Proofs for Observable Propagation

Jacob Dunefsky

1 Proof of Theorem 1

Theorem 1. *Define* $f(x;n) = n \cdot \mathrm{LayerNorm}(x)$. *Define*

$$\theta(x;n) = \arccos\left(\frac{n \cdot \nabla_x f(x;n)}{\|n\| \, \|\nabla_x f(x;n)\|}\right)$$

– that is, $\theta(x;n)$ *is the angle between* n *and* $\nabla_x f(x;n)$. *Then if* x *and* n *are i.i.d.* $\mathcal{N}(0,I)$ *in* $\mathbb{R}^d$, *and* $d \geq 8$ *then*

$$\mathbb{E}\left[\theta(x;n)\right] < 2\arccos\left(\sqrt{1 - \frac{1}{d-1}}\right)$$

To prove this, we will introduce a lemma:

Lemma 1. *Let* y *be an arbitrary vector. Let* $A = I - \frac{vv^T}{\|v\|^2}$ *be the orthogonal projection onto the hyperplane normal to* v. *Then the cosine similarity between* y *and* Ay *is given by* $\sqrt{1 - \cos(\theta)^2}$, *where* $\cos(\theta)$ *is the cosine similarity between* y *and* v.

Proof. Assume without loss of generality that y is a unit vector. (Otherwise, we could rescale it without affecting the angle between y and v, or the angle between y and Ay.)

We have $Ay = y - \frac{y \cdot v}{\|v\|^2} v$. Then,

$$\begin{aligned} y \cdot Ay &= y \cdot (y - \frac{y \cdot v}{\|v\|^2} v) \\ &= \|y\|^2 - \frac{(y \cdot v)^2}{\|v\|^2} \\ &= 1 - \frac{(y \cdot v)^2}{\|v\|^2} \end{aligned}$$

and

$$
\begin{aligned}
\|Ay\|^2 &= (y - \frac{y \cdot v}{\|v\|^2} v) \cdot (y - \frac{y \cdot v}{\|v\|^2} v) \\
&= y \cdot (y - \frac{y \cdot v}{\|v\|^2} v) - \frac{y \cdot v}{\|v\|^2} v \cdot (y - \frac{y \cdot v}{\|v\|^2} v) \\
&= y \cdot Ay - \frac{y \cdot v}{\|v\|^2} v \cdot (y - \frac{y \cdot v}{\|v\|^2} v) \\
&= y \cdot Ay - \frac{(y \cdot v)^2}{\|v\|^2} + \left\| \frac{y \cdot v}{\|v\|^2} v \right\|^2 \\
&= y \cdot Ay - \frac{(y \cdot v)^2}{\|v\|^2} + \frac{(y \cdot v)^2}{\|v\|^4} \|v\|^2 \\
&= y \cdot Ay - \frac{(y \cdot v)^2}{\|v\|^2} + \frac{(y \cdot v)^2}{\|v\|^2} \\
&= y \cdot Ay
\end{aligned}
$$

Now, the cosine similarity between y and Ay is given by

$$
\begin{aligned}
\frac{y \cdot Ay}{\|y\| \|Ay\|} &= \frac{y \cdot Ay}{\|Ay\|} \\
&= \frac{\|Ay\|^2}{\|Ay\|} \\
&= \|Ay\|
\end{aligned}
$$

At this point, note that $\|Ay\| = \sqrt{y \cdot Ay} = \sqrt{1 - \frac{(y \cdot v)^2}{\|v\|^2}}$. But $\frac{y \cdot v}{\|v\|}$ is just the cosine similarity between y and v. Now, if we denote the angle between y and v by θ, we thus have

$$
\|Ay\| = \sqrt{1 - \frac{(y \cdot v)^2}{\|v\|^2}} = \sqrt{1 - \cos(\theta)^2}.
$$

□

Now, we are ready to prove Theorem 1.

