Bayesian inversion of a diffusion evolution equation with application to Biology

Jean-Charles Croix, Nicolas Durrande, Mauricio Alvarez

Bayesian inversion of a diffusion evolution equation with application to Biology.

Jean-Charles Croix[1], Nicolas Durrande[1,2], and Mauricio A. Álvarez[3]

[1]Mines Saint-Etienne, Univ Clermont Auvergne, CNRS, UMR 6158 LIMOS, Institut Henri Fayol, Departement GMI, F-42023 Saint-Etienne, France.
[2]Prowler.io, Cambridge, UK.
[3]Department of Computer Science, University of Sheffield, UK.

June 13, 2018

Abstract

A common task in experimental sciences is to fit mathematical models to real-world measurements to improve understanding of natural phenomenon (reverse-engineering or inverse modeling). When complex dynamical systems are considered, such as partial differential equations, this task may become challenging and ill-posed. In this work, a linear parabolic equation is considered where the objective is to estimate both the differential operator coefficients and the source term at once. The Bayesian methodology for inverse problems provides a form of regularization while quantifying uncertainty as the solution is a probability measure taking in account data. This posterior distribution, which is non-Gaussian and infinite dimensional, is then summarized through a mode and sampled using a state-of-the-art Markov-Chain Monte-Carlo algorithm based on a geometric approach. After a rigorous analysis, this methodology is applied on a dataset of the post-transcriptional regulation of Kni gap gene in the early development of Drosophila Melanogaster where mRNA concentration and both diffusion and depletion rates are inferred from noisy measurement of the protein concentration.

1 Introduction

The problem of diffusion in a porous media, which is ubiquitous in physics, engineering and biology, is usually represented by the following partial differential

equation:

$$\begin{aligned} \frac{\partial y}{\partial t}(x,t) + \lambda(x,t)y(x,t) - D(x,t)\Delta y(x,t) &= f(x,t),\ \forall (x,t) \in \Omega \times]0,T], \\ y(x,t) &= 0,\ \forall (x,t) \in \Omega \times \{t=0\}, \\ y(x,t) &= 0,\ \forall (x,t) \in \partial\Omega \times]0,T]. \end{aligned} \tag{1}$$

where the spatial domain is $\Omega \subset \mathbb{R}^n$ ($n \leq 3$) and the final time is $T \in \mathbb{R}^+$ (other initial and boundary conditions are possible). In real world applications, the quantity of interest y (hereafter called the solution of equation 1) is typically the concentration of some chemical and evolves from a null initial state under three distinct mechanisms: a) direct variation in concentration, given by the source f, b) diffusion at rate D, c) production or depletion at a rate λ. Different hypotheses on the parameters lead to a well-posedness of this solution, which will be detailed later on. Besides the traditional computation of the solution from the parameters, one can use this model for the determination of an optimal control (e.g. source leading to the minimization of a particular cost functional) or the identification of parameters from partial knowledge of the solution in an inverse setting. This is the problem that will be of interest in this paper. The motivation comes from a a challenging identification problem in Biology where the objective is to infer jointly the decay rate λ, the diffusion rate D and the source f given a limited number of noisy observations of the solution y. Note that given their physical interpretation, the two rates and the solution must be positive so this constrain has to be taken into account in our inference scheme.

This problem has already been solved by different approaches, under distinct sets of hypotheses. In [5], the authors use a system of ordinary differential equations instead of equation 1, and minimizes a discrete version of a least-square type functional, while confidence intervals on parameters are given by bootstrapping. Alternative methods to solve this problem, in a Bayesian setting, is to use Latent Force Models [3, 31]. These approaches assume that unknown physical quantities can be modelled with Gaussian Processes [29]. In particular, if f is Gaussian and if the decay and the diffusion are constant, y is Gaussian as well (linear transformation of a Gaussian process). The two constant parameters λ and D can then be estimated through likelihood maximization. The main drawback of this approach is the difficulty to ensure positiveness of the source function f for instance.

In this work, we apply a more general methodology, based on the recent advent of Bayesian Inverse Problems [33] for infinite dimensional spaces. In a sense, it has the advantage of dealing with the ill-posedness while fully integrating the quantification of uncertainties. Moreover, the possibility of taking naturally physical constraints (such as positivity) will be particularly useful. This paper is organized as follows: section 2 presents all the mathematical analysis of the forward model (mapping equation parameters to its solution) and the Bayesian methodology applied in our particular setting. Section 3 focuses on a particular Markov chain Monte-Carlo algorithm that is adapted to the problem and robust to discretization. Finally, section 4 contains all implemen-

tation details and the numerical results obtained on a dataset associated to the developmental biology of the Drosophila Melanogaster.

2 Bayesian inversion

As previously announced, our goal in this work is to infer a source term f (mRNA concentration) jointly with rates of diffusion D and decay λ (i.e. the parameter $u = (\lambda, D, f)$) from noisy and partial measurements of the solution y (gap protein concentration). This problem is ill-posed for multiple reasons: a) the parameter u is infinite dimensional and only finite data are available, b) the solution map is not injective and c) observations are noisy. The typical approach to alleviate this issue is to regularize the problem, usually adding constraints with Tikhonov-Philips functionals, to ensure uniqueness and continuity w.r.t. the observations [26, 32]. Doing so, the regularized problem's solution will be compatible with the dataset, but possibly very different from the *real* parameter (if there is such thing). Additionally, a particularly valuable information is a representation of all parameters u that would lead to similar data, giving precise statement on how the dataset is informative [17, 7, 34]. One approach consists in treating these 2 objectives sequentially, first regularizing then quantifying uncertainty. However, the Bayesian methodology for inverse problems ([33] and more recently [15]) is precisely tailored to complete both tasks at once in an elegant manner. One particularity of these recent contributions is to tackle inverse problems directly in function spaces, postponing discretization at the very end for implementation purposes, leading to algorithms robust to the discretization dimension. Indeed, finite approximations of probability measures may be absolutely continuous while their infinite counterparts are mutually singular. This become particularly problematic in MCMC sampling for instance [12].

In essence, instead of searching for one particular parameter value that would solve the regularized problem, this approach considers the probability distribution of the parameter given the data. Namely, given a prior distribution (see subsection 2.2) and few technical conditions on the forward problem (parameter to data map, see subsection 2.1), Bayes theorem applies and exhibits a unique posterior distribution (see subsection 2.3), which is continuous in the data (w.r.t. Hellinger metric). Finally, one may summarize information from this posterior distribution such as expected value or modes (subsection 2.4). We detail in the next sections how this methodology can be applied to the problem at hand.

2.1 Forward model analysis

The first step is to detail precisely the required regularity of the solution map from equation 1. Using common variational techniques from PDE theory (see [16] or [9]), one can show that this equation has a unique weak solution (see theorem 2.1 which proof is in the appendix) given $u = (\lambda, D, f)$ in a domain $\mathcal{U}$

that will be precised later on. Moreover, this solution evolves smoothly when the parameter varies. Without loss of generality and keeping in mind the Biological application, the underlying physical domain will be $\Omega =]0, L[$ with $L \in \mathbb{R}^+$.

