Lectures on Dynamic Systems and Control

Mohammed Dahleh   Munther A. Dahleh   George Verghese
Department of Electrical Engineering and Computer Science
Massachuasetts Institute of Technology[1]

Chapter 1

Linear Algebra Review

1.1 Introduction

Dynamic systems are systems that evolve with time. Our models for them will comprise coupled sets of ordinary differential equations (ode's). We will study how the internal variables and outputs of such systems respond to their inputs and initial conditions, how their internal behavior can be inferred from input/output (I/O) measurements, how the inputs can be controlled to produce desired behavior, and so on. Most of our attention will be focused on *linear* models (and within this class, on *time invariant* models, i.e. on LTI models), for reasons that include the following:

- linear models describe small perturbations from nominal operation, and most control design is aimed at regulating such perturbations;
- linear models are far more tractable than general nonlinear models, so systematic and detailed control design approaches can be developed;
- engineered systems are often made up of modules that are designed to operate in essentially linear fashion, with any nonlinearities introduced in carefully selected locations and forms.

To describe the interactions of coupled variables in linear models, the tools of *linear algebra* are essential. In the first part of this course (4 or 5 lectures), we shall come up to speed with the "$Ax = y$" or linear equations part of linear algebra, by studying a variety of *least squares* problems. This will also serve to introduce ideas related to dynamic systems — e.g., recursive processing of I/O measurements from a finite-impulse-response (FIR) discrete-time (DT) LTI system, to produce estimates of its impulse response coefficients.

Later parts of the course will treat in considerable detail the representation, structure, and behavior of multi-input, multi-output (MIMO) LTI systems. The "$Av = \lambda v$"

or eigenvalue–eigenvector part of linear algebra enters heavily here, and we shall devote considerable time to it. Along the way, and particularly towards the end of the course, we shall thread all of this together by examining approaches to control design, issues of robustness, etc., for MIMO LTI systems.

What you learn in this course will form a valuable, and even essential, foundation for further work in systems, control, estimation, identification, signal processing, and communication.

We now present a checklist of important notions from linear algebra for you to review, using your favorite linear algebra text. Some of the ideas (e.g. partitioned matrices) may be new.

1.2 Vector Spaces

Review the definition of a **vector space**: vectors, field of scalars, vector addition (which must be associative and commutative), scalar multiplication (with its own associativity and distributivity properties), the existence of a zero vector $\mathbf{0}$ such that $x + \mathbf{0} = x$ for every vector x, and the normalization conditions $0x = \mathbf{0}$, $1x = x$. Use the definition to understand that the first four examples below are vector spaces, while the fifth and sixth are not:

- $\mathbf{R}^n$ and $\mathbf{C}^n$.
- Real continuous functions $f(t)$ on the real line ($\forall t$), with obvious definitions of vector addition (add the functions pointwise, $f(t) + g(t)$) and scalar multiplication (scale the function by a constant, $af(t)$).
- The set of $m \times n$ matrices.
- The set of solutions $y(t)$ of the LTI ode $y^{(1)}(t) + 3y(t) = 0$.
- The set of points $[\ x_1 \quad x_2 \quad x_3\ ]$ in $\mathbf{R}^3$ satisfying $x_1^2 + x_2^2 + x_3^2 = 1$, i.e. "vectors" from the origin to the unit sphere.
- The set of solutions $y(t)$ of the LTI ode $y^{(1)}(t) + 3y(t) = \sin t$.

A **subspace** of a vector space is a subset of vectors that itself forms a vector space. To verify that a set is a subspace, all we need to check is that the subset is *closed under vector addition and under scalar multiplication*; try proving this. Give examples of subspaces of the vector space examples above.

- Show that the *range* of any real $n \times m$ matrix and the *nullspace* of any real $m \times n$ matrix are subspaces of $\mathbf{R}^n$.
- Show that the set of all linear combinations of a given set of vectors forms a subspace (called the subspace *generated by* these vectors, also called their linear *span*).

