TTK4250 Sensor Fusion
Solution to Assignment 2

Task 1: *Transformation of Gaussian random variables*

Let $x \in \mathbb{R}^n$ be $\mathcal{N}(\mu, \Sigma)$. Find the distribution and see if you recognize it:

Hint: they are all given in the book.

(a) $z = \Sigma^{-\frac{1}{2}}(x - \mu)$, where $\Sigma^{\frac{1}{2}}(\Sigma^{\frac{1}{2}})^T = \Sigma$

Hint: If you are using theorem 2.4.1, you might need $\det(A^{\frac{1}{2}}) = \det(A)^{\frac{1}{2}}$, $(A^{-1})^T = (A^T)^{-1}$, and $\det(A^T) = \det(A)$ whenever A has full rank.

Solution: Instead of following the hint we use the linearity of the Gaussian distribution (theorem 3.2.2). First we note that we can rewrite the PDF of x as $\mathcal{N}(x, \mu, \Sigma) = \mathcal{N}(x - \mu; 0, \Sigma)$. We then use linearity of Gaussians to get

$$p(z) = \mathcal{N}(z; 0, \Sigma^{-\frac{1}{2}}\Sigma(\Sigma^{-\frac{1}{2}})^T) = \mathcal{N}(z; 0, I_n)$$

Which is the "standard normal" distribution.

Something to note here is that if $AA^T = I_n$, then $\Sigma^{\frac{1}{2}}A$ also satisfies this, which shows that the matix square root is not unique.

(b) Use transformation of random variables to find $y_i = z_i^2$, where z_i is the i'th variable in the vector z.

Solution: Example 2.11: There are two solutions for the inverse mapping, $z_i = \pm\sqrt{y_i}$. The absolute value of the determinant of the Jacobians of the inverse mappings are $\frac{1}{2\sqrt{y_i}}$ (here simply the derivatives, since z_i is scalar). Using theorem 2.5.1 then gives

$$p(y_i) = 2 \cdot \left(\frac{1}{2\sqrt{y_i}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y_i}{2}}\right) = \frac{1}{\sqrt{2\pi y_i}} e^{-\frac{y_i}{2}} = \chi_1^2(y_i)$$

(c) $y = (x - \mu)^T \Sigma^{-1}(x - \mu) = z^T z = \sum z_i^2 = \sum y_i$.

Hint: The MGF of y_i is given in the book through example 2.8 and 2.10. Example 2.6 might also be handy.

Solution: The MGF of y_i is the MGF of a gamma distribution with shape parameter $\frac{1}{2}$ and scale parameter 2 which from example 2.9 is found to be

$$MGF_{y_i}(t) = \left(\frac{1}{1 - 2s}\right)^{\frac{1}{2}}.$$

Taking the sum corresponds to multiplying the MGFs, which in this case becomes exponentiation since the summands are iid. Hence the MGF of y is

$$MGF_y(t) = \left(\frac{1}{1 - 2s}\right)^{\frac{n}{2}},$$

a gamma distribution with scale 2 and shape $\frac{n}{2}$, from example 2.6 known to be chi squared with n degrees of freedom, $\chi_n^2(y)$.

Task 2: *Sensor fusion*

In this task we want to find out if a boat is above the line $x_2 = x_1 + 5$. In order to do this we will fuse measurements from two sensors with our prior belief: A drone-mounted camera, and a maritime surveillance radar. You have some prior knowledge of the state of the boat. You get 1 measurement from each sensor that are processed so that you know them to be (approximately) Gaussian conditioned on the position.

To be more specific, let us denote the state by x and our prior Gaussian by $\mathcal{N}(x; \bar{x}, P)$. The measurement from the camera is given by $z^c = H^c x + v^c$ and the measurement from the radar by $z^r = H^r x + v^r$, where v^c, v^r denotes the measurement noise and is distributed according to $\mathcal{N}(0, R^c)$ and $\mathcal{N}(0, R^r)$, respectively.

