Math 321: Linear Algebra

T. Kapitula *
Department of Mathematics and Statistics
University of New Mexico

September 8, 2004

Textbook: Linear Algebra, by J. Hefferon

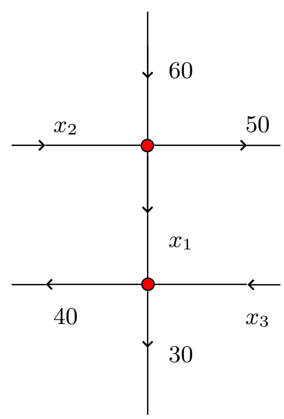
*E-mail: kapitula@math.unm.edu

Contents

1. Linear Systems

Two examples:

(a) Network flow. The assumption is that flow into nodes equals flow out of nodes, and that branches connect the various nodes. The nodes can be thought of as intersections, and the branches can be thought of as streets.



(b) Approximation of data - find the line of best fit (least-squares). For example, find a line which best fits the points $(1, 0), (2, 1), (4, 2), (5, 3)$. The answer is $y = -3/5 + 7/10\,x$.

1.1. Solving Linear Systems

1.1.1. Gauss' Method

Definition 1.1. A linear equation is of the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = d,$$

where

- $x_1, \ldots, x_n$: variables
- $a_1, \ldots, a_n \in \mathbb{R}$: coefficients

- $d \in \mathbb{R}$: constant.

A linear system is a collection of one or more linear equations. An n-tuple $(s_1, \ldots, s_n)$ is a solution if $x_1 = s_1, \ldots, x_n = s_n$ solves the system.

Example. (i) $3x_1 - 4x_2 + 6x_3 = 5$: linear equation

(ii) $x_1 x_2 - 3x_3 = 4$: nonlinear equation

Example. The system

$$2x_1 + x_2 = 8, \quad -2x_1 + 3x_2 = 16$$

has the solution $(x_1, x_2) = (1, 6)$.

Theorem 1.2 (Gauss' Theorem)**.** *If a linear system is changed from one to another by the operations*

(1) one equation is swapped with another (swapping)

(2) an equation is multiplied by a nonzero constant (rescaling)

(3) an equation is replaced by the sum of itself and a multiple of another (pivoting)

then the two systems have the same set of solutions.

Remark 1.3. These operations are known as the elementary reduction operations, row operations, or Gaussian operations.

Why? The idea is to convert a given system into an equivalent system which is easier to solve.

Example. Work the system

$$3x_1 + 6x_2 = 12, \quad x_1 - 2x_2 = 4,$$

which has the solution $(4, 0)$.

Definition 1.4. In each row, the first variable with a nonzero coefficient is the row's leading variable. A system is in echelon form if each leading variable is to the right of the leading variable in the row above it.

Example. Upon using the Gaussian operations one has that

$$\begin{array}{rcr} x_1 \quad\;\; - x_3 & = & 0 \\ 3x_1 + x_2 \quad\;\; & = & 1 \\ x_1 - x_2 - x_3 & = & -4 \end{array} \longrightarrow \begin{array}{rcr} x_1 - x_3 & = & 0 \\ x_2 \quad & = & 4 \\ x_3 & = & -1 \end{array}$$

The solution is $(-1, 4, -1)$.

Theorem 1.5. *A linear system has either*

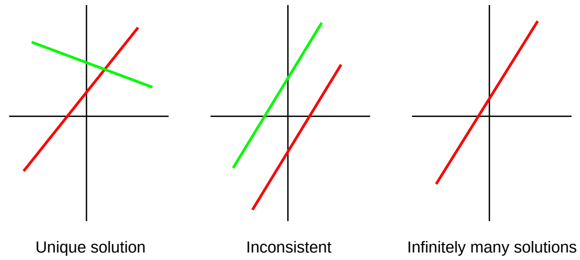
(1) no solution (inconsistent)

(2) a unique solution

(3) an infinite number of solutions.

In the latter two cases the system is consistent.

Illustrate this graphically with systems of two equations in two unknowns.



1.1.2. Describing the Solution Set

Definition 1.6. The variables in an echelon-form linear system which are not leading variables are free variables.

Example. For the system

$$x_1 + 2x_2 - 4x_4 = 2, \quad x_3 - 7x_4 = 8$$

the leading variables are x_1, x_3, and the free variables are x_2, x_4. The solution is parameterized by the free variables via

$$x_1 = 2 - 2x_2 + 4x_4, \quad x_3 = 8 + 7x_4,$$

so that the solution set is

$$\{(2 - 2x_2 + 4x_4, x_2, 8 + 7x_4, x_4) \,:\, x_2, x_4 \in \mathbb{R}\}.$$

Definition 1.7. An $m \times n$ matrix A is a rectangular array of numbers with m rows and n columns. If the numbers are real-valued, then we say that $A = (a_{i,j}) \in \mathbb{R}^{m\times n}$, where $a_{i,j} \in \mathbb{R}$ is the entry in row i and column j.

Example. Find different entries for the matrix

$$A = \begin{pmatrix} 1 & 3 & -4 \\ -2 & 6 & -3 \end{pmatrix}$$

Definition 1.8. A vector (column vector) is a matrix with a single column, i.e., $\vec{a} \in \mathbb{R}^{n\times 1} := \mathbb{R}^n$. A matrix with a single row is a row vector. The entries of a vector are its components.

Remark 1.9. For vectors we will use the notation $\vec{a} = (a_i) \in \mathbb{R}^n$.

Definition 1.10. Consider two vectors $\vec{u} = (u_i), \vec{v} = (v_i) \in \mathbb{R}^n$. The algebraic operations are:

(i) vector sum: $\vec{u} + \vec{v} = (u_i + v_i)$

(ii) scalar multiplication: $r\vec{v} = (rv_i)$ for any $r \in \mathbb{R}$.

Consider the linear system

$$x_1 - 3x_2 + 2x_3 = 8, \quad x_1 - x_3 = 5, \quad -2x_2 + x_3 = 7.$$

Matrices associated with this system are

$$\begin{pmatrix} 1 & -3 & 2 \\ 1 & 0 & -1 \\ 0 & -2 & 1 \end{pmatrix} \quad \text{(coefficient matrix)}, \qquad \left(\begin{array}{ccc|c} 1 & -3 & 2 & 8 \\ 1 & 0 & -1 & 5 \\ 0 & -2 & 1 & 7 \end{array}\right) \quad \text{(augmented matrix)}.$$

The Gaussian operations can be performed on the augmented matrix to put the system into echelon form:

$$\left(\begin{array}{ccc|c} 1 & -3 & 2 & 8 \\ 1 & 0 & -1 & 5 \\ 0 & -2 & 1 & 7 \end{array}\right) \longrightarrow \left(\begin{array}{ccc|c} 1 & -3 & 2 & 8 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & -5 \end{array}\right).$$

Why? It eases the bookkeeping. The solution is $s = (0, -6, -5)$, which can be written in vector notation as

$$\vec{s} = \begin{pmatrix} 0 \\ -6 \\ -5 \end{pmatrix}.$$

Example. The solution to the system

$$x_1 + 2x_2 - 4x_4 = 2, \quad x_3 - 7x_4 = 8$$

is

$$\{(2 - 2x_2 + 4x_4, x_2, 8 + 7x_4, x_4) \, : \, x_2, x_4 \in \mathbb{R}\}.$$

Using vector notation yields

$$\{\begin{pmatrix} 2 \\ 0 \\ 8 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 4 \\ 0 \\ 7 \\ 1 \end{pmatrix} \, : \, x_2, x_4 \in \mathbb{R}\}.$$

It is clear that the system has infinitely many solutions.

1.1.3. General=Particular+Homogeneous

In the previous example it is seen that the solution has two parts: a particular solution which depends upon the right-hand side, and a homogeneous solution, which is independent of the right-hand side. We will see that this feature holds for any linear system. Recall that equation j in a linear system has the form

$$a_{j,1}x_1 + \cdots + a_{j,n}x_n = d_j.$$

Definition 1.11. A linear equation is homogeneous if it has a constant of zero. A linear system is homogeneous is all of the constants are zero.

Remark 1.12. A homogeneous system always has at least one solution, the zero vector $\vec{0}$.

Lemma 1.13. *For any homogeneous linear system there exist vectors $\vec{\beta}_1, \ldots, \vec{\beta}_k$ such that any solution of the system is of the form*

$$\vec{x} = c_1\vec{\beta}_1 + \cdots c_k\vec{\beta}_k, \quad c_1, \ldots, c_k \in \mathbb{R}.$$

Here k is the number of free variables in an echelon form of the system.

Definition 1.14. The set $\{c_1\vec{\beta}_1 + \cdots c_k\vec{\beta}_k \,:\, c_1,\dots,c_k \in \mathbb{R}\}$ is the span of the vectors $\{\vec{\beta}_1,\dots,\vec{\beta}_k\}$.

Proof: The augmented matrix for the system is of the form $(A|\vec{0})$, where $A \in \mathbb{R}^{m\times n}$. Use Gauss' method to reduce the system to echelon form. Furthermore, use the Gauss-Jordan reduction discussed in Section 3.1 to put the system in reduced echelon form. The coefficient associated with each leading variable (the leading entry) will then be one, and there will be zeros above and below each leading entry in the reduced matrix. For example,

$$A \longrightarrow \begin{pmatrix} 1 & 0 & 0 & 6 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

In row j the reduced system is then of the form

$$x_{\ell_j} + a_{j,\ell_j+1}x_{\ell_j+1} + \cdots + a_{j,n}x_n = 0,$$

which can be rewritten as

$$x_{\ell_j} = -a_{j,\ell_j+1}x_{\ell_j+1} - \cdots - a_{j,n}x_n.$$

Since the system is in echelon form, $\ell_{j-1} < \ell_j < \ell_{j+1}$. Suppose that the free variables are labelled as $x_{f_1},\dots,x_{f_k}$. Since the system is in reduced echelon form, one has that in the above equation $a_{j,\ell_j+i} = 0$ for any $i \ge 1$ such that x_{ℓ_j+i} is a leading variable. Thus, after a renaming of the variables the above equation can be rewritten as

$$x_{\ell_j} = \beta_{j,f_1}x_{f_1} + \cdots + \beta_{j,f_k}x_{f_k}.$$

The vectors $\vec{\beta}_j$, $j = 1,\dots,k$, can now be constructed, and in vector form the solution is given by

$$\vec{x} = x_{f_1}\vec{\beta}_1 + \cdots + x_{f_k}\vec{\beta}_k.$$

□

Lemma 1.15. *Let $\vec{p}$ be a particular solution for a linear system. The solution set is given by*

$$\{\vec{p} + \vec{h} \,:\, \vec{h} \text{ is a homogeneous solution}\}.$$

Proof: Let $\vec{s}$ be any solution, and set $\vec{h} = \vec{s} - \vec{p}$. In row j we have that

$$\begin{aligned} a_{j,1}(s_1 - p_1) + \cdots + a_{j,n}(s_n - p_n) &= (a_{j,1}s_1 + \cdots + a_{j,n}s_n) - (a_{j,1}p_1 + \cdots + a_{j,n}p_n) \\ &= d_j - d_j \\ &= 0, \end{aligned}$$

so that $\vec{h} = \vec{s} - \vec{p}$ solves the homogeneous equation.

Now take a vector of the form $\vec{p} + \vec{h}$, where $\vec{p}$ is a particular solution and $\vec{h}$ is a homogeneous solution. Similar to above,

$$\begin{aligned} a_{j,1}(p_1 + h_1) + \cdots + a_{j,n}(p_n + h_n) &= (a_{j,1}p_1 + \cdots + a_{j,n}p_n) + (a_{j,1}h_1 + \cdots + a_{j,n}h_n) \\ &= d_j + 0 \\ &= d_j, \end{aligned}$$

so that $\vec{s} = \vec{p} + \vec{h}$ solves the system. □

Remark 1.16. While a homogeneous solution always exists, this is not the case for a particular solution.

Example. Consider the linear system

$$\begin{array}{rcl} x_1 \quad\quad + \ x_3 + \ x_4 = -1 & & x_1 \quad + \ x_3 + x_4 = \ -1 \\ 2x_1 - x_2 \quad\quad + \ x_4 = \ \ 3 & \longrightarrow & x_2 + 2x_3 + x_4 = \ \ 5 \\ x_1 + x_2 + 3x_3 + 2x_4 = \ \ b & & 0 = b + 1 \end{array}$$

The homogeneous solution is given by

$$\vec{h} = x_3 \begin{pmatrix} -1 \\ -2 \\ 1 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -1 \\ -1 \\ 0 \\ 1 \end{pmatrix}, \quad x_3, x_4 \in \mathbb{R}.$$

If $b \neq -1$, then no particular solution exists. Otherwise, one has that

$$\vec{p} = \begin{pmatrix} -1 \\ 5 \\ 0 \\ 0 \end{pmatrix}.$$

If $b = -1$, the solution set is given by $\vec{x} = \vec{p} + \vec{h}$.

Definition 1.17. A square matrix is nonsingular if it is the coefficient matrix for a homogeneous system with the unique solution $\vec{x} = \vec{0}$. Otherwise, it is singular.

Remark 1.18. In order for a square matrix to be nonsingular, it must be true that for the row-reduced matrix there are no free variables. Consider the two examples:

$$A \longrightarrow \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}, \quad B \longrightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The homogeneous system associated with A has infinitely many solutions, whereas the one associated with B has only one.

1.3. Reduced Echelon Form

1.3.1. Gauss-Jordan Reduction

Definition 1.19. A matrix is in reduced echelon form if

(a) it is in echelon form

(b) each leading entry is a one

(c) each leading entry is the only nonzero entry in its column.

Definition 1.20. The Gauss-Jordan reduction is the process of putting a matrix into reduced echelon form.

Example. Consider

$$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{array}\right) \xrightarrow{\text{echelon}} \left(\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right) \xrightarrow{\text{reduced echelon}} \left(\begin{array}{ccc|c} 1 & 0 & -1 & -2 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

The solution is then given by

$$\vec{x} = \begin{pmatrix} -2 \\ 3 \\ 0 \end{pmatrix} + c_1 \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}, \quad c_1 \in \mathbb{R}.$$

Remark 1.21. The above coefficient matrix is singular.

Definition 1.22. Two matrices are row equivalent if they can be row-reduced to a common third matrix by the elementary row operations.

Example. This can be written as $A \longrightarrow C \longleftarrow B$, i.e., from above

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Remark 1.23. (a) Elementary row operations are reversible.

(b) If two coefficient matrices are row equivalent, then the associated homogeneous linear systems have the same solution.

1.3.2. Row Equivalence

Definition 1.24. A linear combination of the vectors $\vec{x}_1, \ldots, \vec{x}_n$ is an expression of the form

$$\sum_i c_i \vec{x}_i = c_1 \vec{x}_1 + \cdots + c_n \vec{x}_n,$$

where $c_1, \ldots, c_n \in \mathbb{R}$.

Remark 1.25. The span of the set $\{\vec{x}_1, \ldots, \vec{x}_n\}$ is the set of all linear combinations of the vectors.

Lemma 1.26 (Linear Combination Lemma)**.** *A linear combination of linear combinations is a linear combination.*

Proof: Let the linear combinations $\sum_i c_{1,i}\vec{x}_i$ through $\sum_i c_{m,i}\vec{x}_i$ be given, and consider the new linear combination

$$d_1 \left(\sum_{i=1}^n c_{1,i}\vec{x}_i \right) + \cdots + d_m \left(\sum_{i=1}^n c_{m,i}\vec{x}_i \right).$$

Multiplying out and regrouping yields

$$\left(\sum_{i=1}^m d_i c_{i,1} \right) \vec{x}_1 + \cdots + \left(\sum_{i=1}^m d_i c_{i,n} \right) \vec{x}_n,$$

which is again a linear combination of $\vec{x}_1, \ldots, \vec{x}_n$. $\square$

Corollary 1.27. *If two matrices are row equivalent, then each row of the second is a linear combination of the rows of the first.*

Proof: The idea is that one row-reduces a matrix by taking linear combinations of rows. If A and B are row equivalent to C, then each row of C is some linear combination of the rows of A and another linear combination of the rows of B. Since row-reduction is reversible, one can also say that the rows of B are linear combinations of the rows of C. By the Linear Combination Lemma one then gets that the rows of B are linear combinations of the rows of A. $\square$

Definition 1.28. The form of an $m \times n$ matrix is the sequence $\langle \ell_1, \dots, \ell_m \rangle$, where ℓ_i is the column number of the leading entry if row i, and $\ell_i = \infty$ if row i has no leading entry (i.e., it is a zero row).

Example. If

$$A = \begin{pmatrix} 1 & 2 & 0 & 4 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

then the form is $\langle 1, 3, \infty \rangle$.

Lemma 1.29. *If two echelon form matrices are row equivalent, then their forms are equal sequences.*

Remark 1.30. For a counterexample to an "if and only if" statement, consider

$$A = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{pmatrix}.$$

Both have the form $\langle 1, 2, \infty \rangle$, yet the matrices are clearly not row equivalent.

Proof: Let $B, D \in \mathbb{R}^{m \times n}$ be row equivalent. Let the form associated with B be given by $\langle \ell_1, \dots, \ell_m \rangle$, and let the form associated with D by given by $\langle k_1, \dots, k_m \rangle$. Let the rows of B be denoted by $\beta_1, \dots, \beta_m$, with

$$\beta_j = (\beta_{j,1}, \dots, \beta_{j,m}),$$

and the rows of D by $\delta_1, \dots, \delta_m$, with

$$\delta_j = (\delta_{j,1}, \dots, \delta_{j,m}),$$

We need to show that $\ell_i = k_i$ for each i. Let us first show that it holds for $i = 1$. The rest will follow by an induction argument (Problem 2.22). If β_1 is the zero row, then since B is in echelon form the matrix B is the zero matrix. By the above corollary this implies that D is also the zero matrix, so we are then done. Therefore, assume that β_1 and δ_1 are not zero rows. Since B and D are row equivalent, we then have that

$$\beta_1 = s_1\delta_1 + s_2\delta_2 + \cdots + s_m\delta_m,$$

or

$$\beta_{1,j} = \sum_{i=1}^{m} s_i \delta_{i,j};$$

in particular,

$$\beta_{1,\ell_1} = \sum_{i=1}^{m} s_i \delta_{i,\ell_1}.$$

By the definition of the form we have that $\beta_{i,j} = 0$ if $j < \ell_1$, and $\beta_{1,\ell_1} \neq 0$. Similarly, $\delta_{i,j} = 0$ if $j < k_1$, and $\delta_{1,k_1} \neq 0$. If $\ell_1 < k_1$ then the right-hand side of the above equation is zero. Since $\beta_{1,\ell_1} \neq 0$, this then clearly implies that $\ell_1 \geq k_1$. Writing δ_1 as a linear combination of $\beta_1, \dots, \beta_m$ and using the same argument as above shows that $\ell_1 \leq k_1$; hence, $\ell_1 = k_1$. $\square$

Corollary 1.31. *Any two echelon forms of a matrix have the same free variables, and consequently the same number of free variables.*

Lemma 1.32. *Each matrix is row equivalent to a unique reduced echelon form matrix.*

Example. Suppose that

$$A \longrightarrow \begin{pmatrix} 1 & 0 & -3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}, \quad B \longrightarrow \begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{pmatrix},$$

(a) Are A and B row equivalent? No.

