Response to the bound problem

Dingyu Wang
August 24, 2020

Summary

Throughout this response, I introduce a lower bound for the upper bound of $\mathrm{tr}(Y^T ZZ^T Y)$, which I call the "intuitive upper bound" $B_I(a, b; r)$. However this may well be the real upper bound. I've proved that B_I is the maximum under some restrictions.

In short, when $n \geq 2r$ and under certain constraints (see Analysis part), for $\mathrm{tr}(X^T YY^T X) = a$ and $\mathrm{tr}(X^T ZZ^T X) = b$ we have the intuitive upper bound

$$B_I(a, b; r) = \left(\sqrt{ab} + \sqrt{(r-a)(r-b)}\right)^2 \frac{1}{r}.$$

In your notation, the intuitive upper bound of $\frac{1}{r}\|Y^T Z\|_F^2$ is

$$B_I(\eta, \gamma) = (\sqrt{(1-\eta)(1-\gamma)} + \sqrt{\eta\gamma})^2.$$

Notice that this gives correct upper bounds for some known points:

1. $\eta = 1 \implies B_I = \gamma$ and vice versa. This is true since Y lies in the orthogonal space of X and can have maximum affinity with Z equal to $1 - (1 - \gamma) = \gamma$.
2. $\eta = 0 \implies B_I = 1 - \gamma$ and vice versa. This is true since Y conforms with X completely and the affinity between Y and Z should be equal to the affinity between X and Z, i.e. $1 - \gamma$.
3. $\eta = \gamma \implies B_I = 1$. This is true, since if $\eta = \gamma$, Y and Z could potentially be identical, thus having max affinity 1.

I've also done a simple simulation at the end of this response. Welcome any feedback.

Problem

Give that $X, Y, Z \in \mathbb{R}^{n \times r}$ three orthonormal bases for low-dimensional subspaces and

$$\mathrm{tr}(X^T YY^T X) = a$$
$$\mathrm{tr}(X^T ZZ^T X) = b,$$

where $a > b$. Now what is the upper bound of $\mathrm{tr}(Y^T ZZ^T Y)$?

Analysis

First we could derive some equivalent formula of $\mathrm{tr}(X^TYY^TX)$. Denote the columns of X are $x^{(i)}, i \in [r]$ and the columns of Y $y^{(i)}, i \in [r]$. Then we have

$$\begin{aligned}
\mathrm{tr}(X^TYY^TX) &= \mathrm{tr}(XX^TYY^T) \\
&= \mathrm{tr}\left((\sum_{i=1}^{r} x^{(i)}x^{(i)^T})(\sum_{i=1}^{r} y^{(i)}y^{(i)^T}) \right) \\
&= \mathrm{tr}\left(\sum_{i=1}^{r}\sum_{j=1}^{r} x^{(i)}x^{(i)^T}y^{(j)}y^{(j)^T} \right) \\
&= \sum_{i=1}^{r}\sum_{j=1}^{r} \mathrm{tr}(x^{(i)}x^{(i)^T}y^{(j)}y^{(j)^T}) \\
&= \sum_{i=1}^{r}\sum_{j=1}^{r} \mathrm{tr}(x^{(i)^T}y^{(j)}y^{(j)^T}x^{(i)}) \\
&= \sum_{i=1}^{r}\sum_{j=1}^{r} \langle x^{(i)}, y^{(j)} \rangle^2
\end{aligned}$$

Notice that, this formula has an intuitive geometric interpretation- the square sum of the projections of one bases to the other. Therefore, let us consider the two components- one in X and the other in $X^\perp$ of each vector, denote $y^{(i)} = u^{(i)} + u_\perp^{(i)}$, and $z^{(i)} = w^{(i)} + w_\perp^{(i)}$, where $u^{(i)}, w^{(i)} \in X$, $u_\perp^{(i)}, w_\perp^{(i)} \in X^\perp$ for all $i \in [r]$. Then now we have

$$\begin{aligned}
\sum_{i=1}^{r} \left\|u^{(i)}\right\|^2 &= a \\
\sum_{i=1}^{r} \left\|u_\perp^{(i)}\right\|^2 &= r - a \\
\sum_{i=1}^{r} \left\|w^{(i)}\right\|^2 &= b \\
\sum_{i=1}^{r} \left\|w_\perp^{(i)}\right\|^2 &= r - b
\end{aligned}$$

Without loss of generality, we could assume X is just the span of the first r standard unit vectors. Further more, we assume that $n \geq 2r$. Then we could set

$$Y = \left(\begin{array}{cccc} \sqrt{\frac{a}{r}} & 0 & 0 & \cdots \\ 0 & \sqrt{\frac{a}{r}} & 0 & \cdots \\ \vdots & \vdots & \ddots & \cdots \\ \hline \sqrt{\frac{r-a}{r}} & 0 & 0 & \cdots \\ 0 & \sqrt{\frac{r-a}{r}} & 0 & \cdots \\ \vdots & \vdots & \ddots & \cdots \end{array}\right), \qquad Z = \left(\begin{array}{cccc} \sqrt{\frac{b}{r}} & 0 & 0 & \cdots \\ 0 & \sqrt{\frac{b}{r}} & 0 & \cdots \\ \vdots & \vdots & \ddots & \cdots \\ \hline \sqrt{\frac{r-b}{r}} & 0 & 0 & \cdots \\ 0 & \sqrt{\frac{r-b}{r}} & 0 & \cdots \\ \vdots & \vdots & \ddots & \cdots \end{array}\right)$$

