

The sum of squared logarithms inequality in arbitrary dimensions

Lev Borisov¹ and Patrizio Neff² and Suvrit Sra³ and Christian Thiel⁴

November 3, 2015

Abstract

We prove the *sum of squared logarithms inequality* (SSLI) which states that for nonnegative vectors $x, y \in \mathbb{R}^n$ whose elementary symmetric polynomials satisfy $e_k(x) \leq e_k(y)$ (for $1 \leq k < n$) and $e_n(x) = e_n(y)$, the inequality $\sum_i (\log x_i)^2 \leq \sum_i (\log y_i)^2$ holds. Our proof of this inequality follows by a suitable extension to the complex plane. In particular, we show that the function $f: M \subseteq \mathbb{C}^n \rightarrow \mathbb{R}$ with $f(z) = \sum_i (\log z_i)^2$ has nonnegative partial derivatives with respect to the elementary symmetric polynomials of z . This property leads to our proof. We conclude by providing applications and wider connections of the SSLI.

Key Words: elementary symmetric polynomials, fundamental theorem of algebra, polynomials, geodesics, Hencky energy, logarithmic strain tensor, positive definite matrices, algebraic geometry, matrix analysis

AMS 2010 subject classification: 26D05, 26D07, 30C15, 97H20

Contents

1	Introduction	2
2	Proof of Theorem 1.2	3
3	Applications and Connections	12
3.1	Relation to entropy	12
3.2	The SSLI in terms of matrix invariants	13
3.3	Application to geodesic distance on $\text{Sym}_+(n)$	15
3.4	Additional applications	15
4	Alternative proof of Conjecture 2.6	16

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

²Patrizio 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

³Suvrit Sra, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139, United States, email: suvrit@mit.edu

⁴Corresponding author: Christian Thiel, Lehrstuhl für Nichtlineare Analysis und Modellierung, Fakultät für Mathematik, Universität Duisburg-Essen, Thea-Leymann Str. 9, 45127 Essen, Germany, email: christian.thiel@uni-due.de

1 Introduction

The *sum of squared logarithms inequality* (SSLI) arose first as a scientific issue in 2012 [23] while proving the following optimality result

$$\inf_{Q \in \text{SO}(n)} \|\text{sym Log } Q^T F\|^2 = \inf_{Q \in \text{SO}(n)} \inf_{\substack{Y \in \mathbb{R}^{n \times n} \\ \exp(Y) = Q^T F}} \|\text{sym } Y\|^2 = \|\log \sqrt{F^T F}\|^2, \quad (1)$$

where $Y = \text{Log } X$ denotes all solutions of the matrix exponential equation $\exp(Y) = X$, $\|\cdot\|$ denotes the Frobenius matrix norm, and $\text{sym } X := \frac{1}{2}(X + X^T)$.

The SSLI (formally stated in Theorem 1.2) has been investigated in a series of works. In 2013, it was examined more closely by Birsan, Neff and Lankeit in [7], who found a proof for $n \in \{2, 3\}$. For $n = 3$, the inequality can be written as follows: let $x_1, x_2, x_3, y_1, y_2, y_3 > 0$ be positive real numbers such that

$$\begin{aligned} x_1 + x_2 + x_3 &\leq y_1 + y_2 + y_3, \\ x_1 x_2 + x_1 x_3 + x_2 x_3 &\leq y_1 y_2 + y_1 y_3 + y_2 y_3, \\ x_1 x_2 x_3 &= y_1 y_2 y_3. \end{aligned}$$

Then, the sum of their squared logarithms satisfy the following inequality:

$$(\log x_1)^2 + (\log x_2)^2 + (\log x_3)^2 \leq (\log y_1)^2 + (\log y_2)^2 + (\log y_3)^2.$$

In 2015, Pompe and Neff [29] proved the SSLI for $n = 4$, based on a new idea that did not extend to higher dimensions without further complications. To state the SSLI for arbitrary n , we first recall

Definition 1.1. Let $x \in \mathbb{R}^n$. We denote by $e_k(x)$ the k -th *elementary symmetric polynomial*, i.e. the sum of all $\binom{n}{k}$ products of exactly k components of x :

$$e_k(x) := \sum_{1 \leq i_1 < \dots < i_k \leq n} x_{i_1} x_{i_2} \dots x_{i_k} \quad \text{for any } k \in \{1, \dots, n\}.$$

Note that $e_1(x) = x_1 + x_2 + \dots + x_n$ and $e_n(x) = x_1 \cdot x_2 \cdot \dots \cdot x_n$.

We also write $\mathbb{R}_+ := \{x \in \mathbb{R} \mid x > 0\}$ and $\mathbb{R}_- := \{x \in \mathbb{R} \mid x < 0\}$ and set $\mathbb{R}_+^n = (\mathbb{R}_+)^n$.

Theorem 1.2 (Sum of squared logarithms inequality). *Let $n \in \mathbb{N}$ and $x, y \in \mathbb{R}_+^n$ such that*

$$e_k(x) \leq e_k(y) \quad \text{for all } k \in \{1, \dots, n-1\}$$

and

$$e_n(x) = e_n(y).$$

Then

$$\sum_{i=1}^n (\log x_i)^2 \leq \sum_{i=1}^n (\log y_i)^2.$$

This statement can equivalently be expressed as a minimization problem:

For $x \in \mathbb{R}_+^n$, let

$$\mathcal{E}_x := \left\{ y \in \mathbb{R}_+^n \mid e_k(x) \leq e_k(y) \text{ for all } k \in \{1, \dots, n-1\} \text{ and } e_n(x) = e_n(y) \right\}.$$

Then

$$\inf_{y \in \mathcal{E}_x} \left\{ \sum_{i=1}^n (\log y_i)^2 \right\} = \sum_{i=1}^n (\log x_i)^2.$$

Since $(\log y_i)^2 \geq 0$ for all $i \in \{1, \dots, n\}$, the expression is bounded below by 0, so the infimum clearly exists. Note that \mathcal{E}_x is a non-convex set.

Remark 1.3. If the equality assumption $e_n(x) = e_n(y)$ in the last elementary symmetric polynomial is replaced by the weaker requirement $e_n(x) \leq e_n(y)$, then the conclusion no longer holds in general. As a counterexample, consider $x = (e^{-1}, \dots, e^{-1}) \in \mathbb{R}^n$ and $y = (1, \dots, 1) \in \mathbb{R}^n$; then $e_k(x) = \binom{n}{k} e^{-k} \leq \binom{n}{k} = e_k(y)$ for all $k \in \{1, \dots, n\}$, but $\sum_{i=1}^n (\log x_i)^2 = n > 0 = \sum_{i=1}^n (\log y_i)^2$.

Neff, Nakatsukasa and Fischle [26] showed that the SSLI implies (1).

The proof of the SSLI presented in this work was motivated by the second named author, who published the SSLI conjecture (at that point) on the internet platform *MathOverflow* [22]. The first named author extended the problem to the complex plane and presented a sketch of a proof.

Miroslav Šilhavý (Czech Academy of Science) considered the problem after private communication with P. Neff and provided a characterization of functions that satisfy E-monotonicity. Interestingly, shortly after seeing L. Borisov's solution, one of the authors (S. Sra) suggested via email that “a full generalization of this idea should be possible via Pick-Nevalinna theory.” This idea is natural, and the details were independently discovered and worked out by M. Šilhavý [31]; it is also worth noting that actually Jozsa and Mitchison [15] foreshadowed the Pick function based approach to proving such inequalities but did not develop it fully. Our remarks here merely outline the historical sequence of events (to our knowledge), and to highlight the remarkable fact that like many other problems in mathematics, the SSLI also witnessed several essentially simultaneous solutions; each exposing different aspects of it and thus contributing to our understanding.

In this paper we give a self-contained exposition of our new methods towards proving the SSLI.

2 Proof of Theorem 1.2

In our further calculations, we will use the following lemma and the resulting corollary.

