Moment Generating Function

**Definition** Let $X = (X_1, \ldots, X_n)^T$ be a random vector and $t = (t_1, \ldots, t_n)^T \in \mathbb{R}^n$. The moment generating function (MGF) is defined by

$$M_X(t) = E(e^{t^TX})$$

for all $t$ for which the expectation exists (i.e., finite).

**Remarks:**
- $M_X(t) = E(e^{\sum_{i=1}^n t_i X_i})$
- For $0 = (0, \ldots, 0)^T$, we have $M_X(0) = 1$.
- If $X$ is a discrete random variable with finitely many values, then $M_X(t) = E(e^{t^TX})$ is always finite for all $t \in \mathbb{R}^n$.
- We will always assume that the distribution of $X$ is such that $M_X(t)$ is finite for all $t \in (-t_0, t_0)^n$ for some $t_0 > 0$.

**Connection with moments**

- Let $k_1, \ldots, k_n$ be nonnegative integers and $k = k_1 + \cdots + k_n$. Then
  $$\frac{\partial^k}{\partial t_1^{k_1} \cdots \partial t_n^{k_n}} M_X(t) = \frac{\partial^k}{\partial t_1^{k_1} \cdots \partial t_n^{k_n}} E(e^{t_1 X_1 + \cdots + t_n X_n})$$
  $$= E \left( \frac{\partial^k}{\partial t_1^{k_1} \cdots \partial t_n^{k_n}} e^{t_1 X_1 + \cdots + t_n X_n} \right)$$
  $$= E(X_1^{k_1} \cdots X_n^{k_n} (e^{t_1 X_1 + \cdots + t_n X_n}))$$

Setting $t = 0 = (0, \ldots, 0)^T$, we get

$$\frac{\partial^k}{\partial t_1^{k_1} \cdots \partial t_n^{k_n}} M_X(t) \bigg|_{t=0} = E(X_1^{k_1} \cdots X_n^{k_n})$$

- For a (scalar) random variable $X$ we obtain the $k$th moment of $X$:
  $$\frac{d^k}{dt^k} M_X(t) \bigg|_{t=0} = E(X^k)$$
Theorem 2

Assume \( X_1, \ldots, X_m \) are independent random vectors in \( \mathbb{R}^n \) and let \( X = X_1 + \cdots + X_m \). Then

\[
M_X(t) = \prod_{i=1}^{m} M_{X_i}(t)
\]

Proof:

\[
M_X(t) = E(e^{t^T X}) = E(e^{t^T (X_1 + \cdots + X_m)})
= E(e^{t^T X_1} \cdots e^{t^T X_m})
= E(e^{t^T X_1}) \cdots E(e^{t^T X_m})
= M_{X_1}(t) \cdots M_{X_m}(t)
\]

Note: This theorem gives us a powerful tool for determining the distribution of the sum of independent random variables.

Example: MGF for \( X \sim \text{Gamma}(r, \lambda) \) and \( X_1 + \cdots + X_m \) where the \( X_i \) are independent and \( X_i \sim \text{Gamma}(r_i, \lambda) \).

Example: MGF for \( X \sim \text{Poisson}(\lambda) \) and \( X_1 + \cdots + X_m \) where the \( X_i \) are independent and \( X_i \sim \text{Gamma}(\lambda_i) \). Also, use the MGF to find \( E(X) \), \( E(X^2) \), and \( \text{Var}(X) \).

Theorem 3

Assume \( X \) is a random vector in \( \mathbb{R}^n \), \( A \) is an \( m \times n \) real matrix and \( b \in \mathbb{R}^m \). Then the MGF of \( Y = AX + b \) is given at \( t \in \mathbb{R}^n \) by

\[
M_Y(t) = e^{t^T b} M_X(A^T t)
\]

Proof:

\[
M_Y(t) = E(e^{t^T Y}) = E(e^{t^T (AX + b)})
= e^{t^T b} E(e^{t^T AX})
= e^{t^T b} E((e^{t^T} t)^T X)
= e^{t^T b} M_X(A^T t)
\]

Note: In the scalar case \( Y = aX + b \) we obtain

\[
M_Y(t) = e^{bt} M_X(at)
\]

