MIT 18.06 Final Exam **Solutions**, Spring 2022
Johnson

Problem 1 (4+4+6 points):

The matrix A is given by

$$A = LUL^{-1}U^{-1}$$

for

$$L = \begin{pmatrix} 1 & & & \\ -1 & 1 & & \\ 0 & 3 & 1 & \\ 1 & 0 & 0 & 1 \end{pmatrix}, \qquad U = \begin{pmatrix} 2 & 0 & 1 & 1 \\ & -1 & 0 & -1 \\ & & -2 & 1 \\ & & & 1 \end{pmatrix}.$$

(a) Write an expression for A^{-1} in terms of L, U, L^{-1}, and/or U^{-1} (but you **don't** need to actually multiply or invert the terms!).

(b) What is the determinant of A?

(c) Solve $PAx = b$ for x, where P is the 4×4 permutation that swaps the 1st and 4th elements of a vector, and $b = \begin{pmatrix} -5 \\ 4 \\ 11 \\ -3 \end{pmatrix}$. (You can get partial credit by just outlining a reasonable sequence of steps here that doesn't involve a lot of unnecessary calculation.)

Solution:

(a) The inverse of a product is the product of the inverses in reverse order, so $\boxed{A^{-1} = ULU^{-1}L^{-1}}$.

(b) Using the various determinant identities:

$$\begin{aligned} \det A &= \det(LUL^{-1}U^{-1}) = \det(L)\det(U)\det(L^{-1})\det(U^{-1}) \\ &= \det(L)\det(U)\det(L)^{-1}\det(U)^{-1} \\ &= \boxed{1}. \end{aligned}$$

(c) We have

$$x = A^{-1}P^{-1}b = \underbrace{U\,\underbrace{L\,\underbrace{U^{-1}\,\underbrace{L^{-1}P^{-1}b}_{u}}_{v}}_{w}}$$

$$x = A^{-1}P^{-1}b = U\,\underbrace{L\,\underbrace{U^{-1}\,\underbrace{L^{-1}P^{-1}b}_{u}}_{v}}_{w},$$

but each of these steps (labeled u, v, w for convenience below) is easy. To be reasonably efficient (and preserve your sanity), it is crucial to **not** explicitly compute any matrix inverses and to **not** explicitly compute the matrix A (or PA) at all. Instead, we break $ULU^{-1}L^{-1}P^{-1}b$ into a sequence of steps, evaluating **from right-to-left**, exploiting the fact that multiplying by U^{-1} or L^{-1} is just a **triangular solve** and can be done quickly. (**No Gaussian elimination**.)

Another way of saying this is that for $n \times n$ matrices, doing inversion, or Gaussian elimination, or even just multiplying two matrices costs $\sim n^3$ arithmetic operations, whereas a sequence of triangular solves and matrix–vector multiplications costs $\sim n^2$ arithmetic. Here, n is only 4, but even so the $\sim n^2$ approach is *significantly* less work (especially because it turns out to involve no fractions):

(i) P is just a swap, so it is its own inverse: $P^{-1} = P$ and $P^{-1}b = \begin{pmatrix} -3 \\ 4 \\ 11 \\ -5 \end{pmatrix}$.

(ii) $u = L^{-1}(P^{-1}b)$ is equivalent to solving $Lu = P^{-1}b$, which we can do by **forward-substitution** since L is lower-triangular:

$$\begin{pmatrix} 1 & & & \\ -1 & 1 & & \\ 0 & 3 & 1 & \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{pmatrix} = \begin{pmatrix} -3 \\ 4 \\ 11 \\ -5 \end{pmatrix} \implies \begin{array}{r} u_1 = -3 \\ -u_1 + u_2 = 4 \implies u_2 = 1 \\ 3u_2 + u_3 = 11 \implies u_3 = 8 \\ u_1 + u_4 = -5 \implies u_4 = -2 \end{array}$$

$$\implies u = \begin{pmatrix} -3 \\ 1 \\ 8 \\ -2 \end{pmatrix}.$$

(iii) $v = U^{-1}u$ is equivalent to solving $Uv = u$, which we can do by

back-substitution since U is upper-triangular:

