SECTION 1

Affine Sets

Throughout this book, R denotes the real number system, and R^n is the usual vector space of real n-tuples $x = (\xi_1, \ldots, \xi_n)$. Everything takes place in R^n unless otherwise specified. The inner product of two vectors x and x^* in R^n is expressed by

$$\langle x, x^* \rangle = \xi_1 \xi_1^* + \cdots + \xi_n \xi_n^*$$

The same symbol A is used to denote an $m \times n$ real matrix A and the corresponding linear transformation $x \to Ax$ from R^n to R^m. The transpose matrix and the corresponding adjoint linear transformation from R^m to R^n are denoted by A^*, so that one has the identity

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

(In a symbol denoting a vector, * has no operational significance; all vectors are to be regarded as column vectors for purposes of matrix multiplication. Vector symbols involving * are used from time to time merely to bring out the familiar duality between vectors considered as points and vectors considered as the coefficient n-tuples of linear functions.) The end of a proof is signalled by ||.

If x and y are different points in R^n, the set of points of the form

$$(1 - \lambda)x + \lambda y = x + \lambda(y - x), \qquad \lambda \in R,$$

is called the *line through x and y*. A subset M of R^n is called an *affine set* if $(1 - \lambda)x + \lambda y \in M$ for every $x \in M$, $y \in M$ and $\lambda \in R$. (Synonyms for "affine set" used by other authors are "affine manifold," "affine variety," "linear variety" or "flat.")

The empty set $\emptyset$ and the space R^n itself are extreme examples of affine sets. Also covered by the definition is the case where M consists of a solitary point. In general, an affine set has to contain, along with any two different points, the entire line through those points. The intuitive picture is that of an endless uncurved structure, like a line or a plane in space.

The formal geometry of affine sets may be developed from the theorems of linear algebra about subspaces of R^n. The exact correspondence between affine sets and subspaces is described in the two theorems which follow.

THEOREM 1.1. *The subspaces of R^n are the affine sets which contain the origin.*

PROOF. Every subspace contains 0 and, being closed under addition and scalar multiplication, is in particular an affine set.

Conversely, suppose M is an affine set containing 0. For any $x \in M$ and $\lambda \in R$, we have

$$\lambda x = (1 - \lambda)0 + \lambda x \in M,$$

so M is closed under scalar multiplication. Now, if $x \in M$ and $y \in M$, we have

$$\tfrac{1}{2}(x + y) = \tfrac{1}{2}x + (1 - \tfrac{1}{2})y \in M,$$

and hence

$$x + y = 2(\tfrac{1}{2}(x + y)) \in M.$$

Thus M is also closed under addition and is a subspace. ||

For $M \subset R^n$ and $a \in R^n$, the *translate* of M by a is defined to be the set

$$M + a = \{x + a \mid x \in M\}.$$

A translate of an affine set is another affine set, as is easily verified.

An affine set M is said to be *parallel* to an affine set L if $M = L + a$ for some a. Evidently "M is parallel to L" is an equivalence relation on the collection of affine subsets of R^n. Note that this definition of parallelism is more restrictive than the everyday one, in that it does not include the idea of a line being parallel to a plane. One has to speak of a line which is parallel to another line within a given plane, and so forth.

THEOREM 1.2. *Each non-empty affine set M is parallel to a unique subspace L. This L is given by*

$$L = M - M = \{x - y \mid x \in M, y \in M\}.$$

PROOF. Let us show first that M cannot be parallel to two different subspaces. Subspaces L_1 and L_2 parallel to M would be parallel to each other, so that $L_2 = L_1 + a$ for some a. Since $0 \in L_2$, we would then have $-a \in L_1$, and hence $a \in L_1$. But then $L_1 \supset L_1 + a = L_2$. By a similar argument $L_2 \supset L_1$, so $L_1 = L_2$. This establishes the uniqueness. Now observe that, for any $y \in M$, $M - y = M + (-y)$ is a translate of M containing 0. By Theorem 1.1 and what we have just proved, this affine set must be the unique subspace L parallel to M. Since $L = M - y$ no matter which $y \in M$ is chosen, we actually have $L = M - M$. ||

The *dimension* of a non-empty affine set is defined as the dimension of the subspace parallel to it. (The dimension of $\emptyset$ is -1 by convention.) Naturally, affine sets of dimension 0, 1 and 2 are called *points*, *lines* and *planes*, respectively. An $(n - 1)$-dimensional affine set in R^n is called a *hyperplane*. Hyperplanes are very important, because they play a role dual to the role of points in n-dimensional geometry.

