New York University Tandon School of Engineering
Mechanical and Aerospace Engineering

ROB-GY 6013 Mathematics for Robotics

Problem Set 1: Linear Algebra

Due Wednesday, October 7, 2026, before class (5:00 PM)
Submit on Gradescope.

Discussion with other students is allowed. Write your solutions alone.
Use any tool to understand the material, but not to write your solutions.
Solutions typeset in LaTeX or Markdown earn a 10% bonus. Show your reasoning: a bare final answer earns little credit.

Problem 1: Column space, null space, and solution sets (20 points)

An orthographic camera maps a point $x \in \mathbb{R}^3$ in the 3D world to a point $Px \in \mathbb{R}^2$ in an image. If two world points x_1 and x_2 produce the same image point, then the camera cannot distinguish between them. In this problem we use linear algebra to characterise exactly what information a single camera loses, and how adding a second camera removes the ambiguity.

Consider the camera matrix

$$P = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

(a) (4 pts) Let p_1, p_2, p_3 be the columns of P. Write Px as a linear combination of these columns. Find a basis for $\mathrm{Col}(P)$ and determine $\mathrm{rank}(P)$. Does $Px = b$ have a solution for every $b \in \mathbb{R}^2$? Explain.

(b) (4 pts) Solve $Px = 0$. Give a basis for $\mathrm{Null}(P)$, determine $\mathrm{nullity}(P)$, and verify the rank–nullity theorem. Are the columns of P linearly independent? Explain using your answer.

(c) (5 pts) Let $x_1, x_2 \in \mathbb{R}^3$. Show that

$$Px_1 = Px_2 \quad \Longleftrightarrow \quad x_1 - x_2 \in \mathrm{Null}(P).$$

Now suppose the camera observes $b = \left[\begin{smallmatrix} 3 \\ 2 \end{smallmatrix}\right]$. Find one solution x_p to $Px = b$, and describe all solutions in terms of x_p and $\mathrm{Null}(P)$.

(d) (3 pts) Now let P be *any* 2×3 camera matrix. Use the rank–nullity theorem to show that $\mathrm{Null}(P)$ must contain a nonzero vector. Explain why this means no single camera of this kind can distinguish every pair of world points.

(e) (4 pts) A second camera is added:

$$Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Stacking the two images into one observation gives

$$\begin{bmatrix} Px \\ Qx \end{bmatrix} = Ax, \qquad \text{where} \qquad A = \begin{bmatrix} P \\ Q \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \in \mathbb{R}^{4\times 3}.$$

Find rank(A) and Null(A). Can two distinct world points now produce exactly the same observations in both cameras? Explain.

Problem 2: Eigenvectors, diagonalization, and modes (20 points)

A planar robot has a small position error $x_k \in \mathbb{R}^2$ relative to a desired position. Suppose the error evolves at each control step according to

$$x_{k+1} = Ax_k, \qquad A = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}, \qquad x_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

(a) (3 pts) Compute $\det A$. Is A invertible? By what factor does A scale area? Does it reverse orientation?

(b) (5 pts) Find the eigenvalues of A by solving $\det(A - \lambda I) = 0$, and find a corresponding eigenvector for each eigenvalue. Explain geometrically what is special about these two directions under A.

(c) (4 pts) Normalise the two eigenvectors and use them as the columns of $Q = \begin{bmatrix} q_1 & q_2 \end{bmatrix}$. Verify that $Q^\mathsf{T} Q = I$. Compute $\det Q$, and determine whether Q represents a rotation or a reflection. Define $z_k = Q^\mathsf{T} x_k$. What does z_k represent?

(d) (4 pts) Show that $z_{k+1} = \Lambda z_k$ with $\Lambda = Q^\mathsf{T} A Q$. Compute Λ, and explain why the two components of z_k evolve independently.

(e) (4 pts) Find an explicit expression for the position error x_k for the given $x_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Then use it to determine what happens to the error as $k \to \infty$: does it converge to zero or grow, and which eigenvector direction does it eventually align with? *Hint: write x_0 as a combination of the two eigenvectors.*

Problem 3: Projection, normal equations, and least squares (20 points)

A neural network takes an input and produces a prediction. Its behavior is controlled by numerical values called parameters. Training changes these parameters so that the predictions become closer to the desired targets.

Suppose a neural network has two parameters, $\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$. For three training examples, the network currently produces $f(\theta)$, while the desired targets are y; define the current prediction error as r:

$$f(\theta) = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}, \qquad y = \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}, \qquad r = y - f(\theta).$$

The goal of training is to update the parameters θ so as to minimize the squared prediction error $\|y - f(\theta)\|_2^2$.

