

Assignment 1

Computational Intelligence, SS2020

Team Members		
Last name	First name	Matriculation Number
Blöcher	Christian	01573246
Bürgener	Max	01531577

1 Maximum Likelihood Estimation of Model Parameters

1.1 Which measurement in sc.2 is exponentially distributed?

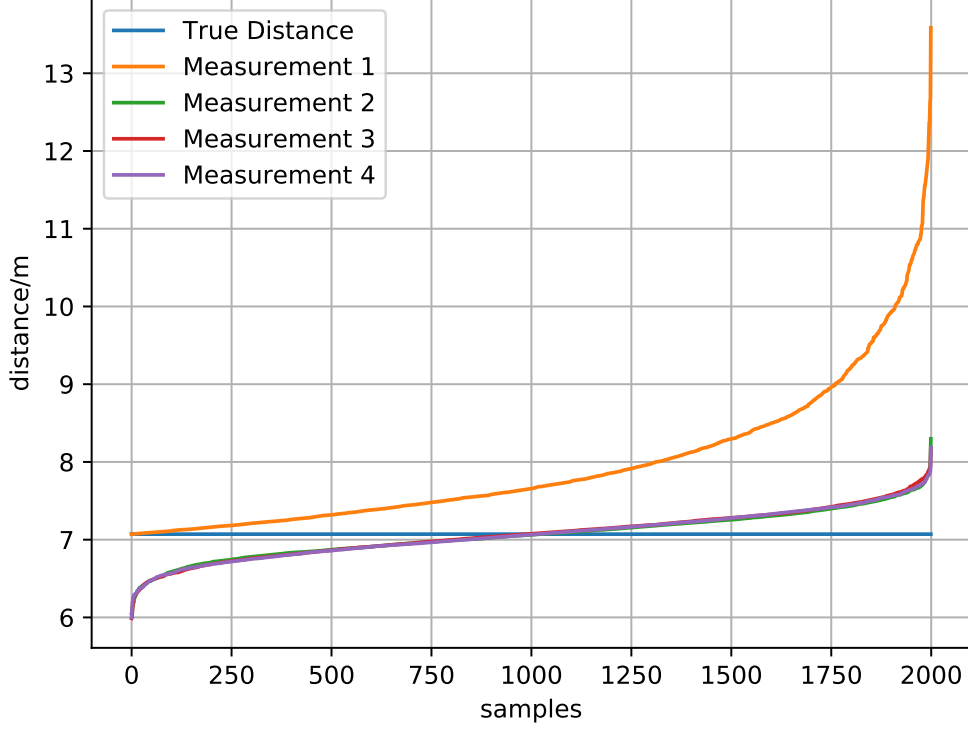


Figure 1: Scenario 2 with mixed measurement models for the anchors

- In figure 1 we can see that in scenario 2 the distance measurements from anchor 1 is exponentially distributed. It is the only distribution which is bigger than the true distance for all samples and its slope is rising exponentially for rising x-values.

1.2 Derivation for the maximum likelihood solution

- Analytical derivation for the Gaussian distribution:

$$p(\tilde{d}_n(a_i, \mathbf{p}) | \mathbf{p}) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{[\tilde{d}_n(a_i, \mathbf{p}) - d(a_i, \mathbf{p})]^2}{2\sigma^2}}$$

The data is independent identically distributed (iid), therefore the likelihood function is the product of all individual likelihoods

$$P(\tilde{d}_n(a_i, \mathbf{p}) | \mathbf{p}) = \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{[\tilde{d}_n(a_i, \mathbf{p}) - d(a_i, \mathbf{p})]^2}{2\sigma^2}}$$

To convert the product to a sum we apply the natural logarithm.

$$L(\tilde{d}_n(a_i, \mathbf{p}) | \mathbf{p}) = \ln \left[\prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{[\tilde{d}_n(a_i, \mathbf{p}) - d(a_i, \mathbf{p})]^2}{2\sigma^2}} \right]$$

$$L(\tilde{d}_n(a_i, \mathbf{p}) | \mathbf{p}) = \sum_{n=0}^{N-1} \left[\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{[\tilde{d}_n(a_i, \mathbf{p}) - d(a_i, \mathbf{p})]^2}{2\sigma^2} \right]$$