Hyperplanes and other affine sets may be represented by linear functions and linear equations. It is easy to deduce this from the theory of orthogonality in R^n. Recall that, by definition, $x \perp y$ means $\langle x, y\rangle = 0$. Given a subspace L of R^n, the set of vectors x such that $x \perp L$, i.e. $x \perp y$ for every $y \in L$, is called the *orthogonal complement* of L, denoted $L^\perp$. It is another subspace, of course, and

$$\dim L + \dim L^\perp = n.$$

The orthogonal complement $(L^\perp)^\perp$ of $L^\perp$ is in turn L. If $b_1, \ldots, b_m$ is a basis for L, then $x \perp L$ is equivalent to the condition that $x \perp b_1, \ldots, x \perp b_m$. In particular, the $(n-1)$-dimensional subspaces of R^n are the orthogonal complements of the one-dimensional subspaces, which are the subspaces L having a basis consisting of a single non-zero vector b (unique up to a non-zero scalar multiple). Thus the $(n-1)$-dimensional subspaces are the sets of the form $\{x \mid x \perp b\}$, where $b \neq 0$. The hyperplanes are the translates of these. But

$$\begin{aligned}\{x \mid x \perp b\} + a &= \{x + a \mid \langle x, b\rangle = 0\}\\ &= \{y \mid \langle y - a, b\rangle = 0\} = \{y \mid \langle y, b\rangle = \beta\},\end{aligned}$$

where $\beta = \langle a, b\rangle$. This leads to the following characterization of hyperplanes.

THEOREM 1.3. *Given $\beta \in R$ and a non-zero $b \in R^n$, the set*

$$H = \{x \mid \langle x, b\rangle = \beta\}$$

is a hyperplane in R^n. Moreover, every hyperplane may be represented in this way, with b and β unique up to a common non-zero multiple.

In Theorem 1.3, the vector b is called a *normal* to the hyperplane H. Every other normal to H is either a positive or a negative scalar multiple of b. A good interpretation of this is that every hyperplane has "two sides," like one's picture of a line in R^2 or a plane in R^3. Note that a plane in R^4 would *not* have "two sides," any more than a line in R^3 has.

The next theorem characterizes the affine subsets of R^n as the solution sets to systems of simultaneous linear equations in n variables.

THEOREM 1.4. *Given $b \in R^m$ and an $m \times n$ real matrix B, the set*

$$M = \{x \in R^n \mid Bx = b\}$$

is an affine set in R^n. Moreover, every affine set may be represented in this way.

PROOF. If $x \in M$, $y \in M$ and $\lambda \in R$, then for $z = (1 - \lambda)x + \lambda y$ one has

$$Bz = (1 - \lambda)Bx + \lambda By = (1 - \lambda)b + \lambda b = b,$$

so $z \in M$. Thus the given M is affine.

On the other hand, starting with an arbitrary non-empty affine set M other than R^n itself, let L be the subspace parallel to M. Let $b_1, \ldots, b_m$ be a basis for $L^\perp$. Then

$$L = (L^\perp)^\perp = \{x \mid x \perp b_1, \ldots, x \perp b_m\}$$
$$= \{x \mid \langle x, b_i \rangle = 0, \quad i = 1, \ldots, m\} = \{x \mid Bx = 0\},$$

where B is the $m \times n$ matrix whose rows are $b_1, \ldots, b_m$. Since M is parallel to L, there exists an $a \in R^n$ such that

$$M = L + a = \{x \mid B(x - a) = 0\} = \{x \mid Bx = b\},$$

where $b = Ba$. (The affine sets R^n and $\emptyset$ can be represented in the form in the theorem by taking B to be the $m \times n$ zero matrix, say, with $b = 0$ in the case of R^n and $b \neq 0$ in the case of $\emptyset$.) ||

