FLIP PROBABILITIES FOR RANDOM PROJECTIONS OF θ-SEPARATED VECTORS

ROBERT J. DURRANT AND ATA KABÁN

Abstract. We give the probability that two vectors in d-dimensional Euclidean space $\mathbf{m}, \mathbf{n} \in \mathbb{R}^d$ which are separated in $\mathbb{R}^d$ by an angle $\theta \in [0, \pi/2]$ have angular separation $\theta_R > \pi/2$ following random projection into a k-dimensional subspace of $\mathbb{R}^d$, $k < d$. This probability, which we call the 'flip probability', has several interesting properties: It is polynomial of order k in θ; it is independent of the original dimensionality d depending only on the projection dimension k and the original separation θ of the vectors; it recovers the existing result for flip probability under random projection when $k = 1$ as a special case, and has a geometric interpretation as the quotient of the surface area of a hyperspherical cap by the area of the corresponding hypersphere which is a natural generalisation of the $k = 1$ case.
We also prove the useful fact that, for all $k \in \mathbb{N}$, the flip probability when projecting to dimension k is greater than the flip probability when projecting to dimension $k + 1$.

1. Statement of Theorem and Proof

Theorem 1 (Flip Probability)**.** *Let* $\mathbf{n}$, $\mathbf{m} \in \mathbb{R}^d$ *with angular separation* $\theta \in [0, \pi/2]$.
Let $R \in \mathcal{M}_{k \times d}$, $k < d$, *be a random projection matrix with entries* $r_{ij} \overset{iid}{\sim} \mathcal{N}(0, 1/d)$ *and let* $R\mathbf{n}$, $R\mathbf{m} \in \mathbb{R}^k$ *be the projections of* $\mathbf{n}$, $\mathbf{m}$ *into* $\mathbb{R}^k$ *with angular separation* θ_R.

(1) *The 'flip probability'* $Pr_R[\theta_R > \pi/2|\theta] = Pr_R[(R\mathbf{n})^T R\mathbf{m} < 0|\mathbf{n}^T\mathbf{m} \geqslant 0]$ *is given by:*

$$Pr_R[(R\mathbf{n})^T R\mathbf{m} < 0|\mathbf{n}^T\mathbf{m} \geqslant 0] = \frac{\Gamma(k)}{(\Gamma(k/2))^2} \int_0^{\psi} \frac{z^{(k-2)/2}}{(1+z)^k} dz \qquad (1.1)$$

where $\psi = (1 - \cos(\theta))/(1 + \cos(\theta))$.

(2) *The expression above can be shown to be of the form of the quotient of the surface area of a hyperspherical cap subtending an*

1991 *Mathematics Subject Classification.* Primary 60D05; Secondary 15B52.
Key words and phrases. sign random projections, dimensionality reduction, random triangles.

angle of 2θ by the surface area of the corresponding hypersphere:

$$Pr_R[\theta_R > \pi/2|\theta] = \frac{\int_0^\theta \sin^{k-1}(\phi)\ d\phi}{\int_0^\pi \sin^{k-1}(\phi)\ d\phi} \tag{1.2}$$

This form recovers Lemma 3.2 of [7] *where the flip probability θ/π for $k = 1$ was given, and extends it for $k \geqslant 1$ showing that the flip probability is polynomial of order k in θ.*

(3) *The flip probability is inversely related to k. Fix $\theta \in [0, \pi/2]$ and define the sequence:*

$$f_k(\theta) = \left\{ \frac{\int_0^\theta \sin^{k-1}(\phi)\ d\phi}{\int_0^\pi \sin^{k-1}(\phi)\ d\phi} \right\} \tag{1.3}$$

Then $f_k(\theta) \geqslant f_{k+1}(\theta)$.

1.1. **Outline of proof and preliminaries.** We start with two vectors $\mathbf{n}, \mathbf{m} \in \mathbb{R}^d$ having angular separation $\theta \in [0, \pi/2]$ and linearly transform them by premultiplying with a random matrix R with entries drawn i.i.d from the Gaussian $\mathcal{N}(0, 1/d)$. Such a random transformation is often referred to as a random projection. As a consequence of the Johnson-Lindenstrauss lemma, the angular separation of the projected vectors $R\mathbf{n}, R\mathbf{m}$ is approximately θ with high probability (see e.g. [2]), so their orientations under random projection are *not* independent. We want to find the probability that following random projection the angle of these vectors $\theta_R > \pi/2$, i.e. switches from being acute to being obtuse. This problem ties in with both sign random projections [12] for dimensionality reduction, and the geometric probability problem regarding whether a Gaussian random triangle is obtuse [6].

