Lecture Notes: Linear Algebra I.
Fall 2014

by Prof. Frederick P. Greenleaf
NYU/Courant Institute of Math Sciences

Prepared with the gratefully acknowledged assistance of

Dr. Sophie Marques

without whose efforts the Latex version of these Course Notes might never have seen the light of day.

Contents

CH I. Vector Spaces over a Field.

1. Basic definitions and examples 1
2. Vector subspaces 5
   Linear systems of equations 6
3. Case Study: Determining the linear span 9
   Further remarks on elementary row operations 13
4. Linear span, independence, and bases 15
   Describing subspaces 18
5. Quotient spaces V/W 22
   Algebraic structure in V/W 23

CH II. Linear Operators.

1. Definitions and general facts 28
2. Invariant subspaces 33
   Decomposition of operators 34
3. Direct sums of vector spaces 37
   Projection operators and direct sums 39
   Direct sums and eigenspaces 42
4. Representing linear operators as matrices 44
   Change of basis and similarity transformations 49
   RST equivalence relations 52

CH III. Dual Spaces and Duality.

1. Definitions and Examples 55
2. Dual bases in V* 58
3. The Transpose Operation 61
   Reflexivity of finite dimensional spaces 64

CH IV. Determinants.

1. The permutation group S_n 68
2. Determinants 73
   Parity of a permutation 71
   Row operations, determinants, and inverses 76
   Computing inverses 78
   Determinants: Matrices vs linear operators 82
   Expansion by minors and Cramer's Rule 84

CH V. The Diagonalization Problem.

1. The Characteristic Polynomial 85
   Factoring polynomials 86
2. Finding Eigenvalues 90

3. Diagonalization and Limits of operators 96
The Cauchy convergence criterion 95
Computing the exponential e^A of a matrix 98
Solving systems of linear ODE 99

CH VI. Inner Product Spaces.

1. Basic definitions and examples 106
Triangle Inequality 109
Hilbert Schmidt norm and L^2 convegrence
2. Orthonormal bases in inner product spaces 111
Bessel's inequality 112
3. Orthonormal complements and projections 113
Orthogonal projections on inner product spaces 114
The Gram-Schmidt construction 115
Fourier series expansions 118
A geometry problem 121
4. Adjoints and orthogonal decompositions 122
Dual spaces of inner product spaces 123
The adjoint operator T^* 124
Normal and self-adjoint operators 127
5. Diagonalization in inner product spaces 128
Orthogonal diagonalization: Schur Normal Form 129
Diagonalizing self-adjoint operators 131
Diagonalizing normal operators 132
Unitary equivalence of operators 134
Unitary operators vs unitary matrices 137
Change of orthonormal bases 138
6. Reflections, rotations and rigid motions on real inner product spaces 140
Rigid motions on $\mathbb{R}^n$ 140
Reflections in inner product spaces 141
Euler's Theorem: rotations in $\mathbb{R}^3$ 143
7. Spectral Theorem for vector and inner product spaces.
The Spectral Theorem 145
Functions of operators: e^T revisited 147
8. Positive operators and the Polar Decomposition 150
Positive square roots 151
The polar decomposition (Invertible T) 152
Polar decomposition for arbitrary T 155
The Singular Value decomposition 156

Chapter 0. A Few Preliminaries.

Course texts:

1. *Linear Algebra*, by S. Friedberg, A. Insel, L. Spence (latest edition). This will be the main backup text to accompany the present *Class Notes* . Various assignments will be taken from it.

2. Schaum's Outline Series: *Linear Algebra*, by Seymour Lipschutz, for a review of matrix algebra, row operations, and solution of linear systems (roughly the first 3-4 chapters). $\mathbb{K}$ is a field (see Appendix *C* of [F/I/S] text; read it). For us, $\mathbb{K} = \mathbb{C}, \mathbb{R}, \mathbb{Q}$, and occasionally the finite field $\mathbb{K} = \mathbb{Z}_p = \mathbb{Z}/p\mathbb{Z}$, for a prime $p > 1$.

Recall that the finite field $\mathbb{Z}_p$ is modeled as $S = \{0, 1, 2, ..., p-1\}$, interpreting $a + b$ and ab (mod p). For example: if $p = 7$ then

$$5 \oplus 6 = 11 \equiv 4 \pmod 7 \qquad \text{and} \qquad 506 \odot 17 \equiv 6 \pmod 7.$$

Elements of $\mathbb{Z}_p$ are the (mod p) congruence classes $[k] = k + p\mathbb{Z} = \{\ell : \ell \equiv k \pmod p\}$. Using this notation the operations in $\mathbb{Z}_p$ take the form

$$[a] \oplus [b] = [a + b] \qquad [a] \odot [b] = [ab] \qquad \text{(add or multiply class representatives)}$$

The system $(\mathbb{Z}_p, \oplus, \odot)$ is a finite number field with additive zero element [0] and multiplicative identity element [1]. All nonzero elements $[k] \neq [0]$ have multiplicative inverses (reciprocals), but it may not be so easy to find the class $[k]^{-1} = [\ell]$, $0 < \ell < p$ such that $[k] \cdot [\ell] = [1]$. If $p = 7$ we have $[3]^{-1} = [5]$ because $3 \odot 5 = 15 = 14 + 1 \equiv 1 \pmod 7$. Notice that in $\mathbb{Z}_p$ the sum $[1] \oplus [1] \oplus .... \oplus [1]$ with p terms is equal to the zero element [0].

Chapter I

Section I.1. Vector Spaces over a Field $\mathbb{K}$

The objects of interest in this chapter will be vector spaces over arbitrary fields.

1.1 Definition. *A* **vector space over a field** $\mathbb{K}$ *is a set* V *equipped with two operations* $(+)$ *and* $(\cdot)$ *from* $V \times V \to V$ *and* $\mathbb{K} \times V \to V$ *having the following properties.*

1. **Axioms for** $(+)$**:**
   COMMUTATIVE LAW: $x + y = y + x$
   ASSOCIATIVE LAW: $(x + y) + z = x + (y + z)$
   ZERO ELEMENT: There exists an element "0" in V such that $0 + v = v$ for all v.
   ADDITIVE INVERSE: For every $v \in V$ there is an element $-v \in V$, such that $v + (-v) = 0$.
2. **Axioms for** $(\cdot)$**:**
   IDENTITY LAW: $1 \cdot v = v$ ($1 =$ the identity in $\mathbb{K}$)
   ASSOCIATIVE LAW: $(ab) \cdot v = a \cdot (b \cdot v)$ for $a, b \in \mathbb{K},\ v \in V$
   DISTRIBUTIVE LAW: $a \cdot (x + y) = (a \cdot x) + (b \cdot y)$
   DISTRIBUTIVE LAW: $(a + b) \cdot x = (a \cdot x) + (b \cdot x)$

As a consequence,

1.2. Lemma. *The zero element is unique: if* $0, 0' \in V$ *are elements such that* $0 + v = v$ *and* $0' + v = v$*, for all* $v \in V$*, then* $0' = 0$*.*

Proof: $0 + 0' = 0'$ and $0 + 0' = 0$, so $0' = 0$. □

1.3. Lemma. *The additive inverse is unique. That is, given* $v \in V$ *there is just one element* $u \in V$ *such that* $u + v = 0$*.*

Proof: Suppose $v \in V$ and we are given u and u' with $u + v = 0$ and $u' + v = 0$. Look at the combination $u + v + u'$; by associativity we get

$$u' = 0 + u' = (u + v) + u' = u + (v + u') = u + 0 = u \quad ,$$

so that $u' = u$. □

1.4. Exercise. From the axioms and previous results prove:

(i) $0 \cdot v = 0_V$ (ii) $\lambda \cdot 0_V = 0_V$ (iii) $\lambda \cdot v = v$ and $v \neq 0_V \Rightarrow \lambda = 1$.

1.5. Exercise. Prove that $-v = (-1) \cdot v$ where -1 is the negative of $1 \in \mathbb{K}$.
Hint: $1 + (-1) = 0$ in $\mathbb{K}$ and $0 \cdot v = 0_V$. Remember: "$-v$" is the unique element that added to v is 0_V; prove that $(-1) \cdot v$ has this property and conclude by uniqueness of additive inverse.

1.6. Exercise. Prove that $-(-v) = v$, for all $v \in V$.
Hint: Same as the previous exercise.

1.7. Exercise (Cancellation Laws). If $a + v = a + w$ for $a, v, w \in V$ prove that $v = w$. Then use this to prove

(i) $\lambda \cdot v = 0_V$ and $v \neq 0_V$ implies that $\lambda = 0$ in $\mathbb{K}$

(ii) $\lambda \cdot v = v$ and $v \neq 0_V$ implies $\lambda = 1$.

1.8. Example. "*Coordinate space*" over the field $\mathbb{K}$ consists of all ordered n-tuples $\mathbb{K}^n = \{\mathbf{x} = (x_1, ..., x_n) : x_k \in \mathbb{K}\}$, equipped with the usual $(+)$ and $(\cdot)$ operations:

(i) $(x_1, ..., x_n) + (y_1, ..., y_n) = (x_1 + y_1, ..., x_n + y_n)$

(ii) $\lambda \cdot (x_1, ..., x_n) = (\lambda \cdot x_n, ..., \lambda \cdot x_n)$ for $\lambda \in \mathbb{K}$. □

1.9. Exercise. Explain why (+) in $\mathbb{R}^2$ is described geometrically by the "parallelogram law" for vector addition shown in Figure 1.1.

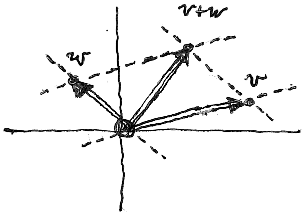

Figure 1.1. The Parallelogram Law for vector addition, illustrated in $\mathbb{R}^2$.

1.10. Example (Matrix Space). The space $\mathrm{M}(n \times m, \mathbb{K})$ of $n \times m$ matrices with entries in $\mathbb{K}$ becomes a vector space when equipped with the operations

$$\text{Addition Operation:} \quad (A + B)_{ij} = A_{ij} + B_{ij}$$

$$\text{Scaling Operation:} \quad (\lambda \cdot A)_{ij} = \lambda A_{ij}$$

The space of square matrices, with $m = n$, is denoted $\mathrm{M}(n, \mathbb{K})$.
Notation: Matrix entry A_{ij} is the one in the i^{th} row and j^{th} column. The pair (i, j) is referred to as its "address." □

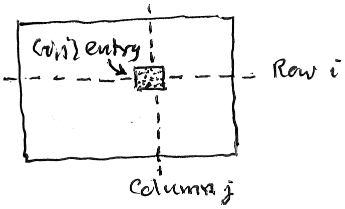

Figure 1.2. The entry in a matrix array with "address" (i, j) is the one in Row i and Column j.

There is also a matrix multiplication that makes $\mathrm{M}(n, \mathbb{K})$ an associative algebra with identity, but the matrix product AB can be defined more generally for non-square matrices as long as they are "compatible," with the number of columns in A equal to the number of rows in B. Thus if A is $m \times q$ and B is $q \times n$ we get an $m \times n$ matrix AB with entries

$$(AB)_{ij} = \sum_{k=1}^{q} A_{ik} B_{kj}$$

The algebra $\mathrm{M}(n, \mathbb{K})$ of square matrices is not commutative unless $n = 1$. □

1.11. Example (Polynomial Ring $\mathbb{K}[x]$). The set $\mathbb{K}[x]$ consists of all finite "formal sums" $a_0 + a_1 x + ... + a_n x^n + ... = \sum_{k \geq 0} a_k x^k$ with $a_i \in \mathbb{K}$, and $a_i = 0$ for all but a finite number of indices. These sums can have arbitrary length. They include the "constant polynomials" which have form $c \cdot \mathbb{1}$ with $c \in \mathbb{K}$, where $\mathbb{1}$ is the particular constant polynomial $1 + 0{\cdot}x + 0{\cdot}x^2 + \ldots$; the zero polynomial $0{\cdot}\mathbb{1}$ is written as "0", which might get confusing.

The algebraic operations in $\mathbb{K}[x]$ are

1. Addition: $(\sum_{k\geq 0} a_k x^k) + (\sum_{k\geq 0} b_k x^k) \;=\; \sum_{k\geq 0}(a_k + b_k)x^k$

2. Scaling: $\lambda \cdot (\sum_{k\geq 0} a_k x^k) \;=\; \sum_{k\geq 0}(\lambda a_k)x^k$.

There is also a multiplication operation, obtained by multiplying terms in the formal sums and gathering together those of the same degree

3. Product: $\Big(\sum_{k\geq 0} a_k x^k\Big) \times \Big(\sum_{l\geq 0} b_l x^l\Big) = \sum_{k,l\geq 0} a_k b_l\, x^{k+l} = \sum_{r\geq 0}\Big(\sum_{k,l\geq 0, k+l=r} a_k b_l\Big)\cdot x^r$

(the sum being finite for each r). This makes $\mathbb{K}[x]$ into a commutative associative algebra over $\mathbb{K}$ with $\mathbb{1}$ as its multiplicative identity.

All information about a polynomial resides in the symbol string $(a_0, a_1, a_2, ...)$ of coefficients, and the algebraic operations on $\mathbb{K}[x]$ can be performed as operations on symbol strings; the zero polynomial is represented by $(0, 0, ...)$, the identity by $\mathbb{1} = (1, 0, ..., 0)$, and x by $x = (0, 1, 0, ...)$, etc. $\square$

1.12. Exercise. If $f(x) = 3 + 3x + x^2$ and $g(x) = 4x^2 - 2x^3 + x^5$, compute the sum $f + g$ and product $f \cdot g$.

The **degree** $\deg(f)$ of $f = \sum_{k\geq 0} a_k x^k$ is n if $a_n \neq 0$ and $a_k = 0$ for all $k > n$. The degree of a constant polynomial $c\mathbb{1}$ is zero, except that no "degree" can be assigned to the zero polynomial 0. (For various reasons, the only possible assignment would be "$-\infty$").

