Conditional Multivariate Gaussian

Gabriel Dernbach

09.05.19

The conditional distribution of a multivariat normal distribution is key to the formulation of Gaussian Processes. In the following we will derive how the conditional can be stated.

The multivariat normal distribution is given by

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

In order to condition the multivariat gaussian onto one of it's components we remember that by

$$p(A|B) = \frac{p(A,B)}{p(B)}$$

the conditional distribution is found by dividing the joint distribution by the respective marignal distribution we which to condition on. Our aim therefore is to find a proper expression for the joint distribution from which the marginal and conditional probabilities immediately follow.

To distinguish between different components of the multivariate gaussian we choose a partitioning as follows

$$\mathbf{x} = \left[\begin{array}{c} \mathbf{x}_1 \\ \mathbf{x}_2 \end{array}\right] \mu = \left[\begin{array}{c} \mu_1 \\ \mu_2 \end{array}\right] \quad \text{and} \quad \boldsymbol{\Sigma} = \left[\begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array}\right]$$

Where $\mathbf{x}$ is of dimension n and $\mathbf{x_1}$,$\mathbf{x_2}$ are of dimension p, q such that $n = p + q$. The marginal distributions of $\mathbf{x_1}$ and $\mathbf{x_2}$ are normal with μ_1, μ_2 and Σ_1, Σ_2 respectively.

By introducing the block notation we hope to find an untangled expression for the joint distribution in that it allows easy marginalization. The joint distribution is

$$\begin{aligned} p(\mathbf{x}) &= p\left(\mathbf{x}_1, \mathbf{x}_2\right) \\ &= \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1}(\mathbf{x}-\mu)\right] \\ &= \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2} Q\left(\mathbf{x}_1, \mathbf{x}_2\right)\right] \end{aligned}$$

where Q is the quadratic form which will take our primary attention from now on. The quadratic form after inserting the partitioning becomes

$$\begin{aligned} Q(\mathbf{x}_1, \mathbf{x}_2) &= (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu) \\ &= \left[(\mathbf{x}_1 - \mu_1)^T, (\mathbf{x} - \mu_2)^T \right] \left[\begin{array}{cc} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{array} \right] \left[\begin{array}{c} \mathbf{x}_1 - \mu_1 \\ \mathbf{x}_2 - \mu_2 \end{array} \right] \\ &= (\mathbf{x}_1 - \mu_1)^T \Sigma^{11} (\mathbf{x}_1 - \mu_1) \\ &\quad + 2 (\mathbf{x}_1 - \mu_1)^T \Sigma^{12} (\mathbf{x}_2 - \mu_2) \\ &\quad + (\mathbf{x}_2 - \mu_2)^T \Sigma^{22} (\mathbf{x}_2 - \mu_2) \end{aligned}$$

where we introduced the following notation for the partitioned inverse

$$\Sigma^{-1} = \left[\begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array} \right]^{-1} = \left[\begin{array}{cc} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{array} \right]$$

As we have assumed a block representation of the inverted block covariance matrix, let us now examine how the original blocks behave under inversion. In this rather general examination of a symmetric block matrix under inversion we will substitute A for the matrix and B for it's inverse in order to facilitate readability.

$$I_n = AA^{-1} = AB = \left[\begin{array}{cc} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{array} \right] \left[\begin{array}{cc} B_{11} & B_{12} \\ B_{12}^T & B_{22} \end{array} \right]$$

$$= \left[\begin{array}{cc} A_{11}B_{11} + A_{12}B_{12}^T & A_{11}B_{12} + A_{12}B_{22} \\ A_{12}^T B_{11} + A_{22}B_{12}^T & A_{12}B_{12} + A_{22}B_{22} \end{array} \right] = \left[\begin{array}{cc} I_p & 0 \\ 0 & I_q \end{array} \right]$$

