Linear Maps and Matrices (Lab 2)

BST 235: Advanced Regression and Statistical Learning

Alex Levis, Fall 2019

1 Brief review and warmup

So far, we have defined and talked about vector spaces, which comprise the *objects* of interest for much of this course: data vectors viewed as fixed quantities

- $\mathbf{X}_i = [X_{i1} \cdots X_{id}]^T \in \mathbb{R}^d$, covariates for i-th individual,
- $\mathbf{X}^{(j)} = [X_{1j} \cdots X_{nj}]^T \in \mathbb{R}^n$, data on j-th covariate,
- $\mathbf{Y} = [Y_1 \cdots Y_n]^T \in \mathbb{R}^n$, the outcome vector,

or as collections of random variables, $X_{ij}, Y_i \in L_2(P)$. The covariate data can be combined succinctly into the *design matrix*, $\mathbb{X} \in \mathbb{R}^{n \times d}$, given by

$$\mathbb{X} = \begin{bmatrix} \mathbf{X}_1^T \\ \vdots \\ \mathbf{X}_n^T \end{bmatrix} = \begin{bmatrix} \mathbf{X}^{(1)} & \cdots & \mathbf{X}^{(d)} \end{bmatrix}$$

Considering vector spaces abstractly, we defined the concepts of linear independence, span, basis, and dimension.

Exercise 1. Let V be a vector space, and $\{v_1, \ldots, v_k\} \subseteq V$ a collection of vectors. Show that $\mathscr{L}(v_1, \ldots, v_k)$ is a subspace of V.

First, $0_V \in \mathscr{L}(v_1, \ldots, v_k)$, since $0_V = 0v_1 + \cdots + 0v_k$. Next, if $\gamma_1, \gamma_2 \in \mathbb{F}$, $u_1 = \sum_{j=1}^k \alpha_j v_j$, $u_2 = \sum_{j=1}^k \beta_j v_j$, for constants $\alpha_j, \beta_j \in \mathbb{F}$, then $\gamma_1 u_1 + \gamma_2 u_2 = \sum_{j=1}^k (\gamma_1 \alpha_j + \gamma_2 \beta_j) v_j \in \mathscr{L}(v_1, \ldots, v_k)$.

We have seen that finite-dimensional vector spaces — essentially by construction — have finite bases which form a coordinate system for the vector space.

Exercise 2. Let V be a vector space, and suppose $\{v_1, \ldots, v_k\} \subseteq V$ is a basis for V. Show that every vector $v \in V$ can be expressed uniquely as a linear combination of elements in $\{v_1, \ldots, v_k\}$.

Let $v \in V$ be arbitrary. Since $\{v_1, \ldots, v_k\}$ is a basis, it must span V, so $\exists\, \alpha_1, \ldots, \alpha_k \in \mathbb{F}$ such that $v = \sum_{j=1}^k \alpha_j v_j$. Suppose some other set of constants $\beta_1, \ldots, \beta_k \in \mathbb{F}$ satisfy $v = \sum_{j=1}^k \beta_j v_j$. Then clearly $\sum_{j=1}^k \alpha_j v_j = \sum_{j=1}^k \beta_j v_j \implies \sum_{j=1}^k (\alpha_j - \beta_j) v_j = 0_V$. By linear independence, we have $\alpha_1 - \beta_1 = \cdots = \alpha_k - \beta_k = 0$, so $\alpha_j = \beta_j$, for $j = 1, \ldots, k$.

With the objects of interest defined, we next study *operations* that can be applied to these vectors. In this lab, we introduce linear maps, which are operations between vector spaces with special properties. We then show that matrices, and especially matrix-vector multiplication, tell the whole story for linear maps between finite-dimensional vector spaces. In lecture, we are beginning to study inner products, orthogonality, and projections. It turns out, as we will see, that projection is linear, and as such we will be able to talk about *projection matrices*.

2 Linear maps

As before, we will develop the theory abstractly, and eventually apply our general results in particular vector spaces.

