ruhr.paD
UA Ruhr Zentrum für
partielle Differentialgleichungen

Optimality of the relaxed polar factors by a characterization of the set of real square roots of real symmetric matrices

L. Borisov, A. Fischle and P. Neff

Preprint 2017-18

Optimality of the relaxed polar factors by a characterization of the set of real square roots of real symmetric matrices

Lev Borisov[1], Andreas Fischle[2], and Patrizio Neff[3]

February 24, 2017

Abstract

We consider the problem to determine the optimal rotations $R \in \mathrm{SO}(n)$ which minimize

$$W : \mathrm{SO}(n) \to \mathbb{R}_0^+, \quad W(R;D) \coloneqq \|\mathrm{sym}(RD - \mathbb{1})\|^2$$

for a given diagonal matrix $D \coloneqq \mathrm{diag}(d_1, \dots, d_n) \in \mathbb{R}^{n\times n}$ with positive entries $d_i > 0$. The objective function W is the reduced form of the Cosserat shear-stretch energy, which, in its general form, is a contribution in any geometrically nonlinear, isotropic, and quadratic Cosserat micropolar (extended) continuum model. We characterize the critical points of the energy $W(R;D)$, determine the global minimizers and compute the global minimum. This proves the correctness of previously obtained formulae for the optimal Cosserat rotations in dimensions two and three. The key to the proof is the result that every real matrix whose square is symmetric can be written in some orthonormal basis as a block-diagonal matrix with blocks of size at most two. This statement does not seem to appear in the literature.

Keywords: Cosserat theory, micropolar media, Grioli's theorem, rotations, special orthogonal group, (non-symmetric) matrix square root, symmetric square, polar decomposition, relaxed-polar decomposition.

AMS 2010 subject classification: 15A24, 22E30, 74A30, 74A35, 74B20, 74G05, 74G65, 74N15.

Contents

[1]Lev Borisov, Department of Mathematics, Rutgers University, 240 Hill Center, Newark, NJ 07102, United States, email: borisov@math.rutgers.edu

[2]Corresponding author: Andreas Fischle, Institut für Numerische Mathematik, TU Dresden, Zellescher Weg 12-14, 01069 Dresden, Germany, email: andreas.fischle@tu-dresden.de

[3]Patrizio Neff, Head of Lehrstuhl für Nichtlineare Analysis und Modellierung, Fakultät für Mathematik, Universität Duisburg-Essen, Thea-Leymann Str. 9, 45127 Essen, Germany, email: patrizio.neff@uni-due.de

1 Introduction

1.1 The problem

In this contribution, we characterize the solutions to the optimality problem stated as

Problem 1.1. *Let $D \coloneqq \operatorname{diag}(d_1, \dots, d_n) > 0$ be a positive definite diagonal matrix and let*

$$W : \mathrm{SO}(n) \times \mathrm{Diag}(n) \to \mathbb{R}^+_0, \quad W(R\,;D) \coloneqq \|\operatorname{sym}(RD - \mathbb{1})\|^2 \,. \tag{1.1}$$

Compute the relaxed polar factors, i.e., the set of energy-minimizing rotations

$$\operatorname{rpolar}(D) \coloneqq \operatorname*{arg\,min}_{R \in \mathrm{SO}(n)} W(R\,;D) = \operatorname*{arg\,min}_{R \in \mathrm{SO}(n)} \|\operatorname{sym}(RD - \mathbb{1})\|^2 \subseteq \mathrm{SO}(n) \,. \tag{1.2}$$

We use the notation $\operatorname{sym}(X) \coloneqq \frac{1}{2}(X + X^T)$, $\operatorname{skew}(X) \coloneqq \frac{1}{2}(X - X^T)$, $\operatorname{dev}(X) \coloneqq X - \frac{1}{n}\operatorname{tr}[X] \cdot \mathbb{1}$, $\langle X, Y \rangle \coloneqq \operatorname{tr}\left[X^T Y\right]$ and we denote the induced Frobenius matrix norm by $\|X\|^2 \coloneqq \langle X, X \rangle = \sum_{1 \le i,j \le n} X_{ij}^2$. We call a rotation $R \in \mathrm{SO}(n)$ optimal for given $D \in \mathrm{Diag}(n)$ if it is a global minimizer for the energy $W(R\,;D)$ defined in (1.1). Furthermore, we denote the spaces of symmetric and skew-symmetric matrices by $\mathrm{Sym}(n) \subset \mathbb{R}^{n \times n}$ and $\mathrm{Skew}(n) \subset \mathbb{R}^{n \times n}$, respectively.

This work is concerned with the derivation of formulae that explicitly characterize the relaxed polar factors $\operatorname{rpolar}(D)$ in arbitrary dimension. It is beyond the scope of the current paper to develop efficient and stable numerical approximations of the relaxed polar factors $\operatorname{rpolar}(D)$. However, this will be a logical next step.

1.2 Previous results

We consider the quadratic Cosserat shear-stretch energy $W_{\mu,\mu_c} : \mathrm{SO}(n) \times \mathrm{GL}^+(n) \to \mathbb{R}^+_0$

$$W_{\mu,\mu_c}(\overline{R}\,;F) \coloneqq \mu \left\|\operatorname{sym}(\overline{R}^T F - \mathbb{1})\right\|^2 + \mu_c \left\|\operatorname{skew}(\overline{R}^T F - \mathbb{1})\right\|^2 \tag{1.3}$$

with weights (material parameters) $\mu > 0$ and $\mu_c \ge 0$.[1] Let us introduce the general weighted form of the relaxed polar factors

$$\operatorname{rpolar}_{\mu,\mu_c}(F) \coloneqq \operatorname*{arg\,min}_{\overline{R} \in \mathrm{SO}(n)} \left(\mu \left\|\operatorname{sym}(\overline{R}^T F - \mathbb{1})\right\|^2 + \mu_c \left\|\operatorname{skew}(\overline{R}^T F - \mathbb{1})\right\|^2 \right) . \tag{1.4}$$

The unique global minimizer $\overline{R} \in \mathrm{SO}(n)$ in the *classical parameter range* $\mu_c \ge \mu > 0$ is the orthogonal factor $\mathrm{R_p}(F)$ in the right polar decomposition of $F \in \mathrm{GL}^+(n)$, see [35]. The *non-classical parameter range* $\mu > \mu_c \ge 0$ of parameters can, surprisingly, be reduced to a single *non-classical limit case* $(\mu, \mu_c) = (1, 0)$, see [7]. The choice of weights (material parameters) $\mu > 0$ and $\mu_c \ge 0$ is crucial since they characterize a pitchfork bifurcation between a classical branch of minimizers, i.e, where $\mathrm{R_p}(F)$ is optimal, and an interesting new type of non-classical minimizers.

Due to the parameter reduction, it suffices to consider the Cosserat shear-stretch energy in the limit case $(\mu, \mu_c) = (1, 0)$ given by

$$W_{1,0}(\overline{R}\,;F) \coloneqq \left\|\operatorname{sym}(\overline{R}^T F - \mathbb{1})\right\|^2 \,. \tag{1.5}$$

A Cosserat strain energy $W(F, \overline{R})$ is called isotropic, if it satisfies the invariance $W(Q_1 F Q_2, Q_1 \overline{R} Q_2) = W(F, \overline{R})$ for all $Q_1, Q_2 \in \mathrm{SO}(n)$. Exploiting the isotropy of the Cosserat shear-stretch energy, it can be equivalently expressed in terms of a rotation $R \in \mathrm{SO}(n)$ acting relative to the polar factor $\mathrm{R_p}(F)$, see the introduction to [8] for details. After this second reduction step, we obtain the equivalent energy

$$W_{1,0}(R,;D) \coloneqq \|\operatorname{sym}(RD - \mathbb{1})\|^2 \,, \tag{1.6}$$

[1] A more in-depth presentation of the interpretation in mechanics is provided in Section 1.4.

where $D = \operatorname{diag}(d_1, \dots, d_n) > 0$ is a positive definite matrix. Its diagonal entries are given by the singular values $d_i = \nu_i > 0$ of $F \in \mathrm{GL}^+(n)$.

Hence, on the basis of the previous works [7] and [8], it suffices to solve Problem 1.1 in order to characterize the global minimizers for the quadratic Cosserat shear-stretch energy in the entire non-classical parameter range. For a short overview of the previous results, see [9].

Explicit formulae for the critical points and the global minimizers $\operatorname{rpolar}^{\pm}_{\mu,\mu_c}(F)$ of $W_{\mu,\mu_c}(\overline{R}\,; F)$ in dimension two have been presented in [7]. The corresponding minimal energy levels were also provided. In dimension three, the following explicit formulae for the solutions to Problem 1.1 were obtained using computer algebra [8, Corollary 2.7]:

Corollary 1.2 (Energy-minimizing relative rotations for $(\mu, \mu_c) = (1, 0)$)**.** *Let $D = \operatorname{diag}(d_1, d_2, d_3)$ such that $d_1 > d_2 > d_3 > 0$. Then the solutions to Problem 1.1 are given by the energy-minimizing relative rotations*

$$\operatorname{rpolar}(D) = \left\{ \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \right\} , \tag{1.7}$$

where $\alpha \in [-\pi, \pi]$ is an optimal rotation angle satisfying

$$\alpha = \begin{cases} 0 \,, & \text{if} \quad d_1 + d_2 \leq 2 \,, \\ \pm \arccos(\frac{2}{d_1 + d_2}) \,, & \text{if} \quad d_1 + d_2 \geq 2 \,. \end{cases} \tag{1.8}$$

In particular, for $d_1 + d_2 \leq 2$, we have $\operatorname{rpolar}(D) = \{\mathbb{1}\}$.

Note that the validation of the minimizers in dimension three, i.e., of the formulae (1.7) and (1.8) in [8] was based on brute force stochastic minimization, since a proof of optimality was out of reach.

With the present contribution, we close this gap in $n = 3$ and generalize the previously obtained formulae $\operatorname{rpolar}^{\pm}_{1,0}(F)$ from [8, 10] to arbitrary dimension n. Note that the parameter transformation proved in [7] allows to recover the general solution in the non-classical parameter range $\operatorname{rpolar}^{\pm}_{\mu,\mu_c}(F)$ from $\operatorname{rpolar}^{\pm}_{1,0}(F)$ by a rescaling of the deformation gradient, but we shall not detail this here.

Our main result is that Problem 1.1 has 2^k global minimizers that are block-diagonal, similar to the $n = 3$ case above. Here, k is the number of blocks of size two. More precisely, we prove

Theorem 5.10. *Let $D \coloneqq \operatorname{diag}(d_1, \dots, d_n) > 0$ with ordered entries $d_1 > d_2 > \ldots > d_n > 0$. Let us fix the maximum $k \in \mathbb{N}_0$ for which $d_{2k-1} + d_{2k} > 2$. Any global minimizer $R \in \mathrm{SO}(n)$ of*

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

corresponds to a partition of the index set $\{1, \dots, n\}$ with $k \geq 0$ leading subsets of size two

$$\underbrace{\{1,2\} \sqcup \{3,4\} \sqcup \ldots \sqcup \{2k-1, 2k\}}_{k \text{ subsets of size two}} \sqcup \underbrace{\{2k+1\} \sqcup \ldots \sqcup \{n\}}_{(n-2k) \text{ subsets of size one}}$$

in the classification of critical points provided by Theorem 5.1. The global minimum of $W(R\,; D)$ is given by

$$W^{\mathrm{red}}(D) \coloneqq \min_{R \in \mathrm{SO}(n)} W(R\,; D) = \frac{1}{2} \sum_{i=1}^{k} (d_{2i-1} - d_{2i})^2 + \sum_{i=2k+1}^{n} (d_i - 1)^2 \,. \tag{1.9}$$

We note in passing that the case of optimal rotations for recurring parameter values d_i, $i = 1, \dots, n$, in the diagonal parameter matrix $D \in \operatorname{Diag}(n)$ has not been treated previously and is accessible with the present approach.[2]

[2] This allows to treat special cases of equal principal stretches ν_i of the deformation gradient $F \in \mathrm{GL}^+(n)$ which may arise, e.g., due to symmetry assumptions.

1.3 Algebraic solution strategy and state of the art

Let us lay out our solution strategy for Problem 1.1 and present the algebraic techniques which lie at the heart of it. The Euler-Lagrange equations for the function $W(R\,;D)$ have been previously derived in [8] and [32]. These equations characterize the critical points of the objective function $W(R\,;D)$ implicitly as solutions of a quadratic matrix equation posed on the manifold of rotations $\mathrm{SO}(n)$ and parametrized by the diagonal matrix $D = \mathrm{diag}(d_1, \ldots, d_n)$. The foundation of our solution approach is the successful analysis of the following equivalent algebraic condition

$$(RD - \mathbb{1})^2 \in \mathrm{Sym}(n)\ . \tag{1.10}$$

This is a *symmetric square condition*

$$\big(X(R)\big)^2 \;=\; S \in \mathrm{Sym}(n), \qquad \text{where} \qquad X(R) \coloneqq RD - \mathbb{1} \in \mathbb{R}^{n\times n}\ . \tag{1.11}$$

Given this condition, one might suspect that the computation of critical points of $W(R\,;D)$ is related to the theory of real matrix square roots of real symmetric matrices. However, the delicate properties of matrix square roots can, for the most part, be avoided and we consider this an intriguing aspect of our solution approach.

