Bayesian parameter identification in Cahn–Hilliard models for biological growth

Christian Kahle ‡ Kei Fong Lam † Jonas Latz ‡ Elisabeth Ullmann ‡

May 10, 2018

Abstract

In this paper we consider the identification of internal parameters in a diffuse interface model for tumour growth. The tumour is modelled by a phase field that acts as a smooth indicator function to distinguish the healthy cells and the tumour cells inside the tissue. Our model is a variant of the model proposed in [Garcke et al., Math. Models Methods Appl. Sci., 26 (2016), pp. 1095–1148] and contains three phenomenological parameters: the proliferation rate for the tumour growth, the consumption rate for the nutrition consumption by the tumour, and a parameter to control the influence of chemotaxis, i.e., the growing of the tumour in the direction of the gradient of the nutrition concentration. These parameters need to be adapted to particular observations, and we apply concepts from Bayesian inversion for this task.

After a brief summary of the analytical properties of the model we discuss the well-posedness of the posterior measure for general prior measures. We approximate the posterior measure by the sequential Monte Carlo approach with tempering, and present test results for two numerical examples. In the first example we simulate a tumour with given parameters and add artificial noise to obtain synthetic data. In the second example we use real-world measurements of the volume of spherical tumours and estimate the corresponding internal variables.

Key words. Tumour modelling, Bayesian inversion, Cahn–Hilliard, Sequential Monte Carlo

AMS subject classification. 65M32, 92B05, 92C17, 35Q92, 35R30

1 Introduction

The aim of mathematical modelling is to reduce a complex physical phenomenon into a description which can be further studied and analyzed. Moreover, it provides practitioners the means to plan and test their experiments while minimizing the use of valuable resources. Until recently, the discipline of medical science relied heavily on experiments

*The first and fourth author gratefully acknowledge the support by the Deutsche Forschungsgemeinschaft (DFG) through the International Research Training Group IGDK 1754 “Optimization and Numerical Analysis for Partial Differential Equations with Nonsmooth Structures”. The third and fourth author gratefully acknowledge the support by the DFG and Technische Universität München through the International Graduate School of Science and Engineering within project 10.02 BAYES.

‡Zentrum Mathematik, Technische Universität München, 85748 Garching bei München, Germany ({Christian.Kahle,jlatz,Elisabeth.Ullmann}@ma.tum.de).

†Department of Mathematics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong (kflam@math.cuhk.edu.hk)
to obtain statistical data as a basis for understanding the behaviour of complex biomedical systems, and to design new drugs for the treatment of diseases. The advent of high-performance computing, big data and bioinformatics, coupled with advances in mathematical and statistical theories and methodologies, lead to the emergence of patient-specific diagnoses, and treatment driven by computational models.

A complex biomedical phenomenon that is still not fully understood is the growth of cancer. A tumour is a mass of tissue that arises when certain inhibition proteins in the cells have been switched off by genetic mutations. This leads to unregulated growth that is limited only by the amount of nutrients in the surrounding environment. Tumours display characteristics that are fundamentally different from normal cells. Tumour cells are able to ignore apoptosis (programmed cell death) signals, remain elusive to attacks from the immune system, and, most dangerously, have the ability to induce the growth of new blood vessels towards itself (angiogenesis). This leads to the spreading of cancer to other parts of the body, and the formation of secondary tumours (metastasis).

The study of tumour growth can be roughly divided according to the physical and chemical phenomena occurring at three scales [Oden et al., 2016]: the tissue scale which is commonly observed in experiments involving movement of cells (such as metastasis and growth into the extracellular matrix) and nutrient diffusion; the cellular scale consisting of activities and interactions between individual cells such as mitosis and the activation of receptors; and sub-cellular scale where genetic mutations and DNA degradation occur. We focus on the tissue-scaled phenomena, as they are the first to be detected in a routine diagnosis, and can be described fairly well with help of continuum models consisting of differential equations.

Since the seminal work in [Burton, 1966] and [Greenspan, 1972] where simple mathematical models for tumour growth are employed, there has been an explosion in the number of models proposed for modelling the multiscale nature of cancer, see for instance [Cristini and Lowengrub, 2010, Deisboeck and Stamatakis, 2010, Oden et al., 2016] and the references cited therein. The diversity of model variants reflects the difficulties when we try to identify key biological phenomena that are responsible for experimental observations.

In this work we restrict our attention to continuum models that can capture metastasis since this is an important hallmark of cancer. Continuum models often rely on a mathematical description to distinguish tumour tissue from healthy host tissues. To be able to capture metastasis the models have to allow for some form of topological change of the separation layers between the tumour and the host tissues. The classical description represents the separation layers as idealized hypersurfaces, known also as the sharp interface approach. In this case complicated boundary conditions have to be enforced to model the mass transfer between tumour and host cells. Unfortunately, at the events of metastasis, the separation layers can no longer be represented as a hypersurface, and the classical sharp interface approach breaks down. To overcome this difficulty some authors proposed a diffuse interface approach (see for example [Cristini et al., 2009, Cristini and Lowengrub, 2010, Frieboes et al., 2010, Garcke et al., 2018, Garcke et al., 2016, Hawkins-Daarud et al., 2013, Lima et al., 2016, Wise et al., 2008] and the references cited therein) which is well-known for being able to handle changes in the topology.

1.1 Diffuse interface model

Now we introduce the diffuse interface model to be studied and outline the phenomenological tumour behaviours that are captured. Let $\Omega \subset \mathbb{R}^2$ denote a bounded domain with boundary $\partial \Omega$. For an arbitrary but fixed constant $T > 0$ we denote by $Q$ the space-time
cylinder \( Q := \Omega \times (0, T) \), and let \( \Sigma := \partial \Omega \times (0, T) \). Let \( n \) denote the outer unit normal vector of \( \partial \Omega \), and let \( \partial_n f := \nabla f \cdot n \) denote the co-normal derivative of a function \( f \). We consider the following generalized Cahn–Hilliard model:

\[
\begin{align}
\varphi_t &= \text{div} (m(\varphi) \nabla (\mu - \chi \sigma)) + \mathcal{P} f(\varphi) g(\sigma) \quad \text{in } Q, \quad (1.1a) \\
\mu &= -\beta \varepsilon \Delta \varphi + \beta \varepsilon^{-1} \Psi'(\varphi) \quad \text{in } Q, \quad (1.1b) \\
\alpha \sigma_t &= \Delta \sigma - C h(\varphi) \sigma \quad \text{in } Q, \quad (1.1c) \\
0 &= \partial_n \varphi = \partial_n \mu = \partial_n \sigma \quad \text{on } \Sigma, \quad (1.1d) \\
\varphi(0) &= \varphi_0, \quad \sigma(0) = \sigma_0 \quad \text{in } \Omega, \quad (1.1e)
\end{align}
\]

with scalar functions \( \varphi, \mu, \sigma, m, f, g \) and \( h \), a double-well potential \( \Psi \), positive constants \( \varepsilon \) and \( \beta \), and non-negative constants \( \mathcal{P}, \chi, C \) and \( \alpha \). In biological applications, in particular tumour growth, the variable \( \sigma \) models the concentration of a nutrient present in the domain \( \Omega \), while \( \varphi \), commonly denoted as the order parameter in diffuse interface models, represents the difference in volume fractions between the tumour and healthy cells.

Equations (1.1a)-(1.1b) comprise of a Cahn–Hilliard system for the pair of variables \((\varphi, \mu)\). In (1.1b), \( \Psi' \) is the derivative of a double-well potential function \( \Psi : \mathbb{R} \to \mathbb{R}_+ \) that has equal minima at \( s = \pm 1 \), signifying the preference of \( \varphi \) attaining values close to 1 and −1. Intuitively, this implies that the domain \( \Omega \) is partitioned into large regions of tumour cells (given by the set \( \{ \varphi = 1 \} \)), and large regions of healthy cells (given by the set \( \{ \varphi = -1 \} \)), separated by thin transition layers \( \{ |\varphi| < 1 \} \) with a thickness proportional to the parameter \( \varepsilon > 0 \). The parameter \( \beta > 0 \) is associated with the surface tension of the interfacial layer, and the variable \( \mu \) is the associated chemical potential for \( \varphi \). The parameter \( \chi \) models the chemotaxis effect, that is, the tendency for tumour cells to move towards regions of high nutrient, see [Garcke et al., 2016] for more details. The term \( \mathcal{P} f(\varphi) g(\sigma) \) models the growth of the tumour cells. The non-negative function \( f \) is a smooth indicator function of the growing front, i.e., of the set \( \{ |\varphi| < 1 \} \), so that the tumour only grows at the interfacial layer between healthy cells and tumour cells, while \( g \) models how nutrient is used to promote growth. Hence \( \mathcal{P} \) can be treated as a tumour proliferation rate.

Equation (1.1c) is a reaction-diffusion equation for the nutrient. The function \( h \) is a non-negative function that smoothly interpolates between \( h(1) = 1 \) and \( h(-1) = 0 \). The term \( C h(\varphi) \sigma \) models the uptake of the nutrient only by the tumour cells. Thus, the constant parameter \( C \) can be interpreted as the nutrient consumption rate. The parameter \( \alpha \in \{0, 1\} \) determines how fast the nutrient evolves compared to the growth of the tumour cells. It is often the case that the nutrient diffusion timescale (\(~\) minutes) is much faster than the tumour doubling timescale (\(~\) days). Hence there are situations where it is appropriate to set \( \alpha = 0 \), leading to a quasi-static evolution. For the upcoming analysis and numerical simulations we consider only the case \( \alpha = 1 \) and \( \beta = 1 \). For further details regarding the model (1.1), we refer the reader to the earlier work [Kahle and Lam, 2018].

### 1.2 Parameter identification and observations

The scientific interest lies in the estimation of the parameter values, such as the proliferation rate \( \mathcal{P} \), the chemotaxis parameter \( \chi \), and the consumption rate \( C \), given observations on the evolution of the tumour. This is also known as model calibration. For meaningful applications of the model to medical research, it is important to be able to calibrate the model parameters in preparation for a comparison between simulations and experimental observations, a process known as model validation. Evaluation and refinements
to the mathematical model, such as including smaller-scale phenomena, elimination of slow processes and changes to boundary conditions, can then be made to improve our understanding of tumour growth.

Two frameworks are currently prevalent in the theory of parameter estimation. The classical framework uses Tikhonov regularization, and formulates the parameter estimation problem as an optimal control problem. Here we wish to obtain the optimal parameters such that the mismatch between the model output and the data is minimized, typically in some form of $L^2$-distance. For the model (1.1) with $\alpha = 1$ this is done in the recent work [Kahle and Lam, 2018]. However, the robustness of the optimal parameters with respect to uncertainty in the measurements is only investigated numerically by considering the optimal parameters for a range of given noise levels. The modern framework is statistical inversion using the Bayesian methodology, and incorporates the uncertainties associated with measurements and the relative probabilities of different optimal parameters given by the data.

In this work we employ the Bayesian approach. We consider observations obtained from (two-dimensional) snapshots of the tumour, either at one time instance or over several time instances. All observations are inherently polluted by some form of noise. More precisely, for fixed $J \in \mathbb{N}$, we model noisy observations as a set of linear functionals $\{l_j\}_{j=1}^J$ of the order parameter $\varphi$, so that $l_j : \varphi \mapsto l_j(\varphi) \in Y^{(j)}$. The data space $Y^{(j)}$ is a separable Banach space and we set $Y := \prod_{j=1}^J Y^{(j)}$. The observations, denoted by $\{y_j\}_{j=1}^J$, $y_j \in Y^{(j)}$, are expressed as

$$y_j = l_j(\varphi) + \eta_j, \quad j = 1, \ldots, J,$$

with the observational noise denoted by $\{\eta_j\}_{j=1}^J$, $\eta_j \in Y^{(j)}$. In this work we consider the following settings for $Y^{(j)}$, $J$ and $l_j$.

