Left and right inverses

In response to a question raised during lecture, we will discuss some formal properties of inverses. In particular, we will see that the theorem that a left inverse of a matrix is also a right inverse is not a consequence of formal manipulation.

Definition. A *monoid* is a set S with an associative binary operation $\cdot$ and a (two-sided) identity element e. In other words, we have the following two axioms

1. For all $a, b, c \in S$, $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.
2. For all $a \in S$, $a \cdot e = e \cdot a = a$.

Example.

– Let S be the set of $n \times n$ matrices, then it is a monoid with the binary operation "multiplication" and identity element I_n.

– Let S be the set of all functions $f : \mathbb{N} \to \mathbb{N}$ (where $\mathbb{N} = \{1, 2, \cdots\}$ is the set of natural numbers). It is a monoid with binary operation "composition" and identity element the identity function $\texttt{id}(n) = n$.

– Let $S = \{e, a, b, c\}$ be a set with four elements. Define the binary operation by the table

$\cdot$	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

This is a monoid with identity element e.

Definition. Let S be a monoid. Suppose $a, b \in S$ satisfies $ab = e$, then b is a *right inverse* of a, and a is a *left inverse* of b.

The first definition of an inverse of a matrix given in lecture was in fact only a left inverse. The more usual definition of an inverse is a two-sided inverse, but they are equivalent by the theorems we looked at on linear systems. In a general monoid, the following is probably the strongest thing you can say.

Lemma. If S is a monoid and a has a left inverse, then a has at most one right inverse. Moreover, if it has a right inverse, then it is equal to the left inverse.

Proof. Let l be a left inverse of a, so $la = e$. Suppose $ab = ab' = e$, then $lab = lab' = l$, so $b = b' = l$. $\square$

Example. We show that the existence of a left inverse does not imply the existence of a right inverse. In the second example, consider the function $f(n) = 2n$. It has a left inverse

$$g(n) = \begin{cases} \frac{n}{2} & \text{if } n \text{ is even} \\ 1 & \text{if } n \text{ is odd} \end{cases}$$

In other words, $g(f(n)) = n$ for all $n \in \mathbb{N}$. But f cannot have a right inverse. We give three proofs of this:

– If h is a right inverse, then $f(h(n)) = n$ for all natural numbers n, but f only takes value in the even numbers, so this cannot hold if n is odd.

– If f has a right inverse, then it must be g by the lemma, but $f(g(1)) = 2 \neq 1$.

– Observe that f has multiple left inverses: just modify the behaviour of g on the odd numbers. This gives a third proof that f has no right inverse using the lemma.

More is true: a function has a left inverse if and only if it is injective (no two numbers get mapped to the same number). It has a right inverse if and only if it is surjective (every element of $\mathbb{N}$ is the value of the function at some point).

But we have the following theorem.

Theorem. If S is a monoid in which every element has a left inverse, then every element has a unique two-sided inverse in S (so S is a *group*).

Proof. Let $a \in S$, and let $b \in S$ be a left inverse of a. We have $ba = e$, so a is a right inverse of b. Since b has a left inverse, it follows from the lemma that a is a two-sided inverse of b, so $ab = ba = e$. Therefore, b is a two-sided inverse of a. $\square$

This theorem does not apply to the first example, since there are non-invertible matrices. It also does not apply to the sub-monoid of elements with left inverses, since it is not guaranteed that the left inverse of an element also has a left inverse.

There is a more general version of inverse called the Moore–Penrose pseudoinverse. Given an $n \times m$ matrix, it is an $m \times n$ matrix A^+ satisfying

- $AA^+A = A, A^+AA^+ = A^+$.
- AA^+ and A^+A are Hermitian.

This pseudoinverse always exists and is unique. The properties are symmetric in A and A^+, so $(A^+)^+ = A$. Using the singular value decomposition, we can write down a formula for the pseudoinverse. Recall that the motivation for introducing inverse was to solve $A\vec{x} = \vec{b}$. In general, this may be inconsistent or it may have infinitely many solutions. The next theorem says that $A^+\vec{b}$ is a good choice of a "solution".

Theorem. Consider the system of equations $A\vec{x} = \vec{b}$ in $\vec{x}$. Let $\vec{z} = A^+\vec{b}$, then

- If the system is consistent, then $\vec{z}$ is the solution with the minimal norm.
- If the system is inconsistent, then $\vec{z}$ is a least-square solution, i.e. the value of $\|A\vec{x} - \vec{b}\|$ is minimized if $\vec{x} = \vec{z}$. Moreover, $\vec{z}$ has the minimal norm among all least square solutions.