Practical work on image restoration

Wiener-Hunt method: unsupervised aspects

Prerequisite: prepare in advance the answer to the questions no. 1, 2, 5 and 6.

In the previous practical work you have faced the difficulties of the deconvolution problem, most notably for the case when information pertaining to the true image is missing in the observations. You have worked with the so-called Wiener-Hunt method: the coupling of a quadratic term quantifying the discrepency of the solution with respect to the data with a quadratic penalty term quantifying the roughness of the solution. It has enabled you to obtain (relatively) satisfying results. However, the downside of it is that it requires the tuning of a hyper-parameter also called the regularisation parameter. You have chosen its value empirically by means of a trial and error process such that the resulting image was neither too irregular nor too smooth. The purpose of this practical work is to present to you an approach which automatically adjusts the value of the hyper-parameter.

1 Hyper-parameters and full a posteriori distribution

The proposed method is based on a Bayesian interpretation of the Wiener-Hunt solution. The interpretation itself is based on two Gaussian probabilistic models: one for the error $\boldsymbol{e}$ and one for the image $\boldsymbol{x}$.

1.1 Error distribution

The error is modelled as white, zero-mean homogeneous Gaussian vector. For the Gaussian distribution an alternative parametrisation involving the so-called precision parameter γ_e (it is the inverse of the variance) is chosen. Its expression under this parametrisation writes:

$$f(\boldsymbol{e} \,|\, \gamma_e) = (2\pi)^{-N/2} \, \gamma_e^{N/2} \exp - \gamma_e \, ||\boldsymbol{e}||^2/2 \, .$$

This results in the following expression for the likelihood of the object of interest $\boldsymbol{x}$ attached to the data $\boldsymbol{y}$:

$$f(\boldsymbol{y} \,|\, \boldsymbol{x}, \gamma_e) = (2\pi)^{-N/2} \gamma_e^{N/2} \exp - \gamma_e \, ||\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}||^2/2 \ ,$$

due to the fact that $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{e}$. Under this expression for the likelihood, the term which quantifies the adequacy of the reconstructed object with respect to the data appears as a co-logarithm:

$$\mathcal{J}_{\text{LS}}(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2 = -k_{\boldsymbol{y}} \log f(\boldsymbol{y} \,|\, \boldsymbol{x}, \gamma_e) + C_{\boldsymbol{y}}$$

up to an additive and a (positive) multiplicative constants, $C_{\boldsymbol{y}}$ and $k_{\boldsymbol{y}}$.

1. *Give the corresponding expression for each of the two constants $k_{\boldsymbol{y}}$ and $C_{\boldsymbol{y}}$.*

1.2 The distribution for the object of interest

Furthermore, the Bayesian interpretation calls for a probability distribution for the object of interest $\boldsymbol{x}$. The model for it is also Gaussian, only that in this case it is not white, that is, there is a correlation between its constituent components. In the following, the correlation is actually modelled through the covariance matrix $\boldsymbol{R}$. It is written through its inverse $\boldsymbol{R}^{-1} = \gamma_x \boldsymbol{\Pi}$: the precision parameter γ_x controls the strength of the correlation whereas the matrix $\boldsymbol{\Pi}$ fixes the structure of it.

$$f(\boldsymbol{x} \,|\, \gamma_x) = (2\pi)^{-N/2} \det\big[\boldsymbol{\Pi}\big]^{1/2} \gamma_x^{N/2} \exp - \gamma_x \, \boldsymbol{x}^{\mathrm{t}} \boldsymbol{\Pi} \boldsymbol{x} / 2 \,.$$

The term which quantifies the adequacy of the object with respect to the prior information appears as the co-logarithm of the density:

$$\mathcal{J}_0(\boldsymbol{x}) = -k_{\boldsymbol{x}} \log f(\boldsymbol{x} \,|\, \gamma_x) + C_{\boldsymbol{x}} = ||\boldsymbol{x}||^2_{\boldsymbol{\Pi}} = \boldsymbol{x}^{\mathrm{t}} \boldsymbol{\Pi} \boldsymbol{x}$$

also up to an additive and a (positive) multiplicative constant. To complete the link with the Wiener-Hunt method from the previous practical work, it suffices to choose $\boldsymbol{\Pi} = \boldsymbol{D}^{\mathrm{t}} \boldsymbol{D}$.

