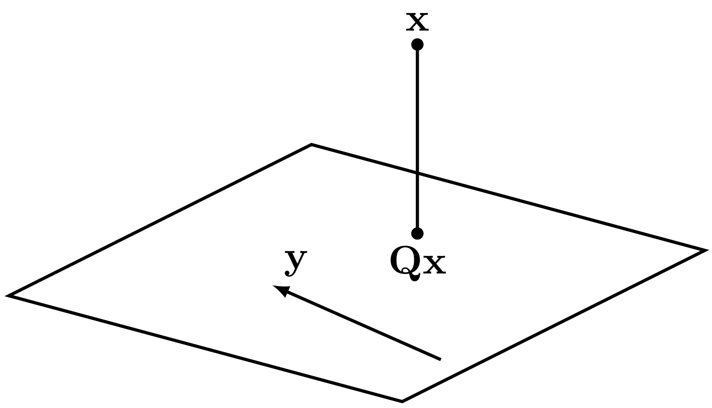


Figure 2.1: Orthogonal Projection

**Definition 2.9.** The matrix $\mathbf{Q} \in \mathbb{R}^{n \times n}$ is called projection matrix, or idempotent, if $\mathbf{Q}^2 = \mathbf{Q}$. It is called orthogonal projection if additionally $\mathbf{Q}^T = \mathbf{Q}$.

The linear transformation $\mathbf{Q}$ maps onto $\mathrm{Im}(\mathbf{Q})$, a $k-$dimensional subspace of $\mathbb{R}^n$. Let $\mathbf{x} \in \mathbb{R}^n$, and $\mathbf{y} = \mathbf{Qx} \in \mathrm{Im}(\mathbf{Q})$. Since $\mathbf{Q}$ is the projection matrix, $\mathbf{Qy} = \mathbf{y}$. For an orthogonal projection, $\mathbf{x} - \mathbf{Qx}$ is orthogonal to all vectors $\mathbf{y}$ in $\mathrm{Im}(\mathbf{Q})$ for every $\mathbf{x} \in \mathbb{R}^n$. To see this, note that there is a vector $\mathbf{z} \in \mathbb{R}^n$ such that $\mathbf{y} = \mathbf{Qz}$. Then we have:

$$\mathbf{y}^T(\mathbf{x} - \mathbf{Qx}) = \mathbf{z}^T\mathbf{Q}^T(\mathbf{x} - \mathbf{Qx}).$$

Since for an orthogonal projection $\mathbf{Q}^T = \mathbf{Q}$ then:

$$\mathbf{z}^T\mathbf{Q}^T(\mathbf{x} - \mathbf{Qx}) = \mathbf{z}^T\mathbf{Q}(\mathbf{x} - \mathbf{Qx}) = \mathbf{z}^T(\mathbf{Qx} - \mathbf{Q}^2\mathbf{x}) = \mathbf{z}^T(\mathbf{Qx} - \mathbf{Qx}) = 0.$$

Therefore $\mathbf{y}^T(\mathbf{x} - \mathbf{Qx}) = 0$ and $\mathbf{x} - \mathbf{Qx}$ is orthogonal to $\mathbf{y}$.

**Lemma 2.10.** *Let $\mathbf{M} = \mathbf{V\Lambda V}^T$ be spectral decomposition of $\mathbf{M} \in \mathbb{R}^{n \times n}$ and symmetric. For $k \leq n$, the matrix $\mathbf{Q} = \sum_{i=1}^{k} \mathbf{v}_i\mathbf{v}_i^T$ is an orthogonal projection onto $\mathrm{Im}(\mathbf{Q}) =< \mathbf{v}_1, \ldots, \mathbf{v}_k >$.*

*Proof.* For $\mathbf{x} \in \mathbb{R}^n$, we have:

$$\mathbf{Qx} = \sum_{i=1}^{k} \mathbf{v}_i\mathbf{v}_i^T x = \sum_{i=1}^{k} (\mathbf{v}_i^T\mathbf{x})\mathbf{v}_i = \sum_{i=1}^{k} \gamma_i\mathbf{v}_i \in \mathrm{Im}(\mathbf{Q}).$$

