FIRST Name MaTricksed LAST Name YuvBin

SID (All Digits): ____________________

- **(5 Points)** In the space provided above, print legibly your name and all digits of your SID.

- **(10 Points) (Pledge of Academic Integrity)** Hand-copy, sign, and date the single-line text (which begins with *I have read, . . .*) of the Pledge of Academic Integrity on page 3 of this document. Your solutions will *not* be evaluated without this.

- **Urgent Contact with the Teaching Staff:** In case of an urgent matter, you can reach us at `eecs16a@berkeley.edu`.

- **The midterm PDF file (distributed separately, at the start of the exam) consists of pages numbered 1 through 6. This document consists of pages numbered 1 through 12.** Verify that your copies of the exam and this answer booklet are free of anomalies, and contain all of the specified number of pages. If you find a defect in either of your copies, produce another printout or contact the teaching staff immediately.

- This midterm is designed to be completed within 70 minutes. However, you may use up to 80 minutes total—*in one sitting*—to tackle the midterm *and* to upload your work to Gradescope.

  The exam starts at 6:40 pm California time. Your allotted window begins with respect to this start time. Students who have official accommodations of $1.5\times$ and $2\times$ time windows have 120 and 160 minutes, respectively.

  With respect to solutions uploaded to Gradescope after 8:00 pm (Regular), 8:40 pm ($1.5\times$ time DSP accommodation), or 9:20 pm ($2\times$ time DSP accommodation) we reserve the right to apply a 5% penalty for each minute of tardiness. For example, if you upload your solutions 10 minutes late, you potentially forfeit 50% of your total exam score.

  If you encounter Gradescope submission difficulty, contact the teaching staff by email immediately, and include, as an attachment to your email, a legible PDF file (preferred) or clear photos of your solutions.

- **This midterm is open book.** So long as you do not violate UC Berkeley, state, or US Federal regulations, you may—during the midterm—access or use (a) any reference in print or electronic form; and (b) any computing, communication, or other electronic device, to advance or verify your solutions.

  Citing the output of any such reference does *not* discharge from you the responsibility to provide sufficient explanation for your work.

- **Collaboration and consultation <u>are</u> permitted**, but
  - *only* with your peers who are currently enrolled in this course;
  - *only* on the public forum of the official Ed Stem page for this course; and
  - *only* to the extent where your peers are unstuck (i.e., not to the level where the solution becomes trivial or otherwise obvious to those who read your hints). Please use good judgment in dispensing hints to each other.

  No private communication is allowed with anyone.

  Communication in any other form—oral, written, or electronic; public or private; direct or indirect—with any human being outside the scope permitted above is *not* permitted.

  You may *not* work on this midterm in the physical proximity of any student currently enrolled in this course.

- Please write neatly and legibly, because *if we can't read it, we can't evaluate it.*

- For each problem, limit your work to the space provided specifically for that problem. *No other work will be considered. For example, we will not evaluate scratch work. No exceptions.*

- Unless explicitly waived by the specific wording of a problem, you must explain your responses (and reasoning) succinctly, but clearly and convincingly.

- In some parts of a problem, we may ask you to establish a certain result—for example, "show this" or "prove that." Even if you're unable to establish the result that we ask of you, you may still take that result for granted—and use it in any subsequent part of the problem.

- If we ask you to provide a "reasonably simple expression" for something, by default we expect your expression to be in closed form—one *not* involving a sum $\sum$ or an integral $\int$—*unless* we explicitly tell you otherwise.

- Noncompliance with these or other instructions from the teaching staff—*including, for example, commencing work prematurely, or continuing it beyond the allocated time window*—is a serious violation of the Code of Student Conduct.

Pledge of Academic Integrity

By my honor, I affirm that

(1) this document—which I have produced for the evaluation of my performance—reflects my original, bona fide work, and that I have neither provided to, nor received from, anyone excessive or unreasonable assistance that produces unfair advantage for me or for any of my peers;

(2) as a member of the UC Berkeley community, I have acted with honesty, integrity, respect for others, and professional responsibility—and in a manner consistent with the letter and intent of the campus Code of Student Conduct;

(3) I have not violated—nor aided or abetted anyone else to violate—the instructions for this exam given by the course staff, including, but not limited to, those on the cover page of this document; and

(4) More generally, I have not committed any act that violates—nor aided or abetted anyone else to violate—UC Berkeley, state, or Federal regulations, during this exam.

(10 Points) In the space below, hand-write the following sentence, verbatim. Then write your name in legible letters, sign, include your full SID, and date before uploading your work to Gradescope:

I have read, I understand, and I commit to adhere to the letter and spirit of the pledge above.

I hereby declare and solemnly pledge to Babakue you on every midterm until you turn into tempered steel, ready for the world ahead.

Full Name: Babakue Chef Signature: BBQC

Date: Happy Nowruz 1403 Student ID: 123456789

MT2.1

Method I: (a) Distinct columns of P have nonoverlapping 1's (and all other entries zero), because each row of P has a single 1. This makes distinct columns of P, each of which is a unit vector, orthonormal. Hence P is an orthogonal matrix.

Method II: Let $Q = P^T$ and $R = QP$. Then $r_{ij} = \sum_{k=1}^{n} q_{ik} p_{kj}$. But $q_{ik} = p_{ki}$ (since $Q = P^T$). Accordingly, $r_{ij} = \sum_{k=1}^{n} p_{ki} p_{kj}$. Due to the structure of P, we note that

$$p_{ki} p_{kj} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \implies r_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \implies R = QP = P^T P = I$$

(Only a single 1 across row k) (Yet more comments on the next page)

(b) This part probes your fluency with matrix multiplication, and knowledge of how pre- and post multiplication plucks out rows and columns of a matrix that is operated on.

