbf{z}\|^2 = 1, \|\mathbf{w}\|^2 = 1 \right\},$$

where g is homogeneous of degree r with respect to both the variables $\mathbf{z}$ and $\mathbf{w}$. Then for $\bar{\alpha} = \bar{\beta} = -\frac{r}{2} g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) \in \mathbb{R}$, we have that (4.13) holds and

$$H_{\bar{\alpha},\bar{\alpha}}(\bar{\mathbf{z}}, \bar{\mathbf{w}}) + 2\bar{\alpha} \begin{bmatrix} (r-2)\bar{\mathbf{z}}\bar{\mathbf{z}}^\top & r\bar{\mathbf{z}}\bar{\mathbf{w}}^\top \\ r\bar{\mathbf{w}}\bar{\mathbf{z}}^\top & (r-2)\bar{\mathbf{w}}\bar{\mathbf{w}}^\top \end{bmatrix} \succeq 0. \tag{4.31}$$

In particular, if $r = 4$, then (4.31) becomes

$$\nabla^2 g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) + 2\bar{\alpha} I_{n+m} + 4\bar{\alpha} \begin{bmatrix} \bar{\mathbf{z}}\bar{\mathbf{z}}^\top & 2\bar{\mathbf{z}}\bar{\mathbf{w}}^\top \\ 2\bar{\mathbf{w}}\bar{\mathbf{z}}^\top & \bar{\mathbf{w}}\bar{\mathbf{w}}^\top \end{bmatrix} \succeq 0, \tag{4.32}$$

where $\nabla^2 g(\bar{\mathbf{z}}, \bar{\mathbf{w}})$ is as in (4.11).

5 Optimality conditions: relations among different formulations

In this section, we consider the one-to-one correspondence among solutions of the original problem and its bi-quartic formulation. For sake of convenience, let us define the two transformations $\mathbf{x} = T_1(\mathbf{z})$ with $x_i = z_i^2$ $(i = 1, \cdots, n)$ and $\mathbf{y} = T_2(\mathbf{w})$ with $y_j = w_j^2$ $(j = 1, \cdots, m)$, respectively. Without loss of generality, we assume that $\mathbf{z} \geq \mathbf{o}$ and $\mathbf{w} \geq \mathbf{o}$. We denote by $\mathbf{z} = T_1^{-1}(\mathbf{x})$ and $\mathbf{w} = T_2^{-1}(\mathbf{y})$ the inverse transformation of T_1 and T_2, respectively, namely $z_i = \sqrt{|x_i|}$ for every $i = 1, \cdots, n$ and $w_j = \sqrt{|y_j|}$ for $j = 1, \cdots, m$.

We readily see that the transformations $\mathbf{x} = T_1(\mathbf{z})$, $\mathbf{y} = T_2(\mathbf{w})$ and their (partial) inverse $\mathbf{z} = T_1^{-1}(\mathbf{x})$, $\mathbf{w} = T_2^{-1}(\mathbf{y})$ are well-defined and continuous. Therefore, we have the following result which can be shown by arguments similar to those employed for proving [5, Theorem 5].

Theorem 5.1 *Let $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ be a feasible solution to (1.1). Then $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is a local solution to (1.1) if and only if $(\bar{\mathbf{z}}, \bar{\mathbf{w}}) = \left(T_1^{-1}(\bar{\mathbf{x}}), T_2^{-1}(\bar{\mathbf{y}})\right)$ is a local solution to (4.7). Further, a point $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is a global solution to (1.1) if and only if $(\bar{\mathbf{z}}, \bar{\mathbf{w}})$ is a global solution to (4.7).*

Theorem 5.2 *Let $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ be a KKT point for (1.1). Then $(\bar{\mathbf{z}}, \bar{\mathbf{w}}) = \left(T_1^{-1}(\bar{\mathbf{x}}), T_2^{-1}(\bar{\mathbf{y}})\right)$ is a KKT point for (4.7).*

Proof. Since $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is a KKT point for (1.1), it follows that there exist $\bar{\lambda}$, $\bar{\mu} \in \mathbb{R}$ such that (3.5) holds. From the first two expressions in (3.5) and the fact that $x_i = z_i^2$ for $i = 1, \cdots, n$, we obtain from the complementarity conditions

$$\left([(\bar{\mathbf{y}}\bar{\mathbf{y}}^\top \mathcal{A})\bar{x}]_i + \bar{\lambda}\right) \bar{z}_i = 0, \quad \text{for } i = 1, \cdots, n,$$

which implies

$$\bar{Z}(\bar{\mathbf{y}}\bar{\mathbf{y}}^\top \mathcal{A})\bar{Z}\bar{\mathbf{z}} + \bar{\lambda}\bar{\mathbf{z}} = \mathbf{o}\,. \tag{5.33}$$

Moreover, by the relation between $\bar{y}$ and $\bar{w}$, it is easy to verify that $R(\mathbf{y}) = \bar{\mathbf{y}}\bar{\mathbf{y}}^\top \mathcal{A} = C(\bar{\mathbf{w}})$. Since $\bar{\lambda} = -p(\bar{\mathbf{x}}, \bar{\mathbf{y}}) = -g(\bar{\mathbf{z}}, \bar{\mathbf{w}})$, it is clear that the first expression in (4.13) with $\bar{\alpha} = 2\bar{\lambda}$ holds. Similarly, we can prove that the second expression in (4.13) with $\bar{\beta} = 2\bar{\mu}$ is also true. Therefore, we obtained the desired result and complete the proof of the theorem. ∎

The converse of Theorem 5.2 is not true in general; this follows from the related result for quartic reformulations of StQPs [5].

Before we proceed to establish equivalence of the second-order optimality conditions, we simplify the Hessian of the objective function p:

$$\nabla^2 p(\mathbf{x}, \mathbf{y}) = 2 \begin{bmatrix} \mathbf{y}\mathbf{y}^\top \mathcal{A} & 2F(\mathbf{x}, \mathbf{y}) \\ 2F(\mathbf{x}, \mathbf{y})^\top & \mathcal{A}\mathbf{x}\mathbf{x}^\top \end{bmatrix} = 2 \begin{bmatrix} C(\mathbf{w}) & 2F(\mathbf{x}, \mathbf{y}) \\ 2F(\mathbf{x}, \mathbf{y})^\top & B(\mathbf{z}) \end{bmatrix}, \tag{5.34}$$

where $F(\mathbf{x}, \mathbf{y}) = \begin{bmatrix} \mathbf{y}^\top A_1(\mathbf{x}) \\ \vdots \\ \mathbf{y}^\top A_n(\mathbf{x}) \end{bmatrix}$ and $A_i(\mathbf{x}) = \left[\sum_{k=1}^{n} a_{ijkl} x_k\right]_{1 \le j,l \le m}$ are $m \times m$ matrices.

Theorem 5.3 *Let $(\bar{\mathbf{x}}, \bar{\mathbf{y}}) = (T_1(\bar{\mathbf{z}}), T_2(\bar{\mathbf{w}})) \in \Delta_n \times \Delta_m$ with $\bar{\mathbf{z}} \ge \mathbf{o}$ and $\bar{\mathbf{w}} \ge \mathbf{o}$. Then the following statements are equivalent:*

(a) $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is a KKT point for (1.1) which satisfies the second-order necessary optimality condition (3.6);

(b) $(\bar{\mathbf{z}}, \bar{\mathbf{w}})$ is a KKT point for (4.7) which satisfies the second-order necessary optimality condition (4.32);