Moreover:

$$\mathbf{Q}^2 = (\sum_{i=1}^{k} \mathbf{v}_i\mathbf{v}_i^T)(\sum_{i=1}^{k} \mathbf{v}_i\mathbf{v}_i^T) = \sum_{i=1}^{k} \mathbf{v}_i\mathbf{v}_i^T = \mathbf{Q}.$$

Finally $\mathbf{Q}$ is symmetric and therefore it is an orthogonal projection. $\square$

- *Let* $\mathbf{Q}$ *be an orthogonal projection on* $\mathrm{Im}(\mathbf{Q})$*, Then* $\mathbf{I}-\mathbf{Q}$*is an orthonormal projection onto* $\ker(\mathbf{Q})$*.*
  $\ker(\mathbf{Q})$ denotes the kernel of $\mathbf{Q}$, and $\mathrm{Im}(\mathbf{Q})$ denotes the image of $\mathbf{Q}$.
  $$(\mathbf{I}-\mathbf{Q})^2 = (\mathbf{I}-\mathbf{Q})(\mathbf{I}-\mathbf{Q}) = \mathbf{I}-2\mathbf{Q}+\mathbf{Q}^2 = \mathbf{I}-\mathbf{Q}.$$
  Therefore $\mathbf{I}-\mathbf{Q}$ is a projection matrix. Since $\mathbf{Q}$ is symmetric, so is $\mathbf{I}-\mathbf{Q}$ and hence an orthogonal projection. Let $\mathbf{y}\in\ker(\mathbf{Q})$, i.e., $\mathbf{Q}\mathbf{y}=\mathbf{0}$. Then:
  $$(\mathbf{I}-\mathbf{Q})\mathbf{y} = \mathbf{y}-\mathbf{Q}\mathbf{y} = \mathbf{y}\in\mathrm{Im}(\mathbf{I}-\mathbf{Q}).$$
  Therefore $\ker(\mathbf{Q})\subseteq\mathrm{Im}(\mathbf{I}-\mathbf{Q})$. On the other hand, suppose that $y\in\mathrm{Im}(\mathbf{I}-\mathbf{Q})$. There is $\mathbf{x}\in\mathbb{R}^n$ such that $\mathbf{y}=(\mathbf{I}-\mathbf{Q})\mathbf{x}$. We have:
  $$\mathbf{Q}\mathbf{y} = \mathbf{Q}(\mathbf{I}-\mathbf{Q})\mathbf{x} = \mathbf{Q}\mathbf{x}-\mathbf{Q}^2\mathbf{x} = \mathbf{Q}\mathbf{x}-\mathbf{Q}\mathbf{x} = \mathbf{0}.$$
  So $\mathbf{y}\in\ker(\mathbf{Q})$ and therefore $\mathrm{Im}(\mathbf{I}-\mathbf{Q})\subseteq\ker(\mathbf{Q})$. So $\mathrm{Im}(\mathbf{I}-\mathbf{Q})=\ker(\mathbf{Q})$.
- *Define* $\mathbf{E}_n$ *as follows:*
  $$\mathbf{E}_n = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_{n\times n} = \begin{bmatrix} 1-\frac{1}{n} & -\frac{1}{n} & \dots & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & \dots & -\frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \dots & 1-\frac{1}{n} \end{bmatrix}$$
  *Then* $\mathbf{E}_n$ *is an orthogonal projection onto* $\mathbf{1}_n^{\perp} = \{\mathbf{x}\in\mathbb{R}^n : \mathbf{1}_n^T\mathbf{x}=0\}$ *where* $\mathbf{1}_n$ *is all one vector in* $\mathbb{R}^n$*.*
  See that for all $\mathbf{x}\in\mathbb{R}^n$:
  $$\mathbf{1}_n^T\mathbf{E}_n\mathbf{x} = \mathbf{1}_n^T(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_{n\times n})\mathbf{x} = (\mathbf{1}_n^T-\mathbf{1}_n^T)\mathbf{x} = \mathbf{0}.$$
  Therefore each vector in $\mathrm{Im}(\mathbf{E}_n)$ is orthogonal to $\mathbf{1}_n$.
  Note that $\frac{1}{n}\mathbf{1}_{n\times n}\times\frac{1}{n}\mathbf{1}_{n\times n} = \frac{1}{n}\mathbf{1}_{n\times n}$ and $\frac{1}{n}\mathbf{1}_{n\times n}$ is symmetric. Therefore it is an orthogonal projection. Moreover its image is a one dimensional subspace spanned by $\mathbf{1}_n$. From the previous item, $\mathbf{I}_n-\frac{1}{n}\mathbf{1}_{n\times n}$ is also an orthogonal projection onto the kernel of $\frac{1}{n}\mathbf{1}_{n\times n}$ which is $\mathbf{1}_n^{\perp}$.