Applications to Normal Distribution

Let \( X \sim N(0,1) \). Then

\[
M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, dx
= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x^2-2tx)} \, dx
= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-t)^2} \, dx
= e^{t^2/2}
\]

We obtain that for all \( t \in \mathbb{R} \)

\[
M_X(t) = e^{t^2/2}
\]
Multivariate Normal Distributions

Linear Algebra Review

- Recall that an \( n \times n \) real matrix \( C \) is called **nonnegative definite** if it is symmetric and

\[
x^T C x \geq 0 \quad \text{for all } x \in \mathbb{R}^n
\]

and **positive definite** if it is symmetric and

\[
x^T C x > 0 \quad \text{for all } x \in \mathbb{R}^n \text{ such that } x \neq 0
\]

- Let \( A \) be an arbitrary \( n \times n \) real matrix. Then \( C = A^T A \) is nonnegative definite. If \( A \) is nonsingular (invertible), then \( C \) is positive definite.

**Proof:** \( A^T A \) is symmetric since \( (A^T A)^T = (A^T)^T A^T = A^T A \).

Thus it is nonnegative definite since

\[
x^T (A^T A)x = x^T A^T Ax = (Ax)^T (Ax) = ||Ax||^2 \geq 0
\]

For any nonnegative definite \( n \times n \) matrix \( C \) the following hold:

1. \( C \) has \( n \) **nonnegative eigenvalues** \( \lambda_1, \ldots, \lambda_n \) (counting multiplicities) and corresponding \( n \) orthogonal unit-length eigenvectors \( b_1, \ldots, b_n \):

\[
C b_i = \lambda_i b_i, \quad i = 1, \ldots, n
\]

where \( b_i^T b_i = 1, i = 1, \ldots, n \) and \( b_i^T b_j = 0 \) if \( i \neq j \).

2. (Spectral decomposition) \( C \) can be written as

\[
C = BDB^T
\]

where \( D = \text{diag}(\lambda_1, \ldots, \lambda_n) \) is the diagonal matrix of the eigenvalues of \( C \), and \( B \) is the orthogonal matrix whose \( i \)th column is \( b_i \), i.e., \( B = [b_1 \ldots b_n] \).

3. \( C \) is **positive definite** \( \iff \) \( C \) is nonsingular \( \iff \) all the eigenvalues \( \lambda_i \) are positive
(4) $C$ has a unique nonnegative definite square root $C^{1/2}$, i.e., there exists a unique nonnegative definite $A$ such that

$$C = AA$$

**Proof:** We only prove the existence of $A$. Let $D^{1/2} = \text{diag}(\lambda_1^{1/2}, \ldots, \lambda_n^{1/2})$ and note that $D^{1/2}D^{1/2} = D$. Let $A = BD^{1/2}B^T$. Then $A$ is nonnegative definite and

$$A^2 = AA = (BD^{1/2}B^T)(BD^{1/2}B^T) = BD^{1/2}B^TB^{1/2}B^T = BD^{1/2}D^{1/2}B^T = C$$

**Remarks:**
- If $C$ is positive definite, then so is $A$.
- If we don’t require that $A$ be nonnegative definite, then in general there are infinitely many solutions $A$ for $AA^T = C$.

### Defining the Multivariate Normal Distribution

Let $Z_1, \ldots, Z_n$ be independent r.v.’s with $Z_i \sim N(0, 1)$. The multivariate MGF of $Z = (Z_1, \ldots, Z_n)^T$ is

$$M_Z(t) = E(e^{tZ}) = E(e^{\sum_{i=1}^n t_i Z_i}) = \prod_{i=1}^n E(e^{t_i Z_i}) = \prod_{i=1}^n e^{t_i^2/2} = e^{\sum_{i=1}^n t_i^2/2} = e^{\frac{1}{2}t^T t}$$

Now let $\mu \in \mathbb{R}^n$ and $A$ an $n \times n$ real matrix. Then the MGF of $X = AZ + \mu$ is

$$M_X(t) = e^{t^T \mu}M_Z(A^T t) = e^{t^T \mu}e^{\frac{1}{2}(A^T t)^T(A^T t)} = e^{t^T \mu}e^{\frac{1}{2}t^T AA^T t} = e^{t^T \mu + \frac{1}{2}t^T \Sigma t}$$

where $\Sigma = AA^T$. Note that $\Sigma$ is nonnegative definite.

