SIAM J. APPL. ALGEBRA GEOMETRY
Vol. 3, No. 1, pp. 31–43

Explicit Global Minimization of the Symmetrized Euclidean Distance by a Characterization of Real Matrices with Symmetric Square*

Patrizio Neff†, Andreas Fischle‡, and Lev Borisov§

Abstract. We determine the optimal orthogonal matrices $R \in \mathrm{O}(n)$ which minimize the symmetrized Euclidean distance $W\colon \mathrm{O}(n) \to \mathbb{R}$, $W(R;D) := ||\mathrm{sym}(RD - \mathbb{1})||^2$, where $\mathbb{1}$ denotes the identity matrix and $\mathrm{sym}(X) = \frac{1}{2}(X + X^T)$ is the symmetric part of X, for a given positive definite diagonal matrix $D = \mathrm{diag}(d_1, \dots, d_n)$ with distinct entries $d_1 > d_2 > \cdots > d_n > 0$. The number of critical points depends on D and can grow faster than exponential in n. In the process, we prove and use a novel result of independent interest: every real matrix whose square is symmetric can be expressed as a block-diagonal matrix composed of blocks of size at most two by a suitable orthonormal change of basis.

Key words. Grioli's theorem, polynomial optimization, orthogonal group, symmetric square, polar decomposition, relaxed-polar decomposition, nonsymmetric matrix square root, Euclidean distance degree

AMS subject classifications. 15A24, 22E30, 26C10

DOI. 10.1137/18M1179663

1. Introduction.

It is a well-known classical problem to characterize the orthogonal matrices $R \in \mathrm{O}(n)$ which minimize the Frobenius distance $||F - R||$ from a given matrix F, where $||X||^2 \coloneqq \sum_{1 \le i,j \le n} X_{ij}^2$. The first solution was published in 1940 by the Italian mathematician Giuseppe Grioli, who solved the Euclidean distance problem in three dimensions; see [14] or [21] for a modern account. Grioli's motivation originated from the interpretation of the Euclidean distance as a quadratic deformation energy density in the context of nonlinear elasticity theory. In the same spirit, but more recently, this connection between matrix distances and deformation energy densities in nonlinear elasticity motivated the study of certain Riemannian matrix distance problems; see, e.g., [23, 19] and also [15].

Due to the orthogonal invariance of the Frobenius norm, the Euclidean distance problem can be restated as

$$\min_{R \in \mathrm{O}(n)} ||F - R|| = \min_{R \in \mathrm{O}(n)} ||R^T F - \mathbb{1}|| = \min_{R \in \mathrm{O}(n)} ||RF - \mathbb{1}||\,,$$

where $\mathbb{1}$ denotes the identity matrix.

*Received by the editors April 10, 2018; accepted for publication (in revised form) November 19, 2018; published electronically February 7, 2019.

http://www.siam.org/journals/siaga/3-1/M117966.html

Funding: The second author was supported by German Research Foundation (DFG) grants SA2130/2-1 and NE902/2-1 (also SCHR570/6-1). The third author was supported by NSF grant DMS-1201466.

†Fakultät für Mathematik, Universität Duisburg-Essen, Essen, 45141 Germany (patrizio.neff@uni-due.de).

‡Institut für Numerische Mathematik, TU Dresden, Dresden, 01069 Germany (andreas.fischle@tu-dresden.de).

§Department of Mathematics, Rutgers University, Newark, NJ 07102 (borisov@math.rutgers.edu).

Remark 1 (optimality of the polar factor for the Euclidean distance). If we also assume that $F \in \mathrm{GL}^+(n)$, i.e., that F is invertible and contained in the identity component of $\mathrm{GL}(n)$, then the unique global minimizer is the polar factor $R_\mathrm{p} \in \mathrm{SO}(n)$ obtained from the right polar decomposition $F = R_\mathrm{p} U$. Here, $U = \sqrt{F^T F}$ denotes the unique symmetric positive definite square root of F; see, e.g., [21] and references therein.

The singular value decomposition of F as a product of orthogonal and positive definite diagonal matrices (carried out in detail in [10]) further reduces this problem to the minimization of the quadratic expression

$$||RD - \mathbb{1}||^2 \tag{1.1}$$

whose critical points are the 2^n diagonal matrices with entries ± 1; see, e.g., [21] or [2]. Recently, problems of this type have attracted renewed interest originating from the modern algebraic viewpoint of the so-called Euclidean distance degree of an algebraic variety; see, e.g., [6, 2, 7, 3].

In this note, we characterize the solutions to a significantly more sophisticated optimization problem that represented the major mathematical obstruction in [9, 10] to the explicit characterization of energy-minimizing (optimal) Cosserat microrotations in solid mechanics; cf. section 2.

Our main objective in this contribution is the solution to the following problem.

Problem 2. *Let $D := \mathrm{diag}(d_1, \dots, d_n)$ be a real diagonal matrix, and let the squared symmetrized Euclidean distance be given by*

$$W : \mathrm{O}(n) \times \mathrm{Diag}(n) \to \mathbb{R}, \quad W(R; D) := ||\mathrm{sym}(RD - \mathbb{1})||^2 , \tag{1.2}$$

where we use the notation $\mathrm{sym}(X) := \frac{1}{2}(X + X^T)$. Compute the set of critical points, the attained critical values, and the global minimizers of the objective function $W(R; D)$ for a fixed choice of D.

The problem can be reduced to the case of positive semidefinite D with $d_1 \geq d_2 \geq \cdots \geq d_n \geq 0$ by multiplication of R and D with a suitable diagonal matrix with entries ± 1.