1.13. Exercise. If $f, g \neq 0$ in $\mathbb{K}[x]$ prove that $fg \neq 0$ and $\deg(fg) = \deg(f) + \deg(g)$.

1.14. Exercise. If $f, g \neq 0$ in $\mathbb{K}[x]$, what (if anything) can you say about $\deg(f + g)$?

1.15. Example (Polynomials in Several Unknowns). The polynomial ring $\mathbb{K}[\mathbf{x}] = \mathbb{K}[x_1, ..., x_n]$ is handled using very efficient "multi-index notation." A **multi-index** is an element $\alpha = (\alpha_1, \ldots \alpha_n)$ of the Cartesian product set $\mathbb{Z}_+^n = \mathbb{Z}_+ \times ... \times \mathbb{Z}_+$ (n factors). Each multi-index determines a **monomial** $x^\alpha = x_1^{\alpha_1} \cdot \ldots \cdot x_n^{\alpha_n}$, in which we interpret $x_k^0 = 1$. Elements of $\mathbb{K}[x_1, \ldots, x_n]$ are finite formal linear combinations of monomials

$$f(x_1, \ldots, x_n) = \sum_{\alpha\in\mathbb{Z}_+^n} c_\alpha x^\alpha \qquad (c_\alpha \in \mathbb{K})$$

The monomial $x^{(0,\ldots,0)}$ is the constant polynomial $\mathbb{1}$ in $\mathbb{K}[x_1, ..., x_n]$. With these ideas in mind,

1. The **total degree** of a multi-index is $|\alpha| = \alpha_1 + \ldots + \alpha_n$ and the degree of the corresponding monomial is $\deg(x^\alpha) = |\alpha|$. Note that many monomials can have same total degree, for example x^2y and xy^2.

2. The **degree of a polynomial** $f \in \mathbb{K}[\mathbf{x}]$ is

$$\deg(f) = \max\{\, |\alpha| : c_\alpha \neq 0 \,\}$$

Nonzero constant polynomials $c\mathbb{1}$ have degree zero: if f is the zero polynomial (all coefficients $c_\alpha = 0$) $\deg(f)$ cannot be defined. The generators $f_k(\mathbf{x}) = x_k$ of the polynomial ring all have degree 1.

The following operations make $V = \mathbb{K}[\mathbf{x}]$ a vector space and a commutative associative algebra with identity $\mathbb{1} = x^{(0,\ldots,0)}$.

1. Sum: $\;f + g = \sum_\alpha (a_\alpha + b_\alpha)\, x^\alpha$

2. Scaling: $\lambda \cdot f = \sum_\alpha (\lambda a_\alpha) x^\alpha$

3. Product Operation:

$$\begin{aligned} f\cdot g &= (\sum_{\alpha} a_\alpha x^\alpha)\cdot(\sum_{\beta} b_\beta x^\beta) \\ &= \sum_{\alpha,\beta\in\mathbb{Z}^n_+} a_\alpha b_\beta x^{\alpha+\beta} \\ &= \sum_{\gamma\in\mathbb{Z}^n_+}\Big(\sum_{\alpha+\beta=\gamma} a_\alpha\cdot b_\beta\Big)\cdot x^\gamma \end{aligned}$$

where we define a "sum of exponents" to be $\alpha+\beta=(\alpha_1+\beta_1,...,\alpha_n+\beta_n)$.

As an example, the monomials of degree 2 in $\mathbb{K}[x_1,x_2,x_3]$ are

multi-index	monomial
$(0,0,2)$	x_3^2
$(0,1,1)$	x_2x_3
$(0,2,0)$	x_2^2
$(1,0,1)$	x_1x_3
$(1,1,0)$	x_1x_2
$(2,0,0)$	x_1^2

Here we have lined up the monomials in "lexicographic"or "dictionary" order (taking $A=0,B=1,C=2,\ldots$), which is a useful way to manage them. This is a strict *linear* ordering of monomials; they are only partially ordered by their "total degree" $\deg(x^\alpha)=|\alpha|$. The system $\mathbb{K}[x_1,\ldots,x_n]$ is a commutative associative algebra with identity element $\mathbb{1}$. Its properties are quite a bit more complicated than those of polynomials $\mathbb{K}[x]$ in one unknown, but they do share two crucial algebraic properties. □

1.16. Exercise. (Hard, but try it) If $f,g\neq 0$ in $\mathbb{K}[x_1,...,x_n]$ prove that

1. Degree Formula: $\deg(f\cdot g)=\deg(f)+\deg(g)$ for all $f,g\neq 0$ in $\mathbb{K}[x_i,\ldots,x_n]$.
2. No Zero Divisors: $f,g\neq 0$ in $\mathbb{K}[x_1,\ldots,x_n]$ $\Rightarrow$ $f\cdot g\neq 0$. This implies we can perform "cancellation" – if $f\neq 0$ and $f\cdot h_1=f\cdot h_2$ then $h_1=h_2$.

Hint: Try it first for $n=1$. For $n=2$ try lexicographic ordering of monomials in $\mathbb{K}[x,y]$.
Note: The maximum possible degree for a nonzero monomial in the product fg is obviously $d=\deg(f)+\deg(g)$. The problem is that the coefficient c_γ of such a monomial will be a sum of products $(\sum_{\alpha+\beta=\gamma} a_\alpha b_\beta)$, and not a simple product as it is when there is just one variable. Such sums could equal zero even if all terms are nonzero, so why couldn't these coefficients (sums) be zero for *all* monomials with the maximum possible degree d, making $\deg(fg)<\deg(f)+\deg(g)$? □

A more complete discussion of the Degree Formula for $n\geq 2$, and especially its proof using lexicographic ordering of monomials, is provided in Appendix A of this chapter.

1.17. Example (Function Spaces). If S is a set, $\mathcal{C}(S)=$ all scalar-valued functions $f:S\to\mathbb{K}$ become a vector space under the usual operations

$$(f+g)(x)=f(x)+g(x), \qquad (\lambda\cdot f)(x)=\lambda f(x),\ \forall x\in S$$

There is also a pointwise multiplication operation

$$(f\cdot g)(x)=f(x)\cdot g(x)\ ,$$

which makes $\mathcal{C}(S)$ a commutative associative algebra over $\mathbb{K}$ with identity element $\mathbb{1}(x) = 1$ for all x, and zero element $0(x) = 0$, for $x \in S$. $\square$

1.18. Example (Polynomial Functions vs Formal Sums). The **polynomial functions** $\mathcal{P}_{\mathbb{K}}$ with values in $\mathbb{K}$ are the functions $\phi_f : \mathbb{K} \to \mathbb{K}$ of the form

$$\phi_f(t) = \left[\, f(x)|_{x=t} \,\right] = \sum_{k \geq 0} a_k\, t^k \qquad (t \in \mathbb{K})$$

for some $f \in \mathbb{K}[x]$. (Thus, $\phi(t) = sin(t)$ is not a polynomial function on $\mathbb{K}$). Note carefully that the elements of $\mathcal{P}_{\mathbb{K}}$ are functions while $\mathbb{K}[x]$ is made up of symbol strings or formal sums. They are not the same thing, though there is a close relation between them implemented by the surjective (="onto") mapping $\Phi : \mathbb{K}[x] \to \mathcal{P}_K$ such that

$$\Phi f(t) = \sum_{k \geq 0} a_k\, t^k \qquad (t \in \mathbb{K})$$

if $f(x) = \sum_{k \geq 0} a_k x^k$ in $\mathbb{K}[x]$. This surjective map is a *homomorphism*: it preserves, or "intertwines," the algebraic operations in $\mathbb{K}[x]$ and in the "target space" $\mathcal{P}_{\mathbb{K}}$, so that

$$\Phi(\lambda \cdot f) = \lambda \cdot \Phi(f) \qquad \Phi(f+g) = \Phi(f) + \Phi(g) \qquad \Phi(f \cdot g) = \Phi(f) \cdot \Phi(g) \quad \square$$

1.19. Exercise. If $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$ explain why Φ is a bijection, hence an "isomorphism" between commutative associative algebras. In fact, prove that this is so for polynomials over any *infinite* field $\mathbb{K}$.
Hint: Φ is linear, hence being one-to-one is equivalent to saying that $\Phi(f) = 0 \Rightarrow f = 0$ in $\mathbb{K}[x]$. If f is nonzero in $\mathbb{K}[x]$ the corresponding polynomial function $\Phi(f) : \mathbb{K} \to \mathbb{K}$ can take on the value zero at no more than $n = \deg(f)$ points – i.e. the number of roots in $\mathbb{K}$ cannot exceed $\deg(f)$. Since $\mathbb{R}$ and $\mathbb{C}$ (and even $\mathbb{Q}$) are infinite we cannot have $\Phi(f) \equiv 0$ on these fields unless f is the zero polynomial. $\square$

The finite fields $\mathbb{Z}_p$ (p a prime) are widely used in number theory, cryptography, image processing, etc. This one-to-one correspondence breaks down for these fields. For example if $\mathbb{K} = \mathbb{Z}_p$ the nonzero polynomial $f = x^p - x$ has value zero for every choice of $x \in \mathbb{Z}_p$ and there are precisely $p = \deg(f)$ roots.

A theorem of Fermat says: if p is a prime then $t^{p-1} = 1$ for all nonzero t in $\mathbb{Z}_p$, but then $t^p - t = t$ is zero at every $t \in \mathbb{Z}_p$ and $\Phi(f) \equiv 0$ (the zero function in $\mathcal{P}_{\mathbb{K}}$).

1.20. Exercise. For $p = 3$, verify that $t^3 - t = 0$ for the three elements $t = [0], [1], [2]$ in $\mathbb{Z}_3$. But the corresponding element of $\mathbb{Z}_3[x]$ is $f = x^3 - x$, whose symbol string $(0, -1, 0, 1, 0, 0, ...)$ differs from that of the zero polynomial in $\mathbb{Z}_3[x]$.

I.2. Vector Subspaces

2.1. Definition. *A nonempty subset W of a vector space V is a* **vector subspace** *if*

1. W is *closed under* $(+)$: $W + W \subseteq W$, so $w_1, w_2 \in W \Rightarrow w_1 + w_2 \in W$.

2. W is *closed under* $(\cdot)$: $\mathbb{K} \cdot W \subseteq W$, so $\lambda \in \mathbb{K}, w \in W \Rightarrow \lambda \cdot w \in W$.

The vector 0 then lies in W, for if $w \in W$ then $-w = (-1) \cdot w$ is also in W and then $0 = w + (-w) \in W$. Thus W becomes a vector space over $\mathbb{K}$ in its own right under the $(+)$ and $(\cdot)$ operations applied to elements of W.

Subspaces of V include the trivial examples $W = (0)$ and $W = V$; all others are "proper" subspaces of V.

2.2. Definition. *Given a non empty set S of vectors in V, its* **linear span** $\langle S \rangle =$

$\mathbb{K}$-span(S) *is the smallest subspace* $W \subseteq V$ *such that* W *contains* S.

It is easy to verify that:

2.3. Exercise. If $\{W_\alpha : \alpha \in I\}$ is any family of subspaces in V, prove that their intersection $W = \bigcap_{\alpha \in I} W_\alpha$ is also a subspace.

Thus Definition 2.2 makes sense: Given S there is at least one subspace containing S, namely V. If $E =$ intersection of all subspace Wthat contain S, then E is a subspace and is obviously the smallest subspace containing S. Thus $\mathbb{K}$-span(S) exists, even if V is "infinite dimensional," for instance $V = \mathbb{K}[x]$.

This "top down" definition has its uses, but an equivalent "bottom-up" version is often more informative.

2.4. Lemma. If $S \neq \emptyset$ in V, its linear span $\mathbb{K}$-span(S) is the set of finite sums

$$\left\{ \sum_{i=1}^{n} a_i v_i : a_i \in \mathbb{K}, v_i \in S, n < \infty \right\}$$

Proof: Let $E = \{\sum_{i=1}^{N} c_i v_i : N < \infty,\ c_i \in \mathbb{K},\ v_i \in S\}$. Since $S \subseteq \mathbb{K}$-span(S), every finite sum lies this span, proving $E \subseteq \mathbb{K}$-span(S). For $(\supseteq)$, it is clear that the family E of finite linear combination is closed under $(+)$ and $(\cdot)$ operations because a linear combination of linear combinations is just one big linear combination of elements of S. It is a subspace of V, and contains S because $1 \cdot s = s$ is a (trivial) linear combination. On the other hand every subspace $W \supseteq S$ must contain all these linear sums, so $S \subseteq E \subseteq W$. Hences E is the smallest subspace containing S and $E = K$-span(S). $\square$

2.5. Exercise. If $K = \mathbb{R}$, $V = \mathbb{R}^3$ show that $W = \{x \in \mathbb{R}^3 : 3x_1 + 2x_2 - x_3 = 0\}$ is a subspace and $W' = \{x \in \mathbb{R}^3 : 3x_1 + 2x_2 - x_3 = 1\}$ is not a subspace.
Hint: For one thing the zero vector $0 = (0, 0, 0)$ is not in W'. The situation is shown in Figure 1.3.

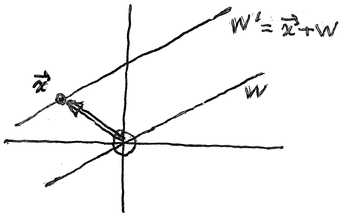

Figure 1.3. The subspace W in Exercise 2.5 and a translate $W' = \mathbf{x}_0 + W$ by some $\mathbf{x}_0 \in V$ such that $3x_1^0 + 2x_2^0 - x_3^0 = 1$, for instance $\mathbf{x}_0 = (0, 1, 1)$. The set W' is not a subspace.

System of Linear Equations.

Systems of n linear equations in m unknowns are of two general types

Homogeneous

$$\begin{cases} a_{11}x_1 + \ldots + a_{1m}x_m = 0 \\ \qquad\qquad \vdots \\ a_{n1}x_1 + \ldots + a_{nm}x_m = 0 \end{cases}$$

Inhomogeneous

$$\left\{\begin{array}{c} a_{11}x_1 + ... + a_{1m}x_m = b_1 \\ \vdots \\ a_{n1}x_1 + ... + a_{nm}x_m = b_n \end{array}\right.$$

with a_{ij} and b_k in $\mathbb{K}$. $\square$

2.6. Exercise. Verify that the solutions $\mathbf{x} = (x_1, \ldots, x_m)$ of the homogeneous system form a vector subspace of $\mathbb{K}^m$. Explain why the solution set of an inhomogeneous system cannot be a vector subspace unless $\mathbf{b} = (b_1, \ldots, b_n) = \mathbf{0}$ in $\mathbb{K}^n$.

If we regard vectors $\mathbf{x} = (x_1, ..., x_m) \in \mathbb{K}^m$ as the entries in an $m \times 1$ column matrix,

$$\mathbf{x} = \mathrm{col}(x_1, \ldots, x_m) = \left(\begin{array}{c} x_1 \\ \vdots \\ x_m \end{array}\right) \quad ,$$

you will recognize that the solutions $\mathbf{x} \in \mathbb{K}^m$ of the homogeneous system of equations are precisely the solutions of the matrix equations

$$A\mathbf{x} = \mathbf{0} \qquad \text{where the zero vector is} \qquad \mathbf{0} = \left(\begin{array}{c} 0 \\ \vdots \\ 0 \end{array}\right)_{n\times 1}$$

and for inhomogeneous systems we must solve

$$A\mathbf{x} = B \qquad \text{where} \qquad B = \left(\begin{array}{c} b_1 \\ \vdots \\ b_n \end{array}\right)_{n\times 1}$$

for $B \in \mathbb{K}^n$.