Matrix square roots are a classical theme in matrix analysis. The most-well known example is the unique symmetric positive definite, so-called principal, matrix square root of a symmetric positive definite real square matrix. However, as it turns out, a given real square matrix can have isolated and non-isolated families of matrix square roots which can be real, but are complex in general. A classification of all complex square roots of a given complex matrix $A \in \mathbb{C}^{n\times n}$ in terms of its Jordan decomposition has been given by Gantmacher, see [14]. This classification can also be adapted to the real case, see, e.g. Higham [17]. Further treatments and results on matrix square roots are given in the extensive monographs [18–20], while [12] provides a compact recent introduction.

Our development is, however, more intuitively phrased in terms of matrix squares, since we do not rely on the classical theory of matrix square roots. For example, the key to the analysis of Problem 1.1 is our Theorem 2.13 which states: *every real matrix $X \in \mathbb{R}^{n\times n}$ whose square $S = X^2$ is symmetric can be written in some orthonormal basis as a block-diagonal matrix with blocks of size at most two.* This statement does not seem to appear in the literature. The construction of this orthonormal basis was originally inspired by the theory of principal angles between linear subspaces, see, e.g. [11]. We emphasize that the *orthogonality* of the associated change of basis matrix $T \in \mathrm{O}(n)$ is of utmost importance for our solution strategy. We require the change of basis to preserve the Frobenius matrix norm in Problem 1.1. The block-diagonal structure in the new basis allows to break the minimization problem down into subproblems of dimension at most two. For example, we shall see that in $n = 3$, for a non-classical minimizer, we have to solve a one-dimensional and a two-dimensional subproblem. The one-dimensional problem determines the rotation axis of the optimal rotations, while the two-dimensional subproblem determines the optimal rotation angles.

Let us briefly illuminate two similar constructions, which are however insufficient for our purposes. We expand on these approaches in the text. First, based on the characterization of the set of complex matrix square roots due to Gantmacher [14], one can construct an invertible change of basis matrix $T_{\mathrm{G}} \in \mathrm{GL}(n)$ which is block-diagonal, but, in general, not orthogonal; see Remark 2.9 for details. Second, the numerical approximation of nonlinear matrix functions, that is currently an important research theme, provides another possible construction. In particular, our solution approach for Problem 1.1 bears some resemblance to the work of Higham underlying the computational approximation of real matrix square roots of real square matrices via their real Schur form, see [16, 17] and [18].[3] Our Example 2.18 and Remark 2.19 illustrate the relation between Theorem 2.13 and the real Schur form.

In the next subsection we outline the mechanical background of our problem. This part may be skipped by readers only interested in the algebraic development.

[3] For a geometric approach and an account of interesting recent developments in the numerical approximation of matrix square roots, see [37] and references therein.

1.4 Optimal Cosserat microrotations and applications in mechanics

The term Cosserat theory describes a class of models in nonlinear solid mechanics incorporating an additional field of rotations. Such models are also referred to as micropolar models; see [4] for an introduction including extensive references. This type of models dates back to the original work of the Cosserat brothers [3] in the early 1900s and was, historically, one of the first generalized continuum theories.[4]

Let us consider a body $\Omega \subset \mathbb{R}^n$ which is deformed by a diffeomorphism $\varphi : \Omega \to \varphi(\Omega) \subset \mathbb{R}^n$ with deformation gradient field $F \coloneqq \nabla\varphi : \Omega \to \mathrm{GL}^+(n)$ and let us denote the additional field of microrotations by $\overline{R} : \Omega \to \mathrm{SO}(n)$. In this context, we introduce the quadratic Cosserat shear-stretch strain energy density

$$\mathrm{W}_{\mu,\mu_c}(\overline{R}\,;F) \;\coloneqq\; \mu \left\|\mathrm{sym}(\overline{R}^T F - \mathbb{1})\right\|^2 \;+\; \mu_c \left\|\mathrm{skew}(\overline{R}^T F - \mathbb{1})\right\|^2 \tag{1.12}$$

which can be evaluated at every point $x \in \Omega$. The function $\mathrm{W}_{\mu,\mu_c} : \mathrm{SO}(n) \times \mathrm{GL}^+(n) \to \mathbb{R}_0^+$ depends on $F = \nabla\varphi$ and $\overline{R} : \Omega \to \mathrm{SO}(n)$. Note that in this text, we consider the deformation gradient F as a parameter, since our interest is to determine energy-minimizing rotations. The weights $\mu > 0$ and $\mu_c \geq 0$ are given by the Lamé shear modulus $\mu > 0$ from linear elasticity and the Cosserat couple modulus $\mu_c \geq 0$, see [28] for a discussion. The chosen quadratic ansatz for $W_{\mu,\mu_c}(\overline{R}\,;F)$ is motivated by a direct extension of the quadratic energy in the linear theory of Cosserat models, see, e.g. [21, 33, 34]. It is a contribution to the variational formulation of *any* geometrically nonlinear, isotropic, and quadratic Cosserat-micropolar continuum model, see [29] and [3, 5, 25].

The polar factor $\mathrm{R_p}(F) \in \mathrm{SO}(n)$ is obtained from the right polar decomposition $F = \mathrm{R_p}(F)\,U(F)$ of the deformation gradient $F \in \mathrm{GL}^+(n)$. It describes the macroscopic rotation of the continuum. Furthermore, $U(F) \coloneqq \sqrt{F^T F} \in \mathrm{PSym}(n)$ describes the stretch and is referred to as the right Biot-stretch tensor. We note that the singular values ν_i, $i = 1, \ldots, n$, of the deformation gradient $F \in \mathrm{GL}^+(n)$ are the eigenvalues of the symmetric positive definite matrix $U \in \mathrm{PSym}(n)$.

It is quite natural to study matrix distance problems in nonlinear continuum mechanics. Let us illustrate this for the example of the Euclidean distance function

$$\mathrm{dist}^2_{\mathrm{Euclid}}(F, \mathrm{SO}(n)) \coloneqq \min_{R \in \mathrm{SO}(n)} \|F - R\|^2 \,. \tag{1.13}$$

Conceptually, this function provides a local measure for the distance of the deformation $\varphi : \Omega \to \varphi(\Omega)$ to the isometric (locally length-preserving) embeddings of the body Ω into $\mathbb{R}^n$. The required invariance properties for isotropy are automatically satisfied. Furthermore, this is consistent with the requirement that a global isometry of a body $\Omega \subset \mathbb{R}^n$, i.e., a rigid body motion, does not produce any deformation energy, because $F = \nabla(Rx + b) = R \in \mathrm{SO}(n)$ implies

$$\int_\Omega \mathrm{dist}^2_{\mathrm{Euclid}}(F, \mathrm{SO}(n)) \;\mathrm{dV} = 0 \,.$$

Variations on this general theme lead to the study of corresponding minimization problems on $\mathrm{SO}(n)$ which have been the subject of multiple contributions, see, e.g., [6–8, 22, 31, 36]. Note that in classical nonlinear continuum models, the local rotation of the specimen at a point is not explicitly accounted for in the strain energy, due to the requirement of frame-indifference. Thus, in a classical theory, the local rotation of the specimen induced by a deformation mapping φ is always given by the continuum rotation $\mathrm{R_p}(\nabla\varphi)$.

In strong contrast, in Cosserat theory and other generalized continuum theories (so-called complex materials) with rotational degrees of freedom, the local rotation $\overline{R} : \Omega \to \mathrm{SO}(n)$ of the material appears explicitly. Accordingly, in such a theory, the computation of locally energy-minimizing rotations provides geometrical insight into the qualitative mechanical behavior of a particular constitutive model. The first result in this area apparently dates back to 1940 when Grioli [15]

[4] The Cosserat brothers established the foundations of continuum mechanics with rotational degrees of freedom and contributed physically necessary invariance requirements for a micropolar continuum theory: the strain energy density W in such a theory must be a function of the first Cosserat deformation tensor $\overline{U} \coloneqq \overline{R}^T F$. They never proposed a specific expression for the local strain energy density $W = W(\overline{U})$ to model specific materials.

proved a remarkable variational characterization of the polar factor $\mathrm{R_p}(F) \in \mathrm{SO}(3)$ in dimension three. We present a generalization [24] to arbitrary dimension

$$\operatorname*{arg\,min}_{\overline{R} \in \mathrm{SO}(n)} \left\| \overline{R}^T F - \mathbb{1} \right\|^2 = \{\mathrm{R_p}(F)\} \ , \quad \text{and} \tag{1.14}$$

$$\min_{\overline{R} \in \mathrm{SO}(n)} \left\| \overline{R}^T F - \mathbb{1} \right\|^2 = W_{\mathrm{Biot}}(F) \coloneqq \|U(F) - \mathbb{1}\|^2 \ . \tag{1.15}$$

Grioli's theorem shows that $\mathrm{R_p}(F)$ is optimal for the Cosserat strain energy minimized in (1.14).

Remark 1.3 (Non-classical rotation patterns in nature)**.** *Grioli's theorem implies that the strain energy minimized in* (1.14) *can only be expected to generate a microrotation field* $\overline{R} \approx \mathrm{R_p}(\nabla\varphi)$*, i.e., approximating the classical macroscopic continuum rotation.*[5] *However, non-classical rotation patterns* $\overline{R}$ *which deviate from* $\mathrm{R_p}(\nabla\varphi)$ *are of interest, since they can be observed in many domains of nature. In metals, for example, the local rotation of the crystal lattice may differ from the continuum rotation considerably which is of importance on the meso- and nano-scale. Non-classical counter-rotations of the crystal lattice have been observed in [38, 39] below nanoindentations in copper single crystals.*[6] *This motivated an analysis of the relaxed polar factors in the setting of an idealized nanoindenation in a copper single crystal [10]. Another application is in geomechanics, where non-classical rotational deformation modes may be observed, e.g., in landslides, see [26] and references therein. In the present work, we prove that the strain energy* $W_{\mu,\mu_c}(R\,;F)$ *defined in* (1.3) *can produce non-classical microrotations* $\overline{R} \approx \mathrm{rpolar}_{\mu,\mu_c}(F)$*, see also [7, 8, 10].*

The energy minimized in (1.14) is isotropic which allows for a quintessential simplification: it allows us to express Grioli's theorem in terms of rotations $R \in \mathrm{SO}(n)$ *relative* to the orthogonal polar factor $\mathrm{R_p}(F)$; see [8] for details. In this relative picture, the deformation gradient F is represented by a diagonal matrix $D \coloneqq \mathrm{diag}(d_1, \dots, d_n)$. The entries $d_i = \nu_i > 0$ are the singular values of $F \in \mathrm{GL}^+(n)$. Grioli's theorem then takes the following equivalent form

$$\operatorname*{arg\,min}_{\overline{R} \in \mathrm{SO}(n)} \|RD - \mathbb{1}\|^2 = \{\mathbb{1}\} \ , \quad \text{and} \tag{1.16}$$

$$\min_{\overline{R} \in \mathrm{SO}(n)} \|RD - \mathbb{1}\|^2 = \|D - \mathbb{1}\|^2 \ . \tag{1.17}$$

The optimal rotation *relative* to $\mathrm{R_p}(F)$ is given by the identity $\mathbb{1}$. Similarly, exploiting the isotropy of $W_{\mu,\mu_c}(\overline{R}\,;F)$ in (1.3), we obtain the expression

$$\mu \ \|\mathrm{sym}(RD - \mathbb{1})\|^2 \ + \ \mu_c \ \|\mathrm{skew}(RD - \mathbb{1})\|^2 \tag{1.18}$$

in terms of a rotation R relative to $\mathrm{R_p}(F)$. For the non-classical parameter range $\mu > \mu_c \geq 0$ (see [7] for details), a non-trivial parameter reduction proved in [7, Lem. 2.2] shows that the corresponding energy-minimizing rotations can be determined by solving Problem 1.1.

