Marginalization, Conditioning, and Bayes Rule

Prof. James Richard Forbes

McGill University, Department of Mechanical Engineering

The Gaussian Distribution

- A continuous random variable $\mathbf{x} \in \mathbb{R}^{n_x}$ is said to have a normal or Gaussian distribution if the pdf associated with $\mathbf{x}$ is given by

$$p(\mathbf{x}) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^{n_x} \det \boldsymbol{\Sigma}}} \exp\left(-\tfrac{1}{2}\left(\mathbf{x} - \boldsymbol{\mu}\right)^{\mathsf{T}} \boldsymbol{\Sigma}^{-1} \left(\mathbf{x} - \boldsymbol{\mu}\right)\right), \tag{1}$$

  where $\bar{\mathbf{x}}$ is the mean and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{\mathsf{T}} > 0$ is the covariance matrix.
  - $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{\mathsf{T}} > 0$ means $\boldsymbol{\Sigma}$ is symmetric and positive definite, thus ensuring $\boldsymbol{\Sigma}$ is not singular, and thus $\boldsymbol{\Sigma}^{-1}$ exists.
- Often it is more convenient to write the Gaussian distribution in information form,

$$p(\mathbf{x}) = \mathcal{N}^{-1}(\boldsymbol{\eta}, \boldsymbol{\Lambda}) = \mathcal{N}^{-1}(\mathbf{x}; \boldsymbol{\eta}, \boldsymbol{\Lambda}) = \frac{\exp(-\frac{1}{2}\boldsymbol{\eta}^{\mathsf{T}}\boldsymbol{\Lambda}^{-1}\boldsymbol{\eta})}{\sqrt{(2\pi)^{n_x} \det \boldsymbol{\Lambda}^{-1}}} \exp\left(-\tfrac{1}{2}\mathbf{x}^{\mathsf{T}}\boldsymbol{\Lambda}\boldsymbol{x} + \mathbf{x}^{\mathsf{T}}\boldsymbol{\eta}\right), \tag{2}$$

  where $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1}$, and $\boldsymbol{\eta} = \boldsymbol{\Lambda}\boldsymbol{\mu}$.

- Equation (2) can be derived from (1). Starting from (1),

$$
\begin{aligned}
\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \frac{1}{\sqrt{(2\pi)^{n_x} \det \boldsymbol{\Sigma}}} \exp\left(-\tfrac{1}{2}\left(\mathbf{x}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\mathbf{x} - 2\mathbf{x}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \boldsymbol{\mu}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right)\right) \\
&= \frac{\exp\left(\boldsymbol{\mu}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right)}{\sqrt{(2\pi)^{n_x} \det \boldsymbol{\Sigma}}} \exp\left(-\tfrac{1}{2}\mathbf{x}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\mathbf{x} + \mathbf{x}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right) \\
&= \frac{\exp(-\tfrac{1}{2}\boldsymbol{\eta}^\mathsf{T}\boldsymbol{\Lambda}^{-1}\boldsymbol{\eta})}{\sqrt{(2\pi)^{n_x} \det \boldsymbol{\Lambda}^{-1}}} \exp\left(-\tfrac{1}{2}\mathbf{x}^\mathsf{T}\boldsymbol{\Lambda}\mathbf{x} + \mathbf{x}^\mathsf{T}\boldsymbol{\eta}\right) \\
&= \mathcal{N}^{-1}(\boldsymbol{\eta}, \boldsymbol{\Lambda}).
\end{aligned} \tag{3}
$$

- A short-hand notation for indicating $\mathbf{x}$ is normally distributed is

$$
\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathcal{N}^{-1}(\boldsymbol{\eta}, \boldsymbol{\Lambda}) = \mathcal{N}^{-1}(\mathbf{x}; \boldsymbol{\eta}, \boldsymbol{\Lambda}).
$$

Three Critical Operations

- Marginalization

$$p(\mathbf{x}) = \int_{-\infty}^{\infty} p(\mathbf{x}, \mathbf{y}) \mathrm{d}\mathbf{y} \tag{4}$$

- Conditioning

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x}|\mathbf{y}) p(\mathbf{y}) \tag{5}$$

- Bayes Rule

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x}) p(\mathbf{x})}{p(\mathbf{y})} \tag{6}$$

Three Critical Operations (Conditional Form)

- Marginalization

$$p(\mathbf{x}|\boldsymbol{\theta}) = \int_{-\infty}^{\infty} p(\mathbf{x}, \mathbf{y}|\boldsymbol{\theta}) \mathrm{d}\mathbf{y} \tag{7}$$

- Conditioning

$$p(\mathbf{x}, \mathbf{y}|\boldsymbol{\theta}) = p(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}) p(\mathbf{y}|\boldsymbol{\theta}) \tag{8}$$

- Bayes Rule

$$p(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}) = \frac{p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) p(\mathbf{x}|\boldsymbol{\theta})}{p(\mathbf{y}|\boldsymbol{\theta})} \tag{9}$$

The Schur Complement

Theorem (1)

[1] Consider the matrix $\mathbf{X} \in \mathbb{R}^{n\times n}$*. Partition* $\mathbf{X}$ *in block matrix form*

$$\mathbf{X} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}, \tag{10}$$

where $\mathbf{A} \in \mathbb{R}^{p\times p}$*,* $\mathbf{B} \in \mathbb{R}^{p\times q}$*,* $\mathbf{C} \in \mathbb{R}^{q\times p}$*, and* $\mathbf{D} \in \mathbb{R}^{q\times q}$*. If both* $\mathbf{D}$ *and* $\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}$ *are invertible then*

$$\mathbf{X}^{-1} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & -(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \end{bmatrix} \tag{11}$$

The matrix $\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}$ *is the Schur Complement of* $\mathbf{D}$ *in* $\mathbf{X}$*.*

Proof.

- Decompose $\mathbf{X}$ into three matrices,

$$\mathbf{X} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{1} & \mathbf{B}\mathbf{D}^{-1} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}}_{\mathbf{P}} \underbrace{\begin{bmatrix} \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix}}_{\mathbf{Q}} \underbrace{\begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{D}^{-1}\mathbf{C} & \mathbf{1} \end{bmatrix}}_{\mathbf{R}} \tag{12}$$

- Note that

$$\mathbf{P}^{-1} = \begin{bmatrix} \mathbf{1} & \mathbf{B}\mathbf{D}^{-1} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{1} & -\mathbf{B}\mathbf{D}^{-1} \\ \mathbf{0} & \mathbf{1} \end{bmatrix},$$

$$\mathbf{Q}^{-1} = \begin{bmatrix} \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^{-1} \end{bmatrix},$$

$$\mathbf{R}^{-1} = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{D}^{-1}\mathbf{C} & \mathbf{1} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ -\mathbf{D}^{-1}\mathbf{C} & \mathbf{1} \end{bmatrix}.$$

- Recall that, for any three invertible matrices $\mathbf{P}, \mathbf{Q}, \mathbf{R} \in \mathbb{R}^{n\times n}$,

$$(\mathbf{P}\mathbf{Q}\mathbf{R})^{-1} = \mathbf{R}^{-1}\mathbf{Q}^{-1}\mathbf{P}^{-1}.$$

- Therefore,

$$\mathbf{X}^{-1} = (\mathbf{P}\mathbf{Q}\mathbf{R})^{-1} = \mathbf{R}^{-1}\mathbf{Q}^{-1}\mathbf{P}^{-1} = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ -\mathbf{D}^{-1}\mathbf{C} & \mathbf{1} \end{bmatrix} \begin{bmatrix} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{1} & -\mathbf{B}\mathbf{D}^{-1} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}$$

$$= \begin{bmatrix} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & -(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \end{bmatrix}. \tag{11}$$

The Schur Complement (Alternative Version)

Theorem (2)

[1] Consider the matrix $\mathbf{X} \in \mathbb{R}^{n\times n}$. Partition $\mathbf{X}$ in block matrix form

$$\mathbf{X} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}, \tag{13}$$

where $\mathbf{A} \in \mathbb{R}^{p\times p}$, $\mathbf{B} \in \mathbb{R}^{p\times q}$, $\mathbf{C} \in \mathbb{R}^{q\times p}$, and $\mathbf{D} \in \mathbb{R}^{q\times q}$. If both $\mathbf{A}$ and $\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$ are invertible then

$$\mathbf{X}^{-1} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \\ -(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \end{bmatrix} \tag{14}$$

The matrix $\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$ is the Schur Complement of $\mathbf{A}$ in $\mathbf{X}$.

Proof.