The homogeneous system always has the zero vector $\mathbf{0} \in \mathbb{K}^m$ as a solution, and the solution set $\{\mathbf{x} \in \mathbb{K}^m : A\mathbf{x} = \mathbf{0}\}$ is a vector subspace in $\mathbb{K}^m$. If $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$ then the number of solutions is either 1 or ∞ for this system. An inhomogeneous system might not have any solutions at all; otherwise, it has just one solution or infinitely many.

If A is an $n \times m$ matrix with entries in $\mathbb{K}$ we will find it useful to let A act by left multiplication as an operator $L_A : \mathbb{K}^m \rightarrow \mathbb{K}^n$ on column vectors

$$\mathbf{y} = L_A(\mathbf{x}) = A \cdot \mathbf{x} \qquad \big(\text{an } (n \times m)\cdot(m \times 1) \text{ matrix product}\big)$$

for $\mathbf{x} \in \mathbb{K}^m$. This is a *linear operator* in the sense that

$$L_A(\mathbf{x} + \mathbf{y}) = L_A(\mathbf{x}) + L_A(\mathbf{y}) \qquad \text{and} \qquad L_A(\lambda \cdot \mathbf{x}) = \lambda \cdot L_A(\mathbf{x})$$

for $\mathbf{x}, \mathbf{y} \in \mathbb{K}^m$ and $\lambda \in \mathbb{K}$. Solving a system of linear equations is then equivalent to finding solutions of $L_A(\mathbf{x}) = 0$ or $L_A(\mathbf{x}) = B$ for $\mathbf{x} \in \mathbb{K}^m$. From this point of view, $A\mathbf{x} = B$ has solutions if and only if B lies in the range $R(L_A) = \{A\mathbf{x} : \mathbf{x} \in \mathbb{K}^m\}$ (a vector subspace in $\mathbb{K}^n$). If $B = \mathbf{0}$ the "homogeneous" equation $A\mathbf{x} = \mathbf{0}$ always has the trivial solution $\mathbf{x} = \mathbf{0}$. $\square$

2.7. Exercise. If A is an $n \times m$ matrix and $L_A; \mathbb{K}^m \rightarrow \mathbb{K}^n$ is defined as above, verify that

1. The *range* $R(L_A) = L_A(\mathbb{K}^m) = \{A\cdot\mathbf{x} : x \in \mathbb{K}^m\}$ is a vector subspace in $\mathbb{K}^n$.

2. The *kernel* $K(L_A) = \ker(L_A) = \{x \in \mathbb{K}^m : L_A(\mathbf{x}) = A\cdot\mathbf{x} = \mathbf{0} \text{ in } \mathbb{K}^n\}$ is a vector subspace in $\mathbb{K}^m$.

2.8. Example. Given a particular solution $\mathbf{x}_0$ of $A\mathbf{x} = B$, the full solution set of this equation consists of the vectors $W_B = \mathbf{x}_0 + W$, where $W = \{x \in \mathbb{K}^m : A\mathbf{x} = 0\}$ is a vector subspace of $\mathbb{K}^m$ because $A\mathbf{x}_1$, $A\mathbf{x}_2 = 0$ implies $A(\mathbf{x}_1 + \mathbf{x}_2) = A\mathbf{x}_1 + A\mathbf{x}_2 = 0 + 0 = 0$ and $A(\lambda \cdot \mathbf{x}) = \lambda \cdot A\mathbf{x} = \lambda \cdot 0 = 0$.
Note: The converse is also true: in $\mathbb{K}^m$ every vector subspace is the solution set of some homogeneous system of linear equation $A\mathbf{x} = 0$, but we are not ready to prove that yet. The situation is shown in Figure 1.4. □

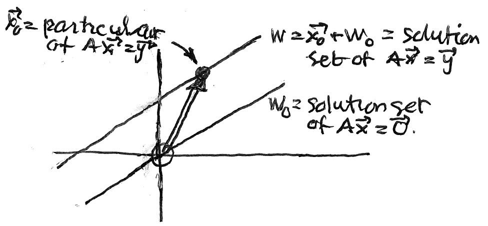
$\vec{x}_0$= particular of $A\vec{x}=\vec{y}$
$W = \vec{x}_0 + W_0$ = solution set of $A\vec{x} = \vec{y}$
W_0= solution set of $A\vec{x} = \vec{0}$.

Figure 1.4. The subspace W_0 is the solution set for a homogeneous equation $A\mathbf{x} = \mathbf{0}$. If the inhomogeneous equation $A\mathbf{x} = \mathbf{y}$ has solutions and if $\mathbf{x}_0$ is a particular solution, so $A\mathbf{x}_0 = y$, the full solution set $W = \{\mathbf{x} : A\mathbf{x} = \mathbf{B}\}$ is the translate $W' = \mathbf{x}_0 + W$ of W_0.

This of course presumes that $A\mathbf{x} = B$ has any solutions at all; if it does not, we say that the system is *inconsistent*. Geometrically, that means B does not lie in the range $R(L_A)$. Here is an example of an inconsistent inhomogeneous system.

$$\begin{pmatrix} 1 & 0 \\ 2 & 0 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

The corresponding system of linear equations

$$\begin{cases} x_1 + 0 \cdot x_2 &= 0 \\ 2x_1 + 0 \cdot x_2 &= 1 \end{cases}$$

implies that $x_1 = 0$ and $2x_1 = 1$, an obvious impossibility.

We will continue discussion of linear systems and their solutions via elementary row operations on A, or on the augmented matrix $[A : B]$, but first a few more examples of vector spaces we will encounter from time to time.

2.9. Example (Sequence Space ℓ^∞). Let ℓ^∞ = all sequences $a = (a_1, a_2, ...)$ with $a + b = (a_1 + b_1, a_2 + b_2, \dots)$ and $\lambda \cdot a = (\lambda a_1, \lambda a_2, ...)$. This infinite dimensional space has the following subspaces:

1. $W_0 = \{$sequences such that $a_n \to 0$ as $n \to \infty\}$;
2. W_n = all sequences of the form $(a_1, ..., a_n, 0, 0, ...)$;
3. $\ell^1 = \{\, a : \sum_{n=1}^\infty |a_n| < \infty \,\}$

2.10. Example. In $M(n, \mathbb{K})$ we have various significant subspaces

1. Symmetric matrices: $A^{\mathrm{t}} = A$ where A^{t} = (transpose of A).

2. Diagonal matrices: $D = \begin{pmatrix} d_1 & & & & 0 \\ & \cdot & & & \\ & & \cdot & & \\ & & & \cdot & \\ 0 & & & & d_n \end{pmatrix}$

3. Block diagonal matrices: $D_{m_1,\ldots,m_r} = \begin{pmatrix} \boxed{B_{m_1\times m_1}} & & & & 0 \\ & \cdot & & & \\ & & \cdot & & \\ & & & \cdot & \\ 0 & & & & \boxed{B_{m_r\times m_r}} \end{pmatrix}$

   for fixed indices $m_1, \ldots, m_r \geq 1$. (The "blocks" are allowed to have arbitrary entries and all other entries are zero; $m_1 + \ldots + m_r = n$.)

4. Upper triangular and Strictly upper triangular matrices.

$$\begin{pmatrix} * & & & & * \\ & \cdot & & & \\ & & \cdot & & \\ & & & \cdot & \\ 0 & & & & * \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & & & & * \\ & \cdot & & & \\ & & \cdot & & \\ & & & \cdot & \\ 0 & & & & 0 \end{pmatrix}$$

2.11. Exercise. Which of these four subspaces, if any, are closed under matrix multiplication as well as $(+)$?

2.12. Exercise. Show that the vector subspace of upper triangular and strictly upper triangular matrices are closed under formation of matrix product AB.

2.13. Exercise. Show that the vector subspaces of upper triangular (or strictly upper triangular) matrices are **Lie algebras**: all **commutators** $[A, B] = AB - BA$ are (strictly) upper triangular if A, B are.

2.14. Exercise. If an $n \times n$ matrix A has the strictly upper triangular form shown in (a), prove that A^2 has the form in (b).

$$\text{(a)} \quad A = \begin{pmatrix} 0 & * & & & * \\ & 0 & * & & * \\ & & \ddots & & \\ & & & 0 & * \\ 0 & & & & 0 \end{pmatrix} \qquad \text{(b)} \quad A^2 = \begin{pmatrix} 0 & 0 & * & & & & & * \\ & 0 & 0 & * & & & & * \\ & & 0 & 0 & * & & & * \\ & & & \ddots & \ddots & & & \\ & & & & & \ddots & \ddots & * \\ & & & & & & 0 & 0 \\ 0 & & & & & & & 0 \end{pmatrix}$$

Note: Further computations show that A^3 has three diagonal files of zeros, etc so that A is a *nilpotent operator*, with $A^n = 0_{n\times n}$.

I.3. Determining Linear Span: A Case Study

Given vectors $\{v_1, \ldots, v_r\} \subseteq V$ and $b \in V$, the basic problem is to decide whether there exist x_1, .., $x_r \in \mathbb{K}$ such that $b = \sum_{i=1}^{r} x_i v_i$ (and if so, for which choices of coefficients x_1, .., x_r). Row operations on matrices are the main tool for resolving such questions.

3.1. Example. Consider the vectors in $\mathbb{K}^3$

$$\mathbf{u}_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \ \mathbf{u}_2 = \begin{pmatrix} -2 \\ -4 \\ -2 \end{pmatrix}, \ \mathbf{u}_3 = \begin{pmatrix} 0 \\ 2 \\ 3 \end{pmatrix}, \ \mathbf{u}_4 = \begin{pmatrix} 2 \\ 0 \\ -3 \end{pmatrix}, \ \mathbf{u}_5 = \begin{pmatrix} -3 \\ 8 \\ 16 \end{pmatrix}$$

and let A be the matrix with these vectors as its columns

$$A = \begin{pmatrix} 1 & -2 & 0 & 2 & -3 \\ 2 & -4 & 2 & 0 & 8 \\ 1 & -2 & 3 & -3 & 16 \end{pmatrix}$$

If $B = \mathrm{col}(2, 6, 8) = \begin{pmatrix} 2 \\ 6 \\ 8 \end{pmatrix}$, determine all column vectors

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix}$$

such that $\sum_i x_i \mathbf{u}_i = \mathbf{0}$ or $\sum_i x_i \mathbf{u}_i = B$ in $\mathbb{K}^3$. (In the second case we are determining whether B lies in the linear span of $\{\mathbf{u}_1, \ldots, \mathbf{u}_5\}$.) Then do this for an arbitrary column vector $B = \mathrm{col}(b_1, b_2, b_3)$ to to get all solutions of $A\mathbf{x} = B$.

Discussion: A solution $\mathbf{x} = \mathrm{col}(x_1, \ldots, x_5)$ of $A\mathbf{x} = B$ statisfies the matrix equation

$$\begin{aligned} B &= \sum_{i=1}^{5} x_i \mathbf{u}_i = x_1 \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} + x_2 \begin{pmatrix} -2 \\ -4 \\ -2 \end{pmatrix} + x_3 \begin{pmatrix} 0 \\ 2 \\ 3 \end{pmatrix} + x_4 \begin{pmatrix} 2 \\ 0 \\ -3 \end{pmatrix} + x_5 \begin{pmatrix} -3 \\ 8 \\ 16 \end{pmatrix} \\ &= \begin{pmatrix} 1 & -2 & 0 & 2 & -3 \\ 2 & -4 & 2 & 0 & 8 \\ 1 & -2 & 3 & -3 & 16 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = A\mathbf{x} \ . \end{aligned}$$

We shall determine the full solution sets of the systems $A\mathbf{x} = 0$ or $A\mathbf{x} = B$ for the 3×5 matrix $A = [\mathbf{u}_1; \mathbf{u}_2; \mathbf{u}_3; \mathbf{u}_4; \mathbf{u}_5]$.

Before analyzing this problem we recall a few basic facts about solving matrix equations using elementary row operations. These methods are based on the following observations with which you should already be familiar: see the early chapters of *Schaum's Outline*. The simple (but important) verification is left as an exercise.

3.2. Proposition. *The following elementary row operations on a matrix A do not change the set of solutions $\mathbf{x}$ of $A\mathbf{x} = 0$.*

1. $R_i \leftrightarrow R_j$: *switch two rows*;

2. $R_i \rightarrow \lambda R_i$: *scale* (*row* i) *by some* $\lambda \neq 0$ *in* $\mathbb{K}$;

3. $R_i \rightarrow R_i + \lambda R_j$: *for* $i \neq j$ *add any scalar multiple of* (*row* j) *to* (*row* i), *leaving* (*row* j) *unaltered.*

Applied to the "augmented matrix" $[A : B]$ associated with an inhomogeneous system $A\mathbf{x} = B$, the system $A'\mathbf{x} = B'$ associated with the modified matrix $[A' : B']$ has the same solution set as $A\mathbf{x} = B$.

The reason is that each of the moves 1.-3. is reversible, with $R_i \rightarrow R_i - \lambda R_j$ the inverse of $R_i \rightarrow R_i + \lambda R_j$. Although row operations do not change the solution set they can greatly simplify the system of equations to be solved, leading to easy systematic solution of matrix equations. For instance, when $\mathbb{K} = \mathbb{Q}, \mathbb{R}, \mathbb{C}$ it is always possible to find a sequence of row operations that put A into upper triangular **echelon form:**

$$\text{(1)}\qquad \textsc{Echelon Form:}\qquad A = \left(\begin{array}{cccccccccc} \boxed{1} & * & . & . & . & & & & & * \\ & & & \boxed{1} & * & . & . & . & . & * \\ 0 & & & & & \boxed{1} & * & . & . & * \\ \hline & & & & & & & & & \\ \multicolumn{10}{c}{\mathbf{0}} \end{array}\right)$$

The same moves put the augmented matrix $[A : B]$ into similar form

$$\text{(2)}\qquad [A' : B'] = \left(\begin{array}{cccccccccc|c} \boxed{1} & * & . & . & . & & & & & * & b'_1 \\ & & & \boxed{1} & * & . & . & . & & * & \vdots \\ 0 & & & & & \boxed{1} & * & . & . & * & b'_r \\ \hline \multicolumn{10}{c|}{} & b'_{r+1} \\ \multicolumn{10}{c|}{\mathbf{0}} & \vdots \\ \multicolumn{10}{c|}{} & b'_m \end{array}\right)$$

Solutions of the systems $A'\mathbf{x} = 0$, $A'\mathbf{x} = B'$ are quickly found by "backsolving" (illustrated bellow). One could go further, forcing A into even simpler form by knocking out all terms $*$ above the "step corners." These additional operations would of course affect B' in the augmented matrix yielding the **reduced echelon form**.

$$[A'' : B''] = \left(\begin{array}{cccccccccc|c} \boxed{1} & * & . & 0 & * & 0 & & & & * & b''_1 \\ & & & \boxed{1} & * & 0 & . & . & . & * & \\ 0 & & & & & \boxed{1} & * & . & . & * & b''_r \\ \hline \multicolumn{10}{c|}{} & b''_{r+1} \\ \multicolumn{10}{c|}{\mathbf{0}} & \vdots \\ \multicolumn{10}{c|}{} & b''_m \end{array}\right)$$

The "step corners" appearing in these echelon displays are often referred to as "*pivots*," and the columns in which they occur are the "*pivot columns*."