Let us briefly present a highly interesting logarithmic minimization problem. To this end, we introduce the logarithmic strain energy[7]

$$W_{\log}(R\,;D) \coloneqq \mu \ \|\mathrm{dev\,sym\log}(RD)\|^2 \ + \ \mu_c \ \|\mathrm{skew\log}(RD)\|^2 \ + \ \frac{\kappa}{2} \ (\mathrm{tr}\,[\log(RD)])^2 \ , \tag{1.20}$$

where κ denotes the so-called bulk modulus. Technicalities aside, one can prove that

$$\operatorname*{arg\,min}_{R \in \mathrm{SO}(n)} W_{\log}(R\,;D) = \{\mathbb{1}\} \, , \quad \text{and} \tag{1.21}$$

$$\min_{R \in \mathrm{SO}(n)} W_{\log}(R\,;D) = \mu \ \|\mathrm{dev}\log D\|^2 + \frac{\kappa}{2} \ (\mathrm{tr}\,[\log D])^2 \ , \tag{1.22}$$

[5] Note also that $\overline{R} = \mathrm{R_p}(\nabla\varphi)$ realizes the classical Biot-energy $\|U - \mathbb{1}\|^2$.

[6] The lattice misorientation can be measured by electron backscattered diffraction analysis (3D-EBSD).

[7] Note that the most general quadratic expression

$$\mu \ \|\mathrm{dev\,sym}(RD - \mathbb{1})\|^2 \ + \ \mu_c \ \|\mathrm{skew}(RD - \mathbb{1})\|^2 \ + \ \frac{\kappa}{2} \ (\mathrm{tr}\,[(RD - \mathbb{1})])^2 \tag{1.19}$$

is also of a certain interest, see [8] for a discussion. The associated optimal rotations are not known to us.

see [2, 22, 36] for proofs and essential details. Note that these results are closely related to the geometric observation that certain natural geodesic distances from $F \in \mathrm{GL}^+(n)$ to the subgroup $\mathrm{SO}(n)$ induce Hencky-type strain energies [30].

The Euclidean distance problem (1.16) and the minimization problem for the logarithmic energy (1.21) share a remarkable property: the identity $\mathbb{1} \in \mathrm{SO}(n)$ is always uniquely optimal for any diagonal positive definite $D > 0$. Equivalently, the polar factor $\mathrm{R_p}(F)$ is always the optimal absolute rotation. In view of Remark 1.3, we note that the logarithmic energy produces only classical microrotation patterns, i.e., $\overline{R} \approx \mathrm{R_p}(F)$.

In strong contrast, for $\mu > \mu_c \geq 0$, the quadratic Cosserat shear-stretch energy density (1.18) can produce interesting non-classical rotations [7, 8, 10], i.e.,

$$\operatorname*{arg\,min}_{R \in \mathrm{SO}(n)} \left(\mu \, \|\mathrm{sym}(RD - \mathbb{1})\|^2 \, + \, \mu_c \, \|\mathrm{skew}(RD - \mathbb{1})\|^2 \right) \, \neq \, \{\mathbb{1}\} \tag{1.23}$$

for the optimal relative rotation. Equivalently, it is possible that

$$\mathrm{rpolar}_{\mu,\mu_c}(F) \, \neq \, \{\mathrm{R_p}(F)\} \tag{1.24}$$

which means that the optimal Cosserat microrotations can produce non-classical rotation patterns, see also Remark 1.3. This interesting property is an immediate consequence of the characterization of $\mathrm{rpolar}(D)$ which we prove in the present work.

This paper is structured as follows: after this introduction in Section 1, we proceed to Section 2 which presents a construction of an orthonormal basis for any real matrix whose square is symmetric such that it takes block-diagonal form with blocks of size at most two. This block structure allows us to characterize the critical points in Section 3 for arbitrary dimension n. This leads to a sequence of decoupled one- and two-dimensional subproblems posed, however, on $\mathrm{O}(1)$ and $\mathrm{O}(2)$ and we continue with the solution of these subproblems in Section 4. In Section 5 we extract the globally energy-minimizing optimal Cosserat rotations from the set of critical points by a comparison of the realized energy levels. It turns out that the optimal rotations and energy levels are entirely consistent with previous results for $n = 2, 3$. We end with a short discussion of the present results in Section 6.

2 A block-diagonal representation of real matrices with a real symmetric square

In this section, we present the construction of an orthogonal change of basis $T \in \mathrm{O}(n)$ for real matrices $X \in \mathbb{R}^{n \times n}$ with a real symmetric square $S \coloneqq X^2 \in \mathrm{Sym}(n)$. The constructed change of basis preserves the Frobenius matrix norm which allows us to reduce Problem 1.1 to lower-dimensional subproblems.

Let us introduce

Definition 2.1. *We say that $X \in \mathbb{R}^{n \times n}$ is a real matrix square root of the real symmetric matrix $S \in \mathrm{Sym}(n)$, if it solves the quadratic matrix equation*

$$X^2 = S \in \mathrm{Sym}(n) \, .$$

For existence of complex matrix square roots and their classification, see [14] and [18]. The theory of real matrix square roots is considered in [17].

Example 2.2. *The identity matrix $\mathbb{1}_2 \in \mathrm{Sym}(2)$ has infinitely many real matrix square roots which are simply size two involution matrices. They fall into three distinct classes according to their trace:*

$$X = \mathbb{1}, \qquad X = -\mathbb{1}, \qquad X \in \left\{ \begin{pmatrix} a & b \\ c & -a \end{pmatrix}, \, a^2 + b\,c = 1 \right\} \, . \tag{2.1}$$

Example 2.3. *The negative identity matrix $-\mathbb{1}_2 \in \mathrm{Sym}(2)$ has the real matrix square roots*

$$X \in \left\{ \begin{pmatrix} a & b \\ c & -a \end{pmatrix}, \, a^2 + b\,c = -1 \right\} \, . \tag{2.2}$$

Example 2.4. *A negative identity matrix of odd size* $-\mathbb{1} \in \mathrm{Sym}(2k-1)$ *does not have real matrix square roots, since its determinant is negative.*

We now provide a simple criterion for $X \in \mathbb{R}^{2\times 2}$ to be a square root of *some* symmetric real matrix, i.e., $X^2 = (X^T)^2$. This will be useful in Section 4.

Lemma 2.5. *A real matrix* $X \in \mathbb{R}^{2\times 2}$ *is a real matrix square root of a real symmetric matrix* $S = X^2 \in \mathrm{Sym}(2)$ *if and only if* $X \in \mathrm{Sym}(2)$ *or* $\operatorname{tr}[X] = 0$.

Proof. The Cayley-Hamilton theorem implies that

$$S = X^2 = \operatorname{tr}[X]\, X - (\det[X])\, \mathbb{1} \, .$$

Since the square $S = X^2$ is symmetric, we have

$$0 = \operatorname{skew}(S) = \operatorname{skew}(\operatorname{tr}[X]\, X - (\det[X])\, \mathbb{1}) = \operatorname{tr}[X]\, \operatorname{skew}(X) \, .$$

This finishes the argument. ∎

In what follows, we write $E_{\lambda_i} \subseteq \mathbb{R}^n$ to denote the maximal real eigenspace of a real symmetric matrix $S \in \mathrm{Sym}(n)$ associated to a given eigenvalue λ_i, $i = 1, \ldots, m$.

Lemma 2.6 (Eigenspaces of $S = X^2 \in \mathrm{Sym}(n)$ are preserved by X)**.** *Let* $X \in \mathbb{R}^{n\times n}$ *with symmetric square* $S \coloneqq X^2 \in \mathrm{Sym}(n)$ *and let* $\lambda \in \mathbb{R}$ *be an eigenvalue of* S. *Then* X *preserves the eigenspace* E_λ *of its square* S, *i.e.,*

$$X E_\lambda \subseteq E_\lambda \, .$$

Proof. The operators X and X^2 commute. ∎

Corollary 2.7. *Let* $X \in \mathbb{R}^{n\times n}$ *be a real matrix square root of* $S = X^2 \in \mathrm{Sym}(n)$ *and let* $T \in \mathrm{O}(n)$ *such that*

$$\widetilde{S} \coloneqq T^{-1} S T = \operatorname{diag}(\underbrace{\lambda_1, \ldots, \lambda_1}_{\dim E_{\lambda_1}}, \underbrace{\lambda_2, \ldots, \lambda_2}_{\dim E_{\lambda_2}}, \ldots, \underbrace{\lambda_m, \ldots, \lambda_m}_{\dim E_{\lambda_m}}) \, .$$

Then the transformed matrix

$$\widetilde{X} \coloneqq T^{-1} X T = \operatorname{diag}(\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_m) \tag{2.3}$$

is block-diagonal with square blocks $\widetilde{X}_i$ *of size* $\dim E_{\lambda_i}$, $i = 1, \ldots, m$, *that satisfy*

$$\widetilde{X}_i^2 = \widetilde{S}_i = \lambda_i \mathbb{1} \, .$$

Remark 2.8. *The preceding Corollary 2.7 reduces the subsequent characterization of real matrix square roots of symmetric matrices significantly, because it shows that it suffices to consider each of the* X*-invariant eigenspaces* E_{λ_i}, $i = 1, \ldots, m$, *of* S individually.

Our next step is to construct a suitable *orthonormal* basis for each individual eigenspace E_{λ_i}, $i = 1, \ldots, m$, of $S \in \mathrm{Sym}(n)$. As we shall see in the following, this yields a change of basis with transition matrix $T \in \mathrm{O}(n)$ which is *adapted* to the structure of Problem 1.1. In particular, it preserves the Frobenius matrix norm.

Before we proceed with our proposed construction, we want to briefly describe an intuitive and *apparently* similar (but completely different) approach relying on the classification of the set of all complex matrix square roots due to Gantmacher [14].

Remark 2.9 (Lack of orthogonality in Gantmacher's representation)**.** *To make our point, it suffices to consider Gantmacher's classification of the complex matrix square roots for non-degenerate* $A \in \mathrm{GL}(n, \mathbb{C})$. *We follow the exposition of Higham [18, Thm. 1.24] including the notation. In our setting* $A = \widetilde{S}$ *is real diagonal, i.e., it is in Jordan canonical form with Jordan blocks of size*

one. Due to Gantmacher's classification, all complex *matrix square roots $X \in \mathbb{C}^{n\times n}$ which satisfy $X^2 = A = \widetilde{S}$ are of the form* [8]

$$X = U\,\mathrm{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_m})U^{-1}\,, \tag{2.4}$$

where $U \in \mathrm{GL}(n, \mathbb{C})$ is arbitrary, but required to commute with the Jordan matrix $A = \widetilde{S}$, i.e.,

$$U\,\mathrm{diag}(\lambda_1, \dots, \lambda_m) = \mathrm{diag}(\lambda_1, \dots, \lambda_m)U\,. \tag{2.5}$$

Consider now a real *matrix square root $X \in \mathbb{R}^{n\times n}$. It is not hard to see that the real Jordan form of X is obtained by an invertible change of basis matrix $T_{\mathrm{G}} \in \mathrm{GL}(n, \mathbb{R})$ with the property that $\widetilde{X} = T_{\mathrm{G}}^{-1}XT_{\mathrm{G}} \in \mathbb{R}^{n\times n}$ is block-diagonal with blocks of size one and two. Here, the blocks of size one correspond to positive eigenvalues of $A = \widetilde{S}$ and the blocks of size two correspond to negative eigenvalues of $A = \widetilde{S}$. Unfortunately, $T_{\mathrm{G}} \in \mathrm{GL}(n, \mathbb{R})$ is in general not orthogonal and the change of basis does not preserve the Frobenius matrix norm in Problem 1.1.*

Let us now proceed with the construction of a suitable *orthogonal* change of basis matrix $T \in \mathrm{O}(n)$ on which our subsequent analysis of Problem 1.1 is based.

We briefly recall the definition of the orthogonal complement $V^\perp$ of a linear subspace $V \subseteq \mathbb{R}^n$,

$$V^\perp \coloneqq \{w \in \mathbb{R}^n \mid w \perp V\} = \{w \in \mathbb{R}^n \mid \forall v \in V : \langle v, w\rangle = 0\}\,,$$

which induces an orthogonal decomposition of $\mathbb{R}^n = V \oplus_\perp V^\perp$. In what follows, we exploit the well-known fact that for $Y \in \mathbb{R}^{n\times n}$,

$$YV^\perp \subseteq V^\perp \quad\Longleftrightarrow\quad Y^TV \subseteq V\,. \tag{2.6}$$

Indeed, let $w \in V^\perp$, then $0 = \langle Yw, v\rangle = \langle w, Y^Tv\rangle$. Since the choice of $w \in V^\perp$ was arbitrary, we have that $Y^Tv \perp V^\perp$ which shows $Y^Tv \in V$, because $\mathbb{R}^n = V \oplus_\perp V^\perp$. The reverse implication is completely analogous.