Proof. First, observe that $\text{LayerNorm}(x) = \frac{Px}{\|Px\|}$, where $P = I - \frac{1}{d}\vec{1}\vec{1}^T$ is the orthogonal projection onto the hyperplane normal to $\vec{1}$, the vector of all ones. Thus, we have

$$
f(x; n) = n^T \left(\frac{Px}{\|Px\|} \right)
$$

Using the multivariate chain rule along with the rule that the derivative of $\frac{x}{\|x\|}$ is given by $\frac{I}{\|x\|} - \frac{xx^T}{\|x\|^3}$ (see section 2.6.1 of The Matrix Cookbook), we thus have that

$$
\begin{aligned}
\nabla_x f(x;n) &= \left(n^T \left(\frac{I}{\|Px\|} - \frac{(Px)(Px)^T}{\|Px\|^3}\right) P\right)^T \\
&= \left(\frac{1}{\|Px\|} n^T \left(I - \frac{(Px)(Px)^T}{\|Px\|^2}\right) P\right)^T \\
&= \frac{1}{\|Px\|} P \left(I - \frac{(Px)(Px)^T}{\|Px\|^2}\right) n \qquad \text{because } P \text{ is symmetric}
\end{aligned}
$$

Denote $Q = I - \frac{(Px)(Px)^T}{\|Px\|^2}$. Note that this is an orthogonal projection onto the hyperplane normal to Px. We now have that $\nabla_x f(x;n) = \frac{1}{\|Px\|} PQn$. Because we only care about the angle between n and $\nabla_x f(x;n)$, it suffices to look at the angle between n and PQn, ignoring the $\frac{1}{\|Px\|}$ term.

Denote the angle between n and PQn as $\theta(x, n)$. (Note that θ is also a function of x because Q is a function of x.) Then if $\theta_Q(x, n)$ is the angle between n and Qn, and $\theta_P(x, n)$ is the angle between Qn and PQn, then $\theta(x, n) \leq \theta_Q(x, n) + \theta_P(x, n)$, so $\mathbb{E}[\theta(x, n)] \leq \mathbb{E}[\theta_Q(x, n)] + \mathbb{E}[\theta_P(x, n)]$.

Using Lemma 1, we have that $\theta_Q(x, n) = \arccos\left(\sqrt{1 - \cos(\phi(n, Px))^2}\right)$, where $\phi(n, Px)$ is the angle between n and Px. Now, because $n \sim \mathcal{N}(0, I)$, we have $\mathbb{E}[\cos(\phi(n, Px))^2] = 1/d$, using the well-known fact that the expected squared dot product between a uniformly distributed unit vector in $\mathbb{R}^d$ and a given unit vector in $\mathbb{R}^d$ is $1/d$.

At this point, define $g(t) = \arccos\left(\sqrt{1-t}\right)$, $h(t) = g'\left(\frac{1}{d-1}\right)\left(t - \frac{1}{d-1}\right) + g\left(\frac{1}{d-1}\right)$. Then if $\frac{1}{d-1} < c$, where c is the least solution to $g'(c) = \frac{\pi - 2g(c)}{2(1-c)}$, then $h(t) \geq g(t)$. (Note that $g(t)$ is convex on $(0, 0.5]$ and concave on $[0.5, 1)$. Therefore, there are exactly two solutions to $g'(c) = \frac{\pi - 2g(c)}{2(1-c)}$. The lesser of the two solutions is the value at which $g'(c)$ equals the slope of the line between $(c, g(c))$ and $(1, \pi/2)$ – the latter point being the maximum of g – at the same time that $g''(c) \geq 0$.) One can compute $c \approx 0.155241\ldots$, so if $d \geq 8$, then $1/(d-1) < c$ is satisfied, so $h(t) \geq g(t)$. Thus, we have the following inequality:

$$
\begin{aligned}
h(1/(d-1)) &> h(1/d) \\
&= h(\mathbb{E}[\cos(\phi(n, Px))^2]) \\
&= \mathbb{E}[h(\cos(\phi(n, Px))^2)] \text{ due to linearity} \\
&\geq \mathbb{E}[g(\cos(\phi(n, Px))^2)] \text{ because } h(t) \geq g(t) \text{ for all } t \\
&= \mathbb{E}[\theta_Q(x, n)]
\end{aligned}
$$

Now, $h(1/(d-1)) = g(1/(d-1)) = \arccos\left(\sqrt{1 - \frac{1}{d-1}}\right)$. Thus, we have that $\arccos\left(\sqrt{1 - \frac{1}{d-1}}\right) > \mathbb{E}[\theta_Q(x, n)]$.