Theorem 2.1. *Let $\mathcal{P} = \mathbb{R}\times]0, +\infty[\times L^2([0,T], H^{-1}(\Omega))$, then for all $u \in \mathcal{P}$, equation 1 has a unique weak solution, defining a map:*

$$y : u \in \mathcal{P} \to y(u) \in W(0, T, L^2, H_0^1).$$

Moreover, this map has the following properties:

1. *it satisfies the following estimate $\forall u \in \mathcal{P}$,*

$$\|y(u)\|_{W(0,T,L^2,H_0^1)} \leq \frac{C}{\sqrt{D}} \|f\|_{L^2([0,T],H^{-1}(\Omega))}$$

   with $C > 0$ constant independent of u,

2. *it is locally Lipschitz, $\forall u \in \mathcal{P}$, $\forall r > 0$ such that $\mathcal{B}(u, r) \subset \mathcal{P}$, $\exists L(r) > 0$, $\forall (u_1, u_2) \in \mathcal{B}(u, r)$,*

$$\|y(u_1) - y(u_2)\|_{W(0,T,L^2,H_0^1)} \leq L(r)\|u_1 - u_2\|_{\mathcal{P}}$$

3. *it is twice Fréchet differentiable on $\mathcal{P}$.*

The proof is given in the appendix. Let us now justify why the properties given in theorem 2.1 are important for the Bayesian inversion (most of them used in section 2.3):

1. The energy estimate will be critical to establish the continuity of the posterior w.r.t. data, because it relates sufficient integrability conditions on y w.r.t. u,
2. Continuity (implied by Fréchet differentiability or local Lipschitz behaviour) will be used to show that the solution map is measurable,
3. Second order Fréchet differentiability will be necessary for geometric methods in optimization (research of modes) and sampling (Markov-Chain Monte-Carlo),
4. The local Lipschitz behaviour is used in the characterization of posterior modes (Maximum a Posteriori).

In the rest of this work, we will restrict ourselves to the subset

$$\mathcal{U} := \mathbb{R}^+\times]0, +\infty[\times\mathcal{C}([0,T] \times [0, L], \mathbb{R}),$$

which is implicitly equipped with the norm

$$\|u\|_{\mathcal{U}} = |\lambda| + |D| + \|f\|_\infty.$$

Since $\mathcal{U} \subset \mathcal{P}$, the solution map is well defined on this subset and keeps all its smoothness properties. Moreover, one can show that a weak solution of equation 1 for $u \in \mathcal{U}$ is also a strong solution [9, 16], but we don't need these regularity results here.

2.2 Choice of prior distribution

The second step is to choose a prior probability distribution on $\mathcal{U}$, encoding all knowledge on the physics at hand, while being simple enough to keep analysis tractable. Here are the constraints given by the biological application:

- (λ, D) must be positive (which imposes decay, not production),
- f must be positive and continuous at all time and position (it is a concentration).

Starting with the depletion and diffusion parameters (λ, D), we choose respectively Borel prior distributions μ_0^λ and μ_0^D on $\mathbb{R}^+$ with densities w.r.t. Lebesgue's measure. Now, since f must be positive, we re-parametrize the problem with

$$f^* = \exp(f), \tag{2}$$

where $f \in \mathcal{C}([0,T] \times [0,L], \mathbb{R})$. By selecting a Borel probability measure μ_0^f on the Banach space $\mathcal{C}([0,T] \times [0,L], \mathbb{R})$, both continuity and positivity of the new source f^* will be ensured almost-surely. In this work, we choose μ_0^f as a Gaussian distribution with covariance operator $\mathcal{C}$ with continuous realizations (see [8] for a presentation of infinite dimensional Gaussian measures). Finally, we assume independence between the three components, leading to the following prior distribution:

$$\mu_0(du) := \mu_0^\lambda(d\lambda) \otimes \mu_0^D(dD) \otimes \mu_0^f(df). \tag{3}$$

These choices clearly ensure positivity of u (in the previous sense) and $\mu_0(\mathcal{U}) = 1$. The exponential map in equation 2 can be replaced with any sufficiently differentiable function from $\mathbb{R}$ to $\mathbb{R}^+$ (to keep the second order Fréchet differentiability of the solution map). Alternative distributions are possible for f (Besov measure from [14] or more general convex measures from [25]) for the regularization. In practice however, our choice is also motivated by the fact that one can find a Gaussian reference measure μ_{ref} in the form

$$\mu_{ref} = \mu_{ref}^\lambda \otimes \mu_{ref}^D \otimes \mu_{ref}^f = \mathcal{N}(\lambda_{ref}, \sigma_\lambda^2) \otimes \mathcal{N}(D_{ref}, \sigma_D^2) \otimes \mathcal{N}(0, \mathcal{C}) = \mathcal{N}(u_{ref}, \mathcal{C}_{ref})$$

where $u_{ref} = (\lambda_{ref}, D_{ref}, 0)$ and

$$\mathcal{C}_{ref} : u \in \mathbb{R}^2 \times L^2([0,L]) \to (\sigma_\lambda^2 \lambda, \sigma_D^2 D, \mathcal{C}f) \in \mathbb{R}^2 \times L^2([0,L]),$$

such that $\mu_0 << \mu_{ref}$. Indeed, choose $(\lambda_{ref}, D_{ref}) \in \mathbb{R}^2$ and $\sigma_\lambda^2, \sigma_D^2 > 0$ then $\mu_0 << \mu_{ref}$ with

$$\frac{d\mu_0}{d\mu_{ref}}(u) = \frac{d\mu_0^\lambda}{d\mu_{ref}^\lambda}(\lambda) \frac{d\mu_0^D}{d\mu_{ref}^D}(D).$$

This reference will be critical for modes analysis (section 2.4) and MCMC sampling (section 3.2).

2.3 Posterior distribution

The third and last step in the theoretical analysis of Bayesian inversion is to show that this particular setting (forward model and prior distribution) leads to a well defined posterior measure using Bayes theorem. This is the purpose of theorem 2.2 which is a direct application of the theory initially developed in [33]. Consider a dataset $z = (z_i)_{i\in[1,n]}$ which corresponds to observations at different times and locations $(t_i, x_i)_{i\in[1,n]}$ and assume they are produced from the following model (in vector notations):

$$z = \mathcal{G}(u) + \eta, \tag{4}$$

where $\eta \sim \mathcal{N}(0, \sigma_\eta^2 I_n)$ (I_n being the identity matrix of dimension n) and $\mathcal{G} : \mathcal{U} \to \mathbb{R}^n$ is the observation operator, mapping directly the PDE parameter u to the value of the associated solution y at measurement locations (composition of solution map y with Diracs). The following theorem, which is proved in the appendix, establishes the existence of a posterior probability measure μ_z (the solution of our inverse problem), expressing how observations z changed prior beliefs μ_0 on the parameter u.

Theorem 2.2. *Let $\mathcal{G}$ be the observation operator defined in equation 4 and μ_0 the probability measure defined in equation 3 (satisfying $\mu_0(\mathcal{U}) = 1$), then there exists a unique posterior measure μ_z for $u|z$. It is characterized by the following Radon-Nikodym density w.r.t. μ_0:*

$$\frac{d\mu_z}{d\mu_0}(u) = \frac{1}{Z(z)} \exp\left(-\phi(u; z)\right),$$

with

$$\phi(u; z) = \frac{1}{2\sigma_\eta^2} \|z - \mathcal{G}(u)\|^2,$$

and

$$Z(z) = \int_{\mathcal{U}} \exp(-\phi(u; z)\mu_0(dz).$$

Furthermore, the two following integrability conditions:

- $\mathbb{E}^{\mu_0^D}\left[D^{-\frac{1}{2}}\right] < +\infty$,
- $\exists \kappa > 0,\ \mathbb{E}^{\mu_0^f}\left[\exp(\kappa \|f\|_\infty)\right] < +\infty$,

imply the continuity of μ_z in the data w.r.t. Hellinger distance.

Theorem 2.2 gives two distinct results: a) the existence and uniqueness of a posterior (as long as μ_0 is Radon and $\mu_0(\mathcal{U}) = 1$), b) well-posedness of the Bayesian inverse problem under additional integrability conditions of certain functions. In particular, the need for an exponential moment under μ_0^f comes from the re-parametrization in equation 2. If one chooses a different map between f^* and f, this condition may be considerably relaxed (using a polynomial map for instance).

