Estimation Laboratory 1

Nathan Magnan, Mathieu Roule

Novembre 2020

Question 1 :

We chose to normalize the amplitude at $a = 1$, and to reproduce a figure visually similar to the one proposed, we set the background at $b = 0.3$ and the noise at $\sigma = 0.1$, that is a 10 % noise.

Just like in the example, we chose a 20×20 image, but to be more general we did not place our star in the center of the image, but a bit offset at position $(r_0, c_0) = (8.5, 8.5)$.

To avoid having too many problems with a part of the information being outside the image, we chose $\alpha = 4$ and a rather steep Moffat function at $\beta = 3$.

Finally, we obtain the following image :

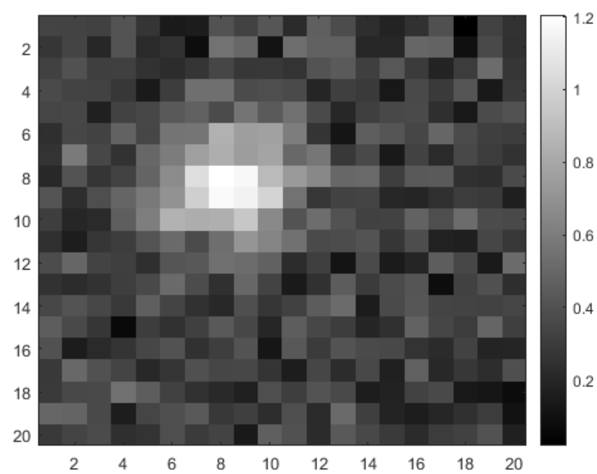

Figure 1: This is the image we want to extract the parameters $a, b, \alpha, \beta, r_0, c_0$ from.

1 Estimation with known star's location and PSF parameters

Question 1 :

Let $\underline{\theta}$ be the parameter vector of size 2 $\left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right)$, and $\underline{d}$ a data vector of size RC representing the image. We also give ourselves a noise vector $\underline{\epsilon}$ of size RC. The forward model can be written in the linear algebra formalism as :

$$\underline{d} = \underline{\underline{H}}\,\underline{\theta} + \underline{\epsilon} \tag{1}$$

Where $\underline{\underline{H}}$ is a matrix of RC lines and 2 columns whose coefficient are defined by the following relation :

$$H_{C(r-1)+c,j} = \begin{cases} PSF(r - r_0, c - c_0) & \text{if } j = 1 \\ 1 & \text{if } j = 2 \end{cases} \tag{2}$$

Question 2 :

We are in the linear additive Gaussian case, hence the likelihood function and the neg-log-likelihood take the following forms :

$$\begin{aligned} L(\underline{\theta};\underline{d}) = & \quad \frac{1}{(2\pi\sigma^2)^{RC/2}} exp\left(-\frac{1}{2}\frac{(\underline{d}-\underline{\underline{H}}\,\underline{\theta})^T(\underline{d}-\underline{\underline{H}}\,\underline{\theta})}{\sigma^2}\right) \\ NLL(\underline{\theta};\underline{d}) = & \quad \frac{1}{2\sigma^2}(\underline{d}-\underline{\underline{H}}\,\underline{\theta})^T(\underline{d}-\underline{\underline{H}}\,\underline{\theta}) + \text{Cst} \end{aligned} \tag{3}$$

The maximum likelihood estimator is the maxima of the likelihood, or equivalently the minima of the neg-log-likelihood. In our case with a centered, independent noise it is simply :

$$\overset{\circ}{\underline{\theta}}_{ML}(\underline{d}) = \quad (\underline{\underline{H}}^T\underline{\underline{H}})^{-1}\underline{\underline{H}}^T\underline{d} \tag{4}$$

Question 3 :

In the additive linear Gaussian case, the maximum likelihood estimator is unbiased. And because our noise is independent, the covariance matrix will be :

$$\underline{\underline{\Gamma}}_{\overset{\circ}{\theta}_{ML}} = \quad \sigma^2(\underline{\underline{H}}^T\underline{\underline{H}})^{-1} \tag{5}$$

$$\begin{aligned} &\approx \begin{pmatrix} \sum_{r,c} PSF_{\alpha,2\beta}(r,c) & \sum_{r,c} PSF_{\alpha,\beta}(r,c) \\ \sum_{r,c} PSF_{\alpha,\beta}(r,c) & 1 \end{pmatrix}^{-1} \times \sigma^2 \\ &= \begin{pmatrix} \iint PSF_{\alpha,2\beta}(r,c)drdc & \iint PSF_{\alpha,\beta}(r,c)drdc \\ \iint PSF_{\alpha,\beta}(r,c)drdc & 1 \end{pmatrix}^{-1} \times \sigma^2 \\ &= \begin{pmatrix} \frac{\pi\alpha^2}{2\beta-1} & \frac{\pi\alpha^2}{\beta-1} \\ \frac{\pi\alpha^2}{\beta-1} & 1 \end{pmatrix}^{-1} \times \sigma^2 \end{aligned} \tag{6}$$

The approximation of the sums by the indefinite integrals is valid if no information is outside the image, and if the image contains enough pixels. When they are valid, they are extremely useful, since we can avoid a big matrix multiplication. But in the general case the solution remains analytical and there is no big matrix inversion so the estimation process remains computationally light.

Question 4 :

We compute the maximum likelihood estimators from questions 2 and 3, and find the following results :

Parameter	Value	Estimation
a	1.0	0.9750
b	0.3	0.3029
Γ	$\underline{\underline{0}}$	$\begin{pmatrix} 0.0012 & -0.0001 \\ -0.0001 & 0.0 \end{pmatrix}$

Table 1: Table of comparison between original and estimated parameters

Question 5 :

The estimated parameters are close to the true values of the parameters, with a maximum relative error of 2.25%.

However, the estimated covariance matrix is too small : it yields a zero variance on parameter b, which is impossible in a noisy problem such as ours. Also, it announces a 1% uncertainty on parameter a. Of course, it could be that our unique test is particularly unlucky (we would be at more than 2 σ), but more likely the covariance matrix estimator is too optimistic.

Question 6 :

The image of the residual does not show traces of the PSF :

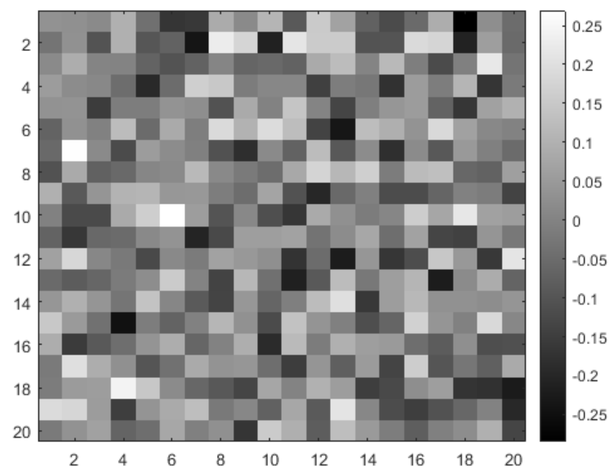

Figure 2: Residuals for the ML estimation with known star's location and PSF parameters. We do not see traces of the PSF on the image.

