

## 2.6 Orthogonal Projections

Recall the discussion of the Gram-Schmidt process, where we saw that any finite-dimensional subspace  $W$  of an inner product space  $V$  has an orthonormal basis  $\beta_W = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$ . In such a situation, we can define the *orthogonal projections* onto  $W$  and  $W^\perp$  via

$$\pi_W : V \rightarrow W : \mathbf{x} \mapsto \sum_{j=1}^n \langle \mathbf{x}, \mathbf{w}_j \rangle \mathbf{w}_j, \quad \pi_W^\perp : V \rightarrow W^\perp : \mathbf{x} \mapsto \mathbf{x} - \pi_W(\mathbf{x})$$

Our previous goal was to use orthonormal bases to ease *computation*. In this section we develop projections more generally. First recall the notion of a direct sum within a vector space  $V$ :

$$V = X \oplus Y \iff \forall \mathbf{v} \in V, \exists \text{ unique } \mathbf{x} \in X, \mathbf{y} \in Y \text{ such that } \mathbf{v} = \mathbf{x} + \mathbf{y}$$

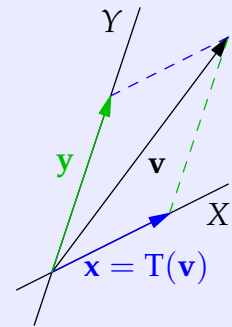
**Definition 2.54.** A linear map  $T \in \mathcal{L}(V)$  is a *projection* if:

$$V = \mathcal{R}(T) \oplus \mathcal{N}(T) \quad \text{and} \quad T|_{\mathcal{R}(T)} = I_{\mathcal{R}(T)}$$

Otherwise said,  $T(\mathbf{r} + \mathbf{n}) = \mathbf{r}$  whenever  $\mathbf{r} \in \mathcal{R}(T)$  and  $\mathbf{n} \in \mathcal{N}(T)$ .

Alternatively, given  $V = X \oplus Y$ , the *projection along  $Y$  onto  $X$*  is the map  $\mathbf{v} = \mathbf{x} + \mathbf{y} \mapsto \mathbf{x}$ .

We call a  $A \in M_n(\mathbb{F})$  a *projection matrix* if  $L_A \in \mathcal{L}(\mathbb{F}^n)$  is a projection.



**Example 2.55.**  $A = \frac{1}{5} \begin{pmatrix} 6 & -2 \\ 3 & -1 \end{pmatrix}$  is a projection matrix with  $\mathcal{R}(A) = \text{Span} \begin{pmatrix} 2 \\ 1 \end{pmatrix}$  and  $\mathcal{N}(A) = \text{Span} \begin{pmatrix} 1 \\ 3 \end{pmatrix}$ . Indeed, it is straightforward to describe all projection matrices in  $M_2(\mathbb{R})$ . There are three cases:

1.  $A = I$  is the identity matrix:  $\mathcal{R}(A) = \mathbb{R}^2$  and  $\mathcal{N}(A) = \{\mathbf{0}\}$ ;
2.  $A = 0$  is the zero matrix:  $\mathcal{R}(A) = \{\mathbf{0}\}$  and  $\mathcal{N}(A) = \mathbb{R}^2$ ;
3. Choose *distinct* subspaces  $\mathcal{R}(A) = \text{Span} \begin{pmatrix} a \\ b \end{pmatrix}$  and  $\mathcal{N}(A) = \text{Span} \begin{pmatrix} c \\ d \end{pmatrix}$ , then

$$A = \frac{1}{ad - bc} \begin{pmatrix} a \\ b \end{pmatrix} (d - c) = \frac{1}{ad - bc} \begin{pmatrix} ad & -ac \\ bd & -bc \end{pmatrix}$$

Think about why this last does what we claim.

It should be clear that every projection  $T$  has (at most) two eigenspaces:

- $\mathcal{R}(T)$  is an eigenspace with eigenvalue 1
- $\mathcal{N}(T)$  is an eigenspace with eigenvalue 0

If  $V$  is finite-dimensional and  $\rho, \eta$  are bases of  $\mathcal{R}(T), \mathcal{N}(T)$  respectively, then the matrix of  $T$  with respect to  $\rho \cup \eta$  has block form

$$[T]_{\rho \cup \eta} = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}$$

where  $\text{rank } I = \text{rank } T$ . In particular, every finite-dimensional projection is diagonalizable.

**Lemma 2.56.**  $T \in \mathcal{L}(V)$  is a projection if and only if  $T^2 = T$ .

*Proof.* Throughout, assume  $\mathbf{r} \in \mathcal{R}(T)$  and  $\mathbf{n} \in \mathcal{N}(T)$ .

( $\Rightarrow$ ) Since every vector in  $V$  has a unique representation  $\mathbf{v} = \mathbf{r} + \mathbf{n}$ , simply compute

$$T^2(\mathbf{v}) = T(T(\mathbf{r} + \mathbf{n})) = T(\mathbf{r}) = \mathbf{r} = T(\mathbf{v})$$

( $\Leftarrow$ ) Suppose  $T^2 = T$ . Note first that if  $\mathbf{r} \in \mathcal{R}(T)$ , then  $\mathbf{r} = T(\mathbf{v})$  for some  $\mathbf{v} \in V$ , whence

$$T(\mathbf{r}) = T^2(\mathbf{v}) = T(\mathbf{v}) = \mathbf{r} \tag{†}$$

Thus  $T$  is the identity on  $\mathcal{R}(T)$ . Moreover, if  $\mathbf{x} \in \mathcal{R}(T) \cap \mathcal{N}(T)$ , (†) says that  $\mathbf{x} = T(\mathbf{x}) = \mathbf{0}$ , whence

$$\mathcal{R}(T) \cap \mathcal{N}(T) = \{\mathbf{0}\}$$

and so  $\mathcal{R}(T) \oplus \mathcal{N}(T)$  is a well-defined subspace of  $V$ .<sup>1</sup> To finish things off, let  $\mathbf{v} \in V$  and observe that

$$T(\mathbf{v} - T(\mathbf{v})) = T(\mathbf{v}) - T^2(\mathbf{v}) = \mathbf{0} \implies \mathbf{v} - T(\mathbf{v}) \in \mathcal{N}(T)$$

so that  $\mathbf{v} = T(\mathbf{v}) + (\mathbf{v} - T(\mathbf{v}))$  is a decomposition into  $\mathcal{R}(T)$ - and  $\mathcal{N}(T)$ -parts. We conclude that  $V = \mathcal{R}(T) \oplus \mathcal{N}(T)$  and that  $T$  is a projection. ■

Thus far the discussion hasn't had anything to do with inner products...

**Definition 2.57.** An *orthogonal projection* is a projection  $T \in \mathcal{L}(V)$  on an inner product space for which we additionally have

$$\mathcal{N}(T) = \mathcal{R}(T)^\perp \quad \text{and} \quad \mathcal{R}(T) = \mathcal{N}(T)^\perp$$

Alternatively, given  $V = W \oplus W^\perp$ , the *orthogonal projection*  $\pi_W$  is the projection along  $W^\perp$  onto  $W$ : that is

$$\mathcal{R}(\pi_W) = W \quad \text{and} \quad \mathcal{N}(\pi_W) = W^\perp$$

The complementary orthogonal projection  $\pi_{W^\perp} = I - \pi_W$  has  $\mathcal{R}(\pi_{W^\perp}) = W^\perp$  and  $\mathcal{N}(\pi_{W^\perp}) = W$ .

**Example (2.55 continued).** The identity and zero matrices are both  $2 \times 2$  orthogonal projection matrices, while those of type 3 are orthogonal if  $\begin{pmatrix} a \\ b \end{pmatrix} \cdot \begin{pmatrix} c \\ d \end{pmatrix} = 0$ : we obtain

$$A = \frac{1}{a^2 + b^2} \begin{pmatrix} a \\ b \end{pmatrix} (a \ b) = \frac{1}{a^2 + b^2} \begin{pmatrix} a^2 & ab \\ ab & b^2 \end{pmatrix}$$

More generally, if  $W \leq \mathbb{F}^n$  has orthonormal basis  $\{\mathbf{w}_1, \dots, \mathbf{w}_k\}$ , then the matrix of  $\pi_W$  is  $\sum_{j=1}^k \mathbf{w}_j \mathbf{w}_j^*$ .

<sup>1</sup>In finite dimensions, the rank-nullity theorem and dimension counting finishes the proof here without having to proceed further:

$$\dim(\mathcal{R}(T) \oplus \mathcal{N}(T)) = \text{rank } T + \text{null } T = \dim V \implies \mathcal{R}(T) \oplus \mathcal{N}(T) = V$$

**Theorem 2.58.** A projection  $T \in \mathcal{L}(V)$  is orthogonal if and only if it is self-adjoint  $T = T^*$ .

*Proof.* ( $\Rightarrow$ ) By assumption,  $\mathcal{R}(T)$  and  $\mathcal{N}(T)$  are orthogonal subspaces. Letting  $\mathbf{x}, \mathbf{y} \in V$  and using subscripts to denote  $\mathcal{R}(T)$ - and  $\mathcal{N}(T)$ -parts, we see that

$$\langle \mathbf{x}, T(\mathbf{y}) \rangle = \langle \mathbf{x}_r + \mathbf{x}_n, \mathbf{y}_r \rangle = \langle \mathbf{x}_r, \mathbf{y}_r + \mathbf{y}_n \rangle = \langle T(\mathbf{x}), \mathbf{y} \rangle \implies T^* = T$$

( $\Leftarrow$ ) Suppose  $T$  is a self-adjoint projection. By the fundamental subspaces theorem,

$$\mathcal{N}(T) = \mathcal{N}(T^*) = \mathcal{R}(T)^\perp$$

Since  $T$  is a projection already, we have  $V = \mathcal{R}(T) \oplus \mathcal{N}(T) = \mathcal{R}(T) \oplus \mathcal{R}(T)^\perp$ , from which

$$\mathcal{R}(T) = (\mathcal{R}(T)^\perp)^\perp = \mathcal{N}(T)^\perp$$

The language of projections allows us to rephrase the Spectral Theorem.

