Vector Spaces

A vector space is a non-empty set V together with the binary operations of addition and scalar multiplication $((u,v) \mapsto u+v : V \times V \to V$, and $(\lambda, v) \mapsto \lambda v : \mathbb{F} \times V \to V)$ satisfying

- $u + v = v + u$ for all $u, v \in V$ (commutativity);
- $u + (v + w) = (u + v) + w$ for all $u, v, w \in V$ (associativity);
- $v + 0_V = v = 0_V + v$ for some $0_V \in V$, for all $v \in V$ (additive identity);
- for all $v \in V$ there exists $w \in V$ such that $v + w = 0_V = w + v$ (additive inverse);
- $\lambda(v + u) = \lambda v + \lambda u$ for all $v, u \in V$, $\lambda \in \mathbb{F}$ (distributivity over vector addition);
- $(\lambda + \mu)v = \lambda v + \mu v$ for all $v \in V$, $\lambda, \mu \in \mathbb{F}$ (distributivity over scalar addition);
- $(\lambda\mu)v = \lambda(\mu v)$ for all $v \in V$, $\lambda, \mu \in \mathbb{F}$;
- $1v = v$ for all $v \in V$ (multiplicative identity).

While the above axioms are normally the means of introduction to the field, it's worth pointing out that this specific system is more of a consequence of mathematicians having already done work on linear problems before gradually noticing that their linearity was a common theme, rather than anyone coming up with axioms out of thin air. The field in general can be roughly described as trying to clamp firmly down on the 'well-behaved' problems in maths, and trying to cast the net of 'well-behaved'-ness as wide as possible.

Reduced Row Echelon Form

Reduced row echelon (RRE) form is a class of matrix satisfying the conditions:
- Any row containing a non-zero element is above all rows which do not.
- The first non-zero element of a row is 1.
- The leading 1 of every non-zero row is to the right of the leading 1s of the rows above.
- In a column that contains a leading 1, all other entries are zero.

The usefulness of this notion stems from the linear independence of non-zero rows in RRE form. The solutions to a linear system are invariant under invertible maps, so any solution of a linear system in RRE form is a solution to the original system.

Lemma 1 (Row space invariance under row-reduction) Let A be an $m \times n$ matrix and B a $k \times m$ matrix, with R the RRE form matrix of A.
1. The non-zero rows of R are independent.
2. The rows of R are linear combinations of the rows of A.
3. $\mathrm{Row}(BA)$ is contained in $\mathrm{Row}(A)$.
4. If $k = m$ and B is invertible then $\mathrm{Row}(BA) = \mathrm{Row}(A)$.
5. $\mathrm{Row}(R) = \mathrm{Row}(A)$.

Existence

Theorem 2 (Existence of RRE form) *Every $m \times n$ matrix A can be reduced to a matrix in RRE form by an invertible matrix.*

The proof of this claim follows by strong induction on the number of rows in an arbitrary matrix. For an arbitrary matrix, take the first column with a non-zero element, and swap its row with the first. Then divide by its value and subtract from the rows below.

By strong induction, row operations may then be performed to place the matrix formed of the lower right quadrant into RRE form. Subtraction of these rows in RRE form from the first row where leading ones are present, and thus the matrix is in RRE form.

Additionally, note that each row operation is represented by an invertible matrix.

Theorem 3 *The system $Ax = b$ is consistent (has a solution) if and only if $\mathrm{rank}(A|b) = \mathrm{rank}(A)$.*

Note that the rank can be observed from the number of non-zero rows of the RRE form of a matrix (as it's just the number of LI columns), so the above essentially just says that b is linearly dependent with the columns of A.

It can be easily shown that the number of leading ones in a matrix's RRE form is equivalent both to the number of linearly independent rows, and linearly independent columns. This is probably more straightforward than the inductive proof of column/row-rank equality given in the lecture notes?

Dimension

Theorem 4 (Steinitz Exchange Lemma) *Let V be a vector space over a field $\mathbb{F}$. Take $X = \{v_1, \ldots, v_n\} \subseteq V$. Suppose that $u \in \langle X \rangle$ but that $u \notin \langle X \setminus \{v_i\} \rangle$ for some $v_i \in X$. Let*
$$Y = (X \setminus \{v_i\}) \cup \{u\}$$
then $\langle Y \rangle = \langle X \rangle$.

This theorem is useful for generating bases, and proving statements about dimensionality. It mainly captures the intuition that each vector space has a well-defined dimension, and so swapping out elements of a basis while preserving linear independence will also preserve the span.

Lemma 5 *Let V and W be vector spaces, with V finite-dimensional. If $T : V \to W$ is linear and $U \leq V$, then*
$$\dim U - \mathrm{null}\, T \leq \dim T(U) \leq \dim U.$$
To prove this, take $S = T(U)$ as the restriction of T to U. By rank-nullity this lemma follows.