We also find the following distribution for the residuals :

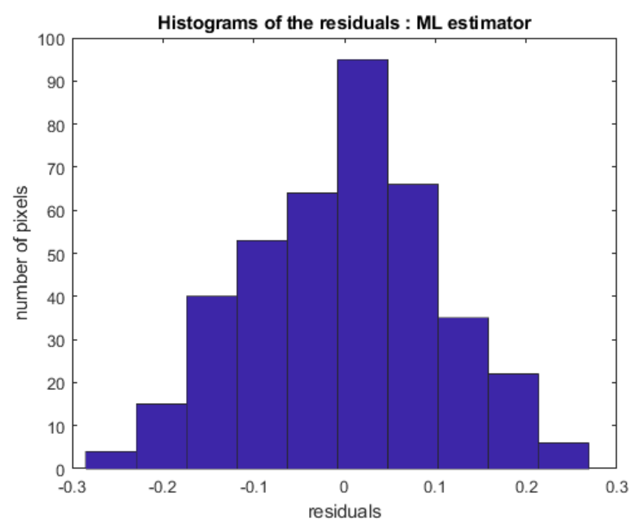

Figure 3: Histogram for the residuals in the case of ML estimation with known star's location and PSF parameters. The distribution seems to be centered. It is hard to say whether it is Gaussian or not.

This distribution seems to be quite well centered, however it is hard to conclude whether it is Gaussian or not. We can only say that these residuals are not obviously in disagreement with the additive Gaussian centered noise assumption we made earlier.

2 Maximum likelihood estimation of all the parameters

Question 1 :

As in the first section, the forward model can be written as :

$$\underline{d} = \underline{\underline{H}}(\underline{\nu})\,\underline{\theta} + \underline{\epsilon} \tag{7}$$

where $\underline{\underline{H}}$ is now unknown since it depends on $\underline{\nu} = (r_o, c_o, \alpha, \beta)^T$ as described by the following relation :

$$H_{C(r-1)+c,j} = \begin{cases} h(r,c;\underline{\nu}) = PSF_{\alpha,\beta}(r - r_0, c - c_0) & \text{if } j = 1 \\ 1 & \text{if } j = 2 \end{cases} \tag{8}$$

and where the size/shape of the vectors/matrix are the same as in part 1 : $\underline{d}$ and $\underline{\epsilon}$ are vectors of size RC, $\underline{\underline{H}}$ a matrix of shape $(RC, 2)$, $\underline{\theta}$ a vector of size 2.

Question 2 :

The main difference with what has been done before is that the matrix $\underline{\underline{H}}$ now depends on unknown parameters we are trying to estimate. Calling $\underline{p} = (\underline{\theta}^T, \underline{\nu}^T)^T$ the vector of all the parameters, the joint likelihood of parameters $\underline{\theta}$ and $\underline{\nu}$ can be written as :

$$\begin{aligned} L(\underline{p};\underline{d}) &= f_{D|P}(\underline{d}|\underline{p}) \\ &= \tfrac{1}{(2\pi\sigma^2)^{RC/2}} exp\left(-\tfrac{1}{2}\tfrac{(\underline{d}-\underline{\underline{H}}(\underline{\nu})\,\underline{\theta})^T(\underline{d}-\underline{\underline{H}}(\underline{\nu})\,\underline{\theta})}{\sigma^2}\right) \end{aligned} \tag{9}$$

Then,

$$J(\underline{\theta},\underline{\nu};\underline{d}) = (\underline{d} - \underline{\underline{H}}(\underline{\nu})\,\underline{\theta})^T(\underline{d} - \underline{\underline{H}}(\underline{\nu})\,\underline{\theta}) \tag{10}$$

is a cost function which is minimal at the maximal likelihood estimation.

$$\overset{\circ}{\underline{p}}_{ML}(\underline{d}) = \underset{(\underline{\theta},\underline{\nu})}{argmin}\; J(\underline{\theta},\underline{\nu};\underline{d}) = \sum_{i=1}^{RC} |d_i - (\underline{\underline{H}}(\underline{\nu})\underline{\theta})_i|^2 \tag{11}$$

We do not have an explicit expression for the maximal likelihood estimator, we are facing a non linear optimization problem as $\underline{\underline{H}}$ depends on the parameters $\underline{\nu}$ through the Point Spread Function which is not linear w.r.t the parameters.

Remark :

Since we know, for a given $\underline{\nu}$, the expression of the maximal likelihood $\overset{\circ}{\underline{\theta}}_{ML}$ through the equation (4), we can rewrite this minization problem as follows :

$$\begin{aligned} \underset{(\underline{\theta},\underline{\nu})}{min}\; J(\underline{\theta},\underline{\nu};\underline{d}) &= \underset{\underline{\nu}}{min}\;\underset{\underline{\theta}}{min}\; J(\underline{\theta},\underline{\nu};\underline{d}) \\ &= \underset{\underline{\nu}}{min}\;[\,\underline{d} - \underline{\underline{H}}(\underline{\nu})\,\overset{\circ}{\underline{\theta}}_{ML}(\underline{\nu},\underline{d})\,]^T[\,\underline{d} - \underline{\underline{H}}(\underline{\nu})\,\overset{\circ}{\underline{\theta}}_{ML}(\underline{\nu},\underline{d})\,] \\ &= \underset{\underline{\nu}}{min}\;[\,\underline{d} - \underline{\underline{H}}_{\underline{\nu}}\left(\underline{\underline{H}}_{\underline{\nu}}^T\underline{\underline{H}}_{\underline{\nu}}\right)^{-1}\underline{\underline{H}}_{\underline{\nu}}^T\underline{d}\,]^T[\,\underline{d} - \underline{\underline{H}}_{\underline{\nu}}\left(\underline{\underline{H}}_{\underline{\nu}}^T\underline{\underline{H}}_{\underline{\nu}}\right)^{-1}\underline{\underline{H}}_{\underline{\nu}}^T\underline{d}\,] \\ &= \underset{\underline{\nu}}{min}\;\underline{d}^T[\,\underline{\underline{Id}} - \underline{\underline{H}}_{\underline{\nu}}\left(\underline{\underline{H}}_{\underline{\nu}}^T\underline{\underline{H}}_{\underline{\nu}}\right)^{-1}\underline{\underline{H}}_{\underline{\nu}}^T\,]^T[\,\underline{\underline{Id}} - \underline{\underline{H}}_{\underline{\nu}}\left(\underline{\underline{H}}_{\underline{\nu}}^T\underline{\underline{H}}_{\underline{\nu}}\right)^{-1}\underline{\underline{H}}_{\underline{\nu}}^T\,]\underline{d} \\ &= \underset{\underline{\nu}}{min}\;\underline{d}^T\underline{d} - \underline{d}^T\underline{\underline{H}}_{\underline{\nu}}\left(\underline{\underline{H}}_{\underline{\nu}}^T\underline{\underline{H}}_{\underline{\nu}}\right)^{-1}\underline{\underline{H}}_{\underline{\nu}}^T\underline{d} \end{aligned}$$

Therefore :

$$\begin{aligned} \overset{\circ}{\underline{\nu}}_{ML}(\underline{d}) &= \underset{\underline{\nu}}{argmax}\;\left(\underline{\underline{H}}_{\underline{\nu}}^T\underline{d}\right)^T\left(\underline{\underline{H}}_{\underline{\nu}}^T\underline{\underline{H}}_{\underline{\nu}}\right)^{-1}\left(\underline{\underline{H}}_{\underline{\nu}}^T\underline{d}\right) \\ \overset{\circ}{\underline{\theta}}_{ML}(\underline{d}) &= \left(\underline{\underline{H}}(\overset{\circ}{\underline{\nu}}_{ML})^T\underline{\underline{H}}(\overset{\circ}{\underline{\nu}}_{ML})\right)^{-1}\underline{\underline{H}}(\overset{\circ}{\underline{\nu}}_{ML})^T\underline{d} \end{aligned} \tag{12}$$

This rewriting of the optimisation problem was an unsuccessful attempt to simplify the solution of the optimization problem. Since the formulation in eq. (10,11) is easier to compute with $\underline{\nu}$ a 4-parameters vector, we will use it for the implementation of the optimization algorithm.

