Matrix Normal Forms

E. Speer

Introduction to Mathematics at Rutgers

September 2008

I. INTRODUCTION TO THE PROBLEM

A. Example: Gauss-Jordan elimination

Problem: solve a system $\boxed{A\vec{x} = \vec{b}}$ of m equations in n unknowns, with coefficients in a field $\mathbb{F}$. Here

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \qquad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} \qquad \vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

Method of solution: *row reduction* of the augmented matrix $[A \mid \vec{b}]$:

$$B = [A \mid \vec{b}] = \left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right] \xrightarrow[\text{operations}]{\text{elementary row}} R,$$

to obtain a matrix R in *reduced row-echelon form (RREF)*.

The relation between B and R is called *row equivalence*. Alternatively, R and B are row equivalent if $\boxed{R = PB}$ with P $m \times m$ invertible.

A matrix in RREF looks like this ($\bullet$ represents any number):

$$\left[\begin{array}{cccccccccc}
0 & \multicolumn{1}{|c}{\mathbf{1}} & \bullet & \bullet & 0 & 0 & \bullet & 0 & \bullet & \cdots \\ \cline{2-4}
0 & 0 & 0 & 0 & \multicolumn{1}{|c}{\mathbf{1}} & 0 & \bullet & 0 & \bullet & \cdots \\ \cline{5-5}
0 & 0 & 0 & 0 & 0 & \multicolumn{1}{|c}{\mathbf{1}} & \bullet & 0 & \bullet & \cdots \\ \cline{6-7}
0 & 0 & 0 & 0 & 0 & 0 & 0 & \multicolumn{1}{|c}{\mathbf{1}} & \bullet & \cdots \\ \cline{8-9}
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots
\end{array}\right]$$

More precisely, a matrix R is in reduced row-echelon form if

- All nonzero rows lie above any zero rows;
- The first nonzero entry (the *pivot*) in any nonzero row is a 1 (this distinguishes the *pivot columns* $j_1 < \cdots < j_r$);
- Each pivot lies to the right of the pivot in the row above it;
- All matrix entries above a pivot are zero.

Remark: the number of nonzero rows in R is the *rank* of R—or of B, since row equivalent matrices have the same rank.

B. A normal form theorem

Theorem: *Every* $m \times n$ *matrix is row equivalent to a unique matrix in reduced row-echelon form.*

This is the standard setup in which *normal* or *canonical forms* for matrices arise. We have:

- An equivalence relation on a set $\mathcal{S}$ of matrices;
- A special class $\mathcal{N}$ of matrices within the set, the *matrices in normal form*, such that every matrix in $\mathcal{S}$ is equivalent to a unique (or essentially unique) matrix in $\mathcal{N}$.

Thus the RREF is a normal form for rectangular matrices under row equivalence.

C. Finish discussion of linear equations

Consider again $A\vec{x} = \vec{b}$. Suppose under row reduction

$$[A \mid \vec{b}] \longrightarrow [S \mid \vec{c}] = \left[\begin{array}{cccccccccc|c}
x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & & x_{n-1} & x_n & \\
0 & 1 & \bullet & \bullet & 0 & 0 & \bullet & \cdots & 0 & \bullet & c_1 \\
0 & 0 & 0 & 0 & 1 & 0 & \bullet & \cdots & 0 & \bullet & c_2 \\
0 & 0 & 0 & 0 & 0 & 1 & \bullet & \cdots & 0 & \bullet & c_3 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & \bullet & c_r \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & c_{r+1} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0
\end{array}\right]$$

The equations $S\vec{x} = \vec{c}$ are equivalent to the original $A\vec{x} = \vec{b}$. Here c_{r+1} is 0 or 1. $\boxed{\text{There is no solution if } c_{r+1} = 1.}$

If $c_{r+1} = 0$, solve for the variables $x_2, x_5, x_6, \ldots, x_{n-1}$ corresponding to the pivot columns; the other variables $x_1, x_3, x_4, x_7, \ldots, x_n$ are free parameters.