**Theorem 2.59 (Spectral Theorem, mk. II).** Let  $V$  be a finite-dimensional complex/real inner product space and  $T \in \mathcal{L}(V)$  be normal/self-adjoint with spectrum  $\{\lambda_1, \dots, \lambda_k\}$  and corresponding eigenspaces  $E_1, \dots, E_k$ . Let  $\pi_j \in \mathcal{L}(V)$  be the orthogonal projection onto  $E_j$ . Then:

1.  $V = E_1 \oplus \dots \oplus E_k$  is a direct sum of orthogonal subspaces, in particular,  $E_j^\perp$  is the direct sum of the remaining eigenspaces.
2.  $\pi_i \pi_j = 0$  if  $i \neq j$ .
3. (Resolution of the identity)  $I_V = \pi_1 + \dots + \pi_k$
4. (Spectral decomposition)  $T = \lambda_1 \pi_1 + \dots + \lambda_k \pi_k$

*Proof.* 1.  $T$  is diagonalizable and so  $V$  is the direct sum of the eigenspaces of  $T$ . Since  $T$  is normal, the eigenvectors corresponding to distinct eigenvalues are orthogonal, whence the eigenspaces are mutually orthogonal. In particular, this says that

$$\hat{E}_j := \bigoplus_{i \neq j} E_i \leq E_j^\perp$$

Since  $V$  is finite-dimensional, we have  $V = E_j \oplus E_j^\perp$ , whence

$$\dim \hat{E}_j = \sum_{i \neq j} \dim E_i = \dim V - \dim E_j = \dim E_j^\perp \implies \hat{E}_j = E_j^\perp$$

2. This is clear by part 1, since  $\mathcal{N}(\pi_j) = E_j^\perp = \hat{E}_j$ .
3. Write  $\mathbf{x} = \sum_{j=1}^k \mathbf{x}_j$  where each  $\mathbf{x}_j \in E_j$ . Then  $\pi_j(\mathbf{x}) = \mathbf{x}_j$ : now add...
4.  $T(\mathbf{x}) = \sum_{j=1}^k T(\mathbf{x}_j) = \sum_{j=1}^k \lambda_j \mathbf{x}_j = \sum_{j=1}^k \lambda_j \pi_j(\mathbf{x})$ .

**Examples 2.60.** We verify the resolution of the identity and the spectral decomposition; for clarity, we index projections and eigenspaces by eigenvalue rather than the natural numbers.

1. The symmetric matrix  $A = \begin{pmatrix} 10 & 2 \\ 2 & 7 \end{pmatrix}$  has spectrum  $\{6, 11\}$  and orthonormal eigenvectors

$$\mathbf{w}_6 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ -2 \end{pmatrix}, \quad \mathbf{w}_{11} = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

The corresponding projections therefore have matrices

$$\pi_6 = \mathbf{w}_6 \mathbf{w}_6^T = \frac{1}{5} \begin{pmatrix} 1 \\ -2 \end{pmatrix} (1 \ -2) = \frac{1}{5} \begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix}, \quad \pi_{11} = \mathbf{w}_{11} \mathbf{w}_{11}^T = \frac{1}{5} \begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix}$$

from which the resolution of the identity and the spectral decomposition are readily verified:

$$\pi_6 + \pi_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad 6\pi_6 + 11\pi_{11} = \frac{1}{5} \begin{pmatrix} 6+44 & -12+22 \\ -12+22 & 24+11 \end{pmatrix} = A$$

2. The normal matrix  $B = \begin{pmatrix} 1+i & 1+i \\ -1-i & 1+i \end{pmatrix} \in M_2(\mathbb{C})$  has spectrum  $\{2, 2i\}$  and corresponding orthonormal eigenvectors

$$\mathbf{w}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} i \\ 1 \end{pmatrix}, \quad \mathbf{w}_{2i} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}$$

The orthogonal projection matrices are therefore

$$\pi_2 = \mathbf{w}_2 \mathbf{w}_2^* = \frac{1}{2} \begin{pmatrix} i \\ 1 \end{pmatrix} (-i \ 1) = \frac{1}{2} \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix} \quad \pi_{2i} = \mathbf{w}_{2i} \mathbf{w}_{2i}^* = \frac{1}{2} \begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix}$$

from which

$$\pi_2 + \pi_{2i} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad 2\pi_2 + 2i\pi_{2i} = \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix} + \begin{pmatrix} i & 1 \\ -1 & i \end{pmatrix} = B$$

3. The matrix  $C = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$  has spectrum  $\{-1, 2\}$ , an orthonormal eigenbasis

$$\{\mathbf{u}, \mathbf{v}, \mathbf{w}\} = \left\{ \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} \right\}$$

and eigenspaces  $E_2 = \text{Span}\{\mathbf{u}\}$  and  $E_{-1} = \text{Span}\{\mathbf{v}, \mathbf{w}\}$ . The orthogonal projections have matrices

$$\begin{aligned} \pi_2 &= \mathbf{u}\mathbf{u}^T = \frac{1}{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} (1 \ 1 \ 1) = \frac{1}{3} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \\ \pi_{-1} &= \mathbf{v}\mathbf{v}^T + \mathbf{w}\mathbf{w}^T = \frac{1}{2} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} (1 \ -1 \ 0) + \frac{1}{6} \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} (1 \ 1 \ -2) \\ &= \frac{1}{2} \begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \frac{1}{6} \begin{pmatrix} 1 & 1 & -2 \\ 1 & 1 & -2 \\ -2 & -2 & 4 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} \end{aligned}$$

It is now easy to check the resolution of the identity and the spectral decomposition:

$$\pi_2 + \pi_{-1} = I \quad \text{and} \quad 2\pi_2 - \pi_{-1} = C$$

## Orthogonal Projections and Minimization Problems

We finish this section with an important observation that drives much of the application of inner product spaces to other parts of mathematics and beyond. Throughout this discussion,  $X$  and  $Y$  denote inner product spaces.

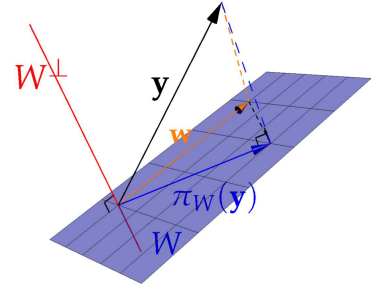
**Theorem 2.61.** Suppose  $Y = W \oplus W^\perp$ . For any  $\mathbf{y} \in Y$ , the orthogonal projection  $\pi_W(\mathbf{y})$  is the unique element of  $W$  which minimizes the distance to  $\mathbf{y}$ :

$$\forall \mathbf{w} \in W, \|\mathbf{y} - \pi_W(\mathbf{y})\| \leq \|\mathbf{y} - \mathbf{w}\|$$

*Proof.* Apply Pythagoras' Theorem: since  $\pi_W^\perp(\mathbf{y}) = \mathbf{y} - \pi_W(\mathbf{y}) \in W^\perp$ ,

$$\begin{aligned} \|\mathbf{y} - \mathbf{w}\|^2 &= \|\mathbf{y} - \pi_W(\mathbf{y}) + \pi_W(\mathbf{y}) - \mathbf{w}\|^2 \\ &= \|\mathbf{y} - \pi_W(\mathbf{y})\|^2 + \|\pi_W(\mathbf{y}) - \mathbf{w}\|^2 \\ &\geq \|\mathbf{y} - \pi_W(\mathbf{y})\|^2 \end{aligned}$$

with equality if and only if  $\mathbf{w} = \pi_W(\mathbf{y})$ .



This set up can be used to compute accurate approximations in many contexts.

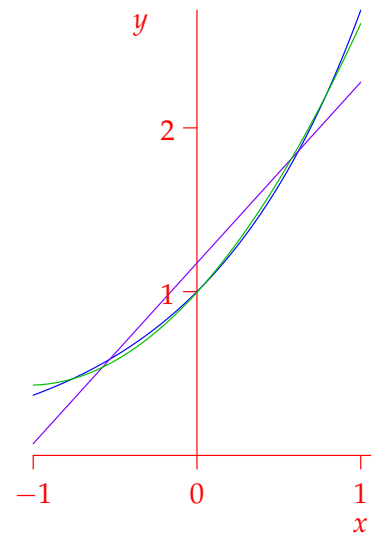
**Examples 2.62.** 1. To obtain a quadratic polynomial approximation  $p(x) = a + bx + cx^2$  to  $e^x$  on the interval  $[-1, 1]$  we choose to minimize the integral  $\int_{-1}^1 |e^x - p(x)|^2 dx$ , namely the squared  $L^2$ -norm  $\|e^x - p(x)\|^2$  on  $C[-1, 1]$ . If we let  $W = \text{Span}\{1, x, x^2\}$ , then the finite-dimensionality of  $W$  means that  $C[-1, 1] = W \oplus W^\perp$ . By the Theorem, the solution is  $p(x) = \pi_W(e^x)$ .

To compute this, recall that we have an orthonormal basis for  $W$ , namely

$$\left\{ \frac{1}{\sqrt{2}}, \sqrt{\frac{3}{2}}x, \sqrt{\frac{5}{8}}(3x^2 - 1) \right\}$$

from which

$$\begin{aligned} p(x) &= \frac{1}{2} \langle 1, e^x \rangle + \frac{3}{2} \langle x, e^x \rangle x + \frac{5}{8} (3x^2 - 1) \langle 3x^2 - 1, e^x \rangle \\ &= \frac{1}{2} \int_{-1}^1 e^x dx + \frac{3}{2} x \int_{-1}^1 x e^x dx \\ &\quad + \frac{5}{8} (3x^2 - 1) \int_{-1}^1 (3x^2 - 1) e^x dx \\ &= \frac{1}{2} (e - e^{-1}) + 3e^{-1}x + \frac{5}{4} (e - 7e^{-1})(3x^2 - 1) \\ &\approx 1.18 + 1.10x + 0.179(3x^2 - 1) \\ &\approx 1 + 1.1x + 0.537x^2 \end{aligned}$$



The **linear** and **quadratic** approximations to  $y = e^x$  are drawn. Compare this with the Maclaurin polynomial  $e^x \approx 1 + x + \frac{1}{2}x^2$  from calculus.