Question 3 :

As we are considering the non-linear case of ML estimation, we do not have an explicit expression for the bias and the variance of the maximum likelihood estimator.

Question 4 :

For the initial state of a, one can choose the maximum luminosity over the pixels, and for the initial value of (r_0,c_0) the maxima of luminosity. For the initial background, one can choose the minimal luminosity over the pixels.

We will work with the arbitrary initial values $\alpha = 2$, $\beta = 2$. We get the model shown in fig. 4.

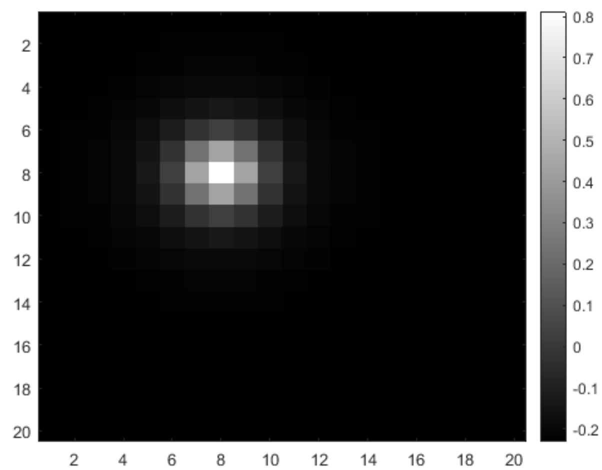

Figure 4: This is the image corresponding to our initial model.

Question 6 :

Since the research space is not explicitly bounded, it contains non physical parameters θ and ν. We cannot assume that the minimum over this complete space will be found in a physically acceptable point. To do so, one must define explicit constraints on the research space. This constraints can be integrated in the function we are trying to minimize through Lagrangian formalism in order to solve an unconstrained optimization problem (over the Lagrangian). One could also introduce a penalisation term to prevent the research algorithm from exploring a forbidden domain.

Question 7 :

	Initial	Estimated	Original
a	1.2053	1.3760	1
b	0.0190	0.1810	0.3
r_0	8	8.4313	8.5
c_0	8	8.3815	8.5
α	2	0.7921	4
β	2	0.4850	3

Table 2: Table of comparison between original and estimated parameters

The estimated parameters seem to be pretty far from the original ones, except for the position (r_0, c_0) which is nearly well captured. The luminosity factors (a, b) are far less good than the ones obtained knowing the parameters of the PSF (position and factors (α, β)) in part 1. There is an error of approximately 40% on a and b which is quite huge compared to the noise to signal ratio of 10%. The last factors (α, β) are the most incorrect ones. The parameter β is even non-physical (we expect $\beta > 1$), which is allowed by the fact we did not introduced any constraint nor penalization term to prevent the optimisation algorithm from exploring this part of the parameters-space.

Question 8 :

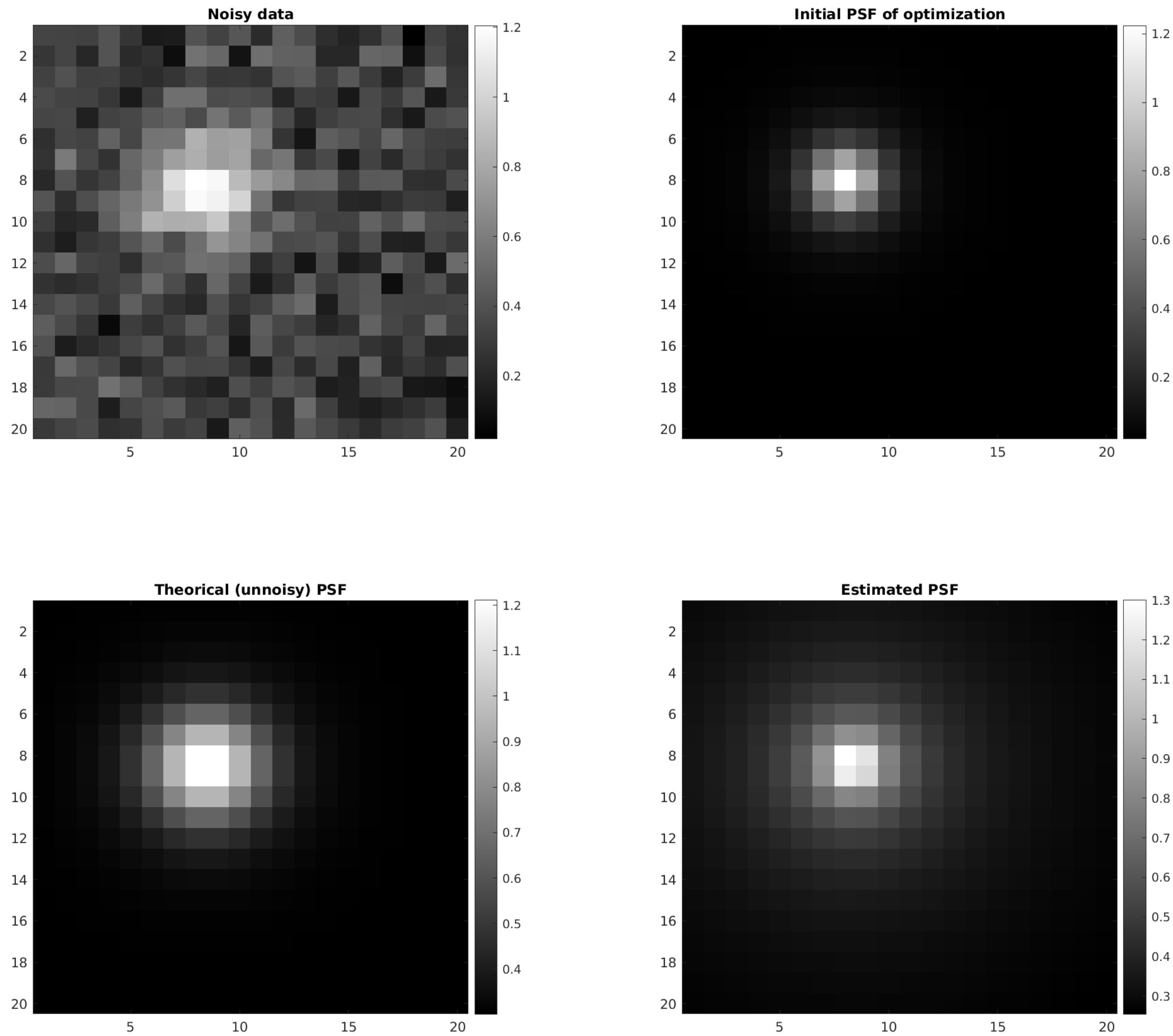

Figure 5: Noisy data (top left), Initial model (top right), Original unnoisy PSF (bottom left) and the estimated one (bottom right).

