3-19-1994

Inner product spaces.

Definition. Let V be a vector space over F, where $F = \mathbb{R}$ or $\mathbb{C}$. An **inner product** on V is a function $(\cdot,\cdot) : V \times V \to F$ which satisfies:

1. (Linearity) $(ax + y, z) = a(x, z) + (y, z)$, for $x, y, z \in V$, $a \in F$.
2. (Symmetry) $(x, y) = \overline{(y, x)}$, for $x, y \in V$. ("$\overline{x}$" denotes the complex conjugate of x.)
3. (Positive-definiteness) If $x \ne 0$, then $(x, x) > 0$.

A vector space with an inner product is an **inner product space**. If $F = \mathbb{R}$, V is a **real inner product space**; if $F = \mathbb{C}$, V is a **complex inner product space**.

Example. Suppose V is a complex inner product space. By properties 1 and 2,

$$(x, ay + z) = \overline{(ay + z, x)} = \overline{a(y, x) + (z, x)} = \overline{a}\overline{(y, x)} + \overline{(z, x)} = \overline{a}(x, y) + (x, z).$$

That is, the inner product is *almost* linear in the second variable — constants pull out, up to complex conjugation. This is sometimes referred to as **sesquilinearity**. $\square$

Example. If $F = \mathbb{R}$, $\overline{x} = x$. In this case, the second property reads $(x, y) = (y, x)$. Moreover, the result I derived in the last example becomes $(x, ay + z) = a(x, y) + (x, z)$.

Thus, a real inner product is linear in each variable.

Why include complex conjugation in the symmetry axiom? If I had used $(x, y) = (y, x)$ in the complex case, then

$$0 < (ix, ix) = i(x, ix) = i(ix, x) = i \cdot i(x, x) = -(x, x).$$

This contradicts $(x, x) > 0$. That is, I can't have *both* pure symmetry and positive definiteness. $\square$

Example.

$$(0, x) = (0 + 0, x) = (0, x) + (0, x), \quad \text{so} \quad 0 = (0, x).$$

By symmetry, $(x, 0) = 0$ as well. $\square$

Example. If $F = \mathbb{C}$, the condition $(x, x) > 0$ implies that $(x, x) \in \mathbb{R}$. $\square$

Example. The **dot product** on $\mathbb{R}^n$ is given by

$$\langle a_1, \ldots, a_n\rangle \cdot \langle b_1, \ldots, b_n\rangle = a_1 b_1 + \cdots + a_n b_n.$$

It's easy to verify that the axioms for an inner product hold. For example,

$$\langle a_1, \ldots, a_n\rangle \cdot \langle a_1, \ldots, a_n\rangle = a_1^2 + \cdot + a_n^2 > 0$$

provided that $\langle a_1, \ldots, a_n\rangle \ne 0$. $\square$

I can use an inner product to define **lengths** and **angles**. You can think of an inner product as an axiomatic way of introducing (metric) geometry into vector spaces.

Definition. Let V be an inner product space, and let $x, y \in V$.

1. The **length** of x is $|x| = (x, x)^{1/2}$.
2. The **distance** between x and y is $|x - y|$.
3. The **angle** between x and y is the smallest positive real number θ satisfying

$$\cos\theta = \frac{(x, y)}{|x||y|}.$$

Remark. The definition of the **angle** between x and y will make sense once I prove the **Schwarz ineqaulity** below. $\square$

Proposition. Let V be a real inner product space, $a \in \mathbb{R}$, $x, y \in V$.

1. $|ax| = |a||x|$. ("$|a|$" denotes the absolute value of a.)
2. $x \neq 0$ if and only if $|x| > 0$.
3. (**Schwarz inequality**) $|(x, y)| \leq |x||y|$.
4. (**Triangle inequality**) $|x + y| \leq |x| + |y|$.

Proof. 1. Since $(ax, ax) = a^2(x, x)$,

$$|ax| = \sqrt{a^2}|x| = |a||x|.$$

2. $x \neq 0$ implies $(x, x) > 0$, and hence $|x| > 0$. Conversely, if $x = 0$, then $(0, 0) = 0$, so $|x| = 0$.