D. Linear equations in abstract vector spaces

V, W finite dimensional vector spaces over $\mathbb{F}$, $L : V \to W$ linear.

$$\boxed{\text{Problem: Given } \mathbf{b} \in W \text{ find } \mathbf{x} \in V \text{ with } L\mathbf{x} = \mathbf{b}.}$$

To solve this problem, introduce:

- *Ordered bases:* $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ for V and $(\mathbf{w}_1, \ldots, \mathbf{w}_m)$ for W
- *Coordinates in each vector space:* If $\mathbf{x} \in V$ has the expansion $\mathbf{x} = \sum_{i=1}^{n} x_i \mathbf{v}_i$, and $\mathbf{b} \in W$ the expansion $\mathbf{b} = \sum_{i=1}^{m} b_i \mathbf{w}_i$, then $\mathbf{x}$ and $\mathbf{b}$ have *coordinate vectors*

$$\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \text{and} \quad \vec{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}.$$

- *The matrix A of the linear operator L:*

$$L\mathbf{v}_i = \sum_{j=1}^{m} a_{ji} \mathbf{w}_j, \; i = 1, \ldots, n$$

- Then $\boxed{L\mathbf{x} = \mathbf{b} \leftrightarrow A\vec{x} = \vec{b}}$ and we solve the problem as above.

Notation: It is convenient to write the bases as column vectors whose components are abstract vectors: $\vec{\mathbf{v}} = \begin{bmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_n \end{bmatrix}$, $\vec{\mathbf{w}} = \begin{bmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_m \end{bmatrix}$.

Then $\mathbf{x} = \vec{\mathbf{v}}^T \vec{x}$, $\mathbf{b} = \vec{\mathbf{w}}^T \vec{b}$ and and $L\mathbf{x} = \vec{\mathbf{w}}^T A \vec{x}$.

Basis change. Consider different choices of bases, say $(\mathbf{v}'_1, \ldots, \mathbf{v}'_n)$ and $(\mathbf{w}'_1, \ldots, \mathbf{w}'_m)$, where $\mathbf{v}'_i = \sum_{j=1}^{n} q_{ji} \mathbf{v}_j$ and $\mathbf{w}'_i = \sum_{j=1}^{m} p_{ji} \mathbf{w}_j$, or

$$\boxed{\vec{\mathbf{v}}' = Q^T \vec{\mathbf{v}} \text{ and } \vec{\mathbf{w}}' = P^T \vec{\mathbf{w}}}\ .$$

There is a corresponding change of coordinates:

$$\mathbf{x} = \vec{\mathbf{v}}'^T \vec{x}' = \vec{\mathbf{v}}^T Q \vec{x}' \quad \text{so} \quad \boxed{\vec{x} = Q\, \vec{x}'}\ .$$

Similarly,

$$\boxed{\vec{b} = P\, \vec{b}'} \quad \text{and} \quad \boxed{A = P\, A'\, Q^{-1}}\ .$$

E. Matrix normal forms in terms of linear transformations

Again look at $L : V \to W$. Consider for the moment a basis change only in W: $\vec{\mathbf{w}}' = P^T\vec{\mathbf{w}}$. Then the matrices A and A' representing L are related by $A = PA'$, that is, they are *row equivalent*. Thus we have the

Theorem: *Given a linear mapping* $L : V \to W$ *and a fixed basis* $\vec{\mathbf{v}}$ *for* V *there is a choice of basis* $\vec{\mathbf{w}}$ *for* W *such that the matrix representing* L *in these bases is in RREF. Moreover, that matrix is uniquely determined by* L*.*

- This is the second way in which normal forms arise in linear algebra: we have a (fixed) abstract object, such as a linear transformation, and we want to choose a basis (or bases) in which the matrix representing it has a standard form.

- Our previous formulation may be looked at as a special case of this one: take $V = \mathbb{F}^n$ and $W = \mathbb{F}^m$, and identify $L : V \to W$ with the matrix A such that $T\vec{x} = A\vec{x}$.

II. Normal forms for $L : V \to W$ or for an $m \times n$ matrix

Equivalence relation on matrices (P, Q invertible)	Basis choice in vector space	Normal form	Invariants
Row equivalence $A \sim PA$	Choice of basis in W	Reduced row echelon form	r, $j_1, \ldots, j_r$ • entries
Column equivalence $A \sim AQ$	Choice of basis in V	Reduced column echelon form	r, $i_1, \ldots, i_r$ • entries
Equivalence $A \sim PAQ$	Choice of basis in both V and W	$\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$	r

III. Normal forms for $L : V \to V$

- V an n dimensional vector space, basis $\vec{\mathbf{v}}^T = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$
- $L : V \to V$ linear with matrix A in this basis: $L\mathbf{x} = \vec{\mathbf{v}}^T A \vec{x}$.