(b) Is either matrix nonsingular? No.

2. Vector Spaces

2.1. Definition of Vector Space

2.1.1. Definition and Examples

Definition 2.1. A vector space over $\mathbb{R}$ consists of a set V along with the two operations '+' and '$\cdot$' such that

(a) if $\vec{u}, \vec{v}, \vec{w} \in V$, then $\vec{v} + \vec{w} \in V$ and

- $\vec{v} + \vec{w} = \vec{w} + \vec{v}$
- $(\vec{v} + \vec{w}) + \vec{u} = \vec{v} + (\vec{w} + \vec{u})$
- there is a $\vec{0} \in V$ such that $\vec{v} + \vec{0} = \vec{v}$ for all $\vec{v} \in V$ (zero vector)
- for each $\vec{v} \in V$ there is a $\vec{w} \in V$ such that $\vec{v} + \vec{w} = \vec{0}$ (additive inverse)

(b) if $r, s \in R$ and $\vec{v}, \vec{w} \in V$, then $r \cdot \vec{v} \in V$ and

- $(r + s) \cdot \vec{v} = r \cdot \vec{v} + s \cdot \vec{v}$
- $r \cdot (\vec{v} + \vec{w}) = r \cdot \vec{v} + r \cdot \vec{w}$
- $(rs) \cdot \vec{v} = r \cdot (s \cdot \vec{v})$
- $1 \cdot \vec{v} = \vec{v}$.

Example. $V = \mathbb{R}^n$ with $\vec{v} + \vec{w} = (v_i + w_i)$ and $r \cdot \vec{v} = (rv_i)$.

Example. $V = R^{m\times n}$ with $A + B = (a_{i,j} + b_{i,j})$ and $r \cdot A = (ra_{i,j})$.

Example. $V = \mathscr{P}_n = \{\sum_{i=0}^n a_i x^i \,:\, a_0, \dots, a_n \in \mathbb{R}\}$ with $(p + q)(x) = p(x) + q(x)$ and $(r \cdot p)(x) = rp(x)$.

Remark 2.2. The set $V = \mathscr{P}_\infty = \{p \in \mathscr{P}_n \,:\, n \in \mathbb{N}\}$ is an infinite-dimensional vector space, whereas $\mathscr{P}_n$ is finite-dimensional.

Example. $V = \{\sum_{i=0}^n a_i \cos(i\theta) \,:\, a_0, \dots, a_n \in \mathbb{R}\}$ with $(f + g)(x) = f(x) + g(x)$ and $(r \cdot f)(x) = rf(x)$.

Example. $V = \mathbb{R}^+$ with $x + y = xy$ and $r \cdot x = x^r$. We have that with $\vec{0} = 1$ this is a vector space.

Example. $V = \mathbb{R}^2$ with

$$\vec{v} + \vec{w} = \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \end{pmatrix}, \quad r \cdot \vec{v} = \begin{pmatrix} rv_1 \\ v_2 \end{pmatrix}.$$

The answer is no, as the multiplicative identities such as $(r + s) \cdot \vec{v} = r \cdot \vec{v} + s \cdot \vec{v}$ are violated.

Example. $V = \mathbb{R}^2$ with

$$\vec{v} + \vec{w} = \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \end{pmatrix}, \quad r \cdot \vec{v} = \begin{pmatrix} rv_1 \\ 0 \end{pmatrix}.$$

The answer is no, as there is no multiplicative identity.

Example. The set $\{f : \mathbb{R} \to \mathbb{R} \,:\, f'' + f = 0\}$ is a vector space, whereas the set $\{f : \mathbb{R} \to \mathbb{R} \,:\, f'' + f = 1\}$ is not.

2.1.2. Subspaces and Spanning Sets

Definition 2.3. A subspace is a subset of a vector space that is itself a vector space under the inherited operations.

Remark 2.4. Any vector space V has the trivial subspace $\{\vec{0}\}$ and the vector space itself as subspaces. These are the improper subspaces. Any other subspaces are proper.

Lemma 2.5. *A nonempty $S \subset V$ is a subspace if $\vec{x}, \vec{y} \in S$ implies that $r \cdot \vec{x} + s \cdot \vec{y} \in S$ for any $r, s \in \mathbb{R}$.*

Proof: By assumption the subset S is closed under vector addition and scalar multiplication. Since $S \subset V$ the operations in S inherit the same properties as those operations in V; hence, the set S is a subspace. $\square$

Example. Examples of subspaces are

(a) $S = \{\vec{x} \in \mathbb{R}^3 : x_1 + 2x_2 + 5x_3 = 0\}$

(b) $S = \{p \in \mathscr{P}_6 : p(3) = 0\}$

(c) $S = \{A \in \mathbb{R}^{n \times n} : a_{i,j} = 0 \text{ for } i > j\}$ (upper triangular matrices)

(d) $S = \{A \in \mathbb{R}^{n \times n} : a_{i,j} = 0 \text{ for } i < j\}$ (lower triangular matrices)

(e) If $A \in \mathbb{R}^{n \times n}$, the trace of A, denoted $\text{trace}(A)$, is given by $\text{trace}(A) = \sum_i a_{i,i}$. The set

$$S = \{A \in \mathbb{R}^{n \times n} : \text{trace}(A) = 0\}$$

Definition 2.6. If $S = \{\vec{x}_1, \dots, \vec{x}_n\}$, the span of S will be denoted by $[S]$.

Remark 2.7. From now on the multiplication $r \cdot \vec{x} \in V$ will be written $r\vec{x} \in V$, where the multiplication will be assumed to be the multiplication associated with V.

Lemma 2.8. *The span of any nonempty subset is a subspace.*

Proof: Let $\vec{u}, \vec{v} \in [S]$ be given. There then exist scalars such that

$$\vec{u} = \sum_i r_i \vec{x}_i, \quad \vec{v} = \sum_i s_i \vec{x}_i.$$

For given scalars $p, q \in \mathbb{R}$ one then sees that

$$p\vec{u} + q\vec{v} = p\left(\sum_i r_i \vec{x}_i\right) + q\left(\sum_i s_i \vec{x}_i\right) = \sum_i (pr_i + qs_i)\vec{x}_i,$$

so that $p\vec{u} + q\vec{v} \in [S]$. Hence, $[S]$ is a subspace. $\square$

Example. The set of solutions to a homogeneous linear system with coefficient matrix A will be denoted by $\mathcal{N}(A)$, the null space of A. It has been previously shown that there is a set of vectors $S = \{\vec{\beta}_1, \dots, \vec{\beta}_k\}$ such that $\mathcal{N}(A) = [S]$. Hence, $\mathcal{N}(A)$ is a subspace.

2.2. Linear Independence

2.2.1. Definition and Examples

Definition 2.9. The vectors $\vec{v}_1, \dots, \vec{v}_n \in V$ are linearly independent if and only if the only solution to

$$c_1\vec{v}_1 + \cdots + c_n\vec{v}_n = \vec{0}$$

is $c_1 = \cdots = c_n = 0$. Otherwise, the vectors are linearly dependent.

Example. (a) $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\} \subset \mathbb{R}^3$, where

$$\vec{v}_1 = \begin{pmatrix} 1 \\ 3 \\ -1 \end{pmatrix}, \quad \vec{v}_2 = \begin{pmatrix} -5 \\ -8 \\ 2 \end{pmatrix}, \quad \vec{v}_3 = \begin{pmatrix} 3 \\ -5 \\ 3 \end{pmatrix},$$

is a linearly dependent set, as $7\vec{v}_1 + 2\vec{v}_2 + \vec{v}_3 = \vec{0}$

(b) $\{1 + x, 1 - x, 1 - 3x + x^2\} \subset \mathscr{P}_2$ is a linearly independent set

Lemma 2.10. *If $S \subset V$ and $\vec{v} \in V$ is given, then*

$$[S] = [S \cup \{\vec{v}\}] \text{ if and only if } \vec{v} \in [S].$$

Proof: If $[S] = [S \cup \{\vec{v}\}]$, then since $\vec{v} \in [S \cup \{\vec{v}\}]$ one must have that $\vec{v} \in [S]$.

Now suppose that $\vec{v} \in [S]$, with $S = \{\vec{x}_1, \dots, \vec{x}_n\}$, so that $\vec{v} = \sum_i c_i\vec{x}_i$. If $\vec{w} \in [S \cup \{\vec{v}\}]$, then one can write $\vec{w} = d_0\vec{v} + \sum_i d_i\vec{x}_i$, which can be rewritten as

$$\vec{w} = \sum_i (d_0c_i + d_i)\vec{x}_i \in [S].$$

Hence, $[S \cup \{\vec{v}\}] \subseteq [S]$. It is clear that $[S] \subseteq [S \cup \{\vec{v}\}]$. □

Lemma 2.11. *If $S \subset V$ is a linearly independent set, then for any $\vec{v} \in V$ the set $S \cup \{\vec{v}\}$ is linearly independent if and only if $\vec{v} \notin [S]$.*

Proof: Let $S = \{\vec{x}_1, \dots, \vec{x}_n\}$. If $\vec{v} \in [S]$, then $\vec{v} = \sum_i c_i\vec{x}_i$, so that $-\vec{v} + \sum_i c_i\vec{x}_i = \vec{0}$. Hence, $S \cup \{\vec{v}\}$ is a linearly dependent set.

Now suppose that $S \cup \{\vec{v}\}$ is a linearly dependent set. There then exist constants $c_0, \dots, c_n$, some of which are nonzero, such that $c_0\vec{v} + \sum_i c_i\vec{x}_i = \vec{0}$. If $c_0 = 0$, then $\sum_i c_i\vec{x}_i = \vec{0}$, which contradicts the fact that the set S is linearly independent. Since $c_0 \neq 0$, upon setting $d_i = -c_i/c_0$ one can then write $\vec{v} = \sum_i d_i\vec{x}_i$, so that $\vec{v} \in [S]$. □

Corollary 2.12. *Let $S = \{\vec{x}_1, \dots, \vec{x}_n\} \subset V$, and define*

$$S_1 = \{\vec{x}_1\}, \quad S_j = S_{j-1} \cup \{\vec{x}_j\}\ j = 2, \dots, n.$$

S is linearly dependent set if and only if there is an $1 \leq \ell \leq n$ such that the set $S_{\ell-1}$ is a linearly independent set and S_ℓ is a linearly dependent set.

Remark 2.13. In other words, $\vec{x}_\ell = \sum_{i=1}^{\ell-1} c_i\vec{x}_i$.

Example. In a previous example we had $S = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ with $S_2 = \{\vec{v}_1, \vec{v}_2\}$ being a linearly independent set and $S_2 \cup \{\vec{v}_3\}$ being a linearly dependent set. The above results imply that $[S_2] = [S]$.

Theorem 2.14. *Let $S = \{\vec{x}_1, \dots, \vec{x}_n\} \subset V$. The set S has a linearly independent subset with the same span.*

Proof: Suppose that S is not a linearly independent set. This implies that for some $2 \le \ell \le n$ that $\vec{x}_\ell = \sum_{i=1}^{\ell-1} c_i\vec{x}_i$. Now define $S' = \{\vec{x}_1, \dots, \vec{x}_{\ell-1}, \vec{x}_{\ell+1}, \dots, \vec{x}_n\}$. Since $S = S' \cup \{\vec{x}_\ell\}$ and $\vec{x}_\ell \in [S']$, one has that $[S] = [S']$. When considering S', remove the next dependent vector (if it exists) from the set $\{\vec{x}_{\ell+1}, \dots, \vec{x}_n\}$, and call this new set S''. Using the same reasoning as above, $[S'] = [S'']$, so that $[S] = [S'']$. Continuing in this fashion and using an induction argument, we can then remove all of the linearly dependent vectors without changing the span. □

2.3. Basis and Dimension

2.3.1. Basis

Definition 2.15. The set $S = \{\vec{\beta}_1, \vec{\beta}_2, \dots\}$ is a basis for V if

(a) the vectors are linearly independent

(b) $V = [S]$

Remark 2.16. A basis will be denoted by $\langle\vec{\beta}_1, \vec{\beta}_2, \dots\rangle$.

Definition 2.17. Set $\vec{e}_j = (e_{j,i}) \in \mathbb{R}^n$ to be the vector which satisfies $e_{j,i} = \delta_{i,j}$. The standard basis for $\mathbb{R}^n$ is given by $\langle\vec{e}_1, \dots, \vec{e}_n\rangle$.

Remark 2.18. A basis is not unique. For example, one basis for $\mathbb{R}^2$ is the standard basis, whereas another is

$$\langle \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 5 \end{pmatrix} \rangle.$$

Example. (a) Two bases for $\mathscr{P}_3$ are $\langle 1, x, x^2, x^3\rangle$ and $\langle 1-x, 1+x, 1+x+x^2, x+x^3\rangle$.

(b) Consider the subspace $S \subset \mathbb{R}^{2\times 2}$ which is given by

$$S = \{A \in \mathbb{R}^{2\times 2} : a_{1,1} + 2a_{2,2} = 0,\ a_{1,2} - 3a_{2,1} = 0\}.$$

Since any $B \in S$ is of the form

$$B = c_1 \begin{pmatrix} -2 & 0 \\ 0 & 1 \end{pmatrix} + c_2 \begin{pmatrix} 0 & 3 \\ 1 & 0 \end{pmatrix},$$

and the two above matrices are linearly independent, a basis for S is given by

$$\langle \begin{pmatrix} -2 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 3 \\ 1 & 0 \end{pmatrix} \rangle.$$

Lemma 2.19. *The set $S = \{\vec{\beta}_1, \dots, \vec{\beta}_n\}$ is a basis if and only if each $\vec{v} \in V$ can be expressed as a linear combination of the vectors in S in a unique manner.*

Proof: If S is a basis, then by definition $\vec{v} \in [S]$. Suppose that $\vec{v}$ can be written in two different ways, i.e.,

$$\vec{v} = \sum_i c_i \vec{\beta}_i, \quad \vec{v} = \sum_i d_i \vec{\beta}_i.$$

One clearly then has that $\sum_i (c_i - d_i)\vec{\beta}_i = \vec{0}$, which, since the vectors in S are linearly independent, implies that $c_i = d_i$ for $i = 1, \dots, n$.

Suppose that $V = [S]$. Since $\vec{0} = \sum_i 0 \cdot \vec{\beta}_i$, and since vectors are expressed uniquely as linear combinations of the vectors of S, by definition the vectors in S are linearly independent. Hence, S is a basis. $\square$

Definition 2.20. Let $B = \langle \vec{\beta}_1, \dots, \vec{\beta}_n \rangle$ be a basis for V. For a given $\vec{v} \in V$ there are unique constants $c_1, \dots, c_n$ such that $\vec{v} = \sum_i c_i \vec{\beta}_i$. The representation of $\vec{v}$ with respect to B is given by

$$\mathrm{Rep}_B(\vec{v}) := \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}_B .$$

The constants $c_1, \dots, c_n$ are the coordinates of $\vec{v}$ with respect to B.

Example. For $\mathscr{P}_3$, consider the two bases $B = \langle 1, x, x^2, x^3 \rangle$ and $D = \langle 1-x, 1+x, 1+x+x^2, x+x^3 \rangle$. One then has that

$$\mathrm{Rep}_B(1 - 2x + x^3) = \begin{pmatrix} 1 \\ -2 \\ 0 \\ 1 \end{pmatrix}_B, \quad \mathrm{Rep}_D(1 - 2x + x^3) = \begin{pmatrix} 2 \\ -1 \\ 0 \\ 1 \end{pmatrix}_D .$$

2.3.2. Dimension

Definition 2.21. A vector space is finite-dimensional if it has a basis with only finitely many vectors.

Example. An example of an infinite-dimensional space is $\mathscr{P}_\infty$, which has as a basis $\langle 1, x, \dots, x^n, \dots \rangle$.

Definition 2.22. The transpose of a matrix $A \in \mathbb{R}^{m\times n}$, denoted by $A^T \in \mathbb{R}^{n\times m}$, is formed by interchanging the rows and columns of A.

Example.

$$A = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix} \longrightarrow A^T = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}.$$

Theorem 2.23. *In any finite-dimensional vector space all of the bases have the same number of elements.*

Proof: Let $B = \langle \vec{\beta}_1, \dots, \vec{\beta}_k \rangle$ be one basis, and let $D = \langle \vec{\delta}_1, \dots, \vec{\delta}_\ell \rangle$ be another basis. Suppose that $k > \ell$. For each $i = 1, \dots, k$ we can write $\vec{\beta}_i = \sum_j a_{i,j}\vec{\delta}_j$, which yields a matrix $A \in \mathbb{R}^{k\times\ell}$. Now let $\vec{v} \in V$ be given. Since B and D are bases, there are unique vectors $\mathrm{Rep}_B(\vec{v}) = \vec{v}^B = (v_i^B) \in \mathbb{R}^k$ and $\mathrm{Rep}_D(\vec{v}) = \vec{v}^D = (v_i^D) \in \mathbb{R}^\ell$ such that

$$\vec{v} = \sum_{i=1}^{k} v_i^B \vec{\beta}_i = \sum_{j=1}^{\ell} v_j^D \vec{\delta}_j.$$

The above can be rewritten as

$$\sum_{j=1}^{\ell} \left(\sum_{i=1}^{k} a_{i,j} v_i^B \right) \vec{\delta}_j = \sum_{j=1}^{\ell} v_j^D \vec{\delta}_j.$$

This then implies that the vector $\vec{v}^B$ is a solution to the linear system with the augmented matrix $(A^T|\vec{v}^D)$, where $A^T = (a_{j,i}) \in \mathbb{R}^{\ell\times k}$ is the transpose of A. Since $\ell < k$, when A^T is row-reduced it will have free variables, which implies that the linear system has an infinite number of solutions. This contradiction yields that $k \le \ell$.

If $k < \ell$, then by writing $\vec{\delta}_i = \sum_j c_{i,j}\vec{\beta}_j$ and using the above argument one gets that $k \ge \ell$. Hence, $k = \ell$. □

Definition 2.24. The dimension of a vector space V, $\dim(V)$, is the number of basis vectors.

Example. (a) Since the standard basis for $\mathbb{R}^n$ has n vectors, $\dim(\mathbb{R}^n) = n$.

(b) Since a basis for $\mathscr{P}_n$ is $\langle 1, x, \dots, x^n \rangle$, one has that $\dim(\mathscr{P}_n) = n + 1$.

