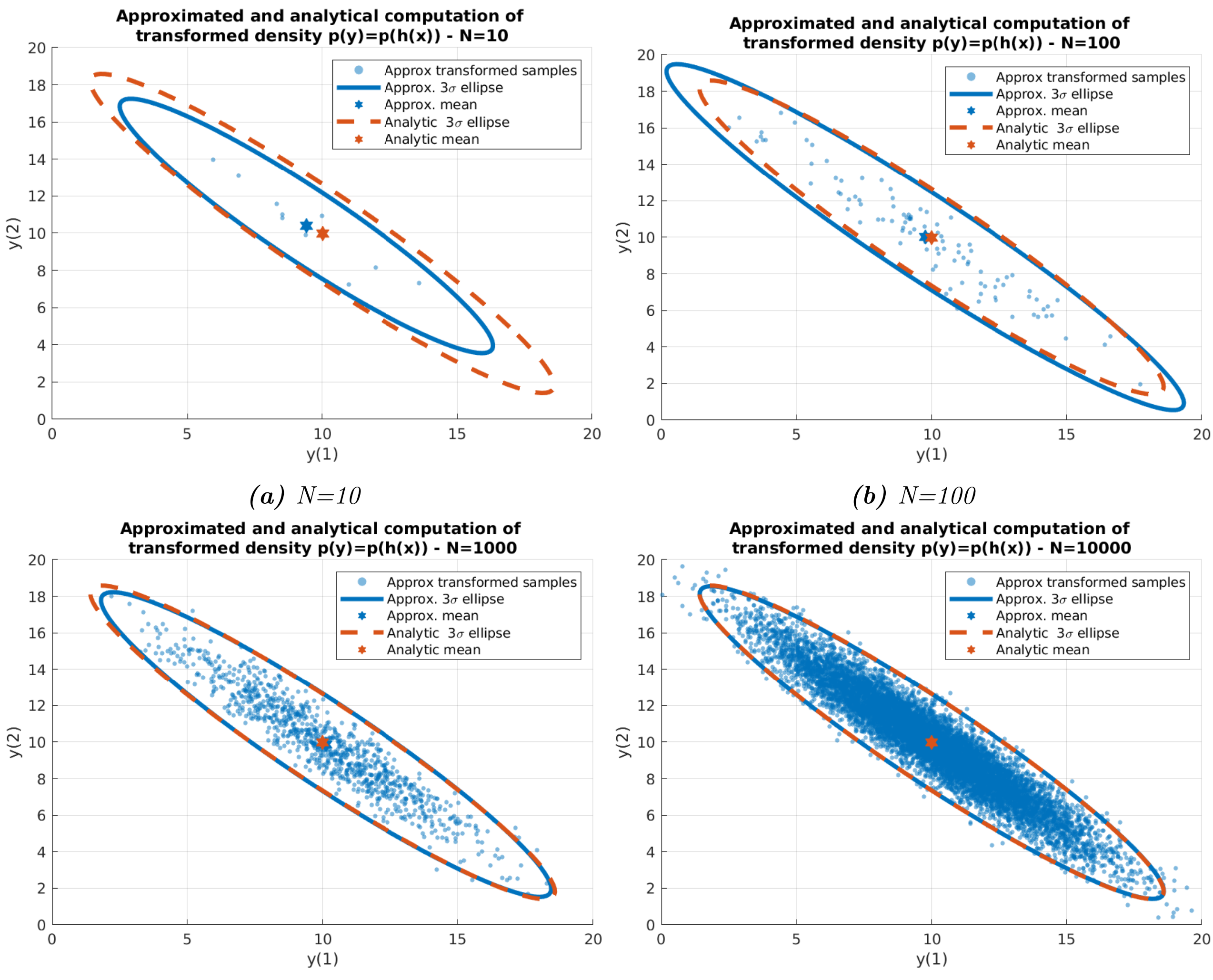
SSY191 - Sensor Fusion and Nonlinear Filtering
Solution to analysis in Home Assignment 01

Lucas Rath

Analysis

In this report I will present my independent analysis of the questions related to home assignment 1. I have not discussed the solution with anyone and I swear that the analysis written here are my own.

