On the Polars of Ordinally Weighted Unitarily Invariant Norms

Govind S. Mudholkar and Yogendra P. Chaubey

Indian Statistical Institute, Delhi Centre
7, SJSS Marg, New Delhi–110 016, India
ON THE POLARS OF ORDINALLY WEIGHTED UNITARILY INVARIANT NORMS

GOVIND S. MUDHOLKAR AND YOGENDRA P. CHAUBEY∗

ABSTRACT. Ordinarily weighted unitarily invariant norms of matrices, or operators, are basically symmetric norms of their singular values weighted by an ordered set of weights. A trimmed version of these norms is obtained by considering only a few of the largest singular values. In this short note polars of these norms are characterized as in Mudholkar and Freimer (1985, Proc. Amer. Math. Soc., 95 331-337). A simple expression for the weighted version of the Ky Fan k−norm is also obtained.

1. INTRODUCTION

Let $H$ be a Hilbert space of finite dimension $n$ and $U(H)$ be a group of unitary operators in the space $\mathcal{B}(H)$ of bounded linear operators on $H$. A norm $\|\cdot\|$ defined on $\mathcal{B}(H)$ is said to be unitarily invariant (UI) if for every $A \in \mathcal{B}(H)$ and unitary operators $U, V \in U(H)$, $\|A\| = \|UAV\|$. If $s_1 \geq s_2 \geq ... s_n \geq 0$ denote the singular values of a given $A \in \mathcal{B}(H)$, then the UI norms are the symmetric gauge functions (SGF) of $(s_1, ..., s_n)$; see von Neumann (1937). In this note we consider a class of UI norms, termed weighted norms (see Bhatia (1997, Chap. IV)) in terms of SGF’s which include the well known norms such as Ky-Fan (1951) $k$-norms and Schatten (1950) $p$-norms as well as trimmed $k$-norms studied in Mudholkar and Friemer (1985).

After introducing some preliminaries introduced in §2, these weighted norms are defined in §3 and §4 addresses computation of their polars. The polar of trimmed norm is given by Mudholkar and Friemer (1985) but no general result is available for the weighted case.

2. PRELIMINARIES

In the seminal paper “Some matrix inequalities and metrization of matric spaces”, von Neumann (1937) used SGF’s for defining norms on spaces of finite dimensional matrices.
Definition 2.1 (von Neumann (1937)). For \( x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \), \( \phi(x) \) is said to be an SGF, if it satisfies,

(i) \( \phi(x) \geq 0 \), with equality iff \( x = 0 \);
(ii) \( \phi(cx) = |c|\phi(x) \) for any scalar \( c \);
(iii) \( \phi(x_1 + x_2) \leq \phi(x_1) + \phi(x_2) \);
(iv) for any combinations of sign changes, \( \epsilon_i = \pm 1 \) and for any permutation \( j_1, j_2, \ldots, j_n \) of \( (1, 2, \ldots, n) \), \( \phi(x_1, \ldots, x_n) = \phi(\epsilon_1 x_{j_1}, \ldots, \epsilon_n x_{j_n}) \).

In other words, \( \phi(x) \) is a norm which is symmetric, i.e. invariant under permutations and arbitrary sign changes of the coordinates.

Theorem 2.1 (von Neumann (1937)). Let \( M_n \) be space of \( n \times n \) matrices. For any symmetric gauge function \( \phi \) of \( n \) real variables, the function \( \| . \| \), defined on \( M_n \) by

\[
\| A \| = \phi(s_1, \ldots, s_n), \quad A \in M_n,
\]

is an unitarily invariant norm, where \( s_1, \ldots, s_n \) are the singular values of \( A \). Conversely, every unitarily invariant norm on \( B(H) \) is obtained in this way; let \( \phi(s_1, s_2, \ldots, s_n) = \| A \| \) where \( A \) is a diagonal matrix with diagonal entries \( s_1, s_2, \ldots, s_n \).

von Neumann (1937) illustrated his results using vector \( p \)-norms,

\[
\| x \|_p = \left( \sum_{i=1}^{n} |x_i|^p \right)^{1/p}.
\]

Schatten (1950, 1960) extended von Neumann’s results to spaces of completely continuous operators on Hilbert spaces of arbitrary dimensions, and also used \( p \)-norms as illustrations. These operator norms are known as Schatten \( p \)-norms. Furthermore, he also clarified the necessity and sufficiency role of the SGF’s by showing that all unitarily invariant operator norms are only in terms of SGF’s of singular values as in von Neumann. Ky-Fan (1951), on the other hand, used his discussion of matrix and operator norms and illustrated, what are now known as Ky-Fan norms. These will be discussed later while discussing the trimmed norms.