Despite the errors on the parameters, the visual reconstruction is quite good (hopefully). Indeed, we are maximizing the probability of having obtained the noisy data from an estimated PSF, it would have been very surprising to find something visually far from the theorical PSF. The estimated PSF is a little bit more spread but is a satisfying approximation of the theorical one. The qualitative results (the image) are better than the quantitative ones (the estimated parameters).

Question 9 :

The image of the residuals does not show any trace of the PSF :

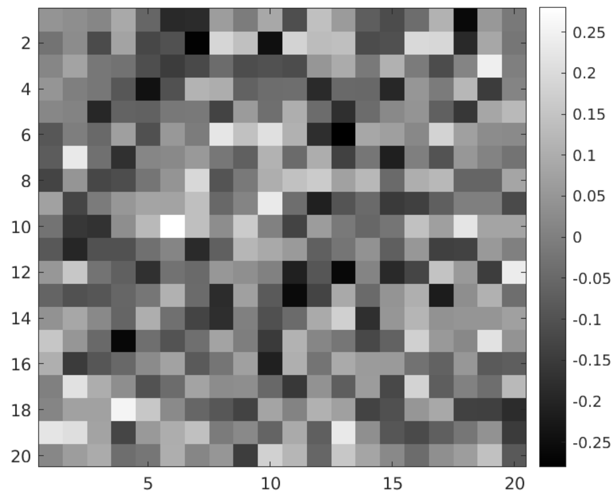

Figure 6: Residuals for the ML estimated PSF with `fminsearch`

We find the following distribution of the residuals :

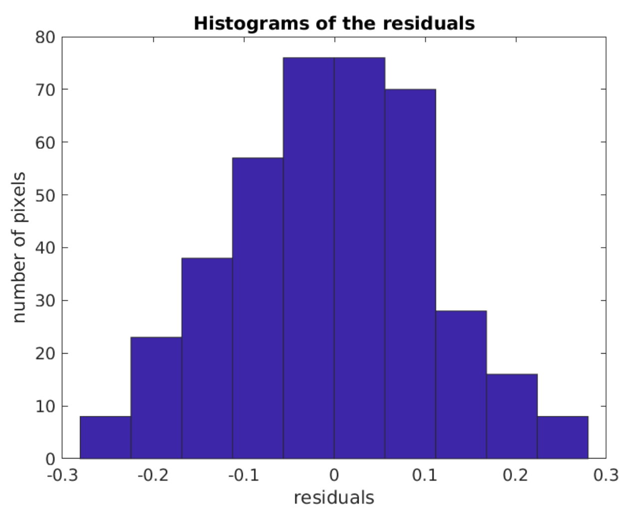

Figure 7: Histogram of the residuals for the ML estimated PSF with `fminsearch`

with a standard deviation of 0.1097 which is coherent with the standard deviation of the noise which has been introduced ($\sigma = 0.1$).

Question 10 :

The results have been summarized in table 3. The first three set of parameters seem to converge to the same local minimum (different from the one we found in question 2 which was obtained with the initialization detailed in question 2). The optimization is then sensitive to initial values but not so much. The fourth set of parameters is a extreme case which converge to a completely different estimation due to the initialization.

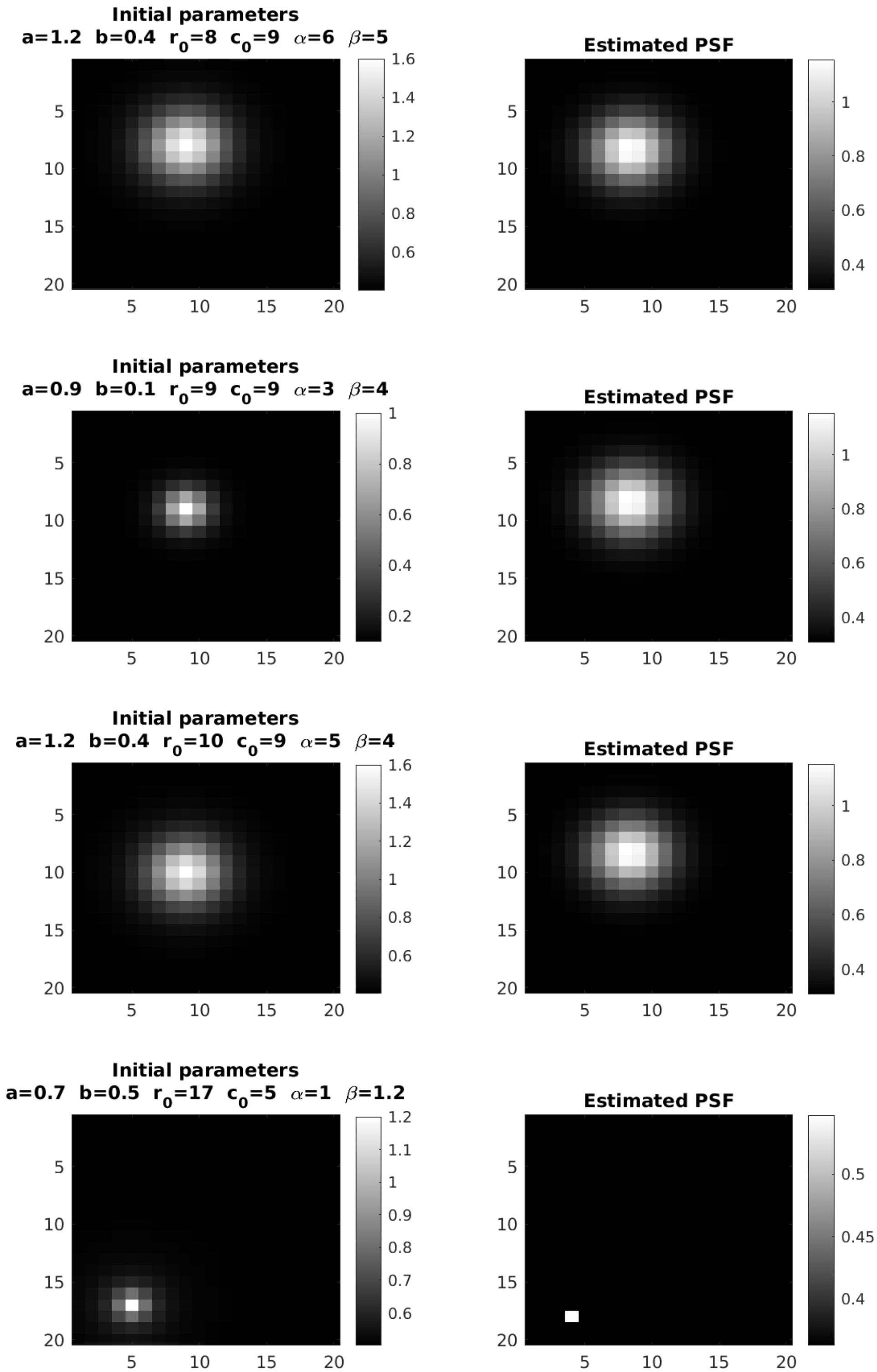

Figure 8: Comparison for different initializations

	Initial	Estimated	Initial	Estimated	Initial	Estimated	Initial	Estimated	Original
a	1.2	0.8803	0.9	0.8803	1.2	0.8803	0.7	0.2986	1
b	0.4	0.3069	0.1	0.3069	0.4	0.3069	0.5	0.3626	0.3
r_0	8	8.3998	9	8.3998	10	8.3998	17	17.9943	8.5
c_0	9	8.3662	9	8.3662	10	8.3662	5	4.0143	8.5
α	6	9.1114	3	9.1119	5	9.1114	1	0.0396	4
β	5	11.2193	4	11.2205	4	11.2194	1.2	3.0769	3