Proof. (a) $\Rightarrow$ (b). Since $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is a KKT point for problem (1.1), by Theorem 5.2, it follows that $(\bar{\mathbf{z}}, \bar{\mathbf{w}})$ is a KKT point for problem (4.7), i.e., (4.13) holds. Now we prove (4.32). For any $(\mathbf{u}, \mathbf{v}) \in \mathbb{R}^n \times \mathbb{R}^m$, we define two vectors $\boldsymbol{\eta} = \bar{Z}(\mathbf{u} - \delta\bar{\mathbf{z}})$ and $\boldsymbol{\zeta} = \bar{W}(\mathbf{v} - \gamma\bar{\mathbf{w}})$, where $\delta = \bar{\mathbf{z}}^\top\mathbf{u}$ and $\gamma = \bar{\mathbf{w}}^\top\mathbf{v}$. It is easy to verify that, $\eta_i = 0$ for every $i \notin I(\bar{\mathbf{x}})$ and $\sum_{i\in I(\bar{\mathbf{x}})}\eta_i = 0$, and $\zeta_j = 0$ for every $j \notin J(\bar{\mathbf{y}})$ and $\sum_{j\in J(\bar{\mathbf{y}})}\zeta_j = 0$, where $I(\bar{\mathbf{x}}) = \{i : \bar{x}_i > 0\}$ and $J(\bar{\mathbf{y}}) = \{j : \bar{y}_j > 0\}$. This shows that $(\boldsymbol{\eta}, \boldsymbol{\zeta}) \in \mathcal{T}(\bar{\mathbf{x}}, \bar{\mathbf{y}})$. Consequently, by the second-order necessary condition (3.6), we have

$$\begin{aligned} 0 \;\leq\; & \begin{bmatrix} \boldsymbol{\eta} \\ \boldsymbol{\zeta} \end{bmatrix}^\top \nabla^2 p(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \begin{bmatrix} \boldsymbol{\eta} \\ \boldsymbol{\zeta} \end{bmatrix} \\ = \; & \begin{bmatrix} \bar{Z}\mathbf{u} \\ \bar{W}\mathbf{v} \end{bmatrix}^\top \nabla^2 p(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \begin{bmatrix} \bar{Z}\mathbf{u} \\ \bar{W}\mathbf{v} \end{bmatrix} - 2 \begin{bmatrix} \delta\bar{Z}\bar{\mathbf{z}} \\ \gamma\bar{W}\bar{\mathbf{w}} \end{bmatrix}^\top \nabla^2 p(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \begin{bmatrix} \bar{Z}\mathbf{u} \\ \bar{W}\mathbf{v} \end{bmatrix} \\ & + \begin{bmatrix} \delta\bar{Z}\bar{\mathbf{z}} \\ \gamma\bar{W}\bar{\mathbf{w}} \end{bmatrix}^\top \nabla^2 p(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \begin{bmatrix} \delta\bar{Z}\bar{\mathbf{z}} \\ \gamma\bar{W}\bar{\mathbf{w}} \end{bmatrix}. \end{aligned} \tag{5.35}$$

By (5.34) for $\mathbf{x} = \bar{\mathbf{x}}$ and $\mathbf{y} = \bar{\mathbf{y}}$, it follows that the first term on the right-hand side of (5.35) amounts to

$$\begin{bmatrix} \bar{Z}\mathbf{u} \\ \bar{W}\mathbf{v} \end{bmatrix}^\top \nabla^2 p(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \begin{bmatrix} \bar{Z}\mathbf{u} \\ \bar{W}\mathbf{v} \end{bmatrix} = 2 \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix}^\top \begin{bmatrix} \bar{Z}C(\bar{\mathbf{w}})\bar{Z} & 2\bar{Z}\bar{F}\bar{W} \\ 2\bar{W}\bar{F}^\top\bar{Z} & \bar{W}B(\bar{\mathbf{z}})\bar{W} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix}, \tag{5.36}$$

where we denote $\bar{F} = F(\bar{\mathbf{x}}, \bar{\mathbf{y}})$. Moreover, it is easy to verify that $\bar{\mathbf{z}}^\top\bar{Z}\bar{F}\bar{W}\bar{\mathbf{w}} = \bar{\mathbf{x}}^\top\bar{F}\bar{\mathbf{y}} = \sum_{i=1}^n \bar{x}_i\bar{\mathbf{y}}^\top A_i(\bar{x})\bar{\mathbf{y}} = p(\bar{\mathbf{x}}, \bar{\mathbf{y}}) = g(\bar{\mathbf{z}}, \bar{\mathbf{w}})$, which implies, together with the fact that $g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) = \bar{\mathbf{z}}^\top\bar{Z}C(\bar{\mathbf{w}})\bar{Z}\bar{\mathbf{z}} = \bar{\mathbf{w}}^\top\bar{W}B(\bar{\mathbf{z}})\bar{W}\bar{\mathbf{w}}$, that the last term on the right-hand side of (5.35) equals

$$\begin{aligned} \begin{bmatrix} \delta\bar{Z}\bar{\mathbf{z}} \\ \gamma\bar{W}\bar{\mathbf{w}} \end{bmatrix}^\top \nabla^2 p(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \begin{bmatrix} \delta\bar{Z}\bar{\mathbf{z}} \\ \gamma\bar{W}\bar{\mathbf{w}} \end{bmatrix} &= 2\left(\delta^2\bar{\mathbf{z}}^\top\bar{Z}C(\bar{\mathbf{w}})\bar{Z}\bar{\mathbf{z}} + 4\delta\gamma\bar{\mathbf{z}}^\top\bar{Z}\bar{F}\bar{W}\bar{\mathbf{w}} + \gamma^2\bar{\mathbf{w}}^\top\bar{W}B(\bar{\mathbf{z}})\bar{W}\bar{\mathbf{w}}\right) \\ &= 2g(\bar{\mathbf{z}}, \bar{\mathbf{w}})\left(\delta^2 + 4\delta\gamma + \gamma^2\right) \\ &= 2g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix}^\top \begin{bmatrix} \bar{\mathbf{z}}\bar{\mathbf{z}}^\top & 2\bar{\mathbf{z}}\bar{\mathbf{w}}^\top \\ 2\bar{\mathbf{w}}\bar{\mathbf{z}}^\top & \bar{\mathbf{w}}\bar{\mathbf{w}}^\top \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix}, \end{aligned} \tag{5.37}$$

where the last equality comes from the fact that $\delta = \bar{\mathbf{z}}^\top\mathbf{u}$ and $\gamma = \bar{\mathbf{w}}^\top\mathbf{v}$. On the other hand, we

have

$$\begin{aligned}\bar{\mathbf{w}}^\top \bar{W}\bar{F}^\top \bar{Z}\mathbf{u} &= \left[\bar{\mathbf{y}}^\top A_1(\bar{\mathbf{x}})\bar{\mathbf{y}}, \cdots, \bar{\mathbf{y}}^\top A_n(\bar{\mathbf{x}})\bar{\mathbf{y}}\right]\bar{Z}\mathbf{u} \\ &= \left([(\bar{\mathbf{y}}\bar{\mathbf{y}}^\top\mathcal{A})\bar{\mathbf{x}}]_1\bar{z}_1, \cdots, [(\bar{\mathbf{y}}\bar{\mathbf{y}}^\top\mathcal{A})\bar{\mathbf{x}}]_n\bar{z}_n\right)\mathbf{u} \\ &= \left[\bar{Z}C(\bar{\mathbf{w}})\bar{Z}\bar{\mathbf{z}}\right]^\top\mathbf{u} \\ &= g(\bar{\mathbf{z}},\bar{\mathbf{w}})(\bar{\mathbf{z}}^\top\mathbf{u}),\end{aligned}$$

where the last equality comes from (4.13), using Theorem 5.2. This implies

$$\gamma\bar{\mathbf{w}}^\top\bar{W}\bar{F}^\top\bar{Z}\mathbf{u} = g(\bar{\mathbf{z}},\bar{\mathbf{w}})(\mathbf{v}^\top\bar{\mathbf{w}})(\bar{\mathbf{z}}^\top\mathbf{u}). \tag{5.38}$$

Similarly, we can prove that

$$\delta\bar{\mathbf{z}}^\top\bar{Z}\bar{F}\bar{W}\mathbf{v} = g(\bar{\mathbf{z}},\bar{\mathbf{w}})(\mathbf{u}^\top\bar{\mathbf{z}})(\bar{\mathbf{w}}^\top\mathbf{v}). \tag{5.39}$$