The proof proceeds by carrying out a whitening transform on each coordinate of the pair of projected vectors, and will make use of techniques inspired from the study of random triangles in [6] to derive the flip probability. We obtain the exact expression for the flip probability in the form of an integral that has no analytic closed form in general. However it turns out to have a natural geometrical interpretation as the quotient of the surface area of a (hyper-)spherical cap by the surface area of the corresponding (hyper-)sphere.

Before commencing the proof proper we make some preliminary observations. First, recall from the definition of the dot product, $\mathbf{n}^T\mathbf{m} = |\mathbf{n}||\mathbf{m}| \cos\theta$, we have $\mathbf{x}^T\mathbf{y} < 0 \Leftrightarrow \cos\theta < 0$ and so the dot product is positive if and only if the angular separation of the vectors $\mathbf{n}$ and $\mathbf{m}$ is $\theta \in [0, \pi/2]$.
Hence, for θ in the original d-dimensional space and θ_R in the k-dimensional randomly projected space we have $\Pr_R[\theta_R > \pi/2|\theta \in [0, \pi/2]] = \Pr_R[(R\mathbf{n})^T R\mathbf{m} < 0|\mathbf{n}^T\mathbf{m} \geqslant 0]$, and this is the probability of our interest. For brevity, we will write $\Pr_R[(R\mathbf{n})^T R\mathbf{m} < 0]$ for this

probability. In fact, as we shall see, the arguments for the proof of the first two parts of our theorem do not rely on the condition $\theta \in [0, \pi/2]$.

Regarding random Gaussian matrices we should note that, for any non-zero vector $\mathbf{x} \in \mathbb{R}^d$, the event: $R\mathbf{x} = \mathbf{0}$ has probability zero with respect to the random choices of R. This is because the null space of R, $\ker(R) = R(\mathbb{R}^d)^\perp$, is a linear subspace of $\mathbb{R}^d$ with dimension $d-k < d$, and therefore $\ker(R)$ has zero Gaussian measure in $\mathbb{R}^d$. Hence $\Pr_R\{\mathbf{x} \in \ker(R)\} = \Pr_R\{R\mathbf{x} = \mathbf{0}\} = 0$. In a similar way, R almost surely has rank k. Denote the i-th row of R by $(r_{i1}, \ldots, r_{id})$, then the event: $\text{span}\{(r_{i1}, \ldots, r_{id})\} = \text{span}\{(r_{i'1}, \ldots, r_{i'd})\}$, $i \neq i'$ has probability zero since $\text{span}\{(r_{i1}, \ldots, r_{id})\}$ is a 1-dimensional linear subspace of $\mathbb{R}^d$ with measure zero. By induction, for finite $k < d$, the probability that the j-th row is in the span of the first $j-1$ rows is likewise zero. In this setting we may therefore safely assume that $\mathbf{n}, \mathbf{m} \notin \ker(R)$ and that R has rank k.

With these preliminaries out of the way, we begin our proof.

1.2. **Proof of Theorem 1.**

Proof of part 1. First we expand out the terms of $(R\mathbf{n})^T R\mathbf{m}$:

$$\Pr_R[(R\mathbf{n})^T R\mathbf{m} < 0] = \Pr_R\left[\sum_{i=1}^{k}\left(\sum_{j=1}^{d} r_{ij}m_j\right)\left(\sum_{j=1}^{d} r_{ij}n_j\right) < 0\right] \quad (1.4)$$

Note that the entries of R are statistically independent and their distribution is $\mathcal{D}(r_{ij},\ i \in \{1, \ldots, k\},\ j \in \{1, \ldots, d\}) = \prod_{i,j} \mathcal{D}(r_{ij}) = \prod_{i,j} \mathcal{N}_{r_{ij}}(0, 1/d)$.

We make the change of variables $u_i = \sum_{j=1}^d r_{ij}m_j$ and $v_i = \sum_{j=1}^d r_{ij}n_j$. The linear combination of Gaussian variables is again Gaussian, however u_i and v_i are no longer independent. In turn, the bivariate vectors (u_i, v_i) and (u_j, v_j) are independent of each other for $i \neq j$ since the bivariate distribution of (u_i, v_i) depends only on the i-th row of R which is independent of the j-th row of R. So the joint distribution of our new variables will have the form $\mathcal{D}((u_i, v_i)^T,\ i \in \{1, \ldots, k\}) = \prod_i \mathcal{D}((u_i, v_i)^T)$ with

$$\mathcal{D}((u_i, v_i)^T) = \mathcal{N}_{(u_i,v_i)^T}\left(\mathrm{E}_R\left[\begin{pmatrix} u_i \\ v_i \end{pmatrix}\right], \mathrm{Cov}_R\left[\begin{pmatrix} u_i \\ v_i \end{pmatrix}\right]\right) \quad (1.5)$$

