Jacobian of the Singular Value Decomposition with Application to the Trace Norm Distribution*

Jason D. M. Rennie
jrennie@gmail.com

February 27, 2006

Abstract

We consider the calculation of the normalization constant for the trace norm distribution. This is an integral over singular values, so we find it beneficial to use a singular value decomposition change of variables. We walk through the steps to compute the Jacobian of the SVD. Finally, we apply the change of variables to our integral and provide a partial evaluation of the integral.

1 Introduction

We are interested in the trace norm distribution [3],

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

where the normalization constant is an integral over matrices, $Z = \int \exp(-\|X\|_\Sigma)dX$. $\|X\|_\Sigma$ is the trace norm of X, which is the sum of its singular values. Evaluation of this integral would clearly benefit from a change of variables to the singular value decomposition (SVD). The SVD of a matrix $X \in \mathbb{R}^{n\times m}$ (wlog $n > m$) is a product of three matrices, $U \in V_{n,m}$, $\Sigma \in \text{diag}(\mathbb{R}^m)$, and $V \in O(m)$, where $V_{n,m}$ is the Stiefel manifold[1] and $O(m)$ is the orthogonal group[2]. See [2] for additional information on these surfaces. The SVD is $X = U\Sigma V^T$. Note that since Σ is diagonal, and rows of U and V must be orthogonal and sum to one, the number of entries in X is equal to the number of free entries of the SVD. We assume that the singular values (diagonal elements of Σ) are ordered. If the singular values are unique, $\sigma_1 > \cdots > \sigma_m$, then SVD is unique up to a change in signs of corresponding columns of U and V.

*Joint work with John Barnett.

[1]Stiefel manifold is tall skinny matrices $Y \in \mathbb{R}^{n\times m}$ with orthogonal columns, $Y^TY = I_m$.

[2]Orthogonal group is orthogonal matrices $Q \in \mathbb{R}^{m\times m}$ ($Q^TQ = I$).

2 Jacobian for the Change of Variables

To change variables to the SVD, we must calcuate the Jacobian. We follow the derivation given in [1], correcting typos and providing additional detail. $X = U\Sigma V^T$. So,

$$dX = U d\Sigma V^T + dU \Sigma V^T + U \Sigma dV^T. \tag{2}$$

Let $H \in \mathbb{R}^{n \times n}$ be the orthogonal matrix with first M columns identical to U. Define $Y = H^T dX V$. Then

$$dY = H^T dX V = I_{n,m} d\Sigma + H^T dU \Sigma - I_{n,m} \Sigma V^T dV. \tag{3}$$

Recall that $V^T V = I$. Hence, $dV^T V = -V^T dV$ or $V^T dV$ is anti-symmetric, which is the reason for the negation of the last term. Similarly, $H^T dU$ is anti-symmetric. Continuing, we take the exterior product[3] of elements of dY. Only the first term of (3) is not anti-symmetric, so the exterior product of the diagonal elements is $d\Sigma$. Let u_i be the i^{th} column of U. Let v_i be the i^{th} column of V. The upper-triangular, $i < j \leq m$, elements of dY are

$$dY_{ij} = \sigma_j u_i^T du_j - \sigma_i v_i^T dv_j, \tag{4}$$

and the lower-triangular elements are

$$dY_{ji} = \sigma_i u_j^T du_i - \sigma_j v_j^T dv_i. \tag{5}$$

Note that $dY_{ij} = -\sigma_j u_j^T du_i + \sigma_i v_j^T dv_i$ due to anti-symmetry. So, the exterior product is

$$dY_{ij} \wedge dY_{ji} = (\sigma_i^2 - \sigma_j^2)(v_j^T dv_i) \wedge (u_j^T du_i). \tag{6}$$

Hence, the product of off-diagonal terms in the upper square part of dY is

$$\prod_{i<j\leq m} (\sigma_i^2 - \sigma_j^2)(V^T dV)^\wedge (U^T dU)^\wedge. \tag{7}$$

For $i > m$,

$$dY_{ij} = \sigma_j h_i^T du_j. \tag{8}$$

Hence, each σ_j appears an additional $n - m$ times in dY. Define $\tilde{H}$ as the portion of H that does not come from U. I.e. $H = [U\tilde{H}]$. Then, the portion of the exterior product from below the top square is

$$\prod_{i\leq m} \sigma_i^{n-m} (\tilde{H}^T dU)^\wedge. \tag{9}$$

Putting everything together, we get

$$dY = \prod_{i<j\leq m} (\sigma_i^2 - \sigma_j^2) \prod_{i\leq m} \sigma_i^{n-m} (d\Sigma)^\wedge (V^T dV)^\wedge (H^T dU)^\wedge. \tag{10}$$

Note that H and V are rotation matrices. They do not affect volume, so we can use dY instead of dX in our integral.

[3] See [1] for a tutorial on the wedge/exterior product.

3 Trace Norm Distribution Integral

Applying the change of variables to our integral, we get

$$Z = \int \exp(-\lambda \|X\|_\Sigma) dX = \frac{1}{2^m} \int \prod_{i<j\leq m} (\sigma_i^2 - \sigma_j^2) \prod_{i\leq m} \sigma_i^{n-m} e^{-\sigma_i} (d\Sigma)^\wedge (V^T dV)^\wedge (H^T dU)^\wedge. \quad (11)$$

Recall that except for a measure zero set, the SVD is unique up to a sign, hence the 2^{-m} term. Note that Z is really the product of three separate integrals,

$$Z = \frac{1}{2^m} \int \prod_{i<j\leq m} (\sigma_i^2 - \sigma_j^2) \prod_{i\leq m} \sigma_i^{n-m} e^{-\sigma_i} (d\Sigma)^\wedge \int (V^T dV)^\wedge \int (H^T dU)^\wedge. \quad (12)$$

Note that $\int (V^T dV)^\wedge$ is the volume of the orthogonal group, $O(m)$, and $\int (H^T dU)^\wedge$ is the volume of the Stiefel manifold, $V_{n,m}$. The Stiefel manifold is a generalization of the orthogonal group. Edelman [2] provides the Stiefel manifold volume,

$$\mathrm{Vol}(V_{n,m}) = \prod_{i=n}^{n-m+1} A_i = \prod_{i=n}^{n-m+1} \frac{2\pi^{i/2}}{\Gamma\left(\frac{i}{2}\right)}, \quad (13)$$

where A_i is the surface area of the sphere in $\mathbb{R}^i$ of radius 1. Note that the remaining integral over singular values,

$$\int_0^\infty \cdots \int_0^{\sigma_{m-1}} \prod_{i<j\leq m} (\sigma_i^2 - \sigma_j^2) \prod_{i\leq m} \sigma_i^{n-m} e^{-\sigma_i} d\sigma_m \ldots d\sigma_1, \quad (14)$$

can be computed analytically. However, exact evaluation is intractible for large n or m.

References

[1] A. Edelman. Jacobians of matrix transforms (with wedge products). http://web.mit.edu/18.325/www/handouts.html, February 2005. 18.325 Class Notes: Finite Random Matrix Theory, Handout #3.

[2] A. Edelman. Volumes and integration. http://web.mit.edu/18.325/www/handouts.html, March 2005. 18.325 Class Notes: Finite Random Matrix Theory, Handout #4.

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