2.4 Maximum a posteriori

In the previous section, we proved the well-posedness of the Bayesian inverse problem under specific integrability conditions. However, the posterior distribution is only known up to a multiplicative constant, through its density w.r.t. μ_0. In our application, we will need to summarize μ_z, which is usually done by the selection of elements of interest in $\mathcal{U}$, such as the conditional mean (CM) or its modes. We saw in theorem 2.2 that the posterior expected value is continuous in the data (consequence of Hellinger continuity, see [15]) but its optimality properties (under the frequentist paradigm) are not yet well-understood in infinite dimension to the best of our knowledge. This is why posterior modes (or Maximum a Posteriori) are more and more considered instead. Indeed, they provide a clear link with Tikhonov-Philips regularization (see [14, 23, 1]) and a practical optimization problem (in case of Gaussian or Besov priors) which can be solved numerically, see theorem 2.3 (which proof is given in the appendix).

Theorem 2.3. *Let μ_0 be the prior probability measure defined in equation 3 and μ_{ref} the Gaussian reference measure from equation 2.2. Suppose additionally that*

$$u \in \mathcal{U} \to ln\left(\frac{d\mu_0}{d\mu_{ref}}(u)\right) \in \mathbb{R}$$

is locally Lipschitz, then the modes of μ_z are exactly the minimizers of the following (generalized) Onsager-Mashlup functional:

$$I(u) := \Phi(u;z) + \frac{1}{2}\|u - u_{ref}\|^2_{\mu_{ref}} - ln\left(\frac{d\mu_0}{d\mu_{ref}}(u)\right),$$

where $\|.\|_{\mu_{ref}}$ and u_{ref} are respectively the norm of the Cameron-Martin space and the mean of μ. A minimizer will be noted $u_{MAP} = (\lambda_{MAP}, D_{MAP}, f_{MAP})$.

The precise application of this theorem to our biological setting is done in section 4.1.

3 Metropolis-Hastings algorithm

As it was previously announced, our motivation for the Bayesian methodology is the quantification of uncertainty, which will be done by simulation. Among the vast catalogue of methods for probability distributions sampling (Sequential Monte-Carlo, Approximate Bayesian Computations, Transport Maps, etc...), Markov chain Monte-Carlo is very popular (MCMC, see [10]) and well defined on function spaces [35] even though ergodicity analysis of such algorithm is still in its infancy [20, 19, 21, 30]. After a short presentation of the functional Metropolis-Hastings algorithm (section 3.1), we will focus on a state-of-the-art Markov kernel designed to sample from Gaussian measures (section 3.2) and adapt it when the prior is not Gaussian, but absolutely continuous with respect to a Gaussian reference.

3.1 Metropolis-Hastings on function spaces

The Metropolis-Hastings algorithm (MH) is a very general [35] method to design Markov chains to sample from a given probability measure. It is based on a two-step process on each iteration:

1. Given a current state $u \in \mathcal{U}$, propose a new candidate v according to a proposal Markov kernel $Q(u, dv)$ (it is a probability distribution on $\mathcal{U}$ for almost any $u \in \mathcal{U}$),
2. Accept the new state v with probability $\alpha(u, v)$ or remain at u.

This algorithm provides a sample distributed under a predefined probability measure μ, if one selects α and Q in a specific way (see [35] for a discussion in general state spaces). For instance, let $\nu(du, dv) = \mu(du)Q(u, dv)$ and $\nu^T(du, dv) = \mu(dv)Q(v, du)$, the Metropolis-Hastings algorithm typically considers the following acceptance probability:

$$\alpha_{MH}(u, v) = \min\left(1, \frac{d\nu^T}{d\nu}(u, v)\right), \tag{5}$$

which, in particular, requires the absolute continuity of ν^T w.r.t. ν (detailed balance of the Markov chain). Contrary to finite dimensional situations, this condition may be difficult to satisfy and a common way to overcome this situation in Bayesian Inverse problems (see [15, 18, 6, 12, 21]) is to select Q revertible w.r.t. μ_0. Indeed, in this case (with $\nu_0(du, dv) = \mu_0(du)Q(u, dv)$):

$$\frac{d\nu^T}{d\nu}(u, v) = \frac{\frac{d\nu^T}{d\nu_0^T}(u, v)}{\frac{d\nu}{d\nu_0}(u, v)} = \frac{\frac{d\mu_z}{d\mu_0}(v)}{\frac{d\mu_z}{d\mu_0}(u)} = \exp\left(\phi(u; z) - \phi(v; z)\right). \tag{6}$$

In theory, the MH algorithm may be implemented with a large family of proposal kernels Q. In practice however, they need to be as efficient as possible and thus adapted to the problem at hand. Two common desirable properties for Q are:

- Adjust the proposal to locally mimic the target distribution μ_z,
- Include a step size to tune acceptance probability.

These two properties may be used to trade-off self-correlation, acceptance rates and convergence speed to high interest areas of the parameter space. The next section presents an algorithm with both properties adapted to Gaussian priors which will then be applied to the problem at hand.

3.2 Geometric MCMC under Gaussian reference

We are now going to detail a specific Markov proposal kernel Q, tailored to sample distributions having a density w.r.t. a Gaussian measure μ_{ref}. Most of

recent work on infinite dimensional MCMC methods are based on the following Langevin stochastic differential equation:

$$\frac{du}{dt} = -\frac{1}{2}K(u)\left(\mathcal{C}_{ref}^{-1}(u - u_{ref}) + \nabla_u \phi(u; z)\right) + \sqrt{K(u)}\frac{dW}{dt}, \tag{7}$$

where $K(u)$ is a (possibly position-dependent) preconditioner, W a cylindrical Brownian motion and $\nabla_u \phi(u; y)$ the gradient in u of the negative log-likelihood. According to [6], a semi-implicit discretization of equation 7 leads to a Markov chain with the following kernel:

$$Q(u, dv) = \mathcal{N}\left(\rho(u - u_{ref}) + u_{ref} + \sqrt{1-\rho^2}\frac{\sqrt{h}}{2}g(u), K(u)\right), \tag{8}$$

where $h > 0$ is a step-size parameter, $\rho = \frac{1-h}{1+h}$ and:

$$g(u) = -K(u)\left[(\mathcal{C}_{ref}^{-1} - K(u)^{-1})(u - u_{ref}) + \nabla_u \phi(u; z)\right].$$

This dynamic explores the parameter space with a balance between Newton-type descent to zones of high density and Gaussian exploration. The philosophy behind this kernel is to use alternative Gaussian reference measures locally adapted to the posterior distribution, since it has been recently showed that higher efficiency is obtained from operator weighted proposals ([28] and later generalized in [6] and [13]). Indeed, highly informative datasets may result in a posterior measure significantly different from the prior in likelihood-informed directions and non-geometric kernels (such as Independent sampler or preconditioned Crank-Nicholson) are ineffective in this case. However, the infinite dimensional manifold Modified Adjusted Langevin Algorithm (∞-mMALA) considers a specific preconditioner:

$$K(u) = \left(\mathcal{C}_{ref}^{-1} + H_\Phi(u)\right)^{-1},$$

where $H_\Phi(u)$ is the Gauss-Newton Hessian matrix of ϕ, which locally adapts to the posterior. This kernel does not preserve the distribution μ_z but is shown to be absolutely continuous w.r.t. the reference measure μ_{ref}, almost-surely in u (under technical assumptions regarding $K(u)$) and the Radon-Nikodym density is:

$$\frac{dQ(u, dv)}{d\mu_{ref}}(v) = \frac{dN\left(\frac{\sqrt{h}}{2}g(u), K(u)\right)}{dN(0, \mathcal{C})}\left(\frac{v - \rho(u - u_{ref}) - u_{ref}}{\sqrt{1-\rho^2}}\right),$$

and noting $w = \frac{v-\rho(u-u_{ref})-u_{ref}}{\sqrt{1-\rho^2}}$ as it is done in [6], it finally comes:

$$\frac{dQ(u, dv)}{d\mu_{ref}}(u, v) = \exp\left(-\frac{h}{8}|K(u)^{-\frac{1}{2}}g(u)|^2 + \frac{\sqrt{h}}{2}\langle K(u)^{-\frac{1}{2}}g(u), K(u)^{-\frac{1}{2}}w\rangle \right.$$
$$\left. -\frac{1}{2}\langle w, H_\Phi(u)w\rangle\right)\left|\mathcal{K}^{\frac{1}{2}}K(u)^{-\frac{1}{2}}\right|.$$