Plucks row 2 of A

Plucks row 3 of A

$$PAP^T = \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}}_{P} \underbrace{\begin{bmatrix} e & 0 & d \\ 0 & a & 0 \\ c & 0 & b \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}}_{P^T}$$

Plucks row 1 of A

Plucks col 2 of PA

$$= \underbrace{\begin{bmatrix} 0 & a & 0 \\ c & 0 & b \\ e & 0 & d \end{bmatrix}}_{PA} \underbrace{\begin{bmatrix} 0 & 0 & \boxed{1} \\ \boxed{1} & 0 & 0 \\ 0 & \boxed{1} & 0 \end{bmatrix}}_{P^T} = \begin{bmatrix} a & 0 & 0 \\ 0 & b & c \\ 0 & d & e \end{bmatrix}$$

Plucks col 1 of PA

Plucks col 3 of PA

$$\boxed{PAP^T = \begin{bmatrix} a & 0 & 0 \\ 0 & b & c \\ 0 & d & e \end{bmatrix}}$$

MT2.1 (a)

Additional Insight: The columns of P are the rows of P^T in the same order. That is,

$$P = [\underline{p}_1 \cdots \underline{p}_n] \Rightarrow P^T = \begin{bmatrix} \underline{p}_1^T \\ \vdots \\ \underline{p}_n^T \end{bmatrix}$$

$$P^TP = \begin{bmatrix} \underline{p}_1^T \\ \vdots \\ \underline{p}_i^T \\ \vdots \\ \underline{p}_n^T \end{bmatrix} [\underline{p}_1 \cdots \underline{p}_j \cdots \underline{p}_n] = \begin{bmatrix} \langle \underline{p}_1, \underline{p}_1 \rangle & \cdots & \langle \underline{p}_1, \underline{p}_n \rangle \\ \vdots & \cdots \langle \underline{p}_i, \underline{p}_j \rangle \cdots & \vdots \\ \langle \underline{p}_n, \underline{p}_1 \rangle & & \langle \underline{p}_n, \underline{p}_n \rangle \end{bmatrix}$$

$$\langle \underline{p}_i, \underline{p}_j \rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

This is because $\underline{p}_i$ and $\underline{p}_j$, for $i \neq j$, don't have overlapping nonzero entries. If they did, say an entry k, then $p_{ki} = p_{kj} = 1$ for $i \neq j$, which means row k of P has at least two 1's, a contradiction. Therefore, we conclude that P^TP has zero off-diagonal entries. And its entries on the diagonal are all 1's. This makes $P^TP = I$.

Fun Factoids:

The matrix P, as described in the problem statement, is called a Permutation Matrix. It consists of an arbitrary rearrangement of the canonical unit vectors $\underline{e}_1, \dots, \underline{e}_n$. Note that $P\mathbb{1} = \mathbb{1}$ (each row sums to 1) and $\mathbb{1}^TP = \mathbb{1}^T$ (each column sums to 1).

Permutation matrices form a subset of Doubly-Stochastic matrices. A doubly-stochastic matrix is a nonnegative matrix each of whose columns and rows sums to 1.

(c) (i) Method I: Multiplication by P simply reorders the entries of $\underline{x}$. Accordingly, $\underline{y} = P\underline{x}$ contains the same entries as those of $\underline{x}$, but possibly in a different order. Therefore, not only is the mean, but also the variance of $\underline{y}$ identical to their counterparts in $\underline{x}$. (This explanation is sufficient for c(i) & c(ii))

(ii) Method I: See the explanation in (c)(i).

Method II:

$$\sigma_y^2 = \frac{\|\underline{y} - \mu_y \mathbb{1}\|^2}{n} = \frac{1}{n}\left(P\underline{x} - \mu_x \mathbb{1}\right)^T\left(P\underline{x} - \mu_x \mathbb{1}\right)$$

Note that we showed $\mu_y = \mu_x$ in (c)(i).

$$\sigma_y^2 = \frac{1}{n}\left(\underline{x}^T P^T - \mu_x \mathbb{1}^T\right)\left(P\underline{x} - \mu_x \mathbb{1}\right) = \frac{1}{n}\Big(\underline{x}^T \underbrace{P^T P}_{I}\underline{x} - \mu_x \underbrace{\underline{x}^T \underbrace{P^T\mathbb{1}}_{\mathbb{1}}}_{n\mu_x} - \mu_x \underbrace{\underbrace{\mathbb{1}^T P}_{\mathbb{1}^T}\underline{x}}_{n\mu_x} + \mu_x^2 \underbrace{\mathbb{1}^T\mathbb{1}}_{n}\Big)$$

$$= \frac{1}{n}\left(\underline{x}^T\underline{x} - n\mu_x^2 - \cancel{n\mu_x^2} + \cancel{n\mu_x^2}\right)$$

(See the next page for a third method.)

$$\sigma_y^2 = \frac{\underline{x}^T\underline{x}}{n} - \mu_x^2 = MS(\underline{x}) - \mu_x^2 = \sigma_x^2$$

