Chapter 7
Bayesian inverse solution versus deterministic inverse solution

Abstract In this note, we are interested in solving inverse problems using statistical techniques.

7.1 Posterior as the solution to Bayesian inverse problems

In this section, we explore the posterior (5.5), the solution of our Bayesian problem, given the likelihood in Section 5.1 and priors in Chapter 6.

To derive results that are valid for all priors discussed so far, we work with the following generic prior

$$\pi_{\text{prior}}(m) \propto \exp\left(-\frac{1}{2\gamma^2}\left\|\Gamma^{-\frac{1}{2}} m\right\|^2\right),$$

where $\Gamma^{-\frac{1}{2}} \in \{L_D, L_A, HL_N\}$, each of which again presents a different belief. The Bayesian solution (5.5) can be now written as

$$\pi_{\text{post}}\left(m|y^{obs}\right) \propto \exp\left(-\underbrace{\left[\frac{1}{2\sigma^2}\left\|y^{obs} - \mathscr{A} m\right\|^2 + \frac{1}{2\gamma^2}\left\|\Gamma^{-\frac{1}{2}} m\right\|^2\right]}_{T(m)}\right),$$

where $T(m)$ is the familiar (to you I hope) *Tikhonov functional*; it is sometimes called the *potential*. We re-emphasize here that the Bayesian solution is the posterior probability density, and if we draw samples from it, we want to know what the most likely function m is going to be. In other words, we ask for the most probable point m in the posterior distribution. This point is known as the *Maximum A Posteriori (MAP)* estimator/point, namely, the point at which the posterior density is maximized. Let us denote this point as m_{MAP}, and we have

$$m_{MAP} \stackrel{\text{def}}{=} \arg\max_m \pi_{\text{post}}\left(m|y^{obs}\right) = \arg\min_m T\left(m\right).$$

Hence, the MAP point is exactly the deterministic solution of the Tikhonov functional!

This is fundamental. If you have not seen this, prove it!

Since both likelihood and prior are Gaussian, the posterior is also a Gaussian. For our case, the resulting posterior Gaussian reads

$$\begin{aligned}\pi_{\text{post}}\left(m|y^{obs}\right) &\propto \exp\left(-\frac{1}{2}\left\|m-\frac{1}{\sigma^2}H^{-1}\mathscr{A}^T y^{obs}\right\|_H^2\right)\\ &= \exp\left(-\frac{1}{2}\left(m-\frac{1}{\sigma^2}H^{-1}\mathscr{A}^T y^{obs}, H\left(m-\frac{1}{\sigma^2}H^{-1}\mathscr{A}^T y^{obs}\right)\right)\right)\\ &\stackrel{\text{def}}{=} \exp\left(-\frac{1}{2}\left(m-\frac{1}{\sigma^2}H^{-1}\mathscr{A}^T y^{obs}, \Gamma_{\text{post}}^{-1}\left(m-\frac{1}{\sigma^2}H^{-1}\mathscr{A}^T y^{obs}\right)\right)\right)\end{aligned}$$

where

$$\boxed{H = \frac{1}{\sigma^2}\mathscr{A}^T\mathscr{A} + \frac{1}{\gamma^2}\Gamma^{-1}},$$

is the Hessian of the Tikhonov functional (aka the regularized misfit), and we have used the weighted norm $\|\cdot\|_H^2 = \left\|H^{\frac{1}{2}}\cdot\right\|^2$.

Exercise 7.1. Show that the posterior is indeed a Gaussian, i.e.,

$$\pi_{\text{post}}\left(m|y^{obs}\right) \propto \exp\left(-\frac{1}{2}\left\|m-\frac{1}{\sigma^2}H^{-1}\mathscr{A}^T y^{obs}\right\|_H^2\right).$$

•

The other important point is that the posterior covariance matrix is precisely the inverse of the Hessian of the regularized misfit, i.e.,

$$\boxed{\Gamma_{\text{post}} = H^{-1}}.$$

Last, but not least, we have showed that the MAP point is given by

$$m_{MAP} = \frac{1}{\sigma^2}H^{-1}\mathscr{A}^T y^{obs} = \frac{1}{\sigma^2}\left(\frac{1}{\sigma^2}\mathscr{A}^T\mathscr{A} + \frac{1}{\gamma^2}\Gamma^{-1}\right)^{-1}\mathscr{A}^T y^{obs},$$

which is, again, exactly the solution of the Tikhonov functional for linear inverse problem.

