A NOTE ON THE CONVEXITY OF $\mathbf{C} \mapsto h(\det \mathbf{C})$

SCOTT J. SPECTOR

Abstract. Recently, Lehmich, Neff, and Lankeit [Math. Mech. Solids **19** (2014), 369–375] obtained necessary and sufficient conditions for the function $\mathbf{C} \mapsto h(\det \mathbf{C})$ to be convex on strictly positive-definite, symmetric $n \times n$ matrices $\mathbf{C}$. In this note an alternate proof of their result is provided.

1. Introduction; Preliminaries

In [7] Lehmich, Neff, and Lankeit proved that, for functions $h : (0, \infty) \to \mathbb{R}$ which have two derivatives, necessary and sufficient conditions for the convexity of the composition map $h \circ \det : \mathbf{C} \mapsto h(\det \mathbf{C})$ on the set of strictly positive-definite, symmetric $n \times n$ matrices, $n \geq 2$, are that

$$nsh''(s) + (n-1)h'(s) \geq 0 \quad \text{and} \quad h'(s) \leq 0 \quad \text{for every } s > 0, \tag{1.1}$$

where $\det \mathbf{C}$ denotes the determinant of the $n \times n$ matrix $\mathbf{C}$. They noted that the convexity of the map $h \circ \det$ may be useful in analyzing stored-energy functions in nonlinear elasticity. For example, a compressible neo-Hookean material can be written as a function of the right Cauchy-Green strain matrix $\mathbf{C} := \mathbf{F}^{\mathrm{T}}\mathbf{F}$, where $\mathbf{F}$ is the (matrix with respect to an orthonormal basis of the) gradient of a deformation at any point in an elastic body (see, e.g., [2, p. 189]):

$$W(\mathbf{C}) = \tfrac{\mu}{2} \operatorname{tr} \mathbf{C} + h(\det \mathbf{C}).$$

Here $\operatorname{tr} \mathbf{C}$ denotes the *trace* of $\mathbf{C}$, i.e., the sum of the diagonal elements of the matrix $\mathbf{C}$.

Our analysis commences with the observation that, for C^2 functions h, (1.1) is equivalent to

$$t \mapsto h(t^n) \text{ is convex on } (0, \infty) \quad \text{and} \quad t \mapsto h(t) \text{ is decreasing on } (0, \infty). \tag{1.2}$$

We will show that these alternative conditions are necessary and sufficient for any function (whether or not it is differentiable) $h : (0, \infty) \to \mathbb{R}$ to satisfy the condition that the composition $h \circ \det$ is convex on the set of strictly positive-definite, symmetric $n \times n$ matrices.

Date: 5 June 2014.

2010 *Mathematics Subject Classification.* 74B20.

Key words and phrases. Convexity, Elasticity.

This material is based upon work supported by the National Science Foundation under Grant No. DMS–1107899. The author thanks K. Pericak-Spector and R. Fosdick for their helpful comments.

Let $\mathbf{E}$ be a symmetric $n \times n$ matrix. Then the *spectral theorem* (see, e.g., [5, §79]) implies that $\mathbf{E}$ has exactly n (not necessarily distinct) eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ each of which is a root of the *characteristic polynomial* (see, e.g., [3, p. 70])

$$p(r) := \det(\mathbf{E} - r\mathbf{I}) = \sum_{k=0}^{n} (-r)^{n-k} I_k(\mathbf{E}) = \prod_{j=1}^{n} (\lambda_j - r), \qquad r \in \mathbb{R}, \tag{1.3}$$

where $\mathbf{I}$ denotes the $n \times n$ identity matrix and $I_k(\mathbf{E})$ denotes the *principal invariants* of $\mathbf{E}$, which we assume are defined by the next to last equality in (1.3). Moreover, since this polynomial has n real roots the last equality in (1.3) implies that

$$I_k(\mathbf{E}) = \sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} \lambda_{i_1} \lambda_{i_2} \cdots \lambda_{i_k}, \tag{1.4}$$

i.e., (see, e.g., [8, Theorem 4.19])