Problem: Find a normal form for the matrix representing L in a suitable basis.

- The matrix representing L in a new basis $\vec{\mathbf{v}}' = Q^T \vec{\mathbf{v}}$ is then $A' = Q^{-1} A\, Q$. We say that A and A' are *similar*.

Equivalent problem: Find a normal form for $n \times n$ matrices under the equivalence relation of similarity.

- In this section we will have to pay attention to the field $\mathbb{F}$ over which our spaces are defined. We will always take $\mathbb{F}$ to be either $\mathbb{R}$ or $\mathbb{C}$.

A. Eigenvalues and eigenvectors

- A *nonzero* vector $\mathbf{x} \in V$ is an *eigenvector* of L, with *eigenvalue* λ, if $L\mathbf{x} = \lambda\mathbf{v}$.

- Suppose we can find a basis $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ of V such that each $\mathbf{v}_i$ is an eigenvalue of L: $L\mathbf{v}_i = \lambda_i\mathbf{v}_i$. Then the matrix A of L in this basis is *diagonal*:

$$A = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

We say that L has been *diagonalized*. Thus if we could always find such a basis our normal form would be a diagonal matrix. But in fact we cannot do so in general.

- In the alternate language: let A be an $n \times n$ matrix. We ask if A can be *diagonalized*, i.e., if there exists an matrix Q such that $A' = Q^{-1} A Q$ is diagonal. But *not all matrices can be diagonalized.*

Diagonal matrices are so pleasant to encounter that we add a new problem to our list:

Problem: Find special classes of linear transformations such that every memeber of the class can be diagonalized.

Remark on uniqueness: Recall that the eigenvalues of a matrix A are precisely the roots of the characteristic polynomial

$$p_A(\lambda) = \det(A - \lambda I).$$

Moreover, if A is diagonalizable then the number of times each eigenvalue occurs in the diagonal form is equal to its multiplicity as a root of the characteristic polynomial. Thus the diagonal form is *unique up to a rearrangement of the diagonal entries.*

- This is an example of what we referred to earlier as an "essentially unique" normal form.

- If $\mathbb{F} = \mathbb{C}$ then L always has at least one eigenvalue; *this may not be true if* $\mathbb{F} = \mathbb{R}$.

B. Normal form for general $L : V \to V$ over $\mathbb{C}$

Theorem: *(a) Suppose that V is a vector space over $\mathbb{C}$, the field of complex numbers. Then every linear transformation $L : V \to V$ is represented in an approriately chosen basis by an essentially unique matrix in Jordan canonical form.*
(b) Equivalently, every complex matrix is similar to a matrix in Jordan canonical form.

Idea: If λ is an eigenvalue of L then we may define the *eigenspace*

$$V_\lambda = \{\,\mathbf{x} \in V \mid (L - \lambda I)\mathbf{x} = 0\,\}.$$

L is diagonalizable if and only if $V = \bigoplus_\lambda V_\lambda$, that is, if and only if every $\mathbf{x} \in V$ can be written as a sum $\mathbf{x} = \sum_\lambda \mathbf{x}_\lambda$, with $\mathbf{x}_\lambda \in V_\lambda$.

Remark: The notation $V = \bigoplus_\lambda V_\lambda$ means that V is the *direct sum* of the subspaces V_λ. In fact, this requires that every $\mathbf{x} \in V$ can be written *uniquely* as a sum $\mathbf{x} = \sum_\lambda \mathbf{x}_\lambda$, with $\mathbf{x}_\lambda \in V_\lambda$. When the V_λ are the eigenspaces, however, this uniqueness is automatically satisfied.

We may also define the *generalized eigenspace*

$$\widehat{V}_\lambda = \{\, \mathbf{x} \in V \mid (L - \lambda I)^k \mathbf{x} = 0 \text{ for some } k \in \mathbb{N} \,\}.$$

Now always $V = \bigoplus_\lambda \widehat{V}_\lambda$; moreover, $\widehat{V}_\lambda$ is *invariant* under L: $L(\widehat{V}_\lambda) \subset \widehat{V}_\lambda$. So it suffices to study the restriction of L to $\widehat{V}_\lambda$.

