Characterization of the Subdifferential of Some Matrix Norms

G. A. Watson
Department of Mathematics and Computer Science
University of Dundee
Dundee DD14HN, Scotland

Submitted by George Phillip Barker

ABSTRACT

A characterization is given of the subdifferential of matrix norms from two classes, orthogonally invariant norms and operator (or subordinate) norms. Specific results are derived for some special cases.

1. INTRODUCTION

Let $\|\cdot\|$ be a norm on the space of $m \times n$ real matrices. Then if A is a given real $m \times n$ matrix, the subdifferential (or set of subgradients) of $\|A\|$ is defined by

$$\partial\|A\| = \left\{G \in \mathbb{R}^{m\times n} : \|B\| \geqslant \|A\| + \operatorname{trace}\left[(B-A)^T G\right], \text{ all } B \in \mathbb{R}^{m \times n}\right\}. \quad (1.1)$$

It is well known (and readily established) that $G \in \partial\|A\|$ is equivalent to the statements

(i) $\|A\| = \operatorname{trace}(G^T A)$,
(ii) $\|G\|^* \leqslant 1$,

where

$$\|G\|^* = \max_{\|B\| \leqslant 1} \operatorname{trace}(B^T G),$$

and $\|\cdot\|^*$ is the polar or dual norm to $\|\cdot\|$. The roles of a norm and its dual

655 Avenue of the Americas, New York, NY 10010 0024-3795/92/$5.00

can be interchanged in this definition. This paper is concerned with a characterization of the subdifferential of some important matrix norms. As well as being of interest for their own sake, results of this kind are of value in the provision of optimality conditions for optimization or approximation problems involving norms of matrices.

For some norms, the structure of the subdifferential follows immediately from known results for the vector case. In particular this is true for the norms defined by

$$\|A\| = \left(\sum_{i,j} |A_{ij}|^p \right)^{1/p}, \qquad p \geqslant 1,$$

because the matrix is being treated as an extended vector in $\mathbb{R}^{mn}$. Two other important classes of matrix norms are considered here: orthogonally invariant norms, which are dealt with in the next section, and operator or subordinate norms, which are treated in Section 3. The results can easily be generalized to complex matrices in $C^{m \times n}$ in an obvious way, but attention here will be restricted to the real case. It will be assumed in what follows (with no loss of generality) that $m \geqslant n$.

2. ORTHOGONALLY INVARIANT NORMS

This class consists of norms such that

$$\|UVA\| = \|A\|$$

for any orthogonal matrices U and V of orders m and n respectively. These matrix norms (or in fact the more general unitarily invariant norms) were introduced by von Neumann [4], and have subsequently generated much interest. Let a given matrix A have the singular value decomposition

$$A = U \Sigma V^T,$$

where U and V are orthogonal matrices and Σ is an $m \times n$ matrix with zeros except down the main diagonal, where there are the singular values in descending order

$$\sigma_1 \geqslant \cdots \geqslant \sigma_n$$

(see, for example, Golub and Van Loan [2]). All such norms can be defined by

$$\|A\| = \phi(\boldsymbol{\sigma}), \tag{2.1}$$

where $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_n)^T$, and ϕ is a symmetric gauge function; such a function satisfies the following conditions:

(i) $\phi(\mathbf{x}) > 0$, $\mathbf{x} \neq 0$,
(ii) $\phi(\alpha \mathbf{x}) = |\alpha| \phi(\mathbf{x})$,
(iii) $\phi(\mathbf{x} + \mathbf{y}) \leqslant \phi(\mathbf{x}) + \phi(\mathbf{y})$,
(iv) $\phi(\varepsilon_1 x_{i_1}, \dots, \varepsilon_n x_{i_n}) = \phi(\mathbf{x})$,

where α is a scalar, $\varepsilon_i = \pm 1$ for all i, and $i_1, \dots, i_n$ is a permutation of $1, 2, \dots, n$. The relationship between symmetric gauge functions and unitarily invariant norms was essentially worked out in [4]; see also Schatten [7], Mirsky [3]. The polar ϕ^* of the symmetric gauge function ϕ is also a symmetric gauge function, and satisfies

$$\phi^*(\mathbf{x}) = \max_{\phi(\mathbf{y}) = 1} \mathbf{x}^T \mathbf{y}.$$

The subdifferential $\partial \phi(\mathbf{x})$ is the set of vectors satisfying the analogue of (1.1), or equivalently those vectors $\mathbf{z} \in \mathbb{R}^n$ such that

