The sum of squared logarithms inequality in arbitrary dimensions

by

Lev Borisov, Patrizio Neff, Suvrit Sra and Christian Thiel

SM-UDE-788 2015
Eingegangen am 18.08.2015
The sum of squared logarithms inequality in arbitrary dimensions

Lev Borisov\textsuperscript{1} and Patrizio Neff\textsuperscript{2} and Suvrit Sra\textsuperscript{3} and Christian Thiel\textsuperscript{4}

August 18, 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 $f: M \subseteq \mathbb{C}^n \to \mathbb{R}$ 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 \hfill 2

2 Proof of Theorem 1.2 \hfill 3

3 Applications and Connections \hfill 11

\hspace{1cm} 3.1 Relation to entropy \hfill 11

\hspace{1cm} 3.2 The SSLI in terms of matrix invariants \hfill 13

\hspace{1cm} 3.3 Application to geodesic distance on $\text{Sym}_+(n)$ \hfill 14

\hspace{1cm} 3.4 Additional applications \hfill 15

4 Alternative proof of Conjecture 2.6 \hfill 16

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

\textsuperscript{2}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

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

\textsuperscript{4}Corresponding 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 [25] while proving the following optimality result

\[ \inf_{Q \in \text{SO}(n)} \| \text{sym} \log Q^T F \|^2 = \inf_{Q \in \text{SO}(n)} \inf_{\exp(Y) = Q^T F} \| \text{sym} Y \|^2 = \| \log \sqrt{F^T F} \|^2, \]

where \( Y = \log X \) 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 Bárás, Neff and Lankeit in [7], who found a proof for \( n = \{2, 3\} \). 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

\[
\begin{align*}
x_1 + x_2 + x_3 & \leq y_1 + y_2 + y_3, \\
x_1 x_2 + x_1 x_3 + x_2 x_3 & \leq y_1 y_2 + y_1 y_3 + y_2 y_3, \\
x_1 x_2 x_3 & = y_1 y_2 y_3.
\end{align*}
\]

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 [31] 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 \( x \):

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

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

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

\[
e_k(x) \leq e_k(y) \quad \text{for all } k \in \{1, \ldots, n-1\}
\]

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

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

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 \(i \in \{1, \ldots, n\}\), the expression is bounded below by 0, so the infimum clearly exists. Note that \(E_x\) 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 \(x = (e^{-1}, \ldots, e^{-1}) \in \mathbb{R}^n\) and \(y = (1, \ldots, 1) \in \mathbb{R}^n\); then \(e_k(x) = \binom{n}{k} e^{-k} \leq \binom{n}{k} = e_k(y)\) for all \(k \in \{1, \ldots, n\}\), but \(\sum_{i=1}^n (\log x_i)^2 = n > 0 = \sum_{i=1}^n (\log y_i)^2\).

Neff, Nakatsukasa and Fischle [28] 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 [24]. The first named author extended the problem to the complex plane and presented a sketch of a proof. Meanwhile, inspired by this new approach, Miroslav Šilhavý (Czech Academy of Science) has found a solution [33] to classifying all functions \(W\) such that \(e_1(x) \leq e_1(y), \ldots, e_{n-1}(x) \leq e_{n-1}(y), e_n(x) = e_n(y)\) implies \(W(x) \leq W(y)\).

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 \(z_1, \ldots, z_n\) be pairwise different complex numbers and \(k \in \{0, \ldots, n-1\}\). Then

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

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

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

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

\[
\frac{t^k}{\prod_{j=1}^n (t-z_j)} = \sum_{i=1}^n \frac{t-z_r}{t-z_i} a_i \quad \text{for all } t \in \mathbb{C} \setminus \{z_1, \ldots, z_n\}. \tag{3}
\]

Taking the limit \(t \to z_r\) on both sides of the equality, we find

\[
\frac{z_r^k}{\prod_{j \neq r}^n (z_r - z_j)} = a_r.
\]
Corollary 2.2. Let \( z_1, \ldots, z_n \in \mathbb{C} \) be pairwise different complex numbers. Then
\[
\sum_{i=1}^{n} \frac{z_i^k}{\prod_{j=1, j \neq i}^{n} (z_i - z_j)} = 0 \quad \text{for all } k \in \{0, \ldots, n - 2\}. \tag{4}
\]

