Linear Algebra Notes

Andi Gu

This is an abridged note on Friedberg's *Linear Algebra*, selected sections from chapters 5 and 7.

Subspace Invariance

We make a few definitions. The characteristic polynomial of a linear operator T is:

$$f(t) = \det([T]_\beta - t\mathbf{I})$$

It turns out that the basis β we chose is completely arbitrary and it does not matter which basis we express T with respect to. To see this, we find the characteristic polynomial of T with respect to β', another basis.

$$f'(t) = \det\bigl([T]'_\beta - t\mathbf{I}\bigr)$$

We are careful to differentiate between the identity matrix $\mathbf{I}$ and the identity linear transformation I.

$$\begin{aligned} &= \det\Bigl([I]_\beta^{\beta'}[T]_\beta[I]_{\beta'}^\beta - t\mathbf{I}\Bigr) \\ &= \det\Bigl([I]_\beta^{\beta'}([T]_\beta - t\mathbf{I})[I]_{\beta'}^\beta\Bigr) \end{aligned}$$

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

$$\begin{aligned} &= \det\Bigl([I]_\beta^{\beta'}[I]_{\beta'}^\beta\Bigr)\det([T]_\beta - t\mathbf{I}) \\ &= \det(\mathbf{I})\det([T]_\beta - t\mathbf{I}) \\ &= \det([T]_\beta - t\mathbf{I}) \end{aligned}$$

Thus we can speak of *the* characteristic polynomial of T without worrying about choosing a basis. The eigenvalues of a linear transformation are found by solving for $f(\lambda) = 0$.

Also, the restriction of a linear operator T to a subspace W is simply the linear transformation T_W with a domain and range restricted to W that outputs $T(w)$ for all $w \in W$.

Eigenspaces

The characteristic polynomial of any $n \times n$ square matrix will always be degree n, and will thus always have n eigenvalues (not necessarily unique). Say a particular eigenvalue λ_i has multiplicity k_i. Let the eigenspace associated with λ_i be E_i. Clearly, the matrix is diagonalizable if and only if $\dim(E_i) = k_i$.

Invariant Subspaces

Let T be a linear operator on a vector space V. A subspace of W of V is said to be a T-invariant subspace of V if $(\forall w \in W)(T(w) \in W)$. We define a T-cyclic subspace of v as $\{v, T(v), T^2(v), \ldots\}$ – clearly, this is T-invariant and is notably the *smallest* T-invariant subspace that contains v. We claim that $\{v, T(v), \ldots, T^{k-1}(v)\}$ is a basis for this T-cyclic subspace, where k is its dimension.[1]

Let W be the T-cyclic subspace generated by v. Then the restriction of T to W, T_W has a characteristic polynomial $(-1)^k \sum_{i=0}^{k} a_i t^i$, where the coefficients a_i are defined as follows: $T^k(v) + \sum_{i=0}^{k-1} a_i T^i(v) = 0$, and $a_k = 1$. This definition of a_i is guaranteed to be valid because $T^k(v) \in W$, and can therefore be expressed as a linear combination of $\beta = \{T^i(v)\}$. This is true because then we have:

$$[T]_\beta = \begin{bmatrix} 0 & 0 & \ldots & 0 & -a_0 \\ 1 & 0 & \ldots & 0 & -a_1 \\ 0 & 1 & \ldots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & -a_{k-1} \end{bmatrix}$$

By working along the last column, we can verify that:

$$f(t) = (-1)^k \sum_{i=0}^{k} a_i t^i \tag{1}$$

Characteristic Polynomial of T_W

Let W be the T-cyclic subspace generated by v, and define a basis β for W, extending it to a basis γ for V. Then:

$$[T]_\gamma = \begin{bmatrix} [T_W]_\beta & X \\ 0 & Y \end{bmatrix} \quad \text{where } X \text{ and } Y \text{ are arbitrary}$$

Using an identity for the determinant of a 'block' form of a matrix, we can see $\det([T]_\gamma - tI) = \det([T]_\beta - tI) \cdot \det(Y - tI)$, so that

$$f(t) = g(t) \cdot \det(Y - tI) \quad \text{where } g \text{ is the characteristic polynomial of } T_W \tag{2}$$

