EE476: Linear Algebra Review

1 Vector paces, subspaces

Definition 1.1 *A* **Vector Space** *(or linear space) over a field K is a non empty set X of elements $x, y, \ldots$ (called vectors) together with two algebraic operations. The first operation is vector addition which associates with any two vectors $x,\ y \in X$ a vector $x + y \in X$, the sum of x and y. The second operation is scalar multiplication which associates to any vector $x \in X$ and any element $\alpha \in K$ a vector $\alpha x \in X$; the scalar multiple of x by α.*

The set X and these operations satisfy the following axioms:

1. $x + y = y + x$ *(commutative law)*
2. $x + (y + z) = (x + y) + z$ *(associative law)*
3. *There exists a vector $\mathbf{0} \in X$ called the zero vector such that $x + \mathbf{0} = x$ for all $x \in X$*
4. $\alpha(\beta x) = (\alpha\beta)x$ *(associative law)*
5. $\left.\begin{array}{l}\alpha(x + y) = \alpha x + \alpha y \\ (\alpha + \beta)x = \alpha x + \beta x\end{array}\right\}$ *(distributive laws)*
6. $0x = \mathbf{0}, \qquad 1x = x$

If $K = \mathbb{R}$ or $K = \mathbf{C}$ then the vector space is called respectively Real or Complex Vector Space.

Definition 1.2 *A* **Subspace** *S of a vector space X is a subset of X which is itself a vector space*

Note that a subspace must contain the 0 vector. To check for a subspace you need to verify that

$$\alpha x + \beta y \in S \qquad \text{for any linear combination of any } x,\ y \in S.$$

Example 1.1 *If $X = \mathbb{R}^3$, and $S = \{x \mid a_1 x(1) + a_2 x(2) = 0\}$, where $x(1)$ and $x(2)$ are the first and second component of the vector $x \in \mathbb{R}^3$, then S is a subspace of $\mathbb{R}^3$.*

1.1 Linear independence, basis, dimension

A set of vectors is **linearly independent** if there is no nontrivial combination of element of the set that add to the zero vector.

A **basis** for a subspace is an independent set of vectors that can be combined linearly to form any other vector in the subspace.

The **dimension** of a subspace is equal to the number of vectors in a basis.

There are many vector spaces, the most common are $\mathbb{R}^n$, $\mathbf{C}^n$. The set $C_{[a,b]}$, of all continuous functions in the interval a, b, is a vector space. The set

$$S = \{f(t) = \alpha_0 + \alpha_1 t + \alpha_{n-1} t^{n-1}, \alpha_i \in \mathbb{R}\}$$

is also a vector space. Note that this vector space is not different from R^n (isomorphic). Every sequence $\alpha_0, \alpha_1 \cdots \alpha_{n-1}$ is uniquely identified with an element of the vector space.

Exercise 1.1 *Show that $\{1,\ t,\ t^2,\ \ldots,\ t^{n-1}\}$ form a basis for S.*
Does the set $\{1,\ (1+t),\ (1+t)^2,\ \ldots,\ (1+t)^{n-1}\}$ form a basis for S?

2 Matrices

A real matrix is denoted as:

$$A \in \mathbb{R}^{m\times n}, \quad \begin{pmatrix} a_{11} & \cdots a_{1m} \\ \vdots & \vdots \\ a_{m1} & \cdots a_{mn} \end{pmatrix}, \quad a_{ij} \in \mathbb{R}^n$$

Complex matrices are defined analogously

$$A \in \mathbf{C}^{m\times n}, \quad \begin{pmatrix} a_{11} & \cdots a_{1m} \\ \vdots & \vdots \\ a_{m1} & \cdots a_{mn} \end{pmatrix}, \quad a_{ij} \in \mathbf{C}^n$$

A matrix $A \in \mathbb{R}^{m\times n}$ is also an element of the vector space of matrices $m \times n$. Matrices are linear operators on vector spaces.

$$A: \quad \mathbf{C}^n \to \mathbf{C}^m$$

$$x \to Ax$$

Exercise 2.1 *Consider the differentiation operator on the set S:*

$$\begin{array}{cccc} \frac{d}{dt}: & S & \to & S \\ & f & \to & \frac{df}{dt} \end{array}.$$

Find a matrix representation for $\left(\frac{d}{dt}\right)$.

2.1 Range and Null Space of A

$\mathcal{R}(A)$, the **Range** of a matrix $A \in \mathbf{C}^{m\times n}$ is the following subspace of $\mathbf{C}^m$

$$\mathcal{R}(A) = \{Ax \,:\, x \in \mathbf{C}^n\}$$

The **rank** of A is defined as: $rank(A) = dim\mathcal{R}(A)$.

$\mathcal{N}(A)$, the **null space** of a matrix $A \in \mathbf{C}^{m\times n}$ is the following subspace of $\mathbf{C}^n$

$$\mathcal{N}(A) = \{x \,:\, Ax = 0\}$$

Exercise 2.2 *What is the range of* $\left(\frac{d}{dt}\right)$*? What is its rank? What is its null space?*

3 Eigenvalues and Eigenvectors

$A \in \mathbf{C}^{n\times n}$ is a square $n \times n$ matrix.

Definition 3.1 *$\lambda \in \mathbf{C}$ is an* **eigenvalue** *of A, if there exists a non zero vector v such that*

$$Av = \lambda v$$

v is called the right eigenvector of A relative to λ.

Interpretation: If we apply A to a vector x in the same direction as v, then A acts on x as a multiplication by a scalar of value λ.

Eigenvectors are not unique; if v is an eigenvector then αv, with $\alpha \in \mathbf{C}$ is also an eigenvector of A relative to the same eigenvalue.

Definition 3.2 *The* **eigenspace** *relative to an eigenvalue λ is the set of all eigenvectors relative to λ, that is:*

$$\{v \neq 0 \mid Av = \lambda v, \lambda \text{ eigenvalue of } A\} = \mathcal{N}(A - \lambda I)$$

Clearly, an eigenspace is a subspace.

3.1 Characterization of the eigenvalues of a matrix A

An eigenvalue must satisfy

$$\begin{aligned} Av = \lambda v \qquad \text{for some } v \neq 0, \quad &\Leftrightarrow (A - \lambda I)v = 0 \qquad \text{for some } v \neq 0. \\ &\Leftrightarrow (A - \lambda I) \quad \text{is not invertible} \\ &\Leftrightarrow det(A - \lambda I) = 0 \end{aligned}$$

$det(A - \lambda I)$ is called the characteristic polynomial of A.