- Decompose $\mathbf{X}$ into three matrices,

$$\mathbf{X} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{B}\mathbf{A}^{-1} & \mathbf{1} \end{bmatrix}}_{\mathbf{P}} \underbrace{\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B} \end{bmatrix}}_{\mathbf{Q}} \underbrace{\begin{bmatrix} \mathbf{1} & \mathbf{A}^{-1}\mathbf{C} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}}_{\mathbf{R}} \tag{15}$$

- Note that

$$\mathbf{P}^{-1} = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{B}\mathbf{A}^{-1} & \mathbf{1} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ -\mathbf{B}\mathbf{A}^{-1} & \mathbf{1} \end{bmatrix},$$

$$\mathbf{Q}^{-1} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0} & (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \end{bmatrix},$$

$$\mathbf{R}^{-1} = \begin{bmatrix} \mathbf{1} & \mathbf{A}^{-1}\mathbf{C} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{1} & -\mathbf{A}^{-1}\mathbf{C} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}.$$

- Recall that, for any three invertible matrices $\mathbf{P}, \mathbf{Q}, \mathbf{R} \in \mathbb{R}^{n \times n}$,

$$(\mathbf{PQR})^{-1} = \mathbf{R}^{-1}\mathbf{Q}^{-1}\mathbf{P}^{-1}.$$

- Therefore,

$$\begin{aligned} \mathbf{X}^{-1} = (\mathbf{PQR})^{-1} = \mathbf{R}^{-1}\mathbf{Q}^{-1}\mathbf{P}^{-1} &= \begin{bmatrix} \mathbf{1} & -\mathbf{A}^{-1}\mathbf{C} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0} & (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ -\mathbf{B}\mathbf{A}^{-1} & \mathbf{1} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \\ -(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \end{bmatrix}. \end{aligned} \tag{14}$$

The Matrix Inversion Lemma

- In the special case where $\mathbf{A}$, $\mathbf{D}$, the Schur complement of $\mathbf{A}$ in $\mathbf{X}$, $\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}$, and the Schur complement of $\mathbf{D}$ in $\mathbf{X}$, $\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$, are all invertible, equating the $(1,1)$ block of each expression for $\mathbf{X}^{-1}$ in Equations (11) and (14) leads to [1]

$$(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} = \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1},$$

  which is called the Woodbury matrix identity.
- The *Matrix Inversion Lemma* is,

$$(\mathbf{A} + \mathbf{B}\mathbf{C})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{1} + \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1}.$$

  which comes from setting $\mathbf{D} = \mathbf{1}$ and changing $\mathbf{B}$ to $-\mathbf{B}$ [1].

Determinant Results

Consider

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix},$$

where $\mathbf{A}$ and $\mathbf{D}$ are square.

- Assuming $\mathbf{A}$ is full rank,

$$\det \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = \det \mathbf{A} \det \left(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B} \right), \tag{16}$$

  which can be derived by taking the determinant of both sides of (15).
- Assuming $\mathbf{D}$ is full rank,

$$\det \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = \det \mathbf{D} \det \left(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C} \right), \tag{17}$$

  which can be derived by taking the determinant of both sides of (12).

Specialized Schur Complement Results

- Consider

$$\mathbf{z} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad \mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{bmatrix}, \quad \boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}_{xx} & \boldsymbol{\Lambda}_{xy} \\ \boldsymbol{\Lambda}_{yx} & \boldsymbol{\Lambda}_{yy} \end{bmatrix}, \tag{18}$$

where $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1}$, $\boldsymbol{\Sigma}_{yx} = \boldsymbol{\Sigma}_{xy}^{\mathsf{T}}$, and $\boldsymbol{\Lambda}_{yx} = \boldsymbol{\Lambda}_{xy}^{\mathsf{T}}$.

- The Schur complement of $\boldsymbol{\Sigma}_{yy}$ in $\boldsymbol{\Sigma}$ is

$$\begin{bmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{\Lambda}_{xx} & \boldsymbol{\Lambda}_{xy} \\ \boldsymbol{\Lambda}_{yx} & \boldsymbol{\Lambda}_{yy} \end{bmatrix} = \begin{bmatrix} (\boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx})^{-1} & -(\boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx})^{-1}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1} \\ -\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}(\boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx})^{-1} & \boldsymbol{\Sigma}_{yy}^{-1} + \boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}(\boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx})^{-1}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1} \end{bmatrix}. \tag{19}$$

- Thus,

$$\boldsymbol{\Lambda}_{xx}^{-1} = \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}. \tag{20}$$

- The Schur complement of $\boldsymbol{\Sigma}_{xx}$ in $\boldsymbol{\Sigma}$ is

$$\begin{bmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{\Lambda}_{xx} & \boldsymbol{\Lambda}_{xy} \\ \boldsymbol{\Lambda}_{yx} & \boldsymbol{\Lambda}_{yy} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{xx}^{-1} + \boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}(\boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy})^{-1}\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1} & -\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}(\boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy})^{-1} \\ -(\boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy})^{-1}\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1} & (\boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy})^{-1} \end{bmatrix}. \tag{21}$$

- Thus,

$$\boldsymbol{\Lambda}_{yy}^{-1} = \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}. \tag{22}$$

- The Schur complement of $\mathbf{\Lambda}_{yy}$ in $\mathbf{\Lambda}$ is

$$\begin{bmatrix} \mathbf{\Lambda}_{xx} & \mathbf{\Lambda}_{xy} \\ \mathbf{\Lambda}_{yx} & \mathbf{\Lambda}_{yy} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{\Sigma}_{xx} & \mathbf{\Sigma}_{xy} \\ \mathbf{\Sigma}_{yx} & \mathbf{\Sigma}_{yy} \end{bmatrix} =$$
$$\begin{bmatrix} (\mathbf{\Lambda}_{xx} - \mathbf{\Lambda}_{xy}\mathbf{\Lambda}_{yy}^{-1}\mathbf{\Lambda}_{yx})^{-1} & -(\mathbf{\Lambda}_{xx} - \mathbf{\Lambda}_{xy}\mathbf{\Lambda}_{yy}^{-1}\mathbf{\Lambda}_{yx})^{-1}\mathbf{\Lambda}_{xy}\mathbf{\Lambda}_{yy}^{-1} \\ -\mathbf{\Lambda}_{yy}^{-1}\mathbf{\Lambda}_{yx}(\mathbf{\Lambda}_{xx} - \mathbf{\Lambda}_{xy}\mathbf{\Lambda}_{yy}^{-1}\mathbf{\Lambda}_{yx})^{-1} & \mathbf{\Lambda}_{yy}^{-1} + \mathbf{\Lambda}_{yy}^{-1}\mathbf{\Lambda}_{yx}(\mathbf{\Lambda}_{xx} - \mathbf{\Lambda}_{xy}\mathbf{\Lambda}_{yy}^{-1}\mathbf{\Lambda}_{yx})^{-1}\mathbf{\Lambda}_{xy}\mathbf{\Lambda}_{yy}^{-1} \end{bmatrix}. \tag{23}$$

- Thus,

$$\mathbf{\Sigma}_{xx}^{-1} = \mathbf{\Lambda}_{xx} - \mathbf{\Lambda}_{xy}\mathbf{\Lambda}_{yy}^{-1}\mathbf{\Lambda}_{yx}. \tag{24}$$

- The Schur complement of $\mathbf{\Lambda}_{xx}$ in $\mathbf{\Lambda}$ is

$$\begin{bmatrix} \mathbf{\Lambda}_{xx} & \mathbf{\Lambda}_{xy} \\ \mathbf{\Lambda}_{yx} & \mathbf{\Lambda}_{yy} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{\Sigma}_{xx} & \mathbf{\Sigma}_{xy} \\ \mathbf{\Sigma}_{yx} & \mathbf{\Sigma}_{yy} \end{bmatrix} =$$
$$\begin{bmatrix} \mathbf{\Lambda}_{xx}^{-1} + \mathbf{\Lambda}_{xx}^{-1}\mathbf{\Lambda}_{xy}(\mathbf{\Lambda}_{yy} - \mathbf{\Lambda}_{yx}\mathbf{\Lambda}_{xx}^{-1}\mathbf{\Lambda}_{xy})^{-1}\mathbf{\Lambda}_{yx}\mathbf{\Lambda}_{xx}^{-1} & -\mathbf{\Lambda}_{xx}^{-1}\mathbf{\Lambda}_{xy}(\mathbf{\Lambda}_{yy} - \mathbf{\Lambda}_{yx}\mathbf{\Lambda}_{xx}^{-1}\mathbf{\Lambda}_{xy})^{-1} \\ -(\mathbf{\Lambda}_{yy} - \mathbf{\Lambda}_{yx}\mathbf{\Lambda}_{xx}^{-1}\mathbf{\Lambda}_{xy})^{-1}\mathbf{\Lambda}_{yx}\mathbf{\Lambda}_{xx}^{-1} & (\mathbf{\Lambda}_{yy} - \mathbf{\Lambda}_{yx}\mathbf{\Lambda}_{xx}^{-1}\mathbf{\Lambda}_{xy})^{-1} \end{bmatrix}. \tag{25}$$

- Thus,

$$\mathbf{\Sigma}_{yy}^{-1} = \mathbf{\Lambda}_{yy} - \mathbf{\Lambda}_{yx}\mathbf{\Lambda}_{xx}^{-1}\mathbf{\Lambda}_{xy}. \tag{26}$$

Marginalization Theorem

- Consider

$$\mathbf{z} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad \mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{bmatrix}, \quad \boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}_{xx} & \boldsymbol{\Lambda}_{xy} \\ \boldsymbol{\Lambda}_{yx} & \boldsymbol{\Lambda}_{yy} \end{bmatrix}. \tag{18}$$

Theorem (3)

Let $\mathbf{z} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *and be partitioned as in* (18). *Then, the marginal pdf* $p(\mathbf{x})$ *is given by*

$$\begin{aligned} p(\mathbf{x}) &= \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx}) \\ &= \mathcal{N}^{-1}(\bar{\boldsymbol{\eta}}, \bar{\boldsymbol{\Lambda}}) \end{aligned}$$

where

$$\bar{\boldsymbol{\eta}} = \boldsymbol{\eta}_x - \boldsymbol{\Lambda}_{xy}\boldsymbol{\Lambda}_{yy}^{-1}\boldsymbol{\eta}_y, \quad \bar{\boldsymbol{\Lambda}} = \boldsymbol{\Lambda}_{xx} - \boldsymbol{\Lambda}_{xy}\boldsymbol{\Lambda}_{yy}^{-1}\boldsymbol{\Lambda}_{yx}.$$

Proof.