This is a quadratic polynomial optimization problem over the field of real numbers with parameters, posed on a linear algebraic group. Toward the computation of critical points, solution methods for polynomial systems of equations are of interest and have recently received much attention; see, e.g., [24].

For simplicity only, we make the following assumption.

Assumption 3. *The real diagonal matrix $D = \mathrm{diag}(d_1, \dots, d_n)$ satisfies*

$$d_1 > d_2 > \cdots > d_n > 0 .$$

The techniques of our paper can be easily adapted to the slightly more general case when $d_1 \geq d_2 \geq \cdots \geq d_n > 0$. In this case, the critical points may no longer be isolated.

Our main result is the following.

Theorem 4. *Let $D = \mathrm{diag}(d_1, \dots, d_n)$ with strictly decreasing entries $d_1 > d_2 > \cdots > d_n > 0$. Denote by k the maximum positive integer for which $d_{2k-1} + d_{2k} > 2$, or set $k = 0$ if $d_1 + d_2 \leq 2$. Any global minimizer $R \in \mathrm{O}(n)$ of*

$$W(R\,;D) \coloneqq ||\mathrm{sym}(RD - \mathbb{1})||^2$$

is given by a block-diagonal matrix, built of k consecutive size-two blocks of the form

$$\begin{pmatrix} \cos\alpha_i & -\sin\alpha_i \\ \sin\alpha_i & \cos\alpha_i \end{pmatrix}$$

for $i = 1, \dots, k$, where $\alpha_i = \pm\arccos(\frac{2}{d_{2i-1}+d_{2i}})$, followed by $(n-2k)$ entries 1 on the main diagonal. The global minimum of $W(R\,;D)$ is given by

$$\min_{R\in\mathrm{O}(n)} W(R\,;D) = \frac{1}{2}\sum_{i=1}^{k}(d_{2i-1} - d_{2i})^2 + \sum_{i=2k+1}^{n}(d_i - 1)^2\,. \tag{1.3}$$

The solution in dimension $n=2$ has been previously presented in [9] and originally in [8]. Results for the case $n = 3$ have been contributed in [10] without proof and have already been used in the context of mechanics [20]. For a qualitative study and visualization of the geometric mechanism in the setting of an idealized nanoindentation, see [12].

Remark 5. We will see in Remark 20 that the number of critical points may grow faster than exponential in n. This is particularly interesting in view of recent results on Euclidean distance degrees presented in [2], showing again that additional complexity is introduced by the symmetrization considered here.

Furthermore, the number of global minimizers of $W(R\,;D)$ as a function of D is interesting due to a pitchfork bifurcation. This property is not shared by other related objective functions. In particular, the expression $W(R\,;D)$ appears naturally if we consider the Taylor expansion of the matrix logarithm at $RD = \mathbb{1}$ in the logarithmic energy

$$||\mathrm{sym}\log(RD)||^2 \;=\; 0 \;+\; ||\mathrm{sym}(RD - \mathbb{1})||^2 \;+\; \text{h.o.t. in } (RD)\,.$$

For this logarithmic energy, however, it has been shown that $R = \mathbb{1}$ is the unique global minimizer (see [23, 16, 21]), by means of a new "sum of squared logarithms"-inequality [4, 5].

This paper is structured as follows. After a brief description of the motivation and possible applications of our result in section 2, we prove a fundamental lemma in section 3 that characterizes real matrices whose square is symmetric. Specifically, we prove the existence of an orthonormal basis with respect to which the matrix attains a block-diagonal form composed of blocks of size at most two. This convenient block structure allows us to characterize the critical points of $W(R\,;D)$ in section 4 for arbitrary dimension n. Further exploiting this block structure, we obtain a sequence of decoupled one- and two-dimensional subproblems, which we solve in section 5. In section 6, we single out the global minimizers among the critical points by a comparison of the function values.

2. Motivation and applications. The minimization problem discussed here plays an important role in Cosserat theory, where a generalized continuum model for idealized solid materials with a particular rigid, purely rotational microstructure is considered. The Cosserat microstructure is represented by a microrotation $R(x) \in \mathrm{SO}(3)$ attached to each material point x in a deformable specimen $\Omega \subset \mathbb{R}^3$, leading to the stored energy

$$I(\varphi, R) \;\coloneqq\; \int_\Omega W(F(x), R(x)) \,\mathrm{dV}$$

for a deformation mapping $\varphi : \Omega \to \varphi(\Omega) \subset \mathbb{R}^3$, where $F \coloneqq D\varphi$ denotes the deformation gradient. The choice of the energy density $W\colon \mathrm{GL}^+(3) \times \mathrm{SO}(3) \to \mathbb{R}$ determines the material response of the quasi-static hyperelastic two-field Cosserat model.

Using a symmetry reduction for the case of planar simple shear, it was recently shown in [22] that the obtained energy-minimizing (optimal) Cosserat microstructure engenders microbands, a phenomenon that can be observed in the nanomechanics of crystalline materials. To the best of our knowledge, this phenomenon cannot be described by any of the established isotropic single-field continuum models.

Unfortunately, the numerical approximation of the solutions exhibiting nontrivial Cosserat microstructure has not yet been successful. In order to improve the numerical approach, a better understanding of the geometric effects produced by the rotation field R of the Cosserat model is therefore required. In particular, the distinguished special case

$$W_{\mathrm{shear}}(F, R) \;\coloneqq\; \mu||\mathrm{sym}(R^T F - \mathbb{1})||^2 + \mu_c||\mathrm{skew}(R^T F - \mathbb{1})||^2$$

of simple shear should be considered in more detail. Note that the choice $(\mu, \mu_c) = (1, 0)$ gives rise to the main objective (Problem 2) of the current contribution. The importance of the mathematical results obtained here are reinforced by the observation that the general case of arbitrary material parameters μ, μ_c can be reduced to this particular choice [9].