Polars

Let \( \Phi_n \) denote the class all SGF’s on \( \mathbb{R}^n \) as defined above. Now consider \( A \in B(H) \) and let \( s_1 \geq \ldots \geq s_n \geq 0 \) be its singular vales. As indicated earlier, for any \( \phi \in \Phi_n \), \( \| A \|_\phi \) defined as

\[
\| A \|_\phi = \phi(s_1, \ldots, s_n),
\]

is a unitarily invariant norm of the operator \( A \). Its polar is defined by

\[
\| A \|_{\phi^0} = \sup_{B \neq 0} \frac{\text{tr}AB}{\| B \|_\phi};
\]

(see Schatten (1950)).
Polars of vector norms may be described in terms of the classical Hölder’s inequality. Thus for vectors \( x = (x_1, \ldots, x_n) \) and \( y = (y_1, \ldots, y_n) \) and a vector norm \( \phi \), there exists a vector norm \( \phi^0 \), known as the polar of \( \phi \), such that the inner product of \( x \) and \( y \) satisfies a sharp inequality,

\[
\sum x_i y_i \leq \phi(x) \phi^0(y).
\]

In other words, the polar \( \phi^0 \) satisfies

\[
\phi^0(y) = \max_{x \neq 0} \left( \frac{x'y}{\phi(x)} \right).
\]

Alternatively, \( \phi^0(y) \) is the maximum of \( x'y \) over the set \( \{ x | \phi(x) \leq 1 \} \). It is easy to see that \( \phi \) is the polar of \( \phi^0 \). In the literature, the polar of \( \phi \) is also referred to as conjugate, dual or associate of norm \( \phi \).

In conventional real analysis the above inequality with \( p \)-norm \( \| x \|_p \) as \( \phi \) and \( q \)-norm as \( \phi^0, 1/p + 1/q = 1 \), is described as the Hölder’s inequality. In the related discussion, it is also observed \( \lim_{p \to \infty} \| x \|_p = \max_i \{|x_i|\} = x_{(1)} \), and \( \sum_{i=1}^n x_{(i)} \) and \( x_{(1)} \) are mutual polars, where \( x_{(1)} \geq \cdots \geq x_{(n)} \geq 0 \), denote the ordered values of \( |x_i|, i = 1, 2, \ldots, n \).

3. Trimmed and Ordinally Weighted Norms

In this section we discuss some basics of the trimmed and weighted unitarily invariant norms. Let \( x_{(1)} \geq \cdots \geq x_{(n)} \geq 0 \), denote the ordered values of \( |x_i|, i = 1, 2, \ldots, n \). Then Ky-Fan (1951) \( k \)-norm, \( k \leq n \), of \( x \) is given by

\[
\phi_K(x) = \sum_{i=1}^k x_{(i)}.
\]

Generalizing the above to the \( p \)-norm, we get the \( p \)-norm extension of Ky-Fan \( k \)-norm

\[
\phi_{Kp}(x) = \left( \sum_{i=1}^k x_{(i)}^p \right)^{1/p}
\]

that gives the Schatten \( p \)-norm for \( k = n \); see Horn and Johnson (1990), Problem 5, pp. 211.

We will refer to the above norm that is based on trimming the smallest \( n - k \) values of the coordinates of \( x \) as trimmed \( p \)-norm. Mudholkar and Friemer (1985) considered following as a generalization:

\[
\phi_T(x) = \phi(x_{(1)}, \ldots, x_{(k)});
\]

for \( k \leq n \), where \( \phi \in \Phi_k \) is a SGF of \( k \) coordinates. We will call \( \phi_T \) as a trimmed SGF, in the sense that it trims away the smallest \( n - k \) coordinates of \( x_{(1)}, \ldots, x_{(n)} \).
This indeed gives Ky-Fan $k$-norm and Schatten $p$-norm as special cases. It can be further generalized to ordinally weighted norms obtained by considering the SGF

$$\phi_{T_w}(x) = \phi(w_1 x_{(1)}, ..., w_k x_{(k)}),$$

for $k \leq n$ where $1 = w_1 \geq w_2 \geq ... \geq w_n \geq 0$.