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \qquad A - \lambda I = \begin{pmatrix} a_{11} - \lambda & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} - \lambda \end{pmatrix}$$

$$\chi(\lambda) = det(A - \lambda I) = \lambda^n + \alpha_{n-1}\lambda^{n-1} + \cdots + \alpha_1\lambda + \alpha_0$$

The eigenvalues of A are the roots of $\chi(\lambda)$.
Note that $\chi(\lambda)$ can have repeated roots. For example, $\chi(\lambda) = (\lambda - 3)^3(\lambda - 2)$. In these cases we say that $\lambda = 3$ is a multiple root of $\chi(\lambda)$.

Example 3.1 *Let's compute the eigenvalues of the following matrix:*

$$A = \begin{pmatrix} 0 & 1 \\ -2 & -3 \end{pmatrix}$$

$$det(A - \lambda I) = det \begin{pmatrix} -\lambda & 1 \\ -2 & -3 - \lambda \end{pmatrix} = \lambda(3 + \lambda) + 2 = \lambda^2 + 3\lambda + 2$$

The roots of $\chi(\lambda) = 0$ are $\lambda_1 = -1$, $\lambda_2 = -2$.

Definition 3.3 *The* **Algebraic Multiplicity AM**(λ) *of an eigenvalue* λ *is its multiplicity as root of* $\chi(\lambda) = 0$
The **Geometric Multiplicity GM**(λ) *of an eigenvalue* λ *is the dimension of its eigenspace.*

Fact 3.1 *For any eigenvalue,*

$$\mathbf{AM}(\lambda) \geq \mathbf{GM}(\lambda)$$

Example 3.2

$$A = \begin{pmatrix} 4 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}; \qquad \chi(\lambda) = (4-\lambda)^3(2-\lambda)$$

$\lambda = 4$ *is an eigenvalue with* **AM** *equal* 3.
$\lambda = 2$ *is an eigenvalue with* **AM** *equal* 1.
What about the **GM** *?*

$$\mathcal{N}(A - 4I) = \mathcal{N}\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -2 \end{pmatrix}$$

The dimension of the eigenspace of $\lambda = 4$ *is clearly* 2*, a possible basis for the eigenspace is:*

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}; \quad \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$

thus
$\lambda = 4$ *is an eigenvalue with* **GM** *equal* 2.
One can easily check that $\lambda = 2$ *is an eigenvalue with* **GM** *= 1.*

Theorem 3.1 *If an eigenvalue has* **AM** *equal* 1 *then its* **GM** *equals 1.*

Proof. $\mathbf{AM}(\lambda) \geq \mathbf{GM}(\lambda)$. But, from the definition of eigenspace, the dimension of the eigenspace is always greater than 1. Therefore, the result follows.

■

Next theorem shows a property of distinct eigenvalues of a matrix A.

Theorem 3.2 *Eigenvectors of distinct eigenvalues are linearly independent.*

Proof. The theorem is proved by contradiction. Let $\lambda_1 \neq \lambda_2$ be two eigenvalues, and v_1, v_2 two respective eigenvectors.
Suppose, to derive the contradiction, that v_1, v_2 are linearly dependent, i.e.,

$$\alpha_1 v_1 + \alpha_2 v_2 = 0 \qquad \text{for some } \alpha_1, \alpha_2 \in \mathbf{C}, \text{ not both equal to } 0 \tag{1}$$

Then

$$0 = A(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 \lambda_1 v_1 + \alpha_2 \lambda_2 v_2$$

Since $\alpha_1 v_1 + \alpha_2 v_2 = 0$ we can subtract $\lambda_2(\alpha_1 v_1 + \alpha_2 v_2)$ from the above equality, and get:

$$0 = \alpha_1(\lambda_1 - \lambda_2) v_1$$

but, $v_1 \neq 0$ by definition of eigenvector, and $(\lambda_1 - \lambda_2) \neq 0$ by hypothesis, thus α_1 must be zero.
In the same way, by subtracting $\lambda_1(\alpha_1 v_1 + \alpha_2 v_2)$ from Equation (1) we obtain:

$$0 = \alpha_2(\lambda_1 - \lambda_2)v_2$$

which implies that, also α_2, must be zero.
But $\alpha_1 = \alpha_2 = 0$ is a conclusion in contradiction with the hypothesis, therefore, v_1 and v_2 cannot be linearly dependent. ■

4 Similar Matrices and Similarity Transformations

Definition 4.1 *Two matrices P and Q are similar if there exists an invertible matrix T such that*

$$P = TQT^{-1}$$

TQT^{-1} is called a similarity transformation of Q.

Similarity transformations represent operations of change of basis or change of variable.

Example 4.1 *Let $y = Ax$ and suppose we want to express the relation between z and w, knowing that $z = Ty$ and $w = Tx$, then $z = TAT^{-1}w$*

Theorem 4.1 *If P and Q are similar, then they have the same eigenvalues.*

Proof.

$$\begin{aligned} det(P - \lambda I) &= det(TQT^{-1} - \lambda TT^{-1}) = det(T(Q - \lambda I)T^{-1}) \\ &= det(T)det(Q - \lambda I)det(T^{-1}) = det(Q - \lambda I) \end{aligned}$$

■

5 Diagonalizable Matrices

Definition 5.1 *A matrix A is said diagonalizable if it is similar to a diagonal matrix $\Lambda = diag(\lambda_1, \cdots, \lambda_n)$ where the λ_i's are the eigenvalues of A.*

When is a matrix diagonalizable?
Suppose we can find n linearly independent eigenvectors for A. then:

$$A \underbrace{[v_1, \ \cdots, \ v_n]}_{V} = \underbrace{[v_1, \ \cdots, \ v_n]}_{V} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

and V is invertible, thus:

$$A = V\Lambda V^{-1} \qquad \Lambda = V^{-1}AV$$

Clearly, from Theorem 3.2 if all the eigenvalues of A are distinct then the respective eigenvectors are linearly independent, therefore the matrix V is invertible. This is a special case of the following general condition.

Theorem 5.1 *A Matrix A is diagonalizable if and only if $\mathbf{AM}(\lambda_i) = \mathbf{GM}(\lambda_i)$ for all $i = 1, \ldots, n$.*

Proof. Left as exercise.

