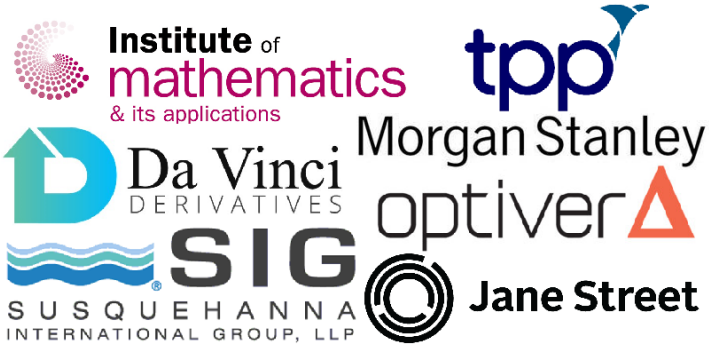
WMS



MA251

Algebra I: Advanced Linear Algebra Revision Guide

Written by David McCormick

Contents

Introduction

This revision guide for MA251 Algebra I: Advanced Linear Algebra has been designed as an aid to revision, not a substitute for it. While it may seem that the module is impenetrably hard, there's nothing in Algebra I to be scared of. The underlying theme is normal forms for matrices, and so while there is some theory you have to learn, most of the module is about doing computations. (After all, this is mathematics, not philosophy.)

Finding books for this module is hard. My personal favourite book on linear algebra is sadly out-of-print and bizarrely not in the library, but if you can nd a copy of Evar Nering's *\Linear Algebra and Matrix Theory"* then it's well worth it (though it doesn't cover the abelian groups section of the course). Three classic seminal books that cover pretty much all rst- and second-year algebra are Michael Artin's *\Algebra"*, P. M. Cohn's *\Classic Algebra"* and I. N. Herstein's *\Topics in Algebra"*, but all of them focus on the theory and not on the computation, and they often take a di erent order (for instance, most do modules before doing JCFs). Your best source of computation questions is old past paper questions, not only for the present module but also for its predecessors MA242 Algebra I and MA245 Algebra II. So practise, practise, PRACTISE, and good luck on the exam!

Disclaimer: Use at your own risk. No guarantee is made that this revision guide is accurate or complete, or that it will improve your exam performance, or that it will make you 20% cooler. Use of this guide *will* increase entropy, contributing to the heat death of the universe.

Authors

Written by D. S. McCormick (d.s.mccormick@warwick.ac.uk). Edited by C. I. Midgley (c.i.midgley@warwick.ac.uk)
Based upon lectures given by Dr. Derek Holt and Dr. Dmitri Rumynin at the University of Warwick, 2006{2008, and later Dr. David Loe er, 2011{2012.
Any corrections or improvements should be entered into our feedback form at http://tinyurl.com/WMSGuides (alternatively email revision.guides@warwickmaths.org).

1 Change of Basis

A major theme in MA106 Linear Algebra is *change of bases*. Since this is fundamental to what follows, we recall some notation and the key theorem here.

Let $T : U \to V$ be a linear map between U and V. To express T as a matrix requires picking a basis $\{\mathbf{e}_i\}$ of U and a basis $\{\mathbf{f}_j\}$ of V. To change between two bases $\{\mathbf{e}_i\}$ and $\{\mathbf{e}'_i\}$ of U, we simply take the identity map $I_U : U \to U$ and use the basis $\{\mathbf{e}'_i\}$ in the domain and the basis $\{\mathbf{e}_i\}$ in the codomain; the matrix so formed is the *change of basis matrix* from the basis of $\mathbf{e}_i$s to the $\mathbf{e}'_i$s. Any such change of basis matrix is invertible | note that this version is the *inverse* of the one learnt in MA106 Linear Algebra!

Proposition 1.1. Let $\mathbf{u} \in U$, and let $\underline{\mathbf{u}}$ and $\underline{\mathbf{u}}'$ denote the column vectors associated with $\mathbf{u}$ in the bases $\mathbf{e}_1, \ldots, \mathbf{e}_n$ and $\mathbf{e}'_1, \ldots, \mathbf{e}'_n$ respectively. Then if P is the change of basis matrix, we have $P\underline{\mathbf{u}}' = \underline{\mathbf{u}}$.

Theorem 1.2. Let A be the matrix of $T : U \to V$ with respect to the bases $\{\mathbf{e}_i\}$ of U and $\{\mathbf{f}_j\}$ of V, and let B be the matrix of T with respect to the bases $\{\mathbf{e}'_i\}$ of U and $\{\mathbf{f}'_j\}$ of V. Let P be the change of basis matrix from $\{\mathbf{e}_i\}$ to $\{\mathbf{e}'_i\}$, and let Q be the change of basis matrix from $\{\mathbf{f}_j\}$ to $\{\mathbf{f}'_j\}$. Then $B = Q^{-1}AP$.

Throughout this course we are concerned primarily with $U = V$, and $\{\mathbf{e}_i\} = \{\mathbf{f}_j\}$, $\{\mathbf{e}'_i\} = \{\mathbf{f}'_j\}$, so that $P = Q$ and hence $B = P^{-1}AP$.

In this course, the aim is to find so-called *normal forms* for matrices and their corresponding linear maps. Given a matrix, we want to know what it does in simple terms, and to be able to compare matrices that look completely different; so, *how should we change bases to get the matrix into a nice form?* There are three very different answers:

1. When working in a finite-dimensional vector space over C, we can always change bases so that T is as close to diagonal as possible, with only eigenvalues on the diagonal and possibly some 1s on the superdiagonal; this is the *Jordan Canonical Form*, discussed in section 2.
2. We may want the change of basis not just to get a matrix into a nice form, but also preserve its geometric properties. This leads to the study of *bilinear* and *quadratic forms*, and along with it the theory of *orthogonal matrices*, in section 3.
3. Alternatively, we may consider matrices with entries in Z, and try and diagonalise them; this leads, perhaps surprisingly, to a classification of all *finitely-generated abelian groups*, which (along with some basic group theory) is the subject of section 4.

2 The Jordan Canonical Form

2.1 Eigenvalues and Eigenvectors

We first recall some facts on eigenvalues and eigenvectors from MA106 Linear Algebra.

Definition 2.1. Let V be a vector space over K and let $T : V \to V$ be a linear map, with associated matrix A. If $T(\mathbf{v}) = \lambda\mathbf{v}$ for some $\lambda \in K$ and $\mathbf{v} \in V$ with $\mathbf{v} \neq \mathbf{0}$, then λ is an *eigenvalue* of T (and of A), and $\mathbf{v}$ a *corresponding eigenvector* of T (and of A). We call the subspace $\{\mathbf{v} \in V \mid T(\mathbf{v}) = \lambda\mathbf{v}\}$ the *eigenspace of T* with respect to λ.

Definition 2.2. For an $n \times n$ matrix A, $c_A(x) := \det(A - xI_n)$ is the *characteristic polynomial* of A.

Theorem 2.3. Let A be an $n \times n$ matrix. Then λ is an eigenvalue of A if and only if $\det(A - \lambda I_n) = 0$.

Recall that $n \times n$ matrices A and B are *similar* if there is an invertible $n \times n$ matrix P such that $B = P^{-1}AP$. Since similar matrices have the same characteristic equation, changing bases does not change the eigenvalues of a linear map.

You have already seen one \normal form", which occurs when the matrix has distinct eigenvalues:

Theorem 2.4. Let $T : V \to V$ be a linear map. Then the matrix of T is diagonal with respect to some basis of V if and only if V has a basis consisting of eigenvectors of T.

Theorem 2.5. Let $\lambda_1, \ldots, \lambda_r$ be distinct eigenvalues of a linear map $T: V \to V$ and let $\mathbf{v}_1, \ldots, \mathbf{v}_r$ be the corresponding eigenvectors. Then $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are linearly independent.

Corollary 2.6. If the linear map $T: V \to V$ has n distinct eigenvalues, where $\dim V = n$, then T is diagonalisable.

Of course, the converse is not true; T may be diagonalisable even though it has repeated eigenvalues.

2.2 Minimal Polynomials

We denote the set of all polynomials in a single variable x with coefficients in a field K by $K[x]$. We recall some properties of polynomials from MA132 FOUNDATIONS; these often resemble properties of $\mathbb{Z}$.

We write $a \mid b$ to mean a divides b; e.g. $(x-4) \mid (x^2-3x-4)$. Given two polynomials $p, q \neq 0$, we can *divide with remainder*, where the remainder has degree less than p. For example, if $p = x^2-4x$ and $q = x^3+2x^2+5$, then $q = sp + r$, where $s = x+6$ and $r = 24x+5$. This is the *Euclidean algorithm*.

Definition 2.7. A polynomial in $K[x]$ is called *monic* if the coefficient of the highest power of x is 1.

Definition 2.8. The *greatest common divisor* of $p, q \in K[x]$ is the unique monic polynomial r such that $r \mid p$ and $r \mid q$, and for any other polynomial r' such that $r' \mid p$ and $r' \mid q$, we have $r' \mid r$. Similarly, the *lowest common multiple* of $p, q \in K[x]$ is the unique monic polynomial r such that $p \mid r$ and $q \mid r$, and for any other polynomial r' such that $p \mid r'$ and $q \mid r'$, we have $r \mid r'$.

We first observe a very interesting fact about characteristic polynomials:

Theorem 2.9 (Cayley–Hamilton Theorem)**.** Let $c_A(x)$ be the characteristic polynomial of an $n \times n$ matrix A over an arbitrary field K. Then $c_A(A) = \mathbf{0}$.

So we know that there is at least some polynomial $p \in K[x]$ such that $p(A) = 0$. The following theorem allows us to define more:

Theorem 2.10. Let A be an $n \times n$ matrix over K representing the linear map $T: V \to V$. Then there is a unique monic non-zero polynomial $p \in K[x]$ with minimal degree such that $p(A) = 0$. Furthermore, if $q \in K[x]$ also satisfies $q(A) = 0$, then $p \mid q$.

Proof. We can assume such a polynomial is monic. The Cayley–Hamilton Theorem tells us that there is $p \in K[x]$ such that $p(A) = 0$. If there were two distinct polynomials p_1, p_2 of minimal degree s.t. $p_1(A) = p_2(A) = 0$, then $p = p_1 - p_2$ would be non-zero and of lower degree, contradicting minimality. Thus p is unique. Furthermore, suppose $q(A) = 0$ but $p \nmid q$. Then we can write $q = sp + r$, with $\deg(r) < \deg(p)$, and $r \neq 0$. But then $r(A) = q(A) - s(A)p(A) = 0$, contradicting minimality of p. □

Definition 2.11. The unique monic non-zero polynomial $\mu_A(x)$ of minimal degree with $\mu_A(A) = 0$ is called the *minimal polynomial* of A, or of the corresponding linear map T.

Combining the last two theorems we observe that $\mu_A(x)$ divides $c_A(x)$. Furthermore, similar matrices have the same minimal polynomial, so the minimal polynomial of a linear map does not depend on bases.

Similarly to above we may define $\mu_{A,\mathbf{v}}$ to be the unique monic polynomial p of minimal degree such that $p(T)(\mathbf{v}) = \mathbf{0}$; since $p(T) = 0$ if and only if $p(T)(\mathbf{v}) = \mathbf{0}$ for all $\mathbf{v} \in V$, μ_A is the least common multiple of $\mu_{A,\mathbf{v}}$ for $\mathbf{v} \in V$. In fact, we only need consider vectors in a basis of V; i.e. if $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is a basis of V then $\mu_A = \operatorname{lcm}\{\mu_{A,\mathbf{e}_i} : 1 \le i \le n\}$. This allows us to calculate μ_A: for $\mathbf{v} \in V$, we compute $\mu_{A,\mathbf{v}}$ by calculating $\mathbf{v}$, $T(\mathbf{v})$, $T^2(\mathbf{v})$, and so on until the sequence becomes linearly dependent.

