5.3 Review of systems of linear equations

A linear equation in n unknowns $x_1, \ldots, x_n$ is of the form $a_1x_1 + \ldots a_nx_n = b$. Given m such equations, we can write the ith equation as $a_{i1}x_1 + \ldots a_{in}x_n = b_i$. We will seek to solve these systems of linear equations.

Example a system of 3 equations in 3 unknowns and its solution is

$$\left\{\begin{array}{rcrcrcr} 4x_1 & + & x_2 & + & 2x_3 & = & 24 \\ 2x_1 & - & x_2 & - & 2x_3 & = & -6 \\ -x_1 & + & 2x_2 & - & x_3 & = & -4 \end{array}\right. \Longrightarrow \left\{\begin{array}{l} x_1 = 3 \\ x_2 = 2 \\ x_3 = 5 \end{array}\right.$$

□

These systems can be represented as a matrix equation $\mathbf{Ax} = \mathbf{b}$ where $\mathbf{A}$ is the $m \times n$ matrix of coefficients, a_{ij}, $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ is the n-dimensional column vector of unknowns and $\mathbf{b}$ is a vector of dimension m.

Example cont. In the example above, $\mathbf{A} = \begin{bmatrix} 4 & 1 & 2 \\ 2 & -1 & -2 \\ -1 & 2 & -1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 24 \\ -6 \\ -4 \end{bmatrix}$ □

We'll initially look at systems of n equations and n unknowns. Systems with $m < n$ are known as *under-determined* as there are less equations than unknowns while systems with $m > n$ are *over-determined* with more equations than there are unknowns.

When A is non-singular (so A^{-1} exists), the system has a unique solution given by $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$.

Recall that $\mathbf{A}$ is nonsingular if and only if:

(*i*) inverse matrix $\mathbf{A}^{-1}$ exists; or
(*ii*) $\det(\mathbf{A}) \neq 0$; or
(*iii*) $\text{rank}(\mathbf{A}) = m$, or
(*iv*) $\mathbf{Ax} \neq \mathbf{0}$ for any vector $\mathbf{x} \neq \mathbf{0}$, or
(*v*) $\text{range}(\mathbf{A}) = \mathbb{R}^m$, or
(*vi*) $\text{null}(\mathbf{A}) = \{\mathbf{0}\}$.

If $\mathbf{A}$ is singular, the system may have infinitely many solutions or no solutions at all, depending on $\mathbf{b}$.

Example If $\mathbf{A} = \begin{bmatrix} 2 & 3 \\ 4 & 6 \end{bmatrix}$, $\mathbf{Ax} = \mathbf{b}$ has no solution if $\mathbf{b} \notin \text{range}(\mathbf{A})$ or infinitely many solutions when $\mathbf{b} \in \text{range}(\mathbf{A})$. Thus, when $\mathbf{b} = \begin{bmatrix} 4 \\ 7 \end{bmatrix}$ there is no solution, while when $\mathbf{b} = \begin{bmatrix} 4 \\ 8 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} \gamma \\ \frac{2}{3}(2-\gamma) \end{bmatrix}$ is a solution for any real γ. □

6 Solving linear equations

In principle, all we need to do to solve the system of equations $\mathbf{Ax} = \mathbf{b}$ is find the inverse of $\mathbf{A}$, $\mathbf{A}^{-1}$. Then $\mathbf{Ax} = \mathbf{b} \implies \mathbf{A^{-1}Ax} = \mathbf{A^{-1}b} \implies \mathbf{x} = \mathbf{A^{-1}b}$. In practice, however, things are more complicated. First, $\mathbf{A}$ only has an inverse if it is square (so $m = n$) and $det(A) \neq 0$. In most cases, $m \neq n$ and often even when $m = n$, $det(A) = 0$ is not unusual. Second, supposing that A is indeed square, m and n are often large (10^4 is common, as are much larger values). In these cases, even calculating $det(A)$ is a hugely expensive and complex computational task while finding $\mathbf{A}^{-1}$ is even harder.

We'll initially concentrate on easily solvable systems and look at how we can coerce other systems into a form where they (or some close approximation) too are easily solvable.

6.1 Easily solvable systems 1: Diagonal matrix

All the simple systems we consider here are assumed to be square, so $m = n$. We want to solve $\mathbf{Ax} = \mathbf{b}$.

