Approximation by orthogonal transform

John Mount*

November 14, 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 Matrix/Linear Algebra

To work the problem we will need some background definitions and facts from linear algebra (most of which we will state without proof). We will not use all of these facts, as we have added a few extras to remind the reader of the important invariant properties of tr() (as having these ideas in mind makes the later proofs easier to anticipate).

Definition 1. *For a m row by n column matrix A the transpose of A written $A^\top$ is a n row by m column matrix such that $(A^\top)_{j,i} = A_{i,j}$ 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 $WW^\top = W^\top W = I$ (I being the identity matrix which has 1s on the diagonal and zeros elsewhere). 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 trace of a n by n matrix X is written as $tr(X)$ and is defined as $\sum_{i=1}^n X_{i,i}$.*

Definition 4. *The squared Frobenius norm of a n by n matrix X is written as $\|X\|_F^2$ and is equal to $\sum_{i=1}^n \sum_{j=1}^n |X_{i,j}|^2$.*

Lemma 1. *For real n by n matrices $A_1, \cdots A_k$ $(A_1 \cdots A_k)^\top = A_k^\top \cdots A_1^\top$. That is: transpose distributes over products by reversing the order.*

Lemma 2. $tr(A) = tr(A^\top)$

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).*

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

Lemma 4. *$tr(XAX^{-1}) = tr(A)$*

Proof.

$$\begin{aligned} \mathrm{tr}(XAX^{-1}) &= \mathrm{tr}(X^{-1}XA) \quad (1) \\ &= \mathrm{tr}(A) \quad (2) \end{aligned}$$

□

Lemma 5. *For real n by n matrices $\|X\|_F^2 = \|X^\top\|_F^2$.*

Lemma 6. *For real n by n matrices $\|X\|_F^2 = tr(XX^\top) = tr(X^\top X)$.*

Lemma 7. *For A, B orthogonal n by n matrices AB is also an orthogonal matrix.*

Lemma 8. *$\|AW\|_F^2 = \|A\|_F^2$ if W is orthogonal.*

Proof.

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

□

Lemma 9. *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 $\text{tr}(W^\top A^\top B)$ will minimize $\|AW - B\|_F^2$.

Let $UDV^\top$ be the singular value decomposition of $A^\top B$.

$$\begin{aligned}\text{tr}(W^\top A^\top B) &= \text{tr}(W^\top U D V^\top) \\ &= \text{tr}(V^\top W^\top U D)\end{aligned}$$

Write $Z = V^\top W^\top U$, notice Z is orthogonal (being the product of orthogonal matrices). The goal is re-stated: maximize $\text{tr}(ZD)$ through our choice of W. Because D is diagonal we have $\text{tr}(ZD) = \sum_{i=1}^n Z_{i,i} D_{i,i}$. 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 blog article [Mount, 2014a] and accompanying iPython notebook [Mount, 2014b].

5 Relation to the Eckart-Young-Mirsky theorem

The more famous claim for singular value decomposition is that it also solves the problem of finding best low-rank approximations for matrices or linear operators. Interestingly enough this is under the induced 2-norm of matrices, not the Frobenius norm.

Just as a refresher we state the definitions and re-derive the classic theorem.

Definition 5. *The squared 2-norm of a real n-vector x is denoted $\|x\|_2^2$ and is equal to $\sum_{i=1}^n x_i^2$.*

Definition 6. *The squared induced 2-norm of a real n by n matrix X is written as $\|X\|_2^2$ and is equal to $\sup_{x \in \mathbb{R}^n : \|x\|_2^2 = 1} \|Ax\|_2^2$.*

Let A be our real n by n matrix with singular value decomposition $A = UDV^\top$. Remember D is a diagonal matrix with $D_{i,i}$ decreasing in i and non-negative. Let D^k denote the diagonal matrix where $D^k_{i,i} = D_{i,i}$ for $i \le k$ and zero otherwise. Let A^k denote the matrix $UD^kV^\top$.

Theorem 2. *$\|A - B\|_2^2$ where B has rank no more than k is minimized at $B = A^k$.*

Proof. Let v_i denote the ith column of $V^\top$.

With some work we can show $\left\|A - A^k\right\|_2^2 = D^2_{k+1,k+1}$ as the maximum $\left\|(A - A^k)x\right\|_2^2$ is achieved at $x = v_{k+1}$.

Now consider a real n by n matrix B with rank no more than k. If we can show $\|A - B\|_2^2 \ge D^2_{k+1,k+1}$ we will have established A^k achieves the minimum and we are done. Because $\|M\|_2^2$ for matrices is defined as a sup() over images of vectors it is enough to exhibit a vector w such that $\|w\|_2^2 = 1$ and $\|(A - B)w\|_2^2 \ge D^2_{k+1,k+1}$.

Since $\text{rank}(B) \le k$ we have the null-space of B is at least $n - k$. This means the null-space of B must intersect the vector space spanned by $v_1, v_2, \cdots, v_{k+1}$ in more places that just at zero (as $\mathbb{R}^n$ is too small to hold a rank $n - k$ subspace that is orthogonal to a rank $k + 1$ subspace). So there is a vector w such that: $\|w\|_2^2 = 1$, $Bw = 0$ and $w \in \text{span}(v_1, v_2, \cdots v_{k+1})$.

With some work we can show $\|w\|_2^2 = 1$ plus $w \in \text{span}(v_1, v_2, \cdots v_{k+1})$ implies that $\|Aw\|_2^2 \ge D^2_{k+1,k+1}$.

So $\|(A - B)w\|_2^2 = \|Aw\|_2^2 \ge D^2_{k+1,k+1}$ and we are done. □

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, 2014a] Mount, J. (2014a). Can we try to make an adjustment? http://www.win-vector.com/blog/2014/11/can-we-try-to-make-an-adjustment/. Accessed: 2014-11-12.

[Mount, 2014b] Mount, J. (2014b). xdrift ipython worksheet. http://winvector.github.io/xDrift/xDrift.html. Accessed: 2014-11-13.

[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.