Example 2.12. Let $K = \mathbb{R}$ and consider $A = \begin{pmatrix} 5 & 0 & 1 \\ 3 & 4 & 3 \\ 1 & 0 & 3 \end{pmatrix}$. Let $\mathbf{e}_1 = (1,0,0)^T$, $\mathbf{e}_2 = (0,1,0)^T$, $\mathbf{e}_3 = (0,0,1)^T$ be the standard basis of $\mathbb{R}^3$. Then:

$A\mathbf{e}_1 = (5,3,1)^T$, $A^2\mathbf{e}_1 = (24,24,8)^T = 8A\mathbf{e}_1 - 16\mathbf{e}_1$, so $(A^2 - 8A + 16)\mathbf{e}_1 = \mathbf{0}$, thus $\mu_{A,\mathbf{e}_1}(x) = x^2 - 8x + 16 = (x-4)^2$.

$A\mathbf{e}_2 = (0,4,0)^T = 4\mathbf{e}_2$, so $(A-4)\mathbf{e}_2 = \mathbf{0}$, thus $\mu_{A,\mathbf{e}_2}(x) = (x-4)$.

$A\mathbf{e}_3 = (-1,3,3)^T$, $A^2\mathbf{e}_3 = (-8,24,8)^T = 8A\mathbf{e}_3 - 16\mathbf{e}_3$, thus $\mu_{A,\mathbf{e}_3}(x) = x^2 - 8x + 16 = (x-4)^2$.

Thus $\mu_A = \operatorname{lcm}\{\mu_{A,\mathbf{e}_1}, \mu_{A,\mathbf{e}_2}, \mu_{A,\mathbf{e}_3}\} = (x-4)^2$. One may compute that $c_A(x) = \det(A - xI) = (4-x)^3$.

Lemma 2.13. $(x - \lambda)$ divides the minimal polynomial $m_A(x)$ if and only if λ is an eigenvalue of A.

Proof. Suppose $(x - \lambda) \mid m_A(x)$; then as $m_A(x) \mid c_A(x)$, we have $(x - \lambda) \mid c_A(x)$, and hence λ is an eigenvalue of A. Conversely, if λ is an eigenvalue of A then there exists $\mathbf{v} \neq 0$ such that $(A - \lambda I)\mathbf{v} = 0$, hence $m_{A,\mathbf{v}} = (x - \lambda)$, and since $m_A = \mathrm{lcm}\{ m_{A,\mathbf{v}} : \mathbf{v} \in V \}$ we have $(x - \lambda) \mid m_A(x)$. □

2.3 Jordan Chains, Jordan Blocks and Jordan Bases

We assume henceforth that $K = \mathbb{C}$, so that all polynomials in $K[x]$ factorise into linear factors (by the Fundamental Theorem of Algebra). We now seek to generalise our notions of eigenvalue and eigenvector in order to be able to find a "normal form" for a matrix with any eigenvalues, not just distinct ones.

Definition 2.14. A *Jordan chain* of length k is a sequence of non-zero vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{C}^{n,1}$ (that is, column vectors of length n with entries in $\mathbb{C}$) that satisfies

$$A\mathbf{v}_1 = \lambda\mathbf{v}_1 \quad \text{and} \quad A\mathbf{v}_i = \lambda\mathbf{v}_i + \mathbf{v}_{i-1} \quad \text{for } 2 \leq i \leq k$$

for some eigenvalue λ of A. Equivalently, $(A - \lambda I_n)\mathbf{v}_1 = 0$ and $(A - \lambda I_n)\mathbf{v}_i = \mathbf{v}_{i-1}$ for $2 \leq i \leq k$, so $(A - \lambda I_n)^i \mathbf{v}_i = \mathbf{0}$ for $1 \leq i \leq k$.

Definition 2.15. A non-zero vector $\mathbf{v} \in V$ such that $(A - \lambda I_n)^i \mathbf{v} = \mathbf{0}$ for some $i > 0$ is called a *generalised eigenvector* of A with respect to the eigenvalue λ. The set $\{\mathbf{v} \in V \mid (A - \lambda I_n)^i \mathbf{v} = \mathbf{0}\}$ is called the *generalised eigenspace* of index i of A with respect to λ; it is the nullspace of $(A - \lambda I_n)^i$. (Note that when $i = 1$, these definitions reduce to ordinary eigenvectors and eigenspaces.)

For example, consider the matrix $A = \begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{pmatrix}$. For the standard basis $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ of $\mathbb{C}^{3,1}$, we have $A\mathbf{e}_1 = 3\mathbf{e}_1$, $A\mathbf{e}_2 = 3\mathbf{e}_2 + \mathbf{e}_1$, $A\mathbf{e}_3 = 3\mathbf{e}_3 + \mathbf{e}_2$. Thus $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ is a Jordan chain of length 3 for the eigenvalue 3 of A. The generalised eigenspaces of index 1, 2 and 3 respectively are $\langle \mathbf{e}_1 \rangle$, $\langle \mathbf{e}_1, \mathbf{e}_2 \rangle$, and $\langle \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 \rangle$.

Note that the dimension of a generalised eigenspace of A is the nullity of $(T - \lambda I_V)^i$, which depends only on the linear map T associated with A; thus the dimensions of corresponding eigenspaces of similar matrices are the same.

Definition 2.16. A *Jordan block* with eigenvalue λ of degree k is the $k \times k$ matrix $J_{\lambda,k} = (a_{ij})$ where $a_{ii} = \lambda$ for $1 \leq i \leq k$, $a_{i,i+1} = 1$ for $1 \leq i < k$, and $a_{ij} = 0$ if $j \neq i, i+1$.

For example, the following are Jordan blocks:

$$J_{2,2} = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}; \qquad J_{(2-i),3} = \begin{pmatrix} 2-i & 1 & 0 \\ 0 & 2-i & 1 \\ 0 & 0 & 2-i \end{pmatrix}.$$

It is a fact that the matrix A of T with respect to the basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of $\mathbb{C}^{n,1}$ is a Jordan block of degree n if and only if $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a Jordan chain for A.

Note that the minimal polynomial of $J_{\lambda,k}$ is $m_{J_{\lambda,k}}(x) = (x - \lambda)^k$, and its characteristic polynomial is $c_{J_{\lambda,k}}(x) = (\lambda - x)^k$.

Definition 2.17. A *Jordan basis* for A is a basis of $\mathbb{C}^{n,1}$ which is a union of disjoint Jordan chains.

For an $m \times m$ matrix A and an $n \times n$ matrix B, we can form the $(m+n) \times (m+n)$ matrix $A \oplus B = \left(\begin{array}{c|c} A & \mathbf{0}_{m,n} \\ \hline \mathbf{0}_{n,m} & B \end{array}\right)$. For example,

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

Suppose A has eigenvalues $\lambda_1, \ldots, \lambda_r$, and suppose $\mathbf{w}_{i,1}, \ldots, \mathbf{w}_{i,k_i}$ is a Jordan chain for A for the eigenvalue λ_i, such that $\mathbf{w}_{1,1}, \ldots, \mathbf{w}_{1,k_1}, \mathbf{w}_{2,1}, \ldots, \mathbf{w}_{2,k_2}, \ldots, \mathbf{w}_{r,1}, \ldots, \mathbf{w}_{r,k_r}$ is a Jordan basis for A. Then

the matrix of the linear map T corresponding to A with respect to this Jordan basis is the direct sum of Jordan blocks $J_{\lambda_1;k_1} \quad J_{\lambda_2;k_2} \quad \cdots \quad J_{\lambda_r;k_r}$.

The main theorem of this section is that we can always find a Jordan matrix for any $n \times n$ matrix A over $\mathbb{C}$; the corresponding matrix which is a direct sum of the Jordan blocks is called the *Jordan canonical form* of A:

Theorem 2.18. Let A be an $n \times n$ matrix over $\mathbb{C}$. Then there exists a Jordan basis for A, and hence A is similar to a matrix J which is a direct sum of Jordan blocks. The Jordan blocks occurring in J are uniquely determined by A, so J is uniquely determined up to the order of the blocks. J is said to be the *Jordan canonical form* of A.

The proof of this theorem is hard and non-examinable. What's far more important is calculating the Jordan canonical form (JCF) of a matrix, and the matrix P whose columns are the vectors of the Jordan basis; then by theorem 1.2, we have that $P^{-1}AP = J$. For 2×2 and 3×3 matrices, the JCF of a matrix is in fact determined solely by its minimal and characteristic polynomials. In higher dimensions, we must consider the generalised eigenspaces.

2.4 Computing the Jordan Canonical Form

Suppose A is an $n \times n$ matrix with eigenvalues $\lambda_1, \ldots, \lambda_r$, and that the Jordan blocks for eigenvalue λ_i are $J_{\lambda_i;k_{i;1}}, \ldots, J_{\lambda_i;k_{i;j_i}}$, where $k_{i;1} \geq k_{i;2} \geq \cdots \geq k_{i;j_i}$. Then the characteristic polynomial of J (and hence of A) is the product of the characteristic polynomials of the Jordan blocks; thus $c_J(x) = \prod_{i=1}^{r} (\lambda_i - x)^{k_i}$, where $k_i = k_{i;1} + \cdots + k_{i;j_i}$; i.e. each $(\lambda_i - x)$ occurs raised to the power of the sum of the sizes of the Jordan blocks of that eigenvalue.

The minimal polynomial of J (hence of A) is the least common multiple of the minimal polynomials of the Jordan blocks; since we have arranged them in descending order of size, $m_J(x) = \prod_{i=1}^{r} (x - \lambda_i)^{k_{i;1}}$; i.e. each $(x - \lambda_i)$ occurs raised to the power of the biggest Jordan block for that eigenvalue.

In 2 and 3 dimensions, this restricts the possible Jordan blocks enough to determine the JCF solely by looking at the minimal and characteristic polynomials. We must then determine the Jordan basis; note that it is often easier to find the vectors in a Jordan chain in reverse order.

2.4.1 2×2 Matrices

For a 2×2 matrix, there are two possibilities for its characteristic polynomial; it must either have two distinct roots, e.g. $(\lambda_1 - x)(\lambda_2 - x)$, or it must have one repeated root, $(\lambda_1 - x)^2$. By lemma 2.13, in the first case the minimal polynomial must be $(x - \lambda_1)(x - \lambda_2)$, and the only possibility is one Jordan block for each eigenvalue of size 1. (This accords with corollary 2.6.) In the second case, we can have $m_A(x) = (\lambda_1 - x)$ or $m_A(x) = (\lambda_1 - x)^2$, which correspond to two Jordan blocks of size 1 and one Jordan block of size 2 for the only eigenvalue. (In fact, when we have two Jordan blocks of size 1 for the same eigenvalue, the JCF is just a scalar matrix $J = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_1 \end{pmatrix}$ which commutes with all matrices, thus $A = PJP^{-1} = J$, i.e. A is its own JCF.) Table 1 summarises the possibilities.

Characteristic Polynomial	Minimal Polynomial	Jordan Canonical Form
$(\lambda_1 - x)(\lambda_2 - x)$	$(x - \lambda_1)(x - \lambda_2)$	$J_{\lambda_1;1} \quad J_{\lambda_2;1} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$
$(\lambda_1 - x)^2$	$(x - \lambda_1)^2$	$J_{\lambda_1;2} = \begin{pmatrix} \lambda_1 & 1 \\ 0 & \lambda_1 \end{pmatrix}$
	$(x - \lambda_1)$	$J_{\lambda_1;1} \quad J_{\lambda_1;1} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_1 \end{pmatrix}$

Table 1: The possible JCFs of a 2×2 matrix.

Example 2.19. $A = \begin{pmatrix} 2 & 4 \\ 5 & 3 \end{pmatrix}$ has characteristic polynomial $c_A(x) = (x + 2)(x - 7)$, so A has eigenvalues -2 and 7, and thus the JCF is $J = \begin{pmatrix} -2 & 0 \\ 0 & 7 \end{pmatrix}$. An eigenvector for the eigenvalue -2 is $(1, -1)^T$, and an eigenvector for the eigenvalue 7 is $(4, 5)^T$; setting $P = \begin{pmatrix} 1 & 4 \\ -1 & 5 \end{pmatrix}$, we may calculate that $P^{-1}AP = J$.

