Journal of Convex Analysis
Volume 9 (2002), No. 1, 269–285

An Elementary Derivation of the Generalized Kohn-Strang Relaxation Formulae

Kewei Zhang
School of Mathematical Sciences, University of Sussex,
Brighton, BN1 9QH, UK
k.zhang@sussex.ac.uk

Received March 21, 2000
Revised manuscript received July 31, 2001

We give elementary proofs of the the quasiconvex relaxation for the generalized Kohn-Strang functions [1, 2] originally studied in an optimal design problem [8]. We show that by using the translation method, we can recover the relaxations without using the homogenization method and the G-closure theory as in [1, 2]. Our calculations give further geometric insight of the relaxation and connections to other related areas.

1. Introduction

In this paper we give self contained elementary derivations (at least for the lower bounds) of various generalizations of Kohn-Strang quasiconvex relaxation formulae [8], [1, Th.2.2], [2, Th.5.3] and show that the quasiconvex envelopes can be obtained by maximizing a family of much simpler quasiconvex functions derived by using single translations. The main new observation is that it is possible to decompose the generalized Kohn-Strang function f into many simpler functions f_α, $\alpha \in \Lambda$ such that $\sup_\alpha f_\alpha = f$ and the quasiconvex envelope Qf_α for each function f_α is easily found. A surprise is that $\sup_\alpha Qf_\alpha$ gives Qf. The idea for each step is simple but the calculations might not be straight forward. However, please keep in mind that the seemingly complicated calculations are just aimed to find maximum for a family of functions.

Our main tools are linear algebra and convexification of simple functions. All the results give new representations of the corresponding quasiconvex relaxations. For $\lambda > 0$, the generalized Kohn-Strang function $f : M^{N\times n} \to \mathbb{R}_+$ is given by

$$f(A) = H(|A|) = \begin{cases} \lambda + |A|^2, & A \neq 0, \\ 0, & A = 0, \end{cases} \tag{1}$$

and its quasiconvex relaxation were studied in a recent paper [1] by Allaire and Francfort. The original function for $n = 2$ was introduced by Kohn and Strang [8] as a model in an optimal design problem in electrostatics. The quasiconvex relaxation of f when $n = 2$ is

$$Qf(A) = \begin{cases} \lambda + |A|^2, & \text{if} \quad |A|^2 + 2|\operatorname{adj}_2 A| \geq \lambda, \\ 2\sqrt{\lambda}\,(|A|^2 + 2|\operatorname{adj}_2 A|^2)^{1/2} - 2|\operatorname{adj}_2 A|, & \text{otherwise}, \end{cases} \tag{2}$$

where $\operatorname{adj}_2 A$ is the $N(N-1)/2$ vector of the 2×2 minors of $A \in M^{N\times 2}$.

ISSN 0944-6532 / $ 2.50 © Heldermann Verlag

The quasiconvex relaxation (or envelope) of the generalized Kohn-Strang function (1) for general n, $N \geq 2$ obtained in [1] is

$$Qf(A) = \begin{cases} \lambda + |A|^2, & \text{if } \sum_{i=1}^{n} a_i \geq \sqrt{\lambda}, \\ |A|^2 - (\sum_{i=1}^{n} a_i)^2 + \sqrt{\lambda} \sum_{i=1}^{n} a_i, & \text{if } \sum_{i=1}^{n} a_i \leq \sqrt{\lambda}, \end{cases} \tag{3}$$

where $a_1 \geq a_2 \geq \cdots \geq a_n \geq 0$ are the eigenvalues of $[A] = \sqrt{A^T A}$.

The results above were generalized to the case of linear strains in [2]. Let $e(A) = (A + A^T)/2$ be the linear strain in $M^{n\times n}$. Let

$$f(e(A)) = \begin{cases} |e(A)|^2 + \lambda, & \text{if } e(A) \neq 0, \\ 0, & \text{if } e(A) = 0. \end{cases} \tag{4}$$

This is a special case of a more general model considered in [2] by taking $\mu_A = 1/2$, $\lambda_A = 0$ under their notation [2,pp.459]. Let $a_1 \geq a_2 \geq \cdots \geq a_n$ be the eigenvalues of $e(A)$, then the quasiconvex relaxation of f given by (4) is

$$Rf(e(A)) = Qf(e(A)) = \begin{cases} |e(A)|^2 + \lambda, & \text{if } g^*(A) \geq \lambda, \\ |e(A)|^2 + 2\sqrt{\lambda g^*(A)} - g^*(A), & \text{if } g^*(A) \leq \lambda, \end{cases} \tag{5}$$

where

$$g^*(A) = \frac{1}{2}\left[\left(\sum_{i=1}^{n} |a_i|\right)^2 + \left(\sum_{i=1}^{n} a_i\right)^2\right]. \tag{6}$$

Let me explain our approach stated at the beginning of this paper more precisely. We borrow the idea of optimal translation method described in [5] to recover Qf. However, we do not apply the approach in [5] direcly. Instead, we introduce a family of functions f_X for each direction $X \in M^{N\times n}$, $|X| = 1$ such that $f_X \leq f$ and f_X agrees with f only on the one dimensional space spanned by X. For each X, we calculate Qf_X easily by using a fixed simple translation. Then we show that the quasiconvex relaxation (3) obtained in [1] is the maximum of these Qf_X. This links formula (3) to the relaxation of the two-matrix problem studied in [7]. In the 2×2 case, we show that the relaxation is the maximum of quasiconvex relaxation of only two simple functions whose relaxations are easy to calculate. For the case of linear strains (4) and (5), we can use a similar method to reach Qf. Our approach does not rely on any knowledge of homogenization, G-closure or bounds of effective moduli which have been developed by many authors since the 1960's.

At this stage, some readers might wonder whether there is such a need to try to recover Qf by using an alternative method given that in principle, the homogenization method and the translation method by using a single translation are equivalent shown in G. Milton's work for studying problems related to optimal bounds of effective moduli [10]. However, even for translation by one single function, the resulting formula provides more geometric and analytic information for the quasiconvex lower bound thus obtained. For a continuous function $f : M^{N\times n} \to \mathbb{R}$ bounded below and for a given (quasi)-convex function, g, the translation lower bound of f by g is given by $h := C(f - g) + g \le f$. If g is a rank-one convex quadratic form (which applies to our case), the first obvious observation is that h is the sum of a convex function and a rank-one convex quadratic form, hence not only h is quasiconvex but also the gradient Dh is quasimonotone [13] in the sense that

$$\int_U Dh(A + D\phi(x)) \cdot D\phi(x)dx \ge 0$$

for every $A \in M^{N\times n}$, $U \subset \mathbb{R}^n$ open and $\phi \in C_0^\infty(U, \mathbb{R}^N)$. This fact is important in the study of the critical points of the variational integral $I(u) = \int_\Omega h(Du) + f \cdot udx$ other than minimizers [14]. Other methods might lead to the same formula in a different form [7] and these facts could not be seen easily. We explain this point further in Remark 3.11 just before the Appendix.

The advantages of the (optimal) translation method are (i) when the translation functions are known, the method will give a direct geometric construction of the bounds and (ii) it applies to some nonlinear functions which are not of quadratic growth. An interesting application of the optimal translation method is concerned with the equality among various semiconvex hulls for compact sets $K \subset M^{N\times n}$ and semiconvex envelopes [15]. Let

$$qr(f) = \sup\{g \le f,\ g \ \text{ rank-one convex quadratic function}\}.$$

Note that $qr(f)$ is equivalent to the optimal translation bound by rank-one convex quadratic forms defined in [5]. For a continuous function $f : M^{N\times n} \to \mathbb{R}$ bounded below: $f(X) \ge c|X|^2 - C_1$, one has $C(f) \le qr(f) \le Q(f) \le R(f) \le f$. Among other results, it was established in [15] that $R(f) = C(f)$ if and only if $qr(f) = C(f)$.