3. First, observe that

$$0 \leq (ax + by, ax + by) = a^2(x, x) + 2ab(x, y) + b^2(y, y) = a^2|x|^2 + 2ab(x, y) + b^2|y|^2.$$

Set $a = |y|$, $b = |x|$. Then

$$0 \leq 2|x|^2|y|^2 + 2|x||y|(x, y),$$
$$0 \leq |x||y| + (x, y),$$
$$|x||y| \geq -(x, y).$$

Replace x with $-x$; this yields

$$|x||y| \geq (x, y).$$

Therefore, $|x||y| \geq |(x, y)|$.

4.

$$|x + y|^2 = (x + y, x + y) = |x|^2 + 2(x, y) + |y|^2 \leq |x|^2 + 2|x||y| + |y|^2 = (|x| + |y|)^2.$$

Hence, $|x + y| \leq |x| + |y|$. $\square$

Example. Let $C[0, 1]$ denote the real vector space of continuous functions on the interval $[0, 1]$. Define an inner product on $C[0, 1]$ by

$$(f, g) = \int_0^1 f(x)g(x)\,dx.$$

Note that $f(x)g(x)$ is integrable, since it's continuous on a closed interval.

The verification that this gives an inner product relies on standard properties of Riemann integrals. For example, if $f \neq 0$,

$$(f, f) = \int_0^1 f(x)^2 \, dx > 0.$$

Given that this is a real inner product, I may apply the preceding proposition to produce some useful results. For example, the Schwarz inequality says that

$$\left(\int_0^1 f(x)^2 \, dx\right)^{1/2} \left(\int_0^1 g(x)^2 \, dx\right)^{1/2} \geq \left|\int_0^1 f(x)g(x) \, dx\right|. \quad \square$$

The Gram-Schmidt algorithm.

Definition. A set of vectors S in an inner product space V is **orthogonal** if $(v_i, v_j) = 0$ for $v_i, v_j \in S$, $v_i \neq v_j$.

An orthogonal set S is **orthonormal** if $|v_i| = 1$ for all $v_i \in S$.

The vectors in an orthogonal set are mutually perpendicular. The vectors in an orthonormal set are mutually perpendicular *unit* vectors.

Notation. If I is an index set, the **Kronecker delta** δ_{ij} (or $\delta(i, j)$) is defined by

$$\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

With this notation, a set $S = \{v_i\}$ is orthonormal if

$$(v_i, v_j) = \delta_{ij}.$$

Note that the $n \times n$ matrix whose (i, j)-th component is δ_{ij} is the $n \times n$ identity matrix.

Example. The standard basis for $\mathbb{R}^n$ is orthonormal. $\square$

Example. Here is an orthonormal basis for $\mathbb{R}^2$:

$$\left\{ \begin{bmatrix} \frac{3}{5} \\ \frac{4}{5} \end{bmatrix}, \begin{bmatrix} -\frac{4}{5} \\ \frac{3}{5} \end{bmatrix} \right\} \quad \square$$

Example. Let $C[0, 2\pi]$ denote the *complex-valued* continuous functions on $[0, 2\pi]$. Define an inner product by

$$(f, g) = \frac{1}{2\pi} \int_0^{2\pi} f(x)\overline{g(x)} \, dx.$$

Let $m, n \in \mathbb{Z}$. Then

$$\frac{1}{2\pi} \int_0^{2pi} e^{imx} e^{-inx} \, dx = \delta_{mn}.$$

It follows that the set

$$\left\{\frac{1}{\sqrt{2\pi}}e^{mix} \mid m = \ldots, -1, 0, 1, \ldots\right\}$$

is orthonormal in $C[0, 2\pi]$. □

Proposition. Let $\{v_i\}$ be an orthogonal set of vectors, $v_i \neq 0$ for all i. Then $\{v_i\}$ is independent.

