## 7.3 Matrix inversion

### 7.3.1 The inverse of a square matrix

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{I} = \mathbf{A}\mathbf{A}^{-1}.$$

1. Unique?

$$B = BI = B(AC) = (BA)C = IC = C$$

2. Exists iff |A| != 0  i.e. rank(A) = n, 向量组线性无关    If $\det(\mathbf{A}) = 0$, it is called a **singular** matrix.

$$|A||A^{-1}| = |I| = 1 \qquad |A| \neq 0$$

$$A^*A = AA^* = |A|I$$ 行列式展开定理

$$A^{-1} = \frac{A^*}{|A|}$$

$\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$ are non-singular:

$$(\mathbf{A}^{-1})^{-1} = \mathbf{A}$$

$$AA^{-1} = E$$

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$$

$$ABB^{-1}A^{-1} = E$$

$$(\mathbf{A}^{-1})^{\mathsf{T}} = (\mathbf{A}^{\mathsf{T}})^{-1} \triangleq \mathbf{A}^{-T}$$

$$A'(A^{-1})' = (A^{-1}A)' = E' = E$$

For the case of a $2 \times 2$ matrix

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|} \boxed{\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}}$$

$$\begin{pmatrix} A_{11} & A_{21} \\ A_{12} & A_{22} \end{pmatrix}$$

For a block diagonal matrix,

$$\begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}^{-1} \end{pmatrix}$$

$$\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \begin{pmatrix} A^{-1} & 0 \\ 0 & B^{-1} \end{pmatrix} = \begin{pmatrix} AA^{-1} & 0 \\ 0 & BB^{-1} \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}$$

### 7.3.2 Schur complements *

$$\mathbf{M} = \begin{pmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{pmatrix}$$

*where we assume* $\mathbf{E}$ *and* $\mathbf{H}$ *are invertible. We have*

$$\boxed{\begin{aligned}\mathbf{M}^{-1} &= \begin{pmatrix} (\mathbf{M}/\mathbf{H})^{-1} & -(\mathbf{M}/\mathbf{H})^{-1}\mathbf{F}\mathbf{H}^{-1} \\ -\mathbf{H}^{-1}\mathbf{G}(\mathbf{M}/\mathbf{H})^{-1} & \mathbf{H}^{-1} + \mathbf{H}^{-1}\mathbf{G}(\mathbf{M}/\mathbf{H})^{-1}\mathbf{F}\mathbf{H}^{-1} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{E}^{-1} + \mathbf{E}^{-1}\mathbf{F}(\mathbf{M}/\mathbf{E})^{-1}\mathbf{G}\mathbf{E}^{-1} & -\mathbf{E}^{-1}\mathbf{F}(\mathbf{M}/\mathbf{E})^{-1} \\ -(\mathbf{M}/\mathbf{E})^{-1}\mathbf{G}\mathbf{E}^{-1} & (\mathbf{M}/\mathbf{E})^{-1} \end{pmatrix}\end{aligned}}$$

*where*

$$\mathbf{M}/\mathbf{H} \triangleq \mathbf{E} - \mathbf{F}\mathbf{H}^{-1}\mathbf{G}$$

$$\mathbf{M}/\mathbf{E} \triangleq \mathbf{H} - \mathbf{G}\mathbf{E}^{-1}\mathbf{F}$$

**partitioned inverse formulae.**

*We say that* $\mathbf{M}/\mathbf{H}$ *is the* **Schur complement** *of* $\mathbf{M}$ *wrt* $\mathbf{H}$, *and* $\mathbf{M}/\mathbf{E}$ *is the Schur complement of* $\mathbf{M}$ *wrt* $\mathbf{E}$.