Definition 1. Let $(V, \oplus_V, \odot_V)$, $(W, \oplus_W, \odot_W)$ be two vector spaces over the field $\mathbb{F}$. Let $T : V \to W$ define a function between V and W. We say that T is *linear* if for any $a \in \mathbb{F}$, $v_1, v_2 \in V$,

(i) $T(a \odot_V v_1) = a \odot_W T(v_1)$, and

(ii) $T(v_1 \oplus_V v_2) = T(v_1) \oplus_W T(v_2)$.

Equivalently, T must satisfy $T((a_1 \odot_V v_1) \oplus_V (a_2 \odot_V v_2)) = (a_1 \odot_W T(v_1)) \oplus_W (a_2 \odot_W T(v_2))$, for any choice of $a_1, a_2 \in \mathbb{F}$, $v_1, v_2 \in V$. When the operations are made less explicit, it becomes clear that a linear map is one that plays well with linear combinations, i.e., $T(a_1v_1 + a_2v_2) = a_1T(v_1) + a_2T(v_2)$. We will denote the set of all linear maps between V and W as $\mathcal{L}(V, W)$.

Remark 1. A straightforward consequence of the definition is that $\mathcal{L}(V, W)$ is itself a vector space. There is additional structure as well: one may compose linear maps together to obtain another linear map. That is, if $T \in \mathcal{L}(V, W)$, $S \in \mathcal{L}(W, U)$, then $S \circ T \in \mathcal{L}(V, U)$, whenever V, W, U are vector spaces, where

$$(S \circ T)(v) \coloneqq S(T(v)), \text{ for any } v \in V.$$

For given vector space V, a special linear map to consider is the identity map, $I_V : V \to V$, defined as $I_V(v) \coloneqq v$, for all $v \in V$. The identity maps have the special property that $T \circ I_V \equiv T$, and $I_W \circ T \equiv T$, for any $T \in \mathcal{L}(V, W)$.

Remark 2. A particular class of linear maps is $\mathcal{L}(V, \mathbb{F})$, which are called *linear functionals* on V. For the mathematically inclined, this forms the so-called algebraic dual space for V. In many cases, we further restrict to the class of *continuous linear maps* between V and W, denoted $\mathcal{CL}(V, W)$. It turns out that when V is finite-dimensional, there is no distinction between these two classes, i.e., $\mathcal{CL}(V, W) = \mathcal{L}(V, W)$. In general vector spaces, the class $\mathcal{CL}(V, \mathbb{F})$ is known as the *topological dual space*, which often plays a very important theoretical role — look up Hilbert spaces and the Riesz representation theorem if you are interested!

Each linear map $T \in \mathcal{L}(V, W)$ is associated with two important subspaces, the *image* or *range*,

$$\text{Im}(T) = \{w \in W \mid \exists v \in V \text{ such that } T(v) = w\} = \{T(v) \mid v \in V\},$$

which is a subspace (**show this**) of W, and the *kernel* or *nullspace*,

$$\text{Ker}(T) = \{v \in V \mid T(v) = 0\},$$

which is a subspace (**show this**) of V. We can think of the range of T as related to the *expressiveness* of T (cf. spanning), and we can think of the nullspace as related to the *redundancy* of T (cf. linear dependence) — check that if $T(v) = w$, and $v_0 \in \text{Ker}(T)$, then $T(v + v_0) = w$.

Before tackling the next exercise, recall the lemma shown in class that demonstrated that dimension is well-defined: if $\{v_1, \dots, v_k\}, \{w_1, \dots, w_s\} \subseteq V$ are two bases of vector space V, then $k = s$. By inspecting the argument closely, we actually proved a slightly more general result: if $\{v_1, \dots, v_k\}$ are linearly independent and $\{w_1, \dots, w_s\}$ span V, then $k \le s$ and $\{v_1, \dots, v_k, w_{j_1}, \dots, w_{j_{s-k}}\}$ spans V; note that we might have to reorder the w's. This fact allows us to extend linearly dependent sets to bases, by adding more vectors.

Exercise 3. Suppose V, W are vector spaces, where V is finite-dimensional, and let $T \in \mathcal{L}(V, W)$. Show that

$$\dim(V) = \underbrace{\dim(\operatorname{Im}(T))}_{\text{"rank"}} + \underbrace{\dim(\operatorname{Ker}(T))}_{\text{"nullity"}}.$$

This famous result is known as the *rank-nullity* theorem.

