On the local and global minimizers of the smooth stress function in Euclidean distance matrix problems[5]

Mengmeng Song[a,b,c,1], Douglas S. Gonçalves[d,2,*], Woosuk L. Jung[c], Carlile Lavor[e,3], Antonio Mucherino[f,4], Henry Wolkowicz[c,5]

[a] *National Frontiers Science Center for Industrial Intelligence and Systems Optimization, Northeastern University, Shenyang, P.R. China*
[b] *School of Mathematical Sciences, Beihang University, Beijing, 100191, P.R. China*
[c] *Department of Combinatorics and Optimization, Faculty of Mathematics, University of Waterloo, N2L 3G1, Waterloo, Canada*
[d] *Department of Mathematics, Universidade Federal de Santa Catarina, 88040-900, Florianopólis, Brazil*
[e] *Department of Applied Mathematics (IMECC), University of Campinas, 13083-859, Campinas, Brazil*
[f] *IRISA, Université de Rennes, F-35042, Rennes, France*

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 whether there are conditions such that the minimization problem admits a local nonglobal min-

[*] Corresponding author.
Email address: `douglas.goncalves@ufsc.br` (Douglas S. Gonçalves)

[1] The work of this author was partially supported by the Academic Excellence Foundation of BUAA for PhD Students, and international joint doctoral education fund of Beihang University.

[2] The work of this author was supported in part by CNPq (Grant 305213/2021-0) and FAPESC (Grant 2024TR002238).

[3] The work of this author is supported in part by FAPESP (Grant 2023/08706-1) and CNPq (Grants 305227/2022-0 and 404616/2024-0).

[4] The work of this author was partially supported by the ANR project EVARISTE (ANR-24-CE23-1621).

[5] Research supported by The Natural Sciences and Engineering Research Council of Canada.

[5] Date: Aug. 13, 2024; Revision: July 19, 2025

imizer, **lngm**. We prove that all second-order stationary points are global minimizers whenever $n \leq d+1$. And, for $d = 1$ and $n \geq 7 > d+1$, we present an example where we can analytically exhibit a local nonglobal minimizer. For more general cases, we numerically find a second-order stationary point and then prove that there indeed exists a nearby **lngm** for the quartic nonconvex minimization problem. Thus, we answer the previously open question about their existence in the affirmative. Our approach to finding the **lngm** is novel in that we first exploit the translation and rotation invariance to remove the singularities of the Hessian, and reduce the size of the problem from nd variables in P to $(n-1)d - d(d-1)/2$ variables. This allows for stabilizing Newton's method, and for finding examples that satisfy the strict second order sufficient optimality conditions.

The motivation for being able to find global minima is to obtain *exact recovery* of the configuration matrix, even in the cases where the data is noisy and/or incomplete, without resorting to approximating solutions from convex (semidefinite programming) relaxations. In the process of our work we present new insights into when **lngm**s of the smooth stress function do and do not exist.

Keywords: distance geometry, Euclidean distance matrices, Gram matrix, local nonglobal minima, Kantorovich Theorem, exact recovery
2000 MSC: 51K05, 90C26, 65K10

Contents

List of Figures

1. Introduction

EDM (completion) problems have long been studied in the scientific literature, see e.g., the surveys, book collections, and some recent papers [7, 9, 13, 3, 2, 5]. 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

$$\bar{G} = \bar{P}\bar{P}^T, \quad \bar{P}^T = \begin{bmatrix}\bar{p}_1, \ldots, \bar{p}_n\end{bmatrix} \in \mathbb{R}^{d\times n}.$$

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

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

where $\mathcal{K} : \mathbb{S}^n \to \mathbb{S}^n$ is the Lindenstrauss operator on symmetric matrix space:

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

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. Moreover, $\|\cdot\|$, $\|\cdot\|_F$, denote the Euclidean and Frobenius norms, respectively.

The optimization problem (1.1) is a Euclidean distance geometry problem (DGP), see e.g., [9]. DGP includes the **EDM** Completion Problem, where $\bar{D}$ can have missing entries as well as noisy entries. This latter problem has been proven to be NP-Hard [15].

The objective function of (1.1) is denoted and expressed as a quartic in P:

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

which is referred to as the *smooth stress* in the multidimensional scaling (MDS) literature. (Throughout this text, for notational convenience, we use $f(P) = \frac{1}{2}\sigma_2(P)$ as our objective function.)

Since the function $\sigma_2(P)$ is nonconvex, and optimization methods generally find local minima, we investigate the possibility for such a quartic function to have all local minimizers as global minimizers. This question has been considered as open and has been widely studied in the MDS literature; see for example, [10, 11, 12, 14]. One of the motivations for the research about the local nonglobal minima (**lngm**) of $\sigma_2(P)$ is that the characterization of the **lngm** is critical to developing efficient algorithms without resorting to convex (semidefinite programming) relaxations.

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

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

Trosset and Mathar [12] analytically verified that the raw stress function $\sigma_1(P)$ admits a **lngm**, where $\bar{P}$ is a square configuration with vertices at the four points $\bar{p}_1 = [1/2, 1/2], \bar{p}_2 = [-1/2, 1/2], \bar{p}_3 = [-1/2, -1/2]$, and $\bar{p}_4 = [1/2, -1/2]$. The authors of [17] applied this example to the smooth function $\sigma_2(P)$ but did not find a **lngm**. Instead, they present an example of the *inexact* **EDM** recovery problem having a **lngm**; specifically, problem (1.1) with $\bar{D}$ replaced by $\Delta \in \mathbb{S}^n$ that is *not* an **EDM**. However, the question about the existence of a **lngm** for the exact problem $\sigma_2(P)$ remained open. Addressing this challenge involves two main difficulties: (1) the apparent lack of simple examples exhibiting **lngm**; and (2) the inherent complexity in proving the existence of a **lngm** for such a complex problem.

In this paper, we provide a definitive answer to this open question. We prove that all second-order stationary points are global minimizers whenever $n \leq d+1$. For $n > d+1$, we present an example in dimension $d = 1$, with a very special structure, for which we can analytically exhibit a **lngm**. For more general cases, we find examples where the function $\sigma_2(P)$ has a **lngm**, and we provide analytic verification using techniques from the local convergence proof of Newton's method. The examples are obtained using the trust region approach with random initializations.

The rest of this paper is arranged as follows. We continue in Section 2 with a description of our main unconstrained minimization problem (1.1). We introduce two additional equivalent problems with reduced numbers of variables. The reduction allows for strict second-order sufficient optimality conditions and thus is necessary for the analytic existence proof. In Section 3, we include various linear transformations, derivatives, and adjoints. Many of these are used throughout this paper. We suggest that they provide a useful addition to the literature on **EDM**, as it emphasizes the use of matrix transformations rather than individual elements or points. In Section 4, we study the optimality conditions and establish that $\sigma_2(P)$ has no **lngm** if $n \leq d+1$. In Section 5.1, we give a special example with $d = 1$ where a **lngm** can be explictly illustrated. In Section 5.2, we provide two examples and use the Kantorovich theorem to (numerically) prove the existence of **lngm**s in this more general setting.

2. Notation and Equivalent Main Problem Formulations

Before presenting the problem formulations, we introduce the necessary notation and background from distance geometry. Further details are provided in [3].

2.1. Notation

We use the *trace inner product* in $m \times n$ matrix space $\mathbb{R}^{m\times n}$, $\langle X, Y\rangle = \operatorname{tr} X^T Y$ with the induced *Frobenius norm,* $\|X\|_F = \sqrt{\langle X, X\rangle}$. Denote $\|X\|_2 := \sqrt{\lambda_{\max}(X^T X)}$, where $\lambda_{\max}(\cdot)$ gives the largest eigenvalue. If no subscript in $\|\cdot\|$ is written, the Frobenius norm $\|\cdot\|_F$ is understood. We note that both $\|\cdot\|_F$ and $\|\cdot\|_2$ reduce to the standard 2-norm in $\mathbb{R}^n$ ($n \times 1$ matrices) $\|x\|$. For a finite dimensional Hilbert space $\mathcal{X}$, we use $B_r(\tilde{x}) := \{x \in \mathcal{X} \mid \|x - \tilde{x}\| \leq r\}$ to denote the ball centered at $\tilde{x}$ with radius $r > 0$.

Let $\mathbb{S}^n$ be the set of symmetric matrices in $\mathbb{R}^{n\times n}$. The cone of positive semidefinite matrices is denoted by $\mathbb{S}^n_+ \subset \mathbb{S}^n$, and we use $S \succeq 0$ for $S \in \mathbb{S}^n_+$. Similarly, for positive definite matrices S, we use $S \in \mathbb{S}^n_{++}$ and $S \succ 0$. Additionally, $S \geq 0$ and $S > 0$ denote that all entries of S are non-negative and positive, respectively. Let $\operatorname{diag}(S) \in \mathbb{R}^n$ denote the vector formed by the diagonal of a matrix $S \in \mathbb{R}^{n\times n}$. The adjoint operator $\operatorname{diag}^*(v) = \operatorname{Diag}(v) \in \mathbb{S}^n$ maps a vector $v \in \mathbb{R}^n$ to $\operatorname{Diag}(v) \in \mathbb{S}^n$, the diagonal matrix with entries from v. For a matrix $C \in \mathbb{R}^{n\times d}$, $\operatorname{vec}(C) \in \mathbb{R}^{nd}$ denotes the column vector formed by stacking the columns of C, and $\operatorname{Mat} \cong \operatorname{vec}^*$ is the adjoint of vec, satisfying $\operatorname{Mat}(\operatorname{vec}(C)) = C$ for all $C \in \mathbb{R}^{n\times d}$. If $F : \mathcal{X} \to \mathcal{Y}$ is a map between finite dimensional Hilbert spaces, $F'(P)$ and $F''(P)$ denote its Fréchet derivatives at $P \in \mathcal{X}$.

For a set of points $p_i \in \mathbb{R}^d,\ i \in [n] := \{1, 2, \cdots, n\}$, denote the *configuration matrix* by

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

Here d is the embedding dimension. Denote the quadratic mapping $\mathcal{M} : \mathbb{R}^{n\times d} \to \mathbb{S}^n$, $\mathcal{M}(P) = PP^T$. Recall that e denotes the column vector of ones of appropriate dimension. Then, the classical result of Schöenberg [16] relates an **EDM**, $\mathcal{D}(P)$, with the corresponding *Gram matrix,* $G = \mathcal{M}(P)$, by applying the linear operator $\mathcal{K} : \mathcal{S}^n_C \to \mathcal{S}^n_H$:

$$\mathcal{K}(G) = \operatorname{diag}(G)e^T + e\operatorname{diag}(G)^T - 2G = \left(\|p_i - p_j\|^2\right)_{ij} =: \mathcal{D}(P), \qquad (2.1)$$

where the *centered subspace,* $\mathcal{S}^n_C$ and the *hollow subspace,* $\mathcal{S}^n_H$ are defined by

$$\mathcal{S}^n_C = \{S \in \mathbb{S}^n \,:\, Se = 0\}, \quad \mathcal{S}^n_H = \{S \in \mathbb{S}^n \,:\, \operatorname{diag}(S) = 0\}.$$

Denote $S_e : \mathbb{R}^n \to \mathbb{S}^n$: $S_e(v) = ve^T + ev^T$. Then, $\mathcal{K}(G) = S_e(\operatorname{diag}(G)) - 2G$. Note that the centered assumption $P^Te = 0 \Leftrightarrow G = PP^T \in \mathcal{S}^n_C$. Also, when the domain of $\mathcal{K}$ is restricted to be $\mathcal{S}^n_C$, the mapping $\mathcal{K}$ is a bijection between $\mathcal{S}^n_C$ and $\mathcal{K}(\mathcal{S}^n_C)$.