- Show that the *intersection* of two subspaces of a vector space is itself a subspace.
- Show that the *union* of two subspaces is in general *not* a subspace. Also determine under what condition the union of subspaces will be a subspace.
- Show that the (Minkowski or) *direct sum* of subspaces, which by definition comprises vectors that can be written as the sum of vectors drawn from each of the subspaces, is a subspace.

Get in the habit of working up small (in $\mathbf{R}^2$ or $\mathbf{R}^3$, for instance) concrete examples for yourself, as you tackle problems such as the above. This will help you develop a feel for what is being stated — perhaps suggesting a strategy for a proof of a claim, or suggesting a counterexample to disprove a claim.
Review what it means for a set of vectors to be **(linearly) dependent** or **(linearly) independent**. A space is *n-dimensional* if every set of more than n vectors is dependent, but there is some set of n vectors that is independent; any such set of n independent vectors is referred to as a **basis** for the space.

- Show that any vector in an n-dimensional space can be written as a *unique* linear combination of the vectors in a basis set; we therefore say that any basis set *spans* the space.
- Show that a basis for a *sub*space can always be augmented to form a basis for the entire space.

If a space has a set of n independent vectors for every nonnegative n, then the space is called *infinite dimensional*.

- Show that the set of functions $f(t) = t^{n-1}\ ,\ n = 1, 2, 3, \cdots$ forms a basis for an infinite dimensional space. (One route to proving this uses a key property of *Vandermonde* matrices, which you may have encountered somewhere.)

Norms

The "lengths" of vectors are measured by introducing the idea of a **norm**. A norm for a vector space $\mathcal{V}$ over the field of real numbers $\mathbf{R}$ or complex numbers $\mathbf{C}$ is defined to be a function that maps vectors x to nonnegative real numbers $\|x\|$, and that satisfies the following properties:

1. Positivity: $\|x\| > 0$ for $x \neq 0$
2. Homogeneity: $\|ax\| = |a|\,\|x\|\ ,\quad$ scalar a.
3. Triangle inequality: $\|x + y\| \leq \|x\| + \|y\|\ ,\quad \forall x, y \in \mathcal{V}$.

- Verify that the usual Euclidean norm on $\mathbf{R}^n$ or $\mathbf{C}^n$ (namely $\sqrt{x'x}$ with $'$ denoting the complex conjugate of the transpose) satisfies these conditions.

- A complex matrix Q is termed **Hermitian** if $Q' = Q$; if Q is real, then this condition simply states that Q is symmetric. Verify that $x'Qx$ is always real, if Q is Hermitian. A matrix is termed **positive definite** if $x'Qx$ is real and positive for $x \neq 0$. Verify that $\sqrt{x'Qx}$ constitutes a norm if Q is Hermitian and positive definite.

- Verify that in $\mathbf{R}^n$ both $\|x\|_1 = \sum_1^n |x_i|$ and $\|x\|_\infty = \max_i |x_i|$ constitute norms. These are referred to as the 1-norm and ∞-norm respectively, while the examples of norms mentioned earlier are all instances of (weighted or unweighted) 2-norms. Describe the sets of vectors that have unit norm in each of these cases.

- The space of continuous fucntions on the interval [0, 1] clearly forms a vector space. One possible norm defined on this space is the ∞-norm defined as:

$$\|f\|_\infty = \sup_{t \in [0,1]} |f(t)|.$$

  This measures the peak value of the function in the interval $[0, 1]$. Another norm is the 2-norm defined as:

$$\|f\|_2 = \left(\int_0^1 |f(t)|^2 dt \right)^{\frac{1}{2}}.$$

  Verify that these measures satisfy the three properties of the norm.

Inner Product

The vector spaces that are most useful in practice are those on which one can define a notion of **inner product**. An inner product is a function of two vectors, usually denoted by $< x, y >$ where x and y are vectors, with the following properties:

1. Symmetry: $< x, y > = < y, x >'$.
2. Linearity: $< x, ay + bz > = \ a < x, y > \ + b < x, z >$ for all scalars a and b.
3. Positivity: $< x, x >$ positive for $x \neq 0$.

