Generally, we like matrices because they're pretty much the cheapest meaningful general op you can get for $\mathbb{R}^n \to \mathbb{R}^m$ transformations. So if we can get away with using them, use them.

But how do we know if we can get away with using linear maps?

1 Linear maps are extrapolation

Recall that linear maps $T : V \to W$ are just functions with a restriction of **linearity**: $T(ax + by) = aT(x) + bT(y)$. This determines what T does to certain inputs $(ax + by)$ based on what T does to other inputs (x, y).

This is essentially an "extrapolation" rule: predefine some input-output pairs like $(x, T(x)), (y, T(y))$ and then linearity extrapolates outputs for all input-output pairs of the form $(ax + by, T(ax + by))$.

More generally, as a consequence of linearity, if you define n input-output pairs $(b_1, T(b_1)), ..., (b_n, T(b_n))$, then you can get an input-output pair $(u, T(u))$ for *every* input u that is a linear combination of the b_is: $u = c_1b_1 + ... + c_nb_n$. So if $b_1, ..., b_n$ form a basis for V, then $T(u)$ is defined for every possible input $u \in V$, since every $u \in V$ is a linear combination of a basis for V.

2 When is the "linearity" extrapolation ok?

The deciding factor on whether or not linear maps are acceptable for some task of mapping $V \to W$ vectors is if T's rule (linearity) extrapolates $T(x)$ for $x \notin \{b_1, ..., b_n\}$ reasonably closely to how we want our V vectors to be mapped to W.

A useful intuition you can use is thinking of a linear map $T : V \to W$ as simply **re-representing** each of its basis vectors $b_1, ..., b_n \in V$ as vectors in W, with all non-basis vectors continuing to use the same coefficients to combine the basis vectors, regardless of if they're the originals in V or if they're re-represented in W. For example, $x = 0.5b_1 + 0.5b_2$ is an average of b_1, b_2, and after transformation, $T(x)$ is still an average between $T(b_1), T(b_2)$: $T(x) = 0.5T(b_1) + 0.5T(b_2)$.

XOR, for example, becomes immediately obvious why it fails: if we try a linear model, we have a linear map $T : \mathbb{R}^2 \to \mathbb{R}$, our basis is $[1, 0], [0, 1]$. Our $(b_i, T(b_i))$ pairs are $([1, 0], 1)$ and $([0, 1], 1)$. For a non-basis input $[1, 1]$, it stays as a mix of $[1, 0]$ and $[0, 1]$ even in the codomain $\mathbb{R}$, yielding $T([1, 1]) = T([1, 0])+T([0, 1]) = 2$, even though it should be 0.