Lemma 2.10 (Block lemma)**.** *Let $Y \in \mathbb{R}^{n\times n}$ with square $S \coloneqq Y^2 = \lambda\mathbb{1}$, $\lambda \in \mathbb{R}$. Then there exists an orthogonal transformation $T \in \mathrm{O}(n)$ such that the matrix*

$$\widetilde{Y} \coloneqq T^{-1}YT = \mathrm{diag}(\widetilde{Y}_1, \widetilde{Y}_2, \dots, \widetilde{Y}_r) = \begin{pmatrix} \widetilde{Y}_1 & 0 & \dots & 0 \\ 0 & \widetilde{Y}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \widetilde{Y}_r \end{pmatrix} \tag{2.7}$$

is block-diagonal, with blocks $\widetilde{Y}_i$, $i = 1, \dots, r$, of size one or two satisfying $\widetilde{Y}_i^2 = \lambda\mathbb{1}$.

Proof. The proof proceeds by induction on n. The base case of induction $n \in \{1, 2\}$ holds, since Y is already block-diagonal with blocks of size one or two. For the induction step let us assume that the statement holds for matrices of size $n - 1$ and $n - 2$.
Our strategy is to prove the existence of a one- or two-dimensional subspace V of $\mathbb{R}^n$ such that both V and its orthogonal complement $V^\perp$ are left invariant by Y, i.e.,

$$\dim V \in \{1, 2\}, \quad YV \subseteq V \quad \text{and} \quad YV^\perp \subseteq V^\perp\,. \tag{2.8}$$

Thus, if we pick an orthonormal basis of V and $V^\perp$, this is equivalent to the statement that orthogonal conjugates of Y and Y^T are block matrices of the form

$$Q^{-1}YQ = \left(\begin{array}{c|c} \widetilde{Y}_1 & 0 \\ \hline 0 & Z \end{array}\right)\,. \tag{2.9}$$

Equivalently, V is invariant under both Y and Y^T. Since $(Q^{-1}YQ)^2 = Q^{-1}Y^2Q = \lambda\mathbb{1}$, we get $Z^2 = \lambda\mathbb{1}$, so by the induction assumption there exists T_0 such that

$$T_0^{-1}ZT_0 = \begin{pmatrix} \widetilde{Z}_1 & 0 & \dots & 0 \\ 0 & \widetilde{Z}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \widetilde{Z}_s \end{pmatrix}\,. \tag{2.10}$$

[8] Here, we use the multi-valued convention to denote the set of all complex square roots by the symbol $\sqrt{\cdot}$.

Then the orthogonal matrix

$$T = Q \begin{pmatrix} \mathbb{1} & 0 \\ 0 & T_0 \end{pmatrix} \in \mathrm{O}(n) \tag{2.11}$$

satisfies

$$T^{-1}YT = \begin{pmatrix} \widetilde{Y}_1 & 0 & \dots & 0 \\ 0 & \widetilde{Z}_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \widetilde{Z}_s \end{pmatrix} , \tag{2.12}$$

which completes the induction step.

To finish the argument, we have to construct a Y- and Y^T-invariant subspace V of $\mathbb{R}^n$ of dimension one or two.

Since $(Y^T)^2 = S^T = S = \lambda\mathbb{1}$, the symmetric matrices YY^T and Y^TY commute

$$(YY^T)(Y^TY) = Y(Y^T)^2Y = Y(\lambda\mathbb{1})Y = \lambda S = \lambda^2\mathbb{1} = (Y^TY)(YY^T). \tag{2.13}$$

Therefore, the operators YY^T and Y^TY are simultaneously diagonalizable and we can find a common eigenvector w of both. Let us normalize w so that $\|w\| = 1$ and note that there exist values $\alpha, \beta \in [0, \infty)$ satisfying

$$Y^TYw \;=\; \alpha w \qquad \text{and} \qquad YY^Tw \;=\; \beta w \,. \tag{2.14}$$

Our next step is to choose the invariant subspace V. We have to distinguish several cases.

Case 1: $Yw \in \mathrm{span}\,(\{w\}), Y^Tw \in \mathrm{span}\,(\{w\})$, in other words, w is an eigenvector of Y and Y^T. We select $V = \mathrm{span}\,(\{w\})$ and construct an orthogonal matrix with first column given by $q_1 = w$, i.e.,

$$Q = (w|q_2|\dots|q_n) \in \mathrm{O}(n) \,. \tag{2.15}$$

An associated change of basis for Y and Y^T introduces the following zero patterns

$$Q^{-1}YQ = \left(\begin{array}{c|c} * & * \\ \hline 0 & \\ \vdots & * \\ 0 & \end{array}\right) \quad \text{and} \quad Q^{-1}Y^TQ = \left(\begin{array}{c|c} * & * \\ \hline 0 & \\ \vdots & * \\ 0 & \end{array}\right) . \tag{2.16}$$

Since $Q^{-1}Y^TQ = (Q^{-1}YQ)^T$ these matrices are transposes of each other which implies that we obtain a block matrix of the form

$$Q^{-1}YQ = \left(\begin{array}{c|ccc} * & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & * & \\ 0 & & & \end{array}\right) , \tag{2.17}$$

which is of the form described in (2.9).

Case 2: $Yw \in \mathrm{span}\,(\{w\}), Y^Tw \notin \mathrm{span}\,(\{w\})$, in other words w is an eigenvector of Y but not of Y^T. Consider the subspace $V = \mathrm{span}\left(\left\{w, Y^Tw\right\}\right)$. Then the image of V under Y satisfies

$$YV = \mathrm{span}\left(\left\{Yw, YY^Tw\right\}\right) \subseteq \mathrm{span}\,(\{w, w\}) \subseteq V \tag{2.18}$$

$$Y^TV = \mathrm{span}\left(\left\{Y^Tw, (Y^T)^2w\right\}\right) \subseteq \mathrm{span}\left(\left\{Y^Tw, \lambda w\right\}\right) \subseteq V \,. \tag{2.19}$$

We now pick an orthonormal basis w_1, w_2 of $V = \operatorname{span}(\{w_1, w_2\}) = \operatorname{span}\left(\left\{w, Y^T w\right\}\right)$ and extend it to an orthogonal matrix

$$Q = (w_1 | w_2 | q_3 | \dots | q_n) \in \mathrm{O}(n) \,. \tag{2.20}$$

Then, similar to Case 1, an associated change of basis for Y and Y^T introduces a zero pattern

$$Q^{-1}YQ = \left(\begin{array}{cc|c} * & * & * \\ * & * & * \\ \hline 0 & 0 & \\ \cdot & \cdot & \\ \cdot & \cdot & * \\ \cdot & \cdot & \\ 0 & 0 & \end{array}\right) \quad \text{and} \quad Q^{-1}Y^TQ = \left(\begin{array}{cc|c} * & * & * \\ * & * & * \\ \hline 0 & 0 & \\ \cdot & \cdot & \\ \cdot & \cdot & * \\ \cdot & \cdot & \\ 0 & 0 & \end{array}\right) \,. \tag{2.21}$$

As before, since $Q^{-1}Y^TQ = (Q^{-1}YQ)^T$ the two matrices are transposes of each other which creates a 2-block in the upper left corner

$$Q^{-1}YQ = \left(\begin{array}{cc|ccc} * & * & 0 & \cdots & 0 \\ * & * & 0 & \cdots & 0 \\ \hline 0 & 0 & & & \\ \cdot & \cdot & & & \\ \cdot & \cdot & & * & \\ \cdot & \cdot & & & \\ 0 & 0 & & & \end{array}\right) \,, \tag{2.22}$$

which is of the form described in (2.9).

Case 3: $Yw \notin \operatorname{span}(\{w\})$, in other words w is *not* an eigenvector of Y. We consider the subspace $V = \operatorname{span}(\{w, Yw\})$. The inclusion

$$YV = \operatorname{span}\left(\left\{Yw, Y^2w\right\}\right) = \operatorname{span}(\{Yw, \lambda w\}) \subseteq V \tag{2.23}$$

is immediate. In order to prove the invariance $Y^TV \subseteq V$, we need to consider the following two subcases:

Case 3a: $\lambda \neq 0$. In this case Y and Y^T are invertible and so $Y^TYw = \alpha w$ with $\alpha > 0$. This allows us to express w as follows

$$\left(\frac{1}{\alpha}Y^TY\right) w = \frac{\alpha}{\alpha}\, w = w \,. \tag{2.24}$$

We have to compute

$$Y^TV = \operatorname{span}\left(\left\{Y^Tw, Y^TYw\right\}\right) = \operatorname{span}\left(\left\{Y^Tw, \alpha w\right\}\right) \,. \tag{2.25}$$

To this end, we expand

$$Y^Tw = Y^T\left(\frac{1}{\alpha}Y^TY\right) w = \frac{1}{\alpha}\,(Y^2)^TYw = \frac{1}{\alpha}\,S^TYw = \frac{1}{\alpha}\,Y^3w = \frac{1}{\alpha}\,YSw = \frac{\lambda}{\alpha}\,Yw \in V \tag{2.26}$$

which shows that $Y^TV \subseteq V$.

Case 3b: $\lambda = 0$. Consider the product

$$\left(Y^TY\right)\left(YY^T\right) w = Y^TSY^Tw = S^2w = \lambda^2 w = 0 \,. \tag{2.27}$$

Since we also have, $Y^TYw = \alpha w$ and $YY^Tw = \beta w$, it follows that $\left(Y^TY\right)\left(YY^T\right) w = \alpha\beta w = 0$. Hence, $\alpha\beta = 0$. If $\beta = 0$, then

$$YY^Tw = 0 \quad \Longrightarrow \quad \left\langle w,\, YY^Tw \right\rangle = 0 \quad \Longrightarrow \quad \left\langle Y^Tw,\, Y^Tw \right\rangle = \|Y^Tw\|^2 = 0 \,. \tag{2.28}$$

Since $Y^T w = 0 \in V$, the subspace V is invariant under both Y and Y^T. The second case $\alpha = 0$ is not possible. To see this, we similarly compute

$$Y^T Y w = 0 \quad \Longrightarrow \quad Y w = 0 \tag{2.29}$$

which shows that w is an eigenvector of Y. This contradicts our assumptions for Case 3 (but note that this situation is handled in Case 1 or 2).

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

Remark 2.11. *The case $\lambda > 0$ of Lemma 2.10 can be deduced from the theory of principal angles (see, e.g, [11]) for the eigenspaces of Y with eigenvalues $\sqrt{\lambda}$ and $-\sqrt{\lambda}$. We are not aware of a similar connection in the case $\lambda \leq 0$.*

Remark 2.12. *The condition on Y is also sufficient, i.e., any matrix with the described block structure is a real matrix square root of a symmetric matrix. It is also possible to show that solutions to $Y^2 = \lambda \mathbb{1}$ exist if and only if $\lambda \geq 0$, or n is even.*

We are now ready to formulate the main result of this section.

Theorem 2.13. *For any real square matrix $X \in \mathbb{R}^{n\times n}$ with symmetric square $X^2 \in \mathrm{Sym}(n)$ there exists an orthogonal change of coordinates $T \in \mathrm{O}(n)$ such that the transformed matrix*

$$\widetilde{X} \coloneqq T^{-1} X T = \mathrm{diag}(\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_r) = \begin{pmatrix} \widetilde{X}_1 & 0 & \ldots & 0 \\ 0 & \widetilde{X}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \widetilde{X}_r \end{pmatrix} \tag{2.30}$$

is block-diagonal with square blocks $\widetilde{X}_j, j = 1, \ldots, r$, that are either of size one or two. Each block $\widetilde{X}_j$ is a real square root of a multiple of an identity matrix $\mathbb{1}$, i.e.,

$$\widetilde{X}_j^2 = \mu_j \mathbb{1}, \quad \mu_j \in \mathbb{R} \ .$$

Proof. It suffices to apply Lemma 2.10 to each eigenblock of $S = X^2$. ■

Remark 2.14. *Each eigenspace E_{λ_i} of S in Lemma 2.6 is possibly decomposed into multiple subspaces by Lemma 2.10. As a result, the eigenvalues μ_j, $1 \leq j \leq r$, of $\widetilde{X}_j^2$ in Theorem 2.13 are equal to the eigenvalues λ_i, $1 \leq i \leq m$, in the notation of Lemma 2.6 with, possibly, different indices. Several μ_j in the statement of Theorem 2.13 may be equal to the same λ_i in the sense of Lemma 2.6. For example, we might have the following*

$$(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5) = (\lambda_1, \lambda_1, \lambda_2, \lambda_3, \lambda_3) \ .$$

Remark 2.15. *An equivalent reformulation of the theorem is the following. For a matrix X whose square is symmetric, there exists a decomposition of $\mathbb{R}^n$ into an orthogonal direct sum of X-invariant subspaces V_i of dimension one or two such that X^2 is a multiple of the identity matrix on each V_i. The list of columns of the change of basis matrix $T \in \mathrm{O}(n)$ in Theorem 2.13 is obtained by concatenation of orthonormal bases of V_i. Note that each V_i is also invariant under X^T.*

Remark 2.16. *Given X and S the decomposition into invariant subspaces is not unique. In particular, a subspace of dimension two can sometimes be further decomposed into two one-dimensional subspaces.*

Remark 2.17. *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, see, e.g., [13, Thm. 12.5,p. 354].*

The following example illustrates the block-diagonal form stated in Theorem 2.13. At the same time it illuminates the relation of the construction to the real Schur form of a real matrix square root described in [17].