In [12], these optimal Cosserat micorotations in dimension three (cf. [10]) have already been applied to the setting of an idealized nanoindentation. A comparison with rotation patterns obtained from nanoindentation in a copper single crystal using the 3D-EBSD (electron backscatter diffraction) is shown in Figure 1.

For a more extensive discussion of optimal Cosserat rotations, we refer the interested reader to [9, 10, 18] and the survey article [11].

3. A block-diagonal representation of real matrices with a real symmetric square. In this section, we prove a linear algebra result that to our surprise does not seem to appear in the literature, or at least in any standard text.

Lemma 6. *Let X be a real $n \times n$ matrix such that $S = X^2$ is symmetric. Then there exists an orthogonal matrix Q such that $Q^{-1}XQ$ is block-diagonal with blocks of size at most two.*

Proof. The claim can be reformulated as saying that there exists an orthogonal decomposition of $\mathbb{R}^n$ into a direct sum of X-invariant subspaces of dimension at most two. We also remark that if we run induction on the dimension, it suffices to find an X-invariant subspace of dimension at most two whose orthogonal complement is also X-invariant. This is equivalent to finding a subspace V of dimension at most two which is both X- and X^T-invariant.

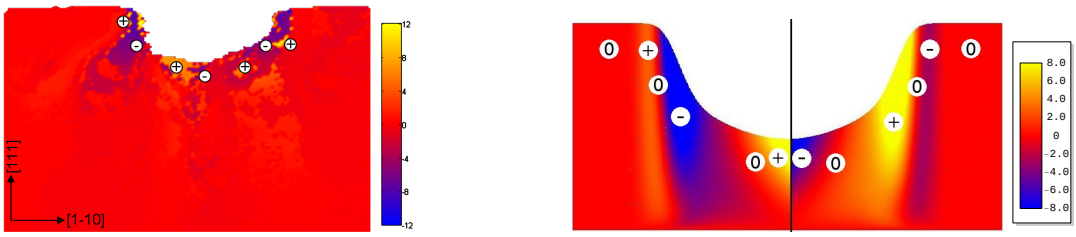

Figure 1. *Left: Rotation angles measured by 3D-EBSD analysis of a nanoindentation into a copper single crystal with a color map scaled to ±8°; note the cross-over zones separating counterrotations (courtesy of Zaafarani et al. [25]). Right: Nonlinear projection of the rotation angles for the optimal Cosserat microrotations* $\mathrm{rpolar}^{\pm}_{1,0}(F_{\mathrm{nano}}) \in \mathrm{SO}(3)$ *obtained in [12] for an idealized nanoindentation mapping. Two optimal branches of rotations are patched together at* $x = 0$.

Moreover, since X commutes with $S = X^2$, the eigenspaces of S are X-invariant; thus for $\widehat{R} \in \mathrm{O}(n)$ with $\widehat{R}^T S \widehat{R} = \mathrm{diag}(\lambda_1, \dots, \lambda_1, \lambda_2, \dots, \lambda_2, \dots, \lambda_m, \dots, \lambda_m)$, where λ_i are the distinct eigenvalues of S, the matrix $\widehat{X} = \widehat{R}^T S \widehat{R}$ is in block-form with symmetric blocks $\widehat{X}_i$ satisfying $\widehat{X}_i^2 = \lambda_i \mathbb{1}$. Therefore, the problem of finding such V reduces to the case $X^2 = \lambda \mathbb{1}$ for some $\lambda \in \mathbb{R}$.

Our main idea is that the self-adjoint operators given by the matrices XX^T and X^TX commute in view of

$$(XX^T)(X^TX) = X(X^T)^2X = X(\lambda\mathbb{1})X = \lambda^2\mathbb{1} = (X^TX)(XX^T).$$

Therefore, XX^T and X^TX are simultaneously diagonalizable and we can find a common eigenvector w of both. We will use this w to build the desired invariant subspace V. We have to distinguish several cases.

Case 1. $Xw \in \mathbb{R}w, X^Tw \in \mathbb{R}w$. In other words, w is an eigenvector of X and X^T. We select $V = \mathbb{R}w$.

Case 2. $Xw \in \mathbb{R}w, X^Tw \notin \mathbb{R}w$. In other words, w is an eigenvector of X but not of X^T. Consider the subspace $V = \mathrm{span}(w, X^Tw)$. Then the image of V under X satisfies

$$XV = \mathrm{span}(Xw, XX^Tw) \subseteq \mathrm{span}(w, w) \subseteq V,$$
$$X^TV = \mathrm{span}(X^Tw, (X^T)^2w) \subseteq \mathrm{span}(X^Tw, \lambda w) \subseteq V.$$

Case 3. $Xw \notin \mathbb{R}w$. In other words w is not an eigenvector of X. We consider the subspace $V = \mathrm{span}(w, Xw)$. The inclusion $XV \subseteq V$ follows from $X^2 = \lambda\mathbb{1}$. In addition, $X^TXw \in V$ by the choice of w. It remains to prove $X^Tw \in V$. We consider the following two subcases:

Case 3a. $\lambda \neq 0$. We use $XX^Tw = \beta w$ to conclude $X^Tw = \frac{1}{\lambda}XXX^Tw = \frac{\beta}{\lambda}Xw \in V$.