$$\begin{pmatrix} 2 & 0 & 1 & 1 \\ & -1 & 0 & -1 \\ & & -2 & 1 \\ & & & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix} = \begin{pmatrix} -3 \\ 1 \\ 8 \\ -2 \end{pmatrix} \implies \begin{array}{c} 2v_1 + v_3 + v_4 = -3 \implies v_1 = 2 \\ -v_2 - v_4 = 1 \implies v_2 = 1 \\ -2v_3 + v_4 = 8 \implies v_3 = -5 \\ v_4 = -2 \end{array}$$

$$\implies v = \begin{pmatrix} 2 \\ 1 \\ -5 \\ -2 \end{pmatrix}.$$

(iv) Finally, we just need to multiply v by L and then U:

$$w = Lv = \begin{pmatrix} 1 & & & \\ -1 & 1 & & \\ 0 & 3 & 1 & \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ -5 \\ -2 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \\ -2 \\ 0 \end{pmatrix},$$

and hence we finally obtain

$$x = Uw = \begin{pmatrix} 2 & 0 & 1 & 1 \\ & -1 & 0 & -1 \\ & & -2 & 1 \\ & & & 1 \end{pmatrix} \begin{pmatrix} 2 \\ -1 \\ -2 \\ 0 \end{pmatrix} = \boxed{\begin{pmatrix} 2 \\ 1 \\ 4 \\ 0 \end{pmatrix}}.$$

If you happened to try to solve the last part by first computing U^{-1} and L^{-1} and multiplying $PA = PLUL^{-1}U^{-1}$ explicitly, in the unlikely event you did it correctly you would have gotten:

$$L^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ -3 & -3 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}, \quad U^{-1} = \begin{pmatrix} 0.5 & 0 & 0.25 & -0.75 \\ 0 & -1 & 0 & -1 \\ 0 & 0 & -0.5 & 0.5 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$PA = PLUL^{-1}U^{-1} = \begin{pmatrix} -1.5 & 3 & -1.25 & 7.75 \\ 1 & -2 & 1 & -6 \\ 2.5 & -3 & 2.25 & -9.75 \\ -1 & 3 & -1 & 6 \end{pmatrix},$$

$$(PA)^{-1} = ULU^{-1}L^{-1}P = \begin{pmatrix} -3.75 & -3.75 & 0.25 & 1.5 \\ 0 & 1 & 0 & 1 \\ 5.25 & 2.25 & 1.25 & -2.5 \\ 0.25 & -0.75 & 0.25 & -0.5 \end{pmatrix}$$

and then you could multiply $(PA)^{-1}$ by b to get the same x. It would be a rare human being who could carry out all those steps without making an arithmetic error, however! (I used Julia to get the matrices above.)

Problem 2 (4+6 points):

(a) If a and x are vectors in $\mathbb{R}^n$, then aa^Tx can be computed using either **left-to-right** as $(aa^T)x$ or **right-to-left** as $a(a^Tx)$, where the parentheses indicate the order of operations. Roughly count the number of arithmetic operations (additions and multiplications) in these two approaches: say whether each approach scales proportional to n, n^2, n^3, etcetera.

(b) A is an $n \times n$ real matrix and x is an n-component real vector. Indicate which of the following **must be equal** to one another:

$$\text{trace}(Axx^T), \quad \text{trace}(xAx^T), \quad \text{trace}(x^TAx), \quad x^TAx,$$

$$\text{trace}(x^TxA), \quad xx^TA, \quad \text{trace}(xx^TA), \quad \text{determinant}(xx^TA).$$

For the expressions that are equal, indicate how you would evaluate this quantity in a cost (in arithmetic operations) proportional to n^2.