$\mathbf{A}$ is *diagonal* all entries the off-diagonal are zero. That is $a_{ij} = 0$ when $i \neq j$. So to specify a diagonal matrix, we need only specify the n diagonal elements. We can thus use the simplifying notation, $\mathbf{A} = \text{diag}\{a_1, \ldots, a_n\}$.

When A is diagonal, $x_i = \frac{b_i}{a_i}$ for all $i = 1, \ldots, n$. That is, $\mathbf{A}^{-1} = \text{diag}\{\frac{1}{a_1}, \ldots, \frac{1}{a_n}\}$. Or, to use less compact notation:

$$\mathbf{A} = \begin{bmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{bmatrix} \Rightarrow \mathbf{A}^{-1} = \begin{bmatrix} \frac{1}{a_1} & & \\ & \ddots & \\ & & \frac{1}{a_n} \end{bmatrix}.$$

6.2 Easily solvable systems 2: Triangular matrix

A matrix is *lower triangular* when all entries above the main diagonal are 0. That is, $\mathbf{A}$ is lower triangular if and only if $a_{ij} = 0$ when $i < j$. Similarly, a matrix is *upper triangular* when all entries above the main diagonal are 0 ($a_{ij} = 0$ for $i > j$). Lower triangular is also called left triangular, and upper called right triangular, for obvious reasons. E.g., a lower triangular matrix:

$$\mathbf{A} = \begin{bmatrix} a_{11} & 0 & \ldots & 0 \\ a_{21} & a_{22} & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix}.$$

The system $\mathbf{Ax} = \mathbf{b}$ is easy to solve for triangular $\mathbf{A}$ and it does not require that we calculate the inverse of $\mathbf{A}$.

For the lower triangular matrix, the solution is given by

$$x_i = \frac{1}{a_{ii}} \left(b_i - \sum_{j=1}^{i-1} a_{ij} x_j \right),$$

so that

$$x_1 = \frac{b_1}{a_{11}};\ x_2 = \frac{b_2 - a_{21}x_1}{a_{22}};\ \ldots;\ x_n = \frac{b_n - a_{n1}x_1 - \ldots - a_{n-1,n}x_{n-1}}{a_{nn}}.$$

A similar simple formula is available for the upper triangular case, this time working backwards from x_n:

$$x_n = \frac{b_n}{a_{nn}}$$

and

$$x_i = \frac{1}{a_{ii}} \left(b_i - a_{i,i+1}x_{i+1} - \ldots - a_{i,n}x_n\right) \text{ for } i = n-1, \ldots, 1.$$

The method of Gaussian elimination, or row reduction, which we assume you have seen before transforms the matrix $\mathbf{A}$ into a triangular one to solve the system. This method is reviewed and discussed in Section **??**

6.3 Easily solvable systems 3: Orthonormal or orthogonal matrix

Matrix $\mathbf{A}$ is *orthogonal* or *orthonormal* if the columns of $\mathbf{A}$ are mutually orthogonal unit vectors.

That is, $\mathbf{A} = [\mathbf{a}_1\ \mathbf{a}_2\ \ldots\ \mathbf{a}_n]$ where $\mathbf{a}_i = [a_{i1}\ a_{i2}\ \ldots\ a_{in}]^\mathsf{T}$ are unit vectors and the set $\{\mathbf{a}_1\ \mathbf{a}_2\ \ldots\ \mathbf{a}_n\}$ is mutually orthogonal (so $\mathbf{a}_i \cdot \mathbf{a}_j = 0$ for $i \neq j$).

When $\mathbf{A}$ is orthonormal, $\mathbf{A}^{-1} = \mathbf{A}^T$. This result is true since

$$\mathbf{A}^\mathsf{T}\mathbf{A} \equiv \begin{bmatrix} \mathbf{a}_1^\mathsf{T} \\ \vdots \\ \mathbf{a}_n^\mathsf{T} \end{bmatrix} [\mathbf{a}_1\ \mathbf{a}_2\ \ldots\ \mathbf{a}_n] = \mathbf{I}_n \equiv \mathrm{diag}\{1, 1, \ldots, 1\}.$$

Also check that $\mathbf{A}\mathbf{A}^\mathsf{T} = \mathbf{I}_n$: $\mathbf{A}\mathbf{A}^\mathsf{T}\mathbf{A} \equiv \mathbf{A}\underbrace{(\mathbf{A}^\mathsf{T}\mathbf{A})}_{\mathbf{I}_n} = \mathbf{A}$ and $\mathbf{A}\mathbf{A}^\mathsf{T}\mathbf{A} \equiv (\mathbf{A}\mathbf{A}^\mathsf{T})\mathbf{A}$

These properties can be taken as a definition of an orthonomal matrix: $\mathbf{A}^{-1} = \mathbf{A}^T$ if and only if $\mathbf{A}$ is orthonormal.