Observe that in Theorem 1.4 one has

$$M = \{x \mid \langle x, b_i \rangle = \beta_i,\ i = 1, \ldots, m\} = \textstyle\bigcap_{i=1}^m H_i,$$

where b_i is the ith row of B, β_i is the ith component of b, and

$$H_i = \{x \mid \langle x, b_i \rangle = \beta_i\}.$$

Each H_i is a hyperplane ($b_i \neq 0$), or the empty set ($b_i = 0$, $\beta_i \neq 0$), or R^n ($b_i = 0$, $\beta_i = 0$). The empty set may itself be regarded as the intersection of two different parallel hyperplanes, while R^n may be regarded as the intersection of the empty collection of hyperplanes of R^n. Thus:

COROLLARY 1.4.1. *Every affine subset of R^n is an intersection of a finite collection of hyperplanes.*

The affine set M in Theorem 1.4 can be expressed in terms of the vectors $b'_1, \ldots, b'_n$ which form the columns of B by

$$M = \{x = (\xi_1, \ldots, \xi_n) \mid \xi_1 b'_1 + \cdots + \xi_n b'_n = b\}.$$

Obviously, the intersection of an arbitrary collection of affine sets is again affine. Therefore, given any $S \subset R^n$ there exists a unique smallest affine set containing S (namely, the intersection of the collection of affine sets M such that $M \supset S$). This set is called the *affine hull* of S and is denoted by aff S. It can be proved, as an exercise, that aff S consists of all the vectors of the form $\lambda_1 x_1 + \cdots + \lambda_m x_m$, such that $x_i \in S$ and $\lambda_1 + \cdots + \lambda_m = 1$.

A set of $m + 1$ points $b_0, b_1, \ldots, b_m$ is said to be *affinely independent* if aff $\{b_0, b_1, \ldots, b_m\}$ is m-dimensional. Of course

$$\text{aff}\,\{b_0, b_1, \ldots, b_m\} = L + b_0,$$

where

$$L = \text{aff}\,\{0, b_1 - b_0, \ldots, b_m - b_0\}.$$

By Theorem 1.1, L is the same as the smallest subspace containing $b_1 - b_0, \ldots, b_m - b_0$. Its dimension is m if and only if these vectors are linearly independent. Thus $b_0, b_1, \ldots, b_m$ are affinely independent if and only if $b_1 - b_0, \ldots, b_m - b_0$ are linearly independent.

All the facts about linear independence can be applied to affine independence in the obvious way. For instance, any affinely independent set of $m + 1$ points in R^n can be enlarged to an affinely independent set of $n + 1$ points. An m-dimensional affine set M can be expressed as the affine hull of $m + 1$ points (translate the points which correspond to a basis of the subspace parallel to M).

Note that, if $M = \text{aff}\,\{b_0, b_1, \ldots, b_m\}$, the vectors in the subspace L parallel to M are the linear combinations of $b_1 - b_0, \ldots, b_m - b_0$. The vectors in M are therefore those expressible in the form

$$x = \lambda_1(b_1 - b_0) + \cdots + \lambda_m(b_m - b_0) + b_0,$$

i.e. in the form

$$x = \lambda_0 b_0 + \lambda_1 b_1 + \cdots + \lambda_m b_m, \qquad \lambda_0 + \lambda_1 + \cdots + \lambda_m = 1.$$

The coefficients in such an expression of x are unique if and only if $b_0, b_1, \ldots, b_m$ are affinely independent. In that event, $\lambda_0, \lambda_1, \ldots, \lambda_m$, as parameters, define what is called a *barycentric coordinate system* for M.

A single-valued mapping $T: x \to Tx$ from R^n to R^m is called an *affine transformation* if

$$T((1 - \lambda)x + \lambda y) = (1 - \lambda)Tx + \lambda Ty$$

for every x and y in R^n and $\lambda \in R$.

THEOREM 1.5. *The affine transformations from R^n to R^m are the mappings T of the form $Tx = Ax + a$, where A is a linear transformation and $a \in R^m$.*

PROOF. If T is affine, let $a = T0$ and $Ax = Tx - a$. Then A is an affine transformation with $A0 = 0$. A simple argument resembling the one in Theorem 1.1 shows that A is actually linear.

Conversely, if $Tx = Ax + a$ where A is linear, one has

$$T((1 - \lambda)x + \lambda y) = (1 - \lambda)Ax + \lambda Ay + a = (1 - \lambda)Tx + \lambda Ty.$$

Thus T is affine. ||

The inverse of an affine transformation, if it exists, is affine.