Problem Set 3 Solutions

Problem 1. *How changing the inner product affects* $\mathbb{R}^2$

Remember that vector spaces on their own are merely *algebraic* objects. If we want to impose any geometric features, we need to give them either a norm or an inner product. Since there is such a variety in the inner products that we can use, we can expect a variety of geometries for the same vector space.

(a) First, show that the bracket

$$\langle \vec{\mathbf{v}}, \vec{\mathbf{w}} \rangle = \vec{\mathbf{v}}^T \begin{bmatrix} 4 & -2 \\ -2 & 9 \end{bmatrix} \vec{\mathbf{w}}$$

defines an inner product on $\mathbb{R}^2$.

We show that $\langle \cdot, \cdot \rangle$ is (i) bilinear, (ii) symmetric, and (iii) positive-definite:

(i) Let $\vec{\mathbf{u}} = (u_1, u_2)^T, \vec{\mathbf{v}} = (v_1, v_2)^T, \vec{\mathbf{w}} = (w_1, w_2)^T \in \mathbb{R}^2$ and $a, b \in \mathbb{R}$. First, consider that

$$\langle \vec{\mathbf{v}}, \vec{\mathbf{w}} \rangle = [v_1 \ v_2] \begin{bmatrix} 4 & -2 \\ -2 & 9 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = 4v_1w_1 - 2v_1w_2 - 2v_2w_1 + 9v_2w_2.$$

Since

$$a\vec{\mathbf{u}} + b\vec{\mathbf{v}} = \begin{bmatrix} au_1 + bv_1 \\ au_2 + bv_2 \end{bmatrix},$$

we can see that

$$\langle a\vec{\mathbf{u}} + b\vec{\mathbf{v}}, \vec{\mathbf{w}} \rangle = 4(au_1 + bv_1)w_1 - 2(au_1 + bv_1)w_2 - 2(au_2 + bv_2)w_1 + 9(au_2 + bv_2)w_2$$

$$= (4au_1w_1 - 2au_1w_2 - 2au_2w_1 + 9au_2w_2) + (4av_1w_1 - 2av_1w_2 - 2av_2w_1 + 9av_2w_2)$$

$$= a\langle \vec{\mathbf{u}}, \vec{\mathbf{v}} \rangle + b\langle \vec{\mathbf{v}}, \vec{\mathbf{w}} \rangle.$$

Therefore, $\langle \cdot, \cdot \rangle$ is linear in its first component. If we show that $\langle \cdot, \cdot \rangle$ is symmetric, then we can omit showing linearity in the second component; hence, we move on to showing symmetry.

(ii) Let $\vec{\mathbf{v}}, \vec{\mathbf{w}} \in \mathbb{R}^2$ be as in part (i) above. Then we see that

$$\langle \vec{\mathbf{v}}, \vec{\mathbf{w}} \rangle = 4v_1w_1 - 2v_1w_2 - 2v_2w_1 + 9v_2w_2 = 4w_1v_1 - 2w_1v_2 - 2w_2v_1 + 9w_2v_2 = \langle \vec{\mathbf{w}}, \vec{\mathbf{v}} \rangle.$$

Therefore, $\langle \cdot, \cdot \rangle$ is symmetric. This also establishes the linearity of the second component.

(iii) Consider first that

$$\langle \vec{\mathbf{u}}, \vec{\mathbf{u}} \rangle = 4u_1^2 - 4u_1u_2 + 9u_2^2.$$

Then, after completing the square on u_1, we see that

$$\langle \vec{\mathbf{u}}, \vec{\mathbf{u}} \rangle = (2u_1 - u_2)^2 + (\sqrt{8}u_2)^2 \geq 0.$$

Moreover, this expression is equal to zero if and only if both $2u_1 - u_2 = 0$ and $\sqrt{8}u_2 = 0$. This last pair of conditions implies that $u_1 = u_2 = 0$. Therefore, $\langle \cdot, \cdot \rangle$ is positive-definite.

This completes the proof that indeed $\langle\cdot,\cdot\rangle$ is an inner product on $\mathbb{R}^2$.

(b) In the standard Euclidean inner product space $(\mathbb{R}^2,\cdot)$ that uses the dot product $\cdot$, the vectors $\hat{\mathbf{i}} = (1,0)^T$ and $\hat{\mathbf{j}} = (0,1)^T$ have unit length. What is $||\hat{\mathbf{i}}||$ and $||\hat{\mathbf{j}}||$ in $(\mathbb{R}^2,\langle\cdot,\cdot\rangle)$? What multiples of $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ actually have unit length in $(\mathbb{R}^2,\langle\cdot,\cdot\rangle)$?