**Theorem 2.11** (Inverse and determinant of partitioned matrix)**.** *Let* $\mathbf{M} = \begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{B}^T & \mathbf{C}\end{bmatrix}$ *be a symmetric, invertible (regular) and* $\mathbf{A}$ *is also invertible (regular). Then:*

*(a) The inverse matrix of* $\mathbf{M}$ *is given by:*
$$\mathbf{M}^{-1} = \begin{bmatrix}\mathbf{A}^{-1}+\mathbf{F}\mathbf{E}^{-1}\mathbf{F}^T & -\mathbf{F}\mathbf{E}^{-1}\\ -\mathbf{E}^{-1}\mathbf{F}^T & \mathbf{E}^{-1}\end{bmatrix}$$
*where* $\mathbf{E}$ *is the Schur complement given by* $\mathbf{E}=\mathbf{C}-\mathbf{B}^T\mathbf{A}^{-1}\mathbf{B}$ *and* $\mathbf{F}=\mathbf{A}^{-1}\mathbf{B}$*.*

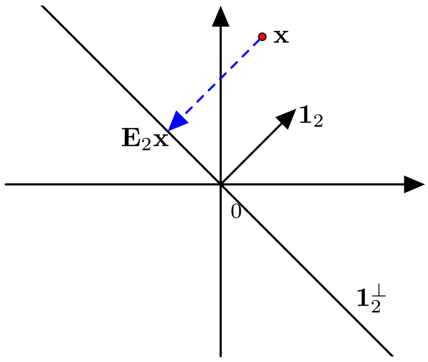


Figure 2.2: Orthogonal Projection of $\mathbf{E}_2$

*(b) The determinant of* $\mathbf{M}$ *is given by:*

$$\det(\mathbf{M}) = \det(\mathbf{A})\det(\mathbf{C} - \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B}).$$

There is also an extension of this theorem for general case where $\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$ (see [Mur12, p.118]).

**Definition 2.12** (Isometry)**.** A linear transformation $\mathbf{M} : \mathbb{R}^n \to \mathbb{R}^n$ is called an isometry if $\mathbf{x}^T\mathbf{x} = (\mathbf{M}\mathbf{x})^T(\mathbf{M}\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$.

Some properties of isometries are as follows:

- If $\mathbf{U}$ and $\mathbf{V}$ are isometries, then the product $\mathbf{U}\mathbf{V}$ is also an isometry.
- If $\mathbf{U}$ is an isometry, $|\det(\mathbf{U})| = 1$.
- If $\mathbf{U}$ is an isometry, then $|\lambda(\mathbf{U})| = 1$ for all eigenvalues of $\mathbf{U}$.

# 3 Multivariate Distributions and Moments

## 3.1 Random Vectors

Let $X_1, \dots, X_n$ be random variables on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$:

$$X_i : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{R}),$$

where $\mathcal{R}$ is the Borel $\sigma$-algebra generated by the open sets of $\mathbb{R}$.