5.1 Diadic Formula

Definition 5.2 *The left eigenvector $w^T \neq 0$ of A relative to the eigenvalue λ is a vector satisfying*

$$w^T A = \lambda w^T$$

When A is diagonalizable, we can let $W = V^{-1}$, then

$$A = V\Lambda V^{-1} \Rightarrow WA = \Lambda W$$

$$W = \begin{pmatrix} w_1^T \\ \vdots \\ w_n^T \end{pmatrix}$$

w_i^T are the left eigenvectors of A. Note that VW=I.

Diadic formula

$$A = V\Lambda W = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} w_1^T \\ \vdots \\ w_n^T \end{bmatrix} = \sum_{i=1}^{n} \lambda_i \underbrace{v_i w_i^T}_{n\times n}$$

The Diadic formula is also referred as modal decomposition of A.

$$y = Ax = \sum_{i=1}^{n} \alpha_i \lambda_i v_i$$

where $\alpha_i = w_i^T x$. The output vector y is expressed as linear combination of the right eigenvectors of A. The coefficients of the linear combination are given by the eigenvalues of A multiplied by the inner product of the input vector x with the left eigenvectors of A.
Note if $x = v_i$ then $y = \lambda_i v_i$, since $WV = I$.

5.2 Power of a Matrix

We want to compute A^N. If A is diagonalizable then A^N can be easily rewritten as

$$A^N = (V\Lambda V^{-1})^N = \underbrace{V\Lambda V^{-1} \cdots V\Lambda V^{-1}}_{N} = V\Lambda^N V^{-1}$$

From the above expression, it results that, A^N and A have the same eigenvectors, and that the eigenvalues of A^N are the Nth power of the eigenvalues of A.

6 Jordan Form

When it is not possible to find n linearly independent eigenvectors of A, the matrix cannot be diagonalized.
The Jordan form is the "closest" form to a diagonal form, to which it is still possible to transform A by similarity transformation. $A = MJM^{-1}$.

The Jordan form relative to a matrix is unique (modulo blocks reordering). A matrix J is in Jordan form if:

$$J = \begin{pmatrix} J_1 & & & \\ & J_2 & 0 & \\ & 0 & \ddots & \\ & & & J_r \end{pmatrix}$$

J is block diagonal and J_i are square matrices with the following structure:

$$J_i = \begin{pmatrix} \lambda_j & 1 & 0 & \cdots & 0 \\ 0 & \lambda_j & 1 & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots 0 & \lambda_j & 1 \\ 0 & \cdots & \cdots & 0 & \lambda_j \end{pmatrix}$$

J_i is called a Jordan block, in this case J_i is Jordan block associated with λ_j. The number of Jordan blocks associated with an eigenvalue λ_j is equal to $\mathbf{GM}(\lambda_j)$.
The sum of the dimensions of all Jordan blocks associated with λ_j is equal to $\mathbf{AM}(\lambda_j)$

Example 6.1 *From the structure of the Jordan form of A one can obtain information about the algebraic and geometric multiplicity of the eigenvalues of A. Consider*

$$J = \left(\begin{array}{ccc|cc|cc|c} \lambda_1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ & \lambda_1 & 1 & 0 & 0 & 0 & 0 & 0 \\ & & \lambda_1 & 0 & 0 & 0 & 0 & 0 \\ \hline & & & \lambda_1 & 1 & 0 & 0 & 0 \\ & & & & \lambda_1 & 0 & 0 & 0 \\ \hline & & 0 & & & \lambda_2 & 1 & 0 \\ & & & & & & \lambda_2 & 0 \\ \hline & & & & & & & \lambda_3 \end{array}\right)$$

Then we can immediately say that: $\mathbf{AM}(\lambda_1) = 5$. $\mathbf{GM}(\lambda_1) = 2$ *(two Jordan block for* λ_1*).*
$\mathbf{AM}(\lambda_2) = 2$ *and* $\mathbf{GM}(\lambda_2) = 1$
$\mathbf{AM}(\lambda_3) = \mathbf{GM}(\lambda_3) = 1$.

The fact that any matrix can be put in its Jordan form by an opportune similarity transformation, makes the Jordan form very useful in linear system theory.
Unfortunately, the computation of the Jordan form is very sensitive to computational errors.

6.1 Generalized eigenvectors

Since there are not n linearly independet eigenvector to form a basis of $\mathbb{R}^n$, one can add linearly independent vectors to the eigenvectors in order to complete the basis. In what follows it is explained what vectors to add so that the associated similarity transformation of A gives J. These vectors are called generalized eigenvectors of A.

Definition 6.1 *A vector $x \neq 0$ is said to be a* generalized eigenvector *of order k of A relative to λ if and only if*

$$\begin{aligned}&(A-\lambda I)^k x = 0\\ &\textit{and}\\ &(A-\lambda I)^{k-1} x \neq 0\end{aligned}$$

Note that if $k = 1$, the above Definition reduces to $(A - \lambda I)x = 0$ and $x \neq 0$, which is the definition of an eigenvector.
Starting from a generalized eigenvector of order k of A relative to λ, denoted by x_k, one can generate a the Jordan chain of generalized eigenvectors as follows:

$$\begin{aligned}&x_k\\ &x_{k-1} = (A-\lambda I)x_k\\ &\vdots\\ &x_i = (A-\lambda I)^{k-i} x_k\\ &\vdots\\ &x_1 = (A-\lambda I)^{k-1} x_k\end{aligned} \tag{2}$$

Note that x_1 is an eigenvector of A, since $(A - \lambda I)x_1 = (A - \lambda I)(A - \lambda I)^{k-1}x_k = 0$. Note also that any x_i, $i = 1, \ldots, k$, is a generalized eigenvector of order i of A. This is true since:

$$\begin{gathered}(A-\lambda I)^i x_i = (A-\lambda I)^i (A-\lambda I)^{k-i} x_k = 0\\ \text{and}\\ (A-\lambda I)^{i-1} x_i = (A-\lambda I)^{k-1} x_k \neq 0\end{gathered}$$

Another immediate consequence of the definition is that all the generalized eigenvectors in the Jordan chain of length k belong to $\mathcal{N}(A - \lambda I)^k$.