This implies that by using the optimal translation method one can tell whether the quasiconvex relaxation is trivial. It is however difficult to establish such a result by using homogenization theory and the G-closure method. Note that in the statement above, we do not assume any upper bound for f. Now we state our main results. Let

$$E_+ = \left\{ \begin{pmatrix} a & b \\ -b & a \end{pmatrix},\ a,\ b \in \mathbb{R} \right\}, \quad E_- = \left\{ \begin{pmatrix} a & b \\ b & -a \end{pmatrix},\ a,\ b \in \mathbb{R} \right\}$$

be the subspaces of conformal and anti-conformal matrices respectively in $M^{2\times 2}$ - the spaces of all 2 × 2 matrices with Euclidean norm. Let $P_{E+} : M^{2\times 2} \to E_+$, $P_{E-} : M^{2\times 2} \to E_-$ be the orthogonal projections respectively. It is well known and easy to check that

$$|P_{E_+}(A)|^2 - |P_{E_-}(A)|^2 = 2\det A, \text{ for } A \in M^{2\times 2}.$$

Theorem 1.1. *Suppose f is given by (1) with $N = n = 2$, then*

$$Qf(A) = \max\{Qf_+(A),\ Qf_-(A)\},$$

with

$$f_+(A) = H(|P_{E_+}(A)|) + |P_{E_-}(A)|^2, \qquad f_-(A) = H(|P_{E_-}(A)|) + |P_{E_+}(A)|^2, \tag{7}$$

and

$$Qf_+(A) = g(|P_{E_+}(A)|) - 2\det A, \qquad Qf_-(A) = g(|P_{E_-}(A)|) + 2\det A, \tag{8}$$

where

$$g(t)=C(H(t)+t^2)=\begin{cases}4\sqrt{\frac{\lambda}{2}}t, & \textit{if } 0\le t\le\sqrt{\frac{\lambda}{2}},\\ t^2+\lambda, & \textit{if } t\ge\sqrt{\frac{\lambda}{2}}.\end{cases}\tag{9}$$

Remark 1.2. In [5], Qf in Theorem 1.1 was obtained by using a single translation

$$q(A)=2\operatorname{tr}(A^TA)^{1/2}-2|\det A|.$$

However, it was not explained in [5] as how this translation is constructed, nor the translation q above sheds any light on how to find such a q for the higher dimensional cases.

To state our result for the general $N\times n$ case, let us introduce some notation and a simple result in linear algebra. Let $X\in M^{N\times n}$ with $|X|=1$, we denote by P_X and $P_{X^\perp}$ the orthogonal projections onto the one-dimensional space span(X) and its orthogonal complement. If X is a rank-one matrix with $|X|=1$, we let $\lambda_X=0$, otherwise, we define

$$\frac{1}{\lambda_X}=\max_{|a|=|b|=1}\frac{|P_X(a\otimes b)|^2}{|P_{X^\perp}(a\otimes b)|^2}<+\infty,$$

where $a\in\mathbb{R}^N$, $b\in\mathbb{R}^n$. It is easy to see that $1/\lambda_X$ is finite because

$$|P_{X^\perp}(a\otimes b)|^2\ge c(X)|a|^2|b|^2$$

for some constant $c(X)>0$ when rank$(X)>1$. We have the following characterization of λ_X. For the proof, see the Appendix.

Proposition 1.3. *Suppose* rank$(X)>1$ *with* $|X|=1$. *Then* $\lambda_X=1/x_1^2-1>0$, *where* $0<x_1<1$ *is the largest eigenvalue of* $[X]=\sqrt{X^TX}$.

For an $N\times n$ real matrix A, we define $\mathcal{R(A)}$ and $\mathcal{R}([\mathcal{A}])$ to be the range of A and $[A]$ respectively. The following result should be well-known. However, for the convenience of the reader, I give a proof in the Appendix.

Proposition 1.4. *Let* $A\in M^{N\times n}$. *Then there is a partial isometry* $R_A\in M^{N\times n}$ *from* $\mathcal{R}([\mathcal{A}])$ *to* $\mathcal{R}(\mathcal{A})$ *such that* $|R_Ay|=|y|$ *for* $y\in\mathcal{R}([\mathcal{A}])$ *and* $R_Ay=0$ *if* $y\in(\mathcal{R}([\mathcal{A}]))^\perp$. *Furthermore,* rank$(R_A)=$ rank$([A])=$ rank(A) *and* $R_A^TR_A=P_{\mathcal{R}([\mathcal{A}])}$ *- the orthogonal projection from* $\mathbb{R}^n$ *to* $\mathcal{R}([\mathcal{A}])$.

Theorem 1.5. *Let* $f: M^{N\times n}\to\mathbb{R}_+$ *be defined by (1) with* $n\ge 2$, *then*

$$Qf(A)=\max_{|X|=1}Qf_X(A),\tag{10}$$

and the maximum is achieved at $X=R_A/|R_A|$, *where* $f_X(A)=f(P_X(A))+|P_{X^\perp}(A)|^2$ *and*

$$\begin{aligned}Qf_X(A)&=g_X\left(\frac{|P_X(A)|}{x_1}\right)+|A|^2-\frac{|P_X(A)|^2}{x_1^2}\\&=\begin{cases}|A|^2+2\sqrt{\lambda}\left(\frac{|X\cdot A|}{x_1}\right)-\left(\frac{|X\cdot A|}{x_1}\right)^2, & \textit{if } 0\le\frac{|X\cdot A|}{x_1}\le\sqrt{\lambda},\\ |A|^2+\lambda, & \textit{if } \frac{|X\cdot A|}{x_1}\ge\sqrt{\lambda}.\end{cases}\end{aligned}\tag{11}$$

with

$$g_X(t) = \begin{cases} 2(\sqrt{\lambda})t & \textit{if}\, 0 \le t \le \sqrt{\lambda}, \\ t^2+\lambda & \textit{if}\, t \ge \sqrt{\lambda} \end{cases} \tag{12}$$

Let $E \subset M^{n\times n}$ be the subspace of all skew-symmetric real matrices. The orthogonal complement $E^\perp$ of E is the subspace of all symmetric matrices. We also see that E does not have rank-one matrices and $e(A) = P_{E^\perp}(A)$ is the orthogonal projection onto $E^\perp$. Let $A \in E^\perp$ and $A \neq 0$, we may write $A = R^T \operatorname{diag}(a_1, \dots, a_n)R$, and let $S_A = R^T \operatorname{diag}(\operatorname{sgn}(a_1), \dots \operatorname{sgn}(a_n))R$, where $\operatorname{diag}(a_1, \dots, a_n)$ denotes a diagonal matrix with diagonal entries $a_1, \dots, a_n$. For a lower semicontinuous function $f : E^\perp \to \mathbb{R}$, we still denote by $Qf(e(A))$ the quasiconvex relaxation of f as a function of A, that is, if we let $F(A) = f(e(A))$, then we let $Qf(e(A)) := QF(A)$. We have

Theorem 1.6. *Let f be given by (4). Then*

$$Qf(e(A)) = \max_{|X|=1,\, X\in E^\perp} Qf_X(e(A)),$$

where $f_X(e(A)) = f(e(P_X(A)) + |e(P_{X^\perp}(A)|^2$ and

$$\begin{aligned} Rf_X(e(A)) &= Qf_X(e(A)) = |e(A)|^2 + H\left(\frac{|X\cdot A|}{\xi_X}\right) \\ &= \begin{cases} |e(A)|^2 + 2\sqrt{\lambda}\left(\frac{|X\cdot A|}{\xi_X}\right) - \left(\frac{|X\cdot A|}{\xi_X}\right)^2, & \textit{if}\ \ 0 \le \frac{|X\cdot A|}{x_1} \le \sqrt{\lambda}, \\ |e(A)|^2 + \lambda, & \textit{if}\ \ \frac{|X\cdot A|}{\xi_X} \ge \sqrt{\lambda}, \end{cases} \end{aligned} \tag{13}$$

with

$$\xi_X^2 = \max_{|a|=|b|=1} \frac{2(Xb\cdot a)^2}{1+(a\cdot b)^2},$$

where $H(t) = 2t\sqrt{\lambda} - t^2$ if $0 \le t \le \sqrt{\lambda}$ and $H(t) = \lambda$ if $t \ge \sqrt{\lambda}$. Further more the maximum is reached at some X.