We simply compute

$$||\hat{\mathbf{i}}||^2 = \langle\hat{\mathbf{i}},\hat{\mathbf{i}}\rangle = 4(1)^2 - 4(1)(0) + 9(0)^2 = 4$$

and

$$||\hat{\mathbf{j}}||^2 = \langle\hat{\mathbf{j}},\hat{\mathbf{j}}\rangle = 4(0)^2 - 4(0)(1) + 9(1)^2 = 9$$

via the calculations in part (iii) from part (a) above. This means that, although they are unit vectors in $(\mathbb{R}^2,\cdot)$, the vectors $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ have lengths $||\hat{\mathbf{i}}|| = 2$ and $||\hat{\mathbf{j}}|| = 3$ in $(\mathbb{R}^2,\langle\cdot,\cdot\rangle)$. Furthermore, the vectors $\frac{1}{2}\hat{\mathbf{i}}$ and $\frac{1}{3}\hat{\mathbf{j}}$ have unit length in $(\mathbb{R}^2,\langle\cdot,\cdot\rangle)$.

(c) In $(\mathbb{R}^2,\cdot)$, $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ are orthogonal. Are they orthogonal in $(\mathbb{R}^2,\langle\cdot,\cdot\rangle)$? Find all vectors that are orthogonal to $\hat{\mathbf{i}}$.

Recall that the "angle" θ between vectors $\vec{\mathbf{u}}$ and $\vec{\mathbf{v}}$ in an inner product space $(V,\langle\cdot,\cdot\rangle)$ is defined by the ratio

$$\cos\theta = \frac{\langle\vec{\mathbf{u}},\vec{\mathbf{v}}\rangle}{||\vec{\mathbf{u}}||\,||\vec{\mathbf{v}}||}.$$

Therefore, we must compute each item in the ratio. We calculate these for $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ to be

$$||\hat{\mathbf{i}}|| = 2, \quad ||\hat{\mathbf{j}}|| = 3,$$
$$\text{and} \quad \langle\hat{\mathbf{i}},\hat{\mathbf{j}}\rangle = 4(1)(0) - 2(1)(1) - 2(0)(0) + 9(0)(1) = -2.$$

The fact that $\langle\hat{\mathbf{i}},\hat{\mathbf{j}}\rangle \neq 0$ is enough to demonstrate that $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ are not orthogonal, but we can continue the computation to show that

$$\theta = \cos^{-1}\left(\frac{-2}{(2)(3)}\right) = \cos^{-1}\left(-\frac{1}{3}\right) \approx 1.91 \text{ radians} \approx 109.47^\circ.$$

(d) The equation for the unit circle in $(\mathbb{R}^2,\cdot)$ is $x^2+y^2=1$. What is the equation for the unit circle in $(\mathbb{R}^2,\langle\cdot,\cdot\rangle)$? What does this unit "circle" actually look like to us?

We are looking for vectors $\vec{\mathbf{v}} = (x,y)^T \in \mathbb{R}^2$ such that

$$||\vec{\mathbf{v}}||^2 = \langle\vec{\mathbf{v}},\vec{\mathbf{v}}\rangle = 1.$$

The expression on the left-hand side simplifies the equation to

$$4x^2 - 4xy + 9y^2 = 1.$$

We can either complete the square to show this is an ellipse, or we can plot this is a computer to convince ourselves. See the figure below.

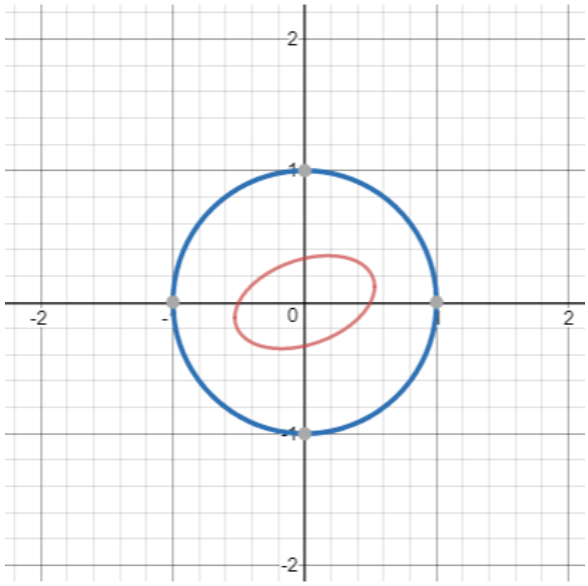

Figure 1: The unit circles in $(\mathbb{R}^2, \cdot)$, colored blue, and in $(\mathbb{R}^2, \langle \cdot, \cdot \rangle)$, colored red.

