# A QUICK REVIEW OF SOME LINEAR ALGEBRA

K. N. RAGHAVAN
THE INSTITUTE OF MATHEMATICAL SCIENCES

CONTENTS



These are notes of lectures given during 2013–14 at various NBHM sponsored workshops. Corrections, comments, suggestions for improvement, criticisms, etc. are welcome. Please send them to the author at `knr.imsc@gmail.com` or `knr@imsc.res.in`. An updated version of these notes is available online at http://www.imsc.res.in/~knr/past/linalg_ed.pdf .

## Preliminaries

Let $\mathbb{F}$ be a field. In what follows, entries of all matrices are supposed to be from $\mathbb{F}$.

Let $\mathbb{F}^m$ denote the vector space of $m \times 1$ matrices ("column matrices"). Let $e^j$ denote the $m \times 1$ column matrix with all entries zero except that on row $j$ which is $1$. Then $e^j$, $1 \leq j \leq m$, form an (ordered) basis for $\mathbb{F}^m$, called the *standard basis*.

Let $\mathbb{F}_m$ denote the vector space of $1 \times m$ matrices ("row matrices"). The map $\mathbb{F}^m \times \mathbb{F}_m \to \mathbb{F}$ given by $(u, v) \mapsto vu$ is *bilinear* (that is, linear separately in the arguments $u$ and $v$ with the other fixed) and *perfect*, that is, $vu = 0$ for all $u$ means $v = 0$; and $vu = 0$ for all $v$ means $u = 0$. This map is sometimes called the *standard pairing* between $\mathbb{F}^m$ and $\mathbb{F}_m$.

**Notions associated with a matrix.** Let $A$ be an $m \times n$ matrix. We may think of $A$ as a linear transformation—also denoted $A$—from $\mathbb{F}^n$ to $\mathbb{F}^m$: the image under the linear transformation $A$ of an element $u$ in $\mathbb{F}^n$ is the matrix product $Au$. The columns of the matrix $A$ are in order the images under the linear transformation $A$ of the standard basis of $\mathbb{F}^n$, so the linear transformation $A$ determines the matrix $A$.

Column space and column rank, row space and row rank. The *column space* of $A$ is the subspace of $\mathbb{F}^m$ spanned by the columns of $A$. It is also the image of the linear transformation $A$. The *column rank of* $A$ is the dimension of this subspace. The *row space* of $A$ is the subspace of $\mathbb{F}_n$ spanned by the rows of $A$. The *row rank* of $A$ is the dimension of this subspace. The kernel of the linear transformation $A$ is (as a subspace of $\mathbb{F}^n$) evidently "the perpendicular" of the row space of $A$: the *perpendicular* of a subspace $V$ of $\mathbb{F}_n$ is by definition the subspace $\{w \in \mathbb{F}^n \mid vw = 0 \;\forall\; v \in V\}$ in $\mathbb{F}^n$.

**Remark 1.** We will soon prove (see Corollary 6) that the row rank and column rank of a matrix are equal. We will then be justified in using the word *rank* to mean either of them.

a matrix

**Proposition 2.** *Let $A$ be an $m \times n$ matrix and $A'$ an $m' \times n$ matrix. If their row spaces are the same, then their column ranks are equal. In fact, a set of columns of $A$ forms a basis for its column space if and only if the corresponding set of columns of $A'$ forms a basis for its column space.*

Proof: The solution space of $Ax = 0$ is the perpendicular to the row space of $A$ (as observed above). The same thing holds also for $A'$, and since the row spaces of $A$ and $A'$ are equal, the solution spaces of $Ax = 0$ and $A'x = 0$ are equal. This means that any linear dependence relation among the columns of $A$ is also one among the corresponding columns of $A'$ and vice versa. In particular, if a particular set of columns of $A$ forms a basis for its column space, then the corresponding set of columns of $A'$ forms a basis for its column space, and vice versa. □

## 1. Row reduced echelon forms

This section is technical, elementary, and fundamental. The definitions involve arbitrary choices and are admittedly contrived. Nevertheless Lemma 5 is the lynchpin of §1-3, which are in turn basic to all that follows.

1.1. **Elementary row operations.** The following simple operations on the rows of a $m \times n$ matrix are called *elementary*:

- interchanging two rows
- multiplying a row by a non-zero scalar
- adding a multiple of one row to another

Every elementary row operation is reversible by means of an elementary row operation: interchanging two specific rows is inverse to itself; multiplying a specific row by the reciprocal of a non-zero scalar is inverse to mulitplication of that row by that scalar; adding $-c$ times row $i$ to row $j$ is inverse to adding $c$ times row $i$ to row $j$.

1.1.1. ER-equivalence. Two $m \times n$ matrices are called *ER-equivalent* if one can be obtained from the other by a sequence of elementary row operations: ER stands for "elementary row". That this is an equivalence relation is clear from what has been said above.

Since the resulting rows after an elementary row operation are all linear combinations of the original rows, and every elementary row operation is reversible by means of an elementary row operation, we have:

**Proposition 3.** *ER-equivalent matrices have the same row space.*

The converse of the above is also true as proved a little later in this section: see Corollary 6.

Whenever there is an equivalence relation, it is natural to look for what are called *sections*. What this amounts to in our present case is:

> Identify a special class of matrices such that there is one and only one special matrix in every row ER-equivalence class.

The row reduced echelon forms defined below are precisely such a special class of matrices: see item (iv) of Lemma 5 below.

1.2. **Definition of row reduced echelon forms.** An $m \times n$ matrix is said to be in *row reduced echelon form* (RREF) if the following conditions are satisfied

- The *leading entry* of any non-zero row—that is, the first non-zero entry from the left on the row—is $1$.
- The column number in which the leading entry of a row appears is a strictly increasing function of the row number.[1] This should be interpreted to mean in particular that zero rows appear only after all non-zero rows.
- If a column contains a leading entry of a row, then all its other entries are zero.

**Remark 4.** (1) A submatrix of a matrix in RREF consisting of the first so many rows or the first so many columns is also in RREF.

(2) If $E$ is a matrix in RREF with $r$ non-zero rows, then its transpose $E^t$ is ER-equivalent to the matrix with top left $r \times r$ block being the identity matrix and zero for the rest of the entries.

[1] It is instructive to note that in military parlance *echelon* means a formation of troops, ships, aircraft, or vehicles in parallel rows with the end of each row projecting further than the one in front.

1.3. **Results about matrices in RREF.** The following technical lemma about matrices in RREF is fundamental to the sequel.

**Lemma 5.** *(i) The non-zero rows of a matrix in RREF form a basis for its row space.*

*(ii) Those columns of a matrix $E$ in RREF in which the leading entries of the non-zero rows appear form a basis of the column space of $E$.*

*(iii) If two matrices of the same size both in RREF have the same row space, then they are equal.*

*(iv) Every matrix is ER-equivalent to a unique matrix in RREF.*

PROOF: Let us first fix some useful notation. Let $E$ be an $m \times n$ matrix in RREF, let $R_1$, ..., $R_r$ be in order its non-zero rows, and let $\ell_1, \ldots, \ell_r$ be the respective column numbers in which the leading entries of the rows $R_1, \ldots, R_r$ occur. Then, by the second condition in the definition of RREFs, we have $\ell_1 < \ldots < \ell_r$.

Suppose that $R$ is a row matrix of size $1 \times n$ in the row space of $E$. Then $R$ is uniquely a linear combination of the non-zero rows of $E$ as follows:

$$R = R(\ell_1)R_1 + \cdots + R(\ell_r)R_r \qquad \text{where } R(\ell_j) \text{ is the entry on column } \ell_j \text{ of } R \tag{1}$$

Indeed the only contribution to $R(\ell_j)$ is from $R_j$, the entry in that column of every other $R_i$ being zero.