Cayley-Hamilton

We prove the remarkable result that $f(T) = T_0$ (T_0 being the zero transformation), where $f(t)$ is the characteristic equation of T. That is, evaluating the characteristic equation of T on itself gives the zero transformation. We do this by showing $(\forall v \in V)(f(T)(v) = 0)$. We look at the T-cyclic subspace generated by v called W, and define the characteristic equation $g(t)$ of T_W. By equation (1):

$$g(T)(v) = (-1)^k \left(\sum_{i=0}^{k} a_i T^i(v) \right)$$

[1] This can be proved by finding the subspace Z spanned by $\beta = \{v, T(v), \ldots T^j(v)\}$ where j is the largest integer that ensures β is linearly independent, so that Z is a subset of the T-cyclic subspace.

We recall that a_i were *defined* by $\sum_{i=0}^{k-1} a_i T^i(v) = T^k(v)$, and $a_k = 1$.

$$= 0$$

We also note that $g(t)$ is a factor of $f(t)$ by equation (2). Then:

$$\begin{aligned} f(T)(v) &= g(T)(v) \cdot h(T)(v) \quad \text{with } h \text{ arbitrary} \\ &= 0 \end{aligned}$$

Jordan Canonical Form

The problem is that not all matrices are diagonalizable – their eigenspaces do not form complete bases for the entire space V. We extend the notion of an eigenspace, which was before $\{v \mid (T - \lambda I)(v) = 0\}$, to a 'generalized' eigenspace corresponding to an eigenvalue λ such that $\{v \mid (T - \lambda I)^p(v) = 0\}$ for any finite positive integer p. We use K_λ to denote this eigenspace.[2] At this point we note something important which is used quite a bit in the proceeding proofs. Composition of linear transformations involving only T and I will generally commute, since $TT = TT$ and $TI = IT$. For example, $(T - \lambda I)^p T = T(T - \lambda I)^p$ – this can be seen by expanding both sides.

First, we show K_λ is a T-invariant subspace of V and contains E_λ (this second assertion is trivially true, the subset of K_λ with $p = 1$ is E_λ). Then, we show that $(\forall v \in K_\lambda) T(x) \in K_\lambda$, by simply showing $(T - \lambda I)^p(T(x)) = 0$ for some p:

$$(T - \lambda I)^p(T(x)) = T((T - \lambda I)^p(x))$$

Clearly there is exists a p such that $(T - \lambda I)^p(x) = 0$ since $x \in W$.

$$\begin{aligned} &= T(0) \\ &= 0 \end{aligned}$$

We also show that for $\mu \neq \lambda$, the restriction of $T - \mu I$ to K_λ is one to one. We do this by showing $(T - \mu I)(x) = 0$ for $x \in K_\lambda$ must imply $x = 0$.[3] So we start by assuming $(T - \mu I)(x) = 0$ for some $x \in K_\lambda$ and $x \neq 0$, and proceed by contradiction. Then let p be the smallest integer that satisfies $0 = (T - \lambda I)^p(x)$, so that $y = (T - \lambda I)^{p-1}(x) \neq 0$. Then:

$$\begin{aligned} (T - \mu I)(y) &= (T - \mu I)(T - \lambda I)^p(x) \\ &= (T - \lambda I)^p(T - \mu I)(x) \\ &= 0 \end{aligned}$$

Thus $y \in E_\mu$ and $y \in E_\lambda$, but we know $E_\lambda \cap E_\mu = \{0\}$, so $y = 0$. But we defined y to be non-zero (by our setting of p).

There are two smaller things to prove before we can proceed to the essential theorem. First, $\dim(K_\lambda) \leq m$ and $K_\lambda = \ker((T - \lambda I)^m)$ where m is the multiplicity of λ. The first claim is easy to

[2] Perhaps a good intuition for this is to interpret $T - \lambda I$ as returning the 'residual' of some vector with respect to the true eigenvector associated with λ. Then the generalized eigenspace is all vectors that are 'close enough' to the true eigenspace of λ, since taking their residual many times results in zero.