Note: In lecture we showed that any two norms on a finite-dimensional vector space are *equivalent*. Given what we've just seen in this exercise, we can definitively say that this notion of equivalence does not take into account the geometry of the inner product space. In fact, the equivalence of norms is more of a statement about the *topology* of the space rather than the geometry.

Problem 2. *Direct computation with the 1-norm*

It is often the case that we use the p-norms both on vectors in $\mathbb{R}^n$ and on matrices in $\mathcal{M}_{m\times n}(\mathbb{R})$. Because of how frequently they are used in computational practice, we should (1) show that these norms are equivalent and (2) come up with some simple formulas that allow us to compute these norms quickly. In particular, we can do this for the 1-, 2-, and ∞-norms quite easily.

(a) We showed in class that the 2-norm and the ∞-norm were equivalent on $\mathbb{R}^n$. Show that the 1-norm and 2-norm are also equivalent on $\mathbb{R}^n$.

First, we should determine what the unit circle in $(\mathbb{R}^n, \|\cdot\|_1)$ looks like. We can get an intuitive sense of this object in $\mathbb{R}^n$ by studying it in $\mathbb{R}^2$ first. For a vector $\vec{\mathbf{v}} = (x, y)^T \in \mathbb{R}^2$, we can compute

$$\|\vec{\mathbf{v}}\|_1 = |x| + |y| = 1$$

as the equation for the unit circle in $(\mathbb{R}^2, \|\cdot\|_1)$. We can solve this first as

$$|y| = 1 - |x|$$

and then as

$$y = \pm(1 - |x|).$$

Graphing each of these functions separately shows that the unit "circle" is actually a *diamond* to our Euclidean eyes. See the figure below. Extending this construction to higher dimensions yields higher-dimensional diamond

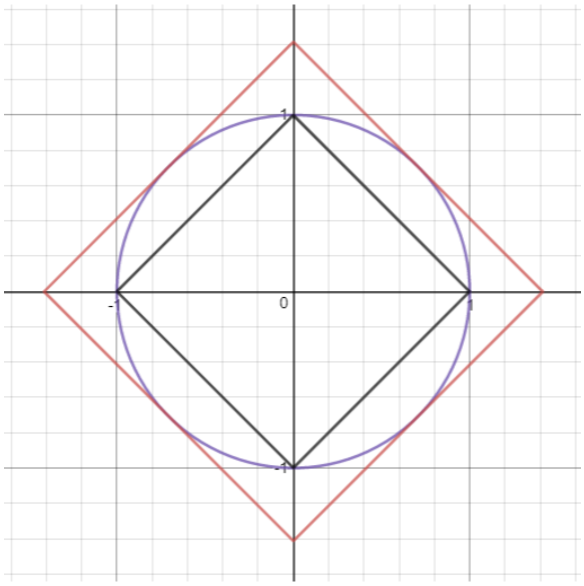

Figure 2: The unit circle of $(\mathbb{R}^2, \|\cdot\|_2)$, colored blue, inscribed (black) and circumscribed (red) by circles of $(\mathbb{R}^2, \|\cdot\|_1)$.

shapes, all of whose corners lie on $\pm\vec{\mathbf{e}}_i$ for some $1 \leq i \leq n$. (In $\mathbb{R}^3$, this is an octahedron - a d8 in tabletop RPGs.)

In this figure, we can easily intuit that $\|\vec{\mathbf{v}}\|_1 \leq \|\vec{\mathbf{v}}\|_2$ for any $\mathbb{R}^n$ so that $m = 1$ in our lower bound. We now aim to find $M > 0$ such that $\|\vec{\mathbf{v}}\|_2 \leq M\|\vec{\mathbf{v}}\|_1$.