- $\mathbf{X} = (X_1, \dots, X_p)^T$ is called a random vector.
- Similarly the matrix $\mathbf{X} = (X_{ij})_{1 \le i \le p, 1 \le j \le m}$ with random variables $X_{ij}$ as its elements is called a random matrix.
- The joint distribution of a random vector is uniquely described by its multivariate distribution function:

$$F(x_1, \dots, x_p) = \mathbb{P}(X_1 \le x_1, \dots, X_p \le x_p), (x_1, \dots, x_p) \in \mathbb{R}^p.$$

- A random vector $\mathbf{X} = (X_1, \dots, X_p)^T$ is called absolutely continuous if there exists an integrable function $f(x_1, \dots, x_n) \ge 0$ such that:

$$F(x_1, \dots, x_p) = \int_{-\infty}^{x_p} \cdots \int_{-\infty}^{x_1} f(x_1, \dots, x_p) \mathrm{d}x_1 \dots \mathrm{d}x_p.$$

  $f$ is called probability density function (pdf) and $F$ is called cumulative distribution function (cdf).

**Example 3.1.** (Multivariate normal distribution) The multivariate normal (or Gaussian) distribution has the following probability density function:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\mathbf{\Sigma}|^{1/2}} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}$$

with parameters $\boldsymbol{\mu} \in \mathbb{R}^p$, $\mathbf{\Sigma} \in \mathbb{R}^{p \times p}$, $\mathbf{\Sigma} \succ 0$.

This is denoted by $\mathbf{X} = (X_1, \dots, X_p)^T \sim N_p(\boldsymbol{\mu}, \mathbf{\Sigma})$. Note that $\mathbf{\Sigma}$ must have full rank. There exists an $n-$dimensional Gaussian random variable, if $\mathrm{rk}(\mathbf{\Sigma}) < p$, however it has no density function with respect to $p-$ dimensional Lebesgue measure.

## 3.2 Expectation and Covariance

Suppose that a random variable $\mathbf{X} = (X_1, \dots, X_p)^T$ is given.

**Definition 3.2.** (a) The expectation (vector) of a random vector $\mathbf{X}$, $\mathbb{E}(\mathbf{X})$, is defined by:

$$\mathbb{E}(\mathbf{X}) = (\mathbb{E}(X_1), \dots, \mathbb{E}(X_1))^T .$$

(b) The covariance matrix of a random vector $\mathbf{X}$, $\mathrm{Cov}(\mathbf{X})$, is defined by:

$$\mathrm{Cov}(\mathbf{X}) = \mathbb{E}\left((\mathbf{X} - \mathbb{E}(\mathbf{X}))(\mathbf{X} - \mathbb{E}(\mathbf{X}))^T\right).$$

Expectation vector is constructed component-wise of expectations $\mathbb{E}(X_i)$. Covariance matrix has as its $(i, j)$th element, the covariance value $\mathrm{Cov}(X_i, X_j)$:

$$(\mathrm{Cov}(\mathbf{X}))_{i,j} = \mathrm{Cov}(X_i, X_j) = \mathbb{E}\left((X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))\right).$$

**Theorem 3.3.** *Given random vectors* $\mathbf{X} = (X_1, \dots, X_p)^T$, $\mathbf{Y} = (Y_1, \dots, Y_p)^T$, *the following statements hold:*

*(a)* $\mathbb{E}(\mathbf{A}\mathbf{X} + \mathbf{b}) = \mathbf{A}\mathbb{E}(\mathbf{X}) + \mathbf{b}$

*(b)* $\mathbb{E}(\mathbf{X} + \mathbf{Y}) = \mathbb{E}(\mathbf{X}) + \mathbb{E}(\mathbf{Y})$

*(c)* $\mathrm{Cov}(\mathbf{A}\mathbf{X} + \mathbf{b}) = \mathbf{A}\mathrm{Cov}(\mathbf{X})\mathbf{A}^T$

*(d)* $\mathrm{Cov}(\mathbf{X} + \mathbf{Y}) = \mathrm{Cov}(\mathbf{X}) + \mathrm{Cov}(\mathbf{Y})$, *if* $\mathbf{X}$ *and* $\mathbf{Y}$ *are stochastically independent.*

*(e)* $\mathrm{Cov}(\mathbf{X}) \succeq 0$, *i.e., the covariance matrix is non-negative definite.*