Example 2.18. *We use the notation of the Block Lemma 2.10 and consider the matrix*

$$Y = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \in \mathbb{R}^{4\times 4} \ . \tag{2.31}$$

Note that $Y^2 = \mathbb{1}$. Clearly, Y is not block-diagonal, but it is in real Schur form [17, Thm. 6]. In this representation, the value of $\|Y\|^2$ cannot be decomposed into diagonal contributions. As we noted before, the orthogonal transformation $T \in \mathrm{O}(n)$ described by Lemma 2.10 is not unique. Let us choose

$$T = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix} \quad \textit{which yields} \quad \widetilde{Y} \coloneqq T^{-1}YT = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & -1 \end{pmatrix} \ . \tag{2.32}$$

Hence, in the orthonormal basis defined by T, we obtain a block-diagonal representation $\widetilde{Y}$ with blocks of size at most two and satisying $\widetilde{Y}^2 = \mathbb{1}$. The Frobenius norm of Y is now composed of diagonal contributions

$$8 = \|Y\|^2 = \left\|T^{-1}YT\right\|^2 = \left\|\widetilde{Y}\right\|^2 = \left\|\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\right\|^2 + \left\|\begin{pmatrix} 1 & 2 \\ 0 & -1 \end{pmatrix}\right\|^2 = 2 + 6 \ . \tag{2.33}$$

In the present example, the transformed matrix $\widetilde{Y}$ is not in real Schur form, since the eigenvalues of the lower right 2×2-block are not complex conjugates. Note that, if we flip the third and fourth columns of T, then $\widetilde{Y}$ is, again, in real Schur form.

Remark 2.19. *The above example shows that the block-diagonal form guaranteed by Theorem 2.13 is not in general a real Schur form. It is, however, always possible to find an orthogonal change of basis that yields the block-diagonal representation of Theorem 2.13 in real Schur form, a fact which we will not use in the paper. Indeed, all blocks of size one and all blocks of size two with $\lambda < 0$ are already in real Schur form. For a size two block with $\lambda \geq 0$, we can pick an orthonormal basis that begins with an eigenvector.*

3 Critical points of the Cosserat shear-stretch energy

In this section we investigate the critical points $R \in \mathrm{SO}(n)$ of the objective function

$$W(R\,;D) \ = \ \|\mathrm{sym}(RD - \mathbb{1})\|^2 \tag{3.1}$$

in Problem 1.1. Since the objective function $W(R\,;D)$ is polynomial, we can proceed by taking derivatives along curves in the matrix group $\mathrm{SO}(n)$. Regarding the diagonal parameter $D = \mathrm{diag}(d_1, \ldots, d_n)$, we make a mild Assumption 3.3, in particular $d_i \neq 0$, and give a complete description of the critical points in that case. For some of our conclusions, we have to make the more restrictive Assumption 3.5 enforcing, in particular, positive $d_i > 0$.

In Lemma 3.1 we show that the matrix $X(R) \coloneqq RD - \mathbb{1}$ satisfies a symmetric square condition precisely when $R \in \mathrm{SO}(n)$ is a critical point of the objective function $W(R\,;D)$. Hence $X(R)$ is *orthogonally* similar to a block-diagonal matrix by Theorem 2.13 at the critical points. This is exploited in the *key result* Lemma 3.4 of this section, which shows that the block-structure of $X(R) = RD - \mathbb{1}$ at the critical points is simultaneously inherited by *both* R and D. This observation then allows to break Problem 1.1 down into decoupled subproblems of dimension one and two.

Since $\mathrm{SO}(n) \subset \mathbb{R}^{n\times n}$ is a Lie matrix group, we can identify the Lie algebra $\mathfrak{so}(n)$ with the tangent space $T_{\mathbb{1}}\,\mathrm{SO}(n) \subset \mathbb{R}^{n\times n}$ of $\mathrm{SO}(n)$ at the identity. It is well-known that the tangent space $\mathfrak{so}(n)$

is given by the skew-symmetric matrices $\mathrm{Skew}(n)$, see, e.g., [1]. Furthermore, the Frobenius inner product $\langle X, Y \rangle \coloneqq \mathrm{tr}\left[X^T Y\right]$ gives rise to an orthogonal decomposition of the vector space of *all* real square matrices

$$\mathbb{R}^{n\times n} = \mathrm{Sym}(n) \oplus_\perp \mathrm{Skew}(n) \ .$$

Our analysis of the critial points is based on the following algebraic stationarity condition obtained from the Euler-Lagrange equations.

Lemma 3.1 (Symmetric square condition)**.** *Let $D \coloneqq \mathrm{diag}(d_1, \dots, d_n) \in \mathbb{R}^{n\times n}$ be a diagonal matrix and let $X(R) \coloneqq RD - \mathbb{1}$. Then a rotation $R \in \mathrm{SO}(n)$ is a critical point of the objective function*

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

if and only if $X(R)$ satisfies the symmetric square condition

$$X(R)^2 \coloneqq (RD - \mathbb{1})^2 \in \mathrm{Sym}(n) \ .$$

Proof. In order to compute critical points in the submanifold $\mathrm{SO}(n) \subset \mathbb{R}^{n\times n}$, we have to locate zeroes of the tangent mapping $\mathrm{d}W : T\,\mathrm{SO}(n) \to T\,\mathbb{R}^{n\times n} \cong \mathbb{R}^{n\times n}$. To this end, we compute the derivatives of the energy $W(R\,; D)$ along a family of smooth curves

$$c_A : (-\varepsilon, \varepsilon) \to \mathrm{SO}(n), \quad c_A(t) \coloneqq \exp(tA)R \in \mathrm{SO}(n), \quad A \in \mathfrak{so}(n), \tag{3.2}$$

in the manifold of rotations. The right-trivialization of the tangent space at $R \in \mathrm{SO}(n)$ allows to identify $T_R\,\mathrm{SO}(n) = \mathfrak{so}(n) \cdot R = \mathrm{Skew}(n) \cdot R$ and so we can always express a tangent vector $\xi \in T_R\,\mathrm{SO}(n)$ in the form $\xi = AR \in \mathrm{Skew}(n) \cdot R$. This family of curves satisfies

$$\forall\, \xi = AR \in T_R\,\mathrm{SO}(n) : \quad \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} c_A(t) = AR = \xi \ . \tag{3.3}$$

Thus, for every possible tangent direction $\xi = AR \in T_R\,\mathrm{SO}(n)$, there is precisely one curve of the family which emanates from $R \in \mathrm{SO}(n)$ into this direction ξ.

A rotation R is a critical point of the energy $W(R\,; D)$ if and only if

$$\forall A \in \mathfrak{so}(n) : \quad \frac{\mathrm{d}}{\mathrm{d}t}\,(W \circ c_A)(t)|_{t=0} \ = \ 0 \ .$$

It is well-known that the matrix exponential is given by $(\mathbb{1} + tA)$ to first order in t and we write $\exp(tA) \sim (\mathbb{1} + tA)$. Thus, by the chain rule, we also have

$$(W \circ c_A)(t) \sim (W \circ (\mathbb{1} + tA)R)(t) \ .$$

We expand the expression

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

and obtain the expression for the first derivative $\mathrm{d}W$ from the term linear in t. In other words

$$\frac{\mathrm{d}}{\mathrm{d}t}\,(W \circ c_A)(t)|_{t=0} = 2\left\langle \mathrm{sym}(RD - \mathbb{1}),\, \mathrm{sym}(ARD)\right\rangle \ . \tag{3.4}$$

Hence, a point R is a critical point for the energy W if and only if it satisfies

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

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

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

Expanding the definition of the Frobenius inner product, we find

$$\begin{aligned} 0 = \left\langle \mathrm{sym}(RD - \mathbb{1}),\, ARD\right\rangle &= \mathrm{tr}\left[\mathrm{sym}(RD - \mathbb{1})^T ARD\right] = \mathrm{tr}\left[RD\,\mathrm{sym}(RD - \mathbb{1})A\right] \\ &= \left\langle \mathrm{sym}(RD - \mathbb{1})(RD)^T,\, A\right\rangle \ . \end{aligned} \tag{3.5}$$

Since this condition must hold for all $A \in \mathrm{Skew}(n)$, it follows that

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

We now multiply by a factor of 2 and expand the definition of $\mathrm{sym}(X) \coloneqq \frac{1}{2}(X + X^T)$ which leads us to

$$\begin{aligned} 2\,\mathrm{sym}(RD - \mathbb{1})DR^T &= (RD + DR^T - 2\mathbb{1})DR^T = RD^2R^T + (DR^T)^2 - 2DR^T \\ &= (DR^T - \mathbb{1})^2 + (RD^2R^T - \mathbb{1}). \end{aligned} \tag{3.6}$$

The second term on the right hand side is always symmetric and the effective condition for a critical point is thus

$$(DR^T - \mathbb{1})^2 \in \mathrm{Sym}(n) \ . \tag{3.7}$$

Finally, observing that symmetry is invariant under transposition, we conclude that

$$\left((DR^T - \mathbb{1})^2\right)^T = (RD - \mathbb{1})^2 \in \mathrm{Sym}(n) \tag{3.8}$$

is a sufficient and necessary condition for a critical point $R \in \mathrm{SO}(n)$ of $W(R\,;D)$. ■

Remark 3.2. *We immediately observe that $R = \mathbb{1}$ solves the condition* (3.8) *and is always a critical point of the energy $W(R\,;D) \coloneqq \|\mathrm{sym}(RD - \mathbb{1})\|^2$. However, in general, it will not be the global minimizer.*

Our next step is to apply Theorem 2.13 and Remark 2.15 to the special case $X(R) = RD - \mathbb{1}$. As we shall see, this implies quite restrictive conditions on $R \in \mathrm{SO}(n)$.

Let us make the following assumption on the diagonal matrix D.

Assumption 3.3. *The entries of the diagonal matrix $D = \mathrm{diag}(d_1, \ldots, d_n)$, which parametrizes the energy $W(R\,;D)$, do not vanish and do not cancel each other additively, i.e.,*

$$d_i \neq 0 \quad \textit{and} \quad d_i + d_j \neq 0, \quad 1 \leq i, j \leq n \ .$$

This ensures that $\ker(D) = \mathbf{0}$ and that any D^2-invariant subspace is also D-invariant. Note that if the entries of $D = \mathrm{diag}(d_1, \ldots, d_n)$ are positive, this assumption is satisfied. For the original problem in Cosserat theory which stimulated the present work [7, 8, 10], the entries of D are the singular values $\nu_i > 0$, $i = 1, \ldots, n$, of the deformation gradient $F \in \mathrm{GL}^+(n)$.

The following insight is a key to our discussion.

Lemma 3.4 (Simultaneous invariance of R and D)**.** *Suppose that the eigenvalues of D satisfy the above assumption. Let V be a subspace invariant under $X(R) = RD - \mathbb{1}$, such that $V^\perp$ is also invariant under $X(R)$. Then both V and $V^\perp$ are invariant under D and R.*

Proof. Recall first that V and $V^\perp$ are both invariant under $X(R)$ if and only if V is invariant under both $X(R)$ and $X(R)^T$; cf. (2.6).

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

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

From the assumption on D, we have $DV \subseteq V$. Since D has only nonzero eigenvalues D is invertible and so $DV = V$. It follows that

$$RDV \subseteq V \quad \Longrightarrow \quad RV \subseteq V \ .$$

Since $R \in \mathrm{SO}(n)$ is invertible, we have $RV = V$. Reversing the roles of V and $V^\perp$, we can apply the same argument to $V^\perp$. ■

By Theorem 2.13, as phrased in Remark 2.15, there exists a sequence of pairwise orthogonal vector spaces V_i, $i = 1, \dots, r$, with $1 \leq \dim V_i \leq 2$ which decompose $\mathbb{R}^n = V_1 \oplus_\perp V_2 \oplus_\perp \dots \oplus_\perp V_r$. These correspond to a block-diagonal representation of $X(R) \coloneqq RD - \mathbb{1}$. The existence of an associated orthogonal change of basis matrix $T \in O(n)$ is also assured by Theorem 2.13. Furthermore, by Lemma 3.4, both R and D are also block-diagonal with respect to this choice of basis. This means, in particular, that *any solution* R satisfying the symmetric square condition $(X(R))^2 = (RD - \mathbb{1})^2 \in \mathrm{Sym}(n)$ admits a block-diagonal representation. Since this condition characterizes the critical points by Lemma 3.1, any critical point of $W(R\,; D)$ admits a representation

$$\widetilde{R} = T^{-1} R T = \mathrm{diag}(\widetilde{R}_1, \dots, \widetilde{R}_r) = \begin{pmatrix} \widetilde{R}_1 & 0 & \dots & 0 \\ 0 & \widetilde{R}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \widetilde{R}_r \end{pmatrix} \in O(n) \subset \mathbb{R}^{n \times n} \tag{3.9}$$

in block-diagonal form. Here, the blocks on the diagonal satisfy $\widetilde{R}_i \in O(n_i)$, $i = 1, \dots, r$, with $n_i \in \{1, 2\}$ and $\sum_i^r n_i = n$.