Table 3: Table of comparison of estimated parameters with different initialization

3 Estimation in the Bayesian framework

3.1 Maximum *a posteriori* estimate

Question 1 :

The posterior distribution $f(\underline{p}|\underline{d})$ is equal to the likelihood of the data given the parameters, multiplied by the prior on the parameters, and re-normalized. Hence we find the following expression :

$$\begin{aligned} f(\underline{p}|\underline{d}) &= \frac{f(\underline{d}|\underline{p})f(\underline{p})}{f(\underline{d})} \\ &\propto \begin{cases} \frac{1}{(2\pi\sigma^2)^{RC/2}} exp\left(-\frac{(\underline{d}-\underline{\underline{H}}_\nu\underline{\theta})^T(\underline{d}-\underline{\underline{H}}_\nu\underline{\theta})}{2\sigma^2}\right) & \text{if } 5 < r, c < 15 \text{ and } 0 < \alpha \text{ and } 1 < \beta \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{13}$$

We cannot give the proportionality coefficient unless we define properly a prior on α, β and a probability distribution over d.

From there, we immediately find that the cost function to minimize is :

$$J(\underline{\theta},\underline{\nu}) = \begin{cases} (\underline{d}-\underline{\underline{H}}_\nu\underline{\theta})^T(\underline{d}-\underline{\underline{H}}_\nu\underline{\theta}) & \text{if } 5 < r, c < 15 \text{ and } 0 < \alpha \text{ and } 1 < \beta \\ +\infty & \text{otherwise} \end{cases} \tag{14}$$

The main difference with the maximum likelihood estimation method is that the new optimisation problem is constrained.

3.2 Posterior mean estimator

Question 1 :

From the form of the posterior distribution in eq. (13), we find :

$$f(\underline{\theta},\underline{\nu}) \propto \begin{cases} exp(-\frac{J(\underline{\theta},\underline{\nu})_{MLL}}{2\sigma^2}) & \text{if } 5 < r, c < 15 \text{ and } 0 < \alpha \text{ and } 1 < \beta \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

Once again, we cannot give the coefficient of proportionality without defining a valid prior on α and β, and a probability distribution on d. But this is not an issue in the Metropolis-Hastings algorithm, as it gets rid of the multiplicative constants during the acceptation step.

This posterior distribution yields the following logarithm :

$$\log\left(f(\underline{\theta},\underline{\nu})\right) = \begin{cases} -\frac{J(\underline{\theta},\underline{\nu})_{MLL}}{2\sigma^2} + C^{st} & \text{if } 5 < r, c < 15 \text{ and } 0 < \alpha \text{ and } 1 < \beta \\ -\infty & \text{otherwise} \end{cases} \tag{16}$$

Question 2 :

We wrote a function that performs a Metropolis-Hastings following the guidelines in the problem sheet. We also chose to represent the prior in the acceptance step not in the cost function, but by a Boolean test. These allowed us to gain some computation time, and also to avoid dealing with infinite values.