Solution:

(a) The arithmetic counts are as follows:

(i) left-to-right $(aa^T)x$ requires $\boxed{\sim n^2}$ operations. Forming the aa^T matrix ($n \times n$) requires n^2 multiplications (of every element of a by every other element of a), while multiplying aa^T by the vector x requires another n^2 multiplications (and $n^2 - n$ additions), for a cost that is proportional to n^2 for large n. Computer scientists would say that the cost is $\Theta(n^2)$.

(ii) right-to-left $a(a^Tx)$ requires $\boxed{\sim n}$ operations. The dot product a^Tx requires n multiplications and $n-1$ additions, while multipling (a^Tx) by a is another n multiplications. Computer scientists would say that the cost is $\Theta(n)$

(b) The key rule to remember here is the **cyclic property** of the trace: $\text{trace}(AB) = \text{trace}(BA)$ for any matrices A and B (whose sizes allow them to be multiplied). Hence

$$\boxed{\text{trace}(Axx^T) = \text{trace}(xx^TA) = \text{trace}(x^TAx) = x^TAx},$$

where for the last equality we have employed the usual 18.06 ambiguity between interpreting x^TAx as a 1×1 matrix or as a scalar. (Note that the expression xAx^T doesn't even makes sense because the shapes are wrong.)

To evaluate this in operations proportional to n^2 , the simplest is perhaps to use $x^TAx = x^T(Ax)$, which requires one matrix-vector product Ax ($\sim n^2$ cost) and one dot product ($\sim n$ cost, negligible for large n). Alternatively, you could (for example) compute Axx^T as $(Ax)x^T$, since both $y = Ax$ and yx^T cost $\sim n^2$ operations, and then find the trace ($\sim n$). One thing you should definitely **not** do is to compute xx^T ($\sim n^2$), multiply it by A ($\sim n^3$ for a matrix–matrix multiplication!) and then take the trace ($\sim n$).

Problem 3 (4+4+4+5 points):

You have a 4×3 matrix $A = \begin{pmatrix} q_1 & 2q_2 & 3q_1 + 4q_2 \end{pmatrix}$, where we have expressed the three columns of A in terms of the orthonormal vectors

$$q_1 = \frac{1}{2}\begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}, \qquad q_2 = \frac{1}{2}\begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}.$$

(a) What is the rank of A?

(b) Give a basis for $N(A)$.

(c) You are asked to calculate the projection matrix P onto $C(A)$. Your friend Harvey Ard suggests applying the formula $P = A(A^TA)^{-1}A^T$ he memorized in linear algebra. Explain why this won't work here, and give an even simpler (correct) formula for P in terms of the quantities above. (You need not evaluate P numerically, just write a formula in terms of products of quantities defined above.)

(d) Find the *closest* vector to $x = \begin{pmatrix} 2 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ in $N(A^T)$.

Solution:

(a) A is $\boxed{\text{rank } 2}$ since its columns are spanned by q_1 and q_2.

(b) The nullspace $N(A)$ is $3 - 2 = 1$ dimensional, so we just need one vector. We don't need to do any elimination, however, because we **are already given** that the third column of A is a linear combination of the first two:

$$\text{(column 3)} = 3\text{(column 1)}+2\text{(column 2)} \implies 3\text{(column 1)}+2\text{(column 2)}-\text{(column 3)} = 0,$$

and hence a basis for $N(A)$ is

$$\boxed{\begin{pmatrix} 3 \\ 2 \\ -1 \end{pmatrix}}.$$

Of course, you could also do Gaussian elimination on A and find the special solution, but this is a lot of wasted effort when we already told you how column 3 depends on the other two columns!

(c) This won't work because A **is not full column rank**, so the square matrix A^TA is **not invertible**. (Note, however, that this formula does *not* require A to be "invertible" or "non-singular," or even be *square*! Nor does it require A to have orthogonal/orthonormal columns...just linearly independent.) Even more precisely, $\text{rank}(A^T A) = \text{rank}(A) = 2$, but $A^T A$ is a 3×3 matrix. Equivalently, $N(A^T A) = N(A) \neq \{\vec{0}\}$, which again implies that $A^T A$ is singular.