This proves (i). Moreover we conclude from (1) that

$$\text{if the leading entry of } R \text{ occurs in column } c\text{, then } c \text{ must equal some } \ell_j \tag{2}$$

In fact, $c$ equals $\ell_j$ where $j$ is least such that $R(\ell_j)$ is non-zero.

Now let $E'$ be another matrix of size $m \times n$ in RREF whose row space is the same as that of $E$. Fix notation for $E'$ analogous to that for $E$ (see above): $R'_1, \ldots, R'_{r'}$ be the non-zero rows of $E'$ and $\ell'_1, \ldots, \ell'_{r'}$ be the column numbers in which the leading entries of the respective rows. Since $E$ and $E'$ have the same row space, every row of either of them is in the span of the rows of the other. Applying (2), we conclude that $r = r'$ and $\ell_j = \ell'_j$ for $1 \leq j \leq r$. Now, applying (1) with $R'_j$ in place of $R$, we get

$$R'_j \;=\; R'_j(\ell_1)R_1 + \cdots + R'_j(\ell_r)R_r \;=\; R'_j(\ell'_1)R_1 + \cdots + R'_j(\ell'_r)R_r \;=\; R'_j(\ell'_j)R_j \;=\; R_j$$

Thus (iii) is proved.

Now suppose that $C$ is a column matrix of size $m \times 1$ in the column space of $E$. Then $C$ is uniquely a linear combination of the columns $C_{\ell_1}, \ldots, C_{\ell_r}$ of $E$ as follows:

$$C = C(1)C_{\ell_1} + \cdots + C(r)C_{\ell_r} \qquad \text{where } C(j) \text{ is the entry on row } j \text{ of } C \tag{3}$$

Indeed the only contribution to $C(j)$ is from $C_{\ell_j}$, the entry in row $j$ of every other $C_{\ell_i}$ being zero. This proves (ii).

For (iv), let $A$ be an arbitrary $m \times n$ matrix. Let us first prove the uniqueness part. Suppose that $E$ and $E'$ are matrices in RREF that are ER-equivalent to $A$. Then by Proposition 3, the row spaces of $E$ and $E'$ are the same. Now, by (iii) above, $E = E'$.

To prove the existence part of (iv), we describe an algorithm to bring $A$ into RREF by performing ER operations on it. The procedure is inductive and has a stepwise description as follows:

- Proceed by induction on the number of columns in $A$. Scan the first column of $A$. Suppose it is zero. If $A$ has a single column, then just return it: there is nothing to do. If $A$ has more than one column, then, letting $A'$ be the matrix of $A$ with first column deleted, we know by induction the conclusion for $A'$. The same operations will do also for $A$: the first column remains zero throughout the process.

- Now suppose that the first column of $A$ has a non-zero entry, say $e$ in row $r$. Interchange rows $1$ and $r$ so as to bring that entry to the first row. Multiply row $1$ by $1/e$ so as to make the entry now in position $(1,1)$ equal to $1$. Add suitable multiples of the first row to other rows so that all other entries in the first column are zero.
- If $A$ has a single column, stop at this point. If it has more than one column, let $A'$ be the submatrix of $A$ (in its present form) obtained by deleting the first row and first column. By induction, $A'$ can be brought to RREF by elementary row operations. Perform this sequence of operations on $A$. The first column of $A$ remains unchanged throughout (the entry in position $(1,1)$ is $1$ and the remaining entries are all $0$) and the resulting matrix is in RREF from the second row onwards.
- Finally add suitable multiples of the non-zero second, third, ... rows to the first so as to ensure vanishing of the entry in the first row of every column containing a leading entry of a non-zero row (second onwards). The resulting matrix is in RREF. □

1.4. **Rank of a matrix.**

**Corollary 6.** *(1) The row rank and column rank of a matrix are equal.*

*(2) Two matrices of the same size are ER-equivalent if and only if they have the same row space.*

*(3) The transpose of a matrix has the same rank as the matrix.*

PROOF: (1) Let $A$ be an arbitrary matrix and $E$ a matrix in RREF that is ER-equivalent to $A$. Such an $E$ exists by (iv) of the lemma. By Proposition 3, $A$ and $E$ have the same row space. So $A$ and $E$ have the same row rank. Moreover, by Proposition 2, they have the same column rank. Finally, by (i) and (ii) of the lemma, the row rank and column rank of $E$ are equal.
(2) Let $A$ and $A'$ be matrices of the same size. If they are ER-equivalent, then they have the same row space by Proposition 3. For the converse, let $E$ and $E'$ be matrices in RREF ER-equivalent to $A$ and $A'$ respectively. Such matrices exist by (iv) of the lemma. Since $A$ and $A'$ have the same row space by hypothesis, it follows from Proposition 3 that $E$ and $E'$ have the same row space. By (iii) of the lemma, $E = E'$. It follows that $A$ and $A'$ are ER-equivalent.
(3) The row space of the transpose of a $m \times n$ matrix is isomorphic to the column space of the matrix via (the restriction of) the standard isomorphism $v \mapsto v^t$ from $\mathbb{F}^m$ to $\mathbb{F}_m$. □

## Exercise Set 1

These exercises are taken from Gilbert Strang's book Introduction to Linear Algebra, South Asian Edition.

(1) For each of the following matrices, find their RREFs. What are their ranks? Here $c$ denotes a variable scalar.

$$\begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & -1 \end{pmatrix} \qquad \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 3 \\ 2 & 4 & 6 \end{pmatrix} \qquad \begin{pmatrix} 1 & 1 & 2 & 2 \\ 2 & 2 & 4 & 4 \\ 1 & c & 2 & 2 \end{pmatrix} \qquad \begin{pmatrix} 1-c & 2 \\ 0 & 2-c \end{pmatrix}$$

(2) Find the RREFs and ranks of the following $3 \times 4$ matrices:
  (a) when all entries are $1$.
  (b) when the entry in position $(i, j)$ is $i + j - 1$.
  (c) when the entry in position $(i, j)$ is $(-1)^{i+j}$.

(3) If $R$ be the RREF of a matrix $A$, what are the RREFs of the following block matrices:

$$\begin{pmatrix} A & A \end{pmatrix} \qquad \begin{pmatrix} A & A \\ A & 0 \end{pmatrix}$$

(4) Fill out the following matrices so that they have rank $1$:

$$\begin{pmatrix} 1 & 2 & 4 \\ 2 & & \\ 4 & & \end{pmatrix} \qquad \begin{pmatrix} & 9 & \\ 1 & & \\ 2 & 6 & -3 \end{pmatrix} \qquad \begin{pmatrix} a & b \\ c & \end{pmatrix}$$

(5) Express the following matrices as the sum of two rank one matrices:

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 4 \\ 1 & 1 & 8 \end{pmatrix} \qquad \begin{pmatrix} 2 & 2 \\ 2 & 3 \end{pmatrix}$$

(6) Write each of following rank $1$ matrices as the product of a column matrix and a row matrix:

$$\begin{pmatrix} 3 & 6 & 6 \\ 1 & 2 & 2 \\ 4 & 8 & 8 \end{pmatrix} \qquad \begin{pmatrix} 2 & 2 & 6 & 4 \\ -1 & -1 & -3 & -2 \end{pmatrix}$$

(7) True or false?: If a matrix is the product of a column matrix and a row matrix then its rank is at most one.

(8) Prove that every $m \times n$ matrix of rank $r$ can be written as the product of an $m \times r$ and an $r \times n$ matrix.

(9) Suppose we allow elementary column operations on a matrix in addition to elementary row operations. What is the *row-and-column reduced form* of an $m \times n$ matrix of rank $r$?

