

1 Flat ...

Sometimes in statistical applications it is useful to consider a linear subspace that is shifted or translated from the origin. This will happen, for example, in models that include an intercept. It is therefore helpful to have the following definition of a space that is displaced from the origin.

Definition 1 (Flat) Suppose $M \subset V$ is a linear subspace and $y_0 \in V$. Then a flat consists of $\{x + y_0 | x \in M\}$. We will write $y_0 + M$ where M is a subspace to indicate a flat.

By considering *translations*, flats are equivalent to vector spaces. If Y is a random variable whose domain is the flat $y_0 + M$, then, if y_0 is fixed, $Y - y_0$ has domain M .

Example. Set $S_4 = \{(1, 1, 1)^T + z, z \in S_2\}$ is a flat because $0 \notin S_4$.

Example. In Ce^2 , consider

$$M = \left\{ \alpha \begin{pmatrix} 1 \\ 2 \end{pmatrix} \mid \alpha \in Ce \right\} \text{ and } y_0 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

Then the flat $y_0 + M$ is given by the set

$$y_0 + M = \left\{ \begin{pmatrix} 2 \\ 2 \end{pmatrix} + \alpha \begin{pmatrix} 1 \\ 2 \end{pmatrix} \mid \alpha \in Ce \right\}$$

which is just a straight line that does not pass through the origin, but rather through the point $(2, 2)$. The choice of y_0 is not unique and it can be any point $y = y_0 + y_\alpha$, where $y_\alpha = \alpha(1, 2)^T$. For example, if $\alpha = -2$, then $y = (0, -2)^T$ and if $\alpha = +1$, then $y = (3, 4)^T$, and so on. For any y_0 not of this form, we simply get a different flat. This is summarized in the next remark.

Theorem 1 *The two spaces*

$$\begin{aligned} F_1 &= \{z \mid z = y_0 + x, y_0 \in V, x \in M \subset V\} \\ F_2 &= \{z \mid z = y_1 + x, y_1 \in F_1, x \in M \subset V\} \end{aligned}$$

are the same subspace, so the representation of the flat is not unique.

Definition 2 (Sum and intersection of subspaces) Let H, K be two linear subspaces. Then:

$$H + K = \{x + y \mid x \in H, y \in K\}$$

is the sum of H and K . The intersection of H and K is

$$H \cap K = \{x \mid x \in H \text{ and } x \in K\}$$

Theorem 2 *Both $H + K$ and $H \cap K$ are linear subspaces.*

Proof: Exercise.

Definition 3 (Disjoint subspaces) *Two subspaces are disjoint if $H \cap K = \{0\}$, the null vector.*

Theorem 3 *If $H \cap K = \{0\}$, and $z \in H + K$, then the decomposition $z = x + y$ with $x \in H$ and $y \in K$ is unique.*

Proof: Suppose $z = x + y$ and $z = x' + y'$. Then, $x - x' \in H$ and $y - y' \in K$. We must have $x + y = x' + y'$ or $x - x' = y - y'$, which in turn requires that $x - x' = y - y' = 0$, since 0 is the only vector common to H and K . Thus, $x = x'$ and $y = y'$. QED

Theorem 4 *If $H \cap K = \{0\}$, then $\dim(H + K) = \dim(H) + \dim(K)$. In general, $\dim(H + K) = \dim(H) + \dim(K) - \dim(H \cap K)$.*

Proof: Exercise.

Definition 4 (Complement of a space) *If M and M^c are disjoint subspaces of V and $V = M + M^c$, then M^c is called a complement of M .*

Remark 1 *The complement is not unique. In \mathbb{R}^2 , a subspace M of dimension 1 consists of a line through the origin. A complement of M is given by any other line $M^c \neq \alpha M$ through the origin, because linear combinations of any two such lines span \mathbb{R}^2 .*

In the linear model $Y = X\beta + \varepsilon$, we have that $\mu = E(Y) = X\beta$, so that $\mu \in \mathcal{C}(X)$. To estimate μ with $\hat{\mu}$, we might want to require that $\hat{\mu} \in \mathcal{C}(X)$ (note: if X includes a constant, then $\mathcal{C}(X)$ is a flat; otherwise, it is a subspace). The estimate would then depend upon Y in a sensible way by “moving” Y to the subspace. The method of moving is via projections. The optimality of moves depends on the way we measure distance — on an inner product defined on the vector space.

2 Solutions to systems of linear equations

Consider the matrix equation $X_{n \times p} \beta_{p \times 1} = y_{n \times 1}$. For a given X and Y does there exist a solution β to these equations? Is it unique? If not unique, can we characterize all possible solutions?

1. If $n = p$ and X is nonsingular, the unique solution is $\beta = X^{-1}y$.
2. If $y \in \mathcal{C}(X)$, y can be expressed as a linear combination of the columns of X . If X is of full column rank, then the columns of X form a basis for $\mathcal{C}(X)$, and the solution β is just the coordinates of y relative to this basis. For any g-inverse X^- , we have $XX^-y = y$ for all $y \in \mathcal{C}(X)$, and so a solution is given by

$$\beta = X^-y \quad (1)$$

If $\rho(X) = \text{rank}(\mathcal{C}(X)) < p$, then the solution is not unique. If β_0 is *any* solution, for example the solution given by (1), then so is $\beta_0 + z$, $z \in N(X) = \text{Null-space of}(X)$. The set of solutions is given by $\beta_0 + N(X)$, which is a flat.

3. If $y \notin \mathcal{C}(X)$, then there is no exact solution. This is the usual situation in linear models, and leads to the estimation problem discussed in the next chapter.

What we might do is get the “closest” solution by replacing Y by another vector \hat{Y} that is as close to Y as possible; if we define close as $\|Y - \hat{Y}\|^2$ making small, we need to solve $X\beta = P_{\mathcal{C}(X)}Y$ instead of the original equation. If X has full column rank, this leads to the familiar solution:

$$\begin{aligned} \beta_0 &= X^+Py \\ &= (X'X)^{-1}X'X(X'X)^{-1}X'Y = (X'X)^{-1}X'Y \end{aligned} \quad (2)$$

which is unique. If X does not have full column rank, then the set of solutions again forms a flat of the form $\beta_0 + N(X)$ with β_0 given by (2).