Convex SO(N) × SO(n)-invariant Functions and Refinements of Von Neumann's Inequality

Bernard Dacorogna* and Pierre Maréchal†

May 19, 2005

Résumé.– Une fonction f sur $M_{N\times n}(\mathbb{R})$ qui est $\mathrm{SO}(N)\times\mathrm{SO}(n)$-invariante est convexe si et seulement si sa restriction au sous-espace des matrices diagonales est convexe. Ceci résulte de variantes de l'inégalité de Von Neumann et fait appel, dans le cas où $N=n$, à la notion de valeur singulière signée.

Abstract.– A function f on $M_{N\times n}(\mathbb{R})$ which is $\mathrm{SO}(N)\times\mathrm{SO}(n)$-invariant is convex if and only if its restriction to the subspace of diagonal matrices is convex. This results from Von Neumann type inequalities and appeals, in the case where $N=n$, to the notion of *signed singular value*.

1 Introduction

A function $f\colon M_n(\mathbb{R})\to[-\infty,\infty]$ is said to be $\mathrm{SO}(n)\times\mathrm{SO}(n)$-invariant if

$$\forall\xi\in M_n(\mathbb{R}),\ \forall Q,R\in\mathrm{SO}(n),\quad f(Q\xi R^t)=f(\xi).$$

The specification of an $\mathrm{SO}(n)\times\mathrm{SO}(n)$-invariant function f is easily seen to be equivalent to that of a function $g\colon\mathbb{R}^n\to\mathbb{R}$ which is invariant under permutation of the components and under change of sign of an even number of components. We will be mostly concerned with following fact:

*EPFL, CH-1015 Lausanne, Switzerland. Email: bernard.dacorogna@epfl.ch.
†Université Paul Sabatier, F-31062 Toulouse cedex 4, France. Email: marechal@mip.ups-tlse.fr.

> An $\mathrm{SO}(n)\times\mathrm{SO}(n)$-invariant function f is convex if and only if its restriction to $D_n(\mathbb{R})$, the subspace of $M_n(\mathbb{R})$ of diagonal matrices, is convex.

This was established by Dacorogna and Koshigoe [3] in the case $n=2$, and later by Vincent [13] in the general case, as a consequence of the convexity theorem of Kostant [5]. An analogous statement, for convex $\mathrm{O}(n)\times\mathrm{O}(n)$-invariant functions, is well known (see Dacorogna and Marcellini [2], for example).

On the other hand, Von Neumann's trace inequality, namely,

$$\mathrm{tr}(\xi\eta^t)\le\sum_{k=1}^{n}\lambda_k(\xi)\lambda_k(\eta), \tag{1}$$

where $\lambda_1(\xi)\le\ldots\le\lambda_n(\xi)$ denote the increasingly ordered singular values of ξ, can be significantly refined. On denoting by $\mu_1(\xi),\ldots,\mu_n(\xi)$ the *signed singular values*, that is,

$$\mu_1(\xi):=\mathrm{sgn}(\det\xi)\lambda_1(\xi)\quad\text{and}\quad\mu_k(\xi):=\lambda_k(\xi)\quad\text{for}\quad k\ge 2,$$

the following holds:

$$\mathrm{tr}(\xi\eta^t)\le\sum_{k=1}^{n}\mu_k(\xi)\mu_k(\eta). \tag{2}$$

This inequality, which was first established by Rosakis [10], is strictly more stringent than that of Von Neumann, and contains it as an immediate consequence.

The purposes of this paper are the following. First, we give a variant of Rosakis' proof of Inequality (2). This variant is self-contained, in the sense that it does not use Von Neumann's inequality. Second, we establish the link between Inequality (2) and the above mentioned result on convex $\mathrm{SO}(n)\times\mathrm{SO}(n)$-invariant functions. Our strategy relies mostly on convex duality rather than Lie theoretic arguments (as in Vincent [13]). Third, we consider analogous results for rectangular matrices. In the latter case, the notion of signed singular value does not make sense but, surprisingly, the notions of $\mathrm{O}(N)\times\mathrm{O}(n)$-invariance and $\mathrm{SO}(N)\times\mathrm{SO}(n)$-invariance coincide when $N\neq n$ (see Proposition 2 below). A rectangular version of Von Neumann's trace inequality then allows to establish the desired properties.

We now introduce some notation. We denote by $M_{N\times n}(\mathbb{R})$ and $D_{N\times n}(\mathbb{R})$ the space of $(N \times n)$-matrices and the subspace of diagonal $(N \times n)$-matrices, respectively. (A matrix $M = (m_{ij}) \in M_{N\times n}(\mathbb{R})$ is said to be diagonal if $m_{ij} = 0$ whenever $i \neq j$.) If $N = n$, we write $M_n(\mathbb{R}) = M_{N\times n}(\mathbb{R})$ and $D_n(\mathbb{R}) = D_{N\times n}(\mathbb{R})$.
We denote by $\langle \cdot, \cdot \rangle$ the standard scalar product in $M_{N\times n}(\mathbb{R})$:

$$\langle M, N \rangle = \sum_{j=1}^{N} \sum_{k=1}^{n} M_{jk} N_{jk} = \operatorname{tr}(MN^t) = \operatorname{tr}(M^t N).$$

For all $\mathbf{x} \in \mathbb{R}^n$, we denote by $\operatorname{diag}_{N\times n}(\mathbf{x})$ the diagonal matrix in $M_{N\times n}(\mathbb{R})$ whose diagonal elements are the components of $\mathbf{x}$. In the square case ($N = n$), we will often write $\operatorname{diag} = \operatorname{diag}_{N\times n}$.
For all $m \in \mathbb{N}^*$, we denote by $\mathrm{GL}(m)$, $\mathrm{O}(m)$ and $\mathrm{SO}(m)$ the group of all invertible $(m \times m)$-matrices, the subgroup of all orthogonal matrices and the subgroup of all orthogonal matrices with determinant 1, respectively. We denote by $\Pi(m)$ the subgroup of $\mathrm{O}(m)$ which consists of the matrices having exactly one nonzero entry per line and per column which belongs to $\{-1, 1\}$, by $\Pi_e(m)$ the subgroup of $\Pi(m)$ which consists of the matrices having an even number of entries equal to -1, and by $\mathrm{S}(m)$ the subgroup of $\Pi_e(m)$ of all permutation matrices. Notice that $\Pi_e(m)$ is the subgroup generated by the permutation matrices and $\operatorname{diag}_{m\times m}(-1, -1, 1, \ldots, 1)$, and that

$$\operatorname{card} \Pi_e(m) = \left(\binom{m}{0} + \binom{m}{2} + \cdots + \binom{m}{2k} \right) \cdot m!,$$

where k is the largest integer such that $2k \leq m$. Notice also that $\mathrm{GL}(m)$, $\mathrm{O}(m)$, $\mathrm{SO}(m)$, $\Pi(m)$, $\Pi_e(m)$ and $\mathrm{S}(m)$ are stable under transposition.