2. The  $n^{\text{th}}$  Fourier approximation of a function  $f(x)$  is its orthogonal projection onto the finite-dimensional subspace

$$\begin{aligned} W_n &= \text{Span}\{1, e^{ix}, e^{-ix}, \dots, e^{inx}, e^{-inx}\} \\ &= \text{Span}\{1, \cos x, \sin x, \cos 2x, \sin 2x, \dots, \cos nx, \sin nx\} \end{aligned}$$

within  $L^2[-\pi, \pi]$ , namely

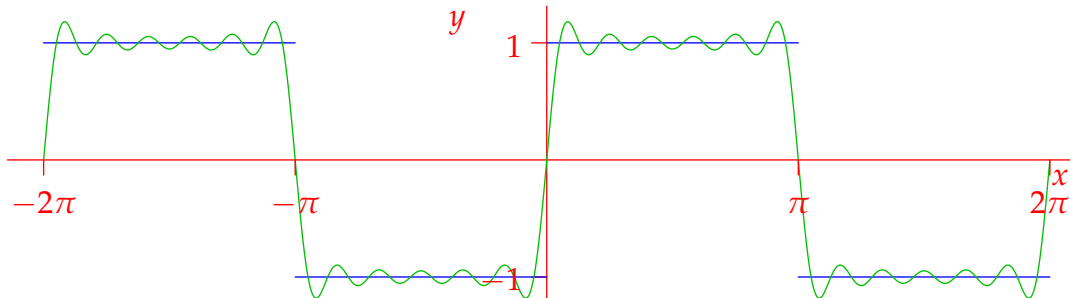
$$\begin{aligned} \mathcal{F}_n(x) &= \frac{1}{2\pi} \sum_{k=-n}^n \langle f(x), e^{ikx} \rangle e^{ikx} \\ &= \frac{1}{2\pi} \langle f(x), 1 \rangle + \frac{1}{\pi} \sum_{k=1}^n \langle f(x), \cos kx \rangle \cos kx + \langle f(x), \sin kx \rangle \sin kx \end{aligned}$$

According to the Theorem, this is the unique function in  $\mathcal{F}_n(x) \in W_n$  minimizing the integral

$$\|f(x) - \mathcal{F}_n(x)\|^2 = \int_{-\pi}^{\pi} |f(x) - \mathcal{F}_n(x)|^2 dx$$

For example, if  $f(x) = \begin{cases} 1 & \text{if } 0 < x \leq \pi \\ -1 & \text{if } -\pi < x \leq 0 \end{cases}$  is extended periodically, then

$$\mathcal{F}_{2n-1}(x) = \frac{4}{\pi} \sum_{j=1}^n \frac{\sin(2j-1)x}{2j-1} = \frac{4}{\pi} \left( \sin x + \frac{\sin 3x}{3} + \frac{\sin 5x}{5} + \dots + \frac{\sin(2n-1)x}{2n-1} \right)$$



$y = f(x)$  and its eleventh Fourier approximation  $y = \mathcal{F}_{11}(x)$

We'll return to related minimization problems in the next section.

**Exercises 2.6** 1. Compute the matrices of the orthogonal projections onto  $W$  viewed as subspaces of the standard inner product spaces  $\mathbb{R}^n$  or  $\mathbb{C}^n$ .

(a)  $W = \text{Span} \left\{ \begin{pmatrix} 4 \\ -1 \end{pmatrix} \right\}$     (b)  $W = \text{Span} \left\{ \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\}$     (c)  $W = \text{Span} \left\{ \begin{pmatrix} i \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ i \\ 1 \end{pmatrix} \right\}$

(d)  $W = \text{Span} \left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \right\}$  (*watch out, these vectors aren't orthogonal!*)

2. For each of the following matrices, compute the projections onto each eigenspace, verify the resolution of the identity and the spectral decomposition.

$$(a) \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \quad (b) \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad (c) \begin{pmatrix} 2 & 3-3i \\ 3+3i & 5 \end{pmatrix} \quad (d) \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

(You should have orthonormal eigenbases from the previous section)

3. If  $W$  be a finite-dimensional subspace of an inner product space  $V$ . If  $T = \pi_W$  is the orthogonal projection onto  $W$ , prove that  $I - T$  is the orthogonal projection onto  $W^\perp$ .
4. Let  $T \in \mathcal{L}(V)$  where  $V$  is finite-dimensional.
- If  $T$  is an orthogonal projection, prove that  $\|T(\mathbf{x})\| \leq \|\mathbf{x}\|$  for all  $\mathbf{x} \in V$ .
  - Give an example of a projection for which the inequality in (a) is *false*.
  - If  $T$  is a projection for which  $\|T(\mathbf{x})\| = \|\mathbf{x}\|$  for all  $\mathbf{x} \in V$ , what is  $T$ ?
  - If  $T$  is a projection for which  $\|T(\mathbf{x})\| \leq \|\mathbf{x}\|$  for all  $\mathbf{x} \in V$ , prove that  $T$  is an *orthogonal* projection.
5. Let  $T$  be a normal operator on a finite-dimensional inner product space. If  $T$  is a projection, prove that it must be an orthogonal projection.
6. Let  $T$  be a normal operator on a finite-dimensional complex inner product space  $V$ . Use the spectral decomposition  $T = \lambda_1 \pi_1 + \cdots + \lambda_k \pi_k$  to prove:
- If  $T^n$  is the zero map for some  $n \in \mathbb{N}$ , then  $T$  is the zero map.
  - $U \in \mathcal{L}(V)$  commutes with  $T$  if and only if  $U$  commutes with each  $\pi_j$ .
  - There exists a normal  $U \in \mathcal{L}(V)$  such that  $U^2 = T$ .
  - $T$  is invertible if and only if  $\lambda_j \neq 0$  for all  $j$ .
  - $T$  is a projection if and only if every  $\lambda_j = 0$  or  $1$ .
  - $T = -T^*$  if and only if every  $\lambda_j$  is imaginary.
7. Find a linear approximation to  $f(x) = e^x$  on  $[0, 1]$  using the  $L^2$  inner product.
8. Consider the  $L^2$  inner product on  $C[-\pi, \pi]$  inner product.
- Explain why  $\langle \sin x, x^{2n} \rangle = 0$  for all  $n$ .
  - Find linear and *cubic* approximations to  $f(x) = \sin x$ .  
(Feel free to use a computer algebra package to evaluate the integrals!)
9. Revisit Example 2.62.2
- Verify that the general complex ( $e^{ikx}$ ) and real ( $\cos kx, \sin kx$ ) expressions for the Fourier approximation are correct.  
(Hint: use Euler's formula  $e^{ikx} = \cos kx + i \sin kx$ )
  - Verify the explicit expression for  $\mathcal{F}_{2n-1}(x)$  when  $f(x)$  is the given step-function. What is  $\mathcal{F}_{2n}(x)$  in this case?

## 2.7 The Singular Value Decomposition and the Pseudoinverse

Given  $T \in \mathcal{L}(V, W)$  between finite-dimensional inner product spaces, the overarching concern of this chapter is the existence and computation of bases  $\beta, \gamma$  of  $V, W$  with two properties:

- That  $\beta, \gamma$  be *orthonormal*, thus facilitating easy calculation within  $V, W$ ;
- That the matrix  $[T]_{\beta}^{\gamma}$  be as simple as possible.

We have already addressed two special cases:

**Spectral Theorem** When  $V = W$  and  $T$  is normal/self-adjoint,  $\exists \beta = \gamma$  such that  $[T]_{\beta}$  is diagonal.

**Schur's Lemma** When  $V = W$  and  $p(t)$  splits,  $\exists \beta = \gamma$  such that  $[T]_{\beta}$  is upper triangular.

In this section we allow  $V \neq W$  and  $\beta \neq \gamma$ , and obtain a result that applies to any linear map between finite-dimensional inner product spaces.

**Example 2.63.** Let  $A = \begin{pmatrix} 3 & 1 \\ 2 & -2 \\ 1 & 3 \end{pmatrix}$  and consider orthonormal bases  $\beta = \{\mathbf{v}_1, \mathbf{v}_2\}$  of  $\mathbb{R}^2$  and  $\gamma = \{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$  of  $\mathbb{R}^3$  respectively:

$$\beta = \left\{ \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right\}, \quad \gamma = \left\{ \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{6}} \begin{pmatrix} -1 \\ -2 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix} \right\}$$

Since  $A\mathbf{v}_1 = 4\mathbf{w}_1$  and  $A\mathbf{v}_2 = 2\sqrt{3}\mathbf{w}_2$ , whence  $[L_A]_{\beta}^{\gamma} = \begin{pmatrix} 4 & 0 \\ 0 & 2\sqrt{3} \\ 0 & 0 \end{pmatrix}$  is almost diagonal.

Our main result says that such bases always exist.

**Theorem 2.64 (Singular Value Decomposition).** Suppose  $V, W$  are finite-dimensional inner product spaces and that  $T \in \mathcal{L}(V, W)$  has rank  $r$ . Then:

1. There exist orthonormal bases  $\beta = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  of  $V$  and  $\gamma = \{\mathbf{w}_1, \dots, \mathbf{w}_m\}$  of  $W$ , and positive scalars  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$  such that

$$T(\mathbf{v}_j) = \begin{cases} \sigma_j \mathbf{w}_j & \text{if } j \leq r \\ \mathbf{0} & \text{otherwise} \end{cases} \quad \text{equivalently} \quad [T]_{\beta}^{\gamma} = \begin{pmatrix} \text{diag}(\sigma_1, \dots, \sigma_r) & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix}$$

2. Any such  $\beta$  is an eigenbasis of  $T^*T$ , whence the scalars  $\sigma_j$  are uniquely determined by  $T$ : indeed

$$T^*T(\mathbf{v}_j) = \begin{cases} \sigma_j^2 \mathbf{v}_j & \text{if } j \leq r \\ \mathbf{0} & \text{otherwise} \end{cases} \quad \text{and} \quad T^*(\mathbf{w}_j) = \begin{cases} \sigma_j \mathbf{v}_j & \text{if } j \leq r \\ \mathbf{0} & \text{otherwise} \end{cases}$$

3. If  $A \in M_{m \times n}(\mathbb{F})$  with  $\text{rank } A = r$ , then  $A = P\Sigma Q^*$  where

$$\Sigma = [L_A]_{\beta}^{\gamma} = \begin{pmatrix} \text{diag}(\sigma_1, \dots, \sigma_r) & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix}, \quad P = (\mathbf{w}_1, \dots, \mathbf{w}_m), \quad Q = (\mathbf{v}_1, \dots, \mathbf{v}_n)$$

Since the columns of  $P, Q$  are orthonormal, these matrices are unitary.



**Definition 2.65.** The numbers  $\sigma_1, \dots, \sigma_r$  are the *singular values* of  $T$ . If  $T$  is not maximum rank, we have additional zero singular values  $\sigma_{r+1} = \dots = \sigma_{\min(m,n)} = 0$ .

*Freedom of Choice* While the singular values are uniquely determined, there is often significant freedom regarding the bases  $\beta$  and  $\gamma$ , particularly if any eigenspace of  $T^*T$  has dimension  $\geq 2$ .

*Special Case (Spectral Theorem)* If  $V = W$  and  $T$  is normal/self-adjoint, we may *choose*  $\beta$  to be an eigenbasis of  $T$ , then  $\sigma_j$  is the *modulus* of the corresponding eigenvalue (see Exercise 7).

*Rank-one decomposition* If we write  $g_j : V \rightarrow \mathbb{F}$  for the linear map  $g_j : \mathbf{v} \mapsto \langle \mathbf{v}, \mathbf{v}_j \rangle$  (recall Riesz's Theorem), then the singular value decomposition says

$$T = \sum_{j=1}^r \sigma_j \mathbf{w}_j g_j \quad \text{that is} \quad T(\mathbf{v}) = \sum_{j=1}^r \sigma_j \langle \mathbf{v}, \mathbf{v}_j \rangle \mathbf{w}_j$$

thus rewriting  $T$  as a linear combination of rank-one maps ( $\mathbf{w}_j g_j : V \rightarrow W$ ).