In Section 2, some basic preliminaries are given. We then establish our main result through several lemmas in Section 3. In the Appendix, we give proofs of some simple results in linear algebra which are used in this note including Propositions 1.3 and 1.4 and our key lemma (Lemma 3.9) for the case of linear strains.

2. Preliminaries

We denote by $M^{N\times n}$ the space of all real $N\times n$ matrices with Euclidean inner product $A\cdot B = \operatorname{tr}(A^TB)$, where tr is the trace operator and A^T the transpose of A. We also let $|A|$ to be the norm of A and define $[A] = \sqrt{A^TA}$, which is a non-negative defined $n\times n$ matrix.

Let $f : M^{N\times n} \to \mathbb{R}$ be a continuous function. f is quasiconvex (c.f. [3, 11, 4]) in $M^{N\times n}$ if for every open and bounded subset Ω of $\mathbb{R}^n$, every $P \in M^{N\times n}$ and every $\phi \in C_0^\infty(\Omega, \mathbb{R}^N)$,

$$\int_\Omega f(P + D\phi(x))dx \ge \int_\Omega f(P)dx.$$

The class of quasiconvex functions is independent of the choice of Ω.

A function $f : M^{N\times n} \to \mathbb{R}$ is called rank-one convex if for any A, $B \in M^{N\times n}$ with $\operatorname{rank}(A-B)=1$ and any $0 \le \alpha \le 0$,

$$f(\alpha A + (1-\alpha)B) \le \alpha f(A) + (1-\alpha) f(B).$$

It is known that Quasiconvexity implies rank-one convexity [11, 3, 4] but the converse is not true [12]. For a given function $f : M^{N\times n} \to \mathbb{R}$, we can consider its quasiconvex relaxation Qf, its rank-one convex relaxation Rf and its convex relaxation Cf [4] as:

$$\begin{aligned}&Qf = \sup\{g \le f;\ g \text{ quasiconvex }\}, \quad Rf = \sup\{g \le f;\ g \text{ rank-one convex }\},\\ &Cf = \sup\{g \le f;\ g \text{ convex }\}.\end{aligned}$$

We need the following result which was essentially due to von Neumann (see [9]). We will deduce the result by using [9].

Proposition 2.1. *Suppose A, $B \in M^{N\times n}$ and $a_1 \ge \cdots \ge a_n \ge 0$ and $b_1 \ge \cdots \ge b_n \ge 0$, are the eigenvalues of $[A]$ and $[B]$ respectively. Then*

$$|\operatorname{tr}(A^T B)| \le \sum_{i=1}^{n} a_i b_i.$$

We conclude this section by introducing the iteration method [8] in calculating Rf for a lower semicontinuous function $f : M^{N\times n} \to R$ bounded below, namely,

$$\begin{cases} R_0 f = f,\\ R_{k+1} f(A) = \inf\{\lambda R_k f(A_1) + (1-\lambda) R_k f(A_2),\\ \lambda A_1 + (1-\lambda)A_2 = A, \quad \operatorname{rank}(A_1 - A_2) \le 1\}.\end{cases}$$

Then it was proved in [8] that $Rf = \lim_{k\to\infty} R_k f$.

Notice that the Kohn-Strang construction above applies to convex relaxation Cf. In the iteration scheme above we only need to drop the restriction $\operatorname{rank}(A_1 - A_2) \le 1$ to obtain a characterization of convexification of f by using the limit of $C_k f$, that is, $Cf = \lim_{k\to\infty} C_k f$. We also notice that $C_k f \le R_k f$ by definition.

3. Proofs of The Results

We prove Theorem 1.1 through Lemma 3.1 and Lemma 3.2.

Lemma 3.1. *Let $\Lambda = \{a \otimes b,\ a,\ b \in \mathbb{R}^2\}$. Then both $P_{E_+} : \Lambda \to E_+$ and $P_{E_-} : \Lambda \to E_-$ are onto mappings.*

Proof. This is easy. Let

$$E_1 = \frac{\sqrt{2}}{2} I,\ E_2 = \frac{\sqrt{2}}{2}\begin{pmatrix} 0 & 1\\ -1 & 0\end{pmatrix},$$

where I is the identity matrix. Then E_1, E_2 form an orthonormal basis of E_+. Let $a = (1,0)$ and $b = \sqrt{2}(t,s)$, we have $P_+(a\otimes b) = tE_1 + sE_2$. The first claim is proved. The proof for P_{E_-} is similar. □

Lemma 3.2. *Let $f_+(\cdot)$ be defined by (7). Then*

$$Qf_+(A) = g(|P_{E_+}(A)|) - 2\det A, \qquad Qf_-(A) = g(|P_{E_-}(A)|) + 2\det A,$$

where $g(t)$ is given by (9).

Proof. It is easy to see that $g(|P_{E_+}(A)|) = C_{E_+}\big(h(P_{E_+}(A))\big)$ where $h(P_{E_+}(A)) = f(P_{E_+}(A) + |P_{E_+}(A)|^2$, and $f_+(A) = h(P_{E_+}(A)) - 2\det A$. Also $f_+(A) \geq g(|P_{E_+}(A)|) - 2\det A$, while the right hand side of the above inequality is quasiconvex (and also rank-one convex), hence $Rf_+(A) \geq Qf_+(A) \geq g(|P_{E_+}(A)|) - 2\det A$. Now, we show that the reversed inequality holds. We apply Proposition 1.3 and consider $R_1 f_+(A)$. Let $0 < \alpha < 1$ and let B_1, $B_2 \in E_+$ be such that $\alpha B_1 + (1-\alpha)B_2 = P_{E_+}(A)$. Let $B = B_1 - B_2$, we have, from Lemma 3.1 that there is some $a \otimes b \in \Lambda$ such that $P_{E_+}(a \otimes b) = B$. Now we set $A_1 = A + (1-\alpha)a \otimes b$, $A_2 = A - \alpha a \otimes b$, we have $A_1 - A_2 = a \otimes b$, $P_+(A_1) = B_1$, $P_+(A_2) = B_2$. Therefore

$$\begin{aligned}
&R_1 f_+(A) \leq \alpha f_+(A_1) + (1-\alpha)f_+(A_2) = \alpha f_+(A + (1-\alpha)a \otimes b) + (1-\alpha)f_+(A - \alpha a \otimes b)\\
&= \alpha h(P_+(A + (1-\alpha)a \otimes b)) + (1-\alpha)h(P_+(A - \alpha a \otimes b))\\
&- 2\alpha \det(A + (1-\alpha)a \otimes b) - 2(1-\alpha)\det(A - \alpha a \otimes b)\\
&= \alpha h(B_1) + (1-\alpha)h(B_2) - 2\det A.
\end{aligned}$$