**Lemma 4**

If $\Sigma$ is the covariance matrix of some random vector $X = (X_1, \ldots, X_n)^T$, then it is nonnegative definite.

**Proof:** We know that $\Sigma = \text{Cov}(X)$ is symmetric. Let $b \in \mathbb{R}^n$ be arbitrary. Then

$$b^T \Sigma b = b^T \text{Cov}(X)b = \text{Cov}(b^T X) = \text{Var}(b^T X) \geq 0$$

so $\Sigma$ is nonnegative definite.

**Remark:** It can be shown that an $n \times n$ matrix $\Sigma$ is nonnegative definite if and only if there exists a random vector $X = (X_1, \ldots, X_n)^T$ such that $\text{Cov}(X) = \Sigma$.

**Definition** Let $\mu \in \mathbb{R}^n$ and let $\Sigma$ be an $n \times n$ nonnegative definite matrix. A random vector $X = (X_1, \ldots, X_n)$ is said to have a multivariate normal distribution with parameters $\mu$ and $\Sigma$ if its multivariate MGF is

$$M_X(t) = e^{t^T \mu + \frac{1}{2}t^T \Sigma t}$$

**Notation:** $X \sim N(\mu, \Sigma)$.

**Remarks:**
- If $Z = (Z_1, \ldots, Z_n)^T$ with $Z_i \sim N(0, 1)$, $i = 1, \ldots, n$, then $Z \sim N(0, I)$, where $I$ is the $n \times n$ identity matrix.
- We saw that if $Z \sim N(0, I)$, then $X = AZ + \mu \sim N(\mu, \Sigma)$, where $\Sigma = AA^T$. One can show the following:

$$X \sim N(\mu, \Sigma) \text{ if and only if } X = AZ + \mu \text{ for a random } n\text{-vector } Z \sim N(0, I) \text{ and some } n \times n \text{ matrix } A \text{ with } \Sigma = AA^T.$$
Mean and covariance for multivariate normal distribution

Consider first \( Z \sim N(0, I) \), i.e., \( Z = (Z_1, \ldots, Z_n)^T \), where the \( Z_i \) are independent \( N(0, 1) \) random variables. Then
\[
E(Z) = (E(Z_1), \ldots, E(Z_n))^T = (0, \ldots, 0)^T
\]
and
\[
E((Z_i - E(Z_i))(Z_j - E(Z_j))) = E(Z_iZ_j) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j \end{cases}
\]
Thus
\[
E(Z) = 0, \quad \text{Cov}(Z) = I
\]

Joint pdf for multivariate normal distribution

**Lemma 5**

If a random vector \( X = (X_1, \ldots, X_n)^T \) has covariance matrix \( \Sigma \) that is not of full rank (i.e., singular), then \( X \) does not have a joint pdf.

**Proof sketch:** If \( \Sigma \) is singular, then there exists \( b \in \mathbb{R}^n \) such that \( b \neq 0 \) and \( \Sigma b = 0 \). Consider the random variable \( b^T X = \sum_{i=1}^n b_i X_i \):
\[
\text{Var}(b^T X) = \text{Cov}(b^T X) = b^T \Sigma b = 0
\]
Therefore \( P(b^T X = c) = 1 \) for some constant \( c \). If \( X \) had a joint pdf \( f(x) \), then for \( B = \{ x : b^T x = c \} \) we should have
\[
1 = P(b^T X = c) = P(X \in B) = \int_B \cdots \int f(x_1, \ldots, x_n) \, dx_1 \cdots dx_n
\]
But this is impossible since \( B \) is an \( (n-1) \)-dimensional hyperplane whose \( n \)-dimensional volume is zero, so the integral must be zero. \( \square \)

If \( X \sim N(\mu, \Sigma) \), then \( X = AZ + \mu \) for a random \( n \)-vector \( Z \sim N(0, I) \) and some \( n \times n \) matrix \( A \) with \( \Sigma = AA^T \).