Example 2.20. $A = \begin{pmatrix} 2 & 1 \\ -1 & 4 \end{pmatrix}$ has $c_A(x) = (3 - x)^2$, and one may calculate $m_A(x) = (x - 3)^2$. Thus its JCF is $J = \begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix}$. To find a Jordan basis we choose any $\mathbf{v}_2$ such that $(A - 3I)\mathbf{v}_2 \neq 0$, and then choose

$\mathbf{v}_1 = (A \quad 3I)\mathbf{v}_2$; for example $v_2 = (0;1)^T$ has $(A \quad 3I)\mathbf{v}_2 = (1;1)^T$, so set $\mathbf{v}_1 = (1;1)^T$; then $A\mathbf{v}_1 = 3\mathbf{v}_1$ and $A\mathbf{v}_2 = 3\mathbf{v}_2 + \mathbf{v}_1$ as required. Thus setting $P = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$, we may calculate that $P \quad {}^1AP = J$.

2.4.2 3 3 Matrices

For 3 3 matrices, we can do the same kind of case analysis that we did for 2 2 matrices. It is a very good test of understanding to go through and derive all the possibilities for yourself, so DO IT NOW! Once you have done so, turn to the next page and check the results in table 2.

Characteristic Polynomial	Minimal Polynomial	Jordan Canonical Form
$(\ _1 \quad x)(\ _2 \quad x)(\ _3 \quad x)$	$(x \quad _1)(x \quad _2)(x \quad _3)$	$J \ _{1;1} \quad J \ _{2;1} \quad J \ _{3;1} = \left(\begin{smallmatrix} {}_1 & 0 & 0 \\ 0 & {}_2 & 0 \\ 0 & 0 & {}_3 \end{smallmatrix}\right)$
$(\ _1 \quad x)^2(\ _2 \quad x)$	$(x \quad _1)^2(x \quad _2)$	$J \ _{1;2} \quad J \ _{2;1} = \left(\begin{smallmatrix} {}_1 & 1 & 0 \\ 0 & {}_1 & 0 \\ 0 & 0 & {}_2 \end{smallmatrix}\right)$
	$(x \quad _1)(x \quad _2)$	$J \ _{1;1} \quad J \ _{1;1} \quad J \ _{2;1} = \left(\begin{smallmatrix} {}_1 & 0 & 0 \\ 0 & {}_1 & 0 \\ 0 & 0 & {}_2 \end{smallmatrix}\right)$
$(\ _1 \quad x)^3$	$(x \quad _1)^3$	$J \ _{1;3} = \left(\begin{smallmatrix} {}_1 & 1 & 0 \\ 0 & {}_1 & 1 \\ 0 & 0 & {}_1 \end{smallmatrix}\right)$
	$(x \quad _1)^2$	$J \ _{1;2} \quad J \ _{1;1} = \left(\begin{smallmatrix} {}_1 & 1 & 0 \\ 0 & {}_1 & 0 \\ 0 & 0 & {}_1 \end{smallmatrix}\right)$
	$(x \quad _1)$	$J \ _{1;1} \quad J \ _{1;1} \quad J \ _{1;1} = \left(\begin{smallmatrix} {}_1 & 0 & 0 \\ 0 & {}_1 & 0 \\ 0 & 0 & {}_1 \end{smallmatrix}\right)$

Table 2: The possible JCFs of a 3 3 matrix.

Example 2.21. Consider $A = \left(\begin{smallmatrix} 5 & 0 & 1 \\ 3 & 4 & 3 \\ 1 & 0 & 3 \end{smallmatrix}\right)$. We previously calculated $c_A(x) = (4 \quad x)^3$, $_A(x) = (x \quad 4)^2$. This tells us that the JCF of A is $J = \left(\begin{smallmatrix} 4 & 1 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{smallmatrix}\right)$. There are two Jordan chains, one of length 2 and one of length 1. For the first we need $\mathbf{v}_1;\mathbf{v}_2$ such that $(A \quad 4I)\mathbf{v}_2 = \mathbf{v}_1$, and $(A \quad 4I)\mathbf{v}_1 = \mathbf{0}$. We calculate $A \quad 4I = \left(\begin{smallmatrix} 1 & 0 & 1 \\ 3 & 0 & 3 \\ 1 & 0 & 1 \end{smallmatrix}\right)$. The minimal polynomial tells us that $(A \quad 4I)^2\mathbf{v} = \mathbf{0}$ for all $\mathbf{v} \ 2 \ V$, so we can choose whatever we like for $\mathbf{v}_2$; say $\mathbf{v}_2 = (1;0;0)^T$; then $\mathbf{v}_1 = (A \quad 4I)\mathbf{v}_2 = (1;3;1)^T$. For the second chain we need an eigenvector $\mathbf{v}_3$ which is linearly independent of $\mathbf{v}_1$; $\mathbf{v}_3 = (1;0;1)^T$ is as good as any. Setting $P = \left(\begin{smallmatrix} 1 & 1 & 1 \\ 3 & 0 & 0 \\ 1 & 0 & 1 \end{smallmatrix}\right)$ we find $J = P \quad {}^1AP$.

Example 2.22. Consider $A = \left(\begin{smallmatrix} 4 & 1 & 1 \\ 4 & 9 & 4 \\ 7 & 10 & 4 \end{smallmatrix}\right)$. One may tediously compute that $c_A(x) = (3 \quad x)^3$, and that

$$A \quad 3I = \begin{pmatrix} 1 & 1 & 1 \\ 4 & 6 & 4 \\ 7 & 10 & 7 \end{pmatrix}; \qquad (A \quad 3I)^2 = \begin{pmatrix} 2 & 3 & 2 \\ 0 & 0 & 0 \\ 2 & 3 & 2 \end{pmatrix}; \qquad (A \quad 3I)^3 = \mathbf{0}:$$

Thus $_A(x) = (x \quad 3)^3$. Thus we have one Jordan chain of length 3; that is, we need nonzero vectors $\mathbf{v}_1;\mathbf{v}_2;\mathbf{v}_3$ such that $(A \quad 3I)\mathbf{v}_3 = \mathbf{v}_2$, $(A \quad 3I)\mathbf{v}_2 = \mathbf{v}_1$, and $(A \quad 3I)\mathbf{v}_1 = \mathbf{0}$. For $\mathbf{v}_3$, we need $(A \quad 3I)\mathbf{v}_3$ and $(A \quad 3I)^2\mathbf{v}_3$ to be nonzero; we may choose $\mathbf{v}_3 = (1;1;0)^T$; we can then compute $\mathbf{v}_2 = (0;2; \quad 3)^T$ and $\mathbf{v}_1 = (1;0;1)^T$. Putting $P = \left(\begin{smallmatrix} 1 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 3 & 0 \end{smallmatrix}\right)$, we obtain $P \quad {}^1AP = \left(\begin{smallmatrix} 3 & 1 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{smallmatrix}\right)$.

2.4.3 Higher Dimensions: The General Case

For dimensions higher than 3, the characteristic polynomial and minimal polynomial do not always determine the JCF uniquely. In 4 dimensions, for example, $J \ _{;2} \quad J \ _{;2}$ and $J \ _{;2} \quad J \ _{;1} \quad J \ _{;1}$ both have $c_A(x) = (\quad x)^4$ and $_A(x) = (x \quad)^2$. In general, we can compute the JCF from the dimensions of the generalised eigenspaces, as follows:

Theorem 2.23. Let be an eigenvalue of A and let J be the JCF of A. Then:
(i) The number of Jordan blocks of J with eigenvalue is equal to nullity$(A \quad I_n)$.

(ii) More generally, for $i > 0$, the number of Jordan blocks of J with eigenvalue λ and degree at least i is equal to $\text{nullity}((A - \lambda I_n)^i) - \text{nullity}((A - \lambda I_n)^{i-1})$.

(Recall that $\text{nullity}(T) = \dim(\ker(T))$.) The proof of this need not be learnt, but the theorem is vital as a tool for calculating JCFs, as the following example shows.

Example 2.24. Let $A = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 \\ 4 & 1 & 3 & 2 & 1 \\ 2 & 1 & 0 & 1 & 1 \\ 3 & 1 & 3 & 4 & 1 \\ 8 & 2 & 7 & 5 & 4 \end{pmatrix}$. One may tediously compute that $c_A(x) = (2-x)^5$, and that

$$A - 2I = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 \\ 4 & 1 & 3 & 2 & 1 \\ 2 & 1 & 2 & 1 & 1 \\ 3 & 1 & 3 & 2 & 1 \\ 8 & 2 & 7 & 5 & 2 \end{pmatrix}; \qquad (A - 2I)^2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \end{pmatrix}; \qquad (A - 2I)^3 = \mathbf{0}.$$

This gives that $\mu_A(x) = (x-2)^3$. Let $\mathbf{r}_j$ denote the j^{th} row of $(A - 2I)$; then one may observe that $\mathbf{r}_4 = \mathbf{r}_1 + \mathbf{r}_3$, and $\mathbf{r}_5 = 2\mathbf{r}_1 + \mathbf{r}_2 + \mathbf{r}_3$, but that $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ are linearly independent, so $\text{rank}(A - 2I) = 3$, and thus by the dimension theorem $\text{nullity}(A - 2I) = 5 - 3 = 2$. Thus there are two Jordan blocks for eigenvalue 2. Furthermore, it is clear that $\text{rank}(A - 2I)^2 = 1$ and hence $\text{nullity}(A - 2I)^2 = 4$, so there are $4 - 2 = 2$ blocks of size at least 2. As $\text{nullity}(A - 2I)^3 = 5$, we have $5 - 4 = 1$ block of size at least 3. Since the largest block has size 3 (by the minimal polynomial), we now know that there are two Jordan blocks, one of size 3 and one of size 2.

To find the Jordan chains, we need $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_5$ such that

$$(A - 2I)\mathbf{v}_5 = \mathbf{v}_4, \quad (A - 2I)\mathbf{v}_4 = \mathbf{0}, \qquad (A - 2I)\mathbf{v}_3 = \mathbf{v}_2, \quad (A - 2I)\mathbf{v}_2 = \mathbf{v}_1, \quad (A - 2I)\mathbf{v}_1 = \mathbf{0}.$$

For the chain of length 3, we may choose $\mathbf{v}_3 = (0, 0, 0, 1, 0)^T$, since then $\mathbf{v}_2 = (A - 2I)\mathbf{v}_3 = (1, 2, 1, 2, 5)^T \neq \mathbf{0}$ and $\mathbf{v}_1 = (A - 2I)^2\mathbf{v}_3 = (0, 0, 1, 1, 1)^T \neq \mathbf{0}$. For the chain of length 2, we must choose $\mathbf{v}_5$ so that $\mathbf{v}_4 = (A - 2I)\mathbf{v}_5 \neq \mathbf{0}$, but so that $(A - 2I)^2\mathbf{v}_5 = \mathbf{0}$, *and* so that all the $\mathbf{v}_i$ are linearly independent. In general there is no easy way of doing this; we choose $\mathbf{v}_5 = (-1, 0, 1, 0, 0)^T$, so that $\mathbf{v}_4 = (A - 2I)\mathbf{v}_5 = (0, 1, 0, 0, 1)^T$. Then, setting $P = \begin{pmatrix} 0 & 1 & 0 & 0 & -1 \\ 0 & 2 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 2 & 1 & 0 & 0 \\ 1 & 5 & 0 & 1 & 0 \end{pmatrix}$, we find $J = P^{-1}AP = \begin{pmatrix} 2 & 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}$.

2.5 Exponentiation of a Matrix

In this section, we define e^A where A is a matrix.

Definition 2.25. If $A \in \mathbb{C}^{n,n}$, we define e^A to be the infinite series

$$e^A = I_n + A + \frac{A^2}{2} + \frac{A^3}{6} + \cdots = \sum_{k=1}^{\infty} \frac{A^k}{k!}$$

Warning: It is not in general true that $e^{A+B} = e^A e^B$ — though this does hold if $AB = BA$.