*Proof.*

block diagonalize $\mathbf{M}$, it would be easier to invert.

$$\begin{pmatrix} \mathrm{I} & -\mathrm{FH}^{-1} \\ 0 & \mathrm{I} \end{pmatrix}\begin{pmatrix} \mathrm{E} & \mathrm{F} \\ \mathrm{G} & \mathrm{H} \end{pmatrix} = \begin{pmatrix} \mathrm{E}-\mathrm{FH}^{-1}\mathrm{G} & 0 \\ \mathrm{G} & \mathrm{H} \end{pmatrix}$$

$$\begin{pmatrix} \mathrm{I} & 0 \\ -GE^{-1} & \mathrm{I} \end{pmatrix}\begin{pmatrix} \mathrm{E} & \mathrm{F} \\ \mathrm{G} & \mathrm{H} \end{pmatrix} = \begin{pmatrix} \mathrm{E} & F \\ 0 & H-GE^{-1}F \end{pmatrix}$$

$$\begin{pmatrix} \mathrm{E}-\mathrm{FH}^{-1}\mathrm{G} & 0 \\ \mathrm{G} & \mathrm{H} \end{pmatrix}\begin{pmatrix} \mathrm{I} & 0 \\ -\mathrm{H}^{-1}\mathrm{G} & \mathrm{I} \end{pmatrix} = \begin{pmatrix} \mathrm{E}-\mathrm{FH}^{-1}\mathrm{G} & 0 \\ 0 & \mathrm{H} \end{pmatrix}$$

$$\begin{pmatrix} \mathrm{E} & F \\ 0 & H-GE^{-1}F \end{pmatrix}\begin{pmatrix} \mathrm{I} & -E^{-1}F \\ 0 & I \end{pmatrix} = \begin{pmatrix} \mathrm{E} & 0 \\ 0 & \frac{M}{E} \end{pmatrix}$$

$$\underbrace{\begin{pmatrix} \mathrm{I} & -\mathrm{FH}^{-1} \\ 0 & \mathrm{I} \end{pmatrix}}_{\mathrm{X}}\underbrace{\begin{pmatrix} \mathrm{E} & \mathrm{F} \\ \mathrm{G} & \mathrm{H} \end{pmatrix}}_{\mathrm{M}}\underbrace{\begin{pmatrix} \mathrm{I} & 0 \\ -\mathrm{H}^{-1}\mathrm{G} & \mathrm{I} \end{pmatrix}}_{\mathrm{Z}} = \underbrace{\begin{pmatrix} \mathrm{E}-\mathrm{FH}^{-1}\mathrm{G} & 0 \\ 0 & \mathrm{H} \end{pmatrix}}_{\mathrm{W}}$$

$$\begin{pmatrix} \mathrm{I} & 0 \\ -GE^{-1} & \mathrm{I} \end{pmatrix}\begin{pmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{pmatrix}\begin{pmatrix} \mathrm{I} & -E^{-1}F \\ 0 & I \end{pmatrix} = \begin{pmatrix} \mathrm{E} & 0 \\ 0 & \frac{M}{E} \end{pmatrix}$$

$$\mathrm{Z}^{-1}\mathrm{M}^{-1}\mathrm{X}^{-1} = \mathrm{W}^{-1}$$
$$\mathrm{M}^{-1} = \mathrm{ZW}^{-1}\mathrm{X}$$

Unique!

$$\begin{pmatrix} \mathrm{I} & -E^{-1}F \\ 0 & I \end{pmatrix}\begin{pmatrix} \mathrm{E}^{-1} & 0 \\ 0 & (\frac{M}{E})^{-1} \end{pmatrix}\begin{pmatrix} \mathrm{I} & 0 \\ -GE^{-1} & \mathrm{I} \end{pmatrix}$$