Without loss of generality, we take $k = n$ by defining $w_{k+1} = ... = w_n$ in case $k$ is strictly less than $n$, and denote the corresponding SGF by $\phi_w(x)$. It is proved below that $\phi_w(x)$ is indeed a norm for $x \in \mathbb{R}^n$, and hence $\phi_w(s)$ defines an unitarily invariant norm for $A \in B(H)$. Such a SGF will be said to generate an ordinally weighted unitarily invariant norm. The value of $w_1$ could be taken to be any positive value, however, without loss of generality it will be taken to be 1.

**Lemma 3.1** (see Exercise IV.1.19, Bhatia (1997)). For $x \in \mathbb{R}^n$, $\phi_w(x)$ defined in (3.4) satisfies the following properties:

(i) $\phi_w(x) \geq 0$, with equality iff $x = 0$;
(ii) $\phi_w(cx) = |c|\phi(x)$ for any scalar $c$;
(iii) $\phi_w(x_1 + x_2) \leq \phi(x_1) + \Phi(x_2)$;
(iv) for any combinations of sign changes, $\epsilon_i = \pm 1$ and for any permutation $(j_1, j_2, ..., j_n)$ of $(1, 2, ..., n)$, $\phi(x_1, ..., x_n) = \phi(\epsilon_1 x_{j_1}, ..., \epsilon_n x_{j_n})$.

**Proof:** Below we verify the properties of the SGF:

(i):

$$\phi_w(x) = 0 \quad \implies \quad w_i x_{(i)} = 0, \forall i$$
$$\implies \quad w_1 x_{(1)} = 0$$
$$\implies \quad x_{(1)} = 0 \text{ since } w_1 > 0.$$ 
$$\implies \quad x_{(i)} = 0 \forall i.$$

(ii):

$$\phi_w(cx) = \phi(cw_1 x_{(1)}, ..., cw_n x_{(n)})$$
$$= |c|\phi(w_1 x_{(1)}, ..., w_n x_{(n)})$$
$$= |c|\phi_w(x_{(1)}, ..., x_{(n)}).$$

(iii): To prove the triangle inequality of $\phi_w$ let us recall the concept of weak majorization and Mirsky’s (1960) theorem (see Bhatia (1997), pp. 45).

**Definition 3.1.** We say that a vector $x \in \mathbb{R}^n$ is weakly majorized by $y \in \mathbb{R}^n$, and write $x \prec_w y$ if

$$\sum_{j=1}^{k} x_{(j)} \leq \sum_{j=1}^{k} y_{(j)}, \quad k = 1, 2, ..., n,$$

where $x_{(i)}, i = 1, 2, ..., n$ denote $n$ ordered values of $x_1, ..., x_n$. 
Theorem 3.1. Let $x, y \in \mathbb{R}_n^+$. Then

$$x \prec_w y$$

is necessary and sufficient condition for the relation

$$\phi(x) \leq \phi(y)$$

to hold for every symmetric gauge function $\phi$.

It follows from the Corollary II.4.3, Bhatia (1997) (see Eq. (II.36)) that for $x, y \in \mathbb{R}^n$

$$(x_1 y_1, \ldots, x_n y_n) \prec_w (x_1 y_1, \ldots, x_n y_n)$$

and hence for any permutation $\sigma = (j_1, j_2, \ldots, j_n)$ of $(1, 2, \ldots, n)$

$$(w_1 x_{j_1}, \ldots, w_n x_{j_n}) \prec_w (w_1 x_1, \ldots, w_n x_n).$$

Therefore using the properties of a symmetric gauge function, it follows from (3.6) and Theorem (3.1) that for

$$\phi_w(x + y) = \phi(w.(x + y)_\sigma)$$

where

$$w.x = (w_1 x_1, \ldots, w_n x_n),$$

we have

$$\phi_w(x + y) \leq \phi(w.x_\sigma) + \phi(w.y_\sigma) \leq \phi(w_1 x_1, \ldots, w_n x_n) + \phi(w_1 y_1, \ldots, w_n y_n) = \phi_w(x) + \phi_w(y).$$

(iv):
Since $|\epsilon.x| = |x|$ and for any permutation matrix $J$, $|Jx|_{(i)} = |x|_{(i)}$, it follows that

$$\phi_w(J(\epsilon.x)) = \phi_w(x).$$

This completes the verification that $\phi_w(x)$ defines a SGF.

As a special case of the SGF (3.4), the norm given by

$$\phi^{k,p}_w(x) = \left( \sum_{i=1}^k w_i x^p_{(i)} \right)^{1/p}$$

with $w_1 = \ldots = w_k = 1, w_{k+1} = \ldots = w_n = 0$ gives Ky-Fan $k, p$ norm.

The choice $w_{k+1} = \ldots = w_n = 0$ gives the ordinally weighted version of Mudholkar-Friemer trimmed norm given in Eq. (3.3).