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 \( k \in \{0, \ldots, n - 2\} \). Using equality (2), we obtain
\[
\sum_{i=1}^{n-1} \frac{z_i^k}{(t - z_i) \prod_{j=1, j \neq i}^{n-1} (z_i - z_j)} = \prod_{j=1}^{n-1} (t - z_j) \quad \text{for all } t \in \mathbb{C} \setminus \{z_1, \ldots, z_{n-1}\}.
\]

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 \( z := (z_1, z_2, \ldots, z_n) \) and the vector of the elementary symmetric polynomials evaluated at \( z \), i.e. \( (e_1(z), \ldots, e_n(z)) \). To this end, we define the characteristic polynomial \( h_e \) of a linear map with the invariants \( e_1, \ldots, e_n \):
\[
h_e(t) := t^n - e_1 t^{n-1} + e_2 t^{n-2} + \ldots + (-1)^n e_n = t^n + \sum_{k=1}^{n} (-1)^k e_k t^{n-k}.
\]

Since if \( h_e \) has the roots \( z_1, \ldots, z_n \), we can write
\[
h_e(t) = (t - z_1)(t - z_2)\ldots(t - z_n)
\]
\[
= t^n - (z_1 + \ldots + z_n) t^{n-1} + (z_1 z_2 + \ldots + z_{n-1} z_n) t^{n-2} + \ldots + (-1)^n z_1 \ldots z_n
\]
\[
= t^n - e_1(z) t^{n-1} + e_2(z) t^{n-2} + \ldots + (-1)^n e_n(z).
\]

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
\[
\mathbb{C}^{n\uparrow} := \{ z \in \mathbb{C}^n \mid \text{Re}(z_1) \geq \ldots \geq \text{Re}(z_n), \text{Re} z_i = \text{Re} z_{i+1} \Rightarrow \text{Im} z_i \geq \text{Im} z_{i+1} \forall i \in \{1, \ldots, n-1\} \},
\]
which contains only ordered vectors and thereby excludes all rearrangements. Furthermore, we define the set
\[
M := \{ z \in \mathbb{C}^{n\uparrow} \mid e_1(z), e_2(z), \ldots, e_n(z) \in \mathbb{R}^+ \}
\]
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 \( z_1, \ldots, z_n \) is continuous and bijective. Its inverse function is continuous as well, and we denote it by
\[
\varphi: \mathbb{R}^n_+ \rightarrow M \subseteq \mathbb{C}^{n\uparrow}, \quad (e_1, \ldots, e_n) \mapsto \varphi(e_1, \ldots, e_n).
\]

Furthermore, each vector \( (z_1, \ldots, z_n) \in M \) contains only positive real numbers and complex conjugate pairs of numbers.
Proof. The elementary symmetric polynomials \( e_1(z), \ldots, e_n(z) \) evaluated at \( z \) are exactly the coefficients \( e_1, \ldots, e_n \) of the polynomial \( h_e \) with the roots \( z_1, \ldots, z_n \). 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 \( \phi \) 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 \( f \circ \phi \) 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 \((f \circ \phi)(e) \in \mathbb{R} \) for all \( e \in T \).

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

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

2.) We find a path \( \gamma: [0, 1] \rightarrow \phi(T) \) 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 \( k \in \{1, \ldots, n-1\} \) 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 [31], Pompe and Neff managed to prove the SSLI for \( n = 4 \) by choosing

\[
T = \{ (e_1(z), \ldots, e_n(z)) \mid z \in \mathbb{R}_+^n \} ;
\]  

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

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

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

$$
T := \mathbb{R}_+^n. \tag{9}
$$

By this choice, satisfying condition 2.) is rather trivial; we take the straight line $\hat{\gamma}$ in $T$ from $e(x)$ to $e(y)$, i.e. $\hat{\gamma}(s) := s \cdot e(x) + (1-s) \cdot e(y)$, and consider the curve $\gamma : [0, 1] \to M$ with $\gamma(s) = \varphi(\hat{\gamma}(s))$. Of course, it is difficult to explicitly characterize $\gamma$, 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

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

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

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

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

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

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

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

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

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

$$
\frac{\partial \varphi_i(e)}{\partial e_k} = \frac{(-1)^{k+1} \varphi_i(e)^{n-k}}{\prod_{j=1, j \neq i}^{n} (\varphi_i(e) - \varphi_j(e))} \quad \text{for any } k \in \{1, \ldots, n\}.
$$
Proof. Using the notation $h(e_1, \ldots, e_n, t) := h_e(t)$, we can characterize the functions $\varphi_i$, $i \in \{1, \ldots, n\}$, implicitly by the equation