- Recall that

$$p(\mathbf{x}) = \int_{-\infty}^{\infty} p(\mathbf{x}, \mathbf{y}) \mathrm{d}\mathbf{y}. \tag{27}$$

- The joint pdf is

$$p(\mathbf{x}, \mathbf{y}) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^{(n_x+n_y)} \det \boldsymbol{\Sigma}}} \exp(\alpha) \tag{28}$$

  where the argument of the exponent, α, is

$$\alpha = -\tfrac{1}{2} \left(\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} - \begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix} \right)^\mathsf{T} \begin{bmatrix} \boldsymbol{\Lambda}_{xx} & \boldsymbol{\Lambda}_{xy} \\ \boldsymbol{\Lambda}_{yx} & \boldsymbol{\Lambda}_{yy} \end{bmatrix} \left(\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} - \begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix} \right). \tag{29}$$

- Using the Schur complement of $\boldsymbol{\Lambda}_{yy}$ in $\boldsymbol{\Lambda}$, Theorem 1, enables α to be written as

$$\begin{aligned} \alpha = -\tfrac{1}{2} \left(\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} - \begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix} \right)^\mathsf{T} \begin{bmatrix} \mathbf{1} & \boldsymbol{\Lambda}_{xy}\boldsymbol{\Lambda}_{yy}^{-1} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}_{xx} - \boldsymbol{\Lambda}_{xy}\boldsymbol{\Lambda}_{yy}^{-1}\boldsymbol{\Lambda}_{yx} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_{yy} \end{bmatrix} \\ \times \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \boldsymbol{\Lambda}_{yy}^{-1}\boldsymbol{\Lambda}_{yx} & \mathbf{1} \end{bmatrix} \left(\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} - \begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix} \right). \end{aligned} \tag{30}$$

- Next, expand α, and write α as the sum of two quadratic terms:
  - a quadratic term in $(\mathbf{y}-\boldsymbol{\mu}_y)+\boldsymbol{\Lambda}_{yy}^{-1}\boldsymbol{\Lambda}_{yx}(\mathbf{x}-\boldsymbol{\mu}_x)$, called α_1, and
  - a quadratic term in $\mathbf{x}-\boldsymbol{\mu}_x$, called α_2.
- It follows that $\alpha=\alpha_1+\alpha_2$ where

$$\alpha_1=-\frac{1}{2}\left((\mathbf{y}-\boldsymbol{\mu}_y)+\boldsymbol{\Lambda}_{yy}^{-1}\boldsymbol{\Lambda}_{yx}(\mathbf{x}-\boldsymbol{\mu}_x)\right)^{\mathsf{T}}\boldsymbol{\Lambda}_{yy}\left((\mathbf{y}-\boldsymbol{\mu}_y)+\boldsymbol{\Lambda}_{yy}^{-1}\boldsymbol{\Lambda}_{yx}(\mathbf{x}-\boldsymbol{\mu}_x)\right), \tag{31}$$

$$\begin{aligned}\alpha_2&=-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_x)^{\mathsf{T}}\left(\boldsymbol{\Lambda}_{xx}-\boldsymbol{\Lambda}_{xy}\boldsymbol{\Lambda}_{yy}^{-1}\boldsymbol{\Lambda}_{yx}\right)(\mathbf{x}-\boldsymbol{\mu}_x)\\&=-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_x)^{\mathsf{T}}\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x}-\boldsymbol{\mu}_x),\end{aligned} \tag{32}$$

  and (24) has been used.
- The joint pdf $p(\mathbf{x},\mathbf{y})$ can now written as

$$p(\mathbf{x},\mathbf{y})=\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})=\frac{1}{\sqrt{(2\pi)^{(n_x+n_y)}\det\boldsymbol{\Sigma}}}\exp(\alpha_1)\exp(\alpha_2) \tag{33}$$

- Next,

$$
\begin{aligned}
p(\mathbf{x}) &= \int_{-\infty}^{\infty} p(\mathbf{x}, \mathbf{y}) \mathrm{d}\mathbf{y} \\
&= \frac{1}{\sqrt{(2\pi)^{(n_x+n_y)} \det \boldsymbol{\Sigma}}} \exp(\alpha_2) \int_{-\infty}^{\infty} \exp(\alpha_1) \mathrm{d}\mathbf{y}, \qquad (34)
\end{aligned}
$$

because $\exp(\alpha_2)$ does not depend on $\mathbf{y}$.

- The integral of $\exp(\alpha_1)$ over $\mathbf{y}$ is

$$
\int_{-\infty}^{\infty} \exp(\alpha_1) \mathrm{d}\mathbf{y} = \sqrt{(2\pi)^{n_y} \det \boldsymbol{\Lambda}_{yy}^{-1}}, \qquad (35)
$$

which follows from the fact that

$$
\int_{-\infty}^{\infty} \frac{1}{\sqrt{(2\pi)^{n_y} \det \boldsymbol{\Lambda}_{yy}^{-1}}} \exp(\alpha_1) \, \mathrm{d}\mathbf{y} = 1 \qquad (36)
$$

to satisfy the axiom of total probability.

- Therefore, the density $p(\mathbf{x})$ can be written as

$$p(\mathbf{x}) = \frac{\sqrt{\det \boldsymbol{\Lambda}_{yy}^{-1}}}{\sqrt{(2\pi)^{n_x} \det \boldsymbol{\Sigma}}} \exp\left(-\tfrac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_x)^\mathsf{T} \boldsymbol{\Sigma}_{xx}^{-1} (\mathbf{x} - \boldsymbol{\mu}_x)\right). \tag{37}$$

- Using (22),

$$\boldsymbol{\Lambda}_{yy}^{-1} = \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx} \boldsymbol{\Sigma}_{xx}^{-1} \boldsymbol{\Sigma}_{xy} \tag{22}$$

  and (16),

$$\det \begin{bmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{bmatrix} = \det \boldsymbol{\Sigma}_{xx} \det(\boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx} \boldsymbol{\Sigma}_{xx}^{-1} \boldsymbol{\Sigma}_{xy}), \tag{38}$$

  it follows that

$$p(\mathbf{x}) = \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx}) = \frac{1}{\sqrt{(2\pi)^{n_x} \det \boldsymbol{\Sigma}_{xx}}} \exp\left(-\tfrac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_x)^\mathsf{T} \boldsymbol{\Sigma}_{xx}^{-1} (\mathbf{x} - \boldsymbol{\mu}_x)\right). \tag{39}$$

- To derive the information form, first recall that

$$\begin{bmatrix} \boldsymbol{\eta}_x \\ \boldsymbol{\eta}_y \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Lambda}_{xx} & \boldsymbol{\Lambda}_{xy} \\ \boldsymbol{\Lambda}_{yx} & \boldsymbol{\Lambda}_{yy} \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Lambda}_{xx}\boldsymbol{\mu}_x + \boldsymbol{\Lambda}_{xy}\boldsymbol{\mu}_y, \\ \boldsymbol{\Lambda}_{yx}\boldsymbol{\mu}_x + \boldsymbol{\Lambda}_{yy}\boldsymbol{\mu}_y \end{bmatrix},$$

  or

$$\boldsymbol{\Lambda}_{xx}\boldsymbol{\mu}_x = \boldsymbol{\eta}_x - \boldsymbol{\Lambda}_{xy}\boldsymbol{\mu}_y \tag{40}$$

$$\boldsymbol{\Lambda}_{yx}\boldsymbol{\mu}_x = \boldsymbol{\eta}_y - \boldsymbol{\Lambda}_{yy}\boldsymbol{\mu}_y. \tag{41}$$

- Next, write the exponent α_2 as

$$\begin{aligned} \alpha_2 &= -\tfrac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_x)^\mathsf{T} \boldsymbol{\Sigma}_{xx}^{-1} (\mathbf{x} - \boldsymbol{\mu}_x) \\ &= -\tfrac{1}{2}\mathbf{x}^\mathsf{T} \boldsymbol{\Sigma}_{xx}^{-1}\mathbf{x} + \mathbf{x}^\mathsf{T}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\mu}_x - \tfrac{1}{2}\boldsymbol{\mu}_x^\mathsf{T}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\mu}_x \end{aligned} \tag{42}$$

