EVERY REAL SYMMETRIC MATRIX IS DIAGONALIZABLE

1. abstract

This proof tries to correct my original "proof." JRN pointed out the error in my proof to me and gave me an alternative one. His proof uses the extreme value theorem for a continuous function defined on a compact domain. His proof takes a slightly different approach than my original, though. This document is an attempt to import his idea (using the extreme value theorem) to correct my original version, while staying true to the intuition/idea I had been pursuing.

2. Proposition

Theorem 1 (Finite Dimensional Spectral Theorem)**.** *For any full-rank real symmetric matrix* $\mathbf{A}$*, there exists an orthogonal (aka "rotation") matrix* $\mathbf{Q}$ *and a diagonal (aka "stretching") matrix* $\mathbf{D}$ *such that:*

$$\mathbf{A} = \mathbf{Q}\mathbf{D}\mathbf{Q}^{\mathsf{T}}$$

Equivalently: any full-rank real symmetric matrix $\mathbf{A}$ *with dimension* n *possesses* n *orthogonal (unit) eigenvectors* $\{\mathbf{u_1}, \ldots \mathbf{u_n}\}$*. That is, for suitable* λ_i*:*

$$\mathbf{A}\mathbf{u_i} = \lambda_i \mathbf{u_i}$$

Note that if we have $\mathbf{Q}$*, we can obtain a set* $\mathbf{u_i}$ *by extracting the columns. Alternatively, given* $\mathbf{u_i}$ *we can produce* $\mathbf{Q}$ *by writing them in as columns.*

3. Original Attempt

My original "proof" was inductive. I first proved a base case: that a full-rank symmetric matrix in $\mathbb{R}^{2\times 2}$ has two orthogonal eigenvectors. I next showed that, if we find even a single eigenvector for $\mathbf{A} \in \mathbb{R}^{n\times n}$, we can reduce the problem to finding $n-1$ orthogonal eigenvectors for a corresponding full-rank symmetric of $\mathbb{R}^{(n-1)\times(n-1)}$. So far, this is correct.

So the problem then becomes to find even a single eigenvector of $\mathbf{A}$. I suggested a process which doesn't quite work. I recall it here.

Our goal will be to find a rotation of basis such that $\mathbf{A}' := \mathbf{Q}^{\mathsf{T}}\mathbf{D}\mathbf{Q}$ has blank first row/column (excepting $\mathbf{A}'_{1,1}$, which cannot also be zero, by assumption that $\mathbf{A}$ is full-rank).

First, find a rotation of basis vectors $\mathbf{e_1}, \mathbf{e_2}$ such that we have a zero at $\mathbf{A}'_{2,1}$. I tacitly assume here that $\mathbf{A}_{2,2} \neq 0$.

Having done this, I want to repeat. I want to find a rotation of $\mathbf{e_1}, \mathbf{e_3}$ to zero out $\mathbf{A}'_{3,1}$. Again, assuming that $\mathbf{A}_{3,3} \neq 0$, this can be done.

The mistake was here though. By zeroing out $\mathbf{A}'_{3,1}$, we may disrupt the zeroing out that was previously performed for $\mathbf{A}'_{2,1}$. That happens exactly if $\mathbf{A}'_{2,3} \neq 0$.

Ideally this wouldn't happen. If it didn't, and we could find a rotated basis in which $\mathbf{A}'_{k,1} = 0, \forall k \neq 1$, then we can invoke a property that, in any rotated basis, a symmetric matrix stays symmetric. That is, we'd also know that $\mathbf{A}'_{1,k} = 0, \forall k \neq 1$.

That would tell us that (1) $\mathbf{e_1}$ is an eigenvector of $\mathbf{A}'$, and (2) any eigenvector $\mathbf{u}'$ of $\mathbf{A}'_{2:n,2:n}$ is also an eigenvector of $\mathbf{A}'$ (when you extend $\mathbf{u}'$ by prefixing with a zero for the first coordinate).

