Estimation laboratory

You have to prepare the subject before the laboratory.

Questions with the † mark should be solved before the laboratory.

The listings of all programs have to be given with your laboratory report.

All results and graphs have to be explained.

I Introduction

The aim of this laboratory is to implement and compare various approaches for the estimation of parameters from data. In practice, we want to estimate the parameters of a simple model for the flux of a galaxy from a noisy image of a galaxy.

The studied approaches for estimation are least squares estimation, maximum likelihood estimation, and Bayesian estimation in the maximum *a posteriori* and *posterior* mean sense. For the optimization we will use the `fminsearch` Matlab function while for computing the *posterior* mean estimator we will implement a stochastic sampling method, more precisely a Metropolis-Hastings algorithm.

II Modeling the problem

From the 1960s, the Sersic profile[1] is often used by astrophysicist as a simple model for the flux of elliptic galaxies observed with a ground-based telescope. The Sersic profile describes the variation of intensity of the galaxy with respect to the distance to his center.

$$I(\ell, c) = \exp(-R(\ell, c)^{\frac{1}{n}})$$

with a shape parameter n, typically in the range $[\frac{1}{2}, 5]$, $n = \frac{1}{2}$ corresponding to a Gaussian model. In practice, for a pixel with coordinates (ℓ, c) in the image, the distance R depends on the coordinates of the galaxy's center (ℓ_0, c_0), on the two galaxy's axes length (σ_ℓ, σ_c) and on the horizontal angle α:

$$R(\ell, c)^2 = \left(\frac{(\ell-\ell_0)\sin(\alpha) - (c-c_0)\cos(\alpha)}{\sigma_\ell}\right)^2 + \left(\frac{(\ell-\ell_0)\cos(\alpha) + (c-c_0)\sin(\alpha)}{\sigma_c}\right)^2 \quad (1)$$

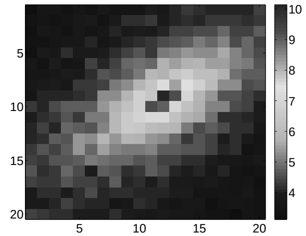
The observed data can thus be modeled as:

$$d(\ell, c) = s + a \cdot I(\ell, c) + n(\ell, c), \quad (2)$$

where a and s correspond respectively to the amplitudes of the galaxy and the sky's background and $n(\ell, c)$ corresponds to the measurements and model's noise. To simplify the problem, we will consider hereafter a white Gaussian noise with known variance σ_n^2. Note that the location and length parameters ℓ_0, c_0, σ_ℓ and σ_c are given in pixels unit and can of course take subpixels values.

The Matlab function `Sersic` alows to compute the image of a galaxy with unit amplitude for a given set of parameters, from rows-wise and column-wise grids, respectively `Lin` and `Col`, defined with the Matlab code (with `L` and `C` the number of rows and columns of the considered image):

```
lin = 1:L; col = 1:C;
[Col,Lin] = meshgrid(col,lin);
nu = [l0; c0; sigma_l; sigma_c; alpha; n];
Gal = Sersic(nu,Lin,Col);
```



1. Choose a set of parameters and noise standard deviation to obtain an image similar to the opposite image.

This image will be considered hereafter as your data set and the chosen parameters as the theoretical parameters to estimate.

It will be useful to distinguish the data in the form of an image $\boldsymbol{D}$ (a L times C matrix), from the column vector $\boldsymbol{d}$ composed with the LC pixels of the image. It is straightforward to switch from one notation to the other with the Matlab codes: `d = D(:);` and `D = reshape(d,L,C);`.

III Estimation with known Galaxy's location and shape parameters

We will consider, in a first time, the estimation of the galaxy's and sky's background amplitudes for known location and shape parameters of the galaxy $(\ell_0, c_0, \sigma_\ell, \sigma_c, \alpha)$.

1.† Let $\boldsymbol{\theta}$ be the vector of parameters to estimate $\boldsymbol{\theta} = [s, a]^T$. Give an expression, with a matrix form, of the model of equation (2), for a known matrix $\mathbf{H}$, and a set of data $\boldsymbol{d}$ specifying the dimensions and components of matrices and vectors of such a model.

2.† Give an expression of the likelihood and of the maximum likelihood estimator of parameters $\boldsymbol{\theta}$ as a function of the data $\boldsymbol{d}$.

3.† Calculate the bias and the covariance matrix of this estimator.

4. With Matlab, for the given data set $\boldsymbol{d}$ of section II, compute the estimated parameters as well as their variance.

5. Compare the estimated parameters with the theoretical ones. Comment these results...

IV Maximum likelihood estimation of all the parameters

We will now consider that the location and shape parameters of the galaxy are unknown, i.e. that parameters $\boldsymbol{\nu} = [\ell_0, c_0, \sigma_\ell, \sigma_c, \alpha, n]^T$ are unknown and have to be estimated simultaneously with parameters $\boldsymbol{\theta}$.