$$\begin{pmatrix} \mathrm{E} & \mathrm{F} \\ \mathrm{G} & \mathrm{H} \end{pmatrix}^{-1} = \begin{pmatrix} \mathrm{I} & 0 \\ -\mathrm{H}^{-1}\mathrm{G} & \mathrm{I} \end{pmatrix}\begin{pmatrix} (\mathrm{M/H})^{-1} & 0 \\ 0 & \mathrm{H}^{-1} \end{pmatrix}\begin{pmatrix} \mathrm{I} & -\mathrm{FH}^{-1} \\ 0 & \mathrm{I} \end{pmatrix}$$
$$= \begin{pmatrix} (\mathrm{M/H})^{-1} & 0 \\ -\mathrm{H}^{-1}\mathrm{G}(\mathrm{M/H})^{-1} & \mathrm{H}^{-1} \end{pmatrix}\begin{pmatrix} \mathrm{I} & -\mathrm{FH}^{-1} \\ 0 & \mathrm{I} \end{pmatrix}$$
$$= \begin{pmatrix} \boxed{(\mathrm{M/H})^{-1}} & -(\mathrm{M/H})^{-1}\mathrm{FH}^{-1} \\ -\mathrm{H}^{-1}\mathrm{G}(\mathrm{M/H})^{-1} & \mathrm{H}^{-1}+\mathrm{H}^{-1}\mathrm{G}(\mathrm{M/H})^{-1}\mathrm{FH}^{-1} \end{pmatrix}$$

$$\begin{pmatrix} \mathrm{E}^{-1} & -E^{-1}F(\frac{M}{E})^{-1} \\ 0 & (\frac{\mathrm{M}}{\mathrm{E}})^{-1} \end{pmatrix}\begin{pmatrix} \mathrm{I} & 0 \\ -GE^{-1} & \mathrm{I} \end{pmatrix}$$

$$\begin{pmatrix} \boxed{\mathbf{E}^{-1}+\mathbf{E}^{-1}\mathbf{F}(\mathbf{M/E})^{-1}\mathbf{GE}^{-1}} & -\mathbf{E}^{-1}\mathbf{F}(\mathbf{M/E})^{-1} \\ -(\mathbf{M/E})^{-1}\mathbf{GE}^{-1} & (\mathbf{M/E})^{-1} \end{pmatrix}$$

$$\begin{pmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{H}^{-1}\mathbf{G} & \mathbf{I} \end{pmatrix} \begin{pmatrix} (\mathbf{M}/\mathbf{H})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{I} & -\mathbf{F}\mathbf{H}^{-1} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$$

$$= \begin{pmatrix} (\mathbf{M}/\mathbf{H})^{-1} & \mathbf{0} \\ -\mathbf{H}^{-1}\mathbf{G}(\mathbf{M}/\mathbf{H})^{-1} & \mathbf{H}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{I} & -\mathbf{F}\mathbf{H}^{-1} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$$

$$= \begin{pmatrix} \boxed{(\mathbf{M}/\mathbf{H})^{-1}} & -(\mathbf{M}/\mathbf{H})^{-1}\mathbf{F}\mathbf{H}^{-1} \\ -\mathbf{H}^{-1}\mathbf{G}(\mathbf{M}/\mathbf{H})^{-1} & \mathbf{H}^{-1} + \mathbf{H}^{-1}\mathbf{G}(\mathbf{M}/\mathbf{H})^{-1}\mathbf{F}\mathbf{H}^{-1} \end{pmatrix}$$

$$\begin{pmatrix} \mathrm{E}^{-1} & -E^{-1}F(\frac{M}{E})^{-1} \\ 0 & (\frac{\mathrm{M}}{\mathrm{E}})^{-1} \end{pmatrix} \begin{pmatrix} \mathrm{I} & 0 \\ -GE^{-1} & \mathrm{I} \end{pmatrix}$$

$$\begin{pmatrix} \boxed{\mathbf{E}^{-1} + \mathbf{E}^{-1}\mathbf{F}(\mathbf{M}/\mathbf{E})^{-1}\mathbf{G}\mathbf{E}^{-1}} & -\mathbf{E}^{-1}\mathbf{F}(\mathbf{M}/\mathbf{E})^{-1} \\ -(\mathbf{M}/\mathbf{E})^{-1}\mathbf{G}\mathbf{E}^{-1} & (\mathbf{M}/\mathbf{E})^{-1} \end{pmatrix}$$

$$\mathbf{M}/\mathbf{H} \triangleq \mathbf{E} - \mathbf{F}\mathbf{H}^{-1}\mathbf{G}$$

$$\mathbf{M}/\mathbf{E} \triangleq \mathbf{H} - \mathbf{G}\mathbf{E}^{-1}\mathbf{F}$$

$$(\mathbf{M}/\mathbf{H})^{-1} = (\mathbf{E} - \mathbf{F}\mathbf{H}^{-1}\mathbf{G})^{-1} = \mathbf{E}^{-1} + \mathbf{E}^{-1}\mathbf{F}(\mathbf{H} - \mathbf{G}\mathbf{E}^{-1}\mathbf{F})^{-1}\mathbf{G}\mathbf{E}^{-1}$$