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

According to the implicit function theorem, $\varphi_i$ 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 $\varphi_1(e), \ldots, \varphi_n(e)$ of $h_e$ are pairwise different; in particular, the root $\varphi_i(e)$ is simple and therefore $\frac{\partial}{\partial t} h(e, \varphi_i(e)) \neq 0$. We differentiate

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

$$= (-1)^k \varphi_i(e)^{n-k} \cdot 1 + \frac{\partial h}{\partial t}(e, \varphi_i(e)) \cdot \frac{\partial \varphi_i}{\partial e_k}(e),$$

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

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

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

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

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 $e = (e_1, \ldots, e_n) \in \mathbb{R}^n_+$ be such that all components of $\varphi(e)$ are pairwise different. Then

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

**Proof.** Because the components of $\varphi(e)$ are pairwise different, $\varphi$ 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 $k \in \{1, \ldots, n\}$ using the chain rule:

$$\frac{\partial (f \circ \varphi)}{\partial e_k}(e) = \sum_{i=1}^{n} \frac{\partial f}{\partial z_i}(\varphi_1(e), \ldots, \varphi_n(e)) \frac{\partial \varphi_i}{\partial e_k}(e_1, \ldots, e_n)$$

$$= \sum_{i=1}^{n} 2 \log \varphi_i(e) \cdot \frac{(-1)^{k+1} \varphi_i(e)^{n-k}}{\prod_{j=1, j \neq i}^{n}(\varphi_i(e) - \varphi_j(e))}$$

$$= 2 \sum_{i=1}^{n} \frac{(-1)^{n+k} \varphi_i(e)^{n-k-1}}{(-1)^{n-1} \prod_{j=1, j \neq i}^{n}(\varphi_i(e) - \varphi_j(e))} \log \varphi_i(e)$$

$$= -2 \sum_{i=1}^{n} \frac{(-\varphi_i(e))^{n-k-1}}{\prod_{j=1, j \neq i}^{n}(\varphi_j(e) - \varphi_i(e))} \log \varphi_i(e).$$ ■
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 [21] that

\[
(-1)^q \sum_{i=1}^{n} \frac{e_i^n}{\prod_{j \neq i}^n (z_j - z_i)} \log z_i \geq 0
\]  

(14)

for all \( q \in 2, \ldots, n \). If we let \( z_i = \varphi_i(e) \), we seem to have already reached our goal. However, Mitchison and Josza prove inequality (14) only for the case that \( z_1, \ldots, z_n \) 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 \( z_1, \ldots, z_n \), while in our case, \( \varphi_1(e), \ldots, \varphi_n(e) \) might also include pairs of complex conjugates.

We close this gap by showing:

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

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

**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 \( r \leq n - 2 \),

\[
\int_0^\infty \left| \frac{t^r}{\prod_{j=1}^n (t + z_j)} \right| \, dt \leq \int_0^1 \frac{t^r}{\prod_{j=1}^n z_j} \, dt + \int_1^\infty \frac{t^{r-n}}{t-1} \, dt \leq \frac{1}{\prod_{j=1}^n z_j} + 1.
\]  

(15)

Thus the integral converges for \( r \in \{0, \ldots, n - 2\} \) 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 \( t = 0 \):

\[
\int_0^\infty \frac{t^r}{\prod_{j=1}^n (t + z_j)} \, dt \overset{(2)}{=} \int_0^\infty \sum_{i=1}^{n} \frac{(-z_i)^r}{(t + z_i) \prod_{j \neq i}^n (z_j - (z_i))} \, dt
\]

\[
= \sum_{i=1}^{n} \frac{(-z_i)^r}{\prod_{j \neq i}^n (z_j - z_i)} \cdot \log(t + z_i) \bigg|_{t=0}^{t=\infty}
\]

\[
= \lim_{t \to \infty} \left( \sum_{i=1}^{n} \frac{(-z_i)^r}{\prod_{j \neq i}^n (z_j - z_i)} \log t + \sum_{i=1}^{n} \frac{(-z_i)^r}{\prod_{j \neq i}^n (z_j - z_i)} \left( \log \left( 1 + \frac{z_i}{t} \right) \right) - \sum_{i=1}^{n} \frac{(-z_i)^r}{\prod_{j \neq i}^n (z_j - z_i)} \log z_i \right).
\]