[1] J. L. Sérsic (1963), *Influence of the atmospheric and instrumental dispersion on the brightness distribution in a galaxy*, Boletin de la Asociacion Argentina de Astronomia, Vol. 6, p.41.

1.† Give an expression, with a matrix form, of the corresponding model, as a function of parameters $\boldsymbol{\theta}$ and $\boldsymbol{\nu}$, matrix $\mathbf{H}(\boldsymbol{\nu})$ being a function of parameters $\boldsymbol{\nu}$, specifying the dimensions and components of matrices and vectors of a such model.

2.† Give an expression of the joint likelihood of parameters $\boldsymbol{\theta}$ and $\boldsymbol{\nu}$ (i.e. the probability distribution of the data when these parameters are known) and of the cost function $J(\boldsymbol{\theta}, \boldsymbol{\nu})$ to minimize to compute the maximum likelihood estimator. Do we have an explicit expression for this estimator? What kind of optimization problem are we faced with?

3.† Do we have an expression for the bias and the variance of this maximum likelihood estimator?

4.† Propose a "very simple" way to initialize parameters a, s, ℓ_0, and c_0 from the data. For parameters σ_ℓ, σ_c, α and n we can fix an arbitrary "reasonable" value. Show the image corresponding to this initial model.

5. Write a Matlab function `crit_J` which computes the cost function to minimize with respect to parameters $\boldsymbol{p} = [\boldsymbol{\theta}^T, \boldsymbol{\nu}^T]^T$: `J = crit_J(p,d);`

We will minimize this cost function with the Matlab function `fminsearch` for initial parameters value `p_init`: `p_opt = fminsearch(@(p) Crit_J(p,d), p_init,options);` The variable `options` helps to monitor the optimization algorithm. It is set to its default value with: `options = optimset('fminsearch');` you can modify some components of such a variable if necessary...

6.† Are we guaranteed with such an approach to obtain physically realistic parameters $\boldsymbol{\theta}$ and $\boldsymbol{\nu}$, in particular $n \in [\frac{1}{2}, 5]$? What should you do in practice for that?

7. Compare the estimated parameters to the theoretical ones. Comment these results...

8. Compare the image of the estimated galaxy to the original (unnoisy) one. Comment these results...

9. Show the image of the residuals (model for the estimated parameters subtracted from the – noisy – data). Estimate the standard deviation of the residuals and show their histogram. Are this residuals compatible with the considered model for the noise?

10. Test this optimization process for various initial values for the parameters. Are the obtained results sensitive to such an initial value?

V Estimation in the Bayesian framework

The Bayesian framework allows to account for prior information on the parameters to estimate, through the prior distribution of these parameters. We will focus hereafter to the joint estimation of all parameters $\boldsymbol{p}$.

We consider that we have only a few prior information on the parameters. Therefore, all the parameters are considered to have a uniform prior distribution on "reasonable" intervals that can be directly deduced from the data. In particular, the data have been selected such

that the galaxy's center is located in the central quarter of the image, parameters (σ_ℓ, σ_c) have a uniform distribution on $[0, \max(\frac{L}{2}, \frac{C}{2})]^2$, α has a uniform distribution on $[0, \pi/2[$ and n a uniform distribution on $[\frac{1}{2}, 5]$. The sky's background and galaxy's amplitude takes positive values, lower that the maximum value of the data and to twice this maximum value respectively.

V.1 Maximum *a posteriori* estimator

1.† Give an expression for the *a posteriori* distribution $f(\boldsymbol{p}|\boldsymbol{d})$ and of the cost function to minimize to compute the maximum *a posteriori* estimator. What is the difference of such an optimization problem compared to the maximum likelihood estimation one?

We will not try to implement such an estimation problem hereafter. Indeed, the considered prior information is to weak to give better results than the maximum likelihood one.

V.2 Posterior mean estimator

We will implement an MCMC (Markov Chain Monte-Carlo) stochastic sampling algorithm to sample the posterior distribution (i.e. to generate realizations of a random vector having this distribution). The estimation step is then straightforward:

- Generate samples $\{\boldsymbol{p}^{(k)}\}_{k=1\ldots K}$ distributed with the posterior distribution $f(\boldsymbol{p}|\boldsymbol{d})$.
- Compute the posterior mean estimator as the samples means $\mathring{\boldsymbol{p}} = \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{p}^{(k)}$.

Note that we can also estimate easily from the samples the variance, the correlation and the marginal posterior probabilities (using the histograms) of the parameters from the samples. The main difficulty of such an estimation scheme is the sampling algorithm. We will implement a generic Metropolis-Hastings algorithm, whose principle is presented Tab 1.