Since we want to find the parameter σ^2 , which maximizes the probability of the distance, we derive $L(\tilde{d}_n(a_i, \mathbf{p}) | \mathbf{p})$ and set it to zero.

$$\begin{aligned} L(\tilde{d}_n(a_i, \mathbf{p}) | \mathbf{p}) &= \sum_{n=0}^{N-1} \left[\ln(1) - \frac{1}{2} \ln(2\pi\sigma^2) - \frac{[\tilde{d}_n(a_i, \mathbf{p}) - d(a_i, \mathbf{p})]^2}{2\sigma^2} \right] \quad \Big| \frac{\partial}{\partial \sigma^2} \\ \frac{\partial}{\partial \sigma^2} L(\tilde{d}_n(a_i, \mathbf{p}) | \mathbf{p}) &= \sum_{n=0}^{N-1} \left[-\frac{1}{\sigma^2} + \frac{[\tilde{d}_n(a_i, \mathbf{p}) - d(a_i, \mathbf{p})]^2}{\sigma^4} \right] \quad \stackrel{!}{=} 0 \\ 0 &= \sum_{n=0}^{N-1} \left[-\frac{1}{\sigma^2} + \frac{[\tilde{d}_n(a_i, \mathbf{p}) - d(a_i, \mathbf{p})]^2}{\sigma^4} \right] \\ \frac{N}{\sigma^2} &= \sum_{n=0}^{N-1} \frac{[\tilde{d}_n(a_i, \mathbf{p}) - d(a_i, \mathbf{p})]^2}{\sigma^4} \\ \sigma^2 &= \frac{\sum_{n=0}^{N-1} [\tilde{d}_n(a_i, \mathbf{p}) - d(a_i, \mathbf{p})]^2}{N} \end{aligned}$$

- Analytical derivation for the Exponential distribution:

$$p(\tilde{d}_n(a_i, \mathbf{p}) | \mathbf{p}) = \begin{cases} \lambda_i e^{-\lambda_i [\tilde{d}_n(a_i, \mathbf{p}) - d(a_i, \mathbf{p})]} & , \tilde{d}_n(a_i, \mathbf{p}) \geq d(a_i, \mathbf{p}) \\ 0 & , \text{else} \end{cases}$$

$$L(\tilde{d}_n(a_i, \mathbf{p}) | \mathbf{p}) = \sum_{n=0}^{N-1} \ln(\lambda_i) - \lambda_i [\tilde{d}_n(a_i, \mathbf{p}) - d(a_i, \mathbf{p})] \quad \Big| \frac{\partial}{\partial \lambda_i}$$

$$\frac{\partial}{\partial \lambda_i} L(\tilde{d}_n(a_i, \mathbf{p}) | \mathbf{p}) = \frac{N}{\lambda_i} - \sum_{n=0}^{N-1} [\tilde{d}_n(a_i, \mathbf{p}) - d(a_i, \mathbf{p})] \quad \stackrel{!}{=} 0$$

$$\frac{N}{\lambda_i} = \sum_{n=0}^{N-1} [\tilde{d}_n(a_i, \mathbf{p}) - d(a_i, \mathbf{p})]$$

$$\lambda_i = \frac{N}{\sum_{n=0}^{N-1} [\tilde{d}_n(a_i, \mathbf{p}) - d(a_i, \mathbf{p})]} \quad , \tilde{d}_n(a_i, \mathbf{p}) \geq d(a_i, \mathbf{p})$$