Notice in the figure that the circumscribed diamond intersects the circle at the center of its edge. This location is either

$$\begin{bmatrix} \pm\frac{1}{\sqrt{2}} \\ \pm\frac{1}{\sqrt{2}} \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} \pm\frac{1}{\sqrt{2}} \\ \mp\frac{1}{\sqrt{2}} \end{bmatrix}.$$

These vectors all have length $\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} = \frac{2}{\sqrt{2}} = \sqrt{2}$ in $(\mathbb{R}^2, \|\cdot\|_1)$. In $\mathbb{R}^2$ then, we have $M = \sqrt{2}$. We can see this generalize to $M = \sqrt{n}$ for $\mathbb{R}^n$. This shows that

$$\|\vec{\mathbf{v}}\|_1 \leq \|\vec{\mathbf{v}}\|_2 \leq \sqrt{n}\|\vec{\mathbf{v}}\|_1$$

so that $\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent norms in $\mathbb{R}^n$.

(b) We have now shown that the 2-norm is equivalent to both the 1- and ∞-norms on $\mathbb{R}^n$. Show how this implies that the 1-norm is equivalent to the ∞-norm on $\mathbb{R}^n$.

We have shown in class that

$$\frac{1}{\sqrt{n}}\|\vec{\mathbf{v}}\|_\infty \leq \|\vec{\mathbf{v}}\|_2 \leq \|\vec{\mathbf{v}}\|_\infty$$

and above that

$$\|\vec{\mathbf{v}}\|_1 \leq \|\vec{\mathbf{v}}\|_2 \leq \sqrt{n}\|\vec{\mathbf{v}}\|_1$$

for all $\vec{\mathbf{v}} \in \mathbb{R}^n$. Our goal now is to find constants $m, M > 0$ such that

$$m\|\vec{\mathbf{v}}\|_\infty \leq \|\vec{\mathbf{v}}\|_1 \leq M\|\vec{\mathbf{v}}\|_\infty$$

for all $\vec{\mathbf{v}} \in \mathbb{R}^n$.
We then see that, for any $\vec{\mathbf{v}} \in \mathbb{R}^n$,

$$\frac{1}{\sqrt{n}}\|\vec{\mathbf{v}}\|_\infty \leq \|\vec{\mathbf{v}}\|_2 \leq \sqrt{n}\|\vec{\mathbf{v}}\|_1 \leq \sqrt{n}\|\vec{\mathbf{v}}\|_2 \leq \sqrt{n}\|\vec{\mathbf{v}}\|_\infty,$$

where the middle inequality is $\|\vec{\mathbf{v}}\|_1 \leq \|\vec{\mathbf{v}}\|_2$ scaled by $\sqrt{n}$ and the final inequality is $\|\vec{\mathbf{v}}\|_2 \leq \|\vec{\mathbf{v}}\|_\infty$ scaled by $\sqrt{n}$ as well.
Therefore, we can set $m = \frac{1}{n}$ and $M = 1$ to show that $\|\cdot\|_1$ and $\|\cdot\|_\infty$ are indeed equivalent.

(c) We came across a simple way of explaining the ∞-norm on $\mathcal{M}_{m\times n}(\mathbb{R})$ as the maximum absolute row sum of a matrix. Show that for $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, the 1-norm $\|A\|_1$ is the maximum absolute column sum of A. [*Hint*: Try to mimic the proof that $\|A\|_\infty$ is the maximum absolute row sum of A.]

For ease of discussion, we first define

$$s_j = \sum_{i=1}^{m} |A_{ij}| \quad \text{and} \quad s = \max_{1\leq j\leq n} s_j.$$

Our proof now comes in two steps: (1) bound $\|A\| \leq s$ and (2) find a (1-norm) unit vector $\hat{\mathbf{u}}$ such that $\|A\vec{\mathbf{v}}\|_1 = s$. This implies that $s \leq \|A\|_1 \leq s$, which establishes the assertion.
First, notating $\vec{\mathbf{v}} = (v_j) \in \mathbb{R}^n$, we compute

