AM 160 - SciML:

Dante Buhl

January 17, 2025

Problem 1: Subproblem A

Show that $\Theta = A^\dagger y$ is the solution such that $||\Theta||_2$ is the minimum out of all the infinite solutions to $y = A\Theta$. Consider $A \in \mathbb{R}^{n\times p}$ where $p \gg n$.

Proof. We begin by writing the SVD of A which is part of the construction of the psuedoinverse $A^\dagger$.

$$A = U\Sigma V^*$$
$$A^\dagger = V\Sigma^{-1}U^*$$

where U and V are unitary matrices, and Σ^{-1} is the transpose of Σ containing the reciprocal of each singular value (in order) along the diagonal, i.e.

$$\Sigma^{-1} = \left[\begin{array}{ccc|c} 1/\sigma_1 & & \mathbf{0} & \\ & \ddots & & \mathbf{0} \\ \mathbf{0} & & 1/\sigma_n & \end{array}\right]$$

Next we need to understand why Θ in this context has infinitely many solutions. Let us assume that A is rank n. The implication is that there are at most n basis vectors in $\mathbb{R}^p$ which are not in the null space of the transformation A. We can write Θ as a linear combination of basis vectors which span $\mathbb{R}^p$

$$\Theta = c_1v_1 + \ldots + c_nv_n + \ldots + c_pv_p$$

Notice though, however, that only n of these vectors are in the kernel of A (and let us assume it is the first n vectors for convenience). We have then that all vectors $v_{n+1}, \ldots, v_p$ in the linear combination of Θ do not affect the solution. Therefore, an infinite number of solutions Θ can be created by adding the linear combination of $v_{n+1}, \ldots, v_p$ to any solution of $y = A\Theta$.

In order to see why the psuedoinverse $A^\dagger$ yields the minimum solution is because it projects the p dimension problem into a n dimension problem, i.e.

$$y = A\Theta$$
$$y = U\Sigma V^*\Theta$$
$$y = UC$$

where C is an $n \times 1$ vector which literally contains the coefficients of the linear combination for the first n basis vectors of Θ scaled by their corresponding singular value σ_i, i.e. $C_i = c_i\sigma_i$.
Finally, we solve for C using the inverse of U which exists since U is unitary.

$$C = U^*y$$

Notice that this constructs Θ out of the minimum number of basis vectors in order to span R^n that is, for any y given. Then looking at the 2-norm of Θ we have,

$$\begin{aligned} ||\Theta||_2 &= ||c_1v_1||_2 + \ldots + ||c_nv_n||_2 \\ &= |c_1|||v_1||_2 + \ldots + |c_n|||v_n||_2 \\ &= |c_1| + \ldots + |c_n| \end{aligned}$$

where we can decompose the 2-norm in this way because each basis vector v_i are orthogonal to each other in the 2-norm. Notice that the addition of any additional $\mathbb{R}^p$ basis vectors will only increase the 2-norm of Θ. We conclude then that solving for Θ using the psuedoinverse of A yields Θ such that the 2-norm of Θ is the minimum out of all possible Θ which solve $y = A\Theta$. $\square$