2 Estimation of the Position

2.1 Least-Squares Estimation of the Position

- Analytical conversion of the ML estimation equation:

$$\begin{aligned}
\hat{\mathbf{p}}_{ML}(n) &= \underset{\mathbf{p}}{\operatorname{argmax}} \prod_{i=0}^{N_A-1} p(\tilde{d}_n(a_i, \mathbf{p}) \mid \mathbf{p}) \\
\hat{\mathbf{p}}_{ML}(n) &= \underset{\mathbf{p}}{\operatorname{argmax}} \ln \left[\prod_{i=0}^{N_A-1} p(\tilde{d}_n(a_i, \mathbf{p}) \mid \mathbf{p}) \right] \\
\hat{\mathbf{p}}_{ML}(n) &= \underset{\mathbf{p}}{\operatorname{argmax}} \sum_{i=0}^{N_A-1} \ln \left(\frac{1}{\sqrt{2\pi\sigma_i^2}} \right) - \frac{[\tilde{d}_n(a_i, \mathbf{p}) - d(a_i, \mathbf{p})]^2}{2\sigma_i^2}
\end{aligned}$$

Because in scenario 1 we only use Gaussian models for all anchors that were calibrated with the same distance to the reference position, we can assume that $\sigma_i^2 = \sigma^2 \forall i$. That means the \ln -term can be neglected since it only shifts the value of the maximum by a constant but does not affect its position. Similarly $\frac{1}{2\sigma^2}$ can be omitted, as it is also just a scaling factor. Furthermore $\underset{\mathbf{p}}{\operatorname{argmax}} (-\dots)$ is equivalent to $\underset{\mathbf{p}}{\operatorname{argmin}} (\dots)$. Thus:

$$\hat{\mathbf{p}}_{ML}(n) = \underset{\mathbf{p}}{\operatorname{argmin}} \sum_{i=0}^{N_A-1} [\tilde{d}_n(a_i, \mathbf{p}) - d(a_i, \mathbf{p})]^2 = \hat{\mathbf{p}}_{LS}(n)$$

2.2 Gauss-Newton Algorithm for Position Estimation

- Analytical solution for the Jacobian matrix.

$$[J(p)]_{i,1} = \frac{\partial}{\partial x} \left[\tilde{d}_n(a_i, \mathbf{p}) - \sqrt{(x_i - x)^2 + (y_i - y)^2} \right]$$

$$[J(p)]_{i,1} = - \frac{[(-)2(x_i - x)]}{2 \cdot \sqrt{(x_i - x)^2 + (y_i - y)^2}}$$

$$[J(p)]_{i,1} = \frac{(x_i - x)}{\sqrt{(x_i - x)^2 + (y_i - y)^2}}$$

$$[J(p)]_{i,2} = \frac{(y_i - y)}{\sqrt{(x_i - x)^2 + (y_i - y)^2}}$$

- Mean and variance

	Scenario 1	Scenario 2		Scenario 3
		with exponential anchor	without exponential anchor	
Error mean μ_e	0.278	0.640	0.399	1.265
Error variance σ_e^2	0.022	0.275	0.054	0.939

Table 1: Error-mean and -variance for different Scenarios.

Table 1 shows that the Gauss-Newton Algorithm is optimized for Gauss distributions. Mean μ_e and variance σ_e^2 are much bigger in the mixed scenario 2 and in scenario 3 which consists of 4 exponential distributions.

- Multivariate Gaussian distributions

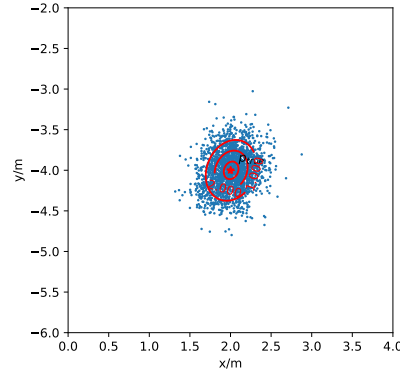


Figure 2: Gaussian distribution of scenario 1

0,04603	0,00869
0,00869	0,05823

Table 2: Covariance matrix

The Gaussian distribution is computed by the Covariance matrix which is next to Figure 2. The distribution for scenario 1 is almost *spherically shaped* due to the small difference in the main diagonal and the very low values in the secondary diagonal.