Lemma 2.26. 1. Let A and $B \in \mathbb{C}^{n,n}$ be similar, so $B = P^{-1}AP$ for some invertible matrix P. Then $e^B = P^{-1}e^A P$.

2. $\frac{\mathrm{d}}{\mathrm{d}t}e^{At} = Ae^{At}$

The first point on this lemma gives us a hint as to how we might compute the exponential of a matrix — using the Jordan form! Given $A = A_1 \oplus \cdots \oplus A_n$, we have that $e^A = e^{A_1} \oplus \cdots \oplus e^{A_n}$, so it will suffice to consider exponentiation of a single Jordan block.

Theorem 2.27. If $J = J_{\lambda,s}$ is a Jordan block, then e^{Jt} is the matrix whose (i, j) entry is given by

$$\begin{cases} \frac{t^{j-i}e^{\lambda t}}{(j-i)!} & \text{if } j \geq i \\ 0 & \text{if } j < i \end{cases}$$

Example 2.28. Given $J = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, we have $e^{Jt} = \begin{pmatrix} e^{2t} & te^{2t} & 0 \\ 0 & e^{2t} & 0 \\ 0 & 0 & e^t \end{pmatrix}$

We can use this to compute the solution to di erential equations. Check your notes for MA133 DIFFERENTIAL EQUATIONS for solution methods, remembering that we can now take exponents and powers of matrices directly. We can also use a slight variation on this method to nd the solution to di erence equations, using matrix powers.

2.6 Powers of a Matrix

Na•vely, we can use a similar strategy to exponentiation, using the Jordan canonical form. Observe that if $A = PJP^{-1}$, where J is the Jordan form of A, then $A^n = PJ^nP^{-1}$. Again, it su ces to consider only a single Jordan block:

Theorem 2.29. If $J = J_{;S}$ is a Jordan block, then J^n is the matrix whose $(i;j)$ entry is given by

$$\begin{cases} \binom{n}{j-i} \ n^{\ (j-i)} \\ 0 \qquad\qquad\qquad \text{if } j < i \end{cases}$$

where $\binom{n}{k}$ is the binomial coe cient $\frac{n!}{k!(n-k)!}$,

Example 2.30. Given $J = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, we have $J^n = \begin{pmatrix} 2^n & n2^{n-1} & 0 \\ 0 & 2^n & 0 \\ 0 & 0 & 1 \end{pmatrix}$

However, if we do not have the Jordan form close to hand, we could be in for a long and annoying computation. Fortunately, we can also nd matrix powers using the Lagrange interpolation polynomial of z^n. Suppose that we know of an equation f that kills A | that is, $f(A) = 0$. The characteristic or minimal polynomials are both good ts. Then dividing z^n by $f(z)$ with remainder gives

$$z^n = f(z)g(z) \quad h(z)$$

which implies that $A^n = h(A)$.

If we know the roots of $f(z)$ (and we likely will, if it's the characteristic or minimal polynomial) we can nd h more easily than simply doing the division. Let f have roots $_1;:::; \ _k$ with multiplicities $m_1;:::;m_k$ respectively. Then h can be found by solving the following system:

$$^{(t)}(\ _j) = h^{(t)}(\ _j); \ 1 \quad j \quad k; 0 \quad t < m_j$$

where $(z) = z^n$ and $^{(n)}$ is the nth derivative.

Example 2.31. Given $J = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, we can see by inspection that the minimal polynomial of J is $f(z) = (z \quad 2)^2(z \quad 1)$. This is of order three, so our Lagrange interpolation polynomial is of order two, and so quadratic. Let $h(z) = \ z^2 + \ z + \ $. The conditions for the coe cients are:

$$\begin{cases} 2^n \ \ = h(2) \ = 4 \ + 2 \ + \\ n2^{n-1} = h'(2) = 4 \ + \\ 1^n \ \ = h(1) \ = \ + \ + \end{cases}$$

Solving gives $= n2^{n-1} \ 2^n + 1, \ = \ 3n2^{n-1} + 4 \ 2^n \ 4$ and $= 2n2^{n-1} \ 3 \ 2^n + 4$. So

$$J^n = (n2^{n-1} \ 2^n + 1)J^2 + (\ 3n2^{n-1} + 4 \ 2^n \ 4)J + (2n2^{n-1} \ 3 \ 2^n + 4)I = \begin{pmatrix} 2^n & n2^{n-1} & 0 \\ 0 & 2^n & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

as we had before.

3 Bilinear Maps and Quadratic Forms

There are many situations in which we wish to consider maps from a vector space into its field of scalars. (We may be interested in \linear forms" or \linear functionals", i.e. linear maps $f : V \to K$; for example, the integral operator which maps a function $g : [a,b] \to \mathbb{R}$ to its integral $\int_a^b g(x)\,dx$ is a linear form on the space of continuous functions $C^0([a,b])$. More in MA3G7 Functional Analysis I.)

In this section we are interested in \quadratic forms". Roughly speaking, a quadratic form $q : V \to K$ is a map from a vector space into its field of scalars which is a homogeneous polynomial of degree 2, i.e. a polynomial in which each term has total degree two, such as $6x^2 + 12xy - 13xz + 7y^2$. These have many applications, such as conic sections; for example, the equation $5x^2 + 5y^2 - 6xy = 2$ defines an ellipse.

3.1 Definitions

In order to actually define quadratic forms, we first introduce \bilinear forms", and the more general \bilinear maps". Bilinear maps are functions $\phi : W \times V \to K$ which take two vectors and spit out a number, and which are linear in each argument, as follows:

Definition 3.1. Let V, W be vector spaces over a field K. A *bilinear map* ϕ on W and V is a map $\phi : W \times V \to K$ such that

$$\phi(\lambda_1\mathbf{w}_1 + \lambda_2\mathbf{w}_2, \mathbf{v}) = \lambda_1\phi(\mathbf{w}_1, \mathbf{v}) + \lambda_2\phi(\mathbf{w}_2, \mathbf{v})$$
$$\text{and} \qquad \phi(\mathbf{w}, \mu_1\mathbf{v}_1 + \mu_2\mathbf{v}_2) = \mu_1\phi(\mathbf{w}, \mathbf{v}_1) + \mu_2\phi(\mathbf{w}, \mathbf{v}_2)$$

for all $\mathbf{w}, \mathbf{w}_1, \mathbf{w}_2 \in W$, all $\mathbf{v}, \mathbf{v}_1, \mathbf{v}_2 \in V$, and $\lambda_1, \lambda_2 \in K$.

By choosing bases $\mathbf{e}_1, \ldots, \mathbf{e}_n$ of V and $\mathbf{f}_1, \ldots, \mathbf{f}_m$ of W, we can set $\alpha_{ij} = \phi(\mathbf{f}_i, \mathbf{e}_j)$ and form the $m \times n$ matrix $A = (\alpha_{ij})$ of ϕ with respect to the above bases. Now, given $\mathbf{v} \in V$ and $\mathbf{w} \in W$, by writing $\mathbf{v} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$ and $\mathbf{w} = y_1\mathbf{f}_1 + \cdots + y_m\mathbf{f}_m$, we can form the column vectors $\underline{\mathbf{v}} = (x_1, \ldots, x_n)^{\mathsf{T}} \in K^{n,1}$ and $\underline{\mathbf{w}} = (y_1, \ldots, y_m)^{\mathsf{T}} \in K^{m,1}$. Then using the linearity properties in the definition, we get

$$\phi(\mathbf{w}, \mathbf{v}) = \sum_{i=1}^{m}\sum_{j=1}^{n} y_i\,\phi(\mathbf{f}_i, \mathbf{e}_j)x_j = \sum_{i=1}^{m}\sum_{j=1}^{n} y_i\,\alpha_{ij}x_j = \underline{\mathbf{w}}^{\mathsf{T}}A\underline{\mathbf{v}}.$$

Suppose we choose new bases $\mathbf{e}'_1, \ldots \mathbf{e}'_n$ of V and $\mathbf{f}'_1, \ldots, \mathbf{f}'_m$ of W, and let P and Q be the associated basis change matrices. Then if $\underline{\mathbf{v}}'$ and $\underline{\mathbf{w}}'$ are the column vectors representing $\mathbf{v}$ and $\mathbf{w}$ with respect to the bases $(\mathbf{e}'_j)$ and $(\mathbf{f}'_j)$, we have $P\underline{\mathbf{v}}' = \underline{\mathbf{v}}$ and $Q\underline{\mathbf{w}}' = \underline{\mathbf{w}}$, so $\underline{\mathbf{w}}^{\mathsf{T}}A\underline{\mathbf{v}} = \underline{\mathbf{w}}'^{\mathsf{T}}Q^{\mathsf{T}}AP\underline{\mathbf{v}}'$, hence:

Theorem 3.2. Let $\phi : W \times V \to K$ be a bilinear map. Let A be the matrix of ϕ w.r.t. bases $(\mathbf{e}_i)$ of V and $(\mathbf{f}_j)$ of W, and let B be the matrix of ϕ w.r.t. bases $(\mathbf{e}'_j)$ of V and $(\mathbf{f}'_j)$ of W. If P is the basis change matrix from $(\mathbf{e}_i)$ to $(\mathbf{e}'_j)$ and Q is the basis change matrix from $(\mathbf{f}_j)$ to $(\mathbf{f}'_j)$, then $B = Q^{\mathsf{T}}AP$.

From now on, we will only consider the case where $V = W$; then a bilinear map $\phi : V \times V \to K$ is called a *bilinear form* on V. Then in the previous theorem, we have that $Q = P$ and thus $B = P^{\mathsf{T}}AP$.

Definition 3.3. The *rank* of a bilinear form ϕ is defined as the rank of the associated matrix A (and such is well-defined and independent of choice of basis).

The kernel of A and of A^T also have special properties in relation to bilinear forms.

Definition 3.4. The kernel of A is equal to the space $\{\mathbf{v} \in V : \phi(\mathbf{w}, \mathbf{v}) = 0\ \forall \mathbf{w} \in V\}$ (the *right radical* of ϕ). The kernel of A^T is equal to the space $\{\mathbf{v} \in V : \phi(\mathbf{v}, \mathbf{w}) = 0\ \forall \mathbf{w} \in V\}$ (the *left radical* of ϕ).

In fact, we are almost exclusively interested in *symmetric bilinear forms*:

Definition 3.5. A bilinear form $\phi : V \times V \to K$ is *symmetric* if $\phi(\mathbf{w}, \mathbf{v}) = \phi(\mathbf{v}, \mathbf{w})$ for all $\mathbf{v}, \mathbf{w} \in V$.

Taking a basis $(\mathbf{e}_i)$, tells us that $\phi(\mathbf{e}_i, \mathbf{e}_j) = \phi(\mathbf{e}_j, \mathbf{e}_i)$ for all $1 \le i, j \le n$, and hence that $\alpha_{ij} = \alpha_{ji}$. Thus $A^{\mathsf{T}} = A$, so a bilinear form is symmetric if and only if its matrix (w.r.t. any basis) is symmetric:

Definition 3.6. A $n \times n$ matrix A is called *symmetric* if $A^{\mathsf{T}} = A$.

Matrices that represent the same bilinear form in different bases are called *congruent*.

Definition 3.7. Symmetric matrices A and B are called *congruent* if there exists an invertible matrix P with $B = P^T AP$.

Given a symmetric bilinear form, we can define a quadratic form:

Definition 3.8. Let V be a vector space over the field[1] K. A *quadratic form* on V is a function $q: V \to K$ defined by $q(\mathbf{v}) = (\mathbf{v}, \mathbf{v})$, where $: V \times V \to K$ is a symmetric bilinear form.