Since all the Gaussians in $\mathcal{D}(u_i, v_i)$ have zero mean, the expectation of this distribution (1.5) is just $(0, 0)^T$, and straightforward calculations (deferred to the Appendix) yield that the covariance matrix of each $\mathcal{D}(u_i, v_i)(i \in \{1, \cdots, k\})$ is:

$$\Sigma_{uv} = \frac{1}{d}\begin{bmatrix} \|\mathbf{n}\|^2 & \mathbf{n}^T\mathbf{m} \\ \mathbf{n}^T\mathbf{m} & \|\mathbf{m}\|^2 \end{bmatrix}$$

and so, $\mathcal{D}(u_i, v_i) \equiv \mathcal{N}(0, \Sigma_{uv})$.

Now we can rewrite the probability in (1.4) as:

$$\Pr_{u_i,v_i \overset{iid}{\sim} \mathcal{N}(0,\Sigma_{uv})} \left\{ \sum_{i=1}^{k} u_i v_i < 0 \right\}$$

Next, it will be useful to rewrite the product $u_i v_i$ in the following form:

$$(u_i, v_i) \begin{bmatrix} 0 & 1/2 \\ 1/2 & 0 \end{bmatrix} \begin{pmatrix} u_i \\ v_i \end{pmatrix}$$

So we want:

$$\Pr_{\mathcal{D}^k(u_i,v_i)} \left\{ \sum_{i=1}^{k} (u_i, v_i) \begin{bmatrix} 0 & 1/2 \\ 1/2 & 0 \end{bmatrix} \begin{pmatrix} u_i \\ v_i \end{pmatrix} < 0 \right\} \tag{1.6}$$

We will make a further change of variables and write:

$$(x_i, y_i) = (u_i, v_i)\Sigma_{uv}^{-1/2} \tag{1.7}$$

where $\Sigma_{uv}^{-1/2} = (\Sigma^{-1})^{1/2} = (\Sigma^{1/2})^{-1} = (\Sigma^{-1/2})^T$ is the unique symmetric positive semi-definite square root of Σ_{uv}^{-1} ([9], Theorem 7.2.6, pg 406).
The new variables x_i, y_i are independent unit variance spherical Gaussian variables, $(x_i, y_i) \overset{iid}{\sim} \mathcal{N}(0, I)$. Solving (1.7) for (u_i, v_i) we get: $(u_i, v_i) = (x_i, y_i)\Sigma_{uv}^{1/2}$, then substituting back into (1.6) we now see that the probability we want to find is:

$$\Pr_{(x_i,y_i) \overset{iid}{\sim} \mathcal{N}(0,I)} \left\{ \frac{1}{2} \sum_{i=1}^{k} (x_i, y_i)\, \Sigma_{uv}^{1/2} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \Sigma_{uv}^{1/2} \begin{pmatrix} x_i \\ y_i \end{pmatrix} < 0 \right\} \tag{1.8}$$

Taking

$$A = \Sigma_{uv}^{1/2} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \Sigma_{uv}^{1/2} \tag{1.9}$$

and carrying out eigendecomposition of A so that $A = QDQ^T$ where Q is an orthogonal matrix and D is a diagonal matrix of the eigenvalues of A, we now have for (1.8):

$$\Pr_{(x_i,y_i) \overset{iid}{\sim} \mathcal{N}(0,I)} \left\{ \frac{1}{2} \sum_{i=1}^{k} (x_i, y_i) QDQ^T \begin{pmatrix} x_i \\ y_i \end{pmatrix} < 0 \right\} \tag{1.10}$$

The Gaussian distribution is invariant under orthogonal transformations, so the form of Q does not affect this probability and without loss of generality we can take it to be the identity matrix. Hence we can rewrite this probability as:

$$\Pr_{(x_i,y_i) \overset{iid}{\sim} \mathcal{N}(0,I)} \left\{ \frac{1}{2} \sum_{i=1}^{k} (x_i, y_i) D \begin{pmatrix} x_i \\ y_i \end{pmatrix} < 0 \right\}$$

Since D is a diagonal matrix of the eigenvalues of A, using the identity $eigenvalues(AB) = eigenvalues(BA)$ ([13], Thm.A.6.2, pg 468) we see from (1.9) that D consists of the eigenvalues of:

$$\Sigma_{uv}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \frac{1}{d}\begin{bmatrix} \|\mathbf{n}\|^2 & \mathbf{n}^T\mathbf{m} \\ \mathbf{n}^T\mathbf{m} & \|\mathbf{m}\|^2 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \frac{1}{d}\begin{bmatrix} \mathbf{n}^T\mathbf{m} & \|\mathbf{n}\|^2 \\ \|\mathbf{m}\|^2 & \mathbf{n}^T\mathbf{m} \end{bmatrix}$$