As near as I can tell, nothing is wrong with the proof, except that the elimination procedure I describe doesn't work.

4. Contrast With Gaussian Elimination

Let's contrast with Gaussian elimination. In Gaussian elimination, our first task is to eliminate the off-diagonal entries of the first row. One-by-one we eliminate $\mathbf{A}_{1,k}, \forall k \neq 1$. For each $k \neq 1$, we subtract an appropriate amount of the first column $\mathbf{A}_{:,1}$ from every other column $\mathbf{A}_{:,k}$.

Each *elementary (column) operation* corresponds to a shearing operation, and shearing preserves determinant. That's why Gaussian elimination is useful for calculation of the determinant.

If you repeat this row-by-row, you can turn the matrix into an upper-triangular matrix. Collecting the operations that undo this gives you the $\mathbf{LU}$ decomposition of $\mathbf{A}$. If one scales the rows/columns of $\mathbf{L}, \mathbf{U}$ appropriately so that the diagonal entries are all 1, collecting these factors as $\mathbf{D}$, then we get the $\mathbf{LDU}$ decomposition. The determinant of $\mathbf{A}$ is simply the determinant of $\mathbf{D}$. That is: the product of the diagonal entries of $\mathbf{D}$.

Of course, the $\mathbf{LDU}$ decomposition doesn't give us information about the eigenvectors, because none of these matrices corresponds to a representation of $\mathbf{A}$ in a rotated basis.

There is a very similar approach that gives a different decomposition. Rather than trying to clear out above-diagonal entries in a row one-by-one, let's try to "orthonormalize" the columns of $\mathbf{A}$. Here, we start with $\mathbf{A}' := \mathbf{A}$ and one-by-one reassign:

$$\mathbf{A}'_{:,k} -= \langle \mathbf{A}'_{:,1}, \mathbf{A}'_{:,k} \rangle \mathbf{A}'_{:,1}$$

Such an approach gives us the $\mathbf{QR}$ decomposition of $\mathbf{A}$. Note: we should also normalize the columns of $\mathbf{Q}$, combining the extracted diagonal matrix with $\mathbf{R}$. Again, the determinant may be calculated by multiplying the diagonal entries of $\mathbf{R}$.

As with the $\mathbf{LDU}$ decomposition before, the $\mathbf{QR}$ doesn't give us access to the eigenvectors.

5. Exact Algorithm For Eigenvectors?

Each column operation in the **LDU** algorithm "fixes" (that is, clears out) exactly one position in $\mathbf{A}$. Once fixed, no subsequent step disturbs that position.

Likewise, each column operation in the **QR** algorithm fixes one column in $\mathbf{A}$. That is, we make a column $\mathbf{A}_{:,i}$ orthogonal to another column $\mathbf{A}_{:,j}$. Again, no subsequent step disturbs this.

Because each step moves us one "unit" of the way toward the final result, the runtime of these algorithms is a function of the number of "units": that is, the number of elements (or columns) in $\mathbf{A}$.

Our problem for zeroing-out the first column of $\mathbf{A}$ with rotations is this. If we try to use rotating operations instead of shearing operations, subsequent operations will always disturb prior operations.

Let's look at a rotated basis which rotates only $\mathbf{e_i}, \mathbf{e_j}$; other basis vectors are left alone. But the problem is that $\mathbf{A}'$, the new representation of $\mathbf{A}$, will see *both* the columns $\mathbf{A}'_{:,i}, \mathbf{A}'_{:,j}$ *and* the corresponding rows $\mathbf{A}'_{i,:}, \mathbf{A}'_{j,:}$ changed.

How can we proceed then? One possibility is to explore an iterative procedure that *converges* to the solution, even if it never terminates. In practice we'll need something like that. Because compactness, an iterative procedure should also imply the existence of an exact solution.