**matrix inversion lemma** Sherman-Morrison-Woodbury formula.

Application 1

$$\boldsymbol{\Sigma}^{-1} \longrightarrow (\boldsymbol{\Sigma} + \mathbf{X}\mathbf{X}^\top)^{-1}$$

A typical application in machine learning is the following. Let $\mathbf{X}$ be an $N \times D$ data matrix, and $\boldsymbol{\Sigma}$ be $N \times N$ diagonal matrix. Then we have (using the substitutions $\mathbf{E} = \boldsymbol{\Sigma}$, $\mathbf{F} = \mathbf{G}^\top = \mathbf{X}$, and $\mathbf{H}^{-1} = -\mathbf{I}$) the following result:

$$\underbrace{\boxed{(\boldsymbol{\Sigma} + \mathbf{X}\mathbf{X}^\top)^{-1}}}_{N * N} = \boxed{\boldsymbol{\Sigma}^{-1}} - \boxed{\boldsymbol{\Sigma}^{-1}}\mathbf{X}\underbrace{\boxed{(\mathbf{I} + \mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}}}_{D * D}\mathbf{X}^\top\boxed{\boldsymbol{\Sigma}^{-1}} \tag{7.123}$$

$$(\mathbf{E} - \mathbf{F}\mathbf{H}^{-1}\mathbf{G})^{-1} = \mathbf{E}^{-1} + \mathbf{E}^{-1}\mathbf{F}(\mathbf{H} - \mathbf{G}\mathbf{E}^{-1}\mathbf{F})^{-1}\mathbf{G}\mathbf{E}^{-1}$$

Faster?

The LHS takes $O(N^3)$ time to compute, the RHS takes time $O(D^3)$ to compute.

Note: matrix products that need $O(n^2)$ , computing the inverse in $O(n^3)$.

## Rank one updates

**Definition** Let $A$ be a $K \times K$ matrix and $u$ and $v$ two $K \times 1$ column vectors. Then, the transformation

$$A + uv^\top$$

is called a rank one update to $A$.

The reason why the transformation is called rank one is that the rank of the $K \times K$ matrix $uv^\top$ is equal to $1$ (because a single vector, $u$, spans all the columns of $uv^\top$).

$$uv^T = \begin{pmatrix} uv_1 & uv_2 & \dots & uv_n \end{pmatrix}$$

Application 2

Another application concerns computing a **rank one update** of an inverse matrix. Let $\mathbf{E} = \mathbf{A}$, $\mathbf{F} = u$, $\mathbf{G} = v^\mathsf{T}$, and $H = -1$. Then we have

$$\boxed{(\mathbf{A} + uv^\mathsf{T})^{-1}} = \mathbf{A}^{-1} + \mathbf{A}^{-1}u\boxed{(-1 - v^\mathsf{T}\mathbf{A}^{-1}u)^{-1}}v^\mathsf{T}\mathbf{A}^{-1} \tag{7.124}$$

Scalar!

$$= \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}uv^\mathsf{T}\mathbf{A}^{-1}}{1 + v^\mathsf{T}\mathbf{A}^{-1}u} \tag{7.125}$$

This is known as the **Sherman-Morrison formula.**

It is clear that the inverse of A can be calculated only once. This allows us to work with matrix products that need $O(n^2)$ rather than computing the inverse in $O(n^3)$.