Lemma 2.1. *Let z_1, \dots, z_n be pairwise different complex numbers and $k \in \{0, \dots, n-1\}$. Then*

$$\sum_{i=1}^n \frac{z_i^k}{(t-z_i) \prod_{\substack{j=1 \\ j \neq i}}^n (z_i - z_j)} = \frac{t^k}{\prod_{j=1}^n (t-z_j)} \quad \text{for all } t \in \mathbb{C} \setminus \{z_1, \dots, z_n\}. \quad (2)$$

For $n = 3$ and $k = 2$, for example, the equality reads

$$\frac{a^2}{(a-b)(a-c)(t-a)} + \frac{b^2}{(b-a)(b-c)(t-b)} + \frac{c^2}{(c-a)(c-b)(t-c)} = \frac{t^2}{(t-a)(t-b)(t-c)}.$$

Proof. Let $k \in \{0, \dots, n-1\}$. Then according to the theorem of partial fraction decomposition, there exist complex numbers a_1, \dots, a_n such that

$$\frac{t^k}{\prod_{j=1}^n (t - z_j)} = \sum_{i=1}^n \frac{a_i}{t - z_i} \quad \text{for all } t \in \mathbb{C} \setminus \{z_1, \dots, z_n\}.$$

For given $r \in \{1, \dots, n\}$, multiplying both sides of the equation with $t - z_r$ yields

$$\frac{t^k}{\prod_{\substack{j=1 \\ j \neq r}}^n (t - z_j)} = \sum_{i=1}^n \frac{t - z_r}{t - z_i} a_i \quad \text{for all } t \in \mathbb{C} \setminus \{z_1, \dots, z_n\}. \quad (3)$$

Taking the limit $t \rightarrow z_r$ on both sides of the equality, we find

$$\frac{z_r^k}{\prod_{\substack{j=1 \\ j \neq r}}^n (z_r - z_j)} = a_r. \quad \blacksquare$$

Corollary 2.2. *Let $z_1, \dots, z_n \in \mathbb{C}$ be pairwise different complex numbers. Then*

$$\sum_{i=1}^n \frac{z_i^k}{\prod_{\substack{j=1 \\ j \neq i}}^n (z_i - z_j)} = 0 \quad \text{for all } k \in \{0, \dots, n-2\}. \quad (4)$$

For example, we find for $n = 4$ and $k = 2$:

$$\frac{a^2}{(a-b)(a-c)(a-d)} + \frac{b^2}{(b-a)(b-c)(b-d)} + \frac{c^2}{(c-a)(c-b)(c-d)} + \frac{d^2}{(d-a)(d-b)(d-c)} = 0.$$

Proof. Let $k \in \{0, \dots, n-2\}$. Using equality (2), we obtain

$$\sum_{i=1}^{n-1} \frac{z_i^k}{(t - z_i) \prod_{\substack{j=1 \\ j \neq i}}^{n-1} (z_i - z_j)} = \frac{t^k}{\prod_{j=1}^{n-1} (t - z_j)} \quad \text{for all } t \in \mathbb{C} \setminus \{z_1, \dots, z_{n-1}\}. \quad (5)$$

Setting $t := z_n$ and rearranging the equation yields the statement. \blacksquare

To introduce the basic idea of our proof, we first recall the relationship between a vector $z := (z_1, z_2, \dots, z_n)$ and the vector of the elementary symmetric polynomials evaluated at z , i.e. $(e_1(z), \dots, e_n(z))$. To this end, we define the characteristic polynomial h_e of a linear map with the invariants e_1, \dots, e_n :

$$h_e(t) := t^n - e_1 t^{n-1} + e_2 t^{n-2} + \dots + (-1)^n e_n = t^n + \sum_{k=1}^n (-1)^k e_k t^{n-k}.$$

Since if h_e has the roots z_1, \dots, z_n , we can write

$$\begin{aligned} h_e(t) &= (t - z_1)(t - z_2) \dots (t - z_n) \\ &= t^n - (z_1 + \dots + z_n)t^{n-1} + (z_1 z_2 + \dots + z_{n-1} z_n)t^{n-2} + \dots + (-1)^n z_1 \dots z_n \\ &= t^n - e_1(z)t^{n-1} + e_2(z)t^{n-2} + \dots + (-1)^n e_n(z). \end{aligned} \quad (6)$$

In this paper we will study different restrictions in the co-domain of the elementary symmetric polynomials. However, we will always assume that this co-domain is positive and real. It is also convenient to introduce an ordering of the complex numbers in order to ensure the uniqueness of the coefficient vector corresponding to a given set of roots. We therefore define the set

$$\mathbb{C}^{n\uparrow} := \{z \in \mathbb{C}^n \mid \operatorname{Re}(z_1) \geq \dots \geq \operatorname{Re}(z_n), \operatorname{Re} z_i = \operatorname{Re} z_{i+1} \Rightarrow \operatorname{Im} z_i \geq \operatorname{Im} z_{i+1} \forall i \in \{1, \dots, n-1\}\},$$

which contains only ordered vectors and thereby excludes all rearrangements. Furthermore, we define the set

$$M := \{z \in \mathbb{C}^{n\uparrow} \mid e_1(z), e_2(z), \dots, e_n(z) \in \mathbb{R}_+\}$$

of all ordered vectors with exclusively positive elementary symmetric polynomials. In contrast to previous work on the SSLI, we extend our view directly to complex roots in M , which provides the crucial advantage.

Lemma 2.3. *The function $M \rightarrow \mathbb{R}_+^n$ that maps each vector $z \in M$ onto the coefficient vector e corresponding to the uniquely determined polynomial h_e with roots z_1, \dots, z_n is continuous and bijective. Its inverse function is continuous as well, and we denote it by*

$$\varphi: \mathbb{R}_+^n \rightarrow M \subseteq \mathbb{C}^{n\uparrow}, \quad (e_1, \dots, e_n) \mapsto \varphi(e_1, \dots, e_n).$$

Furthermore, each vector $(z_1, \dots, z_n) \in M$ contains only positive real numbers and complex conjugate pairs of numbers.

Proof. The elementary symmetric polynomials $e_1(z), \dots, e_n(z)$ evaluated at z are exactly the coefficients e_1, \dots, e_n of the polynomial h_e with the roots z_1, \dots, z_n . The elementary symmetric polynomials are obviously continuous.

On the other hand, applying the fundamental theorem of algebra, we know that h_e has exactly n complex roots, all of which are either real or complex conjugate pairs. It is easy to see that all real roots must be positive: since the polynomial $h_e(-t) = t^n + \sum_{k=1}^n e_k t^{n-k} > 0$ for all $x \in \mathbb{R}_+$, because all e_k are positive. Thus $h_e(-t)$ has no positive and therefore $h_e(t)$ has no negative real roots. A proof of the continuity of φ is shown in [9]. ■

We now come to the proof of the SSLI. The main idea has already been pursued in prior attempts to prove the inequality: instead of directly working with the function $f(z) := \sum_{i=1}^n (\log z_i)^2$ on the set M of roots, we consider the composition $f \circ \varphi$ which depends on the elements $e \in T$ of a suitable set of coefficients $T \subseteq \mathbb{R}_+^n$. Of course, we have to choose T in a way such that $(f \circ \varphi)(e) \in \mathbb{R}$ for all $e \in T$.

The proof of the SSLI now can be divided into two steps:

- 1.) We show that $\frac{\partial(f \circ \varphi)}{\partial e_k} \geq 0$.

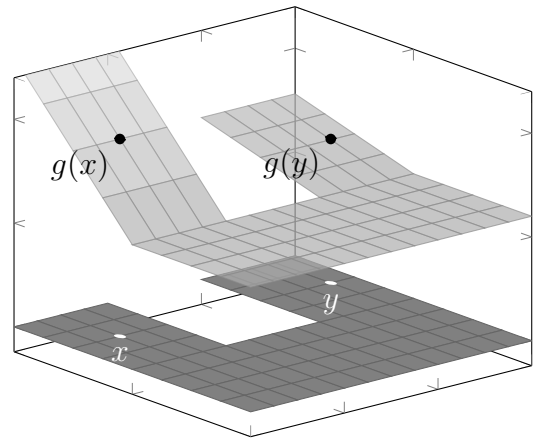


Figure 1: The graph of a function g with non-negative partial derivatives but non-convex domain of definition; note that $g(x) > g(y)$, although $x_i \leq y_i$ for $i \in \{1, 2\}$.

2.) We find a path $\gamma: [0, 1] \rightarrow \varphi(T)$ with $\gamma(0) = x$, $\gamma(1) = y$ such that $\frac{d}{ds}e_k(\gamma(s)) \geq 0$ for all $s \in (0, 1)$ and $k \in \{1, \dots, n-1\}$ as well as $\frac{d}{ds}e_n(\gamma(s)) = 0$ for all $s \in (0, 1)$.

Note carefully that condition 1.) alone is not sufficient. To understand this, let us consider the graph of a function $g: D \rightarrow \mathbb{R}$ with a non-convex domain $D \subseteq \mathbb{R}^n$ as shown in Figure 1.

As we see, even though the function only has non-negative partial derivatives, it is *not* true that $g(x) \leq g(y)$ for each pair $x, y \in D$ such that (componentwise) $x \leq y$. In particular, we cannot find a path connecting x to y which is increasing in all components.

We therefore want to choose an appropriate domain $T \subseteq \mathbb{R}_+^n$ in order to prevent these complications. The problem is that for a too restricted choice of T , we do not easily find suitable paths satisfying condition 2.), while if T is chosen too large, it becomes more difficult to prove that condition 1.) holds on all of T .

