Approximation by orthogonal transform

John Mount*

November 12, 2014

Abstract

We work through the derivation of the standard solution to the orthogonal Procrustes problem.

1 Introduction

The orthogonal Procrustes problem is: given two matrices A, and B then find an orthogonal matrix W such that $\|AW - B\|_F^2$ is minimized. We will limit or work to the case where A, B are real n by n matrices.

The current family of solutions goes back to Peter Schonemann's 1964 thesis[Wikipedia, 2014, Schonemann, 1966] and we adapt the proof from [Bindel, 2012] for this note.

2 Some Linear Algebra

To work the problem we will need some background definitions and facts from linear algebra, which we will state without proof.

Definition 1. *For a m row by n column matrix A the transpose of A written $tr(A)$ is a n row by m column matrix such that $tr(A)_{j,i} = A_{i,j}$ for $i = 1 \cdots m$, $j = 1 \cdots n$.*

Definition 2. *An orthogonal matrix W is a real matrix with n rows and n columns such that $WW^\top = W^\top W = I$. Note that in addition to having orthogonal rows and columns an orthogonal matrix is also full rank and has all rows and columns unit length.*

Definition 3. *The squared Frobenius norm of a n by n matrix X is written as $\|X\|_F^2$ and is equal to $\sum_{i=1}^n \sum_{j=1}^n |X_{i,j}|^2$.*

Definition 4. *The trace of a n by n matrix X is written as $tr(X)$ and is defined as $\sum_{i=1}^n X_{i,i}$.*

Lemma 1. *For real n by n matrices $\|X\|_F^2 = \left\|X^\top\right\|_F^2$.*

Lemma 2. *For real n by n matrices $\|X\|_F^2 = tr(XX^\top) = tr(X^\top X)$.*

Lemma 3. *For real n by n matrices $A_1, \cdots A_k$ $tr(A_1 \cdots A_k) = tr(A_k A_1 \cdots A_{k-1})$. That is: trace is invariant under cyclic re-ordering of a product (though not under arbitrary permutations in general).*

Lemma 4. *$\|AW\|_F^2 = \|A\|_F^2$ if W is orthogonal.*

*email: mailto:jmount@win-vector.com web: http://www.win-vector.com/

Proof.

$$\begin{aligned}\|AW\|_F^2 &= \mathrm{tr}(AWW^\top A^\top)\\ &= \mathrm{tr}(AA^\top)\\ &= \|A\|_F^2\end{aligned}$$

□

Lemma 5. *If A is an m row by n column real matrix then there exists matrices U, D, V such that $A = UDV^\top$ and:*

- *U is a m by m orthogonal matrix*
- *D is a m by n diagonal matrix*
- *$D_{i,i}$ are non-negative and decreasing in i*
- *V is a n by n orthogonal matrix.*

This factorization is called the singular value decomposition, and is available in most linear algebra libraries.

3 The solution

The problem is to find an orthogonal matrix W minimizing $\|AW - B\|_F^2$ where A, B are n by n real matrices.

Theorem 1. *Let $UDV^\top$ be the singular value decomposition of $A^\top B$ where A, B are real n by n matrices. Then $W = UV^\top$ is an orthogonal matrix minimizing $\|AW - B\|_F^2$.*

Proof. To derived the method we first expand $\|AW - B\|_F^2$.

$$\begin{aligned}\|AW - B\|_F^2 &= \sum_{i,j}(AW - B)_{i,j}^2\\ &= \sum_{i,j}(AW)_{i,j}^2 + (B)_{i,j}^2 - 2(AW)_{i,j}(B)_{i,j}\\ &= \|AW\|_F^2 + \|B\|_F^2 - 2\mathrm{tr}(W^\top A^\top B)\\ &= \|A\|_F^2 + \|B\|_F^2 - 2\mathrm{tr}(W^\top A^\top B)\end{aligned}$$

So picking W to maximize $\mathrm{tr}(W^\top A^\top B)$ will minimize $\|AW - B\|_F^2$.
Let $UDV^\top$ be the singular value decomposition of $A^\top B$.

$$\begin{aligned}\mathrm{tr}(W^\top A^\top B) &= \mathrm{tr}(W^\top UDV^\top)\\ &= \mathrm{tr}(V^\top W^\top UD)\end{aligned}$$

Write $Z = V^\top W^\top U$, notice Z is orthogonal (being the product of orthogonal matrices). The goal is re-stated: maximize $\mathrm{tr}(ZD)$ through our choice of W. Because D is diagonal we have $\mathrm{tr}(ZD) = \sum_{i=1}^n Z_{i,i}D_{i,i}$. The $D_{i,i}$ are non-negative and Z is orthogonal for any choice of W. The maximum is achieved by choosing W such that all of $Z_{i,i} = 1$ which implies $Z = I$. So an optimal W is $UV^\top$. □

4 Application

An application of the orthogonal Procrustes solution to machine learning is given in a iPython notebook [Mount, 2014] (and soon to be followed up by a blog post on http://www.win-vector.com/blog).

References

[Bindel, 2012] Bindel, D. (2012). Intro to scientific computing (cs 3220) notes week 5: Wednesday, feb 29. http://www.cs.cornell.edu/~bindel/class/cs3220-s12/notes/lec12.pdf. Accessed: 2014-11-12.

[Mount, 2014] Mount, J. (2014). xdrift ipython worksheet. http://winvector.github.io/xDrift/xDrift.html. Accessed: 2014-11-12.

[Schonemann, 1966] Schonemann, P. H. (1966). A generalized solution of the orthogonal Procrustes problem. *Psychometrika*, 31(1):1–10.

[Wikipedia, 2014] Wikipedia (2014). Orthogonal procrustes problem. http://en.wikipedia.org/wiki/Orthogonal_Procrustes_problem. Accessed: 2014-11-12.