(10) Describe all $2 \times 3$ matrices $A_1$ and $A_2$ such that $R_1 + R_2$ is the RREF of $A_1 + A_2$ (where $R_1$ and $R_2$ are the RREFs of $A_1$ and $A_2$ respectively).

An $m \times n$ matrix $A$ is said to have a *left inverse* (respectively, *right inverse*) if there exists an $n \times m$ matrix $L$ (respectively, $n \times m$ matrix $R$) such that $LA$ (respectively, $AR$) is the identity $n \times n$ matrix (respectively, identity $m \times m$ matrix). If $A$ and $A'$ have respective left (respectively, right) inverses $L$ and $L'$, then $AA'$ has a left (respecitvely, right) inverse $L'L$. If $A$ has a left inverse $L$, then $L$ has $A$ for a right inverse.

Suppose that a matrix $A$ has both a left inverse $L$ and a right inverse $R$. Then

$$L = L \cdot 1 = L \cdot AR = LA \cdot R = 1 \cdot R = R.$$

In particular, $L$ and $R$ are uniquely determined, and (being equal) commonly denoted by $A^{-1}$. We say in this case that $A$ is *invertible* (or that it is a *unit*) and that $A^{-1}$ is its *inverse*. If $A$ is invertible, so is $A^{-1}$ and its inverse is $A$. The product $AB$ of invertible matrices is invertible with inverse $B^{-1}A^{-1}$.

**Proposition 7.** *Let $A$ be an $m \times n$ matrix and $E$ the matrix in RREF that is ER-equivalent to it (see Theorem 5 (iv)).*

*(1) If $A$ has a left inverse, then $m \geq n$ and the first $n$ rows of $E$ form the identity matrix.*
*(2) If $A$ has a right inverse, then $n \geq m$.*
*(3) If $A$ is invertible, then $m = n$ and $E$ is the identity matrix.*

PROOF: (1) Let $L$ be a left inverse of $A$. Since the rows of the product $LA$ are linear combinations of the rows of $A$, the row space of $LA$ is contained in the row space of $A$. But $LA$ being the identity, its row space is $\mathbb{F}_n$, so $A$ has full possible row space $\mathbb{F}_n$. Since $E$ is ER-equivalent to $A$, they have the same row space (Proposition 3). Since the non-zero rows of $E$ form a basis for its row space (Theorem 5 (i)) which is of dimension $n$, it follows that $E$ has exactly $n$ non-zero rows, and so in particular $m \geq n$. Moreover, since any zero row of $E$ appears below all non-zero rows (see the second condition in the definition of RREF), the first $n$ rows of $E$ are its non-zero ones. If $E_n$ be the submatrix of $E$ consisting of its first $n$ rows, then $E_n$ and the $n \times n$ identity matrix are matrices of the same size both in RREF and with a common row space. By Theorem 5 (iii), $E_n$ is the identity matrix.

(2) Let $R$ be a right inverse of $A$. The columns of $AR$ being linear combinations of the columns of $A$, it follows that the column space of $A$ is the full space $\mathbb{F}^m$. But it is spanned by the $n$ columns of $A$, so $n \geq m$. Alternatively, one could observe that $A^t$ has a left inverse and just invoke (1).

(3) follows from (1) and (2). □

2.1. **Elementary row matrices and their invertibility.** To every elemetary row operation (see §1.1) on $m \times n$ matrices there is (as we will presently observe) a corresponding *elementary row matrix* $E$ of size $m \times m$: performing the row operation (on any matrix) corresponds precisely to multiplying on the left by $E$. The matrix $E$ is uniquely determined, as follows from the following elementary observation:

$$\text{for } m, n, \text{ and } p \text{ positive integers, and } A, A' \text{ a } m \times n \text{ matrices,} \\ \text{if } AB = A'B \text{ for every } n \times p \text{ matrix } B, \text{ then } A = A'. \tag{4}$$

Letting $I$ denote the $m \times m$ identity matrix and $E_{ij}$ the $m \times m$ matrix all of whose entries are zero except the one at the spot $(i, j)$, the elementary row matrices corresponding respectively to the three elementary row operations listed in §1.1 are:

- $E_{ji} + E_{ij} + \sum_{k\neq i, k\neq j} E_{kk}$ where the interchange is of rows $i$ and $j$
- $I + (c-1)E_{ii}$ if row $i$ is being multiplied by $c \neq 0$
- $I + cE_{ij}$ where $c$ times row $i$ is being added to row $j$

**Proposition 8.** *Elementary row matrices and their products are invertible.*

PROOF: It suffices to prove that every elementary row matrix in invertible. Let $E$ be such a matrix. As observed in §1.1, every elementary row operation can be reversed by another elementary row operation. So there is an elementary matrix $E'$ such that $E'EB = B$ for every $m \times n$ matrix $B$. By (4), this means that $E'E$ is the identity matrix, in other words that $E'$ is a left inverse for $E$. Since $E'$ is itself an elementary row matrix and thus a left inverse, say $E''$, it has both a left inverse and a right inverse (namely, $E''$ and $E$). We conclude that $E'' = E$. Thus $E'$ is also a right inverse for $E$, and $E$ is invertible. □

### 2.2. Conditions for invertibility of a square matrix.

**Proposition 9.** *(1) Any square matrix having a left inverse is ER-equivalent to the identity matrix and hence a product of elementary row matrices.*
*(2) Any square matrix having a left or right inverse is invertible.*

PROOF: Let $A$ be a square matrix.

(1) If $A$ has a left inverse, then, by Proposition 7 (1), it is ER-equivalent to the identity, which means that it is a product of elementary row matrices.

(2) If $A$ has a left inverse, then it is invertible by part (1) and Proposition 8. If it has a right inverse $R$, then $R$ has a left inverse and so by (1) is invertible. The left inverse of $R$ being unique, $A = R^{-1}$. Alternatively, in the case $A$ has a right inverse, one could observe that its transpose has a left inverse and reduce to (1). □

2.3. **Computing the inverse by row reduction.** Given an invertible $n \times n$ matrix $A$, there is a convenient device to compute its inverse. Form the $n \times 2n$ matrix $(A|I)$, where $I$ is the identity matrix of size $n \times n$. Perform the same ER-operations on $(A|I)$ as you would on $A$ to transform it to the identity matrix $I$. After these operations, the matrix $(A|I)$ would have been transformed to the form $(I|B)$. The matrix $B$ is the inverse of $A$. Indeed, if $E_n \cdots E_1 A = I$, for ER-matrices $E_1, \ldots, E_n$, then, on the one hand, $A^{-1} = E_n \cdots E_1$, and, on the other, $E_n \cdots E_1(A|I) = (I|E_n \cdots E_1)$.

We work out an example:

$$\left(\begin{array}{rrr|rrr} 1 & -1 & 0 & 1 & 0 & 0 \\ -1 & 1 & -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 1 \end{array}\right) \quad \text{is ER-equivalent to} \quad \left(\begin{array}{rrr|rrr} 1 & 0 & 0 & 0 & -1 & -1 \\ 0 & 1 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & -1 & -1 & 0 \end{array}\right)$$

$$\text{and so} \quad \begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & -1 \\ 0 & -1 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 0 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & 0 \end{pmatrix}$$

## 3. LINEAR SYSTEMS OF EQUATIONS

Consider the matrix equation $Ax = b$, where $A$ is an $m \times n$ matrix, $b$ a column matrix of size $m \times 1$, and $x$ a column matrix of unknowns of size $n \times 1$. This is equivalent to a "system" of $m$ linear constraints on the unknown entries of $x$, and hence called a *linear system*.

To *solve* the system means to determine the set of all elements $x$ of $\mathbb{R}^n$ that satisfy the equation. Linear systems of equations are ubiquitous in mathematics and its applications. It is important therefore to be able to solve them. We quickly review the standard method of solution by *row reduction* and, along the way, comment on the structure of the solution set.