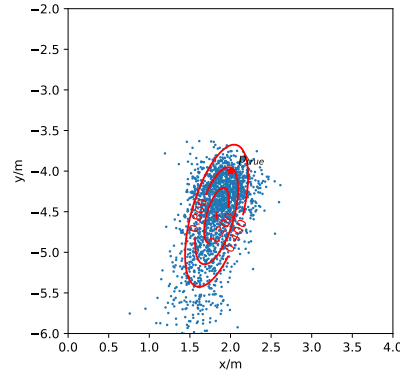


Figure 3: Gaussian distribution of scenario 2

0,05783	0,07271
0,07271	0,29491

Table 3: Covariance matrix

With rising covariance values the variance around the mean value is rising. Therefore we can see a general Gaussian distribution in Figure 4 with the distribution stretched out in the direction leading away from the exponentially modelled anchor. That is because the exponentially distributed distance measurements are prone to take larger values than the Gaussian ones (s. figure 1). This makes the position estimation less precise.

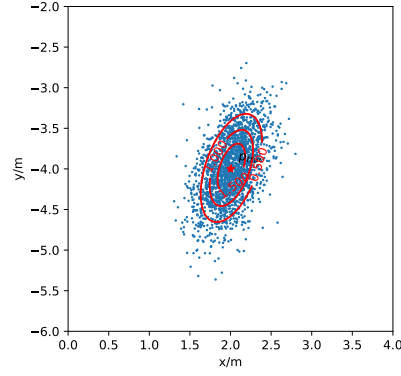


Figure 4: Scenario 2 without the exponential distributed anchor

0,05175	0,04369
0,04369	0,16122

Table 4: Covariance matrix

Omitting the exponential anchor the variance is still bigger than in scenario 2 although the single estimations are more evenly distributed around the mean value which is also very close to the true position. Overall, the measurement is better without the exponential anchor.

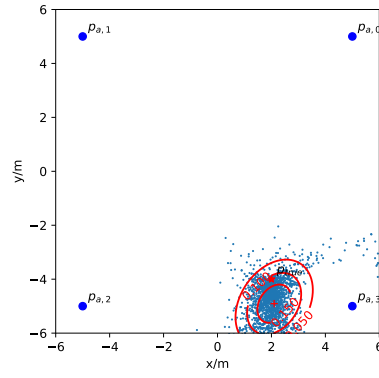


Figure 5: Gaussian distribution of scenario 3

0,73497	0,25884
0,25884	0,98659

Table 5: Covariance matrix

The density function of scenario 3 is also computed by a general covariance matrix with bigger values in all dimensions. Therefore the point estimation is not precise and the variance around the mean value is very high.