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 a centered configuration matrix $\bar{P} \in \mathbb{R}^{n\times d}$, $\bar{P}^Te = 0$. This gives rise to the corresponding Gram matrix $\bar{G} = \bar{P}\bar{P}^T \in \mathcal{S}^n_C$ and **EDM**, $\bar{D} = \mathcal{K}(\bar{G})$. We now present the main problem and two reformulations that reduce the size of variables and help with stability.

Problem 2.1. *Let* $\bar{D} = \mathcal{D}(\bar{P})$ *be a* ***EDM*** *obtained from some given configuration matrix* $\bar{P}$*. Consider the* nonconvex *minimization problem of recovering a corresponding configuration matrix* $\hat{P}$ *given by*

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

thus defining the function $F : \mathbb{R}^{n \times d} \to \mathcal{S}_H^n$*.*

Theorem 2.1 is a nonlinear least squares problem. It has nd variables. By taking advantage of symmetry and the zero-diagonal constraints, the objective function can be seen as a sum of squares of $t(n-1)$ quadratic functions, where $t(n-1) := n(n-1)/2$ is the *triangular number*. Note that $\bar{P}$ is a global minimizer for (2.2) with the optimal value $f(\bar{P}) = 0$. We study whether all stationary points where the second-order necessary optimality conditions hold are global minimizers.

Note that the distance matrix is invariant under translations and rotations of P. Without loss of generality, we assume P is centered ($P^T e = 0$). Let $V \in \mathbb{R}^{n \times (n-1)}$ be such that

$$V^T V = I_{n-1}, \quad V^T e = 0. \tag{2.3}$$

By the fact that VV^T is the orthogonal projection onto $e^\perp$ (the orthogonal complement of e), we have $P^T e = 0$ if, and only if, $P = VL$ for some $L \in \mathbb{R}^{(n-1) \times d}$.[7] We exploit this property for deducing an equivalent problem formulation having a smaller dimension.

Problem 2.2. *Let* $\bar{P}, \bar{D}$ *be as given in Theorem 2.1, and let* V *be as in* (2.3)*. Consider the* nonconvex *minimization problem of recovering a corresponding centered configuration matrix* $\hat{P} = V\hat{L}$ *by finding*

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

thus defining the function $F_L : \mathbb{R}^{(n-1) \times d} \to \mathcal{S}_H^n$*.*

Let $\mathcal{O} = \{Q \in \mathbb{R}^{d \times d} : Q^T Q = I_d\}$ be the orthogonal group of order d. Note that $LL^T = LQQ^T L^T$ holds for all $Q \in \mathcal{O}$. If $L^T = QR$ is the QR factorization, then

$$R^T = LQ \Rightarrow f_L(L) = f_L(R^T Q^T) = f_L(R^T),$$

[7] This is similar to the application of *facial reduction* for the semidefinite relaxation, see [4].

where $R \in \mathbb{R}^{d\times(n-1)}$ is upper triangular (trapezoidal when $d < n-1$). The problem can be further reduced using rotation invariance: $f_L(LQ) = f_L(L), \forall Q \in \mathcal{O}$.

Recall that the linear transformation $\operatorname{svec} : \mathbb{S}^n \to \mathbb{R}^{t(n)}$ is a generalization of the vectorization vec applied to symmetric matrices that avoids the duplication of the lower triangular part. We now extend this idea to triangular (trapezoidal) matrices to avoid the zeros. We define the linear operator that maps a vector $\ell \in \mathbb{R}^{t_\ell}$ to a lower triangular (trapezoidal) matrix in $\mathbb{R}^{(n-1)\times d}$ given by

$$\mathcal{L}\mathrm{Triag}(\ell)_{(i,j)} = \begin{cases} \ell_{nj-n-t(j)+i+1}, & \text{if } j \le i \\ 0, & \text{otherwise,} \end{cases} \tag{2.5}$$

where

$$t_\ell = \begin{cases} t(n-1), & \text{if } d \ge n-1 \\ (n-1)d - t(d-1), & \text{otherwise.} \end{cases} \tag{2.6}$$

For $L \in \mathbb{R}^{(n-1)\times d}$ lower trapezoidal, t_ℓ counts its entries where nonzero values are allowed. For $d < n-1$, $\mathcal{L}\mathrm{Triag}(\ell)$ has $t(d-1)$ zero elements at the top right; whereas for $d \ge n-1$, $\mathcal{L}\mathrm{Triag}(\ell)$ has $t(n-1)$ nonzero elements at the bottom left.

Using the above definition, we define

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

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

Problem 2.3. *Let $\bar{P}, \bar{D}$ be as given in Theorem 2.2 (and in Theorem 2.1), and let $V, t_\ell, f_\ell(\ell)$, be as in* (2.3), (2.6) *and* (2.7)*, respectively. Consider the* nonconvex *minimization problem of* <u>*recovering*</u> *a corresponding centered configuration matrix $\hat{P} = V\hat{L} = V\,\mathcal{L}\mathrm{Triag}(\hat{\ell})\hat{Q}^T$, with $\hat{Q} \in \mathcal{O}$, by finding*

$$\begin{aligned} \hat{\ell} \in \operatorname{argmin}_{\ell\in\mathbb{R}^{t_\ell}} f_\ell(\ell) \;&:=\; \tfrac{1}{2}\|\mathcal{K}(V\,\mathcal{L}\mathrm{Triag}(\ell)(V\,\mathcal{L}\mathrm{Triag}(\ell))^T) - \bar{D}\|_F^2 \\ &= \tfrac{1}{2}\|F_L(\mathcal{L}\mathrm{Triag}(\ell))\|_F^2 =: \tfrac{1}{2}\|F_\ell(\ell)\|_F^2; \end{aligned} \tag{2.8}$$

thus defining the function $F_\ell : \mathbb{R}^{t_\ell} \to \mathcal{S}^n_H$.

Remark 2.4. *Compared to Theorem 2.1, Theorem 2.3 is a nonlinear least squares problem, with fewer variables and the same number of quadratic terms* $\left(\mathcal{K}(V\,\mathcal{L}\mathrm{Triag}(\ell)(V\,\mathcal{L}\mathrm{Triag}(\ell))^T) - \bar{D}\right)_{ij}, i < j$. *For* $d < n-1$, *the underlying system of equations is overdetermined, as* $t_\ell < t(n-1)$. *For* $d \geq n-1$, *from* (2.6), *the number of variables is* $t(n-1)$, *the same as the number of quadratic equations. Thus we no longer have the singularity that arises for the Jacobian of an underdetermined nonlinear least squares problem.*

In order to determine whether a **lngm** exists, 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 above three reformulations.

3.1. Transformations, Derivatives, Adjoints, Range and Null Spaces

Theorem 3.1 below presents a list of auxiliary results. It concerns the following vectors, matrices and functions:

$$P \in \mathbb{R}^{n\times d},\, p = \mathrm{vec}(P) \in \mathbb{R}^{nd}, \Delta P \in \mathbb{R}^{n\times d},\, \Delta p = \mathrm{vec}(\Delta P) \in \mathbb{R}^{nd},$$
$$L \in \mathbb{R}^{(n-1)\times d},\ \ell \in \mathbb{R}^{t_\ell}, t_\ell \text{ in } (2.6),\ S, T \in \mathbb{S}^n;$$

$$\mathcal{M} : \mathbb{R}^{n\times d} \to \mathbb{S}^n,\, \mathcal{K} : \mathbb{S}^n \to \mathbb{S}^n,\, F : \mathbb{R}^{n\times d} \to \mathcal{S}^n_H,\, f : \mathbb{R}^{n\times d} \to \mathbb{R},$$
$$\mathcal{L}\mathrm{Triag} : \mathbb{R}^{t_\ell} \to \mathbb{R}^{(n-1)\times d},\ S_e : \mathbb{R}^n \to \mathbb{S}^n,\ F_L : \mathbb{R}^{(n-1)\times d} \to \mathcal{S}^n_H,\ f_L : \mathbb{R}^{(n-1)\times d} \to \mathbb{R},$$
$$F_\ell : \mathbb{R}^{t_\ell} \to \mathcal{S}^n_H,\ f_\ell : \mathbb{R}^{t_\ell} \to \mathbb{R}.$$

Lemma 3.1. *We have the following first and second Fréchet derivatives and adjoints:*

1. $\mathcal{M}'(P)(\Delta P) = P\Delta P^T + \Delta P P^T$, $\mathcal{M}''(P)(\Delta P, \Delta P) = 2\Delta P \Delta P^T$.
2. $\mathcal{M}'(P)^*(S) = 2SP$.
3. $S_e^*(S) = 2Se$.
4. $\mathcal{K}(G) = S_e(\mathrm{diag}(G)) - 2G$, $\mathrm{range}(\mathcal{K}) = \mathcal{S}^n_H$, $\mathrm{null}(\mathcal{K}) = \mathrm{range}(S_e)$.
5. $\mathcal{K}^*(S) = 2(\mathrm{Diag}(Se) - S)$, $\mathrm{range}(\mathcal{K}^*) = \mathcal{S}^n_C$, $\mathrm{null}(\mathcal{K}^*) = \mathrm{Diag}(\mathbb{R}^n)$. *Moreover,* $S \geq (\leq) 0 \implies \mathcal{K}^*(S) \succeq (\preceq) 0$.
6. $\mathcal{D}(P) = S_e(\mathrm{diag}(\mathcal{M}(P))) - 2\mathcal{M}(P)$.
7. $\mathcal{D}'(P)(\Delta P) = S_e(\mathrm{diag}(\mathcal{M}'(P)(\Delta P)) - 2\mathcal{M}'(P)(\Delta P)$.