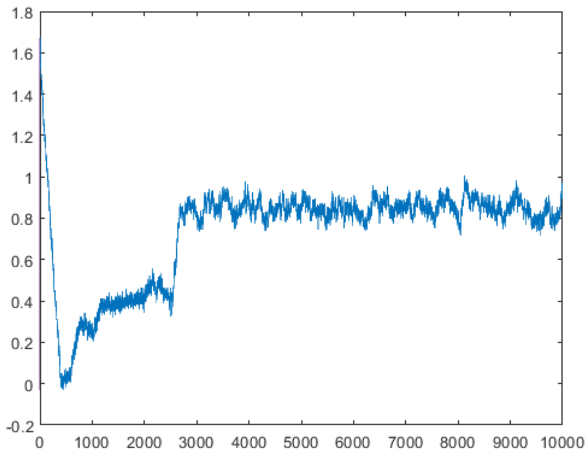

Figure 9: a

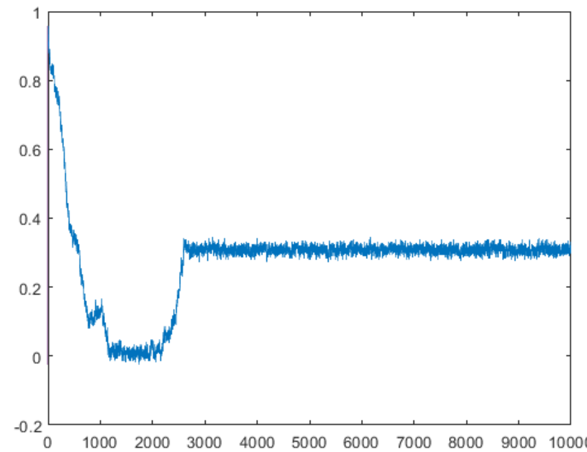

Figure 10: b

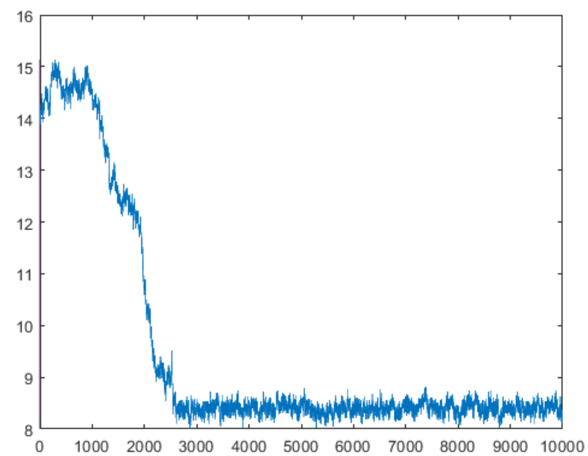

Figure 11: r_0

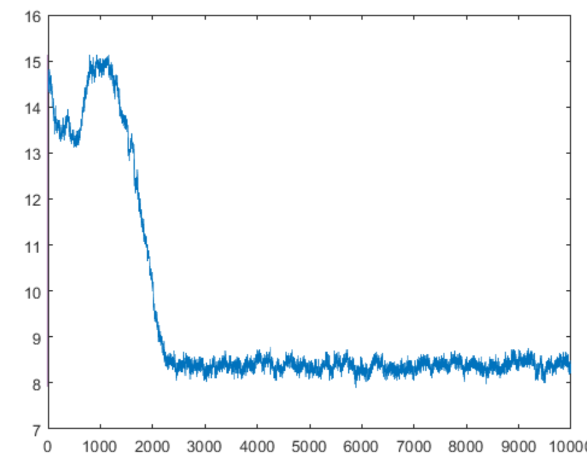

Figure 12: c_0

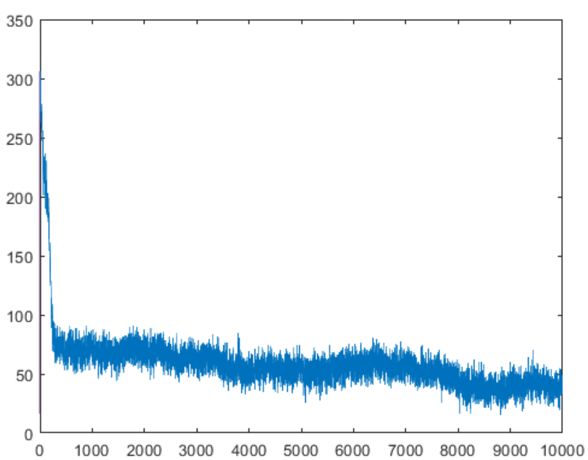

Figure 13: α

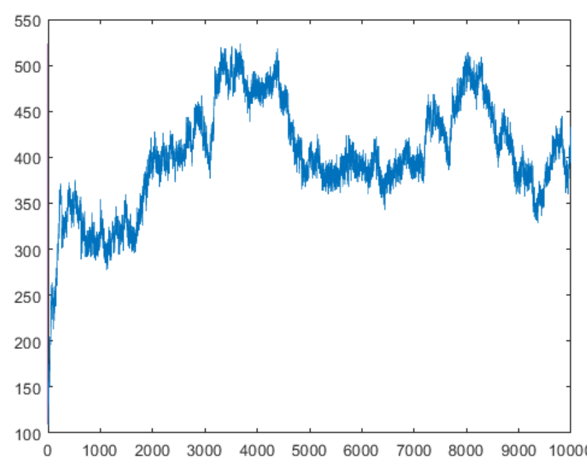

Figure 14: β

Question 3 :

We plot the evolution of the samples with the number of iterations in fig. 9 to 14.
We find that $\tau = 0.007$ is a good value to have an acceptance rate around 20%.

And we find that the algorithm converges to a distribution centered around the expected values for a, b, r_0 and c_0 after an initialization phase that lasts between 2000 and 3000 steps.
However, we find that the algorithm does not converge in distribution at all for parameters α and β. Our interpretation is that the prior interval for these parameters was taken far too large : we took a true value around 1, i.e. 3 orders of magnitude below the prior's standard deviation. When we are very far off, any value of these parameters gives the same output. Hence unless the initialisation step is lucky enough to start with a right value for these parameters, the algorithm has no chance to converge to the right value. Or actually, it will do so incredibly slowly : this is the slow slope we observe for α.
This non-convergence for α and β is a big issue, as it also lowers the quality of the convergence for the 4 other parameters...

Question 4 :

The approach we described here guarantees that we find physically realistic values for the parameters, even α and β, because our prior sets the posterior probability of any parameter at 0 everywhere it is not physically plausible.

We find the following results for the parameters :

Parameter	Value	Estimation
a	1.0	0.8545
b	0.3	0.3087
r_0	8.5	8.4034
c_0	8.5	8.3496
α	4	N.A.
β	3	N.A.

Table 4: Table of comparison between original and estimated parameters

These estimations are much less precise than in the ML case with known star's location and PSF parameters : we reach a 15% error on a, we we used to have a 2% precision. This is not unexpected : we have more parameters to estimate conjointly, hence the problem is much harder. Also, we failed to estimate the PSF parameters, and this reduce the ability of the MCMC to converge precisely to the right values for the other parameters.

Question 5 :

We ran the MCMC pipeline several time and found sensibly the same estimations for parameters a, b, r_0 and c_0. Hence we do not see think there is a sensitivity to the initialization (in the MCMC method) for these 4 parameters. Which was expected, the MCMC sampling is theoretically insensitive to the initialization, provided that the burn-in has been performed properly.

Question 6 :

We regroup in fig. 15 to 17 the images of the estimated stars for the 3 methods we explored.
We find that the estimated image from the MCMC method has a similar background to both former estimations, a maximum amplitude that is slightly lower than both estimates, and spreads further. This last observation is once again due to our failure to estimate the α and β parameters.

Question 7 :

Once again, the image of the residual does not show traces of the PSF (see fig. 18).

We also find the distribution for the residuals presented in fig. 19, with a standard deviation $\sigma_{mcmc} = 0.1038$ that is quite close to the true noise $\sigma = 0.1$ we used :
This distribution seems to be quite well centered, however it is hard to conclude whether it is Gaussian or not. We can

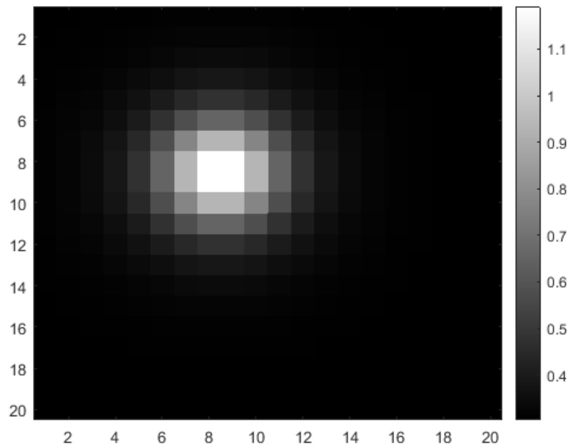

Figure 15: Estimated star's image : ML estimator with known star's location and PSF parameters

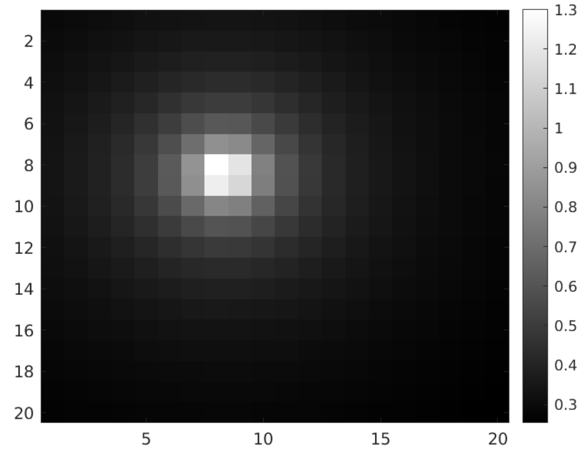

Figure 16: Estimated star's image : complete ML estimator

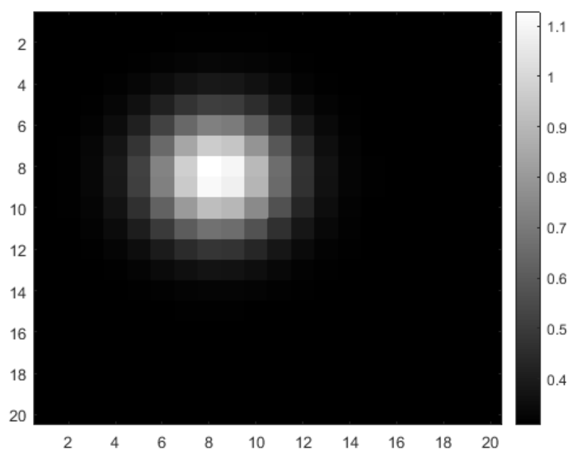

Figure 17: Estimated star's image : Bayesian estimator

only say that these residuals are not obviously in disagreement with the additive Gaussian centered noise assumption we made earlier.