We have
\[
E(AZ + \mu) = AE(Z) + \mu = \mu
\]
Also,
\[
\text{Cov}(AZ + \mu) = \text{Cov}(AZ) = A \text{Cov}(Z) A^T = AA^T = \Sigma
\]
Thus
\[
E(X) = \mu, \quad \text{Cov}(X) = \Sigma
\]

**Theorem 6**

If \( X = (X_1, \ldots, X_n)^T \sim N(\mu, \Sigma) \), where \( \Sigma \) is nonsingular, then it has a joint pdf given by
\[
f_X(x) = \frac{1}{(2\pi)^n \det \Sigma} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}, \quad x \in \mathbb{R}^n
\]

**Proof:** We know that \( X = AZ + \mu \) where \( Z = (Z_1, \ldots, Z_n)^T \sim N(0, I) \) and \( A \) is an \( n \times n \) matrix such that \( AA^T = \Sigma \). Since \( \Sigma \) is nonsingular, \( A \) must be nonsingular with inverse \( A^{-1} \). Thus the mapping
\[
h(z) = Az + \mu
\]
is invertible with inverse \( g(x) = A^{-1}(x - \mu) \) whose Jacobian is
\[
J_g(x) = \det A^{-1}
\]
By the multivariate transformation theorem
\[
f_X(x) = f_Z(g(x)) |J_g(x)| = f_Z(A^{-1}(x - \mu)) |\det A^{-1}|
\]
Proof cont’d: Since $Z = (Z_1, \ldots, Z_n)^T$, where the $Z_i$ are independent $N(0, 1)$ random variables, we have

$$f_Z(z) = \prod_{i=1}^n \left( \frac{1}{\sqrt{2\pi}} \right) e^{-z_i^2/2} = \frac{1}{(2\pi)^n} e^{-\frac{1}{2} \sum_{i=1}^n z_i^2} = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2} z^T z}$$

so we get

$$f_X(x) = f_Z(A^{-1}(x - \mu)) |\det A^{-1}| = \frac{1}{(2\pi)^n} e^{-\frac{1}{2} (A^{-1}(x - \mu))^T A^{-1}(x - \mu)} |\det A^{-1}|$$

$$= \frac{1}{(2\pi)^n} e^{-\frac{1}{2} (x - \mu)^T A^{-1}(x - \mu)} |\det A^{-1}|$$

$$= \frac{1}{(2\pi)^n \det \Sigma} e^{-\frac{1}{2} (x - \mu)^T \Sigma^{-1}(x - \mu)}$$

since $|\det A^{-1}| = \frac{1}{\sqrt{\det \Sigma}}$ and $(A^{-1})^T A^{-1} = \Sigma^{-1}$ (exercise!) □

\[ \begin{align*}
\Sigma^{-1} &= \begin{bmatrix}
\sigma_1^2 & \rho \sigma_1 \sigma_2 \\
\rho \sigma_1 \sigma_2 & \sigma_2^2
\end{bmatrix}^{-1} = \frac{1}{\det \Sigma} \begin{bmatrix}
\sigma_1^2 & -\rho \sigma_1 \sigma_2 \\
-\rho \sigma_1 \sigma_2 & \sigma_2^2
\end{bmatrix} \\
&= \frac{1}{(1 - \rho^2)\sigma_1^2 \sigma_2^2} \begin{bmatrix}
x_1 & x_2 - \mu_1 \\
(1 - \rho^2)\sigma_1^2 \sigma_2^2 & \rho \sigma_1 \sigma_2
\end{bmatrix} \begin{bmatrix}
x_1 - \mu_1 \\
(1 - \rho^2)\sigma_1^2 \sigma_2^2 & \rho \sigma_1 \sigma_2
\end{bmatrix} \\
&= \frac{1}{(1 - \rho^2)\sigma_1^2 \sigma_2^2} \begin{bmatrix}
x_1 & x_2 - \mu_1 \\
\rho \sigma_1 \sigma_2 & \rho \sigma_1 \sigma_2
\end{bmatrix} \begin{bmatrix}
x_1 - \mu_1 \\
\rho \sigma_1 \sigma_2 \\
\rho \sigma_1 \sigma_2 & \rho \sigma_1 \sigma_2
\end{bmatrix} \\
&= \frac{1}{(1 - \rho^2)\sigma_1^2 \sigma_2^2} \left( \frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} - \frac{2\rho(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1 \sigma_2} \right)
\end{align*} \]