We know that for each $\mathbf{x} \in \widehat{V}_\lambda$ there is an integer k, which *a priori* may depend on $\mathbf{x}$, such that $(L - \lambda I)^k \mathbf{x} = 0$. Because V is finite dimensional, however, there must be some one k which has this property for *all* $\mathbf{x} \in \widehat{V}_\lambda$. That is, when restricted to $\widehat{V}_\lambda$, L satisfies $(L - \lambda I)^k = 0$ for some k. How could this happen?

Suppose there are vectors $\mathbf{v}_1^*, \ldots, \mathbf{v}_j^* \in \widehat{V}_\lambda$, with $j \le k$, so that

$$(L - \lambda I)\mathbf{v}_j^* = \mathbf{v}_{j-1}^*, \quad (L - \lambda I)\mathbf{v}_{j-1}^* = \mathbf{v}_{j-2}^*, \quad \ldots \quad (L - \lambda I)\mathbf{v}_1^* = 0.$$

Then $(L-\lambda I)^j \mathbf{v}_i^* = 0 = (L-\lambda I)^k \mathbf{v}_i^*$ for any i. Let V^* be the subspace of V spanned by these vectors; in that subspace, and using $\mathbf{v}_1^*, \ldots, \mathbf{v}_j^*$ as a basis, L has matrix

$$J^\lambda = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 1 & \cdots & 0 & 0 \\ 0 & 0 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{bmatrix}.$$

A matrix of this form is called a *Jordan block*.

Fact: We can always find a basis of $\widehat{V}$ consisting of one or more sequences $\mathbf{v}_1^*, \ldots, \mathbf{v}_j^*$ of vectors of the above type.

Let L have *distinct* eigenvalues $\lambda_1, \ldots \lambda_m$ then we may find a basis for $\widehat{V}_{\lambda_i}$ in which the restriction of L to $\widehat{V}_{\lambda_i}$ has the *block diagonal* form

$$K_i = \begin{bmatrix} J_1^{\lambda_1} & & \\ & \ddots & \\ & & J_{r_1}^{\lambda_1} \end{bmatrix}, \qquad (*)$$

with each $J_r^{\lambda_1}$ a Jordan block. In the basis for V which is the union of all of these bases, L has the form

$$A = \begin{bmatrix} K_1 & & \\ & \ddots & \\ & & K_m \end{bmatrix}. \qquad (**)$$

This is the Jordan form for L.

Uniqueness: L uniquely determines the *number* and *sizes* of the Jordan blocks associated with each eigenvalue. (These, and the eigenvalues themselves, are the invariants.) One may by convention group together all blocks associated with the same eigenvalue, as in $(**)$, and arrange the blocks in each K_i (see $(*)$) in order of (say) decreasing size. However, the order of the K_i in $(**)$ is not unique.

Remark: L satisfies its own characteristic equation, i.e., $p_L(L) = 0$. The monic polynomial $q_L(\lambda)$ of lowest degree for which $q_L(L) = 0$ is called the *minimal poynomial* of L; q_L divides p_L. In fact, if k_i is the size of the matrix K_i—the dimension of $\widehat{V}_{\lambda_i}$—and ℓ_i is the size of the largest Jordan block J^{λ_i} appearing in the Jordan form of L, then

$$p_L(\lambda) = \prod_{i=1}^{m} (\lambda - \lambda_i)^{k_i} \qquad \text{and} \qquad q_L(\lambda) = \prod_{i=1}^{m} (\lambda - \lambda_i)^{\ell_i}.$$

Example: if the Jordan form is (all blank entries are 0)

$$\begin{bmatrix} 3 & 1 & & & & & & \\ & 3 & 1 & & & & & \\ & & 3 & & & & & \\ & & & 3 & 1 & & & \\ & & & & 3 & & & \\ & & & & & 3 & & \\ & & & & & & 0 & \\ & & & & & & & 0 \end{bmatrix}$$

then $p(\lambda) = \lambda^2(\lambda - 3)^6$ and $q(\lambda) = \lambda(\lambda - 3)^3$.