8. $F'(P)(\Delta P) = \mathcal{K}(\mathcal{M}'(P)(\Delta P))$, $F''(P)(\Delta P, \Delta P) = \mathcal{K}(\mathcal{M}''(P)(\Delta P, \Delta P))$.
9. $F'(P)^*(S) = \mathcal{M}'(P)^*(\mathcal{K}^*(S)) = 4(\mathrm{Diag}(Se) - S)P$.
10. *We have*

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

and

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

Proof. 1. It follows 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. Note that

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

3. From the trace inner product,

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

4. See [1, Prop. 2.2].
5. See [1, Prop. 2.2] for the characterization of the nullspace of $\mathcal{K}^*$. The identity $\mathcal{K}^*(S) = 2(\mathrm{Diag}(Se) - S)$ follows from

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

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

6. It follows directly from Item 5.
7. This follows from the linearity of diag and S_e.
8. Both follow from the definitions and linearity of $\mathcal{K}$.
9. It follows from

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

and $\mathcal{M}'(P)^*$ and $\mathcal{K}^*(S)$ presented in Items 2 and 5.
10. From the expansion of $f(P+\Delta P)$,

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

we get (3.1). Then we obtain

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

where the second equality follows from Item 8. Define $\Delta p := \mathrm{vec}(\Delta P)$. Now, we can isolate the matrix representation with

$$\begin{array}{rcl} f''(P)(\Delta P,\Delta P) &=& \langle f''(P)(\mathrm{Mat}\,\mathrm{vec}(\Delta P)), \mathrm{Mat}\,\mathrm{vec}(\Delta P)\rangle \\ &=& \langle [\mathrm{vec}\, f''(P)\,\mathrm{Mat}]\,(\Delta p)\,, (\Delta p)\rangle.\end{array}$$

Denote the symmetrization $\mathcal{S}:\mathbb{R}^{n\times n}\to\mathbb{S}^n, \mathcal{S}(K)=(K+K^T)/2$, and let $\mathcal{T}$ be the self-adjoint transpose operator. The first term in (3.3) is

$$\begin{array}{rl} & 4\langle \mathcal{K}(\mathcal{S}(P(\mathrm{Mat}\,\mathrm{vec}\,\Delta P)^T)), \mathcal{K}(\mathcal{S}(P(\mathrm{Mat}\,\mathrm{vec}\,\Delta P)^T))\rangle \\ = & 4\langle (P^T\mathcal{S}^*\mathcal{K}^*\mathcal{K}\mathcal{S}P)((\mathrm{Mat}\,\mathrm{vec}\,\Delta P)^T), (\mathrm{Mat}\,\mathrm{vec}\,\Delta P)^T\rangle \\ = & 4\langle (P^T\mathcal{S}^*\mathcal{K}^*\mathcal{K}\mathcal{S}P)(\mathcal{T}\,\mathrm{Mat}\,\mathrm{vec}\,\Delta P), (\mathcal{T}\,\mathrm{Mat}\,\mathrm{vec}\,\Delta P)\rangle \\ = & 4\left\langle \left[\mathrm{vec}\,\mathcal{T}^*P^T\mathcal{S}^*\mathcal{K}^*\mathcal{K}\mathcal{S}P\mathcal{T}\,\mathrm{Mat}\right]\Delta p, \Delta p\right\rangle.\end{array} \tag{3.4}$$

The second term in (3.3) is

$$\begin{array}{rl} & 2\langle F(P), \mathcal{K}\left(\Delta P \Delta P^T\right)\rangle \\ = & 2\left\langle \mathcal{K}^*\left(F(P)\right), \Delta P \Delta P^T \right\rangle \\ = & 2\left\langle \Delta P, \mathcal{K}^*(F(P)) \Delta P\right\rangle \\ = & 2\left\langle [\operatorname{vec}\mathcal{K}^* F(P)\operatorname{Mat}]\, \Delta p, \Delta p\right\rangle . \end{array} \tag{3.5}$$

Recall that $F'(P)(\Delta P) = \mathcal{K}(\mathcal{M}'(P)(\Delta P))$. We combine (3.4) and (3.5) and obtain the matrix representation of the Hessian (not necessarily positive semidefinite) :

$$\begin{array}{rcl} [\operatorname{vec} f''(P)\operatorname{Mat}] & = & 4\left[\operatorname{vec}\mathcal{T}^* P^T \mathcal{S}^* \mathcal{K}^* \mathcal{K}\mathcal{S}P\mathcal{T}\operatorname{Mat}\right] \\ & & +2\left[\operatorname{vec}\mathcal{K}^* F(P)\operatorname{Mat}\right] \\ & = & 4\left[J^* J\right] + 2\left[\operatorname{vec}\mathcal{K}^* F(P)\operatorname{Mat}\right], \end{array} \tag{3.6}$$

where

$$J(\Delta p) := \mathcal{K}\mathcal{S}P\mathcal{T}\operatorname{Mat}\Delta p. \tag{3.7}$$

□

Theorem 3.2. *The second-order necessary optimality conditions for* (2.2) *are:*

$$0 \;=\; f'(P) = F'(P)^*(F(P)) = 2\mathcal{K}^*(F(P))P, \tag{3.8}$$

$$0 \;\preceq\; [\operatorname{vec} f''(P)\operatorname{Mat}] = 4\left[J^* J\right] + 2\left[\operatorname{vec}\mathcal{K}^*(F(P))\operatorname{Mat}\right]. \tag{3.9}$$

Proof. The second equality in (3.8) follows from (3.1), and the third equality in (3.8) follows from Items 2 and 9 of Theorem 3.1. In particular, we have

$$f'(P) = F'(P)^*(F(P)) = \mathcal{M}'(P)^*(\mathcal{K}^*(F(P))) = 2\mathcal{K}^*(F(P))P. \tag{3.10}$$

The expression for the second-order term in (3.9) follows from (3.6) in Item 10 of Theorem 3.1. □

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

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

By abuse of notation, H_1 and H_2 represent both the linear maps and their matrix representations. The meaning will be clear from the context. We call P a *stationary point* if (3.8) holds, and we call P a *second-order stationary point* if both (3.8) and (3.9) hold.

3.2. Optimality Conditions of Three Problem Formulations

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

Proposition 3.3. *The derivatives of $f_L(L) : \mathbb{R}^{(n-1)\times d} \to \mathbb{R}$ 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 $f_\ell(\ell) : \mathbb{R}^{t_\ell} \to \mathbb{R}$ are*

$$f_\ell'(\ell) = \mathcal{L}\mathrm{Triag}^* \, f_L'(\mathcal{L}\mathrm{Triag}(\ell)), \tag{3.14}$$

and

$$f_\ell''(\ell) = \mathcal{L}\mathrm{Triag}^* \, f_L''(\mathcal{L}\mathrm{Triag}(\ell)) \, \mathcal{L}\mathrm{Triag}\,. \tag{3.15}$$

In the following, we show that any local minimizer of (2.4) corresponds to a family of local minimizers of (2.2), obtained by translations. Similarly, any local minimizer of (2.8) corresponds to a family of local minimizers of (2.4), derived from rotations.

Proposition 3.5. *The configuration matrix $P_* \in \mathbb{R}^{n\times d}$ is a local minimizer of the function f (see (2.2)) if, and only if, all configurations in $\{P_* + ev^T : v \in \mathbb{R}^d\}$ are local minimizers of the function f.*

Proof. We exploit the fact that the function f is invariant w.r.t. translations: for any point P, we have

$$f(P) = f(P + ev^T).$$

If P_* is a local minimizer of f, there must be a $\delta > 0$ such that:

$$\forall P \,:\, ||P_* - P||_F \le \delta, \quad f(P_*) \le f(P).$$

Then, for all $\hat{P}$ such that $\|\hat{P} - (P_* + ev^T)\|_F \le \delta$, we have $\|(\hat{P} - ev^T) - P_*\|_F \le \delta$, thus

$$f(\hat{P}) = f(\hat{P} - ev^T) \ge f(P_*) = f(P_* + ev^T)$$

implying that $P_* + ev^T$ is also a local minimizer. The other implication follows similarly. □

Proposition 3.6. *The configuration matrix $L_* \in \mathbb{R}^{(n-1)\times d}$ is a local minimizer of the function f_L (see (2.4)) if, and only if, all configurations in $\{L_*Q : Q \in \mathcal{O}\}$ are local minimizers of f_L.*

Proof. We now exploit the fact that the function f_L is invariant w.r.t. rotations: for any configuration L, and $Q \in \mathcal{O}$, we have

$$f_L(L) = f_L(LQ).$$

If L_* is a local minimizer of f_L, there must be a $\delta > 0$ such that:

$$\forall L \,:\, ||L_* - L||_F \le \delta, \quad f_L(L_*) \le f_L(L).$$

Then, for all $\hat{L}$ such that $\|\hat{L} - L_*Q\|_F \le \delta$, we have

$$\|\hat{L}Q^T - L_*\|_F = \|\hat{L} - L_*Q\|_F \le \delta,$$

thus

$$f_L(\hat{L}) = f_L(\hat{L}Q^T) \ge f_L(L_*) = f_L(L_*Q)$$

implying that L_*Q is also a local minimizer. The other implication follows similarly. □

The local minimizers of the two functions in equations (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*

$$v_* = \frac{1}{n}P_*^T e \in \mathbb{R}^d,\ P_{v_*} = P_* - ev_*^T,\ L_* = V^T P_{v_*}.$$

Then, L_* is a local minimizer of (2.4) if, and only if, P_{v_*} and P_* are local minimizers of (2.2).

Proof. First, recall that VV^T is the orthogonal projection onto $e^\perp$ and that the columns of P_{v_*} are centered. Thus, we have $VL_* = VV^TP_{v_*} = P_{v_*}$. Sufficiency: Let P_{v_*} be a local minimizer of (2.2). Then, there exists $\delta > 0$ such that

$$f(P_{v_*}) \le f(P),\ \forall P \,:\, \|P - P_{v_*}\|_F \le \delta. \tag{3.16}$$

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

$$f_L(L_*) = f(VL_*) = f(P_{v_*}) \le f(\hat{P}) = f(VL) = f_L(L),$$

where the inequality is due to $\|\hat{P} - P_{v_*}\|_F = \|VL - VL_*\|_F = \|L - L_*\|_F \leq \delta$ and (3.16), and the equalities hold by the definition of f_L.
Necessity: Suppose L_* is a local minimizer of $f_L(L)$, meaning there exists $\delta > 0$ such that

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

For any configuration P with $\|P - P_{v_*}\|_F \leq \delta$, define its centroid $v = P^T e/n$. Then, there exists $L \in \mathbb{R}^{(n-1)\times d}$ such that the centered configuration can be expressed as $P = VL + ev^T$. This implies that $P - P_{v_*} = V(L - L_*) + ev^T$. As $V(L - L_*)$ and ev^T are orthogonal, we get

$$\|L - L_*\|_F^2 = \|V(L - L_*)\|_F^2 = \|P - P_{v_*}\|_F^2 - \|ev^T\|_F^2 \leq \delta^2. \tag{3.18}$$

Now, from (3.17) and (3.18), we have

$$f(P) = f(VL + ev^T) = f(VL) \geq f(VL_*) = f(P_{v*}),$$

implying that P_{v*} is a local minimizer of $f(P)$. According to Theorem 3.5, P_* is also a local minimizer of $f(P)$. □

From Theorem 3.6, we know that for the case of $d \geq 2$, if L is a local minimizer of $f_L(L)$, then $\{LQ : Q \in \mathcal{O}\}$ is a local minimizer of $f_L(L)$. This means that when $d \geq 2$, any local minimizer of $f_L(L)$ is nonisolated and has a singular Hessian matrix.