For a small parameter update $\Delta\theta$, the Gauss–Newton method uses the linear approximation

$$f(\theta + \Delta\theta) \approx f(\theta) + J\Delta\theta, \qquad \text{where} \qquad J = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}.$$

(a) (4 pts) Compute the current prediction error r. Using the linear approximation above, show that choosing $\Delta\theta$ to minimize the approximate squared prediction error is equivalent to solving $\min_{\Delta\theta}\|r - J\Delta\theta\|_2^2$. Determine whether there exists a $\Delta\theta$ such that $J\Delta\theta = r$. Explain your answer.

(b) (5 pts) Let $\Delta\hat{\theta}$ be the least-squares solution. Using the Projection Theorem, explain why $r - J\Delta\hat{\theta} \perp \mathrm{Col}(J)$, where $\mathrm{Col}(J)$ is the column space of J. Then derive the normal equations $J^\mathsf{T} J\,\Delta\hat{\theta} = J^\mathsf{T} r$.

(c) (5 pts) Solve the normal equations to find the Gauss–Newton update $\Delta\hat{\theta}$.

Then compute the approximate updated predictions

$$f(\theta + \Delta\hat{\theta}) \approx f(\theta) + J\Delta\hat{\theta}.$$

Compute the squared prediction error $\|y - f(\theta)\|_2^2$ before the update and the approximate squared prediction error after the update. Has the prediction error decreased?

(d) (6 pts) Compute the remaining prediction error

$$r - J\Delta\hat{\theta}$$

and verify that

$$J^\mathsf{T}\big(r - J\Delta\hat{\theta}\big) = 0.$$

The current prediction error can be decomposed as

$$r = J\Delta\hat{\theta} + \big(r - J\Delta\hat{\theta}\big).$$

Explain geometrically what $J\Delta\hat{\theta}$ and $r - J\Delta\hat{\theta}$ represent relative to $\mathrm{Col}(J)$.

Finally, is the Gauss–Newton update $\Delta\hat{\theta}$ unique in this example? Explain using the columns or rank of J.

Problem 4: SVD and orthogonalization (20 points)

Muon is a neural-network optimizer that modifies a matrix-valued update before applying it to the model. Its goal is to reduce large differences in update strength across different directions.

Suppose the update matrix at one training step has SVD

$$M = U\Sigma V^\mathsf{T}.$$

The singular values in Σ describe the update strength along different singular-vector directions. Muon approximately replaces M by

$$O = UV^\mathsf{T},$$

which keeps the singular-vector directions but replaces the nonzero singular values by 1.

Consider

$$M = \begin{bmatrix} 3 & 3 \\ -1 & 1 \end{bmatrix}.$$

(a) (1 pt) Compute $M^\mathsf{T}M$.

(b) (4 pts) Find the eigenvalues and orthonormal eigenvectors of $M^\mathsf{T}M$. Use them to determine the singular values σ_1, σ_2 and the corresponding right singular vectors v_1, v_2.

(c) (3 pts) Using $u_i = Mv_i/\sigma_i$, find the left singular vectors u_1, u_2 and write the SVD $M = U\Sigma V^\mathsf{T}$. What does the relation $Mv_i = \sigma_i u_i$ mean geometrically?

(d) (3 pts) Compute $O = UV^\mathsf{T}$. Compare Mv_i and Ov_i for $i = 1, 2$. What happens to the amount of stretching along the two singular-vector directions?

In practice M is very large, and an SVD at every training step is too expensive. Muon instead approximates UV^T by an iteration that uses only matrix multiplications: it leaves U and V unchanged and pushes every singular value toward 1. Before applying the iteration, the matrix is rescaled by its largest singular value, so that all singular values lie in $(0, 1]$.

(e) (3 pts) Suppose $X = U\Sigma V^\mathsf{T}$. Show that

$$XX^\mathsf{T} = U\Sigma^2 U^\mathsf{T} \qquad \text{and hence} \qquad (XX^\mathsf{T})X = U\Sigma^3 V^\mathsf{T}.$$

What does this tell you about the effect of $(XX^\mathsf{T})X$ on the singular vectors and singular values of X?

(f) (6 pts) Consider the simplified Newton–Schulz step

$$X_{\text{new}} = \tfrac{3}{2}X - \tfrac{1}{2}(XX^\mathsf{T})X.$$

Using part (e), show that

$$X_{\text{new}} = U\Big(\tfrac{3}{2}\Sigma - \tfrac{1}{2}\Sigma^3\Big)V^\mathsf{T}.$$

Hence, each singular value is transformed according to

$$\sigma_{\text{new}} = \tfrac{3}{2}\sigma - \tfrac{1}{2}\sigma^3.$$

Show that $\sigma = 1$ remains unchanged. Then, for $0 < \sigma < 1$, show that $\sigma < \sigma_{\text{new}} < 1$. What does this suggest about repeated Newton–Schulz steps and the goal of Muon?

Hint: show that $\sigma_{new} - \sigma > 0$ and that $1 - \sigma_{new} > 0$ by factoring each expression.