Thus the bivariate normal distribution is determined by five scalar parameters $\mu_1$, $\mu_2$, $\sigma_1^2$, $\sigma_2^2$, and $\rho$.

$\Sigma$ is positive definite $\iff \Sigma$ is invertible $\iff \det \Sigma > 0$:

$$\det \Sigma = (1 - \rho^2)\sigma_1^2 \sigma_2^2 > 0 \iff |\rho| < 1 \text{ and } \sigma_1^2 \sigma_2^2 > 0$$

so a bivariate normal random variable $(X_1, X_2)$ has a pdf if and only if the components $X_1$ and $X_2$ have positive variances and $|\rho| < 1$.

\[ \begin{align*}
\text{Special case: bivariate normal} \\
\text{For } n = 2 \text{ we have } \\
\mu &= \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \text{ and } \Sigma = \begin{bmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\
\rho \sigma_1 \sigma_2 & \sigma_2^2 \end{bmatrix} \\
\text{where } \mu_i = E(X_i), \sigma_i^2 = \text{Var}(X_i), i = 1, 2, \text{ and } \rho = \rho(X_1, X_2) = \frac{\text{Cov}(X_1, X_2)}{\sigma_1 \sigma_2} \\
\text{Thus } \rho = 0 \text{ if and only if } \sigma_1^2 \sigma_2^2 > 0 \iff |\rho| < 1 \text{ and } \sigma_1^2 \sigma_2^2 > 0
\end{align*} \]

We have

\[ \begin{align*}
\Sigma^{-1} &= \begin{bmatrix}
\sigma_1^2 & \rho \sigma_1 \sigma_2 \\
\rho \sigma_1 \sigma_2 & \sigma_2^2
\end{bmatrix}^{-1} = \frac{1}{\det \Sigma} \begin{bmatrix}
\sigma_1^2 & -\rho \sigma_1 \sigma_2 \\
-\rho \sigma_1 \sigma_2 & \sigma_2^2
\end{bmatrix} \\
&= \frac{1}{\sigma_1^2 \sigma_2^2} \begin{bmatrix}
x_1 & x_2 - \mu_1 \\
x_1 - \mu_1 & x_2 - \mu_1
\end{bmatrix} \begin{bmatrix}
x_1 - \mu_1 \\
x_2 - \mu_1
\end{bmatrix} \\
&= \frac{1}{\sigma_1^2 \sigma_2^2} \begin{bmatrix}
x_1 & x_2 - \mu_1 \\
x_1 - \mu_1 & x_2 - \mu_1
\end{bmatrix} \begin{bmatrix}
x_1 - \mu_1 \\
x_2 - \mu_1
\end{bmatrix} \\
&= \frac{1}{(1 - \rho^2)\sigma_1^2 \sigma_2^2} \left( \frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} - \frac{2\rho(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1 \sigma_2} \right)
\end{align*} \]

Thus the joint pdf of $(X_1, X_2)^T \sim N(\mu, \Sigma)$ is

$$f(x_1, x_2) = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1 - \rho^2}} e^{-\frac{1}{2} \left( \frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} - \frac{2\rho(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1 \sigma_2} \right)}$$

Remark: If $\rho = 0$, then

$$f(x_1, x_2) = \frac{1}{2\pi \sigma_1 \sigma_2} e^{-\frac{1}{2} \left( \frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right)}$$

Therefore $X_1$ and $X_2$ are independent. It is also easy to see that $f(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2)$ for all $x_1$ and $x_2$ implies $\rho = 0$. Thus we obtain

Two jointly normal random variables $X_1$ and $X_2$ are independent if and only if they are uncorrelated.
In general, the following important facts can be proved using the multivariate MGF:

(i) If \( \mathbf{X} = (X_1, \ldots, X_n)^T \sim N(\mu, \Sigma) \), then \( X_1, X_2, \ldots, X_n \) are independent if and only if they are uncorrelated, i.e., 
\[
\text{Cov}(X_i, X_j) = 0 \quad \text{if} \quad i \neq j, \quad \text{i.e.,} \quad \Sigma \text{ is a diagonal matrix.}
\]

(ii) Assume \( \mathbf{X} = (X_1, \ldots, X_n)^T \sim N(\mu, \Sigma) \) and let
\[
\mathbf{X}_1 = (X_1, \ldots, X_k)^T, \quad \mathbf{X}_2 = (X_{k+1}, \ldots, X_n)^T
\]
Then \( \mathbf{X}_1 \) and \( \mathbf{X}_2 \) are independent if and only if 
\[
\text{Cov}(\mathbf{X}_1, \mathbf{X}_2) = 0_{k \times (n-k)}
\]
the \( k \times (n-k) \) matrix of zeros, i.e., \( \Sigma \) can be partitioned as
\[
\Sigma = \begin{bmatrix}
\Sigma_{11} & 0_{k \times (n-k)} \\
0_{(n-k) \times k} & \Sigma_{22}
\end{bmatrix}
\]
where \( \Sigma_{11} = \text{Cov}(\mathbf{X}_1) \) and \( \Sigma_{22} = \text{Cov}(\mathbf{X}_2) \).

For some \( 1 \leq m < n \) let \( \{i_1, \ldots, i_m\} \subseteq \{1, \ldots, n\} \) such that 
\( i_1 < i_2 < \cdots < i_m \). Let \( \mathbf{e}_j = (0, \ldots, 0, 1, 0, \ldots, 0)^T \) be the \( j \)th unit vector in \( \mathbb{R}^n \) and define the \( m \times n \) matrix \( \mathbf{A} \) by
\[
\mathbf{A} = \begin{bmatrix}
\mathbf{e}_1^T \\
\vdots \\
\mathbf{e}_m^T
\end{bmatrix}
\]
Then
\[
\mathbf{AX} = \begin{bmatrix}
\mathbf{e}_1^T & \mathbf{X}_1 \\
\vdots & \vdots \\
\mathbf{e}_m^T & \mathbf{X}_n
\end{bmatrix}
\]
Thus \( (X_{i_1}, \ldots, X_{i_m})^T \sim N(\mathbf{A}\mu, \mathbf{A}\Sigma\mathbf{A}^T) \).

### Marginals of multivariate normal distributions

Let \( \mathbf{X} = (X_1, \ldots, X_n)^T \sim N(\mu, \Sigma) \). If \( \mathbf{A} \) is an \( m \times n \) matrix and \( \mathbf{b} \in \mathbb{R}^m \), then
\[
\mathbf{Y} = \mathbf{AX} + \mathbf{b}
\]
is a random \( m \)-vector. Its MGF at \( t \in \mathbb{R}^m \) is
\[
M_Y(t) = e^{\mathbf{t}^T \mathbf{b}} M_X(\mathbf{A}^T t)
\]
Since \( M_X(\tau) = e^{\tau^T \mu + \frac{1}{2} \tau^T \Sigma \tau} \) for all \( \tau \in \mathbb{R}^n \), we obtain
\[
M_Y(t) = e^{\mathbf{t}^T \mathbf{b} (\mathbf{A}^T t)^T \mu + \frac{1}{2} (\mathbf{A}^T t)^T \Sigma (\mathbf{A}^T t)}
\]
\[
= e^{\mathbf{t}^T (\mathbf{b} + \mathbf{A}\mu) + \frac{1}{2} \mathbf{t}^T \mathbf{A} \Sigma \mathbf{A}^T \mathbf{t}}
\]
This means that \( \mathbf{Y} \sim N(\mathbf{b} + \mathbf{A}\mu, \mathbf{A}\Sigma\mathbf{A}^T) \), i.e., \( \mathbf{Y} \) is multivariate normal with mean \( \mathbf{b} + \mathbf{A}\mu \) and covariance \( \mathbf{A}\Sigma\mathbf{A}^T \).

