On the local and global minimizers of the smooth stress function in Euclidean Distance Matrix problems

Mengmeng Song[*]   Douglas Gonçalves[†]   Woosuk L. Jung[‡]   Carlile Lavor[§]

Antonio Mucherino[¶]   Henry Wolkowicz[‖]

August 13, 2024, 10:02pm

Abstract

We consider the nonconvex minimization problem, with quartic objective function, that arises in the exact recovery of a configuration matrix $P \in \mathbb{R}^{n \times d}$ of n points when a Euclidean distance matrix, **EDM**, is given with embedding dimension d. It is an open question in the literature under which conditions such a minimization problem admits a local nonglobal minimizer, **lngm**. We prove that all second order stationary points are global minimizers whenever $n \leq d+1$. For $n > d+1$, we numerically find a local nonglobal minimum and show analytically that there indeed exists a nearby **lngm** for the underlying quartic minimization problem. Thus, we answer in the affirmative the previously open question about their existence. Our approach to finding the **lngm** is novel in that we first exploit the translation and rotation invariance to reduce the size of the problem from nd variables in P to $(n-1)d - d(d-1)/2 = d(2n-d-1)/2$ variables. This allows for finding examples that satisfy the strict second order sufficient optimality conditions.

Contents

[*]School of Mathematical Sciences, Beihang University, Beijing, 100191, P.R. China; Department of Combinatorics and Optimization, Faculty of Mathematics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1; songmengmeng@buaa.edu.cn

[†]douglas.goncalves@ufsc.br

[‡]Department of Combinatorics and Optimization, Faculty of Mathematics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1; w2jung@uwaterloo.ca

[§]clavor@unicamp.br

[¶]antonio.mucherino@irisa.fr

[‖]Department of Combinatorics and Optimization, Faculty of Mathematics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1; Research supported by The Natural Sciences and engineering Research Council of Canada.

Key Words: distance geometry, Euclidean distance matrices, Lindenstrauss operator, local nonglobal minima, Kantorovich Theorem **AMS Subject Classification** 51K05, 90C26, 65K10

1   Introduction

EDM (completion) problems have been widely studied in the scientific literature for the past decades, see e.g., the surveys, book collections and some recent papers [2–4, 6, 8, 12]. It is well known that one can obtain the Gram matrix $\bar{G}$ from a given **EDM** $\bar{D}$. Then, a configuration matrix $\bar{P}$ of points $\bar{p}_i \in \mathbb{R}^d$ such that $\bar{D}_{ij} = \|\bar{p}_i - \bar{p}_j\|^2, i, j = 1, \ldots, n$, can be obtained from a full rank factorization

$$\bar{G} = \bar{P}\bar{P}^T, \quad \bar{P}^T = \left[\bar{p}_1, \ldots, \bar{p}_n\right] \in \mathbb{R}^{d \times n}.$$

In this work, we consider the question of exact recovery from the unconstrained minimization problem

$$\min_{P \in \mathbb{R}^{n \times d}} \quad \|\mathcal{K}(PP^T) - \bar{D}\|_F^2 \tag{1.1}$$

that arises from application of the Lindenstrauss operator on symmetric matrix space $\mathcal{K} : \mathbb{S}^n \to \mathbb{S}^n$:

$$\mathcal{K}(G) = \operatorname{diag}(G)e^T + e\operatorname{diag}(G)^T - 2G.$$

Here e is the vector of ones and $\operatorname{diag}(G)$ is the linear mapping providing the vector of diagonal elements of the square matrix G. The objective function of (1.1), given by

$$\sigma_2(P) = \|\mathcal{K}(PP^T) - \bar{D}\|_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} \left(\|p_i - p_j\|^2 - \|\bar{p}_i - \bar{p}_j\|^2\right)^2, \tag{1.2}$$

is quartic in P, known in multidimensional scaling (MDS) literature, e.g. [11], as the *smooth stress*.[1]

Since $\sigma_2(P)$ is nonconvex, and most optimization methods are local, we investigate the possibility for such a quartic to have all its local minimizers being global ones. This is a prevalent question in multidimensional scaling literature [9, 11, 13] with no definitive answer in previous publications.

The question about the existence of a **lngm** was previously considered for another type of stress function, called the *raw stress*:

$$\sigma_1(P) = \sum_{i=1}^{n}\sum_{j=1}^{n} \left(\|p_i - p_j\| - \|\bar{p}_i - \bar{p}_j\|\right)^2. \tag{1.3}$$

---

[1]For notational convenience, we use $f(P) = \frac{1}{2}\sigma_2(P)$ as our objective function.

It was analytically verified that a **lngm** exists for the raw stress function $\sigma_1(P)$ [11]. However, the question about the existence of such a **lngm** for the smooth stress function $\sigma_2(P)$ remained open for a much longer time.

A few years after the result in [11] for the raw stress function, the authors of [9], while studying the problem in the context of "metric" multi-dimensional scaling, already wondered whether the same example could work for the smooth function $\sigma_2(P)$. Unfortunately, the example constructed in [11] for raw stress (1.3) does not produce a nonglobal minimizer for the smooth stress (1.2) as discussed in the technical report [10]. In [15], an example of the *inexact* EDM recovery problem, i.e.,

$$\min_{P \in \mathbb{R}^{n \times d}} \quad \|\mathcal{K}(PP^T) - \Delta\|_F^2,$$

was presented to have a **lngm**, where $\Delta \in \mathbb{S}^n$ is not an **EDM** in $\mathbb{R}^d$ as in (1.1). In a PhD Thesis focusing on EDMs and published in 2013 [13], the question related to the nonglobal minimizers of $\sigma_2(P)$ was mentioned and considered to be open.

In this work, we finally give a definitive answer to this question. We find examples where the function $\sigma_2(P)$ has a local nonglobal minimum, **lngm**, and we provide an analytic verification. In our numerical investigation of stationary points of σ_2, we have used a trust region approach for obtaining points satisfying the second order necessary optimality conditions for (1.1). We also prove that, for $n \le d + 1$, no stationary point of smooth stress with $\sigma_2(P) > 0$ satisfies the second order necessary optimality conditions. In other words, in this latter case all local minimizers are in fact global ones.

We continue in Section 2 with a description of our main unconstrained minimization problem. We include two additional equivalent problems with reduced numbers of variables. The reduction allows for strict optimality conditions that can be used for the analytic existence proof. In Section 3, we include various linear transformations, derivatives, and adjoints. Many of these are used in the paper. It is hoped that these are a useful addition to the literature of **EDM**s as they emphasize the use of linear transformations rather than individual elements or points. In Section 4, we consider the optimality conditions and present a sufficient condition under which $\sigma_2(P)$ has no local nonglobal minimizer. In Section 5, we give two examples and use a modified Kantorovich theorem to prove the existence of **lngm**s.

2   Notation and Main Problem Formulations

Before presenting the problem, we look at some of the required notation and background from distance geometry. Further notation and background can be found in the book [3].

2.1   Notation

We let $S, T \in \mathbb{S}^n$ denote matrices in the space of $n \times n$ symmetric matrices equipped with the *trace inner product* $\langle S, T \rangle = \operatorname{tr} ST$; we use $\operatorname{diag}(S) \in \mathbb{R}^n$ to denote the diagonal of S; the adjoint mapping is $\operatorname{diag}^*(v) = \operatorname{Diag}(v) \in \mathbb{S}^n$. We let $[k] = \{1, 2, \ldots, k\}$ and use $\otimes$ to denote the Kronecker product. Unless stated otherwise, $\| \cdot \|$ denotes the Frobenius norm.

The cone of positive semidefinite matrices is denoted $\mathbb{S}_+^n \subset \mathbb{S}^n$, and we use $S \succeq 0$ for $S \in \mathbb{S}_+^n$. Similarly, for positive definite matrices, we use $\mathbb{S}_{++}^n, S \succ 0$.

For a set of points $p_i \in \mathbb{R}^d$, we let

$$P = \begin{bmatrix} p_1^T \\ p_2^T \\ \dots \\ p_n^T \end{bmatrix} \in \mathbb{R}^{n \times d}$$

denote the *configuration matrix*. Here we denote by d the embedding dimension. We denote the corresponding *Gram matrix*, $G = PP^T$. In addition, we define the quadratic mapping $\mathcal{M} : \mathbb{R}^{n \times d} \to \mathbb{S}^n$, $\mathcal{M}(P) = PP^T$.

The classical result of Schoenberg [14] relates a *Euclidean distance matrix, **EDM***, with a Gram matrix by applying the *Lindenstrauss operator*, $\mathcal{K} : \mathbb{S}^n \to \mathbb{S}^n$, given by

$$G = PP^T, \quad D = \left(\|p_i - p_j\|^2 \right) = \mathcal{K}(G) = \operatorname{diag}(G)e^T + e\operatorname{diag}(G)^T - 2G.$$

We define $\mathcal{D} : \mathbb{R}^{n \times d} \to \mathbb{S}^n$ such that

$$\mathcal{D}(P) = \mathcal{K}(PP^T) = \mathcal{K}(\mathcal{M}(P)).$$

Moreover, the mapping $\mathcal{K}$ is one-one and onto between the *centered subspace, $\mathcal{S}_C^n$* and the *hollow subspace, $\mathcal{S}_H^n$*, given by

$$\mathcal{S}_C^n = \{S \in \mathbb{S}^n \ : \ Se = 0\}, \quad \mathcal{S}_H^n = \{S \in \mathbb{S}^n \ : \ \operatorname{diag} S = 0\}.$$

We ignore the dimension n when the meaning is clear. Note that the centered assumption $P^T e = 0 \implies G = PP^T \in \mathcal{S}_C^n$.

Further detailed properties and a list of (non)linear transformations and adjoints are given in Section 3.

2.2 Main Problem Formulations

Suppose that we are given the data $\bar{P} \in \mathbb{R}^{n \times d}, \bar{P}^T e = 0$, of n (centered) points in the embedding dimension d. This gives rise to the corresponding Gram matrix $\bar{G} = \bar{P}\bar{P}^T \in \mathcal{S}_C^n$ and **EDM**,

$$\bar{D} = \mathcal{K}(\bar{G}) = \operatorname{diag}(\bar{G})e^T + e\operatorname{diag}(\bar{G})^T - 2\bar{G}. \tag{2.1}$$

We now present the main problem and two reformulations that reduce the size and help with stability.

Problem 2.1 (point recovery). *Let $\bar{D}$ be the above given **EDM** in (2.1). Consider the* nonconvex *minimization problem of recovering a corresponding point matrix $\bar{P}$ given by*

$$\min_{P \in \mathbb{R}^{n \times d}} f(P) := \frac{1}{2} \|\mathcal{K}(PP^T) - \bar{D}\|_F^2 =: \frac{1}{2} \|F(P)\|_F^2, \tag{2.2}$$

*which defines the function $F : \mathbb{R}^{n \times d} \to \mathcal{S}_H^n$. Does $P^* \in \mathbb{R}^{n \times d}$ exist such that it is a **lngm** for (2.2)?*

Problem (2.2) is a <u>non</u>linear least squares problem. It has nd variables. By taking advantage of symmetry and the zero diagonal, the objective function can be seen as a sum of squares of a *triangular number*, $t(n-1) := n(n-1)/2$ of quadratic functions. Note that $\bar{P}$ is clearly a global minimum for Problem (2.2) with optimal value $f(\bar{P}) = 0$. The problem is to determine whether or not all stationary points where the second-order necessary optimality conditions hold are global minimizers. If it is true, then the gradient descent method with a small stepsize can find the global minimizer of problem (2.2) with probability one [7].