By comparison of coefficients we can find expressions for the entries $B_{11}, B_{12}, B_{21}, B_{22}$ of our inverse symmetric block matrix, which are

$$\begin{aligned} A_{11}B_{11} + A_{12}B_{12}^T = I_p &\quad\Longrightarrow\quad B_{11} = A_{11}^{-1} - A_{11}^{-1} A_{12} B_{12}^T \\ A_{11}B_{12} + A_{12}B_{22} = 0 &\quad\Longrightarrow\quad B_{12} = -A_{11}^{-1} A_{12} B_{22} \\ A_{12}^T B_{11} + A_{22}B_{12}^T = 0 &\quad\Longrightarrow\quad B_{22} = -A_{22}^{-1} A_{12}^T B_{11} \\ A_{12}^T B_{12} + A_{22}B_{22} = I_q &\quad\Longrightarrow\quad B_{22} = A_{22}^{-1} - A_{22}^{-1} A_{12}^T B_{12} \end{aligned}$$

We continue our pursuit of expressions of B that are purely stated in terms of A. First we can introduce $B_{12}^{\top}$ into the equation of B_{11} to get

$$\begin{aligned}
B_{11} =& A_{11}^{-1} + A_{11}^{-1} A_{12} A_{22}^{-1} A_{12}^{T} B_{11} \\
A_{11}^{-1} =& \left(I - A_{11}^{-1} A_{12} A_{22}^{-1} A_{12}^{T}\right) B_{11} \\
I_p =& \left(A_{11} - A_{12} A_{22}^{-1} A_{12}^{T}\right) B_{11} \\
B_{11} =& \left(A_{11} - A_{12} A_{22}^{-1} A_{12}^{T}\right)^{-1}
\end{aligned}$$

Simillarly we can substitute B_{12} into the equation of B_{22}, as well as using these intermediat results to recover the remaining block matrices B_{12} and $B_{12}^{\top}$

$$\begin{aligned}
B_{12} =& -A_{11}^{-1} A_{12} \left(A_{22} - A_{12}^{T} A_{11}^{-1} A_{12}^{-1}\right) \\
B_{12}^{T} =& -A_{22}^{-1} A_{12}^{T} \left(A_{11} - A_{12} A_{22}^{-1} A_{12}^{T}\right)^{-1} \\
B_{22} =& \left(A_{22} - A_{12}^{T} A_{11}^{-1} A_{12}\right)^{-1}
\end{aligned}$$

After this excursion to the inverse of block matrices we can substitute back to the original notation of Σ to see that

$$\begin{aligned}
\Sigma^{11} &= \left(\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{12}^{T}\right)^{-1} \\
\Sigma^{22} &= \left(\Sigma_{22} - \Sigma_{12}^{T} \Sigma_{11}^{-1} \Sigma_{12}\right)^{-1} \\
\Sigma^{12} &= -\Sigma_{11}^{-1} \Sigma_{12} \left(\Sigma_{22} - \Sigma_{12}^{T} \Sigma_{11}^{-1} \Sigma_{12}\right)^{-1} = \left(\Sigma^{21}\right)^{T}
\end{aligned}$$

As we now have a way to translate a block covariance matrix to it's inverse block matrix we may find an expression for $Q((\mathbf{x_1}, \mathbf{x_2}))$ in terms of the covariance matrix itself. Before substituting our current results into Q however, we can increase flexibility by splitting the product of component Σ^{11} by the Sherman–Morrison–Woodbury formula into a sum. This will later turn out beneficial in cancelling related terms.

The Woodburry matrix identity can be found by comparison of the inverse of two slightly differing decompositions of one matrix. A decomposition that is particularly nice to invert is the LDU decomposition, as the triangular parts will only change their sign [1] and the inverse of the diagonal is found by elementwise inversion. We will therefore consider a block matrix M of the form

$$\begin{bmatrix} A & U \\ V & C \end{bmatrix}$$

The parts of the decomposition can be found by successively multiplying matrices to M such that it's offdiagonal elements are set to zero.

[1] as can directly be shown by an ordinary gauss-jordan elimination.