*Proof.* Prove (a)-(d) as exercise. To prove the last part, let $\mathbf{a} \in \mathbb{R}^p$ be a vector. We have:

$$\mathbf{a}^T \mathrm{Cov}(\mathbf{X})\mathbf{a} \overset{(c)}{=} \mathrm{Cov}(\mathbf{a}^T\mathbf{X}) = \mathrm{Var}(\mathbf{a}^T\mathbf{X}) \geq 0.$$

□

- Show that if $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then:

$$\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu}, \mathrm{Cov}(\mathbf{X}) = \boldsymbol{\Sigma}.$$

**Theorem 3.4** (Steiner's rule)**.** *Given a random vector* $\mathbf{X} = (X_1, \dots, X_p)^T$, *it holds:*

$$\mathbb{E}\left((\mathbf{X} - \mathbf{b})(\mathbf{X} - \mathbf{b})^T\right) = \mathrm{Cov}(\mathbf{X}) + (\mathbf{b} - \mathbb{E}(\mathbf{X}))(\mathbf{b} - \mathbb{E}(\mathbf{X}))^T.$$

*Proof.* Let $\boldsymbol{\mu} = \mathbb{E}(\mathbf{X})$. Note that:

$$\mathbb{E}\left((\mathbf{X} - \boldsymbol{\mu})(\mathbf{b} - \boldsymbol{\mu})^T\right) = \mathbb{E}\left(\mathbf{X} - \boldsymbol{\mu}\right)\left(\mathbf{b} - \boldsymbol{\mu}\right)^T = 0.$$

Using this, we have:

$$\begin{aligned}
\mathbb{E}\left((\mathbf{X} - \mathbf{b})(\mathbf{X} - \mathbf{b})^T\right) &= \mathbb{E}\left((\mathbf{X} - \boldsymbol{\mu} + \boldsymbol{\mu} - \mathbf{b})(\mathbf{X} - \boldsymbol{\mu} + \boldsymbol{\mu} - \mathbf{b})^T\right) \\
&= \mathbb{E}\left((\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T\right) + \mathbb{E}\left((\boldsymbol{\mu} - \mathbf{b})(\boldsymbol{\mu} - \mathbf{b})^T\right) \\
&= \mathrm{Cov}(\mathbf{X}) + (\mathbf{b} - \mathbb{E}(\mathbf{X}))(\mathbf{b} - \mathbb{E}(\mathbf{X}))^T.
\end{aligned}$$

□

**Theorem 3.5.** *Let* $\mathbf{X}$ *be a random vector with* $\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu}$ *and* $\mathrm{Cov}(\mathbf{X}) = \mathbf{V}$. *Then:*

$$\mathbb{P}(\mathbf{X} \in \mathrm{Im}(\mathbf{V}) + \boldsymbol{\mu}) = 1.$$

*Proof.* Let $\ker(\mathbf{V}) = \{\mathbf{x} \in \mathbb{R}^p : \mathbf{V}\mathbf{x} = 0\}$ be the kernel (or null space) of $\mathbf{V}$. Assume a basis for the kernel as $\ker(\mathbf{V}) =< \mathbf{a}_1, \dots, \mathbf{a}_r >$. It holds for $i = 1, \dots, r$:

$$\mathrm{Var}(\mathbf{a}_i^T \mathbf{X}) = \mathrm{Cov}(\mathbf{a}_i^T \mathbf{X}) = \mathbf{a}_i^T \mathbf{V} \mathbf{a}_i = 0.$$

Since the variance of $\mathbf{a}_i^T \mathbf{X}$ is equal to zero, then $\mathbf{a}_i^T \mathbf{X}$ should be almost surely equal to its expectation which is $\mathbf{a}_i^T \boldsymbol{\mu}$. Hence $\mathbb{P}(\mathbf{a}_i^T \mathbf{X} = \mathbf{a}_i^T \boldsymbol{\mu}) = 1$, i.e., $\mathbb{P}(\mathbf{a}_i^T (\mathbf{X} - \boldsymbol{\mu}) = 0) = 1$. Hence:

$$\mathbb{P}((\mathbf{X} - \boldsymbol{\mu}) \in \mathbf{a}_i^\perp) = 1, \forall i = 1, \dots, r.$$