2 Preliminaries

We consider functions of matrices in $M_{N\times n}(\mathbb{R})$ either in the square case ($N = n$) or in the rectangular case ($N \neq n$). In the latter case, we will always assume that $N > n$, the opposite case being entirely analogous.
Throughout, we will write, for all $\xi \in M_{N\times n}(\mathbb{R})$,

$$\boldsymbol{\lambda}(\xi) = (\lambda_1(\xi), \ldots, \lambda_n(\xi)) \quad \text{and} \quad \boldsymbol{\mu}(\xi) = (\mu_1(\xi), \ldots, \mu_n(\xi)).$$

Recall that, for all $\xi \in M_{N\times n}(\mathbb{R})$, we can find $Q \in \mathrm{O}(N)$ and $R \in \mathrm{O}(n)$ such that

$$\xi = Q\Lambda R^t \quad \text{where} \quad \Lambda := \mathrm{diag}_{N\times n}(\lambda_1(\xi), \dots, \lambda_n(\xi))$$

(see [4], Theorem 7.3.5). It is clear that, in the square case ($N = n$), we may choose Q and R in $\mathrm{SO}(n)$ provided that $\lambda_1(\xi)$ is replaced by $\mu_1(\xi)$ in Λ. Given a subgroup G of $\mathrm{GL}(N)$ and a subgroup H of $\mathrm{GL}(n)$, we say that a function $f\colon M_{N\times n}(\mathbb{R}) \to [-\infty, \infty]$ is $G \times H^t$-invariant if

$$\forall \xi \in M_{N\times n}(\mathbb{R}),\ \forall Q \in G,\ \forall R \in H,\ f(Q\xi R^t) = f(\xi).$$

All subgroups G, H encountered in this paper are stable under transposition, so we will equivalently speak of $G \times H$-invariance. For example, a function $f\colon M_{N\times n}(\mathbb{R}) \to [-\infty, \infty]$ is $\mathrm{O}(N) \times \mathrm{O}(n)$-invariant if

$$\forall \xi \in M_{N\times n}(\mathbb{R}),\ \forall Q \in \mathrm{O}(N),\ \forall R \in \mathrm{O}(n),\ f(Q\xi R^t) = f(\xi).$$

Given any subgroup G of $\mathrm{GL}(n)$, we say that a function $g\colon \mathbb{R}^n \to [-\infty, \infty]$ is G-invariant if

$$\forall M \in G, \qquad g(M\mathbf{x}) = g(\mathbf{x}).$$

It is customary to refer to $\mathrm{S}(n)$-invariant functions as *symmetric functions*. The following proposition is an immediate consequence of the Singular Value Decomposition (see [4], Theorem 7.3.5, for example).

Proposition 1 (i) *Let* $f\colon M_n(\mathbb{R}) \to [-\infty, \infty]$. *Then* f *is* $\mathrm{SO}(n) \times \mathrm{SO}(n)$-*invariant if and only if* f *satisfies*

$$f = f \circ \mathrm{diag} \circ \boldsymbol{\mu},$$

and $g := f \circ \mathrm{diag}$ *is then the unique* $\Pi_e(n)$-*invariant function such that* $f = g \circ \boldsymbol{\mu}$.

(ii) *Let* $f\colon M_{N\times n}(\mathbb{R}) \to [-\infty, \infty]$, *where* $N \geq n$. *Then* f *is* $\mathrm{O}(N) \times \mathrm{O}(n)$-*invariant if and only if* f *satisfies*

$$f = f \circ \mathrm{diag}_{N\times n} \circ \boldsymbol{\lambda},$$

and $g := f \circ \mathrm{diag}_{N\times n}$ *is then the unique* $\Pi(n)$-*invariant function such that* $f = g \circ \boldsymbol{\lambda}$.

It is clear that, if $N = n$, the notions of $\mathrm{O}(N) \times \mathrm{O}(n)$, $\mathrm{SO}(N) \times \mathrm{O}(n)$ and $\mathrm{O}(N) \times \mathrm{SO}(n)$-invariance coincide, but differ from that of $\mathrm{SO}(N) \times \mathrm{SO}(n)$-invariance. However, if $N \neq n$, all four notions coincide:

Proposition 2 *Let* $f\colon M_{N\times n}(\mathbb{R}) \to [-\infty, \infty]$, *where* $N > n$. *Then the following are equivalent.*

(i) f *is* $\mathrm{O}(N) \times \mathrm{O}(n)$*-invariant;*

(ii) f *is* $\mathrm{SO}(N) \times \mathrm{SO}(n)$*-invariant.*

PROOF. Obviously, we need only prove that (ii) implies (i). We will see that, if f is $\mathrm{SO}(N) \times \mathrm{SO}(n)$-invariant, then $f = f \circ \mathrm{diag}_{N\times n} \circ \boldsymbol{\lambda}$. The conclusion will then follow from Proposition 1.
Let $\xi \in M_{N\times n}(\mathbb{R})$. By the Singular Value Decomposition, there exists $U \in \mathrm{O}(N)$, $V \in \mathrm{O}(n)$ such that

$$\xi = U\Lambda V^t, \quad \text{where} \quad \Lambda := \mathrm{diag}_{N\times n}(\lambda_1(\xi), \ldots, \lambda_n(\xi))$$

For all $m \geq 1$, let $H_m := \mathrm{diag}(-1, 1 \ldots, 1)$ and $K_m := \mathrm{diag}(1, \ldots, 1, -1)$ in $M_m(\mathbb{R})$.

- If $U \in \mathrm{SO}(N)$ and $V \in \mathrm{SO}(n)$, then
$$f(\xi) = f(\Lambda) = (f \circ \mathrm{diag}_{N\times n} \circ \boldsymbol{\lambda})(\xi). \tag{3}$$
- If $U \in \mathrm{O}(N) \setminus \mathrm{SO}(N)$ and $V \in \mathrm{O}(n) \setminus \mathrm{SO}(n)$, we may write $\Lambda = H_N \Lambda H_n$, so that $U\Lambda V^t = (UH_N)\Lambda(VH_n)^t$, where $UH_N \in \mathrm{SO}(N)$ and $VH_n \in \mathrm{SO}(n)$. Thus Equation (3) holds.
- If $U \in \mathrm{O}(N) \setminus \mathrm{SO}(N)$ and $V \in \mathrm{SO}(n)$, we may write $\Lambda = K_N \Lambda$, so that $U\Lambda V^t = (UK_N)\Lambda V^t$, where $UK_N \in \mathrm{SO}(N)$. Thus Equation (3) holds.
- If $U \in \mathrm{SO}(N)$ and $V \in \mathrm{O}(n) \setminus \mathrm{SO}(n)$, we may write $\Lambda = H_N K_N \Lambda H_n$, so that $U\Lambda V^t = (UH_N K_N)\Lambda(VH_n)^t$, where $UH_N K_N \in \mathrm{SO}(N)$ and $VH_n \in \mathrm{SO}(n)$. Thus Equation (3) holds.

Thus we have shown that $f = f \circ \mathrm{diag}_{N\times n} \circ \boldsymbol{\lambda}$. ∎

3 Von Neumann type inequalities

This section is devoted to Von Neumann type Inequalities (see Theorem 5 below). Our strategy is inspired by Rosakis' paper [10]. It combines a variational argument and the resolution of some discrete optimization problem. The main advantage of our proof is that we get the classical von Neumann inequality as a by product, while Rosakis uses it in his proof. We will need the following technical results.

Lemma 3 (i) *Let $D \in M_n(\mathbb{R})$ be diagonal, with diagonal entries whose absolute values are pairwise distinct. If $M \in M_n(\mathbb{R})$ is such that both MD and DM are symmetric, then M is diagonal.*

(ii) *Let $D \in M_{N\times n}(\mathbb{R})$ be diagonal ($N > n$), with nonzero diagonal entries whose absolute values are pairwise distinct. If $M \in M_{n\times N}(\mathbb{R})$ is such that both MD and DM are symmetric, then M is diagonal.*

Proof.