We start with an identity matrix which we try to alter in a way such that it will set the entry below A to zero. Next we note that the entry at V after the multiplication with the identiy would be the consequence of $0 * A + 1 * V$. It is straight forward to set this equation to zero by finding a coefficient at the 0 position of the identity matrix that cancels V. Such is $-VA^{-1}A + 1V$. Therefore our first elimination results in

$$\begin{bmatrix} I & 0 \\ -VA^{-1} & I \end{bmatrix} \begin{bmatrix} A & U \\ V & C \end{bmatrix} = \begin{bmatrix} A & U \\ 0 & C - VA^{-1}U \end{bmatrix}$$

Similarly the entry above C can be eliminated by

$$\begin{bmatrix} A & U \\ V & C \end{bmatrix} \begin{bmatrix} I & -A^{-1}U \\ 0 & I \end{bmatrix} = \begin{bmatrix} A & 0 \\ V & C - VA^{-1}U \end{bmatrix}$$

Combining the two multiplications already yields the desired diagonal

$$\begin{bmatrix} I & 0 \\ -VA^{-1} & I \end{bmatrix} \begin{bmatrix} A & U \\ V & C \end{bmatrix} \begin{bmatrix} I & -A^{-1}U \\ 0 & I \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & C - VA^{-1}U \end{bmatrix}$$

Moving to the right side yields the LDU decomposition of M

$$\begin{bmatrix} A & U \\ V & C \end{bmatrix} = \begin{bmatrix} I & 0 \\ VA^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & C - VA^{-1}U \end{bmatrix} \begin{bmatrix} I & A^{-1}U \\ 0 & I \end{bmatrix}$$

Inverting both sides we see at the right a decomposition of the total inverse into it's block components.

$$\begin{aligned} \begin{bmatrix} A & U \\ V & C \end{bmatrix}^{-1} &= \begin{bmatrix} I & A^{-1}U \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} A & 0 \\ 0 & C - VA^{-1}U \end{bmatrix}^{-1} \begin{bmatrix} I & 0 \\ VA^{-1} & I \end{bmatrix}^{-1} \\ &= \begin{bmatrix} I & -A^{-1}U \\ 0 & I \end{bmatrix} \begin{bmatrix} A^{-1} & 0 \\ 0 & \left(C - VA^{-1}U\right)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -VA^{-1} & I \end{bmatrix} \\ &= \begin{bmatrix} A^{-1} + A^{-1}U\left(C - VA^{-1}U\right)^{-1} VA^{-1} & -A^{-1}U\left(C - VA^{-1}U\right)^{-1} \\ -\left(C - VA^{-1}U\right)^{-1} VA^{-1} & \left(C - VA^{-1}U\right)^{-1} \end{bmatrix} \end{aligned}$$

Now as a slightly differing approach, we could have equally likely chosen to first set the value above C to zero. The latter approach results in

$$\begin{bmatrix} A & U \\ V & C \end{bmatrix} = \begin{bmatrix} I & UC^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A - UC^{-1}V & 0 \\ 0 & C \end{bmatrix} \begin{bmatrix} I & 0 \\ C^{-1}V & I \end{bmatrix}$$

Inversion of both sides yields

$$\begin{aligned}
\begin{bmatrix} A & U \\ V & C \end{bmatrix}^{-1} &= \begin{bmatrix} I & 0 \\ C^{-1}V & I \end{bmatrix}^{-1} \begin{bmatrix} A - UC^{-1}V & 0 \\ 0 & C \end{bmatrix}^{-1} \begin{bmatrix} I & UC^{-1} \\ 0 & I \end{bmatrix}^{-1} \\
&= \begin{bmatrix} I & 0 \\ -C^{-1}V & I \end{bmatrix} \begin{bmatrix} \left(A - UC^{-1}V\right)^{-1} & 0 \\ 0 & C^{-1} \end{bmatrix} \begin{bmatrix} I & -UC^{-1} \\ 0 & I \end{bmatrix} \\
&= \begin{bmatrix} \left(A - UC^{-1}V\right)^{-1} & -\left(A - UC^{-1}V\right)^{-1} UC^{-1} \\ -C^{-1}V\left(A - UC^{-1}V\right)^{-1} & C^{-1}V\left(A - UC^{-1}V\right)^{-1} UC^{-1} + C^{-1} \end{bmatrix}
\end{aligned}$$

Inspecting the elements of the two approaches we note a discovery of different ways to write an equally valid block of the inverse, that is we found a matrix identity. Most notably is the identitiy given by the respective elements in the first row and column. This identity is commonly referred to as the Woodburry matrix identity, or matrix inversion lemma

$$\left(A - UC^{-1}V\right)^{-1} = A^{-1} + A^{-1}U\left(C - VA^{-1}U\right)^{-1} VA^{-1}$$

A direct proof of validity of our equality can be found by the fact that the following product evaluates to the identity matrix as expected

$$(A + UCV)\left[A^{-1} - A^{-1}U\left(C^{-1} + VA^{-1}U\right)^{-1} VA^{-1}\right]$$

We now can rewrite Σ^{11} as it takes the exact from of the left hand side of the Woodburry Matrix Identity and thereby untangle it's product into a sum.