For matrices,  $g_j$  is simply multiplication by the row vector  $\mathbf{v}_j^*$  and we may write<sup>2</sup>

$$A = \sum_{j=1}^r \sigma_j \mathbf{w}_j \mathbf{v}_j^*$$

**Example (2.63 cont).** We apply the method in the Theorem to  $A = \begin{pmatrix} 3 & 1 \\ 2 & -2 \\ 1 & 3 \end{pmatrix}$ .

The symmetric(!) matrix  $A^T A = \begin{pmatrix} 14 & 2 \\ 2 & 14 \end{pmatrix}$  has eigenvalues  $\sigma_1^2 = 16$ ,  $\sigma_2^2 = 12$  and orthonormal eigenvectors  $\mathbf{v}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ ,  $\mathbf{v}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ . The singular values are therefore  $\sigma_1 = 4$ ,  $\sigma_2 = 2\sqrt{3}$ . Now compute

$$\mathbf{w}_1 = \frac{1}{\sigma_1} A \mathbf{v}_1 = \frac{1}{4\sqrt{2}} A \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{w}_2 = \frac{1}{\sigma_2} A \mathbf{v}_2 = \frac{1}{2\sqrt{6}} A \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{6}} \begin{pmatrix} -1 \\ -2 \\ 1 \end{pmatrix}$$

and observe that these are orthonormal. Finally choose  $\mathbf{w}_3 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}$  to complete the orthonormal basis  $\gamma$  of  $\mathbb{R}^3$ . A singular value decomposition is therefore

$$A = \begin{pmatrix} 3 & 1 \\ 2 & -2 \\ 1 & 3 \end{pmatrix} = P \Sigma Q^* = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{-2}{\sqrt{6}} & \frac{-1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{-1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & 2\sqrt{3} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

By expanding the decomposition,  $A$  is expressed as the sum of rank-one matrices:

$$\sigma_1 \mathbf{w}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{w}_2 \mathbf{v}_2^T = 4 \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} + 2\sqrt{3} \begin{pmatrix} \frac{-1}{\sqrt{6}} \\ \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{pmatrix} \begin{pmatrix} \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 0 & 0 \\ 2 & 2 \end{pmatrix} + \begin{pmatrix} 1 & -1 \\ 2 & -2 \\ -1 & 1 \end{pmatrix}$$

<sup>2</sup>Since  $\beta$  is orthonormal, it is common to write  $\mathbf{v}_j^*$  for the map  $g_j = \langle \cdot, \mathbf{v}_j \rangle$  in general contexts. To those familiar with the dual space  $V^* = \mathcal{L}(V, \mathbb{F})$ , the set  $\{g_1, \dots, g_n\} = \{\mathbf{v}_1^*, \dots, \mathbf{v}_n^*\}$  is the *dual basis* to  $\beta$ . In this course  $\mathbf{v}_j^*$  will only ever mean the *conjugate-transpose* of a column vector in  $\mathbb{F}^n$ . This discussion is part of why physicists write inner products differently!

*Proof.* 1. Since  $T^*T$  is self-adjoint, the spectral theorem says it has an orthonormal basis of eigenvectors  $\beta = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ . If  $T^*T(\mathbf{v}_j) = \lambda_j \mathbf{v}_j$ , then

$$\langle T(\mathbf{v}_j), T(\mathbf{v}_k) \rangle = \langle T^*T(\mathbf{v}_j), \mathbf{v}_k \rangle = \lambda_j \langle \mathbf{v}_j, \mathbf{v}_k \rangle = \begin{cases} \lambda_j & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases} \quad (*)$$

whence every eigenvalue is a non-negative real number:  $\lambda_j = \|T(\mathbf{v}_j)\|^2 \geq 0$ .

Since  $\text{rank } T^*T = \text{rank } T = r$  (Exercise 8), exactly  $r$  eigenvalues are non-zero; by reordering basis vectors if necessary, we may assume

$$\lambda_1 \geq \dots \geq \lambda_r > 0$$

If  $j \leq r$ , define  $\sigma_j := \sqrt{\lambda_j} > 0$  and  $\mathbf{w}_j := \frac{1}{\sigma_j} T(\mathbf{v}_j)$ , then the set  $\{\mathbf{w}_1, \dots, \mathbf{w}_r\}$  is orthonormal (\*). If necessary, extend this to an orthonormal basis  $\gamma$  of  $W$ .

2. If orthonormal bases  $\beta$  and  $\gamma$  exist such that  $[T]_\beta^\gamma = \begin{pmatrix} \text{diag}(\sigma_1, \dots, \sigma_r) & O \\ O & O \end{pmatrix}$ , then  $[T^*]_\gamma^\beta$  is essentially the same matrix just that its dimensions have been reversed:

$$[T^*]_\gamma^\beta = \begin{pmatrix} \text{diag}(\sigma_1, \dots, \sigma_r) & O \\ O & O \end{pmatrix} \implies T^*T = \begin{pmatrix} \text{diag}(\sigma_1^2, \dots, \sigma_r^2) & O \\ O & O \end{pmatrix}$$

whence  $T^*$  and  $T^*T$  are as claimed.

3. This is merely part 1 in the context of  $T = L_A \in \mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$ . The orthonormal bases  $\beta, \gamma$  consist of column vectors and so the (change of co-ordinate) matrices  $P, Q$  are unitary. ■

**Examples 2.66.** 1. The matrix  $A = \begin{pmatrix} 2 & 3 \\ 0 & 2 \end{pmatrix}$  has  $A^T A = \begin{pmatrix} 4 & 6 \\ 6 & 13 \end{pmatrix}$  with eigenvalues  $\sigma_1^2 = 16$  and  $\sigma_2^2 = 1$  and orthonormal eigenbasis

$$\beta = \left\{ \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \frac{1}{\sqrt{5}} \begin{pmatrix} -2 \\ 1 \end{pmatrix} \right\}$$

The singular values are therefore  $\sigma_1 = 4$  and  $\sigma_2 = 1$ , from which we obtain

$$\gamma = \left\{ \frac{1}{\sigma_1} A\mathbf{v}_1, \frac{1}{\sigma_2} A\mathbf{v}_2 \right\} = \left\{ \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{5}} \begin{pmatrix} -1 \\ 2 \end{pmatrix} \right\}$$

and the decomposition

$$A = P\Sigma Q^* = \begin{pmatrix} \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{-2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}$$

Multiplying out, we may write  $A$  as a sum of rank-one matrices

$$A = \frac{4}{5} \begin{pmatrix} 2 \\ 1 \end{pmatrix} (1 \ 2) + \frac{1}{5} \begin{pmatrix} -1 \\ 2 \end{pmatrix} (-2 \ 1) = \frac{4}{5} \begin{pmatrix} 2 & 4 \\ 1 & 2 \end{pmatrix} + \frac{1}{5} \begin{pmatrix} 2 & -1 \\ -4 & 2 \end{pmatrix}$$

2. The decomposition can be very messy to find in non-matrix situations. Here is a classic example where we simply observe the structure directly.

The  $L^2$  inner product  $\langle f, g \rangle = \int_0^1 f(x)g(x) dx$  on  $P_2(\mathbb{R})$  and  $P_1(\mathbb{R})$  admits orthonormal bases

$$\beta = \left\{ \sqrt{5}(6x^2 - 6x + 1), \sqrt{3}(2x - 1), 1 \right\}, \quad \gamma = \left\{ \sqrt{3}(2x - 1), 1 \right\}$$

Let  $T = \frac{d}{dx}$  be the derivative operator. The matrix of  $T$  is already in the required form!

$$[T]_{\beta}^{\gamma} = \begin{pmatrix} 2\sqrt{15} & 0 & 0 \\ 0 & 2\sqrt{3} & 0 \end{pmatrix}$$

thus  $\beta, \gamma$  are suitable bases and the singular values of  $T$  are  $\sigma_1 = 2\sqrt{15}$  and  $\sigma_2 = 2\sqrt{3}$ .

Since  $\beta, \gamma$  are orthonormal, we could have used the adjoint method to evaluate this directly,

$$[T^*T]_{\beta} = ([T]_{\beta}^{\gamma})^T [T]_{\beta}^{\gamma} = \begin{pmatrix} 60 & 0 & 0 \\ 0 & 12 & 0 \\ 0 & 0 & 0 \end{pmatrix} \implies \sigma_1^2 = 60, \sigma_2^2 = 12$$

Up to sign,  $\{[\mathbf{v}_1]_{\beta}, [\mathbf{v}_2]_{\beta}, [\mathbf{v}_3]_{\beta}\}$  is therefore forced to be the standard ordered basis of  $\mathbb{R}^3$ , confirming that  $\beta$  was the correct basis of  $P_2(\mathbb{R})$  all along!

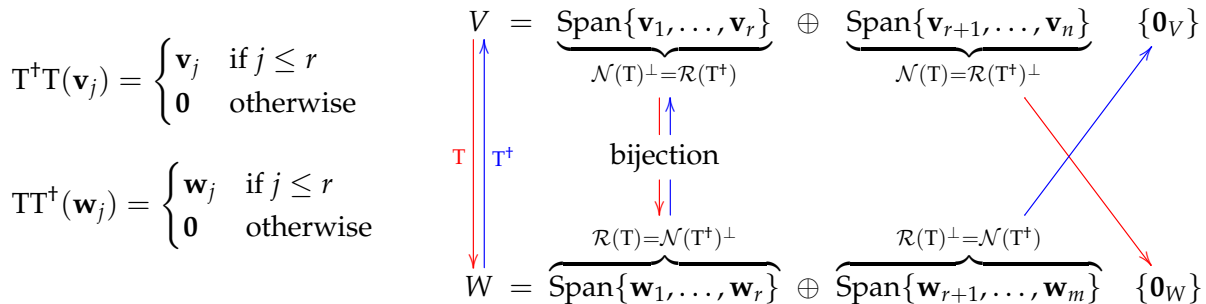
### The Pseudoinverse

The singular value decomposition of a map  $T \in \mathcal{L}(V, W)$  gives rise to a natural map from  $W$  back to  $V$ . This map behaves somewhat like an inverse even when the operator is non-invertible!