Exercise 7.2. Show that m_{MAP} is also the least squares solution of the following over-determined system

$$\begin{bmatrix}\frac{1}{\sigma}\mathscr{A}\\ \frac{1}{\gamma}\Gamma^{-\frac{1}{2}}\end{bmatrix} m = \begin{bmatrix}\frac{1}{\sigma}y^{obs}\\ 0\end{bmatrix}$$

•

Exercise 7.3. Show that the posterior mean, which is in fact the conditional mean, is precisely the MAP point. •

Since the covariance matrix, generalization of the variance in multi-dimensional spaces, represents the uncertainty, quantifying the uncertainty in the MAP estimator is ready by simply computing the inverse the Hessian matrix. Let's us now numerically explore the Bayesian posterior solution.

We choose $\beta = 0.05$, $n = 100$, and $\gamma = 1/n$. The truth underlying function that we would like to invert for is given by

$$f(t) = 10(t-0.5)\exp\left(-50(t-0.5)^2\right) - 0.8 + 1.6t.$$

The noise level is taken to be the 5% of the maximum value of $f(s)$, i.e. $\sigma = 0.05 \max_{s\in[0,1]} |f(s)|$.

We first consider the belief described by π^D_{prior} in which we think that $f(s)$ is zero at the boundaries. Figures 7.1 plots the MAP estimator, the truth function $f(s)$, and the predicted uncertainty. As can be observed, the MAP is in good agreement with the truth function inside the interval $[0,1]$, though it is far from recovering $f(s)$ at the boundaries. This is the price we have to pay for not admitting our ignorance about the boundary values of $f(s)$. The likelihood in fact sees this discrepancy in the prior knowledge and tries to make correction by lifting the MAP away from 0, but not enough to be a good reconstruction. The reason is that our incorrect prior is strong enough such that the information from the data y^{obs} cannot help much.

Exercise 7.4. Can you make the prior less strong? *Change some parameter to make prior contribution less!* Use `BayesianPosterior.m` to test your answer. Is the prediction better in terms of satisfying the boundary conditions? Is the uncertainty smaller? If not, why? •

On the other hand, if we admit this ignorance and use the corresponding prior π^D_{prior}, we see much better reconstruction in Figure 7.2. In this case, we in fact let the information from the data y^{obs} determine the appropriate values for the Dirichlet boundary conditions rather than setting them to zero. By doing this, we allow the likelihood and the prior to be well-balanced leading to good reconstruction and uncertainty quantification.

Exercise 7.5. Play with `BayesianPosterior.m` by varying γ, the data misfit (or the likelihood) contribution, and σ, the regularization (or the prior) contribution. •

Exercise 7.6. Use your favorite deterministic inversion approach to solve the above deconvolution problem and then compare it with the solution in Figure 7.2. •

Now consider the case in which the truth function has a jump discontinuity at $j = 70$. Assume we also know that the magnitude of the jump is 10. In particular, we take the truth function $f(s)$ as the following step function

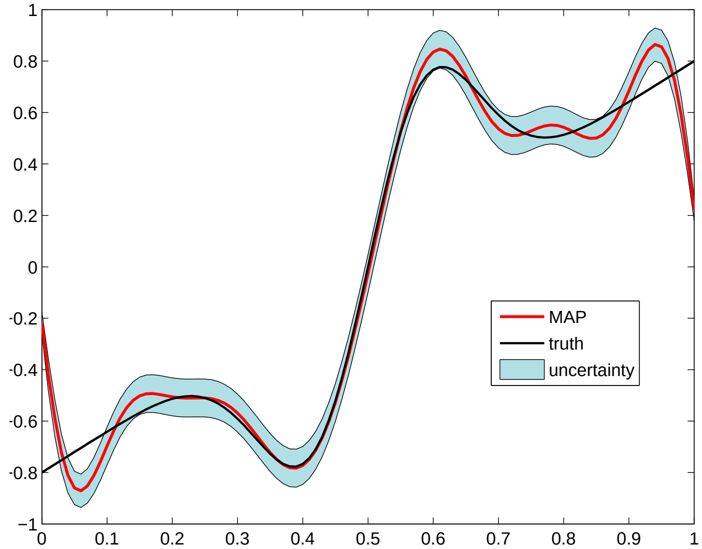

Fig. 7.1 The MAP estimator, the truth function, and the predicted uncertainty (95% credibility region) using π^{D}_{prior}.