The next step is to determine an upper bound for $\mathbb{E}[\theta_P(x,n)]$. By Lemma 1, we have that $\theta_P(x,n) = \arccos\left(\sqrt{1-\cos(\phi(Qn,\vec{1}))^2}\right)$. Now, note that because $n \sim \mathcal{N}(0,I)$, then Qn is distributed according to a unit Gaussian in $\operatorname{Im} Q$, the $(d-1)$-dimensional hyperplane orthogonal to Px. Note that because $\vec{1}$ is orthogonal to Px (by the definition of P) and Px is orthogonal to $\operatorname{Im} Q$, this means that $\vec{1} \in \operatorname{Im} Q$. Now, let us apply the same fact from earlier: that the expected squared dot product between a uniformly distributed unit vector in $\mathbb{R}^{d-1}$ and a given unit vector in $\mathbb{R}^{d-1}$ is $1/(d-1)$. Thus, we have that $\mathbb{E}[\cos(\phi(Qn,\vec{1}))^2] = 1/(d-1)$.

From this, by the same logic as in the previous case, $\arccos\left(\sqrt{1-\frac{1}{d-1}}\right) \geq \mathbb{E}[\theta_P(x,n)]$.

Adding this inequality to the inequality for $\mathbb{E}[\theta_Q(x,n)]$, we have

$$2\arccos\left(\sqrt{1-\frac{1}{d-1}}\right) > \mathbb{E}[\theta_Q(x,n)] + \mathbb{E}[\theta_P(x,n)] \geq \mathbb{E}[\theta(x,n)]$$

. ∎

2 Proof of Theorem 2

Theorem 2. *Let $y_1, y_2 \in \mathbb{R}^d$. Let x be uniformly distributed on the hypersphere defined by the constraints $\|x\| = s$ and $x \cdot y_1 = k$. Then we have*

$$\mathbb{E}[x \cdot y_2] = k\frac{y_1 \cdot y_2}{\|y_1\|^2}$$

and the maximum and minimum values of $x \cdot y_2$ are given by

$$\frac{\|y_2\|}{\|y_1\|}\left(k\cos(\theta) \pm \sin(\theta)\sqrt{s^2\|y_1\|^2 - k^2}\right)$$

where θ is the angle between y_1 and y_2.

Before proving Theorem 2, we will prove a quick lemma.

Lemma 2. *Let $\mathcal{S}$ be a hypersphere with radius r and center c. Then for a given vector y, the mean squared distance from y to the sphere, $\mathbb{E}_{s\in\mathcal{S}}[\|y-c\|^2]$, is given by $\|y-c\|^2 + r^2$.*

Proof. Without loss of generality, assume that $\mathcal{S}$ is centered at the origin (so $\|y-c\|^2 = \|y\|^2$). Induct on the dimension of the $\mathcal{S}$. As our base case, let $\mathcal{S}$ be the 0-sphere consisting of a point in $\mathbb{R}^1$ at $-r$ and a point at r. Then $\mathbb{E}_{s\in\mathcal{S}}[|y-s|^2] = \frac{(y-r)^2+(y-(-r))^2}{2} = y^2 + r^2$.

For our inductive step, assume the inductive hypothesis for spheres of dimension $d-2$; we will prove the theorem of spheres of dimension $d-1$ in an ambient space of dimension d. Without loss of generality, let y lie on the x-axis, so that we have $y = \begin{bmatrix} y_1 & 0 & 0 & \dots \end{bmatrix}^T$. Next, divide $\mathcal{S}$ into slices along the x-axis. Denote the slice at position $x = x_0$ as $\mathcal{S}_{x_0}$. Then $\mathcal{S}_{x_0}$ is a $(d-2)$-sphere centered at $\begin{bmatrix} x_0 & 0 & 0 & \dots \end{bmatrix}^T$, and has radius $\sqrt{r^2 - x_0^2}$. Now, by the law of total expectation, $\mathbb{E}_{s \in \mathcal{S}}[\|y - s\|^2] = \mathbb{E}_{-r \le x \le r}\left[\mathbb{E}_{s' \in \mathcal{S}_x}\left[\|y - s'\|^2\right]\right]$. We then have that

$$\begin{aligned}
\mathbb{E}_{s' \in \mathcal{S}_x}\left[\|y - s'\|^2\right] &= \mathbb{E}\left[(y_1 - x)^2 + s_2^2 + s_3^2 + \cdots\right] \\
&= (y_1 - x)^2 + \mathbb{E}\left[s_2^2 + s_3^2 + \cdots\right]
\end{aligned}$$