The system $Ax = 0$ is called the *corresponding homogeneous system* and its solution set—which is here denoted by $K$—is related closely to that of the original system. Indeed, the solution set of $Ax = b$ equals $x_0 + K$, where $x_0$ is any one particular solution (of $Ax = b$).

In terms of the linear transformation $A$ from $\mathbb{R}^n$ to $\mathbb{R}^m$ (corresponding to the matrix $A$), the solution set of $Ax = b$ is precisely the preimage of $b$. In particular, $Ax = b$ admits a solution if and only if $b$ belongs to the image of the linear transformation $A$ (which equals the column space of the matrix $A$). The solution space $K$ of the homogeneous system $Ax = 0$ is just the kernel of the linear transformation $A$.

3.1. **The method of row reduction.** The observation underlying this method is:

> The solution set remains unchanged when the system $Ax = b$ is modified by mulitplication on the left by an by $m \times m$ invertible matrix $C$. In other words, the system $CAx = Cb$ has the same solution set as $Ax = b$ for $C$ invertible.

We choose an invertible matrix $C$ so that $CA$ is in RREF: note that $A$ may be brought to RREF by ER operations (Lemma 5 (iv)), that each ER operation is equivalent to left multiplication by an ER matrix (§2.1), and that each ER matrix (and so also any product of ER matrices) is invertible (Proposition 8). Left multiplication by such a $C$ transforms $Ax = b$ to $A'x = b'$, where $A' = CA$ is in RREF and $b' = Cb$. We may thus assume, without loss of generality, that $A$ is in RREF.

3.2. **The solution of $Ax = b$ in case $A$ is in RREF.** We claim that, for $A$ in RREF, the system $Ax = b$ admits a solution if and only if $b_{r+1} = \ldots = b_m = 0$ where $r$ is the number of non-zero rows (or the rank) of $A$. Indeed, this condition is necessary, for $(Ax)_{r+1} = \ldots = (Ax)_m = 0$, no matter what $x$ is, since the rows $r+1$ through $m$ of $A$ are zero. For the converse, assuming that $b_{r+1} = \ldots = b_m = 0$, we produce a particular solution as follows: let $\ell_1, \ldots, \ell_r$ be in order the column numbers in which the leading (non-zero) entries of the non-zero rows of $A$ appear; set $x_{\ell_k} = b_k$ for $k$ such that $1 \leq k \leq r$, and $x_j = 0$ if $j \notin \{\ell_1, \ldots, \ell_r\}$.

To get the general solution of $Ax = b$ (still assuming $A$ to be in RREF of rank $r$, with notation as above), now that we have a particular solution in hand (or know that the system admits no solution, in which case there is nothing more to be done), we try to find the space $K$ of solutions of the homogeneous system $Ax = 0$: as observed above, the general solution of $Ax = b$ is the sum of $K$ with any particular solution. Let us first consider an example. Suppose that $A$ is:

$$\begin{pmatrix} 0 & 1 & 2 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 3 & -2 \end{pmatrix}$$

The solution space $K$ in this case consists of all $x \in \mathbb{R}^6$ such that $x_2 = -2x_3 + x_5$ and $x_4 = -3x_5 + 2x_6$. We may think of it as being parametrized by $x_1$, $x_3$, $x_5$, and $x_6$, each of these four being free to take on any scalar as value. For a basis of $K$, we could take $e^1$, $e^3 - 2e^2$, $e^5 + e^2 - 3e^4$, and $e^6 + 2e^4$.

As in the illustration above, the RREF properties of the matrix $A$ enable an explicit description of the solution space $K$ of $Ax = 0$. enable an explicit description of the solution space $K$ of $Ax = 0$. The $r$ equations in the system can be written as: $x_j = -\sum a_{ki}x_i$ for $j \in \{\ell_1, \ldots, \ell_r\}$ where the sum is over $i$ such that $j + 1 \leq i \leq n$, $i \notin \{\ell_1, \ldots, \ell_k\}$. Note that there is a "separation of variables" here: the $x_i$ with $i \in \{\ell_1, \ldots, \ell_r\}$ appear only on the left hand side and the other $x_i$ only on the right hand side (of any of these $r$ equations). We may therefore think of the latter $x_i$ as completely free and the former $x_i$ as determined (by the values of the latter $x_i$). In other words, $K$ has a basis consisting of $e^j - \sum_{1\leq k\leq r} a_{\ell_k j}e^{\ell_k}$, indexed by $j$ such that $1 \leq j \leq n$, $j \notin \{\ell_1, \ldots, \ell_r\}$.

**Corollary 10.** (Rank-nullity theorem) *For an $m \times n$ matrix $A$ of rank $r$, the dimension of the space of solutions of the homogeneous system $Ax = 0$ is $n - r$.*

PROOF: By the method of row reduction (§3.1), we may assume $A$ to be in RREF. But then, as we have just seen, $K$ has a basis consisting of $n - r$ elements. □

**Corollary 11.** *If $S$ is a subspace of dimension $d$ of $\mathbb{F}_n$, then the dimension of its perpendicular space $S^\perp$ (in $\mathbb{F}^n$ or $\mathbb{F}_n$) is $n - d$.*

PROOF: Choose $A$ to be an $m \times n$ matrix the span of whose rows is $S$. Then $A$ had rank $d$ and $S^\perp$ (in $\mathbb{F}^n$) is the solution space $K$ of the homogeneous system $Ax = 0$. Now invoke the previous corollary. □

3.3. **A practical procedure to solve linear systems of equations.** In practice, when we are solving a system $Ax = b$ as above, it is convenient to encode the given information (of $A$ and $b$) in the form of the $m \times (n + 1)$-matrix $(A|b)$. Now apply the same ER operations on $(A|b)$ as you would do on $A$ to bring it to RREF $A'$. Let $(A'|b')$ be the result. If $b'_j \neq 0$ for some $j > r$ (where $r$ is the rank of $A$ or $A'$) then there is no solution. Otherwise, there is a solution. In fact, the set of solutions consists of $x \in \mathbb{R}^n$ such that $x_{\ell_k} = b'_k - \sum_{\ell_k<j\leq n, j\notin\{\ell_1,\ldots,\ell_r\}} a'_{kj}x_j$, where each of $x_j$, $j \notin \{\ell_1, \ldots, \ell_r\}$, is free to take on any scalar value.

## EXERCISE SET 2

Most but not all of these problems are taken from the linear algebra exams available on MIT open courseware.

(1) Given $A$ to be the following $3 \times 4$ matrix, solve the system $Ax = 0$. What is the dimension of the solution space? Find a basis for the column space of $A$. What is the RREF of the $6 \times 8$ matrix $B$?

$$A = \begin{pmatrix} 0 & 1 & 2 & 2 \\ 0 & 3 & 8 & 7 \\ 0 & 0 & 4 & 2 \end{pmatrix} \qquad B = \begin{pmatrix} A & A \\ A & A \end{pmatrix}$$

(2) Consider the linear system $Ax = b$ where $A$ is a $3 \times 3$ matrix. After performing ER operations on both sides, $Ax = b$ is transformed to $Rx = d$ with $R$ in RREF. The complete solution to $Rx = d$ is given by

$$x = \begin{pmatrix} 4 \\ 0 \\ 0 \end{pmatrix} + c_1 \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 5 \\ 0 \\ 1 \end{pmatrix}$$

(a) What is the solution space of the homogeneous system $Rx = 0$? What of $Ax = 0$?
(b) Determine $R$ and $d$.
(c) In moving from $Ax = b$ to $Rx = d$, the ER operations performed in order were: first subtract $3$ times row $1$ from row $2$; then $5$ times row $1$ from row $3$. Determine $A$ and $b$.