- Verify that $\sqrt{< x, x >}$ defines a norm.

- Verify that $x'Qy$ constitutes an inner product if Q is Hermitian and positive definite. The case of $Q = I$ corresponds to the usual Euclidean inner product.

- Verify that

$$\int_0^1 x(t)y(t)dt$$

  defines an inner product on the space of continuous functions. In this case, the norm generated from this inner product is the same as the 2-norm defined earlier.

- **Cauchy-Schwartz Inequality** Verify that for any x and y in an inner product space

$$| < x, y > | \leq \|x\| \|y\|$$

with equality if and only if $x = \alpha y$ for some scalar α. (Hint: Expand $< x+\alpha y, x+\alpha y >$).

Two vectors x, y are said to be **orthogonal** if $< x, y >= 0$; two *sets* of vectors $\mathcal{X}$ and $\mathcal{Y}$ are called orthogonal if *every* vector in one is orthogonal to *every* vector in the other. The **orthogonal complement** of a set of vectors $\mathcal{X}$ is the set of vectors orthogonal to $\mathcal{X}$, and is denoted by $\mathcal{X}^{\perp}$.

- Show that the orthogonal complement of any set is a subspace.

1.3 The Projection Theorem

Consider the following minimization problem:

$$\min_{m \in M} \|y - m\|$$

where the norm is defined through an inner product. The projection theorem (suggested by the figure below), states that the optimal solution $\hat{m}$ is characterized as follows:

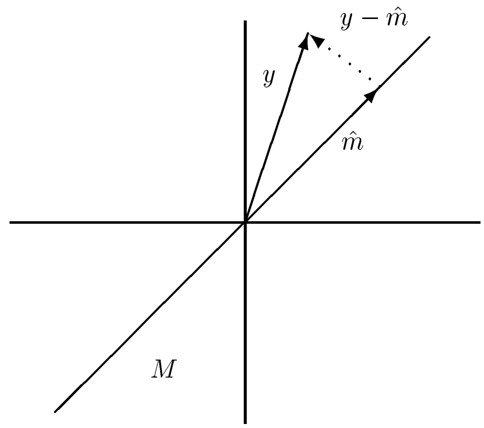
$$(y - \hat{m}) \perp M.$$

To verify this theorem, assume the converse. Then there exists an m_0, $\|m_0\| = 1$, such that $< y - \hat{m}, m_0 >= \delta \neq 0$. We now argue that $(\hat{m} + \delta m_0) \in M$ achieves a smaller value to the above minimization problem. In particular,

$$\begin{aligned} \|y - \hat{m} - \delta m_0\|^2 &= \|y - \hat{m}\|^2 - < y - \hat{m}, \delta m_0 > - < \delta m_0, y - \hat{m} > + |\delta|^2 \|m_0\|^2 \\ &= \|y - \hat{m}\|^2 - |\delta|^2 - |\delta|^2 + |\delta|^2 \\ &= \|y - \hat{m}\|^2 - |\delta|^2 \end{aligned}$$

This conradicts the optimality of $\hat{m}$.

- Given a subspace $\mathcal{S}$, show that any vector x can be *uniquely* written as $x = x_{\mathcal{S}} + x_{\mathcal{S}^{\perp}}$, where $x_{\mathcal{S}} \in \mathcal{S}$ and $x_{\mathcal{S}^{\perp}} \in \mathcal{S}^{\perp}$.



1.4 Matrices

Our usual notion of a matrix is that of a rectangular array of scalars. The definitions of matrix addition, multiplication, etc., are aimed at compactly representing and analyzing systems of equations of the form

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= y_1 \\ \cdots &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= y_m \end{aligned}$$

This system of equations can be written as $Ax = y$ if we define

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ldots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \quad , \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad , \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$$

The rules of matrix addition, matrix multiplication, and scalar multiplication of a matrix remain unchanged if the entries of the matrices we deal with are themselves (conformably dimensioned) *matrices* rather than scalars. A matrix with matrix entries is referred to as a **block** matrix or a **partitioned** matrix.