(a) Let $J = 1$ and $Y^{(1)} := L^2(\Omega)$ with $l_1(\varphi) := \varphi(\cdot, T)$;

(b) Let $J > 1$ and $Y^{(j)} = \mathbb{R}$ with $l_j(\varphi) = \frac{1}{2} \int_{\Omega} (\varphi(x, t_j) + 1)dx$, $j = 1, \ldots, J$, for an increasing sequence $\{t_j\}_{j=1}^J$ in the time interval $[0, T]$.

In (a) we observe the tumour everywhere in the spatial domain at a particular point in time. In (b) we observe the volume of the tumour at $J$ sequential points in time. This setting is an adaptation of the problem in [Collis et al., 2017]. The form of $l_j(\varphi)$ in (b) is motivated by the following observation. Recall that the tumour region is represented by the set $\{\varphi = 1\}$. Hence the function $\frac{1}{2}(\varphi + 1)$ takes the value +1 inside the tumour, and 0 outside, and is thus a smooth indicator function for the tumour.

1.3 Bayesian inversion

We concatenate the parameters $(P, \chi, C)$ to be estimated into the variable $u \in X$. Here $X$ typically denotes a subspace of a separable Banach space associated with the forward model (1.1). Furthermore, we concatenate the noise $(\eta_1, \ldots, \eta_J)$ and the observations $(y_1, \ldots, y_J)$ into the vectors $\eta \in Y$ and $y \in Y$, respectively. Then, equations (1.2) read as

$$y = G(u) + \eta,$$

where we introduced the forward response operator $G : X \to Y$ with

$$G(u) = (l_1(\varphi(u)), \ldots, l_J(\varphi(u))).$$
The operator $G$ is the composition of the forward solution operator $u = (P, \chi, C) \mapsto \varphi$ and the observation functionals $\varphi \mapsto l_j(\varphi)$. Whenever we discuss a particular setting, we denote the forward response operator in setting (a) by $G^a$ and in setting (b) by $G^b$. Throughout the paper, we use the notation $\{\eta^a, l^a, y^a\}$ to denote the noise, observation functional and observations for setting (a), and vice versa for setting (b), but we use the notation $G, \eta, l, y$ if we do not distinguish between setting (a) and (b).

We treat the parameter vector $u$ and the observational noise $\eta$ as stochastically independent random variables. Moreover, $u$ is distributed according to a prior (measure), and we assume that the noise is distributed as $\eta \sim N(0, \Gamma)$. In this framework, the observation of the data can be considered as an event $\{G(u) + \eta = y\}$. In this case, the parameter identification task consists in the computation of the posterior (measure), that is, the conditional measure of $u$, given that $\{G(u) + \eta = y\}$ occurs.

The Bayesian framework for parameter identification has been investigated for the Gompertzian tumour spheroid model in [Achilleos et al., 2014, Collis et al., 2017, Pack and Choi, 2014, Patmanidis et al., 2017], and for reaction-diffusion models in [Chang et al., 2018, Lê et al., 2016, Meghdadi et al., 2017, Menze et al., 2011]. Phase field type models such as (1.1) for tumour growth modelling and prediction have been suggested first in [Oden et al., 2010]. This work also outlines the beginning steps of the development of Bayesian methods for statistical calibration, model validation and uncertainty quantification. In [Hawkins-Daarud et al., 2013], which can be viewed as our closest counterpart in terms of model complexity, the authors consider three models of a similar type to (1.1), and calibrate the parameters $P$, $\chi$, and the diffusion coefficient of a quasi-static nutrient ($\alpha = 0$) using synthetic data. Subsequently, in [Lima et al., 2016, Lima et al., 2017, Oden et al., 2013] the identification of criteria for selecting the most plausible model among several classes of models for given data is discussed. However, to the best of our knowledge, the well-posedness of the Bayesian inverse problem (1.3) for parameter identification with a phase field tumour model remains unaddressed, since previous works [Hawkins-Daarud et al., 2013, Lima et al., 2016, Lima et al., 2017, Oden et al., 2010, Oden et al., 2013] focused on the implementation aspects of the framework. Therefore, the main contribution of our work is to establish the well-posedness of the Bayesian inverse problem. We apply [Stuart, 2010, Thm. 4.1 and 4.2] and [Kahle and Lam, 2018, Thm. 1 and 2], and show that under a uniform or a truncated Gaussian prior, the posterior measure is well-defined and is Lipschitz continuous with respect to the data $y$ in the Hellinger distance. This result is formulated in Theorem 3.1 below.

We perform numerical experiments based on synthetic data in setting (a), and based on real-world data in setting (b). Typically, Bayesian inverse problems are approached with Importance Sampling (see [Agapiou et al., 2017]) or Markov Chain Monte Carlo (see [Beskos et al., 2008, Cotter et al., 2013]). In preliminary tests not reported here we observed that both these methods require a prohibitively large number of evaluations of the expensive phase field model. For this reason we approximate the posterior measure by Sequential Monte Carlo with tempering (see e.g. [Beskos et al., 2015, Kantas et al., 2014] in the context of elliptic partial differential equation, and the Navier-Stokes equation, respectively). We select the tempering parameter adaptively to ensure a fixed effective sample size in every update step throughout the algorithm, see [Beskos et al., 2016] for a discussion and analysis of this approach.

1.4 Outline

The remainder of this article is organized as follows. In Sec. 2 we outline the likelihood for problem setting (a) and (b), and establish some properties of the forward response
operator $\mathcal{G}$. Sec. 3 is dedicated to the well-posedness of the posterior measure. In Sec. 4 we describe a fully discrete finite element approximation of the Cahn–Hilliard system (1.1), and in Sec. 5 we employ sequential Monte Carlo methods with tempering to solve the Bayesian inverse problem. Numerical examples for settings (a) and (b) are given in Sec. 6. Finally, we provide a discussion in Sec. 7.

2 Likelihood and properties of the forward model

2.1 Model for noise and data likelihood

The (data) likelihood $L(y|u)$ is a conditional probability density function (pdf) on the data space $Y$. The observational data is a realisation of the data-generating measure with the pdf $L(y|u)$ where $u$ is fixed. Given data and likelihood the task of a statistical analysis is the identification of $u$. We use the notation $N(\mu,Y)$ for a Gaussian measure with mean function $\mu$ and covariance operator $\Gamma$.

Recall from Sec. 1 that we consider two settings (a) and (b); we first focus on (a). Let $\Gamma_a := \sigma^2_a I_{Y_a}$ be a scaled identity operator on $Y_a$, where $\sigma^2_a \neq 0$. We assume that the observational noise $\eta^a$ is distributed according to a mean-zero Gaussian measure with covariance operator $\Gamma_a$, i.e., $\eta^a \sim N(0,\Gamma_a)$. It is important to note that $Y_a = L^2(\Omega)$ is an infinite dimensional space and that the noise $\eta^a$ is a Gaussian random field. This is in contrast to many settings in the literature where the data space is often finite dimensional.

By assumption $\Gamma_a \propto I_{Y_a}$, the random field $\eta^a$ is a so-called white noise, see [Stein, 1999, pages 40-41] for a definition and review of Gaussian white noise. The inverse of the covariance operator is called precision operator. The precision operator of $\eta^a$ is well-defined on $Y_a$ and is given by $\Gamma_a^{-1} := \sigma^{-2}_a I_{Y_a}$. Furthermore, we assume that $\eta^a$ is independent of the parameter $u$. In this case, the data-generating measure is $N(\mathcal{G}^a(u^1),\Gamma_a)$ and the conditional probability measure of $y$ given $u$ is $N(\mathcal{G}^a(u),\Gamma_a)$.

Let $B_{Y_a}$ denote the Borel-$\sigma$-algebra on $Y_a$. By definition, the likelihood is a pdf of $N(\mathcal{G}^a(u),\Gamma_a)$ with respect to some measure on $(Y_a,B_{Y_a})$ that does not depend on $u$. We can construct such a pdf by applying the Cameron–Martin Theorem. It states that $N(\mathcal{G}^a(u),\Gamma_a)$ and $N(0,\Gamma_a)$ are equivalent, if $\mathcal{G}^a(X)$ is a subset of the so-called Cameron–Martin space associated with $N(0,\Gamma_a)$. If $N(\mathcal{G}^a(u),\Gamma_a)$ and $N(0,\Gamma_a)$ are equivalent, then the likelihood can be defined by the Radon–Nikodym derivative of $N(\mathcal{G}^a(u),\Gamma_a)$ w.r.t. $N(0,\Gamma_a)$. Since $Y_a = L^2(\Omega)$ is a Hilbert space the Cameron–Martin space $CM(N(0,\Gamma_a))$ of $N(0,\Gamma_a)$ is the closure of $\text{Img}(\Gamma_a^{1/2};Y_a)$ with respect to the norm induced by the inner product

$$\langle f,g \rangle_{CM(N(0,\Gamma_a))} := \langle \Gamma_a^{-1/2} f, \Gamma_a^{-1/2} g \rangle_{Y_a},$$

see [Bogachev, 1998, page 44] and [Sullivan, 2015, Def. 2.50].

In our setting it holds that $CM(N(0,\Gamma_a)) = Y_a = L^2(\Omega)$. Moreover, it follows from Lemma 2.1 that $\mathcal{G}^a(u) = \varphi(\cdot,T) \in L^2(\Omega)$ for any $u \in X$. Hence $\mathcal{G}^a(X) \subseteq CM(N(0,\Gamma_a))$, and thus the measures $N(\mathcal{G}^a(u),\Gamma_a)$ and $N(0,\Gamma_a)$ are equivalent. We obtain the following Radon-Nikodym derivative that we will use as likelihood:

$$\frac{dN(\mathcal{G}^a(u),\Gamma_a)}{dN(0,\Gamma_a)}(y^a) = \exp \left( -\frac{\|\Gamma_a^{-1/2} \mathcal{G}^a(u)\|_{Y_a}^2}{2} + \langle \mathcal{G}^a(u),\Gamma_a^{-1} y^a \rangle_{Y_a} \right).$$
After a few simplifications we arrive at the following likelihood \( L_a \) and potential \( \Phi_a \):

\[
L_a(y^a|u) := \exp \left(-\Phi_a(u; y^a)\right),
\]

\[
\Phi_a(u; y^a) := \frac{\|G^a(u)\|_{Y_a}^2 - 2\langle G^a(u), y^a \rangle_{Y_a}}{2\sigma^2} := \frac{\|\varphi(\cdot, T)\|_{Y_a}^2 - 2\langle \varphi(\cdot, T), y^a \rangle_{Y_a}}{2\sigma^2}.
\]

Remark 2.1. It appears to be more intuitive to consider the potential

\[
\Phi'_a(u; y^a) = \frac{1}{2} \|\Gamma_a^{-1/2}(G^a(u) - y^a)\|_{Y_a}^2
\]

instead of \( \Phi_a \) in (2.1). This however would not induce a correct data-generating measure in cases where \( Y_a \) is infinite dimensional. In a finite dimensional setting it can be shown that \( \Phi'_a \) and \( \Phi_a \) lead to an equivalent Bayesian analysis. This is discussed in more detail in [Stuart, 2010, Remark 3.8].