The corresponding $\mathrm{tr}(Y^T Z Z^T Y)$ is just, (let's call it intuitive bound B_I)

$$B_I(a, b; r) = \left(\sqrt{\frac{ab}{r^2}} + \sqrt{\frac{(r-a)(r-b)}{r^2}}\right)^2 r = \left(\sqrt{ab} + \sqrt{(r-a)(r-b)}\right)^2 \frac{1}{r}$$

Now we prove this is the upper bound if Y, Z are of form

$$Y = \left(\begin{array}{cccc} \cos\alpha_1 & 0 & 0 & \cdots \\ 0 & \cos\alpha_2 & 0 & \cdots \\ \vdots & \vdots & \ddots & \cdots \\ \hline \sin\alpha_1 & 0 & 0 & \cdots \\ 0 & \sin\alpha_2 & 0 & \cdots \\ \vdots & \vdots & \ddots & \cdots \end{array}\right), \qquad Z = \left(\begin{array}{cccc} \cos\beta_1 & 0 & 0 & \cdots \\ 0 & \cos\beta_2 & 0 & \cdots \\ \vdots & \vdots & \ddots & \cdots \\ \hline \sin\beta_1 & 0 & 0 & \cdots \\ 0 & \sin\beta_2 & 0 & \cdots \\ \vdots & \vdots & \ddots & \cdots \end{array}\right),$$

where $\sum_{i=1}^r \cos^2 \alpha_i = a$ and $\sum_{i=1}^r \cos^2 \beta_i = b$ (*). Now $\mathrm{tr}(Y^T Z Z^T Y)$ is equal to

$$\sum_{i=1}^r (\cos\alpha_i \cos\beta_i + \sin\alpha_i \sin\beta_i)^2 = \sum_{i=1}^r \cos^2(\alpha_i - \beta_i).$$

Using the method of Lagrange multipliers, we have

$$f = \sum_{i=1}^r \cos^2(\alpha_i - \beta_i) + \lambda_1 \sum_{i=1}^r \cos^2 \alpha_i + \lambda_2 \sum_{i=1}^r \cos^2 \beta_i$$

Then for all $i \in [r]$

$$\frac{\partial f}{\partial \alpha_i} = \sin(2\alpha_i - 2\beta_i) + \lambda_1 \sin 2\alpha_i = 0$$
$$\frac{\partial f}{\partial \beta_i} = \sin(2\alpha_i - 2\beta_i) - \lambda_2 \sin 2\beta_i = 0$$

We observe that $|\lambda_1| \neq |\lambda_2|$ and $\lambda_1\lambda_2 \neq 0$, otherwise $|\cos 2\alpha_i| = |\cos 2\beta_i|$ for all i contradicting (*). Then we have

$$\lambda_1 \sin 2\alpha_i + \lambda_2 \sin 2\beta_i = 0$$
$$\sin(2\alpha_i - 2\beta_i)(\lambda_2 \cos 2\beta_i + \lambda_1 \cos 2\alpha_i + \lambda_1\lambda_2) = 0.$$

The trivial solution is $\alpha_i = \beta_i = 0$ or $\alpha_i = \beta_i = \frac{\pi}{2}$, and except that, we have the equation set

$$\lambda_1 \sin 2\alpha_i + \lambda_2 \sin 2\beta_i = 0$$
$$\lambda_2 \cos 2\beta_i + \lambda_1 \cos 2\alpha_i + \lambda_1\lambda_2 = 0.$$

From this we could solve out that

$$\cos 2\alpha_i = \frac{\lambda_2^2 - \lambda_1^2 - \lambda_1^2\lambda_2^2}{2\lambda_1^2\lambda_2}, \qquad \cos 2\beta_i = \frac{\lambda_1^2 - \lambda_2^2 - \lambda_1^2\lambda_2^2}{2\lambda_1\lambda_2^2}$$

