2. Solving Linear Equations

The prototypic finite dimensional problem we would like to solve is a system of linear equations:

$$Ax = y \tag{3}$$

where $A : \mathbf{R}^n \to \mathbf{R}^m$ is a linear transformation. In case $m = n$ there is a simple criterion for (3) to be solvable for any choice of y:

Theorem 2.1. *: If $A : \mathbf{R}^n \to \mathbf{R}^n$ is linear then (0.1) has a solution for every $y \in \mathbf{R}^n$ if and only if the only solution to*

$$Ax = 0$$

is $x = 0$. More geometrically: A is surjective if and only if it is injective.

Note that this condition for solvability implies that the solution to (3) is unique. The theorem is purely algebraic in character but we can also consider the dependence of the solution, x on the data, y. To that end we need to introduce a topology on $\mathbf{R}^n$. For the purposes of studying linear equations the natural topology on $\mathbf{R}^n$ is that defined by a norm. A norm is a function, $||\cdot|| : \mathbf{R}^n \to \mathbf{R}^n$ which satisfies the conditions:

$$||x|| \geq 0 \quad \forall x \in \mathbf{R}^n,\, ||x|| = 0 \,\text{iff}\, x = 0 \tag{4}$$

$$||x + y|| \leq ||x|| + ||y|| \qquad \text{(triangle inequality)}, \tag{5}$$

$$\forall x \in \mathbf{R}^n,\, \lambda \in \mathbf{R} \quad ||\lambda x|| = |\lambda|\, ||x||. \tag{6}$$

For example if $p \geq 1$ then

$$||x||_p = (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}$$

defines a norm. Note that

$$\lim_{p\to\infty} ||x||_p = \max\{|x_i| : \quad i = 1 \dots n\} = ||x||_\infty.$$

This also a norm

Exercise 2.1. : Show that if $1 \leq p,\, q \leq \infty$ then there are constants c, C such that

$$c||x||_q \leq ||x||_p \leq C||x||_q, \tag{7}$$

Briefly: $||\cdot||_q$ and $||\cdot||_p$ are equivalent norms.

A consequence of (7) is that all the norms, $\{||\cdot||_p\}$ define the same topology on $\mathbf{R}^n$; in fact all norms on $\mathbf{R}^n$ are equivalent and therefore define the same topology. Note in particular that the set $\{x : ||x|| \leq 1\}$ is compact. An especially useful norm is

$$||x||_2^2 = \sum_{i=1}^{n} x_1^2.$$

This is usually called the Euclidean norm. What distinguishes this norm is that it is defined by an inner product. An inner product is a mapping

$$\langle \cdot, \cdot \rangle : \quad \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$$

such that

$$\langle x, y\rangle = \langle y, x\rangle, \tag{8}$$
$$\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle, \tag{9}$$
$$\langle ax, y\rangle = a\langle x, y\rangle, \tag{10}$$
$$\langle x, x\rangle \geq 0 \text{ with equality only if } x = 0. \tag{11}$$

In the case at hand

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i.$$

Evidently

$$||x||_2^2 = \langle x, x\rangle.$$

A collection of vectors $\{e_1 \dots e_n\}$ is a basis for $\mathbf{R}^n$ iff every vector x has a unique representation as

$$x = \sum_{i=1}^{n} x_i e_i.$$

We say that a basis is orthonormal if

$$\langle e_i, e_j\rangle = \begin{cases} 1 & \text{for } i = j, \\ 0 & \text{for } i \neq j. \end{cases}$$

Exercise 2.2. Suppose we are given a basis, $\{f_1 \dots f_n\}$ for $\mathbf{R}^n$ show that there is an orthonormal basis, $\{e_1 \dots e_n\}$ such that for each $1 \leq j \leq n$

$$\{\sum_{i=1}^{j} x_i f_i : \quad x_1 \in \mathbf{R} \quad i = 1 \dots j\} =$$

$$\{\sum_{i=1}^{j} x_i e_i : \quad x_i \in \mathbf{R} \quad 2 = 1 \dots j\}.$$

Give an algorithm to construct $\{e_i\}$ from $\{f_i\}$.