$$\begin{aligned}
I_1(\mathbf{E}) &= \lambda_1 + \lambda_2 + \cdots + \lambda_n = \operatorname{tr} \mathbf{E}, \\
I_2(\mathbf{E}) &= \lambda_1\lambda_2 + \lambda_1\lambda_3 + \cdots + \lambda_1\lambda_n + \lambda_2\lambda_3 + \cdots + \lambda_{n-1}\lambda_n \\
&\qquad \text{(the sum of all 2-term products of the } \lambda\text{'s with distinct indices)}, \\
&\ \ \vdots \\
I_k(\mathbf{E}) &= \lambda_1\lambda_2 \cdots \lambda_k + \cdots + \lambda_2\lambda_3 \cdots \lambda_{k+1} + \cdots + \lambda_{n-k+1}\lambda_{n-k+2} \cdots \lambda_n \\
&\qquad \text{(the sum of all } k\text{-term products of the } \lambda\text{'s with distinct indices)}, \\
&\ \ \vdots \\
I_{n-1}(\mathbf{E}) &= \prod_{i \ne 1} \lambda_i + \prod_{i \ne 2} \lambda_i + \cdots + \prod_{i \ne n} \lambda_i = \operatorname{tr}(\operatorname{adj} \mathbf{E}), \\
I_n(\mathbf{E}) &= \prod_{i=1}^{n} \lambda_i = \lambda_1\lambda_2 \cdots \lambda_n = \det \mathbf{E},
\end{aligned}$$

where $\operatorname{adj} \mathbf{E}$ denotes the *adjugate* matrix ($\mathbf{E} \operatorname{adj} \mathbf{E} = (\det \mathbf{E})\mathbf{I}$). Moreover, such a matrix is (strictly) positive definite if and only if all of its eigenvalues are (strictly) positive.

Our first result is a consequence of the *arithmetic-geometric mean inequality* (see, e.g., [9]): let $m > 0$ be an integer and suppose that $a_1, a_2, \ldots, a_m$ are nonnegative real numbers. Then

$$\frac{1}{m} \sum_{i=1}^{m} a_i \ge \left(\prod_{i=1}^{m} a_i \right)^{1/m}. \tag{1.5}$$

Lemma 1.1. *The principal invariants* $I_k(\mathbf{E})$, $k = 0, 1, 2, \ldots, n$, *of a positive-definite, symmetric* $n \times n$ *matrix* $\mathbf{E}$ *satisfy*

$$I_k(\mathbf{E}) \ge \mathrm{C}_k^n (\det \mathbf{E})^{k/n}, \qquad \mathrm{C}_k^n = \frac{n!}{k!(n-k)!}, \tag{1.6}$$

where the notation C_k^n *denotes the number of ways to* **choose** k *distinct integers from the set* $\{1, 2, 3, \ldots, n\}$ *without regard to the order that each is chosen.*

Remark 1.2. Notice that the constants C_k^n are those that also occur in the *binomial theorem* (see, e.g., [6, pp. 139–140]): let $n \geq 2$ be an integer. Then for all real numbers a and b

$$(a+b)^n = \sum_{k=0}^{n} \mathrm{C}_k^n a^{n-k} b^k. \tag{1.7}$$

Proof of Lemma 1.1. First note that $I_n(\mathbf{E}) = \det \mathbf{E}$ and $I_0(\mathbf{E}) = 1 = \mathrm{C}_0^n(\det \mathbf{E})^0$. Next, fix an integer $k \in [1, n-1]$. Then the summation in (1.4) consists of C_k^n terms. Thus, by the arithmetic-geometric mean inequality (1.5) with $m := \mathrm{C}_k^n$

$$\begin{aligned} \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} \lambda_{i_1} \lambda_{i_2} \cdots \lambda_{i_k} &\geq \mathrm{C}_k^n \Bigg(\prod_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} \lambda_{i_1} \lambda_{i_2} \cdots \lambda_{i_k} \Bigg)^{1/\mathrm{C}_k^n} \\ &= \mathrm{C}_k^n \Bigg(\prod_{i=1}^{n} [\lambda_i]^M \Bigg)^{1/\mathrm{C}_k^n}, \end{aligned} \tag{1.8}$$

where the integer $M = M(k,n)$ denotes the number of times that each eigenvalue λ_i, $i = 1, 2, \ldots, n$, occurs in the sum of products (1.4). We claim that $M(k,n) = \mathrm{C}_k^n(k/n)$, which, together with (1.4), (1.8), and the fact that $\det \mathbf{E} = \lambda_1 \lambda_2 \cdots \lambda_n$, will yield the desired result $(1.6)_1$.