Note that a distance matrix is invariant under translations and rotations. Without loss of generality, we can translate the points and assume they are centered, i.e.,

$$P^T e = 0, \ e \text{ vector of ones}.$$

We let

$$V \in \mathbb{R}^{n \times n-1}, V^T V = I_{n-1}, V^T e = 0. \tag{2.3}$$

Then, $P^T e = 0$ if, and only if, $P = VL$ for some $L \in \mathbb{R}^{n-1 \times d}$. We can take advantage of this to get equivalent smaller dimensional problems.

Problem 2.2 (centered point recovery)**.** *Let $\bar{D}$ be the above given **EDM** in (2.1). Let V be as in (2.3). Consider the* nonconvex *minimization problem of recovering a corresponding centered point matrix $\bar{P} = V\bar{L}$, i.e.,*

$$\min_{L \in \mathbb{R}^{n-1 \times d}} f_L(L) := \frac{1}{2} \| \mathcal{K}(VL(VL)^T) - \bar{D} \|_F^2 =: \frac{1}{2} \| F_L(L) \|_F^2, \tag{2.4}$$

*which defines the function $F_L : \mathbb{R}^{n-1 \times d} \to \mathcal{S}_H^n$. Does $L^* \in \mathbb{R}^{n-1 \times d}$ exist such that it is a **lngm** for (2.4)?*

We let $\mathcal{O} = \{Q \in \mathbb{R}^{d \times d} : Q^T Q = I_d\}$ be the orthogonal group order d. We note that $LL^T = LQQ^T L^T$ holds for all $Q \in \mathcal{O}$. If $L^T = QR$ is the QR-factorization, then $R^T = LQ$, i.e., $f_L(L) = f_L(R^T Q^T) = f_L(R^T)$, where $R \in \mathbb{R}^{d \times n-1}$ is upper triangular(trapezoidal). We can in fact reduce the problem further using the rotation invariance, i.e., $f_L(LQ) = f_L(L)$ for any $Q \in \mathcal{O}$.

We define the linear transformation that takes a vector and changes it into a lower triangular matrix R^T. For $d < n - 1$, R^T is a lower triangular matrix of size $n - 1 \times d$, but with $t(d-1)$ elements being zero at the top right. For $d \geq n-1$, R^T is a lower triangular matrix of size $n-1 \times d$ with $t(n-1)$ elements being nonzero at the bottom left. This yields t_ℓ nonzeros:

$$t_\ell = \begin{cases} t(n-1) & \text{if } d \geq n-1 \\ (n-1)d - t(d-1) & \text{otherwise.} \end{cases} \tag{2.5}$$

We now define

$$\mathcal{L}\text{Triag} : \mathbb{R}^{t_\ell} \to \mathbb{R}^{n-1 \times d}, \ \mathcal{L}\text{Triag}(\ell)_{(i,j)} = \begin{cases} \ell_{nj-n-t(j)+i+1} & \text{if } j \leq i \\ 0 & \text{otherwise,} \end{cases} \tag{2.6}$$

and

$$f_\ell : \mathbb{R}^{t_\ell} \to \mathbb{R}, \quad f_\ell(\ell) := f_L(\mathcal{L}\text{Triag}(\ell)). \tag{2.7}$$

Notice that the adjoint of $\mathcal{L}\text{Triag}$, $\mathcal{L}\text{Triag}^* : \mathbb{R}^{n-1 \times d} \to \mathbb{R}^{t_\ell}$, takes the lower triangular part of $L \in \mathbb{R}^{n-1 \times d}$ and maps it to the corresponding vector ℓ in $\mathbb{R}^{t_\ell}$, such that $\mathcal{L}\text{Triag}^* \mathcal{L}\text{Triag}(\ell) = \ell$. Moreover $\mathcal{L}\text{Triag} \mathcal{L}\text{Triag}^*(L)$ is the projection of L onto the subspace of lower triangular(trapezoidal) matrices.

Problem 2.3 (centered and triangular point recovery). *Let $\bar{D}$ be the above given **EDM** in* (2.1). *Let $V, t_\ell, f_\ell(\ell)$, be as in* (2.3), (2.5) *and* (2.7), *respectively. Consider the* nonconvex *minimization problem of recovering a corresponding centered point matrix $\bar{P} = V\bar{L} = V\,\mathcal{L}\mathrm{Triag}(\bar{\ell})\bar{Q}^T$, with $\bar{Q} \in \mathcal{O}$, i.e.,*

$$\min_{\ell \in \mathbb{R}^{t_\ell}} f_\ell(\ell) := \frac{1}{2}\|\mathcal{K}(V\,\mathcal{L}\mathrm{Triag}(\ell)(V\,\mathcal{L}\mathrm{Triag}(\ell))^T) - \bar{D}\|_F^2 =: \frac{1}{2}\|F_\ell(\mathcal{L}\mathrm{Triag}(\ell))\|_F^2, \qquad (2.8)$$

*which defines the function $F_\ell : \mathbb{R}^{n-1 \times d} \to \mathcal{S}_H^n$. Does $\ell^* \in \mathbb{R}^{t_\ell}$ exist such that it is a **lngm** for* (2.8)?

Remark 2.4. *Analogously to the main Problem* 2.1*, this reduced problem* (2.8) *is a nonlinear least squares problem but with a further reduced number of variables $t_\ell \leq (n-1)d < nd$ in ℓ, and still $t(n-1) = n(n-1)/2$ quadratic functions $\left(\mathcal{K}(V\,\mathcal{L}\mathrm{Triag}(\ell)(V\,\mathcal{L}\mathrm{Triag}(\ell))^T) - \bar{D}\right)_{ij}, i < j$. When $d < n-1$, the underlying system of equations is overdetermined, and it is square otherwise (for $d \geq n-1$, from* (2.5) *we observe that the number of variables is $t(n-1)$, the same as the number of quadratic equations).*

In the following, we shall present conditions for determining whether a **lngm** exists. But first, the next section provides useful formulae for linear transformations and derivatives.

3 Properties and auxiliary results

We now provide appropriate notation and formulae for transformations, adjoints and derivatives involved in **EDM**, and then give the equivalence relationships among local minimizers of the three reformulations.

3.1 Transformations, Derivatives, Adjoints, Range and Null Spaces

We now provide Lemma 3.1 with a list of results including proofs that follow immediately after the result is presented.

Lemma 3.1. *Let*

$$P \in \mathbb{R}^{n \times d},\, p = \mathrm{vec}\, P \in \mathbb{R}^{nd}, \Delta P \in \mathbb{R}^{n \times d},\, \Delta p = \mathrm{vec}\, \Delta P \in \mathbb{R}^{nd},$$

$$L \in \mathbb{R}^{n-1 \times d},\, \ell \in \mathbb{R}^{t_\ell},\, S, T \in \mathbb{S}^n.$$

We use the previously defined functions:

$$\mathcal{M} : \mathbb{R}^{n \times d} \to \mathbb{S}^n,\, \mathcal{K} : \mathbb{S}^n \to \mathbb{S}^n,\, F : \mathbb{R}^{n \times d} \to \mathbb{S}^n,\, f : \mathbb{R}^{n \times d} \to \mathbb{R},$$

$$\mathcal{L}\mathrm{Triag} : \mathbb{R}^{t_\ell} \to \mathbb{R}^{n-1 \times d}, F_L : \mathbb{R}^{n-1 \times d} \to \mathbb{S}^n,\, f_L : \mathbb{R}^{n-1 \times d} \to \mathbb{R},$$

$$F_\ell : \mathbb{R}^{t_\ell} \to \mathbb{S}^n, f_\ell : \mathbb{R}^{t_\ell} \to \mathbb{R}.$$

Then, the following and their proofs hold:

1. $\mathcal{M}'(P)(\Delta P) = P\Delta P^T + \Delta P P^T$, $\mathcal{M}''(P)(\Delta P, \Delta P) = 2\Delta P \Delta P^T$: *follow directly from the expansion*

$$\begin{aligned}\mathcal{M}(P + \Delta P) &= (P + \Delta P)(P + \Delta P)^T \\ &= PP^T + \Delta P P^T + P\Delta P^T + \Delta P \Delta P^T \\ &= \mathcal{M}(P) + \mathcal{M}'(P)(\Delta P) + \tfrac{1}{2}\mathcal{M}''(P)(\Delta P, \Delta P).\end{aligned}$$

2. $\mathcal{M}'(P)^*(S) = 2SP$:

$$
\begin{aligned}
\langle \mathcal{M}'(P)(\Delta P), S \rangle &= \langle P\Delta P^T + \Delta P P^T, S \rangle \\
&= \operatorname{tr}(P\Delta P^T S + \Delta P P^T S) \\
&= \operatorname{tr}(SP\Delta P^T + SP\Delta P^T) \\
&= \langle \mathcal{M}'(P)^*(S), \Delta P \rangle.
\end{aligned}
$$

3. $S_e : \mathbb{R}^n \to \mathbb{S}^n$, $S_e(v) = ve^T + ev^T$: *is a definition.*

4. $S_e^*(S) = 2Se$:

$$
\langle S_e(v), S \rangle = \operatorname{tr}(ev^T S + ve^T S) = \operatorname{tr}(v^T Se) + \operatorname{tr}(Sev^T) = \langle 2Se, v \rangle.
$$

5. $\mathcal{K}(G) = S_e(\operatorname{diag}(G)) - 2G$, $\operatorname{range}(\mathcal{K}) = \mathcal{S}_H^n, \operatorname{null}(\mathcal{K}) = \operatorname{range}(S_e)$:　*see [1, Prop. 2.2].*

6. $\operatorname{Diag} = \operatorname{diag}^*$: *clear from the definitions.*

7. $\mathcal{K}^*(S) = 2(\operatorname{Diag}(Se) - S)$, $\operatorname{range}(\mathcal{K}^*) = \mathcal{S}_C^n$; $\operatorname{null}(\mathcal{K}^*) = \operatorname{Diag}(\mathbb{R}^n)$:　*see [1, Prop. 2.2].*
   Moreover, $S \geq (\leq)0 \implies \mathcal{K}^*(S) \succeq (\preceq)0$.
   $\mathcal{K}^*(S) = 2(\operatorname{Diag}(Se) - S)$:
   This follows from

$$
\begin{aligned}
\langle \mathcal{K}(T), S \rangle &= \langle \operatorname{diag}(T)e^T + e\operatorname{diag}(T)^T - 2T, S \rangle \\
&= 2\operatorname{tr}(e^T S \operatorname{diag}(T)) - 2\operatorname{tr}(TS) \\
&= 2\langle Se, \operatorname{diag}(T) \rangle - 2\langle T, S \rangle \\
&= 2\langle \operatorname{Diag}(Se) - S, T \rangle.
\end{aligned}
$$

   The last equality is due to $\operatorname{Diag} = \operatorname{diag}^*$.
   Moreover, for $S \in \mathbb{S}^n$, we have by diagonal dominance that $S \geq (\leq)0 \implies \mathcal{K}^*(S) \succeq (\preceq)0$.