2. *Give the corresponding expression for each of the two constants $k_{\boldsymbol{x}}$ and $C_{\boldsymbol{x}}$.*

Remark 1 — *Properly speaking the above interpretation is not correct as the matrix $\boldsymbol{D}^{\mathrm{t}} \boldsymbol{D}$ is not invertible: constant images are nothing more than eigenvectors corresponding to the eigenvalue equalling zero (which corresponds to the null frequency). A rigorous development calls for the introduction of a penalty term for the energy at the zero frequency (through a parameter that can be set to an arbitrary small value).*

The above distribution is called *a priori*, or more commonly prior, as it enables one to account for the *a priori* information required to favour images with a higher degree of regularity. The probability for a given image is higher as the image is more regular, that is, as the value of $\mathcal{J}_0(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{t}} \boldsymbol{\Pi} \boldsymbol{x}$ is smaller.

3. *Comment the impact of γ_x on the prior probability distribution of images: more or less probable versus more or less smooth.*

1.3 A posteriori (partial) distribution

With the aid of the two previously defined ingredients and using the product rule for probabilities, one can now construct the joint density for the unknown object $\boldsymbol{x}$ and the data $\boldsymbol{y}$

$$\begin{aligned} f(\boldsymbol{x}, \boldsymbol{y} \,|\, \gamma_e, \gamma_x) &= f(\boldsymbol{y} \,|\, \boldsymbol{x}, \gamma_e) f(\boldsymbol{x} \,|\, \gamma_x) && (1) \\ &= (2\pi)^{-N} \det\big[\boldsymbol{\Pi}\big]^{1/2} \gamma_x^{N/2} \gamma_e^{N/2} \exp - \big[\, \gamma_e \, ||\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}||^2 + \, \gamma_x \, ||\boldsymbol{x}||^2_{\boldsymbol{\Pi}} \big] / 2 && (2) \end{aligned}$$

which is naturally parametrised by the two precision parameters γ_e and γ_x. One can notice that inside the exponential, up to a multiplicative constant, we have the expression of the penalised least squares criterion:

$$\begin{aligned} \mathcal{J}_{\text{PLS}}(\boldsymbol{x}) &= \mathcal{J}_{\text{LS}}(\boldsymbol{x}) \; + \; \mu \, \mathcal{J}_0(\boldsymbol{x}) \\ &= ||\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}||^2 + \mu \, ||\boldsymbol{x}||^2_{\boldsymbol{\Pi}} \end{aligned}$$

where the regularisation parameter is written as the inverse of the signal to noise ratio $\mu = \gamma_x / \gamma_e$.

4. *Comment the role of μ, in relation with γ_x and γ_e.*

The expression of the a posteriori distribution, or more commonly posterior distribution, for the object of interest is determined through the use of Bayes rule:

$$\begin{aligned} f(\boldsymbol{x} \,|\, \boldsymbol{y}, \gamma_e, \gamma_x) &= \frac{f(\boldsymbol{x}, \boldsymbol{y} \,|\, \gamma_e, \gamma_x)}{f(\boldsymbol{y} \,|\, \gamma_e, \gamma_x)} \\ &\propto \exp -\gamma_e \, \mathcal{J}_{\text{PLS}}(\boldsymbol{x})/2 \,, \end{aligned} \tag{3}$$

where $\propto$ stands for "proportional to". It is the distribution for the object of interest given the data and the parameters.