Therefore, the solution of α_i and β_i should be some zeros, some $\pi/2$ and some identical terms. Then for $\cos\alpha_i$ and $\cos\beta_i$ we should have some ones, some zeros and some identical terms. Assume there are $r - m - x$ ones, x zeros and m identical remaining terms, then we have

$$\mathrm{tr}(Y^T Z Z^T Y) = r - m + \left(\sqrt{\frac{(a-r+m+x)(b-r+m+x)}{m^2}} + \sqrt{\frac{(r-a-x)(r-b-x)}{m^2}}\right)^2 m \qquad (*),$$

where $r - m - x < a, b$. By getting the derivative of this w.r.t m and by some manipulations we observe that the derivative is always positive under constraints given above. Therefore, we achieve maximum with $m = r - x$, that is

$$x + \left(\sqrt{\frac{ab}{(r-x)^2}} + \sqrt{\frac{(r-a-x)(r-b-x)}{(r-x)^2}}\right)^2 (r-x),$$

which we could reduce to the solved formula by replacing $m' = r - x, a' = r - a, b' = r - a$, that is

$$r - m' + \left(\sqrt{\frac{(r-a')(r-b')}{m'^2}} + \sqrt{\frac{(m'-r+a')(m'-r+b')}{m'^2}}\right)^2 m',$$

which is just the expression (*) by setting $x = 0$. So this has a positive derivative w.r.t m' and we have maximum when $m' = r$, i.e.

$$\left(\sqrt{\frac{(r-a)(r-b)}{r^2}} + \sqrt{\frac{ab}{r^2}}\right)^2 r,$$

which is just $B_I(a, b; r)$. This corresponds to $x = 0$ and $m = r$, that is, r identical $\cos 2\alpha_i$ and $\cos 2\beta_i$, which conforms with our construction.

Comment

Well, up to now we have get an intuitive upper bound B_I which is equal to the actual upper bound when Y, Z are of certain form. I do believe that this B_I is the actual upper bound (I haven't thought of any counter examples). It is quite difficult for me to prove that, though.

Anyway, this bound should be kind of indicative in practice. I make a simulation (see Figure 1). As observed from this diagram, even using the intuitive upper bound, *the expected affinity is much lower than the upper bound*, i.e. most of sampled data are far from the intuitive upper bound (and real upper bound).

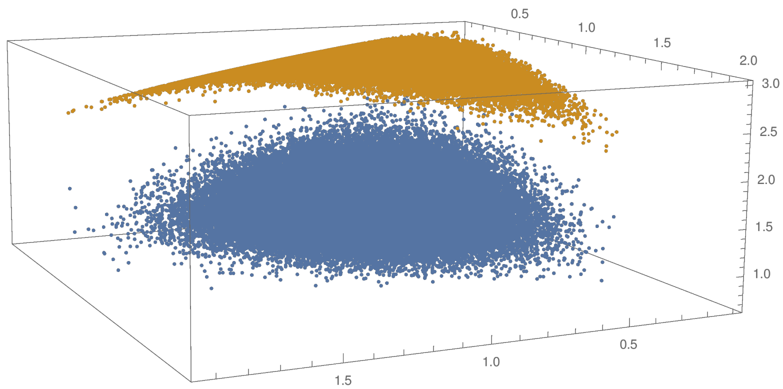

Figure 1: simulation (n=10,r=3), x, y axes are aff(X,Y) and aff(X,Z), blue z is aff(Y,Z), yellow z is hypothesis upper bound, 100,000 random samples

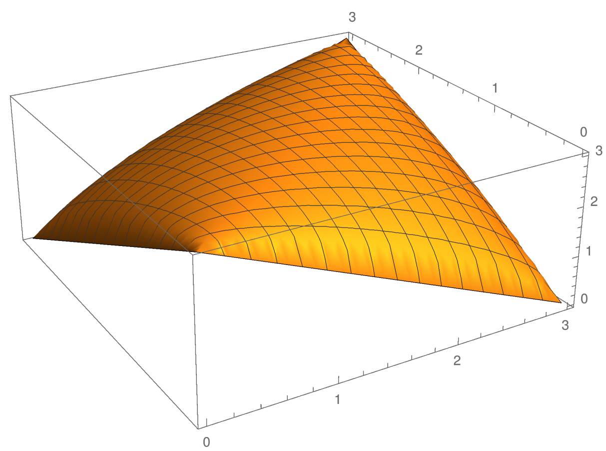

Figure 2: Intuitive upper bound (n=10,r=3)

```
(*code for this tiny simulation, using Mathematica.*)
n=10;
r=3;
iteration=100000;
realdata={};
theodata={};
Do[
a = Orthogonalize[Table[Table[RandomReal[],{j,1,n}],{i,1,r}]];
b = Orthogonalize[Table[Table[RandomReal[],{j,1,n}],{i,1,r}]];
aaff=Norm[a[[Range[r],Range[r]]],"Frobenius"]^2;
baff=Norm[b[[Range[r],Range[r]]],"Frobenius"]^2;
abaff=Norm[b.Transpose[a],"Frobenius"]^2;
theobound=(Sqrt[aaff baff]+Sqrt[(r-aaff)(r-baff)])^2 1/r;
AppendTo[realdata,{aaff,baff,abaff}];
AppendTo[theodata,{aaff,baff,theobound}],iteration];
(*simulation visualization*)
```

```
ListPointPlot3D[{realdata,theodata}]
(*intuitive theoretical upper bound visualization*)
Plot3D[(Sqrt[aaff baff]+Sqrt[(r-aaff)(r-baff)])^2 1/r,{aaff,0,3},{baff,0,3}]
```