We summarize the block inverses

$$\begin{aligned}
\Sigma^{11} &= \left(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^{T}\right)^{-1} \\
&= \Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}\left(\Sigma_{22} - A_{12}^{T}\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1}\Sigma_{12}^{T}\Sigma_{11}^{-1} \\
\Sigma^{22} &= \left(\Sigma_{22} - \Sigma_{12}^{T}\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1} \\
&= \Sigma_{22}^{-1} + \Sigma_{22}^{-1}\Sigma_{12}^{T}\left(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^{T}\right)^{-1}\Sigma_{12}\Sigma_{22}^{-1} \\
\Sigma^{12} &= -\Sigma_{11}^{-1}\Sigma_{12}\left(\Sigma_{22} - \Sigma_{12}^{T}\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1} = \left(\Sigma^{21}\right)^{T}
\end{aligned}$$

We get back to the quadratic $Q(\mathbf{x_1}, \mathbf{x_2})$ and substitute the second equation of

Σ^{11}, the first equation of Σ_{22} and Σ_{12}

$$
\begin{aligned}
Q(\mathbf{x}_1, \mathbf{x}_2) = & (\mathbf{x}_1 - \mu_1)^T \Sigma^{11} (\mathbf{x}_1 - \mu_1) \\
& + 2 (\mathbf{x}_1 - \mu_1)^T \Sigma^{12} (\mathbf{x}_2 - \mu_2) \\
& + (\mathbf{x}_2 - \mu_2)^T \Sigma^{22} (\mathbf{x}_2 - \mu_2)
\end{aligned}
$$

$$
\begin{aligned}
Q(\mathbf{x}_1, \mathbf{x}_2) = & (\mathbf{x}_1 - \mu_1)^T \left[\Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}\left(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1}\Sigma_{12}^T\Sigma_{11}^{-1}\right](\mathbf{x}_1 - \mu_1) \\
& - 2 (\mathbf{x}_1 - \mu_1)^T \left[\Sigma_{11}^{-1}\Sigma_{12}\left(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1}\right](\mathbf{x}_2 - \mu_2) \\
& + (\mathbf{x}_2 - \mu_2)^T \left[\left(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1}\right](\mathbf{x}_2 - \mu_2)
\end{aligned}
$$

And expand the product in the first term

$$
\begin{aligned}
Q(\mathbf{x}_1, \mathbf{x}_2) = & (\mathbf{x}_1 - \mu_1)^T \Sigma_{11}^{-1} (\mathbf{x}_1 - \mu_1) \\
& + (\mathbf{x}_1 - \mu_1)^T \left[\Sigma_{11}^{-1}\Sigma_{12}\left(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1}\Sigma_{12}^T\Sigma_{11}^{-1}\right](\mathbf{x}_1 - \mu_1) \\
& - 2 (\mathbf{x}_1 - \mu_1)^T \left[\Sigma_{11}^{-1}\Sigma_{12}\left(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1}\right](\mathbf{x}_2 - \mu_2) \\
& + (\mathbf{x}_2 - \mu_2)^T \left[\left(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1}\right](\mathbf{x}_2 - \mu_2)
\end{aligned}
$$

The second part resembles a quadratic form already that we only need to factorize. To see more clearly we define it's factors to be

$$
\begin{aligned}
u &\coloneqq (\mathbf{x}_1 - \mu_1)^T \Sigma_{11}^{-1}\Sigma_{12} \\
v &\coloneqq (\mathrm{x}_2 - \mu_2) \\
A &\coloneqq \Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12})^{-1}
\end{aligned}
$$

Which allows the standart factorization as

$$
\begin{aligned}
& u^T Au - 2u^T Av + v^T Av \\
=& u^T Au - u^T Av - u^T Av + v^T Av \\
=& u^T A(u - v) - (u - v)^T Av \\
=& u^T A(u - v) - v^T A(u - v) \\
=& (u - v)^T A(u - v) \\
=& (v - u)^T A(v - u)
\end{aligned}
$$

