1. **Least $C$-norm solution.** Let $A \in \mathbb{R}^{m \times n}$ be full-rank and fat, and $C \in \mathbb{R}^{k \times n}$ be full-rank and tall.

(a) Find the solution to $Ax = b$ that minimizes $\|Cx\|$.

(b) Find the solution to $Ax = b$ that minimizes $\|C(x - x_0)\|$.

(c) Show that $\|Cx - d\|^2 - \|C(x - (C^TC)^{-1}C^Td)\|^2$ does not depend on $x$.

(d) Show that the solution to $Ax = b$ that minimizes $\|Cx - d\|$ is

$$x^* = (C^TC)^{-1}C^Td + (C^TC)^{-1}A^T(A(C^TC)^{-1}A^T)^{-1}(b - A(C^TC)^{-1}C^Td).$$

**Solution:**

(a) Since $A$ is full-rank and fat, $Ax = b$ has at least one solution. From Problem 4 in HW#4, we know that since $C$ is full-rank and tall, $\langle x, y \rangle_C := (Cx)^TCy$ defines a valid inner product. So the problem reduces to one of finding the least-norm solution to $Ax = b$ w.r.t. this inner product. Clearly, the required solution will be found by projecting the zero vector on the affine subspace defined by $Ax = b$, and this projection should be with respect to the inner product $\langle \cdot, \cdot \rangle_C$. Therefore, letting $\hat{x}$ be the required solution, we must have

$$A\hat{x} = b,$$

and from the orthogonality principle,

$$\hat{x} \perp_C (\hat{x} - z) \quad (1)$$

for every $z$ satisfying $Az = b$. $\Box$ shows that $\hat{x}^TC^TC(\hat{x} - z) = 0$ whenever $A(\hat{x} - z) = 0$, which implies that

$$\mathcal{N}(\hat{x}^TC^TC) \supseteq \mathcal{N}(A)$$

or in other words, $\mathcal{R}(C^TC\hat{x}) \subseteq \mathcal{R}(A^T)$, i.e., $C^TC\hat{x} \in \mathcal{R}(A^T)$. Writing

$$C^TC\hat{x} = A^T\tilde{y} \quad (2)$$

and noting that $A\hat{x} = b$, we have

$$A(C^TC)^{-1}A^T\tilde{y} = b.$$
Taking $M$ as a (symmetric) square root of $(C^T C)^{-1}$ (see Problem 5), which exists since $(C^T C)^{-1}$, being symmetric, is diagonalizable, we have $A(C^T C)^{-1} A^T = AM(AM)^T$, therefore
\[
\rank(A(C^T C)^{-1} A^T) = \rank(AM) \geq \rank(A) + \rank(M) - n = m + n - n = m,
\]
which shows that $A(C^T C)^{-1} A^T$ is invertible and
\[
\tilde{y} = (A(C^T C)^{-1} A^T)^{-1} b.
\]
This, together with (2), shows that
\[
\tilde{x} = A^T (A(C^T C)^{-1} A^T)^{-1} b.
\]
(b) The problem is equivalent to finding the solution to $A(x - x_0) = b - Ax_0$ that minimizes $\|C(x - x_0)\|$ and is therefore given by (using the result of part (a))
\[
\tilde{x} = x_0 + A^T (A(C^T C)^{-1} A^T)^{-1} (b - Ax_0).
\]
(c) We have
\[
\|Cx - d\|^2 - \|C(x - (C^T C)^{-1} C^T d)\|^2
= x^T C^T C x + d^T d - 2x^T C^T d - (x - (C^T C)^{-1} C^T d)^T C^T C (x - (C^T C)^{-1} C^T d)
= d^T d - 2x^T C^T d + 2x^T C^T C (C^T C)^{-1} C^T d - d^T C (C^T C)^{-1} (C^T C) (C^T C)^{-1} C^T d
= d^T d - 2x^T C^T d + 2x^T C^T d - d^T C (C^T C)^{-1} C^T d
= d^T (I - C (C^T C)^{-1} C^T) d,
\]
which is independent of $x$.
(d) Part (c) shows that finding the solution to $Ax = b$ that minimizes $\|Cx - d\|$ is equivalent to finding the solution to $Ax = b$ that minimizes $\|C(x - (C^T C)^{-1} C^T d)\|$, which is found immediately by replacing $x_0$ with $(C^T C)^{-1} C^T d$ in the result of part (b):
\[
x^* = (C^T C)^{-1} C^T d + A^T (A(C^T C)^{-1} A^T)^{-1} (b - A(C^T C)^{-1} C^T d).
\]
2. Eigenvalues. Suppose that $A$ has $\lambda_1, \lambda_2, \ldots, \lambda_n$ as its eigenvalues.
(a) Show that $\det(A) = \lambda_1 \cdot \lambda_2 \cdots \lambda_n$.
(b) Show that the eigenvalues of $A^T$ are $\lambda_1, \lambda_2, \ldots, \lambda_n$, that is, $A$ and $A^T$ have the same set of eigenvalues.
(c) Show that the eigenvalues of $A^k$ are $\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k$ for $k = 1, 2, \ldots$.
(d) Show that $A$ is invertible if and only if it does not have a zero eigenvalue.
(e) Suppose that $A$ is invertible. Show that the eigenvalues of $A^{-1}$ are $\lambda_1^{-1}, \lambda_2^{-1}, \ldots, \lambda_n^{-1}$.
(f) Show that $A$ and $T^{-1}AT$ have the same set of eigenvalues, that is, eigenvalues are invariant under a similarity transformation $A \mapsto T^{-1}AT$.

Solution:
(a) Consider the characteristic polynomial of $A$, namely, $\chi_A(\lambda) := \det(\lambda I - A)$. Clearly, the highest power of $\lambda$ in $\chi_A(\lambda)$, i.e., the $n^{th}$ power, occurs only in the term $\prod_{i=1}^{n} (\lambda - A_{ii})$. Therefore, the coefficient of $\lambda^n$ equals 1. The constant term is given by $\chi_A(0) = \det(-A) = (-1)^n \det(A)$. Therefore, we have

$$\lambda_1 \cdot \lambda_2 \cdots \lambda_n = \text{product of all roots of } \{\chi_A(\lambda) = 0\}$$

$$= (-1)^n \text{constant term}$$

$$= \det(A).$$

(b) We have

$$\chi_A^T(\lambda) = \det(\lambda I - A^T) = \det((\lambda I - A)^T) = \det(\lambda I - A) = \chi_A(\lambda),$$

which shows that $A^T$ and $A$ have identical characteristic polynomials and hence, identical eigenvalues.

(c) Consider the Jordan normal form of $A$, i.e., $A = T JT^{-1}$, where $J$ is upper-triangular and has the eigenvalues $\lambda_1, \ldots, \lambda_n$ as its diagonal entries. Then, $A^k = TJ^kT^{-1}$ and the diagonal entries of the upper-triangular matrix $J^k$ are $\lambda_1^k, \ldots, \lambda_n^k$ in the same order. Then, the eigenvalues of $J^k$ (and hence, of $A^k$, see part (f)) are given by $\lambda_1^k, \ldots, \lambda_n^k$.

Alternative proof: For any $\lambda \in \mathbb{C}$, let $\mu_1, \ldots, \mu_k$ be the $k^{th}$ roots of $\lambda$. Then we have

$$\prod_{j=1}^{k} (A - \mu_j I) = (-1)^k c_k I + \sum_{l=0}^{k-1} (-1)^l c_l A^{k-l},$$

where $c_0 = 1$, and for $1 \leq l \leq k$, $c_l$ is the sum of all possible products of the $\mu_j$s, taken $l$ at a time. For example, $c_k$ is simply $\prod_{j=1}^{k} \mu_j$.

Now, $\mu_1, \ldots, \mu_k$ are the roots of the polynomial $p(x) = x^k - \lambda = 0$, therefore $x^k - \lambda$ is identically equal to

$$\prod_{j=1}^{n} (x - \mu_j) = \sum_{l=0}^{k} (-1)^l c_l x^{k-l}.$$ 

Equating the coefficients of like powers of $x$, we therefore conclude that $c_l = 0$ for $l = 1, \ldots, k-1$, $c_0 = 1$, and $c_k = (-1)^{k-1} \lambda$.

Using these relations, (3) becomes

$$\lambda I - A^k = (-1)^{k-1} \prod_{j=1}^{k} (\mu_j I - A).$$

Now, if the characteristic polynomial $\chi_A(\lambda) := \det(\lambda I - A)$ be given by

$$\chi_A(\lambda) = \prod_{i=1}^{n} (\lambda - \lambda_i),$$
we have
\[
\chi_{A^k}(\lambda) = \det(\lambda I - A^k) \\
\quad = (-1)^n \prod_{j=1}^k \det(\mu_j I - A) \\
\quad = (-1)^n \prod_{j=1}^k \prod_{i=1}^n (\mu_j - \lambda_i) \\
\quad = (-1)^n (-1)^n \prod_{i=1}^n \prod_{j=1}^k (\lambda_i - \mu_j) \\
\quad = (-1)^n \prod_{i=1}^n (\lambda_i^k - \lambda) \\
\quad = \prod_{i=1}^n (\lambda - \lambda_i^k),
\]
which shows that the eigenvalues of \( A^k \) are exactly \( \lambda_1^k, \ldots, \lambda_n^k \).

(d) Let \( \lambda_1, \ldots, \lambda_n \) be the eigenvalues of \( A \). We have
\[
A \text{ is invertible } \iff \det(A) \neq 0 \iff \prod_{i=1}^n \lambda_i \neq 0 \iff \lambda_i \neq 0 \text{ for all } i.
\]

(e) If \( A \) is invertible, we know that \( \lambda_i \neq 0 \) for all \( i \). We have
\[
\chi_{A^{-1}}(\lambda) = \det(\lambda I - A^{-1}) \\
\quad = \det((\lambda A - I)A^{-1}) \\
\quad = \det(\lambda A - I) \det(A)^{-1} \\
\quad = \lambda^n \det(A - \lambda^{-1}I) \det(A)^{-1} \\
\quad = (-\lambda)^n \det(\lambda^{-1}I - A) \det(A)^{-1} \\
\quad = (-\lambda)^n \chi_A(\lambda) \det(A)^{-1} \\
\quad = (-1)^n \prod_{i=1}^n \left(\frac{\lambda}{\lambda_i}\right) \chi_A(\lambda^{-1}) \\
\quad = (-1)^n \prod_{i=1}^n \frac{\lambda}{\lambda_i}(\lambda^{-1} - \lambda_i) \\
\quad = \prod_{i=1}^n (\lambda - \lambda_i^{-1}),
\]
which shows that \( \lambda_1^{-1}, \ldots, \lambda_n^{-1} \) are the eigenvalues of \( A^{-1} \).

(f) We have
\[
\chi_{T^{-1}AT}(\lambda) = \det(\lambda I - T^{-1}AT) = \det(T^{-1}(\lambda I - A)T) = \det(T)^{-1} \det(\lambda I - A) \det(T) = \chi_A(\lambda),
\]
which shows that \( T^{-1}AT \) has the same eigenvalues as \( A \).
3. **Trace.** We define the *trace* of \( A \in \mathbb{R}^{n \times n} \) as
\[
\text{tr}(A) = A_{11} + A_{22} + \cdots + A_{nn},
\]
that is, the sum of its diagonal entries.

(a) Suppose that \( A \) has \( \lambda_1, \lambda_2, \ldots, \lambda_n \) as its eigenvalues. Show that
\[
\text{tr}(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n.
\]

(b) Show that
\[
\text{tr}(A^k) = \lambda_1^k + \lambda_2^k + \cdots + \lambda_n^k
\]
for \( k = 1, 2, \ldots \).

**Solution:**

(a) We have
\[
\lambda_1 + \cdots + \lambda_n = \text{-( coefficient of } \lambda^{n-1} \text{ in } \chi_A(\lambda)).
\]
Now, in \( \chi_A(\lambda) = \det(\lambda I - A) \), the only term containing \( \lambda^n \) and \( \lambda^{n-1} \) is \( \prod_{i=1}^{n}(\lambda - A_{ii}) \).
(This is immediately clear by considering the definition of a determinant in terms of permutations.) Therefore, the coefficient of \( \lambda^{n-1} \) in \( \chi_A(\lambda) \) is the same as the coefficient of \( \lambda^{n-1} \) in \( \prod_{i=1}^{n}(\lambda - A_{ii}) \), which is given by \( -\text{tr}(A) \). Therefore,
\[
\text{tr}(A) = \lambda_1 + \cdots + \lambda_n.
\]

(b) Using problems 3(a) and 2(c), the result is immediate.

4. **Matrix exponential.**

(a) Show that \( e^{A+B} = e^A e^B \) if \( A \) and \( B \) *commute*, namely, \( AB = BA \).

(b) Show that \( A \) and \( e^{tA} \) commute for every \( t \in \mathbb{R} \).

(c) Suppose that \( \lambda_1, \lambda_2, \ldots, \lambda_n \) are eigenvalues of \( A \). Show that the eigenvalues of \( e^A \) are \( e^{\lambda_1}, e^{\lambda_2}, \ldots, e^{\lambda_n} \).

(d) Show that \( \det(e^A) = e^{\text{tr}(A)} \).

(e) Show that
\[
\frac{d}{dt} e^{tA} = Ae^{tA} = e^{tA} A.
\]

(f) Show that
\[
e^A = \lim_{k \to \infty} \left( I + \frac{A}{k} \right)^k.
\]

**Solution:**
(a) Consider

\[ e^{(A+B)} = \sum_{n=0}^{\infty} \frac{(A + B)^n}{n!} \]

\[ = \sum_{n=0}^{\infty} \sum_{k=0}^{n} \frac{n!}{n!} A^k B^{n-k} \]

\[ = \sum_{n=0}^{\infty} \sum_{k=0}^{n} \frac{A^k B^{n-k}}{n!} \]

\[ = \sum_{n,k,n \geq k} \frac{A^k B^{n-k}}{n!} \]

\[ = e^B[I + \frac{A^1}{1!} + \frac{A^2}{2!} + \cdots] \]

\[ = e^B e^A, \]

where (a) follows since \( A \) and \( B \) commute and (b) follows by grouping all terms according to the corresponding power of \( A \).

(b) Note that \( AA^i = A^i A = A^{(i+1)} \). We then have \( A e^{tA} = A[I + \frac{A^1}{1!} + \frac{A^2}{2!} + \cdots] = A + \frac{A^2}{2!} + \cdots = [I + \frac{A}{1!} + \frac{A^2}{2!} + \cdots]A = e^{tA}A \). Thus \( A \) and \( e^{tA} \) commute.

(c) Consider any eigenvector \( v_1 \) associated with \( \lambda_1 \). Note that \( A^k v_1 = A^{(k-1)} A v_1 = \lambda_1 A^{(k-1)} v_1 \). Continuing this way, we get \( A^k v_1 = \lambda_1^k v_1 \). We then have \( e^A v_1 = v_1 + \frac{A v_1}{1!} + \frac{A^2 v_1}{2!} + \cdots = [I + \frac{\lambda_1}{1!} + \frac{\lambda_1^2}{2!} + \cdots]v_1 = e^{\lambda_1} v_1 \). Thus, \( e^{\lambda_1} \) is an eigenvalue for \( e^A \) with associated eigenvector \( v_1 \). Similarly, we can show that for all \( i \in \{1, 2, \cdots, n\} \), \( e^{\lambda_i} \) is an eigenvalue of \( e^A \).

More generally, consider the Jordan canonical form of \( A \), i.e., \( A = T^{-1} J T \). Then, we have

\[ e^A = T^{-1} e^J T, \]

and \( e^J \) consists of exponentials of the Jordan blocks arranged in the same order. Now, consider a Jordan block

\[ J = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ \lambda & 1 & \cdots & 0 \\ \lambda & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots \\ \lambda & \cdots & \cdots & \cdots & \lambda \end{bmatrix}. \]