Finally, the acceptance probability associated to the Markov kernel from equation 8 is

$$\alpha(u,v) = \min\left(1, \frac{\frac{dQ}{d\mu_{ref}}(v,u)\frac{d\mu_z}{d\mu_0}(v)}{\frac{dQ}{d\mu_{ref}}(u,v)\frac{d\mu_z}{d\mu_0}(u)}\right).$$

This algorithm is well-defined on function spaces (reversibility is ensured w.r.t. μ_0), thus it is robust to discretization as required. The ∞-mMALA proposal may be computationally expensive, as it requires to compute both gradient $\nabla\Phi$, Gauss-Newton Hessian H_Φ and the Cholesky decomposition of $K(u)^{-1}$ at each step. However, different dimension reduction techniques can be used (split in [6] or likelihood-informed in [13]) to reduce the computational burden. A second alternative is to choose a constant preconditioner, located at a posterior mode for instance (similar to HMALA in [13] and gpCN in [30]).

4 Numerical application

We now turn to the practical implementation of the previous methodology on the problem of reverse-engineering for post-transcriptional gap-gene in Drosophila Melanogaster. First, we precise our choice of distributions for the parameters compatible with previous assumptions, and give a random series representation for f and precise the generalized Onsager-Mashlup functional. Then, we provide quantitative results on the dataset taken from [5], consisting in protein concentration measurements irregularly spread in space and time.

4.1 Choice of measures

We will now specify our choice of prior measure μ_0 with justifications:

- μ_0^λ, μ_0^D. Concerning the decay parameter, the only constraint given in the problem so far is positivity and Lebesgue density. However, the diffusion must also satisfy an integrability condition from theorem 2.2, which is clearly the case for uniform distributions. Finally, we choose $\mu_0^\lambda = \mathcal{U}([0,\lambda_m])$ and $\mu_0^D = \mathcal{U}([0,D_m])$, maximum parameters λ_m and D_m being tuned to 0.5 to be sufficiently large w.r.t. previous estimations from [5].

- μ_0^f. The prior measure μ_0^f will be chosen as a centred Gaussian measure on $L^2([0,T]\times[0,L])$, with covariance operator on $L^2([0,T]\times[0,L])$:

  $$\mathcal{C} = (-\Delta)^{-\alpha},$$

  where $\alpha \in \mathbb{R}^+$ is a smoothness parameter tuned to ensure almost-surely continuity of the samples. The precise eigen-decomposition is given as

follows $i_1, i_2 \geq 1$:

$$\begin{aligned} \varphi_{i_1,i_2}(x,t) &= \frac{1}{\sqrt{LT}} \sin\left(i_1 \frac{\pi}{L} x\right) \sin\left(i_2 \frac{\pi}{T} t\right), \\ \lambda_{i_1,i_2} &= \left(\left(\frac{i_1}{\pi L}\right)^2 + \left(\frac{i_2}{\pi T}\right)^2\right)^{-\alpha}. \end{aligned}$$

- With these choices done, the Radon-Nikodym density of the prior distribution w.r.t. the reference measure is

$$\frac{d\mu_0}{d\mu_{ref}}(u) = \frac{2\pi\sigma_\lambda\sigma_D}{\lambda_m D_m} \exp\left(\frac{(\lambda-\lambda_{ref})^2}{2\sigma_\lambda^2} + \frac{(\lambda - D_{ref})^2}{2\sigma_D^2}\right) \mathbb{1}_{[0,\lambda_m]}(\lambda)\mathbb{1}_{[0,D_m]}(D),$$

which is locally Lipschitz (mean-value theorem). We now tune the parameters of μ_{ref}, by simply choosing $\lambda_{ref} = \frac{\lambda}{2}$, $\sigma_\lambda^2 = \frac{\lambda_m^2}{12}$, $D_{ref} = \frac{D_m}{2}$ and $\sigma_D^2 = \frac{D_m^2}{12}$ (minimizing Kullback-Leibler divergence). Finally, we are capable of specifying the exact form of the generalized Onsager-Mashlup functional using theorem 2.3:

$$I(u) = \frac{1}{2\sigma_\eta^2}\|z - \mathcal{G}(u)\|^2 + \frac{1}{2}\|f\|^2_{\mu_0^f}.$$

In particular, parameters λ and D are only influenced by their range and contribution to the likelihood (the uniform prior is non-informative), contrary to the source f.

4.2 Prior and solution map discretization

The analysis conducted in all previous sections happens to be valid for infinite dimensional quantities. In practice however, one needs to discretize for numerical experiments. In this work, the solution map is approximated using finite elements in space (FEniCS library in Python, see [2] and [27]) and finite differences in time. We use 100 basis functions and 30 time steps on a desktop computer (Intel i7-3770 with 8Gb of RAM memory)[1]. We set $L = 100$ and $T = 100$ is the final time. Concerning the prior measure, we use a truncated Karhunen-Loeve basis of f to simulate from it

$$\tilde{f} = \sum_{1 \leq i_1, i_2 \leq N} \sqrt{\lambda_{i_1,i_2}} \xi_i \varphi_{i_1,i_2}, \tag{9}$$

where $(\xi_{i_1,i_2})_{1\leq i_1,i_2\leq N}$ are i.i.d. $\mathcal{N}(0,1)$ random variables. We consider $N = 10$ (100 basis functions) thus $\tilde{u} = (\lambda, D, \tilde{f})$ is of dimension 102. All quantities related to negative log-likelihood derivatives (Gradient and Gauss-Newton Hessian matrix) are numerically computed using discrete adjoint methods (see [24] or [22]) to keep scalability in N. The initial point in the chain is chosen at the MAP location, obtained by minimization of the functional in equation 4.1 (Prior based initialization results in long burnin phase). Practical optimization is done using L-BFGS-B algorithm from the Scipy library [11].

[1] All codes are available online at `https://github.com/JeanCharlesCroix/2018_Bayesian_estimation`.

4.3 Results

We now turn to our main objective, the inversion and uncertainty quantification of gap-gene protein concentration from [4]. The dataset consists of 508 different measures which are non-uniformly spread in time and space (precise repartition can be seen in figure 4). The noise variance parameter is estimated using the following routine (10 iterations, 3 multi-start each):

1. Find u_{MAP} minimizing I using the current noise level σ_η^2.
2. Update the current variance estimation $\sigma_\eta^2 = \frac{2}{n-1}\|z - \mathcal{G}(u_{MAP})\|^2$.
3. Go back to 1.