(3) Is $\begin{pmatrix} 8 \\ 28 \\ 14 \end{pmatrix}$ in the column space of $\begin{pmatrix} 2 & 1 \\ 6 & 5 \\ 2 & 4 \end{pmatrix}$?

(4) Given that $A$ is a $3 \times 5$ matrix of rank $3$, circle the words that correctly complete the following sentence: the equation $Ax = b$ (always / sometimes but not always) has (no solution / a unique solution/ many solutions). What is the column space of $A$? What is the dimension of its null space?

(5) Let $A$ be an $m \times n$ matrix of rank $r$. Given that $B$ is a matrix of size $n \times p$ whose columns form a basis for the solution space $Ax = 0$, what is $p$ (in terms of $m$, $n$, and $r$)? Describe the solution space of the linear system $xB = 0$.

(6) If $Ax = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ has no solutions and $Ax = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ has precisely one solution, what all can you say about the size and rank of $A$? What is the null space of $A$? Find a matrix $A$ that fits the given description.

(7) Given that the RREF of a $3 \times 5$ matrix $A$ has the following form, what all can you say about the columns of $A$?

$$\begin{pmatrix} 1 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

(8) What is the subspace of $3 \times 3$ matrices spanned by all possible RREFs (of size $3 \times 3$ of course)?

4.1. **Motivation: the line of best fit.** As motivation for what follows later in this section, consider the following situation that occurs routinely in laboratories (presumably!). Suppose that we know that two quantities of interest are related *linearly*—which means that one is a function of the other and that its graph with respect to the other is a straight line—and that we are trying to determine this straight line experimentally. We vary one of the quantities (over a finite set of values) and measure the corresponding values of the other, thereby getting a set of data points. Now if we try to find a straight line running through our set of data points, there is often no such line! This after all should not be so surprising, there being several reasons for the deviation from the ideal behaviour, not the least of which is experimental error. At any rate, our problem now is to find a line that "best fits" the data points.

One way of formulating the problem and the "best fit" criterion is as follows. Let the line of best fit be $y = mx + c$, where $m$ is the slope and $c$ the $y$-intercept.[2] Let $(x_1, y_1)$, ..., $(x_n, y_n)$ be the set of data points. The point $(x_k, y_k)$ lies on $y = mx + c$ if and only if $y_k = mx_k + c$. The ideal situation (which as noted above is rarely the case) would be when we can solve the following system of linear equations for $m$ and $c$ (in other words, when all the data points do lie on a single line):

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} \begin{pmatrix} m \\ c \end{pmatrix} \tag{5}$$

Note that there is a solution to the above system if and only if the column vector on the left side belongs to the column space of the $n \times 2$ matrix on the right. Confronted with the problem of being forced to "solve" this system when no solution exists, a natural thing to do would be to replace the column vector on the left side by the vector "closest" to it in the column space of the $n \times 2$ matrix and then solve. Such a closest vector is given by the "orthogonal projection" (to be defined presently) of the column vector on the left side on to the column space. This approach demands that we know how to compute orthogonal projections.

4.2. **Definition of the orthogonal projection to a subspace.** Suppose that we are given a subspace $V$ of $\mathbb{R}^m$. Define $V^\perp := \{w \in \mathbb{R}^m \mid w^t v = 0 \text{ for all } v \in V\}$. Then $V \cap V^\perp = 0$, for $v^t v = 0$ implies $v = 0$ for $v$ in $\mathbb{R}^m$. Moreover, the dimension of $V^\perp$ is such that $\dim V + \dim V^\perp = m$. Indeed, this follows from the rank-nullity theorem, given the interpretation that $V^\perp$ is the solution space of $A^t w = 0$, where $A$ is a matrix of size $m \times \dim V$ whose column space is $V$.

Putting together a basis of $V$ with a basis of $V^\perp$ therefore gives a basis of $\mathbb{R}^m$. In other words, each element $x$ of $\mathbb{R}^m$ has a unique expression of the form $x = v + v'$ with $v$ in $V$ and $v'$ in $V^\perp$. The association $x \mapsto v$ is a linear transformation from $\mathbb{R}^m$ to $V$ (or to $\mathbb{R}^m$,

[2]This so-called slope-intercept form of the line would not be appropriate if we expect the line of best fit to be vertical. But since, following convention, we plot the "independent variable" on the $x$-axis, and there are many different values of this variable, a vertical line is ruled out, and we are justified in our choice of the form of equation for the required line.

if one prefers). It is characterized by the properties that it is identity on $V$ and vanishes on $V^\perp$. It is called the *orthogonal projection to the subspace* $V$.

4.3. **A formula for the orthogonal projection.** Now suppose that $V$ is specified for us as being the column space of a $m \times n$ matrix $A$ with linearly independent columns. The matrix $A$ of course determines the orthogonal projection—call it $P$—on to $V$. The question now is: how do we write the matrix of $P$ (with respect to the standard basis of $\mathbb{R}^m$) given $A$?

The answer is:

$$\boxed{P = A(A^tA)^{-1}A^t} \tag{6}$$

For the proof, we first observe that the $A^tA$ is an invertible $n \times n$ matrix, so that the inverse in the formula makes sense. Suppose that $(A^tA)x = 0$ for some $x$ in $\mathbb{R}^n$. Then, multiplying by $x^t$ on the left, we get $(x^tA^t)(Ax) = 0$. But this means $||Ax|| = 0$, so $Ax = 0$. Since the columns of $A$ are linearly independent, this in turn means $x = 0$. This proves that the endomorphism of $\mathbb{R}^n$ represented by $A^tA$ is injective and so also bijective. Thus $A^tA$ is invertible.

For an element $v$ of $V$ we have $v = Ax$ for some $x$ in $\mathbb{R}^n$, so that $Pv = A(A^tA)^{-1}A^t(Ax) = A(A^tA)^{-1}(A^tA)x = Ax = v$. And for an element $w$ of $V^\perp$ we have $A^tw = 0$ (because the $m$ entries of $A^tw$ are precisely the inner products of $w$ with the columns of $A$ which span $V$), and so $Pw = 0$. This proves the formula.

4.4. **Remarks.** We make various remarks about the argument in the preceding subsection.

(1) Note that if $m = n$, then we get $P = A(A^tA)^{-1}A^t = AA^{-1}(A^t)^{-1}A^t =$ identity, which makes sense.
(2) In the course of the proof we have shown the following: the map $A^t$ restricted to the image of $A$ is injective (where $A$ is a real matrix). Indeed, if $A^tAx = 0$, then $x^tA^tAx = 0$ and so $||Ax|| = 0$ and $Ax = 0$.
(3) Observe directly (without recourse to formula (6)) that the matrix $P$ representing (with respect to the standard basis of $\mathbb{R}^m$) the orthogonal projection to any subspace of $\mathbb{R}^m$ is symmetric and satisfies $P^2 = P$.
(4) Suppose that $P$ is an $m \times m$ symmetric matrix such that $P^2 = P$. Then $P$ represents with respect to the standard basis of $\mathbb{R}^m$ the orthogonal projection onto its column space.
(5) If in §4.3 the columns of $A$ are orthonormal, then $A^tA$ is the identity matrix, so formula (6) reduces to $P = AA^t$. This motivates the Gram-Schmidt orthogonalization procedure for computing an orthonormal basis for a subspace of $\mathbb{R}^m$ starting from any given basis for that subspace.