By elementary matrix multiplication, we see that for all non-negative integers \( m \), \( J^m \) is an upper triangular matrix with diagonals equal to \( \lambda^m \). Therefore, \( e^J \) is an upper triangular matrix whose diagonals are \( e^{\lambda_1}, \ldots, e^{\lambda_n} \). Therefore, since eigenvalues remain invariant under similarity transformations, and the eigenvalues of an upper-triangular matrix are simply its diagonal entries, the eigenvalues of \( e^A \) are \( e^{\lambda_1}, \ldots, e^{\lambda_n} \).
(d) Since determinant is the product of the eigenvalues, using the result in part (c) we get 
\[ \det(e^A) = \prod_{i=1}^{n} e^{\lambda_i} = e^{\sum_{i=1}^{n} \lambda_i} = e^{\text{tr}(A)} \]
since \( \sum_{i=1}^{n} \lambda_i = \text{tr}(A) \)
(e) 
\[
\frac{d}{dt}e^{tA} = \frac{d}{dt}(I + \sum_{k=1}^{\infty} \frac{t^k}{k!}A^k) = \sum_{k=1}^{\infty} \frac{t^{k-1}}{(k-1)!}A^k = A[I + \sum_{k=1}^{\infty} \frac{t^k}{k!}I] = Ae^{tA} = e^{tA}A,
\]
where the final step follows from part (b).
(f) Consider the Jordan canonical form \( A = T^{-1}JT \). Then for all \( k \in \mathbb{N} \),
\[
\left( I + \frac{A}{k} \right)^k = T^{-1} \left( I + \frac{J}{k} \right)^k T.
\]
Now, consider an \( n \times n \) Jordan block \( J \) such that
\[
J_{ij} = \begin{cases} 
\lambda, & j = i \\
1, & j = i + 1 \\
0, & \text{otherwise}.
\end{cases}
\]
Note that \( kI + J \) is also a Jordan block with \( \lambda + k \) on the diagonal. Therefore, for \( k > n \),
\[
M^{(k)} := \left( I + \frac{J}{k} \right)^k = \frac{(kI + J)^k}{k^k}
\]
is an upper-triangular matrix whose \((i,j)\)th entry is given by (for \( j \geq i \))
\[
\left( \frac{k}{j-i} \right) \frac{(\lambda + k)^{j-i}}{k^k}.
\]
Taking limits, we have
\[
\lim_{k \to \infty} \frac{M^{(k)}_{ij}}{k_i} = \lim_{k \to \infty} \frac{k}{j-i} \frac{(\lambda + k)^{j-i}}{k^k} = \frac{1}{(j-i)!} \lim_{k \to \infty} \left( 1 + \frac{\lambda}{k} \right)^k \frac{k!}{(k+i-j)! (\lambda + k)^{j-i}} = \frac{1}{(j-i)!} \lim_{k \to \infty} \left( 1 + \frac{\lambda}{k} \right)^k \prod_{m=1}^{j-i} \frac{k-1}{k+\lambda} = \frac{e^{\lambda}}{(j-i)!}
\]
for \( j \geq i \). Therefore, \( \lim_{k \to \infty} \left( I + \frac{J}{k} \right)^k = \lim_{k \to \infty} M^{(k)} = e^J \), whence
\[
\lim_{k \to \infty} \left( I + \frac{A}{k} \right)^k = T^{-1} \lim_{k \to \infty} \left( I + \frac{J}{k} \right)^k T = T^{-1}e^J T = e^A.
\]
5. Square root of a matrix. Let \( A \in \mathbb{R}^{n \times n} \) be diagonalizable with eigenvalue decomposition \( T \Lambda T^{-1} \). We say that a matrix \( B \) is a square root of \( A \) if \( B^2 = A \). Show that \( B = T \Lambda^{1/2} T^{-1} \) is a square root of \( A \), where \( \Lambda^{1/2} \) is a diagonal matrix with entries \( \gamma_1, \gamma_2, \ldots, \gamma_n \) such that \( \gamma_i^2 = \lambda_i \).

**Solution:** \( B^2 = T \Lambda^{1/2} T^{-1} T \Lambda^{1/2} T^{-1} = T \Lambda^{1/2} \Lambda^{1/2} T^{-1} \). Since \( \Lambda^{1/2} \) is a diagonal matrix, \( \Lambda^{1/2} \Lambda^{1/2} \) will simply be \( \text{diag}(\gamma_1^2, \gamma_2^2, \ldots, \gamma_n^2) = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) \). Hence, \( T \Lambda^{1/2} \Lambda^{1/2} T^{-1} = T \Lambda T^{-1} = A \). Thus, \( B^2 = A \) and \( B \) is a square root of \( A \).