Now we move on to setting (b). In this case the forward response operator \( G^b \) maps from \( X \) to \( Y_b := \mathbb{R}^J \). The conditional distribution of \( y^b \) given \( u \) is the Gaussian measure \( N(G^b(u), \Gamma_b) \). We define the noise covariance by \( \Gamma_b = \text{diag}(\sigma^2_j(t_j)) : j = 1, ..., J \). Here, \( \sigma^2_j(t_j) \) is the noise variance of the measurement at time \( t_j, j = 1, ..., J \). The noises at different points in time are stochastically independent. Since the data space \( Y_b \) is finite dimensional we can define the likelihood using the pdf of the multivariate Gaussian measure w.r.t. the Lebesgue measure. Moreover, [Collis et al., 2017, §3.2.2] suggest this as a likelihood for their particular estimation problem. However, for reasons of consistency and brevity in the following discussion, we again consider the pdf of \( y^b \) given \( u \) with respect to the probability measure of the noise. This pdf is well-defined since \( \Gamma_b \) is invertible and \( Y_b \) is finite dimensional. We arrive at

\[
L_b(y^b|u) := \exp \left(-\Phi_b(u; y^b)\right),
\]

\[
\Phi_b(u; y^b) := \frac{1}{2} \|\Gamma_b^{-1/2}G^b(u)\|_{Y_b}^2 - \langle G^b(u), \Gamma_b^{-1}y^b \rangle_{Y_b} := \sum_{j=1}^{J} \frac{\phi_j^b(\varphi)^2 - 2y^b_j\phi_j^b(\varphi)}{2\sigma^2_j(t_j)}
\]

\[
:= \sum_{j=1}^{J} \left(\int_{\Omega} \phi_j^b(x, t_j) + 1 dx\right)^2 - y^b_j \left(\int_{\Omega} \phi_j^b(x, t_j) + 1 dx\right).
\]

The likelihood in (2.2) and the likelihood in [Collis et al., 2017, §3.2.2] lead to identical estimation results (cf. Remark 2.1). Finally, we note that when we discuss either of the cases (a) or (b) we drop the subscripts and write

\[
L(y|u) := \exp \left(-\Phi(u; y)\right), \quad \Phi(u; y) := \frac{1}{2} \|\Gamma^{-1/2}G(u)\|_{Y}^2 - \langle G(u), \Gamma^{-1}y \rangle_{Y}.
\]

2.2 Properties of the forward model and potential

For fixed positive constants \( P_\infty, \chi_\infty \) and \( C_\infty \) we define the parameter space \( X \) as

\[
X := [0, P_\infty] \times [0, \chi_\infty] \times [0, C_\infty].
\]

Now we state some useful properties of the forward solution operator \( u \mapsto (\varphi, \mu, \sigma) \). We work under the following assumptions (see [Kahle and Lam, 2018, Assumption 1]):

(A1) \( \Omega \subset \mathbb{R}^2 \) is a bounded domain with \( C^4 \)-boundary \( \partial \Omega \). The initial conditions \( \varphi_0, \sigma_0 \) belong to the space \( H^3(\Omega) \cap H^2_\text{N}(\Omega) \), where \( H^2_\text{N}(\Omega) := \{ f \in H^2(\Omega) \mid \partial_n f = 0 \text{ on } \partial \Omega \} \).
The functions \( m, h, f, g \in C^2(\mathbb{R}) \) are bounded with bounded derivatives, and there exist positive constants \( n_0 \) and \( n_1 \) such that \( 0 \leq h(s) \) and \( n_0 \leq m(s) \leq n_1 \) for all \( s \in \mathbb{R} \).

The potential \( \Psi \in C^3(\mathbb{R}) \) is non-negative and there exist positive constants \( R_1, \ldots, R_5 \) such that for all \( s, t \in \mathbb{R} \), and \( k \in \{1, 2, 3\} \),
\[
\Psi(s) \geq R_1 |s|^2 - R_2, \quad |\Psi'(s)| \leq R_3 (1 + \Psi(s)), \quad |\Psi''(s)| \leq R_4 (1 + |s|^{q-1}),
\]
\[
|\Psi^{(k)}(s) - \Psi^{(k)}(t)| \leq R_5 (1 + |s|^{r-k+1} + |t|^{r-k+1}) |s-t|
\]
for some exponents \( q \in [1, \infty) \) and \( r \in [2, \infty) \).

**Lemma 2.1** ([Kahle and Lam, 2018, Thm. 1]). Under assumptions \([\text{A1}][\text{A4}]\) for any \( u = (\mathcal{P}, \chi, \mathcal{C}) \in X \) and fixed but arbitrary constant \( T > 0 \), there exists a unique triplet of functions \((\varphi, \mu, \sigma)\) that is a strong solution of \((1.1)\) with \( \alpha = 1 \) satisfying the regularities
\[
\varphi \in L^\infty(0, T; H^3(\Omega)) \cap L^2(0, T; H^4(\Omega)) \cap H^1(0, T; L^2(\Omega)) =: Z_1,
\]
\[
\mu \in L^\infty(0, T; H^1(\Omega)) \cap L^2(0, T; H^3(\Omega)) =: Z_2,
\]
\[
\sigma \in L^\infty(0, T; H^3(\Omega)) \cap H^1(0, T; H^2(\Omega)) =: Z_3
\]
as well as the initial and boundary conditions. Furthermore, there exists a positive constant \( C_* > 0 \) not depending on \((\varphi, \mu, \sigma)\) such that
\[
\|\varphi\|_{Z_1} + \|\mu\|_{Z_2} + \|\sigma\|_{Z_3} \leq C_* \tag{2.5}
\]

**Remark 2.2.** In the proof of [Kahle and Lam, 2018, Thm. 1], while it is difficult to keep track of the explicit dependence of the positive constant \( C_* \) on the parameters \((\mathcal{P}, \chi, \mathcal{C})\), we note that \( C_* \) does not depend on the reciprocals of \( \mathcal{P}, \chi \) and \( \mathcal{C} \). Since we consider non-negative parameters \( u \in X \), without loss of generality, we can express \((2.5)\) as
\[
\|\varphi\|_{Z_1} + \|\mu\|_{Z_2} + \|\sigma\|_{Z_3} \leq C_*(1 + |u|), \tag{2.6}
\]
which we will use for the remainder of this work.

We return to the observation settings (a) and (b) in Sec. 1. Using the (compact) embedding
\[
L^\infty(0, T; H^3(\Omega)) \cap H^1(0, T; L^2(\Omega)) \subset C^0([0, T]; C^{1, \delta}(\Omega)), \quad \delta \in (0, 1), \tag{2.7}
\]
we see that the order parameter \( \varphi \) (at least) belongs to \( C^0([0, T]; C^0(\Omega)) \). Hence, for our choice \( l_1(\varphi) = \varphi(\cdot, T) \), which is a function on \( \Omega \), or \( \left\{ l_j(\varphi) \right\}_{j=1}^J = \left\{ \frac{1}{T} \int_0^T \varphi(x, t_j) \, dx \right\}_{j=1}^J \), where \( \{t_j\}_{j=1}^J \) is a discrete set of points in \([0, T]\), we infer from \((2.6)\) that
\[
\|G_a(u)\|_{Y_a} \leq \|\varphi\|_{C^0([0, T]; C^0(\Omega))} \leq |\Omega|^{1/2} C_* (1 + |u|) = E_a(1 + |u|),
\]
\[
\|G_b(u)\|_{Y_b} \leq C J^{1/2} \|\varphi\|_{C^0([0, T]; C^0(\Omega))} + 1 \leq C (C_* (1 + |u|) + 1) \leq E_b(1 + |u|),
\]
for some constants \( E_a, E_b > 0 \). We define \( E := \max\{E_a, E_b\} \) and obtain
\[
\|G(u)\|_Y \leq E(1 + |u|). \tag{2.8}
\]

Now we study the stability of the forward solution operator associated with \((1.1)\) under perturbations in the parameters.
Lemma 2.2 ([Kahle and Lam, 2018 Thm. 2]). Let \( \{ (\varphi_i, \mu_i, \sigma_i) \}_{i=1,2} \) denote two strong solutions of (1.1) with \( \alpha = 1 \) corresponding to data \( \{ \varphi_{0,i}, \sigma_{0,i}, \mathcal{P}_i, \chi_i, C_i \}_{i=1,2} \). Then, there exists a positive constant \( C_{**} > 0 \) not depending on the differences \( \varphi_1 - \varphi_2, \mu_1 - \mu_2, \sigma_1 - \sigma_2, \mathcal{P}_1 - \mathcal{P}_2, \chi_1 - \chi_2, C_1 - C_2, \varphi_0, \varphi_0, \) and \( \sigma_0_1 - \sigma_0_2, \) such that

\[
\| \varphi_1 - \varphi_2 \|_{W_1} + \| \mu_1 - \mu_2 \|_{W_2} + \| \sigma_1 - \sigma_2 \|_{W_3} \leq C_{**} (|\mathcal{P}_1 - \mathcal{P}_2| + |\chi_1 - \chi_2| + |C_1 - C_2|) + C_{**} (\| \varphi_0 - \varphi_0 \|_{H^2(\Omega)} + \| \sigma_0 - \sigma_0 \|_{H^1(\Omega)}) ,
\]

where

\[
\begin{align*}
W_1 &= L^\infty(0, T; H^2(\Omega)) \cap L^2(0, T; H^4(\Omega)) \cap H^1(0, T; H^1(\Omega)'), \\
W_2 &= L^\infty(0, T; L^2(\Omega)) \cap L^2(0, T; H^2(\Omega)), \\
W_3 &= L^\infty(0, T; H^1(\Omega)) \cap L^2(0, T; H^2(\Omega)).
\end{align*}
\]

Remark 2.3. We mention that in the original formulation of [Kahle and Lam, 2018 Thm. 2], the Bochner space \( H^1(0, T; H^4(\Omega)' \) is not mentioned, but this is indeed a consequence, see [Kahle and Lam, 2018 (18)].

By the (compact) embedding

\[
L^\infty(0, T; H^2(\Omega)) \cap H^1(0, T; H^1(\Omega)) \subset \subset C^0([0, T]; C^0(\Omega)),
\]

we infer from (2.9) that for the same initial conditions \( \varphi_{0,1} = \varphi_{0,2} \) and \( \sigma_{0,1} = \sigma_{0,2} \) it holds

\[
\| \varphi_1 - \varphi_2 \|_{C^0([0,T]; C^0(\Omega))} \leq C_{**} (|\mathcal{P}_1 - \mathcal{P}_2| + |\chi_1 - \chi_2| + |C_1 - C_2|) .
\]

(2.10)

Analogously to the derivation of (2.8) we can find positive constants \( E' \) and \( F \) such that

\[
\| \mathcal{G}(u_1) - \mathcal{G}(u_2) \|_{Y} \leq E' (|\mathcal{P}_1 - \mathcal{P}_2| + |\chi_1 - \chi_2| + |C_1 - C_2|) \leq F |u_1 - u_2| .
\]

(2.11)

The following four lemmas show that the negative log-likelihood \( \Phi(u; y) \) defined in (2.1) satisfies the assumptions outlined in [Stuart, 2010 Assumption 2.6]. This is an important step in the proof of the well-posedness of the Bayesian inverse problem, and is required to prove the existence and uniqueness of the posterior measure.

