Self Adjoint Linear Transformations

Francis J. Narcowich

September 2013

1 Definition of the Adjoint

Let V be a complex vector space with an inner product $<\,,\,\rangle$ and norm $\|\,\|$, and suppose that $L : V \to V$ is linear. If there is a function $L^* : V \to V$ for which

$$\langle Lx, y\rangle = \langle x, L^* y\rangle \tag{1.1}$$

holds for every pair of vectors x, y in V, then L^* is said to be the *adjoint* of L. Some of the properties of L^* are listed below.

Proposition 1.1. *Let $L : V \to V$ be linear. Then L^* exists, is unique, and is linear.*

Proof. Introduce an orthonomal basis B for V. Then find the matrix of L, A_L relative to this basis. Also, relative to B, it is easy to show that the inner product becomes

$$\langle x, y\rangle = [y]_B^*[x]_B$$

From this it follows that

$$\langle Lx, y\rangle = [y]_B^* A_L [x]_B = (A_L^*[y]_B)^*[x]_B.$$

Of course, A_L^* exists, is unique, and defines a linear transformation on $\mathbb{C}^n$. If $\Phi : V \to \mathbb{C}^n$, then $\widetilde{L} := \Phi^{-1} A_L^* \Phi$ is a linear transformation on V. From this and the previous equation, $\widetilde{L}$ satisfies

$$(A_L^*[y]_B)^*[x]_B = \langle x, \widetilde{L}y\rangle = \langle Lx, y\rangle,$$

and it follows that $L^* = \widetilde{L}$. ∎

We say that L is *self adjoint.* if $L^* = L$. Self adjoint transformations are extremely important; we will discuss some of their properties later. Before we do that, however, we should look at a few examples of adjoints for linear transformations.

Example 1.2. Consider the usual inner product on $V = \mathbb{C}^n$; this is given by $\langle x, y\rangle_{\mathbb{C}^N} = y^* x$. As noted above, for an $n \times n$ matrix A, $\langle Ax, y\rangle_{\mathbb{C}^N} = y^* Ax = (A^* y)^*$. Thus A^* is the conjugate transpose of A, a fact we tacitly used above.

Example 1.3. Let $V = P_n$, where we allow the coefficients of the polynomials to be complex valued. For an inner product, take

$$\langle p, q\rangle = \int_{-1}^{1} p(x)\overline{q(x)}dx\,, \tag{5}$$

and for L take

$$L(p) = [(1 - x^2)p']'\,. \tag{6}$$

Doing an integration by parts yields $\langle p, Lq\rangle = \langle Lp, q\rangle$. Thus, $L = L^*$ and L is selfadjoint.

Example 1.4. Let V be the set of all complex valued polynomials that are of degree n or less and that are zero at ± 1. Take $L(p) = xp'$, and use (5) as the inner product. Again, an integration by parts shows that $\langle p, Lq\rangle = \langle -p - xp', q\rangle$, so $L^*(p) = -p - xp'$.

What we have said so far is general and applies to any complex vector space with an inner product, no matter what the dimension of the space is. We now want to see, in detail, what happens in a complex, finite dimensional vector space V with an orthonormal basis $B = \{u_1, \ldots, u_n\}$. (Every finite dimensional vector space with an inner product always has an orthonormal basis. To create one from any given basis, just apply the Gram-Schmidt process to the given basis.) Our next result concerns the form of the matrix of the adjoint of a linear transformation on V relative to B.

It is worthwhile to formally state a result that we actually got in the course of establishing the results above.

Proposition 1.5. *Let V and B be as described above. If $L : V \to V$ is a linear transformation whose matrix relative to B is A_L, then the matrix of L^* is $A_{L^*} = A_L^*$.*

2 Selfadjoint Linear Transformations

Having done a few examples, let us return to our discussion of selfadjoint transformations. We begin with the general case where the vector space V is not assumed to be finite dimensional. We have the following important result.