Introducing the factorization in $Q(\mathbf{x_1}, \mathbf{x_1})$ yields

$$
\begin{aligned}
Q(\mathbf{x_1}, \mathbf{x_2}) = & \left(\mathbf{x}_1 - \mu_1\right)^T \Sigma_{11}^{-1} \left(\mathbf{x}_1 - \mu_1\right) \\
& + \left[\left(\mathbf{x}_2 - \mu_2\right) - \Sigma_{12}^T \Sigma_{11}^{-1} \left(\mathbf{x}_1 - \mu_1\right)\right]^T \\
& \quad \left(\Sigma_{22} - \Sigma_{12}^T \Sigma_{11}^{-1} \Sigma_{12}\right)^{-1} \left[\left(\mathbf{x}_2 - \mu_2\right) - \Sigma_{12}^T \Sigma_{11}^{-1} \left(\mathbf{x}_1 - \mu_1\right)\right]
\end{aligned}
$$

As it turns out, the quadratic is itself the sum of two quadratics again. Taking the sum out of the exponential, we can already anticipate the result of a product of gaussians. Whereas the first one has straight forward expected value of μ_1 and covariance Σ_{11}^{-1}, we introduce a short hand for those parameters in the second as

$$
b \coloneqq \mu_2 + \Sigma_{12}^T \Sigma_{11}^{-1} \left(\mathbf{x}_1 - \mu_1\right) \qquad A \coloneqq \Sigma_{22} - \Sigma_{12}^T \Sigma_{11}^{-1} \Sigma_{12}
$$

At this point the joint distribution has become

$$
\begin{aligned}
p(\mathbf{x}) = p\left(\mathbf{x}_1, \mathbf{x}_2\right) = & \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2} Q\left(\mathbf{x}_1, \mathbf{x}_2\right)\right] \\
= & \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}\left(\mathbf{x}_1 - \mu_1\right)^T \Sigma_{11}^{-1} \left(\mathbf{x}_1 - \mu_1\right)\right] \exp\left[-\frac{1}{2}\left(\mathbf{x}_2 - b\right)^T A^{-1} \left(\mathbf{x}_2 - b\right)\right]
\end{aligned}
$$

Lastly we need to split the normalization into the respective blocks. First we note that the determinant of a product is the product of the determinants.

$$
|AB| = |A||B|
$$

Together with the fact that

$$
\left| \begin{array}{cc} B & 0 \\ C & D \end{array} \right| = \left| \begin{array}{cc} B & C \\ 0 & D \end{array} \right| = |B||D|
$$

We again turn to a triangular decomposition, this time however without the diagonal, which for a 2 by 2 matrix is

$$
A = \left[\begin{array}{cc} A_{11} & A_{12} \\ A_{21} & A_{22} \end{array} \right] = \left[\begin{array}{cc} A_{11} & 0 \\ A_{12}^T & I \end{array} \right] \left[\begin{array}{cc} I & A_{11}^{-1} A_{12} \\ 0 & A_{22} - A_{12}^T A_{11}^{-1} A_{12} \end{array} \right]
$$

Combinig the three observations gives the decomposition we need

$$
|A| = \left| \left[\begin{array}{cc} A_{11} & A_{12} \\ A_{21} & A_{22} \end{array} \right] \right| = \left|A_{11}\right| \left|A_{22} - A_{12}^T A_{11}^{-1} A_{12}\right|
$$

We therefore separate the determinant of the covariance matrix by

$$
|\Sigma| = \left| \left[\begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array} \right] \right| = |\Sigma_{22}||\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T| = |\Sigma_{11}||\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}|
$$