Which has eigenvalues $\lambda = \frac{1}{d}(\mathbf{n}^T\mathbf{m} \pm \|\mathbf{n}\|\|\mathbf{m}\|) = \frac{1}{d}\|\mathbf{n}\|\|\mathbf{m}\|(\cos(\theta) \pm 1)$. In the inequality (1.10) we may drop the positive scaling constant $\frac{1}{2}\frac{1}{d}\|\mathbf{n}\|\|\mathbf{m}\|$ since it does not affect the sign of the left hand side, and so the probability to find can now be written as:

$$\begin{aligned}
&\Pr_{\mathcal{N}(0,\,I)}\left\{\sum_{i=1}^{k}(x_i, y_i)^T\begin{bmatrix} \cos(\theta)+1 & 0 \\ 0 & \cos(\theta)-1 \end{bmatrix}\begin{pmatrix} x_i \\ y_i \end{pmatrix} < 0\right\} \\
&= \Pr_{\mathcal{N}(0,\,I)}\left\{\sum_{i=1}^{k}((\cos(\theta)+1)x_i^2 + (\cos(\theta)-1)y_i^2) < 0\right\} \\
&= \Pr_{\mathcal{N}(0,\,I)}\left\{(\cos(\theta)+1)\sum_{i=1}^{k}x_i^2 + (\cos(\theta)-1)\sum_{i=1}^{k}y_i^2 < 0\right\} \\
&= \Pr_{\mathcal{N}(0,\,I)}\left\{\frac{(\cos(\theta)+1)\sum_{i=1}^{k}x_i^2}{(\cos(\theta)-1)\sum_{i=1}^{k}y_i^2} + 1 < 0\right\} \\
&= \Pr_{\mathcal{N}(0,\,I)}\left\{\frac{\sum_{i=1}^{k}x_i^2}{\sum_{i=1}^{k}y_i^2} < \frac{1-\cos(\theta)}{1+\cos(\theta)}\right\} \qquad (1.11)
\end{aligned}$$

Now, x_i and y_i are standard univariate Gaussian variables, hence $x_i^2, y_i^2 \overset{iid}{\sim} \chi^2$, and so the left hand side of (1.11) is F-distributed ([13], Appendix B.4, pg 487). Therefore:

$$\Pr_R[(R\mathbf{n})^T R\mathbf{m} < 0 | \mathbf{n}^T\mathbf{m} \geqslant 0] = \frac{\Gamma(k)}{(\Gamma(k/2))^2}\int_0^{\psi}\frac{z^{(k-2)/2}}{(1+z)^k}\mathrm{d}z$$

where $\psi = (1-\cos(\theta))/(1+\cos(\theta))$. This proves the first part of the theorem. □

Remark. Notice that the flip probability is independent of the lengths of the original vectors. As noted in [12], dependence on the lengths is a key difficulty with deriving useful Johnson-Lindenstrauss type tail bounds on dot products $(R\mathbf{n})^T R\mathbf{m}$, and they show that assuming these lengths are known can tighten their bounds. In turn, whenever preservation of the signs of dot products is sufficient, the probability of this can be computed exactly, cf. our result, without any reference to the length information.

Secondly, it is remarkable that the flip probability is also independent of the original dimensionality d of the data. It only depends on the target space dimension k and the angular separation of the original vectors θ.

Proof of part 2. Note that $\psi = \tan^2(\theta/2)$ and make the substitution $z = \tan^2(\theta/2)$. Then, via the trigonometric identity $\sin(\theta) = 2\tan(\theta)/(1+\tan^2(\theta))$ and $\frac{\mathrm{d}z}{\mathrm{d}\theta} = \tan(\theta/2)(1+\tan^2(\theta/2))$, we obtain:

$$\frac{\Gamma(k)}{2^{k-1}(\Gamma(k/2))^2}\int_0^\theta \sin^{k-1}(\phi)\mathrm{d}\phi \tag{1.12}$$

To put the expression (1.12) in the form of the second part of the theorem, we need to show that the gamma term outside the integral is the reciprocal of $\int_0^\pi \sin^{k-1}(\phi)\mathrm{d}\phi$. We can show this in a straightforward way using the beta function.