A second possibility (which I will pursue here), is to merely prove the *existence* of a solution. But, rather than be too abstract, I will do this by showing that there always exists a series of steps that monotonically "improve" your matrix $\mathbf{A}$. I won't show exactly to find those steps, but my approach is suggestive of an iterative procedure.

6. The Corrected Proof: Optimization Goal

At last, let us begin. We will cast our search for a diagonal matrix $\mathbf{A}' \coloneqq \mathbf{Q}^\mathsf{T}\mathbf{D}\mathbf{Q}$ as an optimization problem. Define:

$$f(\mathbf{A}') = \sum_{i \neq j} (A_{i,j})^2$$

That is, $f(\mathbf{A}')$ is the sum of squares of off-diagonal elements. Of course, $f(\mathbf{A}') = 0$ iff $\mathbf{A}'$ is diagonal. So f is a measure of "badness" that I want to minimize. To show that $\mathbf{A}$ is always diagonalizable is to show that f achieves a minimum of exactly zero on the space of rotated representations $\mathbf{A}'$.

Instead of writing f as a function of $\mathbf{A}'$, we could write f as a function of $\mathbf{Q}$. Or even less redundantly, as a function of $\frac{n(n-1)}{2}$ angles $\theta_{i,j}, \forall i \neq j$.

This last view helps: the space of $\theta_{i,j}$ is really just points on the surface of a unit hypersphere, which is a compact set. (TODO: JRN informs me this intuition is not quite correct.) We can apply the extreme value theorem since f is continuous. That is: f achieves a minimum somewhere.

The only question remains: is $\min_{\mathbf{A}'} f(\mathbf{A}') = 0$?

7. How Rotations Change $\mathbf{A}$

We also have that f is differentiable. Thus we know that if there is a minimum for f at $\mathbf{A}'$, we must have that all derivatives here are zero. Let's consider the simplest rotation between two basis vectors $\mathbf{e_i}, \mathbf{e_j}$:

$$\begin{aligned}\mathbf{Q} \coloneqq \mathbf{e_1}\mathbf{e_1}^\mathsf{T} + \ldots + (\cos\theta\mathbf{e_i} + \sin\theta\mathbf{j})\,\mathbf{e_i}^\mathsf{T} \\ + \ldots + (-\sin\theta\mathbf{e_i} + \cos\theta\mathbf{j})\,\mathbf{e_j}^\mathsf{T} + \ldots + \mathbf{e_n}\mathbf{e_n}^\mathsf{T}\end{aligned}$$

I would like to know what happens to $f(\mathbf{Q}^\mathsf{T}\mathbf{D}\mathbf{Q})$ as we change θ near zero. Let's first study $\mathbf{A}'_{k_1,k_2}(\theta)$ where $k_1, k_2 \notin \{i, j\}$:

$$\begin{aligned}\mathbf{A}'_{k_1,k_2}(\theta) &= \mathbf{e}_{\mathbf{k_1}}^\mathsf{T}(\mathbf{Q}^\mathsf{T}\mathbf{D}\mathbf{Q})\mathbf{e}_{\mathbf{k_2}} \\ &= \mathbf{e}_{\mathbf{k_1}}^\mathsf{T}\mathbf{A}\mathbf{e}_{\mathbf{k_2}} \\ &= \mathbf{A}_{k_1,k_2}\end{aligned} \tag{1}$$

As we see, the only changed elements should be in rows (and columns) i and j. Let's next consider what happens to $\mathbf{A}'_{k,i}(\theta)$ where $k \notin \{i, j\}$.

$$\begin{aligned}\mathbf{A}'_{k,i}(\theta) &= \mathbf{e_k}^\mathsf{T}(\mathbf{Q}^\mathsf{T}\mathbf{D}\mathbf{Q})\mathbf{e_i} \\ &= \mathbf{e_k}^\mathsf{T}\mathbf{A}\,(\cos\theta\mathbf{e_i} + \sin\theta\mathbf{e_j}) \\ &= \cos\theta\mathbf{A}_{k,i} + \sin\theta\mathbf{A}_{k,j}\end{aligned} \tag{2}$$