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

Theorem 3.8. *The following statements hold.*

1. *If L_* is a local minimizer of f_L, then any ℓ_* satisfying*

$$L_* = \mathcal{L}\mathrm{Triag}(\ell_*)Q^T, \tag{3.19}$$

   for some $Q \in \mathcal{O}$, 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 L_* is a local minimizer of f_L, meaning there exists $r > 0$ such that

$$f_L(L) \geq f_L(L_*),\ \forall L\ :\ \|L - L_*\|_F \leq r. \tag{3.20}$$

For any $\ell \in \mathbb{R}^{t_\ell}$ satisfying $\|\ell - \ell_*\| \le r$, we let $L = \mathcal{L}\mathrm{Triag}(\ell)Q^T$ and we have

$$\|L - L_*\|_F = \|\,\mathcal{L}\mathrm{Triag}(\ell) - \mathcal{L}\mathrm{Triag}(\ell_*)\|_F = \|\ell - \ell_*\| \le r.$$

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

$$f_\ell(\ell) = f_L(\mathcal{L}\mathrm{Triag}(\ell)) = f_L(L) \ge f_L(L_*) = f_L(\mathcal{L}\mathrm{Triag}(\ell_*)) = f_\ell(\ell_*).$$

Therefore, ℓ_* is a local minimizer of f_ℓ.

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

$$\lim_{k\to+\infty} L_k = L_*,\ \ f_L(L_k) < f_L(L_*). \tag{3.21}$$

Consider the QR decompositions of $L_k^T, k = 1, 2, \dots$, i.e., there exist $Q_k \in \mathcal{O}, k = 1, 2, \dots$, and upper triangular matrices $R_k, k = 1, 2, \dots$, such that

$$L_k^T = Q_k R_k, k = 1, 2, \dots. \tag{3.22}$$

Since $\|Q_k\|_2 = 1$ for all $k = 1, 2, \dots$, the bounded sequence $\{Q_k\}$ has a convergent subsequence. Without loss of generality, we directly assume that $\lim_{k\to+\infty} Q_k = Q_*$. According to (3.21) and (3.22), we have

$$R_*^T := L_* Q_* = \lim_{k\to+\infty} L_k Q_k = \lim_{k\to+\infty} R_k^T. \tag{3.23}$$

By (3.23), R_*^T is a triangular matrix. Since the first d rows of $L_* = \mathcal{L}\mathrm{Triag}(\ell_*)$ are linearly independent, the QR factorization of L_*^T is unique except for signs in each dimension. Thus, Q_* is a diagonal matrix with diagonal elements being -1 or 1. Let

$$\ell_k = \mathcal{L}\mathrm{Triag}^*(R_k^T Q_*^T), k = 1, 2, \dots.$$

By (3.23), we get

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

By (3.21), we have

$$\begin{aligned} f_\ell(\ell_k) = f_L(R_k^T) = f_L(L_k) < f_L(L_*) &= f_L(R_*^T Q_*^T) \\ &= f_L(R_*^T) = f_\ell(\ell_*). \end{aligned}$$

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

□

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 $v \in \mathbb{R}^d$, we have*

$$\langle f'(P), ev^T \rangle = 0 \text{ and } f''(P)(ev^T, ev^T) = 0.$$

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

$$f(P + tev^T) \;=\; f(P) + t\langle f'(P), ev^T \rangle + \tfrac{t^2}{2}\langle f''(P)(ev^T), ev^T \rangle + o(t^2).$$

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

$$\langle f'(P), ev^T \rangle = \langle f''(P)(ev^T), ev^T \rangle = 0.$$

Moreover, we claim that

$$f''(P)(ev^T) = 0. \tag{3.24}$$

According to (3.6) and (3.7), we have

$$f''(P)(ev^T) \;=\; 4\,[J^*J]\operatorname{vec}(ev^T) + 2\,[\operatorname{vec}\mathcal{K}^*F(P)\operatorname{Mat}]\operatorname{vec}(ev^T).$$

Since $\operatorname{null}(\mathcal{K}) = \operatorname{range}(S_e)$ and $\operatorname{range}(\mathcal{K}^*) = \{S \in \mathbb{S}^n : \; Se = 0\}$ in Items 4 and 5 of Theorem 3.1, we have

$$J(\operatorname{vec}(ev^T)) = \mathcal{K}\left(\frac{Pve^T + ev^TP^T}{2}\right) = 0,\; \mathcal{K}^*F(P)(ev^T) = 0.$$

Thus, (3.24) holds. ☐

Theorem 3.10. *For $P \in \mathbb{R}^{n\times d}$, denote $v = (P^Te)/n \in \mathbb{R}^d$, $P_v = P - ev^T$, $L = V^TP_v$ 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 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^T.*

Proof. Since

$$
\begin{aligned}
& f(P+t\Delta P) = f(P) + t\langle f'(P), \Delta P\rangle + \frac{t^2}{2}\langle f''(P)(\Delta P), \Delta P\rangle + o(t^2) \\
= \quad & f(P_v+t\Delta P) = f(P_v) + t\langle f'(P_v), \Delta P\rangle + \frac{t^2}{2}\langle f''(P_v)(\Delta P), \Delta P\rangle + o(t^2)
\end{aligned}
$$

holds for all $\Delta P \in \mathbb{R}^{n\times d}$ and $t \in \mathbb{R}$, we have

$$f'(P_v) = f'(P), \;\; f''(P_v) = f''(P). \tag{3.25}$$

By

$$
\begin{aligned}
& f_L(L+t\Delta L) = f_L(L) + t\langle f_L'(L), \Delta L\rangle + \frac{t^2}{2}\langle f_L''(L)(\Delta L), \Delta L\rangle + o(t^2) \\
= \quad & f_L(R^T+t\Delta LQ) = f_L(R^T) + t\langle f_L'(R^T), \Delta LQ\rangle + \frac{t^2}{2}\langle f_L''(R^T)(\Delta LQ), \Delta LQ\rangle + o(t^2),
\end{aligned}
$$

we have

$$f_L'(R^T) = 0 \Leftrightarrow f_L'(L) = 0, \;\; f_L''(R^T) \succeq 0 \Leftrightarrow f_L''(L) \succeq 0.$$ [8]

Thus, (i)$\Leftrightarrow$(ii) and (iii)$\Leftrightarrow$(iv).

Now we prove (ii)$\Leftrightarrow$(iii). First, we prove the equivalence of their first-order necessary conditions. According to (3.10) and range$(\mathcal{K}^*) = \mathcal{S}_C^n$ (Theorem 3.1, Item 5), we have

$$e^T f'(P) = 2e^T\mathcal{K}^*(F(P))P = 0. \tag{3.26}$$

By Theorem 3.3, (3.25), (3.26), the definition of V, and Item 5 of Theorem 3.1, we obtain

$$f'(P_v) = 0 \Longleftrightarrow f_L'(L) = V^T f'(P_v) = 0.$$

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

$$f_L''(L)(\Delta L, \Delta L) = V^T f''(VL)V(\Delta L, \Delta L) = f''(VL)(V\Delta L, V\Delta L). \tag{3.27}$$

[8] Note that $f_L''(L)$ is a positive semidefinite linear operator on $\mathbb{R}^{(n-1)\times d}$.

According to (3.24) in Theorem 3.9 and (3.27), we have $f_L''(L)(\Delta L, \Delta L) \geq 0$ if, and only if,

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

(Note that we have proved a slightly stronger statement as the semidefinite condition is treated separately from stationarity.) $\square$

Remark 3.11. *The reduction from* (2.4) *to* (2.8) *may 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, to have a local minimizer correspondence, the assumption that the first* d *rows of* R_*^T *is linear independent in Theorem 3.8, Item 2 is needed.*

4. Second-Order Optimality Conditions

In this section, we present the optimality conditions and derive a sufficient condition such that there is no **lngm**. First of all, the necessary and sufficient characterization for the global minimizer is 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, $F(P) = \mathcal{D}(P) - \bar{D} = 0$. $\square$

In order to further characterize the second-order optimality conditions, we 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. For* H_2*, the following holds:*

- $H_2 \succeq 0$ *when* $F(P)$ *is element-wise nonnegative,*
- $H_2 \preceq 0$ *when* $F(P)$ *is element-wise nonpositive.*

Proof. For any $x \in \mathbb{R}^{nd}$,

$$x^T H_1 x = \langle x, J^* J x\rangle = \langle Jx, Jx\rangle \geq 0.$$

Thus, H_1 is always positive semidefinite. By Theorem 3.1, Item 5, if $F(P) \geq (\leq)\ 0$, then $\mathcal{K}^*(F(P)) \succeq (\preceq)\ 0$, which implies

$$x^T H_2 x = \langle x, \operatorname{Mat}^* \mathcal{K}^* F(P) \operatorname{Mat} x\rangle = \langle \operatorname{Mat} x, \mathcal{K}^*(F(P)) \operatorname{Mat} x\rangle \geq (\leq)\ 0$$

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

Lemma 4.3. *The matrix H_2 is the zero matrix if, and only if, $F(P) = 0$ holds, which is equivalent to P being a global minimizer of* (2.2).

Proof. By Theorem 3.1, Item 5, $\mathcal{K}^*(S) = 2(\operatorname{Diag}(Se) - S)$ and $\operatorname{null}(\mathcal{K}^*) = \operatorname{Diag}(\mathbb{R}^n)$. Since $\operatorname{diag}(F(P)) = \operatorname{diag}(\mathcal{D}(P)) - \operatorname{diag}(\bar{D}) = 0$ is always true, $\mathcal{K}^*(F(P)) = 0$ holds if, and only if, $F(P) = 0$. □

Lemma 4.4. *Let $\bar{P}$, with $\bar{P}^T e = 0$, be a global minimizer of* (2.2). *Suppose that P is a stationary point for* (2.2) *but is not a global optimizer. Then H_2 is not positive semidefinite. Specifically,* $\operatorname{vec}(\bar{P})^T H_2 \operatorname{vec}(\bar{P}) < 0$.