(i) The result is immediate in the case where $n = 2$. We then proceed by induction. Assuming the property established in $M_n(\mathbb{R})$, let

$$D = \begin{pmatrix} x & \mathbf{0}^t \\ \mathbf{0} & Y \end{pmatrix} \quad \text{and} \quad M = \begin{pmatrix} a & \mathbf{b}^t \\ \mathbf{c} & D \end{pmatrix}.$$

Here, $a, x \in \mathbb{R}$, $\mathbf{b}, \mathbf{c} \in \mathbb{R}^n$ and $D, Y \in M_n(\mathbb{R})$ with $Y = \mathrm{diag}(y_1, \ldots, y_n)$, and the absolute values of the diagonal entries of D are pairwise distinct. Then

$$MD = \begin{pmatrix} ax & \mathbf{b}^t Y \\ \mathbf{c}x & DY \end{pmatrix} \quad \text{and} \quad DM = \begin{pmatrix} xa & x\mathbf{b}^t \\ Y\mathbf{c} & YD \end{pmatrix}.$$

If MD and DM are symmetric, then so are DY and YD. Hence D must be diagonal. It remains to show that $\mathbf{b} = \mathbf{c} = \mathbf{0}$. We have:

$$Y\mathbf{b} = x\mathbf{c} \quad \text{and} \quad Y\mathbf{c} = x\mathbf{b}.$$

If $x = 0$, then Y is nonsingular (since, by assumption, $y_i \neq x = 0$ for all i), so that the only solution is $\mathbf{b} = \mathbf{c} = \mathbf{0}$. If $x \neq 0$, then combining the above equations yields $(Y^2 - x^2 I)\mathbf{b} = \mathbf{0}$ and $(Y^2 - x^2 I)\mathbf{c} = \mathbf{0}$. Since $\det(Y^2 - x^2 I) = \prod_j (y_j^2 - x^2) \neq 0$, we must have $\mathbf{b} = \mathbf{c} = \mathbf{0}$.

(ii) Let us write

$$D = \begin{bmatrix} D_1 \\ D_2 \end{bmatrix} \quad \text{with} \quad D_1 \in M_n(\mathbb{R}) \quad \text{and} \quad D_2 = 0 \in M_{(N-n)\times n}(\mathbb{R}),$$

and

$$M = \begin{bmatrix} M_1 \ M_2 \end{bmatrix} \quad \text{with} \quad M_1 \in M_n(\mathbb{R}) \quad \text{and} \quad M_2 \in M_{n\times(N-n)}(\mathbb{R}).$$

Then

$$MD = M_1 D_1 \in M_n(\mathbb{R}) \quad \text{and} \quad DM = \begin{bmatrix} D_1 M_1 & D_1 M_2 \\ 0 & 0 \end{bmatrix} \in M_N(\mathbb{R}).$$

If MD and DM are symmetric, then so are $M_1 D_1$ and $D_1 M_1$, and Part (i) then shows that M_1 is diagonal. In addition, we must have $D_1 M_2 = 0 \in M_{n\times(N-n)}(\mathbb{R})$, and since the diagonal entries of D_1 are nonzero, this implies that $M_2 = 0$. The proof is complete. ∎

The following proposition may be regarded as a primary version of Inequality (2), for diagonal matrices.

Proposition 4 *Let $b_1, \ldots, b_n \in \mathbb{R}$ satisfy $|b_1| \le b_2 \le \ldots \le b_n$. Let $a_1, \ldots, a_n \in \mathbb{R}$, and let τ be a permutation of $\{1, \ldots, n\}$ such that $|a_{\tau(1)}| \le \ldots \le |a_{\tau(n)}|$.*

(i) *If $\prod_{j=1}^n a_j \ge 0$, then $a_1 b_1 + \cdots + a_n b_n \le |a_{\tau(1)}| b_1 + \ldots + |a_{\tau(n)}| b_n$;*

(ii) *if $\prod_{j=1}^n a_j < 0$, then $a_1 b_1 + \cdots + a_n b_n \le -|a_{\tau(1)}| b_1 + \ldots + |a_{\tau(n)}| b_n$.*

In other words, if $\mathbf{b}$ belongs to the set

$$\Gamma_e := \left\{ \mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n \;\middle|\; |x_1| \le x_2 \le \ldots \le x_n \right\},$$

then

$$\max_{M \in \Pi_e(n)} \langle M\mathbf{a}, \mathbf{b} \rangle = \langle \boldsymbol{\mu}(\operatorname{diag} \mathbf{a}), \mathbf{b} \rangle.$$

Proof. The case $n = 2$ is straightforward. It says that, if $|b_1| \le b_2$ and if $\tau \in \mathrm{S}(2)$ is such that $|a_{\tau(1)}| \le |a_{\tau(2)}|$, then

(i') $a_1 a_2 \ge 0$ implies $a_1 b_1 + a_2 b_2 \le |a_{\tau(1)}| b_1 + |a_{\tau(2)}| b_2$, and (ii') $a_1a_2 < 0$ implies $a_1b_1 + a_2b_2 \leq -|a_{\tau(1)}|b_1 + |a_{\tau(2)}|b_2$.

We will use these rules to prove the result in the general case. The given permutation τ will be decomposed as a well chosen product of transpositions, each of them giving rise to an inequality via (i') or (ii'). For example, assuming that $|a_k| \geq |a_{k+1}|$ for some k, we can write, if $a_ka_{k+1} \geq 0$,

$$\begin{aligned} & a_1b_1 + \cdots + a_kb_k + a_{k+1}b_{k+1} + \cdots + a_nb_n \\ \leq\; & a_1b_1 + \cdots + |a_{k+1}|b_k + |a_k|b_{k+1} + \cdots + a_nb_n \end{aligned} \tag{4}$$

or, if $a_ka_{k+1} < 0$,

$$\begin{aligned} & a_1b_1 + \cdots + a_kb_k + a_{k+1}b_{k+1} + \cdots + a_nb_n \\ \leq\; & a_1b_1 + \cdots - |a_{k+1}|b_k + |a_k|b_{k+1} + \cdots + a_nb_n. \end{aligned} \tag{5}$$

Since the b_k will keep the same place throughout, we will symbolize inequalities such as (4), (5) by

$$(a_1, \ldots, a_k, a_{k+1}, \ldots, a_n) \;\to\; (a_1, \ldots, |a_{k+1}|, |a_k|, \ldots, a_n), \tag{6}$$

$$(a_1, \ldots, a_k, a_{k+1}, \ldots, a_n) \;\to\; (a_1, \ldots, -|a_{k+1}|, |a_k|, \ldots, a_n), \tag{7}$$

respectively.