Recall that the beta function is defined by:

$$\beta(w,z) = \frac{\Gamma(w)\Gamma(z)}{\Gamma(w+z)} = 2\int_0^{\pi/2} \sin^{2w-1}(\theta)\cos^{2z-1}(\theta)\mathrm{d}\theta,\ \ \mathrm{Re}(w), \mathrm{Re}(z) > 0 \tag{1.13}$$

Then from (1.13) we have:

$$\frac{1}{2}\beta\left(\frac{k}{2},\frac{1}{2}\right) = \int_0^{\pi/2} \sin^{k-1}(\theta)\mathrm{d}\theta$$

and from the symmetry of the sine function about $\pi/2$, equation (1.13), and using $\Gamma(1/2) = \sqrt{\pi}$ we have:

$$\int_0^{\pi} \sin^{k-1}(\theta)\mathrm{d}\theta = 2\int_0^{\pi/2} \sin^{k-1}(\theta)\mathrm{d}\theta = \beta\left(\frac{k}{2},\frac{1}{2}\right) = \frac{\sqrt{\pi}\ \Gamma(k/2)}{\Gamma((k+1)/2)}$$

Now we just need to show that the left hand side of (1.12):

$$\frac{\Gamma(k)}{2^{k-1}(\Gamma(k/2))^2} = \frac{\Gamma((k+1)/2)}{\sqrt{\pi}\ \Gamma(k/2)} \tag{1.14}$$

To do this we use the duplication formula ([1], 6.1.18, pg 256): $\Gamma(2z) = (2\pi)^{-\frac{1}{2}}2^{2z-\frac{1}{2}}\Gamma(z)\Gamma((2z+1)/2)$ with $z = k/2$.
Then the left hand side of (1.14) is equal to:

$$\frac{2^{k-\frac{1}{2}}\Gamma(k/2)\Gamma((k+1)/2)}{\sqrt{2\pi}\ 2^{k-1}(\Gamma(k/2))^2} = \frac{\Gamma(k/2)\Gamma((k+1)/2)}{\sqrt{\pi}\ (\Gamma(k/2))^2} = \frac{\Gamma((k+1)/2)}{\sqrt{\pi}\ \Gamma(k/2)}$$

as required. Putting everything together, we arrive at the alternative form for (1.12) given in equation (1.2), namely:

$$\Pr_R[(R\mathbf{n})^T R\mathbf{m} < 0 | \mathbf{n}^T\mathbf{m} \geqslant 0] = \frac{\int_0^\theta \sin^{k-1}(\phi)\ \mathrm{d}\phi}{\int_0^\pi \sin^{k-1}(\phi)\ \mathrm{d}\phi} \tag{1.15}$$

This proves the second part of the theorem (and we will give a geometric interpretation of this form of the flip probability in the next section). □

Remark. It is easy to verify that (1.15) recovers the known result for $k = 1$, namely θ/π, as given in [7] (Lemma 3.2). Further, for general k, the expression (1.15) is polynomial of order k in θ. This can be seen by using the fact that $\sin(\phi) \leqslant \phi$, which gives us the upper bound

$\Pr_R[(R\mathbf{n})^T R\mathbf{m} < 0 | \mathbf{n}^T\mathbf{m} \geqslant 0] \leqslant \frac{\theta^k}{k \cdot \int_0^\pi \sin^{k-1}(\phi)\ \mathrm{d}\phi}$.

Proof of part 3. Finally, we prove that the flip probability is inversely related to the projection dimension k. Note that although the value of the expressions in (1.2) and (1.1) can be calculated exactly for any given k and θ (e.g. using integration by parts) there is no general closed form for either the integral or the gamma term and, as k grows, this becomes increasingly inconvenient. The final part of the theorem, bounding the flip probability in the $(k+1)$-dimensional case above by the flip probability in the k-dimensional case, is therefore useful in practice.

To prove the final part of the theorem we will show that for all $\theta \in [0, \pi/2]$, the ratio of successive flip probabilities:

$$\frac{f_{k+1}(\theta)}{f_k(\theta)} = \frac{\left(\dfrac{\int_0^\theta \sin^k(\phi)\ \mathrm{d}\phi}{\int_0^\pi \sin^k(\phi)\ \mathrm{d}\phi}\right)}{\left(\dfrac{\int_0^\theta \sin^{k-1}(\phi)\ \mathrm{d}\phi}{\int_0^\pi \sin^{k-1}(\phi)\ \mathrm{d}\phi}\right)} \leqslant 1 \tag{1.16}$$

which is sufficient.