Lemma 2.3. For every \( \varepsilon > 0 \) and \( r > 0 \), there exists a constant \( M = M(\varepsilon, r) \in \mathbb{R} \) such that for all \( u \in X \) and for all \( y \in Y \) with \( \| y \|_Y < r \) it holds

\[
\Phi(u; y) \geq M - \varepsilon | u |^2 .
\]

Proof. We observe that \( \Gamma^{-1} \) is a strictly positive definite operator with its spectrum bounded below by \( q := 1/ \max \{ \sigma_a^2, \sigma_b^2(t_1), \ldots, \sigma_b^2(t_J) \} > 0. \) Hence, for any \( \xi \in Y \), it holds that \( \langle \xi^*, \Gamma^{-1} \xi \rangle_Y \geq q \| \xi \|_Y^2 . \) By (2.8), it holds that for \( \| y \|_Y < r \) and arbitrary \( \varepsilon > 0, \)

\[
\Phi(u; y) = \frac{1}{2} \| \Gamma^{-1/2} \mathcal{G}(u) \|_Y^2 - \langle \mathcal{G}(u), \Gamma^{-1} y \rangle \geq - q \| \mathcal{G}(u) \|_Y \| y \|_Y \\
\geq - r q E(1 + | u |) \geq - r q E(1 + | u_\infty |) =: M \\
\geq M - \varepsilon | u |^2 ,
\]

where \( u_\infty = (\mathcal{P}_\infty, \chi_\infty, C_\infty) \).
Lemma 2.4. For every $r > 0$, there exists $K = K(r) > 0$ such that for all $u \in X$ and $y \in Y$ with $\max(|u|, |y|) < r$ it holds

$$\Phi(u; y) \leq K.$$  

Proof. Using that the spectrum of $\Gamma^{-1}$ is bounded above by $q' := 1/\min\{\sigma_a^2, \sigma_b^2(t_1), \ldots, \sigma_b^2(t_J)\} < \infty$, we use (2.8) to see that

$$\Phi(u; y) = \frac{1}{2}||\Gamma^{-1/2}G(u)||_Y^2 - \langle G(u), \Gamma^{-1}y \rangle \leq ||\Gamma^{-1/2}G(u)||_Y^2 + \frac{1}{2}||\Gamma^{-1/2}y||_Y^2$$

$$\leq q'||G(u)||_Y^2 + \frac{q'}{2}||y||_Y^2 \leq q'(E^2(1 + |u|)^2 + \frac{q'}{2}||y||_Y^2)$$

$$\leq q'(E^2(1 + r)^2 + \frac{r^2}{2}) =: K.$$  

Lemma 2.5. For every $r > 0$, there exists an $L = L(r) > 0$ such that for all $u_1, u_2 \in X$ and $y \in Y$ with $\max(|u_1|, |u_2|, |y|) < r$,

$$|\Phi(u_1; y) - \Phi(u_2; y)| \leq L |u_1 - u_2|.$$  

Proof. A short computation shows that

$$2(\Phi(u_1; y) - \Phi(u_2; y)) = \langle G(u_1) - G(u_2), \Gamma^{-1}(u_1 - u_2) \rangle_Y$$

$$= \langle \Gamma^{-1/2}G(u_1) - \Gamma^{-1/2}G(u_2), y \rangle + \langle \Gamma^{-1}(G(u_1) - G(u_2)), y \rangle$$

$$= \langle G(u_1) - G(u_2), \Gamma^{-1}(u_1 - u_2) - 2y \rangle_Y.$$  

$$\leq F \cdot q' \cdot (E(1 + r) + r) \cdot |u_1 - u_2| =: L(r)|u_1 - u_2|.$$  

Lemma 2.6. For every $\varepsilon > 0$ and $r > 0$, there exists $H = H(\varepsilon, r) \in \mathbb{R}$ such that for all $y_1, y_2 \in Y$ with $\max(||y_1||_Y, ||y_2||_Y) < r$ and for every $u \in X$,

$$|\Phi(u; y_1) - \Phi(u; y_2)| \leq \exp(\varepsilon |u|^2 + H)||y_1 - y_2||_Y.$$  

Proof. Similar to (2.12), we obtain after a short computation

$$2(\Phi(u; y_1) - \Phi(u; y_2)) = \langle y_2 - y_1, \Gamma^{-1}(2G(u) - y_1 - y_2) \rangle_Y.$$  

Then, due to $\max(||y_1||_Y, ||y_2||_Y) < r$ and (2.8), we see that

$$|\Phi(u; y_1) - \Phi(u; y_2)| \leq q'(E(1 + |u|) + r)||y_1 - y_2||_Y,$$

where we recall $q' := 1/\min\{\sigma_a^2, \sigma_b^2(t_1), \ldots, \sigma_b^2(t_J)\}$. Applying Young’s inequality, for any $\varepsilon > 0$, there exists a positive constant $C = C(\varepsilon, q', E, r)$ such that

$$q'(E(1 + |u|) + r)||y_1 - y_2||_Y \leq \varepsilon |u|^2 + C ||y_1 - y_2||_Y \leq \exp(\varepsilon |u|^2 + C)||y_1 - y_2||_Y.$$  

This yields the desired assertion.
3 Bayesian inversion

3.1 Well-posedness of the posterior measure

Recall that the parameter vector \( u = (P, \chi, C) \in X \subset \mathbb{R}^3 \) is contained in a finite dimensional space. Let \( \mu_0 \) denote a prior probability measure for \( u \) with corresponding probability density function \( \pi_0 \). Our interest is the posterior probability measure of \( u \) given \( y \) which we denote as \( \mu_y \) with probability density function \( \pi_y \). By Bayes’ rule we find the Radon–Nikodym relation

\[
\frac{d\mu_y}{d\mu_0}(u) = \frac{1}{Z(y)} \exp(-\Phi(u; y)),
\]

(3.1)

where the normalization constant \( Z(y) \) is defined as

\[
Z(y) = \int_X \exp(-\Phi(u; y)) \, d\mu_0(u).
\]

(3.2)

To be able to apply the Bayesian framework for inverse problems developed in [Cotter et al., 2009, Stuart, 2010] we require a prior measure \( \mu_0 \) that satisfies \( \mu_0(X) = 1 \). This means that the functions drawn from the prior measure \( \mu_0 \) belong to the space \( X \) almost surely.

For the benefit of the reader we recall that for two probability measures \( \mu_1 \) and \( \mu_2 \) on a measurable space \( (X, B_X) \), both absolutely continuous with respect to the same \( \sigma \)-finite reference measure \( \nu \), i.e., \( \mu_i \ll \nu \) for \( i = 1, 2 \), the Hellinger distance between \( \mu_1 \) and \( \mu_2 \) is

\[
d_{\text{Hell}}(\mu_1, \mu_2) = \left\{ \frac{1}{2} \int_X \left( \sqrt{\frac{d\mu_1}{d\nu}} - \sqrt{\frac{d\mu_2}{d\nu}} \right)^2 \, d\nu \right\}^{1/2},
\]

with Radon–Nikodym derivatives \( \frac{d\mu_1}{d\nu} \) and \( \frac{d\mu_2}{d\nu} \) of \( \mu_1 \) and \( \mu_2 \), respectively.

First we state the abstract result on the well-posedness of the posterior measure. Afterwards we investigate prior measures that are suitable for our inverse problem.

**Theorem 3.1.** Consider the inverse problem of finding parameters \( u = (P, \chi, C) \) from noisy observations of the form (1.2) subject to \( \varphi \) solving (1.1). Suppose \((A1)-(A4)\) hold and the observational noise \( \eta \) is distributed as mean-zero Gaussian with strictly positive definite covariance operator \( \Gamma \). Let \( \mu_0 \) be a prior measure satisfying

\[
\mu_0(X) = 1, \quad \mu_0(\{|u| < r\} \cap X) > 0 \quad \text{for all} \quad r > 0,
\]

where \( X \) is the space defined as in (2.4). Then the posterior measure \( \mu^y \) given by the relation (3.1) is a well-defined probability measure and is Lipschitz continuous in the Hellinger metric with respect to the data, i.e., for any \( r > 0 \) there exists a positive constant \( C = C(r) \) such that for all \( y_1, y_2 \in Y \) with \( \max(\|y_1\|_Y, \|y_2\|_Y) < r \), it holds

\[
d_{\text{Hell}}(\mu^{y_1}, \mu^{y_2}) \leq C\|y_1 - y_2\|_Y.
\]

**Proof.** By the estimate (2.11) we infer that the forward response operator \( G : X \to Y \) is Lipschitz continuous with respect to \( u \). Furthermore, by assumption \( \mu_0(X) = 1 \), it holds that \( G \) is \( \mu_0 \)-almost surely continuous, and thus \( G \) is also \( \mu_0 \)-measurable. By Lemma 2.5 the continuity of \( \Phi(\cdot; y) \) with respect to \( u \) then implies that \( \Phi(\cdot; y) \) is also \( \mu_0 \)-measurable. Thanks to Lemma 2.4 and the assumption \( \mu_0(\{|u| < r\} \cap X) > 0 \) for all \( r > 0 \), the
normalization constant $Z = Z(y)$ defined in (3.2) is bounded from below. Indeed, we see that
\[ Z(y) = \int_X \exp(-\Phi(u; y)) \, d\mu_0(u) \ge \exp(-K(r)) \mu_0(\{|u| < r\} \cap X) > 0. \] (3.3)

Next, choosing $r = \sqrt{3} \max(\mathcal{P}_\infty, \chi_\infty, C_\infty)$, for any $\varepsilon > 0$, we obtain from Lemma 2.3 that
\[ Z(y) \le \int_X \exp(\varepsilon |u|^2 - M(\varepsilon, r)) \, d\mu_0(u) \le \exp(\varepsilon r^2 - M(\varepsilon, r)) \mu_0(X) < \infty. \]

Therefore, the measure $\mu^y$ defined via the relation (3.1) is well-defined on $X$. To show the Lipschitz dependence of $\mu^y$ on the data $y$, let
\[ Z_1 := Z(y_1) = \int_X \exp(-\Phi(u; y_1)) \, d\mu_0(u), \quad Z_2 := Z(y_2) = \int_X \exp(-\Phi(u; y_2)) \, d\mu_0(u), \]
denote the two normalization constants for $\mu^{y_1}$ and $\mu^{y_2}$, respectively. For $\max(\|y_1\|_Y, \|y_2\|_Y) < r$, and any $\varepsilon > 0$, applying the results of Lemma 2.3 and Lemma 2.6 yields
\[ |Z_1 - Z_2| = \left| \int_X \int_0^1 e^{-z\Phi(u; y_1) - (1-z)\Phi(u; y_2)} \, dz (\Phi(u; y_1) - \Phi(u; y_2)) \, d\mu_0(u) \right| \le \left( \int_X e^{\varepsilon |u|^2 - M \varepsilon |u|^2 + H} \, d\mu_0(u) \right) \|y_1 - y_2\|_Y. \]

Let $p := \max(\mathcal{P}_\infty, \chi_\infty, C_\infty)$, then for any $u \in X$, it holds that $|u| \leq \sqrt{3}p$, and so
\[ |Z_1 - Z_2| \leq e^{6p - M(\varepsilon, r) + H(\varepsilon, r)} \|y_1 - y_2\|_Y =: C \|y_1 - y_2\|_Y. \] (3.4)