We first consider the case where $b_1 > 0$. Suppose that $\prod_{j=1}^n a_j \geq 0$. Clearly,

$$(a_1, \ldots, a_n) \to (|a_1|, \ldots, |a_n|).$$

Now, $|a_{\tau(n)}|$ can *migrate* rightward by mean of a transposition of type (6). Thus

$$(|a_1|, \ldots, |a_n|) \to (|a_1|, \ldots, |a_{\tau(n)-1}|, |a_{\tau(n)+1}|, \ldots, |a_{n-1}|, |a_{\tau(n)}|).$$

Repeating this process with $|a_{\tau(n-1)}|$, $|a_{\tau(n-2)}|$ and so on will give rise to the desired inequality. Suppose next that $\prod_{j=1}^n a_j < 0$. In this case, we decide to replace all but one of the negative a_j by their absolute values: for example, if a_k is negative,

$$(a_1, \ldots, a_n) \to (|a_1|, \ldots, |a_{k-1}|, -|a_k|, |a_{k+1}|, \ldots, |a_n|).$$

Now we let $|a_{\tau(n)}|$ migrate rightward, using either a transposition of type (6) or a transposition of type (7) according to the signs of the elements under consideration. Each transposition leaves one negative element. Repeating this process with $|a_{\tau(n-1)}|$, $|a_{\tau(n-2)}|$ and so on will eventually sort the $|a_j|$ according to τ, and give rise to

$$\begin{aligned}&(|a_1|, \ldots, |a_{k-1}|, -|a_k|, |a_{k+1}|, \ldots, |a_n|)\\ &\quad \to \quad (|a_{\tau(1)}|, |a_{\tau(2)}|, \ldots, -|a_{\tau(l)}|, \ldots, |a_{\tau(n-1)}|, |a_{\tau(n)}|).\end{aligned}$$

Finally, it is clear that the minus sign is allowed to migrate leftward, since all elements are now sorted increasingly. Therefore,

$$\begin{aligned}&(|a_{\tau(1)}|, |a_{\tau(2)}|, \ldots, -|a_{\tau(l)}|, \ldots, |a_{\tau(n-1)}|, |a_{\tau(n)}|)\\ &\quad \to \quad (-|a_{\tau(1)}|, |a_{\tau(2)}|, \ldots, |a_{\tau(n)}|)\end{aligned}$$

and we are done.
Finally, the case where $b_1 < 0$ is easily obtained from the above strategy by observing that $a_1b_1 + \cdots + a_nb_n = (-a_1)(-b_1) + a_2b_2 + \cdots + a_nb_n$. ■
We are now ready to prove the main theorem of this section.

Theorem 5 (i) *Let $\xi, \eta \in M_n(\mathbb{R})$. Then*

$$\max_{Q,R \in SO(n)} \{\mathrm{tr}(Q\xi R^t \eta^t)\} = \sum_{j=1}^{n} \mu_j(\xi)\mu_j(\eta).$$

Consequently, $\mathrm{tr}(\xi\eta^t) \leq \sum_{j=1}^{n} \mu_j(\xi)\mu_j(\eta)$.

(ii) *Let $\xi, \eta \in M_{N\times n}(\mathbb{R})$ where $N \geq n$. Then*

$$\max_{\substack{Q\in \mathrm{O}(N)\\ R\in \mathrm{O}(n)}} \{\mathrm{tr}(Q\xi R^t \eta^t)\} = \sum_{j=1}^{n} \lambda_j(\xi)\lambda_j(\eta).$$

Consequently, $\mathrm{tr}(\xi\eta^t) \leq \sum_{j=1}^{n} \lambda_j(\xi)\lambda_j(\eta)$.

PROOF.

(i) As already said, the beginning of our proof follows the one of Rosakis [10]. Observe first that we can assume that η satisfies

$$\eta = \mathrm{diag}\,(\mu_1(\eta), \ldots, \mu_n(\eta)). \tag{8}$$

As a matter of fact, suppose that the result is proved in this case. Let ζ be any element of $M_n(\mathbb{R})$, and let $U, V \in \mathrm{SO}(n)$ be such that $\zeta = UMV^t$,with $M := \mathrm{diag}\,(\mu_1(\zeta), \ldots, \mu_n(\zeta))$. For all $Q, R \in \mathrm{SO}(n)$,

$$\mathrm{tr}(Q\xi R^t\zeta^t) = \mathrm{tr}(Q\xi R^t VMU^t) = \mathrm{tr}((U^tQ)\xi(R^tV)M).$$

Since $U^t\,\mathrm{SO}(n) = \mathrm{SO}(n)\,V = \mathrm{SO}(n)$, we see that

$$\begin{aligned}
\max_{Q,R\in SO(n)}\{\mathrm{tr}(Q\xi R^t\zeta^t)\} &= \max_{Q_1,R_1\in SO(n)}\{\mathrm{tr}(Q_1\xi R_1^t M)\} \\
&= \sum_{j=1}^n \mu_j(\xi)\mu_j(M) \\
&= \sum_{j=1}^n \mu_j(\xi)\mu_j(\zeta),
\end{aligned}$$

where the second equality results from the fact that M satisfies Condition (8).

Notice that we can also assume, in addition to Condition (8), that η satisfies $|\mu_1(\eta)| < \mu_2(\eta) < \ldots < \mu_n(\eta)$, since a continuity argument will then allow to extend the result to the case of wide inequalities.

Since $\mathrm{SO}(n)\times\mathrm{SO}(n)$ is compact and the function $(Q,R)\mapsto \mathrm{tr}(Q\xi R^t\eta^t)$ is continuous, there exist $Q_0, R_0 \in \mathrm{SO}(n)$ such that

$$\mathrm{tr}(Q_0\xi R_0^t\eta^t) = \max_{Q,R\in SO(n)}\{\mathrm{tr}(Q\xi R^t\eta^t)\}. \tag{9}$$

We will prove that Q_0 and R_0 must be such that $Q_0\xi R_0^t$ is diagonal. Let A and B be skew-symmetric matrices, that is, $A^t = -A$ and $B^t = -B$. For all $t \in \mathbb{R}$, let

$$Q(t) := e^{tA}Q_0 \qquad \text{and} \qquad R(t) := e^{tB}R_0.$$

Clearly, $Q(t)$ and $R(t)$ are in $\mathrm{SO}(n)$, and the function

$$\varphi(t) := \mathrm{tr}(Q(t)\xi R(t)^t\eta^t)$$

is differentiable. The optimality condition (9) implies that $t = 0$ maximizes φ. Consequently,

$$0 = \varphi'(0) = \mathrm{tr}(AQ_0\xi R_0^t\eta^t) + \mathrm{tr}(Q_0\xi R_0^t B^t\eta^t).$$

We have therefore shown that, for all skew-symmetric matrices A and B,

$$\mathrm{tr}(AQ_0\xi R_0^t\eta^t) = \langle A, (Q_0\xi R_0^t\eta^t)^t\rangle = 0,$$
$$\mathrm{tr}(\eta^t Q_0\xi R_0^t B^t) = \langle(\eta^t Q_0\xi R_0^t), B\rangle = 0.$$

Recall that $M_n(\mathbb{R})$ is the orthogonal direct sum of $S_n(\mathbb{R})$ and $A_n(\mathbb{R})$, the subspaces of symmetric and skew-symmetric matrices, respectively. Therefore, the above conditions tell us that $Q_0\xi R_0^t\eta^t$ and $\eta^t Q_0\xi R_0^t$ must be symmetric. Lemma 3(i) then implies that $Q_0\xi R_0^t$ is diagonal. We have shown so far that

$$\max_{Q,R\in SO(n)}\{\mathrm{tr}(Q\xi R^t\eta^t)\} = \mathrm{tr}(Q_0\xi R_0^t\eta^t),$$

where $Q_0, R_0 \in \mathrm{SO}(n)$ are such that $Q_0\xi R_0^t$ is diagonal. It remains to see that Q_0 and R_0 are such that

$$Q_0\xi R_0^t = \mathrm{diag}\,(\mu_1(\xi),\dots,\mu_n(\xi)).$$

But this is an immediate consequence of Proposition 4.