(iii)

$$\sigma_{xy} = \frac{(\underline{x} - \mu_x\mathbb{1})^T(\underline{y} - \mu_y\mathbb{1})}{n} = \frac{(\underline{x}^T - \mu_x\mathbb{1}^T)(P\underline{x} - \mu_x\mathbb{1})}{n} = \underline{x}^T P\underline{x} - \mu_x \underline{x}^T\mathbb{1} - \mu_x$$

$\mathbb{1}^T\underline{x} = n\mu_x$

$$= \frac{\underline{x}^T P\underline{x} - \mu_x \underbrace{\underline{x}^T\mathbb{1}}_{n\mu_x} - \mu_x \underbrace{\mathbb{1}^T P}_{\mathbb{1}^T}\underline{x} + \mu_x^2 \underbrace{\mathbb{1}^T\mathbb{1}}_{n}}{n} = \frac{\underline{x}^T P\underline{x} - n\mu_x^2 - \cancel{n\mu_x^2} + \cancel{n\mu_x^2}}{n}$$

$$\Longrightarrow \boxed{\sigma_{xy} = \frac{\underline{x}^T P\underline{x}}{n} - \mu_x^2}$$

MT2.1 (c)(i) Method II:

$\underline{y} = P\underline{x} \implies \mu_y = \frac{1}{n}\mathbb{1}^T\underline{y} = \frac{1}{n}\mathbb{1}^T P\underline{x}$. But $\mathbb{1}^T P = \mathbb{1}^T$, because each column of P has a single 1 (and all other entries zero). So, it sums to 1. Accordingly,

$$\mu_y = \frac{1}{n}\mathbb{1}^T P\underline{x} = \frac{1}{n}\mathbb{1}^T\underline{x} = \mu_x \implies \mu_y = \mu_x .$$

Multiplication by P preserves the mean.

MT2.1(c)(ii) Method III:

$P\mathbb{1} = \mathbb{1}$ b/c P has row sums $= 1$

Factor out P

$$\sigma_y^2 = \frac{\|\underline{y} - \mu_y\mathbb{1}\|^2}{n} = \frac{\|P\underline{x} - \mu_x\mathbb{1}\|^2}{n} = \frac{\|P\underline{x} - \mu_x P\mathbb{1}\|^2}{n} = \frac{\|P(\underline{x} - \mu_x\mathbb{1})\|^2}{n}$$

Transposition reverses the order

$$= \frac{[P(\underline{x} - \mu_x\mathbb{1})]^T P(\underline{x} - \mu_x\mathbb{1})}{n} = \frac{(\underline{x} - \mu_x\mathbb{1})^T \overbrace{P^T P}^{I} (\underline{x} - \mu_x\mathbb{1})}{n}$$

$$= \frac{\|\underline{x} - \mu_x\mathbb{1}\|^2}{n} = \sigma_x^2 \implies \sigma_y^2 = \sigma_x^2$$

MT2.2

(a) $\tilde{\underline{x}} = \underline{x} - \mu_x \mathbb{1} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} \frac{x_1+x_2}{2} \\ \frac{x_1+x_2}{2} \end{bmatrix} = \frac{x_1-x_2}{2}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$. Similarly, $\tilde{\underline{y}} = \frac{y_1-y_2}{2}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$

$\|\tilde{\underline{x}}\| = \frac{|x_1-x_2|}{2}\left\|\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right\| = \frac{|x_1-x_2|}{2}\sqrt{2} = \frac{|x_1-x_2|}{\sqrt{2}}$. Similarly, $\|\tilde{\underline{y}}\| = \frac{|y_1-y_2|}{\sqrt{2}}$

$$\langle \tilde{\underline{x}}, \tilde{\underline{y}} \rangle = \left(\frac{x_1-x_2}{2}\right)\left(\frac{y_1-y_2}{2}\right)\overbrace{\begin{bmatrix} 1 & -1 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix}}^{2} = \frac{(x_1-x_2)(y_1-y_2)}{2}$$

$$\rho_{xy} = \frac{\langle \tilde{\underline{x}}, \tilde{\underline{y}} \rangle}{\|\tilde{\underline{x}}\|\|\tilde{\underline{y}}\|} = \frac{\frac{(x_1-x_2)(y_1-y_2)}{\cancel{2}}}{\frac{|x_1-x_2|}{\cancel{\sqrt{2}}}\frac{|y_1-y_2|}{\cancel{\sqrt{2}}}} = \frac{x_1-x_2}{|x_1-x_2|}\,\frac{y_1-y_2}{|y_1-y_2|} = \pm 1$$

Each factor is either +1 or −1, so the product is ±1 as well.

(b) (i) $\underline{y} = \alpha \underline{x} + \beta \mathbb{1}$. If $\alpha = 0$, then $\underline{y} = \beta\mathbb{1} \Rightarrow y_1 = y_2 = \beta$, which is a contradiction, as we've been told that $y_1 \neq y_2$ in the problem statement.