Well, that makes sense! Naturally, because $\mathbf{A}'(\theta)$ is symmetric, we know that $\mathbf{A}'_{i,k}(\theta) = \mathbf{A}'_{k,i}(\theta)$. Let's next ask about $\mathbf{A}'_{k,j}(\theta)$. Here we go:

$$\begin{aligned}\mathbf{A}'_{k,j}(\theta) &= \mathbf{e_k}^\mathsf{T}(\mathbf{Q}^\mathsf{T}\mathbf{D}\mathbf{Q})\mathbf{e_j} \\ &= \mathbf{e_k}^\mathsf{T}\mathbf{A}\,(-\sin\theta\mathbf{e_i} + \cos\theta\mathbf{e_j}) \\ &= -\sin\theta\mathbf{A}_{k,i} + \cos\theta\mathbf{A}_{k,j}\end{aligned} \tag{3}$$

Naturally, this is just the "opposite" of what is done for $\mathbf{A}'_{k,i}$. Again, we know $\mathbf{A}'_{k,j} = \mathbf{A}'_{j,k}$. Last, we have for $\mathbf{A}'_{i,j}(\theta)$:

$$\begin{aligned}\mathbf{A}'_{i,j}(\theta) &= \mathbf{e_i}^\mathsf{T}(\mathbf{Q}^\mathsf{T}\mathbf{D}\mathbf{Q})\mathbf{e_j} \\ &= (\cos\theta\mathbf{e_i} + \sin\theta\mathbf{e_j})^\mathsf{T}\,\mathbf{A}\,(-\sin\theta\mathbf{e_i} + \cos\theta\mathbf{e_j}) \\ &= -\sin\theta\cos\theta\mathbf{A}_{i,i} + \cos^2\theta\mathbf{A}_{i,j} - \sin^2\theta\mathbf{A}_{j,i} + \sin\theta\cos\theta\mathbf{A}_{j,j} \\ &= (\sin\theta\cos\theta)\,(\mathbf{A}_{j,j} - \mathbf{A}_{i,i}) + \left(\cos^2\theta - \sin^2\theta\right)\mathbf{A}_{i,j}\end{aligned} \tag{4}$$

8. First Order Conditions

Let us now begin examining $\frac{\partial}{\partial\theta} f(\mathbf{Q}^\mathsf{T}\mathbf{D}\mathbf{Q})$. Why? Because, at a minimum, we know that these partials must be zero. We hope to show that a minimum can only occur if only all off-diagonal entries are zero.

First, let's look at:

$$\frac{\partial}{\partial\theta}\left[\left(\mathbf{A}'_{k_1,k_2}(\theta)\right)^2\right] = \frac{\partial}{\partial\theta}\left[\left(\mathbf{A}_{k_1,k_2}\right)^2\right] = 0 \quad \text{[substitution from (1)]}$$

As expected: those elements don't change regardless of θ, so they don't cause any change in f. Next, let's look at:

$$\begin{aligned}
\frac{\partial}{\partial\theta}\left[\left(\mathbf{A}'_{k,i}(\theta)\right)^2\right] &= \frac{\partial}{\partial\theta}\left[\left(\cos\theta\mathbf{A}_{k,i} + \sin\theta\mathbf{A}_{k,j}\right)^2\right] && \text{[substitution from (2)]} \\
&= 2\left(\cos\theta\mathbf{A}_{k,i} + \sin\theta\mathbf{A}_{k,j}\right) \\
&\qquad \tfrac{\partial}{\partial\theta}\left[\cos\theta\mathbf{A}_{k,i} + \sin\theta\mathbf{A}_{k,j}\right] && \text{[chain rule]} \\
&= 2\left(\cos\theta\mathbf{A}_{k,i} + \sin\theta\mathbf{A}_{k,j}\right) \\
&\qquad \left(-\sin\theta\mathbf{A}_{k,i} + \cos\theta\mathbf{A}_{k,j}\right) && (5)
\end{aligned}$$

