Best approximate solutions

We were considering the (least squares) approximate solution to the equation

$$A\vec{x} = \vec{b}, \tag{1}$$

notably when there is no ambiguity, i.e., $\ker(A) = \{\vec{0}\}$. Recall that this kernel condition means that (1) has at most one solution (Exam 1, #8 again); in general $\dim(\ker(A))$ equals the number of free variables in the elimination process. If A is an $n \times m$ matrix, so $T_A : \mathbb{R}^m \to \mathbb{R}^n$, that implies $m \le n$, and that $\dim(\mathrm{im}(A)) = m$. If you like, that is rank-nullity.

The approximate solution is given conceptually as the solution of

$$A\vec{x} = \mathrm{proj}_{\,\mathrm{im}(A)}\vec{b}, \tag{2}$$

by projecting $\vec{b}$ onto the image. This is the true solution when $\mathrm{proj}_{\,\mathrm{im}(A)}\vec{b} = \vec{b}$, i.e., $\vec{b} \in \mathrm{im}(A)$. In the latter case, nothing is changed by our "talking funny".

The matrix of $\mathrm{proj}_{\,\mathrm{im}(A)}$ is $A(A^TA)^{-1}A^T$. A mouthful, maybe, but I sensed too much freaking out over the formula. If you want to see applications of the material, you are almost obliged to bear with complications. Were you being asked to remember it whole during lecture? (Were you even there?) I don't think so. Later? It is always wiser to approach such an expression slowly, figuring out its ingredients. What do you think your professors have to do? (See #4 of *To the Freshmen*, a document in the folder *What Hopkins students should know ...* that you are presumed to have read.) Note that (2) becomes:

$$A\vec{x} = A(A^TA)^{-1}A^T\vec{b}. \tag{3}$$

Our assumption on $\ker(A)$ allows us to cancel the A's on the left—clear?—to yield our approximate solution. By the way, the middle term $(A^TA)^{-1}$ cannot be expanded as $A^{-1}(A^T)^{-1}$ unless A is a square matrix, and it equals I_m when the columns of A are orthonormal. Also, you can infer that $A(A^TA)^{-1}A^T$ is the same for all matrices with the same image, for it is **the** matrix of that projection (w.r.t. the standard basis). Can you deduce that directly? Two $n \times m$ matrices have the same image if and only if ....

The main application was to curve-fitting. By that, we mean doing the above for the linear transformations

$$P_{m-1} \simeq \mathbb{R}^m \to \mathbb{R}^n,$$

given by evaluating polynomials at n distinct points of $\mathbb{R}$. This is used when $n > m$ (it is an isomorphism for $n = m$).