- Using (24) in concert with (40) and (41), the middle term in α_2 can be written

$$\begin{aligned} \mathbf{x}^\mathsf{T}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\mu}_x &= \mathbf{x}^\mathsf{T}\left(\boldsymbol{\Lambda}_{xx} - \boldsymbol{\Lambda}_{xy}\boldsymbol{\Lambda}_{yy}^{-1}\boldsymbol{\Lambda}_{yx}\right)\boldsymbol{\mu}_x & (43) \\ &= \mathbf{x}^\mathsf{T}\left(\boldsymbol{\Lambda}_{xx}\boldsymbol{\mu}_x - \boldsymbol{\Lambda}_{xy}\boldsymbol{\Lambda}_{yy}^{-1}\boldsymbol{\Lambda}_{yx}\boldsymbol{\mu}_x\right) \\ &= \mathbf{x}^\mathsf{T}\left(\boldsymbol{\eta}_x - \boldsymbol{\Lambda}_{xy}\boldsymbol{\mu}_y - \boldsymbol{\Lambda}_{xy}\boldsymbol{\Lambda}_{yy}^{-1}(\boldsymbol{\eta}_y - \boldsymbol{\Lambda}_{yy}\boldsymbol{\mu}_y)\right) \\ &= \mathbf{x}^\mathsf{T}\left(\boldsymbol{\eta}_x - \boldsymbol{\Lambda}_{xy}\boldsymbol{\mu}_y - \boldsymbol{\Lambda}_{xy}\boldsymbol{\Lambda}_{yy}^{-1}\boldsymbol{\eta}_y + \boldsymbol{\Lambda}_{xy}\boldsymbol{\mu}_y\right) \\ &= \mathbf{x}^\mathsf{T}\left(\boldsymbol{\eta}_x - \boldsymbol{\Lambda}_{xy}\boldsymbol{\Lambda}_{yy}^{-1}\boldsymbol{\eta}_y\right) \\ &= \mathbf{x}^\mathsf{T}\bar{\boldsymbol{\eta}}, & (44) \end{aligned}$$

  where

$$\bar{\boldsymbol{\eta}} = \boldsymbol{\eta}_x - \boldsymbol{\Lambda}_{xy}\boldsymbol{\Lambda}_{yy}^{-1}\boldsymbol{\eta}_y. \tag{45}$$

- The last term in α_2 can be written as

$$
\begin{aligned}
& -\tfrac{1}{2}\boldsymbol{\mu}_x^\mathsf{T}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\mu}_x && (46)\\
&= -\tfrac{1}{2}\boldsymbol{\mu}_x^\mathsf{T}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\mu}_x \\
&= -\tfrac{1}{2}(\boldsymbol{\eta}_x - \boldsymbol{\Lambda}_{xy}\boldsymbol{\Lambda}_{yy}^{-1}\boldsymbol{\eta}_y)^\mathsf{T}\left(\boldsymbol{\Lambda}_{xx} - \boldsymbol{\Lambda}_{xy}\boldsymbol{\Lambda}_{yy}^{-1}\boldsymbol{\Lambda}_{yx}\right)^{-1}(\boldsymbol{\eta}_x - \boldsymbol{\Lambda}_{xy}\boldsymbol{\Lambda}_{yy}^{-1}\boldsymbol{\eta}_y) \\
&= -\tfrac{1}{2}\bar{\boldsymbol{\eta}}^\mathsf{T}\bar{\boldsymbol{\Lambda}}\bar{\boldsymbol{\eta}}, && (47)
\end{aligned}
$$

where (45) has been used, and

$$\bar{\boldsymbol{\Lambda}} = \boldsymbol{\Lambda}_{xx} - \boldsymbol{\Lambda}_{xy}\boldsymbol{\Lambda}_{yy}^{-1}\boldsymbol{\Lambda}_{yx}. \tag{48}$$

- Thus, the exponent α_2 can be written as

$$\alpha_2 = -\tfrac{1}{2}\mathbf{x}^\mathsf{T}\boldsymbol{\Sigma}_{xx}^{-1}\mathbf{x} + \mathbf{x}^\mathsf{T}\bar{\boldsymbol{\eta}} - \tfrac{1}{2}\bar{\boldsymbol{\eta}}^\mathsf{T}\bar{\boldsymbol{\Lambda}}\bar{\boldsymbol{\eta}}. \tag{49}$$

- It follows that

$$
\begin{aligned}
p(\mathbf{x}) &= \frac{\exp(-\tfrac{1}{2}\bar{\boldsymbol{\eta}}^\mathsf{T}\bar{\boldsymbol{\Lambda}}\bar{\boldsymbol{\eta}})}{\sqrt{(2\pi)^{n_x}\det\bar{\boldsymbol{\Lambda}}}}\exp\left(-\tfrac{1}{2}\mathbf{x}^\mathsf{T}\bar{\boldsymbol{\Lambda}}\mathbf{x} + \mathbf{x}^\mathsf{T}\bar{\boldsymbol{\eta}}\right) && (50)\\
&= \mathcal{N}^{-1}(\bar{\boldsymbol{\eta}}, \bar{\boldsymbol{\Lambda}}), && (51)
\end{aligned}
$$

where

$$
\begin{aligned}
\bar{\boldsymbol{\eta}} &= \boldsymbol{\eta}_x - \boldsymbol{\Lambda}_{xy}\boldsymbol{\Lambda}_{yy}^{-1}\boldsymbol{\eta}_y, && (45)\\
\bar{\boldsymbol{\Lambda}} &= \boldsymbol{\Lambda}_{xx} - \boldsymbol{\Lambda}_{xy}\boldsymbol{\Lambda}_{yy}^{-1}\boldsymbol{\Lambda}_{yx}. && (48)
\end{aligned}
$$

Conditioning Theorem

- Consider

$$\mathbf{z} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad \mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{bmatrix}, \quad \boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \boldsymbol{\Lambda}_{xx} & \boldsymbol{\Lambda}_{xy} \\ \boldsymbol{\Lambda}_{yx} & \boldsymbol{\Lambda}_{yy} \end{bmatrix}. \tag{18}$$

Theorem (4)

Let $\mathbf{z} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *and be partitioned as in* (18). *Then, the conditional pdf* $p(\mathbf{x}|\mathbf{y})$ *is given by*

$$\begin{aligned} p(\mathbf{x}|\mathbf{y}) &= \mathcal{N}(\boldsymbol{\mu}_{x|y}, \boldsymbol{\Sigma}_{x|y}), \\ &= \mathcal{N}^{-1}(\bar{\boldsymbol{\eta}}_{x|y}, \bar{\boldsymbol{\Lambda}}_{x|y}), \end{aligned}$$

where

$$\boldsymbol{\mu}_{x|y} = \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y), \qquad \boldsymbol{\Sigma}_{x|y} = \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx},$$

and

$$\bar{\boldsymbol{\eta}}_{x|y} = \boldsymbol{\eta}_x - \boldsymbol{\Lambda}_{xy}\mathbf{y}, \qquad \bar{\boldsymbol{\Lambda}}_{x|y} = \boldsymbol{\Lambda}_{xx}.$$

Proof.

- Recall that the relationship between $p(\mathbf{x}, \mathbf{y})$, $p(\mathbf{x}|\mathbf{y})$, and $p(\mathbf{y})$ is given by

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{y})}, \tag{52}$$

- The densities $p(\mathbf{y})$ and $p(\mathbf{x}, \mathbf{y})$ are

$$p(\mathbf{y}) = \mathcal{N}(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_{yy}) = \frac{1}{\sqrt{(2\pi)^{n_y} \det \boldsymbol{\Sigma}_{yy}}} \exp\left(-\tfrac{1}{2}(\mathbf{y} - \boldsymbol{\mu}_y)^\mathsf{T} \boldsymbol{\Sigma}_{yy}^{-1} (\mathbf{y} - \boldsymbol{\mu}_y)\right), \tag{53}$$

$$p(\mathbf{x}, \mathbf{y}) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^{(n_x+n_y)} \det \boldsymbol{\Sigma}}} \exp(\alpha), \tag{54}$$

and the argument of the exponent, α, is

$$\alpha = -\tfrac{1}{2}\left(\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix} - \begin{bmatrix}\boldsymbol{\mu}_x\\ \boldsymbol{\mu}_y\end{bmatrix}\right)^\mathsf{T} \begin{bmatrix}\boldsymbol{\Lambda}_{xx} & \boldsymbol{\Lambda}_{xy}\\ \boldsymbol{\Lambda}_{yx} & \boldsymbol{\Lambda}_{yy}\end{bmatrix} \left(\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix} - \begin{bmatrix}\boldsymbol{\mu}_x\\ \boldsymbol{\mu}_y\end{bmatrix}\right). \tag{55}$$

- Using the Schur complement of $\boldsymbol{\Lambda}_{xx}$ in $\boldsymbol{\Lambda}$, Theorem 2, enables α to be written as

$$\alpha = -\tfrac{1}{2}\left(\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix} - \begin{bmatrix}\boldsymbol{\mu}_x\\ \boldsymbol{\mu}_y\end{bmatrix}\right)^{\mathsf{T}} \begin{bmatrix}\mathbf{1} & \mathbf{0}\\ \boldsymbol{\Lambda}_{yx}\boldsymbol{\Lambda}_{xx}^{-1} & \mathbf{1}\end{bmatrix} \begin{bmatrix}\boldsymbol{\Lambda}_{xx} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Lambda}_{yy} - \boldsymbol{\Lambda}_{yx}\boldsymbol{\Lambda}_{xx}^{-1}\boldsymbol{\Lambda}_{xy}\end{bmatrix} \times \begin{bmatrix}\mathbf{1} & \boldsymbol{\Lambda}_{xx}^{-1}\boldsymbol{\Lambda}_{xy}\\ \mathbf{0} & \mathbf{1}\end{bmatrix}\left(\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix} - \begin{bmatrix}\boldsymbol{\mu}_x\\ \boldsymbol{\mu}_y\end{bmatrix}\right). \tag{56}$$