Case 3b. $\lambda = 0$. We have $X^TXw = \alpha w$, $XX^Tw = \beta w$ with $\alpha\beta = 0$. If $\beta = 0$, then

$$XX^Tw = 0 \Longrightarrow \langle w, XX^Tw\rangle = 0 \Longrightarrow \langle X^Tw, X^Tw\rangle = 0 \Longrightarrow X^Tw = 0 \in V$$

so V is again invariant. The $\alpha = 0$ case similarly leads to $Xw = 0$ in contradiction with the assumption that w is not an eigenvector of X.

This completes the construction of the invariant subspace V and the proof of the lemma. ■

Remark 7. The case $\lambda > 0$ of Lemma 6 can be deduced from the theory of principal angles (see, e.g., [13]) for the eigenspaces of X with eigenvalues $\sqrt{\lambda}$ and $-\sqrt{\lambda}$. We are not aware of a similar connection in the case $\lambda \leq 0$.

Remark 8. Given X with symmetric X^2, the decomposition of $\mathbb{R}^n$ into an orthogonal sum of invariant subspaces is not unique. In particular, a subspace of dimension two can sometimes be further decomposed into two one-dimensional subspaces.

Remark 9. Our description of real matrices which square to a real symmetric matrix resembles the well-known characterization of the group of real orthogonal matrices $\mathrm{O}(n)$: every orthogonal matrix is orthogonally conjugated to a block-diagonal matrix with blocks of size one and two.

We also provide a simple criterion for $X \in \mathbb{R}^{2\times 2}$ to have a symmetric square, i.e., $X^2 = (X^T)^2$, which will be used in section 5 and can easily be obtained by direct computation.

Lemma 10. *A real matrix $X \in \mathbb{R}^{2\times 2}$ has $S = X^2 \in \mathrm{Sym}(2)$ if and only if $X \in \mathrm{Sym}(2)$ or $\mathrm{tr}(X) = 0$.* ∎

4. Critical points: Reduction to low dimension. In this section, we investigate the critical points $R \in \mathrm{O}(n)$ of the function

$$W(R;D) \;=\; ||\mathrm{sym}(RD - \mathbb{1})||^2 \tag{4.1}$$

in Problem 2. Since W is differentiable in R, we can proceed by taking derivatives along curves in the matrix group $\mathrm{O}(n)$. The following derivation applies to any real diagonal matrix D, but we assume here that D is positive definite with distinct eigenvalues.

We first give a simple equivalent description of the set of critical points.

Lemma 11. *A point $R \in \mathrm{O}(n)$ is a critical point of W from (4.1) if and only if $X = RD - \mathbb{1}$ satisfies $X^2 \in \mathrm{Sym}(n)$.*

Proof. We identify the tangent space $\mathfrak{o}(n)$ to $\mathrm{O}(n)$ at $\mathbb{1}$ with the linear subspace of skew symmetric matrices $\mathrm{Skew}(n)$ and use the group structure of $\mathrm{O}(n)$ to describe the tangent space at R. In particular, $T_R\mathrm{O}(n) = \mathfrak{o}(n)R$; thus the criticality condition is equivalent to the statement that for any $A \in \mathrm{Skew}(n)$,

$$0 = D_R W(R;D).(AR) = \frac{d}{dt}\left(W(e^{tA}R\,;D)\right)\big|_{t=0}, \tag{4.2}$$

where we extend the definition of W to all matrices of size n. Up to first order in t, the function $W(e^{tA}R\,;D)$ simplifies to

$$\begin{aligned} &||\mathrm{sym}((1+tA)RD - \mathbb{1})||^2 = ||\mathrm{sym}(RD-\mathbb{1}) + t\,\mathrm{sym}(ARD)||^2 \\ &= ||\mathrm{sym}(RD-\mathbb{1})||^2 + 2t\,\langle \mathrm{sym}(RD-\mathbb{1}), \mathrm{sym}(ARD)\rangle + t^2\,||\mathrm{sym}(ARD)||^2. \end{aligned}$$

Hence, a point R is a critical point for the function W if and only if (cf. [17])

$$\forall A \in \mathrm{Skew}(n): \quad \mathrm{sym}(RD-\mathbb{1}) \perp \mathrm{sym}(ARD).$$

Since $\mathrm{Sym}(n) \perp \mathrm{Skew}(n)$, we may add $\mathrm{skew}(ARD)$ on the right-hand side, which gives us the equivalent condition

$$\forall A \in \mathrm{Skew}(n): \quad \mathrm{sym}(RD - \mathbb{1}) \perp ARD.$$

From the definition of the Frobenius inner product, we obtain the condition

$$\begin{aligned} 0 &= \langle \mathrm{sym}(RD - \mathbb{1}), ARD \rangle = \mathrm{tr}(\mathrm{sym}(RD - \mathbb{1})^T ARD) \\ &= \mathrm{tr}(RD\,\mathrm{sym}(RD - \mathbb{1})A) = \langle \mathrm{sym}(RD - \mathbb{1})(RD)^T, A \rangle \ , \end{aligned}$$

which must hold for all $A \in \mathrm{Skew}(n)$ and is equivalent to

$$\mathrm{sym}(RD - \mathbb{1})\, DR^T \in \mathrm{Sym}(n).$$

It can be further rewritten as

$$\begin{aligned} \mathrm{Sym}(n) \ni (RD + DR^T - 2\,\mathbb{1})DR^T &= RD^2R^T + (DR^T)^2 - 2DR^T \\ &= ((RD - \mathbb{1})^2)^T + \underbrace{RD^2R^T - \mathbb{1}}_{\in \mathrm{Sym}(n)} \ . \end{aligned}$$

■

Remark 12. Note that $R = \mathbb{1}$ is always a critical point of $W(R; D) := ||\mathrm{sym}(RD - \mathbb{1})||^2$. However, in general, it will not be a global minimizer.