(ii)

$$\underline{y} = \alpha\underline{x} + \beta\mathbb{1} \Rightarrow \mu_y = \frac{1}{n}\mathbb{1}^T\underline{y} = \frac{1}{n}\mathbb{1}^T(\alpha\underline{x} + \beta\mathbb{1}) = \alpha\frac{1}{n}\underbrace{\mathbb{1}^T\underline{x}}_{n\mu_x} + \beta\frac{1}{n}\underbrace{\mathbb{1}^T\mathbb{1}}_{n} \Rightarrow$$

$$\mu_y = \alpha\mu_x + \beta \Rightarrow \beta = \mu_y - \alpha\mu_x \Rightarrow \underline{y} = \alpha\underline{x} + (\mu_y - \alpha\mu_x)\mathbb{1}$$

$$\tilde{\underline{y}} = \underline{y} - \mu_y\mathbb{1} = \overbrace{\alpha\underline{x} + \beta\mathbb{1}}^{\underline{y}} - (\overbrace{\alpha\mu_x + \beta}^{\mu_y})\mathbb{1} = \alpha(\underline{x} - \mu_x\mathbb{1}) + \cancel{\beta\mathbb{1}} - \cancel{\beta\mathbb{1}} = \alpha(\underline{x} - \mu_x\mathbb{1}) = \alpha\tilde{\underline{x}}$$

$$\rho_{xy} = \frac{\langle \tilde{\underline{x}}, \tilde{\underline{y}} \rangle}{\|\tilde{\underline{x}}\|\|\tilde{\underline{y}}\|} = \frac{\tilde{\underline{x}}^T\alpha\tilde{\underline{x}}}{\|\tilde{\underline{x}}\|\|\alpha\tilde{\underline{x}}\|} = \frac{\alpha\cancel{\|\tilde{\underline{x}}\|^2}}{|\alpha|\cancel{\|\tilde{\underline{x}}\|^2}} = \frac{\alpha}{|\alpha|} = \begin{cases} 1 & \text{if } \alpha > 0 \\ -1 & \text{if } \alpha < 0 \end{cases}$$

We already know that $\alpha \neq 0$.

(c) Since $y_1 \neq y_2$, it must be that $\{\underline{y}, \mathbb{1}\}$ is a linearly-independent set. Same for $\{\underline{x}, \mathbb{1}\}$. In $\mathbb{R}^2$ a basis consists only of a pair of linearly-independent set of vectors. So, each of $\{\underline{x}, \mathbb{1}\}$ and $\{\underline{y}, \mathbb{1}\}$ forms a basis. This means $\underline{y} \in \text{span}(\underline{x}, \mathbb{1})$ and $\underline{x} \in \text{span}(\underline{y}, \mathbb{1})$. This translates to $\underline{y} = \alpha\underline{x} + \beta\mathbb{1}$ and $\underline{x} = \lambda\underline{y} + \gamma\mathbb{1}$, for some α, β, λ, and $\gamma \in \mathbb{R}$.

MT2.3

(a) $\underline{a}_1$ is a unit vector $\Rightarrow$ $\boxed{\underline{q}_1 = \underline{a}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}}$. We also note that $\underline{a}_1$ and $\underline{a}_2$ have nonoverlapping nonzero entries, so $\langle \underline{a}_1, \underline{a}_2 \rangle = 0 \Rightarrow$ $\underline{a}_2 \perp \underline{a}_1$. This means $\underline{z}_2 = \underline{a}_2 \Rightarrow$ $\boxed{\underline{q}_2 = \frac{\underline{z}_2}{\|\underline{z}_2\|} = \frac{\underline{a}_2}{\|\underline{a}_2\|} = \frac{1}{\sqrt{2}}\begin{bmatrix}0\\1\\1\end{bmatrix}}$

$\|\underline{a}_2\| = \sqrt{0^2+1^2+1^2} = \sqrt{2}$

It's only for $\underline{q}_3$ that we must apply the formula of the Gram-Schmidt Algorithm: $\boxed{\underline{z}_3 = \underline{a}_3 - \langle \underline{a}_3, \underline{q}_1 \rangle \underline{q}_1 - \langle \underline{a}_3, \underline{q}_2 \rangle \underline{q}_2}$ (*)

Note that $\langle \underline{a}_3, \underline{q}_1 \rangle = [1 \ 0 \ 1]\begin{bmatrix}1\\0\\0\end{bmatrix} = 1$ and $\langle \underline{a}_3, \underline{q}_2 \rangle = [1 \ 0 \ 1]\begin{bmatrix}0\\1/\sqrt{2}\\1/\sqrt{2}\end{bmatrix} = \frac{1}{\sqrt{2}}$

So, $\boxed{\underline{z}_3 = \underline{a}_3 - \underline{q}_1 - \frac{1}{\sqrt{2}}\underline{q}_2}$ (**) $= \begin{bmatrix}1\\0\\1\end{bmatrix} - \begin{bmatrix}1\\0\\0\end{bmatrix} - \frac{1}{\sqrt{2}}\begin{bmatrix}0\\1/\sqrt{2}\\1/\sqrt{2}\end{bmatrix} = \begin{bmatrix}0\\-1/2\\1/2\end{bmatrix} \Rightarrow \boxed{\|\underline{z}_3\| = \sqrt{0+\frac{1}{4}+\frac{1}{4}} = \frac{1}{\sqrt{2}}}$

$\underline{q}_3 = \frac{\underline{z}_3}{\|\underline{z}_3\|} = \sqrt{2}\begin{bmatrix}0\\-1/2\\1/2\end{bmatrix} \Rightarrow \boxed{\underline{q}_3 = \begin{bmatrix}0\\-1/\sqrt{2}\\1/\sqrt{2}\end{bmatrix}}$

(b)