- Next, expand α, and write α as the sum of two quadratic terms:
  - a quadratic term in $(\mathbf{x} - \boldsymbol{\mu}_x) + \boldsymbol{\Lambda}_{xx}^{-1}\boldsymbol{\Lambda}_{xy}(\mathbf{y} - \boldsymbol{\mu}_y)$, called α_3, and
  - a quadratic term in $\mathbf{y} - \boldsymbol{\mu}_y$, called α_4.
- It follows that $\alpha = \alpha_3 + \alpha_4$ where

$$\begin{aligned}\alpha_3 &= -\tfrac{1}{2}\left((\mathbf{x} - \boldsymbol{\mu}_x) + \boldsymbol{\Lambda}_{xx}^{-1}\boldsymbol{\Lambda}_{xy}(\mathbf{y} - \boldsymbol{\mu}_y)\right)^{\mathsf{T}} \boldsymbol{\Lambda}_{xx}\left((\mathbf{x} - \boldsymbol{\mu}_x) + \boldsymbol{\Lambda}_{xx}^{-1}\boldsymbol{\Lambda}_{xy}(\mathbf{y} - \boldsymbol{\mu}_y)\right),\\ &= -\tfrac{1}{2}\left(\mathbf{x} - (\boldsymbol{\mu}_x - \boldsymbol{\Lambda}_{xx}^{-1}\boldsymbol{\Lambda}_{xy}(\mathbf{y} - \boldsymbol{\mu}_y))\right)^{\mathsf{T}} \boldsymbol{\Lambda}_{xx}\left(\mathbf{x} - (\boldsymbol{\mu}_x - \boldsymbol{\Lambda}_{xx}^{-1}\boldsymbol{\Lambda}_{xy}(\mathbf{y} - \boldsymbol{\mu}_y))\right)\end{aligned} \tag{57}$$

$$\begin{aligned}\alpha_4 &= -\tfrac{1}{2}(\mathbf{y} - \boldsymbol{\mu}_y)^{\mathsf{T}}\left(\boldsymbol{\Lambda}_{yy} - \boldsymbol{\Lambda}_{yx}\boldsymbol{\Lambda}_{xx}^{-1}\boldsymbol{\Lambda}_{xy}\right)(\mathbf{y} - \boldsymbol{\mu}_y)\\ &= -\tfrac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_x)^{\mathsf{T}}\boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x),\end{aligned} \tag{58}$$

and (26) has been used.

- Next, focus on writing α_3 as

$$\alpha_3 = -\frac{1}{2}\left(\mathbf{x} - \boldsymbol{\mu}_{x|y}\right)^{\mathsf{T}} \boldsymbol{\Sigma}_{x|y}^{-1} \left(\mathbf{x} - \boldsymbol{\mu}_{x|y}\right) \tag{59}$$

  where

$$\boldsymbol{\mu}_{x|y} = \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y), \tag{60}$$

$$\boldsymbol{\Sigma}_{x|y} = \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}. \tag{61}$$

- From (20),

$$\boldsymbol{\Sigma}_{x|y}^{-1} = \boldsymbol{\Lambda}_{xx} = (\boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx})^{-1}. \tag{62}$$

- To write $\boldsymbol{\mu}_{x|y}$ in (60) in terms of $\boldsymbol{\Sigma}_{xy}$ and $\boldsymbol{\Sigma}_{yy}$, consider the following.
  - Using (19), $\boldsymbol{\Lambda}_{xx}$ and $\boldsymbol{\Lambda}_{xy}$ are

$$\boldsymbol{\Lambda}_{xx} = (\boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx})^{-1}, \tag{63}$$

$$\boldsymbol{\Lambda}_{xy} = -(\boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx})^{-1}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}. \tag{64}$$

  - As such, the matrix product $-\boldsymbol{\Lambda}_{xx}^{-1}\boldsymbol{\Lambda}_{xy}$ can be written as

$$\begin{aligned} -\boldsymbol{\Lambda}_{xx}^{-1}\boldsymbol{\Lambda}_{xy} &= (\boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx})(\boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx})^{-1}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1} \\ &= \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}. \end{aligned} \tag{65}$$

  - Thus

$$\boldsymbol{\mu}_{x|y} = \boldsymbol{\mu}_x - \boldsymbol{\Lambda}_{xx}^{-1}\boldsymbol{\Lambda}_{xy}(\mathbf{y} - \boldsymbol{\mu}_y) = \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y). \tag{66}$$

- It follows that α_3 can be written as in (59).

- The joint pdf $p(\mathbf{x}, \mathbf{y})$ can now written as

$$p(\mathbf{x}, \mathbf{y}) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^{(n_x+n_y)} \det \boldsymbol{\Sigma}}} \exp(\alpha_3) \exp(\alpha_4) \tag{67}$$

- It follows that

$$\begin{aligned} p(\mathbf{x}|\mathbf{y}) &= p(\mathbf{x}, \mathbf{y}) \left(\frac{1}{p(\mathbf{y})} \right) \\ &= \frac{1}{\sqrt{(2\pi)^{(n_x+n_y)} \det \boldsymbol{\Sigma}}} \exp(\alpha_3) \exp(\alpha_4) \left(\sqrt{(2\pi)^{n_y} \det \boldsymbol{\Sigma}_{yy}} \exp(-\alpha_4) \right) \\ &= \frac{\sqrt{\det \boldsymbol{\Sigma}_{yy}}}{\sqrt{(2\pi)^{n_x} \det \boldsymbol{\Sigma}}} \exp(-\tfrac{1}{2} \left(\mathbf{x} - \boldsymbol{\mu}_{x|y}\right)^{\mathsf{T}} \boldsymbol{\Sigma}_{x|y}^{-1} \left(\mathbf{x} - \boldsymbol{\mu}_{x|y}\right)). \end{aligned} \tag{68}$$

- Focusing on the normalization constant, using (17), first note that ,

$$\det \begin{bmatrix} \mathbf{\Sigma}_{xx} & \mathbf{\Sigma}_{xy} \\ \mathbf{\Sigma}_{yx} & \mathbf{\Sigma}_{yy} \end{bmatrix} = \det \mathbf{\Sigma}_{yy} \det(\mathbf{\Sigma}_{xx} - \mathbf{\Sigma}_{xy}\mathbf{\Sigma}_{yy}^{-1}\mathbf{\Sigma}_{yx}). \tag{69}$$

- Then, the normalization constant can be simplified as

$$\begin{aligned} \frac{\sqrt{\det \mathbf{\Sigma}_{yy}}}{\sqrt{(2\pi)^{n_x} \det \mathbf{\Sigma}}} &= \frac{\sqrt{\det \mathbf{\Sigma}_{yy}}}{\sqrt{(2\pi)^{n_x} \det \mathbf{\Sigma}_{yy} \det(\mathbf{\Sigma}_{xx} - \mathbf{\Sigma}_{xy}\mathbf{\Sigma}_{yy}^{-1}\mathbf{\Sigma}_{yx})}} \\ &= \frac{1}{\sqrt{(2\pi)^{n_x} \det(\mathbf{\Sigma}_{xx} - \mathbf{\Sigma}_{xy}\mathbf{\Sigma}_{yy}^{-1}\mathbf{\Sigma}_{yx})}}. \end{aligned} \tag{70}$$

- Therefore,

$$p(\mathbf{x}|\mathbf{y}) = \mathcal{N}(\boldsymbol{\mu}_{x|y}, \mathbf{\Sigma}_{x|y}) = \frac{1}{\sqrt{(2\pi)^{n_x} \det \mathbf{\Sigma}_{x|y}}} \exp(-\tfrac{1}{2} \left(\mathbf{x} - \boldsymbol{\mu}_{x|y}\right)^{\mathsf{T}} \mathbf{\Sigma}_{x|y}^{-1} \left(\mathbf{x} - \boldsymbol{\mu}_{x|y}\right)), \tag{71}$$

where

$$\boldsymbol{\mu}_{x|y} = \boldsymbol{\mu}_x + \mathbf{\Sigma}_{xy}\mathbf{\Sigma}_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y), \tag{60}$$

$$\mathbf{\Sigma}_{x|y} = \mathbf{\Lambda}_{xx}^{-1} = \mathbf{\Sigma}_{xx} - \mathbf{\Sigma}_{xy}\mathbf{\Sigma}_{yy}^{-1}\mathbf{\Sigma}_{yx}. \tag{61}$$

- To derive the information form, first recall that

$$\begin{bmatrix} \boldsymbol{\eta}_x \\ \boldsymbol{\eta}_y \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Lambda}_{xx} & \boldsymbol{\Lambda}_{xy} \\ \boldsymbol{\Lambda}_{yx} & \boldsymbol{\Lambda}_{yy} \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Lambda}_{xx}\boldsymbol{\mu}_x + \boldsymbol{\Lambda}_{xy}\boldsymbol{\mu}_y, \\ \boldsymbol{\Lambda}_{yx}\boldsymbol{\mu}_x + \boldsymbol{\Lambda}_{yy}\boldsymbol{\mu}_y \end{bmatrix},$$

or

$$\boldsymbol{\Lambda}_{xx}\boldsymbol{\mu}_x = \boldsymbol{\eta}_x - \boldsymbol{\Lambda}_{xy}\boldsymbol{\mu}_y \tag{40}$$

$$\boldsymbol{\Lambda}_{yx}\boldsymbol{\mu}_x = \boldsymbol{\eta}_y - \boldsymbol{\Lambda}_{yy}\boldsymbol{\mu}_y. \tag{41}$$