Let $k = \dim(\operatorname{Ker}(T))$, $s = \dim(V)$, and $\mathcal{B} = \{u_1, \dots, u_k\} \subseteq \operatorname{Ker}(T)$ be a basis for the kernel of T. It is sufficient to find a basis for $\operatorname{Im}(T)$ that has $s - k$ elements. If $k = 0$, then you can check that Exercise 4(i) below implies $s = \dim(\operatorname{Im}(T))$. Otherwise, taking $\mathcal{B}$ as a linearly independent set in V, and an arbitrary basis for V as a spanning set for V, the above comment tells us that there are some vectors $\mathcal{D} = \{z_1, \dots, z_{s-k}\} \subseteq V$ such that $\mathcal{B} \cup \mathcal{D}$ is a basis for V. We finish by showing that $\{T(z_1), \dots, T(z_{s-k})\}$ is a basis for $\operatorname{Im}(T)$. To show that this set spans $\operatorname{Im}(T)$, note that an arbitrary element of this space is $T(\sum_{j=1}^{k} \alpha_j u_j + \sum_{\ell=1}^{s-k} \beta_\ell z_\ell) = \sum_{j=1}^{k} \alpha_j T(u_j) + \sum_{\ell=1}^{s-k} \beta_\ell T(z_\ell) = \sum_{\ell=1}^{s-k} \beta_\ell T(z_\ell)$, for some constants $\alpha_1, \dots, \alpha_k, \beta_1, \dots, \beta_{s-k} \in \mathbb{F}$, since $u_j \in \operatorname{Ker}(T)$ for all j — note also that if $k = s$, so that $\mathcal{D} = \emptyset$, this shows that $\operatorname{Im}(T) = \{0_W\}$, so the result holds. For linear independence, suppose $0_W = \sum_{\ell=1}^{s-k} \alpha_\ell T(z_\ell) = T(\sum_{\ell=1}^{s-k} \alpha_\ell z_\ell)$. Then $\sum_{\ell=1}^{s-k} \alpha_\ell z_\ell \in \operatorname{Ker}(T)$, so $\exists\, \gamma_1, \dots, \gamma_k \in \mathbb{F}$ such that $\sum_{\ell=1}^{s-k} \alpha_\ell z_\ell = \sum_{j=1}^{k} \gamma_j u_j$. By linear independence of $\mathcal{B} \cup \mathcal{D}$, this implies $\alpha_1 = \dots = \alpha_{s-k} = 0$, so $\{T(z_1), \dots, T(z_{s-k})\}$ are linearly independent.

Before moving on to the connection between linear maps and matrices, we discuss *invertibility* of linear maps. A linear map $T \in \mathcal{L}(V, W)$ is invertible if $\exists T^{-1} : W \to V$ such that $T^{-1} \circ T \equiv I_V$ and $T \circ T^{-1} \equiv I_W$ — inverses are unique, so we can call T^{-1} *the* inverse of T. Any function is invertible if and only if it is bijective, i.e., injective (or one-to-one) and surjective (or onto). For linear maps, this translates to

- Injective: $T(v_1) = T(v_2) \implies v_1 = v_2$, for any $v_1, v_2 \in V$. Equivalently, $\operatorname{Ker}(T) = \{0\}$.
- Surjective: $\forall w \in W, \exists\, v \in V : T(v) = w$. Equivalently, $\operatorname{Im}(T) = W$.

Exercise 4. Suppose that $T \in \mathcal{L}(V, W)$ is an invertible linear map, and V is finite-dimensional. Show that

(i) $\dim(V) = \dim(W)$.

Let $\mathcal{B}_V = \{v_1, \dots, v_k\}$ be a basis for V, and we show that $\mathcal{B}_W = \{T(v_1), \dots, T(v_k)\}$ is a basis for W. It is clear that $\mathcal{B}_W$ spans $W = \operatorname{Im}(T)$, since $T(v)$ is a linear combination of $T(v_j)$, $j = 1, \dots, k$, for any $v \in V$. For linear independence, suppose $\sum_{j=1}^{k} \alpha_j T(v_j) = 0_W$. Then $T(\sum_{j=1}^{k} \alpha_j v_j) = 0_W = T(0_V)$, so $\sum_{j=1}^{k} \alpha_j v_j = 0_V$ by injectivity of T. By linear independence of $\mathcal{B}_V$, $\alpha_1 = \dots = \alpha_k = 0$, hence $\mathcal{B}_W$ is linearly independent.