<u>Method I</u>: Note that $\underline{b} = \begin{bmatrix}2\\1/\sqrt{2}\\1/\sqrt{2}\end{bmatrix} = \begin{bmatrix}2\\0\\0\end{bmatrix} + \begin{bmatrix}0\\1/\sqrt{2}\\1/\sqrt{2}\end{bmatrix} = 2\underbrace{\begin{bmatrix}1\\0\\0\end{bmatrix}}_{\underline{a}_1} + \frac{1}{\sqrt{2}}\underbrace{\begin{bmatrix}0\\1\\1\end{bmatrix}}_{\underline{a}_2}$

$\Rightarrow \underline{b} = 2\underline{a}_1 + \frac{1}{\sqrt{2}}\underline{a}_2 + 0 \cdot \underline{a}_3 \Rightarrow \underbrace{[\underline{a}_1 \ \underline{a}_2 \ \underline{a}_3]}_{A}\underbrace{\begin{bmatrix}2\\1/\sqrt{2}\\0\end{bmatrix}}_{\underline{x}} = \underline{b}$

$\Rightarrow \boxed{\underline{x} = \begin{bmatrix}2\\1/\sqrt{2}\\0\end{bmatrix}}$

For a longer method, using back substitution, see the next page.

MT 2.3 (b) Method II:

In part (a) we determined

$$\underline{q}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \underline{a}_1, \quad \underline{q}_2 = \begin{bmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \frac{1}{\sqrt{2}}\underline{a}_2, \quad \underline{q}_3 = \begin{bmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

Recall Eqn (*), and its simplified, numeric form Eqn (**) from part (a):

$$\underline{z}_3 = \underline{a}_3 - \langle \underline{a}_3, \underline{q}_1 \rangle \underline{q}_1 - \langle \underline{a}_3, \underline{q}_2 \rangle \underline{q}_2$$

$$\left.\begin{array}{l} \underline{z}_3 = \underline{a}_3 - \underline{q}_1 - \frac{1}{\sqrt{2}} \underline{q}_2 \\ \text{Also, } \underline{q}_3 = \frac{\underline{z}_3}{\|\underline{z}_3\|} \Rightarrow \underline{z}_3 = \|\underline{z}_3\| \underline{q}_3 \end{array}\right\} \Rightarrow$$

$$\underline{a}_3 = \underline{q}_1 + \frac{1}{\sqrt{2}} \underline{q}_2 + \underbrace{\|\underline{z}_3\|}_{1/\sqrt{2}} \underline{q}_3$$

We can now write

$$\underbrace{\begin{bmatrix} \underline{a}_1 & \underline{a}_2 & \underline{a}_3 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} \underline{q}_1 & \underline{q}_2 & \underline{q}_3 \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} 1 & 0 & 1 \\ 0 & \sqrt{2} & 1/\sqrt{2} \\ 0 & 0 & 1/\sqrt{2} \end{bmatrix}}_{R}$$

$$\underline{a}_1 = \underline{q}_1$$
$$\underline{a}_2 = \sqrt{2}\, \underline{q}_2$$
$$\underline{a}_3 = \underline{q}_1 + \frac{1}{\sqrt{2}} \underline{q}_2 + \frac{1}{\sqrt{2}} \underline{q}_3$$

$$A\underline{x} = \underline{b} \Rightarrow QR\underline{x} = \underline{b} \Rightarrow \underbrace{Q^TQ}_{I} R\underline{x} = Q^T\underline{b} \Rightarrow R\underline{x} = Q^T\underline{b}$$

We now have an upper-triangular system.

$$\underbrace{\begin{bmatrix} 1 & 0 & 1 \\ 0 & \sqrt{2} & 1/\sqrt{2} \\ 0 & 0 & 1/\sqrt{2} \end{bmatrix}}_{R} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{\underline{x}} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}}_{Q^T} \underbrace{\begin{bmatrix} 2 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}}_{\underline{b}}$$

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & \sqrt{2} & 1/\sqrt{2} \\ 0 & 0 & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$$

Start from the bottom:

$$\frac{1}{\sqrt{2}} x_3 = 0 \implies \underline{x_3 = 0}$$

Penultimate Equation:

$$\sqrt{2}\, x_2 + \frac{1}{\sqrt{2}} \cancelto{0}{x_3} = 1 \implies \underline{x_2 = \frac{1}{\sqrt{2}}}$$

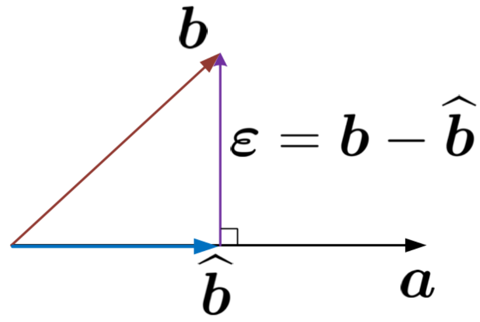
First Equation:

$$x_1 + 0 \cdot \cancelto{0}{x_2} + \cancelto{0}{x_3} = 2 \implies \underline{x_1 = 2}$$

So the solution to $A\underline{x} = \underline{b}$ is $\underline{x} = \begin{bmatrix} 2 \\ 1/\sqrt{2} \\ 0 \end{bmatrix}$, just as we had determine using Method I.

Method II, though much longer, handles a much broader class of problems than Method I. So, it should be in your arsenal.