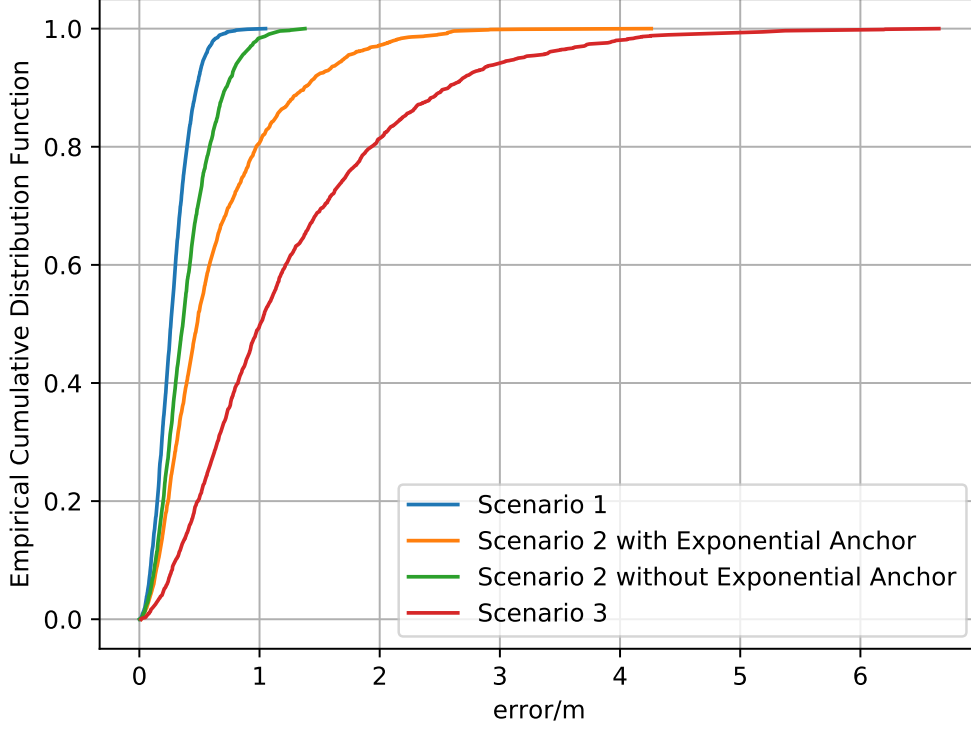


Figure 6: Gaussian distribution of scenario 3

The cumulative distribution functions of the position errors emphasize our finding that the Gauss-Newton Algorithm is optimized for Gauss-distributed data sets. That is why scenario 1 delivers the best performance with a very low maximum error. Only using the three Gaussian anchors in scenario 2 leads to a similar cumulative distribution function with a slightly bigger maximum error. Taking the exponential anchor into account the maximum error greatly increases but 90% of all errors are still smaller than the previous maximum error. Scenario 3 has the worst performance with larger errors overall and by far the biggest maximum error.

2.3 Numerical Maximum Likelihood Estimation of the Position

2.3.1 Single Measurement

- The numerical maximum likelihood estimate is computed by finding the maximum of the joint likelihood of all anchors evaluated within a 2D-grid enclosed by the anchors. Because of the i.i.d.-assumption the joint likelihood can be calculated as the product of all individual exponential likelihoods:

$$p(\tilde{\mathbf{d}}_n(\mathbf{p})|\mathbf{p}) = \begin{cases} \prod_{i=0}^{N_A-1} p(\tilde{d}_n(a_i, \mathbf{p})|\mathbf{p}), & \text{if } \tilde{d}_n(a_i, \mathbf{p}) \geq d(a_i, \mathbf{p}) \forall i \\ 0 & \text{else} \end{cases}$$

- Because of its nonlinearity the joint likelihood function of the first sample $n = 0$ has two local maxima (s. figure 7). If we used a gradient ascent algorithm with a random starting position, it would stop once it reaches any of them, possibly leading to a false estimation of