Consequently, by (4.13), (5.38) and (5.39), it follows that the middle term on the right-hand side of (5.35)

$$\begin{aligned}\begin{bmatrix}\delta\bar{Z}\bar{\mathbf{z}}\\ \gamma\bar{W}\bar{\mathbf{w}}\end{bmatrix}^\top\nabla^2 p(\bar{\mathbf{x}},\bar{\mathbf{y}})\begin{bmatrix}\bar{Z}\mathbf{u}\\ \bar{W}\mathbf{v}\end{bmatrix} &= 2\delta\bar{\mathbf{z}}^\top\bar{Z}C(\bar{\mathbf{w}})\bar{Z}\mathbf{u} + 2\gamma\bar{\mathbf{w}}^\top\bar{W}B(\bar{\mathbf{z}})\bar{W}\mathbf{v} \\ &\quad + 4\delta\bar{\mathbf{z}}^\top\bar{Z}\bar{F}\bar{W}\mathbf{v} + 4\gamma\bar{\mathbf{w}}^\top\bar{W}\bar{F}^\top\bar{Z}\mathbf{u} \\ &= 2g(\bar{\mathbf{z}},\bar{\mathbf{w}})(\mathbf{u}^\top\bar{\mathbf{z}})(\bar{\mathbf{z}}^\top\mathbf{u}) + 2g(\bar{\mathbf{z}},\bar{\mathbf{w}})(\mathbf{v}^\top\bar{\mathbf{w}})(\bar{\mathbf{w}}^\top\mathbf{v}) \\ &\quad + 4g(\bar{\mathbf{z}},\bar{\mathbf{w}})(\mathbf{u}^\top\bar{\mathbf{z}})(\bar{\mathbf{w}}^\top\mathbf{v}) + 4g(\bar{\mathbf{z}},\bar{\mathbf{w}})(\mathbf{v}^\top\bar{\mathbf{w}})(\bar{\mathbf{z}}^\top\mathbf{u}) \\ &= 2g(\bar{\mathbf{z}},\bar{\mathbf{w}})\begin{bmatrix}\mathbf{u}\\ \mathbf{v}\end{bmatrix}^\top\begin{bmatrix}\bar{\mathbf{z}}\bar{\mathbf{z}}^\top & 2\bar{\mathbf{z}}\bar{\mathbf{w}}^\top\\ 2\bar{\mathbf{w}}\bar{\mathbf{z}}^\top & \bar{\mathbf{w}}\bar{\mathbf{w}}^\top\end{bmatrix}\begin{bmatrix}\mathbf{u}\\ \mathbf{v}\end{bmatrix}.\end{aligned} \tag{5.40}$$

By combining (5.35), (5.36), (5.37) and (5.40), we obtain

$$\begin{aligned}0 &\le \begin{bmatrix}\boldsymbol{\eta}\\ \boldsymbol{\zeta}\end{bmatrix}^\top\nabla^2 p(\bar{x},\bar{y})\begin{bmatrix}\boldsymbol{\eta}\\ \boldsymbol{\zeta}\end{bmatrix} \\ &= 2\begin{bmatrix}\mathbf{u}\\ \mathbf{v}\end{bmatrix}^\top\left(\begin{bmatrix}\bar{Z}C(\bar{\mathbf{w}})\bar{Z} & 2\bar{Z}\bar{F}\bar{W}\\ 2\bar{W}\bar{F}^\top\bar{Z} & \bar{W}B(\bar{\mathbf{z}})\bar{W}\end{bmatrix} - g(\bar{\mathbf{z}},\bar{\mathbf{w}})\begin{bmatrix}\bar{\mathbf{z}}\bar{\mathbf{z}}^\top & 2\bar{\mathbf{z}}\bar{\mathbf{w}}^\top\\ 2\bar{\mathbf{w}}\bar{\mathbf{z}}^\top & \bar{\mathbf{w}}\bar{\mathbf{w}}^\top\end{bmatrix}\right)\begin{bmatrix}\mathbf{u}\\ \mathbf{v}\end{bmatrix},\end{aligned}$$

which implies that

$$\begin{bmatrix}\bar{Z}C(\bar{\mathbf{w}})\bar{Z} & 2\bar{Z}\bar{F}\bar{W}\\ 2\bar{W}\bar{F}^\top\bar{Z} & \bar{W}B(\bar{\mathbf{z}})\bar{W}\end{bmatrix} - g(\bar{\mathbf{z}},\bar{\mathbf{w}})\begin{bmatrix}\bar{\mathbf{z}}\bar{\mathbf{z}}^\top & 2\bar{\mathbf{z}}\bar{\mathbf{w}}^\top\\ 2\bar{\mathbf{w}}\bar{\mathbf{z}}^\top & \bar{\mathbf{w}}\bar{\mathbf{w}}^\top\end{bmatrix} \succeq 0. \tag{5.41}$$

On the other hand, since $\sum_{k=1}^n \bar{x}_k C_{ik} = A_i(\bar{\mathbf{x}})$, it is easy to verify via (4.10) that

$$\nabla^2_{\mathbf{zw}} g(\bar{\mathbf{z}},\bar{\mathbf{w}}) = 16\begin{bmatrix}\bar{z}_1\sum_{k=1}^n \bar{z}_k^2\bar{\mathbf{w}}^\top\bar{W}C_{1k}\bar{W}\\ \bar{z}_2\sum_{k=1}^n \bar{z}_k^2\bar{\mathbf{w}}^\top\bar{W}C_{2k}\bar{W}\\ \vdots\\ \bar{z}_n\sum_{k=1}^n \bar{z}_k^2\bar{\mathbf{w}}^\top\bar{W}C_{nk}\bar{W}\end{bmatrix} = 16\bar{Z}\begin{bmatrix}\bar{\mathbf{y}}^\top\left(\sum_{k=1}^n \bar{x}_k C_{1k}\right)\bar{W}\\ \bar{\mathbf{y}}^\top\left(\sum_{k=1}^n \bar{x}_k C_{2k}\right)\bar{W}\\ \vdots\\ \bar{\mathbf{y}}^\top\left(\sum_{k=1}^n \bar{x}_k C_{nk}\right)\bar{W}\end{bmatrix} = 16\bar{Z}\bar{F}\bar{W}.$$

Recall also that $C(\bar{\mathbf{w}})\bar{Z}\bar{\mathbf{z}} + \bar{\lambda}\mathbf{e} \geq \mathbf{o}$ and $B(\bar{\mathbf{z}})\bar{W}\bar{\mathbf{w}} + \bar{\mu}\mathbf{e} \geq \mathbf{o}$ from (3.5), which means that $2\mathrm{Diag}\left(C(\bar{\mathbf{w}})\bar{Z}\bar{\mathbf{z}}\right) + \bar{\alpha}I_n \succeq 0$ and $2\mathrm{Diag}\left(B(\bar{\mathbf{z}})\bar{W}\bar{\mathbf{w}}\right) + \bar{\beta}I_m \succeq 0$. By combining this, (4.11) and (5.41), we know that (4.32) is true.

(b)⇒(a). Since $x_i = z_i^2$ for $i = 1, \cdots, n$ and $y_j = w_j^2$ for $j = 1, \cdots, m$, the first-order condition (4.13) can be rewritten equivalently as

$$\begin{cases} \left(2[C(\bar{\mathbf{w}})\bar{\mathbf{x}}]_i + \bar{\alpha}\right)\bar{z}_i = 0, & \text{for } i = 1, \cdots, n, \\ \left(2[B(\bar{\mathbf{z}})\bar{\mathbf{y}}]_j + \bar{\beta}\right)\bar{w}_j = 0, & \text{for } j = 1, \cdots, m\,. \end{cases}$$