We want to evaluate this derivative at $\theta = 0$, since we want to see what happens when we make small changes to $\mathbf{A}$. Thus, we obtain:

$$\frac{\partial}{\partial\theta}\left[\left(\mathbf{A}'_{k,i}(\theta)\right)^2\right](0) = 2\left(\mathbf{A}_{k,i}\right)\left(\mathbf{A}_{k,j}\right) \tag{6}$$

Let us do the corresponding calculation for $\mathbf{A}'_{k,j}$:

$$\begin{aligned}
\frac{\partial}{\partial\theta}\left[\left(\mathbf{A}'_{k,j}(\theta)\right)^2\right] &= \frac{\partial}{\partial\theta}\left(-\sin\theta\mathbf{A}_{k,i} + \cos\theta\mathbf{A}_{k,j}\right)^2 && \text{[substitution from (3)]} \\
&= 2\left(-\sin\theta\mathbf{A}_{k,i} + \cos\theta\mathbf{A}_{k,j}\right) \\
&\qquad \tfrac{\partial}{\partial\theta}\left[-\sin\theta\mathbf{A}_{k,i} + \cos\theta\mathbf{A}_{k,j}\right] \\
&= 2\left(-\sin\theta\mathbf{A}_{k,i} + \cos\theta\mathbf{A}_{k,j}\right) \\
&\qquad \left(-\cos\theta\mathbf{A}_{k,i} - \sin\theta\mathbf{A}_{k,j}\right) && (7)
\end{aligned}$$

Again, substituting $\theta = 0$, we obtain:

$$\frac{\partial}{\partial\theta}\left[\left(\mathbf{A}'_{k,j}(\theta)\right)^2\right](0) = 2\left(\mathbf{A}_{k,j}\right)\left(-\mathbf{A}_{k,i}\right) \tag{8}$$

Do you see what has happened here? The change to f due to change in $\mathbf{A}'_{k,i}$ (equation 6) be canceled out by the change to f due to change in $\mathbf{A}'_{k,j}$ (equation 8). (And likewise for $\mathbf{A}'_{i,k}$ and $\mathbf{A}'_{j,k}$.)

Thus, when rotating between $\mathbf{e_i}, \mathbf{e_j}$, the only meaningful difference (if any) to f is going to come from change at positions $\mathbf{A}'_{i,j}, \mathbf{A}'_{j,i}$.

9. Calculation Of $\frac{\partial}{\partial\theta}\left[\left(\mathbf{A}'_{i,j}(\theta)\right)^2\right]$

Let's do this. First we simply apply the chain rule.

$$\frac{\partial}{\partial\theta}\left[\left(\mathbf{A}'_{i,j}(\theta)\right)^2\right] = 2\mathbf{A}'_{i,j}(\theta)\frac{\partial}{\partial\theta}\left[\mathbf{A}'_{i,j}(\theta)\right] \tag{9}$$

Next, let's recall our formula for $\mathbf{A}'_{i,j}$.

$$\mathbf{A}'_{i,j}(\theta) = (\sin\theta\cos\theta)(\mathbf{A}_{j,j} - \mathbf{A}_{i,i}) + (\cos^2\theta - \sin^2\theta)\mathbf{A}_{i,j} \qquad \text{[i.e., formula (4)]}$$

The chain rule wants us to calculate the derivative for this:

$$\begin{aligned}\frac{\partial}{\partial\theta}\left[\mathbf{A}'_{i,j}(\theta)\right] &= \frac{\partial}{\partial\theta}\Big[(\sin\theta\cos\theta)(\mathbf{A}_{j,j} - \mathbf{A}_{i,i}) + (\cos^2\theta - \sin^2\theta)\mathbf{A}_{i,j}\Big] \\ &= (\cos^2\theta - \sin^2\theta)(\mathbf{A}_{j,j} - \mathbf{A}_{i,i}) \\ &\qquad + (-2\cos\theta\sin\theta - 2\sin\theta\cos\theta)\mathbf{A}_{i,j} \\ &= (\cos^2\theta - \sin^2\theta)(\mathbf{A}_{j,j} - \mathbf{A}_{i,i}) - 4\cos\theta\sin\theta\mathbf{A}_{i,j}\end{aligned} \tag{10}$$