8. $\mathcal{D} : \mathbb{R}^{n \times d} \to \mathbb{S}^n$, $\mathcal{D}(P) = \mathcal{K}(\mathcal{M}(P)) = S_e(\operatorname{diag}(\mathcal{M}(P))) - 2\mathcal{M}(P)$.

9. $\mathcal{D}'(P)(\Delta P) = S_e(\operatorname{diag}(\mathcal{M}'(P)(\Delta P)) - 2\mathcal{M}'(P)(\Delta P)$: *this follows from the linearity of diag and S_e.*

10. $F : \mathbb{R}^{n \times d} \to \mathbb{S}^n$, $F(P) = \mathcal{K}(\mathcal{M}(P)) - \bar{D}$: *follows from the definitions.*

11. $F'(P)(\Delta P) = \mathcal{K}\left(\mathcal{M}'(P)(\Delta P)\right)$ *and* $F''(P)(\Delta P, \Delta P) = \mathcal{K}\left(\mathcal{M}''(P)(\Delta P, \Delta P)\right)$: *both follow from the definitions and linearity of $\mathcal{K}$.*

12. $F'(P)^*(S) = \mathcal{M}'(P)^*(\mathcal{K}^*(S)) = 4(\operatorname{Diag}(Se) - S)P$:

$$
\begin{aligned}
\langle F'(P)(\Delta P), S \rangle &= \langle \mathcal{K}(\mathcal{M}'(P)(\Delta P)), S \rangle \\
&= \langle \mathcal{M}'(P)(\Delta P), \mathcal{K}^*(S) \rangle \\
&= \langle \Delta P, \mathcal{M}'(P)^*(\mathcal{K}^*(S)) \rangle.
\end{aligned}
$$

Then, the proof is complete by bringing $\mathcal{M}'(P)^*$ and $\mathcal{K}^*(S)$ into the formula.

13. Denote the symmetrization linear transformation $\mathcal{S} : \mathbb{R}^{n \times n} \to \mathbb{S}^n, \mathcal{S}(K) = (K + K^T)/2$, and $\mathcal{T}$ for the transpose operation whose adjoint is transpose again. Then,

$$f'(P) = F'(P)^*(F(P)) = 4[\text{Diag}(F(P)e) - F(P)]P \tag{3.1}$$

and

$$
\begin{aligned}
f''(P)(\Delta P, \Delta P) &= \langle \mathcal{K}(P\Delta P^T + \Delta P P^T), \mathcal{K}(P\Delta P^T + \Delta P P^T) \rangle \\
&\quad + 2\langle F(P), \mathcal{K}(\Delta P \Delta P^T) \rangle.
\end{aligned} \tag{3.2}
$$

From the expansion of $f(P + \Delta P)$,

$$
\begin{aligned}
& f(P + \Delta P) \\
&= \tfrac{1}{2}\langle F(P + \Delta P), F(P + \Delta P) \rangle \\
&= \tfrac{1}{2}\|F(P) + F'(P)(\Delta P) + \tfrac{1}{2}F''(P)(\Delta P, \Delta P) + o(\|\Delta P\|^2)\|^2, \\
&= \tfrac{1}{2}\langle F(P), F(P) \rangle + \langle F(P), F'(P)(\Delta P) \rangle \\
&\quad + \tfrac{1}{2}\langle F'(P)(\Delta P), F'(P)(\Delta P) \rangle + \tfrac{1}{2}\langle F(P), F''(P)(\Delta P, \Delta P) \rangle + o(\|\Delta P\|^2),
\end{aligned}
$$

we get (3.1), and with Item 11, we obtain

$$
\begin{aligned}
& f''(P)(\Delta P, \Delta P) \\
&= \langle F'(P)(\Delta P), F'(P)(\Delta P) \rangle + \langle F(P), F''(P)(\Delta P, \Delta P) \rangle \\
&= \langle \mathcal{K}(\mathcal{M}'(P)(\Delta P)), \mathcal{K}(\mathcal{M}'(P)(\Delta P)) \rangle + \langle F(P), \mathcal{K}(\mathcal{M}''(P)(\Delta P, \Delta P)) \rangle \\
&= \langle \mathcal{K}(P\Delta P^T + \Delta P P^T), \mathcal{K}(P\Delta P^T + \Delta P P^T) \rangle + 2\langle F(P), \mathcal{K}(\Delta P \Delta P^T) \rangle.
\end{aligned} \tag{3.3}
$$

Note that we can isolate the matrix representation with

$$
\begin{aligned}
f''(P)(\Delta P, \Delta P) &= \langle f''(P)(\text{Mat}\,\text{vec}(\Delta P)), \text{Mat}\,\text{vec}(\Delta P) \rangle \\
&= \langle [\text{vec}\, f''(P)\,\text{Mat}]\,(\Delta p), (\Delta p) \rangle.
\end{aligned} \tag{3.4}
$$

The first term in (3.3) is

$$
\begin{aligned}
& 4\langle \mathcal{K}(\mathcal{S}(P(\text{Mat}\,\text{vec}\,\Delta P)^T)), \mathcal{K}(\mathcal{S}(P(\text{Mat}\,\text{vec}\,\Delta P)^T)) \rangle \\
&= 4\langle (P^T\mathcal{S}^*\mathcal{K}^*\mathcal{K}\mathcal{S}P)((\text{Mat}\,\text{vec}\,\Delta P)^T), (\text{Mat}\,\text{vec}\,\Delta P)^T \rangle \\
&= 4\langle (P^T\mathcal{S}^*\mathcal{K}^*\mathcal{K}\mathcal{S}P)(\mathcal{T}\,\text{Mat}\,\text{vec}\,\Delta P), (\mathcal{T}\,\text{Mat}\,\text{vec}\,\Delta P) \rangle \\
&= 4\left\langle [\text{vec}\,\mathcal{T}^*P^T\mathcal{S}^*\mathcal{K}^*\mathcal{K}\mathcal{S}P\mathcal{T}\,\text{Mat}]\,\Delta p, \Delta p \right\rangle.
\end{aligned} \tag{3.5}
$$

The second term in (3.3) is

$$
\begin{aligned}
& 2\langle F(P), \mathcal{K}(\Delta P \Delta P^T) \rangle \\
&= 2\left\langle \mathcal{K}^*(F(P)), \Delta P \Delta P^T \right\rangle \\
&= 2\left\langle \Delta P, \mathcal{K}^*(F(P))\Delta P \right\rangle \\
&= 2\left\langle [\text{vec}\,\mathcal{K}^*F(P)\,\text{Mat}]\,\Delta p, \Delta p \right\rangle.
\end{aligned} \tag{3.6}
$$

Recall that $F'(P)(\Delta P) = \mathcal{K}(\mathcal{M}'(P)(\Delta P))$. We combine (3.5) and (3.6) and obtain the matrix representation of the Hessian (not necessarily positive semidefinite) :

$$
\begin{aligned}
[\text{vec}\, f''(P)\,\text{Mat}] &= 4\left[\text{vec}\,\mathcal{T}^*P^T\mathcal{S}^*\mathcal{K}^*\mathcal{K}\mathcal{S}P\mathcal{T}\,\text{Mat}\right] \\
&\quad + 2\left[\text{vec}\,\mathcal{K}^*F(P)\,\text{Mat}\right] \\
&= 4\left[J^*J\right] + 2\left[\text{vec}\,\mathcal{K}^*F(P)\,\text{Mat}\right],
\end{aligned} \tag{3.7}
$$

where

$$J(\Delta p) := \mathcal{K}\mathcal{S}P\mathcal{T}\,\text{Mat}\,\Delta p. \tag{3.8}$$

8

$\square$

Theorem 3.2. *The second order necessary optimality conditions for Problem 2.1, (2.2), are*

$$0 \;\; = \;\; f'(P) = F'(P)^*(F(P)) = 4\mathcal{K}^*(F(P))P \tag{3.9}$$

and

$$0 \;\; \preceq \;\; [\operatorname{vec} f''(P)\operatorname{Mat}] = 4\,[J^*J] + 2\,[\operatorname{vec}\mathcal{K}^*(F(P))\operatorname{Mat}]. \tag{3.10}$$

Proof. This follows from (3.7) and Lemma 3.1, Item 13. $\square$

Throughout the paper, we denote the images of the following two mappings as two matrices in $\mathbb{S}^{nd}$:

$$H_1 = [J^*J], \;\; H_2 = [\operatorname{vec}(\mathcal{K}^*(F(P))\operatorname{Mat}]. \tag{3.11}$$

We call P a stationary point if (3.9) holds.

3.2 Optimality conditions of three problem formulations

According to the chain rule, the derivatives and optimality conditions of f_L defined in problem (2.4) and f_ℓ defined in problem (2.8) can be easily obtained from that of f.

Proposition 3.3. *The derivatives of $f_L(L)$ are*

$$f'_L(L) = V^T f'(VL) \tag{3.12}$$

and

$$f''_L(L) = V^T f''(VL)V. \tag{3.13}$$

Proposition 3.4. *The derivatives of $f_\ell(\ell)$ are*

$$f'_\ell(\ell) = \mathcal{L}\operatorname{Triag}^* f'_L(\mathcal{L}\operatorname{Triag}(\ell)) \tag{3.14}$$

and

$$f''_\ell(\ell) = \mathcal{L}\operatorname{Triag}^* f''_L(\mathcal{L}\operatorname{Triag}(\ell))\,\mathcal{L}\operatorname{Triag}. \tag{3.15}$$

Before and after the two reduction processes, problems are equivalent in regards to the objective functions and global minimizers. In the following, we show that the optimality conditions of the three problem reformulations (2.2), (2.4), and (2.8) are equivalent to some extent.

Firstly, we find that any local minimizer of (2.2) or (2.4) corresponds to a family of local minimizers which are obtained from translations or rotations.