Notice that $\bar{x}_i > 0$ if and only if $\bar{z}_i > 0$, and $\bar{y}_j > 0$ if and only if $\bar{w}_j > 0$. If $\bar{x}_i > 0$, then $2[C(\bar{\mathbf{w}})\bar{\mathbf{x}}]_i + \bar{\alpha} = 0$, which means $[(\bar{\mathbf{y}}\bar{\mathbf{y}}^\top\mathcal{A})\bar{\mathbf{x}}]_i + \bar{\lambda} = 0$, if we define $\bar{\lambda} = \frac{1}{2}\bar{\alpha}$, because $C(\bar{\mathbf{w}}) = \bar{\mathbf{y}}\bar{\mathbf{y}}^\top\mathcal{A}$. Else, i.e., if $\bar{x}_i = 0$, we choose $\mathbf{u} = \mathbf{e}_i \in \mathbb{R}^n$ and $\mathbf{v} = \mathbf{o} \in \mathbb{R}^m$. By (4.32), we obtain

$$0 \leq \begin{bmatrix} \mathbf{e}_i \\ \mathbf{o} \end{bmatrix}^\top \left(\nabla^2 g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) + 2\bar{\alpha}I_{n+m} + 4\bar{\alpha} \begin{bmatrix} \bar{\mathbf{z}}\bar{\mathbf{z}}^\top & 2\bar{\mathbf{z}}\bar{\mathbf{w}}^\top \\ 2\bar{\mathbf{w}}\bar{\mathbf{z}}^\top & \bar{\mathbf{w}}\bar{\mathbf{w}}^\top \end{bmatrix} \right) \begin{bmatrix} \mathbf{e}_i \\ \mathbf{o} \end{bmatrix},$$

which implies

$$0 \leq 8\mathbf{e}_i^\top\bar{Z}C(\bar{\mathbf{w}})\bar{Z}\mathbf{e}_i + 4\mathbf{e}_i^\top\mathrm{Diag}[C(\bar{\mathbf{w}})\bar{Z}\bar{\mathbf{z}}]\mathbf{e}_i + 2\bar{\alpha} + 4\bar{\alpha}(\bar{\mathbf{z}}^\top\mathbf{e}_i)^2\,.$$

This means that $0 \leq 2[C(\bar{\mathbf{w}})\bar{\mathbf{x}}]_i + \bar{\alpha}$, from the fact that $\bar{Z}\mathbf{e}_i = \bar{z}_i = \bar{\mathbf{z}}^\top\mathbf{e}_i = 0$ and $\bar{Z}\bar{\mathbf{z}} = \bar{\mathbf{x}}$. Consequently, we have $[(\bar{\mathbf{y}}\bar{\mathbf{y}}^\top\mathcal{A})\bar{\mathbf{x}}]_i + \bar{\lambda} \geq 0$. The first two expressions in (3.5) hold. Similarly, we can prove that other two expressions in (3.5) are also true. Therefore, $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is a KKT point for problem (1.1) with the corresponding multipliers $\bar{\lambda} = \bar{\alpha}/2$ and $\bar{\mu} = \bar{\beta}/2 = \bar{\lambda}$. Now let us prove that $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ satisfies also the second-order condition (3.6). Let $[\mathbf{u}^\top, \mathbf{v}^\top]^\top \in \mathcal{T}(\bar{\mathbf{x}}, \bar{\mathbf{y}})$. We define $(\boldsymbol{\eta}, \boldsymbol{\zeta}) \in \mathbb{R}^n \times \mathbb{R}^m$ by

$$\eta_i = \begin{cases} 0 & \text{if } \bar{z}_i = 0, \\ \dfrac{u_i}{\bar{z}_i} & \text{if } \bar{z}_i > 0 \end{cases} \quad \text{and} \quad \zeta_j = \begin{cases} 0 & \text{if } \bar{w}_j = 0\,, \\ \dfrac{v_j}{\bar{w}_j} & \text{if } \bar{w}_j > 0\,. \end{cases}$$

Then we have $\bar{Z}\boldsymbol{\eta} = \mathbf{u}$ and $\bar{W}\boldsymbol{\zeta} = \mathbf{v}$. Moreover, it holds that

$$\bar{\mathbf{z}}^\top\boldsymbol{\eta} = \sum_{i=1}^n \eta_i\bar{z}_i = \sum_{i\in I(\bar{\mathbf{x}})} \eta_i\bar{z}_i = \sum_{i\in I(\bar{\mathbf{x}})} u_i = 0$$

and

$$\bar{\mathbf{w}}^\top\boldsymbol{\zeta} = \sum_{j=1}^m \zeta_j\bar{w}_j = \sum_{j\in J(\bar{\mathbf{y}})} \zeta_j\bar{w}_j = \sum_{j\in J(\bar{\mathbf{y}})} v_j = 0\,.$$

On the other hand, by (4.13), it is easy to prove that

$$\boldsymbol{\eta}^\top\mathrm{Diag}\left[C(\bar{\mathbf{w}})\bar{Z}\bar{\mathbf{z}}\right]\boldsymbol{\eta} = \sum_{i\in I(\bar{\mathbf{x}})} [C(\bar{\mathbf{w}})\bar{Z}\bar{\mathbf{z}}]_i\eta_i^2 = -\frac{\bar{\alpha}}{2}\|\boldsymbol{\eta}\|^2$$

and

$$\boldsymbol{\zeta}^\top \mathrm{Diag}\left[B(\bar{\mathbf{z}})\bar{W}\bar{\mathbf{w}}\right]\boldsymbol{\zeta} = \sum_{j\in J(\bar{\mathbf{y}})}[B(\bar{\mathbf{z}})\bar{W}\bar{\mathbf{w}}]_j\zeta_j^2 = -\frac{\bar{\alpha}}{2}\|\boldsymbol{\zeta}\|^2.$$

Therefore, by (4.32), we obtain

$$\begin{aligned}
0 \;\leq\;& \begin{bmatrix}\boldsymbol{\eta}\\ \boldsymbol{\zeta}\end{bmatrix}^\top \left(\nabla^2 g(\bar{\mathbf{z}},\bar{\mathbf{w}}) + 2\bar{\alpha}I_{n+m} + 4\bar{\alpha}\begin{bmatrix}\bar{\mathbf{z}}\bar{\mathbf{z}}^\top & 2\bar{\mathbf{z}}\bar{\mathbf{w}}^\top\\ 2\bar{\mathbf{w}}\bar{\mathbf{z}}^\top & \bar{\mathbf{w}}\bar{\mathbf{w}}^\top\end{bmatrix}\right)\begin{bmatrix}\boldsymbol{\eta}\\ \boldsymbol{\zeta}\end{bmatrix}\\
=\;& 8\left[\boldsymbol{\eta}^\top\bar{Z}C(\bar{\mathbf{w}})\bar{Z}\boldsymbol{\eta} + \boldsymbol{\zeta}^\top\bar{W}B(\bar{\mathbf{z}})\bar{W}\boldsymbol{\zeta}\right] + 4\left(\boldsymbol{\eta}^\top\mathrm{Diag}\left[C(\bar{\mathbf{w}})\bar{Z}\bar{\mathbf{z}}\right]\boldsymbol{\eta} + \boldsymbol{\zeta}^\top\mathrm{Diag}\left[B(\bar{\mathbf{z}})\bar{W}\bar{\mathbf{w}}\right]\boldsymbol{\zeta}\right)\\
+\;& 32\boldsymbol{\eta}^\top\bar{Z}\bar{F}\bar{W}\boldsymbol{\zeta} + 2\bar{\alpha}\left(\|\boldsymbol{\eta}\|^2 + \|\boldsymbol{\zeta}\|^2\right)\\
=\;& 8\left[\mathbf{u}^\top C(\bar{\mathbf{w}})\mathbf{u} + \mathbf{v}^\top B(\bar{\mathbf{z}})\mathbf{v}\right] + 32\mathbf{u}^\top\bar{F}\mathbf{v}\\
=\;& 4\begin{bmatrix}\mathbf{u}\\ \mathbf{v}\end{bmatrix}^\top \nabla^2 p(\bar{\mathbf{x}},\bar{\mathbf{y}})\begin{bmatrix}\mathbf{u}\\ \mathbf{v}\end{bmatrix},
\end{aligned}$$

using (5.34). This shows that the second-order condition (3.6) holds, and the proof of the theorem is complete. ∎

Theorem 5.3 states the relations among points satisfying the second order necessary conditions of problems (1.1) and (4.7). Hence, if we want to use the bi-quartic formulation to obtain a solution of the original problem (1.1), we need an algorithm that converges to second-order KKT points of problem (4.7).