4.5. **Approximate solution to an overdetermined linear system.** Motivated by the need to find the line of best fit and armed with the formula of the previous subsection, we now proceed to give an approximate solution to an overdetermined linear system of equations. Suppose that we want to solve $Ax = b$ for $x$, where $A$ is an $m \times n$ matrix with linearly independent columns (so, in particular, $m \geq n$). In general $b$ may not be in the column space of $A$. We replace $b$ by its orthogonal projection on to the column space of $A$, which by the formula of the previous subsection is $A(A^tA)^{-1}A^tb$. We get $Ax = A(A^tA)^{-1}A^tb$. But since $A$ has linearly indepedent columns, we can cancel the leading $A$ from both sides, so we get

$$\boxed{x = (A^tA)^{-1}A^tb} \tag{7}$$

4.6. **Illustration.** As an illustration of the method just described, let us work out the line that best fits the three points $(1,1)$, $(2,3)$ and $(3,3)$. The slope $m$ and $y$-intercept $c$ are obtained by an application of (7) as follows.

$$\begin{pmatrix} m \\ c \end{pmatrix} = (A^tA)^{-1}A^tb \quad \text{where } A = \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix} \text{ and } b = \begin{pmatrix} 1 \\ 3 \\ 3 \end{pmatrix}. \text{ We have } A^tA = \begin{pmatrix} 14 & 6 \\ 6 & 3 \end{pmatrix}.$$

Computation of $(A^tA)^{-1}$ (see §2.3):

$$\left(\begin{array}{cc|cc} 14 & 6 & 1 & 0 \\ 6 & 3 & 0 & 1 \end{array}\right) \longrightarrow \left(\begin{array}{cc|cc} 2 & 0 & 1 & -2 \\ 6 & 3 & 0 & 1 \end{array}\right) \longrightarrow \left(\begin{array}{cc|cc} 2 & 0 & 1 & -2 \\ 0 & 3 & -3 & 7 \end{array}\right) \longrightarrow \left(\begin{array}{cc|cc} 1 & 0 & 1/2 & -1 \\ 0 & 1 & -1 & 7/3 \end{array}\right)$$

Thus we have:

$$\begin{pmatrix} m \\ c \end{pmatrix} = \begin{pmatrix} 1/2 & -1 \\ -1 & 7/3 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ 3 \end{pmatrix} = \begin{pmatrix} 1/2 & -1 \\ -1 & 7/3 \end{pmatrix} \begin{pmatrix} 16 \\ 7 \end{pmatrix} = \begin{pmatrix} 1 \\ \frac{1}{3} \end{pmatrix}$$

Thus the line that best fits the points $(1,1)$, $(2,3)$, and $(3,3)$ is $y = x + \frac{1}{3}$.

4.7. **Normal form of a linear system.** Consider a system $Ax = b$ of linear equations over the real numbers. Here $A$ is a real $m \times n$ matrix, $x$ a $n \times 1$ matrix of indeterminates, and $b$ a $m \times 1$ real matrix. The following system is called the *normal form* of $Ax = b$:

$$\boxed{A^tAx = A^tb}$$

This terminology is justified by the following observations:

- The normal form has a solution even if the original one doesn't, for the column space of $A^t$ equals that of $A^tA$. (Proof: Suppose that $A^tx = c$. Write $x = Ay + z$ with $z$ perpendicular to the range space (column space) of $A$. Thus $z^tAu = 0$ for all $u$ in $\mathbb{R}^n$, so $z^tA = 0$ or $A^tz = 0$. Substituting for $x$ in $A^tx = c$, we obtain $A^t(Ay + z) = c$. But $A^tz = 0$, so we obtain $A^tAy = c$.)
- Clearly any solution of the system $Ax = b$ is also a solution of its normal form. Moreover, if $Ax = b$ has a solution, then its solution set and that of its normal form are identical. (Proof: The solutions of the homogeneous systems $A^tAx = 0$ and $Ax = 0$ are identical: $A^tAx = 0$ implies $0 = x^tA^tAx = ||Ax||^2$, so $Ax = 0$.)

The normal form is thus a good proxy for the original. In case $A$ has linearly independent columns, $A^tA$ is invertible and the normal form has a unique solution given by (7).

## 5. The spectral theorem for a real symmetric matrix

Let $A$ be an $n \times n$ real symmetric matrix (RSM, for short). Then:

- *The eigenvalues of $A$ are all real.*
- *Eigenvectors of $A$ corresponding to different eigenvalues are orthogonal.*

- *There exists an orthonormal basis for $\mathbb{R}^n$ each member of which is an eigenvector for $A$. More precisely, if $\lambda_1$, ..., $\lambda_n$ be the eigenvalues of $A$ listed with multiplicity in any given order, and $D$ the diagonal matrix whose diagonal entry on the $i^{th}$ row is $\lambda_i$, then there exists a special orthogonal $n \times n$ matrix $g$ such that* $g \in \mathrm{SO}(n)$ $D = \mathrm{diag}(\underline{\lambda})$

$$\boxed{g^t A g = D}$$

The assertions above are collectively called the *spectral theorem*. In the proof of the third assertion, we will use the following proposition.

**Proposition 12.** *Let $A$ be a symmetric $n \times n$ matrix and $W$ a subspace of $\mathbb{R}^n$ that is $A$-*stable *(that is, $Aw$ belongs to $W$ whenever $w$ does). Then*

*(1) $W^\perp$ is also $A$-stable;*
*(2) the matrix of $A|_W$ with respect to an orthonormal basis of $W$ is symmetric.*

*Proof.* For $y \in W^\perp$ and $w \in W$, we have $w^t A y = y^t A^t w = y^t (Aw) = 0$, so (1) holds. For $w_1$, ..., $w_d$ an orthonormal basis for $W$, the entry in position $(i, j)$ of the matrix of $A|_W$ with respect to this basis is $w_i^t A w_j$. We have $w_i^t A w_j = w_j^t A^t w_i = w_j^t A w_i$, so (2) is proved. □

Now we prove the spectral theorem. We consider the assertions one by one.

Let $\lambda$ be an a priori complex eigenvalue of $A$. Let $v \neq 0$ in $\mathbb{C}^n$ be such that $Av = \lambda v$. We first multiply this equation on the left by $v^*$ to get $v^* A v = \lambda v^* v$. Next we take the conjugate transpose of the same equation and then multiply it by $v$ on the right to get $v^* A^* v = \bar{\lambda} v^* v$. But $A$ being an RSM, we have $A^* = A$, so that the left sides of the two resulting equations are the same. Equating their right sides, we get $\lambda v^* v = \bar{\lambda} v^* v$. Since $v \neq 0$, we have $v^* v \neq 0$ and $\lambda = \bar{\lambda}$, so $\lambda$ is real. $X^*$ = conjugate transpose of $X$

Let $v$ and $w$ be eigenvectors corresponding respectively to distinct eigenvalues $\lambda$ and $\mu$: that is, $Av = \lambda v$ and $Aw = \mu w$. Multiplying on the left the first by $w^*$ and the second by $v^*$ we get $w^* A v = \lambda w^* v$ and $v^* A w = \mu v^* w$. The left sides of these two equations are conjugate transposes of each other. And therefore so are their right sides. Thus $(\lambda - \mu) w^* v = 0$. Since $\lambda \neq \mu$, we have $w^* v = 0$, which means $w$ is orthogonal to $v$.

For the proof of the last statement, let us first observe that, for orthogonal $g$, the statements $g^t A g = D$ and $Ag = gD$ are equivalent, and that the latter statement is just saying that the columns of $g$ form an orthonormal basis of eigenvectors for $A$ with eigenvalues the diagonal entries in the corresponding rows of $D$.

Let $\lambda$ be an eigenvalue of $A$. By the first part, we know $\lambda$ is real. We may choose $v \neq 0$ in $\mathbb{R}^n$ such that $Av = \lambda v$. Such a $v$ exists because the homogeneous linear system $(A - \lambda I)v = 0$ has a non-trivial solution in $\mathbb{R}^n$ (since $A - \lambda I$ is singular and has real entries). Let $P$ be the subspace of vectors perpendicular to $v$.