Next, we must evaluate $\mathbf{A}'_{i,j}(0)$ and $\frac{\partial}{\partial\theta}\mathbf{A}'_{i,j}(0)$.

$$\begin{aligned}\mathbf{A}'_{i,j}(0) &= (\sin 0\cos 0)(\mathbf{A}_{j,j} - \mathbf{A}_{i,i}) + (\cos^2 0 - \sin^2 0)\mathbf{A}_{i,j} \\ &= \mathbf{A}_{i,j}\end{aligned} \tag{11}$$

and

$$\begin{aligned}\frac{\partial}{\partial\theta}\left[\mathbf{A}'_{i,j}(0)\right] &= (\cos^2 0 - \sin^2 0)(\mathbf{A}_{j,j} - \mathbf{A}_{i,i}) - 4\left(\cos 0\sin 0\right)\mathbf{A}_{i,j} \\ &= (\mathbf{A}_{j,j} - \mathbf{A}_{i,i})\end{aligned} \tag{12}$$

We may now evaluate the partial for $(\mathbf{A}'_{i,j})^2$ at $\theta = 0$ by substitution:

$$\begin{aligned}\frac{\partial}{\partial\theta}\left[\left(\mathbf{A}'_{i,j}(0)\right)^2\right] &= 2\mathbf{A}'_{i,j}(0)\frac{\partial}{\partial\theta}\left[\mathbf{A}'_{i,j}(0)\right] \\ &= 2\mathbf{A}_{i,j}(\mathbf{A}_{j,j} - \mathbf{A}_{i,i})\end{aligned} \tag{13}$$

Of course, we may use this to evaluate $\frac{\partial}{\partial\theta} f(0)$, as has been our goal:

$$\begin{aligned}\frac{\partial}{\partial\theta}\left[f(\theta)\right](0) &= \frac{\partial}{\partial\theta}\left[\left(\mathbf{A}'_{i,j}(0)\right)^2\right] + \frac{\partial}{\partial\theta}\left[\left(\mathbf{A}'_{j,i}(0)\right)^2\right] \\ &= 2\frac{\partial}{\partial\theta}\left[\left(\mathbf{A}'_{i,j}(0)\right)^2\right] \\ &= 4\mathbf{A}_{i,j}(\mathbf{A}_{j,j} - \mathbf{A}_{i,i}) \end{aligned} \tag{14}$$

10. Summary Thus Far

Let's sum up. We used these facts:

- there is no change to entries outside the ith and jth rows/columns,
- the effect on f of changes at positions (k, i) cancel with the effect of changes at (k, j) (for $k \notin \{i, j\}$),
- likewise for positions (i, k) and (j, k),
- changes at (i, i) and (j, j) never count toward the error because they are on-diagonal,
- that leaves only positions to consider (i, j) and (j, i).

We may finally say that if $\mathbf{A}$ minimizes f, then:

$$\begin{aligned} &\frac{\partial}{\partial\theta}\left[f(0)\right] = 0 \\ \Rightarrow\ &4\mathbf{A}_{i,j}(\mathbf{A}_{j,j} - \mathbf{A}_{i,i}) = 0 \\ \Rightarrow\ &\mathbf{A}_{i,j}(\mathbf{A}_{j,j} - \mathbf{A}_{i,i}) = 0 \end{aligned} \tag{15}$$

Which of course implies:

$$\mathbf{A}_{i,j} = 0 \quad \text{or} \quad (\mathbf{A}_{j,j} - \mathbf{A}_{i,i}) = 0$$

The first condition is exactly what we want: it would impose a requirement that the matrix $\mathbf{A}$ have zero elements off the diagonal in order to minimize f. That is: f would achieve a minimum only for a diagonal matrix. Since we know f indeed achieves a minimum, it must be that $\mathbf{A}$ is diagonalizable.