Our next step is to apply Lemma 6 to the special case $X = RD - \mathbb{1}$. As we shall see, this implies restrictive conditions on $R \in \mathrm{O}(n)$.

We recall that D is positive definite, i.e., all $d_i > 0$, and prove the following lemma which is key to our discussion.

Lemma 13 (simultaneous invariance of R and D). *Suppose that D is positive definite. Let V be a subspace invariant under $X = RD - \mathbb{1}$ such that $V^\perp$ is also invariant under X. Then both V and $V^\perp$ are invariant under D and R.*

Proof. By assumption, the subspace V is invariant under both $RD = X + \mathbb{1}$ and $(RD)^T = DR^T = X^T + \mathbb{1}$. Therefore,

$$D^2V = (DR^T)(RD)V \subseteq (DR^T)V \subseteq V.$$

It is easy to see that, since D and D^2 have the same one-dimensional eigenspaces due to Assumption 3, V is invariant under D if and only if V is invariant under D^2. Since D is also invertible by assumption, we find that $D^2V \subseteq V$ implies $DV = V$ and thus

$$RDV \subseteq V \implies RV \subseteq V.$$

Since R is invertible, we get $RV = V$. The same argument works for $V^\perp$. ■

By Lemma 6, there exists a sequence of pairwise orthogonal vector spaces V_i, $i = 1, \ldots, r$, with $1 \leq \dim V_i \leq 2$ which decompose

$$\mathbb{R}^n = V_1 \oplus_\perp V_2 \oplus_\perp \cdots \oplus_\perp V_r. \tag{4.3}$$

These correspond to a block-diagonal representation of $X = RD - \mathbb{1}$. By Lemma 13, both R and D are also block-diagonal with respect to this choice of basis, and the latter condition in particular imposes severe restrictions on the V_i.

Remark 14 (implications of D-invariance). With Assumption 3, the D-invariance of the subspaces V_i shown in Lemma 13 implies a strong restriction: the V_i are necessarily coordinate subspaces in the standard basis of $\mathbb{R}^n$. Thus, we can index these data by partitions of the index set $\{1,\dots,n\}$ into disjoint subsets of size one or two.

This decomposition structure allows us to reduce finding critical points of Problem 2 to a finite list of decoupled one- and two-dimensional subproblems. This will be the content of the next section.

5. Analysis of the decoupled subproblems. Let $I \subseteq \{1,\dots,n\}$ be a one-element subset $\{i\}$ or a two-element subset $\{i,j\}$ and let D_I be the associated restriction of D given by

$$\begin{cases} D_I := \begin{pmatrix} d_i \end{pmatrix} & \text{if} \quad I = \{i\}, \\ D_I := \begin{pmatrix} d_i & 0 \\ 0 & d_j \end{pmatrix} & \text{if} \quad I = \{i,j\}\,. \end{cases}$$

In this section, we determine the critical points of the function

$$W(R_I\,;D_I) := ||\mathrm{sym}(R_I D_I - \mathbb{1})||^2$$

for $R_I \in \mathrm{O}(|I|)$ and compute the corresponding critical values. This corresponds to the solution of the decoupled lower-dimensional subproblems as described in the previous section.

Theorem 15 (critical points: size-one blocks). *For $I=\{i\}$ we have the submatrix $D_I=(d_i)$ and $R_I=\pm\mathbb{1}$. The realized critical function values are*

$$W(+\mathbb{1}\,;D_I) = (d_i-1)^2 \quad \textit{and} \quad W(-\mathbb{1}\,;D_I) = (d_i+1)^2\,. \tag{5.1}$$

Proof. $\mathrm{O}(1)=\{\pm\mathbb{1}\}$. ∎

For the case $|I|=2$, we consider the two separate cases $\det R_I = 1$ and $\det R_I = -1$.

Theorem 16 (critical points: size-two blocks with positive determinant). *The critical points R_I with $\det R_I = 1$ are described as follows. For any values d_i and d_j, the matrices $R_I=\pm\mathbb{1}$ are critical points with the critical values $(d_i-1)^2+(d_j-1)^2$ and $(d_i+1)^2+(d_j+1)^2$, respectively. In addition, if $d_i+d_j>2$, then there are two nondiagonal critical points:*

$$R_I = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}, \quad \textit{with} \quad \cos\alpha = \frac{2}{d_i+d_j}, \tag{5.2}$$

both of which attain the critical value

$$W(R_I\,;D_I) = \frac{1}{2}(d_i-d_j)^2\,. \tag{5.3}$$

Proof. By Lemma 11 R_I is a critical point if and only if $(R_I D_I-\mathbb{1})^2$ is symmetric. We may thus apply Lemma 10 to infer $R_I D_I-\mathbb{1}\in\mathrm{Sym}(2)$ or $\mathrm{tr}(R_I D_I-\mathbb{1})=0$. Using the explicit representation

$$R_I = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}\,,$$

we find that the symmetry condition $R_I D_I - \mathbb{1} \in \mathrm{Sym}(2)$ is equivalent to $(d_i + d_j)\sin\alpha = 0$, which has two solutions $R_I = \pm\mathbb{1}$ with the realized function values as claimed.