6 A penalty method for StBQPs

The bi-quartic formulation (4.7) of the StBQP can be solved by a penalty method, which is based upon the use of a continuously differentiable exact penalty function. Our main technique used in this section follows lines similar to that of [5].

6.1 A continuously differentiable penalty function

For the tensor $\mathcal{A}$ in (1.1), we denote by $\overline{a}$ and $\underline{a}$ the maximum and minimum elements in $\mathcal{A}$, respectively. It is clear that $-\|\mathcal{A}\|_F \leq \underline{a} \leq \overline{a} \leq -\frac{1}{(mn)^2}\|\mathcal{A}\|_F < 0 < \|\mathcal{A}\|_F$, from the assumption that all entries of $\mathcal{A}$ are negative. Then, for any $(\mathbf{z},\mathbf{w}) \in \mathbb{R}^n \times \mathbb{R}^m$, we readily verify that

$$-\|\mathcal{A}\|_F\|\mathbf{z}\|^4\|\mathbf{w}\|^4 \leq \underline{a}\|\mathbf{z}\|^4\|\mathbf{w}\|^4 \leq g(\mathbf{z},\mathbf{w}) \leq \overline{a}\|\mathbf{z}\|^4\|\mathbf{w}\|^4 \leq \|\mathcal{A}\|_F\|\mathbf{z}\|^4\|\mathbf{w}\|^4. \tag{6.42}$$

For the bi-quartic optimization problem (4.7), we introduce an exact penalty function, which is defined by

$$\begin{aligned} P(z,w;\varepsilon) &:= g(\mathbf{z},\mathbf{w}) + \tfrac{1}{\varepsilon}\left(\|\mathbf{z}\|^4-1\right)^2 + \tfrac{1}{\varepsilon}\left(\|\mathbf{w}\|^4-1\right)^2 \\ &\quad + \alpha(\mathbf{z},\mathbf{w})\left(\|\mathbf{z}\|^2-1\right) + \alpha(\mathbf{z},\mathbf{w})\left(\|\mathbf{w}\|^2-1\right), \end{aligned}$$

where $\alpha(\mathbf{z},\mathbf{w}) = -2g(\mathbf{z},\mathbf{w})$. Then

$$P(z,w;\varepsilon) = g(\mathbf{z},\mathbf{w})\left(5-2\|\mathbf{z}\|^2-2\|\mathbf{w}\|^2\right) + \frac{1}{\varepsilon}\left(\|\mathbf{z}\|^4-1\right)^2 + \frac{1}{\varepsilon}\left(\|\mathbf{w}\|^4-1\right)^2.$$

Note that the definition of P is similar to (but not the same as) the definition of the penalty function used in [5] and [13], which can ensure that the level set of P is bounded without any assumption other than $g < 0$.

It is easily seen that $P(\cdot,\cdot;\varepsilon)$ is twice continuously differentiable on $\mathbb{R}^n \times \mathbb{R}^m$ for any fixed $\varepsilon > 0$. Moreover, it holds that

$$\begin{aligned} \nabla P(z,w;\varepsilon) &= \nabla g(\mathbf{z},\mathbf{w})\left(5-2\|\mathbf{z}\|^2-2\|\mathbf{w}\|^2\right) - 4g(\mathbf{z},\mathbf{w})\begin{bmatrix}\mathbf{z}\\ \mathbf{w}\end{bmatrix} \\ &\quad + \frac{8}{\varepsilon}\left(\|\mathbf{z}\|^4-1\right)\|\mathbf{z}\|^2\begin{bmatrix}\mathbf{z}\\ \mathbf{o}\end{bmatrix} + \frac{8}{\varepsilon}\left(\|\mathbf{w}\|^4-1\right)\|\mathbf{w}\|^2\begin{bmatrix}\mathbf{o}\\ \mathbf{w}\end{bmatrix} \end{aligned} \tag{6.43}$$

and

$$\begin{aligned} \nabla^2 P(\mathbf{z},\mathbf{w};\varepsilon) &= \nabla^2 g(\mathbf{z},\mathbf{w})\left(5-2\|\mathbf{z}\|^2-2\|\mathbf{w}\|^2\right) - 4g(\mathbf{z},\mathbf{w})I_{n+m} \\ &\quad -4\nabla g(\mathbf{z},\mathbf{w})\begin{bmatrix}\mathbf{z}\\ \mathbf{w}\end{bmatrix}^\top - 4\begin{bmatrix}\mathbf{z}\\ \mathbf{w}\end{bmatrix}\nabla g(\mathbf{z},\mathbf{w})^\top \\ &\quad + \frac{8}{\varepsilon}\left(\|\mathbf{z}\|^4-1\right)\|\mathbf{z}\|^2\begin{bmatrix} I_n & 0\\ 0 & 0\end{bmatrix} + \frac{8}{\varepsilon}\left(\|\mathbf{w}\|^4-1\right)\|\mathbf{w}\|^2\begin{bmatrix} 0 & 0\\ 0 & I_m\end{bmatrix} \\ &\quad + \frac{16}{\varepsilon}\left(3\|\mathbf{z}\|^4-1\right)\begin{bmatrix}\mathbf{z}\mathbf{z}^\top & 0\\ 0 & 0\end{bmatrix} + \frac{16}{\varepsilon}\left(3\|\mathbf{w}\|^4-1\right)\begin{bmatrix} 0 & 0\\ 0 & \mathbf{w}\mathbf{w}^\top\end{bmatrix}. \end{aligned} \tag{6.44}$$

For any fixed $(\mathbf{z}^0,\mathbf{w}^0) \in \mathbb{R}^n \times \mathbb{R}^m$, let us define the sub-level set of P

$$\mathcal{L}_0 = \left\{(\mathbf{z},\mathbf{w}) \in \mathbb{R}^n \times \mathbb{R}^m : P(\mathbf{z},\mathbf{w};\varepsilon) \le P(\mathbf{z}^0,\mathbf{w}^0;\varepsilon)\right\}.$$

Denote

$$\bar{\varepsilon} = \frac{1}{\|\mathcal{A}\|_F}\min\left\{\frac{2\left(C^4-1\right)^2}{C^8(7+5C^8)},\ \frac{2\left(1-\vartheta^4\right)^2}{2+5(C^8+\vartheta^8)},\ \frac{\left(C^4-1\right)^2}{3C^8},\ \frac{\vartheta^4}{3\left(C^6+C^4\right)}\right\}, \tag{6.45}$$

where $\vartheta \in (0,1)$ and $C > 1$ are user-selected constants. If C is large and $\vartheta^2 C \leq 1$, a safe rule of thumb is $\bar{\varepsilon} \approx [3C^8 \|\mathcal{A}\|_F]^{-1}$.

Now we state and prove the following theorem, which characterizes the boundedness of the sub-level set without any assumption. This theorem implies the existence of a global minimizer and the boundedness of the sequence generated by an unconstrained method.

Theorem 6.1 *Let $(\mathbf{z}^0, \mathbf{w}^0) \in \mathbb{R}^n \times \mathbb{R}^m$ be a point such that $\|\mathbf{z}^0\| = 1$ and $\|\mathbf{w}^0\| = 1$. If $\vartheta \in (0,1)$ and $C > 1$ are the constants appearing in (6.45), then for $0 < \varepsilon < \bar{\varepsilon}$*

$$\mathcal{L}_0 \subseteq \{(\mathbf{z}, \mathbf{w}) \in \mathbb{R}^n \times \mathbb{R}^m : \vartheta \leq \|\mathbf{z}\| \leq C,\ \vartheta \leq \|\mathbf{w}\| \leq C\}.$$

Proof. Since $\|\mathbf{z}^0\| = 1$ and $\|\mathbf{w}^0\| = 1$, it follows that

$$P(\mathbf{z}^0, \mathbf{w}^0; \varepsilon) = g(\mathbf{z}^0, \mathbf{w}^0) \leq \|\mathcal{A}\|_F.$$