But the second condition is an "escape hatch." It says that the entry off the diagonal needn't be zero so long as $\mathbf{A}_{j,j} = \mathbf{A}_{i,i}$. Uh-oh?

11. Second Order Conditions

To check out what is happening we should investigate the second order conditions. In particular: we know that f has a maximum, for the same reason it has a minimum: the extreme value theorem!

So let's go ahead and examine the second partial.

$$\begin{aligned}
\frac{\partial}{\partial\theta^2}\left[\left(\mathbf{A}'_{i,j}(\theta)\right)^2\right] &= \frac{\partial}{\partial\theta}\left[\frac{\partial}{\partial\theta}\left[\left(\mathbf{A}'_{i,j}(\theta)\right)^2\right]\right] \\
&= \frac{\partial}{\partial\theta}\left[2\mathbf{A}'_{i,j}(\theta)\frac{\partial}{\partial\theta}\left[\mathbf{A}'_{i,j}(\theta)\right]\right] && \text{[chain rule]} \\
&= 2\left[\left(\frac{\partial}{\partial\theta}\left[\mathbf{A}'_{i,j}(\theta)\right]\right)^2 + \mathbf{A}'_{i,j}(\theta)\frac{\partial}{\partial\theta^2}\left[\mathbf{A}'_{i,j}(\theta)\right]\right] && \text{[product rule]} \\
&= 2\left[\mathbf{A}'_{i,j}(\theta)\frac{\partial}{\partial\theta^2}\left[\mathbf{A}'_{i,j}(\theta)\right]\right] && (16)
\end{aligned}$$

The last step is justified because we wouldn't bother with the second order test unless $\frac{\partial}{\partial\theta}\left[\mathbf{A}'_{i,j}(\theta)\right] = 0$. We've already calculated $\mathbf{A}'_{i,j}(\theta)$ previously, so let's focus on the second derivative: $\frac{\partial}{\partial\theta^2}\left[\mathbf{A}'_{i,j}(\theta)\right]$.

$$\begin{aligned}
\frac{\partial}{\partial\theta^2}\left[\mathbf{A}'_{i,j}(\theta)\right] &= \frac{\partial}{\partial\theta}\left[\frac{\partial}{\partial\theta}\left[\mathbf{A}'_{i,j}(\theta)\right]\right] \\
&= \frac{\partial}{\partial\theta}\left[(\cos^2\theta - \sin^2\theta)(\mathbf{A}_{j,j} - \mathbf{A}_{i,i}) - 4\cos\theta\sin\theta\mathbf{A}_{i,j}\right] && \text{[formula 10]} \\
&= \frac{\partial}{\partial\theta}\left[-4\cos\theta\sin\theta\mathbf{A}_{i,j}\right] && \text{[by presumption } \mathbf{A}_{i,i} = \mathbf{A}_{j,j}\text{]}\ (17)
\end{aligned}$$

Notice that I've used our presumption that $\mathbf{A}_{i,i} = \mathbf{A}_{j,j}$, since otherwise the first-order condition would have already ensured that an off-diagonal partial is zero. We continue:

$$\begin{aligned}
\frac{\partial}{\partial\theta^2}\left[\mathbf{A}'_{i,j}(\theta)\right] &= \frac{\partial}{\partial\theta}\left[-4\cos\theta\sin\theta\mathbf{A}_{i,j}\right] \\
&= -4\left(\cos^2\theta - \sin^2\theta\right)\mathbf{A}_{i,j} && (18)
\end{aligned}$$

And we may now plug formula 18 into formula 16.