Given a basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ of V, we let $A = (a_{ij})$ be the matrix of with respect to this basis, which is symmetric (since is symmetric). Then we can write:

$$q(\mathbf{v}) = \underline{\mathbf{v}}^{\mathsf{T}} A \underline{\mathbf{v}} = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i \, a_{ij} x_j = \sum_{i=1}^{n} a_{ii} x_i^2 + 2\sum_{i=1}^{n}\sum_{j=1}^{i-1} a_{ij} x_i x_j.$$

For this reason we also call A the matrix of q. When $n \geq 3$, we typically write x, y, z for x_1, x_2, x_3.

3.2 Change of Variable under the General Linear Group

Just as we seek to reduce linear transformations into a canonical form in order that we can understand their geometric properties, we can do the same with quadratic forms. Our first aim is simply to eliminate all the $x_i x_j$ terms, leaving only terms of the form $a_{ii} x_i^2$.

Theorem 3.9. Let q be a quadratic form on V. Then there is a basis $\mathbf{e}'_1, \ldots, \mathbf{e}'_n$ of V such that $q(\mathbf{v}) = \sum_{i=1}^{n} a_i (x'_i)^2$, where the x'_i are the coordinates of $\mathbf{v}$ with respect to $\mathbf{e}'_1, \ldots, \mathbf{e}'_n$.

Equivalently, given any symmetric matrix A, there is an invertible matrix P such that $P^{\mathsf{T}} AP$ is a diagonal matrix, i.e. A is congruent to a diagonal matrix.

Sketch proof. This is essentially done by \completing the square": assuming that $a_{11} \neq 0$, we can write

$$q(\mathbf{v}) = a_{11} x_1^2 + 2a_{12} x_1 x_2 + \cdots + 2a_{1n} x_1 x_n + q_0(\mathbf{v}) = a_{11}\left(x_1 + \frac{a_{12}}{a_{11}} x_2 + \cdots + \frac{a_{1n}}{a_{11}} x_n\right)^2 + q_1(\mathbf{v})$$

where q_0 and q_1 are quadratic forms only involving $x_2, \ldots, x_n$. Then we can make the change of coordinates $x'_1 = x_1 + \frac{a_{12}}{a_{11}} x_2 + \cdots + \frac{a_{1n}}{a_{11}} x_n$, $x'_i = x_i$ for $2 \leq i \leq n$, which gets rid of the cross-terms involving x_1 and leaves us with $q(\mathbf{v}) = a_{11}(x'_i)^2 + q_1(\mathbf{v})$, where q_1 only involves $x'_2, \ldots, x'_n$, and we are done by induction; in the case that $a_{11} = 0$, we reduce to the previous case by first changing coordinates. □

The *rank* of a quadratic form is defined to be the rank of its matrix A. Since P and P^{T} are invertible, $P^{\mathsf{T}} AP$ has the same rank as A, so the rank of a quadratic form is independent of the choice of basis, and if $P^{\mathsf{T}} AP$ is diagonal then the number of non-zero entries on the diagonal is equal to the rank.

The rank of a quadratic form gives us one method of distinguishing between different quadratic forms. Depending on the field we are working in, we can often reduce the quadratic form further.

Proposition 3.10. A quadratic form over C has the form $q(\mathbf{v}) = \sum_{i=1}^{r} x_i^2$ with respect to a suitable basis, where $r = \operatorname{rank}(q)$.

Proof. Having reduced q to the form $q(\mathbf{v}) = \sum_{i=1}^{n} a_i (x_i)^2$, permute the axes so the first r coefficients are non-zero and the last $n - r$ are zero, and then make the change $x'_i = \sqrt{a_{ii}}\, x_i$ (for $1 \leq i \leq r$). □

Proposition 3.11 (Sylvester's Theorem)**.** A quadratic form over R has the form $q(\mathbf{v}) = \sum_{i=1}^{t} x_i^2 - \sum_{i=1}^{u} x_{t+i}^2$ with respect to a suitable basis, where $t + u = \operatorname{rank}(q)$.

Proof. Almost the same as over C, except we must put $x'_i = \sqrt{|a_{ii}|}\, x_i$ (and then reorder if necessary). □

Theorem 3.12. Let V be a vector space over R and let q be a quadratic form over V. Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ and $\mathbf{e}'_1, \ldots, \mathbf{e}'_n$ be two bases of V with coordinates x_i and x'_i, such that $q(\mathbf{v}) = \sum_{i=1}^{t} x_i^2 - \sum_{i=t+1}^{u} x_{t+i}^2 = \sum_{i=1}^{t'} (x'_i)^2 - \sum_{i=1}^{u'} (x'_{t'+i})^2$. Then $t = t'$ and $u = u'$, that is t and u are invariants of q.

The tuple (t, u) is called the *signature* of q.

[1] To talk about quadratic forms, we must be able to divide by 2 in K, thus we must assume that $1 + 1 \neq 0$; e.g. K cannot be Z_2, the field of two elements. If you don't like technical details, you can safely assume that K is Q, R or C.

3.3 Change of Variable under the Orthogonal Group

Previously, we allowed any change of coordinates. But if we consider, say, the ellipse $5x^2 + 5y^2 - 6xy = 2$, changing coordinates may not preserve the shape of the ellipse, which is often undesirable. We now study *orthogonal* transformations, which preserve length and angle. Throughout this section we assume $K = \mathbb{R}$.

Definition 3.13. A quadratic form q on V is said to be *positive definite* if $q(\mathbf{v}) > 0$ for all $\mathbf{0} \neq \mathbf{v} \in V$.

It is clear that q is positive definite if and only if q has rank n and signature n. The associated symmetric bilinear form is also called positive definite when q is. If we choose a basis of V such that the matrix of is I_n, then is just the standard scalar (or inner) product on V.

Definition 3.14. A *Euclidean space* is a vector space V over $\mathbb{R}$ together with a positive definite symmetric bilinear form .

We will assume for this section that V is a Euclidean space, and that the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ has been chosen so that the matrix of is I_n. As is then the standard scalar product we will write $\mathbf{v} \cdot \mathbf{w}$ instead of $(\mathbf{v}, \mathbf{w})$; note that $\mathbf{v} \cdot \mathbf{w} = \underline{\mathbf{v}}^{\mathsf{T}} \underline{\mathbf{w}}$ (where $\underline{\mathbf{v}}$, $\underline{\mathbf{w}}$ are the column vectors of $\mathbf{v}$, $\mathbf{w}$ resp.).

It is this scalar product which defines length and angle: for $\mathbf{v} \in V$, the length of $\mathbf{v}$ is $|\mathbf{v}| := \sqrt{\mathbf{v} \cdot \mathbf{v}}$, and the angle between $\mathbf{v}$ and $\mathbf{w}$ is $\arccos \frac{\mathbf{v} \cdot \mathbf{w}}{|\mathbf{v}||\mathbf{w}|}$. Thus for a linear transformation to preserve length and angle – in which case, geometrically, it is a rigid map – it must preserve the inner product on V:

Definition 3.15. A linear map $T : V \to V$ is said to be *orthogonal* if $T(\mathbf{v}) \cdot T(\mathbf{w}) = \mathbf{v} \cdot \mathbf{w}$ for all $\mathbf{v}, \mathbf{w} \in V$.

If A is the matrix of T, i.e. $T(\mathbf{v}) = A\underline{\mathbf{v}}$, then $T(\mathbf{v}) \cdot T(\mathbf{w}) = \underline{\mathbf{v}}^{\mathsf{T}} A^{\mathsf{T}} A \underline{\mathbf{w}}$, hence T is orthogonal if and only if $A^{\mathsf{T}} A = I_n$. Thus a linear map is orthogonal if and only if its matrix is orthogonal:

Definition 3.16. An $n \times n$ matrix A is called *orthogonal* if $A^{\mathsf{T}} A = I_n$, or equivalently if $A^{\mathsf{T}} = A^{-1}$.

We can characterise this more in terms of geometric properties as follows:

Definition 3.17. A set of vectors $\mathbf{f}_1, \ldots, \mathbf{f}_n$ of V which satisfies $\mathbf{f}_i \cdot \mathbf{f}_i = 1$, and $\mathbf{f}_i \cdot \mathbf{f}_j = 0$ when $i \neq j$, is called *orthonormal*. If the $\{\mathbf{f}_i\}$ form a basis, they are called an *orthonormal basis*.

Proposition 3.18. A linear map $T : V \to V$ is orthogonal if and only if $T(\mathbf{e}_1), \ldots, T(\mathbf{e}_n)$ is an orthonormal basis of V (where the $\mathbf{e}_j$ are defined as above).

The Gram–Schmidt algorithm allows us to complete any orthonormal set to an orthonormal basis:

Theorem 3.19 (Gram–Schmidt)**.** Let V be a Euclidean space of dimension n, and suppose $\mathbf{f}_1, \ldots, \mathbf{f}_r$ are orthonormal ($0 \leqslant r \leqslant n$). Then $\mathbf{f}_1, \ldots, \mathbf{f}_r$ can be extended to an orthonormal basis $\mathbf{f}_1, \ldots, \mathbf{f}_n$.

Sketch proof. First we extend $\mathbf{f}_1, \ldots, \mathbf{f}_r$ to any basis $\mathbf{f}_1, \ldots, \mathbf{f}_r, \mathbf{g}_{r+1}, \ldots, \mathbf{g}_n$. The trick is then to define $\mathbf{f}'_{r+1} = \mathbf{g}_{r+1} - \sum_{i=1}^{r} (\mathbf{f}_i \cdot \mathbf{g}_{r+1}) \mathbf{f}_i$. This removes all components of $\mathbf{g}_{r+1}$ in the directions of $\mathbf{f}_1, \ldots, \mathbf{f}_r$, and then the only step left is to normalise it by setting $\mathbf{f}_{r+1} = \frac{\mathbf{f}'_{r+1}}{|\mathbf{f}'_{r+1}|}$, and proceed by induction on $n - r$. □

The main result of this section is to show that we can *always* change coordinates so that the matrix of a quadratic form is diagonal, *and* do so in a way that preserves the geometric properties of the form:

Theorem 3.20. Let q be a quadratic form on a Euclidean space V. Then there is an orthonormal basis $\mathbf{e}'_1, \ldots, \mathbf{e}'_n$ of V such that $q(\mathbf{v}) = \sum_{i=1}^{n} \lambda_i (x'_i)^2$, where the x'_i are the coordinates of $\mathbf{v}$ with respect to $\mathbf{e}'_1, \ldots, \mathbf{e}'_n$, and the λ_i are uniquely determined by q. Equivalently, given any symmetric matrix A, there is an orthogonal matrix P such that $P^{\mathsf{T}} A P$ is a diagonal matrix.

The following two easy lemmas are used both in the proof of the theorem and in doing calculations:

Lemma 3.21. Let A be a real symmetric matrix. Then all complex eigenvalues of A lie in $\mathbb{R}$.

Proof. Suppose $A\mathbf{v} = \lambda \mathbf{v}$ for some $\lambda \in \mathbb{C}$. Then $\mathbf{v}^{\mathsf{T}} \overline{\mathbf{v}} = (A\mathbf{v})^{\mathsf{T}} \overline{\mathbf{v}} = \mathbf{v}^{\mathsf{T}} A \overline{\mathbf{v}} = \mathbf{v}^{\mathsf{T}} \overline{\lambda} \overline{\mathbf{v}} = \overline{\lambda} \mathbf{v}^{\mathsf{T}} \overline{\mathbf{v}}$. As $\mathbf{v} \neq \mathbf{0}$, $\mathbf{v}^{\mathsf{T}} \overline{\mathbf{v}} \neq 0$, and hence $\lambda = \overline{\lambda}$, and thus $\lambda \in \mathbb{R}$. □

Lemma 3.22. Let A be a real symmetric matrix, and let λ_1, λ_2 be two distinct eigenvalues of A, with corresponding eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$. Then $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$.