In the basis provided by $T \in \mathrm{O}(n)$, any critical point $R \in \mathrm{O}(n)$ can be constructed from solutions $\widetilde{R}_i \in \mathrm{O}(n_i)$ of one- and two-dimensional subproblems

$$\left(\widetilde{X}(\widetilde{R}_i)\right)^2 \in \mathrm{Sym}(n_i)\,. \tag{3.10}$$

Note that these subproblems are now posed on the space of *orthogonal*, rather than *special orthogonal* matrices.

Assumption 3.5. *For the purpose of clarity of exposition, we make an additional, stronger assumption on the diagonal matrix* $D = \mathrm{diag}(d_1, \dots, d_n)$*, namely*

$$d_1 > d_2 > \dots > d_n > 0\,.$$

The slightly more general case of possibly non-distinct positive entries d_i can be treated similarly which we will indicate in running commentary.

Remark 3.6 (Implications of D-invariance)**.** *Under the Assumption 3.5, the D-invariance of the subspaces V_i shown in Lemma 3.4 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. Furthermore, by picking a standard coordinate basis for each V_i, we can ensure that the change of basis matrix $T \in \mathrm{O}(n)$ is a permutation matrix.*

We summarize that this particular structure allows to reduce the optimization Problem 1.1 to a finite list of decoupled one- and two-dimensional subproblems. However, we have to consider minimization with respect to *orthogonal* matrices $R \in \mathrm{O}(n)$ instead of $R \in \mathrm{SO}(n)$. This will be the content of the next section.

4 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

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

In this section we solve for critical points of the function

$$W(R_I\,; D_I) \coloneqq \left\| \mathrm{sym}(R_I D_I - \mathbb{1}) \right\|^2$$

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

Theorem 4.1 (Critical points: size one)**.** *For $I = \{i\}$ we have the submatrix $D_I = (d_i)$ and $R_I = \pm\mathbb{1} = (\pm 1)$. The realized critical energy levels are*

$$W(+\mathbb{1}\,; D_I) = (d_i - 1)^2 \quad \textit{and} \quad W(-\mathbb{1}\,; D_I) = (d_i + 1)^2 \,. \tag{4.1}$$

Proof. There are only two orthogonal matrices in dimension one and the result is immediate. ∎

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

Theorem 4.2 (Critical points: size two and 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 $d_i + d_j > 2$, there are two non-diagonal critical points*

$$R_I = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}, \quad \textit{with} \quad \cos\alpha = \frac{2}{d_i + d_j} \tag{4.2}$$

which attain the same critical value

$$W(R_I\,; D_I) = \frac{1}{2}(d_i - d_j)^2 \,. \tag{4.3}$$

Proof. By Lemma 3.1 R_I is a critical point if and only if $(R_I D_I - \mathbb{1})^2$ is symmetric. We may thus apply Lemma 2.5 which implies $R_I D_I - \mathbb{1} \in \mathrm{Sym}(2)$ or $\mathrm{tr}\,[R_I D_I - \mathbb{1}] = 0$. Using the explicit representation

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

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 $R_I = \pm\mathbb{1}$. 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 α if and only if $d_i + d_j \geq 2$. It gives rise to two non-diagonal solutions if and only if $d_i + d_j > 2$.
In the first case $R_I = \pm\mathbb{1}$, the critical values are immediately seen to be $(d_i - 1)^2 + (d_j - 1)^2$ and $(d_i + 1)^2 + (d_j + 1)^2$, respectively.
In the second case, the critical values are calculated as follows. Observing that

$$\mathrm{sym}(R_I D_I - \mathbb{1}) = \begin{pmatrix} d_i\cos\alpha - 1 & \frac{1}{2}(d_j - d_i)\sin\alpha \\ \frac{1}{2}(d_j - d_i)\sin\alpha & d_j\cos\alpha - 1 \end{pmatrix} \tag{4.4}$$

we use $(d_i + d_j)\cos\alpha = 2$ to get

$$\begin{aligned} \|\mathrm{sym}(R_I D_I - \mathbb{1})\|^2 &= (d_i\cos\alpha - 1)^2 + (d_j\cos\alpha - 1)^2 + \frac{1}{2}(d_j - d_i)^2\sin^2\alpha \\ &= (d_i^2 + d_j^2)\cos^2\alpha - 2(d_i + d_j)\cos\alpha + 2 + \frac{1}{2}(d_j - d_i)^2(1 - \cos^2\alpha) \\ &= \frac{1}{2}(d_j - d_i)^2 + \frac{1}{2}(d_i + d_j)^2\cos^2\alpha - 2(d_i + d_j)\cos\alpha + 2 \\ &= \frac{1}{2}(d_i - d_j)^2 + 2 - 4 + 2 = \frac{1}{2}(d_i - d_j)^2 \,. \end{aligned} \tag{4.5}$$

This shows the claim. ∎

Theorem 4.3 (Critical points: size two and 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 non-diagonal critical points*

$$R_I = \begin{pmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{pmatrix}, \quad \textit{with} \quad \cos\alpha = \frac{2}{|d_i - d_j|} \,, \tag{4.6}$$

which attain the same critical value

$$W(R_I\,; D_I) = \frac{1}{2}(d_i + d_j)^2 \,. \tag{4.7}$$

Proof. By Lemma 3.1 R_I is a critical point if and only if $(R_I D_I - \mathbb{1})^2$ is symmetric. We may thus apply Lemma 2.5 which implies $R_I D_I - \mathbb{1} \in \mathrm{Sym}(2)$ or $\mathrm{tr}\,[R_I D_I - \mathbb{1}] = 0$. Using the explicit representation

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

the symmetry condition $R_I D_I - \mathbb{1} \in \mathrm{Sym}(2)$ is equivalent to

$$(d_i - d_j)\sin\alpha = 0 \tag{4.8}$$

which has two solutions $R_I = \pm\,\mathrm{diag}(1,-1)$ since $d_i \neq d_j$ due to Assumption 3.5. 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 α if and only if $|d_i - d_j| \geq 2$. Thus there are two non-diagonal solutions if and only if $|d_i - d_j| > 2$.
In the first case $R_I = \pm\,\mathrm{diag}(1,-1)$, the critical values are immediately seen to be $(d_i-1)^2+(d_j+1)^2$ and $(d_i+1)^2 + (d_j-1)^2$, respectively.
In the second case, the critical values are calculated as follows. Observing that

$$\mathrm{sym}(R_I D_I - \mathbb{1}) = \begin{pmatrix} d_i\cos\alpha - 1 & \frac{1}{2}(d_i + d_j)\sin\alpha \\ \frac{1}{2}(d_i + d_j)\sin\alpha & -d_j\cos\alpha - 1 \end{pmatrix} \tag{4.9}$$

we use $|d_i - d_j|\cos\alpha = 2$ to get

$$\begin{aligned} \|\mathrm{sym}(R_I D_I - \mathbb{1})\|^2 &= (d_i\cos\alpha - 1)^2 + (d_j\cos\alpha + 1)^2 + \frac{1}{2}(d_i + d_j)^2\sin^2\alpha \\ &= (d_i^2 + d_j^2)\cos^2\alpha - 2(d_i - d_j)\cos\alpha + 2 + \frac{1}{2}(d_i + d_j)^2(1 - \cos^2\alpha) \\ &= \frac{1}{2}(d_i + d_j)^2 + \frac{1}{2}(d_i - d_j)^2\cos^2\alpha - 2(d_i - d_j)\cos\alpha + 2 \\ &= \frac{1}{2}(d_i + d_j)^2 + 2 - 4 + 2 = \frac{1}{2}(d_i + d_j)^2\,. \end{aligned} \tag{4.10}$$

This shows the claim. ∎

Remark 4.4 (The positive choice $\det[R_I] = +1$ minimizes energy)**.** *A direct comparison of the energy levels realized by the different choices for the determinant of R_I is instructive. Summarizing our preceding results, we have for $|I| = 1$, i.e., for a block of size one*

$$\det[R_I] = +1 \quad \mapsto \quad (d_i - 1)^2\,, \tag{4.11}$$

$$\det[R_I] = -1 \quad \mapsto \quad (d_i + 1)^2 \quad \geq \quad (d_i - 1)^2\,. \tag{4.12}$$

Similarly, for $|I| = 2$, i.e., for a block of size two, we obtain

$$\det[R_I] = +1 \quad \mapsto \quad \frac{1}{2}\,(d_i - d_j)^2\,, \tag{4.13}$$

$$\det[R_I] = -1 \quad \mapsto \quad \frac{1}{2}\,(d_i + d_j)^2 \quad \geq \quad \frac{1}{2}\,(d_i - d_j)^2\,. \tag{4.14}$$

The estimates follow from our Assumption 3.5 on the entries $d_i > 0$ of the diagonal matrix $D > 0$.

Remark 4.5. *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* (3.9)*.*

Remark 4.6 (On non-distinct entries of D)**.** *If we relax the Assumption 3.5 and allow for*

$$d_1 \geq d_2 \geq \ldots \geq d_n > 0$$

then there are degenerate critical points with $\det[R_I] = -1$ if and only if $d_i = d_j$. The corresponding critical value is the same as that realized by the diagonal matrices $\pm\,\mathrm{diag}(1,-1)$.

5 Global minimization of the Cosserat shear-stretch energy

Combining the results of the two preceding sections, we can now describe the critical values of the Cosserat shear-stretch energy $W(R\,;D)$ which are attained at the critical points. The main result of this section is a procedure (algorithm) which traverses the set of critical points in a way that reduces the energy at every step of the procedure and finally terminates in the subset of global minimizers.

Technically, we label the critical points by certain partitions of the index set $\{1,\dots,n\}$ 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 characterize a critical point $R \in \mathrm{SO}(n)$.

The next theorem expresses the value of $W(R\,;D)$ realized by a critical point in terms of the labeling partition and choice of determinants $\det[R_I]$ which characterize it.

Theorem 5.1 (Characterization of critical points and values)**.** *Let $D \coloneqq \mathrm{diag}(d_1,\dots,d_n) > 0$ satisfy Assumption 3.5, i.e., $d_1 > d_2 > \ldots > d_n > 0$. Then the critical points $R \in \mathrm{SO}(n)$ of the objective function*

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

can be classified according to partitions of the index set $\{1,\dots,n\}$ 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 critical values are given by

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

Proof. A suitable partition of the index set $\{1,\dots,n\}$ can be constructed as detailed in Section 3. The contributions of the subsets I of size one and two are given by the theorems of Section 4. It suffices to consider the non-diagonal 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 4.5. ∎

Remark 5.2 (On non-distinct entries of D)**.** *If we relax the Assumption 3.5 and allow for*

$$d_1 \geq d_2 \geq \ldots \geq d_n > 0$$

then the D- and R-invariant subspaces V_i are not necessarily coordinate subspaces. This produces non-isolated critical points but does not change the formula for the critical values.

It seems instructive to precede our further development with an outline of the scheme which allows us to traverse the set of critical points such that the energy decreases in every step and terminates in a global minimizer. Note that the scheme is conveniently formulated in terms of the labeling partitions which classify the critical points:

Scheme 5.3 (Construction of a minimizing sequence of critical points)**.** *Starting from the labeling partition of an arbitrary critical point:*

1. *Choose the positive sign $\det[R_I] = +1$ for each subset of the partition (cf. Remark 4.4 and Remark 5.4).*
2. *Disentangle all overlapping blocks for $n > 3$ (cf. Lemma 5.9).*
3. *Successively shift all 2×2-blocks to the lowest possible index, i.e., collect the blocks of size two as close to the upper left corner of the matrix R as possible (cf. Lemma 5.5).*
4. *Introduce as many additional 2×2-blocks by joining adjacent blocks of size 1 as the constraint $d_i + d_j > 2$ allows (cf. Lemma 5.5).*

At the end of this section, we provide an Example 5.13.

In order to compute the global minimizers $R \in \mathrm{SO}(n)$ for the Cosserat shear-stretch energy $W(R\,;D)$, we have to compare all the critical values which correspond to the different partitions and choices of the signs of the determinants in the statement of Theorem 5.1. In what follows, we prove the reduction steps of the preceding scheme.