From the definition of the Hellinger distance we obtain
\[ 2d_{\text{Hell}}(\mu^{y_1}, \mu^{y_2})^2 = \int_X \left( Z_1^{-1/2} e^{-\frac{1}{2}\Phi(u; y_1)} - Z_2^{-1/2} e^{-\frac{1}{2}\Phi(u; y_2)} \right)^2 \, d\mu_0(u) \]
\[ = \int_X \left( (Z_1^{-1/2} - Z_2^{-1/2}) e^{-\frac{1}{2}\Phi(u; y_1)} + Z_2^{-1/2} (e^{-\frac{1}{2}\Phi(u; y_1)} - e^{-\frac{1}{2}\Phi(u; y_2)}) \right)^2 \, d\mu_0(u) \]
\[ \leq 2 \left| Z_1^{-1/2} - Z_2^{-1/2} \right|^2 \int_X e^{-\Phi(u; y_1)} \, d\mu_0(u) + \frac{2}{Z_2} \int_X \left( e^{-\frac{1}{2}\Phi(u; y_1)} - e^{-\frac{1}{2}\Phi(u; y_2)} \right)^2 \, d\mu_0(u) \]
\[ =: I_1 + I_2. \]

Using the lower bound (3.3) on the normalization constant, for fixed $r > 0$, there exists a positive constant $G$ such that $\max(Z_1^{-3}, Z_2^{-3}) \leq G$. Hence,
\[ \left| Z_1^{-1/2} - Z_2^{-1/2} \right|^2 = \left| \frac{Z_1}{Z_1^{1/2} - Z_2^{1/2}} - \frac{Z_2}{Z_2^{1/2}} \right|^2 \leq \max(Z_1^{-3}, Z_2^{-3}) \left| Z_1 - Z_2 \right|^2 \leq G \left| Z_1 - Z_2 \right|^2. \]

Together with Lemma 2.3 and (3.4) we find that
\[ I_1 \leq Ge^{3p - M(\varepsilon, r)} \|y_1 - y_2\|_Y^2 \leq C \|y_1 - y_2\|_Y^2 \]
for some positive constant $C$. Next, for $I_2$, with the help of Lemma 2.3 and Lemma 2.6 and the lower bound on the normalization constant $Z_2$, we see that
\[ I_2 = \frac{2}{Z_2} \int_X \left( \int_0^1 e^{-z\Phi(u; y_1) - (1-z)\Phi(u; y_2)} \, dz (\Phi(u; y_1) - \Phi(u; y_2)) \right)^2 \, d\mu_0(u) \]
\[ \leq \frac{2}{Z_2} \int_X |\Phi(u; y_1) - \Phi(u; y_2)|^2 e^{2\varepsilon |u|^2 - M(\varepsilon, r)} \, d\mu_0(u) \]
\[ \leq \frac{2}{Z_2} \int_X e^{2\varepsilon |u|^2 - M(\varepsilon, r)} e^{2\varepsilon |u|^2 + H(\varepsilon, r)} \, d\mu_0(u) \|y_1 - y_2\|_Y^2 \leq C \|y_1 - y_2\|_Y^2 \]
for some positive constant \( C \), as \(|u|^2 \leq 3\eta^2\). The previous two inequalities together imply
\[
d_{\text{Hell}}(\mu^{y_1}, \mu^{y_2}) \leq C\|y_1 - y_2\|_Y
\]
as required.

### 3.2 Uniform and truncated Gaussian priors

A natural candidate for the prior measure \( \mu_0 \) is the following uniform measure on the space \( X \):
\[
\mu_0^{\text{unif}} = \text{Unif}[0, \mathcal{P}_\infty] \otimes \text{Unif}[0, \chi_\infty] \otimes \text{Unif}[0, C_\infty].
\]

It is clear that \( \mu_0^{\text{unif}}(X) = 1 \), in the sense that draws from the uniform prior measure \( \mu_0^{\text{unif}} \) almost surely lie in the space \( X \). Furthermore, the positivity of the corresponding probability density function
\[
\pi_0(u_1, u_2, u_3) = \frac{1}{\mathcal{P}_\infty \chi_\infty C_\infty} \chi_{[0, \mathcal{P}_\infty]}(u_1) \chi_{[0, \chi_\infty]}(u_2) \chi_{[0, C_\infty]}(u_3),
\]
where \( \chi_A \) denotes the indicator function of the set \( A \), shows that for all \( r > 0 \),
\[
\mu_0^{\text{unif}}(\{|u| < r\} \cap X) > 0.
\]

Hence, by Theorem 3.1, the Bayesian inverse problem with the uniform prior \( \mu_0^{\text{unif}} \) is well-posed.

Another natural example for a prior measure \( \mu_0 \) is the truncated Gaussian measure. To construct it, we fix \((m_\mathcal{P}, m_\chi, m_C) \in X\) and select positive constants \((\sigma_\mathcal{P}, \sigma_\chi, \sigma_C)\). Then we define
\[
\phi(b, c; x) := \exp \left( -\frac{1}{2} \frac{(x - b)^2}{c^2} \right), \quad \psi(a, b, c; x) := \frac{\phi(b, c; x) \chi_{[0,a]}(x)}{\int_0^a \phi(b, c; x) \, dx}.
\]

It is easy to see that \( \psi \) is the density function of a truncated Gaussian on the interval \([0, a]\) where the untruncated Gaussian has mean \( b \) and variance \( c^2 \). Our truncated Gaussian prior takes the form
\[
\mu_0^{\text{TGaus}}(A) = \int_{X \cap A} \prod_{i \in \{\mathcal{P}, \chi, C\}} \psi(i_\infty, m_i, \sigma_i; x_i) \, dx_\mathcal{P} \, dx_\chi \, dx_C
\]
for any \( A \subset \mathbb{R}^3 \). Due to the definition of \( \psi \), it is clear that draws from \( \mu_0^{\text{TGaus}} \) lie in \( X \) almost surely. Moreover, the positivity of the density function \( \pi_0 \) guarantees that \( \mu_0^{\text{TGaus}}(\{|u| < r\} \cap X) > 0 \) for all \( r > 0 \). Hence, by Theorem 3.1, the Bayesian inverse problem with the truncated Gaussian prior \( \mu_0^{\text{TGaus}} \) is well-posed.

### 4 The fully discrete tumour model

For the approximation of the tumour model we use the method proposed in [Kahle and Lam, 2018, §6.1, 7.2]. We repeat it here for the convenience of the reader.

Let \( 0 = t_0 < t_1 < \ldots < t_k < \ldots < t_K = T \) denote a subdivision of \( I = [0, T] \). At time instance \( t_k \) we define a subdivision \( T_h^k = \{T^k_i\}_{i=1}^N \) of \( \Omega \) containing closed triangles \( T^k_i \) that exactly represent \( \Omega \), which in the following is assumed to be bounded with polygonal
boundary. While this choice of the domain does not fulfill the required boundary regularity Assumption [(A1)] we argue that as long as the tumour keeps away from \( \partial \Omega \), \( \varphi \) becomes constant close to \( \partial \Omega \). In this situation we can restrict ourselves to some smooth subdomain where Assumption [(A1)] is fulfilled. On \( \mathcal{T}_h^k \) we define the finite element function space \( V_h^k \) as

\[
V_h^k = \{ v \in C(\overline{\Omega}) \mid v|_{\mathcal{T}_h^k} \text{ is linear }, i = 1, \ldots, N^k \},
\]
i.e., the space of piecewise linear and globally continuous finite element functions. We point out that the order parameter \( \varphi \) is constant in large regions of \( \Omega \) and is rapidly changing near the growing front of the tumour. Thus adaptive meshing is necessary and therefore we use a different subdivision of the spatial domain at every time step.

At time instance \( k \) we construct finite element approximations \( \varphi_h^k, \mu_h^k, \sigma_h^k \in V_h^k \) of \( \varphi, \mu, \sigma \), respectively. To this end let \( \varphi^{k-1}, \sigma^{k-1} \in V_h^{k-1} \) be given, let \( I_h^k : C(\Omega) \rightarrow V_h^k \) denote the Lagrangian interpolation operator, and set \( \tau := t_k - t_{k-1} \). At time \( t_k \) we compute \( \varphi_h^k, \mu_h^k, \sigma_h^k \in V_h^k \) such that for all \( v \in V_h^k \) it holds

\[
\begin{align*}
(\varphi_h^k, v) + \tau (m(I_h^k \varphi^{k-1}) \nabla \mu_h^k, \nabla v) &= (I_h^k \varphi^{k-1}, v) + \tau \mathcal{P}(f(I_h^k \varphi^{k-1})) g(\sigma_h^k), v) \\
+ \tau \chi (m(I_h^k \varphi^{k-1}) \nabla \sigma_h^k, \nabla v),
\end{align*}
\]

(4.1a)

\[
\varepsilon (\nabla \varphi_h^k, \nabla v) + \frac{1}{\varepsilon} (\Psi(\varphi_h^k), v)^h = (\mu_h^k, v),
\]

(4.1b)

\[
(\sigma_h^k, v) + \tau (\nabla \sigma_h^k, \nabla v) = (I_h^k \sigma^{k-1}, v) - \tau \mathcal{C}(h(I_h^k \varphi^{k-1}) \sigma_h^k, v),
\]

(4.1c)

where \( (\phi, \psi) = \int_{\Omega} \phi \psi \, dx \) denotes the \( L^2(\Omega) \)-inner product. In (4.1b) we use the lumped integration \( (u, v)^h = \int_{\Omega} I_h^k(uv) \, dx \) for the integral involving \( \Psi(\varphi_h^k) \). For \( k = 1 \) we set \( \varphi^0 := \Pi_h \varphi_0 \) and \( \sigma^0 := \Pi_h \sigma_0 \), where \( \Pi_h \) denotes the \( L^2 \)-projection onto \( V_h^1 \).

For the potential \( \Psi \) we use a relaxed double-obstacle potential (see [Blowey and Elliott, 1991; Hintermüller et al., 2011]):

\[
\Psi(\varphi) = \frac{1}{2} (1 - \varphi^2) + s \Lambda_\rho(\varphi)
\]

for some constant \( s \gg 0 \), where \( \Lambda_\rho(\varphi) := \lambda_\rho(0, \varphi - 1) + \min_\rho(0, \varphi + 1) \). Note that \( \min_\rho \) and \( \max_\rho \) are regularizations of min and max according to [Hintermüller and Kopacka, 2011 (2.5)] such that \( \Psi \in C^{3,1}(\mathbb{R}) \). We fix \( s = 10^4 \) and \( \rho = 0.001 \).

The functions \( f, g, h, m \) are chosen as in [Kahle and Lam, 2018],

\[
\begin{align*}
f(x) &= \frac{1}{2} (\cos(\pi \min(1, \max(x, -1))) + 1), \\
h(x) &= \frac{1}{2} \left( \sin \left( \frac{\pi}{2} \min(1, \max(x, -1)) \right) + 1 \right), \\
m(x) &= (m_1 - m_0) f(x) + m_0, \\
g(x) &= \begin{cases} 
0 & \text{if } x \leq 0, \\
x^2 (-\theta^{-2} x + 2\theta^{-1}) & \text{if } 0 < x < \theta, \\
x & \text{if } \theta \leq x \leq M - \theta, \\
-\theta^{-2}(x - M)^3 - 2\theta^{-1}(x - M)^2 + M & \text{if } M - \theta < x < M, \\
M & \text{if } x \geq M.
\end{cases}
\end{align*}
\]

Here \( M = 10 \) is the maximum amount of nutrition that can be used for proliferation, \( \theta = 0.01, m_1 = 0.05, \) and \( m_0 = 5 \cdot 10^{-6} \). Note that the mobility \( m \) is chosen to be nearly degenerate in order to limit the growth of the tumour due to chemotaxis to the interface. We refer to [Garcke et al., 2016] where a degenerate mobility is used together with the double-obstacle potential.
5 Sequential Monte Carlo with tempering

To solve the Bayesian inverse problem we apply particle based methods. In particular, we approximate the posterior measure $\mu_y$ by a discrete measure of the form
\[ \hat{\mu}_y = \sum_{i=1}^{n} w^{(i)} \delta_{u^{(i)}}. \]

Here, $w^{(i)} > 0, i = 1, \ldots, n,$ are positive weights that sum up to one, and $\{u^{(i)}\}_{i=1}^{n} \in X^n$ is an ensemble of particles. In the following, we briefly review two methods that are popular in Bayesian statistics and Bayesian inversion, namely Importance Sampling and Sequential Monte Carlo (SMC). We refer to [Agapiou et al., 2017, Beskos et al., 2015, Del Moral et al., 2006, Kantas et al., 2014] for more details.