Taking infimum on B_1, B_2 in E_+ with $\alpha B_1 + (1-\alpha)B_2 = P_{E_+}(A)$ for some $0 \leq \alpha \leq 1$, we see that

$$R_1 f_+(A) \leq C_1 h(P_{E_+}(A)) - 2\det A.$$

Repeating this process, we see that

$$R_k f_+(A) \leq C_k h(P_{E_+}(A)) - 2\det A$$

for $k = 1, 2, \ldots$. Passing to the limit $k \to \infty$, we obtain $Rf_+(A) \leq C_{E_+} h(P_{E_+}(A)) - 2\det A$. Hence

$$Rf_+(A) = Qf_+(A) = g(|P_{E_+}A|) - 2\det A.$$

The calculation for $Qf_-(\cdot)$ is similar. □

Proof of Theorem 1.1. We only need to show that $Qf(A) = \max\{Qf_+(A), Qf_-(A)\}$. Let $|P_{E_+}(A)| = t$, $|P_{E_-}(A)| = s$. Then $Qf_+(A) = g(t) + s^2 - t^2$, $Qf_-(A) = g(s) + t^2 - s^2$. We prove that $Qf_+(A) \leq Qf_-(A)$ if and only if $s \geq t$. We have $g(t) + s^2 - t^2 \leq g(s) + t^2 - s^2$ if and only if $g(t) - 2t^2 \leq g(s) - 2s^2$, which is equivalent to that $g(t) - 2t^2$ is increasing. The later statement can be easily checked by differentiating the function. Thus we have established that $\max\{Qf_+(A), Qf_-(A)\} = Qf_+(A)$ if and only if $|P_{E_+}(A)| \geq |P_{E_-}(A)|$, so that $\sqrt{2}|P_{E_+}(A)| = \sqrt{2|P_{E_+}(A)|^2} = (|A|^2 + 2|\det A|)^{1/2}$. It is easy to see then that

$$\max\{Qf_+(A), Qf_-(A)\} = Qf_+(A) = g(|P_{E_+}(A)|) - 2|\det A| = Qf(A).$$

If $|P_{E_+}(A)| \leq |P_{E_-}(A)|$,

$$\max\{Qf_+(A), Qf_-(A)\} = Qf_-(A) = g(|P_{E_+}(A)|) - 2|\det A| = Qf(A).$$

The proof is finished. □

Next we prove Theorem 3.2 through Lemma 3.3 and Lemma 3.4.

Lemma 3.3. *For $X \in M^{N\times n}$ with $|X| = 1$, $Qf_X(\cdot)$ is given by (11) with $g_X(\cdot)$ given by (12).*

Proof. If $\text{rank}(X) > 1$, let x_1 be the greatest eigenvalue of $[X] = \sqrt{X^TX}$. We have, from Proposition 1.3 that $\lambda_X = 1/x_1^2 - 1$. From the definition of λ_X, we also see that $|P_{X^\perp}(\cdot)|^2 - \lambda_X|P_X(\cdot)|^2$ is a rank-one convex quadratic form, so is quasiconvex [3]. In fact, for every rank-one matrix $a \otimes b \in M^{N\times n}$, $|P_{X^\perp}(a \times b)|^2 - \lambda_X|P_X(a \times b)|^2 \geq 0$ and we if we let $\Lambda_X = \{a \otimes b,\ |P_{X^\perp}(a \otimes b)|^2 - \lambda_X|P_X(a \otimes b)|^2 = 0\}$, then $P_X : \Lambda \to \text{span}(X)$ is onto by taking $a = Xu$, $b = tu$ for $t \in \mathbb{R}$. Thus a similar argument as in the proof of Lemma 3.2 gives

$$Rf_X(A) = Qf_X(A) = C_X\left(H(|P_X(A)|^2 + \lambda_X|P_X(A)|^2\right) + (|P_{X^\perp}(A)|^2 - \lambda_X|P_X(A)|^2), \quad (14)$$

where $C_X(H(|P_X(A)|) + \lambda_X|P_X(A)|^2 = g_X(|X \cdot P|/x_1)$ is the convex relaxation along the one dimensional space $\text{span}(X)$. The calculation of the convexification is by examining tangent lines of $t^2 + \lambda$ passing through the origin and is left to the reader.

If $\text{rank}(X) = 1$, $X = a_0 \otimes b_0$, then $\lambda_X = 0$. We can easily show that

$$Rf_X(A) = Qf_X(A) = C_X(H(|P_X(A)|)) + |P_{X^\perp}(A)|^2$$

by using a similar method as in the proof of Lemma 3.2. In fact, if $P_X(A) = t_0X$ and $X_1 = t_1X$, $X_2 = t_2X \in \text{span}(X)$, with $\alpha X_1 + (1-\alpha)X_2 = t_0X = P_X(A)$, for some $0, \alpha < 1$, then $X_1 - X_2$ is obviously a rank-one matrix. Let $A_1 = A + (1-\alpha)X_1$, $A_2 = A - \alpha X_2$, then $\text{rank}(A_1 - A_2) = 1$ and $\alpha A_1 + (1-\alpha)A_2 = A$. Furthermore,

$$P_{X^\perp}(A_1) = P_{X^\perp}(A_2) = P_{X^\perp}(A), \quad P_X(A_1) = X_1, \quad P_X(A_2) = X_2.$$

Hence

$$R_1f_X(A) \leq \alpha f_X(A_1) + (1-\alpha)f(A_2) = \alpha H(|X_1|) + (1-\alpha)H(|X_2|) + |P_{X^\perp}(A)|^2,$$

so that $R_1f_X(A) \leq C_1(H(|P_X(A)|)) + |P_{X^\perp}(A)|^2$. Repeating this we see that

$$Rf_X(A) \leq C_X\left(H(|P_X(A)|) + |P_{X^\perp}(A)|^2.\right.$$

Since the reversed inequality is trivially true, we reach our conclusion.

In both cases we have a unified formula (12) for Qf_X. □

Lemma 3.4. *For a fixed $A \neq 0$, the maximum $\max_{|X|=1} Qf_X(A)$ achieves at $\hat{R}_A = R_A/|R_A|$, where R_A is given by Proposition 1.4 and*

$$Qf(A) = Qf_{\hat{R}_A}(A). \quad (15)$$

Proof. We first establish (15) with Qf given by (3), which implicitly establishes the fact that $Qf_X(A)$ reaches its maximum at $X = \hat{R}_A$. However, we will give a direct proof that the maximum is attained at $\hat{R}_A$ in Proposition 3.5 following the proof of Lemma 3.4.

Recall (11) that

$$Rf_X(A) = Qf_X(A) = \begin{cases} |A|^2 + 2\sqrt{\lambda}\left(\frac{|X\cdot A|}{x_1}\right) - \left(\frac{|X\cdot A|}{x_1}\right)^2, & \text{if } 0 \leq \frac{|X\cdot A|}{x_1} \leq \sqrt{\lambda}, \\ |A|^2 + \lambda, & \text{if } \frac{|X\cdot A|}{x_1} \geq \sqrt{\lambda}. \end{cases}$$

We notice that $\operatorname{rank}(R_A) = \operatorname{rank}([A]) = \operatorname{rank}(A) := m \geq 1$. So $|R_A| = \sqrt{m}$ and the all non-zero eigenvalues of $[\hat{R}_A]$ equal to $1/\sqrt{m}$. In fact $[\hat{R}_A] = \sqrt{\hat{R}_A^T\hat{R}_A} = P_{\mathcal{R}([A])}/\sqrt{m}$. Hence if we substitute $X = \hat{R}_A$ in the formula above and write $A = R_A[A]$, firstly, we have $\lambda_{\hat{R}_A} = m-1$. Secondly, we have

$$|P_{\hat{R}_A}(A)|^2 = \left(\frac{\operatorname{tr}(R_A^T R_A[A])}{\sqrt{m}}\right)^2 = \left(\frac{\operatorname{tr}(P_{\mathcal{R}([A])}[A])}{\sqrt{m}}\right)^2 = \frac{(\operatorname{tr}[A])^2}{m}.$$