Instead, since we already have an orthonormal basis $Q = \begin{pmatrix} q_1 & q_2 \end{pmatrix}$ for $C(A)$, we can use the simpler projection formula for an orthonormal basis:

$$\boxed{P = QQ^T = q_1 q_1^T + q_2 q_2^T}.$$

Note that this QQ^T formula *only* works when Q has ortho*normal* columns, i.e. when $Q^T Q = I$. So, for example, $B = \begin{pmatrix} q_1 & 2q_2 \end{pmatrix}$ (the first two columns of A) would *not* give a projection matrix BB^T.

(d) The closest vector is the orthogonal projection. Since $N(A^T) = C(A)^\perp$, the orthogonal projection is obtained by simply **subtracting** the projection onto $C(A)$:

$$(I - P)x = x - QQ^T x = x - q_1 \underbrace{(q_1^T x)}_{=1} - q_2 \underbrace{(q_2^T x)}_{=1}$$

$$= \begin{pmatrix} 2 \\ 0 \\ 0 \\ 0 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix} = \boxed{\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}}.$$

(Alternatively, you could first compute the matrix $I - P$ and then multiply it by x, but this is a *lot* more work, similar to what you analyzed in problem 2a! An even more wasteful approach is to form A^T and do Gaussian elimination to find its special solutions. In this case, however, you can probably see by inspection that $N(A^T)$, which must be 2-dimensional, is spanned by $(1, 1, 1, 1)$ and by $(1, -1, 1, -1)$, or alternatively by $(1, 0, 1, 0)$ and $(0, 1, 0, 1)$, so you could do orthogonal projection with these vectors pretty easily; many students forgot to normalize them correctly, though.)

Problem 4 (3+4+4+6 points):

The nullspace $N(A)$ of the real matrix A is spanned by the vector $v = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$.

(a) Give as much true information as possible about the size (the number of rows and columns) of A.

(b) Give an eigenvector and eigenvalue of the matrix $B = (3I - A^T A)(3I + A^T A)^{-1}$.

(c) Aside from the eigenvalue identified in the previous part, all *other* eigenvalues λ of B must be (**circle/copy all that apply**): purely real, purely imaginary, zero, negative real part, positive real part, $|\lambda| < 1$, $|\lambda| > 1$, $|\lambda| \le 1$, $|\lambda| \ge 1$.

(d) Give a good approximate formula for of $B^n \begin{pmatrix} 0 \\ -1 \\ 0 \\ 8 \end{pmatrix}$ for large n. (Give an explicit numerical vector, possibly including simple functions of n like 2^n or n^3... no other abstract symbolic formulas.)

Solution:

(a) The matrix A **must have 4 columns** to multiply by v, and it must therefore have rank 3 to have a 1d nullspace. Hence there must be **at least 3 rows**.

(b) $N(A^T A) = N(A)$, so $A^T A$ has an eigenvalue of 0 and an eigenvector of v. Hence B has an eigenvalue of $(3 - 0)/(3 + 0) = \boxed{1}$ and an corresponding eigenvector $\boxed{v}$.

(c) Since the 4×4 matrix $A^T A$ is positive semidefinite (for *any* real A), and we identified its only zero eigenvalue (it has a 1d nullspace), all of its *other three* eigenvalues λ_a must be real and > 0. For any eigenvalue $\lambda_a > 0$ of $A^T A$, the corresponding eigenvalue of B is $\lambda = \frac{3-\lambda_a}{3+\lambda_a}$. these must be **purely real with magnitude** $|\lambda| < 1$**.** Of course, this also means $|\lambda| \le 1$.

(They *could* be zero if there is a $\lambda_a = 3$, or they *could* have positive real parts if $\lambda_a < 3$, but neither of these is *necessarily* true.)