Proof. Suppose

$$a_1 v_{i_1} + \cdots + a_n v_{i_n} = 0.$$

Take the inner product of both sides with v_{i_1}. Since $\{v_i\}$ is orthogonal, this gives $a_1(v_{i_1}, v_{i_1}) = 0$. Now $(v_{i_1}, v_{i_1}) > 0$ since $v_{i_1} \neq 0$, so $a_1 = 0$. Similarly, $a_j = 0$ for all j. Therefore, $\{v_i\}$ is independent. □

An orthonormal set consists of vectors of length 1, so the vectors are obviously nonzero. Hence, an orthonormal set is independent, and forms a basis for the subspace it spans. A basis which is an orthonormal set is called an **orthonormal basis**.

It is very easy to find the components of a vector relative to an orthonormal basis.

Proposition. Let $\{v_i\}$ be an orthonormal basis for V, and let $v \in V$. Then

$$v = \sum_i (v, v_i) v_i.$$

Note: In fact, the sum above is a finite sum — that is, only finitely many terms are nonzero.

Proof. Since $\{v_i\}$ is a basis,

$$v = a_1 v_{i_1} + \cdots + a_n v_{i_n}, \quad \text{where} \quad a_j \in F, v_{i_v} \in \{v_i\}.$$

Let $i_j \in \{i_1, \ldots, i_n\}$. Take the inner product of both sides with v_{i_j}. Then

$$(v, v_{i_j}) = a_1(v_{i_1}, v_{i_j}) + \cdots + a_n(v_{i_n}, v_{i_j}) = a_j,$$

since by orthonormality $(v_{i_k}, v_{i_j}) = 0$ if $i_k \neq i_j$ and $(v_{i_1}, v_{i_j}) = 1$.

Therefore,

$$v = (v, v_{i_1})v_{i_1} + \cdots + (v, v_{i_n})v_{i_n}.$$

Observe that if $i \notin \{i_1, \ldots, i_n\}$, then

$$(v_i, v) = (v_i, a_1 v_{i_1} + \cdots + a_n v_{i_n}) = 0.$$

It follows that the only nonzero terms in the sum $\sum_i (v, v_i)v_i$ are those for which $i \in \{i_1, \ldots, i_n\}$, and

$$v = (v, v_{i_1})v_{i_1} + \cdots + (v, v_{i_n})v_{i_n} = \sum_i (v, v_i)v_i. \ \square$$

Example. Here is an orthonormal basis for $\mathbb{R}^2$:

$$\left\{ \begin{bmatrix} \frac{3}{5} \\ \frac{4}{5} \end{bmatrix}, \begin{bmatrix} -\frac{4}{5} \\ \frac{3}{5} \end{bmatrix} \right\}$$

To express $\langle -7, 6\rangle$ in terms of this basis, take the dot product of the vector with each element of the basis:

$$\begin{bmatrix} -7 \\ 6 \end{bmatrix} = \frac{3}{5}\begin{bmatrix} \frac{3}{5} \\ \frac{4}{5} \end{bmatrix} + \frac{46}{5}\begin{bmatrix} -\frac{4}{5} \\ \frac{3}{5} \end{bmatrix}. \square$$

Example. Let $C[0, 2\pi]$ denote the complex inner product space of complex-valued continuous functions on $[0, 2\pi]$, where the inner product is defined by

$$(f, g) = \frac{1}{2\pi}\int_0^{2\pi} f(x)\overline{g(x)}\, dx.$$

I showed earlier that the following set is orthonormal:

$$S = \left\{\frac{1}{\sqrt{2\pi}}e^{mix} \mid m = \ldots, -1, 0, 1, \ldots\right\}.$$

Suppose I try to compute the "components" of $f(x) = x$ relative to this orthonormal set by taking inner products — that is, using the approach of the preceding example.

For $m = 0$,

$$\frac{1}{\sqrt{2\pi}}\int_0^{2\pi} x\, dx = \pi\sqrt{2\pi}.$$

Suppose $m \neq 0$. Then

$$\frac{1}{\sqrt{2\pi}}\int_0^{2\pi} xe^{-mix}\, dx = \frac{1}{\sqrt{2\pi}}\left[\frac{i}{m}xe^{-mix} - \frac{1}{m^2}e^{-mix}\right]_0^{2\pi} = \frac{i\sqrt{2\pi}}{m}.$$

There are infinitely many nonzero components! Of course, the reason this does not contradict the earlier result is that $f(x) = x$ may not lie in the span of S. S is orthonormal, hence independent, but it is not a basis for $C[0, 2\pi]$.