Hence

$$Qf_{\hat{R}_A}(A) = g_{\hat{R}_A}\left(\frac{\operatorname{tr}[A]}{\sqrt{m}}\right) + |A|^2 - (\operatorname{tr}[A])^2, \tag{16}$$

with

$$g_{\hat{R}_A}\left(\frac{\operatorname{tr}[A]}{\sqrt{m}}\right) = \begin{cases} 2\sqrt{\lambda}\operatorname{tr}[A], & \text{if } 0 \leq \operatorname{tr}[A] \leq \sqrt{\lambda}, \\ (\operatorname{tr}[A])^2 + \lambda, & \text{if } \operatorname{tr}[A] \geq \sqrt{\lambda}. \end{cases} \tag{17}$$

Combining (16) and (17) and recalling our notation $\operatorname{tr}[A] = a_1 + \cdots + a_n$ we may claim that at A, if we compare (3) and (16)-(17), we have $Qf(A) = Qf_{\hat{R}_A}(A)$. Implicitly, we have also proved that $\max_{|X|=1} Qf_X(A) = Qf_{\hat{R}_A}(A)$ because for each X with $|X| = 1$, $f_X(\cdot) \leq f(\cdot)$, so $\max_{|X|=1} Qf_X(A) \leq Qf(A)$. Since at $\hat{R}_A$, the upper bound $Qf(A)$ is reached, we see that $Qf_{\hat{R}_A}(A)$ is a maximum.

The proof of Lemma 3.4 and hence Theorem 1.5 is complete. □

If we did not know formula (3) for Qf, we can still show that

Proposition 3.5. *We have* $\max_{|X|=1} Qf_X(A) = Qf_{\hat{R}_A}(A)$.

Proof. In the last part of (11), if we let $t = \frac{|X\cdot A|}{x_1}$ and define

$$F(t) = \begin{cases} 2\sqrt{\lambda}t - t^2, & \text{if } \ 0 \leq t \leq \sqrt{\lambda}, \\ \lambda & \text{if } \ t \geq \sqrt{\lambda}, \end{cases}$$

we see that $F(t)$ is increasing. So the maximum of $Qf_X(A)$ is attained if we can find $\max_{|X|=1}(|X\cdot A|/x_1)$. Now we apply Proposition 2.1 to X and A to obtain

$$\frac{|X\cdot A|}{x_1} = \frac{|\operatorname{tr}(X^TA)|}{x_1} \leq \sum_{i=1}^n a_i\frac{x_i}{x_1} \leq \sum_{i=1}^n a_i.$$

This is because x_1 is the largest eigenvalue of $[X]$. We can easily check that this upper bound can be reached by taking $X = \hat{R}_A$ and by assuming that $\operatorname{rank}(A) = m \leq n$. Therefore $Qf_X(A)$ is maximized at $\hat{R}_A$. The proof is finished. □

Remark 3.6. Without knowing the quasiconvex relaxation formula (3), we can only claim that the right hand side of (3), which is alternatively obtained by maximizing simple quasiconvex functions, is just a quasiconvex lower bound. So we still need to use 'Step 3' of the one page proof in [1,pp.313–314] to show that the right hand side of (3) is a rank-one convex upper bound to finish the proof. However, that part in [1] is also elementary, by which I mean that it does not involve either G-closure or homogenization arguments. □

Notice that in Lemma 3.3, the function H needs not to be quadratic. In fact we can obtain the quasiconvex relaxation of f_X by using a single 'universal' translation. The following example serves as another justification for using the translation method where I do not know how to use the G-closure theory.

Example 3.7. Suppose $X \in M^{N\times n}$ with $|X| = 1$ and $\operatorname{rank}(X) > 1$. Then for any non-negative continuous function $g : \operatorname{span}(X) \to \mathbb{R}$, the quasiconvex envelope of $G_X : M^{N\times n}$ defined by $G_X(A) = g(P_X(A)) + |P_{X^\perp}(A)|^2$ is given by

$$QG_X(A) = C_X(g(P_X(A)) + \lambda_X|P_X(A)|^2) + [|P_{X^\perp}(A)|^2 - \lambda_X|P_X(A)|^2)].$$

This result follows directly from the proof of Lemma 3.3.

Let us use this general formula to find the quasiconvex relaxation for some slightly more general two-matrix models than that of Kohn [7]. Suppose X be as above and

$$g(t) = \lambda_X \frac{2^{p-1}}{p} \min\{|t-1|^p,\ |t+1||^p\}$$

with $p > 0$. Let us consider $G_X(A) = g(P_X(A)) + |P_{X^\perp}(A)|^2$. Here we have misused notation and identified $g(P_X(A))$ as $g(X \cdot A)$. The coefficient of g is just to make the calculation of $C[g(t)+\lambda_X|t|^2)$ easier. We then have, by using the relaxation formula above that

$$QG_X(A) = \begin{cases} G_X(A) & |P_X(A)| \geq \dfrac{1}{2}, \\ \lambda_X(\dfrac{1}{2^2} + \dfrac{1}{2p}) - \lambda_X|P_X(A)|^2 + |P_X^\perp(A)|^2. \end{cases}$$

Note that both G_X and QG_X vanish precisely at X and $-X$ and along the X-direction the function is of p-th growth at infinity. For example, when $t > 1$, $QG_X(tX) = g(t) = \lambda_X \frac{2^{p-1}}{p}|t-1|^p$. If $0 < p < 1$, $QG_X(tX)$ is of sublinear growth.

Now we turn to the proof of Theorem 1.6.

Lemma 3.8. *The quasiconvex and rank-one relaxations of f_X in Theorem 1.6 are both given by (13), where $X \in E^\perp$, $|X| = 1$.*

Proof. We use the same method as in the proof of Lemma 3.4. We recall that $\operatorname{span}(X)$ has rank-one connections in the linear elasticity setting if either X is a rank-one matrix or X is rank-two while the two non-zero eigenvalues have opposite signs [7]. In this case, there is a rank-one matrix $a \otimes b \in M^{n\times n}$ such that $P_{X^\perp\cap E^\perp}(a\otimes b) = 0$ and we define $\lambda_X = 0$. If $\operatorname{rank}(X) > 2$ or $\operatorname{rank}(X) = 2$ but the nonzero-eigenvalue have the same sign, we see that for some $C_X > 0$, one has $|P_{X^\perp\cap E^\perp}(a\otimes b)|^2 \geq C_X|a|^2|b|^2$ for all rank-one matrices $a \otimes b$. We then define

$$\frac{1}{\lambda_X} = \sup_{|a|=|b|=1} \frac{|P_X(a\otimes b)|^2}{|P_{X^\perp\cap E^\perp}(a\otimes b)|^2} = \frac{1}{\inf_{|a|=|b|=1} \frac{|P_{X^\perp\cap E^\perp}(a\otimes b)|^2}{|P_X(a\otimes b)|^2}}.$$

Notice that

$$\begin{aligned} |P_{X^\perp\cap E^\perp}(a\otimes b)|^2 &= |P_{E^\perp}(a\otimes b)|^2 - |P_X(a\otimes b)|^2 \\ &= \frac{1}{2}\left(|a|^2|b|^2 + 2(a\cdot b)^2\right) - |P_X(a\otimes b)|^2 \geq 0, \end{aligned} \tag{18}$$

we have, when $|a| = |b| = 1$, that

$$\lambda_X = \inf_{|a|=|b|=1} \frac{1}{2}\frac{(1+(a\cdot b)^2)}{|P_X(a\otimes b)|^2} - 1 = \frac{1}{\xi_X^2} - 1$$

where

$$\xi_X^2 = 2 \sup_{|a|=|b|=1} \frac{|P_X(a\otimes b)|^2}{1+(a\cdot b)^2}.$$

We can then claim that the quadratic form

$$q_X(A) = |P_{X^\perp \cap E^\perp}(A)|^2 - \lambda_X |P_X(A)|^2$$

is rank-one convex and there is some rank-one matrix $a\otimes b$ such that $q_X(a\otimes b) = 0$. Now we have

$$Qf_X(e(A)) \geq C\left(f(P_X(A) + \lambda_X|P_X(A)|^2\right) + q_X(A)$$

and the right hand side of the above inequality is quasiconvex. Since $\mathrm{span}(X)$ is one dimensional and $M = \{(a\otimes b),\, q_X(a\otimes b) = 0\} \neq \{0\}$, the projection $P_X : M \to \mathrm{span}(X)$ is onto, so we can show as in the proof of Theorem 1.1 that

$$Rf_X(e(A)) \leq C\left(f(P_X(A) + \lambda_X|P_X(A)|^2\right) + q_X(A).$$

A simple calculation of the above convexification gives (13). □

Now we establish Theorem 1.6. Suppose $x \in \mathbb{R}$, we let $x_+ = x$ if $x \geq 0$ and $x_+ = 0$ if $x < 0$. Similarly, we let $x_- = -(-x)_+$. The following result gives the value of ξ_X in Theorem 1.6.