Proposition 2.1. *Let V be a complex vector space with an inner product. If $L : V \to V$ is a selfadjoint linear transformation, then the eigenvalues of L are real numbers, and eigenvectors of L corresponding to distinct eigenvalues are orthogonal.*

Proof. Suppose that λ is an eigenvalue of L and that x is a corresponding eigenvector. We therefore have $Lx = \lambda x$, and so $\langle Lx, x\rangle = \langle \lambda x, x\rangle = \lambda\langle x, x\rangle$. Similarly, we see that $\langle x, Lx\rangle = \langle x, \lambda x\rangle = \bar{\lambda}\langle x, x\rangle$. Now, because $L = L^*$, we have that $\langle Lx, x\rangle = \langle x, Lx\rangle$, which together with the previous two equations gives us $\lambda\langle x, x\rangle = \bar{\lambda}\langle x, x\rangle$. Finally, since $x \neq 0$, we may divide by $\langle x, x\rangle$; the result is $\lambda = \bar{\lambda}$. This shows that λ is a real number. Now suppose that λ_1 and λ_2 are distinct eigenvalues with eigenvectors x_1 and x_2. Observe that, because L is selfadjoint and the eigenvalues are real, we have $\lambda_1\langle x_1, x_2\rangle = \langle Lx_1, x_2\rangle = \langle x_1, Lx_2\rangle = \lambda_2\langle x_1, x_2\rangle$. Thus, $(\lambda_1 - \lambda_2)\langle x_1, x_2\rangle = 0$ Since $\lambda_1 \neq \lambda_2$, dividing by $\lambda_1 - \lambda_2$ yields $\langle x_1, x_2\rangle = 0$. ∎

Before we go on to our next result, we need to set down a few facts about eigenvalues and eigenvectors, and about bases in general. This we now do.

Lemma 2.2. *Let V be a complex, finite dimensional vector space, with dimension $n \geq 1$. If $L : V \to V$ is linear, then L has at least one eigenvalue.*

Proof. Let B be a basis for V and let A_L be the matrix of L relative to B-coördinates. Because $\det(A_L - \lambda I) = 0$ is a polynomial in λ, it has n roots, if one counts repetitions. In any case, it has between 1 and n distinct roots, all of which are eigenvalues. Thus, L has at least one eigenvalue. ∎

Let us again return to our discussion of selfadjoint linear transformations. This time we will look at the case in which the underlying complex vector space is finite dimensional. In this case, selfadjoint transformations are always diagonalizable. Indeed, we can say even more, as the following result shows.

Theorem 2.3. *Let V be a complex, finite dimensional vector space. If $L : V \to V$ is a selfadjoint linear transformation, then there is an orthonormal basis for V that is composed of eigenvectors of L. The matrix of L relative to this basis is diagonal.*

We will give two proofs for this important theorem. The first is similar to the one given in Keener's book and involves invariant subpaces. The second is one that is more concrete in that it dierectly uses matrix computations. Here is the first proof.

Proof. (Proof 1). A subspace U of V is said to be invariant under $L|V \to V$ if and only if $L : U \to U$. Let $S := \text{span}\{\text{eigenvectors of } L\}$ and let $U = S^{\perp}$. We claim that $S^{\perp}$ is invariant under L. To see this, let $u \in U$ and and $v_j \in S$ be an eigenvector of L. Since $\langle Lu, v_j \rangle = \langle u, Lv_j \rangle = \langle u, \lambda_j v_j \rangle$ and $u \in U = S^{\perp}$. we have $\langle Lu, v_j \rangle = \lambda_j \langle u, v_j \rangle = 0$. Thus, $Lu \in S^{\perp}$, so $S^{\perp}$ is invariant under L. If $S^{\perp} \neq \{0\}$, so that its dimension is one or more, then, by Lemma 2.2, L has an eigenvalue λ and with $v \neq 0$ being its corresponding eigenvector. But then v, being an eigenvector, is also in S. Thus, $v \in S \cap S^{\perp}$. It follows that $\langle v, v \rangle = 0$, and so $v = 0$. This is a contradiction, so $S^{\perp} = \{0\}$ and $V = S$. Using Gram-Schmidt if necessay, we may form an o.n. basis from the eigenvectors of L. ∎