(ii) The case where $N = n$, which results immediately from Part (i), corresponds to Von Neumann's inequality itself. Thus, let us assume that $N > n$. The argument is analogous to that of Part (i), so we merely outline the main steps. We can assume that η satisfies

$$\eta = \mathrm{diag}_{N\times n}(\lambda_1(\eta),\dots,\lambda_n(\eta)), \tag{10}$$

with $0 < \lambda_1(\eta) < \dots < \lambda_n(\eta)$, the case of wide inequalities being deduced by a passage to the limit. The compactness of $\mathrm{O}(N)\times\mathrm{O}(n)$ and the continuity of the function $(Q,R)\mapsto \mathrm{tr}(Q\xi R^t\eta^t)$ imply the existence of $Q_0\in\mathrm{O}(N)$ and $R_0\in\mathrm{O}(n)$ such that

$$\mathrm{tr}(Q_0\xi R_0^t\eta^t) = \max_{\substack{Q\in\mathrm{O}(N)\\ R\in\mathrm{O}(n)}}\{\mathrm{tr}(Q\xi R^t\eta^t)\}. \tag{11}$$

The same variational argument as in Part (i), together with Lemma 3(ii), shows that Q_0 and R_0 must be such that $Q_0\xi R_0^t$ is diagonal. Finally, it is clear that, among all diagonal $(N\times n)$-matrices ξ' with prescribed singular values $\lambda_1(\xi),\dots,\lambda_n(\xi)$, the matrix

$$\mathrm{diag}_{N\times n}(\lambda_1(\xi),\dots,\lambda_n(\xi))$$

maximize $\operatorname{tr}(\xi'\eta^t)$. Thus we must have

$$Q_0 \xi R_0^t = \operatorname{diag}_{N\times n}(\lambda_1(\xi), \ldots, \lambda_n(\xi)),$$

and the result follows. ■

Observe that, in the square case,

$$-\operatorname{tr}(\xi\eta^t) = \operatorname{tr}(-\xi\eta^t) \leq \sum_j \lambda_j(-\xi)\lambda_j(\eta) = \sum_j \lambda_j(\xi)\lambda_j(\eta),$$

so that

$$|\operatorname{tr}(\xi\eta^t)| \leq \sum_j \lambda_j(\xi)\lambda_j(\eta)$$

for all $\xi, \eta \in M_n(\mathbb{R})$. It is worth noticing that the analogous inequality for signed singular values holds as well if n is even.

Corollary 6 *Let $\xi, \eta \in M_n(\mathbb{R})$. If n is even, then*

$$|\operatorname{tr}(\xi\eta^t)| \leq \sum_j \mu_j(\xi)\mu_j(\eta). \tag{12}$$

If n is odd, Inequality (12) *is false in general.*

Proof. If n is even, then $\det(-\xi) = \det\xi$ and $\mu_j(-\xi) = \mu_j(\xi)$ for all $j = 1, \ldots, n$. Since $\operatorname{tr}(-\xi\eta^t) = -\operatorname{tr}(\xi\eta^t)$, we conclude that both $\operatorname{tr}(\xi\eta^t)$ and $-\operatorname{tr}(\xi\eta^t)$ are majorized by $\sum_j \mu_j(\xi)\mu_j(\eta)$.
If n is odd, counterexamples are easy to construct. For example, if $n = 3$, let $\xi := \operatorname{diag}(-1, 1, 1)$ and $\eta := \operatorname{diag}(1, -1, -1)$, Then $\operatorname{tr}(\xi\eta^t) = -3$ and $\sum_j \mu_j(\xi)\mu_j(\eta) = 1$. ■

4 Duality

Recall that if G is a subgroup of $\mathrm{GL}(n)$, then the set $G^t := \{M^t | M \in G\}$ is also a subgroup of $\mathrm{GL}(n)$. The following lemma is elementary.

Lemma 7 *Let $g\colon \mathbb{R}^n \to [-\infty, \infty]$ and let G be any subgroup of $\mathrm{GL}(n)$. Consider the following statements:*

(i) *g is G-invariant;*

(ii) *$g^\star$ is G^t-invariant.*

Then (i) *implies* (ii), *and the converse is true if g is closed proper convex.*

PROOF. Suppose that g is G-invariant, and let $M \in G$. Then

$$\begin{aligned} g^\star(M^t\boldsymbol{\xi}) &= \sup\left\{ \langle M^t\boldsymbol{\xi}, \mathbf{x}\rangle - g(\mathbf{x}) \;\middle|\; \mathbf{x} \in \mathbb{R}^n \right\} \\ &= \sup\left\{ \langle \boldsymbol{\xi}, M\mathbf{x}\rangle - g(M\mathbf{x}) \;\middle|\; \mathbf{x} \in \mathbb{R}^n \right\} \\ &= \sup\left\{ \langle \boldsymbol{\xi}, \mathbf{y}\rangle - g(\mathbf{y}) \;\middle|\; \mathbf{y} \in \mathbb{R}^n \right\} \\ &= g^\star(\boldsymbol{\xi}). \end{aligned}$$

Thus $g^\star$ is G^t-invariant. If g is closed proper convex, the converse follows dually, since $g^{\star\star} = g$ in this case. ∎

Lemma 8 *Let $f\colon M_{N\times n}(\mathbb{R}) \to [-\infty, \infty]$, let G be a subgroup of* GL(N), *and let H be a subgroup of* GL(n). *Consider the following statements:*

(i) *f is $G \times H^t$-invariant;*

(ii) *$f^\star$ is $G^t \times H$-invariant.*

Then (i) *implies* (ii), *and the converse is true if f is closed proper convex.*

PROOF. Suppose that f is $G \times H^t$-invariant, and let $U \in G$ and $V \in H$. For all $\xi, X \in M_{N\times n}(\mathbb{R})$, we have

$$\langle U^t\xi V, X\rangle = \operatorname{tr}(U^t\xi V X^t) = \operatorname{tr}(\xi V X^t U^t) = \langle \xi, UXV^t\rangle.$$

Thus

$$\begin{aligned} f^\star(U^t\xi V) &= \sup\left\{ \langle U^t\xi V, X\rangle - f(X) \;\middle|\; X \in M_n(\mathbb{R}) \right\} \\ &= \sup\left\{ \langle \xi, UXV^t\rangle - f(UXV^t) \;\middle|\; X \in M_n(\mathbb{R}) \right\} \\ &= \sup\left\{ \langle \xi, Y\rangle - f(Y) \;\middle|\; Y \in M_n(\mathbb{R}) \right\} \end{aligned}$$

since $X \mapsto UXV^t$ is bijective. Therefore, $f^\star(U^t\xi V) = f^\star(\xi)$, so that $f^\star$ is $G^t \times H$-invariant. If f is closed proper convex, the converse follows dually, since $f^{\star\star} = f$ in this case. ∎

Theorem 9 (i) *Let* $f\colon M_n(\mathbb{R}) \to (-\infty, \infty]$ *be* $\mathrm{SO}(n) \times \mathrm{SO}(n)$*-invariant, and let* $g\colon \mathbb{R}^n \to (-\infty, \infty]$ *be the unique* $\Pi_e(n)$*-invariant function such that* $f = g \circ \boldsymbol{\mu}$*. Then*

$$f^\star = g^\star \circ \boldsymbol{\mu}.$$

(ii) *Let* $N \geq n$*, let* $f\colon M_{N\times n}(\mathbb{R}) \to (-\infty, \infty]$ *be* $\mathrm{O}(N) \times \mathrm{O}(n)$*-invariant, and let* $g\colon \mathbb{R}^n \to (-\infty, \infty]$ *be the unique* $\Pi(n)$*-invariant function such that* $f = g \circ \boldsymbol{\lambda}$*. Then*

$$f^\star = g^\star \circ \boldsymbol{\lambda}.$$

Proof.