1 Transformation of Gaussian random variables



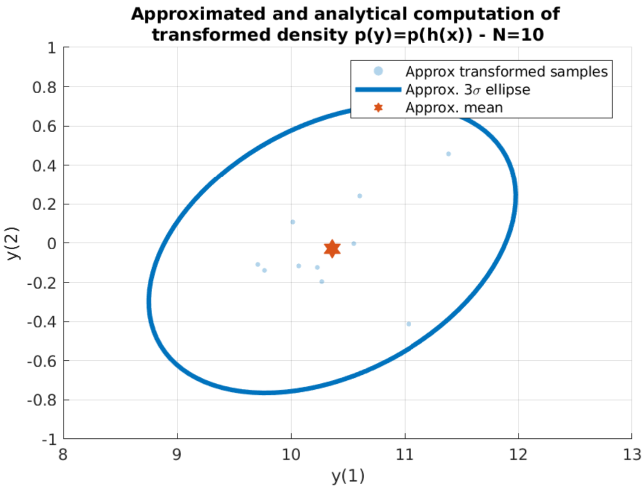
(a) N=10

(b) N=100

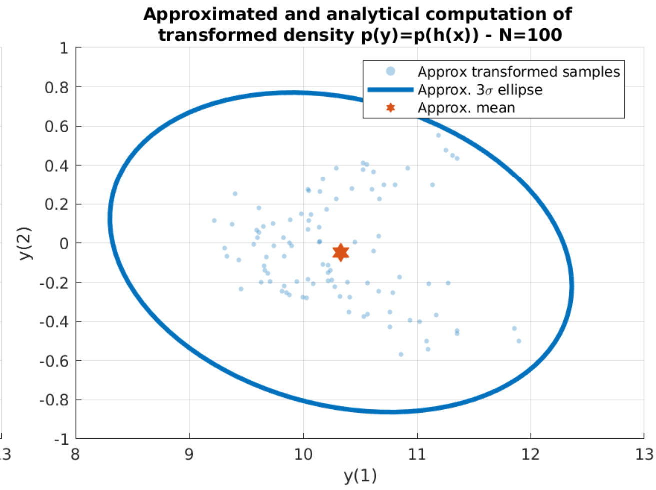
(c) N=1000

(d) N=10000

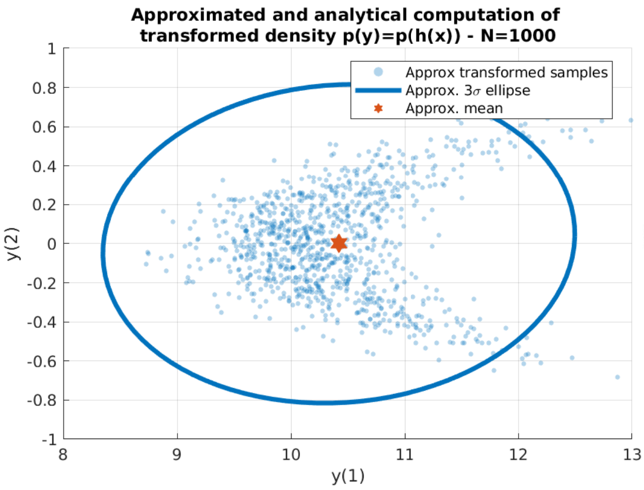
Figure 1: *Given the normal distribution $p(\boldsymbol{x})$ and the linear transformation $\boldsymbol{y} = h(\boldsymbol{x}) = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\boldsymbol{x}$, the mean and the 3σ ellipses of the analytical and approximated transformed density $p(y)$ can be seen in the subfigures (a)-(d) for different number of samples.*
The approximated 3σ ellipses of $\boldsymbol{y}$ match very well the sample points because they are distributed according to a Gaussian distribution. This happens because $\boldsymbol{x}$ is Gaussian and a linear transformation of a Gaussian dist. is also Gaussian. Moreover, the higher the number of samples, the better is the fit of the approximated mean and 3σ ellipses to the analytic values.