Only insert the numbers when asked to. The needed values are given by

$$\bar{x} = \begin{bmatrix} 0 & 0 \end{bmatrix}^T, \qquad P = 25 I_2, \qquad H^c = H^r = I_2,$$

$$R^c = \begin{bmatrix} 79 & 36 \\ 36 & 36 \end{bmatrix}, \qquad R^r = \begin{bmatrix} 28 & 4 \\ 4 & 22 \end{bmatrix}, \qquad z^c = \begin{bmatrix} 2 & 14 \end{bmatrix}^T, \qquad z^r = \begin{bmatrix} -4 & 6 \end{bmatrix}^T$$

(a) What is $p(z^c|x)$?

Solution: Since x is given we can write $v^c = z^c - H^c x$, wich tells us that we can use the PDF for v^c

$$p(z^c - H^c x) = \mathcal{N}(z^c - H^c x; 0, R) = \mathcal{N}(z^c; H^c x, R^c). \tag{1}$$

(b) Show that the joint $p(x, z^c)$ can be written as a Gaussian distribution.

Hint: Use conditional probability and the proof of theorem 3.3.1.

Solution: From chapter 3 we know that we can work with quadratic forms when we are working with Gaussians. To get the quadratic form we scale the logarithm of the distribution by -2, and put all terms that are constants in terms of the random variables into a constant C;

$$\begin{aligned}
-2\ln(p(x, z^c)) &= -2\ln(p(z^c|x)p(x)) = -2\ln(p(z^c|x)) - 2\ln(p(x)) \\
&= (z^c - H^c x)^T R^{-1} (z^c - H^c x) + (x - \bar{x})^T P^{-1} (x - \bar{x}) + C \\
&= \begin{bmatrix} z^c - H^c x \\ x - \bar{x} \end{bmatrix}^T \begin{bmatrix} R & 0 \\ 0 & P \end{bmatrix}^{-1} \begin{bmatrix} z^c - H^c x \\ x - \bar{x} \end{bmatrix} + C \\
&= \begin{bmatrix} z^c - H^c \bar{x} \\ x - \bar{x} \end{bmatrix}^T \begin{bmatrix} I & -H^c \\ 0 & I \end{bmatrix}^T \begin{bmatrix} R & 0 \\ 0 & P \end{bmatrix}^{-1} \begin{bmatrix} I & -H^c \\ 0 & I \end{bmatrix} \begin{bmatrix} z^c - H^c \bar{x} \\ x - \bar{x} \end{bmatrix} + C \\
&= \begin{bmatrix} z^c - H^c \bar{x} \\ x - \bar{x} \end{bmatrix}^T \left(\begin{bmatrix} I & H^c \\ 0 & I \end{bmatrix} \begin{bmatrix} R & 0 \\ 0 & P \end{bmatrix} \begin{bmatrix} I & 0 \\ H^c & I \end{bmatrix} \right)^{-1} \begin{bmatrix} z^c - H^c \bar{x} \\ x - \bar{x} \end{bmatrix} + C \\
&= \begin{bmatrix} z^c - H^c \bar{x} \\ x - \bar{x} \end{bmatrix}^T \begin{bmatrix} H^c P (H^c)^T + R & H^c P \\ P (H^c)^T & P \end{bmatrix}^{-1} \begin{bmatrix} z^c - H^c \bar{x} \\ x - \bar{x} \end{bmatrix} + C.
\end{aligned}$$

If we multiply the last term by $-\frac{1}{2}$ and exponentiate it, we see that it has a Gaussian form. Since it is a distribution by construction, the constant has to be the normalizing constant of that of a Gaussian, and the joint Gaussian result follows.

(c) Find the marginal $p(z^c)$ and the conditional $p(x|z^c)$, using the above and either theorems from the book or calculations.