For example, the a_{ij}, x_j, and y_i in respectively A, x, and y above can be matrices, and the equation $Ax = y$ will still hold, as long as the dimensions of the various submatrices are conformable with the expressions $\sum a_{ij}x_j = y_i$ for $i = 1, \cdots, m$ and $j = 1, \cdots, n$. What this requires is that the number of rows in a_{ij} should equal the number of rows in y_i, the number of columns in a_{ij} should equal the number of rows in x_j, and the number of columns in the x_j and y_i should be the same.

- Verify that

$$\left(\begin{array}{cc|c} 1 & 2 & 2 \\ 0 & 1 & 3 \\ 1 & 1 & 7 \end{array}\right)\left(\begin{array}{cc} 4 & 5 \\ 8 & 9 \\ \hline \\ 2 & 0 \end{array}\right) = \left(\begin{array}{cc} 1 & 2 \\ 0 & 1 \\ 1 & 1 \end{array}\right)\left(\begin{array}{cc} 4 & 5 \\ 8 & 9 \end{array}\right) + \left(\begin{array}{c} 2 \\ 3 \\ 7 \end{array}\right)\left(\begin{array}{cc} 2 & 0 \end{array}\right)$$

In addition to these simple rules for matrix addition, matrix multiplication, and scalar multiplication of partitioned matrices, there is a simple — and simply verified — rule for (complex conjugate) *transposition* of a partitioned matrix: if $[A]_{ij} = a_{ij}$, then $[A']_{ij} = a'_{ji}$, i.e., the (i, j)-th block element of A' is the *transpose* of the (j, i)-th block element of A.

For more involved matrix operations, one has to proceed with caution. For instance, the determinant of the square block-matrix

$$A = \left(\begin{array}{cc} A_1 & A_2 \\ A_3 & A_4 \end{array}\right)$$

is clearly *not* $A_1A_4 - A_3A_2$ unless all the blocks are actually scalar! We shall lead you to the correct expression (in the case where A_1 is square and invertible) in a future Homework.

Matrices as Linear Transformations

T is a transformation or mapping from X to Y, two vector spaces, if it associates to each $x \in X$ a unique element $y \in Y$. This transformation is linear if it satisfies

$$T(\alpha x + \beta y) = \alpha T(x) + \beta T(y).$$

- Verify that an $n \times m$ matrix A is a linear transformation from $\mathbf{R}^m$ to $\mathbf{R}^n$.

Does every linear transformation have a matrix representation? Assume that both X and Y are finite dimensional spaces with respective bases $\{x_1, \ldots x_m\}$ and $\{y_1, \ldots y_n\}$. Every $x \in X$ can be uniquely expressed as: $x = \sum_{i=1}^{m} a_i x_i$. Equivalently, every x is represented uniquely in terms of an element $a \in \mathbf{R}^m$. Similarly every element $y \in Y$ is uniquely represented in terms of an element $b \in \mathbf{R}^n$. Now: $T(x_j) = \sum_{i=1}^{n} b_{ij} y_i$ and hence

$$T(x) = \sum_{j=1}^{m} a_j T(x_j) = \sum_{i=1}^{n} y_i (\sum_{j=1}^{m} a_j b_{ij})$$

A matrix representation is then given by $B = (b_{ij})$. It is evident that a matrix representation is not unique and depends on the basis choice.

1.5 Linear Systems of Equations

Suppose that we have the following system of real or complex linear equations:

$$A^{m\times n}\, x^{n\times 1} = y^{m\times 1}$$

When does this system have a solution x for given A and y?

$$\exists \text{ a solution } x \iff y \in \mathcal{R}(A) \iff \mathcal{R}([A\,|\,y]) = \mathcal{R}(A)$$

We now analyze some possible cases:

(1) If $n = m$, then $\det(A) \neq 0 \Rightarrow x = A^{-1}y$, and x is the unique solution.