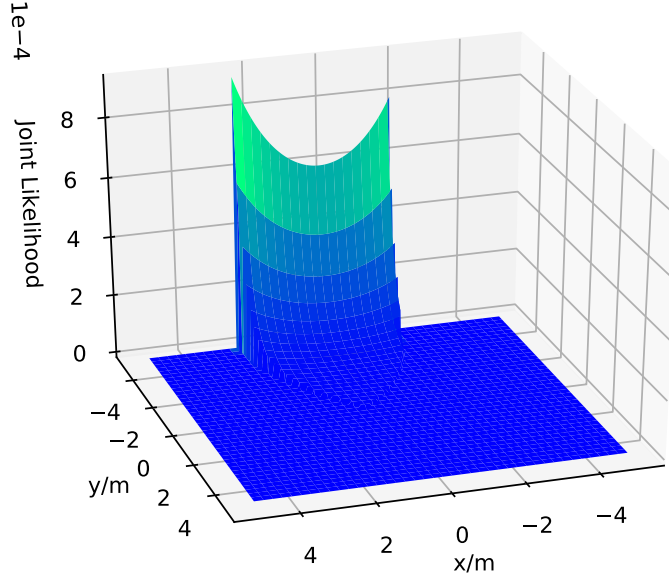


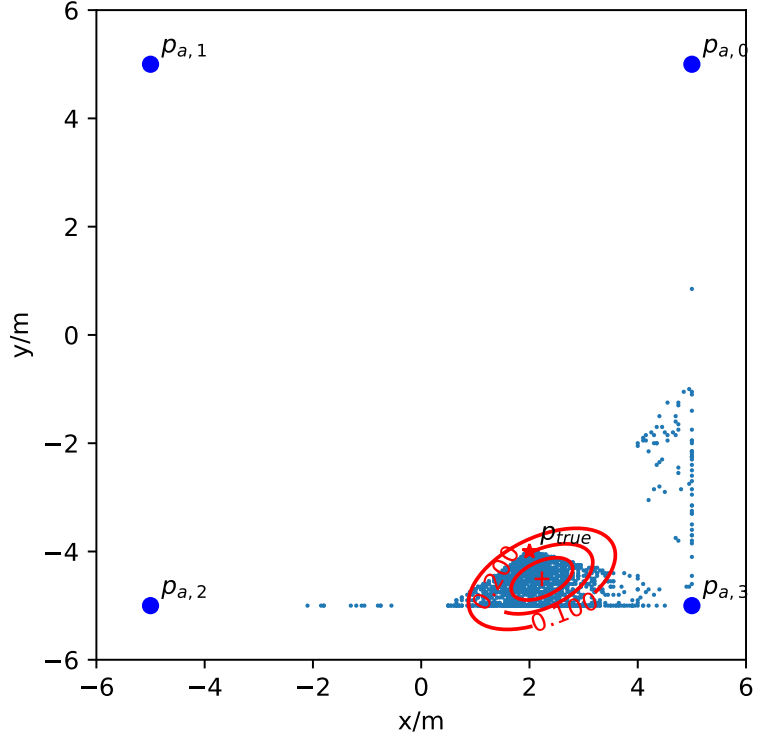
Figure 7: $p(\tilde{\mathbf{d}}_0(\mathbf{p})|\mathbf{p})$ evaluated within a 2D-grid enclosed by the anchors.

the position, if the found maximum is not global.

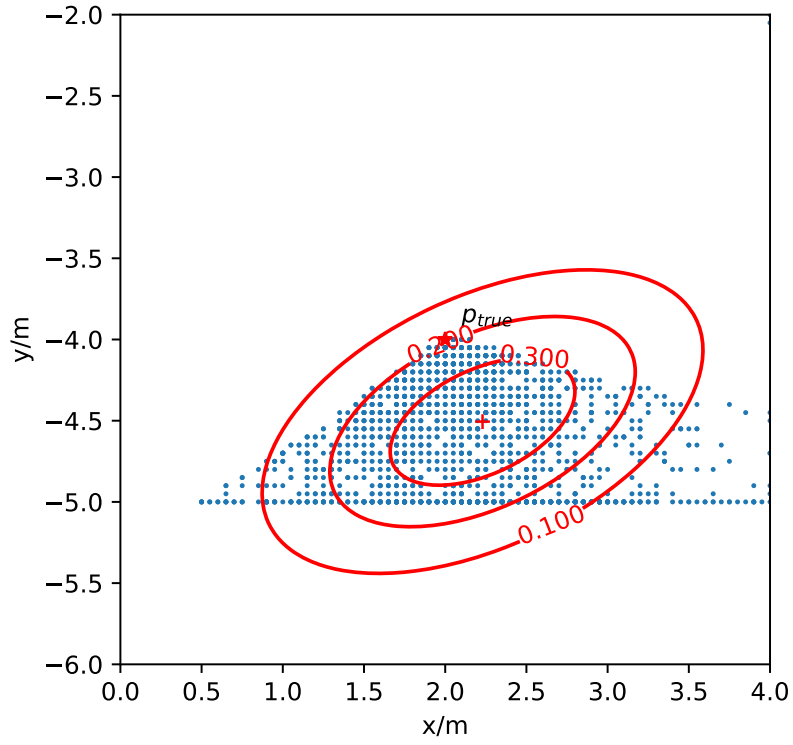
- The found maximum at $\mathbf{p}_{0,NML} = [2.5, -5]$ is not at the true position. This is because the estimation is solely based on the noisy distance measurements $\tilde{\mathbf{d}}_0$: Within the grid the likelihood for each anchor $p(\tilde{d}_0(a_i, \mathbf{p})|\mathbf{p})$ is maximised on a circular arc with radius $\tilde{d}_0(a_i, \mathbf{p})$ and center a_i . The joint likelihood is nonzero within the area enclosed by both the grid borders and all arcs. Its maximum is at a point within this area as close as possible to all arcs - e.g. at the intersection of two arcs in close vicinity to the remaining ones - but because of the noisiness of the data and the restrictions of the grid, this is not necessarily the true position.

