theorem 3 multidimensional case proof writeup

Brandon Dong

July 2023

1 Introduction

1.1 Gaussian Mixture Definition

We say that a random vector X is a (centered) Gaussian vector mixture if there exists a positive random matrix Y and a standard Gaussian random vector Z independent of Y such that X and YZ have the same distribution.

1.2 Theorem 3 in multiple dimensions

Let X be a Gaussian mixture in $\mathbb{R}^d$ and $X_1 \dots X_n$ be independent copies of X. For two vectors $a = (a_1 \dots a_n)$ and $b = (b_1 \dots b_n)$, $a, b \in \mathbb{R}^n$, and $p \geq 2$, we have

$$(a_1^2 \dots a_n^2) \preceq (b_1^2, \dots b_n^2) \Longrightarrow \mathbb{E}\left|\sum_{i=1}^{n} a_i X_i\right|^p \leq \mathbb{E}\left|\sum_{i=1}^{n} b_i X_i\right|^p$$

where for now $|\cdot|$ is the Euclidean norm, and when $p \in (-1, 2)$ the second inequality is reversed (unsure) provided that the expectation is finite

Proof:

First, we fix $p > 2$ and let X be a Gaussian mixture with identical independent copies $X_1 \dots X_n$. Then for each X_i by definition we have $X_i = A_i G_i$ for A_i random matrices and G_i standard d dimensional Gaussian vectors, all independent. Then we know that

$$\mathbb{E}\left|\sum_{i=1}^{n} a_i X_i\right|^p = \mathbb{E}\left|\sum_{i=1}^{n} a_i A_i G_i\right|^p.$$

We introduce the notation X_A to denote the centered Gaussian vector in $\mathbb{R}^d$ with covariance matrix A, i.e $X_A \sim N(0, A)$. Then by the properties of linear transformations and sums of Gaussian vectors,

$$\sum_{i=1}^{n} a_i A_i G_i \sim N\left(0, \sum_{i=1}^{n} a_i^2 A_i A_i^T\right)$$

since each individual element of the sum is a Gaussian vector with mean 0 and covariance matrix $a_i^2 A_i A_i^T$, and we can rewrite the above expression as

$$\mathbb{E}\left|\sum_{i=1}^{n} a_i A_i G_i\right|^p = \mathbb{E}\left|\sqrt{\sum_{i=1}^{n} a_i^2 A_i A_i^T} G_1\right|^p$$

where the square root A of a matrix C is the matrix A such that $C = AA^T$.

Now consider the functional $f(A) = \mathbb{E}|\sqrt{A}G|^p$, where A is a covariance matrix (and therefore symmetric, but not necessarily full rank). We then decompose A into UDU^T with $U, D \in \mathbb{R}^{d\times d}$ where U is orthogonal and D is a diagonal matrix with the eigenvalues of A along its diagonal and potentially zeroes along the diagonal afterwards, so that

$$D = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 & \dots & 0 \\ & & \vdots & & & \\ 0 & \dots & 0 & \lambda_d & \dots & 0 \\ & & \vdots & & \vdots & \\ 0 & \dots & 0 & 0 & 0 & 0 \end{bmatrix}$$

and taking the square root along the diagonal we can set

$$\sqrt{D} = \begin{bmatrix} \sqrt{\lambda_1} & 0 & \dots & 0 & \\ 0 & \sqrt{\lambda_2} & \dots & 0 & \\ & & \vdots & & \\ 0 & \dots & 0 & \sqrt{\lambda_d} & \\ & & \vdots & & \\ 0 & \dots & 0 & 0 & \end{bmatrix}$$

so that $\sqrt{A} = U\sqrt{D}U^T$ and

$$f(A) = \mathbb{E}|\sqrt{A}G|^p = \mathbb{E}|U\sqrt{D}U^T G|^p = \mathbb{E}\left|\sqrt{\sum_{i=1}^{d} \lambda_j g_j^2}\right|^p$$

where the last step is because G is rotationally invariant, and the Euclidean norm is invariant with respect to multiplication by an orthogonal matrix. Here we write g_j to indicate the Gaussian random variable that is the jth component of G.