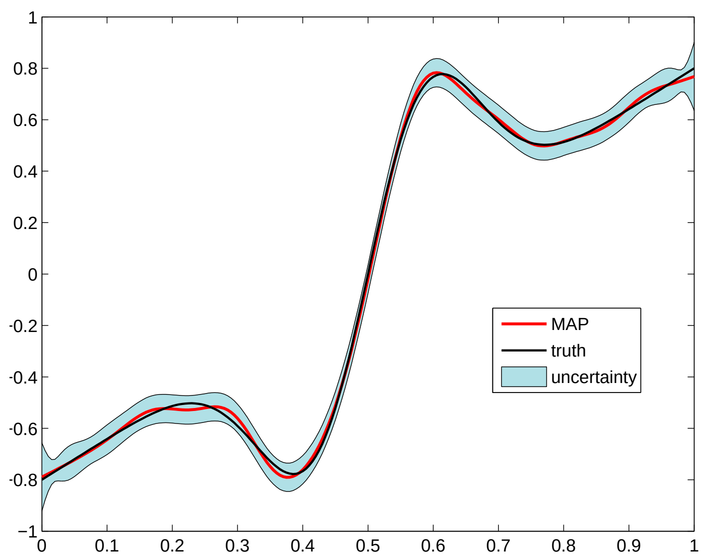

Fig. 7.2 The MAP estimator, the truth function, and the predicted uncertainty (95% credibility region) using π^{A}_{prior}.

$$f(s) = \begin{cases} 0 & \text{if } s \leq 0.7 \\ 10 & \text{otherwise} \end{cases}.$$

Since we otherwise have no further information about $f(s)$, let us be more conservative by choosing $\gamma = 1$ and $\theta = 0.1$ at $j = 70$ in π^{O}_{prior} as we discussed in (6.8). Figure 7.3 shows that we are doing pretty well in recovering the jump and other parts of the truth function.

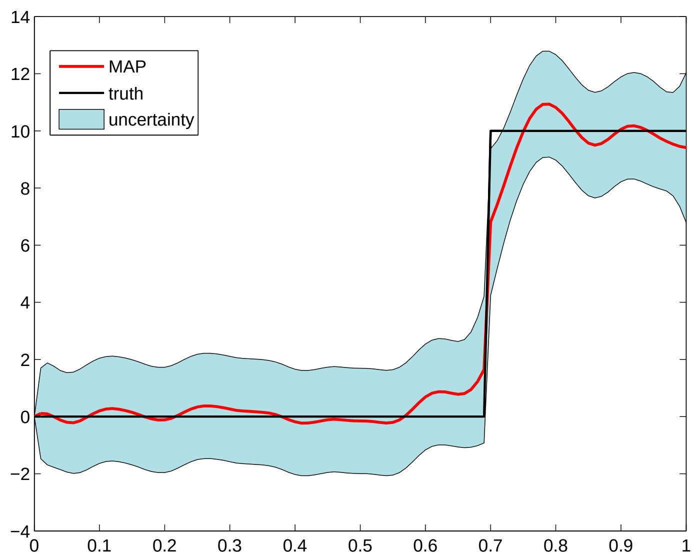

Fig. 7.3 The MAP estimator, the truth function, and the predicted uncertainty (95% credibility region) using π^{O}_{prior}.