Once again, $\mathcal{S}_x$ is a $(d-2)$-sphere defined by $s_2^2 + s_3^2 + \cdots = r^2 - x^2$. This means that by the inductive hypothesis, we have $\mathbb{E}\left[s_2^2 + s_3^2 + \cdots\right] = r^2 - x^2$. Thus, we have

$$\begin{aligned}
\mathbb{E}_{s' \in \mathcal{S}_x}\left[\|y - s'\|^2\right] &= (y_1 - x)^2 + r^2 - x^2 \\
\mathbb{E}_{s' \in \mathcal{S}_x}\left[\|y - s'\|^2\right] &= (y_1 - x)^2 + r^2 - x^2 \\
\mathbb{E}_{s \in \mathcal{S}}[\|y - s\|^2] &= \mathbb{E}_{-r \le x \le r}\left[(y_1 - x)^2 + r^2 - x^2\right] \\
&= \frac{1}{2r}\int_{-r}^{r} (y_1 - x)^2 + r^2 - x^2 dx \\
&= r^2 + y_1^2
\end{aligned}$$

□

We are now ready to begin the main proof.

Proof. First, assume that $\|x\| = 1$. Now, the intersection of the $(d-1)$-sphere defined by $\|x\| = 1$ and the hyperplane $x \cdot y_1 = k$ is a unit hypersphere of dimension $(d-2)$, oriented in the hyperplane $x \cdot y_1 = k$, and centered at $c_1 y_1$ where $c_1 = k / \|y_1\|^2$. Denote this $(d-2)$-sphere as $\mathcal{S}$, and denote its radius by r.

Next, define $c_2 = \frac{k}{y_2 \cdot y_1}$. Then $cy_2 \cdot y_1 = k$, so $c_2 y_2$ lies in the same hyperplane as $\mathcal{S}$. Additionally, because $c_1 y_1$ is in this hyperplane, and $c_1 y_1$ is also the normal vector for this hyperplane, we have that the vectors $c_1 y_1$, $c_2 y_2$, and $c_1 y_1 - c_2 y_2$ form a right triangle, where $c_2 y_2$ is the hypotenuse and $c_1 y_1 - c_2 y_2$ is the leg opposite of the angle θ between y_1 and y_2. As such, we have that $\|c_1 y_1 - c_2 y_2\| = \sin(\theta) \|c_2 y_2\|$.

Furthermore, we have that $c_1 y_1 \cdot c_2 y_2 = \frac{k^2}{\|y_1\|^2}$, that $\|c_1 y_1\| = \frac{|k|}{\|y_1\|^2}$, and that $\|c_2 y_2\| = \frac{|k|}{\|y_1\| |\cos\theta|}$

We will now begin to prove that the maximum and minimum values of $y_2 \cdot x$ are given by $\frac{\|y_2\|}{\|y_1\|}\left(k\cos(\theta) \pm |\sin(\theta)| \sqrt{s^2 \|y_1\|^2 - k^2}\right)$.

To start, note that the nearest point on $\mathcal{S}$ to c_2y_2 and the farthest point on $\mathcal{S}$ from c_2y_2 are located at the intersection of $\mathcal{S}$ with the line between c_2y_2 and c_1y_1.

