

# Analogy using vector transport

Word2grass

September 13, 2020

---

Analogy task looks like this -  $a : b :: c : d$  where  $a, b, c$  are three words represented by points on the grassmanian manifold. Now, let me denote  $a, b, c$  with orthonormal matrices  $A, B, C$  respectively.

Steps to find the unknown subspace  $D$  representing the word  $d$  -

- Use the closed form equation of the geodesic connecting  $a$  and  $b$  to get the tangent vector at point  $A$ . The closed form equation is given by

$$\gamma(t) = XV \cos(\Theta t) + U \sin(\theta t)$$

where  $\gamma(0) = A$ ,  $\gamma(1) = B$  and  $\Theta = \tanh \Sigma$ . The quantities  $U, \Sigma, V$  are gotten after the SVD decomposition of projection of  $B(A^T B)^{-1}$  onto the orthogonal complement of  $X$ , i.e.,

$$U \Sigma V^T = (I - AA^T) B (A^T B)^{-1}$$

The tangent vector that we need is basically  $\dot{\gamma}(0)$ . Let us give it a better name  $T_A$ . So  $T_A$  can be calculated simply using  $U \Theta V^T$ . Infact,  $\Theta$  is the diagonal matrix having all the principal angles between subspace  $A$  and  $B$ .

- We also calculate the geodesic length of the curve connecting the two subspaces. This turns out to be  $\sqrt{\sum_i \theta_i^2}$  where  $\theta_i$  are the principal angles between the two subspaces. Let us call it  $L$ .
- Now that we have  $T_A$ , we use the vector transport equation to parallelly translate it along the geodesic connecting  $A$  and  $C$ . The closed form of the equation is given by

$$T_A(t) = (-AV \sin(\Sigma t) U^T + U \cos(\Sigma t) U^T + (I - UU^T)) T_A$$

where  $T_A(t)$  represents the tangent vector after being transported to a point  $\alpha(t)$  where  $\alpha$  represents the geodesic.  $U, \Sigma, V$  are gotten after SVD decomposition of  $T_A$ , i.e.,  $U \Sigma V^T = T_A$ . We require  $\tau T_A(1)$  and let us call it  $T_C$ .

- Next step is to use this transported vector at subspace  $C$  to move along the geodesic till we have covered a distance of length  $L$ . The closed form of this equation is given by -

$$\beta(t) = XV \cos(\Theta t) + U \sin(\theta t)$$

where  $\beta(0) = C$  and  $U, \Theta, V$  are gotten after SVD decomposition of the parallel transported tangent vector  $T_C$ , i.e.,  $U \Theta V^T = T_C$ .

So, the point we get after traversing the curve  $\beta$  till length  $L$  at a velocity equal to  $T_C$  is our required subspace  $D$ . Now we search for all words that are close to  $D$ .