A question you may ask is whether we can do better? The answer is yes by taking smaller γ if the truth function does not vary much everywhere except at the jump discontinuity. We take this prior information into account by taking $\gamma = 1.e-8$, for example, then our reconstruction is almost perfect in Figure 7.4. *Why?*

Exercise 7.7. Try `BayesianPosteriorJump.m` with γ deceasing from 1 to $1.e-8$ to see the improvement in quality of the reconstruction. •

Exercise 7.8. Use your favorite deterministic inversion approach to solve the above deconvolution problem with discontinuity and then compare it with the solution in Figure 7.4. •

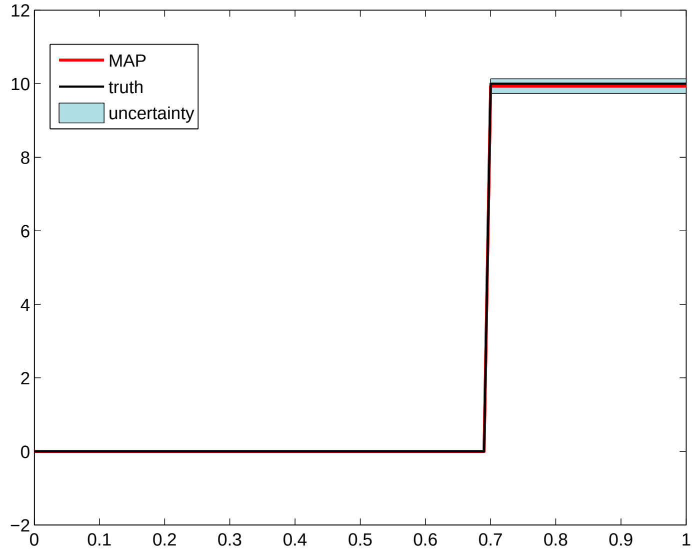

Fig. 7.4 The MAP estimator, the truth function, and the predicted uncertainty (95% credibility region) using π^O_{prior}.

7.2 Connection between Bayesian inverse problems and deterministic inverse problems

We have touched upon the relationship between Bayesian inverse problem and deterministic inverse problem in Section 7.1 by pointing out that the potential of the posterior density is precisely the Tikhonov functional up to a constant. We also point out that the MAP estimator is exactly the solution of the deterministic inverse problem. Note that we derive this relation for a linear likelihood model, but it is in fact true for nonlinear ones (e.g. nonlinear parameter-to-observable map $\mathscr{A}m$).

Can you confirm this?

Up to this point, you may realize that the Bayesian solution contains much more information than its deterministic counterpart. Instead of having a point estimate, the MAP point, we have a complete posterior distribution to explore. In particular, we can talk about a simple uncertainty quantification by examining the diagonal of the posterior covariance matrix. We can even discuss about the posterior correlation structure by looking at the off diagonal elements, though we are not going to do it here in this lecture note. Since, again, both likelihood and prior are Gaussian, the posterior is a Gaussian distribution, and hence the MAP point (the first order moment) and the covariance matrix (the second order moment) are the complete description of the posterior. If, however, the likelihood is not Gaussian, say when the $\mathscr{A}m$ is nonlinear, then one can explore higher moments.

We hope the arguments above convince you that the Bayesian solution provide information far beyond the deterministic counterpart. In the remainder of this section, let us dig into details the connection between the MAP point and the deterministic solution, particularly in the context of the deconvolution problem. Recall the definition of the MAP point

$$
\begin{aligned}
m_{MAP} &\stackrel{\text{def}}{=} \arg\min_m T(m) = \sigma^2 \left(\frac{1}{2} \left\| y^{obs} - \mathscr{A} m \right\|^2 + \frac{1}{2} \frac{\sigma^2}{\gamma^2} \left\| \Gamma^{-\frac{1}{2}} m \right\|^2 \right) \\
&= \arg\min_m T(m) = \sigma^2 \left(\frac{1}{2} \left\| y^{obs} - y \right\|^2 + \frac{1}{2} \kappa \left\| R^{\frac{1}{2}} m \right\|^2 \right),
\end{aligned}
$$

where we have defined $\kappa = \sigma^2/\gamma^2$, $R^{\frac{1}{2}} = \Gamma^{-\frac{1}{2}}$, and $y = \mathscr{A} m$.