- Expanding the exponent, α_3, results in

$$\begin{aligned} \alpha_3 &= -\tfrac{1}{2}\left(\mathbf{x} - \boldsymbol{\mu}_{x|y}\right)^{\mathsf{T}} \boldsymbol{\Sigma}_{x|y}^{-1} \left(\mathbf{x} - \boldsymbol{\mu}_{x|y}\right) \\ &= -\tfrac{1}{2}\mathbf{x}^{\mathsf{T}}\boldsymbol{\Sigma}_{x|y}^{-1}\mathbf{x} + \mathbf{x}^{\mathsf{T}}\boldsymbol{\Sigma}_{x|y}^{-1}\boldsymbol{\mu}_{x|y} - \tfrac{1}{2}\boldsymbol{\mu}_{x|y}^{\mathsf{T}}\boldsymbol{\Sigma}_{x|y}^{-1}\boldsymbol{\mu}_{x|y}. \end{aligned} \tag{72}$$

- Using (62), (66), and (40), the middle term in α_3 can be written as

$$\begin{aligned} \mathbf{x}^{\mathsf{T}}\boldsymbol{\Sigma}_{x|y}^{-1}\boldsymbol{\mu}_{x|y} &= \mathbf{x}^{\mathsf{T}}\boldsymbol{\Lambda}_{xx}\left(\boldsymbol{\mu}_x - \boldsymbol{\Lambda}_{xx}^{-1}\boldsymbol{\Lambda}_{xy}(\mathbf{y} - \boldsymbol{\mu}_y)\right) && (73) \\ &= \mathbf{x}^{\mathsf{T}}\left(\boldsymbol{\Lambda}_{xx}\boldsymbol{\mu}_x - \boldsymbol{\Lambda}_{xy}(\mathbf{y} - \boldsymbol{\mu}_y)\right) \\ &= \mathbf{x}^{\mathsf{T}}\left(\boldsymbol{\eta}_x - \boldsymbol{\Lambda}_{xy}\boldsymbol{\mu}_y - \boldsymbol{\Lambda}_{xy}\mathbf{y} + \boldsymbol{\Lambda}_{xy}\boldsymbol{\mu}_y\right) \\ &= \mathbf{x}^{\mathsf{T}}\left(\boldsymbol{\eta}_x - \boldsymbol{\Lambda}_{xy}\mathbf{y}\right) \\ &= \mathbf{x}^{\mathsf{T}}\bar{\boldsymbol{\eta}}_{x|y}, && (74) \end{aligned}$$

where

$$\bar{\boldsymbol{\eta}}_{x|y} = \boldsymbol{\eta}_x - \boldsymbol{\Lambda}_{xy}\mathbf{y}. \tag{75}$$

- The last term in α_3 can be written as

$$
\begin{aligned}
& -\tfrac{1}{2}\boldsymbol{\mu}_{x|y}^{\mathsf{T}}\boldsymbol{\Sigma}_{x|y}^{-1}\boldsymbol{\mu}_{x|y} && (76)\\
&= -\tfrac{1}{2}\boldsymbol{\mu}_{x|y}^{\mathsf{T}}\boldsymbol{\Lambda}_{xx}\boldsymbol{\Lambda}_{xx}^{-1}\boldsymbol{\Lambda}_{xx}\boldsymbol{\mu}_{x|y} \\
&= -\tfrac{1}{2}\left(\boldsymbol{\mu}_x - \boldsymbol{\Lambda}_{xx}^{-1}\boldsymbol{\Lambda}_{xy}(\mathbf{y}-\boldsymbol{\mu}_y)\right)^{\mathsf{T}}\boldsymbol{\Lambda}_{xx}\boldsymbol{\Lambda}_{xx}^{-1}\boldsymbol{\Lambda}_{xx}\left(\boldsymbol{\mu}_x - \boldsymbol{\Lambda}_{xx}^{-1}\boldsymbol{\Lambda}_{xy}(\mathbf{y}-\boldsymbol{\mu}_y)\right) \\
&= -\bar{\boldsymbol{\eta}}_{x|y}^{\mathsf{T}}\bar{\boldsymbol{\Lambda}}_{x|y}^{-1}\bar{\boldsymbol{\eta}}_{x|y} && (77)
\end{aligned}
$$

  where

$$\bar{\boldsymbol{\Lambda}}_{x|y} = \boldsymbol{\Lambda}_{xx}. \quad (78)$$

- Thus, the exponent α_3 can be written as

$$
\begin{aligned}
\alpha_3 &= -\tfrac{1}{2}\mathbf{x}^{\mathsf{T}}\boldsymbol{\Sigma}_{x|y}^{-1}\mathbf{x} + \mathbf{x}^{\mathsf{T}}\boldsymbol{\Sigma}_{x|y}^{-1}\boldsymbol{\mu}_{x|y} - \tfrac{1}{2}\boldsymbol{\mu}_{x|y}^{\mathsf{T}}\boldsymbol{\Sigma}_{x|y}^{-1}\boldsymbol{\mu}_{x|y} \\
&= -\tfrac{1}{2}\mathbf{x}^{\mathsf{T}}\bar{\boldsymbol{\Lambda}}_{x|y}\mathbf{x} + \mathbf{x}^{\mathsf{T}}\bar{\boldsymbol{\eta}}_{x|y} - \bar{\boldsymbol{\eta}}_{x|y}^{\mathsf{T}}\bar{\boldsymbol{\Lambda}}_{x|y}^{-1}\bar{\boldsymbol{\eta}}_{x|y}. && (79)
\end{aligned}
$$

- It follows that

$$
\begin{aligned}
p(\mathbf{x}|\mathbf{y}) &= \frac{\exp -\bar{\boldsymbol{\eta}}_{x|y}^{\mathsf{T}}\bar{\boldsymbol{\Lambda}}_{x|y}^{-1}\bar{\boldsymbol{\eta}}_{x|y}}{\sqrt{(2\pi)^{n_x}\det\bar{\boldsymbol{\Lambda}}_{x|y}}}\exp(-\tfrac{1}{2}\mathbf{x}^{\mathsf{T}}\bar{\boldsymbol{\Lambda}}_{x|y}\mathbf{x} + \mathbf{x}^{\mathsf{T}}\bar{\boldsymbol{\eta}}_{x|y}) \\
&= \mathcal{N}^{-1}(\bar{\boldsymbol{\eta}}_{x|y}, \bar{\boldsymbol{\Lambda}}_{x|y}),
\end{aligned}
$$

  where

$$
\begin{aligned}
\bar{\boldsymbol{\eta}}_{x|y} &= \boldsymbol{\eta}_x - \boldsymbol{\Lambda}_{xy}\mathbf{y}, && (75)\\
\bar{\boldsymbol{\Lambda}}_{x|y} &= \boldsymbol{\Lambda}_{xx}. && (78)
\end{aligned}
$$

Summary

$$p(\mathbf{x},\mathbf{y}) = \mathcal{N}\left(\begin{bmatrix}\boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy}\end{bmatrix}\right) = \mathcal{N}^{-1}\left(\begin{bmatrix}\boldsymbol{\eta}_x \\ \boldsymbol{\eta}_y\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Lambda}_{xx} & \boldsymbol{\Lambda}_{xy} \\ \boldsymbol{\Lambda}_{yx} & \boldsymbol{\Lambda}_{yy}\end{bmatrix}\right)$$

	Marginalization $p(\mathbf{x}) = \int_{-\infty}^{\infty} p(\mathbf{x},\mathbf{y})\mathrm{d}\mathbf{y}$	Conditioning $p(\mathbf{x},\mathbf{y}) = p(\mathbf{x}\|\mathbf{y})p(\mathbf{y})$
Cov. Form	$\boldsymbol{\mu} = \boldsymbol{\mu}_x$ $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{xx}$	$\boldsymbol{\mu}_{x\|y} = \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)$ $\boldsymbol{\Sigma}_{x\|y} = \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}$
Info. Form	$\bar{\boldsymbol{\eta}} = \boldsymbol{\eta}_x - \boldsymbol{\Lambda}_{xy}\boldsymbol{\Lambda}_{yy}^{-1}\boldsymbol{\eta}_y$ $\bar{\boldsymbol{\Lambda}} = \boldsymbol{\Lambda}_{xx} - \boldsymbol{\Lambda}_{xy}\boldsymbol{\Lambda}_{yy}^{-1}\boldsymbol{\Lambda}_{yx}$	$\bar{\boldsymbol{\eta}}_{x\|y} = \boldsymbol{\eta}_x - \boldsymbol{\Lambda}_{xy}\mathbf{y}$ $\bar{\boldsymbol{\Lambda}}_{x\|y} = \boldsymbol{\Lambda}_{xx}$

(Eustice *et al.* 2006, IEEE Trans. Robotics)

Joint Density with Affine Transformation (Lemma A.1 in Särkkä)

Theorem (5)

Consider $\mathbf{x} \in \mathbb{R}^{n_x}$ *and* $\mathbf{y} \in \mathbb{R}^{n_y}$ *where* $\mathbf{y} = \mathbf{M}\mathbf{x} + \mathbf{b}$*,* $\mathbf{M} \in \mathbb{R}^{n_y \times n_x}$ *is constant, and* $\mathbf{b} \in \mathbb{R}^{n_y}$ *is constant. Let*

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx}), \quad p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}; \mathbf{M}\mathbf{x} + \mathbf{b}, \boldsymbol{\Sigma}_{y|x}).$$

Then, the joint density of $\mathbf{x}$ *and* $\mathbf{y}$*,* $p(\mathbf{x}, \mathbf{y})$*, is*

$$p(\mathbf{x}, \mathbf{y}) = \mathcal{N}\left(\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}; \begin{bmatrix} \boldsymbol{\mu}_x \\ \mathbf{M}\boldsymbol{\mu}_x + \mathbf{b} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xx}\mathbf{M}^\mathsf{T} \\ \mathbf{M}\boldsymbol{\Sigma}_{xx} & \mathbf{M}\boldsymbol{\Sigma}_{xx}\mathbf{M}^\mathsf{T} + \boldsymbol{\Sigma}_{y|x} \end{bmatrix}\right).$$

Proof.