(c) Recall that the subspace

$$S = \{A \in \mathbb{R}^{2\times 2} \,:\, a_{1,1} + 2a_{2,2} = 0,\ a_{1,2} - 3a_{2,1} = 0\}.$$

has a basis

$$\langle \begin{pmatrix} -2 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 3 \\ 1 & 0 \end{pmatrix} \rangle.$$

This implies that $\dim(S) = 2$.

Corollary 2.25. *No linearly independent set can have more vectors than* $\dim(V)$*.*

Proof: Suppose that $S = \{\vec{x}_1, \dots, \vec{x}_n\}$ is a linearly independent set with $n > \dim(V)$. Since $S \subset V$ one has that $[S] \subset V$. Furthermore, since the set is linearly independent, $\dim([S]) = n$. This implies that $\dim(V) \geq n$, which is a contradiction. □

Corollary 2.26. *Any linearly independent set* S *can be expanded to make a basis.*

Proof: Suppose that $\dim(V) = n$, and that $\dim([S]) = k < n$. There then exist $n-k$ linearly independent vectors $\vec{v}_1, \dots, \vec{v}_{n-k}$ such that $\vec{v}_i \notin [S]$. The set $S' = S \cup \{\vec{v}_1, \dots, \vec{v}_{n-k}\}$ is a linearly independent set with $\dim([S']) = n$. As a consequence, S' is a basis for V. □

Corollary 2.27. *If* $\dim(V) = n$*, then a set of* n *vectors* $S = \{\vec{x}_1, \dots, \vec{x}_n\}$ *is linearly independent if and only if* $V = [S]$*.*

Proof: Suppose that $V = [S]$. If the vectors are not linearly independent, then (upon a possible reordering) there is an $\ell < n$ such that for $S' = \{\vec{x}_1, \dots, \vec{x}_\ell\}$ one has $[S] = [S']$ with the vectors in S' being linearly independent. This implies that $V = [S']$, and that $\dim(V) = \ell < n$.

Now suppose that the vectors are linearly independent. If $[S] \subset V$, then there is at least one vector $\vec{v}$ such that $S \cup \{\vec{v}\}$ is linearly independent with $[S \cup \{\vec{v}\}] = V$. This implies that $\dim(V) \geq n+1$. □

Remark 2.28. Put another way, the above corollary states that if $\dim(V) = n$ and the set $S = \{\vec{x}_1, \dots, \vec{x}_n\}$ is linearly independent, then S is a spanning set for V.

2.3.3. Vector Spaces and Linear Systems

Definition 2.29. The null space of a matrix A, denoted by $\mathcal{N}(A)$, is the set of all solutions to the homogeneous system for which A is the coefficient matrix.

Example. Consider

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

A basis for $\mathcal{N}(A)$ is given by

$$\langle \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \rangle.$$

Definition 2.30. The row space of a matrix A, denoted by $\text{Rowspace}(A)$, is the span of the set of the rows of A. The row rank is the dimension of the row space.

Example. If

$$A = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix},$$

then $\text{Rowspace}(A) = [\{\begin{pmatrix} 1 & 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 2 \end{pmatrix}\}]$, and the row rank is 2.

Lemma 2.31. *The nonzero rows of an echelon form matrix make up a linearly independent set.*

Proof: We have already seen that in an echelon form matrix no nonzero row is a linear combination of the other rows. □

Corollary 2.32. *Suppose that a matrix A has been put in echelon form. The nonzero rows of the echelon form matrix are a basis for* $\text{Rowspace}(A)$.

Proof: If $A \longrightarrow B$, where B is in echelon form, then it is known that each row of A is a linear combination of the rows of B. The converse is also true; hence, $\text{Rowspace}(A) = \text{Rowspace}(B)$. Since the rows of B are linearly independent, they form a basis for $\text{Rowspace}(B)$, and hence $\text{Rowspace}(A)$. □

Definition 2.33. The column space of a matrix A, denoted by $\mathcal{R}(A)$, is the span of the set of the columns of A. The column rank is the dimension of the column space.

Remark 2.34. A basis for $\mathcal{R}(A)$ is found by determining $\text{Rowspace}(A^T)$, and the column rank of A is the dimension of $\text{Rowspace}(A^T)$.

Example. Consider

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 3 & 10 \\ 1 & 5 & 18 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 4 \\ 0 & 0 & 0 \end{pmatrix}, \quad A^T = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 3 & 5 \\ 2 & 10 & 18 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

A basis for $\text{Rowspace}(A)$ is $\langle \begin{pmatrix} 1 & 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 4 \end{pmatrix} \rangle$, and a basis for $\mathcal{R}(A)$ is

$$\langle \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \rangle.$$

Theorem 2.35. *The row rank and column rank of a matrix are equal.*

Proof: First, let us note that row operations do not change the column rank of a matrix. If $A = (\vec{a}_1 \ \dots \ \vec{a}_n) \in \mathbb{R}^{m \times n}$, where each column $\vec{a}_i \in \mathbb{R}^m$, then finding the set of homogeneous solutions for the linear system with coefficient matrix A is equivalent to solving

$$c_1\vec{a}_1 + \cdots + c_n\vec{a}_n = \vec{0}.$$

Row operations leave unchanged the the set of solutions $(c_1, \dots, c_n)$; hence, the linear independence of the vectors is unchanged, and the dependence of one vector on the others remains unchanged.

Now bring the matrix to reduced echelon form, so that each column with a leading entry is one of the $\vec{e}_i$'s from the standard basis. The row rank is equal to the number of rows with leading entries. The column rank of the reduced matrix is equal to that of the original matrix. It is clear that the column rank of the reduced matrix is also equal to the number of leading entries. □

Definition 2.36. The rank of a matrix A, denoted by $\text{rank}(A)$, is its row rank.

Remark 2.37. Note that the above statements imply that $\text{rank}(A) = \text{rank}(A^T)$.

Theorem 2.38. *If* $A \in \mathbb{R}^{m\times n}$, *then* $\text{rank}(A) + \dim(\mathcal{N}(A)) = n$.

Proof: Put the matrix A in reduced echelon form. One has that $\text{rank}(A)$ is the number of leading entries, and that $\dim(\mathcal{N}(A))$ is the number of free variables. It is clear that these numbers sum to n. □

2.3.4. Combining Subspaces

Definition 2.39. If $W_1, \dots, W_k$ are subspaces of V, then their sum is given by

$$W_1 + \cdots + W_k = [W_1 \cup \cdots \cup W_k].$$

Let a basis for W_j be given by $\langle \vec{w}_{1,j}, \dots, \vec{w}_{\ell(j),j}\rangle$. If $\vec{v} \in W_1 + \cdots + W_k$, then this implies there are constants $c_{i,j}$ such that

$$\vec{v} = \sum_{i=1}^{\ell(1)} c_{i,1}\vec{w}_{i,1} + \cdots + \sum_{i=1}^{\ell(k)} c_{i,k}\vec{w}_{i,k}.$$

Note that $\sum_i c_{i,j}\vec{w}_{i,j} \in W_j$. For example, when considering $\mathscr{P}_3$ suppose that a basis for W_1 is $\langle 1, 1+x^2\rangle$, and that a basis for W_2 is $\langle x, 1+x, x^3\rangle$. If $p \in W_1 + W_2$, then

$$p(x) = c_1 + c_2(1+x^2) + d_1 x + d_2(1+x) + d_3 x^3.$$

Q: How does the dimension of each W_i relate to the dimension of $W_1 + \cdots + W_k$? In the above example, $\dim(W_1) = 2, \dim(W_2) = 3$, but $\dim(\mathscr{P}_3) = \dim(W_1 + W_2) = 4$. Thus, for this example $\dim(W_1 + W_2) \neq \dim(W_1) + \dim(W_2)$.

Definition 2.40. A collection of subspaces $\{W_1, \dots, W_k\}$ is independent if for each $i = 1, \dots, k$,

$$W_i \cap (\cup_{j\neq i} W_j) = \{\vec{0}\}.$$

Example. Suppose that

$$W_1 = [\begin{pmatrix}1\\0\\0\end{pmatrix}], \quad W_2 = [\begin{pmatrix}0\\1\\0\end{pmatrix}], \quad W_3 = [\begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix}].$$

It is clear that $W_i \cap W_j = \{\vec{0}\}$ for $i \neq j$. However, the subspaces are not independent, as

$$W_3 \cap (W_1 \cup W_2) = [\begin{pmatrix}1\\1\\0\end{pmatrix}].$$

Definition 2.41. A vector space V is the direct sum of the subspaces $W_1, \dots, W_k$ if

(a) the subspaces are independent

(b) $V = W_1 + \cdots + W_k$.

In this case we write $V = W_1 \oplus \cdots \oplus W_k$.

Example. (a) $\mathbb{R}^n = [\vec{e}_1] \oplus [\vec{e}_2] \oplus \cdots \oplus [\vec{e}_n]$

(b) $\mathscr{P}_n = [1] \oplus [x] \oplus \cdots \oplus [x^n]$

Lemma 2.42. *If $V = W_1 \oplus \cdots \oplus W_k$, then*

$$\dim(V) = \sum_i \dim(W_i).$$

Example. In a previous example we had that $\mathbb{R}^3 = W_1 + W_2 + W_3$, with

$$\dim(W_1) = \dim(W_2) = 1, \dim(W_3) = 2.$$

Thus, the sum cannot be direct.

Proof: First show that the result is true for $k = 2$, and then use induction to prove the general result. Let a basis for W_1 be given by $\langle \vec{\beta}_1, \dots, \vec{\beta}_k \rangle$, and a basis for W_2 be given by $\langle \vec{\delta}_1, \dots, \vec{\delta}_\ell \rangle$. This yields that $\dim(W_1) = k$ and $\dim(W_2) = \ell$. Since $W_1 \cap W_2 = \{\vec{0}\}$, the set $\{\vec{\beta}_1, \dots, \vec{\beta}_k, \vec{\delta}_1, \dots, \vec{\delta}_\ell\}$ is linearly independent, and forms a basis for $[W_1 \cup W_2]$. Since $V = [W_1 \cup W_2]$, this then yields that a basis for V is $\langle \vec{\beta}_1, \dots, \vec{\beta}_k, \vec{\delta}_1, \dots, \vec{\delta}_\ell \rangle$; thus,

$$\dim(V) = k + \ell = \dim(W_1) + \dim(W_2).$$

□

Definition 2.43. If $V = W_1 \oplus W_2$, then the subspaces W_1 and W_2 are said to be complements.

Definition 2.44. For vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$, define the dot product (or inner product) to be

$$\vec{u} \cdot \vec{v} = \sum_{i=1}^{n} u_i v_i.$$

The dot product has the properties that

(a) $\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$

(b) $(a\vec{u} + b\vec{v}) \cdot \vec{w} = a\vec{u} \cdot \vec{w} + b\vec{v} \cdot \vec{w}$

(c) $\vec{u} \cdot \vec{u} \geq 0$, with $\vec{u} \cdot \vec{u} = 0$ if and only if $\vec{u} = \vec{0}$.

Definition 2.45. If $U \subset \mathbb{R}^n$ is a subspace, define the orthocomplement of U to be

$$U^\perp = \{\vec{v} \in \mathbb{R}^n : \vec{v} \cdot \vec{u} = 0 \text{ for all } \vec{u} \in U\}.$$

Proposition 2.46. *$U^\perp$ is a subspace.*

Proof: Let $\vec{v}, \vec{w} \in U^\perp$. Since

$$(a\vec{v} + b\vec{w}) \cdot \vec{u} = a\vec{v} \cdot \vec{u} + b\vec{w} \cdot \vec{u} = 0$$

for any $\vec{u} \in U$, this implies that $a\vec{v} + b\vec{w} \in U^\perp$. Hence $U^\perp$ is a subspace. □

Example. If $U = [\vec{e}_1] \subset \mathbb{R}^2$, then $U^\perp = [\vec{e}_2]$, and if $U = [\vec{e}_1, \vec{e}_2] \subset \mathbb{R}^3$, then $U^\perp = [\vec{e}_3]$.

Remark 2.47. If $A = (\vec{a}_1 \ldots \vec{a}_n) \in \mathbb{R}^{m\times n}$, recall that

$$\mathcal{R}(A) = [\vec{a}_1 \ldots \vec{a}_n].$$

Thus, $\vec{b} \in \mathcal{R}(A)$ if and only if $\vec{b} = \sum_i c_i\vec{a}_i$. Furthermore,

$$\mathcal{N}(A) = \{\vec{x} = (x_i) : \sum_i x_i\vec{a}_i = \vec{0}\}.$$

Theorem 2.48. *If $A \in \mathbb{R}^{m\times n}$, then $\mathcal{N}(A^T) = \mathcal{R}(A)^\perp$.*

Proof: Suppose that $\vec{x} \in \mathcal{N}(A^T)$, so that $\vec{x} \cdot \vec{a}_i = 0$ for $i = 1, \ldots, n$. As a consequence, $\vec{x} \cdot (\sum_i c_i\vec{a}_i) = 0$, so that $\vec{x} \in \mathcal{R}(A)^\perp$. Hence, $\mathcal{N}(A^T) \subset \mathcal{R}(A)^\perp$. Similarly, if $\vec{y} \in \mathcal{R}(A)^\perp$, one gets that $\vec{y} \in \mathcal{N}(A^T)$, so that $\mathcal{R}(A)^\perp \subset \mathcal{N}(A^T)$. □

Remark 2.49. Alternatively, one has that $\mathcal{N}(A) = \mathcal{R}(A^T)^\perp$.

Theorem 2.50. *If $U \subset \mathbb{R}^n$ is a subspace, then $\mathbb{R}^n = U \oplus U^\perp$.*

Proof: Let a basis for U be given by $\langle\vec{\beta}_1, \ldots, \vec{\beta}_k\rangle$, and set $A = (\vec{\beta}_1 \ldots \vec{\beta}_k) \in \mathbb{R}^{n\times k}$. By construction, $\text{rank}(A) = k$ (and $\text{rank}(A^T) = k$). By the previous theorem, we have that $U^\perp = \mathcal{N}(A^T)$. Since $\dim(\mathcal{N}(A^T)) + \text{rank}(A^T) = n$, we get that $\dim(U^\perp) = n - k$.

We must now show that $U \cap U^\perp = \{\vec{0}\}$. Let $\vec{\delta} \in U \cap U^\perp$, which implies that $\vec{\delta} = \sum_i c_i\vec{\beta}_i$. By using the linearity of the inner product

$$\vec{\delta} \cdot \vec{\delta} = \sum_{i=1}^{k} c_i(\vec{\beta}_i \cdot \vec{\delta}) = 0,$$

so that $\vec{\delta} = \vec{0}$. □

Remark 2.51. A consequence of the above theorem is that $(U^\perp)^\perp = U$.

Example. Suppose that a basis for U is $\langle\vec{\beta}_1, \ldots, \vec{\beta}_k\rangle$. The above theorem shows us how to compute a basis for $U^\perp$. Simply construct the matrix $A = (\vec{\beta}_1 \ldots \vec{\beta}_k)$, and then find a basis for $\mathcal{N}(A^T) = U^\perp$. For example, suppose that $U = [\vec{a}_1, \vec{a}_2]$, where

$$\vec{a}_1 = \begin{pmatrix} 1 \\ 0 \\ 2 \\ 4 \end{pmatrix}, \quad \vec{a}_2 = \begin{pmatrix} 3 \\ -1 \\ 4 \\ 7 \end{pmatrix}.$$

Since

$$A^T \longrightarrow \begin{pmatrix} 1 & 0 & 2 & 4 \\ 0 & 1 & 2 & 5 \end{pmatrix},$$

one has that $U^\perp = \mathcal{N}(A^T) = [\vec{\delta}_1, \vec{\delta}_2]$, where

$$\vec{\delta}_1 = \begin{pmatrix} -2 \\ -2 \\ 1 \\ 0 \end{pmatrix}, \quad \vec{\delta}_2 = \begin{pmatrix} -4 \\ -5 \\ 0 \\ 1 \end{pmatrix}.$$

Corollary 2.52. *Consider a linear system whose associated augmented matrix is $(A|\vec{b})$. The system is consistent if and only if $\vec{b} \in \mathcal{N}(A^T)^\perp$.*

Proof: If $A = (\vec{a}_1 \ldots \vec{a}_n)$, then the system is consistent if and only if $\vec{b} \in \mathcal{R}(A)$, i.e., $\vec{b} = \sum_i c_i \vec{a}_i$. By the above theorem $\mathcal{R}(A) = \mathcal{N}(A^T)^\perp$. □

Example. As a consequence, the system is consistent if and only if $\vec{b} \cdot \vec{\delta}_i = 0$, where $\langle \vec{\delta}_1, \ldots, \vec{\delta}_k \rangle$ is a basis for $\mathcal{N}(A^T)$. For example, suppose that A is as in the previous example. Then for $\vec{b} = (b_i) \in \mathbb{R}^4$, the associated linear system will be consistent if and only if $\vec{\delta}_1 \cdot \vec{b} = 0, \vec{\delta}_2 \cdot \vec{b} = 0$. In other words, the components of the vector $\vec{b}$ must satisfy the linear system

$$\begin{aligned} -2b_1 - 2b_2 + b_3 \quad\quad &= 0 \\ -4b_1 - 5b_2 \quad\quad + b_4 &= 0 \end{aligned}$$

which implies that $\vec{b} \in S = [\vec{b}_1, \vec{b}_2]$, where

$$\vec{b}_1 = \begin{pmatrix} 1 \\ 0 \\ 2 \\ 4 \end{pmatrix}, \quad \vec{b}_2 = \begin{pmatrix} 0 \\ 1 \\ 2 \\ 5 \end{pmatrix}.$$

Note that $S = \mathcal{R}(A)$, so that a basis for $\mathcal{R}(A)$ is $\langle \vec{b}_1, \vec{b}_2 \rangle$. Further note that this is consistent with the reduced echelon form of A^T.

3. Maps Between Spaces

3.1. Isomorphisms

3.1.1. Definition and Examples

Definition 3.1. Let V and W be vector spaces. A map $f : V \to W$ is one-to-one if $\vec{v}_1 \neq \vec{v}_2$ implies that $f(\vec{v}_1) \neq f(\vec{v}_2)$. The map is onto if for each $\vec{w} \in W$ there is a $\vec{v} \in V$ such that $f(\vec{v}) = \vec{w}$.

Example. (a) The map $f : \mathscr{P}_n \to \mathbb{R}^{n+1}$ given by

$$a_0 + a_1 x + \cdots + a_n x^n \longrightarrow \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix}$$

is one-to-one and onto.

(b) The map $f : \mathbb{R}^{2\times 2} \to \mathbb{R}^4$ given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \longrightarrow \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$$

is one-to-one and onto.