In fact, since $e^{mix} = \cos mx + i \sin mx$, a finite linear combination of elements of S must be periodic.

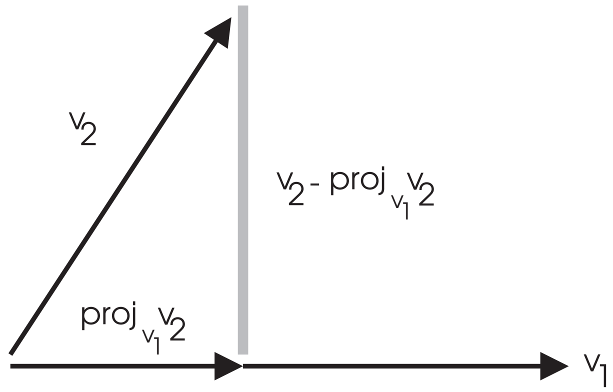
It is still reasonable to ask whether (or in what sense) the infinite sum

$$\pi\sqrt{2\pi} + \sum_{m=1}^{\infty}\left(\frac{i\sqrt{2\pi}}{m}e^{mix} - \frac{i\sqrt{2\pi}}{m}e^{-mix}\right)$$

represents the function $f(x) = x$. For example, it is reasonable to ask whether the series converges uniformly to f at each point of $[0, 2\pi]$. The answers to these kinds of questions would require an excursion into the theory of **Fourier series.** $\square$

Since it's so easy to find the components of a vector relative to an orthonormal basis, it's of interest to have an algorithm which converts a given basis to an orthonormal one.

The **Gram-Schmidt algorithm** converts a basis to an orthonormal basis by "straightening out" the vectors one by one.



The picture shows the first step in the straightening process. Given vectors v_1 and v_2, I want to replace v_2 with a vector perpendicular to v_1. I can do this by taking the component of v_2 perpendicular to v_1, which is

$$v_2 - \frac{(v_1, v_2)}{(v_1, v_1)} v_1.$$

Lemma. (**Gram-Schmidt algorithm**) Let $\{v_1, \ldots, v_k\}$ is a set of nonzero vectors in an inner product space V. Suppose $v_1, \ldots, v_{k-1}$ are pairwise orthogonal. Let

$$v_k' = v_k - \sum_{i<k} \frac{(v_i, v_k)}{(v_i, v_i)} v_i.$$

Then v_k' is orthogonal to $v_1, \ldots, v_{k-1}$.

Proof. Let $j \in \{1, \ldots, k-1\}$. Then

$$(v_j, v_k') = (v_j, v_k) - \sum_{i<k} \frac{(v_i, v_k)}{(v_i, v_i)} (v_j, v_i).$$

Now $(v_j, v_i) = 0$ for $i \neq j$, so the right side collapses to

$$(v_j, v_k) - \frac{(v_j, v_k)}{(v_j, v_j)} (v_j, v_j) = (v_j, v_k) - (v_j, v_k) = 0. \square$$

Suppose that I start with an independent set $\{v_1, \ldots, v_n\}$. Apply the Gram-Schmidt procedure to the set, beginning with $v_1' = v_1$. This produces an orthogonal set $\{v_1', \ldots, v_n'\}$. In fact, $\{v_1', \ldots, v_n'\}$ is a *nonzero* orthogonal set, so it is independent as well.

To see that each v_k' is nonzero, suppose

$$0 = v_k' = v_k - \sum_{i<k} \frac{(v_i, v_k)}{(v_i, v_i)} v_i.$$

Then

$$v_k = \sum_{i<k} \frac{(v_i, v_k)}{(v_i, v_i)} v_i,$$

which contradicts the independence of $\{v_i\}$.

