The Trace Norm Inequality and its Consequences*

Jason D. M. Rennie
jrennie@gmail.com

March 1, 2006

Abstract

We discuss the trace norm inequality and its consequences.

Let $X \in \mathbb{R}^{n \times m}$ be a matrix. Let us construct a partitioning of X into $X_1 \in \mathbb{R}^{k \times m}$ and $X_2 \in \mathbb{R}^{(n-k) \times m}$ $(0 \le k \le n)$ so that each row of X maps to one row of X_1 or one row of X_2 (e.g. $X = [X_1; X_2]$). Let $\|X\|_\Sigma$ be the trace norm of X (the sum of its singular values). Then,

$$\|X\|_\Sigma \le \|X_1\|_\Sigma + \|X_2\|_\Sigma. \tag{1}$$

We call this the trace norm inequality. This is simple result of Corollary 3.4.3 of [1]. Corolarry 3.4.3 states that if $A, B \in M_{m,n}$, then the sum of their first k singular values are no less than the sum of the first k singular values of $A+B$. Let A and B have respective ordered singular values $\sigma_1(A) \ge \cdots \ge \sigma_q(A) \ge 0$ and $\sigma_1(B) \ge \cdots \ge \sigma_q(B) \ge 0$, $q \equiv \min\{m, n\}$; let $\sigma_1(A+B) \ge \cdots \ge \sigma_q(a+B) \ge 0$ be the ordered singular values of $A+B$. Then

$$\sum_{i=1}^{k} \sigma_i(A+B) \le \sum_{i=1}^{k} \sigma_i(A) + \sum_{i=1}^{k} \sigma_i(B), \quad k = 1, \ldots, q. \tag{2}$$

This is slightly more powerful than we need since we only deal with the case k=q. Let A and B be X_1 and X_2 with zero padded rows, $A = [X_1; 0]$, $B = [0; X_2]$. Let $X_1 = U_1 \Sigma_1 V_1^T$ be a singular value decomposition for X_1. Note that $A = [U_1; 0]\Sigma_1 V_1^T$. I.e. $\|X_1\|_\Sigma = \|A\|_\Sigma$. Similarly, $\|X_2\|_\Sigma = \|B\|_\Sigma$. Since $X = A + B$, we have that $\|X\|_\Sigma \le \|X_1\|_\Sigma + \|X_2\|_\Sigma$. $\square$

In earlier work, we have discussed a trace norm distribution [2],

$$P(X) = \frac{1}{Z} \exp(-\lambda \|X\|_\Sigma). \tag{3}$$

*Joint work with John Barnett and Tommi Jaakkola.

We can use the trace norm inequality to establish a relation among normalization constants. Continuing from (1),

$$\exp(-\lambda\|X\|_\Sigma) \geq \exp(-\lambda\|X_1\|_\Sigma)\exp(-\lambda\|X_2\|_\Sigma) \tag{4}$$

$$\int \exp(-\lambda\|X\|_\Sigma)dX \geq \int \exp(-\lambda\|X_1\|_\Sigma)dX_1 \int \exp(-\lambda\|X_2\|_\Sigma)dX_2 \tag{5}$$

$$\log Z \geq \log Z_1 + \log Z_2, \tag{6}$$

where Z, Z_1, and Z_2 are the normalization constants correspondoning to $P(X)$, $P(X_1)$, and $P(X_2)$, respectively. The sum of log-normalization constants for X_1 and X_2 is less than the log-normalization constant for X.

References

[1] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis.* Cambridge University Press, 1991.

[2] J. D. M. Rennie. Learning structure with the trace norm distribution. http://people.csail.mit.edu/jrennie/writing, February 2006.