In [29], Pompe and Neff managed to prove the SSLI for $n = 4$ by choosing

$$T = \{ (e_1(z), \dots, e_n(z)) \mid z \in \mathbb{R}_+^n \}; \quad (7)$$

the authors in fact do not to limit the inequality to the function $f(z) := \sum_{i=1}^n (\log z_i)^2$, but to prove it for a whole class of functions. For these f , they show condition 1.), i.e. the non-negativity of the partial derivatives with respect to the k -th elementary symmetric polynomial, by showing for each point on the path from condition 2.) that

$$\widehat{D} := \sum_{j=1}^n f'(z_j) (-z_j)^{n-k} \left(\prod_{\substack{j=1 \\ j \neq i}}^n (z_j - z_i) \right)^{-1} \geq 0. \quad (8)$$

Although their choice of the domain T is not convex, they demonstrate the existence of paths from x to y without constructing them explicitly: they show that a path which satisfies condition 2.) can be continued and must therefore reach y after starting at x . However, for $n > 4$, this method seems impractical due to the amount of computational effort required.

In this paper, we consider the convex set

$$T := \mathbb{R}_+^n. \quad (9)$$

By this choice, satisfying condition 2.) is rather trivial; we take the straight line $\tilde{\gamma}$ in T from $e(x)$ to $e(y)$, i.e. $\tilde{\gamma}(s) := s \cdot e(x) + (1-s) \cdot e(y)$, and consider the curve $\gamma: [0, 1] \rightarrow M$ with $\gamma(s) = \varphi(\tilde{\gamma}(s))$. Of course, it is difficult to explicitly characterize γ , but this is not necessary for our proof. It therefore only remains to verify condition 1.), which requires some more elaborate methods.

As indicated earlier, we define the function

$$f: (\mathbb{C} \setminus (\mathbb{R}_- \cup \{0\}))^n \rightarrow \mathbb{C} \quad \text{with } f(z_1, \dots, z_n) := \sum_{i=1}^n (\log z_i)^2, \quad (10)$$

where \log denotes the main branch of the complex logarithm, which is defined for $\mathbb{C} \setminus (-\infty, 0]$.

Lemma 2.4. *The composition $f \circ \varphi$ is a real-valued function on \mathbb{R}_+^n , and takes the form*

$$f \circ \varphi: \mathbb{R}_+^n \rightarrow \mathbb{R} \quad \text{with } (f \circ \varphi)(e) := f(\varphi_1(e), \dots, \varphi_n(e)).$$

Proof. From Lemma 2.3 we already know that the components of $(\varphi_1(e), \dots, \varphi_n(e))$ are either real or pairs of complex conjugates. For such pairs,

$$(\log \varphi_i(e))^2 + (\log \overline{\varphi_i(e)})^2 = (\log \varphi_i(e))^2 + \overline{(\log \varphi_i(e))^2} = (\log \varphi_i(e))^2 + \overline{(\log \varphi_i(e))^2} \in \mathbb{R}.$$

Since $(\log \varphi_i(e))^2$ is real-valued for all real components $\varphi_i(e)$ as well, it follows that

$$(f \circ \varphi)(e_1(z), \dots, e_n(z)) = \sum_{i=1}^n (\log z_i)^2 \in \mathbb{R}. \quad \blacksquare$$

In order to prove condition 1.), we need to compute the inner derivative of the composition $f \circ \varphi$:

Proposition 2.5. *Let $e := (e_1, \dots, e_n) \in \mathbb{R}_+^n$ be such that all components of $\varphi(e)$ (i.e. the roots of h_e) are pairwise different. Then φ is differentiable at e with partial derivative*

$$\frac{\partial \varphi_i}{\partial e_k}(e) = \frac{(-1)^{k+1} \varphi_i(e)^{n-k}}{\prod_{\substack{j=1 \\ j \neq i}}^n (\varphi_i(e) - \varphi_j(e))} \quad \text{for any } k \in \{1, \dots, n\}.$$

Proof. Using the notation $h(e_1, \dots, e_n, t) := h_e(t)$, we can characterize the functions φ_i , $i \in \{1, \dots, n\}$, implicitly by the equation

$$0 = h(e_1, \dots, e_n, \varphi_i(e_1, \dots, e_n)) \quad \text{for all } e = (e_1, \dots, e_n) \in \mathbb{R}_+^n.$$

According to the implicit function theorem, φ_i is differentiable at a point $e \in \mathbb{R}_+^n$ if $\frac{\partial h}{\partial t}(e, \varphi_i(e)) \neq 0$. By assumption, the roots $\varphi_1(e), \dots, \varphi_n(e)$ of h_e are pairwise different; in particular, the root $\varphi_i(e)$ is simple and therefore $\frac{\partial h}{\partial t}(e, \varphi_i(e)) \neq 0$. We differentiate

$$\begin{aligned} 0 &= \frac{d}{de_k} h(e, \varphi_i(e)) = \sum_{j=0}^{n-1} \left(\frac{\partial h}{\partial e_j} \right) (e, \varphi_i(e)) \cdot \delta_{j,k} + \frac{\partial h}{\partial t}(e, \varphi_i(e)) \cdot \frac{\partial \varphi_i}{\partial e_k}(e) \\ &= (-1)^k \varphi_i(e)^{n-k} \cdot 1 + \frac{\partial h}{\partial t}(e, \varphi_i(e)) \cdot \frac{\partial \varphi_i}{\partial e_k}(e), \end{aligned} \quad (11)$$

and rearrangement yields the desired equality, since $\frac{\partial h}{\partial t}(e, \varphi_i(e)) = \prod_{\substack{j=1 \\ j \neq i}}^n (\varphi_i(e) - \varphi_j(e))$. \blacksquare

After this preliminary work, we can formulate condition 1.) in our context:

Conjecture 2.6. *Let $e = (e_1, \dots, e_n) \in \mathbb{R}_+^n$. Then the composition $f \circ \varphi$ is differentiable and the partial derivatives $\frac{\partial(f \circ \varphi)}{\partial e_k}(e)$ for $k \in \{1, \dots, n-1\}$ are positive, i.e.*

$$\frac{\partial(f \circ \varphi)}{\partial e_k}(e) > 0 \quad \text{for all } e \in \mathbb{R}_+^n \text{ and any } k \in \{1, \dots, n-1\}.$$

We will prove Conjecture 2.6 under the additional assumption that all components of $\varphi(e)$ are *pairwise different* (i.e. that h_e has only simple roots). With this restriction, the assertion does not imply condition 1.), but we will compensate this drawback with some additional work in Proposition 2.9 as well as Lemmas 2.10 and 2.11.

Lemma 2.7. *Let $e = (e_1, \dots, e_n) \in \mathbb{R}_+^n$ be such that all components of $\varphi(e)$ are pairwise different. Then*

$$\frac{\partial(f \circ \varphi)}{\partial e_k}(e) = -2 \sum_{i=1}^n \frac{(-\varphi_i(e))^{n-k-1}}{\prod_{\substack{j=1 \\ j \neq i}}^n (\varphi_j(e) - \varphi_i(e))} \log \varphi(e). \quad (12)$$

Proof. Because the components of $\varphi(e)$ are pairwise different, φ is differentiable at e according to Proposition 2.5. Then $f \circ \varphi$ is differentiable at e as well, and we can determine the partial derivatives for $k \in \{1, \dots, n\}$ using the chain rule:

$$\begin{aligned} \frac{\partial(f \circ \varphi)}{\partial e_k}(e) &= \sum_{i=1}^n \frac{\partial f}{\partial z_i}(\varphi_1(e), \dots, \varphi_n(e)) \frac{\partial \varphi_i}{\partial e_k}(e_1, \dots, e_n) \\ &= \sum_{i=1}^n \frac{2 \log \varphi_i(e)}{\varphi_i(e)} \frac{(-1)^{k+1} \varphi_i(e)^{n-k}}{\prod_{\substack{j=1 \\ j \neq i}}^n (\varphi_i(e) - \varphi_j(e))} \\ &= 2 \sum_{i=1}^n \frac{(-1)^{n+k} \varphi_i(e)^{n-k-1}}{(-1)^{n-1} \prod_{\substack{j=1 \\ j \neq i}}^n (\varphi_i(e) - \varphi_j(e))} \log \varphi_i(e) \\ &= -2 \sum_{i=1}^n \frac{(-\varphi_i(e))^{n-k-1}}{\prod_{\substack{j=1 \\ j \neq i}}^n (\varphi_j(e) - \varphi_i(e))} \log \varphi_i(e). \quad \blacksquare \end{aligned} \quad (13)$$

We now want to show that the partial derivatives of $f \circ \varphi$ are positive. The expression calculated in Lemma 2.7 also appeared in other work. For example, Mitchison and Josza in 2004 point out in the appendix of [19] that

$$(-1)^q \sum_{i=1}^n \frac{z_i^{n-q} \log z_i}{\prod_{\substack{j=1 \\ j \neq i}}^n (z_j - z_i)} \geq 0 \quad (14)$$

for all $q \in 2, \dots, n$. If we let $z_i = \varphi_i(e)$, we seem to have already reached our goal. However, Mitchison und Josza prove inequality (14) only for the case that z_1, \dots, z_n are the eigenvalues of a Gram matrix, which is necessarily symmetric and positive definite. Their inequality can therefore only be applied to positive real numbers z_1, \dots, z_n , while in our case, $\varphi_1(e), \dots, \varphi_n(e)$ might also include pairs of complex conjugates.