To see this, let x_+ be the at the intersection of $\mathcal{S}$ and the line between c_2y_2 and c_1y_1. We will show that x_+ is the nearest point on $\mathcal{S}$ to c_2y_2. Let $x'_+ \in \mathcal{S} \neq x_+$. Then we have the following cases:

- Case 1: c_2y_2 is outside of $\mathcal{S}$. Then $\|c_2y_2 - c_1y_1\| = \|c_2y_2 - x_+\| + \|x_+ - c_1y_1\|$, because c_2y_2, x_+, and c_1y_1 are collinear – so $\|c_2y_2 - c_1y_1\| = \|c_2y_2 - x_+\| + r$ (because $x_+ \in \mathcal{S}$). By the triangle inequality, we have $\|c_2y_2 - c_1y_1\| \leq \left\|c_2y_2 - x'_+\right\| + \left\|x'_+ - c_1y_1\right\| = \left\|c_2y_2 - x'_+\right\| + r$. But this means that $\|c_2y_2 - x_+\| \leq \left\|c_2y_2 - x'_+\right\|$.

- Case 2: c_2y_2 is inside of $\mathcal{S}$. Then $\|c_2y_2 - c_1y_1\| = \|x_+ - c_1y_1\| - \|c_2y_2 - x_+\|$, because c_2y_2, x_+, and c_1y_1 are collinear – so $\|c_2y_2 - c_1y_1\| = r - \|c_2y_2 - x_+\|$. By the triangle inequality, we have $\left\|x'_+ - c_1y_1\right\| \leq \left\|c_2y_2 - x'_+\right\| + \|c_2y_2 - c_1y_1\|$, so $\left\|x'_+ - c_1y_1\right\| \leq \left\|c_2y_2 - x'_+\right\| + r - \|c_2y_2 - x_+\|$. But since $\left\|x'_+ - c_1y_1\right\| = r$, this means that $\|c_2y_2 - x_+\| \leq \left\|c_2y_2 - x'_+\right\|$.

A similar argument will show that x_-, the farthest point on $\mathcal{S}$ from c_2y_2, is also located at the intersection of $\mathcal{S}$ with the line between c_2y_2 and c_1y_1.

Now, let us find the values of x_+ and x_-. The line between c_2y_2 and c_1y_1 can be parameterized by a scalar t as $c_1y_1 + t(c_2y_2 - c_1y_1)$. Then x_+ and x_- are given by $c_1y_1 + t^*(c_2y_2 - c_1y_1)$, where t^* are the solutions to the equation $\|c_1y_1 + t(c_2y_2 - c_1y_1)\| = 1$.

We have the following:

$$
\begin{aligned}
1 &= \|c_1y_1 + t(c_2y_2 - c_1y_1)\| \\
&= \|c_1y_1\|^2 + 2t(c_1y_1 \cdot (c_2y_2 - c_1y_1)) + t^2 \|c_2y_2 - c_1y_1\|^2 \\
&= \|c_1y_1\|^2 + 2t((c_1y_1 \cdot c_2y_2) - \|c_1y_1\|^2) + t^2 \|c_2y_2\|^2 \sin^2\theta \\
&= \frac{k^2}{\|y_1\|^2} + 2t\left(\frac{k^2}{\|y_1\|^2} - \frac{k^2}{\|y_1\|^2}\right) + t^2 \frac{k^2}{\|y_1\|^2 \cos^2\theta} \sin^2\theta \\
&= \frac{k^2}{\|y_1\|^2}(t^2 \tan^2\theta + 1)
\end{aligned}
$$

Thus, solving for t, we have that $t^* = \frac{\pm\sqrt{\|y_1\|^2 - k^2}}{|k| \tan\theta}$. Therefore, we have that