Theorem 3.5. *The configuration $P^* \in \mathbb{R}^{n\times d}$ is a local minimizer of (2.2) if, and only if, any configuration in $\{P_v^* = P^* + v^T \otimes e : v \in \mathbb{R}^d\}$ is a local minimizer of (2.2).*

Proof. By the definition of the local minimizer, if P^* is a local minimizer, then there exists $\delta > 0$ such that

$$f(P^*) \le f(P), \;\; \forall P \; : \; \|P - P^*\| \le \delta. \tag{3.16}$$

Since $f(\cdot)$ is invariant under any translation, (3.16) is equivalent to

$$f(P_v^*) = f(P^*) \le f(P) = f(\tilde{P}), \;\; \forall \tilde{P} \; : \; \|\tilde{P} - P_v^*\| = \|P - P^*\| \le \delta,$$

where $P = \tilde{P} - v^T \otimes e$. $\square$

Theorem 3.6. *The configuration $L^* \in \mathbb{R}^{n-1 \times d}$ is a local minimizer of* (2.4) *if, and only if,* $\{\tilde{L}^* = L^* Q^T : Q \in \mathcal{O}\}$ *is a local minimizer set of* (2.4).

Proof. By the definition of the local minimizer, L^* is a local minimizer, if, and only if, there exists $\delta > 0$ such that

$$f_L(L^*) \leq f_L(L), \ \forall L \ : \ \|L - L^*\| \leq \delta. \tag{3.17}$$

Since $f_L(\cdot)$ is invariant under any rotation, (3.17) is equivalent to

$$f_L(\tilde{L}^*) = f_L(L^*) \leq f_L(L) = f_L(\tilde{L}), \ \forall \tilde{L} \ : \ \|\tilde{L} - \tilde{L}^*\| = \|L - L^*\| \leq \delta,$$

where $\tilde{L} = LQ^T$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Local minimizers of (2.2) and (2.4) have the following relationships.

Theorem 3.7. *Let $P^* \in \mathbb{R}^{n \times d}$ and V be as defined in* (2.3). *Denote*

$$v^* = \frac{1}{n} P^{*T} e \in \mathbb{R}^d, \ P_v^* = P^* - (v^{*T} \otimes e), \ L^* = V^T P_v^*.$$

Then, L^* is a local minimizer of (2.4) *if, and only if, P_v^* and P^* are local minimizers of* (2.2).

Proof. Firstly, notice that VV^T is the orthogonal projection onto $e^\perp$ and that the columns of P_v^* are centered, thus, we have $VL^* = VV^T P_v^* = P_v^*$. Sufficiency: Let P_v^* be a local minimizer of (2.2), Then, there exists $\delta > 0$ such that

$$f(P_v^*) \leq f(P), \ \forall P \ : \ \|P - P_v^*\| \leq \delta. \tag{3.18}$$

For any $L \in \mathbb{R}^{n-1 \times p}$ such that $\|L - L^*\| \leq \delta$, let $\hat{P} = VL$. Then, we have

$$f_L(L^*) = f(VL^*) = f(P_v^*) \leq f(\hat{P}) = f(VL) = f_L(L),$$

where the inequality is due to $\|\hat{P} - P_v^*\| = \|VL - VL^*\| = \|L - L^*\| \leq \delta$ and (3.18), and the equalities are due to the definition of f_L.

Necessity: Suppose that L^* is a local minimizer of $f_L(L)$, i.e., there exists $\delta > 0$ such that

$$f_L(L^*) \leq f_L(L), \ \forall L \ : \ \|L - L^*\| \leq \delta. \tag{3.19}$$

For all P such that $\|P - P_v^*\| \leq \delta$, let $v = P^T e / n$. Then, there exists $L \in \mathbb{R}^{n-1,d}$ such that $P = VL + v^T \otimes e$, which implies that $P - P_v^* = V(L - L^*) + v^T \otimes e$. As $V(L - L^*)$ and $v^T \otimes e$ are orthogonal,

$$\|L - L^*\|^2 = \|V(L - L^*)\|^2 = \|P - P_v^*\|^2 - \|v^T \otimes e\|^2 \leq \delta^2. \tag{3.20}$$

Now, from (3.19) and (3.20), we have

$$f(P) = f(VL + v^T \otimes e) = f(VL) \geq f(VL^*) = f(P_v^*),$$

implying that P_v^* is a local minimizer of $f(P)$; and, by Theorem 3.5 we also have P^* is a local minimizer of $f(P)$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

From the above analysis, we know that for the case of $d \geq 2$, if L is a local minimizer of $f_L(L)$, then any configuration in $\{LQ : Q \in \mathcal{O}\}$ is also a local minimizer of $f_L(L)$. When $d \geq 2$, any local minimizer of $f_L(L)$ is nonisolate, thus the Hessian matrix at any local minimizer of $f_L(L)$ is singular.

Next, we consider the correspondence between the local minimizers of (2.4) and its rotation-reduction formulation (2.8).

Theorem 3.8. *The following assertions hold.*

1. *If L^* is a local minimizer of f_L, then any ℓ^* satisfying*

$$L^* = \mathcal{L}\mathrm{Triag}(\ell^*)Q^T \tag{3.21}$$

   with $Q \in \mathcal{O}$, is a local minimizer of f_ℓ.

2. *If ℓ^* is a local minimizer of f_ℓ, and the first d rows of $\mathcal{L}\mathrm{Triag}(\ell^*)$ are linearly independent, then $L^* = \mathcal{L}\mathrm{Triag}(\ell^*)$ is a local minimizer of f_L.*

Proof.   1. Suppose that L^* is a local minimizer of f_L, i.e., there exists $r > 0$ such that

$$f_L(L) \geq f_L(L^*), \text{ for all } L \text{ satisfying } \|L - L^*\|_F \leq r. \tag{3.22}$$

   For any $\ell \in \mathbb{R}^{t_\ell}$ satisfying

$$\|\ell - \ell^*\| \leq r,$$

   by taking $L = \mathcal{L}\mathrm{Triag}(\ell)Q^T$ we obtain

$$\|L - L^*\| = \|\mathcal{L}\mathrm{Triag}(\ell) - \mathcal{L}\mathrm{Triag}(\ell^*)\|_F = \|\ell - \ell^*\| \leq r.$$

   Then, from (3.21) and (3.22) we have

$$f_\ell(\ell) = f_L(\mathcal{L}\mathrm{Triag}(\ell)) = f_L(L) \geq f_L(L^*) = f_L(\mathcal{L}\mathrm{Triag}(\ell^*)) = f_\ell(\ell^*).$$

   Thereore, ℓ^* is a local minimizer of f_ℓ.

2. We prove Item 2 by contradiction. Suppose that $L^* = \mathcal{L}\mathrm{Triag}(\ell^*)$ is not a local minimizer of f_L, i.e., there exists a sequence $L_k, k = 1, 2, \cdots$ such that

$$\lim_{k \to +\infty} L_k = L^*, \ f_L(L_k) < f_L(L^*). \tag{3.23}$$

   Considering the QR decomposition of $L_k^T, k = 1, 2, \cdots$, there exist orthogonal matrices $Q_k, k = 1, 2, \cdots$, and upper triangular matrices $R_k, k = 1, 2, \cdots$, such that

$$L_k^T = Q_k R_k, k = 1, 2, \cdots. \tag{3.24}$$

   Since $\|Q_k\| \leq 1$ for all $k = 1, 2, \cdots$, the bounded sequence has a convergent subsequence. Without loss of generality, we directly assume that $\lim_{k \to +\infty} Q_k = Q^*$. According to (3.23) and (3.24), we have

$$R^{*T} := L^* Q^* = \lim_{k \to +\infty} L_k Q_k = \lim_{k \to +\infty} R_k^T.$$

11

Since the first d rows of $L^* := \mathcal{L}\text{Triag}(\ell^*)$ are linear independent, the QR-factorization of L^{*T} is unique except for signs in every dimension, i.e., with diagonal $Q^* = \Lambda$, $\Lambda^2 = I_d$ and $R^* = \Lambda L^{*T}$. Considering

$$\ell^* = \mathcal{L}\text{Triag}^*(L^*) = \mathcal{L}\text{Triag}^*(R^{*T}\Lambda), \ \ell_k := \mathcal{L}\text{Triag}^*(R_k^T \Lambda), k = 1, 2, \cdots,$$

we get

$$\lim_{k \to +\infty} \ell_k = \ell^*.$$

By (3.23), we have

$$f_\ell(\ell_k) = f_L(R_k^T) = f_L(L_k) < f_L(L^*) = f_L(R^{*T}Q^{*T})$$
$$= f_L(R^{*T}) = f_\ell(\ell^*).$$

Thus, ℓ^* is not a local minimizer of f_ℓ, a contradiction.

$\square$

The optimality conditions of (2.2) and (2.4) also have an equivalence relationship. To this end, we first note that the directional derivatives of $f(P)$ are zero in any translation.

Lemma 3.9. *For any $w \in \mathbb{R}^d$, we have*

$$\langle f'(P), w^T \otimes e \rangle = 0, \text{ and } f''(P)(w^T \otimes e) = 0. \tag{3.25}$$

Proof. For $t \in \mathbb{R}$, we have

$$f(P + tw^T \otimes e) = f(P) + t\langle f'(P), w^T \otimes e \rangle + \tfrac{t^2}{2}\langle f''(P)(w^T \otimes e), w^T \otimes e \rangle + o(t^2).$$

Since $f(P + tw^T \otimes e) = f(P)$ holds for all $w \in \mathbb{R}^d$ and $t \in \mathbb{R}$, we get

$$\langle f'(P), w^T \otimes e \rangle = \langle f''(P)(w^T \otimes e), w^T \otimes e \rangle = 0.$$

Moreover, by the form of $f''(P)$ in (3.3) and the fact that $\text{range}(\mathcal{K}^*) \subset \{w^T \otimes e : w \in \mathbb{R}^d\}^\perp$ (Item 7 of Lemma 3.1), we have $f''(P)(w^T \otimes e) = 0$. $\square$

Theorem 3.10. *For $P \in \mathbb{R}^{n \times d}$, denote $v = (P^T e)/n \in \mathbb{R}^d$, $P_v = P - v^T \otimes e$, $L = V^T P_v$ where V is defined in (2.3), denote $L^T = QR$ where R is upper triangular and $Q \in \mathbb{R}^{d \times d}$ is orthogonal. Then, the following conditions are equivalent:*

(i) *the first (resp., second)-order necessary conditions of (2.2) hold at P;*

(ii) *the first (resp., second)-order necessary conditions of (2.2) hold at P_v;*

(iii) *the first (resp., second)-order necessary conditions of (2.4) hold at L;*

(iv) *the first (resp., second)-order necessary conditions of (2.4) hold at R.*

Proof. We first notice that

$$f'(P_v) = f'(P), \ f''(P_v) = f''(P) \tag{3.26}$$

and

$$f'_L(R) = f'_L(L)Q^T, \ f''_L(R) = Qf''_L(L)Q^T \tag{3.27}$$

by the chain rule. So, it is straightforward that (i)⇔(ii) and (iii)⇔(iv). In the following, we prove that (ii)⇔(iii) .