Definition 3.2. The map $f : V \to W$ is an isomorphism if

(a) f is one-to-one and onto

(b) f is linear, i.e.,

- $f(\vec{v}_1 + \vec{v}_2) = f(\vec{v}_1) + f(\vec{v}_2)$
- $f(r\vec{v}) = rf(\vec{v})$ for any $r \in \mathbb{R}$

We write $V \cong W$, and say that V is isomorphic to W.

Remark 3.3. If $V \cong W$, then we can think that V and W are the "same".

Example. (a) In the above examples it is easy to see that the maps are linear. Hence, $\mathscr{P}_(n) \cong \mathbb{R}^{n+1}$ and $\mathbb{R}^{2\times 2} \cong \mathbb{R}^4$.

(b) In general, $\mathbb{R}^{m\times n} \cong \mathbb{R}^{mn}$.

Definition 3.4. If $f : V \to V$ is an isomorphism, then we say that f is an automorphism.

Example. (a) The dilation map $d_s : \mathbb{R}^2 \to \mathbb{R}^2$ given by $d_s(\vec{v}) = s\vec{v}$ for some nonzero $s \in \mathbb{R}$ is an automorphism.

(b) The rotation map $t_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$t_\theta(\vec{v}) = \begin{pmatrix} \cos\theta\, v_1 - \sin\theta\, v_2 \\ \sin\theta\, v_1 + \cos\theta\, v_2 \end{pmatrix}$$

is an automorphism.

Lemma 3.5. *If $f : V \to W$ is linear, then $f(\vec{0}) = \vec{0}$.*

Proof: Since f is linear, $f(\vec{0}) = f(0 \cdot \vec{v}) = 0 \cdot f(\vec{v}) = \vec{0}$. □

Lemma 3.6. *The statement that $f : V \to W$ is linear is equivalent to*

$$f(c_1\vec{v}_1 + \cdots + c_n\vec{v}_n) = c_1 f(\vec{v}_1) + \cdots + c_n f(\vec{v}_n).$$

Proof: Proof by induction. By definition the statement holds for $n = 1$, so now suppose that it holds for $n = N$. This yields

$$\begin{aligned} f(\sum_{i=1}^{N} c_i\vec{v}_i + c_{N+1}\vec{v}_{N+1}) &= f(\sum_{i=1}^{N} c_i\vec{v}_i) + f(c_{N+1}\vec{v}_{N+1}) \\ &= \sum_{i=1}^{N} c_i f(\vec{v}_i) + f(c_{N+1}\vec{v}_{N+1}). \end{aligned}$$

□

Definition 3.7. Let U, V be vector spaces. The external direct sum, $W = U \times V$, is defined by

$$W = \{(\vec{u}, \vec{v}) : \vec{u} \in U,\ \vec{v} \in V\},$$

along with the operations

$$\vec{w}_1 + \vec{w}_2 = (\vec{u}_1 + \vec{u}_2, \vec{v}_1 + \vec{v}_2), \quad r \cdot \vec{w} = (r\vec{u}, r\vec{v}).$$

Lemma 3.8. *The external direct sum $W = U \times V$ is a vector space. Furthermore,* $\dim(W) = \dim(U) + \dim(V)$.

Proof: It is easy to check that W is a vector space. Let $S_u = \{\vec{u}_1, \ldots, \vec{u}_k\}$ be a basis for U, and let $S_V = \{\vec{v}_{k+1}, \ldots, \vec{v}_\ell\}$ be a basis for V. Given a $w = (\vec{u}, \vec{v}) \in W$, it is clear that one can write

$$\vec{w} = (\sum_i c_i\vec{u}_i, \sum_j d_j\vec{v}_j);$$

hence, a potential basis for W is

$$\vec{w}_i = \begin{cases} (\vec{u}_i, \vec{0}), & i = 1, \ldots, k \\ (\vec{0}, \vec{v}_i), & i = k+1, \ldots, \ell. \end{cases}$$

We need to check that the vectors $\vec{w}_1, \dots, \vec{w}_\ell$ are linearly independent. Writing $\sum_i c_i \vec{w}_i = \vec{0}$ is equivalent to the equations

$$\sum_{i=1}^{k} c_i \vec{u}_i = \vec{0}, \qquad \sum_{i=k+1}^{\ell} c_i \vec{v}_i = \vec{0}.$$

Since S_U and S_V are bases, the only solution is $c_i = 0$ for all i. □

Example. A basis for $\mathscr{P}_2 \times \mathbb{R}^2$ is given by

$$\langle (1, \vec{0}), (x, \vec{0}), (x^2, \vec{0}), (0, \vec{e}_1), (0, \vec{e}_2) \rangle,$$

and $\mathscr{P}_2 \times \mathbb{R}^2 \cong \mathbb{R}^5$ via the isomorphism

$$(a_0 + a_1 x + a_2 x^2, c_1 \vec{e}_1 + c_2 \vec{e}_2) \longrightarrow \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ c_1 \\ c_2 \end{pmatrix}.$$

3.1.2. Dimension Characterizes Isomorphism

Note that in all of the examples up to this point, if $U \cong V$, then it was true that $\dim(U) = \dim(V)$. The question: does the dimension of two vector spaces say anything about whether or not they are isomorphic?

Lemma 3.9. *If* $V \cong W$, *then* $\dim(V) = \dim(W)$.

Proof: Let $f : V \to W$ be an isomorphism. Let $\langle \vec{\beta}_1, \dots, \vec{\beta}_n \rangle$ be a basis for V, and consider the set $S_W = \{f(\vec{\beta}_1), \dots, f(\vec{\beta}_n)\}$. First, the set is linearly independent, as

$$\vec{0} = \sum_{i=1}^{n} c_i f(\vec{\beta}_i) = f(\sum_{i=1}^{n} c_i \vec{\beta}_i)$$

implies that $\sum_i c_i \vec{\beta}_i = \vec{0}$ (f is one-to-one), which further implies that $c_i = 0$ for all i. Since f is onto, for each $\vec{w} \in W$ there is a $\vec{v} \in V$ such that $f(\vec{v}) = \vec{w}$. Upon writing $\vec{v} = \sum_i d_i \vec{\beta}_i$ and using the linearity of the function f we get that

$$\vec{w} = \sum_{i=1}^{n} d_i f(\vec{\beta}_i).$$

Hence, S_W is a basis, and we then have the result. □

Lemma 3.10. *If* $\dim(V) = \dim(W)$, *then the two spaces are isomorphic.*

Proof: It will be enough to show that if $\dim(V) = n$, then $V \cong \mathbb{R}^n$. A similar result will yield $W \cong \mathbb{R}^n$, which would then yield $V \cong \mathbb{R}^n \cong W$, i.e., $V \cong W$.

Let $B = \langle \vec{\beta}_1, \dots, \vec{\beta}_n \rangle$ be a basis for V, and consider the map $\text{Rep}_B : V \to \mathbb{R}^n$ given by

$$\text{Rep}_B(\vec{v}) = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}, \quad \vec{v} = \sum_{i=1}^{n} c_i \vec{\beta}_i.$$

The map is clearly linear. The map is one-to-one, for if $\text{Rep}_B(\vec{u}) = \text{Rep}_B(\vec{v})$ with $\vec{u} = \sum_i c_i\vec{\beta}_i$, $\vec{v} = \sum_i d_i\vec{\beta}_i$, then $c_i = d_i$ for all i, which implies that $\vec{u} = \vec{v}$. Finally, the map is clearly onto. Hence, Rep_B is an isomorphism, so that $V \cong \mathbb{R}^n$. □

Theorem 3.11. *$V \cong W$ if and only if $\dim(V) = \dim(W)$.*

Corollary 3.12. *If $\dim(V) = k$, then $V \cong \mathbb{R}^k$.*

Example (cont.)**.** (a) Since $\dim(\mathbb{R}^{m\times n}) = mn$, $\mathbb{R}^{m\times n} \cong \mathbb{R}^{mn}$

(b) Since $\dim(\mathscr{P}_n) = n + 1$, $\mathscr{P}_n \cong \mathbb{R}^{n+1}$

3.2. Homomorphisms

3.2.1. Definition

Definition 3.13. If $h : V \to W$ is linear, then it is a homomorphism.

Example. (a) The projection map $\pi : \mathbb{R}^3 \to \mathbb{R}^2$ given by

$$\pi(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}) = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

is a homomorphism. However, it is not an isomorphism, as the map is not one-to-one, i.e., $\pi(r\vec{e}_3) = \vec{0}$ for any $r \in \mathbb{R}$.

(b) The derivative map $\mathrm{d}/\mathrm{d}x : \mathscr{P}_n \to \mathscr{P}_n$ given by

$$\frac{\mathrm{d}}{\mathrm{d}x}(a_0 + a_1x + \cdots + a_nx^n) = a_1 + 2a_2x + \cdots + na_nx^{n-1}$$

is a homomorphism. However, it is not an isomorphism, as the map is not one-to-one, i.e., $\mathrm{d}/\mathrm{d}x(a_0) = 0$ for any $a_0 \in \mathbb{R}$.

Definition 3.14. If $h : V \to V$, then it is called a linear transformation.

Theorem 3.15. *Let $\langle \vec{v}_1, \ldots, \vec{v}_n\rangle$ be a basis for V, and let $\{\vec{w}_1, \ldots, \vec{w}_n\} \subset W$ be given. There exists a unique homomorphism $h : V \to W$ such that $h(\vec{v}_j) = \vec{w}_j$ for $j = 1, \ldots, n$.*

Proof: Set $h : V \to W$ to be the map given by

$$h(c_1\vec{v}_1 + \cdots + c_n\vec{v}_n) = c_1\vec{w}_1 + \cdots + c_n\vec{w}_n.$$

The map is linear, for if $\vec{u}_1 = \sum_i c_i\vec{v}_i$, $\vec{u}_2 = \sum_i d_i\vec{v}_i$, then

$$\begin{aligned} h(r_1\vec{u}_1 + r_2\vec{u}_2) &= h\left(\sum_i (r_1c_i + r_2d_i)\vec{v}_i\right) \\ &= \sum_i (r_1c_i + r_2d_i)\vec{w}_i \\ &= r_1h(\vec{u}_1) + r_2h(\vec{u}_2). \end{aligned}$$

The map is unique, for if $g : V \to W$ is a homomorphism such that $g(\vec{v}_i) = \vec{w}_i$, then

$$g(\vec{v}) = g(\sum_i c_i\vec{v}_i) = \sum_i c_i g(\vec{v}_i) = \sum_i c_i\vec{w}_i = h(\vec{v});$$

hence, $g(\vec{v}) = h(\vec{v})$ for all $\vec{v} \in V$, so that they are the same map. □

Example. (a) The rotation map $t_\theta ; \mathbb{R}^2 \to \mathbb{R}^2$ is an automorphism which satisfies

$$t_\theta(\vec{e}_1) = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \quad t_\theta(\vec{e}_2) = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}.$$

One then has

$$t_\theta(\vec{v}) = v_1 \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} + v_2 \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}.$$

(b) Suppose that a homomorphism $h : \mathscr{P}_2 \to \mathscr{P}_3$ satisfies

$$h(1) = x, \quad h(x) = \frac{1}{2}x^2, \quad h(x^2) = \frac{1}{3}x^3.$$

Then

$$h(a_0 + a_1x + a_2x^2) = a_0x + \frac{1}{2}a_1x^2 + \frac{1}{3}a_2x^3.$$

The map is not an isomorphism, as it is not onto.

3.2.2. Rangespace and Nullspace

Definition 3.16. Let $h : V \to W$ be a homomorphism. The range space is given by

$$\mathcal{R}(h) := \{h(\vec{v}) : \vec{v} \in V\}.$$

The rank of h, $\text{rank}(h)$, satisfies $\text{rank}(h) = \dim(\mathcal{R}(h))$.

Lemma 3.17. *$\mathcal{R}(h)$ is a subspace.*

Proof: Let $\vec{w}_1, \vec{w}_2 \in \mathcal{R}(h)$ be given. There then exists $\vec{v}_1, \vec{v}_2$ such that $h(\vec{v}_i) = \vec{w}_i$. Since $h(c_1\vec{v}_1 + c_2\vec{v}_2) \in \mathcal{R}(h)$ and $h(c_1\vec{v}_1 + c_2\vec{v}_2) = c_1\vec{w}_1 + c_2\vec{w}_2$, one has that $c_1\vec{w}_1 + c_2\vec{w}_2 \in \mathcal{R}(h)$. Hence, it is a subspace. □

Remark 3.18. (a) $\text{rank}(h) \le \dim(W)$

(b) h is onto if and only if $\text{rank}(h) = \dim(W)$

Example. If $h : \mathbb{R}^{2\times 2} \to \mathscr{P}_3$ is given by

$$h(A) = (a+b)x + cx^2 + dx^3, \quad A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

then a basis for $\mathcal{R}(h)$ is $\langle x, x^2, x^3\rangle$, so that that $\operatorname{rank}(h) = 3$.

Definition 3.19. The inverse map $h^{-1} : W \to V$ is given by

$$h^{-1}(\vec{w}) := \{\vec{v} : h(\vec{v}) = \vec{w}\}.$$

Lemma 3.20. *Let $h : V \to W$ be a homomorphism, and let $S \subset \mathcal{R}(h)$ be a subspace. Then*

$$h^{-1}(S) := \{\vec{v} \in V : h(\vec{v}) \in S\}$$

is a subspace. In particular, $h^{-1}(\vec{0})$ is a subspace.

Definition 3.21. The null space (kernel) of the homomorphism $h : V \to W$ is given by

$$\mathcal{N}(h) := \{\vec{v} \in V : h(\vec{v}) = \vec{0}\} = h^{-1}(\vec{0}).$$

The *nullity* of $\mathcal{N}(h)$ is $\dim(\mathcal{N}(h))$.

Example. Again consider the map $h : \mathbb{R}^{2\times 2} \to \mathscr{P}_3$. It is clear that $h(A) = 0$ if and only if $a + b = 0, c = d = 0$, so that a basis for $\mathcal{N}(h)$ is given by

$$\langle \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \rangle.$$

Note that for this example, $\operatorname{rank}(h) + \dim(\mathcal{N}(h)) = 4$.

Theorem 3.22. *Let $h : V \to W$ be a homomorphism. Then*

$$\operatorname{rank}(h) + \dim(\mathcal{N}(h)) = \dim(V).$$

Remark 3.23. Compare this result to that for matrices, where if $A \in \mathbb{R}^{m\times n}$, then

$$\operatorname{rank}(A) + \dim(\mathcal{N}(A)) = n.$$

Proof: Let $B_{\mathcal{N}} = \langle \vec{\beta}_1, \dots, \vec{\beta}_k\rangle$ be a basis for $\mathcal{N}(h)$, and extend that to a basis $B_V = \langle \vec{\beta}_1, \dots, \vec{\beta}_k, \vec{v}_1, \dots, \vec{v}_\ell\rangle$ for V, where $k + \ell = n$. Set $B_{\mathcal{R}} = \langle h(\vec{v}_1), \dots, h(\vec{v}_\ell)\rangle$. We need to show that $B_{\mathcal{R}}$ is a basis for $\mathcal{R}(h)$.

First consider $\vec{0} = \sum_i c_i h(\vec{v}_i) = h(\sum_i c_i \vec{v}_i)$. Thus, $\sum_i c_i \vec{v}_i \in \mathcal{N}(h)$, so that $\sum_i c_i \vec{v}_i = \sum_i d_i \vec{\beta}_i$. Since B_V is a basis, this yields that $c_1 = \cdots = c_\ell = d_1 = \cdots = d_k = 0$, so that $B_{\mathcal{R}}$ is a linearly independent set.

Now suppose that $h(\vec{v}) \in \mathcal{R}(h)$. Since $\vec{v} = \sum_i a_i \vec{\beta}_i + \sum_i b_i \vec{v}_i$, upon using the fact that h is linear we get that

$$\begin{aligned} h(\vec{v}) &= h(\sum_i a_i \vec{\beta}_i) + \sum_i b_i h(\vec{v}_i) \\ &= \vec{0} + \sum_i b_i h(\vec{v}_i). \end{aligned}$$

Hence, $B_{\mathcal{R}}$ is a spanning set for $\mathcal{R}(h)$.

$B_{\mathcal{N}}$ is a basis for $\mathcal{N}(h)$, and $B_{\mathcal{R}}$ is a basis for $\mathcal{R}(h)$. The result is now clear. □

Remark 3.24. (a) It is clear that $\text{rank}(h) \leq \dim(V)$, with equality if and only if $\dim(\mathcal{N}(h)) = 0$.

(b) If $\dim(W) > \dim(V)$, then h cannot be onto, as $\text{rank}(h) \leq \dim(V) < \dim(W)$.

Lemma 3.25. *Let $h : V \to W$ be a homomorphism.* $\dim(\mathcal{N}(h)) = 0$ *if and only if h is one-to-one.*

Proof: If h is one-to-one, then the only solution to $h(\vec{v}) = \vec{0}$ is $\vec{v} = \vec{0}$. Hence, $\dim(\mathcal{N}(h)) = 0$.

Now suppose that $\dim(\mathcal{N}(h)) = 0$. From the above lemma we have that if $B_V = \langle \vec{v}_1, \dots, \vec{v}_n \rangle$ is a basis for V, then $B_{\mathcal{R}} = \langle h(\vec{v}_1), \dots, h(\vec{v}_n) \rangle$ is a basis for $\mathcal{R}(h)$. Suppose that there is a $\vec{w} \in W$ such that $h(\vec{u}_1) = h(\vec{u}_2) = \vec{w}$. We have that $\vec{u}_1 = \sum_i a_i \vec{v}_i$, $\vec{u}_2 = \sum_i b_i \vec{v}_i$, so that upon using the linearity of h,

$$\sum_i a_i h(\vec{v}_i) = \sum_i b_i h(\vec{v}_i).$$

Since $B_{\mathcal{R}}$ is a basis, this implies that $a_i = b_i$ for all i, so that $\vec{u}_1 = \vec{u}_2$. Hence, h is one-to-one. □

Definition 3.26. A one-to-one homomorphism is nonsingular.

3.3. Computing Linear Maps

3.3.1. Representing Linear Maps with Matrices

Recall that if $B = \langle \vec{v}_1, \dots, \vec{v}_n \rangle$ is a basis for V, then uniquely defined homomorphism $h : V \to W$ is given by

$$h(\vec{v}) = h(\sum_i c_i \vec{v}_i) := \sum_i c_i h(\vec{v}_i),$$

i.e., the homomorphism is determined by its action on the basis.

Definition 3.27. Let $A = (\vec{a}_1\, \vec{a}_2\, \cdots\, \vec{a}_n) \in \mathbb{R}^{m \times n}$, and let $\vec{c} \in \mathbb{R}^n$. The matrix-vector product is defined by

$$A\vec{c} = \sum_i c_i \vec{a}_i.$$

Remark 3.28. (a) Matrix multiplication is a homomorphism from $\mathbb{R}^n \to \mathbb{R}^m$.