The next section is devoted to a discussion of the polars of ordinally weighted SGF’s given by (3.4).
4. Polars of Ordinally Weighted Symmetric Gauge Functions

Mudholkar and Friemer (1985) prove the following important theorem that is useful in computing polar of a trimmed SGF (3.3).

**Theorem 4.1** (Mudholkar and Friemer (1985), Theorem 3.3). Let \( \phi_T \in \Phi_n \) be the SGF derived from \( \phi \in \Phi_k, k \leq n \) according to (3.3). Then the polar \( \phi_0^T \in \Phi_n \) of \( \phi_T \) is given by

\[
\phi_0^T(y) = \phi^0(y_{(1)}, \ldots, y_{(m)}, \bar{y}, \ldots, \bar{y})
\]

where \( \phi^0 \in \Phi_k \) is the polar of \( \phi \), \( y_{(1)} \geq y_{(2)} \geq \ldots, y_{(n)} \geq 0 \) are the ordered values of the magnitudes \( |y_i| \) of the coordinates of \( y \in \mathbb{R}^n \), \( m \) is obtained such that

\[
y_{(m+1)} \leq \sum_{j=m+1}^{n} y_{(j)}/(k-m) < y_{(m)}
\]

and \( \bar{y} = \sum_{j=m+1}^{n} y_{(j)}/(k-m) \).

The uniqueness of the integer \( m \) is guaranteed according to Lemma 3.1 of Mudholkar and Friemer (1985). It readily provides the polar of the ordinally weighted trimmed norm (3.4).

**Corollary 4.1.** Let \( \phi_{Tw} \in \Phi_n \) be the SGF derived from \( \phi \in \Phi_k, k \leq n \) according to (3.4). Then the polar \( \phi_0^Tw \in \Phi_n \) of \( \phi_{Tw} \) is given by

\[
\phi_0^{Tw}(y) = \phi^0(w_1y_{(1)}, \ldots, w_my_{(m)}, \bar{y}_w, \ldots, \bar{y}_w)
\]

where \( \phi^0 \in \Phi_k \) is the polar of \( \phi \), \( y_{(1)} \geq y_{(2)} \geq \ldots, y_{(n)} \geq 0 \) are the ordered values of the magnitudes \( |y_i| \) of the coordinates of \( y \in \mathbb{R}^n \), \( m \) is obtained such that

\[
w_{m+1}y_{(m+1)} \leq \sum_{j=m+1}^{n} w_jy_{(j)}/(k-m) < w_my_{(m)}
\]

and \( \bar{y}_w = \sum_{j=m+1}^{n} w_jy_{(j)}/(k-m) \).

4.1. Polar of weighted Ky-Fan \( k,p \) Norm. Ordinally weighted \( k - p \) norms (for \( p \geq 1 \)) can be defined as a generalization of (3.2). For a vector \( x \in \mathbb{R}^n \), define \( \|x\|_{w,kp} \) as

\[
\|x\|_{w,kp} = \left( \sum_{i=1}^{k} w_i x_{(i)}^p \right)^{1/p},
\]

for \( k \leq n \), and \( 1 = w_1 \geq w_2 \geq \ldots \geq w_k > 0 \).

The computation of the polars of these norms is simplified by recognizing that \( \|x\|_{w,kp} = \phi_{Kp}(y) \) as the usual Ky-Fan \( k,p \) (see (3.2)) norm in variable \( y \) where
\[ y_i = w_i^{1/p} x_{(i)}, \ i = 1, 2, \ldots, n. \] Then we use the results from Mudholkar and Freimer (1985) for the dual \( \phi^0(y) \) of \( \phi(y) = (\sum_{i=1}^{k} y_{(i)}^p)^{1/p} \), as given by

\[ \phi^0(y) = \left( \sum_{i=1}^{r} y_{(i)}^q + (k - r)\bar{y}^q \right)^{1/q}, \]

where \( r \) is an integer such that

\[ y_{(r+1)} \leq \sum_{j=r+1}^{n} y_{(j)} / (k - r) < y_{(r)} \]

and

\[ \bar{y} = \sum_{j=r+1}^{n} y_{(j)} / (k - r). \]

Note that from the above, we find the dual of Ky-Fan \( k \)-norm (for \( p = 1 \)) as

\[ (4.1) \quad \phi^0_K(y) = \max(y_{(1)}, \frac{1}{k} \sum_{j=1}^{n} y_{(j)}), \]

and that for the Ky Fan \( k, p \) norm (for \( p \geq 2 \)) is given by

\[ (4.2) \quad \phi^0_{Kp}(y) = \left( \sum_{i=1}^{r} y_{(i)}^q + (k - r)\bar{y}^q \right)^{1/q}. \]

An explicit form of the expression for the polar of the ordinally weighted version of the Ky-Fan \( n \)-norm is given by

\[ \|x\|_{0,w,n}^0 = \max(x_{(1)}^+ / w_{(i)}^+), \]

where \( x_{(i)}^+ = x_{(1)} + \ldots + x_{(i)} \). This may be obtained using the following lemma.