C. Computation with matrices

The Jordan form is useful for computing powers and exponentials of matrices. For example, suppose we want to solve a linear system of differential equations for an unknown vector function $\vec{x}(t)$:

$$\frac{d\vec{x}}{dt} = A\,\vec{x}, \qquad \vec{x}(0) = \vec{x}_0,$$

where A (an $n \times n$ matrix) and $\vec{x}_0$ are given. Formally the solution is $\vec{x}(t) = e^{tA}\vec{x}_0$, and the matrix exponential may be defined by a power series, but we would like to have an effective way to compute it.

Suppose $A = Q^{-1}KQ$ with K in Jordan form; then

$$e^{tA} = \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k = \sum_{k=0}^{\infty} \frac{t^k}{k!} (Q^{-1}KQ)^k = \sum_{k=0}^{\infty} \frac{t^k}{k!} Q^{-1}K^kQ = Q^{-1}e^{tK}Q,$$

so we need only compute e^{tK}.

But

$$K = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_r \end{bmatrix} \quad \Longrightarrow \quad e^{tK} = \begin{bmatrix} e^{tJ_1} & & \\ & \ddots & \\ & & e^{tJ_r} \end{bmatrix}$$

so we need only to compute the exponential of a single Jordan block.

This turns out to be an easy exercise. Suppose for example that J is an $n \times n$ Jordan block with eigenvalue λ, then

$$e^{tJ} = e^{t\lambda} \begin{bmatrix} 1 & t & \frac{t^2}{2!} & \cdots & \frac{t^{n-1}}{(n-1)!} \\ 0 & 1 & t & \cdots & \frac{t^{n-2}}{(n-2)!} \\ 0 & 0 & 1 & \cdots & \frac{t^{n-3}}{(n-3)!} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

D. Normal form for general $L : V \to V$ over $\mathbb{R}$

Here the normal form is *real canonical form*; this is less important than the Jordan canonical form, and we will be brief. Some eigenvalues of L may be real; others come in complex conjugate pairs $\alpha \pm i\beta$. The normal form again has blocks along the diagonal:

- Corresponding to the real eigenvalues there are Jordan blocks;
- Corresponding to the eigenvalues $\alpha \pm i\beta$ there are blocks

$$\begin{bmatrix} \Lambda & I & & & & \\ & \Lambda & I & & & \\ & & \Lambda & \ddots & & \\ & & & \ddots & I & \\ & & & & \Lambda & I \\ & & & & & \Lambda \end{bmatrix}.$$

where

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \Lambda = \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}.$$

E. When is $L : V \to V$, or a matrix A, necessarily diagonalizable?

Let us first consider matrices. We recall some terminology:

- The *adjoint* A^* of a complex matrix A is the transpose of its complex conjugate: $A^* = \overline{A}^T$. A is *symmetric* if $A^T = A$ and *self-adjont* or *Hermitian symmetric* if $A^* = A$.

- A real matrix P is *orthogonal* if $P^T = P^{-1}$, and a complex matrix P is *unitary* if $P^* = P^{-1}$. We say that a matrix A is *diagonalizable by an orthogonal (or unitary) matrix* if $P^{-1}AP$ is diagonal for some orthogonal (unitary) P.

Theorem: *(a) If A is a real symmetric $n \times n$ matrix then all the eigenvalues of A are real and A is diagonalizable by an orthogonal matrix.*
(b) If A is a self-adjoint $n \times n$ matrix then all the eigenvalues of A are real and A is diagonalizable by a unitary matrix.

To state the corresponding result for $L : V \to V$ we assume that V is an *inner product space*.

- Recall: for $\mathbf{x}, \mathbf{y} \in V$ the inner product $\langle \mathbf{x}, \mathbf{y} \rangle$ lies in $\mathbb{F}$; the inner product is *linear in its first argument*, *positive definite* ($\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$, with equality only if $\mathbf{x} = 0$), and *symmetric* ($\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$) if $\mathbb{F} = \mathbb{R}$, *Hermitian symmetric* ($\langle \mathbf{x}, \mathbf{y} \rangle = \overline{\langle \mathbf{y}, \mathbf{x} \rangle}$) if $\mathbb{F} = \mathbb{C}$.

- Note that when $\mathbb{F} = \mathbb{C}$ the inner product is *antilinear* in its second argument: $\langle \mathbf{x}, a\mathbf{y} + b\mathbf{z} \rangle = \overline{a} \langle \mathbf{x}, \mathbf{y} \rangle + \overline{b} \langle \mathbf{y}, \mathbf{z} \rangle$.