**Definition 2.67.** Given the singular value decomposition of a rank  $r$  map  $T \in \mathcal{L}(V, W)$ , the *pseudoinverse* of  $T$  is the linear map  $T^{\dagger} \in \mathcal{L}(W, V)$  defined by

$$T^{\dagger}(\mathbf{w}_j) = \begin{cases} \frac{1}{\sigma_j} \mathbf{v}_j & \text{if } j \leq r \\ \mathbf{0} & \text{otherwise} \end{cases}$$

Restricted to  $\text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_r\} \rightarrow \text{Span}\{\mathbf{w}_1, \dots, \mathbf{w}_r\}$ , the pseudoinverse really does invert  $T$ :



Otherwise said, the combinations are *orthogonal projections*:  $T^{\dagger}T = \pi_{N(T)}^{\perp}$  and  $TT^{\dagger} = \pi_{R(T)}$

Given the singular value decomposition of a matrix  $A = P\Sigma Q^*$ , its pseudoinverse is that of matrix of  $(L_A)^\dagger$ , namely

$$A^\dagger = \sum_{j=1}^r \frac{1}{\sigma_j} \mathbf{v}_j \mathbf{w}_j^* = Q\Sigma^\dagger P^* \quad \text{where} \quad \Sigma^\dagger = \begin{pmatrix} \text{diag}(\sigma_1^{-1}, \dots, \sigma_r^{-1}) & O \\ O & O \end{pmatrix}$$

**Examples 2.68.** 1. Again continuing Example 2.63,  $A = \begin{pmatrix} 3 & 1 \\ 2 & -2 \\ 1 & 3 \end{pmatrix}$  has pseudoinverse

$$\begin{aligned} A^\dagger &= \frac{1}{\sigma_1} \mathbf{v}_1 \mathbf{w}_1^* + \frac{1}{\sigma_2} \mathbf{v}_2 \mathbf{w}_2^* \\ &= \frac{1}{4\sqrt{2}\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} (1 \ 0 \ 1) + \frac{1}{2\sqrt{3}\sqrt{2}\sqrt{6}} \begin{pmatrix} -1 \\ 1 \end{pmatrix} (-1 \ -2 \ 1) \\ &= \frac{1}{8} \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} + \frac{1}{12} \begin{pmatrix} 1 & 2 & -1 \\ -1 & -2 & 1 \end{pmatrix} = \frac{1}{24} \begin{pmatrix} 5 & 4 & 1 \\ 1 & -4 & 5 \end{pmatrix} \end{aligned}$$

which is exactly what we would have found by computing  $A^\dagger = Q\Sigma^\dagger P^*$ . Observe that

$$A^\dagger A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad AA^\dagger = \frac{1}{3} \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

are the orthogonal projection matrices onto the spaces  $\mathcal{N}(A)^\perp = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\} = \mathbb{R}^2$  and  $\mathcal{R}(A) = \text{Span}\{\mathbf{w}_1, \mathbf{w}_2\} \leq \mathbb{R}^3$  respectively. Both spaces have dimension 2, since  $\text{rank } A = 2$ .

2. The pseudoinverse of  $T = \frac{d}{dx} : P_2(\mathbb{R}) \rightarrow P_1(\mathbb{R})$ , as seen in Example 2.66.2, maps

$$\begin{aligned} T^\dagger(\sqrt{3}(2x-1)) &= \frac{1}{2\sqrt{15}} \sqrt{5}(6x^2-6x+1) = \frac{1}{2\sqrt{3}}(6x^2-6x+1) \\ T^\dagger(1) &= \frac{1}{2\sqrt{3}} \sqrt{3}(2x-1) = x - \frac{1}{2} \\ \implies T^\dagger(a+bx) &= T^\dagger\left(a + \frac{b}{2} + \frac{b}{2\sqrt{3}} \sqrt{3}(2x-1)\right) \\ &= \left(a + \frac{b}{2}\right) \left(x - \frac{1}{2}\right) + \frac{b}{12}(6x^2-6x+1) \\ &= \frac{b}{2}x^2 + ax - \frac{a}{2} - \frac{b}{6} \end{aligned}$$

The pseudoinverse of ‘differentiation’ therefore returns a particular choice of anti-derivative, namely the unique anti-derivative of  $a+bx$  lying in  $\text{Span}\{\sqrt{5}(6x^2-6x+1), \sqrt{3}(2x-1)\}$ .

**Exercises 2.7** 1. Find the ingredients  $\beta, \gamma$  and the singular values for each of the following:

(a)  $T \in \mathcal{L}(\mathbb{R}^2, \mathbb{R}^3)$  where  $T \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x+y \\ x-y \end{pmatrix}$

(b)  $T : P_2(\mathbb{R}) \rightarrow P_1(\mathbb{R})$  and  $T(f) = f''$  where  $\langle f, g \rangle := \int_0^1 f(x)g(x) dx$

(c)  $V = W = \text{Span}\{1, \sin x, \cos x\}$  and  $\langle f, g \rangle = \int_0^{2\pi} f(x)g(x) dx$ , with  $T(f) = f' + 2f$

2. Find a singular value decomposition of each of the matrices:

$$(a) \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ -1 & -1 \end{pmatrix} \quad (b) \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & -1 \end{pmatrix} \quad (c) \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}$$

3. Find an explicit formula for  $T^\dagger$  for each map in Exercise 1.

4. Find the pseudoinverse of each of the matrix in Exercise 2.

5. Suppose  $A = P\Sigma Q^*$  is a singular value decomposition.

(a) Describe a singular value decomposition of  $A^*$ .

(b) Explain why  $A^\dagger = Q\Sigma^\dagger P^*$  isn't a singular value decomposition of  $A^\dagger$ ; what would be the correct decomposition? (Hint: what is wrong with  $\Sigma^\dagger$ ?)

6. Suppose  $T : V \rightarrow W$  is written according to the singular value theorem. Prove that  $\gamma$  is a basis of eigenvectors of  $TT^*$  with the same non-zero eigenvalues as  $T^*T$ , including repetitions.

7. (a) Suppose  $T = \mathcal{L}(V)$  is normal. Prove that each  $\mathbf{v}_j$  in the singular value theorem may be chosen to be an eigenvector of  $T$  and that  $\sigma_j$  is the modulus of the corresponding eigenvalue.

(b) Let  $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ . Show that any orthonormal basis  $\beta$  of  $\mathbb{R}^2$  satisfies the singular value theorem. What is  $\gamma$  here? What is it about the eigenvalues of  $A$  that make this possible?

(Even when  $T$  is self-adjoint, the vectors in  $\beta$  need not also be eigenvectors of  $T$ !)

8. In the proof of the singular value theorem we claimed that  $\text{rank } T^*T = \text{rank } T$ . Verify this by checking explicitly that  $\mathcal{N}(T^*T) = \mathcal{N}(T)$ .

(This is circular logic if you use the decomposition, so you must do without!)

9. Let  $V, W$  be finite-dimensional inner product spaces and  $T \in \mathcal{L}(V, W)$ . Prove:

(a)  $T^*TT^\dagger = T^\dagger TT^* = T^*$ .

(Hint: evaluate on the basis  $\gamma = \{\mathbf{w}_1, \dots, \mathbf{w}_m\}$  in the singular value theorem)

(b) If  $T$  is injective, then  $T^*T$  is invertible and  $T^\dagger = (T^*T)^{-1}T^*$ .

(c) If  $T$  is surjective, then  $TT^*$  is invertible and  $T^\dagger = T^*(TT^*)^{-1}$ .

10. Consider the equation  $T(\mathbf{x}) = \mathbf{b}$ , where  $T$  is a linear map between finite-dimensional inner product spaces. A least-squares solution is a vector  $\mathbf{x}$  which minimizes  $\|T(\mathbf{x}) - \mathbf{b}\|$ .

(a) Prove that  $\mathbf{x}_0 = T^\dagger(\mathbf{b})$  is a least-squares solution and that any other has the form  $\mathbf{x}_0 + \mathbf{n}$  for some  $\mathbf{n} \in \mathcal{N}(T)$ .

(Hint: Theorem 2.61 says that  $\mathbf{x}_0$  is a least-squares solution if and only if  $T(\mathbf{x}_0) = \pi_{\mathcal{R}(T)}(\mathbf{b})$ )

(b) Prove that  $\mathbf{x}_0 = T^\dagger(\mathbf{b})$  has smaller norm than any other least-squares solution.

(c) If  $T$  is injective, prove that  $\mathbf{x}_0 = T^\dagger(\mathbf{b})$  is the unique least-squares solution.

11. Find the minimal norm solution to the first system, and the least-squares solution to the second:

$$\begin{cases} 3x + 2y + z = 9 \\ x - 2y + 3z = 3 \end{cases} \quad \begin{cases} 3x + y = 1 \\ 2x - 2y = 0 \\ x + 3y = 0 \end{cases}$$

### Linear Regression (non-examinable)

Given a data set  $\{(t_j, y_j) : 1 \leq j \leq m\}$ , we may employ the least-squares method to find a *best-fitting line*  $y = c_0 + c_1t$ ; often used to predict  $y$  given a value of  $t$ .

The trick is to minimize the sum of the squares of the **vertical deviations** of the line from the data set.

$$\sum_{j=1}^m |y_j - c_0 - c_1t_j|^2 = \|\mathbf{y} - A\mathbf{x}\|^2 \quad \text{where} \quad A = \begin{pmatrix} t_1 & 1 \\ \vdots & \vdots \\ t_m & 1 \end{pmatrix} \quad \mathbf{x} = \begin{pmatrix} c_1 \\ c_0 \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$$

With the indicated notation, we recognize this as a least-squares problem. Indeed if there are at least two distinct  $t$ -values in the data set, then  $\text{rank } A = 2$  is maximal and we have a unique best-fitting line with coefficients given by

$$\begin{pmatrix} c_1 \\ c_0 \end{pmatrix} = A^\dagger \mathbf{y} = (A^T A)^{-1} A^T \mathbf{y}$$

**Example 2.69.** Given the data set  $\{(0, 1), (1, 1), (2, 0), (3, 2), (4, 2)\}$ , we compute

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 2 \\ 2 \end{pmatrix} \implies \mathbf{x}_0 = \begin{pmatrix} c_1 \\ c_0 \end{pmatrix} = (A^T A)^{-1} A^T \mathbf{y} = \begin{pmatrix} 30 & 10 \\ 10 & 5 \end{pmatrix}^{-1} \begin{pmatrix} 15 \\ 6 \end{pmatrix} = \frac{3}{10} \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

The regression line therefore has equation  $y = \frac{3}{10}(t + 2)$ .