Suppose we have a change in our matrix. We can obtain the following system:

$$Ax = b$$

$$\left(\boldsymbol{A} - \boldsymbol{u}\boldsymbol{v}^T\right)\boldsymbol{x} = \boldsymbol{b}$$

$$\begin{aligned}\boldsymbol{x} &= \left(\boldsymbol{A} - \boldsymbol{u}\boldsymbol{v}^T\right)^{-1}\boldsymbol{b} \\ &= \boxed{\boldsymbol{A}^{-1}\boldsymbol{b}} + \boldsymbol{A}^{-1}\boldsymbol{u}\left(1 - \boldsymbol{v}^T\boldsymbol{A}^{-1}\boldsymbol{u}\right)^{-1}\boldsymbol{v}^T\boxed{\boldsymbol{A}^{-1}\boldsymbol{b}}\end{aligned}$$

Rank one update det?

Application 3

### 7.3.4 Matrix determinant lemma *

an efficient way to compute the determinant of a blockstructured matrix.

$$\mathbf{M/H} \triangleq \mathbf{E} - \mathbf{F}\mathbf{H}^{-1}\mathbf{G}$$
$$\mathbf{M/E} \triangleq \mathbf{H} - \mathbf{G}\mathbf{E}^{-1}\mathbf{F}$$

$$\left|\begin{pmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{pmatrix}\right|$$

$$\left|\begin{pmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{pmatrix}\right| = |\mathbf{E} - \mathbf{F}\mathbf{H}^{-1}\mathbf{G}||\mathbf{H}|$$

$$\underbrace{\begin{pmatrix} \mathbf{I} & -\mathbf{F}\mathbf{H}^{-1} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}}_{\mathbf{X}} \underbrace{\begin{pmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{pmatrix}}_{\mathbf{M}} \underbrace{\begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{H}^{-1}\mathbf{G} & \mathbf{I} \end{pmatrix}}_{\mathbf{Z}} = \underbrace{\begin{pmatrix} \mathbf{E} - \mathbf{F}\mathbf{H}^{-1}\mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{H} \end{pmatrix}}_{\mathbf{W}}$$

$$|\mathbf{M}| = |\mathbf{M/H}||\mathbf{H}|$$

$$|\mathbf{M/H}| = \frac{|\mathbf{M}|}{|\mathbf{H}|}$$

$$|\mathbf{X}||\mathbf{M}||\mathbf{Z}| = |\mathbf{W}| = |\mathbf{E} - \mathbf{F}\mathbf{H}^{-1}\mathbf{G}||\mathbf{H}|$$

$$|\mathbf{M}| = |\mathbf{M/H}||\mathbf{H}| = |\mathbf{M/E}||\mathbf{E}|$$

$$|\mathbf{M/H}| = \frac{|\mathbf{M/E}||\mathbf{E}|}{|\mathbf{H}|}$$

$$\begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -GE^{-1} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{pmatrix} \begin{pmatrix} \mathbf{I} & -E^{-1}F \\ \mathbf{0} & I \end{pmatrix} = \begin{pmatrix} \mathbf{E} & \mathbf{0} \\ \mathbf{0} & \frac{M}{E} \end{pmatrix}$$

$$|\mathbf{E} - \mathbf{F}\mathbf{H}^{-1}\mathbf{G}| = |\mathbf{H} - \mathbf{G}\mathbf{E}^{-1}\mathbf{F}||\mathbf{H}^{-1}||\mathbf{E}|$$

$$|\mathbf{E} - \mathbf{F}\mathbf{H}^{-1}\mathbf{G}| = |\mathbf{H} - \mathbf{G}\mathbf{E}^{-1}\mathbf{F}||\mathbf{H}^{-1}||\mathbf{E}|$$

Hence (setting $\mathbf{E} = \mathbf{A}$, $\mathbf{F} = -u$, $\mathbf{G} = v^{\mathsf{T}}$, $\mathbf{H} = 1$) we have

Rand one update

$$\boxed{|\mathbf{A} + uv^{\mathsf{T}}|} = (1 + v^{\mathsf{T}}\mathbf{A}^{-1}u)|\mathbf{A}|$$

This is known as the **matrix determinant lemma.**

|A| calculated only once!