- Recall that

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x}). \tag{80}$$

- The task at hand is to compute

$$p(\mathbf{x}, \mathbf{y}) \quad \text{given} \quad p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}; \mathbf{M}\mathbf{x} + \mathbf{b}, \boldsymbol{\Sigma}_{y|x}), \quad p(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx}). \tag{81}$$

- By direct computation,

$$\begin{aligned} p(\mathbf{x}, \mathbf{y}) &= p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \\ &= \frac{1}{\sqrt{(2\pi)^{n_y} \det \boldsymbol{\Sigma}_{y|x}}} \exp\left(-\tfrac{1}{2}\left(\mathbf{y} - (\mathbf{M}\mathbf{x} + \mathbf{b})\right)^{\mathsf{T}} \boldsymbol{\Sigma}_{y|x}^{-1} \left(\mathbf{y} - (\mathbf{M}\mathbf{x} + \mathbf{b})\right)\right) \\ &\qquad \times \frac{1}{\sqrt{(2\pi)^{n_x} \det \boldsymbol{\Sigma}_{xx}}} \exp\left(-\tfrac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_x)^{\mathsf{T}} \boldsymbol{\Sigma}_{xx}^{-1} (\mathbf{x} - \boldsymbol{\mu}_x)\right) \\ &= \frac{1}{\sqrt{(2\pi)^{n_x + n_y} \det \boldsymbol{\Sigma}_{y|x} \det \boldsymbol{\Sigma}_{xx}}} \exp(\alpha), \end{aligned} \tag{82}$$

where

$$\alpha = -\tfrac{1}{2}\left(\mathbf{y} - (\mathbf{M}\mathbf{x} + \mathbf{b})\right)^{\mathsf{T}} \boldsymbol{\Sigma}_{y|x}^{-1} \left(\mathbf{y} - (\mathbf{M}\mathbf{x} + \mathbf{b})\right) - \tfrac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_x)^{\mathsf{T}} \boldsymbol{\Sigma}_{xx}^{-1} (\mathbf{x} - \boldsymbol{\mu}_x). \tag{83}$$

- Write $\alpha = \alpha_5 + \alpha_6$ where

$$\alpha_5 = -\tfrac{1}{2}\left(\mathbf{y} - (\mathbf{M}\mathbf{x} + \mathbf{b})\right)^{\mathsf{T}} \boldsymbol{\Sigma}_{y|x}^{-1} \left(\mathbf{y} - (\mathbf{M}\mathbf{x} + \mathbf{b})\right), \tag{84}$$

$$\alpha_6 = -\tfrac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_x)^{\mathsf{T}} \boldsymbol{\Sigma}_{xx}^{-1} (\mathbf{x} - \boldsymbol{\mu}_x). \tag{85}$$

- Add and subtract $\mathbf{M}\boldsymbol{\mu}_x$ in the first term of α_5 and rearrange to yield

$$\begin{aligned}
\alpha_5 &= -\tfrac{1}{2}\left(\mathbf{y} - (\mathbf{M}\mathbf{x} + \mathbf{b}) + \mathbf{M}\boldsymbol{\mu}_x - \mathbf{M}\boldsymbol{\mu}_x\right)^{\mathsf{T}} \boldsymbol{\Sigma}_{y|x}^{-1} \left(\mathbf{y} - (\mathbf{M}\mathbf{x} + \mathbf{b}) + \mathbf{M}\boldsymbol{\mu}_x - \mathbf{M}\boldsymbol{\mu}_x\right) \\
&= -\tfrac{1}{2}\left(\mathbf{y} - (\mathbf{M}\boldsymbol{\mu}_x + \mathbf{b}) - \mathbf{M}(\mathbf{x} - \boldsymbol{\mu}_x)\right)^{\mathsf{T}} \boldsymbol{\Sigma}_{y|x}^{-1} \left(\mathbf{y} - (\mathbf{M}\boldsymbol{\mu}_x + \mathbf{b}) - \mathbf{M}(\mathbf{x} - \boldsymbol{\mu}_x)\right) \\
&= -\tfrac{1}{2}\left(\mathbf{y} - (\mathbf{M}\boldsymbol{\mu}_x + \mathbf{b})\right)^{\mathsf{T}} \boldsymbol{\Sigma}_{y|x}^{-1} \left(\mathbf{y} - (\mathbf{M}\boldsymbol{\mu}_x + \mathbf{b})\right) \\
&\quad + \left(\mathbf{y} - (\mathbf{M}\boldsymbol{\mu}_x + \mathbf{b})\right)^{\mathsf{T}} \boldsymbol{\Sigma}_{y|x}^{-1} \mathbf{M}(\mathbf{x} - \boldsymbol{\mu}_x) \\
&\quad - \tfrac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_x)^{\mathsf{T}} \mathbf{M}^{\mathsf{T}} \boldsymbol{\Sigma}_{y|x}^{-1} \mathbf{M}(\mathbf{x} - \boldsymbol{\mu}_x) \\
&= -\tfrac{1}{2}\begin{bmatrix} \mathbf{x} - \boldsymbol{\mu}_x \\ \mathbf{y} - (\mathbf{M}\boldsymbol{\mu}_x + \mathbf{b}) \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} \mathbf{M}^{\mathsf{T}}\boldsymbol{\Sigma}_{y|x}^{-1}\mathbf{M} & -\mathbf{M}^{\mathsf{T}}\boldsymbol{\Sigma}_{y|x}^{-1} \\ -\boldsymbol{\Sigma}_{y|x}^{-1}\mathbf{M} & \boldsymbol{\Sigma}_{y|x}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{x} - \boldsymbol{\mu}_x \\ \mathbf{y} - (\mathbf{M}\boldsymbol{\mu}_x + \mathbf{b}) \end{bmatrix}.
\end{aligned} \tag{86}$$

- It follows that α can be written as

$$\alpha = -\frac{1}{2}\begin{bmatrix}\mathbf{x}-\boldsymbol{\mu}_x \\ \mathbf{y}-(\mathbf{M}\boldsymbol{\mu}_x+\mathbf{b})\end{bmatrix}^\mathsf{T}\begin{bmatrix}\boldsymbol{\Sigma}_{xx}^{-1}+\mathbf{M}^\mathsf{T}\boldsymbol{\Sigma}_{y|x}^{-1}\mathbf{M} & -\mathbf{M}^\mathsf{T}\boldsymbol{\Sigma}_{y|x}^{-1} \\ -\boldsymbol{\Sigma}_{y|x}^{-1}\mathbf{M} & \boldsymbol{\Sigma}_{y|x}^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{x}-\boldsymbol{\mu}_x \\ \mathbf{y}-(\mathbf{M}\boldsymbol{\mu}_x+\mathbf{b})\end{bmatrix} \tag{87}$$

- Next, partition the information matrix as

$$\mathbf{X}^{-1} = \begin{bmatrix}\boldsymbol{\Sigma}_{xx}^{-1}+\mathbf{M}^\mathsf{T}\boldsymbol{\Sigma}_{y|x}^{-1}\mathbf{M} & -\mathbf{M}^\mathsf{T}\boldsymbol{\Sigma}_{y|x}^{-1} \\ -\boldsymbol{\Sigma}_{y|x}^{-1}\mathbf{M} & \boldsymbol{\Sigma}_{y|x}^{-1}\end{bmatrix} \tag{88}$$

$$= \begin{bmatrix}\mathbf{1} & -\mathbf{M}^\mathsf{T} \\ \mathbf{0} & \mathbf{1}\end{bmatrix}\begin{bmatrix}\boldsymbol{\Sigma}_{xx}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{y|x}^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{1} & -\mathbf{M}^\mathsf{T} \\ -\mathbf{M} & \mathbf{1}\end{bmatrix}. \tag{89}$$

- Using the fact that if $\mathbf{X}=\mathbf{PQR}$, then $\mathbf{X}^{-1}=\mathbf{R}^{-1}\mathbf{Q}^{-1}\mathbf{P}^{-1}$, it follows that the covariance matrix is

$$\mathbf{X} = \begin{bmatrix}\mathbf{1} & \mathbf{0} \\ \mathbf{M} & \mathbf{1}\end{bmatrix}\begin{bmatrix}\boldsymbol{\Sigma}_{xx} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{y|x}\end{bmatrix}\begin{bmatrix}\mathbf{1} & \mathbf{M}^\mathsf{T} \\ \mathbf{0} & \mathbf{1}\end{bmatrix}$$

$$= \begin{bmatrix}\boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xx}\mathbf{M}^\mathsf{T} \\ \mathbf{M}\boldsymbol{\Sigma}_{xx} & \mathbf{M}\boldsymbol{\Sigma}_{xx}\mathbf{M}^\mathsf{T}+\boldsymbol{\Sigma}_{y|x}\end{bmatrix}. \tag{90}$$

- To conclude,

$$\begin{aligned} p(\mathbf{x}, \mathbf{y}) &= p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \\ &= \frac{1}{\sqrt{(2\pi)^{n_x+n_y} \det \boldsymbol{\Sigma}_{y|x} \det \boldsymbol{\Sigma}_{xx}}} \exp(\alpha), \end{aligned} \tag{91}$$

where

$$\alpha = -\frac{1}{2} \begin{bmatrix} \mathbf{x} - \boldsymbol{\mu}_x \\ \mathbf{y} - (\mathbf{M}\boldsymbol{\mu}_x + \mathbf{b}) \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xx}\mathbf{M}^{\mathsf{T}} \\ \mathbf{M}\boldsymbol{\Sigma}_{xx} & \mathbf{M}\boldsymbol{\Sigma}_{xx}\mathbf{M}^{\mathsf{T}} + \boldsymbol{\Sigma}_{y|x} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{x} - \boldsymbol{\mu}_x \\ \mathbf{y} - (\mathbf{M}\boldsymbol{\mu}_x + \mathbf{b}) \end{bmatrix}. \tag{92}$$

Marginal and Cond. Densities with Affine Trans.