$$\begin{aligned}
\|A\|_1 &= \max_{\|\vec{\mathbf{v}}\|_1=1} \|A\vec{\mathbf{v}}\|_1 \\
&= \max_{\|\vec{\mathbf{v}}\|_1=1} \left\| \left(\sum_{j=1}^{n} A_{ij}v_j \right) \right\|_1 \\
&= \max_{\|\vec{\mathbf{v}}\|_1=1} \sum_{i=1}^{m} \left| \sum_{j=1}^{n} A_{ij}v_j \right| \\
&\leq \max_{\|\vec{\mathbf{v}}\|_1=1} \sum_{i=1}^{m} \sum_{j=1}^{n} |A_{ij}v_j| \\
&= \max_{\|\vec{\mathbf{v}}\|_1=1} \sum_{i=1}^{m} \sum_{j=1}^{n} |A_{ij}||v_j| \\
&= \max_{\|\vec{\mathbf{v}}\|_1=1} \sum_{j=1}^{n} \sum_{i=1}^{m} |A_{ij}||v_j| \\
&= \max_{\|\vec{\mathbf{v}}\|_1=1} \sum_{j=1}^{n} \left(|v_j| \sum_{i=1}^{m} |A_{ij}| \right) \\
&\leq \max_{\|\vec{\mathbf{v}}\|_1=1} \sum_{j=1}^{n} (|v_j|s) \\
&\leq s \max_{\|\vec{\mathbf{v}}\|_1=1} \sum_{j=1}^{n} |v_j| \\
&= s \max_{\|\vec{\mathbf{v}}\|_1=1} \|\vec{\mathbf{v}}\|_1 = s \cdot 1 = s.
\end{aligned}$$

This establishes the first step.

We now exhibit a (1-norm) unit vector $\hat{\mathbf{u}}$ that achieves this upper bound. Choose J to be any one of the indices such that $s_J = s$. Define $\hat{\mathbf{u}} = \vec{\mathbf{e}}_J$, the J^{th} standard basis vector of $\mathbb{R}^n$. Then consider that

$$\|A\hat{\mathbf{u}}\|_1 = \|A\vec{\mathbf{e}}_J\|_1 = \|(A_{iJ})\|_1 = \sum_{i=1}^{m} |A_{iJ}| = s_J = s.$$

This implies that $s \leq \|A\|_1$ as well, and we have established the second step. This completes the proof.

Problem 3. *Coefficients of positive-definite matrices*

We have seen many examples of positive-definite matrices in class, and we have determined that it is hard to tell when an arbitrary symmetric matrix is positive-definite. We might wonder if there is a simple "formula" in terms of the coefficients of the matrix that tells us when it is positive-definite, much like the discriminant of a quadratic polynomial indicating the existence of real roots of the polynomial. There is a reason we haven't seen such a formula...

(a) Let's define

$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

to be the form of an arbitrary, symmetric 2×2 matrix. Find a collection of inequalities of a, b, and c that tells us when A is positive-definite.

Let $\vec{\mathbf{v}} = (x, y)^T \in \mathbb{R}^2$ be an arbitrary vector. We aim to show that the associated quadratic form

$$\langle \vec{\mathbf{v}}, \vec{\mathbf{v}} \rangle = \vec{\mathbf{v}}^T A \vec{\mathbf{v}} = ax^2 + 2bxy + cy^2 \geq 0$$

for the correct choices of a, b, and c.

It can't be the case that $a = 0$ or $c = 0$. If this were true, then $\langle \vec{\mathbf{e}}_1, \vec{\mathbf{e}}_1 \rangle = \langle \vec{\mathbf{e}}_2, \vec{\mathbf{e}}_2 \rangle = 0$ so that $\langle \cdot, \cdot \rangle$ can't be definite (let alone positive-definite). In a similar fashion, it can't be the case that $a < 0$ or $c < 0$, lest $\langle \cdot, \cdot \rangle$ is either negative or indefinite. We proceed knowing that at least $a > 0$.

If we complete the square on ax^2, we find

$$\langle \vec{\mathbf{v}}, \vec{\mathbf{v}} \rangle = ax^2 + 2bxy + cy^2 = \left(\sqrt{a}x + \frac{b}{\sqrt{a}}y\right)^2 + \left(c - \frac{b^2}{a}\right)y^2.$$

In order for $\langle \vec{\mathbf{v}}, \vec{\mathbf{v}} \rangle$ to be non-negative, we require

$$Z = c - \frac{b^2}{a} > 0.$$

In that case, we would begin with

$$\begin{cases} \sqrt{a}x + \frac{b}{\sqrt{a}}y & = 0 \\ \sqrt{Z}y & = 0 \end{cases}$$

so that the second equation implies that $y = 0$, which in turn implies $x = 0$. Therefore, $\langle \cdot, \cdot \rangle$ is positive-definite in this case.