Notice that $A\mathbf{x} = B$ has the same solutions as $A'\mathbf{x} = B'$ where $[A' : B']$ is the echelon form of $[A : B]$. Solutions exist if and only if we have $b'_{r+1} = ... = b'_n = 0$ (the terms in B' below the row containing the last "step corner") because the last equations in the new linear system $A'\mathbf{x} = B'$ read $0 = b'_{r+1}$, ..., $0 = b'_n$ (the variables x_1,..., x_m don't appear!) These are inconsistent unless $b'_{r+1} = ... = b'_n = 0$.

Columns $C_i(A)$ that *do not* pass though a step corner correspond to "free variables" x_i in the solutions of the equation $A'\mathbf{x} = 0$; they are also free variables in solutions of $A'\mathbf{x} = B'$ if the consistency conditions $b'_{r+1} = ... = b'_n = 0$ have been met (without which there are no solutions at all.) If $I = \{1 \leq i_1 < ... < i_r \leq m\}$ are the indices labeling the pivot columns, the remaining indices correspond to free variables x_k $(k \notin I)$ in the solution. Once the values of the free variables have been specified, backsolving yields the values of the remaining "dependent" variables x_k $(k \in I)$. We get a unique solution $A'\mathbf{x} = 0$ for *every* choice of the free variables $(k \notin I)$; different choices yield different solutions and all solutions are accounted for. By Proposition 3.2 these are also the solutions of the original equation $A\mathbf{x} = 0$.

Example 3.1 (Resumed). Returning to our discussion, we put the original system into echelon form by applying row operations to

$$\left(\begin{array}{ccccc|c} 1 & -2 & 0 & 2 & -3 & 0 \\ 2 & -4 & 2 & 0 & 8 & 0 \\ 1 & -2 & 3 & -3 & 16 & 0 \end{array}\right)$$

Applying, $R_2 \leftarrow R_2 - 2R_1$ and $R_3 \leftarrow R_3 - R_1$ this becomes

$$\left(\begin{array}{ccccc|c} \boxed{1} & -2 & 0 & 2 & -3 & 0 \\ 0 & 0 & 2 & -4 & 14 & 0 \\ 0 & 0 & 3 & -5 & 19 & 0 \end{array}\right)$$

Now apply $R_3 \leftarrow R_3 - \frac{3}{2}R_2,\ R_2 \leftarrow \frac{1}{2}R_2$, and then $R_3 \leftarrow R_3 - 3R_2$ to get

$$\left(\begin{array}{ccccc|c} \boxed{1} & -2 & 0 & 2 & -3 & 0 \\ 0 & 0 & \boxed{1} & -2 & 7 & 0 \\ 0 & 0 & 0 & \boxed{1} & -2 & 0 \end{array}\right)$$

This is the desired echelon form. Some additional work, needless for most purposes, would yield the reduced echelon form,

$$\rightarrow \left(\begin{array}{ccccc|c} \boxed{1} & -2 & 0 & 0 & * & 0 \\ 0 & 0 & \boxed{1} & 0 & * & 0 \\ 0 & 0 & 0 & \boxed{1} & -2 & 0 \end{array}\right)$$

Recursively backsolving the corresponding system of linear equations, we see that

1. x_2, x_5 are free variables;
2. $x_4 - 2x_5 = 0 \ \Rightarrow\ x_4 = 2x_5$;
3. $x_3 - 2x_4 + 7x_5 = 0 \ \Rightarrow\ x_3 = -7x_5 + 2(2x_5) = -3x_5$;
4. $x_1 - 2x_2 + 2x_4 - 3x_5 = 0 \ \Rightarrow\ x_1 = 2x_2 - 2(2x_5) + 3x_5 = 2x_2 - x_5$.

The solutions of $A'\mathbf{x} = 0$ (which are also the solutions of $A\mathbf{x} = 0$) form a vector subspace in $\mathbb{K}^5$, each of whose points is uniquely labeled (parametrized) by the choice of the free variables x_2, x_5. Setting $x_2 = s$, $x_5 = t$ $(s, t \in \mathbb{K})$ we find that the solution set $W = \{\mathbf{x} \in \mathbb{K}^5 : A\mathbf{x} = 0\} = \{\mathbf{x} \in \mathbb{K}^5 : A'\mathbf{x} = 0\}$ is equal to

$$W = \left\{ \begin{pmatrix} 2s - t \\ s \\ -3t \\ 2t \\ t \end{pmatrix} : s, t \in \mathbb{K} \right\} = \left\{ \begin{pmatrix} 2x_2 - x_5 \\ x_2 \\ -3x_5 \\ 2x_5 \\ x_5 \end{pmatrix} : x_2, x_5 \in \mathbb{K} \right\}$$

These homogeneous solutions can be rewritten in a more instructive form

$$\mathbf{x} = \begin{pmatrix} 2s - t \\ s \\ -3t \\ 2t \\ t \end{pmatrix} = s \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 0 \\ -3 \\ 2 \\ 1 \end{pmatrix} = s\,\mathbf{w}_1 + t\,\mathbf{w}_2 \ \ ,$$

which shows that every solution of $A\mathbf{x} = 0$ is a linear combination of two basic solutions

$$\mathbf{w}_1 = \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \qquad \text{and} \qquad \mathbf{w}_2 = \begin{pmatrix} -1 \\ 0 \\ -3 \\ 2 \\ 1 \end{pmatrix}$$

This approach describes the solution set W of $A\mathbf{x} = 0$ as linear span $\mathbb{K}$-span$\{\mathbf{w}_1, \mathbf{w}_2\}$ of a set of generators $\{\mathbf{w}_1, \mathbf{w}_2\}$. We will later observe that these vectors are a "basis" for the solution set W.

Solving the Inhomogeneous Equation $A\mathbf{x} = B$. The same elementary row operations that put A into echelon form may be applied to the augmented matrix $[A : B]$. We already know what happens to A; applying the same moves to the column vector $B = \text{col}(b_1, b_2, b_3)$ with undetermined coefficients, the operations $R_2 \leftarrow R_2 - 2R_1$ and $R_3 \leftarrow R_3 - R_1$ transform

$$B = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \rightarrow \begin{pmatrix} b_1 \\ b_2 - 2b_1 \\ b_3 - b_1 \end{pmatrix}$$

Then $R_3 \leftarrow R_3 - \frac{3}{2}R_2$; $R_2 \leftarrow \frac{1}{2}R_2$, and $R_3 \leftarrow R_3 - 3R_2$ yield

$$\rightarrow \begin{pmatrix} b_1 \\ \frac{1}{2}b_2 - b_1 \\ b_3 - b_1 - \frac{3}{2}(b_2 - 2b_1) \end{pmatrix} = \begin{pmatrix} b_1 \\ \frac{1}{2}b_2 - b_1 \\ b_3 - \frac{3}{2}b_2 + 2b_1 \end{pmatrix}$$

The augmented matrix becomes

$$[A : B] \rightarrow \left(\begin{array}{ccccc|c} \boxed{1} & -2 & 0 & 2 & -3 & b_1 \\ 0 & 0 & \boxed{1} & -2 & 7 & \frac{1}{2}b_2 - b_1 \\ 0 & 0 & 0 & \boxed{1} & -2 & b_3 - \frac{3}{2}b_2 + 2b_1 \end{array}\right)$$

Again x_2 and x_5 are free variables and the general solution $\mathbf{x} = \text{col}(x_1, x_2, x_3, x_4, x_5)$ of $A\mathbf{x} = B$ can be found by backsolving. Since we have already found the general solutions of $A\mathbf{x} = 0$, all we need is one particular solution $\mathbf{x}_B$. The simplest way to find one is to set $x_2 = x_5 = 0$ and backsolve to get

$$\begin{aligned} &x_2, x_5 = 0 \\ &x_4 = b_3 - \tfrac{3}{2}b_2 + 2b_1 \\ &x_3 - 2x_4 = \tfrac{1}{2}b_2 - b_1 \Rightarrow x_3 = 2(b_3 - \tfrac{3}{2}b_2 + 2b_1) + \tfrac{1}{2}b_2 - b_1 = 2b_3 - \tfrac{5}{2}b_2 + 3b_1 \\ &x_1 - 2 \cdot 0 + 0 + 2x_4 + 0 = b_1 \Rightarrow x_1 = b_1 - 2x_4 = -2(b_3 - \tfrac{3}{2}b_2 + 2b_1) + b_1 = -2b_3 + 3b_2 - 3b_1. \end{aligned}$$

So $\mathbf{x}_B = \text{col}\big(-2b_3 + 3b_2 - 3b_1, 0, 2b_3 - \frac{5}{2}b_2 + 3b_1, b_3 - \frac{3}{2}b_2 + 2b_1, 0\big)$ is a particular solution and the full solution set is

$$W_B = \{x : A\mathbf{x} = B\} = \begin{pmatrix} -2b_3 + 3b_2 - 3b_1 \\ 0 \\ 2b_3 - \frac{5}{2}b_2 + 3b_1 \\ b_3 - \frac{3}{2}b_2 + 2b_1 \\ 0 \end{pmatrix} + \mathbb{K}\mathbf{w}_1 + \mathbb{K}\mathbf{w}_2$$

where $\mathbf{w}_1$ and $\mathbf{w}_2$ are the basis vectors for the space $W = \{x : A\mathbf{x} = 0\}$ of homogeneous solutions determined previously. Writing $s = x_2$, $t = x_5$ for the variable attached to $\mathbf{w}_1$, $\mathbf{w}_2$ we obtain a parametric description of the solution set, with each point in W_B tagged by a unique pair (s, t) in the parameter space $\mathbb{K}^2$.

In the problem originally posed we had $B = \text{col}(2, 6, 8)$. Then the particular solution is $\mathbf{x}_0 = \text{col}(-4, 0, 7, 3, 0)$ and the solution set is

$$W_B = \begin{pmatrix} -4 + 2s - t \\ s \\ 7 - 3t \\ 3 + 2t \\ t \end{pmatrix} = \mathbf{x}_0 + \mathbb{K}\mathbf{w}_1 + \mathbb{K}\mathbf{w}_2$$

That concludes our discussion of the Case Study 3.1. □

Further Remarks about Elementary Row Operations. Row operations can also be used to determine the subspace spanned by any finite set of vectors in $\mathbb{K}^m$. If these have the form $R_1 = (a_{11}, .., a_{1m})$,..., $R_n = (a_{n1}, ..., a_{nm})$ we may regard them as the rows of an $n \times m$ matrix

$$A = \begin{pmatrix} \underline{R_1} \\ \cdot \\ \cdot \\ \cdot \\ \overline{R_n} \end{pmatrix} = \begin{pmatrix} a_{11} & \cdot & \cdot & \cdot & a_{1m} \\ a_{21} & \cdot & \cdot & \cdot & a_{2m} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ a_{n1} & \cdot & \cdot & \cdot & a_{nm} \end{pmatrix}$$

The linear span $\text{Row}(A) = \mathbb{K}\text{-span}\{R_1, ..., R_n\} \subseteq \mathbb{K}^m$ is called the **row space** of A; the linear span of its columns $C_1, \ldots, C_m$ is the **column space** $\text{Col}(A) = \mathbb{K}\text{-span}\{C_1, .., C_m\}$ in $\mathbb{K}^n$. One can show that:

3.3. Lemma. *Elementary row operations on a matrix A do not change the linear span of its rows.*

We leave the proof as a routine exercise. Note, however, that row operations will mess up column space!

As for columns, there is an obvious family of elementary column operations on A.

1. $C_i \leftrightarrow C_i$;

2. $C_i \to \lambda C_i$ for $\lambda \neq 0$ in $\mathbb{K}$;

3. $C_i \to C_i + \lambda C_j$, for $i \neq j$ where λ is any element in $\mathbb{K}$.

These do not change the linear span $\text{Col}(A)$. This can be verified by direct calculation, but it also follows by observing that row and column operations are related via a natural symmetry $A \mapsto A^{\text{t}}$ = the *transpose of* A, given by $(A^{\text{t}})_{ij} = A_{ji}$ (see Figure 1.5). Note that $(A^{\text{t}})^{\text{t}} = A$.

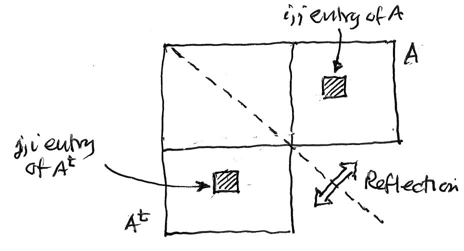

Figure 1.5. A matrix A and its transpose A^{t} are related by a reflection that sends rows in A to columns in A^{t}, and columns to rows.

The transpose operation takes rows of A to columns of A^{t} and vice-versa; elementary row operations on A become the corresponding elementary operations on the columns of A^{t}. It should also be evident that under the transpose operation the row space $\mathrm{Row}(A)$ = (span of the rows, regarded as vectors in $\mathbb{K}^m$) becomes the column space $\mathrm{Col}(A^{\mathrm{t}})$ = (columns in A^{t}, regarded as vectors in $\mathbb{K}^n$) of A^{t}. Invariance of $\mathrm{Col}(A)$ under column operations follows from invariance of $\mathrm{Row}(A^{\mathrm{t}})$ under row operations, discussed earlier.