Proof. Suppose $A\mathbf{v}_1 = \lambda_1\mathbf{v}_1$ and $A\mathbf{v}_2 = \lambda_2\mathbf{v}_2$, with $\lambda_1 \neq \lambda_2$. Transposing the first and multiplying by $\mathbf{v}_2$ gives $\mathbf{v}_1^{\mathsf{T}}A\mathbf{v}_2 = \lambda_1\mathbf{v}_1^{\mathsf{T}}\mathbf{v}_2$ ( ). Similarly $\mathbf{v}_2^{\mathsf{T}}A\mathbf{v}_1 = \lambda_2\mathbf{v}_2^{\mathsf{T}}\mathbf{v}_1$; transposing this yields $\mathbf{v}_1^{\mathsf{T}}A\mathbf{v}_2 = \lambda_2\mathbf{v}_1^{\mathsf{T}}\mathbf{v}_2$ (*y*). Subtracting (*y*) from ( ) gives $(\lambda_2 - \lambda_1)\mathbf{v}_1^{\mathsf{T}}\mathbf{v}_2 = 0$, so as $\lambda_1 \neq \lambda_2$, $\mathbf{v}_1^{\mathsf{T}}\mathbf{v}_2 = 0$. □

Example 3.23. Let $q\colon \mathbb{R}^3 \to \mathbb{R}$ be the quadratic form given by $q(x,y,z) = 3x^2 + 3y^2 + 4xy + 2xz - 2yz$. This has matrix $A = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 3 & -1 \\ 1 & -1 & 0 \end{pmatrix}$. We find $c_A(x) = -x^3 + 6x^2 - 3x - 10 = (5-x)(2-x)(-1-x)$. Thus the eigenvalues are 5, 2, and -1. For $\lambda = 5$, an eigenvector is $(1,-1,0)^{\mathsf{T}}$. For $\lambda = 2$, an eigenvector is $(1,1,-1)^{\mathsf{T}}$. For $\lambda = -1$, an eigenvector[2] is $(1,1,2)^{\mathsf{T}}$. Normalising the eigenvectors, we get

$$P = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \\ -1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \\ 0 & -1/\sqrt{3} & 2/\sqrt{6} \end{pmatrix}; \qquad P^{\mathsf{T}}AP = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Example 3.24 (Conic sections and quadrics)**.** The general second-degree equation in n variables is $\sum_{i=1}^{n} a_i x_i^2 + \sum_{i=1}^{n}\sum_{j=1}^{i-1} b_{ij}x_i x_j + \sum_{i=1}^{n} c_i x_i + d = 0$, which defines a *quadric* curve or surface in $\mathbb{R}^n$. By applying the above theory, we can perform an orthogonal transformation to eliminate the $x_i x_j$ terms, and further orthogonal transformations reduce it so that either exactly one $a_i \neq 0$ or $d \neq 0$.

Being able to take a quadric in two or three variables and sketch its graph is an important skill. The three key shapes to remember in two dimensions are the ellipse $ax^2 + by^2 = 1$, the hyperbola $ax^2 - by^2 = 1$ and the parabola $ax^2 + y = 0$ (where $a,b > 0$). Knowing these should allow you to figure out any three-dimensional quadric, by figuring out the level sets or by analogy. (Lack of space here unfortunately prevents us from including a table of all the possibilities). It helps to keep in mind the degenerate cases (e.g. straight lines, points). Remember also that some surfaces can be constructed using *generators* | straight lines through the origin for which every point on the line is on the surface.

3.4 Unitary, Hermitian and Normal Matrices

The results of the last section only applied for vector spaces over $\mathbb{R}$. Here we generalise the results to vector spaces over $\mathbb{C}$. Recall that $\bar{z}$ denotes the complex conjugate of $z \in \mathbb{C}$, and $\overline{A}$ denotes the result of replacing each entry of A by its complex conjugate. We denote the conjugate transpose of A by $A^* = \overline{A}^{\mathsf{T}}$.

Definition 3.25. Let V be a vector space over $\mathbb{C}$. A *sesquilinear form* is a function $\phi\colon V \times V \to \mathbb{C}$ which is linear in the second argument and conjugate-linear in the first argument, that is

$$\phi(\lambda_1\mathbf{u}_1 + \lambda_2\mathbf{u}_2, \mathbf{v}) = \overline{\lambda_1}\phi(\mathbf{u}_1,\mathbf{v}) + \overline{\lambda_2}\phi(\mathbf{u}_2,\mathbf{v})$$
$$\text{and} \qquad \phi(\mathbf{u}, \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2) = \lambda_1\phi(\mathbf{u},\mathbf{v}_1) + \lambda_2\phi(\mathbf{u},\mathbf{v}_2)$$

for all $\mathbf{u},\mathbf{u}_1,\mathbf{u}_2,\mathbf{v},\mathbf{v}_1,\mathbf{v}_2 \in V$, and $\lambda_1, \lambda_2 \in \mathbb{C}$.

The standard inner product on $\mathbb{C}^n$ is the sesquilinear form defined by $\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^*\mathbf{w}$ (rather than $\mathbf{v}^{\mathsf{T}}\mathbf{w}$). The change is so that $|\mathbf{v}|^2 = \mathbf{v}\cdot\mathbf{v} = \mathbf{v}^*\mathbf{v}$ is still a real number. (Note that this is *not* a bilinear form, nor is it symmetric, since $\mathbf{v}\cdot\mathbf{w} = \overline{\mathbf{w}\cdot\mathbf{v}}$, and hence $(\lambda\mathbf{v})\cdot\mathbf{w} = \overline{\lambda}(\mathbf{v}\cdot\mathbf{w})$.)

The complex analogue of an orthogonal map { one which preserves the complex inner product { is a *unitary map*:

Definition 3.26. A linear map $T\colon \mathbb{C}^n \to \mathbb{C}^n$ is called *unitary* if $T(\mathbf{v})\cdot T(\mathbf{w}) = \mathbf{v}\cdot\mathbf{w}$ for all $\mathbf{v},\mathbf{w} \in \mathbb{C}^n$.

If A is the matrix of T, then $T(\mathbf{v}) = A\underline{\mathbf{v}}$, so $T(\mathbf{v})\cdot T(\mathbf{w}) = \underline{\mathbf{v}}^*A^*A\underline{\mathbf{w}}$; hence T is unitary if and only if $A^*A = I_n$. Thus a linear map is unitary if and only if its matrix is unitary:

Definition 3.27. A matrix $A \in \mathbb{C}^{n,n}$ is called *unitary* if $A^*A = I_n$.

The complex analogue of a symmetric matrix is:

Definition 3.28. A matrix $A \in \mathbb{C}^{n,n}$ is called *Hermitian* if $A^* = A$.

[2] Note that having figured out the first two eigenvectors, we could simply find a vector orthogonal to both, perhaps using the cross product, and take this as our eigenvector (using lemma 3.22).

One would expect to extend theorem 3.20 to Hermitian matrices, but the generalisation in fact applies to the wider class of *normal matrices*, which includes all Hermitian matrices and all unitary matrices:

Definition 3.29. A matrix $A \in \mathbb{C}^{n,n}$ is called *normal* if $AA^* = A^*A$.

Theorem 3.30. Let $A \in \mathbb{C}^{n,n}$ be a normal matrix. Then there exists a unitary matrix $P \in \mathbb{C}^{n,n}$ such that P^*AP is diagonal.

In practice, finding such a P is done in a similar way to the real case.

4 Finitely Generated Abelian Groups

Groups are one of the most fundamental and important constructions in mathematics. When they are first introduced in MA136 Introduction to Abstract Algebra, however, they can seem rather abstract. In this section we apply results from linear algebra to classify all finitely generated abelian groups, thus opening up a large number of concrete examples of groups. (That's not to say that non-abelian groups are unimportant: in fact, they form at least half of MA249 Algebra II: Groups and Rings.) Before we can classify finitely-generated abelian groups, we had better know what they are, so we start by defining everything.

Definition 4.1. A *group* is a set G together with a binary operation $\star$ satisfying:
(i) For every $g, h \in G$, $g \star h \in G$ (closure).
(ii) For every $g, h, k \in G$, $g \star (h \star k) = (g \star h) \star k$ (associativity).
(iii) There exists $e \in G$ such that $e \star g = g \star e = g$ for every $g \in G$ (existence of an identity, e).
(iv) For every $g \in G$ there exists $g' \in G$ such that $g' \star g = g \star g' = e$ (existence of inverses).

Definition 4.2. An *abelian group* is a group G which satisfies the commutative law:
(v) For every $g, h \in G$, $g \star h = h \star g$.

The following result is so fundamental that its use usually goes unreported:

Lemma 4.3 (Cancellation laws)**.** Let G be a group, and let $g, h, k \in G$. If $g \star h = g \star k$, then $h = k$; similarly, if $h \star g = k \star g$, then $h = k$.

Very often, two groups have the same structure, but with the elements \labelled" differently. When this is the case, we call the groups isomorphic. More generally, we can define \structure-preserving maps", or *homomorphisms*, between groups:

Definition 4.4. A function $\varphi : G \to H$ between two groups G and H is called an *homomorphism* if $\varphi(g_1 \star g_2) = \varphi(g_1) \star \varphi(g_2)$ for all $g_1, g_2 \in G$.
If φ is injective, i.e. $\varphi(g_1) = \varphi(g_2) \Rightarrow g_1 = g_2$, then φ is called a *monomorphism*.
If φ is surjective, i.e. $\mathrm{im}(\varphi) = H$, then φ is called an *epimorphism*.
If φ is bijective[3] then φ is called an *isomorphism*; G and H are called *isomorphic*, written $G \cong H$.
If $\varphi : G \to H$ is a homomorphism, then $\varphi(e_G) = e_H$, (i.e. the identity element is always mapped to the identity element), and $\varphi(g') = (\varphi(g))'$, (i.e. the inverse of $\varphi(g)$ is just φ applied to the inverse of g).

For abelian groups, the notation $\star$ for the binary operation is cumbersome. So, from now on:
We will only talk about abelian groups (unless otherwise stated).
We will notate the binary operation of an abelian group by addition (i.e. using + instead of $\star$).
We will notate the identity element by 0 instead of e (or 0_G if we need to specify which group).
We will notate the inverse element of g by $-g$ instead of g'.
(For non-abelian groups, it is more usual to use multiplicative notation, using 1 for the identity and g^{-1} for the inverse of g.)

[3] It might be better to say an isomorphism is a bijective homomorphism *whose inverse is also a homomorphism*, but since it follows from φ being a bijective homomorphism it is usually omitted. For other structure-preserving maps, the distinction is important: a continuous bijection between topological spaces does not necessarily have *continuous* inverse, and thus is not always a homeomorphism; see MA222 Metric Spaces. Interested readers should Google \category theory".

4.1 Generators and Cyclic Groups

For an element g of a group G, we de ne $ng = g + \quad + g$ (n times), where $n \ 2$ N. That is, $1g = g$, $2g = g + g$, $3g = g + g + g$, etc. We extend this to $n \ 2$ Z by de ning $0g = 0$ and $(\ n)g = \ (ng)$ (it then follows that $(\ n)g = n(\ g)$). (In multiplicative notation, this is usually written as g^n instead of ng.) The fundamental observation to make is that *multiples of an element commute with each other*, i.e. $mg + ng = ng + mg = (m + n)g$.

Definition 4.5. Let $g \ 2 \ G$. De ne the *order* of g to be the least integer $n > 0$ with $nx = 0$, if such an n exists, and write $jgj = n$. If there is no such n, g is said to have in nite order and we write $jgj = 1$.

Note that if $: G \ ! \ H$ is an isomorphism, then $jgj = j \ (g)j$, i.e. orders of elements are preserved under isomorphism.

Definition 4.6. A group G is called *cyclic* if there exists $x \ 2 \ G$ such that $G = fnx \ j \ n \ 2 \ Zg$, that is every element of G is of the form nx for some $n \ 2$ Z. We call x a *generator* of G.

By the above remark, a cyclic group is necessarily abelian. Note that Z and Z_n (under addition) are cyclic with generator 1. In fact, up to isomorphism, these are *all* the cyclic groups:

Proposition 4.7. Any cyclic group G is isomorphic either to Z, or to Z_n for some $n \ 2$ N (with $n > 0$).