Otherwise, the trace condition $\mathrm{tr}(R_I D_I - \mathbb{1}) = 0$ is equivalent to $(d_i + d_j)\cos\alpha = 2$, which can be solved for real α if and only if $d_i + d_j \geq 2$. It gives rise to two nondiagonal solutions if and only if $d_i + d_j > 2$. We use $(d_i + d_j)\cos\alpha = 2$ to get $||\mathrm{sym}(R_I D_I - \mathbb{1})||^2 = \frac{1}{2}(d_i - d_j)^2$ by a routine calculation that we leave to the reader. ∎

Theorem 17 (critical points: size-two blocks with negative determinant). *The critical points R_I with $\det R_I = -1$ are described as follows. For any values d_i and d_j the diagonal matrices $R_I = \pm\mathrm{diag}(1, -1)$ are critical points with the critical values $(d_i - 1)^2 + (d_j + 1)^2$ and $(d_i + 1)^2 + (d_j - 1)^2$, respectively. In addition, for $|d_i - d_j| > 2$, there are two nondiagonal critical points*

$$R_I = \begin{pmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{pmatrix}, \quad \textit{with} \quad \cos\alpha = \frac{2}{|d_i - d_j|}\,, \tag{5.4}$$

both of which attain the critical value

$$W(R_I\,; D_I) = \frac{1}{2}(d_i + d_j)^2\,. \tag{5.5}$$

Proof. The calculation is similar to the case of positive determinant and is left to the reader. ∎

Remark 18. The diagonal critical points $R_I = \pm\mathbb{1}$ and $R_I = \pm\mathrm{diag}(1, -1)$ reduce to size-one blocks (or index subsets $|I| = 1$) in the block decomposition (4.3).

6. Global minimization. Combining the results of the two preceding sections, we can now describe the critical values of $W(R\,; D)$, which are attained at the critical points. We label the critical points by partitions of the index set $\{1, \ldots, n\}$ containing only subsets I with one or two elements. In the last section, we have seen that the subsets I and a choice of sign for $\det R_I$ uniquely determine a critical value; cf. Remark 18. These critical values are characterized in the following theorem.

Theorem 19 (characterization of critical points and values). *Let $D = \mathrm{diag}(d_1, \ldots, d_n)$ with $d_1 > d_2 > \cdots > d_n > 0$. Then the critical points $R \in \mathrm{O}(n)$ of the function*

$$W(R\,; D) := ||\mathrm{sym}(RD - \mathbb{1})||^2$$

can be classified according to partitions of the index set $\{1, \ldots, n\}$ into subsets of size one or two and choices of signs for the determinant $\det R_I$ for each subset I. The subsets of size two $I = \{i, j\}$ satisfy

$$\begin{cases} d_i + d_j\ > 2, & \det R_I = +1\,, \quad \textit{and} \\ |d_i - d_j| > 2, & \det R_I = -1\,. \end{cases}$$

The corresponding critical values of $W(R\,; D)$ are given by

$$\sum_{\substack{I=\{i\} \\ \det R_I = 1}} (d_i - 1)^2 + \sum_{\substack{I=\{i\} \\ \det R_I = -1}} (d_i + 1)^2 + \sum_{\substack{I=\{i,j\} \\ \det R_I = 1}} \frac{1}{2}(d_i - d_j)^2 + \sum_{\substack{I=\{i,j\} \\ \det R_I = -1}} \frac{1}{2}(d_i + d_j)^2.$$

Proof. A suitable partition of the index set $\{1,\dots,n\}$ can be constructed as detailed in section 4. The contributions of the subsets I of size one and two are given by the theorems of section 5. It suffices to consider the nondiagonal critical points for the subproblems of size two, because the diagonal cases can be accounted for by splitting the subset $I=\{i,j\}$ into two subsets $\{i\}$ and $\{j\}$ of size one; see Remark 18. ∎

Remark 20. If $d_i - d_{i+1} > 2$ for all i, then all possible splittings of $\{1,2,\dots,n\}$ into subsets of size one and two are possible. In each case the number of critical points is 2^n, so the total number of critical points is $2^n c_n$, where c_n is the well-studied sequence A000085 of self-inverse permutations on n letters [1]. Already in the case of O(2), if we have $d_1 - d_2 > 2$, then the number of critical points is eight, as opposed to the four critical points for the distance problem of (1.1). In particular [1], it grows faster than exponential in n.

For what follows, it will be useful to rewrite $W(R;D)$ in a slightly different form in order to distill the contributions of the size-two blocks in the partition.

Corollary 21. *If* $\det R_I = +1$ *for all* I*, then the following holds:*

$$W(R;D) = \sum_{i=1}^{n}(d_i-1)^2 - \frac{1}{2}\sum_{I=\{i,j\}}(d_i+d_j-2)^2.$$

Proof. If $d_i + d_j > 2$, then the difference between the critical values of $W(R;D)$ corresponding to the choice of a size-two subset $I=\{i,j\}$ as compared to the choice of two size-one subsets $\{i\}$, $\{j\}$ is given by

$$\frac{1}{2}(d_i-d_j)^2 - (d_i-1)^2 - (d_j-1)^2 = -\frac{1}{2}(d_i+d_j-2)^2\,. \qquad ∎$$

We are now ready to prove our main result, Theorem 4. Recall that we assume $d_1 > \cdots d_n > 0$ and that k is the greatest positive integer such that $d_{2k-1} + d_{2k} > 2$ or $k=0$ if $d_1 + d_2 \le 2$.

Proof of Theorem 4. In view of the above calculation we need to argue that the minimum value of $W(R;D)$ on the finite set of critical points R from Theorem 19 is realized for the partition of

$$\{1,2,\dots,n\} = \{1,2\} \sqcup \cdots \sqcup \{2k-1,2k\} \sqcup \{2k+1\} \sqcup \cdots \sqcup \{n\}$$

and all determinants $\det R_I$ picked to be 1.

Suppose that a critical point R is a global minimizer of $W(R;D)$.