(16)

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

\textbf{Proof of Conjecture 2.6 in the case of pairwise different roots $\varphi_i(e)$}. We can now conclude: for all $n - k - 1 \in \{0, \ldots, n - 2\}$, that is for $k \in \{1, \ldots, n - 1\}$,

$$\frac{\partial (f \circ \varphi)}{\partial e_k}(e) \overset{\text{Prop} 2.7}{=} -2 \sum_{i=1}^{n} \prod_{j \neq i}^{i} (\varphi_j(e) - \varphi_i(e)) \log \varphi(e)$$

$$\overset{\text{Prop} 2.8}{=} 2 \int_{0}^{\infty} \frac{t^{n-k-1}}{\prod_{j=1}^{n} (t + \varphi_j(e))} dt > 0.$$ 

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:

\textbf{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.

\textbf{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 $z_1, \ldots, z_n$ and coefficients $a_0, \ldots, a_n$. The discriminant of $p$ is defined as

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

and is zero if and only if $p$ has multiple roots (see also [18, p. 204]). The discriminant is a symmetric homogeneous polynomial in the variables $z_1, \ldots, z_n$ and thus can be expressed as a homogeneous polynomial in $e_k(z_1, \ldots, z_n)$, 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:

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

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

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

9
Proof. Consider the path $e^s = (e_1^s, \ldots, e_n^s) \subseteq \mathbb{R}_+^n$ for $s \in [0, 1]$ with
$$e_k^s := (1 - s) e_k(x) + s e_k(y).$$
Then $e^0 = e(x)$ and $e^1 = e(y)$ as well as $e_k(x) < e_k(y)$ for all $k \in \{1, \ldots, n - 1\}$ and $e_n(x) = e_n(y)$. Since $x_1 > x_2 > \ldots > x_n$, the polynomial $h_{e^x}$ 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
$$\frac{d}{ds} f(\varphi(e_s)) = \sum_{k=1}^n \frac{\partial(f \circ \varphi)}{\partial e_k}(e^s) \cdot \frac{d}{ds} e_k^s = \sum_{k=1}^n \frac{\partial(f \circ \varphi)}{\partial e_k}(e^s) \cdot (e_k(y) - e_k(x)) > 0. \quad (18)$$
Thus the continuity of the mapping $s \mapsto (f \circ \varphi)(e_s)$ implies its strict monotonicity on $[0, 1]$, and therefore
$$f(x) = (f \circ \varphi)(e(x)) = f(\varphi(e_0)) < f(\varphi(e_1)) = (f \circ \varphi)(e(y)) = f(y). \blacksquare$$
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 $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n) \in \mathbb{R}_+^n$ be such that $e_k(x) < e_k(y)$ for all $k \in \{1, \ldots, n - 1\}$ 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 $x = (x_1, \ldots, x_n)$ are pairwise different, then we can assume $x_1 > x_2 > \ldots > x_n$ and $y_1 \geq y_2 \geq \ldots \geq y_n$ 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 $k \in \{1, \ldots, n - 1\}$. 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
$$\begin{align*}
(\log x_i')^2 + (\log x_j')^2 &= (\log x_i + \log(1 + \varepsilon))^2 + (\log x_j - \log(1 + \varepsilon))^2 \\
&= 2(\log x_i)^2 + 2(\log(1 + \varepsilon))^2 > (\log x_i)^2 + (\log x_j)^2.
\end{align*} \quad (19)$$
If we dissolve all pairwise equalities in this way, then we can apply Lemma 2.10 to find
$$\sum_{i=1}^n (\log x_i)^2 < \sum_{i=1}^n (\log x_i')^2 < \sum_{i=1}^n (\log y_i)^2. \blacksquare$$
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 \( x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n \in \mathbb{R}_+ \) such that
\[
e_k(x) \leq e_k(y) \quad \text{for all } k \in \{1, \ldots, n-1\}
\]
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 \( S \subseteq \{1, \ldots, n-1\} \) such that \( e_k(x) = e_k(y) \) for all \( k \in S \) and \( e_k(x) < e_k(y) \) for all \( k \in \{1, \ldots, n-1\} \setminus S \). Furthermore, for \( k \in \{1, \ldots, n\} \) and \( m \in \mathbb{N} \), let
\[
e_k^m = \begin{cases} 
e_k(x) - \frac{1}{m} & \text{if } k \in S, \\ 
e_k(x) & \text{otherwise} \end{cases}
\]
and
\[
x^m := \varphi(e^m).
\]
Then \( 0 < e_k(x^m) < e_k(y) \) for all \( k \in \{1, \ldots, n-1\} \) 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 \( \lim_{m \to \infty} x^m = x \), we find
\[
\sum_{i=1}^{n} (\log x_i)^2 = \lim_{m \to \infty} \sum_{i=1}^{n} (\log x_i^m)^2 \leq \sum_{i=1}^{n} (\log y_i)^2.
\]