Moreover, it holds that for any $(\mathbf{z}, \mathbf{w}) \in \mathbb{R}^n \times \mathbb{R}^m$,

$$P(\mathbf{z}, \mathbf{w}; \varepsilon) \geq \begin{cases} \overline{a}\|\mathbf{z}\|^4\|\mathbf{w}\|^4\left(5 - 2\|\mathbf{z}\|^2 - 2\|\mathbf{w}\|^2\right) + \dfrac{1}{\varepsilon}\left(\|\mathbf{z}\|^4 - 1\right)^2 + \dfrac{1}{\varepsilon}\left(\|\mathbf{w}\|^4 - 1\right)^2, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } 5 < 2\|\mathbf{z}\|^2 + 2\|\mathbf{w}\|^2, \\ \underline{a}\|\mathbf{z}\|^4\|\mathbf{w}\|^4\left(5 - 2\|\mathbf{z}\|^2 - 2\|\mathbf{w}\|^2\right) + \dfrac{1}{\varepsilon}\left(\|\mathbf{z}\|^4 - 1\right)^2 + \dfrac{1}{\varepsilon}\left(\|\mathbf{w}\|^4 - 1\right)^2, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } 5 \geq 2\|\mathbf{z}\|^2 + 2\|\mathbf{w}\|^2, \end{cases}$$

which implies that

$$P(\mathbf{z}, \mathbf{w}; \varepsilon) \geq \begin{cases} 5\overline{a}\|\mathbf{z}\|^4\|\mathbf{w}\|^4 + \dfrac{1}{\varepsilon}\left(\|\mathbf{z}\|^4 - 1\right)^2 + \dfrac{1}{\varepsilon}\left(\|\mathbf{w}\|^4 - 1\right)^2, \text{ if } 5 < 2\|\mathbf{z}\|^2 + 2\|\mathbf{w}\|^2, \\ 5\underline{a}\|\mathbf{z}\|^4\|\mathbf{w}\|^4 + \dfrac{1}{\varepsilon}\left(\|\mathbf{z}\|^4 - 1\right)^2 + \dfrac{1}{\varepsilon}\left(\|\mathbf{w}\|^4 - 1\right)^2, \text{ if } 5 \geq 2\|\mathbf{z}\|^2 + 2\|\mathbf{w}\|^2, \end{cases}$$

since $\underline{a} \leq \overline{a} < 0$. Hence, by (6.42), we have

$$\begin{aligned} P(\mathbf{z}, \mathbf{w}; \varepsilon) &\geq -5\|\mathcal{A}\|_F\|\mathbf{z}\|^4\|\mathbf{w}\|^4 + \frac{1}{\varepsilon}\left(\|\mathbf{z}\|^4 - 1\right)^2 + \frac{1}{\varepsilon}\left(\|\mathbf{w}\|^4 - 1\right)^2 \\ &\geq -\frac{5}{2}\|\mathcal{A}\|_F\left(\|\mathbf{z}\|^8 + \|\mathbf{w}\|^8\right) + \frac{1}{\varepsilon}\left(\|\mathbf{z}\|^4 - 1\right)^2 + \frac{1}{\varepsilon}\left(\|\mathbf{w}\|^4 - 1\right)^2. \end{aligned} \tag{6.46}$$

Now we first prove that $\|\mathbf{z}\| > C > 1$ and $\|\mathbf{w}\| \leq C$ implies $P(\mathbf{z}, \mathbf{w}; \varepsilon) > P(\mathbf{z}^0, \mathbf{w}^0; \varepsilon)$. Since $\|\mathbf{w}\| \leq C$, by (6.46), it follows that

$$P(\mathbf{z}, \mathbf{w}; \varepsilon) \geq -\frac{5}{2}\|\mathcal{A}\|_F\left(\|\mathbf{z}\|^8 + C^8\right) + \frac{1}{\varepsilon}\left(\|\mathbf{z}\|^4 - 1\right)^2. \tag{6.47}$$

On the other hand, since $\varepsilon < \bar{\varepsilon} \leq \frac{2\left(C^4-1\right)^2}{\|\mathcal{A}\|_F C^8(7+5C^8)}$, we obtain

$$\varepsilon < \frac{\left(\|\mathbf{z}\|^4 - 1\right)^2}{\|\mathcal{A}\|_F\left(1 + \frac{5}{2}(1 + C^8)\right)\|\mathbf{z}\|^8},$$

which implies that

$$\varepsilon < \frac{\left(\|\mathbf{z}\|^4 - 1\right)^2}{\|\mathcal{A}\|_F \left(1 + \frac{5}{2}(\|\mathbf{z}\|^8 + C^8)\right)}.$$

By combining this and (6.47), we know that $P(\mathbf{z}, \mathbf{w}; \varepsilon) > \|\mathcal{A}\|_F \geq P(\mathbf{z}^0, \mathbf{w}^0; \varepsilon)$. Similarly, we may prove that $\|\mathbf{w}\| > C > 1$ and $\|\mathbf{z}\| \leq C$ implies $P(\mathbf{z}, \mathbf{w}; \varepsilon) > P(\mathbf{z}^0, \mathbf{w}^0; \varepsilon)$.

Secondly, we prove that $\|\mathbf{z}\| \leq C$ and $\|\mathbf{w}\| < \vartheta < 1$ implies $P(\mathbf{z}, \mathbf{w}; \varepsilon) > P(\mathbf{z}^0, \mathbf{w}^0; \varepsilon)$. By (6.46), we only need to prove

$$-\frac{5}{2}\|\mathcal{A}\|_F \left(C^8 + \vartheta^8\right) + \frac{1}{\varepsilon}\left(\|\mathbf{w}\|^4 - 1\right)^2 > \|\mathcal{A}\|_F, \tag{6.48}$$

under the given conditions. Since $\varepsilon < \bar{\varepsilon} \leq \frac{2\left(1-\vartheta^4\right)^2}{\|\mathcal{A}\|_F(2+5(C^8+\vartheta^8))}$, we obtain

$$\varepsilon < \frac{\left(1 - \|\mathbf{w}\|^4\right)^2}{\|\mathcal{A}\|_F \left(1 + \frac{5}{2}(C^8 + \vartheta^8)\right)},$$

because of $\|\mathbf{w}\| < \vartheta < 1$. Hence (6.48) holds. Similarly, we may prove that $\|\mathbf{z}\| < \vartheta < 1$ and $\|\mathbf{w}\| \leq C$ implies $P(\mathbf{z}, \mathbf{w}; \varepsilon) > P(\mathbf{z}^0, \mathbf{w}^0; \varepsilon)$.

Finally, we prove that $\|\mathbf{z}\| > C > 1$ and $\|\mathbf{w}\| > C > 1$ implies $P(\mathbf{z}, \mathbf{w}; \varepsilon) > P(\mathbf{z}^0, \mathbf{w}^0; \varepsilon)$. To this end, we only need prove

$$\begin{cases} -\frac{5}{2}\|\mathcal{A}\|_F\|\mathbf{z}\|^8 + \frac{1}{\varepsilon}\left(\|\mathbf{z}\|^4 - 1\right)^2 > \frac{1}{2}\|\mathcal{A}\|_F & \text{and} \\ -\frac{5}{2}\|\mathcal{A}\|_F\|\mathbf{w}\|^8 + \frac{1}{\varepsilon}\left(\|\mathbf{w}\|^4 - 1\right)^2 > \frac{1}{2}\|\mathcal{A}\|_F\,. \end{cases} \tag{6.49}$$

From (6.45) and the condition that $\|\mathbf{z}\| > C > 1$, it follows that

$$\varepsilon < \frac{\left(\|\mathbf{z}\|^4 - 1\right)^2}{3\|\mathcal{A}\|_F\|\mathbf{z}\|^8},$$

which implies

$$\varepsilon < \frac{2\left(\|\mathbf{z}\|^4 - 1\right)^2}{\|\mathcal{A}\|_F\left(1 + 5\|\mathbf{z}\|^8\right)}.$$

By this, the first relation in (6.49) holds. The second relation in (6.49) can be similarly proved. Hence we obtain the desired result. ∎

By means of the penalty function P, the problem of locating a constrained global minimizer of problem (4.7) is recast as the problem of locating an unconstrained global minimizer of P. About the one-to-one correspondence between global/local solutions of (4.7) and global/local minimizers of the penalty function P, we have the following two theorems. The arguments are quite standard for the penalty approach, so we omit the proofs here and refer to [5] for details.