Lemma 3.9. *Let $X \in E^\perp$ with $|X| = 1$. Suppose $x_1 \geq x_2 \geq \cdots \geq x_n$ are the eigenvalues of X. Let $a,\, b \in \mathbb{R}^n$. Then*

$$\xi_X^2 = 2 \max_{|a|=|b|=1} \frac{(a^T X b)^2}{1+(a\cdot b)^2} = \max_{1\leq i,j\leq n}\{(x_i)_+^2 + (x_j)_-^2\} := \eta_X^2. \tag{19}$$

We see from (18) that $\xi_X \leq 1$, and the equality holds if and only if X is compatible, that is, either $\mathrm{rank}(X) = 1$ or $\mathrm{rank}(X) = 2$ and the non-zero eigenvalues of X have opposite signs. We establish Lemma 3.9 in the Appendix. Note that Lemma 3.9 is interesting by its own right. If we view $a^T X b$ above as a symmetric bilinear form on $\mathbb{R}^n$, (19) is equivalent to

$$|a^T X b| \leq \frac{\eta_X}{\sqrt{2}}\sqrt{|a|^2|b|^2 + (a\cdot b)^2},$$

and the coefficient is optimal. If X is positive definite, one has $\eta_X = \lambda_{\max}$, the largest eigenvalue of X and the inequality above becomes

$$|a^T X b| \leq \frac{\lambda_{\max}}{\sqrt{2}}\sqrt{|a|^2|b|^2 + (a\cdot b)^2}.$$

This is a sharp generalization of the Cauchy-Schwartz inequality and can give improved Hilbert inequality in l^2 [16].

Lemma 3.10. *For $A \in E^\perp$, $A \neq 0$, $Rf(e(A)) = Qf(e(A)) = \max_{X\in E^\perp, |X|=1} Qf_X(e(A))$ and the maximum is reached at some $X = \hat{S}_A$.*

Proof. Since for each $X \in E^\perp$ with $|X| = 1$, we have $f_X(e(A)) \leq f(e(A))$,

$$\max_{X\in E^\perp, |X|=1} Qf_X(e(A)) \leq Qf(e(A)) = Rf(e(A)).$$

As in the proof of Lemma 3.4, we observe that $H(\cdot)$ defined by (13) is an increasing function of $|X\cdot A|/\xi_A$. Notice that we only need to consider $A \in E^\perp$ so that $e(A) = A$. If the eigenvalues of A are all non-negative or non-positive, we take $S_A = I$ and $\hat{S}_A = S_A/|S_A| = I/\sqrt{n}$ so that $\xi_{S_A} = 1/\sqrt{n}$ and $|\hat{S}_A \cdot A|/\xi_{\hat{S}_A} = \mathrm{tr}[A]$. A direct calculation then gives

$$Qf(A) = Qf_{\hat{S}_A}(A) \leq \max_{|X|=1, X\in E^\perp} Qf_X(A)$$

and the required conclusion follows.

If the eigenvalues of A have different signs, we may assume that they satisfy

$$a_1 \geq \cdots a_p \geq 0 \geq a_{p+1} \geq \cdots \geq a_n.$$

We also have $A = R^T \mathrm{diag}(a_1, \ldots, a_n)R$ where R is a rotation. In this case we take $S_A = R^T \mathrm{diag}(s_1, \cdots, s_n)R$ where

$$s_k = \begin{cases} \displaystyle\sum_{i=1}^p a_i, & \text{if } 1 \leq k \leq p, \\ \displaystyle\sum_{i=p+1}^n a_i, & \text{if } p+1 \leq k \leq n. \end{cases}$$

Observe that $\sum_{i=1}^p a_i > 0$ and $\sum_{i=p+1}^n a_i < 0$, and S_A has only two distinct eigenvalues and they have opposite signs. Let $\hat{S}_A = S_A/|S_A|$, then

$$\xi^2_{\hat{S}_A} = \frac{1}{|S_A|^2}\left((\sum_{i=1}^p a_i)^2 + (\sum_{i=p+1}^n a_i)^2 \right),$$

so that

$$\frac{|\hat{S}_A \cdot A|}{\xi_{S_A}} = \sqrt{\left((\sum_{i=1}^p a_i)^2 + (\sum_{i=p+1}^n a_i)^2 \right)} = \sqrt{2g(A)}.$$

Hence we obtain $Qf_{\hat{S}_A}(A) = Qf(A)$ by a direct calculation. The proof is finished. □

We conclude by examine the well-known explicit relaxation for the double-well energy given by the squared-distance function to a two-point set [7] and explain what further information we may obtain if we use a single translation.

Remark 3.11. Let us consider

$$f(X) = \mathrm{dist}^2(X, \{A_1, A_2\}) = \min\{|X - A_1|^2, |X - A_2|^2\},$$

where dist^2 is the squared distance function from a point $X \in M^{N\times n}$ to the set $\{A_1, A_2\}$. We give a short proof of the relaxation formula Qf first. An explicit formula for Qf was obtained by Kohn [7] through Fourier analysis. We adopt the same method for the proof of Lemma 3.3.

After a simple translation in $M^{N\times n}$, we may consider an equivalent function

$$g(A) = \mathrm{dist}^2(A, \{-X_0, X_0\}).$$

If we let $X = X_0/|X_0|$, we can re-write $g(A)$ as $g(A) = \mathrm{dist}^2(P_X(A), \{-X_0, X_0\}) + |P_{X^\perp}(A)|^2$, so

$$\begin{aligned} &Rg(A) = Qg(A) \\ &= C_X\left(\mathrm{dist}(P_X(A) + \lambda_X|P_X(A)|^2\right) + [|P_{X^\perp}(A)|^2 - \lambda_X|P_X(A)|^2] \\ &= G(A) + H(A), \end{aligned}$$

where λ_X is given by Proposition 1.3. Note that $G(A)$ is a convex function and $H(A)$ a rank-one convex quadratic form. It is easy to see that Qg still has the double-well structure and the gradient of Qg is quasimonotone. Furthermore, if we consider the variational integral $I_\epsilon(u) = \int_\Omega Qg(Du) + \epsilon f \cdot u dx$ under the natural boundary condition, we can show that $I_\epsilon(\cdot)$ satisfies a weak version of the Palais-Smale compactness condition, and for sufficiently small $\epsilon > 0$, $I_\epsilon(\cdot)$ has at least three critical points - a global minimizer, a local minimizer and a mountain pass solution.

Let us compare our relaxation formula with Kohn's original result obtained by using the Fourier analysis and examine the difference.

We still consider the special case $g(A) = \mathrm{dist}^2(A, \{-X_0, X_0\})$ as above. Then the quasiconvex envelope of Qg of g is given by [7]

$$Qg(P) = \min_{0\leq\theta\leq 1}\left\{|X - (1-2\theta)X_0|^2 + 4\theta(1-\theta)[|X_0|^2 - \lambda_{\max}]\right\},$$

where $\lambda_{\max}$ is the largest eigenvalue of the matrix $X_0^T X_0$.