Rearranging (2) immediately follows that the generalized eigenvectors in a Jordan chain satisfy:

$$\begin{aligned}&Ax_1 = \lambda x_1\\ &Ax_2 = \lambda x_2 + x_1\\ &\vdots\\ &Ax_k = \lambda x_k + x_{k-1}\end{aligned}$$

Rewriting the above exression in matrix form reveals the structure of a Jordan block of dimension k relative to an eigenvalue λ

$$A\left[x_1, \ldots, x_k\right] = \left[x_1, \ldots, x_k\right] \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0\\ 0 & \lambda & 1 & 0 & \vdots\\ \vdots & \ddots & \ddots & \ddots & 0\\ 0 & \cdots & \cdots 0 & \lambda & 1\\ 0 & \cdots & \cdots & 0 & \lambda \end{pmatrix}$$

Theorem 6.1 *Let x_k be a generalized eigenvector of A of order k relative to λ, and let $x_1, \ldots x_k$ be the chain of generalized eigenvector generated by x_k. Then $x_1, \ldots x_k$ are linearly independent vectors.*

Proof. We will prove that if

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k = 0 \tag{3}$$

then $\alpha_1, \ldots, \alpha_k = 0$.
First notice that $(A - \lambda I)^{k-1} x_i = 0$ for $1 \leq i \leq k-1$. This follows immediately from the fact that x_{k-1} is a generalized eigenvector of A of order $k-1$. Form the definition of x_k it also follows that $(A - \lambda I)^{k-1} x_k \neq 0$. Now, if $y = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k = 0$, then $(A - \lambda I)^{k-1} y = 0$ but this can only be true if $\alpha_k = 0$. If we multiply y by $(A - \lambda I)^{k-2}$ then it follows that $\alpha_{k-1} = 0$. Continuing the procedure we obtain that $\alpha_1, \ldots, \alpha_k = 0$. ■

6.2 Construction of the Jordan Chains

The following Facts allows to derive a procedure to construct the similarity transformation to put a matrix in its Jordan form.

Fact 6.1 *The generalized eigenvectors of A relative to different eigenvalues are linearly independent.*

To find the dimension of the largest Jordan block(s) relative to λ we use the following:

Fact 6.2 *Let λ be an eigenvalue of A with $\mathbf{AM}(\lambda) = m$. Let k be the smallest integer such that*

$$dim \mathcal{N}(A - \lambda I)^k = m$$

Then k is the dimension of the largest Jordan block(s) relative to λ. k is called the index of λ, denoted by $L(\lambda)$.

To find the Jordan chains of maximal length $k = L(\lambda)$ we use the following:

Fact 6.3 *Suppose that $dim \mathcal{N}(A - \lambda I)^{k-1} = m - s$ and let $\{w_1, \ldots, w_{m-s}\}$ be a basis for it. Then there are s linearly independent generalized vectors of order k, $x^{(k)})_{k,1}, \ldots, x^{(k)}_{k,s}$ such that the vectors*

$$\{x^{(k)}_{k,1}, \ldots, x^{(k)}_{k,s}, w_1, \ldots, w_{m-s}\}$$

form a basis for $\mathcal{N}(A - \lambda I)^k$.

The s generalized eigenvectors obtained in this way generate s Jordan chains of length k according to (2).

To find possible chains of length $k-1$ one must apply the above fact to $\mathcal{N}(A - \lambda I)^{k-2}$ as follows.

If $dim \mathcal{N}(A - \lambda I)^{k-2} = m - s - q$ then there are q Jordan chains of length $k-1$.
If $\{w_1, \ldots, w_{m-s-q}\}$ is a basis for $\mathcal{N}(A - \lambda I)^{k-2}$ The q generalized eigenvectors that generates these q Jordan chains are the vectors $x^{(k-1)}_{k-1,1}, \ldots, x^{(k-1)}_{k-1,q}$ and they are such that

$$\{x^{(k-1)}_{k-1,1}, \ldots, x^{(k-1)}_{k-1,q}, x^{(k)}_{k-1,1}, \ldots, x^{(k)}_{k-1,s}, w_1, \ldots, w_{m-s-q}\}$$

form a basis for $\mathcal{N}(A - \lambda I)^{k-1}$.

Analogously, repeat the above procedure to find possible Jordan chains of length $k-2$ and so forth.

6.3 Procedure for Computing the Jordan Form of A

1. Compute the eigenvalues of A by solving $det(A - \lambda I) = 0$. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of A with Algebraic Multiplicities $n_1, \ldots, n_m$, and Geometric Multiplicities $\gamma_1, \ldots, \gamma_m$, respectively.

2. Use Fact 6.2 compute L_1 the index of λ_1 . The dimension of the largest Jordan block.

3. Use repeatedly Fact 6.3 to find the number of Jordan blocks s_k of dimension k for $k = L_1, \ldots, 1$. Note that the $\sum_{k=1}^{L_1} s_k = \gamma_1$.

4. For each k for which a Jordan block is expected ($s_k > 0$), find a generalized eigenvector $x_{1,k}$ of A of order k, and construct the chain $\{x_{1,k}, \ldots, x_{1,1}\}$ according 2. Find the other $s_k - 1$ linearly independent generalized eigenvectors of order k and their relative Jordan chains. Collect all the vectors in the chains of order k as columns of $M_{1,k} = [x_{1,1}, \ldots, x_{1,k}, \ldots, x_{s_k,1}, \ldots, x_{s_k,k}]$. At the end of the procedure n_1 linearly independent generalized eigenvectors are generated. Let $M_1 = [M_{1,L_1}, \ldots, M_{1,1}]$, where some of the $M_{1,i}$'s may not be present.

5. Repeat step 2 for the rest of the eigenvalues. Let $M = [M_1, \ldots, M_m]$ then the Jordan form J of the matrix A is given by the following similarity transformation:

$$J = M^{-1}AM$$

Example 6.2 *Let*

$$A = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$\chi(\lambda) = \lambda^4$, *therefore A has 4 repeated eigenvalues $\lambda = 0$, i.e., $AM(\lambda) = 4$.*
It easy to see that $GM(\lambda) = 2$. Therefore we expect a Jordan form for A with 2 Jordan blocks. We still do not know their dimensions. By using Fact 6.2 to compute the index of λ, we have that

$$(A - \lambda I)^2 = A^2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The index of λ is 2, and therefore 2 is also the dimension of the largest blocks. It follows that there 2 Jordan blocks both of dimension 2×2.