In general, if the algorithm is applied iteratively to a set of vectors, the span is preserved at each state. That is,

$$\langle v_1, \ldots, v_k \rangle = \langle v_1', \ldots, v_k' \rangle.$$

This is trivially true at the start, since $v_1 = v_1'$. Assume inductively that

$$\langle v_1, \ldots, v_{k-1}\rangle = \langle v_1', \ldots, v_{k-1}'\rangle.$$

The equation

$$v_k' = v_k - \sum_{i<k} \frac{(v_i, v_k)}{(v_i, v_i)} v_i$$

shows that $\langle v_1', \ldots, v_k'\rangle \subset \langle v_1, \ldots, v_k\rangle$.

Conversely,

$$v_k = v_k' + \sum_{i<k} \frac{(v_i, v_k)}{(v_i, v_i)} v_i \subset \langle v_k'\rangle + \langle v_1, \ldots, v_{k-1}\rangle =$$

$$\langle v_k'\rangle + \langle v_1', \ldots, v_{k-1}'\rangle = \langle v_1', \ldots, v_k'\rangle.$$

It follows that $\langle v_1, \ldots, v_k\rangle \subset \langle v_1', \ldots, v_k'\rangle$, so $\langle v_1, \ldots, v_{k-1}\rangle = \langle v_1', \ldots, v_{k-1}'\rangle$, as claimed.

To summarize: If you apply Gram-Schmidt to a set of vectors, the algorithm produces a new set of vectors with the same span as the old set. If the original set was independent, the new set is independent (and orthogonal) as well.

So, for example, if Gram-Schmidt is applied to a *basis* for an inner product space, it will produce an *orthogonal basis* for the space.

Finally, you can always produce *orthonormal* set from a orthogonal set (of nonzero vectors) — merely divide each vector in the orthogonal set by its length. □

Example. Apply Gram-Schmidt to the set

$$\left\{ \begin{bmatrix} 3 \\ 0 \\ 4 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 7 \end{bmatrix}, \begin{bmatrix} 2 \\ 9 \\ 11 \end{bmatrix} \right\}.$$

$$v_1' = v_1 = \langle 3, 0, 4\rangle,$$

$$v_2' = \langle -1, 0, 7\rangle - \frac{\langle -1, 0, 7\rangle \cdot \langle 3, 0, 4\rangle}{\langle 3, 0, 4\rangle \cdot \langle 3, 0, 4\rangle} \langle 3, 0, 4\rangle = \langle -4, 0, 3\rangle,$$

$$v_3' = \langle 2, 9, 11\rangle - \frac{\langle 2, 9, 11\rangle \cdot \langle 3, 0, 4\rangle}{\langle 3, 0, 4\rangle \cdot \langle 3, 0, 4\rangle} \langle 3, 0, 4\rangle - \frac{\langle 2, 9, 11\rangle \cdot \langle -4, 0, 3\rangle}{\langle -4, 0, 3\rangle \cdot \langle -4, 0, 3\rangle} \langle -4, 0, 3\rangle = \langle 0, 9, 0\rangle.$$

(A common mistake here is to project onto v_1, v_2, ... I need to project onto the vectors that have *already been orthogonalized*. That is why I projected onto $\langle 3, 0, 4\rangle$ and $\langle -4, 0, 3\rangle$ rather than $\langle 3, 0, 4\rangle$ and $\langle -1, 0, 7\rangle$.)

The set

$$\left\{ \begin{bmatrix} 3 \\ 0 \\ 4 \end{bmatrix}, \begin{bmatrix} -4 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 9 \\ 0 \end{bmatrix} \right\}$$

is orthogonal.

The correponding orthonormal set is

$$\left\{ \frac{1}{5} \begin{bmatrix} 3 \\ 0 \\ 4 \end{bmatrix}, \frac{1}{5} \begin{bmatrix} -4 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}.$$

Here is how to do the same thing using *Mathematica*. First, load the package:

```
Needs["LinearAlgebra`Orthogonalization`"]
```

Next, enter the vectors:

```
vecs = {{3,0,4}, {-1,0,7}, {2,9,11}}
```

Apply Gram-Schmidt:

```
GramSchmidt[vecs]
```

```
    3     4      4      3
{{-, 0, -}, {-(-), 0, -}, {0, 1, 0}}
    5     5      5      5
```

This assumes that you are using the standard inner product; it's possible to specify a more general inner product as well. □

Representing inner products by matrices.