2.3.2 Multiple Measurements

- Both error mean μ_e and variance σ_e^2 of the NML estimation are much smaller than the corresponding parameters of the LS estimation (s. table 6). In figure 9 a lower maximum error of ca. 5.6m is visible with 90% of all estimates having an error $< 1.5\text{m}$, which is an improvement over the LS errors. The cloud of position estimates displayed in the scatter-plot (s. figure 8) just below \mathbf{p}_{true} is denser than the LS one but is truncated by the bottom grid border. There are also many outliers near the bottom and right grid borders which make the distribution seem even less Gaussian than the the LS results and cause the Gaussian-contour to be more tilted, indicating that the position estimate coordinates x and y are not statistically independent.
- Of course this comparison is not very fair, since we are only looking at maxima within the grid enclosed by the anchors. This prevents position estimates with larger error values - e.g. estimates below $y = -5$ as in the LS-case - which explains the smaller error parameters. But this also means that the numerical approach confined by the grid borders and its limited resolution is not truly a Maximum Likelihood Estimator, as we can not find the actual maximum of the joint likelihood for all samples.



(a) Full view of scatter plot.



(b) Detailed view centered on p_{true} .

Figure 8: Scatter plot of NML-estimated positions with Gaussian-contour.

- The prior knowledge used for the Bayes estimator greatly enhances the accuracy. Here the error mean μ_e is almost half as big as the LS one, the error variance σ_e^2 is cut down to almost one tenth of the LS result (s. table 6). Looking at figure 9 the Bayes estimator is a big improvement over the other approaches, the maximum error being only 1.2m. This is also visible in the scatter-plot (s. figure 10), which shows a very dense estimate-cluster with no outliers. However it is still not centred on \mathbf{p}_{true} but on about the same mean estimate as the NML estimation. The Gaussian-contour is not tilted and (because the cluster is slightly truncated by the bottom grid border) almost circular, indicating stochastically independent estimate coordinates with similar variance.

	Least-Squares	Numerical Maximum Likelihood	Bayes
Error mean μ_e	1.265	0.915	0.680
Error variance σ_e^2	0.939	0.489	0.095

Table 6: Error-mean and -variance of Least-Squares-, Numerical Maximum Likelihood- and Bayes-Estimation.