(i) We have:

$$\begin{aligned} f^\star(\xi) &= \sup_{X\in M_n(\mathbb{R})} \{\langle \xi, X\rangle - f(X)\} \\ &= \sup_{X\in M_n(\mathbb{R})} \{\langle \xi, X\rangle - g(\boldsymbol{\mu}(X))\} \\ &= \sup_{X\in M_n(\mathbb{R})} \left\{ \sup_{Q,R\in \mathrm{SO}(n)} \{\langle \xi, (QXR^t)\rangle - g(\boldsymbol{\mu}(QXR^t))\} \right\} \end{aligned}$$

But

$$\langle \xi, (QXR^t)\rangle = \mathrm{tr}(\xi^t QXR^t) = \mathrm{tr}(QXR^t\xi^t) \quad \text{and} \quad \boldsymbol{\mu}(QXR^t) = \boldsymbol{\mu}(X)$$

for all $Q, R \in \mathrm{SO}(n)$, so that, by Theorem 5(i), the inner supremum is equal to $\sum_{k=1}^n \mu_k(X)\mu_k(\xi) - g(\mu_1(X), \ldots, \mu_n(X))$. Furthermore, $\boldsymbol{\mu}(X)$ runs over

$$\Gamma_e = \left\{ \mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n \;\middle|\; |x_1| \leq x_2 \leq \ldots \leq x_n \right\}$$

as X runs over $M_n(\mathbb{R})$. Therefore,

$$f^\star(\xi) = \sup_{\mathbf{x}\in\Gamma_e} \{\langle \boldsymbol{\mu}(\xi), \mathbf{x}\rangle - g(\mathbf{x})\}. \tag{13}$$

On the other hand, let $\mathbf{y} \in \Gamma_e$. Then, for all $\mathbf{x}'$ in

$$\Pi_e(n)\mathbf{x} = \{M\mathbf{x} | M \in \Pi_e(n)\},$$

$g(\mathbf{x}') = g(\mathbf{x})$ and $\langle \mathbf{y}, \mathbf{x}'\rangle \leq \langle \mathbf{y}, \mathbf{x}\rangle$ by Proposition 4, so that

$$g^\star(\mathbf{y}) := \sup_{\mathbf{x}\in\mathbb{R}^n} \{\langle \mathbf{y}, \mathbf{x}\rangle - g(\mathbf{x})\} = \sup_{\mathbf{x}\in\Gamma_e} \{\langle \mathbf{y}, \mathbf{x}\rangle - g(\mathbf{x})\}. \tag{14}$$

The result follows from Equations (13) and (14).

(ii) We have:

$$
\begin{aligned}
f^\star(\xi) &= \sup_{X\in M_{N\times n}(\mathbb{R})} \{\langle \xi, X\rangle - f(X)\} \\
&= \sup_{X\in M_{N\times n}(\mathbb{R})} \left\{ \sup_{\substack{Q\in \mathrm{O}(N)\\ R\in\mathrm{O}(n)}} \left\{\langle \xi, (QXR^t)\rangle - f(QXR^t)\right\} \right\} \\
&= \sup_{X\in M_{N\times n}(\mathbb{R})} \left\{ \sup_{\substack{Q\in \mathrm{O}(N)\\ R\in\mathrm{O}(n)}} \left\{\langle \xi, (QXR^t)\rangle\right\} - f(X) \right\}
\end{aligned}
$$

By Theorem 5(ii),

$$
\sup_{\substack{Q\in \mathrm{O}(N)\\ R\in\mathrm{O}(n)}} \left\{\langle \xi, (QXR^t)\rangle\right\} = \sup_{\substack{Q\in \mathrm{O}(N)\\ R\in\mathrm{O}(n)}} \left\{\mathrm{tr}(QXR^t\xi^t)\right\} = \sum_{k=1}^{n} \lambda_k(X)\lambda_k(\xi).
$$

Furthermore, $\boldsymbol{\lambda}(X)$ runs over

$$
\Gamma = \left\{ \mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n \;\middle|\; 0 \le x_1 \le \ldots \le x_n \right\}
$$

as X runs over $M_{N\times n}(\mathbb{R})$. Therefore,

$$
f^\star(\xi) = \sup_{\mathbf{x}\in\Gamma} \{\langle \boldsymbol{\lambda}(\xi), \mathbf{x}\rangle - g(\mathbf{x})\} \tag{15}
$$

On the other hand, let $\mathbf{y} \in \Gamma$. Then, for all $\mathbf{x}'$ in

$$
\Pi(n)\mathbf{x} = \{M\mathbf{x} | M \in \Pi(n)\},
$$

$g(\mathbf{x}') = g(\mathbf{x})$ and $\langle \mathbf{y}, \mathbf{x}'\rangle \le \langle \mathbf{y}, \mathbf{x}\rangle$, so that

$$
g^\star(\mathbf{y}) := \sup_{\mathbf{x}\in\mathbb{R}^n} \{\langle \mathbf{y}, \mathbf{x}\rangle - g(\mathbf{x})\} = \sup_{\mathbf{x}\in\Gamma} \{\langle \mathbf{y}, \mathbf{x}\rangle - g(\mathbf{x})\}. \tag{16}
$$

The result follows from Equations (15) and (16). ■

Remark 10 The set of all transformations $\xi \mapsto U\xi V^t$ with $U, V \in \mathrm{SO}(n)$, endowed with the composition, is obviously a group which is isomorphic to the product group $\mathrm{SO}(n) \times \mathrm{SO}(n)$. By abuse of notation, we may denote this group by $\mathrm{SO}(n) \times \mathrm{SO}(n)$. It results from Theorem 5 that the system $(M_n(\mathbb{R}), \mathrm{SO}(n) \times \mathrm{SO}(n), \mathrm{diag} \circ \boldsymbol{\mu})$ satisfies

(i) diag $\circ\,\boldsymbol{\mu}$ is $\mathrm{SO}(n)\times\mathrm{SO}(n)$-invariant;

(ii) for all $\xi \in M_n(\mathbb{R})$, there exists $(U,V) \in \mathrm{SO}(n)\times\mathrm{SO}(n)$ such that $\xi = U\,\mathrm{diag}\,(\boldsymbol{\mu}(\xi))V^t$;

(iii) for all $\xi,\eta \in M_n(\mathbb{R})$, $\mathrm{tr}(\xi\eta^t) \leq \mathrm{tr}(\mathrm{diag}\,(\boldsymbol{\mu}(\xi))\,\mathrm{diag}\,(\boldsymbol{\mu}(\eta)))$.

According to Lewis' terminology [7], $(M_n(\mathbb{R}), \mathrm{SO}(n)\times\mathrm{SO}(n), \mathrm{diag}\circ\boldsymbol{\mu})$ is a *normal decomposition system*. Our preceding results also show that, similarly, $(M_{N\times n}(\mathbb{R}), \mathrm{O}(N)\times\mathrm{O}(n), \mathrm{diag}_{N\times n}\circ\boldsymbol{\lambda})$ is a normal decomposition system. ∎

We are now ready to prove the main theorem.