Application 4

### 7.3.5 Application: deriving the conditionals of an MVN *

Consider a joint Gaussian of the form $p(\mathbf{x}_1, \mathbf{x}_2) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$$

$$p(\mathbf{x}_1|\mathbf{x}_2) = \mathcal{N}(\mathbf{x}_1|\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2),\ \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})$$

Correct?

$$\begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{H}^{-1}\mathbf{G} & \mathbf{I} \end{pmatrix} \begin{pmatrix} (\mathbf{M}/\mathbf{H})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{I} & -\mathbf{F}\mathbf{H}^{-1} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$$

$$p(\boldsymbol{x}_1, \boldsymbol{x}_2) \propto \exp\left\{ -\frac{1}{2} \begin{pmatrix} \boldsymbol{x}_1 - \boldsymbol{\mu}_1 \\ \boldsymbol{x}_2 - \boldsymbol{\mu}_2 \end{pmatrix}^{\mathsf{T}} \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{x}_1 - \boldsymbol{\mu}_1 \\ \boldsymbol{x}_2 - \boldsymbol{\mu}_2 \end{pmatrix} \right\}$$

$$\begin{aligned} p(\boldsymbol{x}_1, \boldsymbol{x}_2) &\propto \exp\left\{ -\frac{1}{2} \begin{pmatrix} \boldsymbol{x}_1 - \boldsymbol{\mu}_1 \\ \boldsymbol{x}_2 - \boldsymbol{\mu}_2 \end{pmatrix}^{\mathsf{T}} \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \mathbf{I} \end{pmatrix} \begin{pmatrix} (\boldsymbol{\Sigma}/\boldsymbol{\Sigma}_{22})^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22}^{-1} \end{pmatrix} \right. \\ &\left. \times \begin{pmatrix} \mathbf{I} & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_1 - \boldsymbol{\mu}_1 \\ \boldsymbol{x}_2 - \boldsymbol{\mu}_2 \end{pmatrix} \right\} \\ &= \exp\left\{ -\frac{1}{2} (\boldsymbol{x}_1 - \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2))^{\mathsf{T}} (\boldsymbol{\Sigma}/\boldsymbol{\Sigma}_{22})^{-1} \right. \\ &\left. (\boldsymbol{x}_1 - \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2)) \right\} \times \exp\left\{ -\frac{1}{2} (\boldsymbol{x}_2 - \boldsymbol{\mu}_2)^{\mathsf{T}} \boldsymbol{\Sigma}_{22}^{-1} (\boldsymbol{x}_2 - \boldsymbol{\mu}_2) \right\} \end{aligned}$$

$$\begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}' \begin{pmatrix} I & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21} & I \end{pmatrix} \begin{pmatrix} (\Sigma/\Sigma_{22})^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{pmatrix} \begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}$$

$$\quad A \qquad\qquad B \qquad\qquad\qquad C \qquad\qquad\qquad D \qquad\qquad E$$

$$A\cdot B = \begin{pmatrix} x_1'-\mu_1' & x_2'-\mu_2' \end{pmatrix} \begin{pmatrix} I & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21} & I \end{pmatrix} = \begin{pmatrix} x_1'-\mu_1'-(x_2'-\mu_2')\Sigma_{22}^{-1}\Sigma_{21} & x_2'-\mu_2' \end{pmatrix}$$

$$A\cdot B\cdot C = \Big( \underbrace{(x_1'-\mu_1'-(x_2'-\mu_2')\Sigma_{22}^{-1}\Sigma_{21})(\Sigma/\Sigma_{22})^{-1}}_{K} \quad (x_2'-\mu_2')\Sigma_{22}^{-1} \Big)$$

$$ABCD = \begin{pmatrix} K & -K\Sigma_{12}\Sigma_{22}^{-1} + (x_2'-\mu_2')\Sigma_{22}^{-1} \end{pmatrix}$$