Remark 5.4. *Notice that under Assumption 3.5, we have 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. Therefore, if R is a critical point which is a global minimizer of $||\operatorname{sym}(RD - \mathbb{1})||^2$, it only contains R_I with determinant $\det[R_I] = 1$.*

This allows us to assume that $\det[R_I] = 1$ for all subsets I without any loss of generality.

The following lemma shows that blocks of size two are always favored *whenever they exist*.

Lemma 5.5 (Comparison lemma)**.** *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*

$$-\frac{1}{2}(d_i + d_j - 2)^2.$$

Proof. We subtract the corresponding contributions of the subsets and simplify

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

This proves the claim. ∎

Let us rewrite $W(R\,;D)$ in a slightly different form in order to distill the contributions of the size two blocks in the partition.

Corollary 5.6. *For the choices of $\det[R_I] = 1$ there holds*

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

Proof. The first term in the formula is the value realized by $W(R\,;D)$ for the trivial partition into n subsets of size one. By virtue of the Comparison Lemma 5.5 each block of size two reduces the critical value by the amount $\frac{1}{2}(d_i + d_j - 2)^2$. ∎

Let us now consider the case of dimension $n = 3$ explicitly in order to prepare the exposition of the higher dimensional case.

Theorem 5.7. *Let $d_1 > d_2 > d_3 > 0$. If $d_1 + d_2 \leq 2$ then the global minimum of*

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

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 $d_1 + d_2 \leq 2$ then $d_i + d_j \leq 2$ for all index pairs (i, j) and there are no blocks of size two at the global minimum. If $d_1 + d_2 > 2$ then the choice of partition $\{1, 2\} \sqcup \{3\}$ is admissible. Corollary 5.6 shows that this is always favorable compared to the partition into three size one subsets $\{1\} \sqcup \{2\} \sqcup \{3\}$. Whether or not other size two subsets are admissible according to the inequalities $d_i + d_j > 2$, the partition $\{1, 2\} \sqcup \{3\}$ is always optimal. This follows from the ordering $d_1 > d_2 > d_3 > 0$ which implies that the partition-dependent term $\frac{1}{2}(d_i + d_j - 2)^2$ in Corollary 5.6 is maximized for $I = \{i, j\} = \{1, 2\}$. ■

In general, a deformation gradient $F \in \mathrm{GL}^+(n)$ can have non-distinct singular values $\nu_i = \nu_j$, $i \neq j$. This situation may arise, e.g., due to a symmetry assumption in mechanics.

Remark 5.8 (On non-distinct entries of D)**.** *Assume $d_1 \geq d_2 \geq d_3 > 0$. Our results imply the following:*
If $d_1 + d_2 \leq 2$, then all V_i are of dimension 1. Since the restriction of a given minimizer R to each V_i satisfies $R|_{V_i} = \mathbb{1}$, we see that $R = \mathbb{1}$. The global minimum of the Cosserat shear-stretch energy 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 for a global minimizer R there is a one-dimensional R-invariant subspace which is also D-invariant with associated eigenvalue d_3. Therefore, R is a rotation with axis in the d_3-eigenspace of D. The rotation angle satisfies the relation $(d_1 + d_2)\cos\alpha = 2$ and the global minimum of the energy is given by

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

This case further splits into several subcases all realizing the same energy level according to the multiplicity of the eigenvalue d_3:
If $d_1 \geq d_2 > d_3$, i.e., the multiplicity of d_3 is one, then there are two isolated global minimizers which are rotations with rotation angle $\arccos(2/(d_1+d_2))$ *with respect to either of the two half-axes in* $\mathrm{span}\,(\{e_3\})$ *(as in the case of distinct entries of D discussed in Theorem 5.7).*
If $d_1 > d_2 = d_3$, i.e., the multiplicity of d_3 is two, then the global minimizers R form a one-dimensional family of rotations with rotation angle $\arccos(2/(d_1 + d_2))$ *and rotation half-axes in the d_3-eigenplane* $\mathrm{span}\,(\{e_2, e_3\})$ *of D.*
If $d_1 = d_2 = d_3$, i.e., the multiplicity of d_3 is three, then there is a two-dimensional family of global minimizers R which are rotations with rotation angle $\arccos(2/(d_1 + d_2))$ *about arbitrary half-axes in* $\mathbb{R}^3$.

It is interesting that the set of global minimizers is *connected* in the last two cases where $d_2 = d_3$. This allows for a continuous transition between minimizers with opposite half-axes which are inverses of each other.

To study the global minimizers for the Cosserat shear-stretch energy in arbitrary dimension $n \geq 4$, we need to investigate the relative location of the size two subsets of the partition.

Lemma 5.9. *Let $R \in \mathrm{SO}(n)$ be a global minimizer for $W(R; D)$. Then R cannot contain overlapping size two subsets, i.e., $I = \{i_1, i_4\}$, $J = \{i_2, i_3\}$, with $i_1 < i_2 < i_3 < i_4$.*

Proof. We assume that R is a global minimizer corresponding to a partition containing two overlapping subsets as described above and derive a contradiction.
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. We recall the ordering $d_1 > d_2 > d_3 > d_4 > 0$.
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 $\{1, 4\} \sqcup \{2, 3\}$. By Corollary 5.6 we have

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

Thus R is not a global minimum of $W(R;D)$.

Case 2: $d_3 + d_4 \leq 2$. In this case, we can not have the size two subset $\{3,4\}$. However, it is possible to decrease the value of $W(R;D)$ by choosing another critical point $\mathring{R}$ corresponding to the partition $\{1,2\} \sqcup \{3\} \sqcup \{4\}$ instead of $\{1,4\} \sqcup \{2,3\}$. By Corollary 5.6 we have

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

In the first inequality we use the fact that for $d_1 + d_4 \geq 2$ the function $(d_1+d_4-2)^2$ is increasing in d_4 and $d_4 \leq 2 - d_3$ by assumption. This shows that R is not a global minimum of $W(R;D)$. We arrive at a contradiction in both cases which proves the statement. ■

We are now ready to state and prove the general n-dimensional case.

Theorem 5.10. *Let $D \coloneqq \operatorname{diag}(d_1,\dots,d_n) > 0$ with ordered entries $d_1 > d_2 > \ldots > d_n > 0$. Let us fix the maximum $k \in \mathbb{N}_0$ for which $d_{2k-1} + d_{2k} > 2$. Any global minimizer $R \in \mathrm{SO}(n)$ of*

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

corresponds to a partition of the index set $\{1,\dots,n\}$ with $k \geq 0$ leading subsets of size two

$$\underbrace{\{1,2\} \sqcup \{3,4\} \sqcup \ldots \sqcup \{2k-1,2k\}}_{k \text{ subsets of size two}} \sqcup \underbrace{\{2k+1\} \sqcup \ldots \sqcup \{n\}}_{(n-2k) \text{ subsets of size one}}$$

in the classification of critical points provided by Theorem 5.1. The global minimum of $W(R;D)$ is given by

$$\begin{aligned} W^{\mathrm{red}}(D) \coloneqq \min_{R \in \mathrm{SO}(n)} W(R;D) &= \sum_{i=1}^{n} (d_i - 1)^2 - \frac{1}{2}\sum_{i=1}^{k} (d_{2i-1} + d_{2i} - 2)^2 \\ &= \frac{1}{2}\sum_{i=1}^{k} (d_{2i-1} - d_{2i})^2 + \sum_{i=2k+1}^{n} (d_i - 1)^2 \,. \end{aligned}$$

Proof. Lemma 5.9 shows that a global minimizer $R \in \mathrm{SO}(n)$ can not have a partition with *overlapping* size two subsets. As in the proof of Theorem 5.7 (the $n = 3$ case) we can decrease the value of $W(R;D)$ by shifting down the indices of all size two subsets as far as possible. Therefore the optimal partition is of the form

$$\{1,2\} \sqcup \{3,4\} \sqcup \ldots \sqcup \{2l-1,2l\} \sqcup \{2l+1\} \sqcup \ldots \sqcup \{n\}$$

for some $l \leq k$. By Corollary 5.6 the global minimum is realized by the critical points corresponding to the maximal possible choice $l = k$. The value of $W(R;D)$ at a global minimizer is computed by inserting the corresponding optimal partition into Theorem 5.1 and Corollary 5.6. ■

Remark 5.11. *The number of global minimizers in the above theorem is 2^k, where k is the number of blocks of size two in the preceding characterization of a global minimizer as a block-diagonal matrix. All global minimizers are block-diagonal similar to the $n = 3$ case (Theorem 5.7).*

Remark 5.12 (On non-distinct entries of D)**.** *If we relax the Assumption 3.5 and allow for*

$$d_1 \geq d_2 \geq \ldots \geq d_n > 0$$

then the global minimizers may or may not be isolated. The formula for the reduced energy as stated in Theorem 5.10 is, however, not affected.

The following example illustrates the energy-minimizing traversal of critical points which always terminates in a global minimizer.

Example 5.13. *Let* $D = \operatorname{diag}\left(4, 2, 1, \frac{1}{2}, \frac{1}{4}\right)$. *Theorem 5.1 shows that the critical points can be characterized by certain partitions*[9] *of the index set* $\{1, 2, 3, 4, 5\}$ *and a choice of a sign for each subset* I *of the partition. Thus, we introduce the convenient notation of a pair of a subset and a sign* $(I, \pm)$, *where the sign encodes a possible choice for the determinant* $\det[R_I] = \pm 1$.

Setup: *We consider a critical point* $R^{(0)}$ *corresponding to the labeling partition*

$$\mathcal{P}^{(0)} = \{(\{1\}, +), (\{2,5\}, -), (\{3\}, -), (\{4\}, -)\} . \tag{5.1}$$

Note that $d_2 + d_5 = 2 + \frac{1}{4} > 2$, *i.e., the* 2×2*-block corresponding to* $I = \{2, 5\}$ *exists, as required for a valid partition characterizing a critical point* $R^{(0)}$. *The corresponding critical value of the summation formula in the statement of Theorem 5.1 is given by*

$$\begin{aligned} W^{(0)} = W(R^{(0)}; D) &= \underbrace{(4-1)^2}_{(\{1\},+)} + \underbrace{(1+1)^2}_{(\{3\},-)} + \underbrace{\left(1+\frac{1}{2}\right)^2}_{(\{4\},-)} + \underbrace{\frac{1}{2}\left(2+\frac{1}{4}\right)^2}_{(\{2,5\},-)} \\ &= \frac{569}{32} \approx 17.78 . \end{aligned} \tag{5.2}$$

Step 1 (Choice of positive sign): *We consistently choose the positive sign for the determinant in the labeling partition which gives*

$$\mathcal{P}^{(1)} = \{(\{1,5\}, +), (\{2\}, +), (\{3\}, +), (\{4\}, +)\} . \tag{5.3}$$

This updated partition characterizes a different *critical point* $R^{(1)}$ *realizing a lower energy level*

$$\begin{aligned} W^{(1)} = W(R^{(1)}; D) &= \underbrace{(4-1)^2}_{(\{1\},+)} + \underbrace{(1-1)^2}_{(\{3\},+)} + \underbrace{\left(1-\frac{1}{2}\right)^2}_{(\{4\},+)} + \underbrace{\frac{1}{2}\left(2-\frac{1}{4}\right)^2}_{(\{2,5\},+)} \\ &= \frac{345}{32} \approx 10.28 . \end{aligned} \tag{5.4}$$

Step 2 (Disentanglement): *The next step of the procedure is to remove overlap of* 2×2*-blocks. In our example, we only have one such block and there is nothing to do, i.e.,* $\mathcal{P}^{(2)} = \mathcal{P}^{(1)}$.

Step 3 (Index shift): *We now decrement the indices of the* 2×2*-blocks as much as possible, i.e., we string them together starting in the upper left corner. Shifting the* $\{2, 5\}$*-block to* $\{1, 2\}$, *we obtain the following new partition*

$$\mathcal{P}^{(3)} = \{(\{1,2\}, +), (\{3\}, +), (\{4\}, +), (\{5\}, +)\} . \tag{5.5}$$

The energy level realized by a corresponding critical point $R^{(3)}$ *is*

$$\begin{aligned} W^{(3)} = W(R^{(3)}; D) &= \underbrace{(1-1)^2}_{(\{3\},+)} + \underbrace{\left(1-\frac{1}{2}\right)^2}_{(\{4\},+)} + \underbrace{\left(1-\frac{1}{4}\right)^2}_{(\{5\},+)} + \underbrace{\frac{1}{2}(4-2)^2}_{(\{1,2\},+)} \\ &= \frac{45}{16} \approx 2.81 . \end{aligned} \tag{5.6}$$

Step 4 (Exhaustion by 2×2**-blocks):** *In this step, we try to create as many* 2×2*-blocks as possible. We first locate the pair of subsets of size one with minimal indices which is* $(\{3\}, \{4\})$. *Since* $d_3 + d_4 = 1 + \frac{1}{2} \leq 2$, *no further* 2×2*-block exists. Thus,* $\mathcal{P}^{(4)} = \mathcal{P}^{(3)}$.