A basis $\{w_1, w_2\}$ for $\mathcal{N}(A - \lambda I)^{k-1} = \mathcal{N}(A)$ is given by:

$$w_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}; \quad w_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

We now look for two generalized eigenvectors to generate the two Jordan chains of length 2. By Fact 6.3 they cannot be any linear combination of w_1 and w_2. If we choose

$$x_{2,1}^{(2)} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \qquad x_{2,2}^{(2)} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}$$

they are <u>not</u> O.K. since $\{x_{2,1}^{(2)}, x_{2,2}^{(2)}, w_1, w_2\}$ do not form a basis for $\mathcal{N}(A-\lambda I)^2$. The vectors:

$$x_{2,1}^{(2)} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \qquad x_{2,2}^{(2)} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

satisfy the conditions in Fact 6.3, therefore are valid generalized eigenvectors. We then compute the rest of the chains. In this case it happens that $(A - \lambda I)x_{2,1}^{(2)} = w_1$ and $(A - \lambda I)x_{2,2}^{(2)} = w_2$. Therefore the matrix $M = [w_1, x_{2,1}^{(2)}, w_2, x_{2,2}^{(2)}]$ of the similarity transformation and the relative Jordan form are

$$M = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad J = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Exercise Find the Jordan form J and the similarity transformation matrix M for the matrix

$$A = \begin{bmatrix} 4 & 2 & -1 \\ 0 & 6 & -1 \\ 0 & -4 & 6 \end{bmatrix}$$

7 Functions of a Square Matrix

7.1 Polynomial of a Square Matrix

We consider a square matrix $A \in \mathbf{C}^{n\times n}$. Define the k^{th} power of A as follows

$$A^k = \begin{cases} \underbrace{AA\cdots A}_{k} & \text{for } k \geq 1 \\ I & \text{for } k = 0 \end{cases}$$

The notion of matrix power allow to naturally define matrix polynomials. If $P(\lambda) = a_m\lambda^m + a_{m-1}\lambda^{m-1} + \cdots + a_1\lambda + a_0$ is an m^{th} degree polynomial in the scalar variable λ then the corresponding matrix polynomial is defined as

$$P(A) = a_mA^m + a_{m-1}A^{m-1} + \cdots + a_1A + a_0I$$

Notice that, if $P(\lambda)$ is written in factored form as

$$P(x) = c(\lambda - a_1)(x - a_2)\cdots(\lambda - a_m)$$

then $P(A)$ is given by

$$P(A) = c(A - a_1I)(A - a_2I)\cdots(A - a_mI)$$

A special polynomial is the characteristic polynomial of A:

$$\chi(\lambda) = det(A - \lambda I)$$

$\chi(\lambda)$ is a polynomial of order n, the dimension of A.

The matrix polynomial $\chi(A)$ has a very special property as stated by the important Cayley-Hamilton theorem.

Theorem 7.1 *Every square matrix A satisfies its own characteristic polynomial, i.e., $\chi(A) = \mathbf{0}$.*

Proof.

$$\chi(\lambda) = c_n\lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0 \tag{4}$$

We use the known result

$$(A - \lambda I)Adj[A - \lambda I] = det(A - \lambda I)I \tag{5}$$

that assumes a more familiar form in the formula for the inverse of a matrix M

$$M^{-1} = \frac{1}{det(M)}Adj[M]$$

where $Adj[A - \lambda I]_{i,j} = (-1)^{i+j}$ times the determinant of the $n-1 \times n-1$ matrix obtained by deleting the j^{th} row and the i^{th} column of $A - \lambda I$

Therefore, the highest power of λ that can be in any element of $Adj[A - \lambda I]$ is λ^{n-1}. This implies that it this possible to write

$$Adj[A - \lambda I] = B_{n-1}\lambda^{n-1} + B_{n-2}\lambda^{n-2} + \cdots + B_1\lambda + B_0 \tag{6}$$

where the B_i terms are $n \times n$ constant matrices not containing λ.

Substituting Equation (6) into the left side of Equation (5) we obtain:

$$\begin{aligned}(A - \lambda I)Adj[A - \lambda I] = \; & -B_{n-1}\lambda^n + (AB_{n-1} - B_{n-2})\lambda^{n-1} + (AB_{n-2} - N_{n-3})\lambda^{n-2} \\ & + \cdots + (AB_2 - B_1)\lambda^2 + (AB_1 + B_0)\lambda + AB_0\end{aligned}$$

Using Equation (4) on the right side of Equation (5) gives

$$\chi(\lambda)I = c_n\lambda^n I + c_{n-1}\lambda^{n-1}I + \cdots + c_1\lambda I + c_0 I$$

The left side equal the right side, and the coefficients of same power of λ on the two sides must be equal. This lead to the following set of equations

$$\begin{aligned} -B_{n-1} &= c_n I \\ AB_{n-1} - B_{n-2} &= c_{n-1} I \\ AB_{n-2} - B_{n-3} &= c_{n-2} I \\ &\vdots \\ AB_2 - B_1 &= c_2 I \\ AB_1 - B_0 &= c_1 I \\ AB_0 &= c_0 I \end{aligned}$$

Premultiply the first of these equations by A^n, the second by A^{n-1}, etc., the sum of the right side terms is equal to $\chi(A)$. The left side sum must be equal to $\chi(A)$ too. It is easy to verify that they add up to the 0 matrix, therefore $\chi(A) = 0$. ■

Cayley-Hamilton theorem can be used to compute the inverse of a matrix when it exists. $\chi(\lambda) = c_n\lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0$ Recall that $c_0 = det(A)$ and it is zero if and only if A is singular. Using $\chi(A) = 0$ and assuming that A^{-1} exists,

$$A\chi(A) = c_nA^{n-1} + c_{n-1}A^{n-2} + \cdots + c_1I + c_0A^{-1} = 0$$

or

$$A^{-1} = -\frac{1}{c_0}\left(c_nA^{n-1} + c_{n-1}A^{n-2} + \cdots + c_1I\right)$$

7.2 Reduction of a Polynomial in A to One of degree $n-1$ or Less

Let $P(\lambda)$ be a scalar polynomial of degree m. Let $P_1(\lambda)$ be another polynomial of degree $n < m$. then $P(\lambda)$ can always be written as $P(\lambda) = Q(\lambda)P_1(\lambda) + R(\lambda)$, where $Q(\lambda)$ (the quotient) is a polynomial of degree $m-n$, and $R(\lambda)$ (the remainder) is a polynomial of degree $n-1$. One can find Q and R by dividing $P(\lambda)/P_1(\lambda) = q(\lambda) + R(\lambda)/P_1(\lambda)$.