The resulting estimated noise level is $\sigma_\eta^2 = 11,77$. With this estimated value, we compute our initial MAP estimate (numerical minimization of the Onsager-Machlup functional) and use it as initial point in the MCMC sampling. The Markov chain is ran for 11000 total iterations and the resulting traceplot is given in figure 1 for negative log-likelihood, decay, diffusion and first three components of f. The first thousand iterations are used as burnin and according to the

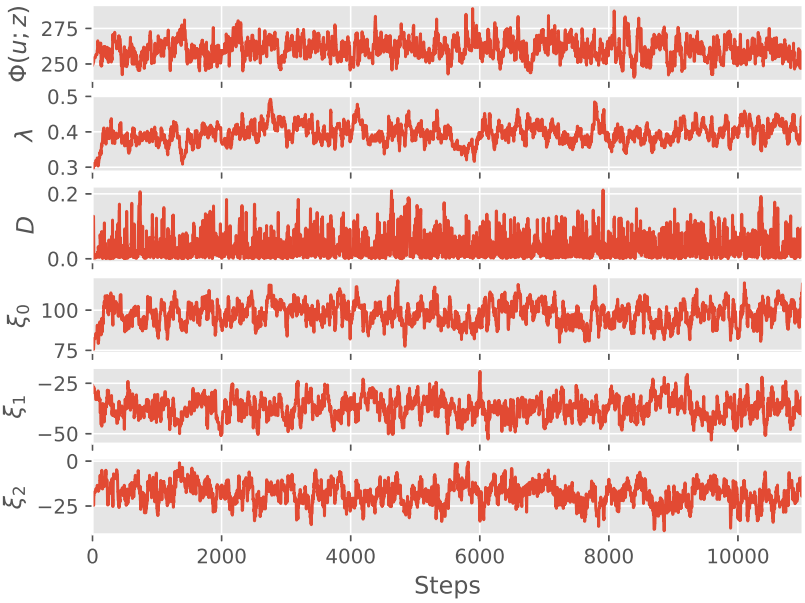

Figure 1: Trace plots of $\Phi(u;z)$, λ, D, ξ_0, ξ_1 and ξ_2 (the first 1000 iterations are burned).

autocorrelation function (figure 2), we choose to keep one iteration out of a

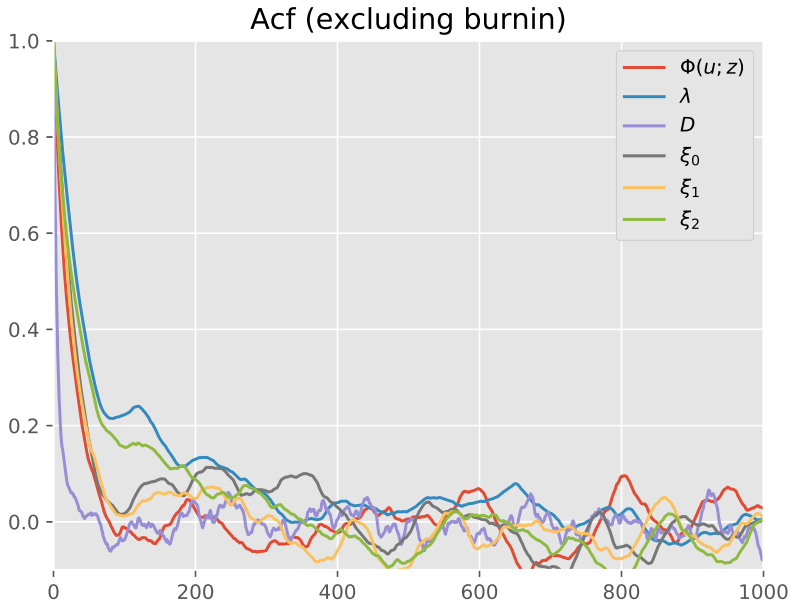

Figure 2: Autocorrelation of negative of $\Phi(u;z)$, λ, D, ξ_0, ξ_1 and ξ_2, excluding burnin sample.

hundred as posterior sample (thinning). From this, we compute both posterior mean and MAP estimates, the precise values of decay, diffusion, negative log-likelihood and Onsager-Machlup functional being given in table 1. Additionally

Parameter	λ	D	$\Phi(u;z)$	$I(u)$
MAP (Onsager-Machlup)	$3.00 * 10^{-1}$	$1.0 * 10^{-8}$	248.91	440.51
MAP (MCMC)	$4.25 * 10^{-1}$	$3.23 * 10^{-2}$	246.72	460.73
Conditional mean (MCMC)	$3.97 * 10^{-1}$	$3.30 * 10^{-2}$	243.54	422.41

Table 1: Values of decay, diffusion, negative log-likelihood and Onsager-Machlup functional for different estimators.

to the estimated values, one can also look at the marginal distribution on figure 3. Concerning the MAP estimator (figure 4), we recover both events described in [4] and [5], that is 2 pikes of protein concentrations. The first happens on the anterior part of the embryo in the early experiment ($x = 35, t = 35$). The second is much more intense and happens in the posterior part during the second half of the experiment. The estimated source explains these with an intense and localized increase in concentration. Finally, the uncertainty on both source and solution around data seems to be really low, which provides a good confidence on

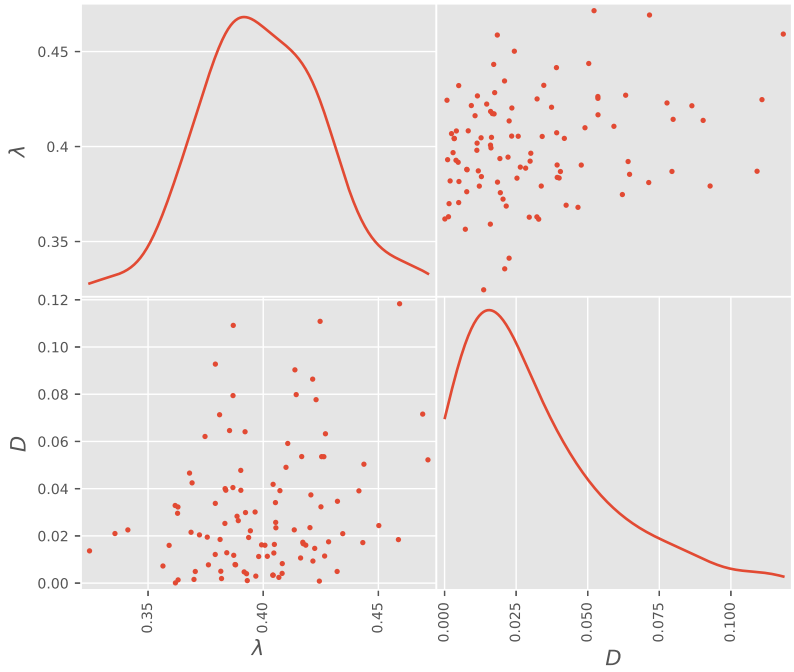

Figure 3: Marginal posterior distributions for $\lambda|z$ and $D|z$.

the level of mRNA at this time and part of the embryo (see figure 5). However, the point-wise variance on the solution y remains important before the first observations.

5 Conclusion

In this work, we applied the Bayesian inverse problem methodology from [33] to a practical Biological dynamical system. Doing so, we provide a rigorous and detailed analysis of the forward model, existence and continuity of the posterior measure, characterization of maximum a posteriori (MAP) estimates and state-of-the-art MCMC methodology. Because the forward MAP is non-linear, the unicity of posterior modes is unclear and it appears that local maximas are present. Nevertheless, the Bayesian methodology provides both a regularized solution to the problem, while giving a quantification of uncertainty. However, the estimation of prior hyper-parameters is still out of reach, giving poor confidence in the estimated variance. This direction requires further research, that we will try to address in a future work.