Let us rewrite the ratio (1.16) above as:

$$\frac{\left(\dfrac{\int_0^\theta \sin^k(\phi)\ \mathrm{d}\phi}{\int_0^\theta \sin^{k-1}(\phi)\ \mathrm{d}\phi}\right)}{\left(\dfrac{\int_0^\pi \sin^k(\phi)\ \mathrm{d}\phi}{\int_0^\pi \sin^{k-1}(\phi)\ \mathrm{d}\phi}\right)} \tag{1.17}$$

Call the numerator of (1.17) $g_k(\theta)$, and notice that the denominator is nothing but $g_k(\pi)$.

Now observe that the denominator, $g_k(\pi) = g_k(\pi/2)$. Since:

$$\frac{\int_0^\pi \sin^k(\phi)\ \mathrm{d}\phi}{\int_0^\pi \sin^{k-1}(\phi)\ \mathrm{d}\phi} = \frac{2\int_0^{\pi/2} \sin^k(\phi)\ \mathrm{d}\phi}{2\int_0^{\pi/2} \sin^{k-1}(\phi)\ \mathrm{d}\phi} = \frac{\int_0^{\pi/2} \sin^k(\phi)\ \mathrm{d}\phi}{\int_0^{\pi/2} \sin^{k-1}(\phi)\ \mathrm{d}\phi}$$

where the first equality follows from the symmetry of the sine function about $\pi/2$. Hence, we see that the whole expression (1.17) is equal to 1 when $\theta = \pi/2$.

It remains now to show that the $g_k(\theta) \leq g_k(\pi/2), \forall\theta \in [0, \pi/2]$ and $k \in \mathbb{N}$. In fact more is true: We show that $\forall k, g_k(\theta)$ is monotonic increasing as a function of θ on $[0, \pi/2]$. From this the required inequality follows, and hence the expression (1.17) has its maximum value of 1 on this domain, from which the result follows.

To show monotonicity, we differentiate the function $g_k(\theta)$ with respect to θ to obtain:

$$\frac{\mathrm{d}}{\mathrm{d}\theta}g_k(\theta) = \frac{\sin^k(\theta)\int_0^\theta \sin^{k-1}(\phi)\mathrm{d}\phi - \sin^{k-1}(\theta)\int_0^\theta \sin^k(\phi)\mathrm{d}\phi}{\left(\int_0^\theta \sin^{k-1}(\phi)\mathrm{d}\phi\right)^2} \tag{1.18}$$

Then (1.18) is greater than zero when its numerator is, and:

$$\sin^k(\theta)\int_0^\theta \sin^{k-1}(\phi)\mathrm{d}\phi - \sin^{k-1}(\theta)\int_0^\theta \sin^k(\phi)\mathrm{d}\phi$$

$$= \sin^{k-1}(\theta)\left[\sin(\theta)\int_0^\theta \sin^{k-1}(\phi)\mathrm{d}\phi - \int_0^\theta \sin^k(\phi)\mathrm{d}\phi\right]$$

$$= \sin^{k-1}(\theta)\left[\int_0^\theta \sin(\theta)\sin^{k-1}(\phi)\mathrm{d}\phi - \int_0^\theta \sin(\phi)\sin^{k-1}(\phi)\mathrm{d}\phi\right] \geqslant 0$$

Where the last step follows from monotonicity of the sine function on $[0, \pi/2]$ and so $\sin(\theta) \geqslant \sin(\phi)$ for $\theta \geqslant \phi > 0$, $\theta \in [0, \pi/2]$. It follows now that the numerator of (1.17) is monotonic increasing with $\theta \in [0, \pi/2]$ and so the whole expression (1.16) takes its maximum value of 1 when $\theta = \pi/2$. This completes the proof of the theorem. ■

Remark. We note that for $\theta \in [\pi/2, \pi]$ it is easy to show, using the symmetry of sine about $\pi/2$, that the sense of the inequality in part 3 of the theorem is reversed. Then: $f_{k+1}(\theta) \geqslant f_k(\theta)$, $\theta \in [\pi/2, \pi]$.

1.3. **Geometric Interpretation.** In the case $k = 1$, the flip probability θ/π (given also in [7]) is the quotient of the length of an arc of 2θ by the circumference of the corresponding circle $r2\pi$. It is interesting to observe that our result written in the form of (1.15) generalises this geometric interpretation in a natural way, as follows.