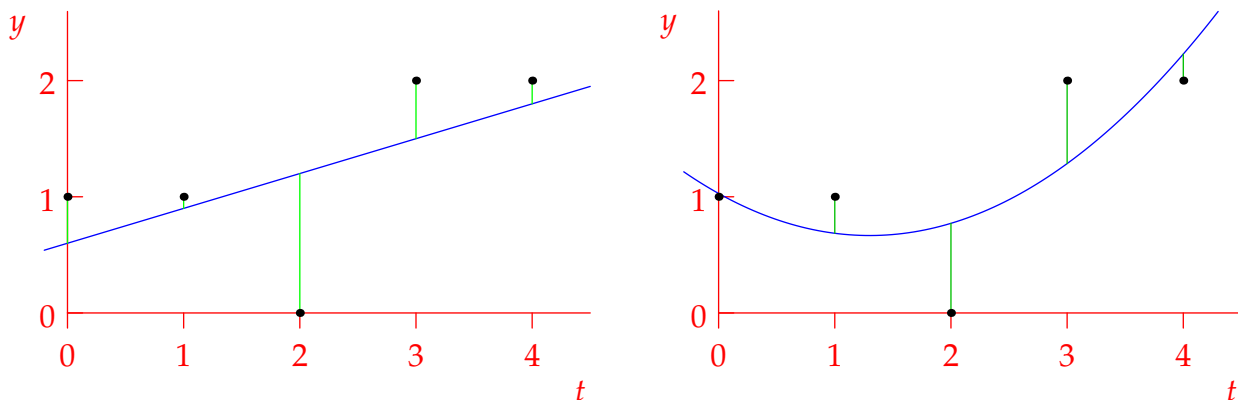
The process can be applied more generally to approximate using other functions. To find the best-fitting quadratic polynomial  $y = c_0 + c_1t + c_2t^2$ , we'd instead work with

$$A = \begin{pmatrix} t_1^2 & t_1 & 1 \\ \vdots & \vdots & \vdots \\ t_m^2 & t_m & 1 \end{pmatrix} \quad \mathbf{x} = \begin{pmatrix} c_2 \\ c_1 \\ c_0 \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \implies \sum_{j=1}^m |y_j - c_0 - c_1t_j - c_2t_j^2|^2 = \|\mathbf{y} - A\mathbf{x}\|^2$$

Provided we have at least three distinct values  $t_1, t_2, t_3$ , the matrix  $A$  is guaranteed to have rank 3 and there will be a best-fitting least-squares quadratic: in this case

$$y = \frac{1}{70}(15t^2 - 39t + 72)$$

This curve and the best-fitting straight line are shown below.



**Optional Problems** Use a computer to invert any  $3 \times 3$  matrices!

1. Check the calculation for the best-fitting least-squares quadratic in Example 2.69.
2. Find the best-fitting least-squares linear and quadratic approximations to the data set

$$\{(1, 2), (3, 4), (5, 7), (7, 9), (9, 12)\}$$

3. Suppose a data set  $\{(t_j, y_j) : 1 \leq j \leq m\}$  has unique regression line  $y = ct + d$ .

(a) Show that the equations  $A^T A \mathbf{x}_0 = A^T \mathbf{y}$  can be written in matrix form

$$\begin{pmatrix} \sum t_j^2 & \sum t_j \\ \sum t_j & m \end{pmatrix} \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} \sum t_j y_j \\ \sum y_j \end{pmatrix}$$

(b) Recover the standard expressions from statistics:

$$c = \frac{\text{Cov}(t, y)}{\sigma_t^2} \quad \text{and} \quad d = \bar{y} - c\bar{t}$$

where

- $\bar{t} = \frac{1}{m} \sum_{j=1}^m t_j$  and  $\bar{y} = \frac{1}{m} \sum_{j=1}^m y_j$  are the *means* (averages),
- $\sigma_t^2 = \frac{1}{m} \sum_{j=1}^m (t_j - \bar{t})^2$  is the *variance*,
- $\text{Cov}(t, y) = \frac{1}{m} \sum_{j=1}^m (t_j - \bar{t})(y_j - \bar{y})$  is the *covariance*.

## 2.8 Bilinear and Quadratic Forms

In this section we slightly generalize the idea of an inner product. Throughout,  $V$  is a vector space over a field  $\mathbb{F}$ ; it need not be an inner product space and  $\mathbb{F}$  can be any field (not just  $\mathbb{R}$  or  $\mathbb{C}$ ).

**Definition 2.70.** A bilinear form  $B : V \times V \rightarrow \mathbb{F}$  is linear in each entry:  $\forall \mathbf{v}, \mathbf{x}, \mathbf{y} \in V, \lambda \in \mathbb{F}$ ,

$$B(\lambda \mathbf{x} + \mathbf{y}, \mathbf{v}) = \lambda B(\mathbf{x}, \mathbf{v}) + B(\mathbf{y}, \mathbf{v}), \quad B(\mathbf{v}, \lambda \mathbf{x} + \mathbf{y}) = \lambda B(\mathbf{v}, \mathbf{x}) + B(\mathbf{v}, \mathbf{y})$$

Additionally,  $B$  is symmetric if  $\forall \mathbf{x}, \mathbf{y} \in V, B(\mathbf{x}, \mathbf{y}) = B(\mathbf{y}, \mathbf{x})$ .

**Examples 2.71.** 1. If  $V$  is a real inner product space, then the inner product  $\langle \cdot, \cdot \rangle$  is a symmetric bilinear form. Note that a complex inner product is *not bilinear!*

2. If  $A \in M_n(\mathbb{F})$ , then  $B(\mathbf{x}, \mathbf{y}) := \mathbf{x}^T A \mathbf{y}$  is a bilinear form on  $\mathbb{F}^n$ . For instance, on  $\mathbb{R}^2$ ,

$$B(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \begin{pmatrix} 1 & 2 \\ 2 & 0 \end{pmatrix} \mathbf{y} = x_1 y_1 + 2x_1 y_2 + 2x_2 y_1$$

defines a symmetric bilinear form, though not an inner product since it isn't positive definite; for example  $B(\mathbf{j}, \mathbf{j}) = 0$ .

As seen above, we often make use of a matrix.

**Definition 2.72.** Let  $B$  be a bilinear form on a finite-dimensional space with basis  $\epsilon = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ . The matrix of  $B$  with respect to  $\epsilon$  is the matrix  $[B]_\epsilon = A \in M_n(\mathbb{F})$  with  $ij^{\text{th}}$  entry

$$A_{ij} = B(\mathbf{v}_i, \mathbf{v}_j)$$

Given  $\mathbf{x}, \mathbf{y} \in V$ , compute their co-ordinate vectors  $[\mathbf{x}]_\epsilon, [\mathbf{y}]_\epsilon$  with respect to  $\epsilon$ , then

$$B(\mathbf{x}, \mathbf{y}) = [\mathbf{x}]_\epsilon^T A [\mathbf{y}]_\epsilon$$

The set of bilinear forms on  $V$  is therefore in bijective correspondence with  $M_n(\mathbb{F})$ . Moreover,

$$B(\mathbf{y}, \mathbf{x}) = [\mathbf{y}]_\epsilon^T A [\mathbf{x}]_\epsilon = \left([\mathbf{y}]_\epsilon^T A [\mathbf{x}]_\epsilon\right)^T = [\mathbf{x}]_\epsilon^T A^T [\mathbf{y}]_\epsilon$$

Finally, if  $\beta$  is another basis of  $V$ , then an appeal to the change of co-ordinate matrix  $Q_\epsilon^\beta$  yields

$$B(\mathbf{x}, \mathbf{y}) = [\mathbf{x}]_\epsilon^T A [\mathbf{y}]_\epsilon = (Q_\beta^\epsilon [\mathbf{x}]_\beta)^T A (Q_\beta^\epsilon [\mathbf{y}]_\beta) = [\mathbf{x}]_\beta^T (Q_\beta^\epsilon)^T A Q_\beta^\epsilon [\mathbf{y}]_\beta \implies [B]_\beta = (Q_\beta^\epsilon)^T [B]_\epsilon Q_\beta^\epsilon$$

To summarize:

**Lemma 2.73.** Let  $B$  be a bilinear form on a finite-dimensional vector space.

1. If  $A$  is the matrix of  $B$  with respect to some basis, then every other matrix of  $B$  has the form  $Q^T A Q$  for some invertible  $Q$ .
2.  $B$  is symmetric if and only if its matrix with respect to any (and all) bases is symmetric.

Naturally, the simplest situation is when the matrix of  $B$  is diagonal...



**Examples 2.74.** 1. Example 2.71.2 can be written

$$\begin{aligned} B(\mathbf{x}, \mathbf{y}) &= \mathbf{x}^T \begin{pmatrix} 1 & 2 \\ 2 & 0 \end{pmatrix} \mathbf{y} = x_1 y_1 + 2x_1 y_2 + 2x_2 y_1 = (x_1 + 2x_2)(y_1 + 2y_2) - 4x_2 y_2 \\ &= (x_1 + 2x_2 \quad x_2) \begin{pmatrix} 1 & 0 \\ 0 & -4 \end{pmatrix} \begin{pmatrix} y_1 + 2y_2 \\ y_2 \end{pmatrix} = \mathbf{x}^T \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}^T \begin{pmatrix} 1 & 0 \\ 0 & -4 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \mathbf{y} \end{aligned}$$

If  $\epsilon$  is the standard basis, then  $[B]_\beta = \begin{pmatrix} 1 & 0 \\ 0 & -4 \end{pmatrix}$  where  $Q_\epsilon^\beta = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ . It follows that  $Q_\beta^\epsilon = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}$  from which  $\beta = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \end{pmatrix} \right\}$  is a diagonalizing basis.

2. In general, we may perform a sequence of *simultaneous* row and column operations to diagonalize *any* symmetric  $B$ ; we require only elementary matrices  $E_{ij}^{(\lambda)}$  of type III.<sup>3</sup> For instance:

$$B(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T A \mathbf{y} = \mathbf{x}^T \begin{pmatrix} 1 & -2 & 3 \\ -2 & 0 & 4 \\ 3 & 4 & -1 \end{pmatrix} \mathbf{y} \quad (A = [B]_\epsilon \text{ with respect to the standard basis})$$

$$E_{21}^{(2)} A E_{12}^{(2)} = \begin{pmatrix} 1 & 0 & 3 \\ 0 & -4 & 10 \\ 3 & 10 & -1 \end{pmatrix} \quad (\text{add twice row 1 to row 2, columns similarly})$$

$$E_{31}^{(-3)} E_{21}^{(2)} A E_{12}^{(2)} E_{13}^{(-3)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -4 & 10 \\ 0 & 10 & -10 \end{pmatrix} \quad (\text{subtract 3 times row 1 from row 3, etc.})$$

$$E_{23}^{(1)} E_{31}^{(-3)} E_{21}^{(2)} A E_{12}^{(2)} E_{13}^{(-3)} E_{32}^{(1)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & -10 \end{pmatrix} \quad (\text{add row 3 to row 2, etc.})$$

If  $\beta$  is the diagonalizing basis, the the change of co-ordinate matrix is

$$Q_\beta^\epsilon = E_{12}^{(2)} E_{13}^{(-3)} E_{32}^{(1)} = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 & -3 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

from which  $\beta = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -3 \\ 0 \\ 1 \end{pmatrix} \right\}$ . If you're having trouble believing this, invert the change of co-ordinate matrix and check that

$$B(\mathbf{x}, \mathbf{y}) = (x_1 - 2x_2 + 3x_3)(y_1 - 2y_2 + 3y_3) + 6x_2 y_2 - 10(-x_2 + x_3)(-y_2 + y_3)$$

**Warning!** If  $\mathbb{F} = \mathbb{R}$  then every symmetric  $B$  may be diagonalized by an orthonormal basis (for the usual dot product on  $\mathbb{R}^n$ ). It is *very* unlikely that our algorithm will produce such! The algorithm has two main advantages over the spectral theorem: it is typically faster and it applies to vector spaces over any field. As a disadvantage, it is highly non-unique.