(ii) $T^{-1} \in \mathcal{L}(W, V)$.

Let $a_1, a_2 \in \mathbb{F}$, and $w_1, w_2 \in W$. By definition of the inverse function T^{-1} and linearity of T,

$$\begin{aligned} T^{-1}(a_1 w_1 + a_2 w_2) = T^{-1}(a_1 T(T^{-1}(w_1)) + a_2 T(T^{-1}(w_2))) &= T^{-1}(T(a_1 T^{-1}(w_1) + a_2 T^{-1}(w_2))) \\ &= a_1 T^{-1}(w_1) + a_2 T^{-1}(w_2). \end{aligned}$$

3 Matrices and their relationship with linear maps

When discussing matrices, the first thing to note is that any matrix can be viewed as a linear map. Namely, for $M \in \mathbb{R}^{n\times d}$, we can consider the linear transformation $T_M \in \mathcal{L}(\mathbb{R}^d, \mathbb{R}^n)$ defined by

$$T_M(\mathbf{x}) \coloneqq M\mathbf{x}, \text{ for any } \mathbf{x} \in \mathbb{R}^d.$$

As an exercise, convince yourself that this is a linear map. The theory we cover in this section will actually show the converse: for any linear map $T \in \mathcal{L}(\mathbb{R}^d, \mathbb{R}^n)$, there is a matrix $M_T \in \mathbb{R}^{n\times d}$ such that $T(\mathbf{x}) = M_T\mathbf{x}$, for all $\mathbf{x} \in \mathbb{R}^d$. Abstractly, what this will mean is that $\mathbb{R}^{n\times d}$ is isomorphic to $\mathcal{L}(\mathbb{R}^d, \mathbb{R}^n)$. Moreover, it turns out that composition of linear maps between finite-dimensional vector spaces is essentially the same thing as matrix multiplication! We begin with the following definition.

Definition 2. Let V be a real vector space, and $\mathcal{B} = \{v_1, \dots, v_k\} \subseteq V$ a basis for V. There exists an invertible (linear) map $[\,\cdot\,]_{\mathcal{B}} : V \to \mathbb{R}^k$, such that for any $v \in V$,

$$v \mapsto [v]_{\mathcal{B}} = \begin{bmatrix} \alpha_{v,1} \\ \vdots \\ \alpha_{v,k} \end{bmatrix},$$

satisfies $v = \sum_{j=1}^k \alpha_{v,j} v_j$. The vector $[v]_{\mathcal{B}}$ is called the *coordinate vector* of v with respect to the basis $\mathcal{B}$, and its entries are called *coordinates.*

Remark 3. Given this definition, justified by Exercise 2, we may note that any finite-dimensional vector space is isomorphic to Euclidian space. It is also helpful to unravel this definition for the canonical basis $\mathcal{E}_k = \{e_1, \dots, e_k\} \subseteq \mathbb{R}^k$. Any $\mathbf{x} = [x_1 \cdots x_k]^T \in \mathbb{R}^k$ satsifies, $\mathbf{x} = \sum_{j=1}^k x_j e_j$, so we simply have $[\mathbf{x}]_{\mathcal{E}_k} = \mathbf{x}$.

With coordinates at our disposal, it becomes natural to define (i.e., construct) the matrix corresponding to a linear map, so long as we do come careful bookkeeping.

Definition 3. Let V, W be vector spaces with bases $\mathcal{B}_V = \{v_1, \dots, v_d\}$, $\mathcal{B}_W = \{w_1, \dots, w_n\}$, respectively, and consider $T \in \mathcal{L}(V, W)$. We will construct the matrix $M_T \in \mathbb{R}^{n\times d}$ corresponding to T, with respect to $\mathcal{B}_V$ and $\mathcal{B}_W$, as follows:

(1) Compute $T(v_j)$ and collect the vectors

$$[T(v_j)]_{\mathcal{B}_W} = \begin{bmatrix} \beta_{1j} \\ \vdots \\ \beta_{nj} \end{bmatrix} \in \mathbb{R}^n,$$

for $j = 1, \dots, d$.