Using the fact that $\mathbb{P}(X \in A) = 1, \mathbb{P}(X \in B) = 1 \implies \mathbb{P}(X \in A \cap B) = 1$ (prove as exercise!), it holds that:

$$\mathbb{P}((\mathbf{X} - \boldsymbol{\mu}) \in \mathbf{a}_1^\perp \cap \dots \cap \mathbf{a}_r^\perp) = 1.$$

But $\mathrm{Im}(\mathbf{V}) = \ker(\mathbf{V})^\perp =< \mathbf{a}_1, \dots, \mathbf{a}_r >^\perp = \mathbf{a}_1^\perp \cap \dots \cap \mathbf{a}_r^\perp$. Therefore:

$$\mathbb{P}((\mathbf{X} - \boldsymbol{\mu}) \in \mathrm{Im}(\mathbf{V})) = 1.$$

□

## 3.3 Conditional Distribution

Let $\mathbf{X} = (X_1, \dots, X_p)^T$ be a random vector and $\mathbf{X} = (\mathbf{Y}_1, \mathbf{Y}_2)^T$ such that $\mathbf{Y}_1 = (X_1, \dots, X_k)$ and $\mathbf{Y}_2 = (X_{k+1}, \dots, X_p)$. Suppose that $\mathbf{X}$ is absolutely continuous with density $f_{\mathbf{X}}$. Then the conditional density of $\mathbf{Y}_1$ given $\mathbf{Y}_2 = \mathbf{y}_2$ is given by:

$$f_{\mathbf{Y}_1|\mathbf{Y}_2}(\mathbf{y}_1|\mathbf{y}_2) = \frac{f_{\mathbf{Y}_1,\mathbf{Y}_2}(\mathbf{y}_1, \mathbf{y}_2)}{f_{\mathbf{Y}_2}(\mathbf{y}_2)}, \quad \mathbf{y}_1 \in \mathbb{R}^k.$$

It also holds that:

$$\mathbb{P}(\mathbf{Y}_1 \in B | \mathbf{Y}_2 = \mathbf{y}_2) = \int_B f_{\mathbf{Y}_1|\mathbf{Y}_2}(\mathbf{y}_1|\mathbf{y}_2) \mathrm{d}\mathbf{y}_1, \quad \forall B \in \mathcal{R}^k.$$

**Theorem 3.6** ([Mur12, Theorem 4.3.1])**.** *Suppose that* $(\mathbf{Y}_1, \mathbf{Y}_2) = N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *and:*

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}, \boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \boldsymbol{\Lambda}_{11} & \boldsymbol{\Lambda}_{12} \\ \boldsymbol{\Lambda}_{21} & \boldsymbol{\Lambda}_{22} \end{bmatrix}.$$

*Then:*

*(a)* $\mathbf{Y}_1 \sim N_k(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ *and* $\mathbf{Y}_2 \sim N_{p-k}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$

(b) *The conditional density* $f_{\mathbf{Y}_1|\mathbf{Y}_2}(\mathbf{y}_1|\mathbf{y}_2)$ *is given by multivariate normal distribution* $N_k(\boldsymbol{\mu}_{1|2}, \boldsymbol{\Sigma}_{1|2})$ *with*

$$\begin{aligned}
\boldsymbol{\mu}_{1|2} &= \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{y}_2 - \boldsymbol{\mu}_2) \\
&= \boldsymbol{\mu}_1 - \boldsymbol{\Lambda}_{11}^{-1}\boldsymbol{\Lambda}_{12}(\mathbf{y}_2 - \boldsymbol{\mu}_2) \\
&= \boldsymbol{\Sigma}_{1|2}(\boldsymbol{\Lambda}_{11}\boldsymbol{\mu}_1 - \boldsymbol{\Lambda}_{12}(\mathbf{y}_2 - \boldsymbol{\mu}_2)) \\
\boldsymbol{\Sigma}_{1|2} &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} = \boldsymbol{\Lambda}_{11}^{-1}.
\end{aligned}$$

Note that $\boldsymbol{\Sigma}_{1|2}$ is the Schur complement, introduced in the previous chapter.