(i) $\phi(\mathbf{x}) = \mathbf{x}^T \mathbf{z}$,
(ii) $\phi^*(\mathbf{z}) \leqslant 1$.

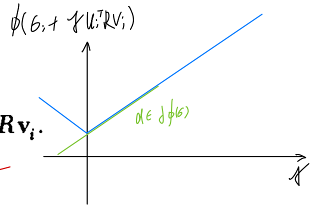
A familiar class of symmetric gauge functions is given by the l_p norms, and this leads to

$$\|A\| = \|\boldsymbol{\sigma}\|_p, \tag{2.2}$$

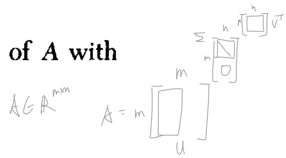
the c_p or Schatten p-norms. Well-known special cases are the l_∞ norm, which gives the spectral norm of A, and the l_2 norm, which gives the Frobenius norm. A characterization of the subdifferential of the spectral norm is given by Berens and Finzel [1] and Ziętak [10]; the latter paper also gives a characterization for the norm defined by the l_1 norm on the right hand side of (2.2). Similar results are also given by So [8]. These are the interesting l_p norms, because the subdifferential is not usually a singleton; when $1 < p < \infty$, the normed linear space is strictly convex, and there is a unique subdifferential, or equivalently the norm is differentiable. In fact the normed space is strictly convex if and only if the symmetric gauge function ϕ is strictly convex (Ziętak [10]). Here a general result for (2.1) is established which contains the above characterizations as special cases. A key feature of the analysis is the following representation of the directional derivative of $\|A\|$. The columns of U (V) will be denoted by $\mathbf{u}_i$ ($\mathbf{v}_i$), ordered so that the ith column corresponds to σ_i.

THEOREM 1. *Let A, R be given $m \times n$ matrices. Then there is a singular value decomposition of A such that*

$$\lim_{\gamma \to 0+} \frac{\|A + \gamma R\| - \|A\|}{\gamma} = \max_{\mathbf{d} \in \partial\phi(\sigma)} \sum_{i=1}^{n} d_i \mathbf{u}_i^T R \mathbf{v}_i.$$



Proof. Let σ_i be a distinct singular value of A with



$$A\mathbf{v}_i = \sigma_i \mathbf{u}_i.$$

If it is assumed that A depends smoothly on a parameter γ, then differentiating through with respect to γ and premultiplying by $\mathbf{u}_i^T$ gives

$$\frac{\partial \sigma_i}{\partial \gamma} = \mathbf{u}_i^T \frac{\partial A}{\partial \gamma} \mathbf{v}_i.$$

For multiple singular values, it is necessary to use the classical result of Rellich [5] that the eigenvalues of a matrix which is an analytic function of a single variable can always be numbered so that they are each analytic functions of the variable; the eigenvectors can be similarly defined. Using the relationship between eigenvalues and singular values, and eigenvectors and singular vectors, it follows that if the singular values of the matrix $A + \gamma R$, where A and R are given $m \times n$ matrices, are denoted by $\sigma_i(\gamma)$, $i = 1, \ldots, n$. Then

$$\sigma_i(\gamma) = \sigma_i + \gamma \mathbf{u}_i^T R \mathbf{v}_i + o(\gamma), \qquad i = 1, \ldots, n, \tag{2.3}$$

where $\sigma_i = \sigma_i(0)$, and $\mathbf{u}_i$ and $\mathbf{v}_i$ are singular vectors of A corresponding to σ_i.