Theorem 11 (A) *Let $f\colon M_n(\mathbb{R}) \to (-\infty,\infty]$ be $\mathrm{SO}(n)\times\mathrm{SO}(n)$-invariant, and let $g\colon \mathbb{R}^n \to (-\infty,\infty]$ be the unique $\Pi_e(n)$-invariant function such that $f = g\circ\boldsymbol{\mu}$. Then the following are equivalent:*

(i) *f is closed proper convex;*

(ii) *the restriction of f to $D_n(\mathbb{R})$, the subspace of $M_n(\mathbb{R})$ of diagonal matrices, is closed proper convex;*

(iii) *g is closed proper convex.*

(B) *Let $N > n$, let $f\colon M_{N\times n}(\mathbb{R}) \to (-\infty,\infty]$ be $\mathrm{SO}(N)\times\mathrm{SO}(n)$-invariant or, equivalently, $\mathrm{O}(N)\times\mathrm{O}(n)$-invariant, and let $g\colon \mathbb{R}^n \to (-\infty,\infty]$ to be the unique $\Pi(n)$-invariant function such that $f = g\circ\boldsymbol{\lambda}$. Then the following are equivalent:*

(i) *f is closed proper convex;*

(ii) *the restriction of f to $D_{N\times n}(\mathbb{R})$, the subspace of $M_{N\times n}(\mathbb{R})$ of diagonal matrices, is closed proper convex;*

(iii) *g is closed proper convex.*

Proof.

(A) The fact that (i) implies (ii) is clear. The fact that (ii) implies (iii) results immediately from the equality $g = f\circ\mathrm{diag}$. Finally, suppose that (iii) holds. Then $g^{\star\star} = g$, and Theorem 9(i) implies that

$$f^{\star\star} = g^{\star\star}\circ\boldsymbol{\mu} = g\circ\boldsymbol{\mu} = f,$$

which shows that f is closed proper convex.

(B) The fact that (i) implies (ii) is clear. The fact that (ii) implies (iii) results immediately from the equality $g = f \circ \mathrm{diag}_{N\times n}$. Finally, suppose that (iii) holds. Theorem 9(ii) then implies that

$$f^{\star\star} = g^{\star\star} \circ \boldsymbol{\lambda} = g \circ \boldsymbol{\lambda} = f,$$

which shows that f is closed proper convex. ∎

In the case of $\mathrm{O}(n) \times \mathrm{O}(n)$-invariant functions, the analogous statement can be derived in several ways from the above results.

Corollary 12 *Let $f\colon M_n(\mathbb{R}) \to (-\infty, \infty]$ be $\mathrm{O}(n) \times \mathrm{O}(n)$-invariant, and let $g\colon \mathbb{R}^n \to (-\infty, \infty]$ be the unique $\Pi(n)$-invariant function such that $f = g \circ \boldsymbol{\lambda}$. Then the following are equivalent:*

(i) *f is closed proper convex;*

(ii) *the restriction of f to $D_n(\mathbb{R})$ is closed proper convex;*

(iii) *g is closed proper convex.*

Remark 13 As a convex $\Pi(n)$-invariant function, the function g appearing in Theorem 11(B) or in Corollary 12 must be such that each partial mapping

$$x_k \mapsto g(x_1, \ldots, x_n), \quad k = 1, \ldots, n$$

is increasing on $\mathbb{R}_+$. As a matter of fact, for all $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ with $x_1 \geq 0$,

$$g(0, x_2, \ldots, x_n) \leq \frac{1}{2} g(-x_1, x_2, \ldots, x_n) + \frac{1}{2} g(x_1, x_2, \ldots, x_n) = g(\mathbf{x}),$$

and if $z > 0$, we see, using the above inequality, that

$$\begin{aligned} g(\mathbf{x}) &\leq \frac{x_1}{x_1 + z} g(x_1 + z, x_2, \ldots, x_n) + \frac{z}{x_1 + z} g(0, x_2, \ldots, x_n) \\ &\leq \frac{x_1}{x_1 + z} g(x_1 + z, x_2, \ldots, x_n) + \frac{z}{x_1 + z} g(x_1 + z, x_2, \ldots, x_n) \\ &= g(x_1 + z, x_2, \ldots, x_n). \end{aligned}$$

Thus $x_1 \mapsto g(x_1, \ldots, x_n)$ is increasing on $\mathbb{R}_+$, and the same reasoning holds for all other partial applications. ∎

5 Concluding comments

The assumption of $\mathrm{SO}(N) \times \mathrm{SO}(n)$-invariance enables to reduce substantially the dimension of the objects whose convexity is studied. This appears clearly in Theorem 11, where the dimension is reduced from Nn to n.
It is worth noticing that the computation of the convex envelope of some $\mathrm{SO}(N) \times \mathrm{SO}(n)$-invariant function f also benefits from this dimension reduction, as one should expect.

Theorem 14 (i) *Let $f = g \circ \boldsymbol{\mu}\colon M_n(\mathbb{R}) \to (-\infty, \infty]$ be $\mathrm{SO}(n) \times \mathrm{SO}(n)$-invariant. Then, on denoting by Cf and Cg the convex envelopes of f and g, respectively, one has*

$$Cf = Cg \circ \boldsymbol{\mu}.$$

(ii) *Let $N \geq n$, and let $f = g \circ \boldsymbol{\lambda}\colon M_{N\times n}(\mathbb{R}) \to (-\infty, \infty]$ be $\mathrm{O}(N) \times \mathrm{O}(n)$-invariant. Then, on denoting by Cf and Cg the convex envelopes of f and g, respectively, one has*

$$Cf = Cg \circ \boldsymbol{\lambda}.$$

Proof. Since both statements are analogous, we prove the first one only. We use the theorem of Carathéodory, which implies that

$$Cf(\xi) = \inf\left\{ \sum_{k=1}^{n^2+1} \alpha_k f(\xi_k) \;\middle|\; (\alpha_1, \ldots, \alpha_{n^2+1}) \in \Delta_{n^2+1},\ \sum_{k=1}^{n^2+1} \alpha_k \xi_k = \xi \right\}$$

and

$$Cg(\mathbf{x}) = \inf\left\{ \sum_{k=1}^{n+1} \alpha_k g(\mathbf{x}_k) \;\middle|\; (\alpha_1, \ldots, \alpha_{n+1}) \in \Delta_{n+1},\ \sum_{k=1}^{n+1} \alpha_k \mathbf{x}_k = \mathbf{x} \right\},$$

in which Δ_m denotes the simplex in $\mathbb{R}^m$ (see [1], Theorem 2.8 and Corollary 2.9, for example).
Let $\tilde{f} := Cg \circ \boldsymbol{\mu}$. For all $\xi \in M_n(\mathbb{R})$, we have

$$\tilde{f}(\xi) = (Cg \circ \boldsymbol{\mu})(\xi) \leq (g \circ \boldsymbol{\mu})(\xi) = f(\xi).$$