**Example:** Let \( a_1, \ldots, a_n \in \mathbb{R} \) and determine the distribution of 
\[
\mathbf{Y} = a_1 X_1 + \cdots + a_n X_n.
\]

Note the following:

\[
\mathbf{A}\mu = \begin{bmatrix}
\mu_{i_1} \\
\vdots \\
\mu_{i_m}
\end{bmatrix}
\]
and the \((j,k)\)th entry of \( \mathbf{A}\Sigma\mathbf{A}^T \) is
\[
(\mathbf{A}\Sigma\mathbf{A}^T)_{jk} = (\mathbf{A} \times (i_{k}\text{th column of } \Sigma))_j
\]
\[
= (\Sigma)_{i_ji_k} = \text{Cov}(X_{i_j}, X_{i_k})
\]
Thus if \( \mathbf{X} = (X_1, \ldots, X_n)^T \sim N(\mu, \Sigma) \), then \( (X_{i_1}, \ldots, X_{i_m})^T \) is multivariate normal whose mean and covariance are obtained by picking out the corresponding elements of \( \mu \) and \( \Sigma \).

**Special case:** For \( m = 1 \) we obtain that \( X_i \sim N(\mu_i, \sigma_i^2) \), where
\[
\mu_i = E(X_i) \quad \text{and} \quad \sigma_i^2 = \text{Var}(X_i), \quad \text{for all } i = 1, \ldots, n.
\]
Conditional distributions

Let \( X = (X_1, \ldots, X_n)^T \sim N(\mu, \Sigma) \) and for \( 1 \leq m < n \) define
\[
X_1 = (X_1, \ldots, X_m)^T, \quad X_2 = (X_{m+1}, \ldots, X_n)^T
\]
We know that \( X_1 \sim N(\mu_1, \Sigma_{11}) \) and \( X_2 \sim N(\mu_2, \Sigma_{22}) \) where
\[
\mu_i = E(X_i), \quad \Sigma_{ii} = \text{Cov}(X_i), \quad i = 1, 2.
\]
Then \( \mu \) and \( \Sigma \) can be partitioned as
\[
\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}
\]
where \( \Sigma_{ij} = \text{Cov}(X_i, X_j), \ i, j = 1, 2 \). Note that \( \Sigma_{11} \) is \( m \times m \), \( \Sigma_{22} \) is \( (n-m) \times (n-m) \), \( \Sigma_{12} \) is \( m \times (n-m) \), and \( \Sigma_{21} \) is \( (n-m) \times m \). Also, \( \Sigma_{21} = \Sigma_{12}^T \).

We assume that \( \Sigma_{11} \) is nonsingular and we want to determine the conditional distribution of \( X_2 \) given \( X_1 = x_1 \).

We want to solve for \( B, C \) and \( D \). First consider \( BB^T = \Sigma_{11} \). We choose \( B \) to be the unique positive definite square root of \( \Sigma_{11} \):
\[
B = \Sigma_{11}^{1/2}
\]
Recall that \( B \) is symmetric and it is invertible since \( \Sigma_{11} \) is. Then \( \Sigma_{21} = CB^T \) implies
\[
C = \Sigma_{21}(B^T)^{-1} = \Sigma_{21}B^{-1}
\]
Then \( \Sigma_{22} = CC^T + DD^T \) gives
\[
DD^T = \Sigma_{22} - CC^T = \Sigma_{22} - \Sigma_{21}B^{-1}B^T(\Sigma_{21})^{-1}
\]
\[
= \Sigma_{22} - \Sigma_{21}BB^{-1}\Sigma_{12} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}
\]
Now note that \( X = AZ + \mu \) gives
\[
X_1 = BZ_1 + \mu_1, \quad X_2 = CZ_1 + DZ_2 + \mu_2
\]
Recall that \( X = AZ + \mu \) for some \( Z = (Z_1, \ldots, Z_n)^T \) where the \( Z_i \) are independent \( N(0, 1) \) random variables and \( A \) is such that \( AA^T = \Sigma \).