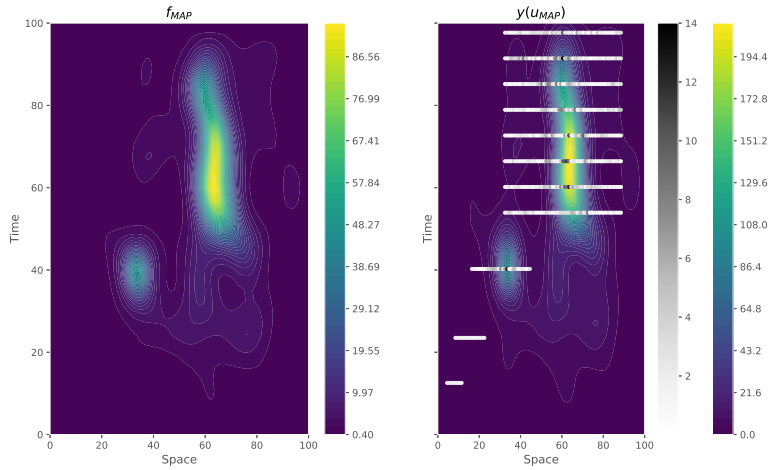

Figure 4: MAP estimator from MCMC sampling. Left: Estimated source term f_{MAP}, right: estimated solution $y(u_{MAP})$ with absolute error at data locations. Grey bar represents the error level between data and $y(u_{MAP})$.

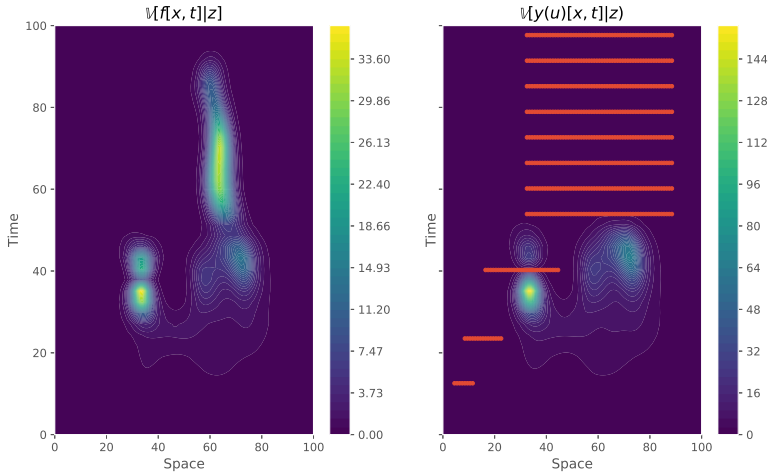

Figure 5: Posterior point-wise variance of source (left figure) and solution (right figure).

6 Aknowledgments

This work has been supported by Colciencias and ECOS-Nord under the grant C15M04. The authors would like to thank Dr. Xavier Bay for his precious advice and proof reading.

A Proofs

Proof of theorem 2.1. Using standard notations in PDE theory, let $\Omega =]0, L[$, $H = L^2(\Omega)$, $V = H_0^1(\Omega)$, $V^* = H^{-1}(\Omega)$, $\mathcal{P} = \mathbb{R}\times]0, \infty[\times L^2([0,T], V^*)$ and

$$W([0,T], H, V) = \left\{ y \in L^2([0,T], H),\ y' \in L^2([0,T], V^*) \right\},$$

equipped with the norm $\|y\|_{W([0,T],H,V)} = \left(\|y\|^2_{L^2([0,T],V)} + \|y'\|^2_{L^2([0,T],V^*)} \right)^{\frac{1}{2}}$, then the weak form associated to equation 1 has the following reduced form:

$$\text{Find } y \in W([0,T], H, V),\ \langle F(y,u), v\rangle = 0,\ \forall v \in L^2([0,T], V),$$

with

$$\begin{aligned}\langle F(y,u), v\rangle &= \int_0^T \langle y_t(t) - f(t), v(t)\rangle_{V^*,V} + \lambda \langle y(t), v(t)\rangle_H + D\langle y_x(t), v_x(t)\rangle_H dt, \\ &= \int_0^T \langle y_t(t) - f(t), v(t)\rangle_{V^*,V} + B(y(t), v(t), \lambda, D) dt.\end{aligned}$$

- **Existence of a solution map**. The PDE operator is uniformly parabolic whenever $D > 0$, thus there exists a unique weak solution of equation 1 in $W(0,T,H,V)$ for every $u \in \mathcal{P}$ (see chapter 7 in [16] for instance). Moreover, we have the following energy estimate:

  $$\|y(u)\|_{W(0,T,H,V)} \leq \frac{C}{\sqrt{D}} \|f\|_{L^2([0,T],V^*)}$$

  with C a constant independent of u.

- **Second order Fréchet differentiability of the solution map**. Let $u, h^u \in \mathcal{P}$ such that $u + h^u \in \mathcal{P}$, and $h^y \in W(0,T,H,V)$ then $\forall v \in L^2([0,T], V)$:

  $$\langle F(y + h^y, u + h^u) - F(y,u), v\rangle = \langle DF(y,u)[h^y, h^u], v\rangle + c(h^u, h^y)$$

  with

  $$|c(h^u, h^y)| = \left| h^\lambda \int_0^T \langle h^y, v\rangle_H dt + h^D \int_0^T \langle h^y_x, v_x\rangle_H dt \right| \leq C \|v\|_{L^2([0,T],V)} \|h\|^2_U,$$

with C an other constant independent from u and:

$$\begin{aligned}\langle DF(y,u)[h^y,h^u],v\rangle = &\int_0^T \langle h_t^y, v\rangle_{V^*,V} dt + \int_0^T h^\lambda \langle y, v\rangle_H dt \\ &+ \int_0^T h^D \langle y_x, v_x\rangle_H dt + \int_0^T \lambda \langle h^y, v\rangle_H dt \\ &+ D\int_0^T \langle h_x^y, v_x\rangle_V dt - \int_0^T \langle f, v\rangle_{V^*,V} dt.\end{aligned}$$

Moreover, we have:

$$|\langle DF(y,u)[h^y,h^u],v\rangle| \leq \|(h^u,h^y)\|\|v\|$$

thus DF is bounded, which shows that F is Fréchet-differentiable on $\mathcal{P}$. Consider F_y the partial derivative of F w.r.t. its first variable:

$$\langle F_y(y,u)h^y, v\rangle = \int_0^T \langle h_t^y, v\rangle + \lambda\langle h^y, v\rangle + D\langle h_x^y, v_x\rangle dt.$$

which defines a unique solution $h \in W(0,T,H,V)$ whenever $D > 0$ (using same arguments than previously). Here, F_y^{-1} is clearly bounded. Because F is differentiable and F_y^{-1} exists and is bounded, the implicit function theorem applies and y is differentiable on $\mathcal{P}$. The second order differentiability uses the same arguments.