\[\blacksquare\]

### 3 Applications and Connections

#### 3.1 Relation to entropy

Jozsa and Mitchison [16] 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
\[
g: (\mathbb{C} \setminus \mathbb{R}_{\leq 0})^n \to \mathbb{C}, \quad g(z_1, \ldots, z_n) = - \sum_{i=1}^{n} z_i \log(-z_i)
\]
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
\[
x_i \log x_i + \overline{x_i} \log \overline{x_i} = x_i \log x_i + \overline{x_i} \cdot \log \overline{x_i} = x_i \log x_i + \overline{x_i} \log \overline{x_i} \in \mathbb{R}.
\]
Analogously to the last section, the function \( g \circ \varphi \) can be expressed as

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

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

We now determine the partial derivatives:

\[
\frac{\partial (g \circ \varphi)}{\partial e_k}(e) = \sum_{i=1}^{n} \frac{\partial g}{\partial z_i}(\varphi_1(e), \ldots, \varphi_n(e)) \frac{\partial \varphi_i}{\partial e_k}(e_1, \ldots, e_n)
\]

\[
\overset{\text{Prop. 2.5}}{=} \sum_{i=1}^{n} \left( \log \varphi_i(e) + 1 \right) \frac{-1^{k+1} \varphi_i(e)^{n-k}}{\prod_{j \neq i}^{n} (\varphi_i(e) - \varphi_j(e))}
\]

\[
= \sum_{i=1}^{n} \frac{(-1)^{n+k-1} \varphi_i(e)^{n-k}}{(-1)^{n-1} \prod_{j \neq i}^{n} (\varphi_i(e) - \varphi_j(e))} \left( \log \varphi_i(e) + 1 \right)
\]

\[
= -\sum_{i=1}^{n} \frac{(-\varphi_i(e))^{n-k}}{\prod_{j \neq i}^{n} (\varphi_j(e) - \varphi_i(e))} \log \varphi_i(e) - \sum_{i=1}^{n} \frac{(-\varphi_i(e))^{n-k}}{\prod_{j \neq i}^{n} (\varphi_j(e) - \varphi_i(e))}.
\]

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

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

for \( n - k \in \{0, \ldots, n - 2\} \) or, equivalently, for \( k \in \{2, \ldots, n\} \).

**Remark 3.1.** It should also be noted that \( \frac{\partial (f \circ \varphi)}{\partial e_k} = 2 \frac{\partial (g \circ \varphi)}{\partial e_k} \).

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 \( x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n > 0 \) such that

\[
e_1(x) = e_1(y)
\]

\[
e_k(x) \leq e_k(y) \quad \text{for all} \ k \in \{2, \ldots, n\}.
\]

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 \( k \in \{1, \ldots, n\} \) again allows for obvious counterexamples similar to the SSLI: For \( x = (e^{-1}, \ldots, e^{-1}) \in \mathbb{R}_+^n \) and \( y = (1, \ldots, 1) \in \mathbb{R}_+^n \) 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 $\lambda_1, \ldots, \lambda_n > 0$. The $k$-th invariant $I_k(U)$ of $U$ is defined as the $k$-th elementary symmetric polynomial of the vector $\lambda(U) = (\lambda_1, \ldots, \lambda_n)$, i.e. $I_k(U) := e_k(\lambda(U))$; (thus, $I_1(U) = \text{tr} X$ 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 $k \in \{1, \ldots, n-1\}$ 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