$$
\begin{aligned}
x_+, x_- &= c_1 y_1 + t^*(c_2 y_2 - c_1 y_1) \\
&= c_1 y_1 + \left(\frac{k^2}{\|y_1\|^2} (t^2 \tan^2 \theta + 1) \right) (c_2 y_2 - c_1 y_1) \\
&= \frac{k y_1}{\|y_1\|^2} + \left(\frac{\pm\sqrt{\|y_1\|^2 - k^2}}{|k| \tan\theta} \right) \left(\frac{k y_2}{y_1 \cdot y_2} - \frac{k y_1}{\|y_1\|^2} \right) \\
&= k \left[\frac{y_1}{\|y_1\|^2} \pm \left(\frac{\sqrt{\|y_1\|^2 - k^2}}{|k| \tan\theta} \right) \left(\frac{y_2}{y_1 \cdot y_2} - \frac{y_1}{\|y_1\|^2} \right) \right] \\
y_2 \cdot x_+, y_2 \cdot x_- &= y_2 \cdot k \left[\frac{y_1}{\|y_1\|^2} \pm \left(\frac{\sqrt{\|y_1\|^2 - k^2}}{|k| \tan\theta} \right) \left(\frac{y_2}{y_1 \cdot y_2} - \frac{y_1}{\|y_1\|^2} \right) \right] \\
&= \frac{k y_1 \cdot y_2}{\|y_1\|^2} \pm \left(\cot\theta \sqrt{\|y_1\|^2 - k^2} \right) \left(\frac{y_2 \cdot y_2}{y_1 \cdot y_2} - \frac{y_1 \cdot y_1}{\|y_1\|^2} \right) \\
&= \frac{k y_1 \cdot y_2}{\|y_1\|^2} \pm \left(\cot\theta \sqrt{\|y_1\|^2 - k^2} \right) \left(\frac{\|y_2\|}{\|y_1\| \cos\theta} - \frac{\|y_2\| \cos\theta}{\|y_1\|} \right) \\
&= \left[\frac{k y_1 \cdot y_2}{\|y_1\|^2} \pm \left(\cot\theta \sqrt{\|y_1\|^2 - k^2} \right) \frac{\|y_2\|}{\|y_1\|} \left(\frac{1}{\cos\theta} - \cos\theta \right) \right] \\
&= \frac{k y_1 \cdot y_2}{\|y_1\|^2} \pm \left(\cot\theta \sqrt{\|y_1\|^2 - k^2} \right) \frac{\|y_2\|}{\|y_1\|} \sin\theta \tan\theta \\
&= \frac{k y_1 \cdot y_2}{\|y_1\|^2} \pm \frac{\|y_2\|}{\|y_1\|} \sin\theta \sqrt{\|y_1\|^2 - k^2} \\
&= \frac{\|y_2\|}{\|y_1\|} \left(k \cos(\theta) \pm \sin(\theta) \sqrt{\|y_1\|^2 - k^2} \right)
\end{aligned}
$$

We will now prove that $\mathbb{E}[y_2 \cdot x] = \frac{y_1 \cdot y_2}{\|y_1\|^2}$. Before we do, note that we can also use our value of t^* to determine the squared radius of $\mathcal{S}$. We have that the squared radius of $\mathcal{S}$ is given by

$$
\begin{aligned}
r^2 &= \|t^*(c_2 y_2 - c_1 y_1)\|^2 \\
&= (t^*)^2 \|(c_2 y_2 - c_1 y_1)\|^2 \\
&= (t^*)^2 \sin^2\theta \|c_2 y_2\|^2 \\
&= \frac{\sin^2(\theta) k^2 / \left(\|y_1\|^2 \cos^2\theta \right)}{k^2 \tan\theta} \left(\|y_1\|^2 - k^2 \right) \\
&= 1 - \frac{k^2}{\|y_1\|^2}
\end{aligned}
$$

We will use this result soon. Now, on to the main event. Begin by noting that $y_2 \cdot x = \|y_2\| \|x\| \cos(y_2, x) = \|y_2\| \cos(y_2, x)$, where $\cos(y_2, x)$ is the cosine of the angle between y_2 and x. Now, $\cos(y_2, x) = \text{signum}(c_2)\cos(c_2 y_2, x)$. And we have that $\|x - cy_2\|^2 = \|x\|^2 + \|cy_2\|^2 - 2\|x\| \|c_2 y_2\| \cos(cy_2, x) = 1 + \|c_2 y_2\|^2 - 2\|c_2 y_2\| \cos(c_2 y_2, x)$. Therefore, we have

$$\begin{aligned}
\cos(y_2, x) &= \text{signum}(c_2)\cos(c_2 y_2, x) \\
&= \text{signum}(c_2)\frac{\|x - c_2 y_2\|^2 - 1 - \|c_2 y_2\|^2}{-2\|c_2 y_2\|} \\
&= \text{signum}(c_2)\frac{1 + \|c_2 y_2\|^2 - \|x - c_2 y_2\|^2}{2\|c_2 y_2\|}
\end{aligned}$$