Sketch proof. Let G be cyclic with generator x; thus $G = fnx \ j \ n \ 2 \ Zg$. Either all the nx (for $n \ 2$ Z) are distinct, in which case $G =$ Z. Otherwise, there must be two which are equal, say $lx = mx$ ($l < m$), so $(m \quad l)x = 0$ with $m > l$; take n to be the least natural number with $nx = 0$, then $G = Z_n$. □

A group is cyclic if it has one generator. We now consider groups generated by more than one element.

Definition 4.8. A group G is *generated* (or *spanned*) by $X \quad G$ if every $g \ 2 \ G$ can be written as $g = \sum_{i=1}^{k} n_i x_i$, with $k \ 2$ N, $x_i \ 2 \ X$ and $n_i \ 2$ Z; we write $G = hX \, i$. In particular, if X is nite, we say that G is *nitely generated*; if $X = fx_1; : : : ; x_n g$ then we write $G = hx_1; : : : ; x_n \, i$.

We have already seen that a group generated by X with $jXj = 1$ (i.e. cyclic) must be isomorphic to Z_n or Z. In order to classify all nitely-generated abelian groups, we introduce the direct sum of groups as a way of putting a group operation on the Cartesian product of groups. (In general group theory this is more often called the direct product.)

Definition 4.9. Given groups $G_1; : : : ; G_n$, we de ne the direct sum $G_1 \quad G_n$ to be the set $G_1 \quad G_n = f(g_1; : : : ; g_n) \ j \ g_i \ 2 \ G_i g$ with the binary operation given by componentwise addition, i.e. $(g_1; : : : ; g_n) + (h_1; : : : ; h_n) := (g_1 + h_1; : : : ; g_n + h_n)$. This is a group with identity element $(0; : : : ; 0)$ and $(g_1; : : : ; g_n) = (\ g_1; : : : ; \ g_n)$.

We will see that all nitely-generated abelian groups are isomorphic to a direct sum of cyclic groups.

4.2 Subgroups and Cosets

Definition 4.10. Let G be a group and $H \quad G$. We call H a *subgroup* of G if H together with the binary operation on G is a group in its own right. (We sometimes write $H \quad G$.)

In practice, we can avoid checking the associativity and identity elements:

Proposition 4.11. A *non-empty* subset $H \quad G$ is a subgroup of G if and only if $h_1; h_2 \ 2 \ H \ =) \ h_1 + h_2 \ 2 \ H$, and $h \ 2 \ H \ =) \quad h \ 2 \ H$.

Note that $f0g$ and G are always subgroups of G. More importantly, if $g \ 2 \ G$, then the set $fng \ j \ n \ 2 \ Zg$ of integer multiples of g is a subgroup of G, called the *cyclic subgroup* generated by g.

Definition 4.12. Let $g \ 2 \ G$, and let H be a subgroup of G. De ne $H + g = fh + g \ j \ h \ 2 \ Hg$, the (right) *coset* of H by g.

Note that since all our groups are abelian, $H + g = g + H$; in general, however, the right and left cosets, denoted Hg and gH respectively, can be di erent.

Example 4.13. Let $G = \mathbb{Z}$, and let $H = 6\mathbb{Z} = \{6n \mid n \in \mathbb{Z}\}$, the cyclic subgroup generated by 6. Then there are six distinct cosets of $6\mathbb{Z}$ in $\mathbb{Z}$, namely $6\mathbb{Z}$, $6\mathbb{Z}+1$, $6\mathbb{Z}+2$, $6\mathbb{Z}+3$, $6\mathbb{Z}+4$, and $6\mathbb{Z}+5$, which correspond to those integers which leave remainder 0, 1, 2, 3, 4 and 5 (respectively) on division by 6.

Proposition 4.14. Two right cosets $H + g_1$ and $H + g_2$ of H in G are either equal or disjoint; hence the cosets of H partition G.

Proposition 4.15. If H is a finite subgroup of G, then all right cosets of H have exactly $|H|$ elements.

From the last two propositions, we get Lagrange's theorem (which also holds for non-abelian groups):

Theorem 4.16 (Lagrange's Theorem)**.** Let H be a subgroup of a finite (abelian) group G. Then the order of H divides the order of G.

By considering the cyclic subgroup of G generated by an element g, i.e. $H = \{ng \mid n \in \mathbb{Z}\}$, we see that if $|g| = n$ then $|H| = n$, and hence:

Proposition 4.17. Let G be a finite (abelian) group, and let $g \in G$. Then $|g|$ divides $|G|$.

4.3 Quotient Groups and the First Isomorphism Theorem

We have already seen two ways of creating new groups from old ones, namely subgroups and direct sums, and we will presently define a third. Let H be a subgroup of an abelian group G, and consider the set G/H of cosets of H in G, i.e. $\{H + g \mid g \in G\}$. Since the cosets of H partition G, each element of G lies in exactly one of the cosets Hg. By dividing G into the cosets of H, we have \quotiented out" the subgroup H. We will now define a binary operation on G/H that turns it into a group.

Definition 4.18. If A and B are subsets of a group G, define the sum[4] $A + B = \{a + b \mid a \in A, b \in B\}$.

The binary operation on G/H rests on the following fundamental lemma:

Lemma 4.19. Let H be a subgroup of an abelian group G, and let $g_1, g_2 \in G$. Then $(H + g_1) + (H + g_2) = H + (g_1 + g_2)$.

Note: This is the first point at which it is crucial that G is abelian. In general, if H is a subgroup of any group G, it is not necessarily the case that $(Hg_1)(Hg_2) = H(g_1g_2)$. In fact, this happens if and only if $gH = Hg$ for every $g \in G$, i.e. every left coset is also a right coset; when this happens, we call H a *normal* subgroup of G. The general case will be done in MA249 Algebra II: Groups and Rings.

Theorem 4.20. Let H be a subgroup of an abelian group G. Then the set G/H of cosets of H in G forms a group under addition of cosets, which is called the *quotient group* of G by H.

Proof. The lemma gives us closure; associativity follows (tediously) from associativity of G; the identity is $H = H + 0$; and the inverse of $H + g$ is $H + (-g)$, which we write as $H - g$. □

Definition 4.21. The *index* of H in G, denoted $|G : H|$, is the number of distinct cosets of H in G.

If G is finite, by Lagrange's theorem we have $|G/H| = |G : H| = \frac{|G|}{|H|}$.

Example 4.22. Consider again $G = \mathbb{Z}$ with $H = \{6n \mid n \in \mathbb{Z}\}$. As there are six distinct cosets of $6\mathbb{Z}$ in $\mathbb{Z}$, we have that $\mathbb{Z}/6\mathbb{Z} = \{6\mathbb{Z}, 6\mathbb{Z}+1, 6\mathbb{Z}+2, 6\mathbb{Z}+3, 6\mathbb{Z}+4, 6\mathbb{Z}+5\}$, with addition given by $(6\mathbb{Z} + m) + (6\mathbb{Z} + n) = 6\mathbb{Z} + (m + n)$. That is, when you add a number whose remainder on division by 6 is m to a number whose remainder on division by 6 is n, you get a number whose remainder on division by 6 is $m + n$.

Note that adding together k copies of $6\mathbb{Z} + 1$ gives you $6\mathbb{Z} + k$, so the quotient group $\mathbb{Z}/6\mathbb{Z}$ is generated by $6\mathbb{Z} + 1$. Hence it is a cyclic group of order 6, and so $\mathbb{Z}/6\mathbb{Z} = \mathbb{Z}_6$. That is, addition of cosets of $6\mathbb{Z}$ is exactly the same as addition modulo 6. The same is true in general[5]: $\mathbb{Z}/n\mathbb{Z} = \mathbb{Z}_n$ for any integer $n > 0$.

[4] In multiplicative notation, this is called the (Frobenius) product and written AB.

[5] By the way, if you feel cheated by this example, as if it has said absolutely nothing, then GOOD!

Understanding quotient groups can be tricky. However, the First Isomorphism Theorem is a very useful way of realising any quotient group G/H as the image of G under a suitable homomorphism. The subgroup H that we \quotient out by" turns out to be the *kernel* of the homomorphism, which is the set of all elements which map to the identity element of the target:

Definition 4.23. Let $\phi : G \to H$ be a homomorphism. We define $\ker(\phi) = \{g \in G \mid \phi(g) = 0_H\} \subseteq G$ to be the *kernel* of ϕ, and $\mathrm{im}(\phi) = \{\phi(g) \mid g \in G\} \subseteq H$ to be the *image* of ϕ.

Lemma 4.24. For a homomorphism $\phi : G \to H$, $\ker(\phi)$ is a subgroup of G and $\mathrm{im}(\phi)$ a subgroup of H.

We know that $\phi(0_G) = 0_H$ for any homomorphism, so the kernel is never empty. When this is the only element that is mapped to 0_H, the homomorphism is injective:

Lemma 4.25. Let $\phi : G \to H$ be a homomorphism. Then ϕ is injective if and only if $\ker(\phi) = \{0_G\}$.

Projecting from the group G to the set of cosets G/H gives a homomorphism with kernel H:

Proposition 4.26. Let H be a subgroup of an abelian group G. Then $\phi : G \to G/H$ defined by $\phi(g) = H + g$ is a surjective homomorphism with kernel H.

The content of the First Isomorphism Theorem is that *every* quotient group appears in such a form:

Theorem 4.27 (First Isomorphism Theorem)**.** Let $\phi : G \to H$ be a homomorphism (of abelian groups) with kernel K. Then there is an isomorphism $\theta : G/K \to \mathrm{im}(\phi)$ defined by $\theta(K + g) = \phi(g)$ for all $g \in G$; that is, $G/K = \mathrm{im}(\phi)$.

That is, given any homomorphism ϕ, the quotient of a group by the kernel is isomorphic to the image.

4.4 Abelian Groups and Matrices Over $\mathbb{Z}$

Definition 4.28. A finitely generated abelian group is called *free abelian* if it is isomorphic to $\mathbb{Z}^n$ for some $n \geq 0$, where $\mathbb{Z}^n$ denotes the direct sum of n copies of $\mathbb{Z}$, $\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}$. (In the case $n = 0$, $\mathbb{Z}^0$ is the trivial group $\{0\}$.)

In order to classify all finitely generated abelian groups, we draw parallels between $\mathbb{Z}^n$ and $\mathbb{R}^n$. Firstly, we define a basis of $\mathbb{Z}^n$, and indeed of any abelian group.

Definition 4.29. Let G be an abelian group. The elements $x_1, \ldots, x_n \in G$ are called *linearly independent* if $\lambda_1 x_1 + \cdots + \lambda_n x_n = 0_G$ with $\lambda_i \in \mathbb{Z}$ implies $\lambda_1 = \cdots = \lambda_n = 0_{\mathbb{Z}}$. Furthermore, the elements $x_1, \ldots, x_n \in G$ form a *free basis* of G if and only if they are linearly independent and generate (span) G.

Lemma 4.30. $x_1, \ldots, x_n$ form a free basis of G if and only if every element $g \in G$ has a unique expression $g = \lambda_1 x_1 + \cdots + \lambda_n x_n$ with $\lambda_i \in \mathbb{Z}$.

For example, $\mathbf{x}_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ where the 1 occurs in the i^{th} place, with $1 \leq i \leq n$, forms a free basis of $\mathbb{Z}^n$. But be careful: $(1, 0)$, $(0, 2)$ does *not* form a free basis of $\mathbb{Z}^2$, since we cannot get $(0, 1)$ from an *integer* linear combination of $(1, 0)$, $(0, 2)$.

Proposition 4.31. An abelian group G is free abelian if and only if it has a free basis $x_1, \ldots, x_n$. If so, there is an isomorphism $\phi : G \to \mathbb{Z}^n$ given by $\phi(x_i) = \mathbf{x}_i$.