Proof. By (3.8), we have $\langle P, \mathcal{K}^* F(P) P\rangle = 0$, and then

$$\begin{array}{rcl} \operatorname{vec}(\bar{P})^T H_2 \operatorname{vec}(\bar{P}) & = & \langle \bar{P}, \mathcal{K}^* F(P) \bar{P}\rangle - \langle P, \mathcal{K}^* F(P) P\rangle \\ & = & \langle \mathcal{K}(\bar{P}\bar{P}^T), F(P)\rangle - \langle \mathcal{K}(PP^T), F(P)\rangle \\ & = & \langle \mathcal{K}(\bar{P}\bar{P}^T) - \mathcal{K}(PP^T), F(P)\rangle \\ & = & \langle \bar{D} - \mathcal{D}(P), F(P)\rangle \\ & = & -\langle F(P), F(P)\rangle \\ & < & 0. \end{array}$$

The last inequality holds because P is not a global minimizer, which implies $F(P) \neq 0$ according to Theorem 4.1. □

Under the condition of Theorem 4.4, we have known that

$$\langle \bar{P}, \mathcal{K}^* F(P) \bar{P}\rangle < 0,$$

which implies that

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

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. Suppose P is a stationary point. As $\bar{D} = 0$, we get $\bar{p}_1 = \ldots = \bar{p}_n$ holds. Since $\mathcal{K}^* F(P)$ is a Laplacian (sum of its columns is zero), we have $\mathcal{K}^* F(P)\bar{P} = 0$. Combining this with the first-order condition (3.8), we have

$$\begin{aligned}\operatorname{vec}(\bar{P})^T H_2 \operatorname{vec}(\bar{P}) &= \langle \bar{P}, \mathcal{K}^*(F(P))\bar{P}\rangle - \langle P, \mathcal{K}^*(F(P))P\rangle \\ &= 0.\end{aligned}$$

From Theorem 4.4, we conclude that any stationary point P for $f(P)$ is a global minimizer. □

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

Theorem 4.6. *For $\bar{D} \neq 0$ and P with $\mathcal{D}(P) = 0$ (i.e., all $p_i \equiv p$), P is a stationary point and has nontrivial negative semidefinite Hessian:*

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

Proof. Since $p_1 = \ldots = p_n = p$ and $\mathcal{K}^*(F(P))$ is a Laplacian, P satisfies the first-order optimality condition (3.8). Since $f(P) = \|\bar{D} - \mathcal{D}(P)\|_F^2 = \|\bar{D}\|_F^2 > 0$, P is not a global minimizer. By

$$P\Delta P^T + \Delta P P^T = ep^T \Delta P^T + \Delta P p e^T = e(\Delta P p)^T + (\Delta P p)e^T,$$

and $\operatorname{null}(\mathcal{K}) = \operatorname{range}(S_e)$ (Theorem 3.1, Item 4), $J = 0$ defined in (3.7) holds, and then $H_1 = 0$. Since $\mathcal{D}(P) = 0$ and $\bar{D} \geq 0$, $F(P) \leq 0$ holds. According to Theorem 4.2 and Theorem 4.3, $0 \neq H_2 \preceq 0$ holds. Therefore, the Hessian matrix satisfies $0 \neq 4H_1 + 2H_2 \preceq 0$. □

Remark 4.7. *If P is a local* **maximizer** *of f, then necessarily $\mathcal{D}(P) = 0$ ($p_1 = \cdots = p_n$). To see this, observe that $t = 1$ must locally maximize $g(t) = f(tP)$. 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, we analyze (2.4) for convenience.

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

Proof. Since $0 = f'_L(L) = 2V^T\mathcal{K}^*F(VL)VL$, where the last equality follows from (3.12) and (3.10), 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 $\text{range}(\mathcal{K}^*) = \mathcal{S}^n_C$ from Theorem 3.1, Item 5, we conclude that $\mathcal{K}^*(F(P)) = 0$, and then $H_2 = 0$. Thus, L is a global minimizer according to Theorem 4.3. □

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

$$\begin{array}{rcl} \text{vec}(aw^T)^T H_2 \,\text{vec}(aw^T) & = & \langle aw^T, \mathcal{K}^*F(P)(aw^T)\rangle \\ & = & \text{Tr}\big(\mathcal{K}^*F(P)(aw^T)(aw^T)^T\big) \\ & = & w^Tw\,\text{Tr}\big(\mathcal{K}^*F(P)aa^T\big) \\ & = & w^Twa^T\mathcal{K}^*F(P)a \\ & < & 0. \end{array}$$

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

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

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

$$J\,\text{vec}(aw^T) = \mathcal{K}\mathcal{S}VL\mathcal{T}(aw^T).$$

By (4.3), we have

$$L\mathcal{T}(aw^T) = L\begin{bmatrix} a_1w & a_2w & \dots & a_{n-1}w \end{bmatrix} = 0.$$

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

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

which implies the second-order necessary optimality condition (3.9) fails. □

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 conditions 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 $\operatorname{rank}(L) = n-1$, then L is globally optimal by Theorem 4.8. If $\operatorname{rank}(L) < n-1$, then $\operatorname{rank}(L) < d$. 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 Theorem 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.

5. lngms Examples

We now provide instances with **lngm**s. The data and codes are available at https://github.com/MengmengSong97/EDM-code.

We first provide an analytical example with $d = 1$ with a specific simple structure for the points in $\mathbb{R}^d$, see Section 5.1. Then in Sections 5.2.1 and 5.2.2 we give two examples where we numerically obtain approximate second-order stationary points. Then, we analytically prove that the assumptions of the Kantorovich theorem hold at these two points, i.e., this implies that there exists **lngm**s in the neighborhoods. We consider the sensitivity analysis needed to analytically prove that our examples have **lngm**s. We exploit the strength of the classical Kantorovich theorem for the convergence of Newton's Method to exact stationary points when using the approximate stationary points that we found as starting points. See Theorem 5.6 and Theorem 5.10, below.

*5.1. An Explicit Example with a **lngm** with $d = 1$*

We now present a simple explicit example with an **lngm**, where we can analytically verify the **lngm**.

Example 5.1. *Let*

$$d = 1, n > 6; \quad \bar{P}^T = [\bar{p}_1, \cdots, \bar{p}_n],\ \tilde{P}^T = [\tilde{p}_1, \cdots, \tilde{p}_n] \in \mathbb{R}^{d \times n},$$

with

$$\bar{p}_1 = 2,\ \bar{p}_2 = 0,\ \bar{p}_3 = \cdots = \bar{p}_n = 1 \quad \textit{and} \quad \tilde{p}_1 = \tilde{p}_2 = 0,\ \tilde{p}_3 = \cdots = \tilde{p}_n = 1.$$

We now continue and illustrate that $\tilde{P}$ is a **lngm** of (2.2) in Theorem 2.1. Let $E_{2,n-2}$ be the $2 \times n-2$ matrix of all ones, and $\begin{bmatrix}0\end{bmatrix}$ denote the matrix of zeros of appropriate size. From the definitions of $F(\cdot)$, $\mathcal{K}(\cdot)$ and $\mathcal{K}^*(\cdot)$, we have:

$$\begin{aligned} F(\tilde{P}) &= \mathcal{D}(\tilde{P}) - \mathcal{D}(\bar{P}) \\ &= \begin{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} & E_{2,n-2} \\ E_{2,n-2}^T & \begin{bmatrix}0\end{bmatrix} \end{bmatrix} - \begin{bmatrix} \begin{bmatrix} 0 & 4 \\ 4 & 0 \end{bmatrix} & E_{2,n-2} \\ E_{2,n-2}^T & \begin{bmatrix}0\end{bmatrix} \end{bmatrix} \qquad (5.1) \\ &= \begin{bmatrix} \begin{bmatrix} 0 & -4 \\ -4 & 0 \end{bmatrix} & 0_{2,n-2} \\ 0_{2,n-2}^T & \begin{bmatrix}0\end{bmatrix} \end{bmatrix}; \qquad (5.2) \end{aligned}$$

$$\mathcal{K}^*(F(\tilde{P})) = 2 \begin{bmatrix} \begin{bmatrix} -4 & 4 \\ 4 & -4 \end{bmatrix} & \begin{bmatrix}0\end{bmatrix} \\ \begin{bmatrix}0\end{bmatrix} & \begin{bmatrix}0\end{bmatrix} \end{bmatrix}.$$

Moreover, we have

$$\begin{aligned} & \langle \mathcal{K}(\tilde{P}\Delta P^T + \Delta P \tilde{P}^T), \mathcal{K}(\tilde{P}\Delta P^T + \Delta P \tilde{P}^T) \rangle \\ = \ & 4 \sum_{i \neq j} [(\tilde{p}_i - \tilde{p}_j)(\Delta p_i - \Delta p_j)]^2 \\ = \ & 8 \Delta P^T \begin{bmatrix} \sum_{i=1}^n (\tilde{p}_1 - \tilde{p}_i)^2 & -(\tilde{p}_1 - \tilde{p}_2)^2 & \cdots & -(\tilde{p}_1 - \tilde{p}_n)^2 \\ -(\tilde{p}_1 - \tilde{p}_2)^2 & \sum_{i=1}^n (\tilde{p}_2 - \tilde{p}_i)^2 & \cdots & -(\tilde{p}_2 - \tilde{p}_n)^2 \\ \vdots & \vdots & \cdots & \vdots \\ -(\tilde{p}_1 - \tilde{p}_2)^2 & -(\tilde{p}_2 - \tilde{p}_n)^2 & \cdots & \sum_{i=1}^n (\tilde{p}_n - \tilde{p}_i)^2 \end{bmatrix} \Delta P \\ = \ & 8 \Delta P^T \begin{bmatrix} (n-2)I & -E_{2,n-2} \\ -E_{2,n-2}^T & 2I \end{bmatrix} \Delta P. \end{aligned}$$

Then we can compute from (3.1) and (3.2) that the derivative (gradient) is

$$f'(\tilde{P}) = 4[\mathrm{Diag}(F(\tilde{P})e) - F(\tilde{P})]\tilde{P} = 0. \qquad (5.3)$$

And the Hessian quadratic form is

$$
\begin{aligned}
& f''(\tilde{P})(\Delta P, \Delta P) \\
= \; & \langle \mathcal{K}(\tilde{P}\Delta P^T + \Delta P \tilde{P}^T), \mathcal{K}(\tilde{P}\Delta P^T + \Delta P \tilde{P}^T)\rangle + 2\langle F(\tilde{P}), \mathcal{K}(\Delta P \Delta P^T)\rangle \\
= \; & 8\Delta P^T \left(\begin{bmatrix} (n-2)I & -E_{2,n-2} \\ -E_{2,n-2}^T & 2I \end{bmatrix} + \begin{bmatrix} \begin{bmatrix} -2 & 2 \\ 2 & -2 \end{bmatrix} & [0] \\ [0] & [0] \end{bmatrix} \right) \Delta P.
\end{aligned}
$$

The Hessian $f''(\tilde{P}) = \nabla^2 f(\tilde{P})$ is a rank 3 update of $16I$. It is positive semidefinite with nullspace $\operatorname{span}(e)$ if, and only if, $n \geq 7$. (It is indefinite when $n < 6$.) Thus, $\tilde{P}$ is a second-order stationary point of the problem with data given above. Next, we prove that $\tilde{P}$ is a **lngm** by considering the reduced formulation (2.4).