In order to determine $M = M(k,n)$ let's fix attention on one eigenvalue, say λ_1, and count the number of terms in which it occurs in the sum (1.4). Note that, in each term, λ_1 must multiply $k-1$ other eigenvalues *chosen*, without regard to order, from the remaining $n-1$, i.e., $\lambda_2, \lambda_3, \ldots, \lambda_n$. Thus, this number is C_{k-1}^{n-1} and so

$$M = \mathrm{C}_{k-1}^{n-1} = \frac{(n-1)!}{(k-1)!([n-1]-[k-1])!} = \frac{k}{n}\frac{n!}{k!(n-k)!} = \frac{k}{n}\mathrm{C}_k^n,$$

as claimed. □

2. Sufficiency

Proposition 2.1. *Let $h : (0, \infty) \to \mathbb{R}$ and let $n \geq 2$ be an integer. Suppose that $t \mapsto h(t^n)$ is convex and $t \mapsto h(t)$ is decreasing on $(0, \infty)$. Then $h \circ \det$ is convex on the set of strictly positive-definite, symmetric $n \times n$ matrices.*

Proof. Let $\mathbf{C}$ and $\mathbf{D}$ be strictly positive-definite, symmetric $n \times n$ matrices. Suppose that $\sigma \in (0,1)$. Define $x := (\det \mathbf{C})^{1/n}$ and $y := (\det \mathbf{D})^{1/n}$. Then by the convexity of the map $t \mapsto h(t^n)$

$$\sigma h(x^n) + (1-\sigma) h(y^n) \geq h\Big([\sigma x + (1-\sigma) y]^n\Big).$$

Therefore,

$$\sigma h(\det \mathbf{C}) + (1-\sigma) h(\det \mathbf{D}) \geq h\bigg(\Big[\sigma (\det \mathbf{C})^{1/n} + (1-\sigma)(\det \mathbf{D})^{1/n}\Big]^n\bigg). \tag{2.1}$$

Next, define $\mathbf{E} := \mathbf{C}^{-1}\mathbf{D}$ and $s := \sigma/(1-\sigma)$. Then by the binomial theorem (1.7), Lemma 1.1, and (1.3) (with $r := -s$)

$$\begin{aligned}\Big[\sigma(\det\mathbf{C})^{1/n} + (1-\sigma)(\det\mathbf{D})^{1/n}\Big]^n &= (1-\sigma)^n(\det\mathbf{C})\Big[s + (\det\mathbf{E})^{1/n}\Big]^n \\ &= (1-\sigma)^n(\det\mathbf{C})\sum_{k=0}^{n} \mathrm{C}^n_k s^{n-k}(\det\mathbf{E})^{k/n} \\ &\le (1-\sigma)^n(\det\mathbf{C})\sum_{k=0}^{n} s^{n-k} I_k(\mathbf{E}) \\ &= (1-\sigma)^n(\det\mathbf{C})\det(s\mathbf{I}+\mathbf{E}) \\ &= \det\Big[\sigma\mathbf{C} + (1-\sigma)\mathbf{D}\Big],\end{aligned}$$

and hence, since h is monotone decreasing,

$$h\bigg(\Big[\sigma(\det\mathbf{C})^{1/n} + (1-\sigma)(\det\mathbf{D})^{1/n}\Big]^n\bigg) \ge h\bigg(\det\Big[\sigma\mathbf{C} + (1-\sigma)\mathbf{D}\Big]\bigg). \tag{2.2}$$

The desired result, the convexity of $h\circ\det$, now follows from (2.1) and (2.2). □

When h is differentiable on $(0,\infty)$ a simpler proof, which does not require the use of Lemma 1.1, the characteristic polynomial, the principal invariants, or the binomial theorem, is possible.

Alternate proof of Prop. 2.1, assuming that h is differentiable on $(0,\infty)$. Let $\mathbf{C}$ and $\mathbf{D}$ be strictly positive-definite, symmetric $n\times n$ matrices and define $\mathbf{U} := \mathbf{D}^{-1}\mathbf{C}$. First note, for future reference, that the arithmetic-geometric mean inequality (1.5) implies that

$$\operatorname{tr}\mathbf{U} = \sum_{i=1}^{n}\hat\lambda_i \ge n\bigg(\prod_{i=1}^{n}\hat\lambda_i\bigg)^{1/n} = n(\det\mathbf{U})^{1/n}, \tag{2.3}$$

where $\hat\lambda_i > 0$, $i = 1, 2, \dots, n$, here denote the eigenvalues of $\mathbf{U}$.