Solution: We can use theorem 3.2.3 and what was found above to find these distributions. Comparing terms in the theorem and above, we get that the marignal for z^c is

$$p(z^c) = \mathcal{N}(z^c; H^c\bar{x}, H^cP(H^c)^T + R^c). \tag{2}$$

Using $W^c = P(H^c)^T(H^cP(H^c)^T + R)^{-1}$, the distribution for x conditioned on z^c is similarly found to be

$$p(x|z^c) = \mathcal{N}(x; \bar{x} + W^c(z^c - H^c\bar{x}), P - W^cH^cP). \tag{3}$$

(d) Given what you found above, what is the marginal $p(z^r)$ and the conditional $p(x|z^r)$?

Solution: The equations are the same so we can simply replace the terms.

$$p(z^r) = \mathcal{N}(z^r; H^r\bar{x}, H^rP(H^r)^T + R^c). \tag{4}$$

Using $W^r = P(H^r)^T(H^rP(H^r)^T + R^r)^{-1}$, the distribution for x conditioned on z^r is similarly found to be

$$p(x|z^r) = \mathcal{N}(x; \bar{x} + W^r(z^r - H^r\bar{x}), P - W^rH^rP). \tag{5}$$

(e) What is the MMSE and MAP estimate of x given z^c? You do not need to do calculations to find the answer, but briefly state what you would do if you had to.

Solution: MMSE can be found as the mean of the posterior, whereas MAP is found at the point where the posterior attains its maximum. Calcultating the mean involves taking the expectation $E[x]$, while the maximum is found through optimization. However, we have already shown the mean of the posterior to be $\bar{x} + W^c(z^c - H^c\bar{x})$. The maximum of a Gaussian also happens to fall on its mean, so we have implicitly calculated that as well.

(f) Finish the `sensor_model.LinearSensorModel2d.get_pred_meas` method that can be used to calculate marginal probabilities $p(z)$.

Solution:

```python
def get_pred_meas(self, state_est: MultiVarGauss2d) -> MultiVarGauss2d:
    """Get the predicted measurement gaussian given a state gaussian.
    That is N(z; z_bar; S) in Theorem 3.3.1 in the book.
    """
    pred_mean = self.H @ state_est.mean
    pred_cov = self.R + self.H @ state_est.cov @ self.H.T

    pred_meas = MultiVarGauss2d(pred_mean, pred_cov)

    return pred_meas
```

(g) Finish the `conditioning.get_cond_state` function that can be used to calculate conditional probabilities $p(x|z)$.

Solution:

```python
def get_cond_state(state: MultiVarGauss2d,
                   sens_modl: LinearSensorModel2d,
                   meas: Measurement2d
                   ) -> MultiVarGauss2d:
    """Get the conditional state estimate
    given previous state estimate, a sensor model and a measurement.
    That is N(x; x_hat; P_hat) in Theorem 3.3.1 in the book.
    """
    pred_meas = sens_modl.get_pred_meas(state)
    kalman_gain = state.cov @ sens_modl.H.T @ np.linalg.inv(pred_meas.cov)
    innovation = meas.value - pred_meas.mean
    cond_mean = state.mean + kalman_gain @ (innovation)
    cond_cov = state.cov - kalman_gain @ sens_modl.H @ state.cov

    cond_state = MultiVarGauss2d(cond_mean, cond_cov)

    return cond_state
```

(h) Finish the `task2.get_conds` function that is used to calculate the conditional probabilities $p(x|z^c)$ and $p(x|z^r)$.