To center as done in Theorem 3.7, we let

$$
\tilde{v} = \frac{\tilde{P}^T e}{n} \in \mathbb{R}, \;\; \tilde{P}_* = \tilde{P} - e\tilde{v}^T, \;\; \tilde{L} = V^T \tilde{P}_*,
$$

to obtain $\tilde{L}$. According to Theorem 3.10, $\tilde{L}$ is a second-order stationary point of the function $f_L(L)$. By Theorem 3.3, we have $f_L''(\tilde{L}) = V^T f''(V\tilde{L})V = V^T f''(\tilde{P})V$. Since $f''(\tilde{P})$ has a one-dimensional nullspace ($\operatorname{span}\{e\}$), its restriction to $\operatorname{span}\{e\}^\perp$ is positive definite. Thus, $f_L''(\tilde{L})$ is positive definite. This implies that $\tilde{L}$ satisfies the second-order sufficient optimality conditions for a strict local minimum of $f_L(L)$. By Theorem 3.7 and Theorem 4.1, $\tilde{P}$ is in fact a **lngm** of $f(P)$. Note that the objective function value $f(\tilde{P}) = 16 > 0 = f(\bar{P})$, thus confirming that $\tilde{P}$ is not a global minimum.

5.2. Examples via Kantorovich Theorem and Sensitivity Analysis

We now present Theorems 5.2 and 5.7, with $d = 1, 2$, respectively, where we first find an approximate second-order stationary point $\tilde{L}$ numerically that has a sufficiently large (positive) objective value; and then we prove that there is a ***lngm*** *nearby* using the Kantorovich theorem and sensitivity analysis.

5.2.1. Case $d = 1$

In this case we analyze a configuration matrix $\tilde{P} = V\tilde{L} \in \mathbb{R}^{n\times 1}, \tilde{L} \in \mathbb{R}^{(n-1)\times 1}$, satisfying:

$$\begin{array}{rll} \text{Objective value :} & f(\tilde{P}) = f_L(\tilde{L}) > \tilde{f}_L, & \text{for some (large) } \tilde{f}_L > 0; \\ \text{Near stationarity :} & \|\nabla f_L(\tilde{L})\| < \tilde{g}_L, & \text{for some (small) } \tilde{g}_L > 0; \\ \text{Local Convexity :} & \lambda_{\min}(\nabla^2 f_L(\tilde{L})) > \tilde{\lambda}_L, & \text{for some (large) } \tilde{\lambda}_L > 0. \end{array} \tag{5.4}$$

While theoretically exact, the floating-point representations introduce round-off errors in finite-precision computations. Our MATLAB implementation performs complete finite-precision arithmetic analysis in order to compute the rigorous bounds $\tilde{f}_L$, $\tilde{g}_L$, and $\tilde{\lambda}_L$ (see Footnote 9).

We apply the classical Kantorovich theorem, e.g., [6, Thm 5.3.1], to show that there is a point *nearby* that satisfies: (i) it is an *exact* stationary point; (ii) the function value is positive; and (iii) the Hessian is still positive definite. This provides an analytic proof that we have a theoretically verified **lngm** near $\tilde{L}$.

Example 5.2. *An example with $n = 50, d = 1$ is given, with data $\bar{P} = V\bar{L}$, $\tilde{P} = V\tilde{L} \in \mathbb{R}^{n\times d}$. (See the footnote below.[9]) 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. $\tilde{L}$ is a numerically convergence point obtained by a trust region method with random initialization, where the objective value is*

$$f_L(\tilde{L}) > 2.65 \times 10^3, \tag{5.5}$$

the absolute and relative gradient norms are

$$\|\nabla f_L(\tilde{L})\| < 10^{-7}, \ \frac{\|\nabla f_L(\tilde{L})\|}{1 + f_L(\tilde{L})} < 3.53 \times 10^{-18}, \tag{5.6}$$

and the least eigenvalue of the Hessian matrix is

$$2.12 \times 10^2 > \lambda_{\min}(\nabla^2 f_L(\tilde{L})) > 2.10 \times 10^2. \tag{5.7}$$

In the following, we will verify that $f_L(L)$ has a **lngm**.

[9] The data and codes are available at https://github.com/MengmengSong97/EDM-code

Remark 5.3. *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, which simplifies the Hessian* $4H1{+}2H_2$ *(see* (3.6) *and* (3.11)*) by keeping only* $4H_1$ *and omitting* $2H_2$ *(same for* L*, cf.* (3.13)*). This approximation relies on the assumption that* $f_L(L)$ *is near zero, a condition we intentionally avoid, as this would yield the global minimum.*

Now, we find an estimate for the Lipschitz constant $\gamma > 0$ for the Hessian matrix of f_L. From our numerical output, we know 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 is now estimated in Theorem 5.4.

Proposition 5.4. *Let* $r > 0$ *and* $\tilde{L} \in \mathbb{R}^{(n-1)\times d}$ *be given. If*

$$\gamma \geq 24\sqrt{2}\left(\sum_{i,j} \|(V\tilde{L})[i,:] - (V\tilde{L})[j,:]\|_F + 2n^{3/2}r\right), \tag{5.8}$$

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

$$\|\nabla^2 f_L(\hat{L}) - \nabla^2 f_L(\check{L})\|_2 \leq \gamma\|\hat{L} - \check{L}\|_F, \quad \textit{for all } \hat{L}, \check{L} \in B_r(\tilde{L}). \tag{5.9}$$

Moreover,

$$\lambda_{\min}(\nabla^2 f_L(L)) \geq \lambda_{\min}(\nabla^2 f_L(\tilde{L})) - \gamma r, \quad \textit{for all } L \in B_r(\tilde{L}). \tag{5.10}$$

Proof. By the definition of the induced norm, (5.9) is equivalent to

$$|f_L''(\hat{L})(\Delta L, \Delta L) - f_L''(\check{L})(\Delta L, \Delta L)| \leq \gamma\|\hat{L} - \check{L}\|, \tag{5.11}$$

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

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

According to (3.3) and Theorem 3.3, we have

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

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

$$\begin{array}{rl} & f''_L(\hat{L})(\Delta L, \Delta L) - f''_L(\check{L})(\Delta L, \Delta L) \\ = & 4\sum_{i,j}[(\hat{p}_i - \hat{p}_j)^T(\Delta p_i - \Delta p_j)]^2 - [(\check{p}_i - \check{p}_j)^T(\Delta p_i - \Delta p_j)]^2 \\ & \quad +2\sum_{i,j}(\|\hat{p}_i - \hat{p}_j\|^2 - \|\check{p}_i - \check{p}_j\|^2)\|\Delta p_i - \Delta p_j\|^2 \\ = & 4\sum_{i,j}(\hat{p}_i - \hat{p}_j - \check{p}_i + \check{p}_j)^T(\Delta p_i - \Delta p_j)(\hat{p}_i - \hat{p}_j + \check{p}_i - \check{p}_j)^T(\Delta p_i - \Delta p_j) \\ & \quad +2\sum_{i,j}(\hat{p}_i - \hat{p}_j - \check{p}_i + \check{p}_j)^T(\hat{p}_i - \hat{p}_j + \check{p}_i - \check{p}_j)\|\Delta p_i - \Delta p_j\|^2. \end{array}$$

Then,

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

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

$$\|\Delta p_i - \Delta p_j\|^2 \leq 2(\|\Delta p_i\|^2 + \|\Delta p_j\|^2) \leq 2.$$

From $\hat{L}, \check{L} \in B_r(\tilde{L})$, it follows that $\hat{P}, \check{P} \in B_r(\tilde{P})$. Then, we have

$$\begin{array}{rl} \|\hat{p}_i - \hat{p}_j - \check{p}_i + \check{p}_j\| \leq & \|\hat{p}_i - \check{p}_i\| + \|\hat{p}_j - \check{p}_j\| \\ \leq & \sqrt{2}\sqrt{\|\hat{p}_i - \check{p}_i\|^2 + \|\hat{p}_j - \check{p}_j\|^2} \\ \leq & \sqrt{2}\|\hat{L} - \check{L}\|_F, \end{array}$$

where the first inequality follows from the triangle inequality, the second from the Cauchy-Schwarz inequality, and the third from the fact that

$$\|\hat{P} - \check{P}\|_F = \|V\hat{L} - V\check{L}\|_F = \|\hat{L} - \check{L}\|_F.$$

We also have

$$\begin{array}{rl} & \sum_{i,j}\|\hat{p}_i - \hat{p}_j + \check{p}_i - \check{p}_j\| \\ = & \sum_{i,j}\|2(\tilde{p}_i - \tilde{p}_j) + (\hat{p}_i - \tilde{p}_i) - (\hat{p}_j - \tilde{p}_j) + (\check{p}_i - \tilde{p}_i) - (\check{p}_j - \tilde{p}_j)\| \\ \leq & 2\sum_{i,j}\|\tilde{p}_i - \tilde{p}_j\| + \sum_{i,j}\|\hat{p}_i - \tilde{p}_i\| + \sum_{i,j}\|\hat{p}_j - \tilde{p}_j\| + \\ & \quad \sum_{i,j}\|\check{p}_i - \tilde{p}_i\| + \sum_{i,j}\|\check{p}_j - \tilde{p}_j\| \\ = & 2\sum_{i,j}\|\tilde{p}_i - \tilde{p}_j\| + n\sum_i\|\hat{p}_i - \tilde{p}_i\| + n\sum_j\|\hat{p}_j - \tilde{p}_j\| + \\ & \quad n\sum_i\|\check{p}_i - \tilde{p}_i\| + n\sum_j\|\check{p}_j - \tilde{p}_j\| \\ \leq & 2\sum_{i,j}\|\tilde{p}_i - \tilde{p}_j\| + 4n^{3/2}r, \end{array}$$

where the last inequality follows from the Hölder inequality. Thus,

$$\begin{array}{rl} |f''_L(\hat{L})(\Delta L, \Delta L) - f''_L(\check{L})(\Delta L, \Delta L)| \leq & 12\sqrt{2}\|\hat{L} - \check{L}\|_F \sum_{i,j}\|\hat{p}_i - \hat{p}_j + \check{p}_i - \check{p}_j\| \\ \leq & 24\sqrt{2}\|\hat{L} - \check{L}\|_F \left(\sum_{i,j}\|\tilde{p}_i - \tilde{p}_j\| + 2n^{3/2}r\right), \end{array}$$

applying (5.8) turns out (5.9). By (5.11), we have

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

Thus, we obtain (5.10). □

To verify the existence of a **lngm** for Theorem 5.2, we calculate the Lipschitz constant estimated in Theorem 5.4. Let $r = 10^{-3}$. Since

$$\sum_{i,j} \|(V\tilde{L})[i,:] - (V\tilde{L})[j,:]\| < 2.13 \times 10^3,$$

(5.8) gives

$$\gamma = 7.24 \times 10^4.$$

Moreover, by (5.10), we have

$$\lambda_{\min}(\nabla^2 f_L(L)) \geq 211 - 7.24 \times 10^4 \times r = 138.6 > 0, \quad \text{for all } L \in B_r(\tilde{L}). \tag{5.12}$$

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 5.5. *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.,*

$$\lambda_{\min}(\nabla^2 f_L(L)) > 0, \quad \textit{for all } L \in B_r(\tilde{L}). \tag{5.13}$$

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

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

Proof. By the positive definiteness assumption of the Hessian in the r-ball $B_r(\tilde{L})$, we can apply convexity of f_L in the ball. Therefore, for all $L \in \mathbb{R}^{(n-1)\times d}$ such that $\|L - \tilde{L}\|_F \leq \min\left\{ r, (f_L(\tilde{L}) - \bar{f}_L)/\|\nabla f_L(\tilde{L})\|_F \right\}$, we have

$$\begin{array}{rcl} f_L(L) & \geq & f_L(\tilde{L}) + \langle \nabla f_L(\tilde{L}), L - \tilde{L} \rangle \\ & \geq & f_L(\tilde{L}) - \|\nabla f_L(\tilde{L})\|_F \|L - \tilde{L}\|_F \\ & > & \bar{f}_L > 0. \end{array}$$

□

By (5.5), (5.6) and considering $\bar{f}_L = 10^3$, we get

$$\frac{f_L(\tilde{L}) - \bar{f}_L}{\|\nabla f_L(\tilde{L})\|} > \frac{2.65 \times 10^3 - 10^3}{10^{-7}} > r.$$

According to Theorem 5.5 with (5.12) we conclude that

$$f_L(L) > \bar{f}_L > 0, \quad \text{for all } L \in B_r(\tilde{L}). \tag{5.14}$$

We now apply the classical Kantorovich theorem to prove the existence of a unique **lngm** point within a certain neighborhood. We reword the version in [6, Thm 5.3.1].