(b) A linear system can be written as $A\vec{x} = \vec{b}$, where A is the coefficient matrix and $\vec{x}$ is the vector of variables.

Example. Suppose that $h : \mathscr{P}_1 \to \mathbb{R}^3$, and that

$$B = \langle 2, 1 + 4x \rangle, \quad D = \langle \vec{e}_1, -2\vec{e}_2, \vec{e}_1 + \vec{e}_3 \rangle$$

are the bases for these spaces. Suppose that

$$h(2) = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad h(1 + 4x) = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}.$$

It is easy to check that

$$\operatorname{Rep}_D(h(2)) = \begin{pmatrix} 0 \\ -1/2 \\ 1 \end{pmatrix}, \quad \operatorname{Rep}_D(h(1+4x)) = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}.$$

Thus, if $p = c_1 \cdot 2 + c_2 \cdot (1+4x)$, i.e.,

$$\operatorname{Rep}_B(p) = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix},$$

and since $h(p) = c_1 h(2) + c_2 h(1+4x)$, by using the fact that Rep_D is linear one gets that $\operatorname{Rep}_D(h(p)) = c_1 \operatorname{Rep}_D(h(2)) + c_2 \operatorname{Rep}_D(h(1+4x))$. If one defines the matrix

$$\operatorname{Rep}_{B,D}(h) := (\operatorname{Rep}_D(h(2)) \; \operatorname{Rep}_D(h(1+4x))),$$

then one has that

$$\operatorname{Rep}_D(h(p)) = \operatorname{Rep}_{B,D}(h) \operatorname{Rep}_B(p).$$

For example, if

$$\operatorname{Rep}_B(p) = \begin{pmatrix} 1 \\ -2 \end{pmatrix} \quad (\Longrightarrow p(x) = -8x),$$

then

$$\begin{aligned} \operatorname{Rep}_D(h(p)) &= \begin{pmatrix} 0 & 1 \\ -1/2 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ -2 \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ -1/2 \\ 1 \end{pmatrix} - 2 \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} -2 \\ 3/2 \\ 1 \end{pmatrix} \end{aligned}$$

so that

$$\begin{aligned} h(p) &= -2 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - \frac{3}{2} \begin{pmatrix} 0 \\ -2 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} -1 \\ -5 \\ 1 \end{pmatrix}. \end{aligned}$$

Definition 3.29. Let $h : V \to W$ be a homomorphism. Suppose that $B = \langle \vec{v}_1, \dots, \vec{v}_n \rangle$ is a basis for V, and $D = \langle \vec{w}_1, \dots, \vec{w}_m \rangle$ is a basis for W. Set

$$\vec{h}_j := \operatorname{Rep}_D(h(\vec{v}_j)), \quad j = 1, \dots, n.$$

The matrix representation of h with respect to B, D is given by

$$\operatorname{Rep}_{B,D}(h) := (\vec{h}_1 \, \vec{h}_2 \, \cdots \, \vec{h}_n) \in \mathbb{R}^{m \times n}.$$

Lemma 3.30. *Let $h : V \to W$ be a homomorphism. Then*

$$\operatorname{Rep}_D(h(\vec{v})) = \operatorname{Rep}_{B,D}(h) \operatorname{Rep}_B(\vec{v}).$$

Remark 3.31. As a consequence, all linear transformations can be thought of as a matrix multiplication.

Example. (a) Suppose that $V = [\mathrm{e}^x, \mathrm{e}^{3x}]$, and that $h : V \to V$ is given by $h(v) = \int v(x)\,\mathrm{d}x$. Since

$$h(\mathrm{e}^x) = \mathrm{e}^x, \quad h(\mathrm{e}^{3x}) = \frac{1}{3}\mathrm{e}^{3x},$$

we have

$$\mathrm{Rep}_{B,B}(h) = \begin{pmatrix} 1 & 0 \\ 0 & 1/3 \end{pmatrix}.$$

(b) Suppose that a basis for V is $B = \langle \vec{v}_1, \vec{v}_2, \vec{v}_3 \rangle$, and that a basis for W is $D = \langle \vec{w}_1, \vec{w}_2, \vec{w}_3 \rangle$. Further suppose that

$$h(\vec{v}_1) = \vec{w}_1 + 3\vec{w}_2, \quad h(\vec{v}_2) = \vec{w}_2 - \vec{w}_3, \quad h(\vec{v}_3) = -\vec{w}_1 + 4\vec{w}_3.$$

We then have that

$$\mathrm{Rep}_{B,D}(h) = \begin{pmatrix} 1 & 0 & -1 \\ 3 & 1 & 0 \\ 0 & -1 & 4 \end{pmatrix}.$$

Thus, if $\vec{v} = 2\vec{v}_1 - \vec{v}_2 + \vec{v}_3$, we have that

$$\mathrm{Rep}_D(h(\vec{v})) = \mathrm{Rep}_{B,D}(h)\,\mathrm{Rep}_B(\vec{v}) = \begin{pmatrix} 1 \\ 5 \\ 5 \end{pmatrix},$$

so that $h(\vec{v}) = \vec{w}_1 + 5\vec{w}_2 + 5\vec{w}_3$.

3.3.2. Any Matrix Represents a Linear Map

Example. Suppose that $h : \mathscr{P}_2 \to \mathbb{R}^3$, with bases $B = \langle 1, x, x^2 \rangle$ and $D = \langle \vec{e}_1, \vec{e}_2 + \vec{e}_3, \vec{e}_1 - \vec{e}_3 \rangle$, is represented by the matrix

$$H = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \\ 2 & -1 & 1 \end{pmatrix}.$$

In order to decide if $\vec{b} = \vec{e}_1 + 3\vec{e}_2 \in \mathcal{R}(h)$, it is equivalent to determine if $\mathrm{Rep}_D(\vec{b}) \in \mathcal{R}(H)$. Since

$$\mathrm{Rep}_D(\vec{b}) = \begin{pmatrix} -2 \\ 3 \\ 3 \end{pmatrix},$$

and

$$(H|\vec{b}) \longrightarrow \left(\begin{array}{ccc|c} 1 & -1 & 0 & -2 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & 0 & 4 \end{array}\right)$$

we have that $\mathrm{Rep}_D(\vec{b}) \notin \mathcal{R}(H)$; hence, $\vec{b} \notin \mathcal{R}(h)$. Note that $\mathrm{rank}(H) = 2$, so that h is neither one-to-one nor onto.

Let us find a basis for $\mathcal{R}(h)$ and $\mathcal{N}(h)$. We have

$$H \longrightarrow \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad H^T \longrightarrow \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$

so that

$$\mathcal{R}(H) = [\begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}], \quad \mathcal{N}(H) = [\begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}].$$

Using the fact that $\vec{b} \in \mathcal{R}(h)$ if and only if $\text{Rep}_D(\vec{b}) \in \mathcal{R}(H)$, and $\vec{v} \in \mathcal{N}(h)$ if and only if $\text{Rep}_B(\vec{v}) \in \mathcal{N}(H)$, then yields

$$\mathcal{R}(h) = [\begin{pmatrix} 3 \\ 0 \\ -2 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}], \quad \mathcal{N}(h) = [-1 - x + x^2].$$

Theorem 3.32. *Let $A \in \mathbb{R}^{m \times n}$. The map $h : \mathbb{R}^n \to \mathbb{R}^m$ defined by $h(\vec{x}) := A\vec{x}$ is a homomorphism.*

Proof: Set $A = (\vec{a}_1\, \vec{a}_2\, \dots\, \vec{a}_n)$, and recall that $A\vec{x} = \sum_i x_i \vec{a}_i$. Since

$$A(r\vec{x} + s\vec{y}) = \sum_{i=1}^{n} (rx_i + sy_i)\vec{a}_i = r \sum_{i=1}^{n} x_i \vec{a}_i + s \sum_{i=1}^{n} y_i \vec{a}_i = rA\vec{x} + sA\vec{y},$$

h is a homomorphism. □

Theorem 3.33. *Let $h : V \to W$ be a homomorphism which is represented by the matrix H. Then* $\text{rank}(h) = \text{rank}(H)$.

Proof: Let $B = \langle \vec{v}_1, \dots, \vec{v}_n \rangle$ be a basis for V, and let W have a basis D, so that

$$H = (\text{Rep}_D(h(\vec{v}_1)) \, \dots \, \text{Rep}_D(h(\vec{v}_n))).$$

The rank of H is the number of linearly independent columns of H, and the rank of h is the number of linearly independent vectors in the set $\{h(\vec{v}_1), \dots, h(\vec{v}_n)\}$. Since $\text{Rep}_D : W \to \mathbb{R}^m$ is an isomorphism, we have that a set in $\mathcal{R}(h)$ is linearly independent if and only if the related set in $\mathcal{R}(\text{Rep}_D(h))$ is linearly independent (problem 3.1.1.28), i.e., $\{h(\vec{v}_1), \dots, h(\vec{v}_k)\}$ is linearly independent if and only if $\{\text{Rep}_D(h(\vec{v}_1)), \dots, \text{Rep}_D(h(\vec{v}_k))\}$ is linearly independent. The conclusion now follows. □

Corollary 3.34. *(a) h is onto if and only if* $\text{rank}(h) = m$

(b) h is one-to-one if and only if $\text{rank}(h) = n$

(c) h is nonsingular if and only if $m = n$ and $\text{rank}(h) = n$

3.4. Matrix Operations

3.4.1. Sums and Scalar Products

Definition 3.35. Let $A = (a_{i,j}), B = (b_{i,j}) \in \mathbb{R}^{m \times n}$. Then

(a) $A + B = (a_{i,j} + b_{i,j})$

(b) $rA = (ra_{i,j})$ for any $r \in \mathbb{R}$.

Lemma 3.36. *Let $g, h : V \to W$ be homomorphisms represented with respect to the bases B and D by the matrices G, H. The map $g + h$ is represented by $G + H$, and the map rh is represented by rH.*

3.4.2. Matrix Multiplication

Definition 3.37. Let $G \in \mathbb{R}^{m\times n}$ and $H = (\vec{h}_1\,\vec{h}_2\,\cdots\,\vec{h}_p) \in \mathbb{R}^{n\times p}$. Then

$$GH = (G\vec{h}_1\,G\vec{h}_2\,\cdots\,G\vec{h}_p) \in \mathbb{R}^{m\times p}.$$

Example. It is easy to check that

$$\begin{pmatrix} 1 & 1 \\ 2 & -1 \\ 3 & -2 \end{pmatrix} \begin{pmatrix} -1 & 2 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 0 & 5 \\ -3 & 1 \\ -5 & 0 \end{pmatrix}.$$

Remark 3.38. Matrix multiplication is generally not commutative. For example:

(a) if $A \in \mathbb{R}^{2\times 3}$ and $B \in \mathbb{R}^{3\times 2}$, then $AB \in \mathbb{R}^{2\times 2}$ while $BA \in \mathbb{R}^{3\times 3}$

(b) if

$$A = \begin{pmatrix} 1 & -1 \\ 2 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 5 & 2 \\ 4 & 4 \end{pmatrix},$$

then

$$AB = \begin{pmatrix} 1 & -2 \\ 10 & 4 \end{pmatrix}, \quad BA = \begin{pmatrix} 9 & -5 \\ 12 & -4 \end{pmatrix}.$$

Lemma 3.39. *Let $g : V \to W$ and $h : W \to U$ be homomorphisms represented by the matrices G, H. The map $h \circ g : V \to U$ is represented by the matrix HG.*

Proof: Give the commutative diagram:

$$\begin{array}{ccccc} V_B & \xrightarrow{g} & W_D & \xrightarrow{h} & U_E \\ {\scriptstyle \mathrm{Rep}_B}\downarrow & & {\scriptstyle \mathrm{Rep}_D}\downarrow & & \downarrow{\scriptstyle \mathrm{Rep}_E} \\ \mathbb{R}^n & \xrightarrow[G]{} & \mathbb{R}^m & \xrightarrow[H]{} & \mathbb{R}^p \end{array}$$

$\square$

Example. Consider the maps $t_\theta, d_r : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$t_\theta(\vec{v}) := \begin{pmatrix} \cos\theta\, v_1 - \sin\theta\, v_2 \\ \sin\theta\, v_1 + \cos\theta\, v_2 \end{pmatrix}, \quad d_r(\vec{v}) := \begin{pmatrix} 3v_1 \\ v_2 \end{pmatrix}.$$

The matrix representations for these maps are

$$T_\theta := \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \quad D_r := \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}.$$

Rotation followed by dilation is represented by the matrix

$$D_r T_\theta = \begin{pmatrix} 3\cos\theta & -3\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

while dilation followed by rotation is represented by the matrix

$$T_\theta D_r = \begin{pmatrix} 3\cos\theta & -\sin\theta \\ 3\sin\theta & \cos\theta \end{pmatrix}.$$

3.4.3. Mechanics of Matrix Multiplication

Definition 3.40. The identity matrix is given by $I = (\vec{e}_1\ \vec{e}_2\ \cdots\ \vec{e}_n) \in \mathbb{R}^{n\times n}$.

Remark 3.41. Assuming that the multiplication makes sense, $I\vec{v} = \vec{v}$ for any vector $\vec{v} \in \mathbb{R}^n$, and consequently $AI = A$, $IB = B$.

Definition 3.42. A diagonal matrix $D = (d_{i,j}) \in \mathbb{R}^{n\times n}$ is such that $d_{i,j} = 0$ for $i \neq j$.

Definition 3.43. An elementary reduction matrix $R \in \mathbb{R}^{n\times n}$ is formed by applying a single row operation to the identity matrix.

Example. Two examples are

$$I \xrightarrow{-2\rho_1+\rho_2} \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad I \xrightarrow{\rho_1\leftrightarrow\rho_2} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Lemma 3.44. *Let R be an elementary reduction matrix. Then RH is equivalent to performing the Gaussian operation on the matrix H.*

Corollary 3.45. *For any matrix H there are elementary reduction matrices $R_1, \dots, R_k$ such that $R_k R_{k_1} \cdots R_1 H$ is in reduced echelon form.*

3.4.4. Inverses

Definition 3.46. Suppose that $A \in \mathbb{R}^{n\times n}$. The matrix is invertible if there is a matrix A^{-1} such that $AA^{-1} = A^{-1}A = I$.

Remark 3.47. If it exists, the matrix A^{-1} is given by the product of elementary reduction matrices. For example,

$$A = \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \xrightarrow{-2\rho_1+\rho_2} \begin{pmatrix} 1 & 1 \\ 0 & -3 \end{pmatrix} \xrightarrow{-1/3\rho_2} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \xrightarrow{-\rho_2+\rho_1} I.$$

Thus, by setting

$$R_1 = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}, \quad R_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1/3 \end{pmatrix}, \quad R_3 = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix},$$

we have that $R_3R_2R_1A = I$, so that $A^{-1} = R_3R_2R_1$.

Lemma 3.48. *A is invertible if and only if it is nonsingular, i.e., the linear map defined by $h(\vec{x}) := A\vec{x}$ is an isomorphism.*

Proof: A can be row-reduced to I if and only if h is an isomorphism. □

Remark 3.49. When computing A^{-1}, do the reduction $(A|I) \rightarrow (I|A^{-1})$ (if possible).

Example. (a) For A given above,

$$(A|I) \rightarrow \left(\begin{array}{cc|cc} 1 & 0 & 1/3 & 1/3 \\ 0 & 1 & 2/3 & -1/3 \end{array}\right).$$

(b) For a general $A \in \mathbb{R}^{2\times 2}$ we have that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

if $ad - bc \neq 0$.

3.5. *Change of Basis*

3.5.1. Changing Representations of Vectors

For the vector space V let one basis be given by $B = \langle \vec{\beta}_1, \dots, \vec{\beta}_n \rangle$, and let another be given by $D = \langle \vec{\delta}_1, \dots, \vec{\delta}_n \rangle$. Define the homomorphism $h : V \rightarrow V$ by $h(\vec{\beta}_j) := \vec{\beta}_j$, i.e., $h = \text{id}$, the identity map. The transformation matrix H associated with the identity map satisfies $\vec{h}_j = \text{Rep}_D(\vec{\beta}_j)$.

Definition 3.50. The change of basis matrix $H = \text{Rep}_{B,D}(\text{id})$ for the bases B, D is the representation of the identity map with respect to these bases, and satisfies $\vec{h}_j = \text{Rep}_D(\vec{\beta}_j)$.

Example. Suppose that

$$B = \langle \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \rangle, \quad D = \langle \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \rangle.$$

We have that

$$\begin{array}{ccc} V_B & \xrightarrow{\text{id}} & V_D \\ {\scriptstyle \text{Rep}_B}\downarrow & & {\scriptstyle \text{Rep}_D}\downarrow \\ \mathbb{R}^2 & \xrightarrow[H=\text{Rep}_{B,D}]{} & \mathbb{R}^2 \end{array}$$

It can be checked that

$$\mathrm{Rep}_D(\begin{pmatrix} 1 \\ 1 \end{pmatrix}) = \begin{pmatrix} 2/3 \\ -1/3 \end{pmatrix}, \quad \mathrm{Rep}_D(\begin{pmatrix} 1 \\ 2 \end{pmatrix}) = \begin{pmatrix} 1 \\ -1 \end{pmatrix},$$

so that the change of basis matrix is

$$H = \begin{pmatrix} 2/3 & 1 \\ -1/3 & -1 \end{pmatrix}.$$

Recall that $H\,\mathrm{Rep}_B(\vec{v}) = \mathrm{Rep}_D(\vec{v})$. For the vector $\vec{v} = (-1, -3)^T$ one has that $\mathrm{Rep}_B(\vec{v}) = (1, -2)^T$, so that

$$\mathrm{Rep}_D(\vec{v}) = H\,\mathrm{Rep}_B(\vec{v}) = \frac{1}{3}\begin{pmatrix} -4 \\ 5 \end{pmatrix}.$$

What about the change of basis matrix from D to B? It must be H^{-1} (use a commutative diagram to show it).