$$ABCDE = \big( K(x_1-\mu_1) - K\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2) + (x_2'-\mu_2')\Sigma_{22}^{-1}(x_2-\mu_2) \big)$$

$$= \big( K(x_1-\mu_1-\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2)) + (x_2-\mu_2)'\Sigma_{22}^{-1}(x_2-\mu_2) \big)$$

$$ABCDE = ( K(x_1-\mu_1) - K\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2) + (x_2'-\mu_2')\Sigma_{22}^{-1}(x_2-\mu_2) )$$

$$= ( K(x_1-\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2)) + (x_2-\mu_2)'\Sigma_{22}^{-1}(x_2-\mu_2) )$$

$$K = (x_1-\mu_1 - \Sigma_{21}'(\Sigma_{22}^{-1})'(x_2-\mu_2))'(\Sigma/\Sigma_{22})^{-1} \qquad (\Sigma_{22}^{-1})' = (\Sigma_{22}')^{-1}$$

$$= (x_1-\mu_1 - \Sigma_{21}(\Sigma_{22}^{-1})(x_2-\mu_2))'(\Sigma/\Sigma_{22})^{-1} \qquad = \Sigma_{22}^{-1}$$

$$ABCDE = (x_1-\mu_1 - \Sigma_{21}\Sigma_{22}^{-1}(x_2-\mu_2))'(\Sigma/\Sigma_{22})^{-1}(x_1-\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2))$$
$$+ (x_2-\mu_2)'\Sigma_{22}^{-1}(x_2-\mu_2)$$

$$= \exp\left\{-\frac{1}{2}(x_1-\mu_1-\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2))^{\mathsf{T}}(\Sigma/\Sigma_{22})^{-1}\right.$$
$$\left.(x_1-\mu_1-\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2))\right\} \times \exp\left\{-\frac{1}{2}(x_2-\mu_2)^{\mathsf{T}}\Sigma_{22}^{-1}(x_2-\mu_2)\right\}$$

$$
\begin{aligned}
p(\boldsymbol{x}_1, \boldsymbol{x}_2) &= p(\boldsymbol{x}_1|\boldsymbol{x}_2)p(\boldsymbol{x}_2) \\
&= \mathcal{N}(\boldsymbol{x}_1|\boldsymbol{\mu}_{1|2}, \boldsymbol{\Sigma}_{1|2})\mathcal{N}(\boldsymbol{x}_2|\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})
\end{aligned}
$$

$$
= \exp\left\{-\frac{1}{2}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2))^{\mathsf{T}}(\boldsymbol{\Sigma}/\boldsymbol{\Sigma}_{22})^{-1}\right.
$$

$$
\left.(\boldsymbol{x}_1 - \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2))\right\} \times \exp\left\{-\frac{1}{2}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2)^{\mathsf{T}}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2)\right\}
$$

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

$$\Sigma_{1|2} = \Sigma / \Sigma_{22}$$

We can also use the fact that $|\mathbf{M}| = |\mathbf{M}/\mathbf{H}||\mathbf{H}|$ to check the normalization constants are correct:

$$(2\pi)^{(d_1+d_2)/2}|\boldsymbol{\Sigma}|^{\frac{1}{2}} = (2\pi)^{(d_1+d_2)/2}(|\boldsymbol{\Sigma}/\boldsymbol{\Sigma}_{22}|\ |\boldsymbol{\Sigma}_{22}|)^{\frac{1}{2}} \tag{7.146}$$

$$= (2\pi)^{d_1/2}|\boldsymbol{\Sigma}/\boldsymbol{\Sigma}_{22}|^{\frac{1}{2}}\ (2\pi)^{d_2/2}|\boldsymbol{\Sigma}_{22}|^{\frac{1}{2}} \tag{7.147}$$

where $d_1 = \dim(\boldsymbol{x}_1)$ and $d_2 = \dim(\boldsymbol{x}_2)$.

$$p(\boldsymbol{x}_1|\boldsymbol{x}_2) = \mathcal{N}(\boldsymbol{x}_1|\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2),\ \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})$$

Correct!