2 PRELIMINARIES

2.1 Norms, inner products, etc.

Definition 2.1. *A real scalar valued function, $\|.\|$, is a norm if it satisfies three properties:*

- $\|x\| \geq 0$, *and* $\|x\| = 0$ *iff* $x = 0$
- $\|\alpha x\| = |\alpha| \, \|x\|, \quad \forall \ \alpha \in \mathcal{C}$
- $\|x + y\| \leq \|x\| + \|y\|$ *(Triangle inequality)*

Definition 2.2. *A complex valued function, $\langle ., . \rangle$, is an inner product if it satisfies the following basic properties:*

- $\langle x, y \rangle = \overline{\langle y, x \rangle}$
- $\langle \alpha x, (\beta_1 y_1 + \beta_2 y_2) \rangle = \overline{\alpha} \, \beta_1 \langle x, y_1 \rangle + \overline{\alpha} \, \beta_2 \langle x, y_2 \rangle \quad \forall \quad \alpha \, , \, \beta_1 \, , \, \beta_2 \quad \in \mathcal{C}$
- $\langle x, x \rangle \geq 0$, *and* $\langle x, x \rangle = 0$ *iff* $x = 0$

Remark 2.3. *A very common norm is the one based on the inner product; i.e.*

$$\|x\|^2 = \langle x, x \rangle \tag{2.1}$$

Definition 2.4. *Two vectors are orthogonal if their inner product is zero; i.e.,*

$$\langle x, y \rangle = 0 \Leftrightarrow x \perp y.$$

Furthermore, two subspaces are orthogonal to each other if any vector from one is orthogonal to any vector from the other; i.e.,

$$\mathcal{X} \perp \mathcal{Y} \Leftrightarrow \langle x, y \rangle = 0 \quad \forall \ x \ \in \mathcal{X} \ and \ \forall \quad y \ \in \ \mathcal{Y}.$$

Example 2.5. *For vectors in $\mathcal{C}^n$, the following are all norms:*

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

$$\|x\|_2 = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{\frac{1}{2}}$$

$$\|x\|_\infty = \max_i \{|x_i|\}$$

where x_i is the i^{th} entry of vector x.

Example 2.6. *Consider scalar functions of time defined for $t \in [0, T]$. We can define the following norms*

$$\|x(.)\|_1 = \int_0^T |x(t)|\, dt$$

$$\|x(.)\|_2 = \Big(\int_0^T |x(t)|^2\, dt \Big)^{\frac{1}{2}}$$

$$\|x(.)\|_\infty = \max_{T \geq t \geq 0} \{|x(t)|\}$$

Example 2.7. *For vectors in $\mathcal{C}^n$, or $\mathcal{R}^n$, the following are all proper inner products*

$$\langle x, y \rangle = \overline{x}^T y$$

$$\langle x, y \rangle_M = \overline{x}^T M y, \quad \textit{for some } M > 0 \ (\textit{positive definite})$$

Example 2.8. *Consider scalar and continuous functions $x(t)$ and $y(t)$ defined for $t \in [0, T]$. We can define the following inner product*

$$\langle x, y \rangle_{L_2} = \int_0^T \overline{x(t)}\, y(t)\, dt$$

and, indeed, $x(t)$ and $y(t)$ can be n-vector as well!

2.2 Linear Operators

Consider two vector spaces $\mathcal{X}$ and $\mathcal{Y}$ (e.g., $\mathcal{R}^n$ and $\mathcal{R}^m$, but it could be more complicated and general). A function T that sends every vector $x \in \mathcal{X}$ to a vector $y = Tx \in \mathcal{Y}$ is called a linear operator (or linear transformation) on - or from - $\mathcal{X}$ to $\mathcal{Y}$ if it preserves linear relations; that is if

$$T(a_1x_1 + a_2x_2) = a_1Tx_1 + a_2Tx_2$$

for all x_1, and x_2 in $\mathcal{X}$, as well as all scalars a_1 and a_2. We often use the notation $T : \mathcal{X} \rightarrow \mathcal{Y}$. Matrices, derivatives, integrals, are examples of linear operators (what are $\mathcal{X}$, $\mathcal{Y}$, x, etc in each case?).