5.1 Importance Sampling

Let $Q: X \to \mathbb{R}$ be a function that is square-integrable with respect to the posterior measure. Importance sampling is based on the following identity
\[ \int Q d\mu_y = \int Q \frac{d\mu_y}{d\mu_0} d\mu_0 = \frac{1}{Z(y)} \int Q(u) \exp(-\Phi(u; y)) d\mu_0(u) \]
which is a consequence of Bayes’ formula (3.1). The above identity tells us that we can replace integrals given w.r.t. the posterior by integrals w.r.t. the prior. While we are typically not able to sample (independently) from the posterior measure, it is often possible to sample from the prior.

We apply standard Monte Carlo techniques to approximate the normalisation constant $Z(y)$ and the integral
\[ \int Q(u) \exp(-\Phi(u; y)) d\mu_0(u) \]
using $n$ samples of the prior $\mu_0$. This is equivalent to integrating $Q$ w.r.t. a specific discrete measure given by
\[ \hat{\mu}_y = \sum_{i=1}^{n} w^{(i)} \delta_{u^{(i)}}, \quad w^{(i)} = \frac{\exp(-\Phi(u^{(i)}; y))}{\sum_{j=1}^{n} \exp(-\Phi(u^{(j)}; y))}, i = 1, \ldots, n, \quad u^{(1)}, \ldots, u^{(n)} \sim \mu_0 \text{ i.i.d.} \]
The random variables $u^{(1)}, \ldots, u^{(n)}$ are measurable functions mapping from a probability space $(\Omega', \mathcal{F}', \mathbb{P})$ to $(X, BX)$. $BX$ denotes the Borel-$\sigma$-algebra of $X$. Hence, the measure $\hat{\mu}_y$ is actually a measure-valued random variable. It is possible to show that $\hat{\mu}_y$ converges weakly to the posterior measure $\mu_y$ as $n \to \infty$. In fact, [Agapiou et al., 2017, Thm. 2.1] shows that
\[ \sup_{\|Q\|_{\infty} \leq 1} \left( \int \left( \int Q d\mu_y - \int Q d\hat{\mu}_y \right)^2 d\mathbb{P} \right)^{1/2} \leq 2 \left( \frac{1 + \overline{cv}^2}{n} \right)^{1/2}, \quad (5.1) \]
where
\[ \overline{cv}^2 := \frac{\int \exp(-\Phi(\cdot; y))^2 d\mu_0}{\left( \int \exp(-\Phi(\cdot; y)) d\mu_0 \right)^2} - 1, \]
and $\overline{cv}$ denotes the coefficient of variation of the update density $\exp(-\Phi)$.
5.2 Sequential Monte Carlo

Importance sampling can be inefficient in Bayesian inversion, in particular, when the parameter space is high-dimensional or the data (resp. the likelihood) is highly informative. In either of these cases the posterior can be concentrated in a small area of the parameter space. In contrast, the prior is typically not concentrated. Hence, a large number of prior samples is required to obtain a reasonable approximation of the posterior measure. For every prior sample we need one evaluation of the potential $\Phi$. In practice, this can lead to a massive number of (expensive) model evaluations.

Sequential Monte Carlo overcomes this issue by constructing a sequence of measures $\{\mu_k\}_{k=0}^K$ starting with the prior, and slowly approaching the posterior $\mu_K \approx \hat{\mu}_y$. The sequence is constructed such that each of the measures $\mu_k$ allows an efficient importance sampling approximation of the measure $\mu_{k+1}$, $k = 0, \ldots, K - 1$.

One can see easily that the noise covariance has an impact on the concentration of the posterior. Hence, we construct our sequence by starting with a highly up-scaled noise covariance and proceed by scaling the noise-level to the actual level, i.e.

$$\frac{d\mu_k}{d\mu_0} \propto \exp\left( -\frac{1}{2} \left\| (\beta_k^{-1} \Gamma)^{-1/2} \mathcal{G}(u) \right\|^2_Y - \langle \mathcal{G}(u), (\beta_k^{-1} \Gamma)^{-1} y \rangle_Y \right),$$

where $\{\beta_k\}_{k=0}^K$ is an increasing sequence starting at $\beta_0 = 0$ and finishing at $\beta_K = 1$. Originating in statistical thermodynamics, this procedure is often referred to as tempering.

Sequential Monte Carlo proceeds in the following way. First, $n$ samples $u^{(1)}, \ldots, u^{(n)}$ are drawn independently from the prior $\mu_0$. Then, the samples are weighted with $\gamma_1$ to approximate $\mu_1$ using the importance sampling idea. This gives the discrete measure $\hat{\mu}_1$. Particles with small weights shall be eliminated from the ensemble. To this end we resample the ensemble, i.e., we draw new, equally weighted samples $u^{(1)}, \ldots, u^{(n)}$ from $\hat{\mu}_1$. To distribute the particles more evenly in the parameter space, we pass the samples through a Markov kernel that is stationary w.r.t. $\mu_1$. This gives a new set of samples $u^{(1)}, \ldots, u^{(n)}$ that is approximately $\mu_1$-distributed. We then proceed iteratively for $k = 2, \ldots, K$. Each sample $u^{(1)}, \ldots, u^{(n)}$ is weighted according to the update density $\gamma_k/\gamma_{k-1}$. Then, we resample the particles and apply a Markov kernel that is stationary w.r.t. to $\mu_k$. We stop, if $k = K$ and obtain $\hat{\mu}_K = \hat{\mu}_y$.

Remark 5.1. The Markov kernel is typically given by a Markov Chain Monte Carlo (MCMC) sampler. We refer to [Robert and Casella, 2004, §§7-10] for a thorough introduction to MCMC and to [Beskos et al., 2008, Cotter et al., 2013] for a discussion of MCMC methods for Bayesian Inversion.

5.3 Adaptivity for the tempering

It is not intuitively clear how to choose the tempering sequence $\{\beta_k\}_{k=0}^K$. A typical approach is induced by the importance sampling error and its connection to the coefficient of variation of the update density in (5.1). If $\text{cv}$ is small, then the accuracy of the importance sampling approximation is high, even if the number of samples is quite small.

Using a simple parameter fitting approach we choose $\{\beta_k\}_{k=0}^K$ adaptively such that the coefficient of variation in each update step equals some target $\text{cv} > 0$ that has been defined a priori. This is in fact a root finding problem in one spatial dimension. We mention that the adaptive algorithm introduces a bias into the estimation of the model evidence. In [Beskos et al., 2016] it is shown that the adaptive SMC method is convergent.
6 Numerical examples

In this section we apply the SMC approach as described in Section 5 for the identification of the parameters in the numerical approximation (4.1) of the tumour model (1.1). We consider two examples. In Section 6.1 we consider the example from [Kahle and Lam, 2018] with synthetic tumour data and the observation setting (a) (see Sec. 1). In Section 6.2 we consider the real world measured data set from [Collis et al., 2017]; this corresponds to the observation setting (b) in Sec. 1.

The implementation is done in C++. We use the finite element library FEniCS [Logg et al., 2012] together with the PETSc linear algebra backend [Balay et al., 2014] and the direct solver MUMPS [Amestoy et al., 2001]. The meshes are generated and adapted using ALBERTA [Schmidt and Siebert, 2005].

In the remainder of this section we always work with fully discrete functions, and neglect the subscript $h$ for the finite element approximations introduced in Section 4.

6.1 The full tumour model compared with synthetic data

As first example we consider the identification of known parameters in artificial data. We use the set-up from [Kahle and Lam, 2018]. The initial conditions are $\sigma_0 = 1$ uniformly in the domain $\Omega = (-5, 5) \times (-5, 5)$, and

\[ z_0 := \arctan(\sqrt{s - 1}), \]

\[ \Phi_0(z) := \begin{cases} -\Phi_0(-z) & \text{if } z < 0, \\ \sqrt{\frac{s}{\pi - 1}} \sin(z) & \text{if } 0 \leq z \leq z_0, \\ \frac{1}{z_0} (s - \exp(\sqrt{s - 1}(z_0 - z))) & \text{else}, \end{cases} \]

\[ \varphi_0(x) = \Phi_0(\varepsilon^{-1}(1 - \|x\|_{l^8})). \]

Here $\Phi_0$ is the first order approximation of $\varphi$ for $\Psi$ (with $\rho = 0$), and $\varphi_0$ describes a rounded square centered at the origin which is realized by the unit circle in the $l^8$-norm, where we set $\varepsilon = 0.05$.

We simulate system (4.1) with $P = 7$, $\chi = 120$, $C = 2$ until $T = 4$ with time steps of size $\tau = 0.05$ to obtain $\varphi_d = \varphi(T)$ and add normally distributed noise with mean 0 and standard deviation 0.1 to every degree of freedom of $\varphi_d$. The resulting function is considered as the data $y$. In Figure 1 we show $\varphi_0$ (left), $\varphi_d$ (middle) and $y$ (right).
Figure 2: The posterior marginal distributions for $\mathcal{P}$, $\chi$, and $C$. The synthetic observation is generated with $\mathcal{P} = 7$, $\chi = 120$, and $C = 2$, and by adding normally distributed noise with standard deviation 0.1 to every node. The posterior mean is $m_\mathcal{P} = 6.84$, $m_\chi = 119.85$, and $m_C = 1.95$. $\chi_\infty = 200$, and $C_\infty = 10$. We use $n = 100$ particles to approximate the posterior measure and $\overline{\sigma} = 0.25$ for the tempering steps. In Figure 2 we depict the marginal posterior distribution of $\mathcal{P}$, $\chi$ and $C$. The corresponding posterior sample means are $m_\mathcal{P} = 6.84$, $m_\chi = 119.85$, and $m_C = 1.95$. We see that these agree quite well with the parameter values that have been used to generate the synthetic observation.

In addition we compute the Maximum-A-Posteriori (MAP), that is, the global maximum of the density of the posterior measure. It is well known that for Gaussian priors the MAP is the solution of a Tikhonov regularised least-squares inverse problem, see [Stuart, 2010] §2.2. Hence the MAP can be computed by the method proposed in [Kahle and Lam, 2018] by choosing suitable values for the individual Tikhonov weights associated with the parameters. For our example we obtain the MAP estimate $\text{MAP}_\mathcal{P} = 7.0188$, $\text{MAP}_\chi = 106.9212$, and $\text{MAP}_C = 2.4836$ using the code from [Kahle and Lam, 2018] with the prior mean as initial value in the optimization. We observe that the MAP estimate for both the chemotaxis parameter $\chi$ and the consumption rate $C$ does not agree very well with the underlying true parameter value. This is in contrast to the Bayesian posterior mean estimate, and we believe that this might be due to a local maximum. We obtain the same MAP estimate when we start the code with the underlying true parameter values.