We begin our discussion with zero Dirichlet boundary condition prior $\pi^D_{\text{prior}}(m)$ in (6.5). Recall in (6.2) and (6.4) that $L_D m$ is proportional to a discretization of the Laplacian operator with zero boundary conditions using second order finite difference method. Therefore, our Tikhonov functional is in fact a discretization, up to a constant, of the following potential in the infinite dimensional setting

$$
T_\infty(f) = \frac{1}{2} \left\| y - y^{obs} \right\|^2 + \frac{1}{2} \kappa \left\| \Delta f \right\|^2_{L^2(0,1)},
$$

where $\|\cdot\|^2_{L^2(0,1)} \stackrel{\text{def}}{=} \int_0^1 (\cdot)^2 \, ds$. Rewrite the preceding equation informally as

$$
T_\infty(f) = \frac{1}{2} \left\| y - y^{obs} \right\|^2 + \frac{1}{2} \kappa \left(f, \Delta^2 f \right)_{L^2(0,1)},
$$

and we immediately realize that the potential in our prior description, namely $\|L_D m\|^2$, is in fact a discretization of Tikhonov regularization using the biharmonic operator. This is another explanation for the smoothness of the prior realizations and the name smooth prior, since biharmonic regularization is very smooth. [1]

The power of the statistical approach lies in the construction of prior $\pi^R_{\text{prior}}(m)$. Here, the interpretation of rows corresponding to interior nodes s_j is still the discretization of the biharmonic regularization, but the design of those corresponding to the boundary points is purely statistics, for which we have no corresponding deterministic counterpart (or at least it is not clear how to construct it from a purely deterministic point of view). As the results in Section 7.1 showed, $\pi^R_{\text{prior}}(m)$ provided much more satisfactory results both in the prediction and in uncertainty quantification.

As for the "non-smooth"priors in Section 6.2, a simple inspection shows that $L_N m$ is, up to a constant, a discretization of ∇f. Similar to the above discussion, the potential in our prior description, namely $\|L_D m\|^2$, is now in fact a discretization

[1] From a functional analysis point of view, $\|\Delta f\|^2_{L^2(0,1)}$ is finite if $f \in H^2(0,1)$, and by Sobolev imbedding theorem we know that in fact $f \in C^{1,1/2-\varepsilon}$, the space of continuous differential functions whose first derivative is in the Hölder space of continuous function $C^{1/2-\varepsilon}$, for any $0 < \varepsilon < \frac{1}{2}$. So indeed f is more than continuously differentiable.

of Tikhonov regularization using the Laplacian operator.[2] As a result, the current prior is less smooth than the previous one with harmonic operator. Nevertheless, all the prior realizations corresponding to $\pi_{\text{pren}}(m)$ are at least continuous, though may have steep gradient at s_j as shown in Figures 7.3 and 7.4. The rigorous arguments for the prior smoothness require the Sobolev embedding theorem, but we avoid the details.

For those who have not seen the Sobolev embedding theorem, you only loose the insight on why $\pi^{O}_{\text{prior}}(m)$ could give very steep gradient realizations (which is the prior belief we start with). Nevertheless, you still can see that $\pi^{O}_{\text{prior}}(m)$ gives less smooth realizations than $\pi^{D}_{\text{prior}}(m)$ does, since, at least, the MAP point corresponding to $\pi^{O}_{\text{prior}}(m)$ only requires finite first derivative of f while second derivative of f needs to be finite at the MAP point if $\pi^{D}_{\text{prior}}(m)$ is used.

[2] Again, Sobolev embedding theorem shows that $f \in C^{1/2-\varepsilon}$ for $\|\nabla f\|^2_{L^2(0,1)}$ to be finite. Hence, all prior realizations corresponding to $\pi_{\text{pren}}(m)$ are at least continuous. The prior $\pi^{O}_{\text{prior}}(m)$ is different, due to the scaling matrix J. As long as θ stays away from zero, prior realizations are still in $H^1(0,1)$, and hence continuous though having steep gradient at s_j as shown in Figures 7.3 and 7.4. But as θ approaches zero, prior realizations are leaving $H^1(0,1)$, and therefore may be no longer continuous. Note that in one dimension, $H^{\frac{1}{2}+\varepsilon}$ is enough to be embedded in the space of C^{ε}-Hölder continuous functions. If you like to know a bit about the Sobolev embedding theorem, see [3].