Firstly, we prove the equivalence of their first-order necessary conditions. According to (3.1) and range$(\mathcal{K}^*) = \mathcal{S}_C^n$ (Lemma 3.1, Item 7), we have

$$e^T f'(P) = 2e^T \mathcal{K}^*(F(P))P = 0. \tag{3.28}$$

By (3.28) and the definition of V, we obtain

$$f'(P_v) = 0 \Longleftrightarrow f'_L(L) = V^T f'(P_v) = 0. \tag{3.29}$$

Secondly, we prove the equivalence of their second-order necessary conditions. According to (3.13), for any $\Delta L \in \mathbb{R}^{n-1 \times d}$, we have

$$f''_L(L)(\Delta L, \Delta L) = V^T f''(VL)V(\Delta L, \Delta L) = f''(VL)(V\Delta L, V\Delta L). \tag{3.30}$$

According to Lemma 3.9 and (3.30), we have $f''_L(L)(\Delta L, \Delta L) \geq 0$ if, and only if,

$$f''(P_v)(\Delta P, \Delta P) = f''(P_v)(V\Delta L + w^T \otimes e, V\Delta L + w^T \otimes e) \geq 0.$$

The proof is complete. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

Remark 3.11. *The reduction from (2.4) to (2.8) can introduce additional stationary points. Let* $\mathcal{L}\mathrm{Triag}(\ell^*) = R^{*T}$. *According to (3.14),* $f'_\ell(\ell) = 0$ *holds if, and only if, the lower triangular part of* $f'_L(R^{*T})$ *is zero. Moreover, the linear independence assumption in Theorem 3.8, Item 2 is needed.*

4 Second Order Conditions

We now consider the optimality conditions and give a sufficient condition such that there is no **lngm**. First of all, the necessary and sufficient characterization for the global minimizer is made clear.

Lemma 4.1. *A matrix* $P \in \mathbb{R}^{n \times d}$ *is a* **global minimizer** *of (2.2) if, and only if,* $\mathcal{D}(P) = \bar{D}$.

Proof. Since $f(P) \geq 0$ holds for all $P \in \mathbb{R}^{n \times d}$ and $f(\bar{P}) = 0$, the global minimum of f is 0. By the definition of f and property of norms , $f(P) = 0$ holds if, and only if, $F(P) = \mathcal{D}(P) - \bar{D} = 0$. □

In order to further characterize the second order optimality conditions we shall discuss essential properties of the matrices H_1 and H_2.

Lemma 4.2. *The matrix* H_1 *defined in (3.11) is always positive semidefinite. The matrix* $H_2 \succeq 0$ *(resp.,* $H_2 \preceq 0$*) holds if every element of* $F(P)$ *is nonnegative (resp., nonpositive), i.e.,* $F(P) \geq 0$ *(resp.,* $F(P) \leq 0$*).*

Proof. For any $x \in \mathbb{R}^{nd}$,
$$x^T H_1 x = \langle x, J^* J x \rangle = \langle J x, J x \rangle \geq 0.$$

Thus, H_1 is always positive semidefinite. By Lemma 3.1, Item 7, if $F(P) \geq 0$, then $\mathcal{K}^*(F(P)) \succeq 0$ holds, i.e.,
$$x^T H_2 x = \langle x, \mathrm{Mat}^* \mathcal{K}^* F(P) \mathrm{Mat}\, x \rangle = \langle \mathrm{Mat}\, x, \mathcal{K}^*(F(P)) \mathrm{Mat}\, x \rangle \geq 0$$

holds for all $x \in \mathbb{R}^{nd}$. Thus, $H_2 \succeq 0$ (resp., $H_2 \preceq 0$) if $F(P) \geq 0$ (resp., $F(P) \leq 0$). $\qquad\square$

Lemma 4.3. *The matrix H_2 is the zero matrix if, and only if, $F(P) = 0$ holds, i.e., P is a global minimizer.*

Proof. By Lemma 3.1, Item 7, $\mathcal{K}^*(S) = 2(\mathrm{Diag}(Se) - S)$ for any $S \in \mathbb{S}^n$ and $\mathrm{null}(\mathcal{K}^*) = \mathrm{Diag}(\mathbb{R}^n)$. Since $\mathrm{diag}(F(P)) = \mathrm{diag}(\mathcal{D}(P)) - \mathrm{diag}(\bar{D}) = 0$ is always true, we get $\mathcal{K}^*(F(P)) = 0$ holds if, and only if, $F(P) = 0$. $\qquad\square$

Lemma 4.4. *Suppose that P is a stationary point for (2.2) but is not a global optimum. Then H_2 is not positive semidefinite, i.e., $\bar{p}^T H_2 \bar{p} < 0$, for $\bar{p} = \mathrm{vec}(\bar{P})$.*

Proof. By (3.9) we have
$$
\begin{aligned}
\bar{p}^T H_2 \bar{p} &= \langle \bar{P}, \mathcal{K}^* F(P) \bar{P} \rangle - \langle P, \mathcal{K}^* F(P) P \rangle \\
&= \langle \mathcal{K}(\bar{P}\bar{P}^T), F(P) \rangle - \langle \mathcal{K}(PP^T), F(P) \rangle \\
&= \langle \mathcal{K}(\bar{P}\bar{P}^T) - \mathcal{K}(PP^T), F(P) \rangle \\
&= \langle \bar{D} - \mathcal{D}(P), F(P) \rangle \\
&= -\langle F(P), F(P) \rangle \\
&< 0.
\end{aligned}
$$

The last inequality holds since P is not a global minimizer, i.e., $F(P) \neq 0$ by Lemma 4.1. $\qquad\square$

Under the condition of Lemma 4.4, we have known that
$$\langle \bar{P}, \mathcal{K}^* F(P) \bar{P} \rangle < 0,$$

which implies that
$$\mathcal{K}^* F(P) \not\succeq 0. \tag{4.1}$$

Otherwise,
$$\langle \bar{P}, \mathcal{K}^* F(P) \bar{P} \rangle = \mathrm{Tr}\big(\mathcal{K}^* F(P) \bar{P}\bar{P}^T\big) \geq 0,$$

which contradicts Lemma 4.4.

We analyze the extreme case of $\bar{D} = 0$.

Corollary 4.5. *If $\bar{D} = 0$, then every stationary point is a global minimizer.*

Proof. As $\bar{D} = 0$, we get $\bar{p}_1 = \cdots = \bar{p}_n$ holds. Since $\mathcal{K}^* F(P)$ is a Laplacian[2], we have $\mathcal{K}^* F(P)\bar{P} = 0$. Combining this with the first-order condition (3.9), we have
$$
\begin{aligned}
\bar{p}^T H_2 \bar{p} &= \langle \bar{P}, \mathcal{K}^*(F(P))\bar{P} \rangle - \langle P, \mathcal{K}^*(F(P))P \rangle \\
&= 0.
\end{aligned}
$$

From Lemma 4.4 we conclude P is a global minimizer. $\qquad\square$

---

[2]sum of its columns is zero

Next, we consider another extreme case, that of $\bar{D} \neq 0$ and $\mathcal{D}(P) = 0$.

Theorem 4.6. *Let $\bar{D} \neq 0$, and let P satisfy $\mathcal{D}(P) = 0$, i.e., $p_1 = \cdots = p_n$. Then P is a stationary point but the Hessian matrix is nonzero and negative semidefinite, i.e.,*

$$0 \neq 4H_1 + 2H_2 \preceq 0.$$

Proof. Since $p_1 = \cdots = p_n$ and $\mathcal{K}^*(F(P))$ is a Laplacian, P satisfies the first-order optimality condition (3.9). Since $f(P) = \|\bar{D} - \mathcal{D}(P)\|_F^2 = \|\bar{D}\|_F^2 > 0$, P is not a global minimizer. By $p_1 = \cdots = p_n$, and Lemma 3.1, Item 5, $J = 0$ defined in (3.8) holds, and then $H_1 = 0$. Since $\mathcal{D}(P) = 0$ and $\bar{D} \geq 0$, $F(P) \leq 0$ holds. According to Lemma 4.2 and Lemma 4.3, $0 \neq H_2 \preceq 0$ holds. Therefore, the Hessian matrix satisfies $0 \neq 4H_1 + 2H_2 \preceq 0$. $\square$

Remark 4.7. *In fact, one can show that if P is a local **maximizer** of f, then $\mathcal{D}(P) = 0$, i.e., $p_1 = \cdots = p_n$. Indeed, if P is a local **maximizer** of f, then $t = 1$ is a local maximizer of $g(t) = f(tP)$. By the second order necessary conditions, we have $g'(1) = 0$ and $g''(1) \leq 0$. Since $g'(t) = 4t^3 \|\mathcal{D}(P)\|_F^2 - 4t\langle \bar{D}, \mathcal{D}(P) \rangle$, $0 = g'(1)$ implies that $\|\mathcal{D}(P)\|_F^2 = \langle \bar{D}, \mathcal{D}(P) \rangle$. Then, $0 \geq g''(1) = 8\|\mathcal{D}(P)\|_F^2$ and we conclude that $\mathcal{D}(P) = 0$.*

In the following, we present the condition under which there is no **lngm**. Recall the equivalence between local minimizers of (2.4) and (2.2) in Theorem 3.7. For convenience, we use (2.4) for the analysis instead of (2.2).

Theorem 4.8. *Any stationary point L of (2.4) satisfying $\text{rank}(L) = n - 1$ is a global minimizer.*

Proof. Since $f_L'(L) = V^T \mathcal{K}^* F(VL) VL = 0$, the span of columns of L is an $n-1$ dimensional eigenvector space corresponding to the zero eigenvalue of the $(n-1) \times (n-1)$ matrix $V^T \mathcal{K}^* F(VL) V$. Therefore $V^T \mathcal{K}^*(F(P))V = 0$. Combining this with $\text{range}(\mathcal{K}^*) = \mathcal{S}_C^n$ from Lemma 3.1, Item 7, we conclude that $\mathcal{K}^*(F(P)) = 0$, and moreover $H_2 = 0$. Thus, L is a global minimizer according to Lemma 4.3. $\square$

As $L \in \mathbb{R}^{n-1 \times d}$, the condition in Theorem 4.8 holds only in the case of $d \geq n - 1$ and L full row rank. Next, we consider another case where L is not full rank.

Theorem 4.9. *Suppose that L is a non-globally-optimal stationary point of (2.4) and*

$$\text{rank}(L) < d. \tag{4.2}$$

Then, the second-order necessary optimality conditions fail at L.