Any estimator that we may wish to construct for the object of interest will be based on the above distribution. The most common estimators are the mean, median or the mode (*i.e.,* the maximizer) of the posterior distribution. In the current case where the posterior is Gaussian, all three are equal. Given the expression in (3), the mode or Maximizer A Posteriori (MAP), denoted $\widehat{\boldsymbol{x}}_{\text{MAP}}$, is the one which minimises the criterion $\mathcal{J}_{\text{PLS}}(\boldsymbol{x})$:

$$\widehat{\boldsymbol{x}}_{\text{MAP}} \;=\; \arg\max_{\boldsymbol{x}} f(\boldsymbol{x} \,|\, \boldsymbol{y}, \gamma_e, \gamma_x) \;=\; \arg\min_{\boldsymbol{x}} \mathcal{J}_{\text{PLS}}(\boldsymbol{x}) \;=\; \widehat{\boldsymbol{x}}_{\text{PLS}} \,.$$

To conclude, the solution $\widehat{\boldsymbol{x}}_{\text{PLS}}$ to the penalised least squares criterion, derived in the previous practical work, turns out to be also the mode $\widehat{\boldsymbol{x}}_{\text{MAP}}$ of the posterior distribution.

1.4 Extended posterior distribution

So far, the Bayesian approach allows us only to give a probabilistic interpretation to an already existing estimator for fixed hyper-parameter values. Extending the previous framework to include also the estimation of the two hyper-parameters requires the introduction of a prior distribution for each of the two precision parameters γ_e and γ_x. Several choices are available, in the following we will focus only on the Gamma distribution:

$$f(\gamma) = \frac{\beta^\alpha}{\Gamma(\alpha)} \gamma^{\alpha-1} \exp\left[-\beta\gamma\right] \; \mathbb{1}_{\mathbb{R}_+}(\gamma) \,. \tag{4}$$

It is driven by two positive and real parameters (α, β), having the mean α/β and the variance α/β^2. This choice can be motivated by two arguments of different nature.

1. Practical arguments concerning the implementation. The choice of a prior Gamma distribution ensures that the conditional posterior distribution is also Gamma (we are talking about a conjugate prior). Algorithmically, this means one has just to update the values of the parameters of the distribution. See sections 1.5.1 and 1.5.2.

2. Arguments from an informational point of view. This choice allows one to also cater for the case where the available information with regard to the value of the parameters could be poor (also said "non-informative prior") or could be precise, *e.g.,* a nominal value and/or an uncertinty. Of particular practical interest in this work is the limit case of "non-informative prior" that is $(\alpha, \beta) = (0, 0)$.

In addition, regarding the couple of variable γ_e and γ_x, they are modeled as independent.

Starting from the expression of the Gamma distribution (4) and of the partial joint distribution (1)-(2), we derive the expression of the full joint distribution for $\boldsymbol{y}$, $\boldsymbol{x}$, γ_x and γ_e as:

$$f(\boldsymbol{y}, \boldsymbol{x}, \gamma_e, \gamma_x) = f(\boldsymbol{y}, \boldsymbol{x} \,|\, \gamma_e, \gamma_x) f(\gamma_e) f(\gamma_x)$$

which written in an explicit manner yields :

$$\begin{aligned} f(\boldsymbol{x}, \boldsymbol{y}, \gamma_e, \gamma_x) \;=\; & (2\pi)^{-N} \det\left[\boldsymbol{\Pi}\right]^{1/2} \frac{\beta_e^{\alpha_e} \beta_x^{\alpha_x}}{\Gamma(\alpha_e)\Gamma(\alpha_x)} \\ & \gamma_e^{\alpha_e + N/2 - 1} \gamma_x^{\alpha_x + N/2 - 1} \exp - \left[\gamma_e \left(\beta_e + ||\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}||^2 / 2 \right) + \gamma_x \left(\beta_x + ||\boldsymbol{x}||^2_{\boldsymbol{\Pi}} / 2 \right) \right] \,. \end{aligned} \tag{5}$$

Remark 2 — *This density is crucial as it allows one to derive all other related densities, particularly it allows one to derive all possible joint, marginal and conditional densities.*