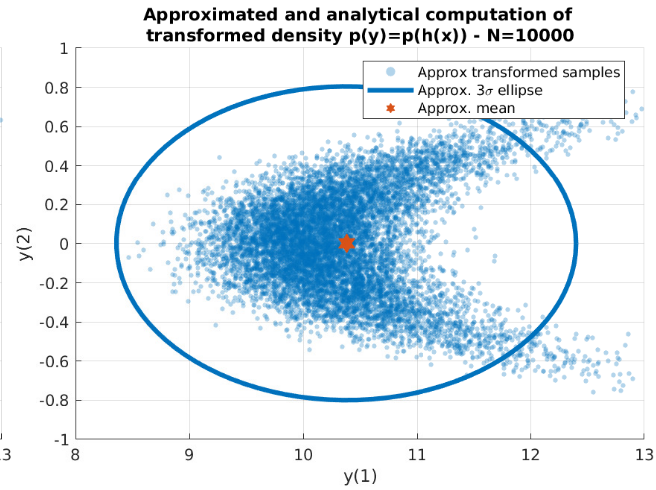
(a) *N=10*



(b) *N=100*



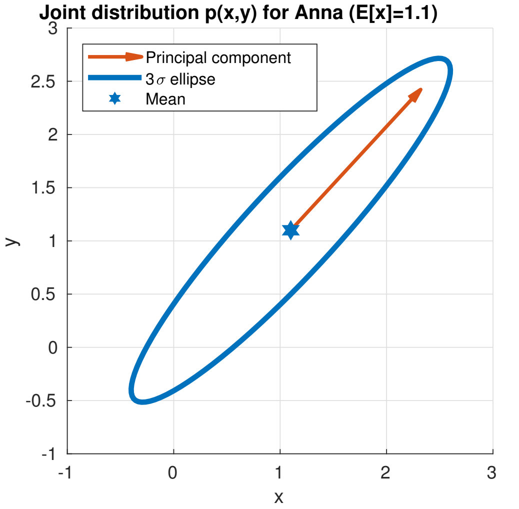
(c) *N=1000*



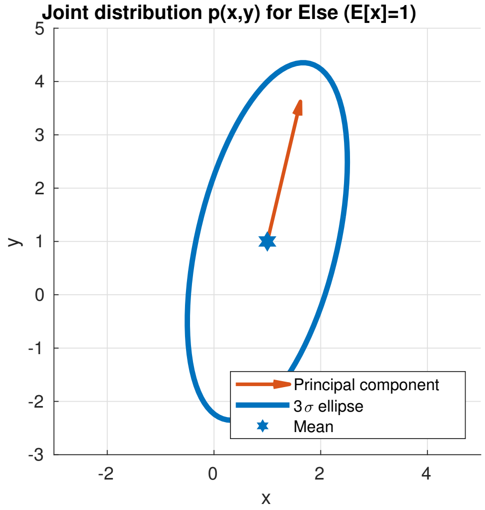
(d) *N=10000*

Figure 2: *Given the normal distribution $p(\boldsymbol{x})$ and the linear transformation $\boldsymbol{y} = h(\boldsymbol{x}) = \begin{bmatrix} ||\boldsymbol{x}|| \\ atan2(x_2, x_1) \end{bmatrix}$, the mean and the 3σ ellipses of the analytical and approximated transformed density $p(y)$ can be seen in the subfigures (a)-(d) for different number of samples.*
The approximated 3σ ellipses of $\boldsymbol{y}$ do not match the sample points because they are not distributed according to a Gaussian distribution. This happens because since the transformation $h(\boldsymbol{x})$ is not linear, $p(\boldsymbol{y})$ will not be a Gaussian distribution even if $\boldsymbol{x}$ is Gaussian. In this case, the 3σ ellipses do not have a relevant meaning or are not enough to represent this distribution. Moreover, the higher the number of samples, the better is the approximation of the real mean and variance of the samples.