Init. $k = 0$	Generate a initial parameter set $\boldsymbol{p}^{(0)}$ with the *a priori* distribution.
Iterations $k = 1 \ldots K$	*a)* Generate a random parameters set $\boldsymbol{p}'$ with the proposal distribution $Q(\boldsymbol{p}'; \boldsymbol{p}^{(k-1)})$. *b)* Generate a random variable u uniformly distributed on the interval $[0, 1]$ $\boldsymbol{p}^{(k)} = \begin{cases} \boldsymbol{p}' & \text{if } u < \dfrac{f(\boldsymbol{p}')Q(\boldsymbol{p}^{(k-1)}; \boldsymbol{p}')}{f(\boldsymbol{p}^{(k-1)})Q(\boldsymbol{p}'; \boldsymbol{p}^{(k-1)})}, \text{ (acceptation)} \\ \boldsymbol{p}^{(k-1)} & \text{otherwise.} \end{cases}$
Estimation	Suppress the K_0 first samples which depend on the initialization ("burn-in" period) and estimated the parameters from the resulting samples.

Table 1: Principle of a Metropolis-Hastings algorithm to sample the distribution $f(\boldsymbol{p})$.

As only few information is taken into account in the *a priori* distributions, we don't expect to have very different results, in terms of estimator, than with the maximum likelihood estimator (note that both estimators are strictly equivalent in theory only for uniform prior distributions and if the maximum and mean of the distribution are the same, which is not likely to happen in practice). However we expect to get rid off the problems of local minima in the cost function (the MCMC sampling is theoretically insensitive to the initialization) and to take advantage of additional information on the parameters such as their variances, correlations and shape of the *a posteriori* marginal distribution.

The cornerstone in the implementation of such an algorithm is the choice of the proposal distribution and the tuning of its parameters. We will choose a Gaussian proposal distribution: $Q(\cdot;\boldsymbol{p}^{(k-1)}) = \mathcal{N}(\boldsymbol{p}^{(k-1)};\Gamma)$ (Gaussian random walk) with a diagonal covariance matrix $\Gamma = \mathrm{diag}\{\boldsymbol{\gamma}^2\}$, where the components of vector $\boldsymbol{\gamma}$ are tuned to have an acceptance ratio around 20%. The larger are the components of $\boldsymbol{\gamma}$, the smaller will be the acceptance ratio. In practice, the components of $\boldsymbol{\gamma}$ can be set proportionally to the standard deviation of the prior distribution of each parameter.

The proposal distribution being symmetric, the acceptance test can be computed as $\log(u) < \log(f(\boldsymbol{p}'|\boldsymbol{d})) - \log(f(\boldsymbol{p}^{(k-1)}|\boldsymbol{d}))$. The logarithmic scale allows to avoid numerical instabilities problems. In practice, it can be useful to monitor the samples evolution over the iterations to tune the algorithm's parameters, that is K (the number of samples necessary for convergence), K_0 (the number of iterations of the "burn-in" period) and the components of vector $\boldsymbol{\gamma}$ (variances of the proposal distribution).

1.† Give an expression of the posterior distribution $f(\boldsymbol{p}|\boldsymbol{d})$ (up to a multiplicative constant) and of its logarithm $\log(f(\boldsymbol{p}|\boldsymbol{d}))$ (up to an additive constant) as a function of the cost function $J(\boldsymbol{\theta},\boldsymbol{\nu})$ of section IV.

2. Write a Matlab script implementing the Metropolis-Hastings algorithm using your Matlab function `crit_J`.

3. Show the samples generated with your algorithm. Analyze the tuning of the proposal distribution's parameters on the samples and on the acceptance ratio to find a correct trade-off... How many iterations seems to be necessary to reach the convergence toward a stationary distribution of the algorithm?

4. Compare the estimated parameters to the theoretical ones and to those estimated with the previous methods. Comment these results... Are we guaranteed with such an approach to obtain physically realistic parameters $\boldsymbol{\theta}$ and $\boldsymbol{\nu}$, in particular $n \in [\frac{1}{2}, 5]$?

5. Are the obtained results sensitive to initial values of the parameters?

6. Compare the image of the estimated galaxy to the original one and the previous estimated one. Comment these results...

7. Show the image of the residuals. Estimate the standard deviation of the residuals and show their histogram. Are this residuals compatible with the considered model for the noise?

8. Give an estimation of the estimated parameters variance. Do these values seems realistic?

9. Show the shape of the posterior marginal distribution estimated with the histogram of the samples. Comment these results...

10. How can you study the correlation between the estimated parameters? What information can that brings?

11. Show as a scatter graph the estimated parameters in pairs. How can you interpret theses scatter graphs in terms of correlation? What information can that brings in general and on your samples?

VI Conclusion

Conclude on the various estimation approaches used during this laboratory for the estimation of parameters, focusing on the pros and cons of each approaches...