Question 8 :

We estimate the standard deviation of the estimators by taking the standard deviation of the posterior distribution for each parameter. We find the results in tab. 3.2:

Parameter	Estimated standard deviation	Error
a	0.0410	0.1539
b	0.0104	0.0084
r_0	0.1154	0.1165
c_0	0.1227	0.1372

Table 5: Table of comparison between the estimated standard deviation and the reported errors.

By comparing the estimated standard deviation with the reported errors, we find that these estimated values seem realistic, except on a where we would be beyond 3σ. We have a single test so this could be due to chance, but most likely not.

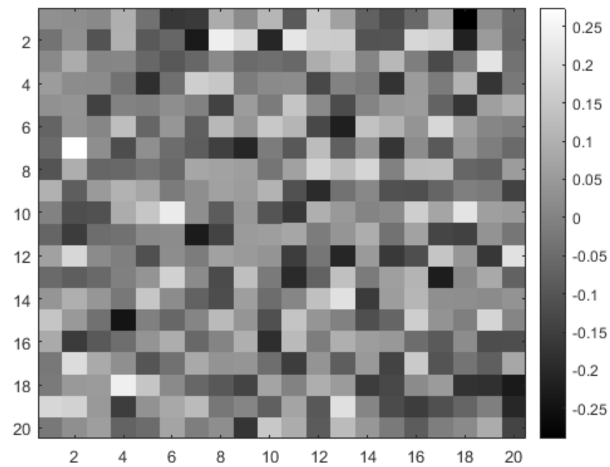

Figure 18: Residuals for the MCMC estimation. We do not see traces of the PSF on the image.

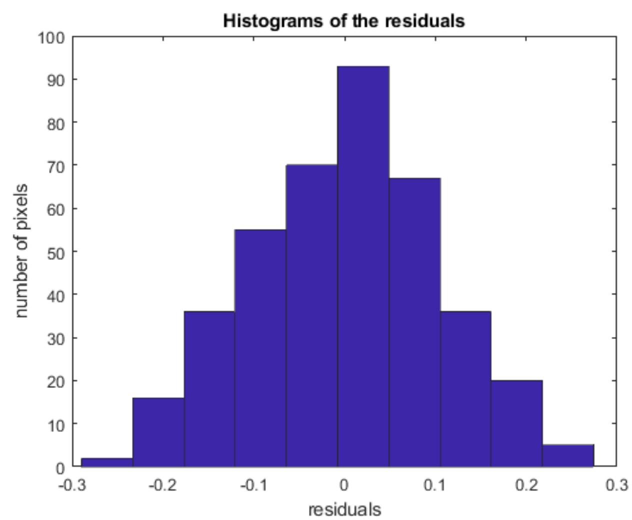

Figure 19: Histogram for the residuals in the case of MCMC estimation. The distribution seems to be centered. It is hard to say whether it is Gaussian or not.

However, we have no explanation for this phenomenon : this estimator computed from the posterior distribution has no reason (that we know of) to underestimate the uncertainty from the estimator.

Question 9 :

For this question, we use the diagonal plots of fig. 20. We find that the posterior marginal distributions for b, r_0 and c_0 have a seemingly Gaussian shape with a peak that is centered on the expected value. Hence our Bayesian pipeline will be able to predict a star position and the level of background.

Then, the distribution for a seems Gaussian as well, but it is not centered at all on the expected value. We are not quite sure how this could have happened : the Markov chain has explored the correct value, so it is not a problem of local minima and duration of exploration. This means that the minimum of the cost function is indeed around $a = 0.85$, at least for the values of α and β the MCMC has explored. Anyway, this means that our Bayesian pipeline will not be able to estimate the star's intensity. This is inacceptable, as the magnitude is often the most interesting quantity when considering a star.

And finally, the distribution for α and even more so β do not have a Gaussian shape and are not centered anywhere close the correct value. Here, we think it is because the correct value has not been explored at all during the MCMC, because the prior region is too large so we need luck to explore the right region. Hence, our pipeline will not be able to estimate the atmosphere's influence on the PSF.

Question 10 :

We find the following covariance matrix, after renormalization :

$$\begin{pmatrix} 1 & -0.0037 & 0.0234 & 0.0446 & -0.873 & 0.1203 \\ & 1 & 0.0020 & -0.0319 & -0.0749 & -0.208 \\ & & 1 & -0.0418 & 0.0101 & 0.0266 \\ & & & 1 & 0.0531 & 0.0897 \\ & & & & 1 & 0.5556 \\ & & & & & 1 \end{pmatrix} \tag{17}$$

From this matrix, we should be able to deduce the estimations of which pairs of parameters are most correlated.

We find that the correlation coefficients of the 3 parameters that are correctly estimated are all small (below 5%). This means that the error on these 3 estimators are nearly independent (pairwise !).
We find that the correlation coefficient between a and α is 0.87. This is extremely high, in fact nearly half of the uncertainty on a comes from the uncertainty on α. Here we confirm what we said earlier, that failing to estimate α causes an imprecision on the other dimensions.
Leading to the same conclusion, we also observe that β is correlated with a and b at a non-negligible level (above 10%).
Finally, we find that the samples for α and β are very correlated. We do not know what that means.

Question 11 :

We get the corner plot in fig. 20 :