A note on Matrix-Vector Multiplication

The definition of matrix-vector multiplication for a matrix $A \in M_{m \times n}(\mathbb{F})$, vector $x \in M_n(\mathbb{F})$ specifies that for any $i \in 1, \ldots, m$:
$$(Ax)_i = \sum_{k=1}^{n} A_{ik} x_k$$

The most obvious way of conceptualising this is as just being the dot product between A's ith row and x, however for many applications it is more useful to see that

$$\begin{bmatrix} | & | & & | \\ \boldsymbol{a_1} & \boldsymbol{a_2} & \cdots & \boldsymbol{a_n} \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \sum_{k=1}^{n} x_k \boldsymbol{a_k}.$$

The main thing this exposes is that matrix vector multiplication is the same as taking a linear combination of vectors.

Linear Transformations

Where V and W are vector spaces, a transformation $T : V \to W$ is linear if $T(v + \lambda u) = T(v) + \lambda T(u)$ for all $v, u \in V$, $\lambda \in \mathbb{F}$.

Matrix Representation

There is a one-to-one correspondence between an n by m matrix and a linear transformation from a vector space of dimension m to one of dimension n with respect to a specific ordered bases for each vector space. This follows easily from just applying any arbitrary linear transformation to the elements of a basis, noting that the corresponding outputs are sufficient to determine all other outputs of the transformation via linearity.

Theorem 6 (Change of basis) *Let V, W be finite-dimensional vector spaces over $\mathbb{F}$ with ordered bases $\mathcal{V}$, $\widetilde{\mathcal{V}}$, and $\mathcal{W}$, $\widetilde{\mathcal{W}}$ respectively. Let $T : V \to W$ be a linear map. Then*
$$_{\widetilde{\mathcal{W}}}T_{\widetilde{\mathcal{V}}} = (_{\widetilde{\mathcal{W}}}I_{\mathcal{W}})(_{\mathcal{W}}T_{\mathcal{V}})(_{\mathcal{V}}I_{\widetilde{\mathcal{V}}})$$
An immediate consequence of this is that any two matrices representing the same linear transformation from a vector space back to itself within different bases are similar.

Bilinear Forms

It's often useful to have some notion of vector 'distance'. In Euclidean geometry we tend to use the dot product for this purpose, but a more general notion of this can be found from the inner product, which is itself a special case of a bilinear form.

We define a bilinear form as a function $B : V \times V \to \mathbb{F}$ such that $B(v + \alpha u, w) = B(v, w) + \alpha B(u, w)$ and $B(v, w + \alpha u) = B(v, w) + \alpha B(v, u)$ for any $u, v, w \in V$, $\alpha \in \mathbb{F}$. Otherwise put, the function is linear in both of its arguments. Helpfully this gives a nice general form when taking V with respect to an ordered basis $\{e_1, \ldots, e_n\}$. For any $v, u \in V$:
$$v = \sum_{i=1}^{n} x_i e_i$$
$$u = \sum_{i=1}^{n} y_i e_i$$
for some $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{F}^n$, and consequently by linearity

$$\begin{aligned} B(v, u) &= B\left(\sum_{i=1}^{n} x_i e_i, \sum_{j=1}^{n} y_j e_j \right) \\ &= \sum_{i=1}^{n} x_i B\left(e_i, \sum_{j=1}^{n} y_j e_j \right) \\ &= \sum_{i=1}^{n} x_i \sum_{j=1}^{n} y_j B(e_i, e_j) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} x_i a_{ij} y_j \qquad \text{writing } a_{ij} = B(e_i, e_j) \\ &= \boldsymbol{x}^\top A \boldsymbol{y} \qquad\qquad \text{with } A = (a_{ij}) \end{aligned}$$

A in this case is called the gram matrix. Every unique bilinear form with respect to a certain ordered basis has a unique gram matrix, and to every gram matrix corresponds a bilinear form.

Inner Products

As already mentioned, an inner product is a special case of a bilinear form. Specifically, it is a bilinear form that is positive definite, meaning that for non-zero $v \in V$, $B(v, v) > 0$ (at least, this is its definition for real inner product spaces), and symmetric, so for any $v, u \in V$, $B(v, u) = B(u, v)$. We normally write inner products using the notation $\langle v, u \rangle := B(v, u)$.

Like many properties that can apply to bilinear forms, positive definiteness is basis invariant, and thus also a property of any corresponding gram matrix. Immediate consequences of positive definiteness include that the matrix has full rank, and that all the eigenvectors of the matrix are positive (the latter being equivalent).

There isn't in all honesty a great deal that one can do with inner product spaces at this point in the course, apart from noting some of their properties. The lecture notes introduce orthogonality in the context of bilinear forms, but this again is something that isn't obviously significant until later on in Linear Algebra II.