3.4. Example. Let $v_1, .., v_n \in \mathbb{K}^m$. To find a basis for $W = \mathbb{K}\text{-span}\{v_1, .., v_n\}$, view the v_i as $1 \times m$ row vectors and assemble them as the $n \times m$ matrix

$$A = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}_{n\times m}$$

If we perform row operations to put A in echelon form, this does not change row space $\mathrm{Row}(A) = \mathbb{K}\text{-span}\{v_1, .., v_n\}$, but it is now easy to pick out a minimal set of vectors with the same linear span, namely the rows R_1', ..., R_k' that meet the step corners in the array.

$$A' = \begin{array}{c} R_1' \\ R_2' \\ \cdot \\ \cdot \\ \cdot \\ R_k' \\ \\ \\ \end{array} \left(\begin{array}{cccccccccc} \boxed{1} & * & * & * & * & & & & & * \\ & & \boxed{1} & * & * & & & & & * \\ & & & & \boxed{1} & * & * & & & * \\ & & & & & \ddots & & & & \vdots \\ & & & & & & \boxed{1} & * & & * \\ 0 & & & & & & & & \boxed{1} & * \\ \hline \\ 0 & & & & & & & & & 0 \end{array}\right)$$

We will say more about this in the next section. □

3.5. Exercise. By invariance of row space $\mathrm{Row}(A)$ under row operations, the rows $R_1', ..., R_k'$ also span $\mathrm{Row}(A)$. They are a *basis* for row space if they are also linearly independent in the following sense.

> Linear Independence: *If $\sum_{i=1}^k c_i R_i' = \mathbf{0}$ in $\mathbb{K}^m$ for coefficients $c_1, .., c_k$ in $\mathbb{K}$, we must have $c_1 = ... = c_k = 0$ in $\mathbb{K}$.*

Explain why the row vectors R_1',..., R_k' in the previous example must have this independence property.
Hint: If $\sum_{i=1}^n c_i R_i' = (0, ..., 0)$ in $\mathbb{K}^m$, what conclusion can you draw about the first coefficient c_1? Etc.

A set of vectors $\{v_1, ..., v_n\} \subseteq V$ is a basis for V if they span V and are linearly independent. We will now show that this happens if and only if every $v \in V$ has a unique expansion $v = \sum_{i=1}^n c_i v_i$ with $c_i \in \mathbb{K}$. Independence simply says that the zero vector $v = 0$ in V has the unique expansion $0 = 0{\cdot}v_1 + .. + 0{\cdot}v_n$. But if some vector had two expansions $v = \sum_i c_i v_i = \sum_i d_i v_i$ then $0 = v - v = \sum (c_i - d_i) v_i$, so independence of the v_i implies $c_i = d_i$, and v has a unique expansion.

I.4. Linear Span, Independence and Bases

We now explain how to solve arbitrary systems of linear equations.

4.1. Definition. *A set of vectors $S = \{v_1, .., v_r\}$ in a vector space V* **spans** *a subspace W if*

$$W = \mathbb{K}\text{-span}\{S\} = \Big\{ \sum_{i=1}^r c_i v_i : c_i \in \mathbb{K} \Big\}$$

The vectors are **linearly independent** *if the only linear combination $\sum_i c_iv_i = \mathbf{0}$ adding up to zero in V is the trivial combination with $c_1 = ... = c_r = 0$. The vectors are a* **basis** *for W if they span W and are independent, so every $w \in W$ has a unique representation as $\sum_{i=1}^{n} \lambda_iv_i$ $(\lambda_i \in \mathbb{K})$.*

4.2. Exercise. If $\mathfrak{X} = \{v_1, \dots, v_n\}$ span V and are independent, explain why every $v \in V$ has a unique representation as $\sum_{i=1}^{n} \lambda_iv_i$ $(\lambda_i \in K)$, so $\mathfrak{X}$ is a basis for V.

The next result exhibits two ways to construct a basis in a vector space. One starts with a spanning set and "prunes" it, deleting redundant vectors until we arrive at an independent subset with the same span as the original vectors. This yields a basis for V. The other constructs a basis recursively by adjoining "outside vectors" to an initial family of independent vectors in V. The initial family might consist of a single nonzero vector (obviously an independent set).

4.3. Proposition. *Every finite spanning set $\{v_1, ..., v_n\}$ in a vector space can be made into a basis by deleting suitably chosen entries from the list.*

Proof: We argue by induction on $n = \#$(vectors in list). There is nothing to prove if $n = 1$; then $V = \mathbb{K} \cdot v_1$ and $\{v_1\}$ is already a basis. The induction hypotheses (one for each index $n = 1, 2, ...$) are:

> Hypothesis $P(n)$: *For any vector space V containing a spanning set of n vectors, we can delete vectors from the list to get a basis for V.*

We have proved this for $n = 1$. It is true for all n if we can prove $P(n + 1)$ is true, using only the information that $P(n)$ is true – i.e. if we can verify that

$$P(n) \text{ true } \Rightarrow P(n + 1) \text{ true}$$

(Remember: This is a *conditional* statement owing to the presence of the word "*If...*" It does not assert that $P(n)$ is actually true.)

So, assuming $P(n)$ true consider a spanning set $\mathfrak{X} = \{v_1, ..., v_n, v_{n+1}\}$ in V. If these vectors are already independent (which could be checked using row operations if $V = \mathbb{K}^m$), we already have a basis for V without deleting any vectors. If $\mathfrak{X}$ is not independent there must be coefficients $c_1, ..., c_{n+1} \in \mathbb{K}$ (not all equal to 0) such that $\sum_{i=1}^{n+1} c_iv_i = \mathbf{0}$. Relabeling, we may assume $c_{n+1} \neq 0$, and then ($\mathbb{K}$ being a field)

$$-c_{n+1}v_{n+1} = \sum_{i=1}^{n} c_iv_i \quad \text{and} \quad v_{n+1} = \sum_{i=1}^{n} -\big(c_i/c_{n+1}\big)\cdot v_i$$

Thus $v_{n+1} \in \mathbb{K}\text{-span}\{v_1, ..., v_n\}$ and $\mathbb{K}\text{-span}\{v_1, ..., v_{n+1}\} = \mathbb{K}\text{-span}\{v_1, ..., v_n\}$ is all of V. By the induction hypotheses we may thin out $\{v_1, .., v_n\}$ to get a basis for V. □

4.4. Proposition. *If $\{v_1, ..., v_n\}$ are independent in a vector space V, and v_{n+1} is a vector not in $W_0 = \mathbb{K}\text{-span}\{v_1, ..., v_n\}$ then*

1. *$\{v_1, ..., v_n, v_{n+1}\}$ are independent*;
2. $W_0 \subsetneqq W_1 = \mathbb{K}\text{-span}\{v_1, ..., v_n, v_{n+1}\}$;
3. *$\{v_1, ..., v_n, v_{n+1}\}$ is a basis for W_1.*

Proof: If $v_1 \dots, v_{n+1}$ are not independent there would be $c_i \in \mathbb{K}$ (not all zero) such that $\sum_{i=1}^{n+1} c_iv_i = \mathbf{0}$. We can't have $c_{n+1} = 0$, otherwise $\sum_{i=1}^{n} c_iv_i = \mathbf{0}$ contrary to assumed independence of $\{v_1, ..., v_n\}$. Thus $v_{n+1} = \sum_{i=1}^{n} -\big(c_i/c_{n+1}\big) \cdot v_i$ is in W_0, which contradicts the assumption $v_{n+1} \notin W_0$. Conclusion: $v_1, ..., v_{n+1}$ are independent. It follows immediately that $\{v_1, ..., v_{n+1}\}$ is a basis for $W_1 = \mathbb{K}\text{-span}\{v_1, ..., v_{n+1}\}$.

Note: This is an example of a "proof by contradiction," in which the assumption that $v_1, ..., v_n$ are *not* independent leads to an impossible conclusion. Therefore the statement "$v_1, ..., v_n$ are independent" must be true. □

Important Remark: This process of "adjoining an outside vector" can be iterated to construct larger and larger independent sets and subspaces

$$\begin{aligned} W_0 = \mathbb{K}\text{-span}\{v_1, ..., v_n\} \; &\underset{\neq}{\subseteq} \; W_1 = \mathbb{K}\text{-span}\{v_1, ..., v_n, v_{n+1}\} \\ &\underset{\neq}{\subseteq} \; .... \underset{\neq}{\subseteq} W_r = \mathbb{K}\text{-span}\{v_1, ..., v_{n+r}\} \end{aligned}$$

Since $\{v_1, \dots, v_n\}$ are independent they are a basis for the initial space W_0, and by Lemma 4.4 $v_1, \dots, v_n, \dots, v_{n+r}$ will be a basis for W_r. If this process stops in finitely many steps (because $W_r = V$ and we can no longer find a vector outside W_r), we have produced a basis for V. If the process never stops, no finite subset of vectors can span V and in this case we say V is **infinite dimensional**. To begin the process we need an initial set of independent vectors, but if $V \neq (0)$ we could start with any $v_1 \neq 0$ and $W_0 = \mathbb{K}{\cdot}v_1$. Then apply Lemma 4.4 recursively as above. □

4.5. Definition. *A vector space V is* **finite dimensional** *if there is a finite set of vectors $S = \{v_1, ..., v_n\}$ that span V. Otherwise V is said to be* **infinite dimensional**, *which we indicate by writing* $\dim(V) = \infty$.

Coordinate space $\mathbb{K}^n$ and matrix spaces $\mathrm{M}(m \times n, \mathbb{K})$ are finite dimensional; the spaces of polynomials $\mathbb{K}[x]$ and $\mathbb{K}[x_1, \dots, x_n]$ are infinite dimensional.

4.6. Example. Coordinate space $\mathbb{K}^n$ is finite dimensional and is spanned by the **standard basis vectors** $\mathfrak{X} = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$

$$\mathbf{e}_1 = (1, 0, \dots, 0), \quad \mathbf{e}_2 = (0, 1, 0, \dots, 0), \ \dots \ , \mathbf{e}_n = (0, \dots, 0, 1)$$

In fact $\mathfrak{X}$ is a basis for $\mathbb{K}^n$.

Discussion: Obviously $v = (a_1, ..., a_n) = a_1\mathbf{e}_1 + \ldots + a_n\mathbf{e}_n$ so the $\mathbf{e}_i$ span $\mathbb{K}^n$. But if $\sum_i c_i\mathbf{e}_i = \mathbf{0} = (0, ..., 0)$, that means $(c_1, ..., c_n) = (0, ..., 0)$ and $c_i = 0$ for all i. □

4.7. Example. Polynomial space $\mathbb{K}[x]$ is infinite dimensional. Given any finite set of nonzero vectors $\mathfrak{X} = \{f_1, ..., f_r\}$, let $d_i = \deg(f_i)$. All coefficients of f_i are zero if $i > N = \max\{d_1, ..., d_r\}$, and the same is true for all linear combinations $\sum_{i=1}^{r} c_i f_i$. But then $\mathfrak{X}$ cannot span $\mathbb{K}[x]$ because x^{N+1} is not in $\mathbb{K}$-span$\{f_1, ..., f_r\}$.

Actually the vectors $f_0 = \mathbb{1}, f_1 = x, f_2 = x^2, ...$ *are* a basis for $\mathbb{K}[x]$. This (infinite) set of vectors clearly spans $\mathbb{K}[x]$, but it is also independent, for if $\sum_{i=0}^{r} c_i f_i = 0$ that means $c_0 + c_1x + ... + c_rx^r = 0$ as a polynomial, so the symbol string $(c_0, ..., c_r, 0, 0, ...)$ is equal to $(0, 0, 0, ...)$. □

4.8. Corollary. *Every finite dimensional vector space has a basis.*

Proof: If $\{v_1, ..., v_r\}$ span V, hen by Proposition 4.3 we may delete some of the vectors to get an independent set with the same linear span. □

4.9. Lemma. *If $S \subseteq V$ is an independent set of vectors in V and T a finite set of vectors that span V, we can adjoin certain vectors from T to S to get a basis for V containing the original set of independent vectors S.*

Proof: Let $W = \mathbb{K}$-span$\{S\}$. If $W = V$, S is already a basis. If $W \neq V$, there exists some $v_1 \in T$ such that $v_1 \notin W$ and then $S \cup \{v_1\}$ is an independent set, a basis for the larger space $W_1 = \mathbb{K}$-span$\{S \cup \{v_1\}\} \underset{\neq}{\supseteq} W$. Continuing, we get vectors $v_1, ..., v_r$ in T such that $W \underset{\neq}{\subseteq} W_1 \underset{\neq}{\subseteq} W_2 \underset{\neq}{\subseteq} .... \underset{\neq}{\subseteq} W_r$ for $0 \leq i \leq r$, where $W_i = \mathbb{K}$-span$\{v_1, ..., v_i\}$. The process must terminate when no vector $v_{r+1} \in T$ can be found outside of W_r. Then

$T \subseteq W_r$, so $\mathbb{K}$-span$\{T\} = V \subseteq W_r$ and $W_r = V$. Therefore $S \cup \{v_1, ..., v_r\}$ is a basis for $V = W_r$ (and $S \cup \{v_1, ..., v_k\}$ is a basis for W_k for each $1 \leq k \leq r$). $\square$

4.10. Theorem (Dimension Defined). *All bases in a finite dimensional vector space have the same cardinality. More generally, if V is finite dimensional, and S is a finite spanning set (with $|S| = n$), every independent set of vectors $L \subseteq V$ has cardinality $|L| \leq |S|$. In other words, the size of any independent set is always less than or equal to that of any spanning set.*

Proof: We can eliminate vectors from S to get an independent spanning set $S' \subseteq S$, which is then a basis for V. We will show that $|L| \leq |S'| \leq |S|$. Let $S' = \{u_1, ..., u_n\}$ and $L = \{v_1, ..., v_m\}$. Every $v_i \in L$ can be written $v_i = \sum_{i=1}^n a_{ji}u_j$ since the $u_i \in S'$ are a basis for V. On the other hand, if $c_1, ..., c_m$ are scalars such that $0 = \sum_{j=1}^m c_j v_j$, we must have $c_1 = ... = c_m = 0$ because the v_j are independent. But the identity $\sum_{j=1}^m c_j v_j = 0$ can be written another way, as

$$0 = \sum_{i=1}^m c_i \Big(\sum_{j=1}^n a_{ji} u_j \Big) \; = \sum_{j=1}^n \Big(\sum_{i=1}^m a_{ji} c_i \Big) \, u_j$$

Since the $u_j \in S'$ span V and are independent each expression $(\ldots)$ is $= 0$ so the coefficients $c_1, ..., c_m$ satisfy the system of n equations in m unknowns

$$(3) \qquad \sum_{i=1}^m a_{ji} c_i = 0, \qquad \text{for } 1 \leq j \leq n$$

(a solution $C = \mathrm{col}\big(c_1, \ldots, c_m\big)$ of the matrix equation $AC = 0$).