In Figure 3 we show the mean order parameter $\varphi^K_m$ at final time $T$ (left) together with the zero level lines of $\varphi^K_m$ (black) and $\varphi^K_d$ (white). On the right we show the variance of $\varphi^K$ together with the $\varphi^K_m = \pm 1$ isolines. The magnitude of $\varphi^K$ is of order 0.4 in the black regions and $10^{-11}$ in the white regions.

Finally, in Table 1 we list the covariances of the parameters. We observe that $\mathcal{P}$ and $\chi$ are negatively correlated, as are $\chi$ and $C$, while $\mathcal{P}$ and $C$ have a positive correlation. We attribute the negative correlation between $\mathcal{P}$ and $\chi$ to the fact that both parameters control the growth of the tumour, although the growth mechanisms differ somewhat. Namely, $\mathcal{P}$ leads to undirected growth, while $\chi$ gives directional growth depending on nutrient concentration. It is likely that neither parameters can be large at the same time in order to obtain a tumour of a comparable size to the data. The positive correlation between $\mathcal{P}$ and $C$ can be explained as follows. Larger values of $C$ imply that nutrients are consumed at a faster rate. Thus on the growing front the level of nutrients is lower compared to regions away from the growing front, and so, in order to maintain growth in regions of lower nutrient concentration a larger value of $\mathcal{P}$ is desirable. Similarly, larger values of $C$ leads to a larger nutrient gradient across the interface between the tumour and its surroundings. Hence, large values of $\chi$ would likely amplify the directed growth of the tumour, and the negative correlation between $C$ and $\chi$ is a means to match the simulations more closely with the data.
Table 1: The covariances of the parameters $\mathcal{P}$, $\chi$, $\mathcal{C}$ for the first numerical example.

|   | $\mathcal{P}$ | $\chi$ | $\mathcal{C}$ |
|---|---|---|---|
| $\mathcal{P}$ | 0.10475 | -1.1818 | 0.039662 |
| $\chi$ | -1.1818 | 55.975 | -0.9936 |
| $\mathcal{C}$ | 0.039662 | -0.9936 | 0.045525 |

Figure 3: The pointwise mean $\varphi^K_m$ of the order parameter at final time $T$ together with the zero level lines for $\varphi^K_m$ in black and for $\varphi_d$ in white (nearly identical) on the left for the first numerical example. On the right we show the pointwise variance $\varphi^K_d$ of $\varphi^K$ together with the $\pm 1$ isolines of $\varphi^K_m$. The magnitude of $\varphi^K_d$ is of order 0.4 in black regions. Due to symmetry, we only show one quarter of the computational domain.
Table 2: The covariance of the parameters $P$ and $y^b_0$ for the 4th tumour data set.

|     | $P$         | $y^b_0$     |
|-----|-------------|-------------|
| $P$ | 1.406e-06   | -4.1007e-06 |
| $y^b_0$ | -4.1007e-06 | 1.7446e-05  |

6.2 The reduced tumour model compared with measured data

In [Collis et al., 2017, Connor, 2016] measurements for the volume of 10 individual spherical growing tumours are provided. In this section we aim to identify the parameters associated with the tumour evolution. This requires small changes in our set-up as we explain next.

First of all, in the model (1.1) we set $\chi \equiv 0$ and $C \equiv 0$. The latter is due to the fact that no information on nutrient concentration is provided in the data in [Connor, 2016]. For the former we remark that in the absence of chemotaxis the growth of the tumours remains radial provided the initial tumour is radial. This is supported by a linear stability analysis and numerical simulations showing that small perturbations around the radial interface decay when $\chi$ is zero or small, see for example Cristini et al., 2009 Garcke et al., 2016 Garcke et al., 2018. Therefore, it remains to identify the proliferation rate $P$.

Further in [Collis et al., 2017, §3.2.2] a specific likelihood function is defined, that we replace with the equivalent likelihood in (2.2). See §2.1 for a discussion of this function. Here $J$ is the number of volume measurements $y^b_j$ taken from the growing tumour at times $t_j$. The experimental noise is assumed normally distributed about 0 with variance $\sigma^2(t_j)$. These data are provided in [Connor, 2016] and contains triples $(t_j, y^b_j, \sigma^2(t_j))$ for every individual tumour. However, the data in [Connor, 2016] does not contain the initial volume of the tumour. For this reason we consider the initial volume $y^b_0$ as additional parameter to be identified. Since the tumours are all spherical, the initial tumour $\varphi_0$ can be completely described by the radius of the tumour which makes this approach feasible. Note that under this assumption, in Lemma 2.2 we can estimate $\|\varphi_0 - \varphi_0\|_{H^2(\Omega)} \leq C|r_1 - r_2|$, where $r_i$ is the radius of the initial tumour represented by $\varphi_{0,i}$, $i = 1, 2$. Here we define $\varphi_0$ as in (6.1) but with the $l^2$-norm instead of the $l^8$-norm, and with corresponding radius $r_i$. We set $\sigma_0 = 1.0$, $\tau = 10$, $T = 700$, $\Omega = (-1.5, 1.5) \times (-1.5, 1.5)$, and $\varepsilon = 0.05$. The time unit is one hour, and the spatial unit is one millimeter.

We use $n = 200$ particles and $cv = 0.25$ for the tempering in SMC. The prior is the product of truncated Gaussian distributions with $m_P = 0.025$, $\sigma_P = 0.015$ and $m_{y^b_0} = 0.04$, $\sigma_{y^b_0} = 0.01$, see Sec. 3.2. The upper bounds are $P_\infty = 0.1$ and $(y^b_0)_\infty = 0.06$.

In Figure 4 we depict the marginals of the posterior distributions for the 4th data set from [Connor, 2016] in the top row. In the bottom row we depict the same quantities for the 5th data set. In Tables 2 and 3 the corresponding covariances of the parameters are shown. Both indicate that $P$ and $y^b_0$ are negatively correlated. This is expected as a large initial tumour volume coupled with a large proliferation rate would lead to larger tumour evolutions. In [Collis et al., 2017] a Gompertzian model for tumour growth is used, which is different to model (1.1). Hence we can not perform quantitative comparisons between our estimation result with the results from [Collis et al., 2017].
7 Discussion

In this paper we formulated a Bayesian inverse problem which is suitable for the identification of internal parameters in a diffuse interface model for tumour growth. We proved the well-posedness of the posterior measure for finite and infinite dimensional data spaces associated with our observation model. We implemented Sequential Monte Carlo with tempering in combination with a finite element discretization to approximate the posterior measure of the unknown parameters. We tested our implementation with a synthetic data set observing the full phase field, and a real-world data set observing the volume of spherical tumours. To finish we discuss directions of further work.

7.1 Quasi-static model

The analysis in this work is performed for the case where the diffusion of the nutrient is comparable to the growth of the tumour, i.e., the case $\alpha = 1$ in (1.1), as we mentioned in the introduction. However, it might be more realistic to consider the quasi-static model variant $\alpha = 0$ since typical nutrient diffusion timescales are much shorter than tumour growth timescales. In the proof of Theorem 3.1 the crucial component is the continuous dependence of $\varphi$ in the $L^\infty(0, T; H^2(\Omega)) \cap H^1(0, T; H^1(\Omega)')$ norm on the parameters $(\mathcal{P}, \chi, \mathcal{C})$, which then leads to (2.10) and (2.11). Therefore, an analogous well-posedness result for the Bayesian parameter identification problem with the quasi-static model can be obtained if one is able to show a similar continuous dependence result. We leave this verification for future research and remark that ideas in [Garcke and Lam, 2016, Kahle and Lam, 2018] may be helpful in establishing the required continuous dependence result.

Table 3: The covariance of the parameters $\mathcal{P}$ and $y_0^b$ for the 5th tumour data set.
7.2 Cahn–Hilliard models with velocity

We point out that many of the earlier diffuse interface tumour models include a notion of cellular velocity, which affixes the system (1.1) with a Darcy system or a Navier–Stokes system, and also appends the Cahn–Hilliard equation and the nutrient equation with convection terms. We point out that the divergence of the velocity field need not be equal to zero in the presence of mass transfer, as the gain or loss in volume resulting from mass transition leads to sources or sinks in the mass balance.

It is reported in [Garcke et al., 2018] that models with a fluid velocity produce biologically more realistic results compared to models without fluid velocity for the situation involving multiple species of cells. We do not consider the fluid velocity in our presenting setting with two components (tumour and host cells), since the differences in biological effects are less significant compared to the multispecies case. In addition, analytical results for models with velocity are not as complete compared to simpler ones such as (1.1). However, we envision that a similar analysis for the Bayesian inverse problem can be done if analogous results to (2.9) are available for these types of models. This would also enable a study of model selection as outlined in [Lima et al., 2016, Lima et al., 2017, Oden et al., 2013].

7.3 Surrogates for the order parameter

Bayesian inversion for the Cahn–Hilliard model (1.1) is very expensive since the repeated evaluation of the likelihood requires forward solves of (1.1) for many different parameter configurations and initial states. The computational burden can be reduced by constructing surrogates for $\varphi$ which can be evaluated cheaply without the need to run an expensive forward solve. For Bayesian inversion a number of surrogates have been studied, e.g. Gaussian process models [Kennedy and O’Hagan, 2001], or generalized polynomial chaos surrogates [Marzouk et al., 2007, Marzouk and Najm, 2009]. We point out that the solution of the Cahn–Hilliard model (1.1) depends continuously on the parameters $(P, \chi, C)$ and initial conditions, see [Kahle and Lam, 2018, Theorem 2]. Thus it is feasible to construct smooth, polynomial based surrogates, or sparse grid surrogates. While this has been investigated in depth for the classical elliptic Bayesian inverse problem, see e.g. [Schwab and Stuart, 2012, Schillings and Schwab, 2013], this is not the case for Cahn–Hilliard models such as (1.1). In addition, the error and convergence analysis for surrogates in Bayesian inversion is far from complete (see e.g. [Yan and Zhang, 2017, Stuart and Teckentrup, 2018] for recent studies), and requires further work.

References

[Achilleos et al., 2014] Achilleos, A., Loizides, C., Hadjiantereou, M., Stylianopoulos, T., and Mitsis, G. (2014). Multiprocess Dynamic Modeling of Tumor Evolution with Bayesian Tumor-Specific predictions. *Ann. Biomed. Eng.*, 42(5):1095–1111.

[Agapiou et al., 2017] Agapiou, S., Papaspiliopoulos, O., Sanz-Alonso, D., and Stuart, A. (2017). Importance sampling: intrinsic dimension and computational cost. *Statist. Sci.*, 32(3):405–431.

[Amestoy et al., 2001] Amestoy, P., Duff, I., L’Excellent, J.-Y., and Koster, J. (2001). A fully asynchronous multifrontal solver using distributed dynamic scheduling. *SIAM J. Matrix Anal. & Appl*, 23(1):15–41.
[Balay et al., 2014] Balay, S., Abhyankar, S., Adams, M., Brown, J., Brune, P., Buschelman, K., Dalcin, L., Eijkhout, V., Gropp, W., Kaushik, D., Knepley, M., McInnes, L., Rupp, K., Smith, B., Zampini, S., and Zhang, H. (2014). PETSc Web page. http://www.mcs.anl.gov/petsc

[Beskos et al., 2016] Beskos, A., Jasra, A., Kantas, N., and Thiery, A. (2016). On the convergence of adaptive sequential Monte Carlo methods. Ann. Appl. Probab., 26(2):1111–1146.

[Beskos et al., 2015] Beskos, A., Jasra, A., Muzaffer, E., and Stuart, A. (2015). Sequential Monte Carlo methods for Bayesian elliptic inverse problems. Stat. Comput., 25(4):727–737.