We close this gap by showing:

Lemma 2.8. *Let $(z_1, \dots, z_n) \in M$. Then for all $r \in \{0, \dots, n-2\}$,*

$$-\sum_{i=1}^n \frac{(-z_i)^r}{\prod_{\substack{j=1 \\ j \neq i}}^n (z_j - z_i)} \log z_i = \int_0^\infty \frac{t^r}{\prod_{j=1}^n (t + z_j)} dt > 0.$$

Proof. We observe that

$$\prod_{j=1}^n (t + z_j) = t^n + \sum_{k=2}^{n-1} e_k(z) t^{n-k} + \prod_{j=1}^n z_j \geq \min \left\{ t^n, \prod_{j=1}^n z_j \right\} > 0$$

for all $t \in \mathbb{R}_+$. Therefore, since $r \leq n - 2$,

$$\int_0^\infty \left| \frac{t^r}{\prod_{j=1}^n (t + z_j)} \right| dt \leq \int_0^1 \frac{t^r}{\prod_{j=1}^n z_j} dt + \int_1^\infty \underbrace{t^{r-n}}_{\leq t^{-2}} dt \leq \frac{1}{\prod_{j=1}^n z_j} + 1. \quad (15)$$

Thus the integral converges for $r \in \{0, \dots, n - 2\}$ and, since $\frac{t^r}{\prod_{j=1}^n (t + z_j)} > 0$, its value is positive.

We now use (2) and exchange the order of summation and integration, where in evaluating the integral at its ‘‘upper boundary’’ we use that $\log(t + z_i) = \log t + \log(1 + \frac{z_i}{t})$, whereas this decomposition is not applied at $t = 0$:

$$\begin{aligned} \int_0^\infty \frac{t^r}{\prod_{j=1}^n (t + z_j)} dt &\stackrel{(2)}{=} \int_0^\infty \sum_{i=1}^n \frac{(-z_i)^r}{(t + z_i) \prod_{\substack{j=1 \\ j \neq i}}^n ((-z_i) - (z_j))} dt \\ &= \sum_{i=1}^n \frac{(-z_i)^r}{\prod_{\substack{j=1 \\ j \neq i}}^n (z_j - z_i)} \cdot \underbrace{\log(t + z_i)}_{=\log t + \log(1 + \frac{z_i}{t})} \Big|_{t=0}^{t \rightarrow \infty} \\ &= \lim_{t \rightarrow \infty} \left(\sum_{i=1}^n \frac{(-z_i)^r}{\prod_{\substack{j=1 \\ j \neq i}}^n (z_j - z_i)} \log t + \sum_{i=1}^n \frac{(-z_i)^r}{\prod_{\substack{j=1 \\ j \neq i}}^n (z_j - z_i)} \left(\log \left(1 + \frac{z_i}{t} \right) \right) - \sum_{i=1}^n \frac{(-z_i)^r}{\prod_{\substack{j=1 \\ j \neq i}}^n (z_j - z_i)} \log z_i \right). \end{aligned} \quad (16)$$

It is easy to see that $\lim_{t \rightarrow \infty} \log \left(1 + \frac{z_i}{t} \right) = 0$.

This immediately implies that $\sum_{i=1}^n \frac{(-z_i)^r}{\prod_{\substack{j=1 \\ j \neq i}}^n (z_j - z_i)} = 0$, because otherwise $\lim_{t \rightarrow \infty} \sum_{i=1}^n \frac{(-z_i)^r}{\prod_{\substack{j=1 \\ j \neq i}}^n (z_j - z_i)} \log t$ would diverge, in contradiction to the already established convergence of the integral. This completes the proof. \blacksquare

Proof of Conjecture 2.6 in the case of pairwise different roots $\varphi_i(e)$.

We can now conclude: for all $n - k - 1 \in \{0, \dots, n - 2\}$, that is for $k \in \{1, \dots, n - 1\}$,

$$\begin{aligned} \frac{\partial(f \circ \varphi)}{\partial e_k}(e) &\stackrel{\text{Prop. 2.7}}{=} -2 \sum_{i=1}^n \frac{(-\varphi_i(e))^{n-k-1}}{\prod_{\substack{j=1 \\ j \neq i}}^n (\varphi_j(e) - \varphi_i(e))} \log \varphi(e) \\ &\stackrel{\text{Prop. 2.8}}{=} 2 \int_0^\infty \frac{t^{n-k-1}}{\prod_{j=1}^n (t + \varphi_j(e))} dt > 0. \end{aligned} \quad \blacksquare$$

This shows that condition 1.) holds on nearly the entire domain T ; the only problems occur for those points $e \in T$ for which h_e has multiple roots. The set T is indeed convex, meaning that we can easily construct a path as described in condition 2.), but since such a path may pass through points with multiple roots, it is not necessarily differentiable everywhere. However, as the next proposition shows, under suitable assumptions a straight line in T contains at most finitely many of these problematic points:

Proposition 2.9. *Let p_0 and p_1 be polynomials of degree n such that at least one of them has only simple roots. Then there are only finitely many $s \in [0, 1]$ such that the polynomial $p_s := (1 - s)p_0 + sp_1$ has multiple roots.*

Proof. Let $p = \alpha \prod_{i=1}^n (X - x_i) = \sum_{i=0}^n a_i X^{n-i}$ be a polynomial of degree n with roots z_1, \dots, z_n and coefficients a_0, \dots, a_n . The discriminant of p is defined as

$$D(p) := \prod_{1 \leq i < j \leq n} (z_i - z_j)^2 \quad (17)$$

and is zero if and only if p has multiple roots (see also [17, p. 204]). The discriminant is a symmetric homogeneous polynomial in the variables z_1, \dots, z_n and thus can be expressed as a homogeneous polynomial in $e_k(z_1, \dots, z_n)$, i.e. in the coefficients a_k .¹

The coefficients of p_s are polynomials in s , so $D(s) := D(p_s)$ is also a polynomial in s . If $D(s)$ is the zero polynomial, then both p_0 and p_1 must have multiple roots. On the other hand, if both p_0 and p_1 do not have multiple roots, then $D(s)$ is not the zero polynomial and is zero only for finitely many $s \in [0, 1]$, and thus p_s can have multiple roots only for finitely many $s \in [0, 1]$. ■

For vectors $x, y \in \mathbb{R}_+^n$, we can therefore directly prove the SSLI using conditions (1) and (2) only under the additional assumption that at least one of the two vectors has pairwise different components. In order to circumvent this limitation, we first show a strict version of the SSLI:

Lemma 2.10. *Let $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n) \in \mathbb{R}_+^n$ be such that $x_1 > x_2 > \dots > x_n$, $y_1 \geq y_2 \geq \dots \geq y_n$, $e_k(x) < e_k(y)$ for all $k \in \{1, \dots, n-1\}$ and $e_n(x) = e_n(y)$. Then $f(x) < f(y)$ with f as in (10), i.e.*

$$f(x) = \sum_{i=1}^n (\log x_i)^2 < \sum_{i=1}^n (\log y_i)^2 = f(y).$$

Proof. Consider the path $e^s = (e_1^s, \dots, e_n^s) \subseteq \mathbb{R}_+^n$ for $s \in [0, 1]$ with

$$e_k^s := (1 - s)e_k(x) + se_k(y).$$

Then $e^0 = e(x)$ and $e^1 = e(y)$ as well as $e_k(x) < e_k(y)$ for all $k \in \{1, \dots, n-1\}$ and $e_n(x) = e_n(y)$. Since $x_1 > x_2 > \dots > x_n$, the polynomial h_{e^0} has pairwise different roots. According to Proposition 2.9, there is only a finite number of $s \in [0, 1]$ such that h_{e^s} has pairwise different roots. Then by Conjecture 2.6, $f \circ \varphi$ is differentiable at all but finitely many points along the path $s \mapsto e_s$, and

$$\frac{d}{ds} f(\varphi(e_s)) = \sum_{k=1}^n \frac{\partial(f \circ \varphi)}{\partial e_k}(e^s) \cdot \frac{d}{ds} e_k^s = \sum_{k=1}^n \underbrace{\frac{\partial(f \circ \varphi)}{\partial e_k}(e^s)}_{>0} \cdot \underbrace{(e_k(y) - e_k(x))}_{>0} > 0. \quad (18)$$

Thus the continuity of the mapping $s \mapsto (f \circ \varphi)(e_s)$ implies its strict monotonicity on $[0, 1]$, and therefore

$$f(x) = (f \circ \varphi)(e(x)) = f(\varphi(e_0)) < f(\varphi(e_1)) = (f \circ \varphi)(e(y)) = f(y). \quad \blacksquare$$

¹For example, the polynomial $q(t) = t^3 - at^2 + bt - c$ has the discriminant $D(q) = a^2 b^2 - 4b^3 - 4a^3 c + 18abc - 27c^2$.