Let $v = \langle v_1, \ldots, v_n \rangle$ and $w = \langle w_1, \ldots, w_n \rangle$, where the vectors are expressed in terms of the standard basis. The standard inner product on $\mathbb{R}^n$ is

$$v \cdot w = v_1 w_1 + \cdots + v_n w_n.$$

I can write this in matrix form as

$$\begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}.$$

From this point of view, it is natural to say that the **matrix of the standard inner product with respect to the standard basis** is the $n \times n$ identity matrix I. I will write

$$v \cdot w = v_{\text{std}}^T \cdot I \cdot w_{\text{std}}.$$

Recall that A^T denotes the **transpose** of the matrix A. As usual, a vector in $\mathbb{R}^n$ is represented as an $n \times 1$ matrix.

The obvious question is: What is the matrix of the standard inner product with respect to an arbitrary basis $\mathcal{B}$?

It is easy to prove that if A and B are matrices, then $(AB)^T = B^T A^T$. So

$$v_{\text{std}}^T = ([\mathcal{B} \to \text{std}] \cdot v_{\mathcal{B}})^T = v_{\mathcal{B}}^T [\mathcal{B} \to \text{std}]^T.$$

Moreover,

$$w_{\text{std}} = [\mathcal{B} \to \text{std}] \cdot w_{\mathcal{B}}.$$

Hence,

$$v \cdot w = v_{\mathcal{B}}^T [\mathcal{B} \to \text{std}]^T \cdot I \cdot [\mathcal{B} \to \text{std}] w_{\mathcal{B}} = v_{\mathcal{B}}^T [\mathcal{B} \to \text{std}]^T \cdot [\mathcal{B} \to \text{std}] w_{\mathcal{B}}.$$

Thus, the **matrix of the standard inner product relative to** $\mathcal{B}$ is $[\mathcal{B} \to \text{std}]^T \cdot [\mathcal{B} \to \text{std}]$.

Example. If A is a matrix, $A^T A$ is a **symmetric matrix**. For

$$(A^T A)_{ij} = \sum_{k=1}^{n} A_{ik}^T A_{kj} = \sum_{k=1}^{n} A_{ki} A_{kj} = \sum_{k=1}^{n} A_{jk}^T A_{ki} = (A^T A)_{ji}.$$

It follows that the matrix of the standard inner product relative to an arbitrary basis is symmetric. □

Example. More generally, let $f : V \times V \to F$ be an inner product. Let $\mathcal{B} = \{v_1, \ldots, v_n\}$ be a basis for V. Then the matrix of F relative to $\mathcal{B}$ is the $n \times n$ matrix whose i-j-th entry is

$$A_{ij} = f(v_i, v_j).$$

If $F = \mathbb{R}$, then

$$A_{ij} = f(v_i, v_j) = f(v_j, v_i) = A_{ji},$$

and the matrix $A = (A_{ij})$ is symmetric. That is, the matrix of an *arbitrary* inner product relative to an arbitrary basis is symmetric.

Explicitly,

$$f(v, w) = v_{\mathcal{B}}^T A w_{\mathcal{B}}.$$

It works the other way, too — you can start with a real symmetric matrix and use this equation to construct a real inner product. □

Example. The symmetric matrix

$$A = \begin{bmatrix} 2 & 3 \\ 3 & 1 \end{bmatrix}$$

defines a real inner product on $\mathbb{R}$. Interpret this as the matrix of the inner product relative to the standard basis.

If $v = \langle v_1, v_2 \rangle$ and $w = \langle w_1, w_2 \rangle$ are vectors written in terms of the standard basis, then

$$f(v, w) = \begin{bmatrix} v_1 & v_2 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = 2v_1w_1 + 3v_1w_2 + 3v_2w_1 + v_2w_2. \; \square$$

Example. Here is a basis for $\mathbb{R}^2$:

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$$

Then

$$[\mathcal{B} \to \mathrm{std}] = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

The matrix of the standard inner product relative to this basis is

$$\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}^T \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}. \; \square$$