A COMPARISON OF THE MOMENTS OF A QUADRATIC FORM INVOLVING ORTHONORMALISED AND NORMALISED RANDOM PROJECTION MATRICES.

ROBERT J. DURRANT AND ATA KABÁN

Abstract. We compare the moments of a quadratic form involving normalised and orthonormalised random projection matrices, and prove that the moments are smaller when the rows of the matrix are orthonormalised than when they are only normalised. Specifically, we show that the expected value of the quadratic form is identical in both cases, but for higher moments strict inequality holds almost surely.

1. Introduction

In our recent papers [3, 4, 2] we asserted without proof that two particular exponentiated quadratic forms were (in each respective case) bounded above by the moment generating function of a chi-squared distribution. This report fills that gap by proving that those quadratic forms satisfy their corresponding inequalities, and that moreover such a bound holds for each moment individually.

2. Results

Theorem 1. *Let R be a random matrix, $R \in \mathcal{M}_{k\times d}$, $k < d$, with entries $r_{ij} \overset{i.i.d}{\sim} \mathcal{N}(0,\sigma^2)$. Let $\mathbf{v}$ be any vector in $\mathbb{R}^d$. Then the moments of $\mathbf{v}^T R^T \left(RR^T\right)^{-1} R\mathbf{v}$ are no greater than the moments of $\mathbf{v}R^T R\mathbf{v}/\sigma^2 d$. Specifically:*

$$E_R\left[\left(\mathbf{v}^T R^T \left(RR^T\right)^{-1} R\mathbf{v}\right)^i\right] \leqslant E_R\left[\left(\frac{1}{\sigma^2 d}\cdot \mathbf{v}^T R^T R\mathbf{v}\right)^i\right],\ \forall i \in \mathbb{N} \text{ and } \mathbf{v} \in \mathbb{R}^d$$

with equality only when $i = 1$.

Corollary 1. *Let $R, \mathbf{v}$ be as given in theorem 1. Then:*

$$E_R\left[\exp\left(\mathbf{v}^T R^T \left(RR^T\right)^{-1} R\mathbf{v}\right)\right] \leqslant E_R\left[\exp\left(\frac{1}{\sigma^2 d}\cdot \mathbf{v}^T R^T R\mathbf{v}\right)\right],\ \forall \mathbf{v} \in \mathbb{R}^d$$

3. Proofs

3.1. **Proof of Theorem 1.** We want to show that:

$$\mathrm{E}_R\left[\left(\mathbf{v}^T R^T \left(RR^T\right)^{-1} R\mathbf{v}\right)^i\right] \leqslant \mathrm{E}_R\left[\left(\frac{1}{\sigma^2 d}\cdot \mathbf{v}^T R^T R\mathbf{v}\right)^i\right],\ \forall i \in \mathbb{N} \text{ and } \mathbf{v} \in \mathbb{R}^d \tag{3.1}$$

Date: 20th January 2011.

where $R \in \mathcal{M}_{k \times d}$ is a $k \times d$ random matrix with entries $r_{ij} \sim \mathcal{N}(0, \sigma^2)$ and so $R/\sigma\sqrt{d}$ is a $k \times d$ random matrix with entries $r_{ij} \sim \mathcal{N}(0, 1/d)$ and normalised rows. The normalisation term $(1/\sigma^2 d)$ on RHS of the inequality (3.1) is required to make the LHS and RHS comparable, since the matrix $\left(RR^T\right)^{-1/2} R$ already has orthonormal rows[1].
Note that by the method of construction of R its rows are almost surely linearly independent, and hence $\text{rank}\,(R) = k$ with probability 1. In the following we may therefore safely assume that $\text{rank}\,(R) = \text{rank}\left(R^T R\right) = \text{rank}\left(RR^T\right) = k$.

The proof now proceeds via eigendecomposition of the matrix $R^T R$.
When $\mathbf{v} = 0$ there is nothing to prove, so let $\mathbf{v}$ be a non-zero vector in $\mathbb{R}^d$ and let $\mathbf{x}_j$ be a unit eigenvector of $R^T R$ with $\lambda(\mathbf{x}_j)$ its corresponding eigenvalue. Since $R^T R \in \mathcal{M}_{d \times d}$ there exists an orthonormal eigenvector basis for $\mathbb{R}^d$, $\mathcal{B}_R = \{\mathbf{x}_1, \ldots, \mathbf{x}_d\}$. Furthermore, since $\text{rank}\left(R^T R\right) = k < d$ we know that k of the eigenvalues $\lambda(\mathbf{x}_j)$ are non-zero and the remaining $d-k$ of the $\lambda(\mathbf{x}_j)$ are zero. Writing $\mathbf{v} = \sum_{j=1}^{d} \alpha_j \mathbf{x}_j$ we then have:

$$
\begin{aligned}
\frac{1}{\sigma^2 d} \cdot \mathbf{v}^T R^T R \mathbf{v} \quad &= \frac{1}{\sigma^2 d} \cdot \sum_{j=1}^{d} \lambda(\mathbf{x}_j) \alpha_j \mathbf{x}_j^T \mathbf{x}_j \beta_j = \frac{1}{\sigma^2 d} \cdot \sum_{j=1}^{d} \lambda(\mathbf{x}_j) \alpha_j^2 \|\mathbf{x}_j\|^2 \\
&= \frac{1}{\sigma^2 d} \left(\sum_{\{j : \lambda(\mathbf{x}_j) \neq 0\}} \lambda(\mathbf{x}_j) \alpha_j^2 \|\mathbf{x}_j\|^2 + \sum_{\{j : \lambda(\mathbf{x}_j) = 0\}} 0 \cdot \alpha_j^2 \|\mathbf{x}_j\|^2 \right) \\
&= \frac{1}{\sigma^2 d} \cdot \sum_{\{j : \lambda(\mathbf{x}_j) \neq 0\}} \lambda(\mathbf{x}_j) \alpha_j^2 \|\mathbf{x}_j\|^2 \\
&= \frac{1}{\sigma^2 d} \cdot \sum_{\{j : \lambda(\mathbf{x}_j) \neq 0\}} \lambda(\mathbf{x}_j) \alpha_j^2 \qquad (3.2)
\end{aligned}
$$

Next, note that if $\mathbf{x}_j$ is an eigenvector of $R^T R$ with non-zero eigenvalue $\lambda(\mathbf{x}_j)$, then $R\mathbf{x}_j$ is an eigenvector of RR^T with the same non-zero eigenvalue, since:

$$
R^T R \mathbf{x}_j = \lambda(\mathbf{x}_j) \mathbf{x}_j \Longrightarrow RR^T R \mathbf{x}_j = \lambda(\mathbf{x}_j) R \mathbf{x}_j
$$

There are k such non-zero eigenvalues, and as $\text{rank}(RR^T) = k$ the non-zero eigenvalues of $R^T R$ are the eigenvalues of RR^T. Furthermore, since $RR^T \in \mathcal{M}_{k \times k}$, RR^T is invertible. It now follows that if $\mathbf{x}_j$ is an eigenvector of $R^T R$ with non-zero eigenvector $\lambda(\mathbf{x}_j)$, then $R\mathbf{x}_j$ is an eigenvector of $\left(RR^T\right)^{-1}$ with non-zero eigenvalue $1/\lambda(\mathbf{x}_j)$. Hence:

$$
\begin{aligned}
\mathbf{v}^T R^T \left(RR^T\right)^{-1} R \mathbf{v} \quad &= \sum_{j=1}^{k} \frac{1}{\lambda(\mathbf{x}_j)} \cdot \lambda(\mathbf{x}_j) \alpha_j^2 \|\mathbf{x}_j\|^2 \\
&= \sum_{\{j : \lambda(\mathbf{x}_j) \neq 0\}} \alpha_j^2 \qquad (3.3)
\end{aligned}
$$

---

[1]Since $\left(RR^T\right)^{-1/2} RR^T \left(\left(RR^T\right)^{-1/2} \right)^T = \left(RR^T\right)^{-1/2} \left(RR^T\right)^{1/2} \left(RR^T\right)^{1/2} \left(RR^T\right)^{-1/2} = I$

We can now rewrite the inequality (3.1) to be proved as the following equivalent problem. For all $i \in \mathbb{N}$:

$$\begin{aligned}
&\mathrm{E}_R\left[\left(\sum_{\{j:\lambda(\mathbf{x}_j)\neq 0\}} \alpha_j^2\right)^i\right] \leqslant \mathrm{E}_R\left[\left(\frac{1}{\sigma^2 d}\sum_{\{j:\lambda(\mathbf{x}_j)\neq 0\}} \lambda(\mathbf{x}_j)\alpha_j^2\right)^i\right] \\
\iff &\mathrm{E}_{\lambda,\alpha}\left[\left(\sum_{\{j:\lambda(\mathbf{x}_j)\neq 0\}} \alpha_j^2\right)^i\right] \leqslant \mathrm{E}_{\lambda,\alpha}\left[\left(\frac{1}{\sigma^2 d}\sum_{\{j:\lambda(\mathbf{x}_j)\neq 0\}} \lambda(\mathbf{x}_j)\alpha_j^2\right)^i\right] \\
\iff &\mathrm{E}_{\lambda,\alpha}\left[\left(\sum_{\{j:\lambda(\mathbf{x}_j)\neq 0\}} \alpha_j^2\right)^i\right] \leqslant \mathrm{E}_{\alpha}\left[\mathrm{E}_{\lambda|\alpha}\left[\left(\sum_{\{j:\lambda(\mathbf{x}_j)\neq 0\}} \frac{1}{\sigma^2 d}\lambda(\mathbf{x}_j)\alpha_j^2\right)^i\right]\right]
\end{aligned} \tag{3.4}$$