Define $x := (\det\mathbf{C})^{1/n}$ and $y := (\det\mathbf{D})^{1/n}$. Then a well-known consequence (see, e.g., [1, §3.1.3]) of the convexity of the map $t \mapsto h(t^n)$ is that it lies above its tangent lines:

$$h(x^n) \ge h(y^n) + ny^{n-1}h'(y^n)(x-y).$$

Therefore,

$$\begin{aligned}h(\det\mathbf{C}) &\ge h(\det\mathbf{D}) + n(\det\mathbf{D})^{1-1/n}h'(\det\mathbf{D})\Big[(\det\mathbf{C})^{1/n} - (\det\mathbf{D})^{1/n}\Big] \\ &= h(\det\mathbf{D}) + (\det\mathbf{D})h'(\det\mathbf{D})\Big[n(\det\mathbf{U})^{1/n} - n\Big] \\ &\ge h(\det\mathbf{D}) + (\det\mathbf{D})h'(\det\mathbf{D})\Big[\operatorname{tr}\mathbf{U} - n\Big],\end{aligned} \tag{2.4}$$

where we have made use of (2.3) and the fact that $h'(\det\mathbf{D}) \le 0$, which is a consequence of the assumption that h is decreasing and differentiable.

Next, note that the mapping $\mathbf{F} \mapsto \det \mathbf{F}$ is differentiable (on the set of $n \times n$ matrices with strictly positive determinant) with derivative given by (see, e.g., [4, p. 23])

$$\frac{d}{d\mathbf{F}}(\det \mathbf{F})[\mathbf{H}] = (\det \mathbf{F}) \operatorname{tr}\Big[\mathbf{F}^{-1}\mathbf{H}\Big].$$

Thus, by the chain rule

$$\begin{aligned} \frac{d}{d\mathbf{D}} h(\det \mathbf{D})[\mathbf{C} - \mathbf{D}] &= (\det \mathbf{D}) h'(\det \mathbf{D}) \operatorname{tr}\Big[\mathbf{D}^{-1}(\mathbf{C} - \mathbf{D})\Big] \\ &= (\det \mathbf{D}) h'(\det \mathbf{D})[\operatorname{tr} \mathbf{U} - n]. \end{aligned} \tag{2.5}$$

Finally, (2.4) and (2.5) imply that the differentiable mapping $h \circ \det$ lies above all of its tangent hyperplanes. A well-known consequence (see, e.g., [1, §3.1.3]) of that property is the convexity of $h \circ \det$. □

3. Necessity

Proposition 3.1. *Let $h : (0, \infty) \to \mathbb{R}$ and let $n \geq 2$ be an integer. Suppose that $h \circ \det$ is convex on the set of strictly positive-definite, symmetric $n \times n$ matrices. Then $t \mapsto h(t^n)$ is convex and $t \mapsto h(t)$ is decreasing on $(0, \infty)$.*

Proof. We first show that $t \mapsto h(t^n)$ is convex on $(0, \infty)$. Towards that end let $x \neq y$ be strictly positive real numbers and suppose that $\sigma \in (0, 1)$. Define $\mathbf{C}$ and $\mathbf{D}$ to be the strictly positive-definite, symmetric $n \times n$ matrices given by $\mathbf{C} := x\mathbf{I}$ and $\mathbf{D} := y\mathbf{I}$. Then $\det \mathbf{C} = x^n$, $\det \mathbf{D} = y^n$,

$$\det\Big(\sigma\mathbf{C} + (1 - \sigma)\mathbf{D}\Big) = \det\Big([\sigma x + (1 - \sigma)y]\mathbf{I}\Big) = [\sigma x + (1 - \sigma)y]^n,$$

and hence the convexity of the map $h \circ \det$ yields

$$\begin{aligned} \sigma h(x^n) + (1 - \sigma)h(y^n) &= \sigma h(\det \mathbf{C}) + (1 - \sigma)h(\det \mathbf{D}) \\ &\geq h\Big(\det\Big[\sigma\mathbf{C} + (1 - \sigma)\mathbf{D}\Big]\Big) \\ &= h\Big([\sigma x + (1 - \sigma)y]^n\Big), \end{aligned}$$

which establishes the convexity of the map $t \mapsto h(t^n)$.