Proof. (Proof 2). We will work with the matrix of L relative to an orthonormal basis. We denote this matrix by A. Of course this will be an $n \times n$ self adjoint matrix. By Lemma 2.2, A has an eigenvalue λ_1 with corresponding eigenvector x_1, normalized so that $\|x_1\| = 1$. Use Gram-Schmidt, if necessary, to form a orthonormal basis $B_1 = \{x_1, y_2, y_3, \ldots, y_n\}$. If we change coordinates to B_1, it is easy to show that A, in the new coordinates, becomes

$$A_1 = \begin{pmatrix} \lambda_1 & 0_{n-1}^T \\ 0_{n-1} & \widetilde{A}_1 \end{pmatrix}$$

where $\widetilde{A}_1$ is a selfadjoint $(n-1) \times (n-1)$ matrix. Repeating the argument for $\widetilde{A}_1$, we see that the matrix for $\widetilde{A}_1$ becomes A_2, where

$$A_2 = \begin{pmatrix} \lambda_1 & 0 & 0_{n-2}^T \\ 0 & \lambda_2 & 0_{n-2}^T \\ 0_{n-2} & 0_{n-2} & \widetilde{A}_2 \end{pmatrix}$$

Again $\widetilde{A}_2$ is self adjoint and put in a form similar to those above. Continuing in this way, we can find an orthonormal system of coordinates in $\mathbb{C}^n$ relative to which the matrix A is diagonal. The corresponding basis for V is also orthonormal and is composed eigenvectors of L. ∎

Of course, for a self-adjoint matrix A, Theorem 2.3 implies that there is a matrix $S = [x_1 \ \cdots x_n]$, whose columns are an o.n. set of eigenvectors of A, such that $A = S\Lambda S^{-1}$, where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. – note that he eigenvalues are listed in the same order as the eigenvectors. Since the columns of S are an o.n. set, it is easy to show that $S^{-1} = S^*$. In this form. we have

$$A = S\Lambda S^*, \ S^*S = I. \tag{2.1}$$

3 Applications

There are many important applications of what was discussed in the previous sections. The theory of adjoints and of self-adjoint linear transformations comes up in the study of partial differential equations and the eigenvalue problems that result when the method of separation of variables is used to solve them. (Partial differential equations arise in connection with heat conduction, wave propagation, fluid flow, electromagnetic fields, quantum mechanics, and many other areas as well.) Selfadjoint linear transformations play a fundamental role in formulating quantum mechanics; they represent the physical quantities that can be observed in a laboratory—the observables of physics. In this section we will provide a few examples.

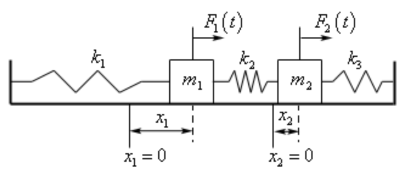

Figure 1: Coupled spring system

Example 3.1. *Normal modes for a coupled spring system.* In the coupled spring system showm in Fig. 3, let $m_1 = m_2 = m$ and $k_1 = k_2 = k_3 = k$. Using Hooke's law and Newton's law, the equation of motion for the spring system is, in matrix, form

$$\ddot{\mathbf{x}} = -\frac{k}{m}A\mathbf{x}, \quad A := \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}, \tag{3.1}$$

A normal mode for the system is a solution of the form $\mathbf{x}(t) = \text{function}(t)\mathbf{x}_0$, where $\mathbf{x}_0$ is independent of t. The usual way to treat this problem is to make the *Ansatz* $\mathbf{x}(t) = e^{i\omega t}\mathbf{x}_0$, where ω is a constant angular frequency. Plugging this solution into (3.1) and canceling the time factor, we obtain

$$\tfrac{m\omega^2}{k}\mathbf{x}_0 = A\mathbf{x}_0.$$

It follows that $m\omega^2/k = \lambda$ is an eigenvalue of A, with $\mathbf{x}_0$ being the corresponding eigenvector. In this case, we have two eigenvalues $\lambda = 1$, $\mathbf{x}_1 = (1\ 1)^T$ and $\lambda = 3$, $\mathbf{x}_3 = (1\ -1)^T$. The *eigenfrequencies* are thus $\omega_1 = \sqrt{k/m}$

and $\omega_3 = \sqrt{3k/m}$ and the corresponding normal modes are

$$e^{\pm i\sqrt{k/m}\,t}\begin{pmatrix}1\\1\end{pmatrix} \text{ and } e^{\pm i\sqrt{3k/m}\,t}\begin{pmatrix}1\\-1\end{pmatrix}.$$