- A basis for $\mathbf{v}_1, \ldots, \mathbf{v}_n$ for V is *orthnormal* if $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \delta_{ij}$.

- If $\vec{\mathbf{v}}$ and $\vec{\mathbf{v}}' = P^T \vec{\mathbf{v}}$ are two orthonormal bases then P is (i) orthogonal if $\mathbb{F} = \mathbb{R}$ and (ii) unitary if $\mathbb{F} = \mathbb{C}$.

- If $\langle L\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, L\mathbf{y} \rangle$ for all $\mathbf{x}, y \in V$ then L is *symmetric* when $\mathbb{F} = \mathbb{R}$ and *self-adjoint* or *Hermitian symmetric* when $\mathbb{F} = \mathbb{C}$.

- If $\vec{\mathbf{v}}$ is an orthonormal basis then the matrix A of L in this basis is (i) symmetric if $\mathbb{F} = \mathbb{R}$ and (ii) self-adjoint if $\mathbb{F} = \mathbb{C}$.

Thus the previous theorem is equivalent to

Theorem: *Suppose that either $\mathbb{F} = \mathbb{R}$ and L is symmetric, or that $\mathbb{F} = \mathbb{C}$ and L is self-adjoint. Then all the eigenvalues of L are real and there exists an orthnormal basis for V composed of eigenvectors of L.*

We will sketch the proof of this theorem for the case $\mathbb{F} = \mathbb{C}$.

- First we show that all the eigenvalues of L are real. If $L\mathbf{x} = \lambda\mathbf{x}$ then

$$\begin{aligned} \langle L\mathbf{x}, \mathbf{x}\rangle &= \langle \lambda\mathbf{x}, \mathbf{x}\rangle = \lambda\langle \mathbf{x}, \mathbf{x}\rangle \\ &= \langle \mathbf{x}, L\mathbf{x}\rangle = \langle \mathbf{x}, \lambda\mathbf{x}\rangle = \overline{\lambda}\langle \mathbf{x}, \mathbf{x}\rangle \end{aligned}$$

Thus $\lambda\langle \mathbf{x}, \mathbf{x}\rangle = \overline{\lambda}\langle \mathbf{x}, \mathbf{x}\rangle$ and since $\langle \mathbf{x}, \mathbf{x}\rangle \neq 0$, $\lambda = \overline{\lambda}$.

- Now we use induction on the dimension of V to find an orthonormal basis of eigenvectors. Such a basis obviously exists if $\dim V = 1$. If $\dim V = d > 1$ then we know that L has at least one eigenvalue and corresponding eigenvector, say $L\mathbf{x} = \lambda\mathbf{x}$. Let W be the subspace of V consisting of all vectors orthogonal to $\mathbf{x}$:

$$W = \{\, \mathbf{y} \in V \mid \langle \mathbf{x}, \mathbf{y} \rangle = 0 \,\}.$$

If $\mathbf{y} \in W$ then $L\mathbf{y} \in W$, since

$$\langle \mathbf{x}, L\mathbf{y} \rangle = \langle L\mathbf{x}, \mathbf{y} \rangle = \langle \lambda\mathbf{x}, \mathbf{y} \rangle = \lambda\langle \mathbf{x}, \mathbf{y} \rangle = 0.$$

- Thus the restriction L_W of L to W is a mapping $L_W : W \to W$; L_W is easily seen to be self-adjoint and thus there is an orthonormal basis for W consisting of eigenvectors of L_W. Adding $\mathbf{x}$ to this basis gives us the require basis of V. ■

There is an important consequence. A (complex) matrix A is *normal* if A and its adjoint A^* commute: $AA^* = A^*A$.

Corollary: *If A is a normal $n \times n$ matrix then A is diagonalizable by a unitary matrix.*

Proof sketch: Let $A_R = (A + A^*)/2$ and $A_I = (A - A^*)/2i$; it is easy to see that A_R and A_I are self-adjoint and that $A = A_R + i\,A_I$. Because A and A^* commute, so do A_R and A_I. We know that we may diagonalize each of these separately with a unitary matrix, but *because they commute we can find a single unitary matrix that diagonalizes both*, and that is all we need.