MT2.4



(a)

$$\underline{a} \perp \underline{\varepsilon} \Rightarrow \langle \underline{a}, \underline{\varepsilon} \rangle = \underline{a}^T \underline{\varepsilon} = 0 \Rightarrow \left.\begin{matrix} \underline{a}^T(\underline{b} - \hat{\underline{b}}) = 0 \\ \hat{\underline{b}} = \mu \underline{a} \end{matrix}\right\} \Rightarrow$$

$$\underline{a}^T(\underline{b} - \mu \underline{a}) = 0 \Rightarrow \underline{a}^T \underline{b} - \mu \underline{a}^T \underline{a} = 0 \Rightarrow \mu = (\underline{a}^T \underline{a})^{-1} \underline{a}^T \underline{b}$$

$$\hat{\underline{b}} = \mu \underline{a} = \underline{a} \mu = \underbrace{\underline{a}(\underline{a}^T \underline{a})^{-1} \underline{a}^T}_{P} \underline{b} \Rightarrow \hat{\underline{b}} = P \underline{b}$$, where

$$\boxed{P = \underline{a}(\underline{a}^T \underline{a})^{-1} \underline{a}^T}$$

(b) $P = \underline{a}(\underline{a}^T \underline{a})^{-1} \underline{a}^T = \dfrac{1}{\|\underline{a}\|^2} \underline{a}\,\underline{a}^T$

$\|\underline{a}\|^2 = \underline{a}^T \underline{a}$

The outer product $\underline{a}\,\underline{a}^T$ is a rank-1 matrix. Every column of $\underline{a}\,\underline{a}^T$ is a scaled version of $\underline{a}$.

P is a scaled version of a rank-1 matrix, so it, too, has rank 1.

$$\boxed{\operatorname{rank}(P) = 1}$$

MT2.4

(c) (i) $P = \frac{1}{\|\underline{a}\|^2}\underline{a}\,\underline{a}^T \Rightarrow P^T = \left(\frac{1}{\|\underline{a}\|^2}\underline{a}\,\underline{a}^T\right)^T = \frac{1}{\|\underline{a}\|^2}\underbrace{(\underline{a}^T)^T}_{\underline{a}}\underline{a}^T = P$

$$\boxed{\Rightarrow P^T = P}$$

(ii) $P^2 = \underline{a}(\underline{a}^T\underline{a})^{-1}\overbrace{\underline{a}^T\,\underline{a}(\underline{a}^T\underline{a})^{-1}}^{1}\underline{a}^T = \underline{a}(\underline{a}^T\underline{a})^{-1}\underline{a}^T = P$

Now look at P^3: $P^3 = P\underbrace{P^2}_{P} = PP = P^2 = P \Rightarrow P^3 = P$

Continuing, we have $\boxed{P^k = P,\ \forall k \in \{1,2,3,\ldots\}}$. Geometrically, this means that once a vector $\underline{b}$ is projected onto span($\underline{a}$) through the transformation $\hat{\underline{b}} = P\underline{b}$, repeated applications of P to $\hat{\underline{b}}$ will not move it. In fact, any matrix P that satisfies $P^2 = P$ is called a projection matrix. If, in addition, $P^T = P$, it's called an orthogonal projection!

(d)

$P\underline{y} = \underline{0} \Rightarrow \underset{\neq \underline{0}}{\underline{a}}\underbrace{(\underline{a}^T\underline{a})^{-1}}_{>0}\underline{a}^T\underline{y} = 0 \Rightarrow \underline{a}^T\underline{y} = 0 \Rightarrow \underline{a} \perp \underline{y}$

$\underline{a}$ is orthogonal to $\underline{y}$.

(e) (i) $S = I - P \Rightarrow S^2 = (I-P)(I-P) = I - P - P + \overbrace{P^2}^{P} = I - 2P + P$

$\Rightarrow S^2 = I - P$. By a similar reasoning as was used in (c)(ii), $S^k = S \quad \forall k \in \{1, 2, \ldots\}$.

In fact $S^T = I^T - P^T = I - P = S$.
So, $S = I - P$ is an orthogonal projection matrix as well.

(ii) $P\underline{y} = \underline{0}$

$S\underline{y} = (I-P)\underline{y} = \underbrace{I\underline{y}}_{\underline{y}} - \underbrace{P\underline{y}}_{\underline{0}} = \underline{y} \Rightarrow \boxed{S\underline{y} = \underline{y}}$

We found earlier that $\underline{y} \perp \underline{a}$ if $P\underline{y} = \underline{0}$. Here we learn that $S = I - P$ projects vectors onto the subspace of $\mathbb{R}^n$ that is orthogonal to span$(\underline{a})$.

(iii)

As the hint suggests, let $\underline{v} \in \mathbb{R}^n$ be such that $S\underline{v} = \underline{0}$ (i.e., $\underline{v} \in \mathcal{N}(S)$). This means $(I-P)\underline{v} = 0 \Rightarrow \underline{v} = P\underline{v}$. From here, we can argue in two ways: (I) $P\underline{v} \in$ span$(\underline{a})$ since P projects any vector onto span$(\underline{a})$. So if after that projection, $\underline{v} = P\underline{v}$, it means $\underline{v} \in$ span$(\underline{a})$ to begin with. (II) $\underline{v} = P\underline{v} = \underline{a}\underbrace{(\underline{a}^T\underline{a})^{-1}}_{\text{scalar}}\underbrace{\underline{a}^T\underline{v}}_{\text{scalar}} = \underbrace{(\underline{a}^T\underline{a})^{-1}\underline{a}^T\underline{v}}_{\text{scalar}}\,\underline{a}$.