Similarly the matrix polynomial $P(A)$ can be written as

$$P(A) = Q(A)P_1(A) + R(A)$$

Now we have the following theorem:

Theorem 7.2 *Let $P(A)$ be a matrix polynomial, then*

$$P(A) = \sum_{i=1}^{n-1} \alpha_i A^i$$

Any matrix polynomial (of any order) is equal to a matrix polynomial of degree at most $n-1$.
Proof. $P(\lambda) = Q(\lambda)P_1(\lambda)+R(\lambda)$, choose $P_1(\lambda) = \chi(\lambda)$. Then by Cayley-Hamilton theorem $\chi(A) = 0$ thus $P(A) = Q(A)0 + R(A) = R(A)$.

■

Corollary 7.1 *Suppose $A \in \mathbf{C}^{n\times n}$ has l distinct eigenvalues $\lambda_1, \lambda_2 \ldots, \lambda_l$, let μ_i be the algebraic multiplicity of the eigenvalue λ_i for $i = 1...l$. Finally let $P(\lambda)$ be the polynomial of degree m and $R(\lambda)$ the polynomial of degree at most $n-1$ such that $P(A) = R(A)$ or*

$$P(\lambda) = Q(\lambda)\chi(\lambda) + R(\lambda).$$

Then

$$P^{(j)}(\lambda_i) = R^{(j)}(\lambda_i) \quad \text{for } j = 1, ..., \mu_i - 1 \qquad i = 1, ..., l \tag{7}$$

where $P^{(j)}(\lambda_i) = \left.\dfrac{d^jP(\lambda)}{d\lambda^j}\right|_{\lambda=\lambda_i}$.

Proof. Left as an exercise.
This corollary give a way to compute the coefficients of the matrix polynomial $R(A)$

Example 7.1 *Compute A^{100} where*

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$$

In other words, given $P(\lambda) = \lambda^{100}$ compute $P(A)$. The characteristic polynomial of A is $\chi(\lambda) = (\lambda - 1)^2$. Let $R(\lambda)$ be a polynomial of degree $n - 1 = 1$, say

$$R(\lambda) = \alpha_0 + \alpha_1 \lambda$$

Now from the previous corollary $P(A) = R(A)$ if

$$\begin{array}{ll} P(1) = R(1); & 1^{100} = \alpha_0 + \alpha 1 \\ P'(1) = R'(1); & 100 \cdot 1^{99} = \alpha 1 \end{array}$$

Solving these two equations we obtain $\alpha_1 = 100$ $\alpha_0 = -99$. Hence

$$A^{100} = R(A) = \alpha_0 I + \alpha_1 A = -99 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + 100 \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 200 \\ 0 & 1 \end{bmatrix}$$

7.3 Function of a Matrix

Definition 7.1 *Let $A \in \mathbf{C}^{n \times n}$ have l distinct eigenvalues $\lambda_1, \lambda_2 \ldots, \lambda_l$, let μ_i be the algebraic multiplicity of the eigenvalue λ_i for $i = 1...l$.*

Let $f(\lambda)$ be a function (not necessarily a polynomial) such that $f^{(j)}(\lambda_i)$ is well defined for all $j = 1, ...\mu_i$ and $i = 1, ..., l$. If $g(\lambda)$ is a polynomial such that

$$\begin{array}{ll} f^{(j)}(\lambda_i) = g^{(j)}(\lambda_i) & \text{for } j = 1, ..., \mu_i - 1 \\ & \quad i = 1, ..., l \end{array} \tag{8}$$

then the matrix-valued function $f(A)$ is defined as $f(A) \stackrel{\triangle}{=} g(A)$.

This definition is an extension of Corollary 7.1 to include polynomial as well as functions. If A is an $n \times n$ matrix, given the n values of $f(\lambda)$ on the spectrum of A, we can find a polynomial of degree $n - 1$,

$$g(\lambda) = \alpha_0 + \alpha_1 \lambda + \cdots + \alpha_{n-1} \lambda^{n-1}$$

which is equal to $f(\lambda)$ on the spectrum of A. Hence from this definition we know that every function of A can be expressed as

$$f(A) = \alpha_0 I + \alpha_1 A + \cdots + \alpha_{n-1} A^{n-1}$$

We summarize the procedure of computing a function of a matrix:
Given $A \in \mathbf{C}^{n \times n}$ and a function $f(\lambda)$, we first compute the eigenvalues of A and their algebraic multiplicity. Let

$$g(\lambda) = \alpha_0 + \alpha_1 \lambda + \cdots + \alpha_{n-1} \lambda^{n-1}$$

where the α's are unknown constants. Next use equations (**??**) to compute these α's in terms of the values of f on the spectrum of A. Finally have $f(A) = g(A)$.

Exercise 7.1 *Let*

$$A = \begin{bmatrix} 0 & 0 & -2 \\ 0 & 1 & 0 \\ 1 & 0 & 3 \end{bmatrix}$$

Compute e^{At}.

7.4 Functions of Matrix in Jordan form

One of the reasons to use the Jordan-form matrix is that if

$$J = \begin{bmatrix} J_1 & 0 \\ 0 & J_2 \end{bmatrix}$$

where J_1 and J_2 are square matrices, then

$$f(J) = \begin{bmatrix} f(J_1) & 0 \\ 0 & f(J_2) \end{bmatrix}$$

This can be easily verified by observing that

$$J^k = \begin{bmatrix} J_1^k & 0 \\ 0 & J_2^k \end{bmatrix}$$

Moreover for a Jordan block J_{ij} relative to an eigenvalue λ_i of dimension $n_{ij} \times n_{ij}$ we have that:

$$(J_{ij} - \lambda_i I) = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

$$(J_{ij} - \lambda_i I)^2 = \begin{bmatrix} 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

$$(J_{ij} - \lambda_i I)^{n_{ij}-1} = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

and $(J_{ij} - \lambda_i I)^k = 0$ for any integer $k \geq n_{ij}$.

To simplify the notation assume J consists of only one block of dimension $n \times n$, the extension to multiblock is immediate given the block-diagonal structure of the Jordan form.