(d) B is diagonalizable, so we imagine expanding $(0, -1, 0, 8)$ in the **basis of eigenvectors**. Then B^n multiplies each eigenvector by λ^n, and for large n this is dominated by the $\lambda = 1$ term (since all of the other eigenvalues have magnitude < 1, they give terms that are exponentially decaying with n). Furthermore, B is Hermitian with **orthogonal** eigenvectors (the same eigenvectors as those of $A^T A$), so we can find the $\lambda = 1$ component of $(0, -1, 0, 8)$ simply by **orthogonal projection** onto the corresponding eigenvector v:

$$B^n \begin{pmatrix} 0 \\ -1 \\ 0 \\ 8 \end{pmatrix} \approx 1^n \frac{vv^T}{v^T v} \begin{pmatrix} 0 \\ -1 \\ 0 \\ 8 \end{pmatrix} = \underbrace{\frac{1}{1^2+2^2+3^2+4^2}}_{=30} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \underbrace{\begin{pmatrix} 1 & 2 & 3 & 4 \end{pmatrix} \begin{pmatrix} 0 \\ -1 \\ 0 \\ 8 \end{pmatrix}}_{=30} = \boxed{\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}}.$$

Problem 5 (10 points):

Describe (give an explicit numerical result with as few unknowns as possible) all possible linear combinations of the vectors

$$a_1 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \; a_2 = \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}, \; a_3 = \begin{pmatrix} 1 \\ -1 \\ 3 \end{pmatrix}, \; a_4 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

that give the vector $x = \begin{pmatrix} 4 \\ -1 \\ 5 \end{pmatrix}$.

Solution:

We are trying to find all coefficients (c_1, c_2, c_3, c_4) of the linear combinations $c_1a_1 + c_2a_2 + c_3a_3 + c_4a_4$ such that

$$c_1a_1 + c_2a_2 + c_3a_3 + c_4a_4 = \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & -1 & 1 \\ 2 & 4 & 3 & 1 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix}}_{c} = \underbrace{\begin{pmatrix} 4 \\ -1 \\ 5 \end{pmatrix}}_{x}.$$

But this is exactly the **complete solution** to $Ac = x$. So, we need to follow the usual steps from exam 1, and it's good to remind ourselves of the process before we start chugging through arithmetic.

(a) Do Gaussian elimination on A. Determine the rank and pivot/free columns.

(b) Solve for a basis of the nullspace $N(A)$, e.g. the special solutions. (We expect at least a 1d nullspace since the rank of A *must* be ≤ 3). This will tell us the *non-uniqueness* of the solution.

(c) Solve for a particular solution c_p by setting the free variables to zero.

(d) The complete solution is then c_p plus any vector in $N(A)$, which will have *one unknown* (one free parameter) per special solution.

Now, let's carry this out. Gaussian elimination gives:

$$\underbrace{\begin{pmatrix} \boxed{1} & 1 & 1 & 1 \\ 0 & 2 & -1 & 1 \\ 2 & 4 & 3 & 1 \end{pmatrix}}_{A} \xrightarrow[r_3 - 2r_1]{} \begin{pmatrix} \boxed{1} & 1 & 1 & 1 \\ 0 & \boxed{2} & -1 & 1 \\ 0 & 2 & 1 & -1 \end{pmatrix} \xrightarrow[r_3 - r_2]{} \underbrace{\begin{pmatrix} \boxed{1} & 1 & 1 & 1 \\ 0 & \boxed{2} & -1 & 1 \\ 0 & 0 & \boxed{2} & -2 \end{pmatrix}}_{U}.$$

where we have put boxes around the pivots. A has rank 3 (it is **full row rank**), so we will **always** have solutions to $Ac = x$, and we expect a **1d nullspace**. To find a vector in the nullspace, we just set the free variable (c_4) to $= 1$ and then solve for the pivot variables by **back-substitution**:

$$U \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ 1 \end{pmatrix} = 0 \implies \begin{pmatrix} 1 & 1 & 1 \\ & 2 & -1 \\ & & 2 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \underbrace{\begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}}_{-(\text{free column})}$$

$$\implies \begin{array}{c} c_1 + c_2 + c_3 = -1 \implies c_1 = -2 \\ 2c_2 - c_3 = -1 \implies c_2 = 0 \\ c_3 = 1 \end{array}$$

$$\implies \text{special solution } c_n = \begin{pmatrix} -2 \\ 0 \\ 1 \\ 1 \end{pmatrix}.$$