(2) If $m > n$, then there are more equations than unknowns, i.e. the system is "overconstrained". If A and/or y reflect actual experimental data, then it is quite likely that the n-component vector y does *not* lie in $\mathcal{R}(A)$, since this subspace is only n-dimensional (if A has full column rank) or less, but lives in an m-dimensional space. The system will then be *inconsistent*. This is the sort of situation encountered in estimation or identification problems, where x is a parameter vector of low dimension compared to the dimension of the measurements that are available. We then look for a choice of x that comes closest to achieving consistency, according to some error criterion. We shall say quite a bit more about this shortly.

(3) If $m < n$, then there are fewer equations than unknowns, and the system is "underconstrained". If the system has a particular solution x_p (and when $\text{rank}(A) = m$, there is guaranteed to be a solution for any y) then there exist an infinite number of solutions. More specifically, x is a solution iff (if and only if)

$$x = x_p + x_h\ , \quad Ax_p = y\ , \quad Ax_h = 0 \quad \text{i.e. } x_h \in \mathcal{N}(A)$$

Since the nullspace $\mathcal{N}(A)$ has dimension at least $n - m$, there are at least this many degrees of freedom in the solution. This is the sort of situation that occurs in many control problems, where the control objectives do not uniquely constrain or determine the control. We then typically search among the available solutions for ones that are optimal according to some criterion.

Exercises

Exercise 1.1 Partitioned Matrices

Suppose

$$A = \begin{pmatrix} A_1 & A_2 \\ 0 & A_4 \end{pmatrix}$$

with A_1 and A_4 *square*.

(a) Write the determinant $\det A$ in terms of $\det A_1$ and $\det A_4$. (Hint: Write A as the product

$$\begin{pmatrix} I & 0 \\ 0 & A_4 \end{pmatrix} \begin{pmatrix} A_1 & A_2 \\ 0 & I \end{pmatrix}$$

and use the fact that the determinant of the product of two *square* matrices is the product of the individual determinants — the individual determinants are easy to evaluate in this case.)

(b) Assume for this part that A_1 and A_4 are *nonsingular* (i.e., square and invertible). Now find A^{-1}. (Hint: Write $AB = I$ and partition B and I commensurably with the partitioning of A.)

Exercise 1.2 Partitioned Matrices

Suppose

$$A = \begin{pmatrix} A_1 & A_2 \\ A_3 & A_4 \end{pmatrix}$$

where the A_i are matrices of conformable dimension.

(a) What can A be premultiplied by to get the matrix

$$\begin{pmatrix} A_3 & A_4 \\ A_1 & A_2 \end{pmatrix} \quad ?$$

(b) Assume that A_1 is nonsingular. What can A be premultiplied by to get the matrix

$$\begin{pmatrix} A_1 & A_2 \\ 0 & C \end{pmatrix}$$

where $C = A_4 - A_3 A_1^{-1} A_2$?

(c) Suppose A is a square matrix. Use the result in (b) — and the fact mentioned in the hint to Problem 1(a) — to obtain an expression for $\det(A)$ in terms of determinants involving only the submatrices A_1, A_2, A_3, A_4.

Exercise 1.3 Matrix Identities

Prove the following *very useful* matrix identities. In proving identities such as these, see if you can obtain proofs that make as few assumptions as possible beyond those implied by the problem statement. For example, in (1) and (2) below, neither A nor B need be square, and in (3) neither B nor D need be square — so avoid assuming that any of these matrices is (square and) invertible!.

(a) $\det(I - AB) = \det(I - BA)$, if A is $p \times q$ and B is $q \times p$. (Hint: Evaluate the determinants of

$$\begin{pmatrix} I & A \\ B & I \end{pmatrix} \begin{pmatrix} I & -A \\ 0 & I \end{pmatrix}, \qquad \begin{pmatrix} I & -A \\ 0 & I \end{pmatrix} \begin{pmatrix} I & A \\ B & I \end{pmatrix}$$

to obtain the desired result). One common situation in which the above result is useful is when $p > q$; why is this so?

(b) Show that $(I - AB)^{-1}A = A(I - BA)^{-1}$.