Recall that the surface area of a hypersphere that lives in a $(k+1)$-dimensional space and having radius r, is given by [10]:

$$r^k \cdot 2\pi \cdot \prod_{i=1}^{k-1}\int_0^\pi \sin^i(\phi)\mathrm{d}\phi$$

(which is is also $(k+1)/r$ times the volume of a the same hypersphere.) This expression is seen to be the extension of the standard 'integrating slabs' approach to finding the volume of the 3-dimensional sphere S_2, and so the surface area of the hyperspherical cap subtending angle 2θ is simply:

$$r^k \cdot 2\pi \cdot \prod_{i=1}^{k-2}\int_0^\pi \sin^i(\phi)\mathrm{d}\phi \cdot \int_0^\theta \sin^{k-1}(\phi)\mathrm{d}\phi$$

If we now take the quotient of these two areas, all but the last terms cancel, and we obtain:

$$\frac{r^k \cdot 2\pi \cdot \prod_{i=1}^{k-2} \int_0^\pi \sin^i(\phi)\mathrm{d}\phi \cdot \int_0^\theta \sin^{k-1}(\phi)\mathrm{d}\phi}{r^k \cdot 2\pi \cdot \prod_{i=1}^{k-1} \int_0^\pi \sin^i(\phi)\mathrm{d}\phi} = \frac{\int_0^\theta \sin^{k-1}(\phi)\ \mathrm{d}\phi}{\int_0^\pi \sin^{k-1}(\phi)\ \mathrm{d}\phi}$$

which is exactly our flip probability as given in (1.15).
Hence, the probability that a dot product flips from being positive to being negative (equivalently the angle flips from acute to obtuse) after Gaussian random transformation is given by the ratio of the surface area in $\mathbb{R}^{k+1}$ of a hyperspherical cap to the surface area of the corresponding hypersphere.

2. Discussion

The problem of finding the flip probability for general k arose in the context of evaluating the error of a linear classification algorithm in randomly projected data spaces [5], specifically for studying a particular finite sample effect. However, flip probabilities for $k = 1$ using the result derived initially for semidefinite programming [7] have found use is a wide range of applications including: Certain classes of hash functions [4], sign random projections for storage-efficient data sketches and recovery of angles between high dimensional points [12], approximate similarity search methods in high dimensions [11], and data classification with a better-than-chance guarantee [3]. We believe therefore that our results may also have utility in some such areas. For example, the latter two areas exploit the idea that, if $\mathbf{n}$ is closer to the query point than $\mathbf{m}$ in the data space, then there is greater than 0.5 chance that this is also the case following projection onto a random line. This chance may be not high enough though, and one way to increase it in practice has been to inspect several independent trials of the projection [3]. Knowing the flip probability for a general k may therefore be used to control the above chance by choosing k to ensure a specified flip probability δ as low (or as high) as desired.

We also find it interesting to relate our results with [6]. There, the authors give the probability that a random Gaussian triangle (i.e. the vertices are k-dimensional vectors with components selected randomly from the standard Gaussian $\mathcal{N}(0, 1)$) is obtuse – this was also found to be an upper bound to the probability previously found by [8] for the case when the vertices are drawn uniformly in a k-dimensional ball. Comparing our result with theirs, we see that the probability that a Gaussian triangle is obtuse is three times our flip probability evaluated at $\theta = \pi/3$ – that is, three times the probability that a d-dimensional *equilateral* triangle becomes obtuse following a random Gaussian projection into k dimensions. (The factor of three comes in because a triangle can be obtuse in three different and mutually exclusive ways, i.e. if any one of its angles is obtuse.) Hence, this interpretation generalises the random triangle problem considered there (if restricted to one designated angle), in the sense that a range of θ other than just $\pi/3$ can be considered this way for the 'original' deterministic triangle that lives in the d-dimensional space. Following up on this observation it may be interesting, as future research, to consider the (expected) geometry of random projections of other deterministic geometric objects. For example, random projection of the convex hull of a point set in $\mathbb{R}^d$.

References

[1] Abramowitz, M. and Stegun, I. (1972). *Handbook of mathematical functions with formulas, graphs, and mathematical tables* 10th ed. Dover, New York.

[2] Arriaga, R. and Vempala, S. (2006). An algorithmic theory of learning. *Machine Learning* **63,** 161–182.

[3] Blum, A. (2006). Random projection, margins, kernels, and feature-selection. In *SLSFS 2005.* ed. Saunders et al. No. 3940 in LNCS. pp. 55–68.

[4] Charikar, M. (2002). Similarity estimation techniques from rounding algorithms. In *Proceedings of the thirty-fourth annual ACM symposium on Theory of computing.* ACM. pp. 380–388.

[5] Durrant, R. J. and Kabán, A. (2010). Compressed Fisher Linear Discriminant Analysis: Classification of Randomly Projected Data. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2010).* ACM.

[6] Eisenberg, B. and Sullivan, R. (1996). Random triangles in n dimensions. *American Mathematical Monthly* **103,** 308–318.