If $Z = 0$, then $\langle \cdot, \cdot \rangle$ would be positive-semidefinite instead of positive-definite (e.g., try using $\vec{\mathbf{v}} = (a^{-1/2}, -a^{1/2}b^{-1})^T$).

Therefore, we only require

$$a > 0 \quad \text{and} \quad Z = c - \frac{b^2}{a} > 0.$$

Remark: Notice that these conditions are actually equivalent to $tr(A) > 0$ and $det(A) > 0$.

(b) Do the same for the arbitrary, symmetric 3×3 matrix

$$A = \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}.$$

Let $\vec{\mathbf{v}} = (x, y, z)^T \in \mathbb{R}^3$ be an arbitrary vector. We aim to show that the associated quadratic form

$$\langle \vec{\mathbf{v}}, \vec{\mathbf{v}} \rangle = ax^2 + 2bxy + 2cxz + dy^2 + 2eyz + fz^2 \geq 0$$

for the correct choices of a, b, c, d, e, and f.

Similar to part (a), it cannot be the case that $a = 0$, $d = 0$, or $f = 0$. This would allow for $\langle \vec{\mathbf{e}}_i, \vec{\mathbf{e}}_i \rangle = 0$ for $i = 1, 2, 3$ so that $\langle \cdot, \cdot \rangle$ cannot be definite (let alone positive-definite). Also, lest $\langle \cdot, \cdot \rangle$ is negative or indefinite, we can't have $a < 0$, $b < 0$, or $c < 0$. We proceed knowing that at least $a > 0$.

As before, we complete the square on ax^2 first. This can be done as

$$\langle \vec{\mathbf{v}}, \vec{\mathbf{v}} \rangle = \left(\sqrt{a}x + \frac{b}{\sqrt{a}}y + \frac{c}{\sqrt{a}}z \right)^2 + \left(d - \frac{b^2}{a} \right) y^2 + 2 \left(e - \frac{bc}{a} \right) yz + \left(f - \frac{c^2}{a} \right) z^2.$$

Similar to part (a), we require

$$Z_1 = d - \frac{b^2}{a} > 0.$$

Again, if $Z_1 = 0$, $\langle \cdot, \cdot \rangle$ is positive-semidefinite.

We now complete the square on the $Z_1 y^2$ term as

$$\langle \vec{\mathbf{v}}, \vec{\mathbf{v}} \rangle = \left(\sqrt{a}x + \frac{b}{\sqrt{a}}y + \frac{c}{\sqrt{a}}z \right)^2 + \left(\sqrt{Z_1}y + \frac{e - \frac{bc}{a}}{\sqrt{Z_1}}z \right)^2 + \left(f - \frac{c^2}{a} - \frac{\left(e - \frac{bc}{a}\right)^2}{Z_1} \right) z^2.$$

This imposes the requirement that

$$Z_2 = f - \frac{c^2}{a} - \frac{\left(e - \frac{bc}{a}\right)^2}{Z_1} > 0$$

as well. In this case, we would begin with

$$\begin{cases} \sqrt{a}x + \frac{b}{\sqrt{a}}y + \frac{c}{\sqrt{a}}z & = 0 \\ \sqrt{Z_1}y + \ z & = 0 \\ \sqrt{Z_2}z & = 0 \end{cases}$$

so that this upper-triangular system of equations has only the trivial solution $x = y = z = 0$. Therefore, $\langle \cdot, \cdot \rangle$ is positive-definite in this case.
In total, we require five inequalities

$$a > 0, \quad , b > 0, \quad , c > 0,$$

$$Z_1 = d - \frac{b^2}{a} > 0, \quad \text{and} \quad Z_2 = f - \frac{c^2}{a} - \frac{(ae - bc)^2}{a(ad - b^2)} > 0.$$

Remark: As before, it can be shown that these five inequalities are equivalent to $tr(A) > 0$ and $det(A) > 0$, though it is a fair bit less trivial to do so.

(c) Does this method seem like a practical way to check if an $n \times n$ matrix is positive-definite when $n \geq 4$?

The rabbit hole only gets deeper from here. We expect to see some form of recursion relation as we complete more and more squares in the associated quadratic forms. This will make for uglier and uglier computations for higher dimensions. This is highly impractical to detect positive-definiteness.

Remark: However, it might actually be simpler to show if we can find a quick method of calculating $det(A)$. Our knowledge currently has at a complexity of $\mathcal{O}(n^3)$ since we compute it through Gaussian elimination. Maybe there's a quicker, more reliable algorithm..?