We observe that $|d_i - d_j| > 2$ implies that $d_i + d_j > 2$. Therefore, it is always possible to replace negative determinant choices by positive ones. In the process the value of $W(R;D)$ is reduced, since $(d_i-1)^2 < (d_i+1)^2$ and $(d_i-d_j)^2 < (d_i+d_j)^2$. Therefore, R only contains blocks R_I with $\det R_I = 1$.

We now argue that the blocks of size two do not intersect in the sense that the partition of R contains size-two subsets $I=\{i_1,i_3\}$, $J=\{i_2,i_4\}$, with $i_1 < i_2 < i_3 < i_4$. It suffices to consider the case $i_1=1, i_2=2, i_3=3$, and $i_4=4$ with the general case being completely analogous.

There are two cases to consider.

Case 1. $d_3 + d_4 > 2$. In this case, we can consider another critical point $\mathring{R}$ corresponding to the partition $\{1,2\} \sqcup \{3,4\}$ instead of $\{1,3\} \sqcup \{2,4\}$. We have

$$\begin{aligned} W(R;D) - W(\mathring{R};D) &= \frac{1}{2}(d_1 - d_3)^2 + \frac{1}{2}(d_2 - d_4)^2 - \frac{1}{2}(d_1 - d_2)^2 - \frac{1}{2}(d_3 - d_4)^2 \\ &= d_1 d_2 + d_3 d_4 - d_1 d_3 - d_2 d_4 = (d_1 - d_4)(d_2 - d_3) > 0, \end{aligned}$$

which contradicts the assumed minimality of R.

Case 2. $d_3 + d_4 \leq 2$. In this case, the subset $\{3,4\}$ cannot appear in the partition and we consider the matrix $\mathring{R}$ with partition $\{1,2\} \sqcup \{3\} \sqcup \{4\}$ instead of $\{1,3\} \sqcup \{2,4\}$. The difference of the corresponding function values is found by comparing the previous calculation and Corollary 21:

$$W(R;D) - W(\mathring{R};D) = (d_1 - d_4)(d_2 - d_3) - \frac{1}{2}(d_3 + d_4 - 2)^2.$$

The latter is an increasing function of d_2, and we know that $d_2 + d_4 > 2$; therefore,

$$\begin{aligned} W(R;D) - W(\mathring{R};D) &> (d_1 - d_4)(2 - d_4 - d_3) - \frac{1}{2}(d_3 + d_4 - 2)^2 \\ &= \frac{1}{2}(2 - d_3 - d_4)((d_1 + d_3 - 2) + (d_1 - d_4)) \geq 0 \end{aligned}$$

by assumptions on d_i. This is again a contradiction.

We also observe that the partition of R does not have blocks that contain each other in the sense of having size-two subsets $I = \{i_1, i_4\}$, $J = \{i_2, i_3\}$, with $i_1 < i_2 < i_3 < i_4$. It again suffices to consider $i_1 = 1, i_2 = 2, i_3 = 3$, and $i_4 = 4$. We proceed similarly to the case $I = \{1,3\}$, $J = \{2,4\}$ above; as we never used $d_3 > d_4$, we can simply switch d_3 and d_4 in all formulas.

Since different size-two blocks do not overlap and are not contained in one another, we observe that a minimizing partition of R corresponds to size-two blocks built from consecutive indices. Indeed, if we have $i < j < k$ and blocks $\{i,k\}$ and $\{j\}$, we can replace them with $\{i,j\}$ and $\{k\}$ to decrease the value of W by Corollary 21. Similarly, it is impossible that a size-one block precedes a size-two block. To see this, assume that $i < j < k$ and consider blocks $\{i\}$ and $\{j,k\}$. Since we can further decrease W by replacing them with $\{i,j\}$ and $\{k\}$ this contradicts minimality. Therefore, the partition of R must start with size-two blocks and end with size-one blocks (which have to be 1). It remains to observe that Corollary 21 implies that all possible size-two blocks must be realized. ∎

Remark 22 (optimality of $\mathbb{1}$). Our results imply that the identity matrix $\mathbb{1} \in \mathrm{O}(n)$ is globally optimal for $W(R;D)$ with $D > 0$ if and only if there exists no 2×2-block with a positive choice of $\det R_I$, i.e.,

$$\max_{1 \leq i \neq j \leq n} (d_i + d_j) \;\leq\; 2.$$

Finally, for the case of dimension $n = 3$, we recover the result of [10] which was originally obtained and verified by a completely different numerical approach that did not allow for a rigorous proof.

Corollary 23. *Let $d_1 > d_2 > d_3 > 0$. If $d_1 + d_2 \leq 2$, then the global minimum of*

$$W(R;D) = ||\mathrm{sym}(RD - \mathbb{1})||^2$$

occurs at $R = \mathbb{1}$ and is given by

$$W(R;D) = (d_1 - 1)^2 + (d_2 - 1)^2 + (d_3 - 1)^2.$$

If $d_1 + d_2 > 2$, then the global minimum is realized by either of two critical points of the form

$$R = \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \textit{with} \quad (d_1 + d_2)\cos\alpha = 2\ .$$

In this case the global minimum is

$$W(R;D) = (d_1 - 1)^2 + (d_2 - 1)^2 + (d_3 - 1)^2 - \frac{1}{2}(d_1 + d_2 - 2)^2 = \frac{1}{2}(d_1 - d_2)^2 + (d_3 - 1)^2.$$

Proof. If $d_1 + d_2 > 2$, then $k = 1$. Otherwise, $k = 0$. ∎

Further reducing this result to dimension $n=2$, we can also recover the results of [9].