Given

$$J = \begin{bmatrix} \lambda_1 & 1 & 0 & \cdots & 0 \\ 0 & \lambda_1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & \lambda_1 \end{bmatrix}$$

The characteristic polynomial of J is $\chi(\lambda) = (\lambda - \lambda_1)^n$. Let the polynomial $g(\lambda)$ be of the form

$$g(\lambda) = \beta_0 + \beta_1(\lambda - \lambda_1) + \beta_2(\lambda - \lambda_1)^2 + \cdots + \beta_{n-1}(\lambda - \lambda_1)^{n-1}$$

(Note that $g(\lambda)$ can always be rewritten as $g(\lambda) = \alpha_0 + \alpha_1\lambda + \alpha_2\lambda^2 + \cdots + \alpha_{n-1}\lambda^{n-1}$.)

Then the conditions in (**??**) give immediately

$$\beta_0 = f(\lambda_1); \quad \beta_1 = f'(\lambda_1); \cdots \beta_{n-1} = \frac{f^{(n-1)}(\lambda_1)}{(n-1)!}$$

Hence,

$$f(J) = g(J) = f(\lambda_1)I + \frac{f'(\lambda_1)}{1!}(J - \lambda_1 I) + \cdots + \frac{f^{(n-1)}(\lambda_1)}{(n-1)!}(J - \lambda_1 I)^{n-1} \tag{9}$$

$$f(j) = \begin{bmatrix} f(\lambda_1) & \frac{f'(\lambda_1)}{1!} & \frac{f''(\lambda_1)}{2!} & \cdots & \frac{f^{(n-1)}(\lambda_1)}{(n-1)!} \\ 0 & f(\lambda_1) & \frac{f'(\lambda_1)}{1!} & \cdots & \frac{f^{(n-2)}(\lambda_1)}{(n-2)!} \\ 0 & 0 & f(\lambda_1) & \cdots & \frac{f^{(n-3)}(\lambda_1)}{(n-3)!} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & f(\lambda_1) \end{bmatrix}$$

If $f(\lambda) = e^{\lambda t}$ then

$$e^{Jt} = \begin{bmatrix} e_1^{\lambda} & te^{\lambda_1 t} & t^2 \frac{e^{\lambda_1 t}}{2!} & \cdots & t^{n-1} \frac{e^{\lambda_1 t}}{(n-1)!} \\ 0 & e^{\lambda_1 t} & te^{\lambda_1 t} & \cdots & t^{n-2} \frac{e^{\lambda_1 t}}{(n-2)!} \\ 0 & 0 & e^{\lambda_1 t} & \cdots & t^{n-3} \frac{e^{\lambda_1 t}}{(n-3)!} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & e^{\lambda_1 t} \end{bmatrix}$$

Note that the derivatives are taken with respect to λ not t. Finally if $A = MJM^{-1}$ then $f(A) = Mf(J)M^{-1}$.

8 Functions of a Matrix Defined by Means of Power Series

We have used a polynomial of finite degree to define a function of a matrix. An alternative expression of a function of a matrix can be given as infinite power series. This is actually the way the matrix exponential is defined.

$$e^{At} = \sum_{i=0}^{\infty} \frac{A^i}{i!}$$

since it directly comes from the definition of the exponential function as infinite Taylor series.

Definition 8.1 *Let the power series representation of a function f be:*

$$f(\lambda) = \sum_{i=0}^{\infty} \alpha_i \lambda^i$$

with radius of convergence ρ. Then the function $f(A)$ is defined as

$$f(A) = \sum_{i=0}^{\infty} \alpha_i A^i$$

if the absolute values of all the eigenvalues of A are smaller than ρ, the radius of convergence, or the matrix has the property $A^k = 0$ for some positive integer k.

Note that this definition is meaningful only when the infinite series that defines $f(\lambda)$ converges. If the absolute values of all the eigenvalues of A are smaller then ρ, it can be shown that the infinite series that defines $f(A)$ converges. Definition 8.1 and Definition 7.1 lead exactly to the same matrix function.

Example 8.1 *Consider the Jordan form matrix J in the previous section. Let*

$$f(\lambda) = f(\lambda_1) + f'(\lambda_1)(\lambda - \lambda_1) + \frac{f''(\lambda_1)}{2!}(\lambda - \lambda_1)^2 + \cdots$$

then

$$f(J) \triangleq f(\lambda_1)I + f'(\lambda_1)(J - \lambda_1 I) + \cdots \frac{f^{(n-1)}(\lambda_1)}{(n-1)!}(J - \lambda_1 I)^{n-1} + \cdots \tag{10}$$

Since $(J - \lambda_1 I)^k = 0$ for $k \geq n$, the matrix function (10) reduces immediately to (9).

Example 8.2 *Consider the function*

$$f(\lambda) = \frac{1}{1 - \lambda}$$

$f(\lambda)$ has a singularity at $\lambda = 1$. For all λ in the complex plane inside the circle $|\lambda| < 1$, $f(\lambda)$ can be defined by the following infinite series

$$f(\lambda) = 1 + \lambda + \lambda^2 + \lambda^3 + \cdots$$

If all the eigenvalues of A satisfy λ_i satisfy $|\lambda_i| < 1$, then $f(A) = (I - A)^{-1}$ exists and can be written as a convergent series

$$f(A) = I + A + A^2 + A^3 + \cdots$$

Note that $(I - A)f(A) = I$ as required. When A has an eigenvalue at $\lambda = 1$, then $(I - A)$ is singular and the inverse does not exist. If $\lambda = 1$ is not an eigenvalue, but at least an eigenvalue of A has magnitude greater than 1, then $(I - A)^{-1}$ exists, but cannot be represented by the above infinite series. Note that $(I - A)^{-1}$ can always be found using Definition 7.1 when A has no eigenvalues at $\lambda = 1$.

Exercise 8.1 *Let*

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

Compute cos(At) in closed form.

9 Norms

Definition 9.1 *A norm is a function from a vector space X to the nonnegative real numbers that satisfies:*

a) $\|x\| \geq 0$ *and* $\|x\| = 0 \Rightarrow x = 0$

b) $\|\alpha x\| = |\alpha| \, \|x\| \quad \alpha \in \mathbf{C}$.

c) $\|x + y\| \leq \|x\| + \|y\| \quad \forall x, y \in X$.