Thus, if $\mathbf{A}$ is orthonormal, the solution to $\mathbf{Ax} = \mathbf{b}$ is simply $\mathbf{x} = \mathbf{A}^T\mathbf{b}$.

Example: Find the solution to the set of equations

$$\begin{aligned} 0.48x_1 + 0.64x_2 + 0.60x_3 &= 3.56 \\ 0.36x_1 + 0.48x_2 - 0.80x_3 &= -1.08 \\ 0.80x_1 - 0.60x_2 &= -0.40 \end{aligned}$$

or

$$x_1 \overbrace{\begin{bmatrix} 0.48 \\ 0.36 \\ 0.80 \end{bmatrix}}^{\mathbf{a}_1} + x_2 \overbrace{\begin{bmatrix} 0.64 \\ 0.48 \\ -0.60 \end{bmatrix}}^{\mathbf{a}_2} + x_3 \overbrace{\begin{bmatrix} 0.60 \\ -0.80 \\ 0.00 \end{bmatrix}}^{\mathbf{a}_3} = \begin{bmatrix} 3.56 \\ -1.08 \\ -0.40 \end{bmatrix}.$$

So $\mathbf{A} = [\mathbf{a}_1\, \mathbf{a}_2\, \mathbf{a}_3]$.

Solution: By checking that $\mathbf{a}_i \cdot \mathbf{a}_j = 1$ for $i = j = 0$ for $i \neq j$, it is easy to see that $\mathbf{A}$ is orthonormal. So we have the solution

$$\mathbf{x} = \mathbf{A}^{\mathsf{T}}\mathbf{b} = \begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \mathbf{a}_3^T \end{bmatrix} \mathbf{b}$$

and

$$\begin{aligned} x_1 = \mathbf{a}_1^T\mathbf{b} &= 0.48 \cdot 3.56 - 0.36 \cdot 1.08 - 0.80 \cdot 0.40 \\ &= 1.7088 - 0.3888 - 0.3200 && = 1.0 \\ x_2 = \mathbf{a}_2^T\mathbf{b} &= 0.64 \cdot 3.56 - 0.48 \cdot 1.08 + 0.60 \cdot 0.40 \\ &= 2.2784 - 0.5184 + 0.2400 && = 2.0 \\ x_3 = \mathbf{a}_3^T\mathbf{b} &= 0.60 \cdot 3.56 + 0.80 \cdot 1.08 \\ &= 2.136 + 0.864 && = 3.0. \end{aligned}$$

$\square$

7 Factorising matrices

As we saw in the previous section, matrices with special forms are often much easier to work with than arbitrary matrices. The remainder of this part of the course is focused on how we can manipulate an arbitrary given matrix into a form that is convenient for a stated problem. This is known as *factorising* or *decomposing* matrices.

There are 3 factorisations we will study in various degrees of depth: LU-factorisation, Singular Value Decomposition (SVD) and QR decomposition. A brief summary is given here:

- **Elimination** (LU decomposition): $\mathbf{A} = \mathbf{LU}$
  - Lower triangular matrix  $\times$  Upper triangular matrix
- **Singular Value Decomposition** (SVD): $\mathbf{A} = \mathbf{UDV}^{\mathsf{T}}$

-  $\times$  diag(singular values) $\times$  Orthogonal (rows)
- Orthonormal columns in $\mathbf{U}$ and $\mathbf{V}$:
  the left and right **singular vectors**, respectively
- Left singular vector: an **eigenvector** of the square $m \times m$ matrix $\mathbf{A}\mathbf{A}^\top$
- Right singular vector: an **eigenvector** of the square $n \times n$ matrix $\mathbf{A}^\top\mathbf{A}$
- Singular value: the square root of an eigenvalue of $\mathbf{A}^\top\mathbf{A}$ (or $\mathbf{A}\mathbf{A}^\top$).

• **Orthogonalisation** (QR decomposition): $\mathbf{A} = \mathbf{QR}$

- Orthogonal matrix (columns)  $\times$