By the proposition, $P$ is $A$-stable and the matrix of $A|_P$ with respect to an orthonormal basis of $P$ is symmetric. By induction on $n$, there exists an orthonormal basis of $P$ each member of which is an eigenvector for $A|_P$ (and hence of $A$). Putting these together with $v$ we get an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of $A$. Let $g$ be the matrix whose columns are these basis in any given order. Then $g$ is orthogonal. Thus there exists orthogonal $g$ such that $g^t A g = D$. Finally, multiplying one of the columns of $g$ by $-1$ if necessary (that is, replacing one of the eigenvectors by its negative), we may take $g$ to be special orthogonal.

5.1. **Symmetric positive definite and semi-definite matrices.** Let $A$ be an $n \times n$ RSM. The assignment $(v, w) \mapsto v^t A w$ defines on $\mathbb{R}^n$ a symmetric bilinear form. We call $A$ *positive semi-definite* if $v^t A v \geq 0$ for all $v$; we call it *positive definite* if moreover $v^t A v = 0$ only for $v = 0$. We observe:

(1) These notions are preserved under the passage $A \leftrightarrow g^t A g$ with $g$ orthogonal.
(2) These notions extend to linear endomorphisms of finite dimensional inner product spaces (in particular, to those of subspaces of $\mathbb{R}^n$): an endomorphism is said to have the property if its matrix with respect to an orthonormal basis has the corresponding property. The choice of the orthonormal basis is immaterial by (1).
(3) As is easily seen, a diagonal matrix is positive definite (respectively, semi-definite) if and only if all its diagonal entries are positive (respectively, non-negative). Together with (1) and the spectral theorem, this gives: $A$ is positive definite if and only if all its eigenvalues are positive; positive semi-definite if and only if all its eigenvalues are non-negative.
(4) If $A$ is positive semi-definite, then $B^t A B$ is also symmetric positive semi-definite, for $B$ any real matrix of size $n \times m$. Further, if $A$ is positive definite, then the restriction $B^t A B|_{\text{range}(B^t)}$ of $B^t A B$ to the range of $B^t$ is positive definite (here $B^t A B$ denotes the endomorphism of $\mathbb{R}^m$ whose matrix with respect to the standard basis is $B^t A B$): indeed if $z^t B^t A B z = 0$ for $z = B^t y$ for some $y$, then, since $A$ is positive definite, $Bz = BB^t y = 0$, which in turn means $y^t B B^t y = 0$ or $||B^t y|| = 0$, so $z = 0$.
(5) An important special case of (4) is the following. Let $W$ be a subspace of $\mathbb{R}^n$ and $P$ the $n \times n$ matrix representing with respect to the standard basis the orthogonal projection on to $W$. (A formula for $P$ is given in §4.3.) Then $P$ is symmetric and $\text{range}(P) = W$. So, by (4), $PAP$ is symmetric positive semi-definite if $A$ is so; morevoer $PAP|_W$ is positive definite if $A$ is so. In particular, the principal submatrices of a positive definite (respectively, semi-definite) matrix are themselves positive definite (respectively, semi-definite) RSMs.
(6) Combining (5) with (3) shows that the determinants of all principal submatrices of $A$ are positive (respectively, non-negative) if $A$ is positive definite (respectively, semi-definite). Conversely, if the determinants of the top left corner submatrices of $A$ of sizes $1 \times 1, \ldots, n-1 \times n-1, n \times n$ are all positive, then $A$ is positive definite: see Exercise 6. (If we assume that these determinants are only non-negative, is $A$ semi-definite? Again see Exercise 6.)
(7) A positive semi-definite RSM has a unique positive semi-definite RSM as $n^{\text{th}}$ root, for $n$ a positive integer. Every RSM has a unique RSM as $n^{\text{th}}$ root, for $n$ an odd positive integer. See Exercise 8.

## 6. Singular Value Decomposition (SVD)

This is a version of spectral theorem for a non-square matrix. Let $A$ be an $m \times n$ matrix. We think of $A$ also as a linear transformation from $\mathbb{R}^n \to \mathbb{R}^m$, the columns of $A$ being the images in order of the standard basis elements of $\mathbb{R}^n$. SVD is about nice bases for the domain $\mathbb{R}^n$ and the codomain $\mathbb{R}^m$ with respect to which the matrix of the linear transformation $A$ has a particularly simple form.

the
out?

We first find a nice basis for $\mathbb{R}^n$. Observe that $A^t A$ is a positive semi-definite $n \times n$ real symmetric matrix. By the spectral theorem for a real symmetric matrix (§5), there exists an orthonormal basis, say $u_1, \ldots, u_n$, of $\mathbb{R}^n$, each member of which is an eigenvector for $A^tA$. Since $A^tA$ is positive semi-definite, its eigenvalues are all non-negative, and we may write $A^tAu_i = \sigma_i^2 u_i$ with $\sigma_i$ non-negative (the $\sigma_i$ are uniquely determined). After a rearrangement of the basis $u_i$, we may assume that $\sigma_i$ are strictly positive for $1 \leq i \leq r$, and zero for $i > r$, where $r$ is the rank of $A^tA$ (and also of $A$).

Orthonormal basis for the domain

Next we observe that orthogonality is preserved by $A$. More precisely, we have:

- For $i \neq j$, $Au_i$ and $Au_j$ are orthogonal: indeed
$$(Au_j)^t Au_i = (u_j^t A^t)Au_i = u_j^t(A^tAu_i) = u_j^t(\sigma_i^2 u_i) = \sigma_i^2(u_j^t u_i) = 0$$
- $||Au_i|| = \sigma_i$, for
$$(Au_i)^t Au_i = (u_i^t A^t)Au_i = u_i^t(A^tAu_i) = u_i^t(\sigma_i^2 u_i) = \sigma_i^2(u_i^t u_i) = \sigma_i^2$$

Taking a cue from the above observations, we set $v_i := Au_i/\sigma_i$, for $1 \leq i \leq r$. Then $v_1$, $\ldots$, $v_r$ are orthonormal. We choose $v_{r+1}, \ldots, v_m$, so that $v_1, \ldots, v_m$ form an orthonormal basis of $\mathbb{R}^m$.

Orthonormal basis for the codomain

Finally we express in a matrix equation the action of the linear transformation $A$. Let $U := (u_1, u_2, \ldots, u_n)$ be the $n \times n$ matrix whose columns in order are the basis vectors $u_1$, $\ldots$, $u_n$, and similarly $V := (v_1, v_2, \ldots, v_m)$ the $m \times m$ matrix whose columns in order are the basis vectors $v_1, \ldots, v_m$. Then
$$AU = \Sigma V$$
where $\Sigma$ is a $m \times n$ matrix all of whose entries are zero except those at positions $(1,1)$, $(2,2)$, $\ldots$, $(r,r)$, which respectively are $\sigma_1, \sigma_2, \ldots, \sigma_r$. Since the $u_i$ form an orthonormal basis, the inverse of the matrix $U$ is $U^t$, multiplying by which the last equation on the right, we get:

This is the SVD.

$$\boxed{A = V\Sigma U^t} \tag{8}$$

The numbers $\sigma_1, \ldots, \sigma_r$ being the positive square roots of the positive eigenvalues of $A^tA$, they are unique. Thus the matrix $\Sigma$ is uniquely determined if we impose the condition that $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r$.

Uniqueness

## Exercise Set 3

(1) Consider the $4 \times 4$ RSM all of whose entries are $1$. What are its eigenvalues? Find an orthonormal basis (for $\mathbb{R}^4$) consisting of eigenvectors of this matrix.

(2) The range of a real symmetric matrix is the orthogonal complement of its kernel.