Theorem 6.2 *(Correspondence of global minimizers). For $0 < \varepsilon < \bar{\varepsilon}$ as in (6.45). Every global minimizer of problem (4.7) is a global minimizer of $P(z, w, \varepsilon)$ and conversely.*

Theorem 6.3 *(Correspondence of local minimizers). For $0 < \varepsilon < \bar{\varepsilon}$ as in (6.45). Let $(\bar{\mathbf{z}}, \bar{\mathbf{w}})$ be a local minimizer of $P(\mathbf{z}, \mathbf{w}; \varepsilon)$. Then $(\bar{\mathbf{z}}, \bar{\mathbf{w}})$ is a local solution to problem (4.7), and the associated KKT multipliers are $(\alpha(\bar{\mathbf{z}}, \bar{\mathbf{w}}), \alpha(\bar{\mathbf{z}}, \bar{\mathbf{w}}))$.*

The following theorem describes the relationship between the stationary points of $P(\mathbf{z}, \mathbf{w}; \varepsilon)$ and (4.7).

Theorem 6.4 *(First-order exactness property). For $0 < \varepsilon < \bar{\varepsilon}$ as in (6.45), a point $(\bar{\mathbf{z}}, \bar{\mathbf{w}}) \in \mathcal{L}_0$ is a stationary point of $P(\mathbf{z}, \mathbf{w}; \varepsilon)$ if and only if $(\bar{\mathbf{z}}, \bar{\mathbf{w}})$ is a KKT point for problem (4.7), and the associated KKT multipliers are $(\alpha(\bar{\mathbf{z}}, \bar{\mathbf{w}}), \alpha(\bar{\mathbf{z}}, \bar{\mathbf{w}}))$.*

Proof. ($\Leftarrow$). Since $(\bar{\mathbf{z}}, \bar{\mathbf{w}})$ is a KKT point for problem (4.7), i.e., (4.12) holds, it is in particular feasible. Consequently, by (6.43), we have

$$\nabla P(\bar{\mathbf{z}}, \bar{\mathbf{w}}; \varepsilon) = \nabla g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) + 2\alpha(\bar{\mathbf{z}}, \bar{\mathbf{w}}) \begin{bmatrix} \bar{\mathbf{z}} \\ \bar{\mathbf{w}} \end{bmatrix} = \mathbf{o}\,.$$

($\Rightarrow$). If $\nabla P(\bar{\mathbf{z}}, \bar{\mathbf{w}}; \varepsilon) = \mathbf{o}$, then we have

$$\frac{\varepsilon}{4}\bar{\mathbf{z}}^\top \nabla_{\mathbf{z}} P(\bar{\mathbf{z}}, \bar{\mathbf{w}}; \varepsilon) = 0 \quad \text{and} \quad \frac{\varepsilon}{4}\bar{\mathbf{w}}^\top \nabla_{\mathbf{w}} P(\bar{\mathbf{z}}, \bar{\mathbf{w}}; \varepsilon) = 0\,.$$

Now let $\tau_1 = \|\bar{\mathbf{z}}\|^2 - 1$ and $\tau_2 = \|\bar{\mathbf{w}}\|^2 - 1$. Then the above equations imply together with (6.43), that

$$\left.\begin{array}{rcl} \left[2\|\bar{\mathbf{z}}\|^4\left(\|\bar{\mathbf{z}}\|^2+1\right) - 3\varepsilon g(\bar{\mathbf{z}}, \bar{\mathbf{w}})\right]\tau_1 - 2\varepsilon g(\bar{\mathbf{z}}, \bar{\mathbf{w}})\tau_2 &=& 0 \\ -2\varepsilon g(\bar{\mathbf{z}}, \bar{\mathbf{w}})\tau_1 + \left[2\|\bar{\mathbf{w}}\|^4\left(\|\bar{\mathbf{w}}\|^2+1\right) - 3\varepsilon g(\bar{\mathbf{z}}, \bar{\mathbf{w}})\right]\tau_2 &=& 0 \end{array}\right\} \tag{6.50}$$

Notice that the determinant of this homogeneous system of linear equations in $[\tau_1, \tau_2]^\top$ is

$$\begin{aligned} \Delta &= \begin{vmatrix} 2\|\bar{\mathbf{z}}\|^4\left(\|\bar{\mathbf{z}}\|^2+1\right) - 3\varepsilon g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) & -2\varepsilon g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) \\ -2\varepsilon g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) & 2\|\bar{\mathbf{w}}\|^4\left(\|\bar{\mathbf{w}}\|^2+1\right) - 3\varepsilon g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) \end{vmatrix} \\ &= 4\|\bar{\mathbf{z}}\|^4\|\bar{\mathbf{w}}\|^4\left(\|\bar{\mathbf{z}}\|^2+1\right)\left(\|\bar{\mathbf{w}}\|^2+1\right) + 5\varepsilon^2 g^2(\bar{\mathbf{z}}, \bar{\mathbf{w}}) \\ &\quad - 6\varepsilon g(\bar{\mathbf{z}}, \bar{\mathbf{w}})\left[\|\bar{\mathbf{z}}\|^4\left(\|\bar{\mathbf{z}}\|^2+1\right) + \|\bar{\mathbf{w}}\|^4\left(\|\bar{\mathbf{w}}\|^2+1\right)\right] \\ &\geq 2\left\{2\|\bar{\mathbf{z}}\|^6\|\bar{\mathbf{w}}\|^6 - 3\varepsilon\|\mathcal{A}\|_F\|\bar{\mathbf{z}}\|^4\|\bar{\mathbf{w}}\|^4\left[\|\bar{\mathbf{z}}\|^4\left(\|\bar{\mathbf{z}}\|^2+1\right) + \|\bar{\mathbf{w}}\|^4\left(\|\bar{\mathbf{w}}\|^2+1\right)\right]\right\} \\ &\geq 2\vartheta^8\left\{2\vartheta^4 - 6\varepsilon\|\mathcal{A}\|_F\left(C^6 + C^4\right)\right\} \\ &> 0\,, \end{aligned}$$

where the first inequality comes from (6.42), and the last inequality is due to (6.45). Therefore, it follows from (6.50) that $\tau_1 = \tau_2 = 0$, i.e., $\|\bar{\mathbf{z}}\|^2 = 1$ and $\|\bar{\mathbf{w}}\|^2 = 1$. Consequently, from (6.43),

we obtain

$$\nabla g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) + 2\alpha(\bar{\mathbf{z}}, \bar{\mathbf{w}}) \begin{bmatrix} \bar{\mathbf{z}} \\ \bar{\mathbf{w}} \end{bmatrix} = \mathbf{o}\,,$$

which means that $(\bar{\mathbf{z}}, \bar{\mathbf{w}})$ is a KKT point for problem (4.7) with the multipliers $\bar{\alpha} = \bar{\beta} = \alpha(\bar{\mathbf{z}}, \bar{\mathbf{w}})$, and the proof is complete. ∎

For the second-order optimality condition of (4.7), we have

Theorem 6.5 *(Second-order exactness property). For $0 < \varepsilon < \bar{\varepsilon}$ as in (6.45), let $(\bar{\mathbf{z}}, \bar{\mathbf{w}}) \in \mathcal{L}_0$ be a stationary point of $P(\mathbf{z}, \mathbf{w}; \varepsilon)$ satisfying the standard second-order necessary conditions for unconstrained optimality. Then $(\bar{\mathbf{z}}, \bar{\mathbf{w}})$ satisfies the second-order necessary conditions for problem (4.7).*