Theorem 5.6. *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*

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

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

$$\beta := \|\nabla^2 f_L(\tilde{L})^{-1}\|_2 \quad \text{and} \quad \eta := \|\nabla^2 f_L(\tilde{L})^{-1}\nabla f_L(\tilde{L})\|.$$

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* $B_{r_1}(\tilde{L})$*, where*

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

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

Proof. The proof is a direct application of the Kantorovich theorem, e.g., [6, Thm 5.3.1], along with the above lemmas and corollaries in this section. □

As mentioned previously in this section, the conditions required in Theorem 5.5 and Theorem 5.6 are fulfilled with

$$r = 10^{-3},\ \gamma = 7.24 \times 10^4,\ \bar{f}_L = 10^3.$$

Plugging them and (5.5), (5.6), (5.7) into Theorem 5.6, we have

$$\begin{array}{l} 1/212 < \beta < 1/210, \\ \eta \le \|\nabla^2 f_L(\tilde{L})^{-1}\|_2 \|\nabla f_L(\tilde{L})\| \le 1/210 \times 10^{-7}, \\ \gamma_R = \beta\gamma < 1/210 \times 7.24 \times 10^4, \\ \alpha = \gamma_R \eta < (1/210 \times 7.24 \times 10^4) \times (1/210 \times 10^{-7}) < 1/2, \\ r_0 = \frac{(1-\sqrt{1-2\alpha})}{\beta\gamma} < \frac{1-\sqrt{1-2\times 1/210^2 \times 7.24 \times 10^{-3}}}{1/212 \times 7.24 \times 10^4} < r. \end{array}$$

Combining these with (5.12) and (5.14) via Theorem 5.6, we conclude that f_L has a **lngm** in $B_r(\tilde{L})$.

In Figure 5.1 we plot the known centered global minimizer $\bar{P}$ and the centered numerical found $\tilde{P}$ near which it has been proved there exists a **lngm**. It is interesting to observe that the $\tilde{p}_i \approx \bar{p}_i$ for all $i \neq i_0$, all except the one $\bar{p}_{i_0}$ with the biggest absolute value, i.e., $|\bar{p}_{i_0}| > |\bar{p}_i|$, for all $i \neq i_0$; the corresponding $\tilde{p}_{i_0}$ has the biggest absolute value among all $\tilde{p}_i$ for $i = 1, \ldots, n = 50$, and has an opposite sign to $\bar{p}_{i_0}$.

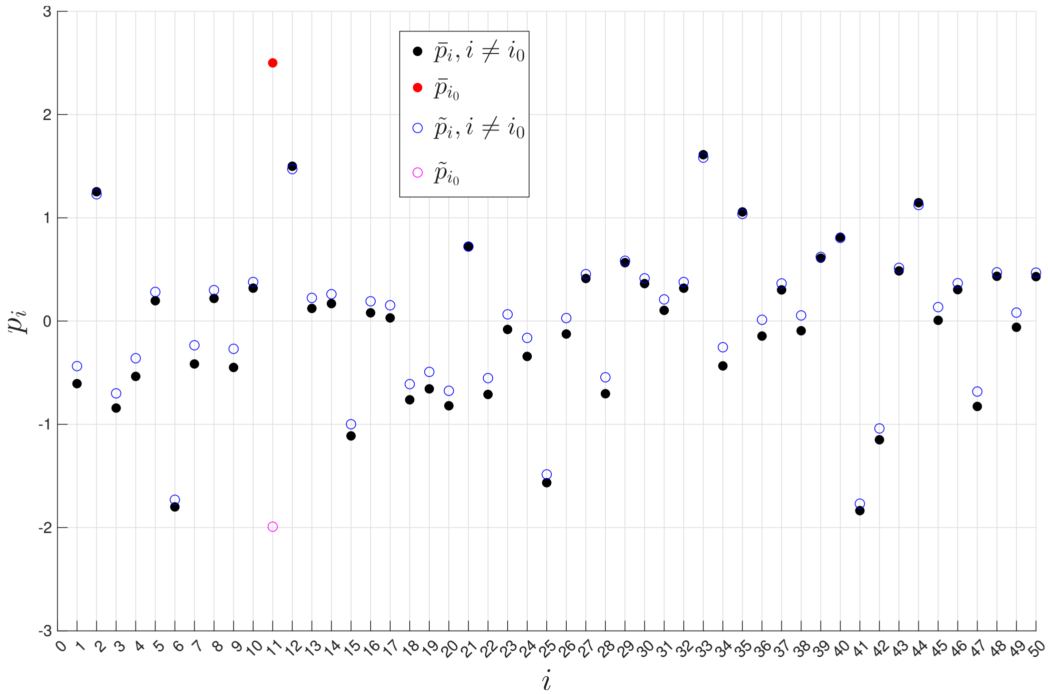

Figure 5.1: values of global min and numerical near **lngm**: $\bar{P}, \tilde{P} \in \mathbb{R}^{50\times 1}$; in Theorem 5.2.

Based on these observations, we generated examples with $d = 1$ and $n = 100$. We randomly generated $\bar{L}$, then get centered $\bar{P}$ by setting $\bar{P} = V\bar{L}$. Next, we find $\bar{p}_{i_0}$, which has the biggest absolute value among all $\bar{p}_i$ for $i = 1, \ldots, n$. We define the starting point $\hat{P}$ by setting $\hat{p}_i = \bar{p}_i$ for $i \neq i_0$ and $\hat{p}_{i_0} = -\bar{p}_{i_0}$. Starting from $\hat{L} = V^T\hat{P}$, the trust region method converges to a nonglobal second-order stationary point of $f_L(L)$ with high frequency.

5.2.2. Case $d = 2$

As mentioned in Section 2 for the case of $d > 1$, all local minimizers of $f_L(L)$ are nonisolated, 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 $d = 2$, and we analyze a configuration $\tilde{\ell} \in \mathbb{R}^{t_\ell}$ satisfying equivalent conditions to (5.4) but for f_ℓ:

$$\begin{array}{rlll} \text{Objective value}: & f_\ell(\tilde\ell) = f_\ell(\tilde\ell) > \tilde f_\ell, & \text{for some (large) } \tilde f_\ell > 0; \\ \text{Near stationarity}: & \|\nabla f_\ell(\tilde\ell)\| < \tilde g_\ell, & \text{for some (small) } \tilde g_\ell > 0; \\ \text{Local Convexity}: & \lambda_{\min}(\nabla^2 f_\ell(\tilde\ell)) > \tilde\lambda_\ell, & \text{for some (large) } \tilde\lambda_\ell > 0. \end{array} \tag{5.15}$$

We use Kantorovich Theorem to verify that this example has a strict local nonglobal minimizer ℓ_*, where the first 2 rows of $\mathcal{L}\mathrm{Triag}(\ell_*)$ are linearly independent. According to Theorems 3.7 and 3.8, $\mathcal{L}\mathrm{Triag}(\ell_*)$ is a local nonglobal minimizer of $f_L(L)$, and $V\,\mathcal{L}\mathrm{Triag}(\ell_*)$ is a local nonglobal minimizer of $f(P)$.

Example 5.7. *An example with $n = 100, d = 2$ is given, with data $\bar\ell, \tilde\ell \in \mathbb{R}^{197}$ presented, see Footnote 9. Matrix $\bar D$ is the distance matrix obtained from $\bar\ell$ by*

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

Thus, $\bar\ell$ is a global minimizer; $\tilde\ell$ is a numerically convergence point obtained by a trust region method with random initialization. The objective value is

$$f_\ell(\tilde\ell) > 9.99 \times 10^3, \tag{5.16}$$

the absolute and relative gradient norms are

$$\|\nabla f_\ell(\tilde\ell)\| < 9.23 \times 10^{-8},\ \frac{\|\nabla f_\ell(\tilde\ell)\|}{1 + f_\ell(\tilde\ell)} < 9.24 \times 10^{-12}, \tag{5.17}$$

and the least eigenvalue of the Hessian matrix is

$$8.58 > \lambda_{\min}(\nabla^2 f_\ell(\tilde\ell)) > 7.93. \tag{5.18}$$

In the following Theorem 5.8, we verify that there exists a **lngm** of $f_\ell(\ell)$.

Corollary 5.8. *Let $r > 0, \tilde\ell \in \mathbb{R}^{(n-1)\times d}$ be given. If*

$$\gamma \geq 24\sqrt{2}\left(\sum_{i,j}\|(V\,\mathcal{L}\mathrm{Triag}(\tilde\ell))[i,:] - (V\,\mathcal{L}\mathrm{Triag}(\tilde\ell))[j,:]\|_F + 2n^{3/2}r\right), \tag{5.19}$$

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

$$\|\nabla^2 f_\ell(\hat{\ell}) - \nabla^2 f_\ell(\check{\ell})\|_2 \le \gamma\|\hat{\ell} - \check{\ell}\|, \quad \textit{for all } \hat{\ell}, \check{\ell} \in B_r(\tilde{\ell}).$$

Moreover,

$$\lambda_{\min}(\nabla^2 f_\ell(\ell)) \ge \lambda_{\min}(\nabla^2 f_\ell(\tilde{\ell})) - \gamma r, \quad \textit{for all } \ell \in B_r(\tilde{\ell}).$$

Proof. The results follow from the fact that, according to (3.15), the Hessian matrix of f_ℓ at $\tilde{\ell}$ is a submatrix of the Hessian matrix of f_L at $\mathcal{L}\mathrm{Triag}(\tilde{\ell})$. The steps are similar to the proof of Theorem 5.4. □

In Theorem 5.7, we have

$$\sum_{i,j} \|(V\,\mathcal{L}\mathrm{Triag}(\tilde{\ell}))[i,:] - (V\,\mathcal{L}\mathrm{Triag}(\tilde{\ell}))[j,:]\| < 1.803 \times 10^4.$$

Let $r = 10^{-5}$. Then $\gamma = 6.12 \times 10^5$ satisfies (5.19). According to Theorem 5.8 and (5.18), we have

$$\lambda_{\min}(\nabla^2 f_\ell(\ell)) \ge \lambda_{\min}(\nabla^2 f_\ell(\tilde{\ell})) - \gamma r > 0, \quad \text{for all } \ell \in B_r(\tilde{\ell}). \tag{5.20}$$

We now continue to extend the results from Section 5.2.1 to this $d = 2$ case.