In the next step, we use the continuity of the elementary symmetric polynomials in order to show the strict inequality for those x, y whose components are not pairwise different.

Lemma 2.11. *Let $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n) \in \mathbb{R}_+^n$ be such that $e_k(x) < e_k(y)$ for all $k \in \{1, \dots, n-1\}$ and $e_n(x) = e_n(y)$. Then $f(x) < f(y)$ with f as in (10), i.e.*

$$f(x) = \sum_{i=1}^n (\log x_i)^2 < \sum_{i=1}^n (\log y_i)^2 = f(y).$$

Proof. If the components of $x = (x_1, \dots, x_n)$ are pairwise different, then we can assume $x_1 > x_2 > \dots > x_n$ and $y_1 \geq y_2 \geq \dots \geq y_n$ after rearrangement. In this case, Lemma 2.10 shows that the inequality holds.

Otherwise, we need to slightly change the identical component pairs of x in order to apply Lemma 2.10. Because of the continuity of the elementary symmetric polynomials, if the changes to the components of x are sufficiently small, then the resulting vector x' still satisfies the inequality $e_k(x') < e_k(y)$ for all $k \in \{1, \dots, n-1\}$. In order to preserve the equality $e_n(x') = e_n(y)$, we choose $x'_i := x_i(1 + \varepsilon)$ and $x'_j := x_j \frac{1}{1 + \varepsilon}$ for each component pair $x_i = x_j$ and small $\varepsilon > 0$. So we find

$$\begin{aligned} (\log x'_i)^2 + (\log x'_j)^2 &= (\log x_i + \log(1 + \varepsilon))^2 + (\log x_j - \log(1 + \varepsilon))^2 \\ &= 2(\log x_i)^2 + 2(\log(1 + \varepsilon))^2 > (\log x_i)^2 + (\log x_j)^2. \end{aligned} \quad (19)$$

If we dissolve all pairwise equalities in this way, then we can apply Lemma 2.10 to find

$$\sum_{i=1}^n (\log x_i)^2 < \sum_{i=1}^n (\log x'_i)^2 \stackrel{2.10}{<} \sum_{i=1}^n (\log y_i)^2. \quad \blacksquare$$

Using the continuity of the elementary symmetric polynomials and the logarithmic function we are finally able to extend Lemma 2.11 and thus prove the SSLI without any restrictions:

Theorem 1.2 (Sum of squared logarithms inequality). *Let $n \in \mathbb{N}$ and $x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n \in \mathbb{R}_+$ such that*

$$e_k(x) \leq e_k(y) \quad \text{for all } k \in \{1, \dots, n-1\}$$

and

$$e_n(x) = e_n(y).$$

Then

$$\sum_{i=1}^n (\log x_i)^2 \leq \sum_{i=1}^n (\log y_i)^2.$$

Proof. Choose $S \subseteq \{1, \dots, n-1\}$ such that $e_k(x) = e_k(y)$ for all $k \in S$ and $e_k(x) < e_k(y)$ for all $k \in \{1, \dots, n-1\} \setminus S$. Furthermore, for $k \in \{1, \dots, n\}$ and $m \in \mathbb{N}$, let

$$e_k^m = \begin{cases} e_k(x) - \frac{1}{m} & \text{if } k \in S, \\ e_k(x) & \text{otherwise} \end{cases} \quad \text{and} \quad x^m := \varphi(e^m).$$

Then $0 < e_k(x^m) < e_k(y)$ for all $k \in \{1, \dots, n-1\}$ and all sufficiently large $m \in \mathbb{N}$ as well as $e_n(x^m) = e_n(y)$, thus Proposition 2.11 yields $\sum_{i=1}^n (\log x_i^m)^2 < \sum_{i=1}^n (\log y_i)^2$ for all sufficiently large $m \in \mathbb{N}$. Since $\lim_{m \rightarrow \infty} x^m = x$, we find

$$\sum_{i=1}^n (\log x_i)^2 = \lim_{m \rightarrow \infty} \sum_{i=1}^n (\log x_i^m)^2 \leq \sum_{i=1}^n (\log y_i)^2. \quad \blacksquare$$

3 Applications and Connections

3.1 Relation to entropy

Jozsa and Mitchison [15] study entropy and “subentropy” from the perspective of quantum information theory. They introduce and investigate partial derivatives of these quantities with respect to e_k , and establish their unrestricted nonnegativity. They view entropy as a function of the symmetric polynomials e_k and use analytic continuation to extend the definition to the entire set of nonnegative real e_k . Consequently, they obtain integral representations of entropy (and subentropy) from which the desired nonnegativity properties follow. More interestingly, they study higher partial derivatives w.r.t. the elementary symmetric polynomials and establish complete monotonicity of entropy and subentropy. Similar higher order monotonicity properties can be established for $f(x) = \sum_{i=1}^n (\log x_i)^2$.

In [10] Dannan, Neff und Thiel discussed applications of the SSLI towards the entropy of probability distributions. We now prove a statement very similar to the entropy expression of two vectors: we repeat our procedure from the last section with

$$g: (\mathbb{C} \setminus \mathbb{R}_{\leq 0})^n \rightarrow \mathbb{C}, \quad g(z_1, \dots, z_n) = - \sum_{i=1}^n z_i \log(-z_i) \quad (20)$$

instead of f , but otherwise identical notation and definitions. The composition $g \circ \varphi$ on \mathbb{R}_+^n is once more a real-valued function: $x \in \mathbb{R}_-$ and $y \in \mathbb{R}_{\neq 0}$ imply $x \log(-x) \in \mathbb{R}$, while for complex conjugate pairs we find

$$x_i \log x_i + \overline{x_i} \log \overline{x_i} = x_i \log x_i + \overline{x_i} \cdot \overline{\log x_i} = x_i \log x_i + \overline{x_i \log x_i} \in \mathbb{R}.$$

Analogously to the last section, the function $g \circ \varphi$ can be expressed as

$$(g \circ \varphi): \mathbb{R}_+^n \rightarrow \mathbb{R}, \quad (f \circ \varphi)(e) = f(\varphi_1(e), \dots, \varphi_n(e)) = - \sum_{i=1}^n \varphi_i(e) \log \varphi_i(e),$$

and for $x_1, \dots, x_n \in \mathbb{R}_+$ we have $(g \circ \varphi)(e_1(x), e_2(x), \dots, e_n(x)) = - \sum_{i=1}^n x_i \log x_i$.

We now determine the partial derivatives:

$$\begin{aligned} \frac{\partial(g \circ \varphi)}{\partial e_k}(e) &= \sum_{i=1}^n \frac{\partial g}{\partial z_i}(\varphi_1(e), \dots, \varphi_n(e)) \frac{\partial \varphi_i}{\partial e_k}(e_1, \dots, e_n) \\ &\stackrel{\text{Prop. 2.5}}{=} \sum_{i=1}^n (\log \varphi_i(e) + 1) \frac{(-1)^{k+1} \varphi_i(e)^{n-k}}{\prod_{\substack{j=1 \\ j \neq i}}^n (\varphi_i(e) - \varphi_j(e))} \\ &= \sum_{i=1}^n \frac{(-1)^{n+k-1} \varphi_i(e)^{n-k}}{(-1)^{n-1} \prod_{\substack{j=1 \\ j \neq i}}^n (\varphi_i(e) - \varphi_j(e))} (\log \varphi_i(e) + 1) \\ &= - \sum_{i=1}^n \frac{(-\varphi_i(e))^{n-k}}{\prod_{\substack{j=1 \\ j \neq i}}^n (\varphi_j(e) - \varphi_i(e))} \log \varphi_i(e) - \sum_{i=1}^n \frac{(-\varphi_i(e))^{n-k}}{\prod_{\substack{j=1 \\ j \neq i}}^n (\varphi_j(e) - \varphi_i(e))}. \end{aligned} \quad (21)$$

By (4), the second sum is zero for $n - k \in \{0, \dots, n - 2\}$, i.e. for $k \in \{2, \dots, n\}$. Thus we obtain

$$\frac{\partial(g \circ \varphi)}{\partial e_k}(e) = - \sum_{i=1}^n \frac{(-\varphi_i(e))^{n-k}}{\prod_{\substack{j=1 \\ j \neq i}}^n (\varphi_j(e) - \varphi_i(e))} \log \varphi_i(e) \stackrel{\text{Lemma 2.8}}{>} 0$$

for $n - k \in \{0, \dots, n - 2\}$ or, equivalently, for $k \in \{2, \dots, n\}$.

Remark 3.1. It should also be noted that $\frac{\partial(f \circ \varphi_i)}{\partial e_k} = 2 \frac{\partial(g \circ \varphi_i)}{\partial e_{k+1}}$.

As with the SSLI, we can now infer the monotonicity of $g \circ \varphi$ along straight lines, which yields the following result.