So such a T can be considered an operator. It can also be considered an element (vector) of a large collection of operators that share the main properties (i.e., linear operators from $\mathcal{X}$ to $\mathcal{Y}$). The big set or collection of these operators actually is a linear vector space itself (why?), often denoted by $\mathcal{L}(\mathcal{X}, \mathcal{Y})$. So T can be considers a vector in $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ as well, for which a vector norm can be obtained as well!

Example 2.9. *Every matrix is a linear operator (work all details)*

Example 2.10. *The operation $\frac{d}{dt}$ is a linear operator ($y = \dot{x} = \mathcal{L}x$). Note that clearly we have x from the space of continuous function while y may not be. Similarly, simple integrals are linear operators.*

Since linear operators are elements of sets (i.e., vectors in subspaces), one can define vector norms (or metric) for them. As an example, consider matrices that are $m \times n$: a common vector norm for this matrix is

$$\|A\|_F = (\sum_i \sum_j |a_{i,j}|^2)^{.5}$$

where $a_{i,j}$ is the (i,j) element of the matrix A. This norm, often called the Frobenius norm, may indicate the 'size' of a given matrix compared to other matrices.

A more interesting, and useful, way to 'size' a linear operator is by examining what it does to the vectors it operates on. We define the operator norm for the space of linear operators (i.e, $\mathcal{L}$).

Definition 2.11. *The operator norm, or the induced norm, of an operator A is the supremum of the ratio of $\|Ax\|$ to $\|x\|$, over all non zero x; i.e.*

$$\|A\|_i = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|}$$

where the subscript i is often used to emphasize the fact that the induced norm is used. When clear from the context, this subscript is dropped.

Remark 2.12. *The norms in the fraction are vectors norm, from potentially different spaces. While they are often the same; e.g., both are 2-norms, they could be mixed. A common example is energy to peak norm; i.e.,* $\frac{\|Ax\|_\infty}{\|x\|_2}$ *which measures the worst peak per unit of input energy.*

Remark 2.13. *The form used in* definition 2.11 *also applies to nonlinear operators and is used in many nonlinear analysis problems. For the linear case, we can also use* $\sup_{\|x\|=1} \|Ax\|$.

Remark 2.14. *'Sup' is the lowest upper bound. When it is achieved, it is 'max' (which is 99% of problems!). To calculate the induced norm, a three step process is followed: (1) Come up with an estimate (guess), (2) show it is a legitimate upper bound (i.e., bigger that the ratio for all possible x), and (3) show that with a clever choice of x, one can either achieve it or get arbitrarily close to it.*

Example 2.15. *For $m \times n$ matrices, we have*

$$\sup_{x\neq 0} \frac{\|Ax\|_1}{\|x\|_1} = \max_j \sum_{i=1}^{m} |a_{i,j}|$$

$$\sup_{x\neq 0} \frac{\|Ax\|_2}{\|x\|_2} = [\,\lambda_{max}(\bar{A}^T A)\,]^{\frac{1}{2}}$$

$$\sup_{x\neq 0} \frac{\|Ax\|_\infty}{\|x\|_\infty} = \max_i \sum_{j=1}^{n} |a_{i,j}|$$

Before we leave this subsection, it is useful to review the concept of adjoint operator.

Definition 2.16. *Consider the linear operator A from $\mathcal{X}$ to $\mathcal{Y}$, with inner products $\langle .,. \rangle_x$ and $\langle .,. \rangle_y$ defined on $\mathcal{X}$ and $\mathcal{Y}$, respectively. Then the adjoint operator A^* is defined by the operator that satisfies the following*

$$\langle Ax, y\rangle_y = \langle x, A^* y\rangle_x \;\forall\; x \in \mathcal{X}, \quad y \in \mathcal{Y}.$$

Example 2.17. *The simplest case to consider would be A in the space of real matrices of dimension $m \times n$ with basic l_2 inner product on both $\mathcal{R}^n$ and $\mathcal{R}^m$. It that case, $A^* = A^T$. What if the inner product was $\langle x, y\rangle = x^T M y$, for some $M > 0$? (show that $A^* = M^{-1} A^T M$).*

2.3 Orthogonal Projection

Let v_1, v_2, …, v_m be in H. Define H_m to be the span of these vectors (i.e., all possible linear combination of v_i's). For a vector $x \in H$, its orthogonal projection onto the span of v_i's is defined as the vector $\hat{x} \in H_m$ such that one of the following (equivalent) properties holds: (1) the error (i.e., $x - \hat{x}$) is orthogonal to H_m, and (2) among all possible vectors in H_m, $\hat{x}$ is the one that minimizes the norm of the error (i.e., $\|x - \hat{x}\|$ is minimized).