We next show that $t \mapsto h(t)$ is decreasing on $(0, \infty)$. Suppose that x and y are real numbers that satisfy $y > x > 0$. Define the real number $a > 0$ by

$$a := \Big(\sqrt{y} - \sqrt{y - x}\Big)/x$$

so that a satisfies the quadratic equation $xa^2 - 2a\sqrt{y} + 1 = 0$ and hence

$$ax + a^{-1} = 2\sqrt{y}. \tag{3.1}$$

Define $\mathbf{C}$ and $\mathbf{D}$ to be the strictly positive-definite, symmetric $n \times n$ *diagonal* matrices given by

$$\mathbf{C} := \operatorname{diag}\{ax, a^{-1}, 1, 1, 1, \ldots, 1\}, \qquad \mathbf{D} := \operatorname{diag}\{a^{-1}, ax, 1, 1, 1, \ldots, 1\}.$$

Then

$$\det \mathbf{C} = x = \det \mathbf{D}, \tag{3.2}$$

$$\tfrac{1}{2}(\mathbf{C}+\mathbf{D}) = \operatorname{diag}\{\tfrac{1}{2}(ax+a^{-1}), \tfrac{1}{2}(a^{-1}+ax), 1, 1, 1, \ldots, 1\},$$

and hence, in view of (3.1),

$$\det\Big[\tfrac{1}{2}\mathbf{C} + \tfrac{1}{2}\mathbf{D}\Big] = \tfrac{1}{4}\Big[ax + a^{-1}\Big]^2 = y. \tag{3.3}$$

The desired result now follows from (3.3), (3.2), and the convexity of the composition $h \circ \det$, i.e.,

$$h(y) = h\Big(\det\Big[\tfrac{1}{2}\mathbf{C} + \tfrac{1}{2}\mathbf{D}\Big]\Big) \leq \tfrac{1}{2}h(\det \mathbf{C}) + \tfrac{1}{2}h(\det \mathbf{D}) = h(x). \qquad \square$$

4. Concluding Remarks

Simple examples show that (1.1) does not yield the convexity of $h \circ \det$ on the set of $n \times n$ matrices with strictly positive determinant. However, if one first fixes any matrix $\mathbf{G}_0$ that satisfies $\det \mathbf{G}_0 > 0$, then (1.1) implies that $\phi(\mathbf{F}) := h(\det \mathbf{F})$ is convex at $\mathbf{G}_0$ in certain directions, that is, assuming h is twice differentiable at $\mathbf{G}_0$,

$$\mathbf{C} : \frac{d^2\phi}{d\mathbf{F}^2}(\mathbf{G}_0)[\mathbf{C}] \geq 0$$

for all strictly positive-definite, symmetric $n \times n$ matrices $\mathbf{C}$.

Note that $(1.2)_1$ is also equivalent to the convexity of the map $t \mapsto th(t^{-n})$. Thus, one could establish a similar result with $(1.2)_1$ replaced by this alternate condition. Finally, we note that any function h that satisfies (1.2) is itself convex and continuous.

References

[1] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
[2] P. G. Ciarlet, *Mathematical Elasticity, vol. 1*, North-Holland, 1988.
[3] F. R. Gantmacher, *The Theory of Matrices, vol. 1*, Chelsea, 1959.
[4] M. E. Gurtin, *An Introduction to Continuum Mechanics*, Academic Press, 1981.
[5] P. R. Halmos, *Finite-Dimensional Vector Spaces, 2^{nd} ed.*, Van Nostrand, 1958.
[6] J. M. Harris, J. L. Hirst, and M. J. Mossinghoff, *Combinatorics and Graph Theory, 2^{nd} ed.*, Springer, 2008.
[7] S. Lehmich, P. Neff, and J. Lankeit, On the convexity of the function $\mathbf{C} \mapsto f(\det \mathbf{C})$ on positive-definite matrices, *Math. Mech. Solids* **19** (2014), 369–375.
[8] A. D. Martin and V. J. Mizel, *Introduction to Linear Algebra*, McGraw-Hill, 1966.
[9] R. F. Muirhead, Proof's that the arithmetic mean is greater than the geometric mean, *Math. Gazette* **2** (1903), 283–287 (http://www.jstor.org/stable/3603556).

Department of Mathematics, Southern Illinois University, Carbondale, IL 62901, USA

E-mail address: sspector@siu.edu