**Lemma 4.1.** For \( x, y \in \mathbb{R}^n \), we have the following sharp inequality:

\[ x'y \leq \|x\|_{w,n1} \max(y_{(i)}^+ / w_{(i)}^+) \]

where \( y_{(i)}^+ = y_{(1)} + \ldots + y_{(i)} \).

**Proof:** First note the following alternative way of representing the sum \( \sum_{i=1}^{n} a_i b_i \) as

\[ \sum_{i=1}^{n} a_i b_i = \sum_{i=1}^{n} a_i^+ b_i^- , \]

where

\[ a_i^+ = a_1 + \ldots + a_i; \ i = 1, \ldots, n \]
\[ b_i^- = b_i - b_{i+1}; \ b_{n+1} = 0. \]
This implies
\[ \|\mathbf{x}\|_{w,k} = \sum_{i=1}^{k} w_i^+ x_{(i)}^- , \]
where it is explicitly assumed that \( w_{k+1} = ... = w_n = 0 \). Then by the standard rearrangement inequality
\[ \sum_{i=1}^{n} x_i y_i \leq \sum_{i=1}^{n} x_{(i)} y_{(i)} \]
and since necessarily \( w_1 > 0, w_i^+ > 0, \forall i = 1, 2, ..., n \) we have
\[ \sum_{i=1}^{n} x_i y_i \leq \sum_{i=1}^{n} (w_i^+ x_{(i)}^-)(y_{(i)}/w_i^+) \]
\[ \leq \|\mathbf{x}\|_{w,n} \max(y_{(i)}/w_i^+) \].

The sharpness of the inequality follows by choosing
\[ (4.3) \quad y_i = w_i \max(x_i^+) \]
as both sides of the inequality equal to \( \sum_{i=1}^{n} x_i w_i \max(x_i^+) \). \( \square \)

4.2. Particular Cases.

(1). For \( p = 1 \), and \( w_1 = ... = w_k = 1 \) and \( w_j = 0, j > k \), we get the Ky Fan \( k \)-norm.

In this case \( w_i^+ = i, i = 1, ..., k \) and \( w_i^+ = k \) for \( i > k \). Thus,
\[ \max(x_{(i)}^- / w_i^+) = \max(\frac{1}{k} \sum_{j=1}^{i} x_{(j)}, i = 1, 2, ..., k; \frac{1}{k} \sum_{j=1}^{n} x_{(j)}) . \]

Since
\[ \frac{x_i^+}{w_i^+} - \frac{x_{i+1}}{w_{i+1}} = \frac{1}{i(i+1)} \{(x_{(1)} + ... + x_{(i)} - ix_{(i+1)} \}
\[ = \frac{1}{i(i+1)} \sum_{j=1}^{i} (x_{(j)} - x_{(i+1)}) \]
\[ \geq 0, \quad \text{for} \quad i = 1, 2, ..., k. \]

we have
\[ \max_i \{\frac{1}{k} \sum_{j=1}^{i} x_{(j)}, i = 1, 2, ..., k\} = x_{(1)} , \]
and we get
\[ \max_i (x_i^+ / w_i^+) = \max \left( x_{(1)}, \frac{1}{k} \sum_{j=1}^{n} x_{(j)} \right) . \]
The same expression is obtained using the formula of Mudholkar and Friemer (1985).

(2). Consider \( k = 2 \) and \( p = 2 \) then \( w_j = 0, j \geq 3 \). In this case from Eq. (2.26) of Friemer and Mudholkar (1984), we get

\[
\|x\|_{0,2}^2 = \begin{cases} 
\frac{(x_1^2 + w_2x_2^2)^{1/2}}{\sqrt{2}} & \text{if } x_1 \geq \sqrt{w_2x_2} \\
\frac{(x_1 + \sqrt{w_2x_2})/\sqrt{2}}{\sqrt{2}} & \text{if } x_1 < \sqrt{w_2x_2}.
\end{cases}
\]

5. References