Corollary 3.2. *Let $n \in \mathbb{N}$ and $x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n > 0$ such that*

$$e_1(x) = e_1(y)$$

and

$$e_k(x) \leq e_k(y) \quad \text{for all } k \in \{2, \dots, n\}.$$

Then

$$-\sum_{i=1}^n x_i \log x_i \leq -\sum_{i=1}^n y_i \log y_i.$$

Weakening the equality in the first condition to $e_k(x) \leq e_k(y)$ for all $k \in \{1, \dots, n\}$ again allows for obvious counterexamples similar to the SSLI: For $x = (e^{-1}, \dots, e^{-1}) \in \mathbb{R}_+^n$ and $y = (1, \dots, 1) \in \mathbb{R}_+^n$ the weakened condition is true, yet $-\sum_{i=1}^n x_i \log x_i = \frac{n}{e} > 0 = -\sum_{i=1}^n y_i \log y_i$.

See also the discussion in Section 3.

3.2 The SSLI in terms of matrix invariants

Let $U \in \text{Sym}_+(n)$, where $\text{Sym}_+(n) \subset \mathbb{R}^{n \times n}$ denotes the set of positive definite symmetric $n \times n$ -matrices. Then U is orthogonally diagonalizable with real eigenvalues $\lambda_1, \dots, \lambda_n > 0$. The k -th invariant $I_k(U)$ of U is defined as the k -th elementary symmetric polynomial of the vector $\lambda(U) = (\lambda_1, \dots, \lambda_n)$, i.e. $I_k(U) := e_k(\lambda(U))$; (thus, $I_1(U) = \text{tr } U$ and $I_n(U) = \det U$).

The SSLI can be equivalently expressed in terms of these invariants of positive definite symmetric matrices.

Theorem 3.3. *Let $U, \tilde{U} \in \text{Sym}_+(n)$. If $I_k(U) \leq I_k(\tilde{U})$ for all $k \in \{1, \dots, n-1\}$ and $\det U = \det \tilde{U}$, then $\|\log U\|^2 \leq \|\log \tilde{U}\|^2$, where \log is the principal matrix logarithm on $\text{Sym}_+(n)$ and $\|\cdot\|$ denotes the Frobenius matrix norm.*

Proof. Since $\|\log U\|^2 = \sum_{i=1}^n (\log \lambda_i(U))^2$, using the SSLI we immediately have

$$\|\log U\|^2 = \sum_{i=1}^n (\log \lambda_i(U))^2 \leq \sum_{i=1}^n (\log \lambda_i(\tilde{U}))^2 = \|\log \tilde{U}\|^2. \quad \blacksquare$$

Theorem 3.3 can be applied directly to the *quadratic Hencky energy*

$$W_H(F) = \mu \|\operatorname{dev}_n \log U\|^2 + \frac{\kappa}{2} [\operatorname{tr}(\log U)]^2 = \mu \|\log U\|^2 + \frac{\lambda}{2} [\log(\det U)]^2,$$

which was introduced into the theory of nonlinear elasticity in 1929 by H. Hencky [13]. Here, $F \in \operatorname{GL}^+(n)$ is the deformation gradient, $\operatorname{GL}^+(n)$ is the set of invertible $n \times n$ -matrices with positive determinant, $U = \sqrt{F^T F}$ is the right stretch tensor and $\operatorname{dev}_n \log U = \log U - \frac{1}{n} \operatorname{tr}(\log U) \cdot \mathbf{1}$ is the deviatoric part of the Hencky strain tensor $\log U$. The material parameters μ, λ with $\mu > 0$ and $3\lambda + 2\mu \geq 0$ are called the Lamé constants, while $\kappa \geq 0$ is known as the bulk modulus. The Hencky energy has recently been characterized by a unique geometric property [27, 24, 28]: it measures the squared geodesic distance of F to the special orthogonal group $\operatorname{SO}(n)$ with respect to a left- $\operatorname{GL}(n)$ -invariant, right- $\operatorname{SO}(n)$ -invariant Riemannian metric on $\operatorname{GL}(n)$.

In terms of the quadratic Hencky energy, Theorem 3.3 can be stated as follows:

Corollary 3.4. *Let $F, \tilde{F} \in \operatorname{GL}^+(n)$ with $U = \sqrt{F^T F}$ and $\tilde{U} = \sqrt{\tilde{F}^T \tilde{F}}$. If $\det U = \det \tilde{U}$ and $I_k(U) \leq I_k(\tilde{U})$ for all $k \in \{1, \dots, n-1\}$, then $W_H(F) \leq W_H(\tilde{F})$.*

Remark 3.5. Corollary 3.4 means that W_H satisfies a version of Truesdell's empirical inequalities [32, pages 158, 171].

In a similar way, Corollary 3.2 can be stated in terms of matrix invariants.

Theorem 3.6. *Let $U, \tilde{U} \in \operatorname{Sym}_+(n)$. If $\operatorname{tr} U = \operatorname{tr} \tilde{U}$ and $I_k(U) \leq I_k(\tilde{U})$ for all $k \in \{2, \dots, n\}$, then $\langle U, \log U - \mathbf{1} \rangle \geq \langle \tilde{U}, \log \tilde{U} - \mathbf{1} \rangle$, where $\langle X, Y \rangle = \operatorname{tr}(X^T Y)$ denotes the canonical inner product of two $n \times n$ -matrices X and Y .*

Proof. Corollary 3.2 implies $\sum_{i=1}^n \lambda_i(U) \log \lambda_i(U) \geq \sum_{i=1}^n \log \lambda_i(\tilde{U}) \log \lambda_i(\tilde{U})$. Using the condition $\operatorname{tr} U = \operatorname{tr} \tilde{U}$, we compute

$$\begin{aligned} \langle U, \log U - \mathbf{1} \rangle &= \langle U, \log U \rangle - \langle U, \mathbf{1} \rangle = \sum_{i=1}^n \lambda_i(U) \cdot \log \lambda_i(U) - \operatorname{tr} U \\ &\geq \sum_{i=1}^n \lambda_i(\tilde{U}) \cdot \log \lambda_i(\tilde{U}) - \operatorname{tr} \tilde{U} = \langle \tilde{U}, \log \tilde{U} - \mathbf{1} \rangle. \quad \blacksquare \end{aligned}$$

Remark 3.7. The constitutive law induced by the hyperelastic energy potential

$$W_B(F) = \langle U, \log U - \mathbf{1} \rangle = \langle \sqrt{F^T F}, \log \sqrt{F^T F} - \mathbf{1} \rangle$$

is a special case of *Becker's law* of elasticity, which was introduced by the geologist G.F. Becker in 1893 [3] in a way remarkably similar to H. Hencky's deduction of the quadratic Hencky energy [25]. Becker's elastic law is hyperelastic (i.e. admits an energy potential) only in the lateral contraction free case, which is described by the energy function W_B .

Since $\operatorname{tr}(X \log X) = \sum_{i=1}^n \lambda_i(X) \log \lambda_i(X)$, we can translate the last Theorem 3.6 into a statement for the quantum von Neumann entropy.

Corollary 3.8. *Let $X, Y \in \operatorname{Sym}_+(n)$ be density matrices, so that $\operatorname{tr} X = \operatorname{tr} Y = 1$. If $I_k(X) \leq I_k(Y)$ for all $k \in \{2, \dots, n\}$, then the von Neumann entropy $-\operatorname{tr}(X \log X) \leq -\operatorname{tr}(Y \log Y)$.*

3.3 Application to geodesic distance on $\text{Sym}_+(n)$

The convex cone $\text{Sym}_+(n)$ is frequently also viewed as a *Riemannian manifold* endowed with the Riemannian metric [6, 11, 20, 21, 30]

$$g_C(X, Y) = \text{tr}(C^{-1}XC^{-1}Y), \quad (22)$$

where $C \in \text{Sym}_+(n)$ and $X, Y \in \text{Sym}(n) = T_C \text{Sym}_+(n)$, the tangent space at C . This manifold is geodesically complete, and the unique geodesic joining $C_1, C_2 \in \text{Sym}_+(n)$ has the closed form [5, 6.1.6]

$$\gamma(t) = C_1^{1/2}(C_1^{-1/2}C_2C_1^{-1/2})^t C_1^{1/2}, \quad t \in [0, 1]. \quad (23)$$

The *geodesic distance* between $C_1, C_2 \in \text{Sym}_+(n)$ is given by [5, 6.1.6]

$$\text{dist}_{\text{geod}, \text{Sym}_+(n)}(C_1, C_2) = \|\log(C_2^{-1/2} C_1 C_2^{-1/2})\|.$$

In particular, for $C_2 = \mathbb{1}$, we obtain the simple formula

$$\text{dist}_{\text{geod}, \text{Sym}_+(n)}^2(C_1, \mathbb{1}) = \|\log C_1\|^2.$$

The sum-of-squared-logarithm inequalities can therefore be stated in terms of the geodesic distance of positive definite symmetric matrices to the identity matrix $\mathbb{1}$.