Now

$$\|A\| = \phi(\boldsymbol{\sigma}) \geqslant \boldsymbol{\sigma}^T \mathbf{d}(\gamma) \qquad \text{for any} \quad \mathbf{d}(\gamma) \in \partial\phi(\boldsymbol{\sigma}(\gamma))$$

$$= \boldsymbol{\sigma}(\gamma)^T \mathbf{d}(\gamma) - \gamma \sum_{i=1}^{n} d_i(\gamma) \mathbf{u}_i^T R \mathbf{v}_i + o(\gamma)$$

$$= \|A + \gamma R\| - \gamma \sum_{i=1}^{n} d_i(\gamma) \mathbf{u}_i^T R \mathbf{v}_i + o(\gamma). \tag{2.4}$$

Also

$$\|A + \gamma R\| = \phi(\boldsymbol{\sigma}(\gamma)) \geqslant \boldsymbol{\sigma}(\gamma)^T \mathbf{d} \qquad \text{for any} \quad \mathbf{d} \in \partial\phi(\boldsymbol{\sigma})$$

$$= \|A\| + \gamma \sum_{i=1}^{n} d_i \mathbf{u}_i^T R \mathbf{v}_i + o(\gamma). \tag{2.5}$$

From (2.4) and (2.5), it follows that if $\gamma > 0$,

$$\sum_{i=1}^{n} d_i \mathbf{u}_i^T R \mathbf{v}_i + o(1) \leqslant \frac{\|A + \gamma R\| - \|A\|}{\gamma} \leqslant \sum_{i=1}^{n} d_i(\gamma) \mathbf{u}_i^T R \mathbf{v}_i + o(1).$$

Letting $\gamma \to 0+$, the result follows, because (going to a subsequence if necessary) $\mathbf{d}(\gamma) \to \bar{\mathbf{d}} \in \partial\phi(\boldsymbol{\sigma})$ (for example, Rockafellar [6]). ■

Just as the vector $\boldsymbol{\sigma}$ is related to the diagonal elements of the diagonal matrix Σ, it will be assumed in what follows that there exists the same relationship between diagonal matrices and the corresponding lowercase letters. The notation conv$\{\cdot\}$ will signify, as usual, the convex hull of a set.

THEOREM 2. *Let D denote an $m \times n$ diagonal matrix. Then*

an intuitive proof using the relationship between $\|A\|$ and $\phi(\sigma)$

$$\partial\|A\| = \text{conv}\{UDV^T,\ A = U\Sigma V^T,\ \mathbf{d} \in \partial\phi(\boldsymbol{\sigma})\}. \tag{2.6}$$

$\|A + U\,\text{diag}(x)V^T\| = \|U(\Sigma + \text{diag}(x))V^T\|$

$= \phi(\sigma + x) \approx \phi(\sigma) + x^T \nabla\phi(\sigma)$

$= \|A\| + tr[\text{diag}(x)\ \text{diag}(\nabla\phi(\sigma))]$

$= \|A\| + tr[\text{diag}(x) V^T V\, \text{diag}(\nabla\phi(\sigma)) U^T U]$

$= \|A\| + tr[U\,\text{diag}(x) V^T V\, \text{diag}(\nabla\phi(\sigma)) U^T]$

$= \|A\| + \langle U\,\text{diag}(x) V^T,\ U\,\text{diag}(\nabla\phi(\sigma)) V^T \rangle$

$\therefore \nabla\|A\| = U\,\text{diag}(\nabla\phi(\sigma)) V^T$

$\partial\|A\| = \{U\,\text{diag}(d) V^T \mid d \in \text{conv}\{\nabla\phi(\sigma)\}\}$