To prove the theorem form here it would be sufficient to show that f is matrix convex for symmetric matrices, i.e. $f(\frac{A+B}{2}) \leq \frac{1}{2}f(A)+\frac{1}{2}f(B)$ with A and B symmetric. To see why, note that if we define the function $\Psi(a) = \mathbb{E}\left|\sum_{j=1}^n \sqrt{a_i}X_j\right|^p$ with the X_j defined as above, and follow the logic above, we see that $\Psi(a) = \mathbb{E}_{A,G}\left|\sqrt{\sum_{j=1}^n a_j A_j A_j^T} G_1\right|^p$, where the subscript on the expected value indicates that it's the expected value over both A and G. Then using the above functional and applying Fubini's/Tonelli's, this becomes $\Psi(a) = \mathbb{E}_A[f(\sum_{j=1}^n a_j A_j A_j^T)]$.

Now, if we want the Schur convexity of Ψ, we have $a \preceq b$ and want $\Psi(a) \leq \Psi(b)$. By an equivalent condition for Schur majorization we know that $a = \sum_{i=1}^n \lambda_\sigma b_\sigma$ for some permutation σ of $\{1 \dots n\}$ and $\lambda_\sigma \geq 0$, $\sum \lambda_\sigma = 1$.

Then we know that

$$\begin{aligned}
\Psi(a) &= \mathbb{E}_A\left[f\left(\sum_{j=1}^n a_j A_j A_j^T\right)\right] \\
&= \mathbb{E}_A\left[f\left(\sum_{j=1}^n \left(\sum_\sigma \lambda_\sigma b_{\sigma(j)}\right) A_j A_j^T\right)\right] \\
&= \mathbb{E}_A\left[f\left(\sum_\sigma \lambda_\sigma \left(\sum_j b_{\sigma(j)} A_j A_j^T\right)\right)\right] \\
&\leq \mathbb{E}_A\left(\sum_\sigma \lambda_\sigma f\left(\sum_j b_{\sigma(j)} A_j A_j^T\right)\right) && \text{(assuming convexity of } f\text{)} \\
&= \sum_\sigma \lambda_\sigma \mathbb{E}_A f\left(\sum_j b_{\sigma(j)} A_j A_j^T\right) \\
&= \sum_\sigma \lambda_\sigma \mathbb{E}_A f\left(\sum_j b_{\sigma(j)} A_{\sigma(j)} A_{\sigma(j)}^T\right) && (A_j \text{ iid}) \\
&= \mathbb{E}_A f\left(\sum_j b_{\sigma(j)} A_{\sigma(j)} A_{\sigma(j)}^T\right) \\
&= \Psi(b)
\end{aligned}$$

with the last step because $\sum_\sigma \lambda_\sigma = 1$, so all we care about is the convexity of f.

To do show its convexity, then, we can note that $f(A) = g(\lambda(A)) = \mathbb{E}\sqrt{\sum \lambda_i g_i^2}^p$ where $\lambda(A)$ is the vector of eigenvalues of A. However, then using Exercise II.1.14 from Bhatia's Matrix Analysis and that A and B are real symmetric we get that $\lambda(A+B) \prec \lambda(A) + \lambda(B)$, so we can use the Marshall and Proschan result to get that g is Schur convex and therefore f is matrix convex, since we have that

$$\begin{aligned}
f\left(\frac{A+B}{2}\right) &= g\left(\frac{1}{2}\lambda(A+B)\right) \\
&= \mathbb{E}\left|\sqrt{\sum_{i=1}^{d} \frac{1}{2}\lambda_j(A+B) g_j^2}\right|^p \\
&\leq \mathbb{E}\left|\sqrt{\sum_{i=1}^{d} \frac{1}{2}\lambda_j(A) + \frac{1}{2}\lambda_j(B) g_j^2}\right|^p && \text{(M\&P)} \\
&\leq \mathbb{E}\left|\sqrt{\sum_{i=1}^{d} \frac{1}{2}\lambda_j(A) g_j^2}\right|^p + \mathbb{E}\left|\sqrt{\sum_{i=1}^{d} \frac{1}{2}\lambda_j(B) g_j^2}\right|^p && \text{(convexity of } x \mapsto x^{p/2} \text{ for } p > 2) \\
&\leq \frac{1}{2}\mathbb{E}\left|\sqrt{\sum_{i=1}^{d} \lambda_j(A) g_j^2}\right|^p + \frac{1}{2}\mathbb{E}\left|\sqrt{\sum_{i=1}^{d} \lambda_j(B) g_j^2}\right|^p = \frac{f(A)+f(B)}{2}
\end{aligned}$$

and that completes the proof. For $p \in (-1, 2), p \neq 0$, each inequality above flips, giving concavity instead.