We can now derive the *full* posterior distribution, that is the distribution for the object of interest and for the hyper-parameters given the observations, with the aid of Bayes rule:

$$\begin{aligned} f(\boldsymbol{x}, \gamma_e, \gamma_x \,|\, \boldsymbol{y}) &= \frac{f(\boldsymbol{x}, \boldsymbol{y}, \gamma_e, \gamma_x)}{f(\boldsymbol{y})} \\ &\propto \gamma_e^{\alpha_e + N/2 - 1} \gamma_x^{\alpha_x + N/2 - 1} \exp - \left[\gamma_e \left(\beta_e + ||\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}||^2 / 2 \right) + \gamma_x \left(\beta_x + ||\boldsymbol{x}||_\Pi^2 / 2 \right) \right] . \end{aligned} \tag{6}$$

It summarises all the available information with respect to the object of interest and the hyper-parameters in view of the data: for all triplet $\boldsymbol{x}, \gamma_e, \gamma_x$ it quantifies the posterior density, that is the probability of the triplet given the observations. The estimators for the object of interest and the hyper-parameters are constructed starting from it. We can look at the mean, median or mode of the posterior distribution. Each has its specificity, its share of advantages and disadvantages. In what follows we will focus on the mean.

1.5 Posterior mean

Given the complexity of the posterior distribution (6), obtaining an analytical formula for the mean is out of hand. Several techniques are available to compute it and in this practical work we will focus on stochastic sampling techniques. Ultimately, it boils down to a random sampling of the posterior distribution followed by taking the empirical mean of the drawn samples, which will then approximate the posterior mean.

Sampling the posterior distribution is possible thanks to a series of techniques collectively known under the name of Markov Chain Monte Carlo (MCMC). They require constructing an iterative procedure to generate the random samples which, after a certain time known as burn-in, will be distributed according to the target distribution. Constructing such a procedure is not something trivial, however for the current case there exists a standard algorithm which we can put at use with ease: the Gibbs sampling algorithm. It transforms the problem of sampling the posterior distribution for the triplet $(\boldsymbol{x}, \gamma_e, \gamma_x)$ into the problem of sampling three simpler distributions, one for each of the member of the triplet. Each of the three distribution is in fact the conditional distribution for one of the member given the remaining members. The working principle for this algorithm is given in the table below with each of the three steps being detailed in the following sections.

Table 1: Gibbs algorithm for $f(\boldsymbol{x}, \gamma_e, \gamma_x \,|\, \boldsymbol{y})$

- Initialize $\boldsymbol{x}^{[0]} = \boldsymbol{y}$
- For $k = 1, 2, \ldots$ repeat
  - (a) sample $\gamma_e^{[k]}$ under $f(\gamma_e | \gamma_x^{[k-1]}, \boldsymbol{x}^{[k-1]}, \boldsymbol{y})$
  - (b) sample $\gamma_x^{[k]}$ under $f(\gamma_x | \gamma_e^{[k]}, \boldsymbol{x}^{[k-1]}, \boldsymbol{y})$
  - (c) sample $\boldsymbol{x}^{[k]}$ under $f(\boldsymbol{x} | \gamma_e^{[k]}, \gamma_x^{[k]}, \boldsymbol{y})$

1.5.1 Sampling the inverse error power

Sampling the hyper-parameter γ_e, corresponding to step (a) from the algorithm in Table 1, requires sampling the conditional posterior distribution $f(\gamma_e|\boldsymbol{x}, \gamma_x, \boldsymbol{y})$. It is obtained from the full joint distribution (5) as follows:

$$f(\gamma_e|\boldsymbol{x}, \gamma_x, \boldsymbol{y}) = \frac{f(\boldsymbol{x}, \boldsymbol{y}, \gamma_e, \gamma_x)}{f(\boldsymbol{x}, \gamma_x, \boldsymbol{y})} .$$