Solution:

```python
def get_conds(state: MultiVarGauss2d,
              sens_model_c: LinearSensorModel2d, meas_c: Measurement2d,
              sens_model_r: LinearSensorModel2d, meas_r: Measurement2d
              ) -> Tuple[MultiVarGauss2d, MultiVarGauss2d]:
    """Get the conditionals cond_c=p(x|z_c) and cond_r=p(x|z_r)"""
    cond_c = get_cond_state(state, sens_model_c, meas_c)
    cond_r = get_cond_state(state, sens_model_r, meas_r)
    return cond_c, cond_r
```

(i) Finish the `task2.get_double_conds` function that is used to calculate the conditional probabilities $p(x|z^c, z^r)$, i.e. the posterior of x conditioned on z^c then z^r, and $p(x|z^r, z^c)$, i.e. the posterior of x conditioned on z^r then z^c. Does it matter which order we condition?

Solution:

```
def get_double_conds(state: MultiVarGauss2d,
                     sens_model_c: LinearSensorModel2d, meas_c: Measurement2d,
                     sens_model_r: LinearSensorModel2d, meas_r: Measurement2d
                     ) -> Tuple[MultiVarGauss2d, MultiVarGauss2d]:
    """Get the conditionals cond_cr=p(x|z_c, z_r) and cond_rc=p(x|z_r, z_c)"""
    cond_c, cond_r = get_conds(state,
                               sens_model_c, meas_c,
                               sens_model_r, meas_r)
    cond_cr = get_cond_state(cond_c, sens_model_r, meas_r)
    cond_rc = get_cond_state(cond_r, sens_model_c, meas_c)
    return cond_cr, cond_rc
```

(j) Finish the `gaussian.MultiVarGauss2d.get_transformed` method that is used to calculate the probability $p(Tx)$, where T is a linear transformation.

Solution:

```
def get_transformed(self, lin_transform: np.ndarray) -> 'MultiVarGauss2d':
    """Get the transformed distribution from a given linear tranform.
    See 3.2.2 in the book."""
    transformed_mean = lin_transform @ self.mean
    transformed_cov = lin_transform @ self.cov @ lin_transform.T
    transformed = MultiVarGauss2d(transformed_mean, transformed_cov)
    return transformed
```

(k) You now want to know the probability that the boat is above the line, $x_2 = x_1 + 5$.
Finish `task2.get_prob_over_line` using the appropriate linear transform and the CDF.

Hint: This is the same as finding $\Pr(x_2 - x_1 > 5) = \Pr(\begin{bmatrix}-1 & 1\end{bmatrix} x > 5)$

To get the cdf you can use `from scipy.stats import norm`, and then use `norm.cdf(value, mean, std)`. Note that it takes the standard deviation (std) and not the variance as input.

Solution: Se python script for more details.

The value that one should get is $\Pr(x_2 - x_1 > 5) = \Pr(x_2 > x_1 + 5) = 0.87103$.

The linear transform to get $\xi = x_2 - x_1$ is $w^T x = \begin{bmatrix}-1 & 1\end{bmatrix} x$, which gives the scalar Gaussian $\mathcal{N}(\xi; w^T\hat{x}, w^T\hat{P}w)$. w can be seen as a normal vector to the line. This translates our problem into $\Pr(x_2 - x_1 > 5) = \Pr(\xi > 5) = 1 - \Pr(\xi \leq 5) = 1 - P_\xi(5)$.

```python
def get_prob_over_line(gauss: MultiVarGauss2d) → float:
    """Get the probability that a sample from the gaussian is above the line
    x2 = x1 + 5"""
    mat = np.array([[-1, 1]])
    transformed = gauss.get_transformed(mat)
    loc = transformed.mean
    scale = np.sqrt(transformed.cov)
    prob = 1-norm.cdf(5, loc, scale)
    return prob
```

Task 3: *Working with the canonical form*

In Section 3.3 the fundamental product identity was studied using a moment-based parametrization. Clearly, it must also be possible to establish an equivalent result using the canonical representation. In this exercise we shall therefore consider the product

$$\mathcal{N}^{-1}(\mathbf{x}\,;\,\mathbf{a},\mathbf{B})\mathcal{N}^{-1}(\mathbf{y}\,;\,\mathbf{C}\mathbf{x},\mathbf{D}). \tag{6}$$