Just as in linear algebra, we can define change of basis matrices. Writing $\mathbf{y}_1, \ldots, \mathbf{y}_m \in \mathbb{Z}^n$ with $\mathbf{y}_j = \sum_{i=1}^{n} \alpha_{ij} \mathbf{x}_i$, it can be shown that $\mathbf{y}_1, \ldots, \mathbf{y}_m$ is a free basis of $\mathbb{Z}^n$ if and only if $n = m$ and P is an invertible matrix such that P^{-1} has entries in $\mathbb{Z}$, or equivalently $\det P = \pm 1$. An $n \times n$ matrix P with entries in $\mathbb{Z}$ is called *unimodular* if $\det P = \pm 1$. Let A be an $m \times n$ matrix over $\mathbb{Z}$. We define *unimodular elementary row and column operations* as follows:
(UR1) Replace some row $\mathbf{r}_i$ of A by $\mathbf{r}_i + t\mathbf{r}_j$, where $j \neq i$ and $t \in \mathbb{Z}$;
(UR2) Interchange two rows of A;
(UR3) Replace some row $\mathbf{r}_i$ of A by $-\mathbf{r}_i$.
(UC1), (UC2) and (UC3) are defined analogously for column operations.

Our strategy for classifying nitely-generated abelian groups is to express a group as a matrix, reduce that matrix to a \normal form" by means of row and column operations, and read o what the group is from that. The reason this works is the following theorem:

Theorem 4.32. Let A be an $m \times n$ matrix over $\mathbb{Z}$ with rank r. Then, by using a sequence of unimodular elementary row and column operations, we can reduce A to a matrix $B = (b_{ij})$ such that $b_{ii} = d_i$ for $1 \leq i \leq r$ and $b_{ij} = 0$ otherwise, where the integers d_i satisfy $d_i > 0$ and $d_i \mid d_{i+1}$ for $1 \leq i < r$. Furthermore, the d_i are uniquely determined by A.

The resulting form of the matrix is called *Smith Normal Form*. The strategy by which we do this is to reduce the size of entries in the rst row and column, until the top left entry divides all the other entries in the rst row and column, so we can turn them all to zeroes, and proceed to do the same to the second row and column and so on. An example is, as always, worth a thousand theorems.

Example 4.33. Let $A = \left(\begin{smallmatrix} 12 & 6 & 42 \\ 18 & 24 & 18 \end{smallmatrix}\right)$. The entry smallest in absolute value is seen to be 6, so we proceed as follows:

$$\begin{pmatrix} 12 & 6 & 42 \\ 18 & 24 & 18 \end{pmatrix} \mathbf{c}_2 \leftrightarrow \mathbf{c}_1 \begin{pmatrix} 6 & 12 & 42 \\ 24 & 18 & 18 \end{pmatrix} \mathbf{c}_2 \to \mathbf{c}_2 - 2\mathbf{c}_1, \mathbf{c}_3 \to \mathbf{c}_3 + 7\mathbf{c}_1 \begin{pmatrix} 6 & 0 & 0 \\ 24 & 30 & 150 \end{pmatrix}$$

$$\mathbf{r}_2 \to \mathbf{r}_2 - 4\mathbf{r}_1 \begin{pmatrix} 6 & 0 & 0 \\ 0 & 30 & 150 \end{pmatrix} \mathbf{c}_3 \to \mathbf{c}_3 + 5\mathbf{c}_2 \begin{pmatrix} 6 & 0 & 0 \\ 0 & 30 & 0 \end{pmatrix} \mathbf{c}_2 \to -\mathbf{c}_2 \begin{pmatrix} 6 & 0 & 0 \\ 0 & 30 & 0 \end{pmatrix}$$

Then $6 \mid 30$, and all the other entries are zero, so we are done.

To put matrices over $\mathbb{Z}$ to work classifying nitely-generated abelian groups, we need one more technical lemma:

Lemma 4.34. Any subgroup of a nitely generated abelian group is nitely generated.

So, how do we set up the link between nitely-generated abelian groups and matrices over $\mathbb{Z}$? Well, let H be a subgroup of the free abelian group $\mathbb{Z}^n$, and suppose that H is generated by $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{Z}^n$. Then we can represent H as a an $n \times m$ matrix A whose columns are $\mathbf{v}_1^{\mathsf{T}}, \ldots, \mathbf{v}_m^{\mathsf{T}}$.

Now, what e ect does applying unimodular row and column operations to A have on the subgroup H? Well, since unimodular elementary row operations are invertible, when we apply unimodular elementary row operations to A, we may apply the inverse of the operation to the free basis of $\mathbb{Z}^n$; thus the resulting matrix represents the same subgroup H of $\mathbb{Z}^n$ using a di erent free basis of $\mathbb{Z}^n$.

The unimodular elementary column operations (UC1), (UC2) and (UC3) respectively amount to replacing a generator $\mathbf{v}_i$ by $\mathbf{v}_i + t\mathbf{v}_j$, interchanging two generators, and changing $\mathbf{v}_i$ to $-\mathbf{v}_i$. All of these operations do not change the subgroup being generated. Summing up, we have:

Proposition 4.35. Let $A, B \in \mathbb{Z}^{n,m}$, and suppose B is obtained from A by a sequence of unimodular row and column operations. Then the subgroups of $\mathbb{Z}^n$ represented by A and B are the same, using (possibly) di erent free bases of $\mathbb{Z}^n$.

In particular, given a subgroup H of $\mathbb{Z}^n$, we may reduce its matrix A to Smith Normal Form:

Theorem 4.36. Let H be a subgroup of $\mathbb{Z}^n$. Then there exists a free basis $\mathbf{y}_1, \ldots, \mathbf{y}_n$ of $\mathbb{Z}^n$ such that $H = \langle d_1\mathbf{y}_1, \ldots, d_r\mathbf{y}_r \rangle$, where each $d_i > 0$ and $d_i \mid d_{i+1}$ for all $1 \leq i < r$.

Example 4.37. Let $H = \langle 12\mathbf{x}_1 + 18\mathbf{x}_2, 6\mathbf{x}_1 + 24\mathbf{x}_2, 42\mathbf{x}_1 + 18\mathbf{x}_2 \rangle$. The associated matrix is $A = \left(\begin{smallmatrix} 12 & 6 & 42 \\ 18 & 24 & 18 \end{smallmatrix}\right)$. As above, its Smith normal form is $\left(\begin{smallmatrix} 6 & 0 & 0 \\ 0 & 30 & 0 \end{smallmatrix}\right)$. Looking at the row operations we performed, we did $\mathbf{r}_2 \to \mathbf{r}_2 - 4\mathbf{r}_1$, so we must do the inverse operation to the basis matrix to get the new basis matrix; thus doing $\mathbf{r}_2 \to \mathbf{r}_2 + 4\mathbf{r}_1$ to $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ yields $\left(\begin{smallmatrix} 1 & 0 \\ 4 & 1 \end{smallmatrix}\right)$. Thus, letting $\mathbf{y}_1 = (1, 4)$ and $\mathbf{y}_2 = (0, 1)$, we see that $H = \langle 6\mathbf{y}_1, 30\mathbf{y}_2 \rangle$. (Check it!)

Now, let $G = \langle x_1, \ldots, x_n \rangle$ be any nitely-generated abelian group, and let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be the standard free basis of $\mathbb{Z}^n$. Then by de ning $\varphi : \mathbb{Z}^n \to G$ by the following (where $\lambda_i \in \mathbb{Z}$):

$$\varphi(\lambda_1\mathbf{x}_1 + \cdots + \lambda_n\mathbf{x}_n) = \lambda_1 x_1 + \cdots + \lambda_n x_n,$$

we get an epimorphism onto G. By the First Isomorphism Theorem, $G = Z^n/K$, where $K = \ker(\ )$. By definition, $K = f(\ _1;\dots;\ _n) \in Z^n \mid \ _1x_1 + \dots + \ _nx_n = 0_G g$. Since this is a subgroup of a finitely-generated abelian group, it is finitely generated by lemma 4.34, say $K = \langle \mathbf{v}_1;\dots;\mathbf{v}_m \rangle$ for $\mathbf{v}_i \in Z^n$. The notation[6] $\langle \mathbf{x}_1;\dots;\mathbf{x}_n \mid \mathbf{v}_1;\dots;\mathbf{v}_m \rangle$ is used to denote Z^n/K, so we have $G = \langle \mathbf{x}_1;\dots;\mathbf{x}_n \mid \mathbf{v}_1;\dots;\mathbf{v}_m \rangle$. Applying theorem 4.36 to K we see there is a free basis $\mathbf{y}_1;\dots;\mathbf{y}_n$ of Z^n such that $K = \langle d_1\mathbf{y}_1;\dots;d_r\mathbf{y}_r \rangle$, where each $d_i > 0$ and $d_i \mid d_{i+1}$ for all $1 \quad i < r$, so $G = \langle \mathbf{y}_1;\dots;\mathbf{y}_n \mid d_1\mathbf{y}_1;\dots;d_r\mathbf{y}_r \rangle$. It can be shown that $\langle \mathbf{y}_1;\dots;\mathbf{y}_n \mid d_1\mathbf{y}_1;\dots;d_r\mathbf{y}_r \rangle = Z_{d_1} \quad Z_{d_r} \quad Z^{n\ r}$. Hence we end up at the main theorem:

Theorem 4.38 (Fundamental Theorem of Finitely Generated Abelian Groups)**.** If $G = \langle x_1;\dots;x_n \rangle$ is a finitely-generated abelian group, then for some $r \quad n$ there are $d_1;\dots;d_r \in Z$ such that each $d_i > 0$ with $d_1 > 1$ and $d_i \mid d_{i+1}$ for all $1 \quad i < r$, such that

$$G = Z_{d_1} \quad Z_{d_r} \quad Z^{n\ r}:$$

and the integers r, $d_1;\dots;d_r$ are uniquely determined.

Example 4.39. Let $G = \langle \mathbf{x}_1;\mathbf{x}_2 \mid 12\mathbf{x}_1 + 18\mathbf{x}_2;\ 6\mathbf{x}_1 + 24\mathbf{x}_2;\ \ 42\mathbf{x}_1 \quad 18\mathbf{x}_2 \rangle$. The associated matrix is $A = \left(\begin{smallmatrix} 12 & 6 & \ \ 42 \\ 18 & 24 & \ \ 18 \end{smallmatrix}\right)$, and its Smith normal form is $\left(\begin{smallmatrix} 6 & 0 & 0 \\ 0 & 30 & 0 \end{smallmatrix}\right)$. Writing $H = \langle 6\mathbf{y}_1;30\mathbf{y}_2 \rangle$, we may simply read off that $G = \langle \mathbf{x}_1;\mathbf{x}_2 \rangle / H = \langle \mathbf{y}_1;\mathbf{y}_2 \mid 6\mathbf{y}_1;30\mathbf{y}_2 \rangle = Z_6 \quad Z_{30}$, a finite group of order 180.

In addition, the classification theorem allows us to classify abelian groups of a given finite order n (up to isomorphism), since they correspond to decompositions $n = d_1d_2\dots d_r$ where $d_i \mid d_{i+1}$ for all i:

Example 4.40. Let G be an abelian group of order 360. Then, as $360 = 2^3 \quad 3^2 \quad 5$, the possible decompositions are 360, $2 \quad 180$, $2 \quad 2 \quad 90$, $2 \quad 6 \quad 30$, $3 \quad 120$, and $6 \quad 60$. Thus G must be isomorphic to one of Z_{360}, $Z_2 \quad Z_{180}$, $Z_2 \quad Z_2 \quad Z_{90}$, $Z_2 \quad Z_6 \quad Z_{30}$, $Z_3 \quad Z_{120}$, or $Z_6 \quad Z_{60}$.

To prove none of these are isomorphic, one uses orders of elements: Z_{360} has an element of order 360, but none of the others do, so it cannot be isomorphic to any of the others; then $Z_2 \quad Z_{180}$ has an element of order 180, which the other four do not, and so on.

[6] In general $\langle x_i \mid y_j \rangle$ is called a *presentation* of a group, and it denotes the free group generated by the x_i, quotiented by the normal subgroup generated by the y_j. Since we are dealing only with abelian groups, we are abusing notation to mean the free *abelian* group generated by the x_i, quotiented by the abelian subgroup generated by the y_j.

This guide would not be possible without our wonderful sponsors: