1. Find the $n \times 1$ vector $\hat{x}$ that minimizes $||x - \hat{x}||^2 = (x - \hat{x})^T (x - \hat{x})$ for the two cases below.
   
   (a) Under the constraint $\hat{x} = H \theta$ where $H$ is $n \times p$ and of rank $p$.
   
   (b) Under the constraint $A^T x = 0$ with $A$ an $n \times p$ matrix of rank $p$.

2. Consider the quadratic form

   $$x^T Q x = \begin{bmatrix} x_1^T & x_2^T \end{bmatrix} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$ 

   Show that for fixed $x_2$, $x^T Q x$ is minimized when $\hat{x}_1 = -Q_{11}^{-1} Q_{12} x_2$. Similarly, show that $\hat{x}_2 = -Q_{22}^{-1} Q_{21} x_1$.

3. Begin with the identity

   $$I(a) = \int_{-\infty}^{\infty} e^{-ax^2} \, dx = \sqrt{\pi/a}.$$ 

   Differentiate $n$ times with respect to $a$ to derive the identity

   $$\int_{-\infty}^{\infty} x^{2n} e^{-ax^2} \, dx = \frac{1 \cdot 3 \cdot \cdots \cdot (2n-1)}{2^n} \sqrt{\pi/a^{2n+1}}.$$ 

   Use this result to find the even moments $E X^{2n}$ for an $N(0, \sigma^2)$ random variable. Now find the even and odd moments for an $N(\theta, \sigma^2)$ random variable.

4. Let $X = \begin{bmatrix} X_1 & X_2 \end{bmatrix}$ denote a bivariate normal random vector. Assume that

   $$E X = 0 \quad \text{and} \quad E X X^T = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$ 

   Define $Y_1 = X_1 + X_2$ and $Y_2 = -X_1 + X_2$.

   (a) Find the joint distribution of $Y_1$ and $Y_2$; find the marginal distributions of $Y_1$ and $Y_2$.
   
   (b) Find the conditional density of $X_1$ given $Y_1$; find the conditional density of $X_1$ given $Y_2$.
   
   (c) Find the conditional mean and variance of $X_1$ given $Y_1$; find the conditional mean and variance of $X_1$ given $Y_2$.

5. If $X_1, X_2, X_3, \text{and} \ X_4$ are zero mean and jointly Gaussian random variables, prove the following moment factoring property

   $$E\{X_1 X_2 X_3 X_4\} = E\{X_1 X_2\} E\{X_3 X_4\} + E\{X_1 X_3\} E\{X_2 X_4\} + E\{X_2 X_3\} E\{X_1 X_4\}.$$ 

   Use the moment generating property of the characteristic function.

6. For a scalar rv $X$ the characteristic function is defined to be $\Phi_X(\lambda) = E\{e^{j \lambda X}\}$, this is the Fourier transform of the pdf of $X$ (with a sign change). For a vector rv $Y$ the characteristic function is defined to be $\Phi_Y(\lambda) = E\{e^{j \lambda^T Y}\}$, this is the multidimensional Fourier transform of the joint pdf of $Y$ (with a sign change).

   (a) If $X \sim N(0,1)$ find its characteristic function by a direct computation.
   
   (b) If $X \sim N(\mu, \sigma^2)$ find its characteristic function using the result of (a).
   
   (c) If $Y \sim N(0, I)$ (a vector rv) find its multidimensional characteristic function using the result of (a).
(d) If $Y \sim \mathcal{N}(\mu, \Sigma)$ (a vector rv) find its multidimensional characteristic function using the the previous results.

7. An $n$-dimensional rv $X$ is said to be Gaussian if its multi-dimensional characteristic function is of the form
\[ \Phi_X(\lambda) = e^{i\lambda^T\mu - 0.5\lambda^T\Sigma\lambda} \]
where $\mu$ is the mean vector and $\Sigma$ is the covariance matrix (possibly singular).

(a) When the covariance matrix is non-singular the pdf of $X$ has the familiar form. Find that form and write it down.

(b) On the otherhand when the covariance matrix is singular the pdf does not exist. Suppose that rank $\Sigma = r < n$. Show that there exists an $r$-dimensional Gaussian random vector $Y$ and matrices $A, b$ such that
\[ X = AY + b \]
with probability 1.
Problem: Find the n x 1 vector $\hat{x}$ that minimizes

$$\|x - \hat{x}\|^2 = (x - \hat{x})^T(x - \hat{x})$$

for

(a) Case where $\hat{x} = H\Theta$ with $H$ n x p of rank p

(b) Case where $A^T\hat{x} = 0$ with $A$ n x p of rank p.

(a) Can think of this problem with the language of Hilbert space as was done in class or more directly using ideas from linear algebra.

Let $H$ be n x p of rank p. Then of necessity p ≤ n so that an equation of the form

$$H\Theta = x$$

is known; $\Theta$ to be found; is overdetermined i.e. there are more equations than unknowns. For a given $x$ there may not be a solution.

If there is no solution (i.e. if $x \notin \mathbb{R}(H)$ = column space) then one can try to minimize the residual i.e.

$$\|r\|^2 = \|x - H\Theta\|^2$$

$$= (x - H\Theta)^T(x - H\Theta)$$

$$= x^Tx - x^TH\Theta - \Theta^TH^Tx + \Theta^TH^TH\Theta$$

Can take the gradient to find a stationary point of the error $\|r(\Theta)\|^2$ as a function of $\Theta$. 
\[
\frac{\partial}{\partial \theta} \{ \| \theta \|^2 \} = \begin{bmatrix}
\frac{\partial \| \theta \|^2}{\partial \theta_1} \\
\frac{\partial \| \theta \|^2}{\partial \theta_2} \\
\vdots \\
\frac{\partial \| \theta \|^2}{\partial \theta_p}
\end{bmatrix}^T = 0_{p \times 1}
\]

Then we can show (working it out by coordinates to check) that
\[
\frac{\partial}{\partial \theta} \{ \| \theta \|^2 \} = -2H^T x + 2H^TH \theta \\
= -2H^T(x - H \theta)
\]

A value of \( \theta \) minimizing the form \( \| \theta \|^2 \) is called the least squares solution. It solves
\[
H^T(x - H \theta) = 0 \quad (\star)
\]

Note also that the Hessian
\[
\frac{\partial}{\partial \theta} \left( \frac{\partial \| \theta \|^2}{\partial \theta} \right)^T = 2H^TH \geq 0
\]

is non-negative definite and, in fact, strictly positive definite if \( H \) is full rank as we have here assumed. A basic result of multi-variate calculus can be used to conclude that solutions of \( (\star) \) minimize \( \| \theta \|^2 \).

Another way to write \((\star)\) is to say that a least squares solution \( \hat{\theta}_L \) must solve the square \( p \times p \) system
\[
H^T H \hat{\theta}_L = H^T x
\]
called the Normal Equations.
It can be shown that \( \text{rank}(H^TH) = \text{rank}(H) \). Therefore, subject to assumption that \( H \) has full column rank, then \( H^TH \) must be invertible. So

\[
\hat{\Theta}_{LS} = (H^TH)^{-1} H^T x
\]

whence

\[
\hat{x} = H \hat{\Theta}_{LS} = H (H^TH)^{-1} H^T x \quad \text{QED} \]

6. The standard and direct approach is to recognize the problem is a constrained optimization

\[
\min_{\hat{x} \in \mathbb{R}^n} \| x - \hat{x} \|^2 \quad \text{subject to} \quad H^T \hat{x} = 0
\]

where \( H \) is \( n \times p \) of rank \( p \). Thus it is equivalent to \( p \) linear constraints.

Approach is to form Lagrangian:

\[
L(\hat{x}, \lambda) = (x - \hat{x})^T(x - \hat{x}) + \lambda^T H^T \hat{x}
\]

and seek a stationary point from

\[
\frac{\partial L}{\partial \hat{x}} = 0 \quad \frac{\partial L}{\partial \lambda} = 0 \quad \hat{x} \in \mathbb{R}^n \quad \lambda \in \mathbb{R}^p
\]

\[
\frac{\partial L}{\partial \hat{x}} = \frac{2}{\partial \hat{x}} \left\{ x^T - \hat{x}^T x - \hat{x}^T \hat{x} + \hat{x}^T \hat{x} + \lambda^T H^T \hat{x} \right\} = -2x + 2\hat{x} + H\lambda = 0
\]
\[
\frac{\partial l}{\partial \lambda} = H^T \hat{x} = 0
\]

Need to solve the above equations simultaneously

\[H\lambda = 2(x - \hat{x}) \Rightarrow H^TH\lambda = 2H^T(x - \hat{x}) = 2H^Tx\]

\[\Rightarrow \lambda = 2(H^TH)^{-1}H^Tx\]

Substituting back into the first equation

\[0 = -2x + 2\hat{x} + 2H(H^TH)^{-1}H^Tx\]

\[\Rightarrow \hat{x} = \left( I - H(H^TH)^{-1}H^T \right)x\]

Note that this is the orthogonal projection of \(x\) onto

\[\mathcal{R}(H)^\perp = \eta(H^T)\]

\(\text{QED}\)
Consider the quadratic form

\[ x^T Q x = \begin{bmatrix} x_1^T & x_2^T \end{bmatrix} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}. \]

Show that for fixed \( x_2 \), \( x^T Q x \) is minimized when \( \hat{x}_1 = -Q_{11}^{-1} Q_{12} x_2 \). Similarly, show that \( \hat{x}_2 = -Q_{22}^{-1} Q_{21} x_1 \).
Quadratic Forms
A symmetric matrix \( Q \) is said to be nonneg definite if
\[
x^T Q x \geq 0 \quad \text{for all vectors } x
\]
It is said to be positive definite if
\[
x^T Q x > 0 \quad \text{for all vectors } x \neq 0.
\]
Consider the following partition of the quadratic form induced by partitioning the vector \( x \)
\[
x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}
\]
\[
x^T Q x = \begin{bmatrix} x_1^T \\ x_2^T \end{bmatrix} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12}^T & Q_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}
\]
\[
= x_1^T Q_{11} x_1 + x_1^T Q_{12} x_2 + x_2^T Q_{12}^T x_1 + x_2^T Q_{22} x_2
\]
Then consideration of \( x = \begin{bmatrix} 0 \\ x_2 \end{bmatrix} \) and \( x = \begin{bmatrix} x_1 \\ 0 \end{bmatrix} \) makes it easy to see:

\( Q \) nonneg def. \( \Rightarrow \) \( Q_{11}, Q_{22} \) nonneg def.
\( Q \) pos. def. \( \Rightarrow \) \( Q_{11}, Q_{22} \) pos. def.

For simplicity in following assume that \( Q \) is pos. definite. Consider the problem of minimizing
\[
f(x) = x^T Q x = f(x_1, x_2)
\]
with respect to \( x_2 \) for a fixed value of \( x_1 \). To find the value of \( x_2 \), minimizing the form we use our above expression, take the gradient and set it equal to zero:
\[
\frac{\partial f(x_1, x_2)}{\partial x_2} = 2 Q_{22} x_2 + 2 Q_{12}^T x_1 = 0
\]
\[
\therefore x_2 = -Q_{22} Q_{12}^T x_1.
\]
If this choice \( \hat{x}_2 \) \((= \hat{x}_2(x_1))\) is put back into the quadratic form we can see

\[
\begin{align*}
\hat{f}(x_1, \hat{x}_2) &= x_1^T Q_{11} x_1 + x_1^T Q_{12} (-Q_{22}^{-1} Q_{12}^T x_1) + (-Q_{22}^{-1} Q_{12}^T x_1) Q_{12}^T x_1 \\
& \quad + (-Q_{22}^{-1} Q_{12}^T x_1)^T Q_{22} (-Q_{22}^{-1} Q_{12}^T x_1) \\
& = x_1^T Q_{11} x_1 - x_1^T Q_{12} Q_{22}^{-1} Q_{12}^T x_1 - x_1^T Q_{12} Q_{22}^{-1} Q_{22} x_1 \\
& \quad + x_1^T Q_{12} Q_{22}^{-1} Q_{22} x_1 \\
& = x_1^T (Q_{11} - Q_{12} Q_{22}^{-1} Q_{12}^T) x_1 \geq 0
\end{align*}
\]

The matrix \( Q_{11} - Q_{12} Q_{22}^{-1} Q_{12}^T \) is called a Schur complement of the partitioned matrix \( Q \). When it exists it must be nonneg. def. and under our simplifying hypothesis \( Q \succ 0 \)

\[
Q \succ 0 \Rightarrow Q_{11} \succ 0, \quad Q_{11} - Q_{12} Q_{22}^{-1} Q_{12}^T \succ 0
\]

In addition can write the original quadratic form as:

\[
f(x) = f(x_1, x_2) = (x_2 - \hat{x}_2)^T Q_{22} (x_2 - \hat{x}_2) + x_1^T (Q_{11} - Q_{12} Q_{22}^{-1} Q_{12}^T) x_1
\]

which can be seen simply by multiplying it all out. Another way to say the same thing is as the block diagonal representation

\[
Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12}^T & Q_{22} \end{bmatrix}
\]

\[
= \begin{bmatrix} I & Q_{12} Q_{22}^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} Q_{11} - Q_{12} Q_{22}^{-1} Q_{12}^T & 0 \\ 0 & Q_{22} \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & Q_{22}^{-1} Q_{12}^T \end{bmatrix}
\]

Again just multiply it out.
Begin with the identity

\[ I(a) = \int_{-\infty}^{\infty} e^{-ax^2} \, dx = \sqrt{\pi/a}. \]

Differentiate \( n \) times with respect to \( a \) to derive the identity

\[ \int_{-\infty}^{\infty} x^{2n} e^{-ax^2} \, dx = \frac{1 \cdot 3 \cdots (2n-1)}{2^n} \frac{\sqrt{\pi/a^{2n+1}}}{\sqrt{a^{2n+1}}}. \]

Use this result to find the even moments \( \text{E}X^{2n} \) for an \( \mathcal{N}(0, \sigma^2) \) random variable. Now find the even and odd moments for an \( \mathcal{N}(\theta, \sigma^2) \) random variable.
\begin{align*}
I(a) &= \int_{-\infty}^{\infty} e^{-ax^2} \, dx = (\pi a)^{1/2} \\
\frac{d}{da} I(a) &= -\int_{-\infty}^{\infty} x^2 e^{-ax^2} \, dx = -\frac{1}{2} \sqrt{\pi a}^{-3/2} \Rightarrow \int_{-\infty}^{\infty} x^2 e^{-ax^2} \, dx = \frac{1}{2} \sqrt{\pi a}^{-3/2} \\
\frac{d^2}{da^2} I(a) &= -\int_{-\infty}^{\infty} x^4 e^{-ax^2} \, dx = -\frac{3}{2} \sqrt{\pi a}^{-5/2} \Rightarrow \int_{-\infty}^{\infty} x^4 e^{-ax^2} \, dx = \frac{3}{2} \sqrt{\pi a}^{-5/2}
\end{align*}

Suppose that
\[
\int_{-\infty}^{\infty} x^k e^{-ax^2} \, dx = \frac{1 \cdot 3 \cdots (2k-1)}{2^k} \sqrt{\pi/a}^{-(2k+1)/2}
\]

Then taking the derivative of above wrt \( a \)
\[
\int_{-\infty}^{\infty} x^{k+2} e^{-ax^2} \, dx = \frac{1 \cdot 3 \cdots (2k-1)}{2^k} \sqrt{\pi/a}^{-(2k+1)/2} \\
\int_{-\infty}^{\infty} x^{2k+2} e^{-ax^2} \, dx = \frac{1 \cdot 3 \cdots (2k-1)(2k+1)}{2^{k+1}} \sqrt{\pi/a}^{-(2k+3)/2}
\]

Result holds for all \( k \geq 1 \) integer by induction. Note that differentiation under integral operator is justified by continuous differentiability of integrand.

Now consider a random variable \( X \sim \mathcal{N}(0, \sigma^2) \). The even moments are given by
\[
EX^{2n} = \frac{1}{\sqrt{2\pi} \sigma} \int_{-\infty}^{\infty} x^{2n} e^{-x^2/2\sigma^2} \, dx
\]
Using the formula with $a = \frac{1}{2a^2}$ we should have:

$$E X^{2n} = \frac{1}{(2\pi)^n} \frac{1 \cdot 3 \cdots (2n-1)}{2^n} \frac{1}{\sqrt{2\pi (2a^2)^{n+1}}}$$

$$= 1 \cdot 3 \cdots (2n-1) a^{2n}$$

The odd moments must vanish because $e^{-x^2/(2a^2)}$ is even and $x^{2n+1}$ is odd.

Now suppose that $X \sim \eta(\theta, a^2)$. It is easy to see that $\tilde{X} = X - \theta$ is distributed $\eta(0, \sigma^2)$. Then obviously have

$$E \tilde{X}^{2n+1} = E \{(X - \theta)^{2n+1}\} = 0 \quad \text{for } n \geq 1$$

$$E \tilde{X}^{2n} = E \{(X - \theta)^{2n}\} = 1 \cdot 3 \cdots (2n-1) \sigma^{2n} \quad n \geq 1.$$ 

One could clearly compute any of the moments $EX^n$ from the above but it does not result in a simple formula as in the case of zero mean.
Let \( X = \begin{bmatrix} X_1 & X_2 \end{bmatrix} \) denote a bivariate normal random vector. Assume that

\[
\mathbb{E}X = 0 \quad \text{and} \quad \mathbb{E}XX^T = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.
\]

Define \( Y_1 = X_1 + X_2 \) and \( Y_2 = -X_1 + X_2 \).

(a) Find the joint distribution of \( Y_1 \) and \( Y_2 \); find the marginal distributions of \( Y_1 \) and \( Y_2 \).

(b) Find the conditional density of \( X_1 \) given \( Y_1 \); find the conditional density of \( X_1 \) given \( Y_2 \).

(c) Find the conditional mean and variance of \( X_1 \) given \( Y_1 \); find the conditional mean and variance of \( X_1 \) given \( Y_2 \).
\[
X = [X_1 \; X_2]^T \ \text{bivariate Normal with } \ E X = 0 \text{ and } \ E X X^T = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}
\]

Then define \( Y_1 = X_1 + X_2 \) \( Y_2 = -X_1 + X_2 \).

(a) Find joint dist. of \( Y_1 \) and \( Y_2 \) and marginal distributions of \( Y_1 \) and \( Y_2 \). Notice that

\[
Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \Rightarrow Y \text{ is bivariate normal since it is a linear transform of biv. normal.}
\]

\[
EY = \begin{bmatrix} 1 \\ -1 \end{bmatrix} E X = 0
\]

\[
E Y Y^T = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 2 \begin{bmatrix} \rho + 1 & 0 \\ 0 & 1 - \rho \end{bmatrix}
\]

\[
\therefore \ Y \sim \mathcal{N}(0, 2 \begin{bmatrix} \rho + 1 & 0 \\ 0 & 1 - \rho \end{bmatrix}) \text{ Note that } Y_1 \text{ and } Y_2 \text{ are indep.}
\]

For marginal densities clear that \( E Y_1 = E Y_2 = 0 \). Also

\[
E Y_1^2 = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 2(\rho + 1) \text{ or directly from above.}
\]

\[
E Y_2^2 = 2(1 - \rho).
\]

\[
\therefore Y_1 \sim \mathcal{N}(0, 2^{\rho + 2}) \quad Y_2 \sim \mathcal{N}(0, 2^{2 - 2\rho}).
\]

(b) Conditional densities
(c) Cond. mean & variances

\( X_1 | Y_1 \) and \( X_1 | Y_2 \) are Normally distributed so its enough to find the conditional means and variances. This can be done directly or by using the formulas from class.

\[
X_1 | Y_1 \sim E \{X_1 Y_1 \} \frac{1}{2^{\rho + 2}} Y_1 = \frac{\rho + 1}{z(\rho + 1)} Y_1 = Y_1 / z \quad (\text{cond. mean})
\]

\[
1 - \left( \frac{\rho + 1}{2(\rho + 1)} \right)^2 = 1 - \frac{1}{2} (\rho + 1) = \frac{1}{2} (1 - \rho) \quad (\text{cond. var}).
\]

\[
X_1 | Y_2 \sim \frac{\rho - 1}{2(1 - \rho)} Y_2 = -\frac{Y_2}{z} \quad (\text{cond. mean})
\]

\[
\frac{1}{2} (\rho + 1) \quad (\text{cond. variance})
\]
If $X_1, X_2, X_3,$ and $X_4$ are zero mean and jointly Gaussian random variables, prove the following moment factoring property

$$E\{X_1 X_2 X_3 X_4\} = E\{X_1 X_2\}E\{X_3 X_4\} + E\{X_1 X_3\}E\{X_2 X_4\} + E\{X_2 X_3\}E\{X_1 X_4\}.$$

Use the moment generating property of the characteristic function.
\[ X = [x_1, x_2, x_3, x_4]^T \sim \eta(0, \Sigma) \Rightarrow \Phi_X(x) = E\{e^{i\mathbf{ax}^T \Sigma^{-1} x}\} = \exp\left\{-\frac{1}{2} \mathbf{a}^T \Sigma \mathbf{a} \right\}\]

The simplest way to go is to use the fact that

\[ E\{x_1 x_2 x_3 x_4\} = \left. \Phi_X(x_1, x_2, x_3, x_4) \right|_{x=0} \]

Now

\[ -\frac{1}{2} \mathbf{a}^T \Sigma \mathbf{a} = -\frac{1}{2} \sum_{k=1}^{4} \sum_{l=1}^{4} Z_{kl} a_k a_l \quad Z_{kl} = E\{x_k x_l\} \]

Of course the gradient of the form above is:

\[ \frac{\partial}{\partial \mathbf{a}} \left\{ -\frac{1}{2} \mathbf{a}^T \Sigma \mathbf{a} \right\} = \left[ \begin{array}{c} \frac{\partial}{\partial a_1} \\ \vdots \\ \frac{\partial}{\partial a_4} \end{array} \right] \left\{ 0 \right\} = -\mathbf{Z} \mathbf{a} \]

Furthermore

\[ \frac{\partial}{\partial a_k} \Phi(x) = \frac{\partial}{\partial a_k} \left\{ -\frac{1}{2} \mathbf{a}^T \Sigma \mathbf{a} \right\} \exp\left\{-\frac{1}{2} \mathbf{a}^T \Sigma^{-1} \mathbf{a} \right\} \]

\[ = - \left[ 0 \cdots 1 \cdots 0 \right] Z_{kl} \mathbf{a} \exp\left\{-\frac{1}{2} \mathbf{a}^T \Sigma^{-1} \mathbf{a} \right\} \]

\[ \text{at } k^{th} \text{ place.} \]

\[ \frac{\partial^2}{\partial a_k^2} \Phi(x) = - \left[ 0 \cdots 0 \cdots 1 \cdots 0 \right] Z_{kl} \mathbf{a} \exp\left\{-\frac{1}{2} \mathbf{a}^T \Sigma^{-1} \mathbf{a} \right\} \]

\[ = \left( -\frac{1}{2} \sum_{l=1}^{4} Z_{kl} a_k a_l \right) \exp\left\{-\frac{1}{2} \mathbf{a}^T \Sigma^{-1} \mathbf{a} \right\} \]
\[
\frac{\partial^2}{\partial x_3 \partial x_4} \Phi(x) = -Z_{43} \exp\left\{-\frac{1}{2} \omega^T \Gamma \omega\right\} + \left( -\frac{1}{2} \sum_{l=1}^{4} Z_{4l}^2 \omega_l \right) \exp\left\{-\frac{1}{2} \omega^T \Gamma \omega\right\}
\]

\[
= \left( -Z_{43} + \sum_{l=1}^{4} Z_{4l}^2 \omega_l \right) \exp\left\{-\frac{1}{2} \omega^T \Gamma \omega\right\}
\]

\[
= \left[ \sum_{l=1}^{4} \left( \frac{Z_{4l}^2}{\omega_l} \right) \right] \exp\left\{-\frac{1}{2} \omega^T \Gamma \omega\right\}
\]

\[
+ \left( -Z_{43} + \sum_{l=1}^{4} Z_{4l}^2 \omega_l \right) \exp\left\{-\frac{1}{2} \omega^T \Gamma \omega\right\}
\]

\[
= \left[ \sum_{l=1}^{4} \left( \frac{Z_{4l}^2}{\omega_l} \right) \right] \exp\left\{-\frac{1}{2} \omega^T \Gamma \omega\right\}
\]