For example in $\mathbf{C}^n$ we have that

$$\|x\|_2 = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2} = \sqrt{x^* x}, \quad x^* \text{ is the conjugate transpose of } x, \;\; x^* = (\overline{x}_1, \overline{x}_2 \cdots \overline{x}_n)$$

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$

$$\|x\|_\infty = \max_i |x_i|.$$

Using the above definitions, it is possible to define a norm on the space of all $m \times n$ matrices $A \in \mathbf{C}^{m\times n}$ by looking at $\mathbf{C}^{m\times n}$ as $\mathbf{C}^{mn}$, $\mathbf{C}^{m\times n} \simeq \mathbf{C}^{m\,n}$. As an example, the Frobenius norm of A is defined as follows:

$$\|A\|_F = \left(\sum_{i=1}^{m} \sum_{i=1}^{n} |a_{ij}|^2\right)^{1/2}$$

Norms defined for a matrix seen as an element of a vector space, are not so interesting. The so called induced norms are more important for us.
When the matrix is seen as a linear operator between vector spaces, the induced norm characterizes a measure of the maximum "gain" or amplification of the operator.

10 Inner Product

Definition 10.1 *An inner product is a bilinear function on a vector space X, denoted $\langle x, y\rangle$, with the following properties:*

1) $\langle x, x\rangle \geq 0$ *if* $\langle x, x\rangle = 0$ *then* $x = 0$

2) $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle$

3) $\langle x, y\rangle = \overline{\langle y, x\rangle}$

4) $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$.

Example 10.1 *Let ℓ_2 denote the space of all infinite sequences such that $\sum_{k=0}^{\infty} |x(k)|^2 < \infty$. This is an inner product space with inner product defined as $\langle x, y\rangle = \sum_{k=0}^{\infty} x(k)y(k)$.*

For any inner product space there is a well defined norm given by $\|x\|_2 = \sqrt{\langle x, x\rangle}$. We say that the 2-norm is compatible with an inner product.

Example 10.2 *For $\mathbf{C}^n$* $\quad \langle x, y\rangle = y^* x, \quad \|x\|_2 = \sqrt{\langle x, x\rangle}$.

Working in inner product spaces makes certain optimization problems more tractable.

Orthogonality

$x \perp y$ (x is orthogonal to y) if $\langle x, y\rangle = 0$

$x \perp S$ (x is orthogonal to the set or subspace S) if $\langle x, s\rangle = 0\ \forall\ s_1 \in S$

$S_1 \perp S_2$ (both are subspace) if $\langle s_1, s_2\rangle = 0\ \forall\ s_1 \in S_1$ and $s_2 \in S_2$

An inner product and its compatible 2-norm satisfy the following important inequality:
Cauchy Schwarz inequality $\quad |\langle x, y\rangle| \leq \|x\|_2 \, \|y\|_2$

(Recall that $|\langle x, y\rangle|^2 = \langle x, y\rangle\overline{\langle x, y\rangle}$)

Using the orthogonality condition it is very easy to prove the Pythagorean theorem:

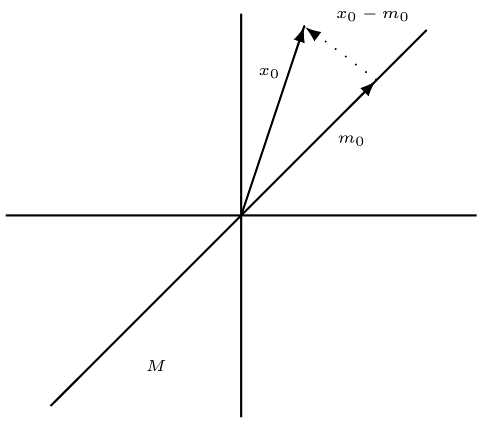
Theorem 10.1 *(Pythagorean Theorem)*

$$\text{If } x \perp y, \text{ then } \|x + y\|_2^2 = \|x\|_2^2 + \|y\|_2^2$$

The proof is left as an exercise.

11 Projection Theorem

This theorem is valid in any inner product space.



Theorem 11.1 *Let x_0 be a fixed element in an inner product space X, M is a closed subspace of X. Then*

$$\min_{m \in M} \|x_0 - m\|_2 = \|x_0 - m_0\|_2$$

and m_0 satisfies $(x_0 - m_0) \perp M$. Also m_0 is unique.

Proof. We will prove the theorem by contradiction. Assume that m_0 is the optimal solution but $x_0 - m_0$ is not orthogonal to M, i.e., $\langle x_0 - m_0, \tilde{m} \rangle = \delta$ for some $\tilde{m} \in M$ with $\|\tilde{m}\|_2 = 1$. Define $m_1 = m_0 + \delta\tilde{m}$, will show that

$$\|x_0 - m_1\|_2 < \|x_0 - m_0\|_2 \,,$$

and hence m_0 cannot be optimal. Note that

$$\begin{aligned}
\|x_0 - m_1\|_2^2 &= \langle x_0 - m_1, x_0 - m_1 \rangle \\
&= \langle x_0 - m_0 - \delta\tilde{m}, x_0 - m_0 - \delta\tilde{m} \rangle \\
&= \|x_0 - m_0\|_2^2 - \delta\langle \tilde{m}, x_0 - m_0 \rangle - \overline{\delta}\langle x_0 - m_0, \tilde{m} \rangle + |\delta|^2 \|\tilde{m}\|_2^2
\end{aligned}$$

but

$$\langle \tilde{m}, x_0 - m_0 \rangle = \overline{\delta} \quad \text{and } \langle x_0 - m_0, \tilde{m} \rangle = \delta$$

thus, since $\|\tilde{m}\|_2^2 = 1$, it follows that:

$$\|x_0 - m_1\|^2 = \|x_0 - m_0\|^2 - |\delta|^2$$

which contradicts the hypothesis that m_0 is optimal.

To show uniqueness, let m_0 and m_1 be two solutions, then

$$\begin{aligned}
\|x_0 - m_0\|_2^2 &= \|(x_0 - m_1) + (m_1 - m_0)\|_2^2 \\
&= \|(x_0 - m_1)\|_2^2 + \|m_1 - m_0\|_2^2 \qquad \text{since } \ (x_0 - m_1) \perp M \\
&\Rightarrow \|m_1 - m_0\| = 0 \Rightarrow m_1 = m_0
\end{aligned}$$

Exercise 11.1 *Does the projection theorem hold for* $\|\cdot\|_1$, $\|\cdot\|_\infty$ *?*
(Hint: there is no inner product compatible with these norms).

References

W. L. Brogan “Modern Control Theory” Prentice-Hall 1982.

T. C. Chen “Introduction to Linear System Theory” HRW 1970.

D. Luenberger ”Optimization by Vector Space Methods”