To understand the key step here we may first diagonalize A_R. Let V_λ be an eigenspace of A_R; from $A_R A_I = A_I A_R$ we find that A_I maps V_λ into itself, and therefore can be diagonalized within V_λ. Putting together these diagonalizations in each eigenspace of A_R gives us the joint diagonalization of A_R and A_I.

IV. Bilinear and quadratic forms

Let V be a finite dimensional vector space; for the moment take $\mathbb{F} = \mathbb{R}$.

- A *bilinear form* B on V is a mapping $B : V \times V \to \mathbb{F}$ which is linear in each argument:

$$\begin{aligned} B(a\mathbf{x} + b\mathbf{y}, \mathbf{z}) &= aB(\mathbf{x}, \mathbf{z}) + bB(\mathbf{y}, \mathbf{z}) \\ B(\mathbf{x}, a\mathbf{y} + b\mathbf{z}) &= aB(\mathbf{x}, \mathbf{y}) + bB(\mathbf{x}, \mathbf{z}) \end{aligned}$$

B is *symmetric* if $B(\mathbf{x}, \mathbf{y}) = B(\mathbf{y}, \mathbf{x})$ for all $\mathbf{x}, \mathbf{y} \in V$ and *skew symmetric* or *alternating* if $B(\mathbf{x}, \mathbf{y}) = -B(\mathbf{y}, \mathbf{x})$ for all $\mathbf{x}, \mathbf{y}$.

- A *quadratic form* Q on V is a mapping $Q : V \to \mathbb{F}$ obtained from a symmetric bilinear form B via $Q(\mathbf{x}) = B(\mathbf{x}, \mathbf{x})$.

- Remark: one may give an intrinsic definition of a quadratic form and then show that, *because $\mathbb{F}$ does not have characteristic 2*, quadratic forms and symmetric bilinear forms are equivalent in the sense above. Over a field of characteristic 2 these are not equivalent concepts.

- If B is a bilinear form on V and $\mathbf{v}_1, \dots, \mathbf{v}_n$ is a basis for V then one may associate with B the matrix A defined by $a_{ij} = B(\mathbf{v}_i, \mathbf{v}_j)$.
- If $\mathbf{x}$ and $\mathbf{y}$ are vectors in V and $\vec{x}$ and $\vec{y}$ their coordinate vectors ($\mathbf{x} = \vec{\mathbf{v}}^T \vec{x}$, $\mathbf{y} = \vec{\mathbf{v}}^T \vec{y}$) then $B(\mathbf{x}, \mathbf{y}) = \vec{x}^T A \vec{y}$. B is symmetric if $A = A^T$ and skew symmetric if $A = -A^T$.
- Under a change of basis $\vec{\mathbf{v}}' = P^T \vec{\mathbf{v}}$ we have $\vec{x} = P\vec{x}'$, $\vec{y} = P\vec{y}'$, and so $B(\mathbf{x}, \mathbf{y}) = \vec{x}^T \, A \, \vec{y} = \vec{x}'^T \, P^T \, A \, P \, \vec{y}'$. Thus

$$\boxed{A' = P^T A P} \qquad (*)$$

When A and A' are related as in $(*)$, with P an invertible matrix, we say that they are *congruent*.

Problem: Find normal forms for real symmetric matrices, and for real skew symmetric matrices, under the equivalence relation of congruence.

Equivalent problem: Find normal forms for symmetric and skew symmetric bilinear forms, on a real vector space, under change of basis.

A. Symmetric bilinear forms

- Let A be a real symmetric matrix. We know that A can be diagonalized by an orthogonal matrix P:

$$P^{-1}AP = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}.$$

Since $P^{-1} = P^T$, this implies that every symmetric matrix is congruent, via an orthogonal matrix, to a diagonal matrix. The matrix entries on the diagonal are of course the eigenvalues of A.

- Equivalently: if B is a bilinear form on a real vector space V then there is an orthonormal basis $\vec{\mathbf{v}}$ for V such that for any $\mathbf{x}, \mathbf{y} \in V$

$$B(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} \lambda_i x_i y_i,$$

where as usual $\mathbf{x} = \vec{\mathbf{v}}^T \vec{x}$ and $\mathbf{y} = \vec{\mathbf{v}}^T \vec{y}$.