Next, we find the particular solution by setting $c_4 = 0$ and solving for the pivot variables, again by back-substitution. However, remember that the right-hand-side has changed: we need to **apply the original elimination steps** to x:

$$\underbrace{\begin{pmatrix} 4 \\ -1 \\ 5 \end{pmatrix}}_{x} \xrightarrow[r_3 - 2r_1]{} \begin{pmatrix} 4 \\ -1 \\ -3 \end{pmatrix} \xrightarrow[r_3 - r_2]{} \begin{pmatrix} 4 \\ -1 \\ -2 \end{pmatrix}.$$

(We could have also done this by "augmenting" A with x originally, the usual trick for hand calculation.) We finally solve

$$U \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ & 2 & -1 \\ & & 2 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 4 \\ -1 \\ -2 \end{pmatrix}$$

$$\implies \begin{array}{c} c_1 + c_2 + c_3 = 4 \implies c_1 = 6 \\ 2c_2 - c_3 = -1 \implies c_2 = -1 \\ c_3 = -1 \end{array} \implies c_p = \begin{pmatrix} 6 \\ -1 \\ -1 \\ 0 \end{pmatrix}.$$

Hence, the complete solution is given by $c = c_p + \alpha c_n$ for any scalar α or, more explicitly, by the linear combinations

$$\boxed{6a_1 - a_2 - a_3 + \alpha(-2a_1 + a_3 + a_4)}$$

for any scalar α.

Problem 6 (8+8 points):

Professor May Trix is trying to construct an 18.06 homework question in which $\frac{dx}{dt} = Ax$ has the solution

$$x(t) = v_1 \cos(2t) + v_2 e^{-t} + v_3 \sin(2t)$$

for some **nonzero real** constant vectors v_1, v_2, v_3, and some initial condition $x(0)$. Help May construct A, v_1, v_2, v_3, and $x(0)$:

(a) Write down a numerical formula for a possible **real** matrix A such that A is as **small in size as possible** and where A contains **no zero entries**. Your formula can be left as a product of some matrices and/or matrix inverses — you *don't* need to multiply them out or invert any matrices, but you should give possible numeric values for all of the matrices in your formula. (You *don't* need to explicitly check that your A has no zero entries as long as zero entries seem unlikely. e.g. the inverse of a matrix with no special structure probably has no zero entries.)

(Note that there are many possible answers here, but they will all have certain things in common.)

(b) Using the numbers that you chose from the formula in your previous part, give possible corresponding (numeric) values for $x(0)$, v_1, v_2, and v_3.

Solution:

(a) **One approach** here that we can construct a possible A **from its diagonalization**, given what we know of the eigenvalues and by *choosing* appropriate eigenvectors. (We're doing exactly the same steps that we carry out in going from eigenvectors to the solution $x(t)$, just in reverse order!)

From $x(t)$, we know that A must have at **least 3 eigenvalues** $\lambda = -1, +2i, -2i$, each of which give $e^{\lambda t}$ terms in $x(t) = e^{At}x(0)$. (Note that $\cos(2t) = \frac{e^{2it}+e^{-2it}}{2}$ and $\sin(2t) = \frac{e^{2it}-e^{-2it}}{2i}$, which corresponds to the $\pm 2i$ eigenvalues—which must come in a complex-conjugate pair since A is real.) Therefore, the smallest possible matrix is 3×3. For the matrix A to be real, we must also have **complex-conjugate eigenvectors** for the $\pm 2i$ eigenvalues. Putting it all together, we can just pick some eigenvectors x_1, x_2, x_3 arbitrarily (anything we want as long as they are **linearly independent** and have a complex-conjugate pair $x_2 = \overline{x_3}$!) and then define A by its diagonalization:

$$\boxed{A = X\Lambda X^{-1} = \underbrace{\begin{pmatrix} 1 & 1+i & 1-i \\ 1 & i & -i \\ 1 & -i & i \end{pmatrix}}_{X = \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix}} \underbrace{\begin{pmatrix} -1 & & \\ & 2i & \\ & & -2i \end{pmatrix}}_{\Lambda} X^{-1}}.$$