2 Snow depth in Norway

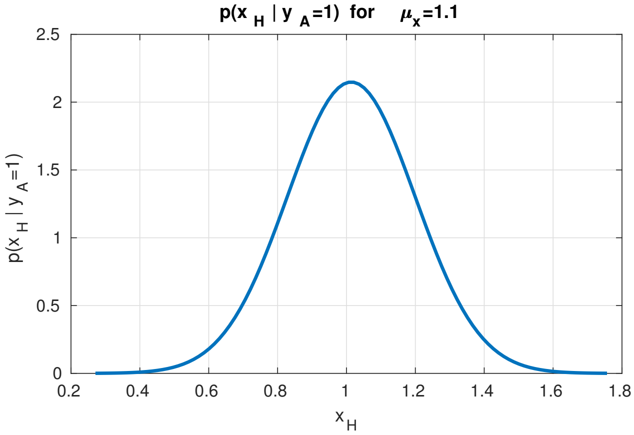


(a) Joint distribution $p(x, y)$ for Anna such that $p(x)=\mathcal{N}(x; \mu=1.1, 0.5^2)$ and $y=x+r$, $p(r)=\mathcal{N}(0, 0.2^2)$

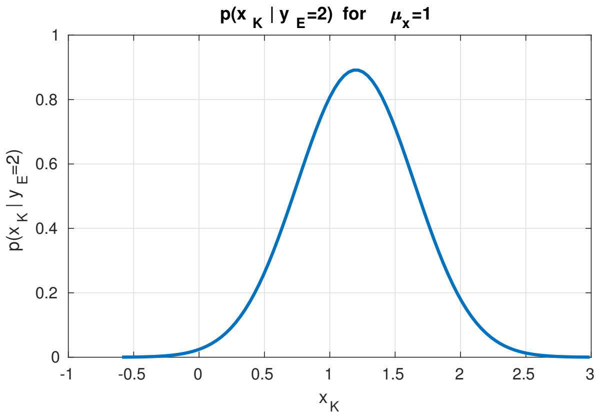


(b) Joint distribution $p(x, y)$ for Else such that $p(x)=\mathcal{N}(x; \mu=1, 0.5^2)$ and $y=x+r$, $p(r)=\mathcal{N}(0, 1^2)$

Figure 3: *Joint Gaussian distribution $p(x, y)$ of snow depth for (a) Anna at Hafjell and (b) Else at Kvitfjell. By analyzing the dependency of x and y with the 3σ ellipses, one can easily verify that in (a) y has similar covariance in comparison to x, which implies that p(x,y) depends almost equally on x and y. Also important to analyze is the direction of the principal component of p(x,y), which tells the direction of greatest variance of the distribution. In the case of figure (a) the slope of the principal component of p(y,x) is around 45 degrees, which means that the conditional distributions $p(x|y)$ and $p(y|x)$, which are proportional to p(x,y), will be equally sensitive to variations of the given value x or y, respectively. In addition, the estimators $\hat{x}_{MAP}$ and $\hat{x}_{MMSE}$ concerning $p(x|y)$ will be always located, in this case of a nicely-behaved Gaussian distribution, on the intersection of the given y value and the principal component line. This translates into "how important" is the given y value for the estimation of $\hat{x}$, since the slope of the principal component changes how much the given y influences the estimated value for the snow depth.*
On the other hand, in (b), the variance of y is much higher than the variance of x, which implies that p(x,y) is much more sensitive to variations of x than y. The principal component, which is almost aligned with the y-axis results in a distribution $p(x|y)$ that do not change that much for variations of the given y, since $E[x|y]$ and $Var[x|y]$ basically remain the same. As a result, the estimated snow depth $\hat{x}_{MAP}$ and $\hat{x}_{MMSE}$ of $p(x|y)$ will depend very little on the given value of y.