Let \( Z_1 = (Z_1, \ldots, Z_m)^T \) and \( Z_2 = (Z_{m+1}, \ldots, Z_n)^T \). We want to determine such \( A \) in a partitioned form with dimensions corresponding to the partitioning of \( \Sigma \):
\[
A = \begin{bmatrix} B & 0_{m \times (n-m)} \\ C & D \end{bmatrix}
\]
We can write \( \Sigma = AA^T \) as
\[
\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} B & 0_{m \times (n-m)} \\ C & D \end{bmatrix}
\begin{bmatrix} B^T & C^T \\ 0_{m \times m} & D^T \end{bmatrix}
\]
\[
= \begin{bmatrix} BB^T & BC^T \\ CB^T & CC^T + DD^T \end{bmatrix}
\]
Since \( B \) is invertible, given \( X_1 = x_1 \), we have \( Z_1 = B^{-1}(x_1 - \mu_1) \). So given \( X_1 = x_1 \), we have that the conditional distribution of \( X_2 \) and the conditional distribution of
\[
CB^{-1}(x_1 - \mu_1) + DZ_2 + \mu_2
\]
are the same.

But \( Z_2 \) is independent of \( X_1 \), so given \( X_1 = x_1 \), the conditional distribution of \( CB^{-1}(x_1 - \mu_1) + DZ_2 + \mu_2 \) is the same as its unconditional distribution.

We conclude that the conditional distribution of \( X_2 \) given \( X_1 = x_1 \) is multivariate normal with mean
\[
E(X_2|X_1 = x_1) = \mu_2 + CB^{-1}(x_1 - \mu_1)
\]
and covariance matrix
\[
\Sigma_{22|1} = DD^T = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}
\]
**Special case: bivariate normal**

Suppose \( X = (X_1, X_2)^T \sim N(\mu, \Sigma) \) with
\[
\mu = \begin{bmatrix} \mu_1 \\ \mu_1 \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{bmatrix}
\]

We have
\[
\mu_2 + \Sigma_{21} \Sigma_{11}^{-1} (x_1 - \mu_1) = \mu_2 + \rho \frac{\sigma_2}{\sigma_1} (x_1 - \mu_1)
\]

and
\[
\Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} = \sigma_2^2 - \frac{\rho^2 \sigma_1^2 \sigma_2^2}{\sigma_1^2} = \sigma_2^2 (1 - \rho^2)
\]

Thus the conditional distribution of \( X_2 \) given \( X_1 = x_1 \) is normal with (conditional) mean
\[
E(X_2|X_1 = x_1) = \mu_2 + \rho \frac{\sigma_2}{\sigma_1} (x_1 - \mu_1)
\]

and variance
\[
\text{Var}(X_2|X_1 = x_1) = \sigma_2^2 (1 - \rho^2)
\]

Equivalently, the conditional distribution of \( X_2 \) given \( X_1 = x_1 \) is
\[
N \left( \mu_2 + \rho \frac{\sigma_2}{\sigma_1} (x_1 - \mu_1), \sigma_2^2 (1 - \rho^2) \right)
\]

If \( |\rho| < 1 \), then the conditional pdf exists and is given by
\[
f_{X_2|X_1}(x_2|x_1) = \frac{1}{\sigma_2 \sqrt{2\pi(1 - \rho^2)}} e^{-\frac{(x_2 - \mu_2 - \rho \frac{\sigma_2}{\sigma_1} (x_1 - \mu_1))^2}{2\sigma_2^2 (1 - \rho^2)}}
\]

**Remark:** Note that \( E(X_2|X_1 = x_1) = \mu_2 + \rho \frac{\sigma_2}{\sigma_1} (x_1 - \mu_1) \) is a linear (affine) function of \( x_1 \).

**Example:** Recall the MMSE estimate problem for \( X \sim N(0, \sigma_X^2) \) from the observation \( Y = X + Z \), where \( Z \sim N(0, \sigma_Z^2) \) and \( X \) and \( Z \) are independent. Use the above the find \( g^*(y) = E[X|Y = y] \) and compute the minimum mean square error \( E[(X - g^*(Y))^2] \).