Proof. Denote the set described inside the braces on the right hand side of (2.6) by $S(A)$, and let $G \in \text{conv}\{S(A)\}$. Then

$$\text{trace}(G^T A) = \text{trace}\left(A^T \sum_i \lambda_i U_i D_i V_i^T\right),$$

where $\lambda_i \geqslant 0$, $\Sigma_i \lambda_i = 1$, and for each i, $\mathbf{d}_i \in \partial\phi(\boldsymbol{\sigma})$, $A = U_i \Sigma V_i^T$ is a singular value decomposition. Thus

$$\begin{aligned}\text{trace}(G^T A) &= \text{trace}\left(\sum_i \lambda_i V_i \Sigma U_i^T U_i D_i V_i^T\right)\\ &= \sum_i \lambda_i \mathbf{d}_i^T \boldsymbol{\sigma}\\ &= \|A\|.\end{aligned}$$

Further

$$\begin{aligned}\|G\|^* &= \max_{\|R\| \leqslant 1} \text{trace}(G^T R)\\ &= \max_{\|R\| \leqslant 1} \text{trace}\left(R^T \sum_i \lambda_i U_i D_i V_i^T\right), \qquad \text{as above.}\end{aligned}$$

Now for each i,

$$\|U_i D_i V_i^T\|^* = \|D_i\|^* = \phi^*(\mathbf{d}_i) = 1,$$

using the known fact that

$$\|A\|^* = \phi^*(\boldsymbol{\sigma}).$$

Thus

$$\text{trace}\left(R^T U_i D_i V_i^T\right) \leqslant \|R\|,$$

and $\|G\|^* \leqslant 1$, showing that $G \in \partial\|A\|$.

Now assume that $G \in \partial\|A\|$ but $G \notin \operatorname{conv} S(A)$. Then by a well-known separation result (see, for example, Watson [9, p. 13]) there exists $R \in \mathbb{R}^{m \times n}$ such that

$$\operatorname{trace}(R^T H) < \operatorname{trace}(R^T G) \qquad \text{for all} \quad H \in S(A),$$

so that

$$\max_{H \in S(A)} \operatorname{trace}(R^T H) < \max_{G \in \partial\|A\|} \operatorname{trace}(R^T G),$$

or for any singular value decomposition

$$\max_{\mathbf{d} \in \partial\phi(\boldsymbol{\sigma})} \sum_{i=1}^{n} d_i \mathbf{u}_i^T R \mathbf{v}_i < \max_{G \in \partial\|A\|} \operatorname{trace}(R^T G).$$

But the right hand side is just the standard expression for the directional derivative of the convex function $\|A\|$ in the direction R (for example, Rockafellar [6]), and so Theorem 1 is contradicted. The proof is completed. ∎

EXAMPLE 1. Let $\phi(\boldsymbol{\sigma}) = \|\boldsymbol{\sigma}\|_\infty$, giving rise to the spectral norm of A. Then

$$\partial\|\boldsymbol{\sigma}\|_\infty = \operatorname{conv}\{\mathbf{e}_i, i : \sigma_i = \sigma_1\}.$$

Let $A = U\Sigma V^T$ be any singular value decomposition, and let the multiplicity of σ_1 be t, with

$$U = \left[U^{(1)} \,\vdots\, U^{(2)}\right], \qquad V = \left[V^{(1)} \,\vdots\, V^{(2)}\right], \tag{2.7}$$

where $U^{(1)}$ and $V^{(1)}$ have t columns. Then

$$A = \sigma_1 U^{(1)} V^{(1)T} + U^{(2)} \Sigma^{(2)} V^{(2)T}, \quad \text{say.}$$

Any element of the set $\partial\|A\|$ can be written as

$$G = \sum_i \mu_i U_i^{(1)} D_i^{(1)} V_i^{(1)T},$$

where $\mu_i \geqslant 0$, $\Sigma_i \mu_i = 1$, and for each i, $A = U_i \Sigma V_i^T$ is a singular value decomposition, $\mathbf{d}_i \in \partial\|\sigma\|_\infty$. The superscripts have the same meaning as in (2.7). Expressing $U_i^{(1)}$ and $V_i^{(1)}$ in terms of $U^{(1)}$ and $V^{(1)}$, it follows that