- **Local Lipschitz continuity of the solution map**. Let $u \in \mathcal{P}$, $r > 0$ such that $\mathcal{B}(u,r) \subset \mathcal{P}$ and $(u_1,u_2) \in \mathcal{B}(u,r)$. There exists two unique solution with respect to (u_1,u_2) satisfying $\forall i \in \{1,2\}$, $\forall v \in L^2([0,T],V)$:

$$\langle y_i', v\rangle_{V^*,V} + B(y_i, v, \lambda_i, D_i) = \langle f_i, v\rangle_{V^*,V} \text{ for almost every } t \in [0,T],$$

which leads to $\forall v \in L^2([0,T],V)$ and almost-every $t \in [0,T]$:

$$\langle y_1' - y_2', v\rangle_{V^*,V} + B(y_1, v, \lambda_1, D_1) - B(y_2, v, \lambda_2, D_2) = \langle f_1 - f_2, v\rangle_{V^*,V}.$$

and equivalently:

$$\begin{aligned}&\langle y_1' - y_2', v\rangle_{V^*,V} + B(y_1 - y_2, v, \lambda_1, D_1) \\ &= \langle f_1 - f_2, v\rangle_{V^*,V} + B(y_2, v, \lambda_1, D_1) - B(y_2, v, \lambda_2, D_2).\end{aligned}$$

Now, letting $v = y_1 - y_2$ and using the coercivity of B we get:

$$\begin{aligned}&\langle y_1' - y_2', y_1 - y_2\rangle_{V^*,V} + D_1\|y_1 - y_2\|_V^2 \\ &\leq \langle f_1 - f_2, y_1 - y_2\rangle_{V^*,V} + B(y_2, y_1 - y_2, \lambda_1, D_1) - B(y_2, y_1 - y_2, \lambda_2, D_2).\end{aligned} \tag{10}$$

Now, we use the identity $\frac{d}{dt}\left(\frac{1}{2}\|y_1 - y_2\|_H^2\right) = \langle y_1' - y_2', y_1 - y_2\rangle_{V^*,V}$ and drop the positive term in the left hand side to get

$$\begin{aligned}&\frac{d}{dt}\left(\frac{1}{2}\|y_1 - y_2\|_H^2\right) \\ &\leq \langle f_1 - f_2, y_1 - y_2\rangle_{V^*,V} + B(y_2, y_1 - y_2, \lambda_1, D_1) - B(y_2, y_1 - y_2, \lambda_2, D_2).\end{aligned}$$

We integrate between 0 and t, use Cauchy-Schwartz and Poincarré's inequalities and $y_1(0) = y_2(0) = 0$ to obtain the following estimation

$$\frac{1}{2}\|y_1 - y_2\|_H^2 \leq C\|y_1 - y_2\|_{L^2([0,T],V)}\|u_1 - u_2\|,\ t - a.e.$$

The same reasoning applies now to equation 10 to get

$$\|y_1 - y_2\|_{L^2([0,T],V)} \leq C\|u_1 - u_2\|.$$

We start back from equation A to get:

$$\begin{aligned}&\langle y_1' - y_2', v\rangle_{V^*,V} = -B(y_1 - y_2, v, \lambda_1, D_1)\\ &+ \langle f_1 - f_2, v\rangle_{V^*,V} + B(y_2, v, \lambda_1, D_1) - B(y_2, v, \lambda_2, D_2),\end{aligned}$$

and taking the supremum on the unit ball of V before integrating:

$$\|y_1' - y_2'\|_{L^2([0,T],V^*)} \leq C_2\|u_1 - u_2\|$$

which completes the proof.

□

Proof of theorem 2.2. Following the program in [15], let us first show that μ_z exists and is unique and then the continuity of the posterior with respect to the Hellinger distance.

- (Existence and unicity of μ_z). Let μ_0 be a probability measure defined as in equation 3 and consider the following Gaussian negative log-likelihood:

$$\Phi(u; z) = \frac{1}{2\sigma_\eta^2}\|z - \mathcal{G}(u)\|^2.$$

Since $\mathcal{G}$ is differentiable (theorem 2.1), Φ is measurable w.r.t. μ_0 for all $z \in \mathbb{R}^n$. Consider now the following integral

$$Z(z) = \int_{\mathcal{U}} \exp(-\Phi(u; z))\mu_0(du).$$

The negative log-likelihood being positive, the integral is finite. Furthermore, we have $\forall (u, z) \in \mathcal{U} \times \mathbb{R}^n,\ \Phi(u; z) < +\infty$ thus $Z(z) > 0$ for every $z \in \mathbb{R}^n$. In other words, the following function

$$\frac{\exp(-\Phi(u; z))}{Z(z)}$$

defines a probability density function w.r.t. μ_0, and the associated measure is μ_z, the (unique) posterior distribution of $u|z$.

- (Continuity in z). It remains to show that the posterior distribution is continuous in z w.r.t. the Hellinger distance. Following the lines of [33, 15] we proove the following sufficient condition

$$|\Phi(u;z_1) - \Phi(u;z_2)| \leq g(u)\|z_1 - z_2\|,$$

with $g \in L^1(\mu_0)$.

$$\begin{aligned} |\Phi(u;z_1) - \Phi(u;z_2)| &= \frac{1}{2\sigma_\epsilon^2}\left|\|z_1\|^2 - \|z_2\|^2 + 2\langle z_2 - z_1, \mathcal{G}(u)\rangle\right|, \\ &= \frac{1}{2\sigma_\epsilon^2}\left|(\|z_1\| - \|z_2\|)(\|z_1\| + \|z_2\|) + 2\langle z_2 - z_1, \mathcal{G}(u)\rangle\right|, \\ &\leq \frac{1}{\sigma_\eta^2}(r + \|\mathcal{G}(u)\|)\|z_1 - z_2\|. \end{aligned}$$

Now, using continuous injection between spaces, we have

$$\|\mathcal{G}(u)\| \leq n\|y(u)\|_\infty \leq nC_1\|y(u)\|_W \leq nC_2\frac{\|f\|_{L^2}}{D},$$

the last inequality is given by the energy estimate of theorem 2.1. It remains to see that

$$g(u) := \frac{1}{\sigma_\eta^2}\left(r + \frac{\|f\|}{D}\right)$$

is in $L^1(\mu_0)$, that is

$$\mathbb{E}^{\mu_0}\left[g(u)\right] = \mathbb{E}^{\mu_0^D}\left[\frac{1}{D}\right]\mathbb{E}^{\mu_0^f}[\|f\|_{L^2}].$$

The first term in the right hand side product is finite by assumption on μ_0^D. The second term is finite as well since

$$\|\exp(f^*)\|_{L^2} \leq \|\exp(f^*)\|_\infty \leq \exp(\|f^*\|_\infty)$$

which is μ_0^f integrable by the assumption on μ_0^f. Finally, Fubini's theorem gives the result on the integrability of g. Now, it remains to follow the proof from [33] as is.

□

Proof of theorem 2.3. The posterior measure μ_z is absolutely continuous w.r.t. μ with the following Radon-Nikodym density:

$$\frac{d\mu_z}{d\mu_{ref}}(u) = \frac{d\mu_z}{d\mu_0}(u)\frac{d\mu_0}{d\mu_{ref}}(u) = \frac{1}{Z(z)}\exp(-\tilde{\Phi}(u;z)).$$

where $\tilde{\Phi}(u;z) = \Phi(u;z) - \ln\left(\frac{d\mu_0}{d\mu_{ref}}(u)\right)$. Let us now show that Φ is locally Lispchitz in its first argument:

$$\begin{aligned}
|\Phi(u_1;z) - \Phi(u_2;z)| &= \frac{1}{2\sigma_\eta^2}\left|\|z-\mathcal{G}(u_1)\|^2 - \|z-\mathcal{G}(u_2)\|^2\right|, \\
&= \frac{1}{2\sigma_\eta^2}\left|\|\mathcal{G}(u_1)\|^2 - \|\mathcal{G}(u_2)\|^2 + 2\langle z, \mathcal{G}(u_2)-\mathcal{G}(u_1)\rangle\right|, \\
&\leq \frac{\|\mathcal{G}(u_1)\| + \|\mathcal{G}(u_2)\| + 2\|z\|}{2\sigma_\eta^2}\|\mathcal{G}(u_1)-\mathcal{G}(u_2)\|,
\end{aligned}$$

and since $\|\mathcal{G}(u_1)-\mathcal{G}(u_2)\| \leq C\|y(u_1)-y(u_2)\|$, we conclude that Φ and thus $\tilde{\Phi}$ are locally Lipschitz. Now, we follow the lines of [14] it comes:

$$\begin{aligned}
\frac{J^\delta(u_1)}{J^\delta(u_2)} &= \frac{\int_{B^\delta(u_1)}\exp(-\tilde{\Phi}(u;z))\mu_{ref}(du)}{\int_{B^\delta(u_2)}\exp(-\tilde{\Phi}(v;z))\mu_{ref}(dv)}, \\
&= \frac{\int_{B^\delta(u_1)}\exp(-\tilde{\Phi}(u;z)+\tilde{\Phi}(u_1;z))\exp(-\tilde{\Phi}(u_1;z))\mu_{ref}(du)}{\int_{B^\delta(u_2)}\exp(-\tilde{\Phi}(v;z)+\tilde{\Phi}(u_2;z))\exp(-\tilde{\Phi}(u_2;z))\mu_{ref}(dv)}.
\end{aligned}$$