And see that the joint probability is indeed the product of two gaussians

$$
\begin{aligned}
p(\mathbf{x}_1, \mathbf{x}_2) = & \frac{1}{(2\pi)^{p/2} |\Sigma_{11}|^{1/2}} \exp\left[-\frac{1}{2}(\mathrm{x}_1 - \mu_1)^T \Sigma_{11}^{-1} (\mathrm{x}_1 - \mu_1)\right] \\
& \frac{1}{(2\pi)^{q/2} |A|^{1/2}} \exp\left[-\frac{1}{2}(\mathrm{x}_2 - b)^T A^{-1} (\mathrm{x}_2 - b)\right] \\
= & \mathcal{N}(\mathbf{x}_1, \mu_1, \Sigma_{11}) \mathcal{N}(\mathbf{x}_2, b, A)
\end{aligned}
$$

The marginal distribution over x_1 is

$$
p_1(\mathbf{x}_1) = \int p(\mathbf{x}_1, \mathbf{x}_2)\, d\mathbf{x}_2 = \frac{1}{(2\pi)^{p/2} |\Sigma_{11}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}_1 - \mu_1)^T \Sigma_{11}^{-1} (\mathbf{x}_1 - \mu_1)\right]
$$

As $p(x_2)$ is normalized and therefore integrates to 1. Altough we find a dependence on x_1 in the second gaussian, it is only in the mean vector b that shifts the gaussian but lets its integral stay unchanged of course.

Finally by dividing the joint by the marginal we arrive at the conditional distribution

$$
p_{2|1}(\mathbf{x}_2|\mathbf{x}_1) = \frac{p(\mathbf{x}_1, \mathbf{x}_2)}{p(\mathbf{x}_1)} = \frac{1}{(2\pi)^{q/2} |A|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}_2 - b)^T A^{-1} (\mathbf{x}_2 - b)\right]
$$

with

$$
\begin{aligned}
b &= \mu_2 + \Sigma_{12}^T \Sigma_{11}^{-1} (\mathbf{x}_1 - \mu_1) \\
A &= \Sigma_{22} - \Sigma_{12}^T \Sigma_{11}^{-1} \Sigma_{12}
\end{aligned}
$$

For a gaussian process regression we assume $\mu_1 = \mu_2 = 0$. An example of such a regression is given in matlab and python. We denote $x1, y1$ as the training data and $x2, y2$ as the predictions.

```matlab
% specify distribution
x1 = [-1, -0.9, -0.4, -0.25, 0.5, 1];
y1 = [0, -1, 2, 0.5, 2, 0];
x2 = linspace(-1,1,1000);

% estimate posterior
[y2,sigma2] = gaussian_process(x1,y1,x2,width,lambda);


function [y2,sigma2] = gaussian_process(x1,y1,x2,width,lambda)
    k11 = kernel(x1,x1,width,lambda); % auto covariance of x1
    k12 = kernel(x2,x1,width,lambda); % cross covariance x1 to x2
    k22 = kernel(x2,x2,width,lambda); % auto covariance of x2

    y2 = k12' / k11 * y1'; % posterior mean prediction
    sigma2 = k22 - k12' / k11 * k12; % posterior variance prediction
end

function kab = kernel(a,b,width,lambda) % gaussain kernel
    kab = exp(-(a-b').^2 ./ (2*width^2)); % exponential pairwise
    kab = kab + eye(size(kab)) .* lambda; % regularization
end
```

Listing 1: matlab implementation of a gaussian process

```matlab
figure()
plot(x1,y1,'o','MarkerSize',7)
hold('on')
plot(x2,y2,'Color',[0, 0.4470, 0.7410],'Linewidth',3)
std2 = 1.96.*sqrt(diag(sigma2)); % extract 95 percent quantile
plot(x2,y2+std2,'Color',[0.8500, 0.3250, 0.0980],'Linewidth',0.05)
plot(x2,y2-std2,'color',[0.8500, 0.3250, 0.0980],'Linewidth',0.05)
title('mean prediction and 95 percentile')
legend('prior','mean prediction',' 95 percentile')

figure()
for i = 1:8
    plot(x2,mvnrnd(y2,sigma2)); hold('on')
    title('8 realizations of the distribution')
end
```