<sup>3</sup>Recall that  $E_{ij}^{(\lambda)}$  is the identity matrix with an additional  $\lambda$  in the  $ij^{\text{th}}$  entry.

- As a *column* operation (right-multiplication),  $A \mapsto A E_{ij}^{(\lambda)}$  adds  $\lambda$  times the  $i^{\text{th}}$  column to the  $j^{\text{th}}$ .
- As a *row* operation (left-multiplication),  $A \mapsto E_{ji}^{(\lambda)} A = (E_{ij}^{(\lambda)})^T A$  adds  $\lambda$  times the  $i^{\text{th}}$  row to the  $j^{\text{th}}$ .

**Example 2.75.** We diagonalize  $B(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \begin{pmatrix} 1 & 6 \\ 6 & 3 \end{pmatrix} \mathbf{y} = x_1y_1 + 6x_1y_2 + 6x_2y_1 + 3x_2y_2$  in three ways.

- $\begin{pmatrix} 1 & 0 \\ -6 & 1 \end{pmatrix} \begin{pmatrix} 1 & 6 \\ 6 & 3 \end{pmatrix} \begin{pmatrix} 1 & -6 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -33 \end{pmatrix} = [B]_\beta$  where  $\beta = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -6 \\ 1 \end{pmatrix} \right\}$ . This corresponds to

$$B(\mathbf{x}, \mathbf{y}) = (x_1 + 6x_2)(y_1 + 6y_2) - 33x_2y_2$$

- $\begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 6 \\ 6 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix} = \begin{pmatrix} -11 & 0 \\ 0 & 3 \end{pmatrix} = [B]_\gamma$  where  $\gamma = \left\{ \begin{pmatrix} 1 \\ -2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$ . This corresponds to

$$B(\mathbf{x}, \mathbf{y}) = -11x_1y_1 + 3(2x_1 + x_2)(2y_1 + y_2)$$

- If  $\mathbb{F} = \mathbb{R}$ , we may apply the spectral theorem to see that  $[B]_\eta = \begin{pmatrix} 2+\sqrt{37} & 0 \\ 0 & 2-\sqrt{37} \end{pmatrix}$  is diagonal with respect to  $\eta = \left\{ \begin{pmatrix} 6 \\ 1+\sqrt{37} \end{pmatrix}, \begin{pmatrix} -1-\sqrt{37} \\ 6 \end{pmatrix} \right\}$ . The expression for  $B(\mathbf{x}, \mathbf{y})$  in these co-ordinates is disgusting, so we omit it; that  $B$  can be diagonalized orthogonally doesn't mean it should be!

**Theorem 2.76.** Suppose  $B$  is a bilinear form of a finite-dimensional space  $V$  over  $\mathbb{F}$ .

1. If  $B$  is diagonalizable, then it is symmetric.
2. If  $B$  is symmetric and  $\mathbb{F}$  does not have characteristic two (see aside), then  $B$  is diagonalizable.

*Proof.* 1. If  $B$  is diagonalizable,  $\exists \beta$  such that  $[B]_\beta$  is diagonal and thus symmetric.

2. Suppose  $B$  is non-zero (otherwise the result is trivial). We prove by induction on  $n = \dim V$ .

If  $n = 1$ , the result is trivial:  $B(x, y) = axy$  for some  $a \in \mathbb{F}$  is clearly symmetric.

Fix  $n \in \mathbb{N}$  and assume that every non-zero symmetric bilinear form on a dimension  $n$  vector space over  $\mathbb{F}$  is diagonalizable. Let  $\dim V = n + 1$ . By the discussion below,  $\exists \mathbf{x} \in V$  such that  $B(\mathbf{x}, \mathbf{x}) \neq 0$ . Consider the linear map

$$T : V \rightarrow \mathbb{F} : \mathbf{v} \mapsto B(\mathbf{x}, \mathbf{v})$$

Clearly  $\text{rank } T = 1 \implies \dim \mathcal{N}(T) = n$ . Moreover,  $B$  is symmetric when restricted to  $\mathcal{N}(T)$ ; by the induction hypothesis there exists a basis  $\beta$  of  $\mathcal{N}(T)$  such that  $[B_{\mathcal{N}(T)}]_\beta$  is diagonal. But then  $B$  is diagonal with respect to the basis  $\beta \cup \{\mathbf{x}\}$ . ■

**Aside: Characteristic two fields** This means  $1 + 1 = 0$  in  $\mathbb{F}$ ; this holds, for instance, in the field  $\mathbb{Z}_2 = \{0, 1\}$  of remainders modulo 2. We now see the importance of  $\text{char } \mathbb{F} \neq 2$  to the above result.

- The proof uses the existence of  $\mathbf{x} \in V$  such that  $B(\mathbf{x}, \mathbf{x}) \neq 0$ . If  $B$  is non-zero,  $\exists \mathbf{u}, \mathbf{v}$  such that  $B(\mathbf{u}, \mathbf{v}) \neq 0$ . If both  $B(\mathbf{u}, \mathbf{u}) = 0 = B(\mathbf{v}, \mathbf{v})$ , then  $\mathbf{x} = \mathbf{u} + \mathbf{v}$  does the job whenever  $\text{char } \mathbb{F} \neq 2$ :

$$B(\mathbf{x}, \mathbf{x}) = B(\mathbf{u}, \mathbf{v}) + B(\mathbf{v}, \mathbf{u}) = 2B(\mathbf{u}, \mathbf{v}) \neq 0$$

- To see that the requirement isn't idle, consider  $B(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \mathbf{y}$  on the finite vector space  $\mathbb{Z}_2^2 = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$  over  $\mathbb{Z}_2$ . Every element of this space satisfies  $B(\mathbf{x}, \mathbf{x}) = 0$ ! Perhaps surprisingly, the matrix of  $B$  is identical with respect to *any* basis of  $\mathbb{Z}_2^2$ , whence  $B$  is symmetric but non-diagonalizable.

In Example 2.75 notice how the three diagonal matrix representations have something in common: one each of the diagonal entries are positive and negative. This is a general phenomenon:

**Theorem 2.77 (Sylvester's Law of Inertia).** Suppose  $B$  is a symmetric bilinear form on a real vector space  $V$  with diagonal matrix representation  $\text{diag}(\lambda_1, \dots, \lambda_n)$ . Then the number of entries  $\lambda_j$  which are positive/negative/zero is independent of the diagonal representation.

**Definition 2.78.** The *signature* of a symmetric bilinear form  $B$  is the triple  $(n_+, n_-, n_0)$  representing how many positive, negative and zero terms are in any diagonal representation. Sylvester's Law says that the signature is an *invariant* of a symmetric bilinear form.

Positive-definiteness says that a real inner product on an  $n$ -dimensional space has signature  $(n, 0, 0)$ . Practitioners of relativity often work in *Minkowski spacetime*:  $\mathbb{R}^4$  equipped with a signature  $(1, 3, 0)$  bilinear form, typically

$$B(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \begin{pmatrix} c^2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \mathbf{y} = c^2 x_1 y_1 - x_2 y_2 - x_3 y_3 - x_4 y_4$$

where  $c$  is the speed of light. Vectors are *time-*, *space-*, or *light-like* depending on whether  $B(\mathbf{x}, \mathbf{x})$  is positive, negative or zero. For instance  $\mathbf{x} = 3c^{-1}\mathbf{e}_1 + 2\mathbf{e}_2 + 2\mathbf{e}_3 + \mathbf{e}_4$  is light-like.

*Proof.* For simplicity, let  $V = \mathbb{R}^n$  and write  $B(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T A \mathbf{y}$  where  $A$  is symmetric.

1. Define  $\text{rank } B := \text{rank } A$  and observe this is independent of basis (exercises).
2. Let  $\beta, \gamma$  be diagonalizing bases, ordered according to whether  $B$  is **positive**, **negative** or **zero**.

$$\begin{aligned} \beta &= \{\mathbf{v}_1, \dots, \mathbf{v}_p, \mathbf{v}_{p+1}, \dots, \mathbf{v}_r, \mathbf{v}_{r+1}, \dots, \mathbf{v}_n\} \\ \gamma &= \{\mathbf{w}_1, \dots, \mathbf{w}_q, \mathbf{w}_{q+1}, \dots, \mathbf{w}_r, \mathbf{w}_{r+1}, \dots, \mathbf{w}_n\} \end{aligned}$$

Here  $r = \text{rank } B$  in accordance with part 1: our goal is to prove that  $p = q$ .

3. Assume  $p < q$ , define the matrix  $C$  and check what follows:

$$C = \begin{pmatrix} \mathbf{v}_1^T A \\ \vdots \\ \mathbf{v}_p^T A \\ \mathbf{w}_{q+1}^T A \\ \vdots \\ \mathbf{w}_r^T A \end{pmatrix} \in M_{(r-q+p) \times n}(\mathbb{R})$$

- (a)  $\text{rank } C \leq r - q + p < r \implies \text{null } C > n - r$ , thus

$$\exists \mathbf{x} \in \mathbb{R}^n \text{ such that } C\mathbf{x} = \mathbf{0} \text{ and } \mathbf{x} \notin \text{Span}\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$$

- (b) The first  $p$  entries of  $C\mathbf{x} = \mathbf{0}$  mean that  $\mathbf{x} \in \text{Span}\{\mathbf{v}_{p+1}, \dots, \mathbf{v}_r, \mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$  and so  $B(\mathbf{x}, \mathbf{x}) < 0$ . Note how we use part (a) to get a strict inequality here.
- (c) Now write  $\mathbf{x}$  with respect to  $\gamma$ : this time we see that  $\mathbf{x} \in \text{Span}\{\mathbf{w}_1, \dots, \mathbf{w}_q, \mathbf{w}_{r+1}, \dots, \mathbf{w}_n\}$ , whence  $B(\mathbf{x}, \mathbf{x}) \geq 0$ : a contradiction. ■

## Quadratic Forms & Diagonalizing Conics

**Definition 2.79.** To every symmetric bilinear form  $B : V \times V \rightarrow \mathbb{F}$  is associated a *quadratic form*

$$K : V \rightarrow \mathbb{F} : \mathbf{x} \mapsto B(\mathbf{x}, \mathbf{x})$$

A function  $K : V \rightarrow \mathbb{F}$  is termed a quadratic form when such a symmetric bilinear form exists.