Example 3.2. *Inertia tensor.* The kinetic energy of a rigid body freely rotating about its center of mass is

$$T = \frac{1}{2}\omega^T \mathcal{I}\omega.$$

The vector ω is the angular velocity of the body and $\mathcal{I}$ is the 3×3 inertia tensor. If $\rho(\mathbf{x})$ is the mass density of the body, which occupies the region $\Omega \subset \mathbb{R}^3$, then

$$\mathcal{I} = \int_\Omega \left(|\mathbf{x}|^2 I_{3\times 3} - \mathbf{x}\mathbf{x}^T\right)\rho(\mathbf{x})d^3\mathbf{x}.$$

The eigenvalues and eigenvectors of $\mathcal{I}$ play an important role in the equation of motion for a rigid body. (For details, see: H. Goldstein, *Classical Mechanics*, Addison-Wesley, 1965.)

Example 3.3. *Principal axis theorem.* Consider the conic $3x^2 - 2xy + 3y^2 = 1$. We want to rotate axes to find new coordinates x', y' relative to which the conic is in standard form. Let's put the equation in matrix form:

$$\begin{pmatrix}x & y\end{pmatrix}\underbrace{\begin{pmatrix}3 & -1\\-1 & 3\end{pmatrix}}_{A}\begin{pmatrix}x\\y\end{pmatrix} = 1.$$

It is straightforward to show that the eigenvalues of A are 2 and 4, with corresponding orthonormal eigenvectors $\frac{1}{\sqrt{2}}(1\ 1)^T$ and $\frac{1}{\sqrt{2}}(-1\ 1)^T$. In the factored form in (2.1), we have

$$S = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & -1\\1 & 1\end{pmatrix} \text{ and } \Lambda = \begin{pmatrix}2 & 0\\0 & 4\end{pmatrix}.$$

Let $\mathbf{x} = (x\ y)^T$. The original form of the conic was $\mathbf{x}^T A\mathbf{x} = 1$. If we set $\mathbf{x}' = S^T\mathbf{x}$, then the equation of the conic becomes $\mathbf{x}'^T\Lambda\mathbf{x}' = 1$ or, in the new coordinates, $2x'^2 + 4y'^2 = 1$. The matrix S is actually changes from x'-y' to x-y coordinates. In effect, it gets the x'-y' axes by rotating the x-y axes counter clockwise through an angle $\pi/4$.

4 The Courant-Fischer Theorem

It is simple to calculate the eigenvalues for small matrices, with $n = 2$ or 3. However, direct calculation is not possible for large systems. Thus, we need a method for approximating them. This is supplied by the Courant-Fischer Theorem, which we now state.

Theorem 4.1 (Courant-Fischer). *Let A be a real $n \times n$ self-adjoint matrix having eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Then,*

$$\lambda_k = \min_{C \in \mathbb{R}^{k-1 \times n}} \max_{\substack{\|\mathbf{x}\|=1 \\ C\mathbf{x}=\mathbf{0}}} \mathbf{x}^T A \mathbf{x}. \tag{4.1}$$

Proof. Use (2.1) to get $\mathbf{x}^T A \mathbf{x} = \mathbf{y}^T \Lambda \mathbf{y}$, where $\mathbf{y} = S^T \mathbf{x}$. Because S is orthogonal, we have $\|\mathbf{x}\| = \|S\mathbf{y}\| = \|\mathbf{y}\|$. In addition, CS runs over all matrices in $\mathbb{R}^{k-1 \times n}$ if $C \in \mathbb{R}^{k-1 \times n}$ does. Thus, we are now trying to show that

$$\lambda_k = \min_{C \in \mathbb{R}^{k-1 \times n}} \max_{\substack{\|\mathbf{y}\|=1 \\ C\mathbf{y}=\mathbf{0}}} \mathbf{y}^T \Lambda \mathbf{y}. \tag{4.2}$$