[Beskos et al., 2008] Beskos, A., Roberts, G., Stuart, A., and Voss, J. (2008). An MCMC method for diffusion bridges. Stoch. Dyn., 8:319–350.

[Blowey and Elliott, 1991] Blowey, J. and Elliott, C. (1991). The Cahn–Hilliard gradient theory for phase separation with non-smooth free energy. Part I: Mathematical analysis. European J. Appl. Math., 2(3):233–280.

[Bogachev, 1998] Bogachev, V. (1998). Gaussian measures, volume 62 of Mathematical Surveys and Monographs. AMS, Providence, RI.

[Burton, 1966] Burton, A. (1966). Rate of growth of solid tumors as a problem of diffusion. Growth, 30:157–176.

[Chang et al., 2018] Chang, Y., Sharp, G., Li, Q., Shih, H., El Fakhri, G., Ra, J., and Woo, J. (2018). Subject-specific brain tumor growth modelling via an efficient Bayesian inference framework. In Angelini, E. and Landman, B., editors, Medical Imaging 2018: Image Processing, Houston, Texas, United States, 10-15 February 2018, volume 10574 of SPIE Proceedings, page 105742I.

[Collis et al., 2017] Collis, J., Connor, A., Paczkowski, M., Kannan, P., Pitt-Francis, J., Byrne, H., and Hubbard, M. (2017). Bayesian Calibration, Validation and Uncertainty Quantification for Predictive Modelling of Tumour Growth: A Tutorial. Bull. Math. Biol., 79(4):939–974.

[Connor, 2016] Connor, A. (2016). Calibration, validation and uncertainty quantification. https://figshare.com/articles/Calibration_validation_and_uncertainty_quantification/3406876

[Cotter et al., 2009] Cotter, S., Dashti, M., Robinson, J., and Stuart, A. (2009). Bayesian inverse problems for functions and applications to fluid mechanics. Inverse Problems, 25:115008.

[Cotter et al., 2013] Cotter, S., Roberts, G., Stuart, A., and White, D. (2013). MCMC Methods for Functions: Modifying Old Algorithms to Make Them Faster. Statist. Sci., 28(3):424–446.

[Cristini et al., 2009] Cristini, V., Li, X., Lowengrub, J., and Wise, S. (2009). Nonlinear simulations of solid tumor growth using a mixture model: invasion and branching. J. Math. Biol., 58:723–763.
[Cristini and Lowengrub, 2010] Cristini, V. and Lowengrub, J. (2010). *Multiscale Modeling of Cancer. An Integrated Experimental and Mathematical Modeling Approach*. Cambridge University Press.

[Deisboeck and Stamatakos, 2010] Deisboeck, T. and Stamatakos, G. (2010). *Multiscale cancer modeling*. Chapman & Hall/CRC, Taylor & Francis Group, London.

[Del Moral et al., 2006] Del Moral, P., Doucet, A., and Jasra, A. (2006). Sequential Monte Carlo samplers. *J. R. Statist. Soc. B*, 68(3):411–436.

[Frieboes et al., 2010] Frieboes, H., Jin, F., Chuang, Y.-L., Wise, S., Lowengrub, J., and Cristini, V. (2010). Three-dimensional multispecies nonlinear tumor growth - II: Tumor invasion and angiogenesis. *J. Theor. Biol.*, 264:1254–1278.

[Garcke and Lam, 2016] Garcke, H. and Lam, K. (2016). Well-posedness of a Cahn–Hilliard system modelling tumour growth with chemotaxis and active transport. *European J. Appl. Math.*, 28(2):284–316.

[Garcke et al., 2018] Garcke, H., Lam, K., Nürnberg, R., and Sitka, E. (2018). A multiphase Cahn–Hilliard–Darcy model for tumour growth with necrosis. *Math. Models Methods Appl. Sci.*, 28:525–578.

[Garcke et al., 2016] Garcke, H., Lam, K., Sitka, E., and Styles, V. (2016). A Cahn–Hilliard–Darcy model for tumour growth with chemotaxis and active transport. *Math. Models. Methods. Appl. Sci.*, 26:1095–1148.

[Greenspan, 1972] Greenspan, H. (1972). Models for the growth of a solid tumor by diffusion. *Studies in Applied Mathematics*, 51:317–340.

[Hawkins-Daarud et al., 2013] Hawkins-Daarud, A., Prudhomme, S., van der Zee, K., and Oden, J. (2013). Bayesian calibration, validation, and uncertainty quantification of diffuse interface models of tumor growth. *J. Math. Biol.*, 67:1457–1485.

[Hawkins-Daarud et al., 2012] Hawkins-Daarud, A., van der Zee, K., and Oden, J. (2012). Numerical simulation of a thermodynamically consistent fourspecies tumor growth model. *Int. J. Numer. Methods Biomed. Eng.*, 28:3–24.

[Hintermüller et al., 2011] Hintermüller, M., Hinze, M., and Tber, M. (2011). An adaptive finite element Moreau–Yosida-based solver for a non-smooth Cahn–Hilliard problem. *Optim. Methods Softw.*, 25(4-5):777–811.

[Hintermüller and Kopacka, 2011] Hintermüller, M. and Kopacka, I. (2011). A smooth penalty approach and a nonlinear multigrid algorithm for elliptic MPECs. *Comput. Optim. Appl.*, 50(1):111–145.

[Kahle and Lam, 2018] Kahle, C. and Lam, K. (2018). Parameter identification via optimal control for a Cahn–Hilliard-chemotaxis system with a variable mobility. *Appl. Math. Optim.*, doi:10.1007/s00245-018-9491-z. Published electronically 22 March 2018.

[Kantas et al., 2014] Kantas, N., Beskos, A., and Jasra, A. (2014). Sequential Monte Carlo Methods for High-Dimensional Inverse Problems: A case study for the Navier-Stokes equations. *SIAM/ASA J. Uncertain. Quantif.*, 2(1):464–489.

[Kennedy and O’Hagan, 2001] Kennedy, M. C. and O’Hagan, A. (2001). Bayesian calibration of computer models. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 63(3):425–464.
[Lê et al., 2016] Lê, M., Delingetter, H., Kalpathy-Cramer, J., Gerstner, E., Batchelor, T., Unkelbach, J., and Ayache, N. (2016). MRI Based Bayesian Personalization of a Tumor Growth Model. *IEEE T. Med. Imaging*, 35(10):2329–2339.

[Lima et al., 2016] Lima, E., Oden, J., Hornuth II, D., Yankeelov, T., and Almeida, R. (2016). Selection, calibration, and validation of models of tumor growth. *Math. Models Methods Appl. Sci.*, 26(12):2341–2368.

[Lima et al., 2017] Lima, E., Oden, J., Wohlmuth, B., Shahmoradi, A., Hornuth II, D., Yankeelov, T., Scarabosio, L., and Horger, T. (2017). Selection and Validation of Predictive Models of Radiation Effects on Tumor Growth Based on Noninvasive Imaging Data. *Comput. Methods Appl. Mech. Eng.*, 327:277–305.

[Logg et al., 2012] Logg, A., Mardal, K.-A., and Wells, G. (2012). *Automated Solution of Differential Equations by the Finite Element Method - The FEniCS Book*, volume 84 of *Lecture Notes in Computational Science and Engineering*. Springer–Verlag Berlin Heidelberg.

[Marzouk and Najm, 2009] Marzouk, Y. M. and Najm, H. N. (2009). Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems. *J. Comput. Phys.*, 228(6):1862–1902.

[Marzouk et al., 2007] Marzouk, Y. M., Najm, H. N., and Rahn, L. A. (2007). Stochastic spectral methods for efficient Bayesian solution of inverse problems. *J. Comput. Phys.*, 224(2):560–586.

[Meghdadi et al., 2017] Meghdadi, N., Niroomand-Oscuii, H., Soltani, M., Ghalichi, F., and Pourgolmohammad, M. (2017). Brain tumor growth simulation: model validation through uncertainty quantification. *Int. J. Syst. Assur. Eng. Manag.*, 8(3):655–662.

[Menze et al., 2011] Menze, B., Van Leemput, K., Honkela, A., Konukoglu, E., Weber, M.-A., Ayache, N., and Golland, P. (2011). A generative approach for image-based modeling of tumor growth. *Inf. Process Med. Imaging*, 22:735–747.

[Oden et al., 2010] Oden, J., Hawkins, A., and Prudhomme, S. (2010). General diffuse-interface theories and an approach to predictive tumour growth modeling. *Math. Models Methods Appl. Sci.*, 20(3):477–517.

[Oden et al., 2016] Oden, J., Lima, E., Almeida, R., Feng, Y., Rylander, M., Fuentes, D., Faghihi, D., Rahman, M., DeWitt, M., Gadde, M., and Zhou, J. (2016). Toward predictive multiscale modeling of vascular tumor growth: Computational and experimental oncology for tumor prediction. *Arch. Computat. Methods Eng.*, 23:735–779.

[Oden et al., 2013] Oden, J., Prudencio, E., and Hawkins-Daarud, A. (2013). Selection and assessment of phenomenological models of tumor growth. *Math. Models Methods Appl. Sci.*, 23(7):1309–1338.

[Paek and Choi, 2014] Paek, J. and Choi, I. (2014). Bayesian inference of the stochastic Gompertz growth model for tumor growth. *Commun. Stat. Appl. Methods*, 21(6):521–528.

[Patmanidis et al., 2017] Patmanidis, S., Charalampidis, A., Kordonis, I., Mitsis, G., and Papavassilopoulos, G. (2017). Comparing methods for parameter estimation of the Gompertz tumor growth model. *IFAC PapersOnLine*, 50(1):12203–12209.
[Robert and Casella, 2004] Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer.

[Schillings and Schwab, 2013] Schillings, C. and Schwab, C. (2013). Sparse, adaptive Smolyak quadratures for Bayesian inverse problems. *Inverse Problems*, 29(6):065011, 28.

[Schmidt and Siebert, 2005] Schmidt, A. and Siebert, K. (2005). *Design of adaptive finite element software: The finite element toolbox ALBERTA*, volume 42 of *Lecture Notes in Computational Science and Engineering*. Springer–Verlag Berlin Heidelberg.

[Schwab and Stuart, 2012] Schwab, C. and Stuart, A. M. (2012). Sparse deterministic approximation of Bayesian inverse problems. *Inverse Problems*, 28(4):045003, 32.

[Stein, 1999] Stein, M. (1999). *Interpolation of Spatial Data - Some Theory for Kriging*. Springer Series in Statistics. Springer–Verlag New York.

[Stuart, 2010] Stuart, A. (2010). Inverse problems: A Bayesian perspective. *Acta Numer.*, 19:451–559.

[Stuart and Teckentrup, 2018] Stuart, A. M. and Teckentrup, A. L. (2018). Posterior consistency for Gaussian process approximations of Bayesian posterior distributions. *Math. Comp.*, 87(310):721–753.

[Sullivan, 2015] Sullivan, T. (2015). *Introduction to Uncertainty Quantification*. Springer International Publishing, Switzerland.

[Wise et al., 2008] Wise, S., Lowengrub, J., Frieboes, H., and Cristini, V. (2008). Three-dimensional multispecies nonlinear tumor growth - I: model and numerical method. *J. Theor. Biol.*, 253:524–543.

[Yan and Zhang, 2017] Yan, L. and Zhang, Y.-X. (2017). Convergence analysis of surrogate-based methods for Bayesian inverse problems. *Inverse Problems*, 33(12):125001, 20.