**Examples 2.80.** 1. If  $B$  is a real inner product, then  $K(\mathbf{v}) = \langle \mathbf{v}, \mathbf{v} \rangle = \|\mathbf{v}\|^2$  is the square of the norm.

2. Let  $\dim V = n$  and  $A$  be the matrix of  $B$  with respect to a basis  $\beta$ . By the symmetry of  $A$ ,

$$K(\mathbf{x}) = \mathbf{x}^T A \mathbf{x} = \sum_{i,j=1}^n x_i A_{ij} x_j = \sum_{1 \leq i \leq j \leq n} \tilde{a}_{ij} x_i x_j \quad \text{where } \tilde{a}_{ij} = \begin{cases} A_{ij} & \text{if } i = j \\ 2A_{ij} & \text{if } i \neq j \end{cases}$$

E.g.,  $K(\mathbf{x}) = 3x_1^2 + 4x_2^2 - 2x_1x_2$  corresponds to the bilinear form  $B(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \begin{pmatrix} 3 & -1 \\ -1 & 4 \end{pmatrix} \mathbf{y}$

As a fun application, we consider the diagonalization of conics in  $\mathbb{R}^2$ . The general non-zero conic has equation

$$ax^2 + 2bxy + cy^2 + dx + ey + f = 0$$

where the first three terms comprise a quadratic form

$$K(\mathbf{x}) = ax^2 + 2bxy + cy^2 \iff B(\mathbf{v}, \mathbf{w}) = \mathbf{v}^T \begin{pmatrix} a & b \\ b & c \end{pmatrix} \mathbf{w}$$

If  $\{\mathbf{v}_1, \mathbf{v}_2\}$  is a diagonalizing basis, then there exist scalars  $\lambda_1, \lambda_2$  for which

$$K(t_1\mathbf{v}_1 + t_2\mathbf{v}_2) = \lambda_1 t_1^2 + \lambda_2 t_2^2$$

whence the general conic becomes

$$\lambda_1 t_1^2 + \lambda_2 t_2^2 + \mu_1 t_1 + \mu_2 t_2 = \eta, \quad \lambda_1, \lambda_2, \mu_1, \mu_2, \eta \in \mathbb{R}$$

If  $\lambda_i \neq 0$ , we may complete the square via the linear transformation  $s_j = t_j + \frac{\mu_j}{2\lambda_j}$ . The canonical forms are then recovered:

*Parabola:*  $\lambda_1$  or  $\lambda_2 = 0$  (but not both).

*Ellipse:*  $\lambda_1 s_1^2 + \lambda_2 s_2^2 = k \neq 0$  where  $\lambda_1, \lambda_2, k$  have the same sign.

*Hyperbola:*  $\lambda_1 s_1^2 + \lambda_2 s_2^2 = k \neq 0$  where  $\lambda_1, \lambda_2$  have opposite signs.

Since  $B$  is symmetric, we may take  $\{\mathbf{v}_1, \mathbf{v}_2\}$  to be an *orthonormal* basis of  $\mathbb{R}^2$ , whence any conic may be put in canonical form by applying only a rotation/reflection and translation (completing the square). Alternatively, we could diagonalize  $K$  using our earlier algorithm; this additionally permits shear transforms. By Sylvester's Law, the diagonal entries will have the same number of  $(+, -, 0)$  terms regardless of the method, so the canonical form will be unchanged.

**Examples 2.81.** 1. We describe and plot the conic with equation  $7x^2 + 24xy = 144$ .

The matrix of the associated bilinear form is  $\begin{pmatrix} 7 & 12 \\ 12 & 0 \end{pmatrix}$  which has orthonormal eigenbasis

$$\beta = \{\mathbf{v}_1, \mathbf{v}_2\} = \left\{ \frac{1}{5} \begin{pmatrix} 4 \\ 3 \end{pmatrix}, \frac{1}{5} \begin{pmatrix} -3 \\ 4 \end{pmatrix} \right\}$$

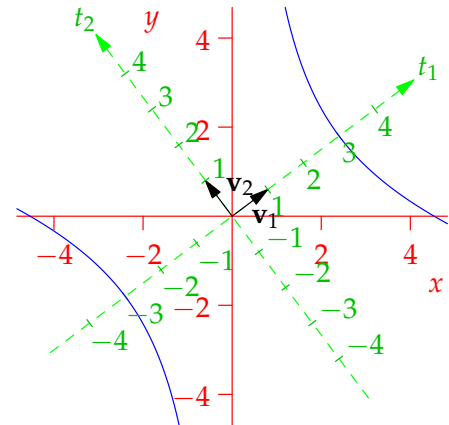
with eigenvalues  $(\lambda_1, \lambda_2) = (16, -9)$ . In the rotated basis, this is the canonical hyperbola

$$16t_1^2 - 9t_2^2 = 144 \iff \frac{t_1^2}{3^2} - \frac{t_2^2}{4^2} = 1$$

which is easily plotted. In case this is too fast, use the change of co-ordinate matrix to compute directly:

$$Q_\beta^\epsilon = \frac{1}{5} \begin{pmatrix} 4 & -3 \\ 3 & 4 \end{pmatrix} \implies \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} = [\mathbf{x}]_\beta = Q_\beta^\epsilon \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{5} \begin{pmatrix} 4x + 3y \\ -3x + 4y \end{pmatrix}$$

which quickly recovers the original conic by substitution.



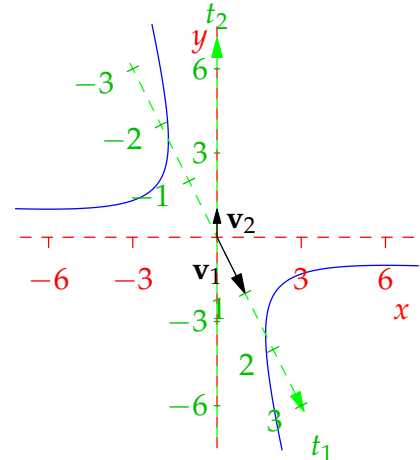
2. The conic defined by  $K(\mathbf{x}) = x^2 + 12xy + 3y^2 = -33$  defines a hyperbola in accordance with Example 2.75. With respect to the basis  $\gamma = \left\{ \begin{pmatrix} 1 \\ -2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$ , we see that

$$\begin{aligned} K(\mathbf{x}) &= -11x^2 + 3(2x + y)^2 = -11t_1^2 + 3t_2^2 = -33 \\ \iff \frac{t_1^2}{3} - \frac{t_2^2}{11} &= 1 \end{aligned}$$

Even though  $\gamma$  is non-orthogonal, we can still plot the conic! If we instead used the orthonormal basis  $\eta$ , we'd obtain

$$K(\mathbf{x}) = (\sqrt{37} + 2)s_1^2 - (\sqrt{37} - 2)s_2^2 = -33$$

however the calculation to find  $\eta$  is time-consuming and the expressions for  $s_1, s_2$  are extremely ugly.



A similar approach can be applied to higher degree quadratic equations/manifolds: e.g. ellipsoids, paraboloids and hyperboloids in  $\mathbb{R}^3$ .

**Exercises 2.8** 1. Prove that the sum of any two bilinear forms is bilinear, and that any scalar multiple of a bilinear form is bilinear: thus the set of bilinear forms on  $V$  is a vector space.

(You can't use matrices here, since  $V$  could be infinite-dimensional!)

2. Compute the matrix of the bilinear form

$$B(\mathbf{x}, \mathbf{y}) = x_1y_1 - 2x_1y_2 + x_2y_1 - x_3y_3$$

on  $\mathbb{R}^3$  with respect to the basis  $\beta = \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$ .

3. Check that the function  $B(f, g) = f'(0)g''(0)$  is a bilinear form on the vector space of twice-differentiable functions. Find the matrix of  $B$  with respect to  $\beta = \{\cos t, \sin t, \cos 2t, \sin 2t\}$  when restricted to the subspace  $\text{Span } \beta$ .

4. For each matrix  $A$ , find a diagonal matrix  $D$  and an invertible matrix  $Q$  such that  $Q^T A Q = D$ .

$$(a) \begin{pmatrix} 1 & 3 \\ 3 & 2 \end{pmatrix} \quad (b) \begin{pmatrix} 3 & 1 & 2 \\ 1 & 4 & 0 \\ 2 & 0 & -1 \end{pmatrix}$$

5. If  $K$  is a quadratic form and  $K(\mathbf{x}) = 2$ , what is the value of  $K(3\mathbf{x})$ ?

6. If  $\mathbb{F}$  does not have characteristic 2, and  $K(\mathbf{x}) = B(\mathbf{x}, \mathbf{x})$  is a quadratic form, prove that we can recover the bilinear form  $B$  via

$$B(\mathbf{x}, \mathbf{y}) = \frac{1}{2} (K(\mathbf{x} + \mathbf{y}) - K(\mathbf{x}) - K(\mathbf{y}))$$

7. If  $B(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \mathbf{y}$  is a bilinear form on  $\mathbb{F}^2$ , compute the quadratic form  $K(\mathbf{x})$ .

8. Suppose  $B$  is a symmetric bilinear form on a real finite-dimensional space. With reference to the proof of Sylvester's Law, explain why  $\text{rank } B$  is independent of the choice of diagonalizing basis.

9. If  $\text{char } \mathbb{F} \neq 2$ , apply the diagonalizing algorithm to the symmetric bilinear form  $B = \mathbf{x}^T \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \mathbf{y}$  on  $\mathbb{F}^2$ . What goes wrong if  $\text{char } \mathbb{F} = 2$ ?

10. Describe and plot the following conics:

(a)  $x^2 + y^2 + xy = 6$

(b)  $35x^2 + 120xy = 4x + 3y$

11. Suppose that a non-empty, non-degenerate<sup>4</sup> conic  $C$  in  $\mathbb{R}^2$  has the form  $ax^2 + 2bxy + cy^2 + dx + ey + f = 0$ , where at least one of  $a, b, c \neq 0$ , and define  $\Delta = b^2 - ac$ . Prove that:

- $C$  is a parabola if and only if  $\Delta = 0$ ;
- $C$  is an ellipse if and only if  $\Delta < 0$ ;
- $C$  is a hyperbola if and only if  $\Delta > 0$ .

(Hint:  $\lambda_1, \lambda_2$  are the eigenvalues of a symmetric matrix, so...)

---

<sup>4</sup>The conic contains at least two points and cannot be factorized as a product of two straight lines: for example, the following are disallowed;

- $x^2 + y^2 + 1 = 0$  is empty (unless one allows conics over  $\mathbb{C}$ ...)
- $x^2 + y^2 = 0$  contains only one point;
- $x^2 - xy - x + y = (x - 1)(x - y) = 0$  is the product of two lines.