Let $\{e_i \dots e_n\}$ be an orthonormal basis for $\mathbf{R}^n$. For each $x \in \mathbf{R}$ we can express $x = \sum_{i=1}^{n} x_i e_i$ and thus

$$Ax = \sum_{i=1}^{n} x_i A e_i.$$

By the triangle inequality

$$||Ax|| \leq \sum_{i=1}^{n} |x_1| \, ||Ae_i||$$

So if $M = \max\{||Ae_1||, \dots ||Ae_n||\}$ then

$$||Ax|| \leq M ||x||_1.$$

Since all norms on $\mathbf{R}^n$ are equivalent there is a constant C s.t. $||x||_1 \leq C||x||$ and therefore we've shown:

Lemma 2.1. *If $A : \mathbf{R}^n \to \mathbf{R}^n$ is a linear transformation and $||\cdot||$ is a norm on $\mathbf{R}^n$ then there is a constant C such that*

$$||Ax|| \leq C||x|| \qquad \forall x \in \mathbf{R}^n.$$

Corollary 2.1. *: If $A : \mathbf{R}^n \to \mathbf{R}^n$ is a linear transformation then $A : \mathbf{R}^n \to \mathbf{R}^n$ is continuous.*

Proof. Since A is linear $Ax - Ay = A(x - y)$, thus $||Ax - Ay|| = ||A(x - y)|| \leq C||x - y||$, for some constant C. □

Proposition 2.1. *: If $A : \mathbf{R}^n \to \mathbf{R}^n$ is a linear transformation then A is surjective iff there is a constant, $C > 0$ such that*

$$c||x|| \leq ||Ax||. \tag{12}$$

Proof. By theorem 1.1 A is surjective iff A is injective. Thus $||Ax|| \neq 0, \quad \forall x \neq 0$. Since $S = \{x : ||x|| = 1\}$ is compact and A is continuous the function $x \to ||Ax||$ assumes its minimum value at some point of $x_0 \in S_1$. As $||x_0|| = 1$ it is clear that $||Ax_0|| = c > 0$. For $X \in S_1$ we have

$$c||x|| \leq ||Ax||,$$

since $||\lambda x|| = |\lambda|\,||x||$ and $A\lambda x = \lambda Ax$ this shows that if A is surjective then (12) holds. If (12) holds then evidently $Ax = 0$ iff $x = 0$ and so by Theorem 1.1 A is surjective. □

Corollary 2.2. *: If $A : \mathbf{R}^n \to \mathbf{R}^n$ is a surjective linear transformation then*

$$||A^{-1}x|| \leq \frac{1}{c}||x|| \tag{13}$$

where c is the constant appearing in (13).

In summary we see that

$$\max_{\{||x||=1\}} ||Ax||$$

gives a quantitative measure of the continuity of A whereas

$$\min_{\{||x||=1\}} ||Ax||$$

gives a quantitative measure of the continuity of A^{-1}.

Now we consider the equation (3) when $n \neq m$. If for example $m > n$ then it seems quite unlikely that (3) could be solvable for arbitrary $y \in \mathbf{R}^m$. In order to obtain conditions on y for (3) to be solvable we need to consider the space of linear functions on $\mathbf{R}^n$. A map $\ell : \mathbf{R}^n \to \mathbf{R}$ is linear if

$$\ell(x + y) = \ell(x) + \ell(y), \tag{14}$$

$$\ell(ax) = a\ell(x).$$

The set of such linear maps clearly is itself a linear space which we denote by $(\mathbf{R}^n)'$. For example, if $y \in \mathbf{R}^n$ then

$$\ell_y(x) = \langle x, y \rangle$$

defines an element of $(\mathbf{R}^n)'$. In fact it is quite easy to prove

Proposition 2.2. *: The map $y \to \ell_y$ is an isomorphism of the linear spaces $\mathbf{R}^n$ and $(\mathbf{R}^n)'$.*

If $A : \mathbf{R}^n \to \mathbf{R}^m$ is a linear transformation then for each $y \in \mathbf{R}^m$ we can think of the map

$$x \to \langle Ax, y \rangle$$

as defining an element of $(\mathbf{R}^n)'$. Hence by Proposition 1.2 there is a unique vector $z_y \in \mathbf{R}^n$ such that

$$\langle Ax, y \rangle = \langle x, z_y \rangle.$$

Exercise 2.3. : Show that the map $y \to z_y$ is a linear transformation.

We call this linear transformation the transpose, dual or adjoint of A, it is denoted A^t:

$$\langle Ax, y \rangle = \langle x, A^t y \rangle.$$

For any linear transformation, $A : \mathbf{R}^n \to \mathbf{R}^m$ we define

image of $A = \operatorname{Im} A = \{Ax : x \in \mathbf{R}^n\}$,

kernel of $A = \ker A = \{x : Ax = 0\}$,

cokemel of $A = \operatorname{coker} A = \mathbf{R}^m / \operatorname{Im} A$.