Corollary 5.9. *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 $\tilde{\ell}$:*

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

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

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

Proof. The results follow as in Theorem 5.5. □

Let $\bar{f}_L = 10^3$. By (5.16) and (5.17), we have

$$\frac{f_\ell(\tilde{\ell}) - \bar{f}_\ell}{\|\nabla f_\ell(\tilde{\ell})\|} > \frac{8.99 \times 10^3}{9.23 \times 10^{-8}} > r.$$

Thus, according to Theorem 5.9, we have

$$f_\ell(\ell) > \bar{f}_\ell > 0, \quad \text{for all } \ell \in B_r(\tilde{\ell}). \tag{5.21}$$

Corollary 5.10. *Let $\tilde{\ell} \in \mathbb{R}^{t_\ell}$ be given and $r \in \mathbb{R}_{++}$ be found such that*

$$\nabla^2 f_\ell(\ell) \succ 0, \quad \textit{for all } \ell \in B_r(\tilde{\ell}), \tag{5.22}$$

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

$$f_\ell(\ell) > \bar{f}_\ell > 0, \quad \textit{for all } \ell \in B_r(\tilde{\ell}).$$

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

$$\beta := \|\nabla^2 f_\ell(\tilde{\ell})^{-1}\|_2, \quad \textit{and} \quad \eta := \|\nabla^2 f_\ell(\tilde{\ell})^{-1} \nabla f_\ell(\tilde{\ell})\|.$$

Define $\gamma_R = \beta\gamma$ and $\alpha = \gamma_R \eta$. If $\alpha \leq \frac{1}{2}$ and $r \geq r_0 := \frac{1-\sqrt{1-2\alpha}}{\beta\gamma}$, 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

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

Moreover, ℓ_* is a ***lngm****.*

Proof. As in Theorem 5.6, the proof is a direct application of the Kantorovich theorem. □

Plugging (5.16), (5.17), (5.18), and

$$r = 10^{-5},\ \gamma = 6.12 \times 10^5,\ \bar{f}_L = 10^3,$$

into Theorem 5.10, we have

$$\begin{array}{l} 1/7.93 > \beta > 1/8.58, \\ \eta \leq \|\nabla^2 f_\ell(\tilde{\ell})^{-1}\|_2 \|\nabla f_\ell(\tilde{\ell})\| < 1/7.93 \times 9.23 \times 10^{-8}, \\ \gamma_R = \beta\gamma < 1/7.93 \times 6.12 \times 10^5, \\ \alpha = \gamma_R \eta < (1/7.93 \times 6.12 \times 10^5) \times (1/7.93 \times 9.23 \times 10^{-8}) < 1/2, \\ r_0 = \frac{1-\sqrt{1-2\alpha}}{\beta\gamma} < \frac{1-\sqrt{1-2\times 1/7.93^2 \times 6.12 \times 9.23 \times 10^{-3}}}{1/8.58 \times 6.12 \times 10^5} < r. \end{array}$$

Combining this with (5.20) and (5.21), we conclude from Theorem 5.10 that there exists a **lngm** in $B_r(\tilde{\ell})$.

In Figure 5.2 we plot: the points $\bar{p}_i \in \mathbb{R}^2,\ i = 1, \dots, n = 100$, of the global configuration $\bar{P} \in \mathbb{R}^{100\times 2}$, and the corresponding points $\tilde{p}_i \in \mathbb{R}^2,\ i = 1, \dots, n$, of the numerical configuration $\tilde{P} \in \mathbb{R}^{100\times 2}$ near a proven **lngm**. We note that $\bar{p}_i \approx \tilde{p}_i, \forall i = 1, \dots, n$, except for two indices i_0 and i_1; while $\tilde{p}_{i_0}$ and $\tilde{p}_{i_1}$ appear to be the reflections of $\bar{p}_{i_0}$ and $\bar{p}_{i_1}$. This interesting observation is similar to what happens in Theorem 5.2 as seen in Figure 5.1.

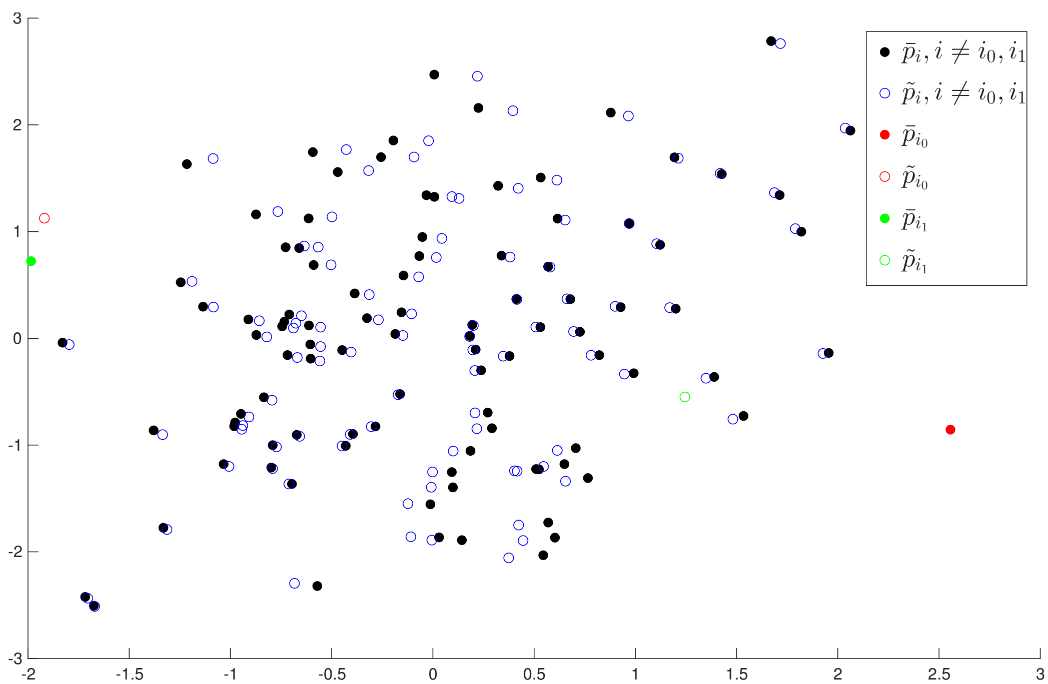

Figure 5.2: coordinates of global min and numerical near proven **lngm**, resp.: $\bar{P}, \tilde{P} \in \mathbb{R}^{100\times 2}$; in Theorem 5.7.

6. Conclusion

In this paper, we addressed the nonconvex optimization problem arising from the exact recovery of points from a given **EDM**. Our investigation led to significant advancements in understanding the conditions under which the smooth stress function (as known in the MDS literature) in **EDM** problems has a **lngm**. 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 $n \leq d + 1$. For $n > d + 1$, we not only identified **lngm** through numerical methods, but also used Kantorovich's theorem to provide rigorous analytical proofs confirming their existence. Moreover, based on the special patterns in those **lngm**s, we were able to build the analytical Theorem 5.1. Our methodology includes two reduction techniques based on translation and rotation invariance. These reductions are necessary for the application of Kantorovich's theorem.

The findings of this research resolve a longstanding open question regarding the existence of **lngm**s in the context of MDS. Additionally, our research highlights the importance of second-order methods for minimizing the smooth stress function.

For the future we plan to explore the possibility of further characterizing the properties of **lngm**s, e.g., with respect to embedding dimensions and ranks. Moreover, our goal is to study conditions for the existence of **lngm** when $\bar{D}$ has inexact and missing entries, as this is closer to the so called **EDM** Completion Problem, e.g., [7, 9]. This work is a first step in this direction as here we assume $\bar{D}$ is complete and a true **EDM**.

Acknowledgments

We thank the anonymous referees for the careful reading and meaningful suggestions that helped improve the paper. We also thank José Mario Martínez, Emeritus Professor of the University of Campinas, for fruitful discussions. The work of M.S. was supported by the international joint doctoral education fund of Beihang University. The work of D.G. was supported in part by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) Grant 305213/2021-0 and by FAPESC (Fundação de Amparo a Pesquisa e Inovação do Estado de Santa Catarina) Grant 2024TR002238. W.J. and H.W. research is supported by the Natural Sciences and Engineering Research Council of Canada. A.M. was partially supported by the ANR project EVARISTE (ANR-24-CE23-1621). C.L. is supported in part by FAPESP (Grant 2023/08706-1) and CNPq (Grants 305227/2022-0 and 404616/2024-0).

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.

[2] J. Alencar, C. Lavor and L. Liberti,*Realizing Euclidean distance matrices by sphere intersection*, Discrete Appl. Math., 256 (2019), pp. 5–10

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

[4] A. Alfakih, A. Khandani and H. Wolkowicz, *Solving Euclidean distance matrix completion problems via semidefinite programming*, Comput. Optim. Appl., 12 (1999), pp. 13–30. A tribute to Olvi Mangasarian.

[5] A. D. Baez-Sanchez 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.

[6] 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.

[7] 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.

[8] 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.

[9] L. Liberti, C. Lavor, N. Maculan, and A. Mucherino, *Euclidean distance geometry and applications*, SIAM Rev., 56 (2014), pp. 3–69.

[10] S. Malone and M. Trosset, *A study of the critical points of the SSTRESS criterion in metric multidimensional scaling*, Technical Report 00-06, Department of Computational & Applied Mathematics, Rice University, Houston, TX, USA (2000)

[11] S. Malone, P. Tarazaga, and M. Trosset, *Better initial configurations for metric multidimensional scaling*, Comput. Statist. Data Anal., 41 (2002), pp. 143–156.

[12] M. Trosset and R. Mathar , *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.

[13] A. Mucherino, C. Lavor, L. Liberti, and N. Maculan, eds., *Distance geometry*, Springer, New York, 2013. Theory, methods, and applications.

[14] R. Parhizkar, *Euclidean Distance Matrices Properties, Algorithms and Applications*, PhD thesis, ISC, Lausanne, 2013.

[15] J. B. Saxe, *Embeddability of weighted graphs in k-space is strongly np-hard*, in Proceedings of 17^{th} Allerton Conference in Communications, Control and Computing, 1979, pages 480–489.

[16] 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.

[17] A. Žilinskas and A. Podlipskytė, *On multimodality of the SSTRESS criterion for metric multidimensional scaling*, Informatica, 14 (2003), p. 121–130.