(a) *Posterior distribution $p(x_H|y_A = 1)$*



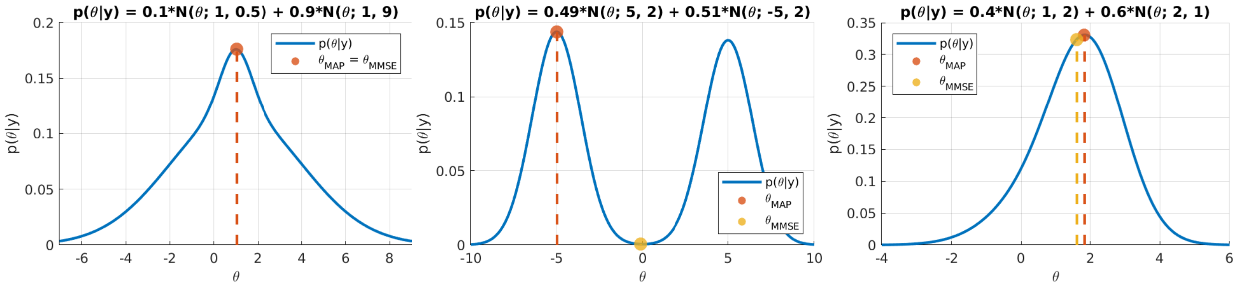
(b) *Posterior distribution $p(x_K|y_E = 2)$*

Figure 4: *Posterior distributions for (a) Anna at Hafjell and (b) Else at Kvitfjell. The posterior distributions $p(x|y)$ are proportional to the correspondent joint distribution $p(x, y)$, and therefore $p(x|y)$ can be obtained by simply intersecting the correspondent 2D distribution $p(x, y)$ with the given value of $y = 1$ for Anna and $y = 2$ for Else. By analyzing the two distributions, we can conclude that the expectation to find more snow is higher at Kvitfjell considering the forecasts and his friends' report since $E(x_K|y_E = 2) > E(x_H|y_A = 1)$.*

In sum, Anders should go ski with Else at Kvitfjell if he bases his decision on the MAP and MMSE estimators, since $x_{MAP} = x_{MMSE} = E(x_K|y_E = 2) = 1.2m$ are greater than $x_{MAP} = x_{MMSE} = E(x_H|y_A = 1) = 1.014m$. In this way, the expected value of the snow depth given the forecasts and his friends' reports is greater at Kvitfjell.

However, the expected values $E(x_K|y_E = 2) = 1.2000$ and $E(x_H|y_A = 1) = 1.0138$ do not differ that much and Anders could adopt a more conservative strategy going to Hafjell, since the covariance $Cov(x_K|y_E = 2) = 0.0345$ is much smaller than $Cov(x_H|y_A = 1) = 0.2000$. We can also visually verify that the probability of the snow depth being less than 0.5 meters is almost 0 at Hafjell, but is significantly large at Kvitfjell, which might be a crucial information for his decision. In conclusion, his decision depends on the cost function chosen by Anders, which could take into account not only the mean or the maximum a-posteriori but also another aspects of the distribution $p(x|y)$.

3 MMSE and MAP estimates for Gaussian mixture posteriors



(a) *Posterior dist.* $p_a(\theta|y)$ ***(b)*** *Posterior dist.* $p_b(\theta|y)$ ***(c)*** *Posterior dist.* $p_c(\theta|y)$

Figure 5: *MAP and MMSE estimators for different posterior distributions* $p(\theta|y)$.
In sub-figure a) both MAP and MMSE estimators are the same because the expected value and the argument of the maximum likelihood of the distribution $p(\theta|y)$ *will lead to the same value of* $\hat{\theta}$.
In sub-figure b) the MAP and MMSE estimators lead to completely different values of $\hat{\theta}$. *Clearly, the mean of* $p(\theta|y)$ *corresponds to a very low likelihood, which in many applications might result in a poor estimation.*
In sub-figure c) both MAP and MMSE are very similar, but not identical. In this case, the distribution $p(\theta|y)$ *has slightly more weight to values of* θ *below the peak. Therefore, the MMSE estimator leads to a* $\hat{\theta}_{MMSE}$ *slightly smaller than* $\hat{\theta}_{MAP}$.