DS-GA 3001.001 Special Topics in Data Science: Modeling Time Series
For the curious mind: Proof of formula for multivariate gaussian conditioning.

Given a joint multivariate gaussian distribution for $\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}$ with mean $\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ and covariance $\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$ we want to compute the conditional distribution $\mathbf{x}_1|\mathbf{x}_2$.

In principle this can be achieved by brute force: using the definition of the conditional distribution, the functional form of the multivariate gaussian and lots of linear algebra, see e.g. `http://fourier.eng.hmc.edu/e161/lectures/gaussianprocess/node7.html`. Here I am taking for granted the fact that the resulting distribution is also normal and just compute the corresponding moments. This derivation has the advantage that it uses only the fact that the (co)variance distributes for linear combinations (something we have repeatedly used in the lecture) and very simple linear algebra.

The trick is to define an auxiliary variable $\mathbf{y} = \mathbf{x}_1 + \mathbf{A}\mathbf{x}_2$ where $\mathbf{A} = \Sigma_{12}\Sigma_{22}^{-1}$ is set so that $\mathbf{y}$ and $\mathbf{x}_2$ are jointly gaussian and independent. To check that this is true, let's compute the covariance:

$$\begin{aligned}
\mathrm{cov}(\mathbf{y}, \mathbf{x}_2) &= \mathrm{cov}(\mathbf{x}_1 + \mathbf{A}\mathbf{x}_2, \mathbf{x}_2) \\
&= \mathrm{cov}(\mathbf{x}_1, \mathbf{x}_2) + \mathbf{A}\mathrm{cov}(\mathbf{x}_2, \mathbf{x}_2) \\
&= \Sigma_{12} + \mathbf{A}\Sigma_{22} \\
&= \Sigma_{12} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22} \\
&= 0
\end{aligned}$$

Using $\mathbf{z}$ we start deriving the mean and covariance of $\mathbf{x}_1|\mathbf{x}_2$, as:

$$\begin{aligned}
\mathbb{E}[\mathbf{x}_1|\mathbf{x}_2] &= \mathbb{E}[\mathbf{z} - \mathbf{A}\mathbf{x}_2|\mathbf{x}_2] \\
&= \mathbb{E}[\mathbf{z}|\mathbf{x}_2] - \mathbb{E}[\mathbf{A}\mathbf{x}_2|\mathbf{x}_2] \\
&= \mathbb{E}[\mathbf{z}] - \mathbf{A}\mathbf{x}_2 \\
&= \mu_1 + \mathbf{A}\mu_2 - \mathbf{A}\mathbf{x}_2 \\
&= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \mu_2)
\end{aligned}$$

where we have used the fact that $\mathbf{y}$ and $\mathbf{x}_2$ are independent so the conditional disappears in $\mathbf{z}|\mathbf{x}_2$.

Lastly, we need to compute the variance:

$$\begin{aligned}
\mathrm{var}(\mathbf{x}_1|\mathbf{x}_2) &= \mathrm{var}(\mathbf{z} - \mathbf{A}\mathbf{x}_2|\mathbf{x}_2) \\
&= \mathrm{var}(\mathbf{z}) + \mathrm{var}(\mathbf{A}\mathbf{x}_2|\mathbf{x}_2) - \mathbf{A}\mathrm{cov}(\mathbf{z}, -\mathbf{x}_2) - [\mathbf{A}\mathrm{cov}(\mathbf{z}, -\mathbf{x}_2)]^{\mathrm{t}} \\
&= \mathrm{var}(\mathbf{z})
\end{aligned}$$

where we have used the fact that $\mathrm{var}(a + b) = \mathrm{var}(a) + \mathrm{var}(b) - \mathrm{cov}(a, b) - \mathrm{cov}(b, a)$ (easy to check using the definition of variance), $\mathrm{var}(\mathbf{A}\mathbf{x}_2|\mathbf{x}_2) = 0$ and the independence of $\mathbf{y}$ and $\mathbf{x}_2$. Continuing the derivation yields:

$$\begin{aligned}
\mathrm{var}(\mathbf{x}_1|\mathbf{x}_2) &= \mathrm{var}(\mathbf{z}) = \mathrm{var}\left(\mathbf{x}_1 + \mathbf{A}\mathbf{x}_2\right) \\
&= \mathrm{var}(\mathbf{x}_1) + \mathbf{A}\mathrm{var}(\mathbf{x}_2)\mathbf{A}^{\mathrm{t}} - \mathbf{A}\mathrm{cov}(\mathbf{x}_1, \mathbf{x}_2) - \mathrm{cov}(\mathbf{x}_2, \mathbf{x}_1)\mathbf{A}^{\mathrm{t}} \\
&= \Sigma_{11} + \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22}\Sigma_{22}^{-1\mathrm{t}}\Sigma_{12}^{\mathrm{t}} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12} - \Sigma_{21}\Sigma_{22}^{-1\mathrm{t}}\Sigma_{12}^{\mathrm{t}} \\
&= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}
\end{aligned}$$

where we have used $\Sigma_{12}^{\mathrm{t}} = \Sigma_{21}$ and $\Sigma_{22}^{\mathrm{t}} = \Sigma_{22}$.