(a) Show that (6) is identical to the Gaussian

$$\mathcal{N}^{-1}\left(\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix};\begin{bmatrix}\mathbf{a}\\ \mathbf{0}\end{bmatrix},\begin{bmatrix}\mathbf{B}+\mathbf{C}^\mathsf{T}\mathbf{D}^{-1}\mathbf{C} & -\mathbf{C}^\mathsf{T}\\ -\mathbf{C} & \mathbf{D}\end{bmatrix}\right) \tag{7}$$

Hint:

Taking the logarithm of the form (3.17) in the book with (3.20) inserted give a relatively simple way to the goal, after the terms constant in $\mathbf{x}$ and $\mathbf{y}$ are subtracted. Also

$$\mathbf{a}^\mathsf{T}\mathbf{A}\mathbf{a}+\mathbf{b}^\mathsf{T}\mathbf{B}\mathbf{b}+2\mathbf{a}^\mathsf{T}\mathbf{C}\mathbf{b}=\mathbf{a}^\mathsf{T}\mathbf{A}\mathbf{a}+\mathbf{b}^\mathsf{T}\mathbf{B}\mathbf{b}+\mathbf{a}^\mathsf{T}\mathbf{C}\mathbf{b}+\mathbf{b}^\mathsf{T}\mathbf{C}^\mathsf{T}\mathbf{a}=\begin{bmatrix}\mathbf{a}\\ \mathbf{b}\end{bmatrix}^\mathsf{T}\begin{bmatrix}\mathbf{A} & \mathbf{C}\\ \mathbf{C}^\mathsf{T} & \mathbf{B}\end{bmatrix}\begin{bmatrix}\mathbf{a}\\ \mathbf{b}\end{bmatrix},$$

is handy, and valid for any vectors $\mathbf{a}$ and $\mathbf{b}$ and matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ of interest (variable names are not related to the task).

Solution: We follow the hint with the constant $A = -((\dim(\mathbf{x})+\dim(\mathbf{y}))\log(2\pi)-\log(|\mathbf{B}|)-\log(|\mathbf{D}|)+\mathbf{a}^\mathsf{T}\mathbf{B}^{-1}\mathbf{a})$. We thus have

$$\begin{aligned}&2\log\left(\mathcal{N}^{-1}(\mathbf{x}\,;\,\mathbf{a},\mathbf{B})\mathcal{N}^{-1}(\mathbf{y}\,;\,\mathbf{C}\mathbf{x},\mathbf{D})\right)-A\\ &=2\mathbf{a}^\mathsf{T}\mathbf{x}-\mathbf{x}^\mathsf{T}\mathbf{B}\mathbf{x}+2\mathbf{x}^\mathsf{T}\mathbf{C}^\mathsf{T}\mathbf{y}-\mathbf{y}^\mathsf{T}\mathbf{D}\mathbf{y}-\mathbf{x}^\mathsf{T}\mathbf{C}^\mathsf{T}\mathbf{D}^{-1}\mathbf{C}\mathbf{x}\\ &=2\mathbf{a}^\mathsf{T}\mathbf{x}-\mathbf{x}^\mathsf{T}(\mathbf{B}+\mathbf{C}^\mathsf{T}\mathbf{D}^{-1}\mathbf{C})\mathbf{x}+2\mathbf{x}^\mathsf{T}\mathbf{C}^\mathsf{T}\mathbf{y}-\mathbf{y}^\mathsf{T}\mathbf{D}\mathbf{y}.\end{aligned}$$

We now finish the quadratic form as in the hint (note the need for a sign change on the cross term)

$$=2\begin{bmatrix}\mathbf{a}\\ \mathbf{0}\end{bmatrix}^\mathsf{T}\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix}-\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix}^\mathsf{T}\begin{bmatrix}\mathbf{B}+\mathbf{C}^\mathsf{T}\mathbf{D}^{-1}\mathbf{C} & -\mathbf{C}^\mathsf{T}\\ -\mathbf{C} & \mathbf{D}\end{bmatrix}\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix}.$$

This mathces the pattern in (3.17) in the book to be the canonical form in (7).