- If we ask for a normal form under congruence by an arbitrary invertible matrix then we can go further. If A is symmetric we may first transform A into a diagonal matrix D, as on the previous slide; suppose that the eigenvalues are arranged so that

$$\lambda_1, \ldots, \lambda_r > 0, \qquad \lambda_{r+1}, \ldots, \lambda_{r+s} < 0, \qquad \lambda_{r+s+1} = \cdots = \lambda_n = 0.$$

Now let R be the diagonal matrix

$$R = \begin{bmatrix} \lambda_1^{-1/2} & & & & & & & & \\ & \ddots & & & & & & & \\ & & \lambda_r^{-1/2} & & & & & & \\ & & & |\lambda_{r+1}|^{-1/2} & & & & & \\ & & & & \ddots & & & & \\ & & & & & |\lambda_{r+s}|^{-1/2} & & & \\ & & & & & & 1 & & \\ & & & & & & & \ddots & \\ & & & & & & & & 1 \end{bmatrix}$$

Then $R = R^T$ and

$$R^T D R = \begin{bmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & & -1 & & & & & \\ & & & & \ddots & & & & \\ & & & & & -1 & & & \\ & & & & & & 0 & & \\ & & & & & & & \ddots & \\ & & & & & & & & 0 \end{bmatrix}$$

This is the desired normal form.

- *Sylvester's Law of Inertia* states that the numbers r and s of $+1$'s and -1's on the diagonal are uniquely determined by the original matrix A.

B. An application to mechanics

Consider a system of n particles

- coordinates $x_1(t), \ldots, x_n(t)$, velocities $\dot{x}_1(t), \ldots, \dot{x}_n(t)$.
- kinetic energy $T = \dot{\vec{x}}^T M \dot{\vec{x}}$, potential energy $V = \vec{x}^T Q \vec{x}$, with M and Q symmetric matrices and M positive definite (all eigenvalues strictly positive).
- Equations of motion: $\boxed{M\ddot{\vec{x}} = -Q\vec{x}}$

  n *coupled* equations in n unknowns.

Idea: Introduce new coordinates $\vec{y}$ with $\vec{x} = P\vec{y}$ (and hence $\dot{\vec{x}} = P\dot{\vec{y}}$) so that in the new coordinates the kinetic and potential energies are given by

$$T = \dot{\vec{y}}^T P^T M P \dot{\vec{y}} = \dot{\vec{y}}^T M' \dot{\vec{y}}, \qquad V = \vec{y}^T P^T Q P \vec{y} = \vec{y}^T Q' \vec{x},$$

with diagonal matrices $M' = P^T M P$ and $Q' = P^T Q P$.

How?

- First find P_1 with $\boxed{P_1^T M P_1 = I}$ (possible because M is positive definite). Q is now replaced by $\boxed{Q_1 = P_1^T Q P_1}$.

- Now find an orthogonal P_2 with $Q' = P_2^T Q_1 P_1 = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$ diagonal, and take $P = P_2 P_1$, $\vec{x} = P\vec{y}$. Then

$$T = \dot{\vec{y}}^T P_2^T I P_2 \dot{\vec{y}} = \dot{\vec{y}}^T \dot{\vec{y}} = \sum_i \dot{y}_i^2, \qquad V = \vec{y}^T Q' \vec{y} = \sum_i \lambda_i y_i^2$$

- Now the equations of motion decouple: for $i = 1, \ldots, n$, if $\lambda_i < 0$,

$$\ddot{y}_i = -\lambda_i y_i \qquad \Longrightarrow \qquad y_i(t) = A_i \cos \omega_i t + B_i \sin \omega_i t$$

for $\omega_i = \sqrt{-\lambda_i}$. Thus

$$\vec{x}_j(t) = \sum_{i=1}^{n} p_{ij}(A_i \cos \omega_i t + B_i \sin \omega_i t)$$

The motion of $\vec{x}$ is a superposition of “normal modes” of vibration.

C. Skew symmetric bilinear forms

Theorem: *Every real* $n \times n$ *skew symmetric matrix* A *is congruent to a unique matrix with a block decomposition of the form*

$$\begin{bmatrix} 0 & I & 0 \\ -I & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Here I is the identy matrix of some dimension r with $2r \leq n$.