Equivalent Ways of Expressing the Trace Norm of a Matrix

Jason D. M. Rennie
jrennie@gmail.com

September 8, 2005

Srebro (§5.3.1 of [2]) provides the following lemma and definition:

Lemma 1 *For any matrix X, the following are all equal:*

1. $\min_{U,V|X=UV^T} \|U\|_{Fro}\|V\|_{Fro}$
2. $\min_{U,V|X=UV^T} \frac{1}{2}\left(\|U\|_{Fro}^2 + \|V\|_{Fro}^2\right)$
3. *The sum of the singular values of X, i.e. trace S, where $X = USV^T$ is the singular value decomposition of X.*

Furthermore, if $X = USV^T$ is the singular value decomposition of X, then the matrices $U\sqrt{S}$ and $V\sqrt{S}$ minimize the first quantity.

Definition 1 *The trace norm $\|X\|_{tr}$ of a matrix is given by the three quantities in Lemma 1.*

Here we provide discussion of and a partial proof of this Lemma.

First, we give names to the two minimization objectives. Let

$$J_1(U,V) = \|U\|_{\text{Fro}}\|V\|_{\text{Fro}}, \quad \text{and} \tag{1}$$

$$J_2(U,V) = \frac{1}{2}\left(\|U\|_{\text{Fro}}^2 + \|V\|_{\text{Fro}}^2\right). \tag{2}$$

We begin with the end note of the Lemma. Let $X = USV^T$ be the singular value decomposition (SVD) of X. Then, if we use the factorization $(U\sqrt{S}, V\sqrt{S})$, then $J_1(U\sqrt{S}, V\sqrt{S}) = J_2(U\sqrt{S}, V\sqrt{S}) = \|X\|_{\text{tr}}$. To see this, note that U and V from the SVD are orthogonal matrices (rows are orthonormal) and if Q is orthogonal, then $\|QA\|_{\text{Fro}} = \|A\|_{\text{Fro}}$ [1]. Then,

$$J_1(U\sqrt{S}, V\sqrt{S}) = \|U\sqrt{S}\|_{\text{Fro}}\|V\sqrt{S}\|_{\text{Fro}} = \|\sqrt{S}\|_{\text{Fro}}^2 = \|X\|_{\text{tr}}, \quad \text{and} \tag{3}$$

$$J_2(U\sqrt{S}, V\sqrt{S}) = \frac{1}{2}\left(\|U\sqrt{S}\|_{\text{Fro}}^2 + \|V\sqrt{S}\|_{\text{Fro}}^2\right) = \|\sqrt{S}\|_{\text{Fro}}^2 = |X\|_{\text{tr}}. \tag{4}$$

Now, we consider other factorizations. Let A and B be diagonal matrices so that $AB = S$. We consider factorizations of the form (UA, VB). Note that for such a factorization,

$$J_2(UA, VB) = \sum_i A_i^2 + B_i^2 = \sum_i A_i^2 + \left(\frac{S_i}{A_i}\right)^2 \tag{5}$$

Solving for the zero gradient, we find the minimum at $A_i = \sqrt{S_i}$,

$$\frac{\partial J_2(UA, VB)}{\partial A_i} = 2A_i - 2S_i^2 A_i^{-3} = \sum_i 1 - S_i^2 A_i^{-4}. \tag{6}$$

This can be seen intuitively by realizing that the minimization of each A_i is separate, that $A_i B_i = S_i$ traces a parabola in the first quadrant, that J_2 is the (squared) L_2 distance from the origin of the concatenated diagonals of A & B (treated as a single vector), and that $A = B$ minimizes this distance. Now, we look at J_1. Note that

$$J_1(UA, VB) = \sqrt{\sum_i A_i^2} \sqrt{\sum_i \left(\frac{S_i}{A_i}\right)^2}, \tag{7}$$

and

$$\frac{\partial J_1(UA, VB)}{\partial A_i} = \frac{2A_i}{\sqrt{\sum_{i'} A_{i'}^2}} \sqrt{\sum_{i'} \left(\frac{S_{i'}}{A_{i'}}\right)^2} - \sqrt{\sum_{i'} A_{i'}^2} \frac{2S_i^2 A_i^{-3}}{\sqrt{\sum_{i'} \left(\frac{S_{i'}}{A_{i'}}\right)^2}} \tag{8}$$

$$= \sum_{i'} \left(\frac{S_{i'}}{A_{i'}}\right)^2 - S_i^2 A_i^{-4} \sum_{i'} A_{i'}^2. \tag{9}$$

Again, solving for the zero gradient, we find the minimum at $A_i = \sqrt{S_i}$.

Left to be shown is that either (1) arbitrary factorizations yield values $\geq \|X\|_{\mathrm{tr}}$, or (2) (UA, VB) decompositions give all possible Frobenius norm values.

References

[1] M. Brookes. The matrix reference manual. http://www.ee.ic.ac.uk/hp/staff/dmb/matrix/intro.html, 1998.

[2] N. Srebro. *Learning with Matrix Factorizations*. PhD thesis, Massachusetts Institute of Technology, 2004.