Example (cont.)**.** The change of basis matrix from D to B is

$$H^{-1} = \begin{pmatrix} 3 & 3 \\ -1 & -2 \end{pmatrix}.$$

Note that this implies that

$$\mathrm{Rep}_B(\begin{pmatrix} 2 \\ 1 \end{pmatrix}) = \begin{pmatrix} 3 \\ -1 \end{pmatrix}, \quad \mathrm{Rep}_B(\begin{pmatrix} 1 \\ -1 \end{pmatrix}) = \begin{pmatrix} 3 \\ -2 \end{pmatrix}.$$

3.5.2. Changing Map Representations

Consider the homomorphism $h : V \to W$, and suppose that V has bases $B, \hat{B}$, while W has bases $D, \hat{D}$. With respect to B, D there is a transformation matrix H, while with respect to $\hat{B}, \hat{D}$ there is a transformation matrix $\hat{H}$, i.e.,

$$\begin{array}{ccc} V_B & \xrightarrow[H]{h} & W_D \\ {\scriptstyle \mathrm{id}}\downarrow & & \downarrow{\scriptstyle \mathrm{id}} \\ V_{\hat{B}} & \xrightarrow[\hat{H}]{h} & W_{\hat{D}} \end{array}$$

We have that

$$\begin{aligned} \hat{H} &= \mathrm{Rep}_{D,\hat{D}}(\mathrm{id}) \cdot H \cdot \mathrm{Rep}_{\hat{B},B}(\mathrm{id}) \\ &= \mathrm{Rep}_{D,\hat{D}}(\mathrm{id}) \cdot H \cdot \mathrm{Rep}_{B,\hat{B}}(\mathrm{id})^{-1} \\ &= \mathrm{Rep}_{\hat{D},D}^{-1}(\mathrm{id}) \cdot H \cdot \mathrm{Rep}_{\hat{B},B}(\mathrm{id}). \end{aligned}$$

The idea is that, if possible, we wish to choose bases $\hat{B}, \hat{D}$ such that $h(\vec{\beta}_j) = a_j\vec{\delta}_j$ for some $a_j \in \mathbb{R}$.

Definition 3.51. $H, \hat{H} \in \mathbb{R}^{m\times n}$ are matrix equivalent if there are nonsingular matrices P, Q such that $\hat{H} = PHQ$.

Lemma 3.52. *Matrix equivalent matrices represent the same map with respect to appropriate pairs of bases.*

Example. In the standard basis $\mathscr{E}_2$ consider the homomorphism

$$h(\vec{x}) = \begin{pmatrix} 4 & -2 \\ 1 & 1 \end{pmatrix}\vec{x}, \quad H = \begin{pmatrix} 4 & -2 \\ 1 & 1 \end{pmatrix}.$$

Consider the basis

$$D = \langle \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix} \rangle.$$

We have that

$$\text{Rep}_{D,\mathscr{E}_2}(\text{id}) = \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix},$$

so that

$$\hat{H} = \text{Rep}_{D,\mathscr{E}_2}(\text{id})^{-1} H \, \text{Rep}_{D,\mathscr{E}_2}(\text{id}) = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix};$$

hence, the homomorphism has the desired property with the basis D, i.e.,

$$h(\begin{pmatrix} 1 \\ 1 \end{pmatrix}) = 2\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad h(\begin{pmatrix} 2 \\ 1 \end{pmatrix}) = 3\begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Remark 3.53. The above example does not really change if, with respect to the standard basis $B = \langle 1, x \rangle$, the homomorphism $h : \mathscr{P}_2 \to \mathscr{P}_2$ satisfies $h(1) = 4 + x, h(x) = -2 + x$, so that the matrix representing the homomorphism is that given above. If we set $D = \langle 1 + x, 2 + x \rangle$, then one then has that

$$h(1 + x) = 2(1 + x), \quad h(2 + x) = 3(2 + x).$$

3.6. Projection

3.6.1. Orthogonal Projection into a Line

Let the line ℓ be given by $\ell := [\vec{s}]$, and let $\vec{v} \in \mathbb{R}^n$ be a given vector. The orthogonal projection of $\vec{v}$ onto ℓ is given by $\vec{v}_{\vec{p}} = c_{\vec{p}}\vec{s}$, where $c_{\vec{p}}$ is chosen so that

$$\vec{v} = (\vec{v} - c_{\vec{p}}\vec{s}) + c_{\vec{p}}\vec{s}, \quad (\vec{v} - c_{\vec{p}}\vec{s}) \cdot \vec{s} = 0$$

(give a picture). Note that the second condition implies that $\vec{v} - c_{\vec{p}}\vec{s} \in \ell^\perp$. The second condition yields that

$$c_{\vec{p}} = \frac{\vec{v} \cdot \vec{s}}{\vec{s} \cdot \vec{s}}.$$

Definition 3.54. The orthogonal projection of $\vec{v}$ onto the line spanned by $\vec{s}$ is the vector

$$\text{proj}_{[\vec{s}]}(\vec{v}) = \frac{\vec{v} \cdot \vec{s}}{\vec{s} \cdot \vec{s}}\vec{s}.$$

Remark 3.55. (a) By construction, $\vec{v} - \text{proj}_{[\vec{s}]}(\vec{v}) \in [\vec{s}]^\perp$.

(b) Since $\text{proj}_{[\vec{s}]} : \mathbb{R}^n \to \mathbb{R}^n$ is a homomorphism, it is represented by a matrix $\mathbb{P}_{[\vec{s}]}$, which is given by

$$\begin{aligned}\mathbb{P}_{[\vec{s}]} &:= \frac{1}{\vec{s} \cdot \vec{s}}(s_1\vec{s}, s_2\vec{s}, \ldots, s_n\vec{s}) \\ &= (\vec{s})(\frac{s_1}{\vec{s} \cdot \vec{s}}, \ldots, \frac{s_n}{\vec{s} \cdot \vec{s}}).\end{aligned}$$

Note that $\text{rank}(\mathbb{P}_{[\vec{s}]}) = 1$, and hence $\dim(\mathcal{N}(\mathbb{P}_{[\vec{s}]})) = n - 1$. Further note that $[\vec{s}]^\perp = \mathcal{N}(\mathbb{P}_{[\vec{s}]})$.

Example. Suppose that

$$\vec{v} = \begin{pmatrix} 1 \\ 1 \\ 4 \end{pmatrix}, \quad \vec{s} = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}.$$

Then

$$\text{proj}_{[\vec{s}]}(\vec{v}) = -\frac{1}{6}\vec{s}.$$

3.6.2. Gram-Schmidt Orthogonalization

Definition 3.56. The vectors $\vec{v}_1, \dots, \vec{v}_k \in \mathbb{R}^n$ are mutually orthogonal if $\vec{v}_i \cdot \vec{v}_j = 0$ for any $i \neq j$.

Theorem 3.57. *Suppose that the nonzero vectors $\vec{v}_1, \dots, \vec{v}_k \in \mathbb{R}^n$ are mutually orthogonal. The set $\{\vec{v}_1, \dots, \vec{v}_k\}$ is then linearly independent.*

Proof: Suppose that $\sum_i c_i\vec{v}_i = \vec{0}$. For each j take the dot product with $\vec{v}_j$, so that

$$\vec{v}_j \cdot \vec{0} = \vec{v}_j \cdot (\sum_i c_i\vec{v}_i) = \sum_i c_i(\vec{v}_j \cdot \vec{v}_i) = c_j\vec{v}_j \cdot \vec{v}_j.$$

Since the vectors are nonzero, this implies that $c_j = 0$. Hence, the vectors are linearly independent. □

Corollary 3.58. *The set $\{\vec{v}_1, \dots, \vec{v}_k\}$ is a basis.*

Definition 3.59. The basis given above is an orthogonal basis.

Recall that if $\vec{v} \in \mathbb{R}^n$, then $\|\vec{v}\|^2 = \vec{v} \cdot \vec{v}$.

Lemma 3.60. *Suppose that $B = \langle\vec{\kappa}_1, \dots, \vec{\kappa}_\ell\rangle$ is an orthogonal basis for the subspace $S \subset \mathbb{R}^n$. If $\vec{v} = \sum_i c_i\vec{\kappa}_i \in S$, then*

(a) $c_i = \dfrac{\vec{v} \cdot \vec{\kappa}_i}{\vec{\kappa}_i \cdot \vec{\kappa}_i}$

(b) $\|\vec{v}\|^2 = \displaystyle\sum_{i=1}^{\ell} c_i^2\vec{\kappa}_i \cdot \vec{\kappa}_i.$

Proof: Follows immediately from the fact that B is an orthogonal basis. For part (a) consider $\vec{v} \cdot \vec{\kappa}_i$, and for part (b) simply look at $\vec{v} \cdot \vec{v}$. □

Lemma 3.61. *Let $\langle\vec{\kappa}_1,\dots,\vec{\kappa}_\ell\rangle$ be an orthogonal basis for a subspace S. For a given $\vec{v}$ set*

$$\vec{p}=\sum_i c_i\vec{\kappa}_i,\quad c_i=\frac{\vec{v}\cdot\vec{\kappa}_i}{\vec{\kappa}_i\cdot\vec{\kappa}_i}.$$

Then $\vec{p}-\vec{v}\in S^\perp$.

Proof: First note that

$$\begin{aligned}\vec{\kappa}_j\cdot(\vec{p}-\vec{v})&=\vec{\kappa}_j\cdot\vec{p}-\vec{\kappa}_j\cdot\vec{v}\\&=\vec{\kappa}_j\cdot(\sum_i c_i\vec{\kappa}_i)-c_j\vec{\kappa}_j\cdot\vec{\kappa}_j\\&=(\sum_i c_i\vec{\kappa}_j\cdot\vec{\kappa}_i)-c_j\vec{\kappa}_j\cdot\vec{\kappa}_j\\&=c_j\vec{\kappa}_j\cdot\vec{\kappa}_j-c_j\vec{\kappa}_j\cdot\vec{\kappa}_j\\&=0,\end{aligned}$$

so that $\vec{p}-\vec{v}$ is orthogonal to each $\vec{\kappa}_j$. If $\vec{x}\in S$, so that $\vec{x}=\sum_i d_i\vec{\kappa}_i$, by using the linearity of the dot product it is clear that $(\vec{p}-\vec{v})\cdot\vec{x}=0$. Hence, $\vec{p}-\vec{v}\in S^\perp$. □

Definition 3.62. Define the homomorphism $\text{proj}_S:\mathbb{R}^n\to\mathbb{R}^n$ by

$$\text{proj}_S(\vec{v})=\sum_i \text{proj}_{[\vec{\kappa}_i]}(\vec{v}).$$

Remark 3.63. From the above one has that $\vec{v}-\text{proj}_S(\vec{v})\in S^\perp$. The matrix representation for proj_S, $\mathbb{P}_S$, is given by

$$\mathbb{P}_S=\sum_i\mathbb{P}_{[\vec{\kappa}_i]}.$$

Since $\langle\vec{\kappa}_1,\dots,\vec{\kappa}_\ell\rangle$ is an orthogonal basis, one has that $\text{rank}(\mathbb{P}_S)=\ell$, and hence that $\dim(S^\perp)=\dim(\mathcal{N}(\mathbb{P}_S))=n-\ell$.

Recall that if $S=[\vec{\beta}_1,\dots,\vec{\beta}_k]$, then a basis for S is not at all unique. The question is how to find an orthogonal basis for S. For example, suppose that

$$S=[\begin{pmatrix}1\\1\\1\end{pmatrix},\begin{pmatrix}1\\2\\3\end{pmatrix},\begin{pmatrix}5\\6\\7\end{pmatrix}].$$

Keep in mind that $\dim(S)\le 3$. Set

$$\vec{\kappa}_1=\begin{pmatrix}1\\1\\1\end{pmatrix}.$$

Define $\vec{\kappa}_2$ by

$$\vec{\kappa}_2=\begin{pmatrix}1\\2\\3\end{pmatrix}-\text{proj}_{[\vec{\kappa}_1]}(\begin{pmatrix}1\\2\\3\end{pmatrix})=\begin{pmatrix}-1\\0\\1\end{pmatrix}.$$

If $S = [\vec{\kappa}_1, \vec{\kappa}_2]$, now define

$$\begin{aligned}\vec{\kappa}_3 &= \begin{pmatrix} 5 \\ 6 \\ 7 \end{pmatrix} - \text{proj}_S(\begin{pmatrix} 5 \\ 6 \\ 7 \end{pmatrix}) \\ &= \begin{pmatrix} 5 \\ 6 \\ 7 \end{pmatrix} - \text{proj}_{[\vec{\kappa}_1]}(\begin{pmatrix} 5 \\ 6 \\ 7 \end{pmatrix}) - \text{proj}_{[\vec{\kappa}_2]}(\begin{pmatrix} 5 \\ 6 \\ 7 \end{pmatrix}) \\ &= \vec{0}.\end{aligned}$$

An orthogonal basis for S is then

$$\langle \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \rangle,$$

so that $\dim(S) = 2$.

Theorem 3.64 (Gram-Schmidt orthogonalization)**.** *Suppose that $S = [\vec{\beta}_1, \dots, \vec{\beta}_k]$. An orthogonal basis for S is constructed via*

$$\begin{aligned}\vec{\kappa}_1 &= \vec{\beta}_1 \\ \vec{\kappa}_{\ell+1} &= \vec{\beta}_{\ell+1} - \text{proj}_{S_\ell}(\vec{\beta}_{\ell+1}),\end{aligned}$$

where $S_\ell = [\vec{\kappa}_1, \dots, \vec{\kappa}_\ell]$.

Remark 3.65. The procedure can alternatively be written as

$$\begin{aligned}\vec{\kappa}_1 &= \vec{\beta}_1 \\ \vec{\kappa}_2 &= \vec{\beta}_2 - \text{proj}_{[\vec{\kappa}_1]}(\vec{\beta}_2) \\ \vec{\kappa}_3 &= \vec{\beta}_3 - \text{proj}_{[\vec{\kappa}_1]}(\vec{\beta}_3) - \text{proj}_{[\vec{\kappa}_2]}(\vec{\beta}_3) \\ &\vdots \\ \vec{\kappa}_k &= \vec{\beta}_k - \text{proj}_{[\vec{\kappa}_1]}(\vec{\beta}_k) - \cdots - \text{proj}_{[\vec{\kappa}_{k-1}]}(\vec{\beta}_k).\end{aligned}$$

Proof: The argument will be by induction. It is clear that $[\vec{\kappa}_1] = [\vec{\beta}_1]$. Define $S_\ell = [\vec{\kappa}_1, \dots, \vec{\kappa}_\ell]$. Suppose that for some $1 < \ell < k$ that $S_\ell = [\vec{\beta}_1, \dots, \vec{\beta}_\ell]$. Set

$$\vec{p}_\ell = \text{proj}_{S_\ell}(\vec{\beta}_{\ell+1}) = \text{proj}_{[\vec{\kappa}_1]}(\vec{\beta}_{\ell+1}) + \cdots + \text{proj}_{[\vec{\kappa}_\ell]}(\vec{\beta}_{\ell+1}).$$

It is clear that $\vec{p}_\ell \in S_\ell$. By supposition, $\vec{\kappa}_{\ell+1} = \vec{\beta}_{\ell+1} - \vec{p} \in [\vec{\beta}_1, \dots, \vec{\beta}_{\ell+1}]$. Since $\vec{p} \in S_\ell$, it is clear that $\vec{\beta}_{\ell+1} = \vec{p} + \vec{\kappa}_{\ell+1} \in S_{\ell+1}$. Thus, $S_{\ell+1} = [\vec{\beta}_1, \dots, \vec{\beta}_{\ell+1}]$. From the above lemma $\vec{\kappa}_{\ell+1} = \vec{\beta}_{\ell+1} - \vec{p} \in S_\ell^\perp$, so that the basis for $S_{\ell+1}$ is orthogonal. It then follows that an orthogonal basis for S is given by $\langle \vec{\kappa}_1, \dots, \vec{\kappa}_k \rangle$. □

3.6.3. Projection into a Subspace

Let a basis (not necessarily orthogonal) for a subspace M be given by $\langle \vec{v}_1, \dots, \vec{v}_\ell \rangle$. Given a $\vec{b} \in \mathbb{R}^n$, one could compute $\text{proj}_M(\vec{b})$ by first using the Gram-Schmidt procedure to get an orthogonal basis for M, and then using the projection formula. However, that is a lengthy process. Let us try another approach. Set $A = (\vec{v}_1\, \vec{v}_2\, \cdots\, \vec{v}_\ell) \in \mathbb{R}^{n\times\ell}$. Since $\text{proj}_M(\vec{b}) \in M$, by definition $\text{proj}_M(\vec{b}) \in \mathcal{R}(A)$, i.e., there is an $\vec{x}^*$ such that $A\vec{x}^* = \text{proj}_M(\vec{b})$. Recall that the projection also satisfies $\vec{b} - \text{proj}_M(\vec{b}) = \vec{b} - A\vec{x}^* \in \mathcal{R}(A)^\perp = \mathcal{N}(A^T)$. Thus, $A^T(\vec{b} - A\vec{x}^*) = \vec{0}$, so that the vector $\vec{x}^*$ is given as the solution to $A^TA\vec{x} = A^T\vec{b}$.

Is the vector $\vec{x}^*$ unique? We must show that given $\text{rank}(A) = \ell$, then $\dim(\mathcal{N}(A^TA)) = 0$. If $\vec{x} \in \mathcal{N}(A)$, it is clear that $\vec{x} \in \mathcal{N}(A^TA)$; hence, $\mathcal{N}(A) \subset \mathcal{N}(A^TA)$. Now suppose that $\vec{x} \in \mathcal{N}(A^TA)$. This clearly implies that $A\vec{x} \in \mathcal{N}(A^T)$. It is equally clear that $A\vec{x} \in \mathcal{R}(A)$. Since $\mathcal{R}(A)^\perp = \mathcal{N}(A^T)$, this necessarily implies that $A\vec{x} = \vec{0}$, i.e., $\vec{x} \in \mathcal{N}(A)$. Hence, $\mathcal{N}(A^TA) \subset \mathcal{N}(A)$, so that $\mathcal{N}(A) = \mathcal{N}(A^TA)$. Since $\text{rank}(A) = \ell$ implies that $\dim(\mathcal{N}(A)) = 0$, one then gets that $\dim(\mathcal{N}(A^TA)) = 0$. Thus, $\vec{x}^*$ is unique.

Lemma 3.66. *For a given $A \in \mathbb{R}^{n\times\ell}$ with $\text{rank}(A) = \ell$, and for a given $\vec{b} \in \mathbb{R}^n$, the projection of $\vec{b}$ onto $\mathcal{R}(A)$ is given by $A\vec{x}^*$, where $\vec{x}^*$ is the unique solution to the system $A^TA\vec{x} = A^T\vec{b}$.*

Definition 3.67. The least-squares solution of $A\vec{x} = \vec{b}$ is the solution to $A^TA\vec{x} = A^T\vec{b}$.

Remark 3.68. (a) By construction, the least-squares solution satisfies $\|A\vec{x}^* - \vec{b}\| \le \|A\vec{x} - \vec{b}\|$ for any $\vec{x} \neq \vec{x}^*$.

(b) If $\vec{b} \in \mathcal{R}(A)$, then the least-squares solution is the true solution.