By keeping only the terms which contain γ_e and since the denominator does not depend on γ_e we get:

$$\begin{aligned} f(\gamma_e|\boldsymbol{x}, \gamma_x, \boldsymbol{y}) &\propto f(\boldsymbol{x}, \boldsymbol{y}, \gamma_e, \gamma_x) \\ &\propto \gamma_e^{\alpha_e + N/2 - 1} \exp -\gamma_e \left(\beta_e + ||\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}||^2/2\right) . \end{aligned}$$

The conditional distribution that is obtained is in fact a Gamma distribution with parameters (α, β):

$$\alpha = \alpha_e + N/2 \quad \text{and} \quad \beta = \beta_e + ||\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}||^2/2 .$$

In the limit case when the parameters of the prior (α_e, β_e) equal $(0, 0)$, the parameters of the posterior are:

$$\alpha = N/2 \quad \text{and} \quad \beta = ||\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}||^2/2 .$$

5. *Give the expression for the mean and variance of this law and comment upon the resulting expressions, explain how it is connected to the power of the outpt error $\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}$.*

(a)

The samples for $\gamma_e^{[k]}$ are drawn from a Gamma distribution with the values of its two parameters `Alpha` *and* `Beta` *given by the above two equations. For that purpose, one can use the provided Matlab function* `RNDGamma` *as follows:* `RNDGamma(Alpha,Beta).`

Remark 3 — *Computing the parameter β involves computing the norm of the modelling error $||\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}||$. The cost of computing the norm in the spatial domain is usually high, however it is possible to compute it in the Fourier domain to ease off the cost.*

1.5.2 Sampling the inverse power of the object of interest

We shall now focus on sampling the hyper-parameter γ_x, corresponding to step (b) in the algorithm from Table 1. This requires sampling the conditional posterior $f(\gamma_x|\boldsymbol{x}, \gamma_e, \boldsymbol{y})$. Using a similar approach as in the previous section we obtain that

$$f(\gamma_x|\boldsymbol{x}, \gamma_e, \boldsymbol{y}) \propto \gamma_x^{\alpha_x + N/2 - 1} \exp -\gamma_x \left(\beta_x + ||\boldsymbol{x}||^2_{\boldsymbol{\Pi}}/2\right) .$$

As it can be noticed this conditional posterior distribution is also a Gamma distribution. In the limit case where (α_x, β_x) equal to $(0, 0)$ the posterior parameters simply equal:

$$\alpha = N/2 \quad \text{and} \quad \beta = ||\boldsymbol{x}||^2_{\boldsymbol{\Pi}}/2$$

6. *Give the expression for the mean and variance of this law. Comment upon the resulting expressions.*

(b)

The samples $\gamma_x^{[k]}$ are drawn from a Gamma distribution having the values of its parameters given by the above two equations. They can be obtained using the `RNDGamma` *function:* `RNDGamma(Alpha,Beta).`

Remark 4 — *Computing β requires computing the norm with respect to the regularisation matrix which can and should be computed in the Fourier domain for similar reasons as the ones exposed in the previous section.*

1.6 Sampling the object of interest

Last but not least, we shall tackle the sampling of the object of interest $\boldsymbol{x}$ corresponding to the final step (c) from the algorithm in Table 1. This implies sampling the conditional posterior distribution $f(\boldsymbol{x}|\gamma_x, \gamma_e, \boldsymbol{y})$, the expression of which has already been derived, see Eq. (3), and which is conveniently rewritten below:

$$\begin{aligned} f(\boldsymbol{x}|\gamma_e, \gamma_x, \boldsymbol{y}) &\propto \exp - \left[\gamma_e \, ||\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}||^2 + \gamma_x \, ||\boldsymbol{x}||^2_{\boldsymbol{\Pi}} \right] /2 \\ &= \exp - \gamma_e \, \mathcal{J}_{\text{PLS}}(\boldsymbol{x})/2 \, . \end{aligned}$$

The density itself is Gaussian as the argument inside the exponential is a positive definite quadratic equation in $\boldsymbol{x}$. It is fully characterised by its mean and its covariance matrix.