[3] Again going with our intuition of K_λ, this means that if one vector is close to an eigenvector with eigenvalue λ, it is close to that eigenvector *only*. In other words, K_λ and K_μ are mutually exclusive.

show – the characteristic polynomial $g(t)$ of T_{K_λ} is of degree $\dim(K_\lambda)$ and has λ as its only root, as shown above. Also, it is a factor of the characteristic polynomial of T, which has the factor $(t-\lambda)$ exactly m times – thus $\dim(K_\lambda) \leq m$, the result we sought. The second is not much harder: clearly $\ker((T-\lambda I)^m) \subseteq K_\lambda$ by definition. Going the other way, $g(T_{K_\lambda}) = 0$ – since g only has λ as roots, $T_{K_\lambda} - \lambda I = 0$, and $(T-\lambda I)^{\dim K_\lambda}(x) = 0$ for $x \in K_\lambda$, and since $\dim K_\lambda \leq m$, this implies that $(T-\lambda I)^m(x) = 0$ for $x \in K_\lambda$, therefore $K_\lambda \subseteq \ker((T-\lambda I)^m)$.[4]

Forming a Basis with General Eigenspaces

This is an essential proof. We wish to show that $x \in V$ can be expressed as a sum $\sum_{i=1}^k v_i$, where $v_i \in K_{\lambda_i}$ (there are k distinct eigenvalues of T). Essentially, that $\bigcup_i K_{\lambda_i} = V$.

We proceed by induction over k, the number of distinct eigenvalues of T. In the base case of $k=1$, if the multiplicity of λ_1 is m, then $f(T) = (T-\lambda_1 I)^m = 0$ by Cayley-Hamilton, so $\ker((T-\lambda I)^m) = V = K_{\lambda_1}$.

Next, the induction step is as follows. Assume the result holds for all $k' < k$. Then the characteristic polynomial of T is $f(t) = (t-\lambda_k)^m g(t)$. Let $W = \operatorname{im}((T-\lambda_k I)^m)$. Then we note that $K_{\lambda_i} \subseteq W$, since clearly $(T-\lambda_k I)^m$ maps K_{λ_i} *into* itself (i.e. $\forall x \in K_{\lambda_i} : (T-\lambda_k I)^m x \in K_{\lambda_i}$), and the restriction of $T - \lambda_k I$ to K_{λ_i} is one-to-one (as proved above), then $(T-\lambda_k I)^m$ must map K_{λ_i} *onto* itself (so that $(T-\lambda_k I)^m x \in K_{\lambda_i} \iff x \in K_{\lambda_i}$). Therefore $E_{\lambda_i} \subseteq K_{\lambda_i} \subseteq W$ for $i < k$, so $\lambda_i | i < k$ is among the eigenvalues of T_W.

Now, we show that λ_k is *not* an eigenvalue of T_W. We proceed by contradiction – assume there is a $v \in W$ such that $Tv = \lambda_k v \implies (T-\lambda_k I)v = 0 \implies (T-\lambda_k I)^m v = 0 \implies v \notin W$. Thus, we have a contradiction, and in short, the eigenvalues of T_W are $\lambda_1, \ldots, \lambda_{k-1}$. Since T_W has $k-1$ distinct eigenvalues, we can apply the induction hypothesis and so $\forall x \in V : \exists v_i \in K_{\lambda_i} : (T-\lambda_k I)^m x = \sum_{i=1}^{k-1} v_i$. Since $(T-\lambda_k I)^m$ maps K_{λ_i} onto itself, $\exists w_i \in K_{\lambda_i} : (T-\lambda_k I)^m w_i = v_i$, so:

$$(T-\lambda_k I)^m x = \sum_{i=1}^{k-1} (T-\lambda_k I)^m w_i$$

$$0 = (T-\lambda_k I)^m \left(x - \sum_{i=1}^{k-1} w_i \right)$$

Thus $x - \sum_{i=1}^{k-1} w_i \in K_{\lambda_k}$, so if we let $w_k = x - \sum_{i=1}^{k-1} w_i$ where $w_k \in K_{\lambda_k}$, we get our desired result:

$$x = \sum_{i=1}^{k-1} w_i + w_k = \sum_{i=1}^{k} w_i$$

[4] Of course, we can make a weaker statement that $K_\lambda = \ker((T-\lambda I)^p)$ for any $p \geq m$, but $(T-\lambda I)^p$ does not grow after $p > m$.