Now take \( \frac{\partial^4}{\partial x_1 \partial x_2 \partial x_3 \partial x_4} \Phi(x) \) and to save work note terms that will be zero when \( \omega = 0 \). One obtains

\[
\left. \frac{\partial^4}{\partial x_1 \partial x_2 \partial x_3 \partial x_4} \Phi(x) \right|_{\omega=0} = Z_{42} Z_{31} + Z_{32} Z_{41} + Z_{43} Z_{21}.
\]

This is it!
For a scalar rv $X$ the characteristic function is defined to be $\Phi_X(\lambda) = \mathbb{E}\{e^{i\lambda X}\}$, this is the Fourier transform of the pdf of $X$ (with a sign change). For a vector rv $Y$ the characteristic function is defined to be $\Phi_Y(\lambda) = \mathbb{E}\{e^{i\lambda^T Y}\}$, this is the multidimensional Fourier transform of the joint pdf of $Y$ (with a sign change).

(a) If $X \sim \mathcal{N}(0, 1)$ find its characteristic function by a direct computation.
(b) If $X \sim \mathcal{N}(\mu, \sigma^2)$ find its characteristic function using the result of (a).
(c) If $Y \sim \mathcal{N}(0, I)$ (a vector rv) find its multidimensional characteristic function using the result of (a).
(d) If $Y \sim \mathcal{N}(\mu, \Sigma)$ (a vector rv) find its multidimensional characteristic function using the previous results.
\( Y \sim N(0, I) \)

\[
\Phi_Y(\lambda) = E\left\{ e^{j\lambda^T Y} \right\} = E\left\{ \exp\left( j \sum_{k=1}^{n} \lambda_k Y_k \right) \right\} \\
= E\left\{ \prod_{k=1}^{n} e^{j\lambda_k Y_k} \right\} = \prod_{k=1}^{n} E\left\{ e^{j\lambda_k Y_k} \right\} \\
\text{since the components of vector } Y \text{ are statistically independent. Also since they are iid } N(0,1): \\
\Phi_Y(\lambda) = \prod_{k=1}^{n} \Phi_{Y_k}(\lambda_k) = \prod_{k=1}^{n} e^{-\lambda_k^2/2} \\
= e^{-\frac{1}{2} \sum_{k=1}^{n} \lambda_k^2} \quad \lambda_k \in \mathbb{R} \quad k = 1, 2, \ldots, n
\]

(4) \( Y \sim N(\mu, \Sigma) \)

Suppose for the moment that \( Y =HX + b \) where \( H \) is \( n \times m \) of rank \( m \) and \( X \) is an \( m \times 1 \) \( N(0, I) \) random vector. Then can show:

\[
EY = HEX + b = b \\
E\{(Y-b)(Y-b)^T\} = E\{(HX)(HX)^T\} = HE\{XX^T\}H^T \\
= HH^T
\]

Calculating the char. function for \( Y \) in terms of that for \( X \)

\[
\Phi_Y(\lambda) = E\left\{ \exp\left( j \lambda^T Y \right) \right\} \\
= E\left\{ \exp\left( j \lambda^T (HX + b) \right) \right\}
\]
An $n$-dimensional rv $X$ is said to be Gaussian if its multi-dimensional characteristic function is of the form

$$\Phi_X(\lambda) = e^{i\lambda^T \mu - 0.5\lambda^T \Sigma \lambda}$$

where $\mu$ is the mean vector and $\Sigma$ is the covariance matrix (possibly singular).

(a) When the covariance matrix is non-singular the pdf of $X$ has the familiar form. Find that form and write it down.

(b) On the otherhand when the covariance matrix is singular the pdf does not exist. Suppose that rank $\Sigma = r < n$. Show that there exists an $r$-dimensional Gaussian random vector $Y$ and matrices $A$, $b$ such that

$$X = AY + b$$

with probability 1.

(See class notes for solution).