(c) Show that $(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$. (Hint: Multiply the right side by $A + BCD$ and cleverly gather terms.) This is perhaps the most used of matrix identities, and is known by various names — the matrix inversion lemma, the $ABCD$ lemma (!), Woodbury's formula, etc. It is rediscovered from time to time in different guises. Its noteworthy feature is that, if A^{-1} is known, then the inverse of a modification of A is expressed as a modification of A^{-1} that may be simple to compute, e.g. when C is of small dimensions. Show, for instance, that evaluation of $(I - ab^T)^{-1}$, where a and b are column vectors, only requires inversion of a scalar quantity.

Exercise 1.4 Range and Rank

This is a practice problem in linear algebra (except that you have perhaps only seen such results stated for the case of real matrices and vectors, rather than *complex* ones — the extensions are routine).

Assume that $A \in \mathbf{C}^{m \times n}$ (i.e., A is a complex $m \times n$ matrix) and $B \in \mathbf{C}^{n \times p}$. We shall use the symbols $\mathcal{R}(A)$ and $\mathcal{N}(A)$ to respectively denote the range space and null space (or kernel) of the matrix A. Following the Matlab convention, we use the symbol A' to denote the transpose of the *complex conjugate of* the matrix A; $\mathcal{R}^{\perp}(A)$ denotes the subspace *orthogonal* to the subspace $\mathcal{R}(A)$, i.e. the set of vectors x such that $x'y = 0\,, \quad \forall y \in \mathcal{R}(A)$, etc.

(a) Show that $\mathcal{R}^{\perp}(A) = \mathcal{N}(A')$ and $\mathcal{N}^{\perp}(A) = \mathcal{R}(A')$.

(b) Show that

$$\text{rank}(A) + \text{rank}(B) - n \le \text{rank}(AB) \le \min\{\text{rank}(A), \text{rank}(B)\}$$

This result is referred to as *Sylvester's inequality.*

Exercise 1.5 Vandermonde Matrix

A matrix with the following structure is referred to as a *Vandermonde* matrix:

$$\begin{pmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{n-1} \\ 1 & \lambda_2 & \lambda_2^2 & \cdots & \lambda_2^{n-1} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & \lambda_n & \lambda_n^2 & \cdots & \lambda_n^{n-1} \end{pmatrix}$$

This matrix is clearly singular if the λ_i are *not* all distinct. Show the converse, namely that if all n of the λ_i are distinct, then the matrix is nonsingular. One way to do this — although **not** the easiest! — is to show by induction that the determinant of the Vandermonde matrix is

$$\prod_{i=1\,;\,j>i}^{i,j=n} (\lambda_j - \lambda_i)$$

Look for an easier argument first.

Exercise 1.6 Matrix Derivatives

(a) Suppose $A(t)$ and $B(t)$ are matrices whose entries are differentiable functions of t, and assume the product $A(t)B(t)$ is well-defined. Show that

$$\frac{d}{dt}\Big(A(t)B(t)\Big) = \frac{dA(t)}{dt}B(t) \,+\, A(t)\frac{dB(t)}{dt}$$

where the derivative of a matrix is, by definition, the matrix of derivatives — i.e., to obtain the derivative of a matrix, simply replace each entry of the matrix by its derivative. (Note: The ordering of the matrices in the above result is important!).

(b) Use the result of (a) to evaluate the derivative of the *inverse* of a matrix $A(t)$, i.e. evaluate the derivative of $A^{-1}(t)$.

Exercise 1.7 Suppose T is a linear transformation from X to itself. Verify that any two matrix representations, A and B, of T are related by a nonsingular transformation; i.e., $A = R^{-1}BR$ for some R. Show that as R varies over all nonsingular matrices, we get all possible representations.

Exercise 1.8 Let X be the vector space of polynomials of order less than or equal to M.

(a) Show that the set $B = \{1, x, \ldots x^M\}$ is a basis for this vector space.

(b) Consider the mapping T from X to X defined as:

$$f(x) = Tg(x) = \frac{d}{dx}g(x)$$

1. Show that T is linear.
2. Derive a matrix representation for T in terms of the basis B.
3. What are the eigenvalues of T.
4. Compute one eigenvector associated with one of the eighenvalues.