Topic: Line of Best Fit

Suppose that you have a collection of point $(x_1, y_1), \dots, (x_n, y_n)$ that you wish to fit with a curve of the form

$$y = c_0f_0(x) + c_1f_1(x) + \cdots + c_kf_k(x).$$

If the points were to lie exactly on the proposed curve, then the following system of equations would be satisfied:

$$y_j = c_0f_0(x_j) + c_1f_1(x_j) + \cdots + c_kf_k(x_j), \quad j = 1, \dots, n,$$

i.e.,

$$A\vec{c} = \vec{y}, \quad A = \begin{pmatrix} f_0(x_1) & \cdots & f_k(x_1) \\ \vdots & & \vdots \\ f_0(x_n) & \cdots & f_k(x_n) \end{pmatrix}.$$

It is highly unlikely, however, that the points will lie exactly on the proposed curve. The curve of best fit would be given by choosing the vector $\vec{c}$ so that $\|A\vec{c} - \vec{y}\|$ is as small as possible. The vector $\vec{c}$ is the solution to the least-squares problem $A^TA\vec{c} = A^T\vec{y}$.

For the first example, suppose that you wish to fit the data $(1, 0), (2, 1), (4, 2), (5, 3)$ with a line. One then gets that

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 4 \\ 1 & 5 \end{pmatrix}, \quad \vec{y} = \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \end{pmatrix}.$$

The least-squares solution is $\vec{c} = (-3/5, 7/10)^T$, so that the line of best fit is

$$y = -\frac{3}{5} + \frac{7}{10}x.$$

Now consider the following example. Among the important inputs in weather forecasting models are data sets consisting of temperature values at various parts of the atmosphere. These are either measured directly using weather balloons or inferred from remote soundings taken by weather satellites. A typical set of RAOB (weather balloon) data is given below:

p	1	2	3	4	5	6	7	8	9	10
T	222	227	223	233	244	253	260	266	270	266

The temperature T in kelvins may be considered as a function of p, the atmospheric pressure measured in decibars. Pressures in the range from 1 to 3 decibars correspond to the top of the atmosphere, and those in the range from 9 to 10 correspond to the lower part of the atmosphere. The linear and cubic least-squares fits are given below.

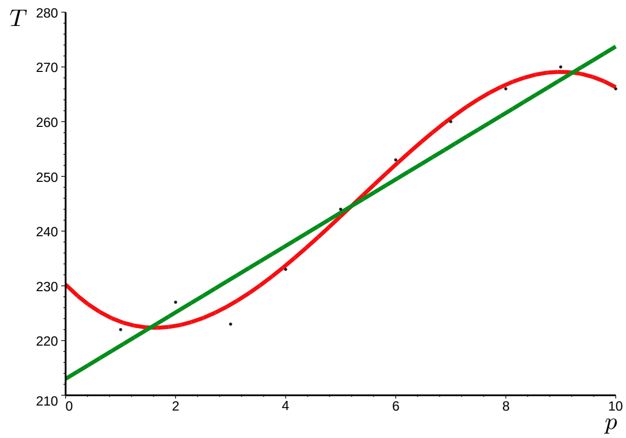


Figure 1: Comparison of a linear fit with a cubic fit for the temperature vs. pressure data.

4. Determinants

4.1. Definition

4.1.1. Exploration

Consider the matrix $A \in \mathbb{R}^{2\times 2}$ given by

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

It is known that A is nonsingular if and only if $ad - bc \neq 0$.

Definition 4.1. If $A \in \mathbb{R}^{2\times 2}$, define the determinant of A to be

$$\det(A) = ad - bc.$$

Remark 4.2. Often one writes $|A| := \det(A)$.

The following properties are easy to verify:

(a) $\det(A^T) = \det(A)$

(b) $\det(AB) = \det(A)\det(B)$.

Recall that Gaussian elimination is equivalent to multiplication by one or more elementary reduction matrices. Using this fact and the above properties yields the further properties:

(a) if

$$\hat{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} A,$$

then $\det(\hat{A}) = -\det(A)$

(b) if

$$\hat{A} = \begin{pmatrix} r & 0 \\ 0 & 1 \end{pmatrix} A,$$

then $\det(\hat{A}) = r\det(A)$

(c) if

$$\hat{A} = \begin{pmatrix} 1 & 0 \\ r & 1 \end{pmatrix} A,$$

then $\det(\hat{A}) = \det(A)$

4.1.2. Properties of Determinants

Let

$$A = (\vec{\rho}_1, \vec{\rho}_2, \dots, \vec{\rho}_n) \in \mathbb{R}^{n\times n},$$

where $\vec{\rho}_j$ represents row j of A.

Definition 4.3. A determinant is a function $\det : \mathbb{R}^{n\times n} \to \mathbb{R}$ such that

(a) $\det(\vec{\rho}_1, \dots, r\vec{\rho}_i + \vec{\rho}_j, \dots, \vec{\rho}_n) = \det(\vec{\rho}_1, \dots, \vec{\rho}_j, \dots, \vec{\rho}_n)$ for $i \neq j$

(b) $\det(\vec{\rho}_1, \dots, \vec{\rho}_i, \dots, \vec{\rho}_j, \dots, \vec{\rho}_n) = -\det(\vec{\rho}_1, \dots, \vec{\rho}_j, \dots, \vec{\rho}_i, \dots, \vec{\rho}_n)$ for $i \neq j$

(c) $\det(\vec{\rho}_1, \dots, r\vec{\rho}_i, \dots, \vec{\rho}_n) = r\det(\vec{\rho}_1, \dots, \vec{\rho}_i, \dots, \vec{\rho}_n)$

(d) $\det(I) = 1$, where $I \in \mathbb{R}^{n\times n}$ is the identity matrix

Lemma 4.4. *The determinant satisfies the following properties:*

(a) if A has two identical rows, then $\det(A) = 0$

(b) if A has a zero row, then $\det(A) = 0$

(c) A is nonsingular if and only if $\det(A) \neq 0$

(d) if A is in echelon form, then $\det(A)$ is the product of the diagonal entries.

Proof: First suppose that A has two equal rows. Using property (b) yields that $\det(A) = -\det(A)$, so that $\det(A) = 0$.

Now suppose that A has a zero row. Using property (a) gives

$$\det(\dots, \vec{\rho}_i, \dots, \vec{0}, \dots) = \det(\dots, \vec{\rho}_i, \dots, \vec{\rho}_i + \vec{0}, \dots).$$

The matrix on the right has two identical rows, so the above result gives that $\det(A) = 0$.

Let $\hat{A}$ be the Gauss-Jordan reduction of A. As an application of the first three properties one has that $\det(A) = \alpha\det(\hat{A})$, where α is some nonzero constant. If A is nonsingular, $\hat{A} = I$, so that by property (d) $\det(\hat{A}) = 1$, and hence $\det(A) = \alpha \neq 0$. If A is singular, then $\hat{A}$ has at least one zero row, so by the above result $\det(\hat{A}) = 0$, and hence $\det(A) = 0$.

Finally, suppose that A is in echelon form. If one of the diagonal entries is zero, then in reduced echelon form the matrix will have a zero row, and hence $\det(A) = 0$. This is exactly the product of the diagonal entries. Suppose that none of the diagonal entries is zero, i.e., $a_{i,i} \neq 0$ for $i = 1, \dots, n$. Upon using property (c) and factoring $a_{i,i}$ from row i one has that

$$\det(A) = a_{1,1}a_{2,2}\cdots a_{n,n}\det(\hat{A}),$$

where $\hat{A}$ is such that $\hat{a}_{i,i} = 1$ for $i = 1, \dots, n$. The matrix $\hat{A}$ can clearly be row-reduced to I; furthermore, this reduction only requires property (a), so that $\det(\hat{A}) = \det(I) = 1$. This yields the desired result. □

Example. The above result gives us an indication as to how to calculate $\det(A)$ if $n \geq 3$. Simply row-reduce A to echelon form keeping track of the row-swapping and row multiplication, and then use result (d) from the above. For example,

$$\begin{vmatrix} 1 & 2 & 3/2 \\ -3 & 4 & -2 \\ 1 & 0 & -2 \end{vmatrix} = \frac{1}{2}\begin{vmatrix} 2 & 4 & 3 \\ -3 & 4 & -2 \\ 1 & 0 & -2 \end{vmatrix} = -\frac{1}{2}\begin{vmatrix} 1 & 0 & -2 \\ -3 & 4 & -2 \\ 2 & 4 & 3 \end{vmatrix}$$
$$= -\frac{1}{2}\begin{vmatrix} 1 & 0 & -2 \\ 0 & 4 & -8 \\ 0 & 4 & 7 \end{vmatrix} = -\frac{1}{2}\begin{vmatrix} 1 & 0 & -2 \\ 0 & 4 & -8 \\ 0 & 0 & 15 \end{vmatrix} = -30.$$

Lemma 4.5. *The determinant has the following properties:*

(a) $\det(A^T) = \det(A)$

(b) $\det(AB) = \det(A)\det(B)$

(c) if A is nonsingular, then $\det(A^{-1}) = \det(A)^{-1}$.

4.3. Other Formulas

4.3.1. Laplace's Expansion

Definition 4.6. For a given $A \in \mathbb{R}^{n\times n}$, let $A_{i,j} \in \mathbb{R}^{(n-1)\times(n-1)}$ be the matrix formed by deleting row i and column j from A. The matrix $A_{i,j}$ is the i,j minor of A. The i,j cofactor is given by $(-1)^{i+j}\det(A_{i,j})$.

Example. If

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix},$$

then

$$A_{1,2} = \begin{pmatrix} 4 & 6 \\ 7 & 9 \end{pmatrix}, \quad A_{2,3} = \begin{pmatrix} 1 & 2 \\ 7 & 8 \end{pmatrix}$$

Theorem 4.7 (Laplace Expansion of Determinants)**.** *If $A \in \mathbb{R}^{n\times n}$, then*

$$\begin{aligned} \det(A) &= \sum_{k=1}^{n}(-1)^{i+k}a_{i,k}\det(A_{i,k}) \quad \textit{(across row } i\textit{)} \\ &= \sum_{k=1}^{n}(-1)^{k+j}a_{k,j}\det(A_{k,j}) \quad \textit{(down column } j\textit{)}. \end{aligned}$$

Example. One has that

$$\begin{aligned} \begin{vmatrix} 1 & 0 & -5 \\ 2 & 1 & 3 \\ -3 & 0 & 6 \end{vmatrix} &= (-1)^{1+2}\cdot 0 \cdot \begin{vmatrix} 2 & 3 \\ -3 & 6 \end{vmatrix} + (-1)^{2+2}\cdot 1 \cdot \begin{vmatrix} 1 & -5 \\ -3 & 6 \end{vmatrix} + (-1)^{3+2}\cdot 0 \cdot \begin{vmatrix} 1 & -5 \\ 2 & 3 \end{vmatrix} \\ &= -0\cdot(21) + 1\cdot(-9) - 0\cdot(13) = -9. \end{aligned}$$

Topic: Cramer's Rule

Consider the linear system $A\vec{x} = \vec{b}$, where $A \in \mathbb{R}^{n\times n}$ is such that $\det(A) \neq 0$. Since A is nonsingular, the solution is given by $\vec{x} = A^{-1}\vec{b}$. Let $B_j \in \mathbb{R}^{n\times n}$ be the matrix formed by substituting $\vec{b}$ into column j of A.

Theorem 4.8 (Cramer's Rule)**.** *The solution vector $\vec{x}$ is given by*

$$x_j = \frac{\det(B_j)}{\det(A)}.$$

Example. If

$$A = \begin{pmatrix} 2 & 1 & 1 \\ 3 & 0 & 1 \\ 1 & -1 & -1 \end{pmatrix}, \quad \vec{b} = \begin{pmatrix} 1 \\ 4 \\ 2 \end{pmatrix},$$

then for the system $A\vec{x} = \vec{b}$ one has that

$$B_3 = \begin{pmatrix} 2 & 1 & 1 \\ 3 & 0 & 4 \\ 1 & -1 & 2 \end{pmatrix}.$$

Hence,

$$x_3 = \frac{\det(B_j)}{\det(A)} = \frac{3}{3} = 1.$$

5. Similarity

5.2. Similarity

5.2.1. Definition and Examples

Example. Consider the following table:

Married	Single	
0.85	0.10	Married
0.15	0.90	Single

Set $\vec{x}_j = (m_j, s_j)^T$, where m_j represents the number of married men in year j , and s_j is the number of single men in year j. The above table gives the percentage of married men who will stay married, etc. If

$$A = \begin{pmatrix} 0.85 & 0.10 \\ 0.15 & 0.90 \end{pmatrix},$$

we have the relationship $\vec{x}_{n+1} = A\vec{x}_n$. This is a dynamical system. Since

$$\vec{x}_1 = A\vec{x}_0, \quad \vec{x}_2 = A\vec{x}_1 = A^2\vec{x}_0, \dots, \vec{x}_n = A\vec{x}_{n-1} = A^n\vec{x}_0,$$

if we wish to understand the final distribution of married men vs. single men, we need to understand $\lim_{n\to\infty} A^n$. How do we accomplish this task without actually doing the multiplication?

Recall that two matrices H and $\hat{H}$ are equivalent if $\hat{H} = PHQ$. From the commutative diagram

$$\begin{array}{ccc} V_B & \xrightarrow[H]{h} & W_D \\ {\scriptstyle \text{id}}\downarrow & & \downarrow{\scriptstyle \text{id}} \\ V_{\hat{B}} & \xrightarrow[\hat{H}]{h} & W_{\hat{D}} \end{array}$$

we have that

$$\begin{aligned} \hat{H} &= \text{Rep}_{D,\hat{D}}(\text{id}) \cdot H \cdot \text{Rep}_{\hat{B},B}(\text{id}) \\ &= \text{Rep}_{D,\hat{D}}(\text{id}) \cdot H \cdot \text{Rep}_{B,\hat{B}}(\text{id})^{-1}. \end{aligned}$$

If we restrict ourselves to the following situation,

$$\begin{array}{ccc} V_B & \xrightarrow[H]{h} & V_B \\ {\scriptstyle \text{id}}\downarrow & & \downarrow{\scriptstyle \text{id}} \\ V_D & \xrightarrow[\hat{H}]{h} & V_D \end{array}$$

we have that

$$\hat{H} = \text{Rep}_{B,D}(\text{id}) \cdot H \cdot \text{Rep}_{B,D}(\text{id})^{-1}.$$

Definition 5.1. The matrices T and S are similar if there is a nonsingular matrix P such that $T = PSP^{-1}$.

Remark 5.2. (a) The above can be rewritten as $TP = PS$, which, as we will see, allows us to choose the optimal P.

(b) Two matrices are also similar if $QTQ^{-1} = S$, as $Q = P^{-1}$.

Example. Consider the matrices

$$A = \begin{pmatrix} 4 & -2 \\ 1 & 1 \end{pmatrix}, \quad P = \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix}.$$

It is easy to check that

$$P^{-1}AP = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}.$$

Proposition 5.3. *Suppose that* $T = PSP^{-1}$. *Then* $\det(T) = \det(S)$.

Proof: Recall that $\det(AB) = \det(A)\det(B)$, and that $\det(A^{-1}) = \det(A)^{-1}$. Using these properties

$$\det(T) = \det(P)\det(S)\det(P^{-1}) = \det(S)\det(P)\frac{1}{\det(P)} = \det(S).$$ □

Proposition 5.4. *Suppose that* $T = PSP^{-1}$, *and further suppose that* S *is nonsingular. Then* T *is nonsingular.*

Proof: From the above $\det(T) = \det(S)$. If S is nonsingular, then $\det(T) = \det(S) \neq 0$, so that T is nonsingular. □

Proposition 5.5. *If* T *is similar to* S, *i.e.,* $T = PSP^{-1}$. *Then* T^k *is similar to* S^k *for any* $k \geq 1$. *Furthermore,* $T^k = PS^kP^{-1}$.

Proof: Homework problem. □

5.2.2. Diagonalizability

Definition 5.6. A homomorphism $h : V \to V$ is diagonalizable if there is a basis $\langle \vec{\beta}_1, \ldots, \vec{\beta}_n \rangle$ for V such that $h(\vec{\beta}_i) = \lambda_i \vec{\beta}_i$ for $i = 1, \ldots, n$. A diagonalizable matrix is one that is similar to a diagonal matrix.

Example. We have already seen that the following are similar;

$$A = \begin{pmatrix} 4 & -2 \\ 1 & 1 \end{pmatrix}, \quad D = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}.$$

Hence, A is diagonalizable.

Remark 5.7. Not every matrix is diagonalizable. If $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ is a diagonal matrix, then it is easy to check that $D^k = \mathrm{diag}(\lambda_1^k, \ldots, \lambda_n^k)$ for each $k \geq 1$. If A is similar to D, then we know that A^k is similar to D^k. Consider

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \Longrightarrow A^2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

A clearly cannot be similar to the zero matrix, for this would imply that $A = 0$. Since $\det(A) = 0$, $\det(D) = 0$, so that we can assume that $D = \text{diag}(\lambda, 0)$. If

$$P = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is invertible, it is easy to check that

$$PD^2P^{-1} = \frac{\lambda^2}{\det(P)} \begin{pmatrix} ad & -ab \\ cd & -bc \end{pmatrix}.$$

In order to get the zero matrix, we need $ad = bc = 0$, which would require that $\det(P) = 0$. Hence, no such P exists, so that A is not diagonalizable.

5.2.3. Eigenvalues and Eigenvectors

Definition 5.8. A homomorphism $h : V \to V$ has a scalar eigenvalue $\lambda \in \mathbb{C}$ if there is a nonzero eigenvector $\vec{\zeta} \in V$ such that $h(\vec{\zeta}) = \lambda\vec{\zeta}$.

Example. Consider the homomorphism $h : \mathscr{P}_1 \to \mathscr{P}_1$ given by

$$h(c_0 + c_1x) = (4c_0 - 2c_1) + (c_0 + c_1)x.$$

We wish to solve $h(p(x)) = \lambda p(x)$. With respect to the standard basis this can be rewritten as $H\vec{\zeta} = \lambda\vec{\zeta}$, where

$$H = \text{Rep}_{\mathscr{E}_2,\mathscr{E}_2}(h) = \begin{pmatrix} 4 & -2 \\ 1 & 1 \end{pmatrix}, \quad \vec{\zeta} = \text{Rep}_{\mathscr{E}_2}(p).$$

The system can again be rewritten as $(H - \lambda I)\vec{\zeta} = \vec{0}$. In other words, we need $\vec{\zeta} \in \mathcal{N}(H - \lambda I)$. In order for $\dim(\mathcal{N}(H - \lambda I)) \geq 1$, we need $\det(H - \lambda I) = 0$. Thus, to find the eigenvalues we must solve

$$\det(H - \lambda I) = \lambda^2 - 5\lambda + 6 = (\lambda - 2)(\lambda - 3) = 0.$$

For $\lambda = 2$ one has $\vec{\zeta} = (1,1)^T$, and for $\lambda = 3$ one has $\vec{\zeta} = (2,1)^T$. Thus, the eigenvalues for h are $\lambda = 2, 3$, and the associated eigenvectors are $p(x) = 1 + x, 2 + x$.