$$G = \sum_i \mu_i U^{(1)} X_i D_i^{(1)} X_i^T V^{(1)T},$$

where each X_i is a $t \times t$ orthogonal matrix. Thus

$$G = U^{(1)} H V^{(1)T},$$

where $H \geqslant 0$, that is, H is a symmetric positive semidefinite $t \times t$ matrix. In addition, trace $H = 1$. Thus given any singular value decomposition of A, the subdifferential is defined by

$$\partial\|A\| = \left\{ U^{(1)} H V^{(1)T} \text{ for all } H \in \mathbb{R}^{t \times t},\ H \geqslant 0, \text{ trace } H = 1 \right\}.$$

EXAMPLE 2. Let $\phi(\sigma) = \|\sigma\|_1$. For given A let there be s zero singular values, and let $A = U \Sigma V^T$ be any singular value decomposition with the matrices partitioned so that

$$U = \left[U^{(1)} \vdots\ U^{(2)} \right], \qquad V = \left[V^{(1)} \vdots\ V^{(2)} \right], \tag{2.8}$$

with $U^{(1)}$ and $V^{(1)}$ having $n - s$ columns. (Notice that this is not the same partitioning as in Example 1.) Recall that

$$\partial\|\sigma\|_1 = \left\{ \mathbf{x} \in \mathbb{R}^n : |x_i| \leqslant 1,\ x_i = 1,\ i = 1, \ldots, n - s \right\}.$$

Let $G \in \partial\|A\|$. Then

$$G = \sum_i \lambda_i U_i D_i V_i^T,$$

where $\lambda_i \geqslant 0$, $\Sigma_i \lambda_i = 1$, and for each i, $\mathbf{d}_i \in \partial\|\sigma\|_1$, and $A = U_i \Sigma V_i^T$ are singular value decompositions. Thus

$$G = U^{(1)} V^{(1)T} + \sum_i \lambda_i U_i^{(2)} W_i V_i^{(2)T},$$

where W_i is an $(m - n + s) \times s$ diagonal matrix with diagonal elements $\leqslant 1$

in modulus, and the partitioning is consistent with (2.8). Therefore

$$G = U^{(1)}V^{(1)T} + \sum_i \lambda_i U^{(2)} Y_i W_i Z_i^T V^{(2)T},$$

where the matrices Y_i and Z_i are orthogonal matrices of dimension $m - n + s$ and s respectively, and so

$$G = U^{(1)}V^{(1)T} + U^{(2)}TV^{(2)T},$$

where T is $(m - n + s)\times s$. If $\sigma_1(\cdot)$ denotes the largest singular value of a given matrix, then

$$\begin{aligned}\sigma_1(T) &= \sigma_1\Big(\sum_i \lambda_i Y_i W_i Z_i^T\Big)\\ &\leqslant \sum_i \lambda_i \sigma_i(W_i)\\ &\leqslant 1.\end{aligned}$$

Thus given any singular value decomposition of A, a characterization of the subdifferential in this case is given by

$$\partial\|A\| = \Big\{U^{(1)}V^{(1)T} + U^{(2)}TV^{(2)T} \text{ for all } T \in \mathbb{R}^{(m-n+s)\times s},\ \sigma_1(T) \leqslant 1\Big\}.$$

3. OPERATOR NORMS

Let $\|\cdot\|_A$ and $\|\cdot\|_B$ be norms on $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively. Then a norm on $m \times n$ matrices may be defined by

$$\|A\| = \max_{\|\mathbf{x}\|_B = 1} \|A\mathbf{x}\|_A. \tag{3.1}$$

The required subdifferential characterization can be established by arguments similar to those of the previous section. It is convenient to define the set of vector pairs

$$\Phi(A) = \{\mathbf{v} \in \mathbb{R}^n, \mathbf{w} \in \mathbb{R}^m : \|\mathbf{v}\|_B = 1,\ A\mathbf{v} = \|A\|\mathbf{u},\ \|\mathbf{u}\|_A = 1,\ \mathbf{w} \in \partial\|\mathbf{u}\|_A\}. \tag{3.2}$$

Clearly, this set contains vectors v where the norm is attained in the expression (3.1).