Proof. Denote $P = VL$. According to Lemma 4.4 and the subsequent discussion, (4.1) holds. Thus there exists $a \in \mathbb{R}^n$ such that $a^T \mathcal{K}^* F(P) a < 0$. Then, for any nonzero $w \in \mathbb{R}^d$,

$$
\begin{aligned}
\text{vec}(a \otimes w^T)^T H_2 \, \text{vec}(a \otimes w^T) &= \langle a \otimes w^T, \mathcal{K}^* F(P)(a \otimes w^T) \rangle \\
&= \text{Tr}\big(\mathcal{K}^* F(P)(a \otimes w^T)(a \otimes w^T)^T\big) \\
&= w^T w \, \text{Tr}\big(\mathcal{K}^* F(P) a a^T\big) \\
&= w^T w \, a^T \mathcal{K}^* F(P) a \\
&< 0.
\end{aligned}
$$

By (4.2), there exists a nonzero $w \in \mathbb{R}^d$ such that $w \in \text{null}(L)$, i.e.,

$$Lw = 0. \tag{4.3}$$

We claim that $H_1 \text{vec}(a \otimes w^T) = 0$ holds. First, we have

$$J \text{vec}(a \otimes w^T) = \mathcal{KSV}L\mathcal{T}(a \otimes w^T).$$

By (4.3), we have

$$L\mathcal{T}(a \otimes w^T) = L \begin{bmatrix} a_1 w & a_2 w & \cdots & a_{n-1} w \end{bmatrix} = 0.$$

Thus, $H_1 \text{vec}(a \otimes w^T) = 0$ holds. In sum, we have

$$\text{vec}(a \otimes w^T)^T (4H_1 + 2H_2) \text{vec}(a \otimes w^T) < 0,$$

i.e., the second-order necessary optimality conditions (3.9)-(3.10) fail. □

Combining Theorem 4.8 and Theorem 4.9, we present the main result of this section.

Theorem 4.10. *If $n \leq d + 1$, then any stationary point satisfying the second-order necessary optimality condition is a global minimizer.*

Proof. Suppose that $n \leq d + 1$, and L is a stationary point satisfying the second-order necessary optimality condition. If $\text{rank}(L) = n - 1$, then L is globally optimal by Theorem 4.8. If $\text{rank}(L) < n - 1$, then $\text{rank}(L) < d$. If we assume L is not a global minimizer, according to Theorem 4.9, L does not satisfy the second-order necessary optimality condition, a contradiction. □

Recalling Remark 2.4, we note that $n \leq d + 1$ is exactly the condition such that the underlying system of equations is square. When $n > d + 1$ (overdetermined), it is possible to find local nonglobal minimizers. We prove this claim below by presenting two examples where we are able to numerically obtain second order stationary points. Then, we will analytically prove that the assumptions of the Kantorovich theorem hold at these two points. This implies that in a neighborhood there exist strict **lngm**s.

5 Kantorovich Theorem and Sensitivity Analysis

We now consider the sensitivity analysis needed to analytically prove that we have a **lngm**. We exploit the strength of the classical Kantorovich theorem for convergence of Newton's Method to a stationary point.

For $d = 1$, we consider an example where we have the following: rational matrices $\tilde{L} \in \mathbb{Q}^{n-1 \times 1}, \tilde{P} = V\tilde{L} \in \mathbb{Q}^{n \times 1}$ with the function value $f(\tilde{P}) = f_L(\tilde{L}) > \tilde{f}_L > 0$; the gradient is approximately zero, i.e., we have a near stationary point $\|\nabla f_L(\tilde{L})\| < \tilde{g}_L$, for small $\tilde{g}_L$; and the Hessian is positive definite, $\lambda_{\min}(\nabla^2 f_L(\tilde{L})) > \tilde{\lambda}_L > 0$. Then for $d = 2$, we consider another case with a $\tilde{\ell}$ with similar properties.

We emphasize that $\tilde{P}, \tilde{L}$ have rational entries and that we are doing finite precision arithmetic for which both sensitivity and roundoff error analysis can be done. In our calculations, the roundoff error in every step is less than 10^{-15}, and all of the absolute values of elements in the original data $\tilde{L}$ and $\bar{L}$ are greater than 5×10^{-4}. This yields a relative error of at most 2×10^{-12}. As the number of arithmetic calculations increases, this relative error can accumulate. However, as the number of

all arithmetic operations in this section, including the calculations of gradients, Hessians and all parameters used later, is definitely less than 10^{10}, the relative error in the end will be less than $(2 \times 10^{-12} + 1)^{10^{10}} - 1 < 10^{-1}$. In this section, we choose 10^{-1} as a bound on the relative error contained in the calculated data.

In fact, for our first numerical example, $\tilde{f}_L \approx 2.6 \times 10^3 > 0$, $\tilde{g}_L \approx 1.8 \times 10^{-3}$, $\tilde{\lambda}_L \approx 2.1 \times 10^2 > 0$. We apply sensitivity analysis to guarantee that these properties hold in a r-neighbourhood of $\tilde{L}$ and then apply the classical Kantorovich theorem, e.g., [5, Thm 5.3.1], to show that there is a stationary point nearby where the function value is positive and the Hessian is still positive definite. This provides an analytic proof that we have a proper **lngm** near $\tilde{L}$.

Example 5.1. [3] *An example with $n = 50, d = 1$ is given, with data $\bar{L}, \tilde{L} \in \mathbb{R}^{n-1 \times d}$, see Footnote 3. Matrix $\bar{D}$ is the distance matrix obtained from $\bar{L}$ by $\bar{D} = \mathcal{K}(V\bar{L}(V\bar{L})^T)$. Thus, $\bar{L}$ is a global minimizer, and we verify that $\tilde{L}$ is a numerically convergence point, where the objective value is*

$$f_L(\tilde{L}) > 2.6 \times 10^3, \tag{5.1}$$

the absolute and relative gradient norms are

$$\|\nabla f_L(\tilde{L})\| < 1.8 \times 10^{-3}, \ \frac{\|\nabla f_L(\tilde{L})\|}{1 + f_L(\tilde{L})} \approx 6 \times 10^{-5}, \tag{5.2}$$

and the least eigenvalue of the Hessian matrix is

$$\lambda_{\min}(\nabla^2 f_L(\tilde{L})) > 211. \tag{5.3}$$

Remark 5.2. *The problem to find a **lngm** is a nonlinear least squares problem. The standard approach for nonlinear least squares is to use the Gauss-Newton method rather than the Newton method, i.e., relax the Hessian and only use $4H_1$ (or the corresponding one for L) by discarding the second order terms from $2H_2$ (or the corresponding one for L) in (3.11) and (3.13). However, one major reason for the success of Gauss-Newton is that one expects $f_L(L)$ to be near 0, a root. In our case, we want the opposite as we do not want to be near a root of f_L as that yields the global minimum.*

Now, we calculate a Lipschitz constant estimate $\gamma > 0$ of the Hessian matrix of f_L. From our numerical output, we have that the smallest eigenvalue $\lambda_{\min}(\nabla^2 f_L(\tilde{L})) > 0$. By continuity of eigenvalues, we are guaranteed that this holds in a neighbourhood of $\tilde{L}$, which can be estimated out by Proposition 5.3.

Proposition 5.3. *Let $r > 0, \tilde{L} \in \mathbb{R}^{n-1 \times d}$ be given. If*

$$\gamma \geq 48\sqrt{2}r \left(\sum_{i,j} \|(V\tilde{L})[i,:] - (V\tilde{L})[j,:]\| + 2n\sqrt{n}r \right), \tag{5.4}$$

then γ is a Lipschitz constant for the Hessian of f_L in the radius-r neighborhood of $\tilde{L}$, i.e.,

$$\|\nabla^2 f_L(\hat{L}) - \nabla^2 f_L(\check{L})\| \leq \gamma \|\hat{L} - \check{L}\|, \quad \text{if } \hat{L}, \check{L} \in B_r(\tilde{L}). \tag{5.5}$$

Moreover,

$$\lambda_{\min}(\nabla^2 f_L(L)) \geq \lambda_{\min}(\nabla^2 f_L(\tilde{L})) - \gamma r, \quad \text{if } L \in B_r(\tilde{L}). \tag{5.6}$$

---

[3] The preprint and a link to the data and codes are available at www.math.uwaterloo.ca/~hwolkowi/henry/reports/ABSTRACTS.html

Proof. By the definition of the induced norm, we see that (5.5) is equivalent to

$$|f_L''(\hat{L})(\Delta L, \Delta L) - f_L''(\check{L})(\Delta L, \Delta L)| \leq \gamma \|\hat{L} - \check{L}\|, \tag{5.7}$$

for all $\hat{L}, \check{L} \in B_r(\tilde{L}), \|\Delta L\|_F = 1$. Let

$$\hat{P} = V\hat{L}, \check{P} = V\check{L}, \Delta P = V\Delta L.$$

According to Lemma 3.1 and Proposition 3.3, we have

$$
\begin{aligned}
& f_L''(\hat{L})(\Delta L, \Delta L) \\
= \; & f''(\hat{P})(\Delta P, \Delta P) \\
= \; & \|\mathcal{K}(\hat{P}\Delta P^T + \Delta P \hat{P}^T)\|^2 + 2\langle F(\hat{P}), \mathcal{K}(\Delta P \Delta P^T)\rangle \\
= \; & \sum_{i,j}[2\hat{p}_i^T \Delta p_i + 2\hat{p}_j^T \Delta p_j - 2\hat{p}_i^T \Delta p_j - 2\hat{p}_j^T \Delta p_i]^2 + 2\sum_{i,j}\|\hat{p}_i - \hat{p}_j\|^2 \|\Delta p_i - \Delta p_j\|^2 \\
= \; & 4\sum_{i,j}[(\hat{p}_i - \hat{p}_j)^T(\Delta p_i - \Delta p_j)]^2 + 2\sum_{i,j}\|\hat{p}_i - \hat{p}_j\|^2 \|\Delta p_i - \Delta p_j\|^2.
\end{aligned}
$$

The calculations about $\check{L}$ are similar, implying

$$
\begin{aligned}
& f_L''(\hat{L})(\Delta L, \Delta L) - f_L''(\check{L})(\Delta L, \Delta L) \\
= \; & 4\sum_{i,j}\left\{[(\hat{p}_i - \hat{p}_j)^T(\Delta p_i - \Delta p_j)]^2 - [(\check{p}_i - \check{p}_j)^T(\Delta p_i - \Delta p_j)]^2\right\} \\
& \quad + 2\sum_{i,j}(\|\hat{p}_i - \hat{p}_j\|^2 - \|\check{p}_i - \check{p}_j\|^2)\|\Delta p_i - \Delta p_j\|^2 \\
= \; & 4\sum_{i,j}(\hat{p}_i - \hat{p}_j - \check{p}_i + \check{p}_j)^T(\Delta p_i - \Delta p_j)(\hat{p}_i - \hat{p}_j + \check{p}_i - \check{p}_j)^T(\Delta p_i - \Delta p_j) \\
& \quad + 2\sum_{i,j}(\hat{p}_i - \hat{p}_j - \check{p}_i + \check{p}_j)(\hat{p}_i - \hat{p}_j + \check{p}_i - \check{p}_j)\|\Delta p_i - \Delta p_j\|^2.
\end{aligned}
$$

Then,

$$
\begin{aligned}
& |f_L''(\hat{L})(\Delta L, \Delta L) - f_L''(\check{L})(\Delta L, \Delta L)| \\
\leq \; & 6\sum_{i,j}\|\hat{p}_i - \hat{p}_j - \check{p}_i + \check{p}_j\|\|\hat{p}_i - \hat{p}_j + \check{p}_i - \check{p}_j\|\|\Delta p_i - \Delta p_j\|^2.
\end{aligned}
$$