Corollary 3.9. *Let $I_1(C), \dots, I_n(C), I_1(\tilde{C}), \dots, I_n(\tilde{C})$ denote the principal invariants of $C, \tilde{C} \in \text{Sym}_+(n)$. If $I_k(C) \leq I_k(\tilde{C})$ for all $k \in \{1, \dots, n-1\}$ and $I_n(C) = \det C = \det \tilde{C} = I_n(\tilde{C})$, then*

$$\text{dist}_{\text{geod}, \text{Sym}_+(n)}(C, \mathbb{1}) \leq \text{dist}_{\text{geod}, \text{Sym}_+(n)}(\tilde{C}, \mathbb{1}).$$

Since $\mathbb{1}$ commutes with every matrix, $\text{dist}_{\text{geod}, \text{Sym}_+(n)}$ actually reduces to the *log-Euclidean distance*

$$\text{dist}_{\text{log-Euclid}}(C, \tilde{C}) := \|\log C - \log \tilde{C}\|,$$

i.e., the Euclidean distance between the principal logarithms of two matrices [2]. Thus, the SSLI may be equivalently stated for $\text{dist}_{\text{log-Euclid}}$.

3.4 Additional applications

The SSLI may also find applications in other fields. In the following we list some of these potential applications.

- Recall that $I_k(X) = \text{tr} \wedge^k X$, where \wedge denotes the antisymmetric tensor product (Grassmann product). Thus, the SSLI for matrices can be equivalently stated in terms of this tensor product. This notation immediately suggests an interesting generalization, namely to the dual case of *symmetric tensor product* denoted $\text{tr} \vee^k X = h_k(\lambda(X))$, where h_k denotes the k -th complete symmetric polynomial. Thus, we can consider inequalities of the form

$$\begin{aligned} \text{tr} \vee^k X \leq \text{tr} \vee^k Y \quad \text{for all } k \in \{1, \dots, n-1\} \quad \text{and} \quad \text{tr} \vee^n X = \text{tr} \vee^n Y \\ \implies F(X) \leq F(Y). \end{aligned}$$

More generally, extensions of the SSLI to other matrix monotone functions may be obtained by building on [31].

- It might also be of interest to characterize the matrix functions $\mathcal{F}: \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n}$ for which

$$\inf_{Q \in \text{SO}(n)} \|\text{sym } \mathcal{F}(Q^T Z)\| = \inf_{Q \in \text{SO}(n)} \|\mathcal{F}(Q^T Z)\| = \|\mathcal{F}(R^T Z)\| = \|\mathcal{F}(H)\|,$$

where $Z = RH$ is the polar decomposition of Z , see [18, 26] and compare to (1).

- The final connection that we mention is perhaps the most interesting. Nonnegative elementary symmetric polynomials of matrices have been studied under the guise of Q -matrices (Q_0 -matrices) [14]. These are complex matrices, whose elementary symmetric polynomials are positive (nonnegative). These matrices are much more tractable than the better known class of P -matrices (i.e. matrices with positive principal minors), which have been extensively studied in matrix analysis and optimization [12, 4, 8]. The following theorem of Kellogg applies to P matrices, and also to Q matrices [14]:

Theorem 3.10 ([16]). *Let $X \in \mathbb{C}^{n \times n}$ be a matrix for which $I_k(X) \geq 0$ for $k \in \{1, \dots, n\}$. Then, the spectrum of A is contained in the set*

$$D := \left\{ z : |\arg z| \leq \pi - \frac{\pi}{n} \right\}.$$

Moreover, if any eigenvalue of X lies on the boundary of D , then necessarily all symmetric functions $I_k(X)$, except $I_n(X)$, are equal to zero.

4 Alternative proof of Conjecture 2.6

In Section 2, we have only shown Conjecture 2.6 under the additional assumption that $e \in \mathbb{R}_+^n$ corresponds to pairwise different roots $\varphi(e)$ which was sufficient to prove the SSLI. Nevertheless, it is striking that the expression of the partial derivatives

$$\frac{\partial(f \circ \varphi)}{\partial e_k}(e) = 2 \int_0^\infty \frac{t^{n-k-1}}{\prod_{j=1}^n (t + \varphi_j(e))} dt \quad (24)$$

can be extended continuously to $e \in \mathbb{R}_+^n$ with multiple roots. This strongly suggests that Conjecture 2.6 indeed holds for all $e \in \mathbb{R}_+^n$. We now restate the conjecture as a lemma and give a short proof for the general case.

Lemma 4.1. *For $a = (a_1, \dots, a_{n-1}) \in \mathbb{R}_+^{n-1}$ let $\hat{h}_a(z) = z^n + a_{n-1}z^{n-1} + \dots + a_1z + 1$. Define the function $\hat{f}: \mathbb{R}_+^{n-1} \rightarrow \mathbb{R}$ by*

$$\hat{f}(a_1, \dots, a_{n-1}) = \sum_{\substack{\hat{z} \in \mathbb{C} \\ \hat{h}_a(\hat{z})=0}} (\log(-\hat{z}))^2,$$

where \log denotes the principal branch of the complex logarithm on $\mathbb{C} \setminus \mathbb{R}_{\leq 0}$ and roots are counted with multiplicities. Then all partial derivatives $\frac{\partial \hat{f}}{\partial a_k}(a)$ for any $a \in \mathbb{R}_+^{n-1}$ are positive.

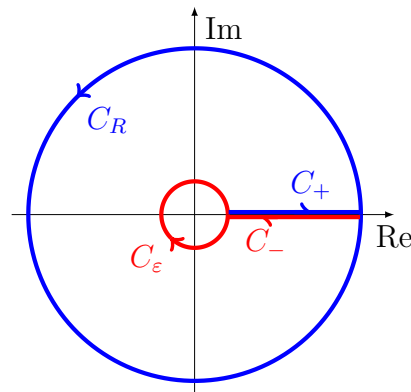


Figure 2: The path C consists of the following four pieces C_e , C_+ , C_R and C_- .

Note carefully that the roots z_1, \dots, z_n of h_a in the original formulation of Conjecture 2.6 correspond to the negative $-\widehat{z}_1, \dots, -\widehat{z}_n$ of the roots of \widehat{h}_a in Lemma 4.1.

The function \widehat{f} is well defined, since none of the roots of \widehat{h}_a are positive real numbers. Furthermore, \widehat{f} is real-valued, because the roots \widehat{z} of \widehat{h}_a only occur as real numbers or complex conjugate pairs, which means that the values of $(\log(-\widehat{z}))^2$ are also real or complex conjugate pairs.

Proof. We write the function \widehat{f} as a contour integral. In order to accommodate the multiple values of $\log(-z)$ we consider the Riemann surface with boundary obtained by gluing two copies of \mathbb{R}_+ to $\mathbb{C} \setminus \mathbb{R}_{\geq 0}$, where $\mathbb{R}_{\geq 0} := \{x \in \mathbb{R} \mid x \geq 0\}$. We will view the two copies as *north* and *south* copies \mathbb{R}_+ depending on whether they are approached from above and from below respectively. Note that the $\log(-z)$ function extends to the following

$$\log(-r_{\text{north}}) = \log r - \pi i, \quad \log(-r_{\text{south}}) = \log r + \pi i$$

on the north and south copies of \mathbb{R}_+ .

In a neighborhood of fixed $a \in (\mathbb{R}_{\geq 0})^{n-1}$, for any sufficiently large R and small enough $\varepsilon > 0$, we can apply the generalized argument principle [1] to find²

$$\widehat{f}(a) = \frac{1}{2\pi i} \int_C (\log(-z))^2 \frac{\widehat{h}'_a(z)}{\widehat{h}_a(z)} dz, \quad (25)$$

where the path C consists of the following four pieces C_ε , C_+ , C_R and C_- , see also Figure 2:

- C_ε is the circle of radius ε around 0 travelled clockwise from $\varepsilon_{\text{south}}$ to $\varepsilon_{\text{north}}$;
- C_+ is the line segment from $\varepsilon_{\text{north}}$ to R_{north} on the north copy of the positive real axis;
- C_R is the circle of radius R around 0 travelled counter-clockwise from R_{north} to R_{south} ;
- C_- is the line segment from R_{south} to $\varepsilon_{\text{south}}$ on the south copy of the positive axis.