Although the two relaxation formulae are in fact the same, It would be more difficult to see that the second relaxation formula obtained in [7] naturally gives quasimonotone gradient, the fact that $I_\epsilon(\cdot)$ satisfies the weak PS condition is even less obvious.

Appendix

We give proofs of Propositions 1.3, 1.4, 2.1 and Lemma 3.9 in this section.

Proof of Proposition 1.3. If $\mathrm{rank}(X) > 1$, let $x_1 \geq x_2 \geq \cdots \geq x_n \geq 0$ be the eigenvalues of $[X] = \sqrt{X^T X}$. From the definition of λ_X, we see that

$$\frac{1}{\lambda_X} = \max_{|a|=|b|=1} \frac{|P_X(a\otimes b)|^2}{|P_{X^\perp}(a\otimes b)|^2} = \max_{|a|=|b|=1} \frac{|P_X(a\otimes b)|^2}{1 - |P_X(a\otimes b)|^2},$$

hence

$$\lambda_X = \frac{1}{\max_{|a|=|b|=1}|P_X(a\otimes b)|^2} - 1.$$

If we view $a \in \mathbb{R}^N$, $b \in \mathbb{R}^n$ as column vectors, we have

$$|P_X(a \otimes b)|^2 = (a^T X b)^2 \leq |a|^2 |Xb|^2 \leq x_1^2.$$

New we show that the maximum can be reached. Let u be a unit eigenvector of $[X]$ corresponding to the largest eigenvalue x_1 and let $a = Xu/x_1$, $b = u$. Then

$$|P_X(a \otimes b)|^2 = (a^T X b)^2 = (u^T X^T X u)^2 / x_1^2 = x_1^2.$$

Therefore, $\lambda_X = \frac{1}{x_1^2} - 1$. □

Proof of Proposition 1.4. We follow [6] for the $n \times n$ case and notice that $|[A]x| = |Ax|$ for every $x \in \mathbb{R}^n$. We first define R_A on $\mathcal{R}([\mathcal{A}])$ as a mapping to $\mathcal{R}(\mathcal{A})$. If $y = [A]x$, we let $R_A y = Ax$. This mapping is well-defined because if $y = [A]x_1 = [A]x_2$, we can easily seen that $Ax_1 = Ax_2$. In fact, $|A(x_1 - x_2)| = |[A](x_1 - x_2)| = 0$. It is also easy to check that $R_A : \mathcal{R}([\mathcal{A}]) \to \mathcal{R}(\mathcal{A})$ is linear and $|R_A y| = |y|$, hence R_A is an isometry. Since both $\mathcal{R}([\mathcal{A}])$ and $\mathcal{R}(\mathcal{A})$ are finite dimensional, we have $\dim(\mathcal{R}([\mathcal{A}])) = \dim(\mathcal{R}(\mathcal{A})$, hence $\mathrm{rank}(A) = \mathrm{rank}([A])$. Next we define $R_A y = 0$ if $y \in (\mathcal{R}([\mathcal{A}]))^\perp$. Thus R_A is a linear transform from $\mathbb{R}^n$ to $\mathbb{R}^N$ with $\mathrm{rank}(R_A) = \dim(\mathcal{R}(\mathcal{R}_A)) = \dim(\mathcal{R}(\mathcal{A})) = \mathrm{rank}(\mathcal{A})$.

Finally, we show that $R_A^T R_A = P_{\mathcal{R}([\mathcal{A}])}$. By the parallelogram law, we see that $(R_A y_1) \cdot (R_A z_1) = y_1 \cdot z_1$ for y_1, $z_1 \in \mathcal{R}([\mathcal{A}])$. Now for y, $z \in \mathbb{R}^n$, we have the orthogonal decomposition $y = y_1 + y_2$, $z = z_1 + z_2$ with y_1, $z_1 \in \mathcal{R}([\mathcal{A}])$ and y_2, $z_2 \in (\mathcal{R}([\mathcal{A}]))^\perp$. So

$$(R_A^T R_A y) \cdot z = (R_A y) \cdot (R_A z) = (R_A y_1) \cdot (R_A z_1) = y_1 \cdot z_1 = (P_{\mathcal{R}([\mathcal{A}])} y) \cdot z$$

for all $z \in \mathbb{R}^n$, hence $R_A^T R_A y = P_{\mathcal{R}([\mathcal{A}])} y$ for all $y \in \mathbb{R}^n$. The conclusion follows. □

Proof of Proposition 2.1. By a result in [9, pp.173]: $|\mathrm{tr}(AB)| \leq \sum_{i=1}^n a_i b_i$ for A, $B \in M^{n \times n}$, where $a_1 \geq \cdots \geq a_n \geq 0$, $b_1 \geq \cdots \geq b_n \geq 0$ are the eigenvalues of $[A]$ and $[B]$ respectively. To apply this result to not necessarily square matrices, we first notice the well-known fact that for $A \in M^{N \times n}$, $A^T A$ and AA^T have the same non-zero eigenvalues. Next we can always enlarge A and B to be square matrices by adding $n - N$ rows with zero entries if $N < n$, or $N - n$ columns if $n < N$, say

$$\hat{A} = \begin{pmatrix} A \\ 0 \end{pmatrix}, \quad \hat{B} = \begin{pmatrix} B \\ 0 \end{pmatrix}, \text{ if } N < n; \quad \hat{A} = (A, 0), \quad \hat{B} = (B, 0), \text{ if } N > n.$$

in both cases we see that the non-zero eigenvalues of $[\hat{A}^T]$ and $[A]$, and those of $[\hat{B}]$ and $[B]$ are the same respectively. Moreover, $\mathrm{tr}\, \hat{A}^T \hat{B} = \mathrm{tr}\, A^T B$. So the conclusion follows. □

We conclude this section by establishing Lemma 3.9.

Proof of Lemma 3.9. Up to a simple rotation, we may assume that $|X| = 1$ and $X = \mathrm{diag}(x_1, \ldots, x_n)$ is a diagonal matrix. For a fixed $a \in \mathbb{R}^n$, $|a| = 1$, we consider two different cases:

(i) Xa is parallel to a;

(ii) Xa is not parallel to a.

If (i) happens, we see that either $Xa = 0$ and such an a cannot get us the maximum, so we can safely ignore the case. If $Xa \neq 0$, all the eigenvalues are the same and equals $\pm 1/\sqrt{n}$, so

$$2\frac{(a^T Xb)^2}{1+(a,\, b)^2} = \frac{2}{n}\frac{(a,\, b)^2}{1+(a,\, b)^2} \leq \frac{1}{n} = \max_{1\leq i,\, j\leq n}\{(x_i)_+^2 + (x_j)_-^2\}, \tag{A1}$$

and the maximum is achieved at $a = b = (1, 0, \ldots, 0)^T$.

Let us consider Case (ii) which is less as trivial. Our plan is to maximize the left hand side of (19) with respect to b first, then we maximize the resulting quantity with respect to a. We apply the Gram-Schmidt process to a and Xa. Let

$$u = \frac{Xa - (Xa,\, a)a}{|Xa - (Xa,\, a)a|},$$

then a, u form an orthonormal basis of span$(a,\, Xa)$, and $Xa = (Xa,\, a)a + (Xa,\, u)u$. We can also write $b = \alpha a + \beta u + v$ where v is orthogonal to span$(a,\, Xa)$. Since $|b| = 1$, we have $\alpha^2 + \beta^2 \leq 1$ and

$$2\frac{(a^T Xb)^2}{1+(a,\, b)^2} = 2\frac{\left(\alpha(Xa,\, a) + \beta(Xa,\, u)\right)^2}{1+\alpha^2}.$$

To simplify our calculation, we may consider the function

$$g(\alpha, \beta) = \sqrt{2}\frac{\alpha(Xa,\, a) + \beta(Xa,\, u)}{\sqrt{1+\alpha^2}}, \qquad \alpha^2 + \beta^2 \leq 1,$$

and maximize it, then square the resulting maximum. The reason is that the function is odd, hence we only need to consider its maximum.