Now,

$$\frac{J^\delta(u_1)}{J^\delta(u_2)} \leq \exp\left(\delta C - \tilde{\Phi}(u_1,z) + \tilde{\Phi}(u_2;z)\right)\frac{\int_{B^\delta(u_1)}\mu_{ref}(du)}{\int_{B^\delta(u_2)}\mu_{ref}(dv)}$$

and finally

$$\limsup_{\delta\to 0}\frac{J^\delta(u_1)}{J^\delta(u_2)} \leq \exp\left(-I(u_1)+I(u_2)\right).$$

A similar argument leads to

$$\liminf_{\delta\to 0}\frac{J^\delta(u_1)}{J^\delta(u_2)} \geq \exp\left(-I(u_1)+I(u_2)\right).$$

We conclude that $\lim_{\delta\to 0}\frac{J^\delta(u_1)}{J^\delta(u_2)} = \exp\left(-I(u_1)+I(u_2)\right)$. For a fixed value u_2, this quantity is maximized when u_1 is a minimizer of I, the proof is then complete. □

References

[1] Sergios Agapiou, Martin Burger, Masoumeh Dashti, and Tapio Helin. Sparsity-promoting and edge-preserving maximum a posteriori estimators in non-parametric Bayesian inverse problems. pages 1–36, 2017.

[2] Martin S. Alnaes, Jan Blechta, Johan Hake, August Johansson, Benjamin Kehlet, Anders Logg, Chris Richardson, Johannes Ring, Marie E. Rognes, and Garth N. Wells. The FEniCS Project Version 1.5. *Archive of Numerical Software*, 3(100):9–23, 2015.

[3] Mauricio A Alvarez, David Luengo-Garcia, and Neil D. Lawrence. Latent Forces Models using Gaussian Processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11):1–20, 2013.

[4] Kolja Becker. *A Quantitative Study of Translational Regulation in Drosophilia Segment Determination*. PhD thesis, University of Mainz, 2012.

[5] Kolja Becker, Eva Balsa-Canto, Damjan Cicin-Sain, Astrid Hoermann, Hilde Janssens, Julio R. Banga, and Johannes Jaeger. Reverse-Engineering Post-Transcriptional Regulation of Gap Genes in Drosophilia melanogaster. *PLOS Computational Biology*, 9(10), 2013.

[6] Alexandros Beskos, Mark A Girolami, Shiwei Lan, Patrick E. Farrell, and Andrew M. Stuart. Geometric MCMC for infinite-dimensional inverse problems. *Journal of Computational Physics*, 335:327–351, 2017.

[7] Lorenz Biegler, George Biros, Omar Ghattas, Matthias Heinkenschloss, David Keyes, Bani Mallick, Youssef M. Marzouk, Luis Tenorio, Bart van Bloemen Waanders, and Karen E. Willcox. *Large-Scale Inverse Problems and Quantification of Uncertainty*. Wiley edition, 2011.

[8] Vladimir Igorevich Bogachev. *Gaussian Measures*, volume 62. 1998.

[9] Haim Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, volume 2010. Springer, 2011.

[10] Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng. *Handbook of Markov Chain Monte Carlo*. Crc press edition, 2011.

[11] Richard H. Byrd, Peihuang Lu, Jorge Nocedal, and Ciyou Zhu. A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208, 1995.

[12] S. L. Cotter, Gareth O. Roberts, Andrew M. Stuart, and D. White. MCMC Methods for Functions: Modifying Old Algorithms to Make Them Faster. *Statistical Science*, 28(3):424–446, aug 2013.

[13] Tiangang Cui, Kody J. H. Law, and Youssef M. Marzouk. Dimension-independent likelihood-informed MCMC. *Journal of Computational Physics*, 304:109–137, 2016.

[14] Masoumeh Dashti, Kody J. H. Law, Andrew M. Stuart, and J. Voss. MAP estimators and their consistency in Bayesian nonparametric inverse problems. *Inverse Problems*, 29(9):095017, 2013.

[15] Masoumeh Dashti and Andrew M. Stuart. The Bayesian Approach To Inverse Problems, 2015.

[16] Lawrence C. Evans. *Partial differential equations*. American Mathematical Society, 2010.

[17] Roger Ghanem, David Higdon, and Houman Owhadi. *Handbook of Uncertainty Quantification*. 2017.

[18] Mark A Girolami and Ben Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2):123–214, mar 2011.

[19] M. Hairer, Andrew M. Stuart, and J. Voss. Analysis of SPDEs arising in path sampling part II: The nonlinear case. *Annals of Applied Probability*, 17(5-6):1657–1706, 2007.

[20] M. Hairer, Andrew M. Stuart, J. Voss, and P. Wiberg. Analysis of SPDEs arising in path sampling part I: The Gaussian case. *Communications in Mathematical Sciences*, 3(4):587–603, 2005.

[21] Martin Hairer, Andrew M. Stuart, and Sebastian J. Vollmer. Spectral gaps for a Metropolis-Hastings algorithm in infinite dimensions. *Annals of Applied Probability*, 24(6):2455–2490, 2014.

[22] Matthias Heinkenschloss. Numerical solution to implicitely constrained optimization problems. Technical report, 2008.

[23] Tapio Helin and Martin Burger. Maximum a posteriori probability estimates in infinite-dimensional Bayesian inverse problems. *Inverse Problems*, 31(8), 2015.

[24] Michael Hinze, René Pinnau, Michael Ulbrich, and Stephan Ulbrich. *Optimization with PDE Constraints*. Springer, 2009.

[25] Bamdad Hosseini and Nilima Nigam. Well-Posed Bayesian Inverse Problems: Priors with Exponential Tails. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):436–465, 2017.

[26] Victor Isakov. *Inverse Problems for Partial Differential Equations*, volume 127 of *Applied Mathematical Sciences*. Springer International Publishing, Cham, 2017.

[27] Hans Petter Langtangen and Anders Logg. *Solving PDEs in Python - The FEniCS Tutorial Volume I*. 2017.

[28] Kody J. H. Law. Proposals which speed up function-space MCMC. *Journal of Computational and Applied Mathematics*, 262:127–138, 2014.

[29] Carl E. Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning.*, volume 14. 2004.

[30] Daniel Rudolf and Björn Sprungk. On a generalization of the preconditioned Crank-Nicolson Metropolis algorithm. *Foundations of Computational Mathematics*, 18(2):309–343, 2018.

[31] Simo Särkkä, Mauricio A Alvarez, and Neil D. Lawrence. Gaussian Process Latent Force Models for Learning and Stochastic Control of Physical Systems. 2017.

[32] Thomas Schuster, Barbara Kaltenbacher, Bernd Hofmann, and Kamil S. Kazimierski. *Regularization Methods in Banach Spaces*. De Gruyter, 2012.

[33] Andrew M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numerica*, 19:451–559, may 2010.

[34] Tim J Sullivan. *Introduction to uncertainty quantification*. Springer, 2015.

[35] Luke Tierney. A note on Metropolis-Hastings kernels for general state spaces. *Annals of Applied Probability*, 8(1):1–9, 1998.