(2) Aggregate the above coordinate vectors, and define

$$M_T \coloneqq \left[[T(v_1)]_{\mathcal{B}_W} \cdots [T(v_d)]_{\mathcal{B}_W}\right] = \begin{bmatrix} \beta_{11} & \cdots & \beta_{1d} \\ \vdots & \ddots & \vdots \\ \beta_{n1} & \cdots & \beta_{nd} \end{bmatrix}$$

Exercise 5. Taking the setting and notation of the previous definition, show that for any $v \in V$,

$$[T(v)]_{\mathcal{B}_W} = M_T[v]_{\mathcal{B}_V}.$$

We unravel the left and right sides of the equality. Let $[v]_{\mathcal{B}_V} = [\alpha_1 \cdots \alpha_d]^T$, so that by linearity and the definition of the β_{ij} constants,

$$T(v) = \sum_{j=1}^{d} \alpha_j T(v_j) = \sum_{j=1}^{d} \alpha_j \sum_{i=1}^{n} \beta_{ij} w_i = \sum_{i=1}^{n} \left(\sum_{j=1}^{d} \alpha_j \beta_{ij} \right) w_i,$$

so $[T(v)]_{\mathcal{B}_W} = [\sum_{j=1}^{d} \alpha_j \beta_{1j} \cdots \sum_{j=1}^{d} \alpha_j \beta_{nj}]^T$. On the other hand,

$$M_T[v]_{\mathcal{B}_V} = \begin{bmatrix} \beta_{11} & \cdots & \beta_{1d} \\ \vdots & \ddots & \vdots \\ \beta_{n1} & \cdots & \beta_{nd} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_d \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{d} \alpha_j \beta_{1j} \\ \vdots \\ \sum_{j=1}^{d} \alpha_j \beta_{nj} \end{bmatrix}.$$

Exercise 6. Let V, W, Z be vector spaces with bases $\mathcal{B}_V = \{v_1, \ldots, v_d\}$, $\mathcal{B}_W = \{w_1, \ldots, w_n\}$, $\mathcal{B}_Z = \{z_1, \ldots, z_p\}$, respectively, and let $T \in \mathcal{L}(V, W)$, $S \in \mathcal{L}(W, Z)$. Show that

$$M_{S \circ T} = M_S M_T.$$

The matrix corresponding to $S \circ T$ with respect to $\mathcal{B}_V$ and $\mathcal{B}_Z$ is, by definition,

$$M_{S \circ T} = \big[[S(T(v_1))]_{\mathcal{B}_Z} \cdots [S(T(v_d))]_{\mathcal{B}_Z}\big] = \big[M_S[T(v_1)]_{\mathcal{B}_W} \cdots M_S[T(v_d)]_{\mathcal{B}_W}\big],$$

by Exercise 5. Pulling out the matrix M_S, this simplifies to

$$M_{S \circ T} = M_S \big[[T(v_1)]_{\mathcal{B}_W} \cdots [T(v_d)]_{\mathcal{B}_W}\big] = M_S M_T.$$

Remark 4. Combined, the previous two exercises show that linear maps are in a sense equivalent to matrices, which we can describe more precisely through an isomorphism. Moreover, the special structure of linear map composition is preserved by matrices, in the form of matrix multiplication. We will most often be dealing with the case that $V = \mathbb{R}^d$, $W = \mathbb{R}^n$, and will use the canonical bases $\mathcal{E}_d$ and $\mathcal{E}_n$. In this case, our theory tells us that not only are $(n \times d)$ matrices linear maps, but they are the *only* linear maps between $\mathbb{R}^d$ and $\mathbb{R}^n$. That is, for any $T \in \mathcal{L}(\mathbb{R}^d, \mathbb{R}^n)$, there is a unique matrix $M_T \in \mathbb{R}^{n \times d}$ (use above construction) such that

$$T(\mathbf{x}) = [T(\mathbf{x})]_{\mathcal{E}_n} = M_T[\mathbf{x}]_{\mathcal{E}_d} = M_T\mathbf{x},$$

for all $\mathbf{x} \in \mathbb{R}^d$. One important matrix is that corresponding to the identity map, $M_{I_{\mathbb{R}^d}} \equiv I_d$, the familiar $(d \times d)$ identity matrix.