THEOREM 3.

$$\lim_{\gamma \to 0+} \frac{\|A + \gamma R\| - \|A\|}{\gamma} = \max_{(\mathbf{v},\mathbf{w}) \in \Phi(A)} \mathbf{w}^T R \mathbf{v}.$$

Proof. We have

$$\begin{aligned} \|A\| &= \max_{\|\mathbf{x}\|_B = 1} \|A\mathbf{x}\|_A \\ &\geqslant \|A\mathbf{v}(\gamma)\|_A, \\ &\geqslant \mathbf{w}(\gamma)^T A\mathbf{v}(\gamma) \qquad \text{for any} \quad (\mathbf{v}(\gamma), \mathbf{w}(\gamma)) \in \Phi(A + \gamma R) \\ &= \|A + \gamma R\| - \gamma \mathbf{w}(\gamma)^T R\mathbf{v}(\gamma). \end{aligned} \tag{3.3}$$

Also

$$\begin{aligned} \|A + \gamma R\| &\geqslant \|(A + \gamma R)\mathbf{v}\|_A \\ &\geqslant \mathbf{w}^T (A + \gamma R)\mathbf{v} \qquad \text{for any} \quad (\mathbf{w}, \mathbf{v}) \in \Phi(A) \\ &= \|A\| + \gamma \mathbf{w}^T R\mathbf{v}. \end{aligned} \tag{3.4}$$

From (3.3) and (3.4) it follows that if $\gamma > 0$,

$$\mathbf{w}^T R\mathbf{v} \leqslant \frac{\|A + \gamma R\| - \|A\|}{\gamma} \leqslant \mathbf{w}(\gamma)^T R\mathbf{v}(\gamma).$$

Now define, for all γ, $\mathbf{u}(\gamma)$ by

$$(A + \gamma R)\mathbf{v}(\gamma) = \|A + \gamma R\| \mathbf{u}(\gamma).$$

Then letting $\gamma \to 0+$ along a subsequence if necessary, it follows that

$$\begin{aligned} \mathbf{v}(\gamma) &\to \bar{\mathbf{v}}, & \|\bar{\mathbf{v}}\|_B &= 1, \\ \mathbf{w}(\gamma) &\to \bar{\mathbf{w}}, & \|\bar{\mathbf{w}}\|_A^* &= 1, \\ \mathbf{u}(\gamma) &\to \bar{\mathbf{u}}, & \|\bar{\mathbf{u}}\|_A &= 1. \end{aligned}$$

Further

$$A\bar{\mathbf{v}} = \|A\|\bar{\mathbf{u}},$$

and since $\mathbf{w}(\gamma)^T\mathbf{u}(\gamma) = \|\mathbf{u}(\gamma)\|_A$, it follows that, taking limits,

$$\bar{\mathbf{w}} \in \partial\|\bar{\mathbf{u}}\|_A.$$

Thus $(\bar{\mathbf{w}}, \bar{\mathbf{v}}) \in \Phi(A)$, and the result follows. ■

THEOREM 4.

$$\partial\|A\| = \operatorname{conv}\{\mathbf{w}\mathbf{v}^T : (\mathbf{v},\mathbf{w}) \in \Phi(A)\}.$$

Proof. Let $S(A)$ be the set described in braces on the right hand side, and let $G \in \operatorname{conv} S(A)$. Now

$$\operatorname{trace}(G^TA) = \operatorname{trace}\left(\sum_i \lambda_i \mathbf{v}_i \mathbf{w}_i^T A\right),$$

where $\lambda_i \geqslant 0$, $\Sigma_i\lambda_i = 1$, and for each i, $(\mathbf{v}_i, \mathbf{w}_i) \in \Phi(A)$. Thus

$$\begin{aligned}\operatorname{trace}(G^TA) &= \sum_i \lambda_i \|A\| \mathbf{w}_i^T \mathbf{u}_i \\ &= \|A\|.\end{aligned}$$