In the discussion above, orthogonality is defined by having the inner product to be zero and the definition of the norm used here is the square root of the inner product of a vector by itself (i.e., (2.1)). As a result, everything here depends on the specific definition of the inner product used.

Solution: Since $\hat{x} \in H_m$, it must be a combination of v_i's. Therefore

$$\hat{x} = a_1 v_1 + a_2 v_2 + \cdots + a_m v_m \tag{2.2}$$

for some set of a_i's. We are looking for the set of a_i's that result in $\hat{x}$ becoming the orthogonal projection. We will use the first property: For each v_i we must have

$$\langle v_i, x - \hat{x} \rangle = 0 \;\Rightarrow \langle v_i, \hat{x} \rangle = \langle v_i, x \rangle.$$

Using (2.2) for $\hat{x}$ on the left hand side, we get

$$\langle v_i, a_1 v_1 \rangle + \langle v_i, a_2 v_2 \rangle + \cdots \langle v_i, a_m v_m \rangle = \langle v_i, x \rangle. \tag{2.3}$$

As a result, we have m equations of the form in (2.3), for v_1, v_2, $\cdots v_m$. These m-linear equations can be solved for the unknowns, a_i's. Stacking these m equations on top of one another, we can form a matrix equation of the form

$$Aa = b$$

where A is a matrix whose (i,j) entry is $\langle v_i, v_j \rangle$, a is the vector of a_i's, and b is a m-vector whose j^{th} entry is $\langle v_j, x \rangle$.

If the vector x and basis v_i's are known, then the development above can be used to solve for a_i's and form the projection $\hat{x} = \sum a_i v_i$.

Remark 2.18. *The matrix A is often called the gramian matrix. It can be shown that it is nonsingular if and only if vectors v_i's are linearly independent. Also, note that we have not assumed that v_i's belong to $\mathcal{R}^n$ or similar spaces. Indeed, the set up here applies to a great many problems. (Think about Fourier series! The basis are orthogonal to one another - in L_2 inner product- so A will be diagonal! Can you finish this ?)*

Definition 2.19. *The relationship $\hat{x} = Px$ defines the projection operator P.*

Remark 2.20. *It can be proven that $P^2 = P$ for all projections (the so called idempotency property) and $P = P^*$ for all orthogonal projections (self adjoint property).*

When applicable, we can find the matrix representation of this operator. Consider the following example where x and all of the v_i's are in $\mathcal{R}^n$

Example 2.21. *For $x, v_i \in \mathcal{R}^n$, define $V = [v_1 \ v_2 \ldots v_m]$, which is a $n \times m$ matrix. Then, $A = V^T V$ and $b = V^T x$, if we use $\langle x, y \rangle = x^T y$ for inner product. Then we have,*

$$\hat{x} = Va = V(V^T V)^{-1} V^T x = Px \quad (i.e., \ P = V(V^T V)^{-1} V^T \ )$$

where P is the matrix representation of the projection operator (onto the span of v_i's).

Remark 2.22. *The previous exercise is the famous least squares solution. The next level up would be weighted least squares, which is done by modifying the inner product. For example, by using $\langle x_1, x_2 \rangle = x_1^T W x_2$.*

2.4 Rank, Range, Null Space, etc

Let us start with linear operator A (from $\mathcal{X}$ to $\mathcal{Y}$)

Definition 2.23. *Range of A is the linear space defined by*

$$R(A) = \{y : \exists\ x\ s.t.\ y = Ax\} = A(\mathcal{X}).$$

Definition 2.24. *The null space of A, or kernel of A, is the space defined by*

$$N(A) = \{x \in \mathcal{X}\ s.t.\ Ax = 0\}.$$

A very important property of these spaces is the following

$$R(A) \perp N(A^*) \quad , \quad R(A^*) \perp N(A) \tag{2.4}$$

where A^* is the adjoint consistent with the inner products used. So far, everything we have talked about applies to all linear operators. We can do more for matrices:

Definition 2.25. *For a given $m \times n$ matrix, we have*

$$R(A) = span\ of\ columns\ of\ A$$

$$\rho(A) = rank\ of\ A = dimension\ of\ R(A)$$

$$\rho(A) = min\{\#of\ indep\ rows\ of\ A\ , \#of\ indep\ columns\ of\ A\ \}$$

$$\nu(A) = nulity\ of\ A = dimension\ of\ N(A)$$

Remark 2.26. *An important property of matrices are that for any $A \in \mathcal{C}^{m\times n}$*

$$\rho(A) + \nu(A) = n$$

consequently,

$$\mathcal{R}^n = R(A^*) \overset{\perp}{\oplus} N(A)$$

$$\mathcal{R}^m = R(A) \overset{\perp}{\oplus} N(A^*)$$

2.5 Stability and Related Topics

Consider the dynamical system described by

$$\dot{x} = Ax \tag{2.5}$$

if A is time varying, eigenvalues tell you *next to nothing* about the stability of this system. Recall that the study of stability is, typically, the study of equilibrium points, which for linear systems boil down to study of the point $x = 0$. The most common way to study the stability of such systems is through the 'Lyapunov second - or direct- method'. Its most important results can be summarized by the following:

- If $\exists$ a continuously differentiable function $V(x)$ that has the following properties , (1) $V(x) > 0$ for all nonzero x, $V(x)$ is radially unbounded (i.e, as x gets large, so does V), and (3) $\dot{V} < 0$ for all nonzero x, then the system in (2.5) is globally asymptotically stable.
- If we could only guarantee $\dot{V} \leq 0$, then it is marginally stable, or stable in the sense of Lyapunov. If $\dot{V}$ can change sign, then the system is not stable.

Remark 2.27. *This is a sufficient condition only. Also, a great deal more can be said about this method. We leave the details and embellishments to the study of nonlinear systems.*

Remark 2.28. *In most applications, the Lyapunov function is chosen to be $V(x) = x^T Px$ for some $P > 0$. While we discuss the notion of positive definiteness later in some details, for now it suffices to know that $P > 0$ means that P is positive definite, which means that for* any *non-zero x, $x^T Px > 0$.*

If the matrix A in (2.5) is constant, a great deal of simplification can be made:

- If the real parts of every eigenvalue of A is strictly negative (i.e., it is on the open left half plane), then the system is globally asymptotically (and exponentially) stable. (i.e, starting with any x_o, $x(t)$ will go to zero as time goes by).
- If the real part of one or more eigenvalues is zero then the system is at best marginally stable (stability in the sense of Lyapunov). In particular, if zero real parts correspond to Jordan blocks of dimension 2 or more, then the system is *unstable*.

A few odds and ends should be reviewed. Recall that for constant A, the solution of (2.5) is $x(t) = e^{A(t-t_o)}x_o$ where x_o is the state at time t_o and $e^{A(t-t_o)}$ is the state transition matrix. Also, recall that if all eigenvalues of A have negative real parts, then as $t \to \infty$, $e^{A(t-t_o)} \to 0$. Indeed, its norm is bounded by $Me^{\alpha(t-t_o)}$, where M is a fixed constant and α is the real part of the least stable eigenvalue. As a result, integrals of the form

$$\lim_{t\to\infty} \int_0^t e^{At}\, dt$$

remain bounded, and converge, if and only if A is stable (i.e., real part λ is strictly negative).

2.5.1 Lyapunov Equation

Consider the following linear equation

$$PA + A^T P = -Q \tag{2.6}$$

where both P and Q are symmetric. From now on, by stable we mean all eigenvalues in the open left half plane. The basic results can be summarized

- A is stable iff for every $Q > 0$, (2.6) has a unique positive definite solution
- A is stable iff for every $Q \geq 0$ and (A, Q) observable, (2.6) has a unique positive definite solution

Proof: (Sketch)
$\Rightarrow$ $P = \int_0^\infty e^{A^T t} Q e^{At}\, dt$ is the solution to the Lyapunov equation. Check with Leibniz rule, show uniqueness of the solution, prove positive definiteness.
$\Leftarrow$ Use $V(x) = x^T P x$ and show $\dot{V} < 0$, where you may end up using LaSalle's lemma!