[7] Goemans, M. and Williamson, D. (1995). Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)* **42,** 1145.

[8] Hall, G. (1982). Acute triangles in the n-ball. *Journal of Applied Probability* **19,** 712–715.

[9] Horn, R. A. and Johnson, C. R. (1985). *Matrix Analysis.* CUP, New York.

[10] Kendall, M. (2004). *A Course in the Geometry of n Dimensions.* Dover Pubns, New York.

[11] Kleinberg, J. (1997). Two algorithms for nearest-neighbor search in high dimensions. In *Proceedings of the twenty-ninth annual ACM Symposium on Theory of Computing (STOC 1997).* ACM. p. 608.

[12] Li, P., Hastie, T. and Church, K. (2006). Improving random projections using marginal information. *Learning Theory* **4005,** 635–649.

[13] Mardia, K., Kent, J. and Bibby, J. (1979). *Multivariate analysis.* Academic Press, London.

Appendix

$$\mathrm{Cov}\left[\begin{pmatrix} u_i \\ v_i \end{pmatrix}\right] = \begin{bmatrix} \mathrm{Var}(u_i) & \mathrm{Cov}(u_i, v_i) \\ \mathrm{Cov}(u_i, v_i) & \mathrm{Var}(v_i) \end{bmatrix}$$

Then:

$$\mathrm{Var}(u_i) = \mathrm{E}[(u_i - \mathrm{E}(u_i))^2]$$
$$= \mathrm{E}[(u_i)^2 - 2u_i\mathrm{E}(u_i) + (\mathrm{E}(u_i))^2], \text{ but } \mathrm{E}(u_i) = 0 \text{ and so:}$$

$$
\begin{aligned}
&= \mathrm{E}[(u_i)^2] \\
&= \mathrm{E}\left[\left(\sum_{j=1}^{d} r_{ij} n_j\right)^2\right] \\
&= \mathrm{E}\left[\sum_{\substack{j=1\\j'=1}}^{d} r_{ij} r_{ij'} n_j n_{j'}\right] \\
&= \sum_{\substack{j=1\\j'=1}}^{d} n_j n_{j'} \mathrm{E}\left[r_{ij} r_{ij'}\right]
\end{aligned}
$$

Now, when $j \neq j'$, r_{ij} and $r_{ij'}$ are independent, and so $\mathrm{E}[r_{ij}r_{ij'}] = \mathrm{E}[r_{ij}]\mathrm{E}[r_{ij'}] = 0$. On the other hand, when $j = j'$ we have $\mathrm{E}[r_{ij}r_{ij'}] = \mathrm{E}[r_{ij}^2] = \mathrm{Var}(r_{ij}) = 1/d$, since $r_{ij} \sim \mathcal{N}(0, 1/d)$. Hence:

$$
\mathrm{Var}(u_i) = \sum_{j=1}^{d} \frac{n_j^2}{d} = \frac{1}{d}\|\mathbf{n}\|^2
$$

and a similar argument then gives $\mathrm{Var}(v_i) = \|\mathbf{m}\|^2/d$.

To find the covariance $\mathrm{Cov}(u_i, v_i)$, we write:

$$
\begin{aligned}
\mathrm{Cov}(u_i, v_i) &= \mathrm{E}\left[(u_i - \mathrm{E}[u_i])(v_i - \mathrm{E}[v_i])\right] \\
&= \mathrm{E}[u_i v_i] \\
&= \mathrm{E}\left[\left(\sum_{j=1}^{d} r_{ij} n_j\right)\left(\sum_{j=1}^{d} r_{ij} m_j\right)\right] \\
&= \sum_{\substack{j=1\\j'=1}}^{d} n_j m_{j'} \mathrm{E}[r_{ij} r_{ij'}] \qquad (2.1)
\end{aligned}
$$

When $j \neq j'$ the expectation is zero, as before, and similarly when $j = j'$ we have for (2.1):

$$
= \sum_{j=1}^{d} n_j m_j \mathrm{E}[(r_{ij})^2] = \sum_{j=1}^{d} n_j m_j \mathrm{Var}(r_{ij}) = \frac{1}{d}\sum_{j=1}^{d} n_j m_j = \frac{1}{d}\mathbf{n}^T\mathbf{m}
$$

E-mail address, RJD: `r.j.durrant@cs.bham.ac.uk`
URL: `http://www.cs.bham.ac.uk/~durranrj`

E-mail address, AK: `a.kaban@cs.bham.ac.uk`
URL: `http://www.cs.bham.ac.uk/~axk`

(RJD and AK) School of Computer Science,
University of Birmingham
Edgbaston, UK,
B15 4TT