Proof. By Theorem 6.4, the first-order optimality conditions hold. Therefore, it follows that $\|\bar{\mathbf{z}}\|^2 = 1$ and $\|\bar{\mathbf{w}}\|^2 = 1$. Moreover, we obtain

$$\nabla g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) + 2\alpha(\bar{\mathbf{z}}, \bar{\mathbf{w}}) \begin{bmatrix} \bar{\mathbf{z}} \\ \bar{\mathbf{w}} \end{bmatrix} = \mathbf{o}\,,$$

which implies, together with (6.44), that

$$\nabla^2 P(\bar{\mathbf{z}}, \bar{\mathbf{w}}; \varepsilon) = \nabla^2 g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) - 4g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) I_{n+m} + 16\alpha(\bar{\mathbf{z}}, \bar{\mathbf{w}}) \begin{bmatrix} \bar{\mathbf{z}}\bar{\mathbf{z}}^\top & \bar{\mathbf{z}}\bar{\mathbf{w}}^\top \\ \bar{\mathbf{w}}\bar{\mathbf{z}}^\top & \bar{\mathbf{w}}\bar{\mathbf{w}}^\top \end{bmatrix} + \frac{32}{\varepsilon} \begin{bmatrix} \bar{\mathbf{z}}\bar{\mathbf{z}}^\top & 0 \\ 0 & \bar{\mathbf{w}}\bar{\mathbf{w}}^\top \end{bmatrix}.$$

Consequently, for every $(\mathbf{u}, \mathbf{v}) \in \mathbb{R}^n \times \mathbb{R}^m$ with $\bar{\mathbf{z}}^\top \mathbf{u} = \bar{\mathbf{w}}^\top \mathbf{v} = 0$, we have

$$0 \leq \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix}^\top \nabla^2 P(\bar{\mathbf{z}}, \bar{\mathbf{w}}; \varepsilon) \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix}^\top \left(\nabla^2 g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) + 2\alpha(\bar{\mathbf{z}}, \bar{\mathbf{w}}) I_{n+m}\right) \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix},$$

which shows that (4.16) holds and the proof is complete. ∎

6.2 Penalty method guarantees improvement

Theorems 6.2 and 6.5 show that we may generate a sequence $\{(\mathbf{z}^k, \mathbf{w}^k)\}$ via an unconstrained method for the minimization of the penalty function P, which converges to a point $(\bar{\mathbf{z}}, \bar{\mathbf{w}})$ satisfying the second order necessary conditions. Indeed, by Theorem 6.5, stationary points of P in $\mathcal{L}_0$ satisfying the second order necessary conditions, are points satisfying the second-order

necessary conditions for problem (4.7) which, in turn, by Theorem 5.3 are points satisfying the second-order necessary condition (3.6) for the StBQP (1.1).

We observe that given a feasible starting point $(\mathbf{z}^0, \mathbf{w}^0)$ any reasonable unconstrained minimization algorithm is able to locate a KKT point with a lower value of the objective function. In fact, any of these algorithms obtains a stationary point $(\bar{\mathbf{z}}, \bar{\mathbf{w}})$ for P such that $P(\bar{\mathbf{z}}, \bar{\mathbf{w}}; \varepsilon) \leq P(\mathbf{z}^0, \mathbf{w}^0; \varepsilon)$. Then Theorem 6.5 ensures that $(\bar{\mathbf{z}}, \bar{\mathbf{w}})$ is a KKT point of problem (4.7). On the other hand, if $(\mathbf{z}^0, \mathbf{w}^0)$ is a feasible point for (4.7), recalling the definition of P, we get that

$$g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) = P(\bar{\mathbf{z}}, \bar{\mathbf{w}}; \varepsilon) < P(\mathbf{z}^0, \mathbf{w}^0; \varepsilon) = g(\bar{\mathbf{z}}, \bar{\mathbf{w}}) \,.$$

In conclusion, by using an unconstrained optimization algorithm, we get a KKT point $(\bar{\mathbf{z}}, \bar{\mathbf{w}})$ of the problem (4.7) with a value of the objective function lower than the value at the starting point $(\mathbf{z}^0, \mathbf{w}^0)$. However, in general, the point $(\bar{\mathbf{z}}, \bar{\mathbf{w}})$ obtained by the algorithm mentioned above is not a global minimizer of the problem (1.1). In order to obtain a global solution of the problem (1.1), we may use some appropriate global technique to 'escape' from local solutions.

References

[1] K. Anstreicher and S. Burer, "D.C.Versus copositive bounds for standard QP", *Journal of Global Optimization*, Vol.33 (2005), pp.299-312.

[2] A. Bagchi and B. Kalantari, "New optimality conditions and algorithms for homogeneous and polynomial optimization over spheres," Rutcor Research Report n. **40-90**, 1990.

[3] I.M. Bomze, "On Standard Quadratic Optimization Problems," *Journal of Global Optimization*, Vol.13 (1998), pp.369-387.

[4] I.M. Bomze, M. Budinich, P. Pardalos, and M. Pelillo (1999), "The maximum clique problem," in: D.-Z. Du, P.M. Pardalos (eds.), *Handbook of Combinatorial Optimization (supp. Vol. A)*, 1-74. Kluwer, Dordrecht.

[5] I.M. Bomze and L. Palagi, "Quartic Formulation of Standard Quadratic Optimization Problems", *Journal of Global Optimization*, Vol.32 (2005), pp.181-205.

[6] I.M. Bomze and W. Schachinger, "Multi-Standard Quadratic optimization problems: interior point methods and cone programming reformulation", to appear in *Computational Optimization and Applications* (2009).

[7] G. Dahl, J.M. Leinaas, J. Myrheim and E. Ovrum, "A tensor product matrix aproximation problem in quantum physics", *Linear Algebra and applications,* Vol. 420 (2007), pp.711-725.

[8] A. Einstein, B. Podolsky and N. Rosen, "Can quantum-mechanical description of physical reality be considered complete?" *Physical Review,* Vol. 47 (1935), pp.777.

[9] D. Han, H.H. Dai and L. Qi, "Conditions for strong ellipticity of anisotropic elastic materials", to appear in: *Journal of Elasticity*.

[10] J. Jahn, "Introduction to the theory of nonlinear optimization", Springer, New York (2006).

[11] J.K. Knowles and E. Sternberg, "On the ellipticity of the equations of the equations for finite elastostatics for a special material", *Journal of Elasticity,* Vol. 5 (1975), pp.341-361.

[12] C. Ling, J. Nie, L. Qi and Y. Ye, "Bi-quadratic optimization over unit spheres and semidefinite programming relaxations", to appear in: *SIAM Journal on Optimization.*

[13] S. Lucidi and L. Palagi, "Solution of the trust region problem via a smooth Unconstrained Reformulation", In: Pardalos, P. and Wolkowicz, H. (eds.), *Topics in Semidefinite and Interior-Point methods*, Fields Institute Communications, Vol. 18, pp.237-250, AMS.

[14] H.M. Markowitz, "Portfolio selection", *The Journal of Finance*, Vol.7 (1952), pp.77-91.

[15] H.M. Markowitz, "The general mean-variance portfolio selection problem," in: S.D. Howison, F.P. Kelly and P. Wilmott (eds.), *Mathematical Models in Finance*, pp.93-99. London (1995).

[16] J.S. Pang "A new and efficient algorithm for a class of portfolio selection problems", *Operations Research*, Vol. 8 (1980), pp. 754-767.

[17] L. Qi, H.H. Dai and D. Han, "Conditions for strong ellipticity and M-eigenvalues", *Frontiers of Mathematics in China,* Vol. 4 (2009), pp.349-364.

[18] P. Rosakis, "Ellipticity and deformations with discontinuous deformation gradients in finite elastostatics", *Archive Rational Mechanics Analysis,* Vol. 109 (1990), pp.1-37.

[19] W. Schachinger and I.M. Bomze, "A conic duality Frank-Wolfe type theorem via exact penalization in quadratic optimization", *Mathematics of Operations Research*, Vol. 34 (2009), pp.83-91.

[20] P. Tseng, I.M. Bomze, and W. Schachinger, "A first-order interior-point method for linearly constrained smooth optimization", to appear in: *Mathematical Programming* (2009).

[21] Y. Wang and M. Aron, "A reformulation of the strong ellipticity conditions for unconstrained hyperelastic media", *Journal of Elasticity,* Vol. 44 (1996), pp.89-96.

[22] Y. Wang, L. Qi and X. Zhang, "A practical method for computing the largest M-eigenvalue of a fourth-order partially symmetric tensor", *Numerical Linear Algebra with Applications,* Vol. 16 (2009), pp.589-601.