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

Theorem 3.3 can be applied directly to the quadratic Hencky energy

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

which was introduced into the theory of nonlinear elasticity in 1929 by H. Hencky [14]. Here, $F \in \text{GL}^+(n)$ is the deformation gradient, $\text{GL}^+(n)$ is the set of invertible $n \times n$-matrices with positive determinant, $U = \sqrt{F^T F}$ is the right stretch tensor and $\text{dev}_n \log U = \log U - \frac{1}{n} \text{tr}(\log U) \cdot \mathbb{1}$ is the deviatoric part of the Hencky strain tensor $\log U$. The material parameters $\mu, \lambda$ 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 [29, 26, 30]: it measures the squared geodesic distance of $F$ to the special orthogonal group $\text{SO}(n)$ with respect to a left-$\text{GL}(n)$-invariant, right-$\text{SO}(n)$-invariant Riemannian metric on $\text{GL}(n)$.

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

**Corollary 3.4.** Let $F, \tilde{F} \in \text{GL}^+(n)$ 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 $k \in \{1, \ldots, n-1\}$, 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 [34, pages 158, 171].

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

**Theorem 3.6.** Let $U, \tilde{U} \in \text{Sym}_+(n)$. If $\text{tr} U = \text{tr} \tilde{U}$ and $I_k(U) \leq I_k(\tilde{U})$ for all $k \in \{2, \ldots, n\}$, then $\langle U, \log U - \mathbb{1} \rangle \geq \langle \tilde{U}, \log \tilde{U} - \mathbb{1} \rangle$, where $\langle X, Y \rangle = \text{tr}(X^T Y)$ 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 \( \text{tr} \ U = \text{tr} \ \tilde{U} \), we compute

\[
\langle U, \log U - \mathbb{1} \rangle = \langle U, \log U \rangle - \langle U, \mathbb{1} \rangle = \sum_{i=1}^{n} \lambda_{i}(U) \cdot \log \lambda_{i}(U) - \text{tr} \ U
\]

\[
\geq \sum_{i=1}^{n} \lambda_{i}(\tilde{U}) \cdot \log \lambda_{i}(\tilde{U}) - \text{tr} \ \tilde{U} = \langle \tilde{U}, \log \tilde{U} - \mathbb{1} \rangle.
\]

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

\[
W_{B}(F) = \langle U, \log U - \mathbb{1} \rangle = \langle \sqrt{FF^{T}}, \log \sqrt{FF^{T}} - \mathbb{1} \rangle
\]

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 [27]. 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 \( \text{tr}(X \log X) = \sum_{i=1}^{n} \lambda_{i}(X) \log \lambda_{i}(X) \), we can translate the last Theorem 3.6 into a statement for the quantum von Neumann entropy.

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

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, 22, 23, 32]

\[
g_{C}(X,Y) = \text{tr}(C^{-1}XC^{-1}Y),
\]

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]

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

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 \( I_{1}(C), \ldots, I_{n}(C), I_{1}(\tilde{C}), \ldots, I_{n}(\tilde{C}) \) denote the principal invariants of \( C,\tilde{C} \in \text{Sym}_{+}(n) \). If \( I_{k}(C) \leq I_{k}(\tilde{C}) \) for all \( k \in \{1, \ldots, n-1\} \) 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 \( \mathbf{1} \) 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

\[
\text{tr} \vee^k X \leq \text{tr} \vee^k Y \quad \text{for all } k \in \{1, \ldots, n-1\} \quad \text{and} \quad \text{tr} \vee^n X = \text{tr} \vee^n Y
\]

\[
\implies F(X) \leq F(Y) .
\]

More generally, extensions of the SSLI to other matrix monotone functions may be obtained by building on [33].

- It might also be of interest to characterize the matrix functions \( F : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n} \) 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 [19, 28] 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) [15]. 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 Kellog applies to \( P \) matrices, and also to \( Q \) matrices [15]:

**Theorem 3.10** ([17]). Let \( X \in \mathbb{C}^{n \times n} \) be a matrix for which \( I_k(X) \geq 0 \) for \( k \in \{1, \ldots, n\} \). 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

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

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 $a = (a_1, \ldots, a_{n-1}) \in \mathbb{R}^n_{+1}$ let $\hat{h}_a(z) = z^n + a_{n-1}z^{n-1} + \ldots + a_1z + 1$. Define the function $\hat{f} : \mathbb{R}^n_{+1} \to \mathbb{R}$ by

$$\hat{f}(a_1, \ldots, a_{n-1}) = \sum_{\hat{z} \in \mathbb{C} \setminus \mathbb{R} \leq 0} \left( \log(-\hat{z}) \right)^2,$$

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.