(b) Show that the marginal distribution of $\mathbf{y}$, from the joint density (7), is

$$\mathcal{N}^{-1}\left(\mathbf{y}\,;\,\mathbf{C}^\mathsf{T}(\mathbf{B}+\mathbf{C}^\mathsf{T}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{a}, \mathbf{D}-\mathbf{C}(\mathbf{B}+\mathbf{C}^\mathsf{T}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{C}^\mathsf{T}\right). \tag{8}$$

Hint: Theorem 3.4.1

Solution: Comparing terms with 3.4.1 we are interested in $\boldsymbol{\eta}_*$ and $\boldsymbol{\Lambda}_*$. For this we associate $\boldsymbol{\eta}_b \to \mathbf{0}$, $\boldsymbol{\eta}_a \to \mathbf{a}$, $\boldsymbol{\Lambda}_{xx} \to \mathbf{B}+\mathbf{C}^\mathsf{T}\mathbf{D}^{-1}\mathbf{C}$, $\boldsymbol{\Lambda}_{yy} \to \mathbf{D}$ and $\boldsymbol{\Lambda}_{xy} \to -\mathbf{C}^\mathsf{T}$. Inserting yields the marginal information state $\mathbf{C}(\mathbf{B}+\mathbf{C}^\mathsf{T}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{a}$ and marginal information matrix $D-\mathbf{C}(\mathbf{B}+\mathbf{C}^\mathsf{T}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{C}^\mathsf{T}$ directly.

(c) Show that the conditional distribution of $\mathbf{x}$ given $\mathbf{y}$ is

$$\mathcal{N}^{-1}\left(\mathbf{x}\,;\,\mathbf{a}+\mathbf{C}^\mathsf{T}\mathbf{y}, \mathbf{B}+\mathbf{C}^\mathsf{T}\mathbf{D}^{-1}\mathbf{C}\right). \tag{9}$$

Hint: Theorem 3.4.1

Solution: Same procedure as above gives that we are interested in the conditional information state $\boldsymbol{\eta}_{x|y}$ and the conditional information matrix $\boldsymbol{\Lambda}_{x|y}$ from theorem 3.4.1. With the same associations we get the information state $\mathbf{b}+\mathbf{C}^\mathsf{T}\mathbf{y}$ and the information matrix $\mathbf{B}+\mathbf{C}^\mathsf{T}\mathbf{D}^{-1}\mathbf{C}$, which we wanted to show.

(d) Let us now return to the original formulation of the product identity in Theorem 3.3.1. Use the result from c) to show that

$$\widehat{\mathbf{P}}^{-1} = \mathbf{P}^{-1} + \mathbf{H}^\mathsf{T}\mathbf{R}^{-1}\mathbf{H}. \tag{10}$$

Hint: Match variables in (3.21) with (3.10) in the book.

Solution: Matching variables we see that $\mathbf{B} = \mathbf{P}^{-1}$, $\mathbf{C} = \mathbf{R}^{-1}\mathbf{H}$, $\mathbf{D} = \mathbf{R}^{-1}$ and $\widehat{\mathbf{P}}^{-1} = \boldsymbol{\Lambda}_{x|y} = \mathbf{B}+\mathbf{C}^\mathsf{T}\mathbf{D}^{-1}\mathbf{C} = \mathbf{P}^{-1} + (\mathbf{R}^{-1}\mathbf{H})^\mathsf{T}\mathbf{R}\mathbf{R}^{-1}\mathbf{H} = \mathbf{P}^{-1} + \mathbf{H}^\mathsf{T}\mathbf{R}^{-1}\mathbf{H}$ as claimed.