Since $\tilde{f} \leq f$ and $\tilde{f}$ is convex by Theorem 11(A), we deduce that $\tilde{f} \leq Cf$. Conversely,

$$\begin{aligned}
\tilde{f}(\xi) &= Cg(\boldsymbol{\mu}(\xi)) \\
&= \inf\left\{ \sum_{k=1}^{n+1} \alpha_k f(\operatorname{diag}(\mathbf{x}_k)) \;\middle|\; (\alpha_1, \ldots, \alpha_{n+1}) \in \Delta_{n+1},\ \sum_{k=1}^{n+1} \alpha_k \mathbf{x}_k = \boldsymbol{\mu}(\xi) \right\} \\
&\geq \inf\left\{ \sum_{k=1}^{n+1} \alpha_k Cf(\operatorname{diag}(\mathbf{x}_k)) \;\middle|\; (\alpha_1, \ldots, \alpha_{n+1}) \in \Delta_{n+1},\ \sum_{k=1}^{n+1} \alpha_k \mathbf{x}_k = \boldsymbol{\mu}(\xi) \right\} \\
&\geq \inf\left\{ \sum_{k=1}^{n+1} \alpha_k Cf(\xi_k) \;\middle|\; (\alpha_1, \ldots, \alpha_{n+1}) \in \Delta_{n+1},\ \sum_{k=1}^{n+1} \alpha_k \xi_k = \operatorname{diag}(\boldsymbol{\mu}(\xi)) \right\} \\
&\geq CCf(\operatorname{diag} \boldsymbol{\mu}(\xi)) \;=\; Cf(\xi),
\end{aligned}$$

in which the last equality results from the obvious fact that $CCf = Cf$ and from the $\mathrm{SO}(n) \times \mathrm{SO}(n)$-invariance of Cf, whose proof is easy and left to the reader. ∎

Remark 15 In the case where $f^{\star\star} = Cf$ and $g^{\star\star} = Cg$, which happens notably when f and g are finite, the above result is an immediate consequence of Theorem 9. ∎

Another noteworthy dimension reduction occurs in the computation of the inf-convolution of two convex invariant functions. If f_1 and f_2 are two extended real-valued functions on $M_{N\times n}(\mathbb{R})$, their inf-convolution is defined by

$$(f_1 \Box f_2)(\xi) = \inf_{\eta \in M_{N\times n}(\mathbb{R})} \big\{ f_1(\xi - \eta) + f_2(\eta) \big\}.$$

Recall that, in essence, inf-convolution and addtion are dual operations. More precisely, if f_1 and f_2 are proper, then

$$(f_1 \Box f_2)^\star = f_1^\star + f_2^\star,$$

and consequently the formula

$$f_1 \Box f_2 = (f_1^\star + f_2^\star)^\star$$

holds whenever $f_1 \Box f_2 = (f_1 \Box f_2)^{\star\star}$, that is, whenever $f_1 \Box f_2$ is closed proper convex. This duality, combined with Theorem 9, gives rise to the following result.

Theorem 16 (i) *For* $i = 1, 2$, *let* $f_i = g_i \circ \boldsymbol{\mu}\colon M_n(\mathbb{R}) \to (-\infty, \infty]$ *be closed proper convex and* $\mathrm{SO}(n) \times \mathrm{SO}(n)$*-invariant. If* f_1 *or* f_2 *is inf-compact, then*

$$f_1 \Box f_2 = (g_1 \Box g_2) \circ \boldsymbol{\mu}. \tag{17}$$

(ii) *Let* $N \geq n$. *For* $i = 1, 2$, *let* $f_i = g_i \circ \boldsymbol{\lambda}\colon M_{N\times n}(\mathbb{R}) \to (-\infty, \infty]$ *be closed proper convex and* $\mathrm{O}(N) \times \mathrm{O}(n)$*-invariant. If* f_1 *or* f_2 *is inf-compact, then*

$$f_1 \Box f_2 = (g_1 \Box g_2) \circ \boldsymbol{\lambda}.$$

Proof. Again, we restrict attention to the first statement. Recall that, by definition, f_i is inf-compact if

$$f_i(\xi) \to \infty \quad \text{as} \quad \|\xi\| \to \infty.$$

The relationships $f_i = g_i \circ \boldsymbol{\mu}$ and $g_i = f_i \circ \mathrm{diag}$ imply that f_i is inf-compact if and only if g_i is inf-compact. Note that the $\Pi_e(n)$-invariance of g_i and $g_i^\star$ implies that $\mathrm{dom}\, g_i$, $\mathrm{dom}\, g_i^\star$, $\mathrm{dom}\, f_i$ and $\mathrm{dom}\, f_i^\star$ contain the origin. We may assume that $g_i \not\equiv 0$, $i = 1, 2$, for otherwise Equation (17) holds trivially. The $\Pi_e(n)$-invariance of $g_i^\star$ then implies that $\mathrm{int}\,\mathrm{dom}\, g_i^\star$ and $\mathrm{int}\,\mathrm{dom}\, f_i^\star$ contain the origin, and that $g_i^\star$ and $f_i^\star$ are continuous at the origin. By [6], Theorem 6.5.7, $g_1 \Box g_2$ and $f_1 \Box f_2$ are closed proper convex. Theorem 9 then implies that

$$\begin{aligned}
f_1 \Box f_2 &= (f_1^\star + f_2^\star)^\star \\
&= (g_1^\star \circ \boldsymbol{\mu} + g_2^\star \circ \boldsymbol{\mu})^\star \\
&= ((g_1^\star + g_2^\star) \circ \boldsymbol{\mu})^\star \\
&= (g_1^\star + g_2^\star)^\star \circ \boldsymbol{\mu} \\
&= (g_1 \Box g_2) \circ \boldsymbol{\mu}. \ \blacksquare
\end{aligned}$$

Acknowledgements: This work was performed while the second author was visiting professor at EPFL and we thank P. Metzener for interesting discussions. Moreover, after presenting Theorem 5(i) with our proof to Denis Serre, he found a new proof of the result. He will include it in the electronic version of his book [12].

References

[1] B. Dacorogna, *Direct Methods in the Calculus of Variations*, Springer-Verlag, 1989.

[2] B. Dacorogna & P. Marcellini, *Implicit Partial Differential Equations*, Birkhäuser, 1999.

[3] B. Dacorogna & H. Koshigoe, *On the different notions of convexity for rotationally invariant functions*, Annales de la Faculté des Sciences de Toulouse, II(2), pp. 163-184, 1993.

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

[5] B. Kostant, *On convexity, the Weyl group and the Iwasawa decomposition*, Annales Scientifiques de l'Ecole Normale Suprieure, 6, pp.413-455, 1973.

[6] P. J. Laurent, *Approximation et Optimisation*, Hermann, 1972.

[7] A. Lewis, *Group invariance and convex matrix analysis*, SIAM Journal of Matrix Analysis and Applications, 17, pp.927-949, 1996.

[8] A. Lewis, *Convex analysis on Cartan subspaces*, Nonlinear Analysis, 42, pp. 813-820, 2000.

[9] A. Lewis, *The mathematics of eigenvalue optimization*, Mathematical Programming, Series B 97, pp. 155-176, 2003.

[10] P. Rosakis, *Characterization of convex isotropic functions*, Journal of Elasticity, 49, pp. 257-267, 1997.

[11] A. Seeger, *Convex analysis of spectrally defined matrix functions*, SIAM Journal on Optimization, 7(3), pp. 679-696 (1997).

[12] D. Serre, *Matrices: Theory and Applications*, Grad. Text in Math. 216, Springer-Verlag, 2002. See also http://www.umpa.ens-lyon.fr/∼serre/publi.html.

[13] F. Vincent, *Une note sur les fonctions convexes invariantes*, Annales de la Faculté des Sciences de Toulouse, pp. 357-363, 1997.