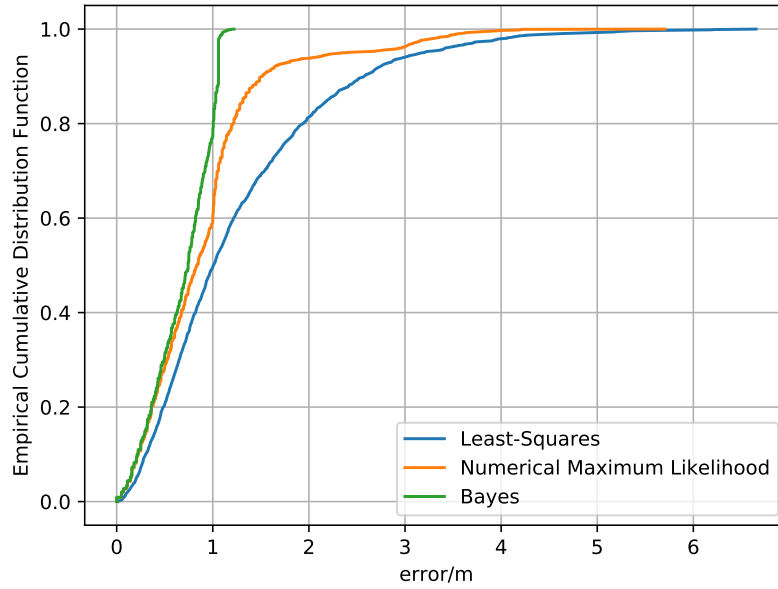
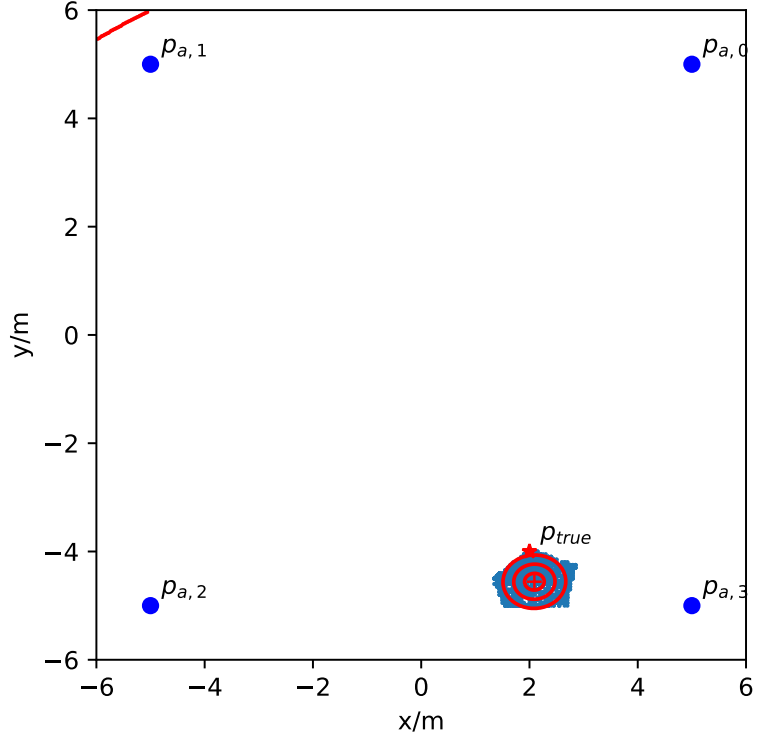
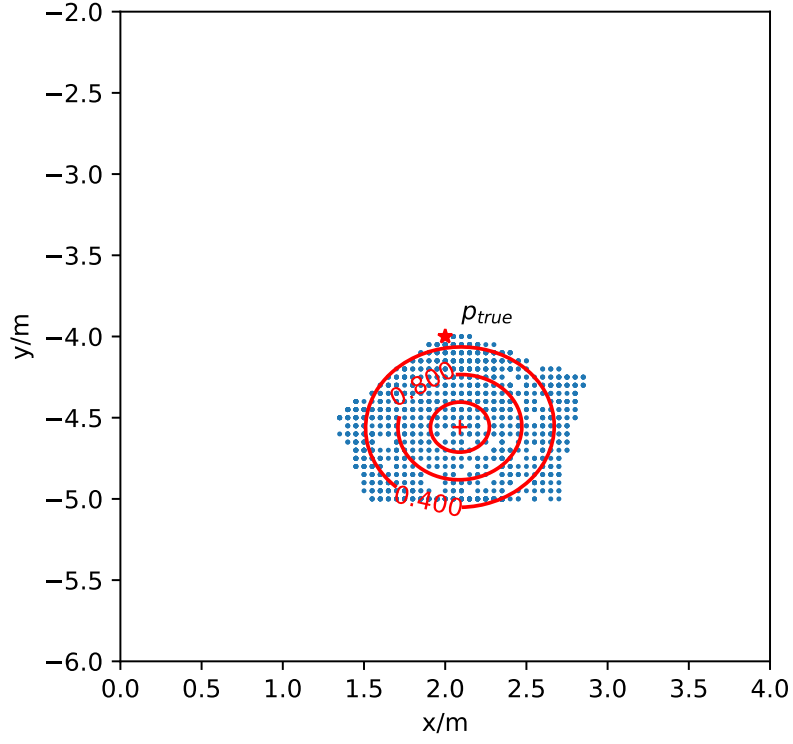


Figure 9: ECDF of Least-Squares-, Numerical Maximum Likelihood- and Bayes-Estimation.



(a) Full view of scatter plot.



(b) Detailed view centered on \mathbf{p}_{true} .

Figure 10: Scatter plot of Bayes-estimated positions with Gaussian-contour.