Now, in RHS of (3.4) the α_j depend only on $\mathbf{v}$ and the $\mathbf{x}_j$ and these are both independent of $\lambda(\mathbf{x}_j)$ (e.g. [1] Proposition 4.18, [5] Lemma 2.6) and so α is independent of λ. Hence we can rewrite this term as:

$$\mathrm{E}_{\alpha}\left[\mathrm{E}_{\lambda}\left[\left(\sum_{\{j:\lambda(\mathbf{x}_j)\neq 0\}} \frac{1}{\sigma^2 d}\lambda(\mathbf{x}_j)\alpha_j^2\right)^i\right]\right] \tag{3.5}$$

and since (3.5) is the expectation of a convex function, applying Jensen's inequality to the inner term we see that:

$$\mathrm{E}_{\alpha}\left[\left(\mathrm{E}_{\lambda}\left[\sum_{\{j:\lambda(\mathbf{x}_j)\neq 0\}} \frac{1}{\sigma^2 d}\lambda(\mathbf{x}_j)\alpha_j^2\right]\right)^i\right] \leqslant \mathrm{E}_{\alpha}\left[\mathrm{E}_{\lambda}\left[\left(\sum_{\{j:\lambda(\mathbf{x}_j)\neq 0\}} \frac{1}{\sigma^2 d}\lambda(\mathbf{x}_j)\alpha_j^2\right)^i\right]\right] \tag{3.6}$$

Note that when $i = 1$ we have equality in (3.6) and strict inequality when $i > 1$. If we can show that the LHS of (3.6) above is no less than the LHS of (3.1) then we are done. Now, equation (3.3) implies that all terms in LHS of (3.1) are positive, and so in order to prove the theorem it is sufficient to show that $\mathrm{E}_R\left[\lambda(\mathbf{x}_j)\right]/\sigma^2 d \geqslant 1$. But this is certainly so since:

$$\mathrm{E}_R\left[\lambda(\mathbf{x}_j)\right] = \frac{1}{k}\sum_{\{j:\lambda(\mathbf{x}_j)\neq 0\}} \mathrm{E}_R\left[\lambda(\mathbf{x}_j)\right] = \frac{1}{k}\mathrm{E}_R\left[\mathrm{Tr}\left(RR^T\right)\right] = \frac{1}{k}\sum_{j=1}^{k}\mathrm{E}_R\left[\mathbf{r}_j^T\mathbf{r}_j\right] \tag{3.7}$$

where $\mathbf{r}_j$ is the j-th row of R. Then, since the $\mathbf{r}_j \overset{\text{i.i.d}}{\sim} \mathcal{N}\left(0, \mathrm{diag}(\sigma^2)\right)$ we have $\mathbf{r}_j^T\mathbf{r}_j/\sigma^2 \overset{\text{i.i.d}}{\sim} \chi_d^2$ and so $\mathrm{E}_R\left[\mathbf{r}_j^T\mathbf{r}_j/\sigma^2\right] = d$. Finally, it then follows that $\frac{1}{k}\sum_{j=1}^{k}\mathrm{E}_R\left[\lambda(\mathbf{x}_j)\right]/\sigma^2 d = 1$ and this completes the proof.

3.2. **Proof of Corollary 1.** To prove the corollary we rewrite the inequality (3.1) using the Taylor series expansion for exp to see that:

$$
\begin{aligned}
\mathrm{E}_R\left[\sum_{i=0}^{\infty}\frac{\left(\mathbf{v}^T R^T \left(RR^T\right)^{-1} R\mathbf{v}\right)^i}{i!}\right] &\leqslant \mathrm{E}_R\left[\sum_{i=0}^{\infty}\frac{\left(\frac{1}{\sigma^2 d}\cdot\mathbf{v}^T R^T R\mathbf{v}\right)^i}{i!}\right] \quad (3.8)\\
\Longrightarrow \quad \mathrm{E}_R\left[\exp\left(\mathbf{v}^T R^T \left(RR^T\right)^{-1} R\mathbf{v}\right)\right] &\leqslant \mathrm{E}_R\left[\exp\left(\frac{1}{\sigma^2 d}\cdot\mathbf{v}^T R^T R\mathbf{v}\right)\right]
\end{aligned}
$$

Since by theorem 1 we have the required inequality for each of the i-th powers in the summations in equation (3.8), the result follows immediately.

References

[1] M. Artin. *Algebra*. Pearson Education, 2010.
[2] R. J. Durrant and A. Kaban. A tight bound on the performance of Fisher's linear discriminant in randomly projected data spaces. *Pattern Recognition Letters (Submitted)*, 2011.
[3] R.J. Durrant and A. Kabán. Compressed Fisher Linear Discriminant Analysis: Classification of Randomly Projected Data. In *Proceedings16th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2010)*, 2010.
[4] R.J.Durrant and A. Kabán. A bound on the performance of LDA in randomly projected data spaces. In *Proceedings 20th International Conference on Pattern Recognition (ICPR 2010)*, pages 4044–4047, 2010.
[5] A. Tulino and S. Verdú. *Random matrix theory and wireless communications.* Now Publishers Inc, 2004.