Listing 2: code to create the plots in fig. 1 and 2

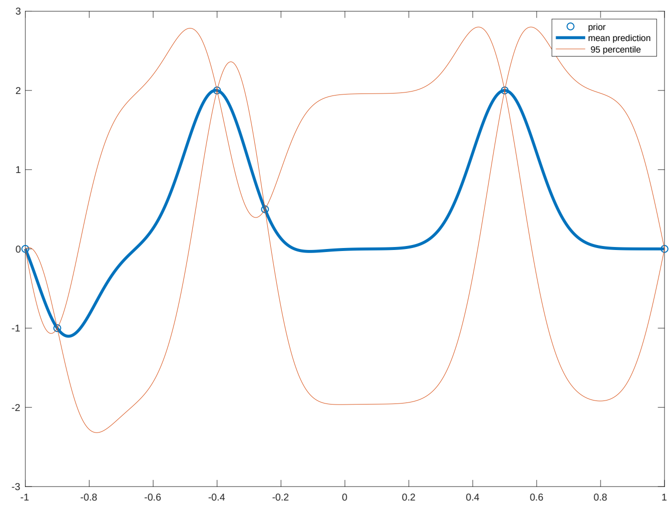

Figure 1: A gaussian process that fulfills the specified points and has increased variance where the process is less determined.

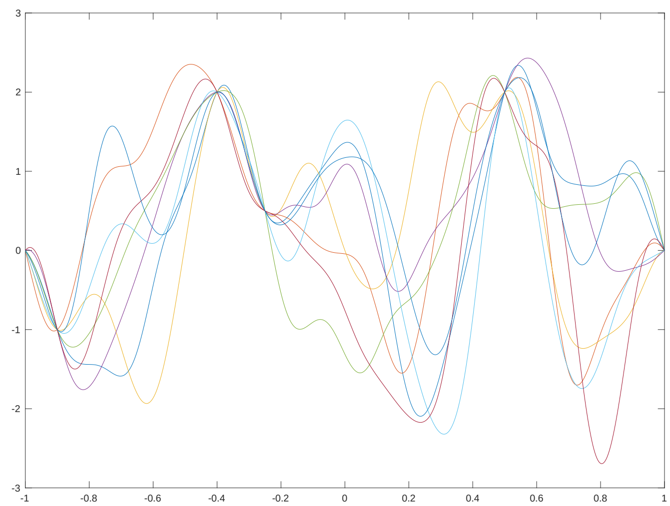

Figure 2: The 8 samples drawn from the process specified by the posterior distribution show strong deviations in parts where the variance is high while taking equal values for the points that we specified.

```python
import numpy as np

x1 = np.array([-1, -0.9, -0.4, -0.25, 0.5, 1]);
y1 = np.array([0, -1, 2, 0.5, 2, 0]);

x2 = np.linspace(-1,1,1000);

sigma = 0.3
llambda = 0

def kernel(x,y,sigma,llambda):
    k = np.exp(-(x[:,None] - y[None,:])**2 / (2*sigma**2))
    k = k + np.eye(k.shape[0],k.shape[1]) * llambda
    return k

def gaussian_process(x2,x1,y1,sigma,llambda):
    k11 = kernel(x1,x1,sigma,llambda)
    k12 = kernel(x1,x2,sigma,llambda)
    k22 = kernel(x2,x2,sigma,llambda)

    k11i = np.linalg.inv(k11)
    y2 = k12.T @ k11i @ y1.T
    sig2 = k22 - k12.T @ k11i @ k12
    return y2, sig2

y2,sig2 = gaussian_process(x2,x1,y1,sigma,llambda)
instances8 = np.random.multivariate_normal(y2,sig2,8)
```

Listing 3: python implementation of a gaussian process

```python
import matplotlib.pyplot as plt
plt.subplot(211)
plt.plot(x1,y1,'o',mfc='None',c='C0',label='data')
plt.plot(x2,y2,'-',c='C0',label='mean prediction')
std2 = 1.96 * np.sqrt(np.diag(sig2))
plt.plot(x2,y2-std2,
  '-.',c='C1',linewidth=0.5,label='95 percentile')
plt.plot(x2,y2+std2,'-.',c='C1',linewidth=0.5)
plt.legend(); plt.title('mean prediction given data')

plt.subplot(212)
plt.plot(instances8.T)
plt.title('random instances drawn from the posterior distribution')

plt.tight_layout()
plt.show()
```

Listing 4: python plotting results

References

[1] Christopher KI Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT Press Cambridge, MA, 2006.

[2] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

[3] Wang Ruye Drakos Nikos, Moore Ross. gaussian process.