You were *not* required to multiply this out, but if you did so you would find $A = \frac{1}{2}\begin{pmatrix} -4 & 5 & -3 \\ -4 & 3 & -1 \\ 4 & -5 & -1 \end{pmatrix}$ in this particular example. As could have been expected, this has no nonzero entries—generically we don't expect to see zeros appearing in a product like this unless X has a very special structure (e.g. triangular or diagonal).

Of course, there are **infinitely many possible answers here**, but they should all have the same basic pattern, just with different choices of eigenvectors. Any X matrix with linearly independent columns, two complex-conjugate columns and one real column (in the right spots), and no special pattern of zero entries (not triangular or diagonal) was acceptable. Note also that the eigenvectors cannot be purely real *or* purely imaginary, because otherwise their complex conjugates would not be linearly independent.

There is also a **completely different approach** which does not depend on eigenvalues and diagonalization at all! Instead, we could just write out $\frac{dx}{dt}$ for this $x(t)$ and try to find a matrix A such that $\frac{dx}{dt} = Ax$. One way to carry this out is to first write out $x(t)$ in matrix form:

$$x(t) = \underbrace{\begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix}}_{V} \begin{pmatrix} \cos(2t) \\ e^{-t} \\ \sin(2t) \end{pmatrix},$$

which gives

$$\begin{aligned} \frac{dx}{dt} &= V \begin{pmatrix} -2\sin(2t) \\ -e^{-t} \\ 2\cos(2t) \end{pmatrix} \\ &= \underbrace{V \begin{pmatrix} & & -2 \\ & -1 & \\ 2 & & \end{pmatrix}}_{=AV?} \begin{pmatrix} \cos(2t) \\ e^{-t} \\ \sin(2t) \end{pmatrix}. \end{aligned}$$

By comparing to $Ax(t)$, we immediately obtain

$$A = V \begin{pmatrix} & & -2 \\ & -1 & \\ 2 & & \end{pmatrix} V^{-1}.$$

which is the desired matrix for *any* invertible choice of V—just pick some random real numbers for the entries V (and as long as they have no special pattern you are almost certain to get all nonzero entries of A). Note that this also **immediately gives the answer for part (b)**, since the columns of V are the desired vectors, and $x(0) = v_1 + v_2$. Note also that this means that A is *similar* to the "anti-diagonal matrix" in the middle, whose eigenvalues you can straightforwardly compute to be -1 and $\pm 2i$ as above: its characteristic polynomial is $\lambda^2(\lambda+1)+4(\lambda+1) = (\lambda^2+4)(\lambda+1)$.

(b) For such a matrix A, we generically expect $x(t) = e^{At}x(0)$ to look like

$$x(t) = c_1 e^{-t} x_1 + c_2 e^{2it} x_2 + c_3 e^{-2it} x_3$$

for some coefficients c_1, c_2, c_3 determined by expanding $x(0)$ in the basis of the eigenvectors, i.e. $c = X^{-1}x(0)$. If $x(0)$ is real, then the coefficients c_2 and c_3 must be complex conjugates, leading to solutions of the form

$$x(t) = c_1 e^{-t} x_1 + 2\text{Re}\left[c_2 e^{2it} x_2\right].$$

We can pick c_1 and c_2 to be whatever we want (as long as they are nonzero, since we have to get both terms), so let's pick $c_1 = 1$ and $c_2 = 1/2$ (to cancel the 2). (You can pick something else of course.) This gives

$$x(t) = \underbrace{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}}_{v_2 = x_1} e^{-t} + \text{Re}\left[e^{2it} \begin{pmatrix} 1+i \\ i \\ -i \end{pmatrix} \right].$$