$$\begin{aligned}
\frac{\partial}{\partial\theta^2}\left[\left(\mathbf{A}'_{i,j}(\theta)\right)^2\right] &= 2\left[\mathbf{A}'_{i,j}(\theta)\frac{\partial}{\partial\theta^2}\left[\mathbf{A}'_{i,j}(\theta)\right]\right] \\
&= 2\mathbf{A}'_{i,j}(\theta)\Big(-4\left(\cos^2\theta - \sin^2\theta\right)\mathbf{A}_{i,j}\Big) \\
&= -8\mathbf{A}'_{i,j}(\theta)\left(\cos^2\theta - \sin^2\theta\right)\mathbf{A}_{i,j} && (19)
\end{aligned}$$

There is nothing left to do but evaluate at $\theta = 0$:

$$\begin{aligned}\frac{\partial}{\partial\theta^2}\left[\left(\mathbf{A}'_{i,j}(0)\right)^2\right] &= -8\left(\mathbf{A}'_{i,j}(0)\right)\left(\cos^2 0 - \sin^2 0\right)\mathbf{A}_{i,j} \\ &= -8\mathbf{A}_{i,j}\mathbf{A}_{i,j} \\ &= -8\left(\mathbf{A}_{i,j}\right)^2 \end{aligned} \tag{20}$$

And there you have it. This second derivative must always be negative, since it is -8 (always negative) times $(\mathbf{A}_{i,j})^2$ (always positive). Note that we assumed $\mathbf{A}_{i,j} \neq 0$, otherwise there was no reason to go down this road. That saves us from any further derivative testing.

12. Conclusion

Let's recap what we have done.

(1) We proved (elsewhere, previously) that rotation of basis keeps symmetric matrices symmetric.
(2) We then defined an "error" function f to minimize. The error is minimal (zero) precisely when all off-diagonal entries are zeroed.
(3) Since the space of rotations of $\mathbf{A}$ is compact, and because f is continuous, we know it achieves a minimum in this space.
(4) We were unsure whether the minimum of f on the space of rotations is actually zero.
(5) We use the first derivative test to discover that for any optimal $\mathbf{A}$, either $\mathbf{A}_{i,j} = \mathbf{A}_{j,i} = 0$ (as desired) OR $\mathbf{A}_{i,i} = \mathbf{A}_{j,j}$ (leaving open the possibility that $\mathbf{A}_{i,j} \neq 0$).
(6) We then examined what the second derivative would be at a first-order critical point, in the undesired case that $\mathbf{A}_{i,j} \neq 0$, and thus necessarily $\mathbf{A}_{i,i} = \mathbf{A}_{j,j}$.
(7) The second derivative test showed us this must be a local *maximum*. Thus we can preclude this scenario at a global minimum.
(8) We have foreclosed any escape. A global minimum must have $\mathbf{A}_{i,j} = \mathbf{A}_{j,i} = 0$, else it would be possible to improve.

Is this proof really any better than the much more succinct proof JRN gave me? For me, this proof accords with an intuition that you can always iteratively keep improving your matrix by doing little rotations. It shows that you can never get stuck. Indeed, you can use gradient descent if you want. So this proof is much more clearly tied to a notion of *procedure*, which is certainly related to intuition.

Second, the proof says that, if you believe things work in two dimensions, you can just keep picking pairs of basis vectors to fix, and do your best. Remember, you might "disrupt" previous work, but to the extent you disrupt previous work, you're helping other work. That's what we know from the cancelation of the $\mathbf{A}_{k,i}, \mathbf{A}_{k,j}$ terms.

Probably that would have been a simpler proof right there. You have this iterative process now, and it must converge to an answer, so the answer must exist…

One goal was to limit the magical use of symmetry. Here it is still used somewhat: you use symmetry when you presume that rotation of basis preserves symmetry. Likewise, you use symmetry when you assume that fixing things in one subspace won't have an overall negative impact.