Since $\|\Delta P\| = \|V\Delta L\| = \|\Delta L\| = 1$,

$$
\begin{aligned}
& \|\Delta p_i - \Delta p_j\|^2 \\
\leq \; & 2(\|\Delta p_i\|^2 + \|\Delta p_j\|^2) \\
\leq \; & 2.
\end{aligned}
$$

From $\hat{L}, \check{L} \in B_r(\tilde{L})$ and the Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
\|\hat{p}_i - \hat{p}_j - \check{p}_i + \check{p}_j\| \leq \; & \|(\hat{p}_i - \tilde{p}_i) - (\hat{p}_j - \tilde{p}_j) - (\check{p}_i - \tilde{p}_i) + (\check{p}_j - \tilde{p}_j)\| \\
\leq \; & \|\hat{p}_i - \tilde{p}_i\| + \|\hat{p}_j - \tilde{p}_j\| + \|\check{p}_i - \tilde{p}_i\| + \|\check{p}_j - \tilde{p}_j\| \\
\leq \; & \sqrt{(\|\hat{p}_i - \tilde{p}_i\|^2 + \|\hat{p}_j - \tilde{p}_j\|^2 + \|\check{p}_i - \tilde{p}_i\|^2 + \|\check{p}_j - \tilde{p}_j\|^2)} \\
& \quad \times \sqrt{1^2 + 1^2 + 1^2 + 1^2} \\
\leq \; & \sqrt{2r^2} \cdot 2 \\
= \; & 2\sqrt{2}r,
\end{aligned}
$$

and

$$
\begin{aligned}
& \sum_{i,j}\|\hat{p}_i - \hat{p}_j + \check{p}_i - \check{p}_j\| \\
= \; & \sum_{i,j}\|2(\tilde{p}_i - \tilde{p}_j) + (\hat{p}_i - \tilde{p}_i) - (\hat{p}_j - \tilde{p}_j) + (\check{p}_i - \tilde{p}_i) - (\check{p}_j - \tilde{p}_j)\| \\
\leq \; & 2\sum_{i,j}\|\tilde{p}_i - \tilde{p}_j\| + \sum_{i,j}\|\hat{p}_i - \tilde{p}_i\| + \sum_{i,j}\|\hat{p}_j - \tilde{p}_j\| + \\
& \quad \sum_{i,j}\|\check{p}_i - \tilde{p}_i\| + \sum_{i,j}\|\check{p}_j - \tilde{p}_j\| \\
= \; & 2\sum_{i,j}\|\tilde{p}_i - \tilde{p}_j\| + n\sum_i\|\hat{p}_i - \tilde{p}_i\| + n\sum_j\|\hat{p}_j - \tilde{p}_j\| + \\
& \quad n\sum_i\|\check{p}_i - \tilde{p}_i\| + n\sum_j\|\check{p}_j - \tilde{p}_j\| \\
\leq \; & 2\sum_{i,j}\|\tilde{p}_i - \tilde{p}_j\| + 4n\sqrt{n}r.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\|f_L''(\hat{L})(\Delta L, \Delta L) - f_L''(\check{L})(\Delta L, \Delta L)\| &\leq \quad 24\sqrt{2}r \sum_{i,j} \|\hat{p}_i - \hat{p}_j + \check{p}_i - \check{p}_j\| \\
&\leq \quad 48\sqrt{2}r \left(\sum_{i,j} \|\tilde{p}_i - \tilde{p}_j\| + 2n\sqrt{n}r \right),
\end{aligned}
$$

implying (5.4) and (5.5). By (5.7), we have

$$
\begin{aligned}
f_L''(L)(\Delta L, \Delta L) = \quad & f_L''(\tilde{L})(\Delta L, \Delta L) - (f_L''(\tilde{L})(\Delta L, \Delta L) - f_L''(L)(\Delta L, \Delta L)) \\
\geq \quad & f_L''(\tilde{L})(\Delta L, \Delta L) - |f_L''(L)(\Delta L, \Delta L) - f_L''(\tilde{L})(\Delta L, \Delta L)| \\
\geq \quad & \lambda_{\min}(\nabla^2 f_L(\tilde{L})) - \gamma\|\hat{L} - \tilde{L}\| \\
\geq \quad & \lambda_{\min}(\nabla^2 f_L(\tilde{L})) - \gamma r, \quad \text{for all } L \in B_r(\tilde{L}), \|\Delta L\| = 1.
\end{aligned}
$$

Thus, we obtain (5.6). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

To verify the existence of a **lngm** for Example 5.1, we calculate the Lipschitz constant estimated in Proposition 5.3. Let $r = 10^{-3}$. Since $2130 > \sum_{i,j} \|(V\tilde{L})[i,:] - (V\tilde{L})[j,:]\| \in (2127.8, 2127.9)$, (5.4) gives

$$
\gamma = 145.
$$

Moreover, by (5.6),

$$
\lambda_{\min}(\nabla^2 f_L(L)) \geq 211 - 145r > 0, \quad \text{for all } L \in B_r(\tilde{L}). \tag{5.8}
$$

That is, we find a neighbourhood where the Hessian stays positive semidefinite. Next, we prove that the objective stays sufficiently positive in a region around $\tilde{L}$.

Lemma 5.4. *Let the configuration* $\tilde{P} = V\tilde{L} \in \mathbb{R}^{n-1\times d}, \tilde{L} \in \mathbb{R}^{n-1\times d}$ *and positive parameters* $\bar{f}_L, r \in \mathbb{R}_{++}$, *be given. Suppose that* $f_L(\tilde{L}) > \bar{f}_L$ *and that the Hessian* $\nabla^2 f_L$ *is uniformly positive definite in the* r-*ball around* $\tilde{L}$, *i.e.,*

$$
\lambda_{\min}(\nabla^2 f_L(\tilde{L})) > 0, \quad \text{for all } \tilde{L} \in B_r(\tilde{L}). \tag{5.9}
$$

Then, f_L *is positively uniformly bounded below in* $B_r(\tilde{L})$, *i.e.,*

$$
f_L(L) > \bar{f}_L > 0, \quad \text{for all } \|L - \tilde{L}\| \leq \min\left\{ r, \frac{f_L(\tilde{L}) - \bar{f}_L}{\|\nabla f_L(\tilde{L})\|} \right\}.
$$

Proof. By the positive definiteness of the Hessian in the r-ball $B_r(\tilde{L})$, we can apply the convexity of f_L in the ball, implying that

$$
\begin{aligned}
f_L(L) \quad \geq \quad & f_L(\tilde{L}) + \langle \nabla f_L(\tilde{L}), L - \tilde{L} \rangle, && \text{if } \tilde{L} \in B_r(\tilde{L}), \\
\geq \quad & f_L(\tilde{L}) - \|\nabla f_L(\tilde{L})\|\|L - \tilde{L}\|, && \text{if } \tilde{L} \in B_r(\tilde{L}), \\
> \quad & \bar{f}_L > 0, && \text{if } \|L - \tilde{L}\| \leq \min\left\{ r, \frac{f_L(\tilde{L}) - \bar{f}_L}{\|\nabla f_L(\tilde{L})\|} \right\}.
\end{aligned}
$$

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$$

By (5.1), (5.2) and considering $\bar{f}_L = 10^3$, we get

$$\frac{f_L(\tilde{L}) - \bar{f}_L}{\|\nabla f_L(\tilde{L})\|} > \frac{2.6 \times 10^3 - 1 \times 10^3}{1.8 \times 10^{-3}} > r. \tag{5.10}$$

According to Lemma 5.4,

$$f_L(L) > \bar{f}_L > 0, \quad \text{for all } L \in B_r(\tilde{L}).$$

We now apply the classical Kantorovich theorem to obtain a unique **lngm** point (within a certain neighborhood) without needing the assumption of the existence of a stationary point. We reword the version in [5, Thm 5.3.1].

Theorem 5.5. *Let the configuration matrix $\tilde{P} = V\tilde{L} \in \mathbb{R}^{n \times d}, \tilde{L} \in \mathbb{R}^{n-1 \times d}$ be given. Let $r \in \mathbb{R}_{++}$ be found such that*

$$\nabla^2 f_L(L) \succ 0, \quad \text{for all } L \in B_r(\tilde{L}),$$

and $\bar{f}_L$ satisfying

$$f_L(\tilde{L}) > \bar{f}_L > 0, \quad \text{for all } L \in B_r(\tilde{L}).$$

Let γ be a Lipschitz constant for the Hessian of f_L in the r-ball about $\tilde{L}$. Set

$$\beta := \|\nabla^2 f_L(\tilde{L})^{-1}\| \quad \text{and} \quad \eta := \|\nabla^2 f_L(\tilde{L})^{-1}\nabla f_L(\tilde{L})\|.$$

Define $\gamma_R = \beta\gamma$ and $\alpha = \gamma_R\eta$. If $\alpha \leq \frac{1}{2}$ and $r \geq r_0 := \frac{1-\sqrt{1-2\alpha}}{\beta\gamma}$, then the sequence $L_0 = \tilde{L}, L_1, L_2, \ldots$, produced by

$$L_{k+1} = L_k - \nabla^2 f_L(L_k)^{-1}\nabla f_L(L_k), \ k = 0, 1, \ldots,$$

is well defined and converges to L^*, a unique root of the gradient ∇f_L in the closure of $B_{r_0}(\tilde{L})$. If $\alpha < \frac{1}{2}$, then L^* is the unique zero of ∇f_L in the closure of $B_{r_1}(\tilde{L})$,

$$r_1 := \min\left\{r, \frac{1 + \sqrt{1 - 2\alpha}}{\beta\gamma}\right\},$$

*and $\|L_k - L^*\| \leq (2\alpha)^{2k}\frac{\eta}{\alpha}, \quad k = 0, 1, \ldots$. Moreover, L^* is a **lngm**.*

Proof. The proof is a direct application of the Kantorovich theorem, e.g., [5, Thm 5.3.1], along with the above lemmas and corollaries in this section. The fact that $\nabla^2 f_L$ is invertible and positive definite follows from finding r in Proposition 5.3. The strict positive lower bound follows from Lemma 5.4. $\qquad\square$

As mentioned previously in this section, the conditions required in Lemma 5.4 and Theorem 5.5 are fulfilled with

$$r = 10^{-3}, \ \gamma = 145, \ \bar{f}_L = 10^3.$$

Plugging into Theorem 5.5, we have

$$\begin{aligned}
&\beta \approx 5.7 \times 10^{-4}, \\
&\eta \approx 2.4 \times 10^{-6}, \\
&\gamma_R \approx 8.2 \times 10^{-2}, \\
&\alpha = \gamma_R\eta \approx 2.0 \times 10^{-7}, \\
&r_0 \approx 2.4 \times 10^{-4} < r, \\
&r_1 = r = 10^{-3}.
\end{aligned}$$

According to Theorem 5.5, there exists a **lngm** in B_{r_1}.