- In this case, the mean $\boldsymbol{\mu}_{x|*}$ is also the mode and as such, also the minimiser of $\mathcal{J}_{\text{PLS}}(\boldsymbol{x})$, that is to say the penalised least squares solution, *i.e.,* the Wiener-Hunt solution discussed in the previous practical work.
- The covariance matrix $\boldsymbol{\Sigma}_{x|*}$ is obtained by calculating the Hessian (the matrix of second order derivatives) of $\mathcal{J}_{\text{PLS}}(\boldsymbol{x})$.

We obtain the following expressions:

$$\boldsymbol{\Sigma}_{x|*} = \left[\gamma_e \boldsymbol{H}^{\mathrm{t}} \boldsymbol{H} + \gamma_x \boldsymbol{\Pi} \right]^{-1} , \tag{7}$$

$$\text{and } \boldsymbol{\mu}_{x|*} = \gamma_e \, \boldsymbol{\Sigma}_{x|*} \, \boldsymbol{H}^{\mathrm{t}} \boldsymbol{y} \, . \tag{8}$$

The numerical problem one now faces is the sampling of a potentially high dimensional Gaussian distribution. This high dimensionality prohibits the inversion or factorisation of the covariance matrix $\boldsymbol{\Sigma}_{x|*}$, which in turn means that there is no simple sampling scheme available. To overcome this issue, we employ a circulant approximation of the covariance matrix to gain access to fast matrix operations performed in the Fourier domain. This yields the following expression:

$$\boldsymbol{\Lambda}_{x|*} = \left[\gamma_e \boldsymbol{\Lambda}_H^{\dagger} \boldsymbol{\Lambda}_H + \gamma_x \boldsymbol{\Lambda}_D^{\dagger} \boldsymbol{\Lambda}_D \right]^{-1} , \tag{9}$$

$$\mathring{\boldsymbol{\mu}}_{x|*} = \gamma_e \, \boldsymbol{\Lambda}_{x|*} \, \boldsymbol{\Lambda}_H^{\dagger} \mathring{\boldsymbol{y}} \, . \tag{10}$$

In the Fourier domain, the covariance matrix is diagonal which means that its components are decorrelated. As such, each component is independent of the rest which enables one to sample them in parallel.

7. *Optional question. Prove that (9)-(10) can be deduced from (7)-(8) based on diagonalisation of circulant matrices. Hint: $\boldsymbol{H}$ is a real matrix, so it is also its complex conjugate, and $\boldsymbol{H}^{\mathrm{t}} = \boldsymbol{H}^{\dagger}$; same idea for $\boldsymbol{D}$.*

(c) *The samples $\boldsymbol{x}^{[k]}$ for the image are to be drawn from the Gaussian distribution having first and second moments in the Fourier domain given by the above two equations. For that, one can use the provided Matlab function* `RNDGauss` *as follows:* `RNDGauss(Moy,Cov)`*, where* `Moy` *and* `Cov` *must be given in the Fourier domain.*

2 Implementation

For the practical part, we shall work with the same data sets as for the previous practical work and one should obtain similar results. This time though, one does not need to manually tune the regularisation parameter. Table 1 summarizes the algorithm one has to implement. It involves the repeated sampling of the conditional distributions for two hyper-parameters, γ_e and γ_x, and of the conditional distribution for the image $\boldsymbol{x}$, where the practical aspects regarding the sampling part are discussed in the framed text.

For the practical part, two *Matlab* function are provided.

1. The function `RNDGamma` generates a scalar sample (realisation) from a Gamma distribution. More precisely, the command `Precision = RNDGamma(Alpha,Beta)` returns in the variable `Precision` a sample from the Gamma distribution having the parameters given by `Alpha` and `Beta`

2. The function `RNDGauss` generates an image sample under a gaussian distribution. More precisely, the command `Image = RNDGauss(Moy,Cov)` returns in the variable `Image` an image drawn from the gaussian distribution with the mean given by `Moy` and the covariance matrix given by `Cov`. The two parameters `Moy` and `Cov` and the image `Image` are in the Fourier domain and not in the spatial one.