$$\begin{aligned}
y_2 \cdot x &= \|y_2\| \cos(y_2, x) \\
&= \text{signum}(c_2)\|y_2\| \frac{1 + \|c_2 y_2\|^2 - \|x - c_2 y_2\|^2}{2\|c_2 y_2\|}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[y_2 \cdot x] &= \mathbb{E}\left[\text{signum}(c_2)\|y_2\| \frac{1 + \|c_2 y_2\|^2 - \|x - c_2 y_2\|^2}{2\|c_2 y_2\|}\right] \\
&= \text{signum}(c_2)\|y_2\| \frac{1 + \|c_2 y_2\|^2 - \mathbb{E}\left[\|x - c_2 y_2\|^2\right]}{2\|c_2 y_2\|} \\
&= \text{signum}(c_2)\|y_2\| \frac{1 + \|c_2 y_2\|^2 - \left(1 - \frac{k^2}{\|y_1\|^2} + \|c_1 y_1 - c_2 y_2\|^2\right)}{2\|c_2 y_2\|}
\end{aligned}$$

This last line uses Lemma 2: $c_1 y_1$ is the center of $\mathcal{S}$, so the expected squared distance between $c_2 y_2$ and a point on $\mathcal{S}$ is given by $1 - \frac{k^2}{\|y_1\|^2} + \|c_1 y_1 - c_2 y_2\|^2$, where $1 - \frac{k^2}{\|y_1\|^2}$ is the squared radius of $\mathcal{S}$ and $\|c_1 y_1 - c_2 y_2\|^2$ is the squared distance from $c_2 y_2$ to the center. We can use this lemma because $c_2 y_2$ is in the same hyperplane as $\mathcal{S}$, so we can treat this situation as being set in a space of dimension $d - 1$.

Now, continue to simplify:

$$
\begin{aligned}
\mathbb{E}\left[y_2 \cdot x\right] &= \operatorname{signum}(c_2) \left\|y_2\right\| \frac{1 + \left\|c_2 y_2\right\|^2 - \left(1 - \frac{k^2}{\left\|y_1\right\|^2} + \left\|c_1 y_1 - c_2 y_2\right\|^2\right)}{2 \left\|c_2 y_2\right\|} \\
&= \operatorname{signum}(c_2) \left\|y_2\right\| \frac{\left\|c_2 y_2\right\|^2 + \frac{k^2}{\left\|y_1\right\|^2} - \sin^2\theta \left\|c_2 y_2\right\|^2}{2 \left\|c_2 y_2\right\|} \\
&= \operatorname{signum}(c_2) \left\|y_2\right\| \frac{\left\|c_2 y_2\right\|^2 \cos^2\theta + \frac{k^2}{\left\|y_1\right\|^2}}{2 \left\|c_2 y_2\right\|} \\
&= \operatorname{signum}(c_2) \left\|y_2\right\| \frac{1}{2} \left(\left\|c_2 y_2\right\| \cos^2\theta + \frac{|k| \cos\theta}{\left\|y_1\right\|}\right) \\
&= \operatorname{signum}(c_2) \left\|y_2\right\| \frac{1}{2} \left(\frac{|k| \left|\cos\theta\right|}{\left\|y_1\right\|} + \frac{|k| \left|\cos\theta\right|}{\left\|y_1\right\|}\right) \\
&= \operatorname{signum}(c_2) \left|k\right| \frac{\left\|y_2\right\|}{\left\|y_1\right\|} \left|\cos\theta\right| \\
&= k \frac{\left\|y_2\right\|}{\left\|y_1\right\|} \cos\theta \\
&= k \frac{y_1 \cdot y_2}{\left\|y_1\right\|^2}
\end{aligned}
$$

The last thing to do is to note that the above formulas are only valid when $\|x\| = 1$. But if $\|x\| = s$, this is equivalent to the case when $\|x\| = 1$ if we scale y_1 and y_2 by s. Scaling those two vectors by s gives us the final formulas in Theorem 2. □