Let $q(\mathbf{y}) = \mathbf{y}^T \Lambda \mathbf{y}$. Of course, q can be written as

$$q(\mathbf{y}) = \sum_{j=1}^{n} \lambda_j y_j^2.$$

The proof proceeds in two steps. First, to satisfy $C_0 \mathbf{y} = \mathbf{0}$ when $C_0 = [e_1 \ \cdots \ e_{k-1}]^T$, we need only take $\mathbf{y} = \sum_{j=k}^{n} y_j e_j$. In that case, we have, since $\|\mathbf{y}\|^2 = \sum_{j=k}^{n} y_j^2 = 1$,

$$q(\mathbf{y}) = \sum_{j=k}^{n} \lambda_j y_j^2 \leq \lambda_k \sum_{j=k}^{n} y_j^2 = \lambda_k \cdot 1 = \lambda_k,$$

and so, for C_0, we have $\max_{\substack{\|\mathbf{y}\|=1 \\ C_0\mathbf{y}=\mathbf{0}}} q(\mathbf{y}) = \lambda_k$. The second step is to show that for any C we can find a $\mathbf{y}'$ such that $q(\mathbf{y}') \geq \lambda_k$. If we can do that, then

$$\max_{\substack{\|\mathbf{y}\|=1 \\ C\mathbf{y}=\mathbf{0}}} q(\mathbf{y}) \geq q(\mathbf{y}') \geq \lambda_k = \max_{\substack{\|\mathbf{y}\|=1 \\ C_0\mathbf{y}=\mathbf{0}}} q(\mathbf{y}),$$

and (4.2) follows immediately. To show that such a $\mathbf{y}$ exists, start by augmenting C by adding rows e_j^T, $j = k+1, \ldots n$:

$$\widetilde{C} = \begin{pmatrix} C \\ e_{k+1}^T \\ \vdots \\ e_n^T \end{pmatrix} \in \mathbb{R}^{(n-1)\times n}.$$

Note that since $\text{rank}(\widetilde{C}) \leq n-1$, so $\text{nullity}(\widetilde{C}) \geq 1$. Thus there is a vector $\mathbf{y}' \neq \mathbf{0}$ such that $\widetilde{C}\mathbf{y}' = \mathbf{0}$. This is equivalent to the equations $C\mathbf{y}' = \mathbf{0}$ and $y_j' = e_j^T \mathbf{y}' = 0$, $j = k+1, \ldots, n$. Moreover, this implies that

$$q(\mathbf{y}') = \sum_{j=1}^{k} \lambda_j {y_j'}^2 \geq \lambda_k \sum_{j=1}^{k} {y_j'}^2 = \lambda_k \cdot 1 = \lambda_k.$$

This completes the proof. ∎

Example 4.2. *Estimating an eigenvalue.* Show that $\lambda_2 \leq 0$, for the matrix

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

Because A has positive entries, we expect that the eigenvector for λ_1 will have all positive entries. (This is in fact a consequence of the Perron-Frobenius Theorem.) Thus, since we want to get an estimate of the minimum of the maximum of $\mathbf{x}^T A\mathbf{x}$, we guess that $C = (1\ 1\ 1)$ and so $C\mathbf{x} = x_1 + x_2 + x_3 = 0$. The quadratic form

$$q(\mathbf{x}) := \mathbf{x}^T A\mathbf{x} = x_1x_2 + 2x_1x_3 + 3x_2x_3.$$

Let's solve $x_1+x_2+x_3 = 0$ for x_1 and put the result in the expression above for $q(\mathbf{x})$. Doing so yields

$$q(\mathbf{x}) = -(x_2 + x_3)(x_2 + 2x_3) + +3x_2x_3 = -x_2^2 - 2x_3^2 \leq 0.$$

From this, we see that for $C = (1\ 1\ 1)$ and $C\mathbf{x} = 0$, $\lambda_2 \leq \max_{C\mathbf{x}=0} q(\mathbf{x}) \leq 0$.