Since Xa is not parallel to a, $(Xa,\, u) \neq 0$, so g does not have any interior stationary points. Hence the maximum is on the boundary $\alpha^2 + \beta^2 = 1$. We may let $\alpha = \cos\theta$, $\beta = \sin\theta$ and find stationary points of $g(\cos\theta,\, \sin\theta)$. A simple calculation leads to the condition for stationary points:

$$-(Xa,\, a)\sin\theta + 2(Xa,\, u)\cos\theta = 0.$$

so we may choose

$$\sin\theta = 2(Xa,\, u)/\sqrt{(Xa,\, a)^2 + (2(Xa,\, u))^2},\ \cos\theta = (Xa,\, a)/\sqrt{(Xa,\, a)^2 + (2(Xa,\, u))^2}$$

and the maximum is $\max g = \sqrt{(Xa,\, a)^2 + 2(Xa,\, u)^2}$. From the definition of u, we see that $(Xa \cdot u)^2 = |Xa|^2 - (Xa,\, a)^2$, hence

$$\max_{|b|=1} 2\frac{(a^T Xb)^2}{1+(a,\, b)^2} = 2|Xa|^2 - (Xa,\, a)^2.$$

Notice now that if

$$\max_{|a|=1}\{2|Xa|^2 - (Xa,\, a)^2\} \leq \eta_X^2,$$

we reach our conclusion.

Now we maximize the right hand side of the above equality by using the Lagrangian multipliers with the constraint $|a|^2 = 1$: Let $a = (a_1, \dots, a_n)^T$, we have

$$4x_i^2 a_i - 4(Xa,\, a)x_i a_i = 2\tau a_i.$$

If $x_i = 0$ for some i while $a_i \neq 0$, we may claim that $\tau = 0$. So if we sum up the above equality from 1 to n, we obtain $4|Xa|^2 - 4(Xa,\, a)^2 = 0$ which implies that Xa is parallel to a which contradicts to our assumption for Case (ii).

Now we claim that there are less than three non-zero a_i's corresponding to distinct eigenvalues of X respectively. If the claim was not true, there are $x_i \neq x_j \neq x_k$, such that $a_l \neq 0$, with $l = i,\, j,\, k$, we have $x_i^2 - (Xa,\, a)x_i = \tau/2$, we subtract the equality corresponding to a_j to obtain

$$x_i + x_j - (Xa,\, a) = 0. \tag{A2}$$

Applying this to $i,\, j,\, k$, we have $x_i + x_j = x_i + x_k = x_k + x_j$, hence $x_i = x_j = x_k$, a contradiction.

If there is only one x_i such that a is the corresponding eigenvector, with $|a| = 1$, then $2|Xa|^2 - (Xa,\, a)^2 = x_i^2 \leq \eta_X^2$.

If there are two distinct eigenvalues x_i and x_j such that $a = v + w$ where $Xv = x_i v$ and $Xw = x_j w$ such that $|v|^2 + |w|^2 = 1$, $v \neq 0$, $w \neq 0$, then from (A2),

$$0 = x_i + x_j - (Xa,\, a) = x_i + x_j - (x_i|v|^2 + x_j|w|^2) = x_i(1 - |v|^2) + x_j(1 - |w|^2).$$

We see that x_i and x_j must have different signs. Otherwise, we have $|v|^2 = 1$ and $|w|^2 = 1$ which contradicts to $|v|^2 + |w|^2 = 1$ and $v \neq 0$, $w \neq 0$. hence we may assume that $x_i > 0$, $x_j < 0$ and solve $|v|^2$ and $|w|^2$ by x_i and x_j. We have $(1 - |w|^2)/x_i = -(1 - |v|^2)/x_j := t$, so that $1 - |w|^2 = tx_i$, $1 - |v|^2 = -tx_j$. Thus $1 = t((x_i - x_j)$, hence $t = 1/(x_i - x_j)$. We then have

$$|v|^2 = 1 + \frac{x_j}{x_i - x_j}, \qquad |w|^2 = 1 - \frac{x_i}{x_i - x_j}.$$

Consequently,

$$\begin{aligned} 2|Xa|^2 - (Xa,\, a)^2 &= 2\left(x_i^2\left(1 + \frac{x_j}{x_i - x_j}\right) + x_j^2\left(1 - \frac{x_i}{x_i - x_j}\right)\right) - (x_i + x_j)^2 \\ &= 2(x_i^2 + 2x_j^2 + x_i x_j) - (x_i + x_j)^2 = x_i^2 + x_j^2 = (x_i)_+^2 + (x_j)_-^2. \end{aligned}$$

Therefore, we may claim that

$$\max_{|a|=1}\max_{|b|=1} 2\frac{(b^T Xa)^2}{1 + (a,\, b)^2} = \max_{i,j}\{(x_i)_+^2 + (x_j)_-^2\}.$$

Now it is easy to see that $\xi_X^2 \leq \eta_X^2$, and η_X^2 can be reached by taking a and b properly. The proof is finished. □

Acknowledgements. This work is supported by the Australian Research Council through the ARC Grant Scheme.

References

[1] G. Allaire, G. Francfort: Existence of minimizers for non-quasiconvex functionals arising in optimal design, Anal. Non-Lin. H. Poincaré Inst. 15 (1998) 301–339.

[2] G. Allaire, V. Lods: Minimizers for a double-well problem with affine boundary conditions, Proc. Royal Soc. Edin. 129A (1999) 439–466.

[3] J. M. Ball: Convexity conditions and existence theorems in nonlinear elasticity, Arch. Rational Mech. Anal. 63 (1977) 337–403.

[4] B. Dacorogna: Direct Methods in the Calculus of Variations, Springer-Verlag (1989).

[5] N. B. Firoozye: Optimal use of the translation method and relaxations of variational problems, Comm. Pure Appl. Math. 44 (1991) 643–678.

[6] P. Halmos: Finite-Dimensional Vector Spaces, Van Nostrand Reinold (1958).

[7] R. V. Kohn: The relaxation of a double-well energy, Cont. Mech. Therm. 3 (1991) 981–1000.

[8] R. V. Kohn, D. Strang: Optimal design and relaxation of variational problems I, II, III, Comm. Pure Appl. Math. 39 (1986) 113–137, 139–182, 353–377.

[9] L. Mirsky: On the trace of a matrix product, Math. Nachr. 20 (1959) 171–174.

[10] G. Milton: On characterizing the set of possible effective tensors of composites: the variational method and the translation method, Comm. Pure Appl. Math. 43 (1990) 63–125.

[11] C. B. Morrey Jr.: Multiple Integrals in the Calculus of Variations, Springer-Verlag (1966).

[12] V. Šverák: Rank one convexity does not imply quasiconvexity, Proc. Royal Soc. Edin. 120A (1992) 185–189.

[13] K.-W. Zhang: On the Dirichlet problem for a class of quasilinear elliptic systems of partial differential equations in divergence form, in: Partial Differential Equations, Proc. Symp. Tianjin 1986, Lect. Notes Math. 1306, S. S. Chern (ed.), Springer-Verlag (1988) 262–277.

[14] K.-W. Zhang: A two-well structure and intrinsic mountain pass points, Calc. Var. PDEs 13 (2001) 231–246.

[15] K.-W. Zhang: On conditions for equality of relaxations in the calculus of variations, J. Nonlinear Convex Anal., to appear.

[16] K.-W. Zhang: A bilinear inequality, J. Math. Anal. Appl., to appear.