So $\underline{v} = \beta\underline{a}$, where $\beta = (\underline{a}^T\underline{a})^{-1}\underline{a}^T\underline{v} \in \mathbb{R} \Rightarrow \underline{v} \in$ span$(\underline{a})$.

(iv) $\mathcal{N}(S) =$ span$(\underline{a}) \Rightarrow \dim \mathcal{N}(S) = 1$. Claim: rank$(S) = n - \dim\mathcal{N}(S) = n-1$

Construct a basis $\{\underline{z}_0 = \underline{a}, \underline{z}_1, \ldots, \underline{z}_{n-1}\}$ for $\mathbb{R}^n$. By Gram-Schmidt, we can construct this basis to be orthogonal — that is, $\underline{z}_i \perp \underline{z}_j$ if $i \neq j$. We don't need them to be unit norm in this part. Any vector $\underline{x}$ in $\mathbb{R}^n$ can be writen as $\underline{x} = \sum_{i=0}^{n-1} \alpha_i \underline{z}_i \Rightarrow \underline{y} = S\underline{x} = \alpha_0 \underbrace{\cancel{S\underline{a}}}_{0} + \sum_{i=1}^{n-1} \alpha_i S\underline{z}_i = \sum_{i=1}^{n-1}\alpha_i S\underline{z}_i$

Note that if $\underline{z} \perp \underline{a} \Rightarrow P\underline{z} = \underline{0} \Rightarrow S\underline{z} = (I-P)\underline{z} = \underline{z} - \underbrace{\cancel{P\underline{z}}}_{0} = \underline{z} \Rightarrow S\underline{z}_i = \underline{z}_i \quad i = 1, \ldots, n-1$.

So any $\underline{y} \in$ Column Span of S is expressible as $\underline{y} = \sum_{i=1}^{n-1} \alpha_i \underline{z}_i \Rightarrow \underline{z}_1, \ldots, \underline{z}_n$ form a basis for the column span of $S \Rightarrow S$ has $n-1$ lin indep cols $\Rightarrow$ rank $S = n-1$.

MT2.5

(a) $W = I \Rightarrow \langle \underline{x}, \underline{y} \rangle_W = \underline{x}^T I \underline{y} = \underline{x}^T \underline{y}$

(b) $W = \begin{bmatrix} 0 & 3 \\ 1 & 0 \end{bmatrix}$

Let $\underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \Rightarrow \langle \underline{x}, \underline{x} \rangle_W = [x_1 \;\; x_2] \begin{bmatrix} 0 & 3 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = [x_1 \;\; x_2] \begin{bmatrix} 3x_2 \\ x_1 \end{bmatrix} = 3x_1x_2 + x_1x_2$

$\Rightarrow \langle \underline{x}, \underline{x} \rangle_W = 4x_1x_2$

Let $\underline{x} = \begin{bmatrix} 0 \\ x_2 \end{bmatrix}$ where $x_2 \neq 0$. Hence $\underline{x} \neq \underline{0}$ but $\langle \underline{x}, \underline{x} \rangle_W = 0$.

Alternative, we can make x_1 and x_2 have opposite signs so, for example $x_2 < 0 < x_1$. This means $x_1x_2 < 0 \Rightarrow$ $\langle \underline{x}, \underline{x} \rangle_W = 4x_1x_2 < 0$, which violates the nonnegativity property of an inner product.

So $\langle \underline{x}, \underline{y} \rangle_W = \underline{x}^T \begin{bmatrix} 0 & 3 \\ 1 & 0 \end{bmatrix} \underline{y}$ is not an inner product.

This problem explores weighted inner products $\langle \underline{x}, \underline{y} \rangle_W = \underline{x}^T W \underline{y}$, where W is a weighting matrix that must be chosen properly.

MT2.5

(c) $W = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. The first three properties of an inner product are satisfied trivially for this matrix W. So, we focus on nonnegativity. (See next page for bonus)

Let $\underline{x} = \begin{bmatrix} a \\ b \end{bmatrix}$

$$\langle \underline{x}, \underline{x} \rangle_W = [a \ \ b] \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = [a \ \ b] \begin{bmatrix} 2a+b \\ a+2b \end{bmatrix} = a(2a+b) + b(a+2b)$$
$$= 2a^2 + ab + ba + 2b^2 = 2a^2 + 2ab + 2b^2$$

The hint suggests that we rewrite this as

$$\langle \boldsymbol{z}, \boldsymbol{z} \rangle_{\mathbf{W}} = \lambda(a-b)^2 + \mu(a+b)^2 = \lambda(a^2 - 2ab + b^2) + \mu(a^2 + 2ab + b^2)$$
$$= (\lambda+\mu)a^2 + 2(\mu-\lambda)ab + (\lambda+\mu)b^2$$
$$= 2\ a^2 + 2\ \ ab + 2\ b^2$$

$$\boxed{\begin{cases} \lambda + \mu = 2 \\ -\lambda + \mu = 1 \end{cases} \quad 2\mu = 3 \rightarrow \mu = 3/2, \quad \lambda = 1/2}$$

$\langle \underline{x}, \underline{x} \rangle_W = 2a^2 + 2ab + 2b^2 = \frac{1}{2}(a-b)^2 + \frac{3}{2}(a+b)^2 \geq 0$ w/ equality iff $a = b = 0$.

(d) $W = \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ & & \\ w_{n1} & \cdots & w_{nn} \end{bmatrix}$. Let $w_{ii} = 0$, $\exists i \in \{1, \dots, n\}$.