A linear system such as (3) always has nontrivial solutions if the number of unknowns $m = |L|$ exceeds the number of equations $n = |S'|$; it follows that $|L| \leq |S'|$, as claimed. In fact, row operations on the coefficient matrix A yield an echelon form shown below. There are at most n step corners and if $M > n$ there must be at least one column that fails to meet one of these pivots.

$$\left(\begin{array}{cccccccccc} \boxed{1} & * & . & . & . & & & & & * \\ & & & . & . & . & . & . & . & \\ & & & . & . & . & . & . & . & \\ & & & \boxed{1} & * & . & . & . & . & * \\ 0 & & & & & \boxed{1} & * & . & . & * \\ \hline 0 & . & & . & & . & & . & . & 0 \end{array} \right)_{n \times m}$$

Hence there is at least one free variable and the system $AC = 0$ has nontrivial solutions. But we showed above that $C = 0$ is the only solution, so we obtain a contradiction unless $|L| \leq |S'| \leq |S|$. The theorem is proved. $\square$

4.11. Corollary. *In a finite dimensional vector space all bases have the same cardinality, which we refer to hereafter as the* **dimension** $\dim_{\mathbb{K}}(V)$.

Notation: We will often simplify notation when the underlying ground field $\mathbb{K}$ is understood, by writing $\dim(V)$ or even $|V|$ for the dimension of V. $\square$

4.12. Example. We have already seen that $\dim_{\mathbb{K}}(\mathbb{K}^n) = n$, with the standard basis vectors $\mathbf{e}_1 = (1, 0, ..., 0)$, ..., $\mathbf{e}_n = (0, .., 0, 1)$. We may view $\mathbb{C}^n$ (or any vector space over $\mathbb{C}$) as a vector space over $\mathbb{R}$ by restricting scalars in $\lambda \cdot v$ to be real. As a vector space over $\mathbb{C}$ we have $\dim_{\mathbb{C}}(V) = n$, but as a vector space over $\mathbb{R}$ we have $\dim_{\mathbb{R}}(V) = 2n$.

Discussion: In fact, any $v \in \mathbb{C}^n$ can be written as a complex sum $v = \sum_{j=1}^{n} z_j\mathbf{e}_j$, and if $z_j = x_j + iy_j$ we may write

$$v = x_1\mathbf{e}_1 + \ldots + x_n\mathbf{e}_n + y_1(i\mathbf{e}_1) + \ldots + y_n(i\mathbf{e}_n) \qquad \text{with } x_i, y_j \in \mathbb{R}.$$

Thus the vectors $\{\mathbf{e}_1, ..., \mathbf{e}_n, i\mathbf{e}_1, ..., i\mathbf{e}_n\} \subseteq \mathbb{C}^n$ span $\mathbb{C}^n$ as a vector space over $\mathbb{R}$. They are also independent over $\mathbb{R}$, for if

$$0 = \sum a_j\mathbf{e}_j + \sum b_j(i\mathbf{e}_j) = \sum(a_j + ib_j)\mathbf{e}_j \ ,$$

we must have $a_j + ib_j = 0$ and $a_j = b_j = 0$ because $\{\mathbf{e}_j\}$ is a basis over $\mathbb{C}$. $\square$

4.13. Exercise. If V is a finite dimensional vector space and $W \subseteq V$ a subspace, explain why W must also be finite dimensional.

4.14. Exercise. If V_1, V_2 are finite dimensional vector spaces prove that

1. If $V_1 \subseteq V_2$ then $\dim(V_1) \leq \dim(V_2)$;

2. If $\dim(V_1) = \dim(V_2)$ and $V_1 \subseteq V_2$, then $V_1 = V_2$ as sets.

4.15. Exercise. Explain why $W \subseteq V \Rightarrow \dim(W) \leq \dim(V)$, even if one or both of these spaces is infinite dimensional.

Describing Subspaces.

How can a subspace W in a vector space be specified? Every V of dimension n can be identified in a natural way with $\mathbb{K}^n$ once a basis $\{f_1, ..., f_n\}$ in V has been determined, so we may as well restrict attention to describing subspaces W of coordinate space $\mathbb{K}^n$. (Given a basis $\mathfrak{X} = \{f_i\}$ in V the map $j_{\mathfrak{X}} : \mathbb{K}^n \rightarrow V$ given by

$$\mathbf{x} = (x_1, ..., x_n) \mapsto j_{\mathfrak{X}}(\mathbf{x}) = \sum_{i=1}^{n} x_i f_i$$

is a bijection that respects all vector space operations in the sense that

$$j_{\mathfrak{X}}(\lambda\cdot\mathbf{x}) = \lambda\cdot j_{\mathfrak{X}}(\mathbf{x}) \qquad \text{and} \qquad j_{\mathfrak{X}}(\mathbf{x}+\mathbf{y}) = j_{\mathfrak{X}}(\mathbf{x}) + j_{\mathfrak{X}}(\mathbf{y})$$

It is an *isomorphism* between $\mathbb{K}^n$ and V, by which properties of one space can be matched with those of the other.

Subspaces $W \subseteq \mathbb{K}^n$ can be described in two ways.

1. By exhibiting a basis $\mathfrak{X} = \{f_1, ..., f_r\}$ in W, so $W = \mathbb{K}\text{-span}\{\mathfrak{X}\}$ and $\dim_{\mathbb{K}}(W) = r$. This is a "**parametric description**" of W since each $w \in W$ is labeled by a coordinate r-tuple $\mathbf{c} = (c_1, \ldots, c_r)$.

2. By finding a set of linear equations

$$\begin{aligned} a_{11}x_1 + ... + a_{1m}x_m &= 0 \\ \vdots \qquad\qquad & \quad \vdots \\ a_{n1}x_1 + ... + a_{nm}x_m &= 0 \end{aligned}$$

   described by a matrix equation $A\mathbf{x} = \mathbf{0}$ ($A = n \times m$, $\mathbf{0} = n \times 1$, $\mathbf{x} = m \times 1$) whose solution set $\{\mathbf{x} \in \mathbb{K}^m : A\mathbf{x} = \mathbf{0}\}$ is equal to W. Such an **"implicit description"** may include redundant equations. When there are no redundant equations we will see that $W = \{\mathbf{x} \in \mathbb{K}^n : A\mathbf{x} = \mathbf{0}\}$ has dimension $m - n = \dim(V) - \#(\mathit{equations})$.

We illustrate this with some computational examples.

4.16. Example. Determine the dimension of the subspace $W = \mathbb{R}\text{-span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ in $\mathbb{R}^3$ if

$$\mathbf{u}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \qquad \mathbf{u}_2 = \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} \qquad \mathbf{u}_3 = \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix}$$

Find a basis for W. Then describe W as the solution set of a system of linear equations:

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 & = & 0 \\ \vdots & & \vdots \\ a_{n1}x_1 + a_{n2}x_2 + a_{n3}x_3 & = & 0 \end{array}$$

where $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$.

Solution: We write the vectors as the rows of the 3×3 matrix

$$A = \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \mathbf{u}_3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{pmatrix}$$

Row space $W = \text{Row}(A)$, the span of the rows, is unaffected by elementary row operations. These yield the echelon form

$$A \rightarrow \begin{pmatrix} 1 & 2 & 3 \\ 0 & -1 & -2 \\ 0 & -2 & -4 \end{pmatrix} \rightarrow \begin{pmatrix} \boxed{1} & 2 & 3 \\ 0 & \boxed{1} & 2 \\ 0 & 0 & 0 \end{pmatrix}$$

Therefore $\mathbf{w}_1 = (1, 2, 3)$ and $\mathbf{w}_2 = (0, 1, 2)$ span W; they are also independent because $0 = c_1\mathbf{w}_1 + c_2\mathbf{w}_2 = (c_1\,,\, 2c_1 + c_2\,,\, 3c_1 + 2c_2)$ implies

$$\begin{cases} c_1 = 0 \\ 2c_1 + c_2 = 0 \\ 3c_1 + 2c_2 = 0 \end{cases} \Rightarrow \; c_1 = c_2 = c_3 = 0 \;\; .$$

Thus $\{\mathbf{w}_1, \mathbf{w}_2\}$ is a basis and $\dim(W) = 2$. A typical vector in W can be written (uniquely) as

$$s\,\mathbf{w}_1 + t\,\mathbf{w}_2 = (s\,,\, 2s + t\,,\, 3s + 2t) = (x_1, x_2, x_3) \qquad \text{with } s, t \in \mathbb{R}$$

To describe W as the solution set of a system of equations in x_1, x_2, x_3 we need to "eliminate" s, t from this parametric description of W. This can be done by writing

$$\begin{cases} x_1 = s & \Rightarrow s = x_1 \\ x_2 = 2s + t & \Rightarrow x_2 = 2s + t = 2x_1 + t \Rightarrow t = x_2 - 2x_1 \\ x_3 = 3s + 2t & \end{cases}$$

The last equation yields the "constraint" identity that determines W,

$$x_3 = 3s + 2t = 3x_1 + 2(x_2 - 2x_1) = -x_1 + x_2$$

or $x_1 - x_2 + x_3 = 0$ (1 equation in 3 unknows). Thus $W = \{\mathbf{x} \in \mathbb{R}^3 : x_1 - 2x_2 + x_3 = 0\}$, which has dimension $\dim(\mathbb{R}^3) - 1 = 2$. □

4.17. Example. Let $W \subseteq \mathbb{R}^4$ be the solution set for the system of linear equations:

$$\begin{cases} x_1 + x_2 - x_3 + 2x_4 & = & 0 \\ 3x_1 - x_2 + x_4 & = & 0 \end{cases}$$

so $A\mathbf{x} = \mathbf{0}$ ($\mathbf{x} \in \mathbb{R}^4$) where

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

Find a basis for W and determine $\dim_{\mathbb{R}}(W)$. Do the answers change if we replace $\mathbb{R}$ by $\mathbb{Q}$ or $\mathbb{C}$?

Solution: Elementary row operations yield

$$A \to \begin{pmatrix} 1 & 1 & -1 & 2 \\ 0 & -4 & 3 & -2 \end{pmatrix} \to \begin{pmatrix} \boxed{1} & 1 & -1 & 2 \\ 0 & \boxed{1} & -\frac{3}{4} & \frac{1}{2} \end{pmatrix}$$

and for any solution of $A\mathbf{x} = 0$, $\mathbf{x} = \mathrm{col}(x_1, x_2, x_3, x_4)$ has x_3, x_4 as free variables. Backsolving yields the dependent variables

$$\begin{aligned} x_2 &= \tfrac{3}{4}x_3 - \tfrac{1}{2}x_4 \\ x_1 &= -x_2 + x_3 - 2x_4 = (-\tfrac{3}{4}x_3 + \tfrac{1}{2}x_4) + x_3 - 2x_4 \; = \; \tfrac{1}{4}x_3 - \tfrac{3}{2}x_4 \end{aligned}$$

Thus solutions have the form

$$\mathbf{x} = \begin{pmatrix} \frac{1}{4}x_3 - \frac{3}{2}x_4 \\ \frac{3}{4}x_3 - \frac{1}{2}x_4 \\ x_3 \\ x_4 \end{pmatrix} = x_3 \begin{pmatrix} \frac{1}{4} \\ \frac{3}{4} \\ 1 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -\frac{3}{2} \\ -\frac{1}{2} \\ 0 \\ 1 \end{pmatrix} = x_3\mathbf{f}_1 + x_4\mathbf{f}_2$$

for every $x_3, x_4 \in \mathbb{K}$. The solution set is equal to the $\mathbb{R}$-span$\{(1,3,4,0)\,,\,(3,1,0,-2)\}$ = $\mathbb{R}$-span$\{\mathbf{f}_1, \mathbf{f}_2\}$. The vectors $\mathbf{f}_1$, $\mathbf{f}_2$ span the solution set W, but are also independent because

$$c_1(1,3,4,0) + c_2(3,1,0,-2) = (c_1 + 3c_2\,,\, 3c_1 + c_2\,,\, 4c_1\,,\, -2c_2) = (0,0,0,0)$$

implies that $c_1 = c_2 = 0$. Thus $\{\mathbf{f}_1, \mathbf{f}_2\}$ is a basis and $\dim_{\mathbb{R}}(W) = 2$. The result is the same if we replace the ground field $\mathbb{R}$ with $\mathbb{Q}$ or $\mathbb{C}$. □

As a “rule of thumb,” each constraint equation $a_{i1}x_1 + ... + a_{im}x_m = 0$ on $\mathbb{K}^m$ reduces the dimension of the solution set $W = \{\mathbf{x} \in \mathbb{K}^m : A\mathbf{x} = 0\}$ by 1, but this is not always the case.

4.18. Exercise. Consider the special case of one constraint equation

$$W = \{\mathbf{x} : \sum_{i=1}^{n} c_i x_i = 0\} \qquad \text{with } c_1, ..., c_n \in \mathbb{K}$$

1. Under what condition on $\{c_1, ..., c_n\}$ do we have $\dim_{\mathbb{K}}(W) = n - 1$?
2. Explain why $\dim(W) < n - 1$ is impossible.

4.19. Exercise. Same question but now with two constraint equations

$$\begin{cases} a_{11}x_1 + ... + a_{1m}x_m &= 0 \\ a_{21}x_1 + ... + a_{2m}x_m &= 0 \end{cases}$$

(or $A\mathbf{x} = \mathbf{0}$ with $A = 2 \times m$, $\mathbf{x} = m \times 1$, $\mathbf{0} = 2 \times 1$.) Now what condition on A make

1. $\dim_{\mathbb{K}}(W) = 0$
2. $\dim_{\mathbb{K}}(W) = 1$,

for the subspace $W = \{\mathbf{x} \in \mathbb{K} : A\mathbf{x} = \mathbf{0} \text{ in } \mathbb{K}^2\}$?

4.20. Example (Lagrange Interpolation Formula). For any infinite field such as $\mathbb{K} = \mathbb{Q}, \mathbb{R}, \mathbb{C}$, the problem of finding a polynomial $f \in \mathbb{K}[x]$ having specified values $f(p_j) = \lambda_j$ at a given set of distinct points $p_1, ..., p_n$ in $\mathbb{K}$ always has a solution. The solution is nonunique (the problem is underdetermined) unless we require that $\deg(f) = n-1$; there may be no solution if $\deg(f) < n-1$.

Discussion: The product $h(x) = \prod_{j=1}^{n}(x - p_j)$ has degree equal to n and is zero at each p_j (and zero nowhere else), so the solution to the interpolation problem cannot be unique without restrictions on $f(x)$: one can add h (or any scalar multiple thereof) to any proposed solution f. It is reasonable to ask for a solution $f(x)$ of minimal degree to reduce the ambiguity. The polynomial

$$f(x) = \sum_{i=1}^{n} \lambda_i \cdot \frac{\prod_{j \neq i}(x - p_j)}{\prod_{j \neq i}(p_i - p_j)} \tag{4}$$

has nonzero denominator, is equal to λ_i at p_i for each i, and has $\deg(f) = n-1$.