Evidently if $y \in \operatorname{Im} A$ and $z \in \operatorname{Ker} A^t$ then

$$\langle y, z \rangle = \langle Ax, z \rangle = \langle x, A^t z \rangle = 0.$$

Thus we have a necessary condition for $Ax = y$ to be solvable. This turns out also to be a sufficient condition:

Theorem 2.2. *: The equation $Ax = y$ is solvable iff $\langle y, z \rangle = 0$ for all $z \in \ker A^t$.*

Exercise 2.4. : A map $B : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$ is called a bilinear form if

$$B(x + y, z) = B(x, z) + B(y, z)$$

$$B(x, y + z) = B(x, y) + B(x, z)$$

$$B(ax, y) = B(x, ay) = aB(x, y).$$

A bilinear form is non degenerate provided $B(x, y) = 0$ for all $x \in \mathbf{R}^n$ iff $y = 0$.

(a) Show that $y \to B(\cdot, y)$ defines an isomorphism, $\mathbf{R}^n \to (\mathbf{R}^n)'$
(b) If B_1 and B_2 are nondegenerate bilinear on $\mathbf{R}^n$ and $\mathbf{R}^m$, respectively forms show that there is a uniquely defined linear transformation A^t such that $B_2(Ax, y) = B_1(x, A^t y)$.
(c) Show that $Ax = y$ is solvable iff $B_2(y, z) = 0$ for every $z \in \operatorname{Ker} A^t$.

Exercise 2.5. : If $A : \mathbf{R}^n \to \mathbf{R}^m$ then show that:

$$\dim \operatorname{Im} A = n - \dim \operatorname{Ker} A,$$

$$\dim \operatorname{Im} A = m - \dim \operatorname{Ker} A'$$

therefore:

$$\dim \operatorname{Ker} A - \dim \operatorname{Ker} A' = n - m.$$

Exercise 2.6. : If $S \leq \mathbf{R}^n$ is a subspace and $||\cdot||$ is a norm on $\mathbf{R}^n$ then we can define a function $N([x])$ on the quotient vector space $\mathbf{R}^n/S$ by setting:

$$N([x_0]) = \inf_{x \in [x_0]} ||x||.$$

Prove that $N(\cdot)$ is a norm on $\mathbf{R}^n/S$

Exercise 2.7. : Let $A : \mathbf{R}^n \to \mathbf{R}^m$ be a linear map and suppose that $||\cdot||$ denotes a norm on $\mathbf{R}^n$ or $\mathbf{R}^m$. Show that there exists a constant C such for $y \in \operatorname{Im} A$ there is an $x \in \mathbf{R}^n$ with $Ax = y$ and $||x|| \leq C||y||$.

Good references for this material are

1. *Linear Algebra* by Peter D. Lax
2. *Intro. to Matrix Analysis* by Richard Bellman
3. *Calculus*, vol II by Tom M. Apostol.

3. Basic Functional Analysis

In finite dimensions the problem of solving linear equations is purely algebraic. That is: there is no necessity to introduce a topology to give necessary and sufficient conditions for the solvability of $Ax = y$. In infinite dimensions there is a simi lar analysis but it requires a topology on the domain and range of the linear map. In finite dimensions there is a unique norm topology, this is closely related to the fact that the unit sphere, with respect to any norm, is compact. To compare two norms , $||\cdot||_1, \quad ||\cdot||_2$ we simply compute

$$c_1 = \inf_{\{||x||_1=1\}} ||x||_2 \quad \text{and} \quad c_2 = \sup_{\{||x||_1=1\}} ||x||_2$$

Then

$$c_1||x||_1 \leq ||x||_2 \leq c_2||x||_1.$$

In infinite dimensions the unit sphere is never compact and there are many different normed linear spaces. Recall that in the analysis of $Ax = y$ the dual space $(\mathbf{R}^m)^*$ played an important role. This feature becomes even more pronounced in infinite dimensions.

Let's briefly consider normed linear spaces in general. Let X be a vector space. We need to specify an underlying field, the field of scalars. It will usually be $\mathbf{C}$ but occasionally we use $\mathbf{R}$. A norm is a map:

$$||\cdot|| : x \to \mathbf{R} \qquad \text{such that}$$