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

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

Proof. We write the function $\hat{f}$ 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 $R \geq 0 := \{ x \in \mathbb{R} | x \geq 0 \}$. 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 find

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

A related representation to (25) has been used recently in [16] in connection with the entropy function.
where the path $C$ consists of the following four pieces $C_\varepsilon$, $C_+$, $C_R$ and $C_-$, see also Figure 2:

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

By straightforward computation, we find

\[
\frac{\partial}{\partial a_k} \hat{h}_a'(z) = \frac{\frac{\partial}{\partial a_k} \hat{h}_a'(z) \cdot \hat{h}_a(z) - \hat{h}_a'(z) \cdot \frac{\partial}{\partial a_k} \hat{h}_a(z)}{(\hat{h}_a(z))^2} = \frac{k \cdot z^k \cdot \hat{h}_a(z) - z^k \cdot \hat{h}_a'(z)}{(\hat{h}_a(z))^2} = \left( \frac{z^k}{\hat{h}_a(z)} \right)'.
\]

(26)

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

\[
\frac{\partial}{\partial a_k} f(a) = \frac{1}{2\pi i} \int_C (\log(-z))^2 \frac{\partial}{\partial a_k} \hat{h}_a'(z) \frac{\partial}{\partial a_k} \hat{h}_a(z) \, dz = \frac{1}{2\pi i} \int_C (\log(-z))^2 \left( \frac{z^k}{\hat{h}_a(z)} \right)' \, dz
\]

\[
= -\frac{1}{2\pi i} \int_C \left( (\log(-z))^2 \right)' \frac{z^k}{\hat{h}_a(z)} \, dz = -\frac{1}{\pi i} \int_C \log(-z) \frac{z^{k-1}}{\hat{h}_a(z)} \, dz.
\]

(27)

We can now take the limits as $R \to +\infty$ and $\varepsilon \to 0$. 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 $C_\varepsilon$ also tends to zero, since for $k \geq 1$, the integrand is of order $O(\log \varepsilon)$ and the length of $C_\varepsilon$ is $2\pi \varepsilon$.

We conclude:

\[
\frac{\partial}{\partial a_k} f(a) = -\frac{1}{\pi i} \lim_{\varepsilon \searrow 0, R \to \infty} \left( \int_{C_+ \cup C_-} \log(-z) \frac{z^{k-1}}{\hat{h}_a(z)} \, dz \right)
\]

\[
= -\frac{1}{\pi i} \lim_{\varepsilon \searrow 0, R \to \infty} \left( \int_{[\varepsilon, R]} \left( (\log t - \pi i) - (\log t + \pi i) \right) \frac{t^{k-1}}{\hat{h}_a(t)} \, dt \right)
\]

\[
= 2 \lim_{\varepsilon \searrow 0, R \to \infty} \int_{\varepsilon}^{R} \frac{t^{k-1}}{\hat{h}_a(t)} \, dt = 2 \int_{0}^{+\infty} \frac{t^{k-1}}{\hat{h}_a(t)} \, dt > 0.
\]

(28)

\[\square\]

**Acknowledgement**

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


IN DER SCHRIFTENREIHE DER FAKULTÄT FÜR MATHEMATIK ZULETZT ERSCHIENENE
BEITRÄGE:


Nr. 778: Pozzi, P.: Computational Anisotropic Willmore Flow, 2014


Nr. 780: Blatt, S., Reiter, Ph.: Modeling repulsive forces on fibres via knot energies, 2014


Nr. 784: Müller, F.: On $C^{1,1/2}$-regularity of $H$-surfaces with a free boundary, 2014


Nr. 786: Bauer S., Pauly, D.: On Korn’s First Inequality for Tangential or Normal Boundary Conditions with Explicit Constants, 2015

Nr. 787: Neff, P., Eidel, B., Martin, R.J.: Geometry of logarithmic strain measures in solid mechanics, 2015

Nr. 788: Borisov, L., Neff, P., Sra, S., Thiel, Chr.: The sum of squared logarithms inequality in arbitrary dimensions, 2015