This is the **Lagrange Interpolation Formula**, determined by direct methods. It is a bit complicated to rewrite this sum of products in the form $f = c_0 + c_1 x + ... + c_{n-1}x^{n-1}$. But the coefficients $c_0, \ldots, c_{n-1}$ can also be found directly as the solution of a system of linear equations

$$\lambda_j = f(p_j) = \sum_{k=0}^{n-1} p_j^k c_k \qquad \text{for } 1 \leq j \leq n-1 \ ,$$

which is equivalent to the matrix equation $A\mathbf{c} = \lambda$ in which

$$A = \begin{pmatrix} p_1^0 & . & . & . & p_1^{n-1} \\ . & & & & . \\ . & & & & . \\ . & & & & . \\ p_n^0 & . & . & . & p_n^{n-1} \end{pmatrix}_{n \times n} \quad \text{and} \quad \mathbf{c} = \begin{pmatrix} c_0 \\ . \\ . \\ . \\ c_{n-1} \end{pmatrix}_{n \times 1} \quad \lambda = \begin{pmatrix} \lambda_0 \\ . \\ . \\ . \\ \lambda_{n-1} \end{pmatrix}_{n \times 1}$$

I.5 Quotient Spaces V/W.

If V is a vector space and W a subspace, the **additive cosets** of W are the translates of W by various vectors in V. They are the subsets $x + W = \{x + w : w \in W\}$ for some $x \in V$, which we shall often denote by $[x]$ when the subspace W is understood. In particular, W itself is the "**zero coset**": $[0] = 0 + W = W$. The key observation is that the whole space V gets partitioned into disjoint cosets that fill V. The collection of all cosets $[x]$ is the **quotient space** V/W. Observe that *points* in the space V/W are at the same time *subsets* in V.

5.1. Lemma. *If W is a subspace in V and $x, y \in V$,*

1. *Two cosets $x + W$ and $y + W$ either coincide or are disjoint, hence the distinct cosets of W partition the space V.*
2. *An additive coset can have various* **representatives** *$x \in V$. We have $y + W = x + W \Leftrightarrow$ there is some $w \in W$ such that $y = x + w$ (or $y - x \in W$).*
3. *If $y \in x + W$ then $y + W = x + W$.*

Proof: We start with an observation about sums $A + B = \{a + b : a \in A, b \in B\}$ of sets $A, B \subseteq V$ that will be invoked repeatedly in what follows.

5.2. Exercise. If W is a subspace of a vector space V and $w \in W$, prove that

1. $w + W = W$, for all $w \in W$;
2. $W + W = \{w_1 + w_2 : w_1, w_2 \in W\}$ is equal to W;
3. $W - W = W$.

Resuming the proof of Lemma 5.1, if cosets $x + W$ and $y + W$ have a point p in common there are $w_1, w_2 \in W$ such that $x + w_1 = p = y + w_2$, hence $y = x + (w_1 - w_2)$. By Exercise 5.2 the cosets are equal:

$$y + W = (x + (w_1 - w_2)) + W = x + \big((w_1 - w_2) + W\big) = x + W$$

For (2.), $x + W = y + W \Rightarrow y = y + 0 = x + w$ for some $w \in W$. Conversely, if $y = x + w$ for $w \in W$, then $y + W = x + (w + W) = x + W$ again by the Exercise. For (3.), it follows from (1.) that $y \in x + W \Rightarrow (y + W) \cap (x + W) \neq \emptyset \Rightarrow y + W = x + W$.

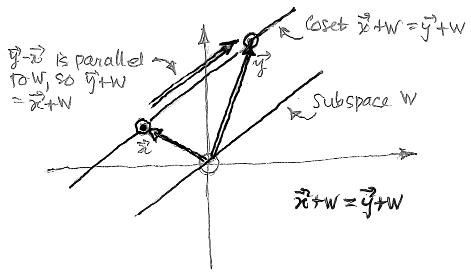
Coset $\vec{x}+W = \vec{y}+W$
$\vec{y}-\vec{x}$ is parallel to W, so $\vec{y}+W$ $= \vec{x}+W$
$\vec{y}$
Subspace W
$\vec{x}$
$\vec{x}+W = \vec{y}+W$

Figure 1.6. Additive cosets $\mathbf{x}+W$ of a subspace W are a family of parallel "hyperplanes" in a vector space V. When $V = \mathbb{R}^2$ and W a line through the origin, *all* lines parallel to W are cosets. Two vectors $\mathbf{x}, \mathbf{y}$ in the same coset yield the same translate of W: $\mathbf{x} + W = \mathbf{y} + W$ because $\mathbf{y} - \mathbf{x}$ is parallel to the subspace W.

As an example, if $V = \mathbb{R}^2$ and $W = \{(x, y) : x = y\}$ the cosets of W are precisely the distinct lines in the plane that make an angle of 45° with the positive x-axis. These lines are the "points" in the quotient space V/W, see Figure 1.6.

5.3. Definition. *There is a natural surjective* **quotient map** $\pi : V \to V/W$*, such that*

$$\pi(x) = [x] = x + W \tag{5}$$

If C is a coset, any point $v \in C$ such that $C = [v] = v + W$ is called a **representative of the coset**. *Part* (2.) *of Lemma 5.1 tells us when two vectors x, y represent the same coset.*

Algebraic Structure in V/W.

There are natural sum and scalar multiplication operations in V/W, inherited from the overlying vector space V.

5.4. Definition. *For any $x, y \in V$ and $\lambda \in \mathbb{K}$ we define operations in V/W*

1. Addition: $[x] \oplus [y] = [x + y]$;
2. Scalar Multiplication: $\lambda \odot [x] = [\lambda \cdot x]$

To spell out what is involved, this definition tells us how to form the sum $X \oplus Y$ of two cosets $X, Y \in V/W$ via the following algorithm:

1. Pick representatives $x, y \in V$ such that $X = [x], Y = [y]$.

2. Add the representatives to get $x + y \in V$.

3. Form the coset $[x + y] = (x + y) + W$ and report the output: $X \oplus Y = [x + y]$

But why should this make sense? The outcome depends on a choice of representatives for each coset X, Y and if different choices yield different outputs, everything written above is nonsense. Fortuately the outcome is independent of the choice of representatives and the operation $(\oplus)$ is *well-defined.* In fact, if $[x] = [x']$ and $[y] = [y']$ there must exist $w_1, w_2 \in W$ such that $x' = x + w_1,\ y' = y + w_2$, and

$$[x' + y'] = (x' + y') + W = (x + y) + \big((w_1 + w_2) + W\big) = (x + y) + W = [x + y]$$

Similarly, the scaling operation is well-defined: if $[x'] = [x]$ we have $x' = x + w$ for some $w \in W$, and then

$$[\lambda{\cdot}x'] = (\lambda{\cdot}x') + W = (\lambda{\cdot}x) + (\lambda w + W) = (\lambda{\cdot}x) + W = [\lambda{\cdot}x]$$

Once we know the operations $(\oplus)$ and $(\odot)$ make sense, direct calculations involving representatives show that all vector space axioms are satisfied by the system $(V/W, \oplus, \odot)$. For instance,

1. Associativity of $\oplus$ on V/W follows from associativity of $(+)$ on V: since $x + (y + z) = (x + y) + z$ in V we get

$$\begin{array}{rcl} [x] \oplus \big([y] \oplus [z]\big) & = & [x] \oplus [y + z] \ = \ [x + (y + z)] \\ & = & [(x + y) + z] \ = \ [x + y] \oplus [z] \ = \ \big([x] \oplus [y]\big) \oplus [z] \end{array}$$

2. The zero element is $[0] = 0 + W = W$ because $[0] \oplus [x] = [0 + x] = [x]$

3. The additive inverse $-[x]$ of $[x] = x + W$ is $[-x] = (-x) + W$ since $[x] \oplus [-x] = [x + (-x)] = [0]$.

5.5. Exercise. Verify the remaining vector space axioms for $(V/W, \oplus, \odot)$. Then show that the quotient map $\pi : V \to V/W$ with $\pi(x) = [x] = x + W$ "*intertwines*" the algebraic operations in $(V, +, \cdot\,)$ with those in $(V/W, \oplus, \odot)$ in the sense that: for any $v_1, v_2 \in V$ and $\lambda \in \mathbb{K}$ we have

1. $\pi(v_1 + v_2) = \pi(v_1) \oplus \pi(v_2)$

2. $\pi(\lambda \cdot v_1) = \lambda \odot \pi(v_1)$

Thus $\pi : V \to V/W$ is a *linear operator* between these vector spaces. □

When $W = (0)$ the quotient space consists of single points $[v] = v + W = \{v\}$, and V/W has a natural identification with V under the quotient map which is now a bijection. When $W = V$, there is just one coset, $v + W = v + V = V$; the quotient space reduces to a single point, the zero element $[0] = 0 + V = V$.

5.6. Exercise. Let $V = \mathbb{R}^3$ and $W = \{(x_1, x_2, x_3) : x_3 = 0\}$ = the x, y-plane in 3-dimensional space. The cosets in V/W are the distinct planes parallel to the x, y-plane: if $v = (v_1, v_2, v_3)$ then

$$\begin{array}{rcl} v + W & = & \{v + w : w \in W\} \\ & = & \{(v_1, v_2, v_3) + (w_1, w_2, 0) : w_1, w_2 \in \mathbb{R}\} \\ & = & \{(v_1 + s, v_2 + t, v_3) : s, t \in \mathbb{R}\} \\ & = & \{(x_1, x_2, x_3) : x_1, x_2 \in \mathbb{R}, x_3 = v_3\} \end{array}$$

(the plane parallel to W passing through $(0, 0, v_3)$). Each value of $v_3 \in \mathbb{R}$ gives a different coset.

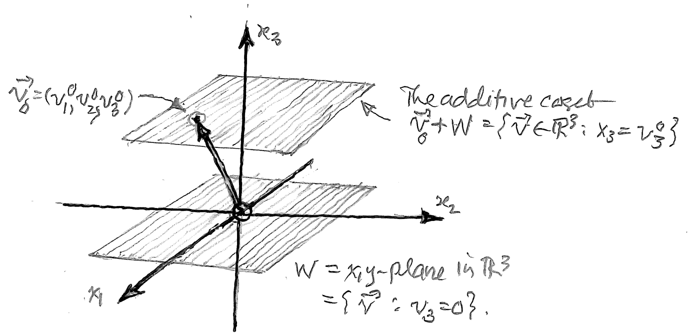
x_3
$\vec{v}_0=(v_1^0,v_2^0,v_3^0)$
The additive coset
$\vec{v}_0+W=\{\vec{v}\in\mathbb{R}^3 : x_3=v_3^0\}$
x_2
x_1
$W=x_1y$-plane in $\mathbb{R}^3$
$=\{\vec{v} : v_3=0\}$.

Figure 1.7. Additive cosets of $W = \{\mathbf{v} \in \mathbb{R}^3 : v_3 = 0\}$ are planes parallel to W in $\mathbb{R}^3$. A typical coset $\mathbf{v}_0 + W$ is shown.

One important viewpoint is to think of the quotient map $\pi : v \rightarrow V/W$ as "erasing" inessential aspects of the original vector space, retaining only those relevant to the problem at hand. Whole "bunches" of vectors in V, the cosets $v+W$, collapse to single points in the target space V/W (the planes in the last example become points in V/W). A lot of detail is lost in this collapse, but if W is suitably chosen the quotient map space will retain information that is buried in a lot of superfluous detail when we look at what is happening in the larger space V. We will soon give many examples of this, once we start looking at the structure of "linear operators" between vector spaces. For the moment we assemble a few more basic facts about quotients of vector spaces.

5.7. Theorem (Dimension Theorem for Quotients). *If V is finite dimensional and W is a subspace. Then:*

1. $\dim(V/W) \leq \dim(V) < \infty$;
2. $\dim(W) \leq \dim(V) < \infty$;

and

$$\dim(V) = \dim(W) + \dim(V/W) \tag{6}$$

By our notational conventions this identity can also be written in the abbreviated form $|V| = |W| + |V/W|$.

Proof: The quotient map $\pi : V \rightarrow V/W$ preserves linear combinations in the sense that

$$\pi\Big(\sum_{i=1}^{m} \lambda_i v_i\Big) = \sum_{i=1}^{m} \lambda_i \pi(v_i).$$

(recall Exercise 5.5), so if vectors $\{v_i\}$ span V their images $\overline{v}_i = \pi(v_i)$ span V/W. That proves

$$\dim(V/W) \leq \#\{\overline{v}_i\} \leq \#\{v_i\} = \dim(V) < \infty$$

as claimed in (1.).

As for item (2.), we know $\dim(V) < \infty$ but have no *a priori* information about W, but we showed earlier that no independent set in V can have more than $\dim(V)$ elements, and a basis for W would be such a set.

The identity (6) is proved by constructing a basis in V/W aligned with a specially chosen basis in V. Since $\dim(W) < \infty$ there is a basis $\{w_1, ..., w_m\}$ in W. If $W = V$ then

V/W is trivial and there is nothing more to do, but otherwise we can find an "outside vector" $v_{m+1} \notin W$ such that the larger set $\{w_1, ..., w_m, v_{m+1}\}$ is independent, and hence a basis for

$$W_1 = \mathbb{K}\text{-span}\{w_1, ..., w_m, v_{m+1}\} \underset{\neq}{\supseteq} W_0 = W.$$

If $W_1 \neq V$, we can adjoin one more vector $v_{m+2} \notin W_1$ to get an independent set $\{w_1, ..., w_m, v_{m+1}, v_{m+2}\}$ with

$$W_0 \underset{\neq}{\subseteq} W_1 \underset{\neq}{\subseteq} W_2 = \mathbb{K}\text{-span}\{w_1, ..., w_m, v_{m+1}, v_{m+2}\}$$

This process must terminate, otherwise we would have arbitrary large independent sets in the finite dimensional space V. When the construction terminates we get an independent spanning set $\{w_1, ..., w_m, v_{m+1}, ..., v_{m+k}\}$ in $W_k = V$. This is a basis for V so $\dim(V) = m + k = \dim(W) + k$.