Also

$$\begin{aligned}\|G\|^* &= \max_{\|R\| \leqslant 1} \operatorname{trace}(G^TR) \\ &= \max_{\|R\| \leqslant 1} \sum_i \lambda_i \mathbf{w}_i^T R \mathbf{v}_i \qquad \text{(as above)} \\ &\leqslant 1,\end{aligned}$$

using the fact that for all i,

$$\mathbf{w}_i^T R \mathbf{v}_i \leqslant \|R\|.$$

Now let $G \in \partial\|A\|$, but assume that $G \notin \operatorname{conv} S(A)$. Then, as in Theorem 2, there exists R such that

$$\operatorname{trace}[R^T(\mathbf{w}\mathbf{v}^T - G)] < 0 \qquad \text{for all}(\mathbf{w},\mathbf{v}) \in \Phi(A),$$

so that

$$\mathbf{w}^T R\mathbf{v} < \operatorname{trace}(R^T G) \qquad \text{for all} \quad (\mathbf{w},\mathbf{v}) \in \Phi(A),$$

and therefore

$$\max_{(\mathbf{v},\mathbf{w}) \in \Phi(A)} \mathbf{w}^T R\mathbf{v} < \max_{G \in \partial\|A\|} \operatorname{trace}(R^T G).$$

The fact that the right hand side is just the directional derivative of $\|A\|$ in the direction R leads to a contradiction of Theorem 3, and the result is proved. ■

EXAMPLE 3. The most common operator norm is the one with both vector norms l_2 norms. This is just the spectral norm, and corresponds to the l_∞ case treated in Example 1 of the previous section, but the recovery of the subdifferential will be repeated from the operator norm point of view. Here

$$\partial\|A\| = \operatorname{conv}\{\mathbf{u}\mathbf{v}^T : \|\mathbf{u}\|_2 = \|\mathbf{v}\|_2 = 1,\ A\mathbf{v} = \|A\|\mathbf{u}\}.$$

It is readily seen that any element of the subdifferential has the form

$$\sum_i \lambda_i \mathbf{u}_i \mathbf{v}_i^T,$$

where $\lambda_i \geqslant 0$, $\Sigma_i \lambda_i = 1$, and $\mathbf{u}_i$ and $\mathbf{v}_i$ are any left and right singular vectors of A corresponding to σ_1. The form established in Example 1 follows in a straightforward manner.

REFERENCES

1 H. Berens and M. Finzel, A continuous selection of the metric projection in matrix spaces, in *Numerical Methods of Approximation Theory*, Vol. 8 (L. Collatz, G. Meinardus, and G. Nurnberger, Eds.) ISNM 81, Birkhäuser-Verlag, 1987, pp. 21–29.

2 G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd ed., Johns Hopkins U.P., Baltimore, 1989.
3 L. Mirsky, Symmetric gauge functions and unitarily invariant norms, *Quart. J. Math. Oxford Ser.* (2) 11:50–59 (1960).
4 J. von Neumann, Some matrix inequalities and metrization of metric spaces, *Tomsk Univ. Rev.* 1:286–300 (1937).
5 F. Rellich, *Perturbation Theory of Eigenvalue Problems*, Gordon and Breach, New York, 1969.
6 R. T. Rockafellar, *Convex Analysis*, Princeton U.P., Princeton, 1970.
7 R. Schatten, *Norm Ideals of Completely Continuous Operators*, Springer-Verlag, Berlin, 1960.
8 W. So, Facial structures of Schatten p-norms, *Linear and Multilinear Algebra* 27:207–212 (1990).
9 G. A. Watson, *Approximation Theory and Numerical Methods*, Wiley, Chichester, 1980.
10 K. Ziętak, On the characterization of the extremal points of the unit sphere of matrices, *Linear Algebra Appl.* 106:57–75 (1988).

Received 1 October 1990; final manuscript accepted 1 May 1991