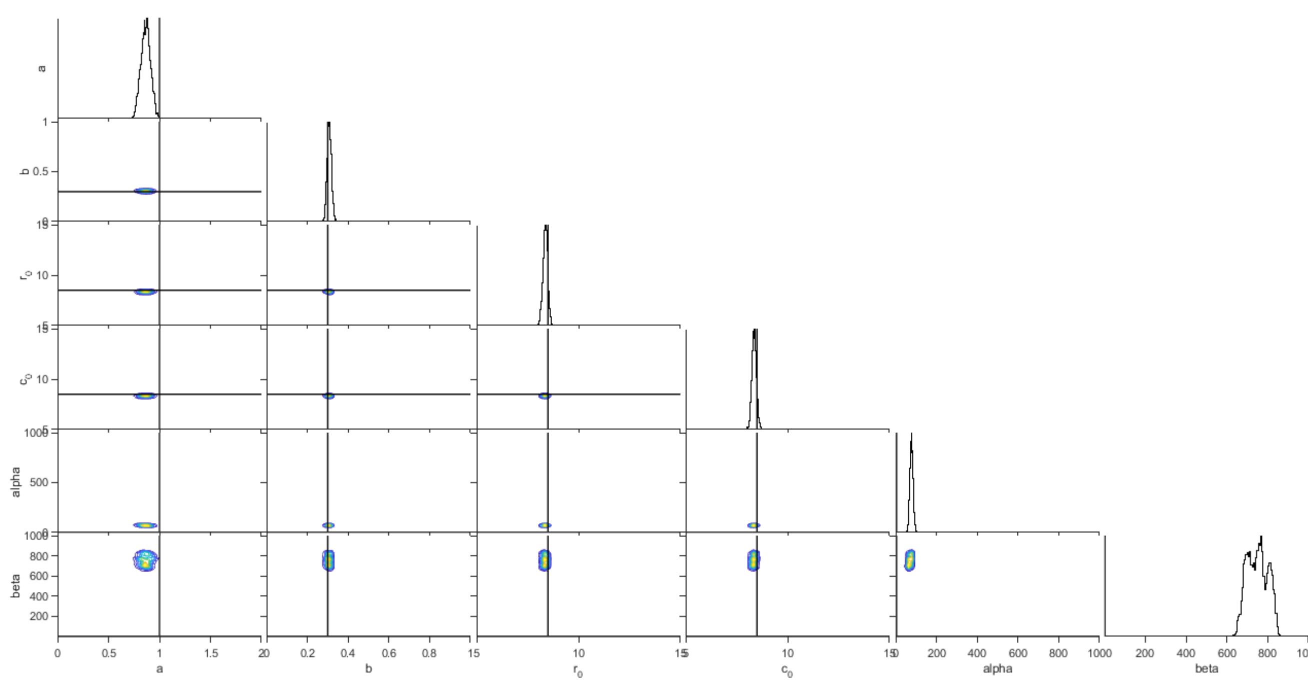

Figure 20: Corner plot for the posterior distribution of the 6 parameters. The diagonal plot represent the 1D marginal posterior distribution, and the sub-diagonal plots the 2D marginal posterior distributions. The black lines represent the true values of the parameters.

We will only consider the 2D marginal plots of the parameters that converged to the right value, as the other parameters

are not reliable anyway.
We observe that these 3 2D plots have an ellipsoidal shape, with the particular directions along the x and y axes of the plots. These particular direction means that the errors of the 3 estimators are (sensibly) independent. And the ellipsoidal shape could confirm that the parameters follow a joint Gaussian law, but it is hard to say if we see a pure ellipse or not.

4 Conclusion

In this laboratory, we have experienced different solutions to extract data from a noisy picture of a star searching in the space of the possible PSF (*Point Spread Functions*). The parameters we wanted to determine were the luminosity factors (a, b) of the star and the background, the position (r_0, c_0) of the star and the parameters (α, β) of the PSF which give together the theoretical luminosity (without noise):

$$\begin{aligned} m((r,c)\,;\,(a,b,r_0,c_0,\alpha,\beta)) &= a \times PSF((r,c)\,;\,(r_0,c_0,\alpha,\beta)) + b \\ &= a \times \left(1 + \frac{(r-r_0)^2+(c-c_0)^2}{\alpha^2}\right)^{-\beta} + b \end{aligned}$$

In fact, a noise $\underline{\epsilon}$ which we suppose to be centered gaussian is introduced at the measurement. Therefore, there is no set of parameter that fit perfectly the data measured, we are trying, in this laboratory to find the set of parameters which is the most likely to produced such a measurement, making the hypothesis that the noise in centered gaussian (fundamental hypothesis which allow us to define precisely what we call "likely to produced such a measurement"). We first explored a case where we know the four parameters linked to the PSF $(r_0, c_0, \alpha, \beta)$ and where we are trying to estimate the luminosity factors (a, b). Since the dependency of the measurement on this factors is linear, we were able to explicit the maximum likelihood estimator (ML estimator). Then, the numerical estimation on our particular case is only sensible to computation precision and to the validity of the hypothesis of centered gaussian noise (since there is a finite number of pixel/measure the true residuals - hence the noise - cannot be perfectly gaussian). We plot the true residuals and their histogram in figures 21 and 22.

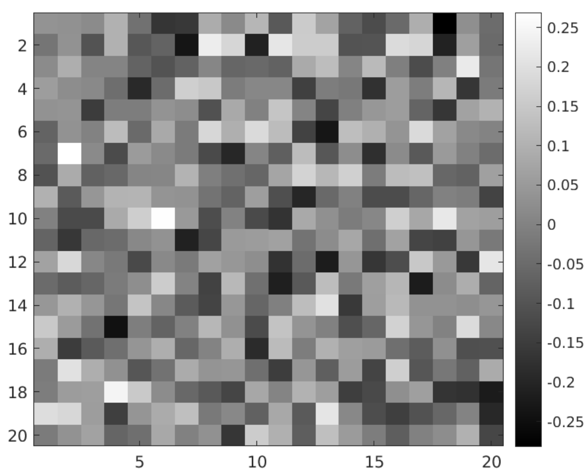

Figure 21: True residuals

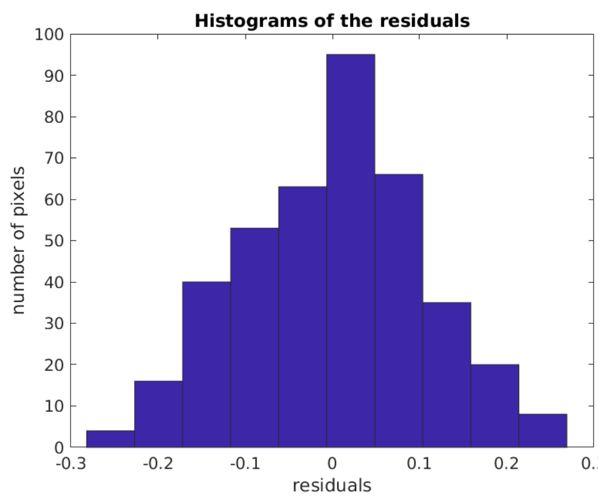

Figure 22: Histogram of the true residuals

We note that the distribution of the noise (true residuals) is not perfectly centered. With the computation of the analytical solution (with 4 known parameters) we have a relative error of 2.25% on the luminosity factors. And the image is very well reconstructed, it is difficult to find a difference between the estimated residuals (19 and 18) and the true one (21 and 22).
Then, we tried to estimate the all set of parameters, first without *a priori* on them by maximizing the likehood (minimizing the associated cost function) thanks to the function `fminsearch`, and after with a bounded uniform *a priori* on the parameters using Metropolis-Hastings algorithm.
The first solution gives good qualitative results with a short computation time but quantitatively, the estimated parameters are inaccurate. Except for the position which is determine with a relative error of $1-2\%$, we have huge relative errors for the luminosity factors $\sim 40\%$ and completely false PSF factors. Furthermore, this method can return non-physical parameters since did not introduced any constraint or penalization term. The second one gives better results, especially on

the estimation of the parameter b. This is particularly visible on the distribution of residuals (and the image of residuals) which are very similar to the true residuals. Furthermore, this method take into account the physical constraint on the parameters. Nevertheless, this method still doesn't capture correctly the star luminosity (15% relative error) even if it is better than the first method. And this method encounters huge difficulties with the PSF parameters (α, β) for which it does not give a consistent result.
The two solutions we explored to estimate the all 6 parameters $(a, b, r_0, c_0, \alpha, \beta)$ have been struggling with the PSF parameters (α, β). And the estimation of the star's luminosity is not nearly satisfying knowing the utility of this parameter in the study of a star.