To conclude the proof we demonstrate that the $k = \dim(V/W)$ by showing that the π-images $\bar{v}_{m+1}, \ldots, \bar{v}_{m+k} \in V/W$ of the "outside vectors" are a basis for V/W. Since π is surjective the images $\pi(w_1), \ldots, \pi(v_{m+k})$ span V/W. But π "kills" all vectors in W, so

$$\pi(w_1) = \ldots = \pi(w_m) = [0] \quad \text{in } V/W\ ,$$

and the remaining images $\bar{v}_{k+i} = \pi(v_{m+i})$ span V/W. They are also linearly independent. In fact, if some linear combination $\sum_{i=1}^{k} c_{m+i}\bar{v}_{m+i} = [0]$ in V/W, then by linearity of the quotient map π we get

$$[0] = \sum_{j=1}^{k} c_{m+j}\pi(v_{m+j}) = \pi\Big(\sum_{j=1}^{k} c_{m+j}v_{m+j}\Big)$$

But $\pi(v) = [0]$ for a vector $v \in V \Leftrightarrow [v] = v + W$ is equal to the zero coset $[0] = W$. Furthermore $v + W = W \Leftrightarrow v \in W$, so we can find coefficients $c_1, \ldots, c_m$ such that

$$\sum_{i=1}^{m} c_i w_i \ = \ v \ = \ \sum_{j=1}^{k} c_{m+j}v_{m+j}$$

or

$$0 = \sum_{i=1}^{m} c_i w_i + \sum_{j=1}^{k} (-1)c_{m+j}v_{m+j} \qquad \text{in } V.$$

Since $w_1, \ldots, w_m, v_{m+1}, \ldots v_{m+k}$ is a basis for V this can only happen if all coefficients in this sum are zero, and in particular $c_{m+1}, \ldots, c_{m+k} = 0$. Thus the $\{\bar{v}_i\}$ are independent and a basis for V/W, and $\dim(V/W) = k = \dim(V) - \dim(W)$. $\square$

Remark: The construction developed in proving Theorem 5.7 shows how to find bases in a quotient space V/W, and perform effective calculations with them. The key was to find representatives v_i back in V so we can transfer calculations involving cosets in V/W to calculations in V involving actual vectors v_i. The proof of Theorem 5.7 describes an explicit procedure for finding independent vectors $\{v_i\}$ outside of W, whose images $\pi(v_i) = \overline{v}_i$ are the desired basis in the quotient space.

5.8. Exercise. Find explicit bases for the following quotient spaces

1. $V = \mathbb{R}^3$, $W = \mathbb{R}\mathbf{e}_1 + \mathbb{R}\mathbf{e}_2$.
2. $V = \mathbb{R}^3$, $W = \mathbb{R}\text{-span}\{\mathbf{w}_1 = (1, 2, 3), \mathbf{w}_2 = (0, 1, -1)\}$;
3. $V = \mathbb{C}^4$, $W = \mathbb{C}\text{-span}\{\mathbf{z}_1 = (1\,, 1+i\,, 3-2i\,, -i),\ \mathbf{z}_2 = (4-i\,, 0\,, -1\,, 1+i)\}$;

4. $V = \mathbb{R}^4$, $W = \{\mathbf{x} : x_1 + x_2 - x_3 + x_4 = 0 \text{ and } 4x_1 - 3x_2 + 2x_3 + x4 = 0\}$.

Here is a simple example involving bases in a quotient space V/W.

5.9. Example. Let $V = \mathbb{R}^4$ and $W = \{x \in \mathbb{R} : 2x_1 - x_2 + x_4 = 0\}$. The subspace W is the solution set of the matrix equation

$$A\mathbf{x} = \mathbf{0} \qquad \text{where} \qquad A = [\,2, -1, 0, 1\,]_{1\times 4}$$

that imposes a single linear constraint on $\mathbb{R}^4$. Find a basis for V/W

Solution: Row operations yield

$$A \rightarrow A' = \left[\ \boxed{1}\ , -\tfrac{1}{2}\ , 0\ , \tfrac{1}{2}\ \right]$$

The free variable are x_2, x_3, x_4 and $x_1 = \frac{1}{2}x_2 - \frac{1}{2}x_4$, so the solutions have the form

$$\mathbf{x} = \begin{pmatrix} \frac{1}{2}x_2 - \frac{1}{2}x_4 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = x_2 \begin{pmatrix} \frac{1}{2} \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -\frac{1}{2} \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

for $x_2, x_3, x_4 \in \mathbb{R}$. Thus the solution set for $A\mathbf{x} = \mathbf{0}$ is the linear span of the column vectors

$$\mathbf{u}_1 = \text{col}(1, 2, 0, 0) \qquad \mathbf{u}_2 = \text{col}(0, 0, 1, 0) \qquad \mathbf{u}_3 = \text{col}(-1, 0, 0, 2)$$

These are a basis for W since they are easily seen to be linearly independent. Just row reduce the 3×4 matrix M that has these vectors as its rows

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

and see if you get a row of zeros; you do not. Therefore $\dim(\text{Row}(M)) = 3$ and the vectors are independent.

Since $\dim(V) = \dim(W) + \dim(V/W)$ and $\dim(W) = 3$, we need only find one "outside" vector $\mathbf{u}_4 \notin W$ to complete a basis for $V = \mathbb{R}^4$; then $\pi(\mathbf{u}_4) = \mathbf{u}_4 + W$ will be nonzero, and a basis vector for the 1-dimensional quotient space. The vector $\mathbf{u}_4 = \mathbf{e}_4 = (0, 0, 0, 1)$ is not in W because it fails to satisfy the constraint equation $2x_1 - x_2 + x_4 = 0$. Thus the single vector $[\mathbf{e}_4] = \pi(\mathbf{e}_4) = \mathbf{e}_4 + W$ is a basis for V/W, and $\dim(V/W) = 1$. $\square$

5.10. Exercise (Another Dimension Formula). If E, F are subspaces in a finite-dimensional vector space V and $E + F = \{e + f : e \in E, f \in F\}$ is their linear span, prove that

$$\dim(E + F) = \dim(E) + \dim(F) - \dim(E \cap F)$$

Hint: Choose appropriate bases related to E, F and $E \cap F$.

Appendix A: The Degree Formula for $\mathbb{K}[x_1, \ldots x_N]$.

Let $\mathbb{K}[\mathbf{x}] = \mathbb{K}[x_1, \ldots, x_N]$ be the unital ring of polynomials with coefficients in an integral domain. Using the multi-index notation introduced in Section 10.1 we can write any such polynomial as a finite sum (finitely many nonzero coefficients)

$$f(\mathbf{x}) = \sum_{\alpha \in \mathbb{Z}_+^N} a_\alpha \, x^\alpha \qquad (x^\alpha = x_1^{\alpha_1} \cdot \ldots \cdot x_N^{\alpha_N}, \; c_\alpha \in R) \tag{7}$$

The degree of a monomial x^α is $|\alpha| = \alpha_1 + \ldots + \alpha_N$ and if $f \in \mathbb{K}[\mathbf{x}]$ is not the zero polynomial (all $a_\alpha = 0$) its degree is

$$m = \deg(f) = \max\{\, |\alpha| : c_\alpha \neq 0\}$$

When $N > 1$ there may be several different monomials x^α of the same total degree $|\alpha| = m$ with nonzero coefficients.

Let $f, g \neq 0$ in $\mathbb{K}[\mathbf{x}]$ with degrees $m = \deg(f), n = \deg(g)$. Their product is

$$\begin{aligned} (f \cdot g)(\mathbf{x}) &= \Big(\sum_\alpha a_\alpha \, x^\alpha\Big) \cdot \Big(\sum_\beta b_\beta \, x^\beta\Big) = \sum_{\alpha,\beta} a_\alpha b_\beta \, x^{\alpha+\beta} \\ &= \sum_\gamma \Big(\sum_{\alpha+\beta=\gamma} a_\alpha b_\beta \Big) \, x^\gamma = \sum_\gamma c_\gamma \, x^\gamma \end{aligned} \tag{8}$$

where $\alpha + \beta = (\alpha_1 + \beta_1, \ldots, \alpha_N + \beta_n)$. If $a_\alpha b_\beta x^{\alpha+\beta} \neq 0$ in (36) we must have $|\alpha| \leq m$ and $|\beta| \leq n$, so that $|\alpha + \beta| \leq m + n$; consequently $\deg(f \cdot g) \leq \deg(f) + \deg(g)$.

Let us split off the monomials of maximum degree, writing

$$\begin{aligned} f(\mathbf{x}) &= \sum_{|\alpha|=m} a_\alpha \, x^\alpha + (\cdots) \\ g(\mathbf{x}) &= \sum_{|\beta|=n} b_\beta \, x^\beta + (\cdots) \\ (f \cdot g)(\mathbf{x}) &= \sum_{|\gamma|=m+n} c_\gamma \, x^\gamma + (\cdots) \end{aligned}$$

where $(\cdots)$ are terms of lower degree. To prove the degree formula

$$\text{(9)} \quad \text{Degree Formula:} \qquad \deg(f \cdot g) = \deg(f) + \deg(g) \qquad \text{for } f, g \neq 0 \text{ in } \mathbb{K}[\mathbf{x}]$$

it suffice to show there is at least one monomial x^{γ_0} of maximal degree $m + n$ such that the coefficient

$$c_{\gamma_0} = \sum_{\alpha+\beta=\gamma_0} a_\alpha b_\beta \qquad \text{is nonzero.} \tag{10}$$

This is trivial for $N = 1$, but problematic when $N \geq 2$ because this sum of products can be zero if there is more than one term, even if the individual terms are nonzero. On the other hand the degree formula (37) follows immediately if we can prove

(11) *There exists some monomial x^γ of maximal degree $m + n$ for which the sum* (38) *consists of a single nonzero term.*

The key to proving (39) is to introduce a ranking of the monomials x^γ, $\gamma \in \mathbb{Z}_+^N$, more refined than ranking by total degree $\deg(x^\gamma) = |\gamma|$, which cannot distinguish between the various monomials of the same degree. The tool for doing this is "**lexicographic**," or "**lexical**," ordering of the indices in $\mathbb{Z}_+^N$, an idea that has proved useful in many parts of mathematics.

A.1. Definition (Lexicographic Order). *For $\alpha, \beta \in \mathbb{Z}_+^N$ we define the relation $\alpha \succ \beta$ to mean*

$$\alpha_i > \beta_i \text{ at the first index } i = 1, 2, \ldots, N \text{ at which } \alpha_i \text{ differs from } \beta_i$$

Thus

$$\alpha_1 = \beta_1, \ldots, \alpha_{i-1} = \beta_{i-1} \text{ and } \alpha_i > \beta_i \qquad (\text{other entries in } \alpha, \beta \text{ are irrelevant})$$

This is a **linear ordering** *of multi-indices: given α, β exactly one of the possibilities*

$$\alpha \succ \beta \qquad \alpha = \beta \qquad \beta \succ \alpha$$

holds. We write $\alpha \succeq \beta$ when the possibility $\alpha = \beta$ is allowed. □

Obviously $\alpha = (0, \ldots, 0)$ is the lowest multi-index in lexicographic order, and any finite set of multi-indices has a unique highest element. Note carefully that $\alpha \succ \beta$ does not imply that $|\alpha| \geq |\beta|$. For instance we have

$$\alpha = (1, 0, 0) \succ \beta = (0, 2, 2) \text{ in lexicographic order, but } |\beta| = 4 > |\alpha| = 1\,.$$

Other elementary properties of lexicographic order are easily verified once you understand the definitions.

A.2. Exercise. For lexicographic order in $\mathbb{Z}_+^N$ verify that

1. LINEAR ORDERING. For any pair α, β we have exactly one of the possibilities $\alpha \succ \beta$, $\alpha = \beta$, $\beta \succ \alpha$.

2. TRANSITIVITY OF ORDER. If $\alpha \succ \beta$ and $\beta \succ \gamma$ then $\alpha \succ \gamma$.

3. If $\alpha \succ \alpha'$ then $\alpha + \beta \succ \alpha' + \beta$ for all indices β.

4. If $\alpha \succ \alpha'$ and $\beta \succ \beta'$ then $\alpha + \beta \succ \alpha' + \beta'$.

HINT: It might help to make diagrams showing how the various N-tuples are related. You will have to do some "casework" in (3.) □

We now outline how the crucial fact (39) might be proved, leaving the final details as an exercise for the reader. If $f \neq 0$ with $m = \deg(f)$, so $f = \sum_{|\alpha| \leq m} a_\alpha \, x^\alpha$, there may be several monomials having maximal degree m with $a_\alpha \neq 0$, but just one of these is maximal with respect to lexicographic order, namely

$$\alpha_0 = \max_{\succ}\{\alpha : |\alpha| = m \text{ and } a_\alpha \neq 0\}$$

Likewise there is a unique index

$$\beta_0 = \max_{\succ}\{\beta : |\beta| = n \text{ and } b_\beta \neq 0\}$$

The multi-index $\gamma_0 = \alpha_0 + \beta_0$ has $|\gamma_0| = m + n$, and is a likely candidate for the solution to (39); note that $a_{\alpha_0} b_{\beta_0} \neq 0$ *by definition.* We leave the reader to verify a few simple properties of this particular multi-index.

A.3. Exercise. Explain why $\alpha_0 = \max_{\succ}\{\alpha : |\alpha| = m \text{ and } a_\alpha \neq 0\}$ might not be the same as $\alpha_0' = \max_{\succ}\{\alpha : a_\alpha \neq 0\}$. Is there any reason to expect α_0' to have maximal degree

$|\alpha_0'| = m$? □

A.4. Exercise. In $\gamma_0 = \alpha_0 + \beta_0$ we have $|\alpha_0| = m$ and $|\beta_0| = n$, and $a_{\alpha_0} b_{\beta_0} \neq 0$, by definition. If α, β are any indices such that

$$|\alpha + \beta| = |\alpha_0 + \beta_0| = m + n \qquad \text{and} \qquad a_\alpha b_\beta \neq 0$$

prove that we must have $|\alpha| = |\alpha_0| = m$ and $|\beta| = |\beta_0| = n$. □

Defining $\alpha_0, \beta_0, \gamma_0 = \alpha_0 + \beta_0$ as above, we make the following claim:

> Claim: If $\alpha + \beta = \alpha_0 + \beta_0$ and $a_\alpha b_\beta \neq 0$ then $\alpha = \alpha_0$ and $\beta = \beta_0$. Hence the sum
>
> $$c_{\gamma_0} = \sum_{\alpha+\beta=\gamma_0} a_\alpha b_\beta \tag{$A.1$}$$
>
> reduces to the single nonzero term $a_{\gamma_0} b_{\beta_0}$

A.5. Exercise. Prove the claim made in (A.1) using the facts assembled in the preceding discussion. □

That will complete the proof of the Degree Formula.