Remark 5. When describing solutions to the least squares problem, we will need to talk about inverses of square matrices. By our characterization, we can simply define $W \in \mathbb{R}^{d \times d}$ to be invertible iff $T_W \in \mathcal{L}(\mathbb{R}^d, \mathbb{R}^d)$ is invertible as a linear map. In this case, we write $W^{-1} = W_{T_W^{-1}}$, and this must uniquely satisfy

$$W^{-1}W = WW^{-1} = I_d.$$

4 Special subspaces related to matrices (if we have time)

Given the correspondence between matrices and linear maps, we can predict that two subspaces will play a special role in characterizing a matrix. Let $M \in \mathbb{R}^{n\times d}$, and write

$$M = \begin{bmatrix} M_1^T \\ \vdots \\ M_n^T \end{bmatrix} = \begin{bmatrix} M^{(1)} & \cdots & M^{(d)} \end{bmatrix}$$

We will define the *column space* of M as the space spanned by the columns of M,

$$\mathcal{C}(M) \coloneqq \mathscr{L}(M^{(1)}, \ldots, M^{(d)}) = \left\{ \sum_{j=1}^{d} x_j M^{(j)} \,\middle|\, x_1, \ldots, x_d \in \mathbb{R} \right\} = \left\{ M\mathbf{x} \,\middle|\, \mathbf{x} \in \mathbb{R}^d \right\} = \mathrm{Im}(T_M) \subseteq \mathbb{R}^n,$$

and the *nullspace* of M as

$$\mathcal{N}(M) \coloneqq \left\{ \mathbf{x} \in \mathbb{R}^d \,\middle|\, M\mathbf{x} = \mathbf{0} \right\} = \mathrm{Ker}(T_M) \subseteq R^d.$$

Conveniently, any results about linear maps can be translated directly into results about matrices. For example we can restate the rank-nullity theorem in terms of $M \in \mathbb{R}^{n\times d}$:

$$d = \dim(\mathbb{R}^d) = \dim(\mathrm{Im}(T_M)) + \dim(\mathrm{Ker}(T_M)) = \dim(\mathcal{C}(M)) + \dim(\mathcal{N}(M))$$

Indeed, the *rank* of a matrix M is defined as the dimension of its column space, and we write $\mathrm{rank}(M) \coloneqq \dim(\mathcal{C}(M))$.

As a preview for things to come — look forward to spectral decomposition and singular value decompositions of matrices — it is useful to take the transpose of M and consider the two subspaces,

$$\mathcal{C}(M^T) = \left\{ M^T\mathbf{y} \,\middle|\, \mathbf{y} \in \mathbb{R}^n \right\} = \left\{ \sum_{i=1}^{n} y_i M_i \,\middle|\, y_1, \ldots, y_n \in \mathbb{R} \right\} = \mathscr{L}(M_1, \ldots, M_n) \eqqcolon \mathcal{R}(M) \subseteq \mathbb{R}^d,$$

also known as the *row space* of M, and

$$\mathcal{N}(M^T) = \left\{ \mathbf{y} \in \mathbb{R}^n \,\middle|\, M^T\mathbf{y} = \mathbf{0} \right\} \subseteq \mathbb{R}^n,$$

also known as the *left nullspace* of M. Together the four fundamental subspaces characterize the matrix M.

Once we have developed more theory of inner product spaces, including projections, orthogonal complements and direct sums, we will be able to show a nice result about matrix rank:

$$\mathrm{rank}(M) = \mathrm{rank}(M^T) = \mathrm{rank}(M^TM) = \mathrm{rank}(MM^T).$$

Put differently, the dimension of the column space is always the same as the dimension of the row space! The square matrices $M^TM \in \mathbb{R}^{d\times d}$ and $MM^T \in \mathbb{R}^{n\times n}$ may appear mysterious now, but their importance will become clear over the coming weeks.