Let $\underline{x} = \underline{e}_i$: ith canonical unit vector. Then

$\langle \underline{x}, \underline{x} \rangle_W = \underline{e}_i^T W \underline{e}_i$. Note $W\underline{e}_i = \underline{w}_i$: the ith column of W. So

$\underline{e}_i^T W \underline{e}_i = \underline{e}_i^T \underline{w}_i = w_{ii} = 0$. Even though $\underline{x} = \underline{e}_i \neq \underline{0}$, we've found that $\langle \underline{x}, \underline{x} \rangle_W = \langle \underline{e}_i, \underline{e}_i \rangle_W = w_{ii} = 0$. Violates the nonnegativity property.

(e) If $W = [\underline{w}_1 \cdots \underline{w}_n] \in \mathbb{R}^{n \times n}$ is not invertible, then its columns are linearly dependent. Accordingly, there exists a a nonzero vector $\underline{x}$ such that $W\underline{x} = \sum_{i=1}^{n} x_i \underline{w}_i = \underline{0}$. Therefore, $\langle \underline{x}, \underline{x} \rangle_W = \underline{x}^T W \underline{x} = \underline{x}^T \underline{0} = 0$ even though $\underline{x} \neq \underline{0}$. The nonnegativity property doesn't hold.

Recall that any inner product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ must satisfy the following properties:

(i) $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$

(ii) $\langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \langle \boldsymbol{y}, \boldsymbol{z} \rangle$

(iii) $\langle c\, \boldsymbol{x}, \boldsymbol{y} \rangle = c \langle \boldsymbol{x}, \boldsymbol{y} \rangle$

(iv) $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geq 0$, with equality if, and only if, $\boldsymbol{x} = 0$.

Consider a proposed inner product

$$\langle \underline{x}, \underline{y} \rangle_W = \underline{x}^T W \underline{y}$$

Properties (ii) and (iii) hold regardless of W.

(ii) $\langle \underline{x} + \underline{y}, \underline{z} \rangle_W = (\underline{x} + \underline{y})^T W \underline{z} = (\underline{x}^T + \underline{y}^T) W \underline{z}$

$$= \underline{x}^T W \underline{z} + \underline{y}^T W \underline{z}$$

$$= \langle \underline{x}, \underline{z} \rangle_W + \langle \underline{y}, \underline{z} \rangle_W$$

(iii) $\langle c\underline{x}, \underline{y} \rangle_W = (c\underline{x})^T W \underline{y} = c\, \underline{x}^T W \underline{y} = c \langle \underline{x}, \underline{y} \rangle_W$

What about Property (i)?

$$\langle \underline{y}, \underline{x} \rangle_W = \underline{y}^T W \underline{x} = (\underline{x}^T W^T \underline{y})^T = \underline{x}^T W^T \underline{y}$$

because what's being transposed is a scalar.

But $\langle \underline{x}, \underline{y} \rangle_W = \underline{x}^T W \underline{y}$

If $W = W^T$ (i.e., if W is symmetric), then clearly

$$\langle \underline{y}, \underline{x} \rangle_W = \underline{y}^T W \underline{x} = \underline{x}^T W^T \underline{y} = \underline{x}^T W \underline{y} = \langle \underline{x}, \underline{y} \rangle_W$$

if $W = W^T$

But $W = W^T$ is also a necessary condition.

To see this, assume $W^T \neq W \Rightarrow W^T - W \neq 0$ ← $n \times n$ matrix of zeroes.

$$\left.\begin{aligned} \langle \underline{y}, \underline{x} \rangle_W &= \underline{x}^T W^T \underline{y} \\ \langle \underline{x}, \underline{y} \rangle_W &= \underline{x}^T W \underline{y} \end{aligned}\right\} \Rightarrow \langle \underline{y}, \underline{x} \rangle_W - \langle \underline{x}, \underline{y} \rangle_W = \underline{x}^T W^T \underline{y} - \underline{x}^T W \underline{y} \Rightarrow$$

$$\langle \underline{y}, \underline{x} \rangle_W - \langle \underline{x}, \underline{y} \rangle_W = x^T (W^T - W) \underline{y} = x^T \Delta \underline{y}, \text{ where } \Delta = W^T - W$$

For $\langle \underline{x}, \underline{y} \rangle_W = \underline{x}^T W \underline{y}$ to be an inner product, $\underline{x}^T \Delta \underline{y}$ must be zero for all $\underline{x}$ and $\underline{y}$ in $\mathbb{R}^n$, not merely for a cherry-picked set of $\underline{x}$ and $\underline{y}$.

Since $\Delta = W^T - W \neq 0$, assume, without loss of generality, that its ij-th entry $\delta_{ij} \neq 0$. Then use $\underline{y} = \underline{e}_j$ and $\underline{x} = \underline{e}_i$ in

$$\langle \underline{y}, \underline{x} \rangle_W - \langle \underline{x}, \underline{y} \rangle_W = \underline{x}^T \Delta \underline{y} = \underline{e}_i^T \Delta \underline{e}_j = \delta_{ij} \neq 0$$

This shows that if $W^T \neq W$, $\langle \underline{x}, \underline{y} \rangle_W$ does not satisfy Property (i), and hence is not an inner product.

$$\langle \underline{x}, \underline{y} \rangle_W = \langle \underline{y}, \underline{x} \rangle_W \iff W^T = W$$