Theorem (6)

Consider $\mathbf{x} \in \mathbb{R}^{n_x}$ *and* $\mathbf{y} \in \mathbb{R}^{n_y}$ *where* $\mathbf{y} = \mathbf{M}\mathbf{x} + \mathbf{b}$*,* $\mathbf{M} \in \mathbb{R}^{n_y \times n_x}$ *is constant, and* $\mathbf{b} \in \mathbb{R}^{n_y}$ *is constant. Let*

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx}), \quad p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}; \mathbf{M}\mathbf{x} + \mathbf{b}, \boldsymbol{\Sigma}_{y|x}).$$

Then, the marginal density of $\mathbf{y}$*,* $p(\mathbf{y})$*, is*

$$p(\mathbf{y}) = \mathcal{N}\left(\mathbf{y}; \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y\right),$$

where

$$\begin{aligned} \boldsymbol{\mu}_y &= \mathbf{M}\boldsymbol{\mu}_x + \mathbf{b}, \\ \boldsymbol{\Sigma}_y &= \boldsymbol{\Sigma}_{y|x} + \mathbf{M}\boldsymbol{\Sigma}_{xx}\mathbf{M}^{\mathsf{T}}, \end{aligned}$$

and the conditional density of $\mathbf{x}$ *given* $\mathbf{y}$*,* $p(\mathbf{x}|\mathbf{y})$*, is*

$$p(\mathbf{x}|\mathbf{y}) = \mathcal{N}\left(\mathbf{x}; \boldsymbol{\mu}_{x|y}, \boldsymbol{\Sigma}_{x|y}\right),$$

where

$$\begin{aligned} \boldsymbol{\mu}_{x|y} &= \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xx}\mathbf{M}^{\mathsf{T}}\boldsymbol{\Sigma}_y^{-1}(\mathbf{y} - (\mathbf{M}\boldsymbol{\mu}_x + \mathbf{b})), \\ \boldsymbol{\Sigma}_{x|y} &= \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xx}\mathbf{M}^{\mathsf{T}}\boldsymbol{\Sigma}_y^{-1}\mathbf{M}\boldsymbol{\Sigma}_{xx}. \end{aligned}$$

Proof.

- Recall that

$$p(\mathbf{x},\mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x}). \tag{93}$$

- Using Theorem 5 (Joint Density with Affine Transformation Theorem),

$$p(\mathbf{x},\mathbf{y}) = \mathcal{N}\left(\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix}; \begin{bmatrix}\boldsymbol{\mu}_x \\ \mathbf{M}\boldsymbol{\mu}_x + \mathbf{b}\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xx}\mathbf{M}^\mathsf{T} \\ \mathbf{M}\boldsymbol{\Sigma}_{xx} & \mathbf{M}\boldsymbol{\Sigma}_{xx}\mathbf{M}^\mathsf{T} + \boldsymbol{\Sigma}_{y|x}\end{bmatrix}\right). \tag{94}$$

- Using Theorem 3 (Marginalization Theorem),

$$p(\mathbf{y}) = \mathcal{N}\left(\mathbf{y}; \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y\right) \tag{95}$$

  where

$$\begin{aligned}\boldsymbol{\mu}_y &= \mathbf{M}\boldsymbol{\mu}_x + \mathbf{b}, \\ \boldsymbol{\Sigma}_y &= \boldsymbol{\Sigma}_{y|x} + \mathbf{M}\boldsymbol{\Sigma}_{xx}\mathbf{M}^\mathsf{T}.\end{aligned}$$

- Next, recall that

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x}|\mathbf{y})p(\mathbf{y}), \quad \text{or} \quad p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{y})}. \tag{96}$$

- Recall that

$$p(\mathbf{y}) = \mathcal{N}\left(\mathbf{y}; \mathbf{M}\boldsymbol{\mu}_x + \mathbf{b}, \boldsymbol{\Sigma}_{y|x} + \mathbf{M}\boldsymbol{\Sigma}_{xx}\mathbf{M}^{\mathsf{T}}\right), \tag{97}$$

$$p(\mathbf{x}, \mathbf{y}) = \mathcal{N}\left(\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}; \begin{bmatrix} \boldsymbol{\mu}_x \\ \mathbf{M}\boldsymbol{\mu}_x + \mathbf{b} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xx}\mathbf{M}^{\mathsf{T}} \\ \mathbf{M}\boldsymbol{\Sigma}_{xx} & \mathbf{M}\boldsymbol{\Sigma}_{xx}\mathbf{M}^{\mathsf{T}} + \boldsymbol{\Sigma}_{y|x} \end{bmatrix}\right). \tag{98}$$

- Using Theorem 4 (Conditioning Theorem),

$$p(\mathbf{x}|\mathbf{y}) = \mathcal{N}(\boldsymbol{\mu}_{x|y}, \boldsymbol{\Sigma}_{x|y}),$$

where

$$\boldsymbol{\mu}_{x|y} = \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xx}\mathbf{M}^{\mathsf{T}}(\boldsymbol{\Sigma}_{y|x} + \mathbf{M}\boldsymbol{\Sigma}_{xx}\mathbf{M}^{\mathsf{T}})^{-1}(\mathbf{y} - (\mathbf{M}\boldsymbol{\mu}_x + \mathbf{b})),$$
$$\boldsymbol{\Sigma}_{x|y} = \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xx}\mathbf{M}^{\mathsf{T}}(\mathbf{M}\boldsymbol{\Sigma}_{xx}\mathbf{M}^{\mathsf{T}} + \boldsymbol{\Sigma}_{y|x})^{-1}\mathbf{M}\boldsymbol{\Sigma}_{xx}.$$

Bayes' Rule

- The joint pdf $p(\mathbf{x}, \mathbf{y})$ can be written as

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x}|\mathbf{y})p(\mathbf{y}), \quad \text{or} \tag{99}$$

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x}). \tag{100}$$

- Setting (99) equal to (100) and rearranging leads to Bayes' rule,

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}. \tag{101}$$

- Write $p(\mathbf{y})$ as

$$\begin{aligned} p(\mathbf{y}) = p(\mathbf{y}) \underbrace{\int_{\mathbf{a}}^{\mathbf{b}} p(\mathbf{x}|\mathbf{y}) \, \mathrm{d}\mathbf{x}}_{=1} = \int_{\mathbf{a}}^{\mathbf{b}} p(\mathbf{x}|\mathbf{y})p(\mathbf{y}) \, \mathrm{d}\mathbf{x} = \int_{\mathbf{a}}^{\mathbf{b}} p(\mathbf{x}, \mathbf{y}) \, \mathrm{d}\mathbf{x} \\ = \int_{\mathbf{a}}^{\mathbf{b}} p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \, \mathrm{d}\mathbf{x}. \end{aligned}$$

- It follows that Bayes' rule can then be written as

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{\int_{\mathbf{a}}^{\mathbf{b}} p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \, \mathrm{d}\mathbf{x}}.$$

- This applies to non-Gaussian, as well as Gaussian, distributions.

- In the Gaussian case, this is in fact exactly what was done in Theorem 6, the Marginal and Conditional Densities with Affine Transformation Theorem.
- Given $p(\mathbf{y}|\mathbf{x})$ and $p(\mathbf{x})$, construct the joint density $p(\mathbf{x},\mathbf{y})$ using

$$p(\mathbf{x},\mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x}).$$

- Using Theorem 3,the Marginalization Theorem, marginalize out $\mathbf{x}$ from $p(\mathbf{x},\mathbf{y})$ to get $p(\mathbf{y})$.
- Given the joint density $p(\mathbf{x},\mathbf{y})$ and $p(\mathbf{y})$, each of which have just been computed, compute

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x},\mathbf{y})}{p(\mathbf{y})}$$

  using Theorem 4, the Conditioning Theorem.

References

Slides are based on [2–6].

[1] J. Gallier, *The Schur Complement and Symmetric Positive Semidefinite (and Definite) Matrices*. Course Notes, 2010.

[2] T. B. Schön and F. Lindsten, "Manipulating the multivariate gaussian density," *Division of Automatic Control, Linköping University*, 2011.

[3] R. M. Eustice, H. Singh, and J. J. Leonard, "Exactly Sparse Delayed-State Filters for View-Based SLAM," *IEEE Trans. Robotics*, vol. 22, no. 6, pp. 1100–1114, 2006.

[4] S. Särkkä, *Bayesian Filtering and Smoothing*. New York, NY: Cambridge University Press, 2013.

[5] M. Kok, "Bayesian State Estimation for Sensor Fusion Using Inertial Sensors," *Guest Lectures, TU Berlin, January 23-24, 2019*,

[6] T. D. Barfoot, *State Estimation for Robotics*. New York, New York: Cambridge University Press, 2017.