[9] More precisely, a labeling partition uniquely characterizes sets of critical points which generate the same critical value. A block of size two, for example, characterizes two different symmetric solutions corresponding to the choice of sign for the rotation angle α. Both choices, however, yield the same value for the energy.

Result: *The finally obtained labeling partition*

$$\mathcal{P} = \mathcal{P}^{(4)} = \{(\{1,2\}\,,+)\,,(\{3\},+)\,,(\{4\},+)\,,(\{5\},+)\} \tag{5.7}$$

characterizes a global minimizer. With the notation of Theorem 5.10 the maximal number of 2×2*-blocks is* $k=1$ *and we have* $2^k = 2$ *global minimizers of the form*

$$\operatorname{rpolar}(D) = \left(\begin{array}{|cc|ccc} \hline \cos\alpha_1 & -\sin\alpha_1 & 0 & 0 & 0 \\ \sin\alpha_1 & \cos\alpha_1 & 0 & 0 & 0 \\ \hline 0 & 0 & \boxed{1} & 0 & 0 \\ 0 & 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & 0 & \boxed{1} \end{array}\right), \quad \textit{with} \quad \cos(\alpha_1) = \frac{2}{d_1+d_2} = \frac{1}{3}\,. \tag{5.8}$$

Inserting the global minimizers into the energy, we obtain the reduced energy

$$W^{\mathrm{red}}(D) \coloneqq W(\operatorname{rpolar}(D)\,;D) = \frac{45}{16} \approx 2.81\,. \tag{5.9}$$

Just to give a comparison, the identity matrix $\mathbb{1}\in \mathrm{SO}(n)$ *realizes the energy level*

$$\begin{aligned} W(\mathbb{1}\,;D) &= \underbrace{(4-1)^2}_{(\{1\},+)} + \underbrace{(2-1)^2}_{(\{2\},+)} + \underbrace{(1-1)^2}_{(\{3\},+)} + \underbrace{\left(1-\frac{1}{2}\right)^2}_{(\{4\},+)} + \underbrace{\left(1-\frac{1}{4}\right)^2}_{(\{5\},+)} \\ &= \frac{173}{16} \approx 10.81\,. \end{aligned} \tag{5.10}$$

Thus, the identity $\mathbb{1}\in\mathrm{SO}(n)$ *is not a global minimizer.*

Remark 5.14 (Optimality of $\mathbb{1}$)**.** *Our results imply that the identity matrix* $\mathbb{1}\in\mathrm{SO}(n)$ *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.,*

$$\max_{1\le i\neq j\le n}(d_i+d_j) \quad \le \quad 2\,.$$

This corresponds to the tension-compression asymmetry described in [7, 8, 10] for dimensions $n=2,3$*.*

6 Concluding remarks

For the sake of clarity of exposition, we have restricted our attention to the case of a diagonal and positive definite parameter matrix $D>0$, i.e., $d_i>0$. Our technical approach, however, readily carries over to the more general case $d_i\neq 0$ with minor modifications. The construction

$$\left\| \operatorname{sym}\left\{ \left[R \left(\begin{array}{c|c} \mathbb{1} & \\ \hline & -\mathbb{1} \end{array}\right)\right] \left[\left(\begin{array}{c|c} \mathbb{1} & \\ \hline & -\mathbb{1} \end{array}\right) D\right] - \mathbb{1} \right\} \right\|^2 \tag{6.1}$$

allows to reduce such a parameter matrix D to $|D| \coloneqq \operatorname{diag}(|d_1|\,,\ldots,|d_n|) > 0$ which is positive definite. Note that the minimization must then be carried out in the appropriate connected component of the orthogonal matrices $\mathrm{O}(n)$. We also expect that the degenerate case where some $d_i=0$ can be handled with our techniques as well.

The matrix group of rotations $\mathrm{SO}(3)$ equipped with its natural bi-invariant Riemannian metric

$$g(\xi,\eta)|_R \coloneqq g(R^T\xi, R^T\eta)|_{\mathbb{1}} \coloneqq \langle R^T\xi,\, R^T\eta\rangle = \langle \xi,\,\eta\rangle \tag{6.2}$$

is a Riemannian manifold $(\mathrm{SO}(3), g)$. In [27], the dynamics of the following Riemannian gradient flow[10] was investigated

$$R^T\dot{R} = \operatorname{skew}(R^T D) \quad \Longleftrightarrow \quad \dot{R} = -\operatorname{grad}\left(\frac{1}{2}\,\|RD-\mathbb{1}\|^2\right)\,. \tag{6.3}$$

[10] For an introductory exposition of gradient flows on Riemannian manifolds, see, e.g., [23].

The flow (6.3) converges to $R = \mathbb{1}$ for appropriate initial conditions which is consistent with Grioli's theorem; cf. Section 1. Similarly, one can study the gradient flow for the energy $\frac{1}{2}\left\|\operatorname{sym}(RD - \mathbb{1})\right\|^2$ given by

$$R^T\dot{R} = -\frac{1}{2}\operatorname{skew}\left((R^TD - \mathbb{1})^2\right) \quad \Longleftrightarrow \quad \dot{R} = -\operatorname{grad}\left(\frac{1}{2}\left\|\operatorname{sym}(RD - \mathbb{1})\right\|^2\right) . \tag{6.4}$$

Our present results on critical points of $W(R\,;D)$ determines the possible asymptotic solutions for the gradient flow (6.4). A characterization of *local* minimizers is currently missing. For example, it is not clear whether every local minimizer is automatically a global minimizer which holds in dimension $n = 2$. It seems likely, that this holds in $n = 3$ as well. The classification of local extrema of $W(R\,;D)$ is a completely open question in $n \geq 4$.

Acknowledgments: Lev Borisov was partially supported by NSF grant DMS-1201466. Andreas Fischle was supported by German Research Foundation (DFG) grant SA2130/2-1 and, previously, partially supported by DFG grant NE902/2-1 (also: SCHR570/6-1).

References

[1] A. Baker. *Matrix Groups: An Introduction to Lie Group Theory*. Undergraduate Mathematics. Springer, 2012.

[2] L. Borisov, P. Neff, S. Sra, and C. Thiel. The sum of squared logarithms inequality in arbitrary dimensions. *arXiv preprint arXiv:1508.04039*, 2015. `http://arxiv.org/abs/1508.04039`, to appear in Lin. Alg. Appl.

[3] E. Cosserat and F. Cosserat. *Théorie des corps déformables*. Librairie Scientifique A. Hermann et Fils (engl. translation by D. Delphenich 2007, available online at `https://www.uni-due.de/~hm0014/Cosserat_files/Cosserat09_eng.pdf`), reprint 2009 by Hermann Librairie Scientifique, ISBN 978 27056 6920 1, Paris, 1909.

[4] V. A. Eremeyev, L. P. Lebedev, and H. Altenbach. *Foundations of Micropolar Mechanics*. Springer, 2012.

[5] A. C. Eringen. *Microcontinuum Field Theories. Vol. I: Foundations and Solids*. Springer, 1999.

[6] A. Fischle. The planar Cosserat model: minimization of the shear energy on SO(2) and relations to geometric function theory. (diploma thesis). 2007. (available online: `http://www.uni-due.de/~hm0014/Supervision_files/dipl_final_online.pdf`).

[7] 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. *arXiv preprint arXiv:1507.05480*, 2015. `http://arxiv.org/abs/1507.05480`, to appear in Z. angew. Math. Mechanik.

[8] 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. *arXiv preprint arXiv:1509.06236*, 2015. `http://arxiv.org/pdf/1509.06236v1`.

[9] A. Fischle and P. Neff. Grioli's Theorem with weights and the relaxed-polar mechanism of optimal Cosserat rotations. *accepted by Atti Accad. Naz. Lincei Rend. Lincei Mat. Appl.*, 2017.

[10] 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. *arXiv preprint arXiv:1603.06633*, 2016. `http://arxiv.org/abs/1603.06633`.

[11] A. Galántai. *Projectors and projection methods*, volume 6. Springer Science & Business Media, 2013.

[12] J. Gallier. Logarithms and square roots of real matrices. *arXiv preprint arXiv:0805.0245*, 2008. `http://arxiv.org/abs/0805.0245`.

[13] J. Gallier. *Geometric methods and applications: for computer science and engineering*, volume 38. Springer Science & Business Media, 2. edition, 2011.

[14] F. R. Gantmacher. Matrix Theory, Vol. 1. *New York*, 1959.

[15] G. Grioli. Una proprieta di minimo nella cinematica delle deformazioni finite. *Boll. Un. Math. Ital.*, 2:252–255, 1940.

[16] N. J. Higham. Newton's method for the matrix square root. *Mathematics of Computation*, 46(174):537–549, 1986.

[17] N. J. Higham. Computing real square roots of a real matrix. *Lin. Alg. Appl*, 88:405–430, 1987.

[18] N. J. Higham. *Functions of Matrices: Theory and Computation*. SIAM, Philadelphia, PA, USA, 2008.

[19] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, New York, 1985.

[20] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, New York, 1991.

[21] J. Jeong, H. Ramézani, I. Münch, and P. Neff. A numerical study for linear isotropic Cosserat elasticity with conformally invariant curvature. *Z. Angew. Math. Mech.*, 89(7):552–569, 2009.

[22] J. Lankeit, P. Neff, and Y. Nakatsukasa. The minimization of matrix logarithms: On a fundamental property of the unitary polar factor. *Lin. Alg. Appl.*, 449:28–42, 2014.

[23] J. M. Lee. *Introduction to Smooth Manifolds*. Graduate Texts in Mathematics. Springer, 2002.

[24] L. C. Martins and P. Podio-Guidugli. An elementary proof of the polar decomposition theorem. *Amer. Math. Month.*, 87:288–290, 1980.

[25] G. A. Maugin. On the structure of the theory of polar elasticity. *R. Soc. Lond. Philos. Trans. Ser. A Math. Phys. Eng. Sci.*, 356(1741):1367–1395, 1998.

[26] I. Münch. *Ein geometrisch und materiell nichtlineares Cosserat-Modell - Theorie, Numerik und Anwendungsmöglichkeiten*. Dissertation in der Fakultät für Bauingenieur-, Geo- und Umweltwissenschaften, Karlsruhe, 2007. `http://digbib.ubka.uni-karlsruhe.de/volltexte/1000007371`.

[27] P. Neff. Local existence and uniqueness for quasistatic finite plasticity with grain boundary relaxation. *Quart. Appl. Math.*, 63:88–116, 2005.

[28] P. Neff. The Cosserat couple modulus for continuous solids is zero viz the linearized Cauchy-stress tensor is symmetric. *Z. Angew. Math. Mech.*, 86:892–912, 2006.

[29] P. Neff. A finite-strain elastic-plastic Cosserat theory for polycrystals with grain rotations. *Int. J. Engng. Sci.*, 44:574–594, 2006.

[30] P. Neff, B. Eidel, and R. J. Martin. Geometry of logarithmic strain measures in solid mechanics. *arXiv preprint arXiv:1505.02203*, 2015. `http://arxiv.org/pdf/1505.02203v1` to appear in Arch. Rat. Mech. Analysis.

[31] P. Neff, B. Eidel, F. Osterbrink, and R. Martin. A Riemannian approach to strain measures in nonlinear elasticity. *C. R. Acad. Sci. Paris (Mecanique)*, 342(4):254–257, 2014.

[32] 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.*, 197:19–30, 2008.

[33] P. Neff and J. Jeong. A new paradigm: the linear isotropic Cosserat model with conformally invariant curvature energy. *Z. Angew. Math. Mech.*, 89(2):107–122, 2009.

[34] P. Neff, J. Jeong, and A. Fischle. Stable identification of linear isotropic Cosserat parameters: bounded stiffness in bending and torsion implies conformal invariance of curvature. *Acta Mech.*, 211(3-4):237–249, 2010.

[35] P. Neff, J. Lankeit, and A. Madeo. On Grioli's minimum property and its relation to Cauchy's polar decomposition. *Int. J. Engng. Sci.*, 80:209–217, 2014.

[36] 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(3):1132–1154, 2014.

[37] S. Sra. On the matrix square root via geometric optimization. *arXiv preprint arXiv:1507.08366*, 2015. `http://arxiv.org/abs/1507.08366`.

[38] N. Zaafarani, D. Raabe, F. Roters, and S. Zaefferer. On the origin of deformation-induced rotation patterns below nanoindents. *Acta Mater.*, 56(1):31 – 42, 2008.

[39] 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.*, Volume 54/7:1863–1876, 2006. `http://www.sciencedirect.com/science/article/B6TW8-4J91NNJ-1/2/76ab02ad11c9d01d545627eeb5df081b`.