(3) For any real $m \times n$ matrix $A$ and any integer $p \geq 0$, the range of $A(A^tA)^p$ is the range of $A$; the range of $(A^tA)^pA^t$ is the range of $A^t$. In particular, the ranks of all these matrices are the same. Specializing to the case of a symmetric matrix $A$, we conclude that the ranges of all its powers $A^n$, for $n \geq 1$, coincide. In particular $\operatorname{rank} A = \operatorname{rank} A^2 = \ldots$. Zero is the only nilpotent real symmetric matrix.

(4) If $A$ is an invertible RSM and $A^2 = A$, what is $A$?

(5) Suppose that $A$ is a $n \times n$ real symmetric matrix with distinct eigenvalues. The number of elements $g$ with $g^tAg$ diagonal and $g$ orthogonal is $2^n n!$. For a fixed diagonal, the corresponding number is $2^n$. The corresponding numbers of special orthogonal $g$ are $2^{n-1}n!$ and $2^{n-1}$.

(6) Prove the assertion in item (6) of §5.1 (the converse part).

(7) Given commuting real symmetric matrices $A$ and $B$, there exists orthogonal $g$ such that both $g^tAg$ and $g^tBg$ are diagonal.
(8) Prove the assertions in item (7) of §5.1.
(9) Can a real $m \times n$ matrix $A$ be recovered from $AA^tA$? (Hint: Can an RSM be recovered from its cube?)
(10) State and prove the spectral theorem for normal matrices. Recall that a complex $n \times n$ matrix $N$ is called *normal* if it commutes with its adjoint $N^*$.

## 7. Min-max characterisation of the eigenvalues of a RSM

Let $A$ be a $n \times n$ RSM.

7.1. **Alternative proof of the spectral theorem.** The crucial starting point in the proof of the spectral theorem in §5 above is that $A$ has a real eigenvalue. This was proved by showing that any complex eigenvalue (which exists by the fundamental theorem of algebra) must be real. Here we give a different proof, using some basic topology and calculus but avoiding the use of the fundamental theorem of algebra, that $A$ has a real eigenvalue.

Consider the continuous function $f(x) = x^tAx$ on the sphere $S^{n-1} := \{x \in \mathbb{R}^n \mid ||x|| = 1\}$. Since $S^{n-1}$ is compact, this attains a maximum, say at $v \in S^{n-1}$. Fix arbitrarily a unit vector $w$ perpendicular to $v$. Consider the unit circle $S^1$ in the plane spanned by $v$ and $w$. This circle is parametrized as $v \cos t + w \sin t$, $t \in \mathbb{R}$, and the function $f$ restricted to $S^1$ is given in terms of $t$ by:

$$v^tAv\cos^2 t + v^tAw \sin t \cos t + w^tAv \sin t \cos t + w^tAw \sin^2 t$$

Since $A$ is symmetric, we have $v^tAw = w^tAv$, and so the function value above may be rewritten as:

$$v^tAv\cos^2 t + 2v^tAw \sin t \cos t + w^tAw \sin^2 t$$

The smooth function $f(t)$ attains a maximum at $t = 0$, and so its derivative vanishes at $t = 0$. As an easy calculation shows, this means that $v^tAw = 0$, or in other words that $Av$ is perpendicular to $w$.

Since this is true for any $w$ that is perpendicular to $v$, we conclude that $Av$ must be a multiple of $v$, and so $v$ is an eigenvector for $A$.

The rest of the proof remains the same: consider the linear transformation $x \mapsto Ax$ on the subspace $P$ of vectors in $\mathbb{R}^n$ that are perpendicular to $v$, observe that it is represented by a symmetric matrix with respect to any orthonormal basis of $P$, and invoke induction on the size $n$ of $A$.

7.2. **A characterization of the eigenvalues of $A$.** Let $\lambda_1 \geq \ldots \geq \lambda_n$ be the eigenvalues of $A$. The eigenvalue $\lambda_k$ may be characterised thus:

$$\boxed{\lambda_k = \max_{\substack{W \subseteq \mathbb{R}^n \\ \dim W = k}} \ \min_{\substack{x \in W \\ ||x|| = 1}} \ x^tAx}$$

Let $v_1, \ldots, v_n$ be an orthonormal basis for $\mathbb{R}^n$ with $v_i$ being an eigenvector with eigenvalue $\lambda_i$ for $A$ for $1 \leq i \leq n$. The minimum value of $x^tAx$ for $x$ of norm one in the span of $v_1, \ldots, v_k$ is easily seen to be $\lambda_k$. On the other hand, any subspace $W$ of $\mathbb{R}^n$ of dimension $k$ intersects non-trivially the $n-k+1$ dimensional subspace generated by $v_k, \ldots,$

$v_n$, and so the minimum value of $x^tAx$ for $x$ of norm one in $W$ is at most $\lambda_k$. The above characterisation is thus proved.

7.3. **An application: Sylvester's criterion for positive definiteness.** As an application of the characterisation in §7.2, we prove the following sufficiency criterion due to Sylvester for positive definiteness (see item (6) in §5.1; also item (6) in Exercise set 3): $A$ is positive definite if all the top left corner submatrices of $A$ of sizes $1\times 1$, $2\times 2$, ..., $(n-1)\times(n-1)$, and $n\times n$ have positive determinant (note that the top left corner $n\times n$ submatrix is $A$ itself). As seen in item (5) of §5.1, the condition is clearly necessary: in fact, all principal minors of a positive definite matrix are positive.

As seen in item (3) of §5.1, $A$ is positive definite if and only if all its eigenvalues are positive. Suppose that $A$ satisfies the hypothesis in Sylvester's criterion. To show that it is positive definite, it is enough to show that $\lambda_{n-1}$ is positive (where as in §7.2 above $\lambda_1 \geq \ldots \geq \lambda_n$ are the eigenvalues of $A$ arranged in weakly decreasing order), for then $\lambda_1$, ..., $\lambda_{n-2}$ are also positive, and so is $\lambda_n = \det A/\lambda_1 \ldots \lambda_{n-1}$.

By induction on the size $n$ of $A$, we conclude that the top left corner $(n-1)\times(n-1)$ submatrix of $A$ is positive definite. All its eigenvalues are therefore positive. This means (proof?) that the least value is positive of $x^tAx$ as $x$ runs over vectors of norm one in the span of the first $n-1$ standard basis vectors $e_1$, ..., $e_{n-1}$. This in turn means $\lambda_{n-1}$ is positive by the characterization in §7.2.

## Exercise Set 4

(1) Let $\lambda_1 \leq \ldots \leq \lambda_n$ be the eigenvalues (which are all real) of a RSM $A$, arranged in weakly increasing order. Observe that

$$\boxed{\lambda_k = \min_{\substack{W \subseteq \mathbb{R}^n \\ \dim W = k}} \max_{\substack{x \in W \\ ||x|| = 1}} x^tAx}$$

(2) Suppose that $A$ is a $n\times n$ real skew-symmetric matrix. Observe that $v^tAv = 0$ for all $v \in \mathbb{R}^n$. Prove that this property characterises skew-symmetry: if $A$ is a $n\times n$ real matrix such that $v^tAv = 0$ for all $v \in \mathbb{R}^n$, then $A$ is skew-symmetric.

(3) Let $A$ be a real $n\times n$ matrix (not necessarily symmetric). Show the existence of a vector $v$ in $\mathbb{R}^n$ such that $v^tAw + w^tAv = 0$ for all vectors $w$ in $\mathbb{R}^n$ perpendicular to $v$. In fact, there exist at least two linearly independent such vectors $v$ for $n \geq 2$. (Observe that $v^tAw + w^tAv = 0$ for all $v$, $w$ in case $A$ is skew-symmetric.)