2.6 Controllability and Observability

In this subsection, we will focus on different definitions of observability for linear time invariant systems. Recall that controllability is dual of observability and all relevant results can be obtained by replacing A with A^T and C with B^T. We start with

$$\dot{x} = Ax + Bu \quad , \quad y = Cx \tag{2.7}$$

where A is a $n \times n$ matrix and B and C are matrices of appropriate dimension. The system in (2.7) is observable if any of the following equivalent conditions hold

1. Given $u(t)$ and $y(t)$ for $t \in [0, T]$, $x(0)$ can be determined.
2. For $u(t) \equiv 0$, $Ce^{At}x_o = 0$ for any interval implies the initial condition was zero ($x_o = 0$).
3. For $u(t) \equiv 0$, $W_o(t) = \int_0^t e^{A^T\tau}C^TCe^{A\tau}\,d\tau \; > 0$, for any $t > 0$
4. $\text{rank} \begin{pmatrix} C \\ CA \\ \cdot \\ \cdot \\ \cdot \\ CA^{n-1} \end{pmatrix} = n$
5. $\text{rank} \begin{pmatrix} \lambda I - A \\ C \end{pmatrix} = n$ for all complex λ
6. $Ax = \lambda x$ and $Cx = 0$, (together) imply $x = 0$

Definition 2.29. *The system in* (2.7) *is detectable if all of the unstable modes are observable; i.e.,* $Ax = \lambda x$ *and* $Cx = 0$ *imply either* $x = 0$ *or* $Real(\lambda) < 0$*. Similarly, the system in* (2.7) *is stabilizable if all of the unstable modes are controllable; i.e.,* $A^T x = \lambda x$ *and* $B^T x = 0$ *imply either* $x = 0$ *or* $Real(\lambda) < 0$ *. Some books define stabilizability as: The system is stabilizable if and only if there exists a matrix* K*, of appropriate dimension, such that* $A - BK$ *is stable (strictly). Similar definitions also hold for detectability.*

2.7 PROBLEM SET

Exercise 2.30. *Prove the famous Cauchy-Schwarz inequality:*

$$|\langle x, y\rangle| \leq \langle x, x\rangle^{\frac{1}{2}} \langle y, y\rangle^{\frac{1}{2}}$$

where equality holds only when one of the vectors is zero or the vectors are linearly dependent.

Exercise 2.31. *Confirm* (2.1) *is an appropriate norm.*

Exercise 2.32. *Confirm that the inner products in* example 2.7 *and* example 2.8 *are proper inner products. What is the corresponding norm in each case?*

Exercise 2.33. *Show that for any $x \in \mathcal{R}^n$, we have*

$$\|x\|_2 \leq \|x\|_1 \leq \sqrt{n}\|x\|_2$$

$$\|x\|_\infty \leq \|x\|_2 \leq \sqrt{n}\|x\|_\infty$$

$$\|x\|_\infty \leq \|x\|_1 \leq n\|x\|_\infty$$

Exercise 2.34. *Show that the operator norm satisfies the norm properties. Furthermore, show that $\|AB\|_i \leq \|A\|_i\|B\|_i$. This property is not true for non-induced norms of linear operators!*

Exercise 2.35. *In* example 2.15*, prove the expression given for induced 1-norm and ∞-norm are correct.*

Exercise 2.36. *Show that the gramian matrix encountered in orthogonal projection is nonsingular iff the basis vectors are linearly independent.*

Exercise 2.37. *In orthogonal projection of a vector in $\mathcal{R}^n$ onto the span of other vectors (also in $\mathcal{R}^n$), show that: (a) If $x \in H_n$ then $x = \hat{x}$, (b) $P = P^2$ and $P = P^*$ (i.e., is self-adjoint), and (c) $(I - P)$ is also an orthogonal projection.*

Exercise 2.38. *Show that the property* (2.4) *is true.*

Exercise 2.39. *Provide a compete proof of the Lyapunov equation, for the case of $Q > 0$. First, show that if the equation holds for $P > 0$ and $Q > 0$, then A must be stable. Next, show that if the equation holds, A is stable and $Q > 0$, then $P > 0$. (NOTE: it is not true that if A is stable and $P > 0$, the resulting Q is necessarily positive definite!).*

Exercise 2.40. *Write the controllability equivalence of properties 1-6 of the observability.*

Exercise 2.41. *In the observability properties, show 5 $\Leftrightarrow$ 6 and 2 $\Leftrightarrow$ 3.*

Exercise 2.42. *Given that the system in* (2.7) *is observable and given $y(t)$ and $u(t)$, how would you calculate $x(0)$? What is the dual problem in controllability?*

Exercise 2.43. *For a stable A, prove that the observability matrix $W_o(\infty)$ in property 3 satisfies $PA + A^T P = -C^T C$*