Acknowledgments. We thank Bernd Sturmfels and the anonymous referee for radical suggestions on exposition which improved the paper and Robert Martin for further helpful suggestions.

REFERENCES

[1] OEIS Foundation Inc., *The On-Line Encyclopedia of Integer Sequences*, 2017, http://oeis.org/A000085.
[2] J. A. Baaijens and J. Draisma, *Euclidean distance degrees of real algebraic groups*, Linear Algebra Appl., 467 (2015), pp. 174–187.
[3] A. Bik and J. Draisma, *A note on ED degrees of stable subvarieties in polar representations*, Israel J. Math., 228 (2018), pp. 353–377.
[4] M. Bîrsan, P. Neff, and J. Lankeit, *Sum of squared logarithms—an inequality relating positive definite matrices and their matrix logarithm*, J. Inequal. Appl., 2013 (2013), 168.
[5] L. Borisov, P. Neff, S. Sra, and C. Thiel, *The sum of squared logarithms inequality in arbitrary dimensions*, Linear Algebra Appl., 528 (2017), pp. 124–146.
[6] J. Draisma, E. Horobeţ, G. Ottaviani, B. Sturmfels, and R. R. Thomas, *The Euclidean distance degree of an algebraic variety*, Found. Comput. Math., 16 (2016), pp. 99–149.
[7] D. Drusvyatskiy, H.-L. Lee, G. Ottaviani, and R. R. Thomas, *The Euclidean distance degree of orthogonally invariant matrix varieties*, Israel J. Math., 221 (2017), pp. 291–316.
[8] A. Fischle, *The Planar Cosserat Model: Minimization of the Shear Energy on* SO(2) *and Relations to Geometric Function Theory*, Diploma Thesis, Technische Universität Darmstadt, Darmstadt, Germany, 2007.
[9] A. Fischle and P. Neff, *The geometrically nonlinear Cosserat micropolar shear–stretch energy.* Part I: *A general parameter reduction formula and energy-minimizing microrotations in 2D*, ZAMM Z. Angew. Math. Mech., 97 (2017), pp. 828–842.
[10] A. Fischle and P. Neff, *The geometrically nonlinear Cosserat micropolar shear–stretch energy. Part* II: *Non-classical energy-minimizing microrotations in 3D and their computational validation*, ZAMM Z. Angew. Math. Mech., 97 (2017), pp. 843–871.

[11] A. Fischle and P. Neff, *Grioli's theorem with weights and the relaxed-polar mechanism of optimal Cosserat rotations*, Atti Accad. Naz. Lincei Rend. Lincei Mat. Appl., 28 (2017), pp. 573–600.

[12] A. Fischle, P. Neff, and D. Raabe, *The relaxed-polar mechanism of locally optimal Cosserat rotations for an idealized nanoindentation and comparison with 3D-EBSD experiments*, Z. Angew. Math. Phys., 68 (2017), 90.

[13] A. Galántai, *Projectors and Projection Methods*, Adv. Math. (Dordr.) 6, Kluwer, Boston, 2004.

[14] G. Grioli, *Una proprieta di minimo nella cinematica delle deformazioni finite*, Boll. Un. Mat. Ital. (2), 2 (1940), pp. 452–455.

[15] R. Kupferman and A. Shachar, *On strain measures and the geodesic distance to $SO(n)$ in the general linear group*, J. Geom. Mech., 8 (2016), pp. 437–460.

[16] J. Lankeit, P. Neff, and Y. Nakatsukasa, *The minimization of matrix logarithms: On a fundamental property of the unitary polar factor*, Linear Algebra Appl., 449 (2014), pp. 28–42.

[17] P. Neff, *Finite multiplicative plasticity for small elastic strains with linear balance equations and grain boundary relaxation*, Contin. Mech. Thermodyn., 15 (2003), pp. 161–195.

[18] P. Neff, M. Bîrsan, and F. Osterbrink, *Existence theorem for geometrically nonlinear Cosserat micropolar model under uniform convexity requirements*, J. Elasticity, 121 (2015), pp. 119–141.

[19] P. Neff, B. Eidel, and R. J. Martin, *Geometry of logarithmic strain measures in solid mechanics*, Arch. Ration. Mech. Anal., 222 (2016), pp. 507–572.

[20] P. Neff, A. Fischle, and I. Münch, *Symmetric Cauchy-stresses do not imply symmetric Biot-strains in weak formulations of isotropic hyperelasticity with rotational degrees of freedom*, Acta Mech., 197 (2008), pp. 19–30.

[21] P. Neff, J. Lankeit, and A. Madeo, *On Grioli's minimum property and its relation to Cauchy's polar decomposition*, Internat. J. Engrg. Sci., 80 (2014), pp. 209–217.

[22] P. Neff and I. Münch, *Simple shear in nonlinear Cosserat elasticity: Bifurcation and induced microstructure*, Contin. Mech. Thermodyn., 21 (2009), pp. 195–221.

[23] P. Neff, Y. Nakatsukasa, and A. Fischle, *A logarithmic minimization property of the unitary polar factor in the spectral and Frobenius norms*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 1132–1154, https://doi.org/10.1137/130909949.

[24] B. Sturmfels, *Solving Systems of Polynomial Equations*, CBMS Reg. Conf. Ser. Math. 97, AMS, Providence, RI, 2002.

[25] N. Zaafarani, N. Raabe, R. N. Singh, F. Roters, and S. Zaefferer, *Three-dimensional investigation of the texture and microstructure below a nanoindent in a Cu single crystal using 3D EBSD and crystal plasticity finite element simulations*, Acta Mater., 54 (2006), pp. 1863–1876.