As mentioned above, for the case of $d = 2$, any local minimizer of $f_L(L)$ is nonisolate, implying that the Hessian matrix at any local minimizer of $f_L(L)$ is singular. We consider the model $f_\ell(\ell)$ for an example with $d = 2$. We use Kantorovich Theorem to verify that this example has a strict local nonglobal minimizer ℓ^*, where the first 2 rows of $\mathcal{L}\text{Triag}(\ell^*)$ are independent. According to Theorems 3.7 and 3.8, this local nonglobal minimizer corresponds to local nonglobal minimizers of $f_L(L)$ and $f(P)$.

Example 5.6. *The details for the data and codes are given in Footnote 3. An example with $n = 100, d = 2$ is given, with data $\bar{\ell}, \tilde{\ell} \in \mathbb{R}^{197}$ presented, see Footnote 3. Matrix $\bar{D}$ is the distance matrix obtained from $\bar{\ell}$ by*

$$\bar{D} = \mathcal{K}(V \mathcal{L}\text{Triag}(\bar{\ell}) \mathcal{L}\text{Triag}(\bar{\ell})^T V^T).$$

Thus, $\bar{\ell}$ is a global minimizer, and we verify that $\tilde{\ell}$ is a numerically convergence point. The objective value is

$$f_\ell(\tilde{\ell}) > 9 \times 10^3, \tag{5.11}$$

the absolute and relative gradient norms are

$$\|\nabla f_\ell(\tilde{\ell})\| < 1.7 \times 10^{-2}, \quad \frac{\|\nabla f_\ell(\tilde{\ell})\|}{1 + f_\ell(\tilde{\ell})} < 10^{-6}, \tag{5.12}$$

and the least eigenvalue of the Hessian matrix is

$$\lambda_{\min}(\nabla^2 f_\ell(\tilde{\ell})) > 8.$$

The analysis process for $f_\ell(\ell)$ is similar to $f_L(L)$. For completeness, we present it as follows.

Proposition 5.7. *Let $r > 0, \tilde{\ell} \in \mathbb{R}^{n-1 \times d}$ be given. If*

$$\gamma \geq 48\sqrt{2}r \left(\sum_{i,j} \|(V \mathcal{L}\text{Triag}(\tilde{\ell}))[i,:] - (V \mathcal{L}\text{Triag}(\tilde{\ell}))[j,:]\| + 2n\sqrt{n}r \right),$$

then γ is a Lipschitz constant for the Hessian of f_ℓ in the radius-r neighborhood of $\tilde{\ell}$, i.e.,

$$\|\nabla^2 f_\ell(\hat{\ell}) - \nabla^2 f_\ell(\check{\ell})\| \leq \gamma \|\hat{\ell} - \check{\ell}\|, \quad \text{if } \hat{\ell}, \check{\ell} \in B_r(\tilde{\ell}).$$

Moreover,

$$\lambda_{\min}(\nabla^2 f_\ell(\ell)) \geq \lambda_{\min}(\nabla^2 f_\ell(\tilde{\ell})) - \gamma r, \quad \text{if } \ell \in B_r(\tilde{\ell}).$$

Lemma 5.8. *Suppose that $f_\ell(\tilde{\ell}) > \bar{f}_\ell$ and that the Hessian $\nabla^2 f_\ell$ is uniformly positive definite in the r-ball around $\tilde{\ell}$, i.e.,*

$$\lambda_{\min}(\nabla^2 f_\ell(\tilde{\ell})) > 0, \quad \text{for all } \tilde{\ell} \in B_r(\tilde{\ell}).$$

Then, f_ℓ is positively uniformly bounded below in $B_r(\tilde{\ell})$, i.e.,

$$f_\ell(\ell) > \bar{f}_\ell > 0, \quad \text{for all } \|\ell - \tilde{\ell}\| \leq \min \left\{ r, \frac{f_\ell(\tilde{\ell}) - \bar{f}_\ell}{\|\nabla f_\ell(\tilde{\ell})\|} \right\}.$$

Theorem 5.9. *Let $\tilde{\ell} \in \mathbb{R}^{t_\ell}$ be given and $r \in \mathbb{R}_{++}$ be found such that*

$$\nabla^2 f_\ell(\ell) \succ 0, \quad \text{for all } \ell \in B_r(\tilde{\ell}), \tag{5.13}$$

and $\bar{f}_\ell$ satisfy

$$f_\ell(\tilde{\ell}) > \bar{f}_\ell, \quad \text{for all } \tilde{\ell} \in B_r(\tilde{\ell}).$$

Let γ be a Lipschitz constant for the Hessian of f_ℓ in the r-ball about $\tilde{\ell}$. Set

$$\beta := \|\nabla^2 f_\ell(\tilde{\ell})^{-1}\|, \quad \text{and} \quad \eta := \|\nabla^2 f_\ell(\tilde{\ell})^{-1} \nabla f_\ell(\tilde{\ell})\|.$$

Define $\gamma_R = \beta\gamma$ and $\alpha = \gamma_R \eta$. If $\alpha \le \frac{1}{2}$ and $r \ge r_0 := \frac{1-\sqrt{1-2\alpha}}{\beta\gamma}$, then the sequence $\ell_0 = \tilde{\ell}, \ell_1, \ell_2, \ldots,$ produced by

$$\ell_{k+1} = \ell_k - \nabla^2 f_\ell(\ell_k)^{-1} \nabla f_\ell(\ell_k),\ k = 0, 1, \ldots,$$

is well defined and converges to ℓ^*, a unique root of the gradient ∇f_ℓ in the closure of $B_{r_0}(\tilde{\ell})$. If $\alpha < \frac{1}{2}$, then ℓ^* is the unique zero of ∇f_ℓ in the closure of $B_{r_1}(\tilde{\ell})$,

$$r_1 := \min\left\{r, \frac{1 + \sqrt{1-2\alpha}}{\beta\gamma}\right\}$$

and

$$\|\ell_k - \ell^*\| \le (2\alpha)^{2k}\frac{\eta}{\alpha}, \quad k = 0, 1, \ldots.$$

*Moreover, ℓ^* is a **lngm**.*

The conditions required in Proposition 5.7 and Lemma 5.8 are fulfilled with

$$r = 10^{-3}, \ \gamma = 651, \ \bar{f}_L = 10^3.$$

Plugging into Theorem 5.5, we have

$$\begin{aligned}
&\beta \approx 1.0 \times 10^{-4}, \\
&\eta \approx 1.3 \times 10^{-5}, \\
&\gamma_R \approx 6.5 \times 10^{-2}, \\
&\alpha = \gamma_R \eta \approx 8.7 \times 10^{-7} < \tfrac{1}{2}, \\
&r_0 = \tfrac{1-\sqrt{1-2\alpha}}{\beta\gamma} \approx 1.3 \times 10^{-5} < r, \\
&r_1 = r = 10^{-3}.
\end{aligned}$$

According to Theorem 5.5, there exists a **lngm** in B_r.

6 Final remarks

In this paper, we addressed the nonconvex optimization problem arising from the exact recovery of points from a given Euclidean Distance Matrix (EDM). Our investigation led to significant advancements in understanding the conditions under which local nonglobal minima (**lngm**) exist for the smooth stress function (as known in MDS literature) in EDM problems. We established that for the smooth stress function, which is a quartic in $P \in \mathbb{R}^{n \times d}$, all second-order stationary points are global minimizers when $n \le d+1$. For $n > d+1$, we not only identified **lngm** through numerical

methods but also provided rigorous analytical proofs (via Kantorovich's theorem) confirming their existence.

Our methodology was characterized by a reduction in the problem's dimensionality, leveraging translation and rotation invariance, which simplified the analytical process and computational efforts.

The findings of this research settle longstanding open questions regarding the existence of **lngm** in the context of multidimensional scaling and highlight the importance of second-order methods for minimizing the smooth stress function.

Acknowledgments

Index

References

[1] S. AL-HOMIDAN AND H. WOLKOWICZ, *Approximate and exact completion problems for Euclidean distance matrices using semidefinite programming*, Linear Algebra Appl., 406 (2005), pp. 109–141. 7

[2] J. ALENCAR, C. LAVOR, AND L. LIBERTI, *Realizing Euclidean distance matrices by sphere intersection*, Discrete Appl. Math., 256 (2019), pp. 5–10. 2

[3] A. ALFAKIH, *Euclidean distance matrices and their applications in rigidity theory*, Springer, Cham, 2018. 2, 3

[4] A. D. BÁEZ SÁNCHEZ AND C. LAVOR, *On the estimation of unknown distances for a class of Euclidean distance matrix completion problems with interval data*, Linear Algebra Appl., 592 (2020), pp. 287–305. 2

[5] J. DENNIS JR. AND R. SCHNABEL, *Numerical methods for unconstrained optimization and nonlinear equations*, vol. 16 of Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996. Corrected reprint of the 1983 original. 17, 20

[6] N. KRISLOCK AND H. WOLKOWICZ, *Euclidean distance matrices and applications*, in Handbook on semidefinite, conic and polynomial optimization, vol. 166 of Internat. Ser. Oper. Res. Management Sci., Springer, New York, 2012, pp. 879–914. 2

[7] J. D. LEE, I. PANAGEAS, G. PILIOURAS, M. SIMCHOWITZ, M. I. JORDAN, AND B. RECHT, *First-order methods almost always avoid strict saddle points*, Mathematical Programming, 176 (2019), pp. 311–337. 5

[8] L. LIBERTI, C. LAVOR, N. MACULAN, AND A. MUCHERINO, *Euclidean distance geometry and applications*, SIAM Rev., 56 (2014), pp. 3–69. 2

[9] S. MALONE, P. TARAZAGA, AND M. TROSSET, *Better initial configurations for metric multidimensional scaling*, Comput. Statist. Data Anal., 41 (2002), pp. 143–156. 2, 3

[10] S. MALONE AND M. TROSSET, *A study of the critical points of the SSTRESS criterion in metric multidimensional scaling*, Tech. Rep. 00-06, Department of Computational & Applied Mathematics, Rice University, Houston, TX, USA, 2000. 3

[11] R. MATHAR AND M. TROSSET, *On the existence of nonglobal minimizers of the stress criterion for metric multidimensional scaling*, in Proceedings of the Statistical Computing Section, 1997, pp. 158–162. 2, 3

[12] A. MUCHERINO, C. LAVOR, L. LIBERTI, AND N. MACULAN, eds., *Distance geometry*, Springer, New York, 2013. Theory, methods, and applications. 2

[13] R. PARHIZKAR, *Euclidean Distance Matrices Properties, Algorithms and Applications*, PhD thesis, ISC, Lausanne, 2013. 2, 3

[14] I. SCHOENBERG, *Remarks to Maurice Fréchet's article "Sur la définition axiomatique d'une classe d'espace distanciés vectoriellement applicable sur l'espace de Hilbert"*, Ann. Math., 36 (1935), pp. 724–732. 4

[15] A. ŽILINSKAS AND A. PODLIPSKYTUNDEFINED, *On multimodality of the sstress criterion for metric multidimensional scaling*, Informatica, 14 (2003), p. 121–130. 3