Shashank Singh[1]

2 Subgradients of matrix norms [30 points, 8+9+9+4] (Yifei)

We use $\langle \cdot, \cdot \rangle$ to denote the dot product and $\mathcal{N}$ and $\mathcal{R}$ to denote the nullspace and range.

(a) We use a more convenient definition of the operator norm:

$$\|A\|_{\text{op}} := \max_{\|x\|=1} \|Ax\|_2.$$

It is easy to check that this definition is identical to the given one:

$$\max_{\|x\|=1} \|Ax\|_2 = \sqrt{\max_{\|x\|=1} \|Ax\|_2^2} = \sqrt{\max_{\|x\|=1} \langle Ax, Ax \rangle} = \sqrt{\max_{\|x\|=1} \langle A^T Ax, x \rangle} = \|A\|_{\text{op}},$$

where the last equality follows from the Spectral Theorem, since $A^T A$ is symmetric.

Since $U^T W = 0$, $\mathcal{R}(U)$ and $\mathcal{R}(W)$ are orthogonal. Since $WV = 0$, $\mathcal{R}(V) \subseteq \mathcal{N}(W)$, so that $\mathcal{N}(V^T) + \mathcal{N}(W) = \mathbb{R}^n$ (i.e., any $x \in \mathbb{R}^n$ has a decomposition $x = x_1 + x_2$ with $x_2 \in \mathcal{N}(V^T)$, $x_1 \in \mathcal{N}(W), \langle x_1, x_2 \rangle = 0$). Since U and V are orthogonal, $\|UV^T\|_{\text{op}} = 1$. Thus, by the Pythagorean Theorem, $\forall x \in \mathbb{R}^n$,

$$\begin{aligned}\|(UV^T + W)x\|_2^2 = \|UV^T x\|_2^2 + \|Wx\|_2^2 &= \|UV^T x_1\|_2^2 + \|Wx_2\|_2^2 \\ &\leq \|UV^T\|_{\text{op}}^2 \|x_1\|_2^2 + \|W\|_{\text{op}}^2 \|x_2\|_2^2 \\ &\leq \|x_1\|_2^2 + \|x_2\|_2^2 = \|x\|_2^2,\end{aligned}$$

so that $\|UV^T + W\|_{\text{op}} \leq 1$. $\blacksquare$

(b) Since U and V are orthogonal and Σ is diagonal,

$$VU^T A = VU^T U \Sigma V^T = V \Sigma V^T = \Sigma.$$

Since $U^T W = 0$,

$$W^T A = W^T U \Sigma V^T = (U^T W)^T \Sigma V^T = 0.$$

Thus,

$$\operatorname{tr}((UV^T + W)^T A) = \operatorname{tr}(VU^T A + W^T A) = \operatorname{tr}(\Sigma). \quad \blacksquare$$

(c) Suppose $W \in \mathbb{R}^{m \times n}$ with $\|W\|_{\text{op}} \leq 1$, $U^T W = 0$, and $WV = 0$. Then, by parts (a) and (b),

$$\begin{aligned}\operatorname{tr}((UV^T + W)^T (B - A)) &= \operatorname{tr}((UV^T + W)^T B) - \operatorname{tr}((UV^T + W)^T A) \\ &\leq \max_{\|C\|_{\text{op}} \leq 1} \operatorname{tr}(C^T B) - \operatorname{tr}(\Sigma) = \|B\|_* - \|A\|_*,\end{aligned}$$

where the last equality follows from duality of the trace and operator norms and the definitions of Σ and the trace norm. Thus, $UV^T + W \in \partial \|A\|_*$. $\blacksquare$

[1]sss1@andrew.cmu.edu

(d) Suppose $G = u_j v_j^T$, where j satisfies $\Sigma_{jj} = \Sigma_{11}$.

We first prove two lemmas, analogous to parts (a) and (b) above.

Lemma 1: $\|G\|_* = 1$.

Proof:
$$G^T G = (u_j v_j^T)^T u_j v_j^T = v_j u_j^T u_j v_j^T = \|u\|_2^2 v_j v_j^T = v_j v_j^T.$$
Thus, $\|G\|_* = \mathrm{tr}(G^T G) = \|v_j\|_2^2 = 1$, proving the lemma.

Lemma 2: $\mathrm{tr}(G^T A) = \|A\|_{\mathrm{op}}$.

Proof: $G^T A = v_j u_j^T U \Sigma V^T$. Using the fact that U and V are orthogonal, it can be checked that this reduces to the matrix $\Sigma[1_{jj}]$, where $[1_{jj}] \in \mathbb{R}^{r \times r}$ denotes the matrix with $\|u\|_2^2 \|v\|_2^2 = 1$ in the index (j, j) and zeros elsewhere. Thus, $\mathrm{tr}(G^T A) = \sigma_j(A) = \|A\|_{\mathrm{op}}$, proving the lemma.

By these lemmas and the duality of the trance and operator norms,

$$\begin{aligned}
\mathrm{tr}(G^T(B - A)) &= \mathrm{tr}(G^T B) - \mathrm{tr}(G^T A) \\
&\leq \max_{\|C\|_* \leq 1} \mathrm{tr}(C^T B) - \|A\|_{\mathrm{op}} = \|B\|_{\mathrm{op}} - \|A\|_{\mathrm{op}},
\end{aligned}$$

Thus, $G \in \partial \|A\|_*$. The desired result follows, since $\partial \|A\|_{\mathrm{op}}$ is convex. ■

(e) If the result in (d) is an equality, then $\partial \|A\|_{\mathrm{op}}$ contains exactly one element (and hence $\| \cdot \|_{\mathrm{op}}$ is differentiable at A) precisely when there is a unique j such that $\Sigma_{jj} = \Sigma_{11}$ (i.e., the largest singular value of A is not repeated). ■