Here's how I understand rank-nullity intuitively.

1 Definition

Rank-nullity says for a linear map $T : V \to W$ over the field F,

$$\dim(V) = \dim(\mathrm{im}(T)) + \dim(\ker(T))$$

2 Short version

If you have a basis $b = \{b_1, ..., b_n\}$ for V, you have a $\dim(\mathrm{im}(T))$-dimensional subspace of V which is bijective across T to $\mathrm{im}(T)$. So $\dim(\mathrm{im}(T))$ of your basis vectors are important for T's output, and $\dim(V) - \dim(\mathrm{im}(T))$ of your basis vectors are not; it turns out that these unimportant vectors have **free coefficients** then, because their mapped versions live inside the span of the mapped useful vectors and thus can be cancelled out. Since you have $\dim(V) - \dim(\mathrm{im}(T))$ free coefficients, that's the dimensionality of your failure of injectivity; in other words, $\dim(\ker(T))$, so

$$\dim(V) - \dim(\mathrm{im}(T)) = \dim(\ker(T))$$

This is equivalent to rank-nullity.

3 Reasoning / full proof

Rank-nullity essentially says our "expressible dimensions" of the output of T are lost directly to the dimensions of injectivity failure.

More concretely, let $b = \{b_1, ..., b_n\}$ be the basis for V $(b_i \in V)$.

- Choose a basis such that $B_1 = \{b_1, ..., b_k\}$ $(k = \dim(\mathrm{im}(T)))$ forms a basis for the subspace of V that is bijective across T to the subspace $\mathrm{im}(T)$ of W. Note that $T(B_1)$ forms a basis for $\mathrm{im}(T)$.
- The rest of the V-basis set can go in $B_2 = \{b_{k+1}, ..., b_n\}$.

Notice that since $\mathrm{span}(T(B_1)) = \mathrm{span}(T(B_1 \cup B_2)) = \mathrm{im}(T)$, we know none of the $n - k$ vectors in $T(B_2)$ introduce any new information / dimensionality; any of them can be obtained from a combination in $T(B_1)$.

- Seems like there's already a connection to $\ker(T)$: injectivity failure! If we have redundant "building blocks" in $T(B_2)$ when mapping out $\mathrm{im}(T)$, then we no longer have only one way to move along any dimension.
- Specifically, if we map all $v \in \mathrm{im}(T)$ using linear combos of b to feed into T, we have multiple $\alpha = \{\alpha_1, ..., \alpha_n\}$ $(\alpha_i \in F)$ such that
  $$v = T\left(\sum_{i=1}^{n} \alpha_i b_i\right)$$
  for any particular v.
- This is because $\alpha_{k+1}, ..., \alpha_n$ are **free variables** for a fixed target output: any change in these coefficients can be cancelled out by changes to $\alpha_1, ..., \alpha_k$. Conceptualize $\alpha_1, ..., \alpha_k$ as the "responsible coefficients" which manage to keep the output the same by accommodating for the chaotic $\alpha_{k+1}, ..., \alpha_n$ which don't follow any rules.
  - Also note: there always exists a unique solution $\alpha_1, ..., \alpha_k$ for any set of $\alpha_{k+1}, ..., \alpha_n$ because $T\left(\sum_{i=k+1}^{n} \alpha_i b_i\right) \in \mathrm{im}(T)$ and $T(B_1)$ forms a basis for $\mathrm{im}(T)$ (which consequently means there is a unique set of coefficients to reach any member of $\mathrm{im}(T)$, and B_1's coefficients are $\alpha_1, ..., \alpha_k$).
- To make it concrete, we have that $\sum_{i=1}^{n} \alpha_i b_i$ can take on multiple values because of α changes, yet these changes don't reflect in $T\left(\sum_{i=1}^{n} \alpha_i b_i\right)$. This is exactly what it means to fail injectivity; two inputs lead to the same output.

(and a footnote on "$n - k$ free coefficients $= n - k$-dimensional kernel": each $T(b_j)$ for $b_j \in B_2$ can be represented as a combination of $T(B_1)$, so rearrange and get $T(b_j) - T\left(\sum_{b_i \in B_1} (\gamma_i b_i)\right) = 0_W$, so $b_j - \sum_{b_i \in B_1} (\gamma_i b_i)$ is in the kernel. Since each of these are LI for any $b_j \in B_2$, each $b_j \in B_2$ adds one dimension to the kernel.)

4 Matrix-vector framing

Tie this back to finding the nullspace of A.

In this scenario, $T : \mathbb{R}^n \to \mathbb{R}^m$ and $T(x) = Ax$.

- A basis for the domain is the standard basis $e_1, ..., e_n$ where e_i is a column vector of n zeros and a one at the ith element.
- We express any x as $x = \sum_{i=1}^{n} \alpha_i e_i$ where $\alpha_i \in \mathbb{R}$.

Then, when we find the nullspace, we're doing $A \cdot \left(\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \alpha_1 + \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} \alpha_2 + \cdots + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \alpha_n \right) = 0$. Or more simply, $A \cdot \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} = 0$. We solve for α values in terms of each other from the system of equations the matmul gives us.