By straightforward computation, we find

$$\begin{aligned} \frac{\partial}{\partial a_k} \frac{\widehat{h}'_a(z)}{\widehat{h}_a(z)} &= \frac{\frac{\partial}{\partial a_k} \widehat{h}'_a(z) \cdot \widehat{h}_a(z) - \widehat{h}'_a(z) \cdot \frac{\partial}{\partial a_k} \widehat{h}_a(z)}{(\widehat{h}_a(z))^2} \\ &= \frac{k z^k \cdot \widehat{h}_a(z) - z^k \cdot \widehat{h}'_a(z)}{(\widehat{h}_a(z))^2} = \left(\frac{z^k}{\widehat{h}_a(z)} \right)'. \end{aligned} \quad (26)$$

The integrand is analytic on the compact path C , hence we can differentiate inside the integral:

$$\begin{aligned} \frac{\partial \widehat{f}(a)}{\partial a_k} &= \frac{1}{2\pi i} \int_C (\log(-z))^2 \frac{\partial}{\partial a_k} \frac{\widehat{h}'_a(z)}{\widehat{h}_a(z)} dz \stackrel{(26)}{=} \frac{1}{2\pi i} \int_C (\log(-z))^2 \left(\frac{z^k}{\widehat{h}_a(z)} \right)' dz \\ &= -\frac{1}{2\pi i} \int_C \left((\log(-z))^2 \right)' \frac{z^k}{\widehat{h}_a(z)} dz = -\frac{1}{\pi i} \int_C \log(-z) \frac{z^{k-1}}{\widehat{h}_a(z)} dz. \end{aligned} \quad (27)$$

²A related representation to (25) has been used recently in [15] in connection with the entropy function.

We can now take the limits as $R \rightarrow +\infty$ and $\varepsilon \rightarrow 0$. Since $k \leq n - 1$, the integral over C_R tends to zero (the length of C_R is $2\pi R$, while the function is of order $O(R^{-2} \log R)$). The integral over C_ε also tends to zero, since for $k \geq 1$, the integrand is of order $O(\log \varepsilon)$ and the length of C_ε is $2\pi\varepsilon$.

We conclude:

$$\begin{aligned}
\frac{\partial f(a)}{\partial a_k} &= -\frac{1}{\pi i} \lim_{\varepsilon \searrow 0, R \rightarrow \infty} \left(\int_{C_+ \cup C_-} \log(-z) \frac{z^{k-1}}{\widehat{h}_a(z)} dz \right) \\
&= -\frac{1}{\pi i} \lim_{\varepsilon \searrow 0, R \rightarrow \infty} \left(\int_{[\varepsilon, R]} ((\log t - \pi i) - (\log t + \pi i)) \frac{t^{k-1}}{\widehat{h}_a(t)} dt \right) \\
&= 2 \lim_{\varepsilon \searrow 0, R \rightarrow \infty} \int_\varepsilon^R \frac{t^{k-1}}{\widehat{h}_a(t)} dt = 2 \int_0^{+\infty} \frac{t^{k-1}}{\widehat{h}_a(t)} dt > 0. \quad \blacksquare
\end{aligned} \tag{28}$$

Acknowledgement

L.B. was partially supported by NSF grant DMS-1201466. We are grateful to Johannes Lankeit (Universität Paderborn) for fruitful discussions on the SSLI. We also want to thank Prof. Georg Hein (Universität Duisburg-Essen) for his suggestions on algebraic topics. Our special thanks go out to Robert Martin (Universität Duisburg-Essen), who is always a great support with his commitment and considerable technical competence.

References

- [1] H. Ammari, H. Kang, and H. Lee. *Layer potential techniques in spectral analysis*, volume 153. American Mathematical Society Providence, 2009. chapter 1 available at www.ams.org/bookstore/pspdf/surv-153-prev.pdf.
- [2] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 29(1):328–347, 2007.
- [3] G. Becker. The Finite Elastic Stress-Strain Function. *American Journal of Science*, 46:337–356, 1893. newly typeset version available at www.uni-due.de/imperia/md/content/mathematik/ag_neff/becker_latex_new1893.pdf.
- [4] A. Berman and R. J. Plemmons. Nonnegative matrices. *The Mathematical Sciences, Classics in Applied Mathematics*, 9, 1979.
- [5] R. Bhatia. *Positive definite matrices*. Princeton University Press, 2009.
- [6] R. Bhatia and J. Holbrook. Riemannian geometry and matrix geometric means. *Linear Algebra and its Applications*, 413(2):594–618, 2006.
- [7] M. Birsan, P. Neff, and J. Lankeit. Sum of squared logarithms – an inequality relating positive definite matrices and their matrix logarithm. *Journal of Inequalities and Applications*, 2013(168):1–16, 2013. open access.
- [8] G. E. Coxson. The P-matrix problem is co-NP-complete. *Mathematical Programming*, 64(1-3):173–178, 1994.
- [9] F. Cucker and A. G. Corbalan. An alternate proof of the continuity of the roots of a polynomial. *American Mathematical Monthly*, pages 342–345, 1989.

- [10] F. Dannan, P. Neff, and C. Thiel. On the sum of squared logarithms inequality and related inequalities. *Journal of Mathematical Inequalities*, 2015. open access, available at arXiv:1411.1290.
- [11] Z. Fiala. Geometrical setting of solid mechanics. *Annals of Physics*, 326(8):1983–1997, 2011.
- [12] M. Fiedler and V. Ptak. On matrices with non-positive off-diagonal elements and positive principal minors. *Czechoslovak Mathematical Journal*, 12(3):382–400, 1962.
- [13] H. Hencky. Welche Umstände bedingen die Verfestigung bei der bildsamen Verformung von festen isotropen Körpern? *Zeitschrift für Physik*, 55:145–155, 1929. available at www.uni-due.de/imperia/md/content/mathematik/ag_neff/hencky1929.pdf.
- [14] D. Hershkowitz. On the spectra of matrices having nonnegative sums of principal minors. *Linear Algebra and its Applications*, 55:81–86, 1983.
- [15] R. Jozsa and G. Mitchison. Symmetric polynomials in information theory: Entropy and subentropy. *Journal of Mathematical Physics*, 56(6):062201, 2015.
- [16] R. Kellogg. On complex eigenvalues of M and P matrices. *Numerische Mathematik*, 19(2):170–175, 1972.
- [17] S. Lang. *Algebra*. Springer, revised third edition, 2002.
- [18] J. Lankeit, P. Neff, and Y. Nakatsukasa. The minimization of matrix logarithms: On a fundamental property of the unitary polar factor. *Linear Algebra and its Applications*, 449:28–42, 2014.
- [19] G. Mitchison and R. Jozsa. Towards a geometrical interpretation of quantum-information compression. *Physical Review A*, 69(3):032304, 2004.
- [20] M. Moakher. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 26(3):735–747, 2005.
- [21] M. Moakher and A. N. Norris. The closest elastic tensor of arbitrary symmetry to an elasticity tensor of lower symmetry. *Journal of Elasticity*, 85(3):215–263, 2006.
- [22] P. Neff. The sum of squared logarithms conjecture, May 2015. www.mathoverflow.net/questions/207845/the-sum-of-squared-logarithms-conjecture.
- [23] P. Neff, B. Eidel, F. Osterbrink, and R. Martin. The Hencky strain energy $\|\log U\|^2$ measures the geodesic distance of the deformation gradient to $SO(n)$ in the canonical left-invariant Riemannian metric on $GL(n)$. *Proceedings in Applied Mathematics and Mechanics*, 13(1):369–370, 2013.
- [24] P. Neff, B. Eidel, F. Osterbrink, and R. J. Martin. A Riemannian approach to strain measures in nonlinear elasticity. *Comptes Rendus Mécanique*, 342(4):254–257, 2014.
- [25] P. Neff, I. Münch, and R. Martin. Rediscovering G.F. Becker’s early axiomatic deduction of a multiaxial nonlinear stress-strain relation based on logarithmic strain. *to appear in Mathematics and Mechanics of Solids*, doi: 10.1177/1081286514542296, 2014. doi: 10.1177/1081286514542296. available at arXiv:1403.4675.
- [26] P. Neff, Y. Nakatsukasa, and A. Fischle. A logarithmic minimization property of the unitary polar factor in the spectral norm and the Frobenius matrix norm. *SIAM Journal on Matrix Analysis and Applications*, 35(3):1132–1154, 2014.
- [27] P. Neff, B. Eidel, and R. J. Martin. Geometry of logarithmic strain measures in solid mechanics. *arXiv preprint arXiv:1505.02203*, submitted, 2015.
- [28] P. Neff, I. D. Ghiba, and J. Lankeit. The exponentiated Hencky-logarithmic strain energy. Part I: Constitutive issues and rank-one convexity. *to appear in Journal of Elasticity*, 2015. available at arXiv:1403.3843.
- [29] W. Pompe and P. Neff. On the generalized sum of squared logarithms inequality. *Journal of Inequalities and Applications*, 2015. open access, available at arXiv:1410.2706.

- [30] P. Rougée. An intrinsic Lagrangian statement of constitutive laws in large strain. *Computers and Structures*, 84(17):1125–1133, 2006.
- [31] M. Šilhavý. A functional inequality related to analytic continuation. *preprint Institute of Mathematics AS CR IM-2015-37*, 2015. available at www.math.cas.cz/fichier/preprints/IM_20150623102729_44.pdf.
- [32] C. Truesdell and W. Noll. *The non-linear field theories of mechanics*. Springer, 2004. originally published as Volume III/3 of the Encyclopedia of Physics in 1965.