If $B = \langle 1 + x, 2 + x\rangle$, then

$$H = \text{Rep}_{B,B}(h) = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix},$$

i.e., $h(c_1(1 + x) + c_2(2 + x)) = 2c_1(1 + x) + 3c_2(2 + x)$.

Definition 5.9. If $H \in \mathbb{R}^{n\times n}$, then H has an eigenvalue λ with associated nonzero eigenvector $\vec{\zeta}$ if $H\vec{\zeta} = \lambda\vec{\zeta}$.

Definition 5.10. The eigenspace of the homomorphism h associated with the eigenvalue λ is

$$V_\lambda := \{\vec{\zeta} : h(\vec{\zeta}) = \lambda\vec{\zeta}\}.$$

Lemma 5.11. *The eigenspace V_λ is a subspace.*

Proof: Just check that $h(c_1\vec{\zeta}_1 + c_2\vec{\zeta}_2) = \lambda(c_1\vec{\zeta}_1 + c_2\vec{\zeta}_2)$, so that $\vec{\zeta}_1, \vec{\zeta}_2 \in V_\lambda$ implies that $c_1\vec{\zeta}_1 + c_2\vec{\zeta}_2 \in V_\lambda$. □

Definition 5.12. The characteristic polynomial of $H \in \mathbb{R}^{n\times n}$ is given by

$$p(x) := \det(H - xI).$$

The characteristic equation is $p(x) = 0$. The characteristic polynomial of a homomorphism h is the characteristic polynomial of any $\mathrm{Rep}_{B,B}(h)$.

Remark 5.13. $p(x)$ is a polynomial of degree n. As a consequence, there will be n eigenvalues (counting multiplicity).

Lemma 5.14. *The characteristic polynomial of h is independent of the basis B.*

Proof: If H_B and H_D are representations of h with respect to the bases B and D, there is a nonsingular P such that $H_B = PH_DP^{-1}$. Noting that

$$PH_DP^{-1} - xI = P(H_D - xI)P^{-1}$$

yields that

$$\det(H_B - xI) = \det(PH_DP^{-1} - xI) = \det(P)\det(H_D - xI)\det(P^{-1}) = \det(H_D - xI).$$ □

Lemma 5.15. *Let $H \in \mathbb{R}^{n\times n}$ be given, and let $\lambda_1, \dots, \lambda_\ell$ be distinct eigenvalues. If $\vec{\zeta}_j \in V_{\lambda_j}$ for $j = 1, \dots, \ell$, then the set $\{\vec{\zeta}_1, \dots, \vec{\zeta}_\ell\}$ is linearly independent.*

Proof: Proof by induction. The statement is clearly true if $\ell = 1$. Assume that it is true for some value $1 < k < \ell$, so that the set $\{\vec{\zeta}_1, \dots, \vec{\zeta}_k\}$ is linearly independent. Now consider

$$c_1\vec{\zeta}_1 + \cdots + c_k\vec{\zeta}_k + c_{k+1}\vec{\zeta}_{k+1} = \vec{0}.$$

Multiplying both sides by λ_{k+1} yields

$$c_1\lambda_{k+1}\vec{\zeta}_1 + \cdots + c_k\lambda_{k+1}\vec{\zeta}_k + c_{k+1}\lambda_{k+1}\vec{\zeta}_{k+1} = \vec{0},$$

and multiplying both sides by H and using the fact that $H\vec{\zeta}_j = \lambda_j\vec{\zeta}_j$ yields

$$c_1\lambda_1\vec{\zeta}_1 + \cdots + c_k\vec{\zeta}_k\vec{\zeta}_k + c_{k+1}\lambda_{k+1}\vec{\zeta}_{k+1} = \vec{0}.$$

Subtracting the first from the second gives

$$c_1(\lambda_1 - \lambda_{k+1})\vec{\zeta}_1 + \cdots + c_k(\lambda_k - \lambda_{k+1})\vec{\zeta}_k = \vec{0}.$$

Since the eigenvalues are distinct, this implies that $c_1 = \cdots = c_k = 0$. Since $\vec{\zeta}_{k+1} \neq \vec{0}$, this further implies that $c_{k+1} = 0$. Hence, the set $\{\vec{\zeta}_1, \dots, \vec{\zeta}_k, \vec{\zeta}_{k+1}\}$ is linearly independent. □

Corollary 5.16. *Let $H \in \mathbb{R}^{n\times n}$ be given, and let $\lambda_1, \dots, \lambda_n$ be distinct eigenvalues. The set $\{\vec{\zeta}_1, \dots, \vec{\zeta}_n\}$ is a basis.*

Corollary 5.17. *Suppose that $H \in \mathbb{R}^{n\times n}$ has n distinct eigenvalues. H is diagonalizable.*

Proof: Suppose that $H\vec{\zeta}_j = \lambda_j\vec{\zeta}_j$ for $j = 1, \dots, n$. As a consequence of the above lemma, the matrix $P = (\vec{\zeta}_1, \vec{\zeta}_2, \dots, \vec{\zeta}_n)$ is nonsingular. Setting $D = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$, it is easy to check that $HP = PD$; hence, $P^{-1}HP = D$. □

Remark 5.18. (a) The statement can be relaxed to say that if there are n linearly independent eigenvectors, then H is diagonalizable.

(b) The matrix P depends on the ordering of the eigenvalues.

Example. Suppose that

$$\begin{pmatrix} 4 & -2 \\ 1 & 1 \end{pmatrix}.$$

If

$$P = \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix},$$

then $D = \operatorname{diag}(2, 3)$, while if

$$P = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix},$$

then $D = \operatorname{diag}(3, 2)$

Topic: Stable Populations

Consider the example at the beginning of this chapter. Setting

$$\vec{\zeta}_1 = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \quad \vec{\zeta}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix},$$

we have that with the basis $B = \langle\vec{\zeta}_1, \vec{\zeta}_2\rangle$ the homomorphism $h(\vec{x}) = A\vec{x}$ is represented by the matrix $D = \operatorname{diag}(1, 3/4)$. Thus, with the basis B we have that dynamical system becomes $\vec{y}_{n+1} = D\vec{y}_n$, where $\vec{y} = \operatorname{Rep}_B(\vec{x})$. If $\vec{y}_0 = (a, b)^T$, the solution is

$$(\vec{y}_n)_1 = a, \quad (\vec{y}_n)_2 = \left(\frac{3}{4}\right)^n b,$$

so that as $n \to \infty$ we have that $\vec{y}_n \to (a, 0)^T$. For example, suppose that $\vec{x}_0 = (14, 36)^T = 10\vec{\zeta}_1 + 6\vec{\zeta}_2$, so that $\vec{y}_0 = (10, 6)^T$. We have that $\lim_{n\to\infty} \vec{y}_n = (10, 0)$, so that $\lim_{n\to\infty} \vec{x}_n = 10\vec{\zeta}_1 = (20, 30)^T$.

Remark 5.19. The fact that $\lambda = 1$ is an eigenvalue is *not* a coincidence. Problems 5.2.3.33 and 5.2.3.42 discuss this issue in more detail.

Topic: Method of Powers

For many applications it is important to find only the largest eigenvalue of the matrix $H \in \mathbb{R}^{n\times n}$. Why? As in the previous example, consider dynamical system $\vec{x}_{n+1} = H\vec{x}_n$. Suppose that H has n distinct eigenvalues $\lambda_1, \dots, \lambda_n$ with associated eigenvectors $\vec{\zeta}_1, \dots, \vec{\zeta}_n$. Since $\langle \vec{\zeta}_1, \dots, \vec{\zeta}_n \rangle$ is a basis, we can write $\vec{v} \sum_i c_i \vec{\zeta}_i$. Using linearity we see that

$$\begin{aligned} H\vec{v} &= \sum_i c_i \lambda_i \vec{\zeta}_i \\ H^2\vec{v} &= \sum_i c_i \lambda_i^2 \vec{\zeta}_i \\ &\vdots \\ H^\ell \vec{v} &= \sum_i c_i \lambda_i^\ell \vec{\zeta}_i. \end{aligned}$$

If one of the eigenvalues, say λ_1, is such that $|\lambda_k| < |\lambda_1|$ for $k = 2, \dots, n$, then as ℓ becomes large the dominant behavior will be $H^\ell \vec{v} \approx c_1 \lambda_1^\ell \vec{\zeta}_1$. Note that

$$\frac{H^\ell \vec{v}}{\lambda_1^\ell} = c_1 \vec{\zeta}_1 + \sum_{i=2}^n c_i \left(\frac{\lambda_i}{\lambda_1} \right)^\ell \vec{\zeta}_i,$$

so that

$$\lim_{\ell \to \infty} \frac{H^\ell \vec{v}}{\lambda_1^\ell} = c_1 \vec{\zeta}_1.$$

How do we determine the eigenvalue λ_1? If n is large, then the most efficient way to accomplish this task is *not* by finding all of the roots of the characteristic equation. Given a $\vec{v}_0$, consider the algorithm for $j = 0, 1, \dots,$

$$\vec{w}_j = \frac{\vec{v}_j}{\|\vec{v}_j\|}, \quad \vec{v}_{j+1} = H\vec{w}_j.$$

Note that $\|\vec{w}_j\| = 1$ for all j. If $\vec{w}_j$ is an eigenvector, then we will have that $\vec{v}_{j+1} = \lambda \vec{w}_j$ for the associated eigenvalue λ, so that $|\lambda| = \|\vec{v}_{j+1}\| / \|\vec{w}_j\| = \|\vec{v}_{j+1}\|$. From the above discussion we expect that as j gets large we will have that $|\lambda_1| \approx \|\vec{v}_{j+1}\|$. A stopping criterion could then be

$$\left| \frac{\|\vec{v}_{j+1}\| - \|\vec{v}_j\|}{\|\vec{v}_j\|} \right| < \epsilon, \quad 0 < \epsilon \ll 1.$$

Remark 5.20. If $|\lambda_2/\lambda_1|$ is sufficiently small, where λ_2 is the second largest eigenvalue in absolute value, then this algorithm will converge fairly quickly. Otherwise, there are other techniques that one could use.

Topic: Symmetric Matrices

The equations of motion for a coupled mass-spring system can be written in the form

$$M\vec{x}'' = K\vec{x},$$

where $M = \mathrm{diag}(m_1, \dots, m_n)$ and $K = K^T$, i.e., K is a symmetric matrix. Here $m_i > 0$ represents the mass on the end of spring i, $\vec{x}_i$ is the distance of mass i from the equilibrium position, and the entries of K are various combinations of the individual spring constants. For example, when considering a system of two masses with only one of the springs being connected to a wall, one has that

$$K = \begin{pmatrix} -(k_1 + k_2) & k_2 \\ k_2 & -k_2 \end{pmatrix},$$

where k_i is the spring constant for spring i.

In order to solve the ODE, set $\vec{x}(t) = \mathrm{e}^{\omega t}\vec{v}$. Upon substitution this yields the linear system

$$(K - \omega^2 M)\vec{v} = \vec{0},$$

i.e., we have a generalized eigenvalue problem. Note that if ω_0 is an eigenvalue with eigenvector $\vec{v}_0$, then $-\omega_0$ is also an eigenvalue with the same eigenvector. Further note that since K and M are real, then ω_0^* is an eigenvalue with eigenvector $\vec{v}_0^*$. Thus, eigenvalues come in the quartets $\{\omega, -\omega, \omega^*, -\omega^*\}$. If $\mathrm{Re}\,\omega > 0$, then the oscillations of the masses will grow exponentially, while if $\mathrm{Re}\,\omega = 0$ the motion will be oscillatory with frequency $\mathrm{Im}\,\omega$.

This problem has a great deal of structure. Some interesting questions are:

(a) What in general is known about the eigenvalue/eigenvector decomposition of symmetric matrices?

(b) How do the eigenvalues for the full problem relate to those of K and M?

Definition 5.21. Let $\vec{v} = \vec{a} + \mathrm{i}\vec{b} \in \mathbb{C}^n$ be given, where $\vec{a}, \vec{b} \in \mathbb{R}^n$. An inner product $(\cdot,\cdot)$ on $\mathbb{C}^n$ can be defined by

$$(\vec{v}, \vec{w}) = \vec{v}^T\vec{w}^*,$$

where $\vec{w}^* = \vec{c} - \mathrm{i}\vec{d}$.

Remark 5.22. (a) The inner product $(\cdot,\cdot)$ is the standard one for $\vec{v} \in \mathbb{R}^n$.

(b) Note that $(r\vec{v}, \vec{w}) = r(\vec{v}, \vec{w})$, while $(\vec{v}, r\vec{w}) = r^*(\vec{v}, \vec{w})$.

Proposition 5.23. *The inner product $(\cdot,\cdot)$ on $\mathbb{C}^n$ satisfies the properties:*

(a) $(\vec{v}, \vec{w}) = (\vec{w}, vv)^*$

(b) $(\vec{v}, \vec{v}) \geq 0$, *with equality only if* $\vec{v} = \vec{0}$

(c) $(r\vec{v} + s\vec{w}, \vec{x}) = r(\vec{v}, \vec{x}) + s(\vec{w}, \vec{x})$.

Lemma 5.24. *If $A \in \mathbb{R}^{n\times n}$ is symmetric, then all of the eigenvalues are real. Furthermore, the eigenvectors associated with distinct eigenvalues are orthogonal.*

Proof: It is not difficult to show that $(A\vec{v}, \vec{w}) = (\vec{v}, A^T\vec{w})$. Since A is symmetric, this implies that $(A\vec{v}, \vec{w}) = (\vec{v}, A\vec{w})$.

First suppose that $A\vec{v} = \lambda\vec{v}$. One then has that

$$(A\vec{v}, \vec{v}) = (\lambda\vec{v}, \vec{v}) = \lambda(\vec{v}, \vec{v}),$$

and

$$(A\vec{v}, \vec{v}) = (\vec{v}, A\vec{v}) = (\vec{v}, \lambda\vec{v}) = \lambda^*(\vec{v}, \vec{v}).$$

Since $\vec{v} \neq \vec{0}$, this implies that $\lambda = \lambda^*$, so that $\lambda \in \mathbb{R}$.

Suppose that $A\vec{v}_i = \lambda_i\vec{v}_i$ for $i = 1, 2$ with $\lambda_1 \neq \lambda_2$. One has that

$$(A\vec{v}_1, \vec{v}_2) = (\lambda_1\vec{v}_1, \vec{v}_2) = \lambda_1(\vec{v}_1, \vec{v}_2),$$

and

$$(A\vec{v}_1, \vec{v}_2) = (\vec{v}_1, A\vec{v}_2) = (\vec{v}_1, \lambda_2\vec{v}_2) = \lambda_2(\vec{v}_1, \vec{v}_2).$$

Since $\lambda_1 \neq \lambda_2$, this implies that $(\vec{v}_1, \vec{v}_2) = 0$. $\square$

Definition 5.25. $O \in \mathbb{R}^{n\times n}$ is an orthogonal matrix if its column vectors are an orthogonal with length one, i.e., they form an orthonormal set.

Proposition 5.26. *Suppose that O is an orthogonal matrix. Then O is nonsingular, and $O^{-1} = O^T$.*

Proof: The fact that O is nonsingular follows immediately from the fact that the column vectors are linearly independent. If $O = (\vec{o}_1, \dots, \vec{o}_n)$, then upon using the fact that

$$\vec{o}_j^{\,T} \vec{o}_i = \begin{cases} 0, & j \neq i \\ 1, & j = i \end{cases}$$

it is easy to check that $O^T O = I$. □

Proposition 5.27. *Suppose that O is an orthogonal matrix. Then $(O\vec{x}, O\vec{y}) = (\vec{x}, \vec{y})$.*

Proof: We have that

$$(O\vec{x}, O\vec{y}) = (\vec{x}, O^T O\vec{y}) = (\vec{x}, \vec{y}).$$ □

Remark 5.28. As a consequence, multiplication by an orthogonal matrix preserves both the angle between vectors and the length of vectors.

Theorem 5.29 (Schur's Theorem)**.** *If $A \in \mathbb{R}^{n\times n}$ has only real eigenvalues, then there is an orthogonal matrix such that $O^T AO$ is upper triangular.*

Theorem 5.30. *If $A \in \mathbb{R}^{n\times n}$ is symmetric, then there is an orthogonal matrix which diagonalizes A.*

Proof: Recall that $(AB)^T = B^T A^T$. By Schur's Theorem there is an orthogonal matrix O and upper triangular matrix T such that $O^T AO = T$. Since A is symmetric,

$$T^T = (O^T AO)^T = O^T A^T O = O^T AO,$$

so that $T^T = T$, i.e., T is symmetric. Since T is upper triangular, T^T is lower triangular. Since T is symmetric, this implies that T is a diagonal matrix. □

Example. Consider

$$A = \begin{pmatrix} 0 & 2 & -1 \\ 2 & 3 & -2 \\ -1 & -2 & 0 \end{pmatrix}.$$

The eigenvalues are $\lambda = -1, 5$. We have that

$$\mathcal{N}(A + I) = [\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}], \quad \mathcal{N}(A - 5I) = [\begin{pmatrix} -1 \\ -2 \\ 1 \end{pmatrix}].$$

Using the Gram-Schmidt procedure, an orthogonal basis for $\mathcal{N}(A+I)$ is given by

$$\mathcal{N}(A+I) = [\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}].$$

Normalizing all these vectors to length one gives the matrix O such that $O^T AO = \text{diag}(5, -1, -1)$.

Again consider the eigenvalue problem $(K - \omega^2 M)\vec{v} = \vec{0}$. First suppose that $M = I$. There is then an orthogonal matrix O_K such that $O_K^T K O_K = D_K$, where $D_K = \text{diag}(\lambda_1, \dots, \lambda_n)$ with each $\lambda_i \in \mathbb{R}$. We have that $\omega^2 = \lambda$. If $\lambda > 0$, then there will be exponential growth for the system, while if $\lambda < 0$ the motion will be oscillatory with frequency ω. If one sets $\vec{u} = O_K^T \vec{x}$, so that $\vec{x} = O_K \vec{u}$, then the ODE uncouples and becomes

$$\vec{u}'' = O_K^T K O_K \vec{u} = D_K \vec{u},$$

i.e., $\vec{u}_i'' = \lambda_i \vec{u}_i$. Now suppose that M is merely diagonal with $m_i > 0$. Let $\text{n}(A, B)$ represents the number of negative eigenvalues for the eigenvalue problem $(A - \lambda B)\vec{v} = \vec{0}$. matrix A. Because all of the eigenvalues of M are positive, it can be shown that $\text{n}(K, M) = \text{n}(K, I)$. Thus, if oscillatory motion is predicted for the equal mass case, then this feature will continue to hold true as the masses are varied.