Using $\text{Re}\left[e^{2it}\right] = \cos(2t)$ and $\text{Re}\left[ie^{2it}\right] = -\sin(2t)$, we finally obtain

$$\boxed{x(t) = \underbrace{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}}_{v_2} e^{-t} + \underbrace{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}}_{v_1} \cos(2t) + \underbrace{\begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}}_{v_3} \sin(2t)},$$

where we have identified v_1, v_2, v_3. And of course $\boxed{x(0) = v_2 + v_1}$ will hold for any choices of numbers; in this particular case we get $\boxed{x(0) = (2, 1, 1)}$. More generally, if we choose $c_2 = 1/2$ as above, then

$$2\text{Re}\left[c_2 e^{2it} x_2\right] = \text{Re}\left[e^{2it} x_2\right] = \underbrace{\text{Re}\left[x_2\right]}_{v_1} \cos(2t) - \underbrace{\text{Im}\left[x_2\right]}_{-v_3} \sin(2t)$$

with $x(0) = v_2 + \text{Re}\left[x_2\right]$.

Problem 7 (8+8 points):

Suppose that we have a sequence of m data points (x_i, y_i) coming from a physics experiment that we want to fit to a line $cx + d$, where the coefficients c and d are chosen to minimize the sum of the squares of the errors. But, because some of the data points have more measurement error than others, we don't weight the errors equally in minimizing the error. In particular, suppose that we want to minimize:

$$E = \text{weighted error} = w_1(cx_1+d-y_1)^2+w_2(cx_2+d-y_2)^2+\cdots = \sum_{i=1}^{m} w_i(cx_i+d-y_i)^2.$$

where $w_1, w_2, \ldots, w_m > 0$ are some positive weights associated with each data point (more uncertain data points have smaller weight).

(a) To **convert this into a standard least-squares** problem, show that we can rewrite E in the form $E = \|Mu-v\|^2$ for some matrix M, an unknown vector u, and a known vector v — give explicit expressions for M, u, and v in terms of the points (x_i, y_i), the weights w_i, and the unknowns c and d.

(b) More generally, suppose that we are minimizing $E = (Ax-b)^T W(Ax-b)$ over $x \in \mathbb{R}^n$ where A is an $m \times n$ real matrix, b is an m-component real vector, and W is an $m \times m$ real-symmetric *positive-definite* "weight" matrix. Using the **properties of positive-definite matrices** from class, show that we can rewrite E as a **standard least-squares problem** $E = \|Mx - v\|^2$ for some matrix M and vector v: that is, explain how M and v could be related to A, W, and b.

Solution:

(a) To make it look like a standard least-squares problem, we can just rewrite each term in the sum as

$$w_i(cx_i+d-y_i)^2 = (cx_i\sqrt{w_i}+d\sqrt{w_i}-y_i\sqrt{w_i})^2 = \left([\text{linear in } c,d] - [\text{constant}]\right)^2,$$

which is the form of an ordinary least-squares problem. In particular, it leads to $E = \|Mu - v\|^2$ where:

$$\boxed{u = \begin{pmatrix} c \\ d \end{pmatrix}, \quad M = \begin{pmatrix} x_1\sqrt{w_1} & \sqrt{w_1} \\ x_2\sqrt{w_2} & \sqrt{w_2} \\ \vdots & \vdots \\ x_m\sqrt{w_m} & \sqrt{w_m} \end{pmatrix}, \quad v = \begin{pmatrix} y_1\sqrt{w_1} \\ y_2\sqrt{w_2} \\ \vdots \\ y_m\sqrt{w_m} \end{pmatrix}}.$$

(b) If W is positive-definite, then we said in class that it must be possible to factorize W in the form $\boxed{W = B^T B}$ for some (full column rank) B. hence

$$E = (Ax-b)^T B^T B(Ax-b) = \|B(Ax-b)\|^2 = \|(BA)x-(Bb)\|^2 = \|Mx-v\|^2$$

where $\boxed{M = BA, \quad v = Bb}$.

For example, in class we mentioned the possible choice $B = \sqrt{W}$. (Computationally, an easier choice would be a Cholesky factorization, but that's beyond the scope of 18.06.)