PREPRINT 485

NOVEMBER 2018

Kinetic Methods for Inverse Problems

Michael Herty and Giuseppe Visconti

Institut für Geometrie und Praktische Mathematik
Templergraben 55, 52062 Aachen, Germany

Institut für Geometrie und Praktische Mathematik, RWTH Aachen University, Templergraben 55, D-52056 Aachen, Germany;
Email: herty@googlemail.com.de, visconti@igpm.rwth-aachen.de

Kinetic Methods for Inverse Problems

Michael Herty and Giuseppe Visconti

Institut für Geometrie und Praktische Mathematik (IGPM)
RWTH Aachen University
Templergraben 55, 52062 Aachen, Germany

November 22, 2018

Abstract

The Ensemble Kalman Filter method can be used as an iterative numerical scheme for parameter identification or nonlinear filtering problems. We study the limit of infinitely large ensemble size and derive the corresponding mean-field limit of the ensemble method. The kinetic equation allows in simple cases to analyze stability of the solution to inverse problems as mean of the distribution of the ensembles. Further, we present a slight but stable modification of the method which leads to a Fokker-Planck-type kinetic equation. The kinetic methods proposed here are able to solve the problem with a reduced computational complexity in the limit of a large ensemble size. We illustrate the properties and the ability of the kinetic model to provide solution to inverse problems by using examples from the literature.

Mathematics Subject Classification (2010) 35Q84, 65N21, 93E11, 65N75

Keywords Kinetic Partial Differential Equations, Nonlinear Filtering Methods, Inverse Problems

1 Introduction

We are concerned with the following abstract inverse problem or parameter identification problem

$$y = \mathcal{G}(u) + \eta \tag{1}$$

where $\mathcal{G} : X \to Y$ is the (possible nonlinear) forward operator between Hilbert spaces X and Y, $u \in X$ is the control, $y \in Y$ is the observation and η is observational noise. Given noisy measurements or observations y and the known mathematical model $\mathcal{G}$, we are interested in finding the corresponding control u. Typically, the observational noise η is not explicitly known but only information on its distribution is available. Inverse problems, in particular in view of a possible ill-posedness, have been discussed in vast amount of literature and we refer to [16] for an introduction and further references. In the following we will investigate a particular numerical method for solving problem (1), namely, Ensemble Kalman Filtering (EnKF). While this method has already been introduced more than ten years ago citeEvensen1994, recent theoretical progress [38] is the starting point of this work.

In order to set up the mathematical formulation we consider the case $X = \mathbb{R}^d$ and $Y = \mathbb{R}^K$, with $d, K \in \mathbb{N}$. As in [38] we aim to solve the inverse problem by minimizing the least squares functional

$$\Phi(u, y) := \frac{1}{2} \left\| \Gamma^{\frac{1}{2}} (y - \mathcal{G}(u)) \right\|_Y^2 \tag{2}$$

where Γ normalizes the so-called model-data misfit. This is defined as the covariance of the noise η. Note that so far there is no regularization of the control u added towards the minimization problem, see e.g. [4, 20, 22, 28] for examples of Tikhonov and other regularizations.

Later we also explore the link of EnKF to Bayesian inversion and therefore we briefly recall a Bayesian inversion formulation for problem (1). Following [13, 40] a solution to the inverse problem is obtained by treating the unknown control u, the data y and the noise η as random variables. Then, the conditional probability measure of the control u given the observation y, denoted by $u|y$, called posterior, is computed via Bayes Theorem. Typically, there is an interest in moments of the posterior, e.g. choosing the point of maximal probability (MAP estimator). For further details concerning Bayesian inversion, such as the modeling of the unknown prior distributions and other choices of estimators, see e.g. [3, 8, 13, 17] and references therein.

Before finally stating the aim of this work, we briefly recall some references on the EnKF method without aiming to give a complete list. Iterative filtering methods have also been successfully applied to inverse problems since many years. A particular successful method has been originally proposed in [27] to estimate state variables, parameters, etc. of stochastic dynamical systems. This method has been extended to EnKF in [18]. The EnKF sequentially updates each member of an ensemble of random elements in the space X by means of the Kalman update formula and using the knowledge of the model $\mathcal{G}$ and of given observational data y. It is important to note that *no* information on the derivative of $\mathcal{G}$ is required. Some examples in mathematical literature of the application of the filtering method to inverse problems are given in the incomplete list [5, 6, 24, 25, 38, 39]. Our starting point is [38] where the time-continuous limit of the EnKF has been studied as a regularization technique for minimization of the least squares functional (2) with a finite ensemble size. Note that the EnKF can also formally be derived within the Bayesian framework [17, 26, 29, 30, 32]. In the cited references the ensemble size was fixed and, due to the possible associated high computational cost, limited to a small number of ensembles. The analysis of the method for a large ensemble size limit has been investigated in [17, 31]. However, to the best of our knowledge, an evolution equation for the probability distribution of the ensembles has not been derived. We aim to provide a continuous representation of the EnKF method that also holds in the limit of infinitely many ensembles.

We proceed as follows: We start from the continuous time limit equation derived in [38] and interpret the method as interacting particle system. Then, we study the mean-field limit for large ensemble sizes. From a mathematical point of view, this technique have been widely used to reduce the computational complexity and to analyze interacting particle models e.g. in socio-economic dynamics or gas dynamics [10, 11, 12, 21, 23, 36, 41, 42]. The kinetic equation evolves in time the probability distribution of the control and the solution to the inverse problem is shown to be the mean of this distribution. We link this formulation to the Bayesian approach and analyze linear stability of the EnKF. Further, we present suitable modifications of the method based on the kinetic formulation in order to obtain a computational gain in the numerical simulations using a Monte Carlo approach, similar to [1, 7, 19, 33, 34, 35, 36], and to modify the stability pattern.

2 From the Ensemble Kalman Filter to the gradient descent equation

The Ensemble Kalman Filter (EnKF) has been introduced [18] as a discrete time method to estimate state variables, parameters, etc, of stochastic dynamical systems. The estimations are based on system dynamics and measurement data that is possibly perturbed by known noise. The EnKF is a generalization and improved version of the classical Kalman Filter method [27]. In the following, we briefly review the definition of the EnKF which is based on a sequential update of an ensemble of states and parameters. Then we recover the continuous time limit equation derived in the recent work [38]. This will be the starting point to introduce and compute in the next sections a mean-field limit for the limit of a large ensemble size. The arising kinetic partial differential equations allows subsequent analysis on the nature of the method.

As in [38] we consider a control $\mathbf{u} \in \mathbb{R}^d$, a given state $\mathbf{y} \in \mathbb{R}^K$ coupled by the system dynamic $\mathcal{G}$ as stated by equation (1). The problem is to identify the unknown control $\mathbf{u}$ given possibly perturbed measurements of the state $\mathbf{y}$. Hence, the observation of the system dynamic $\mathcal{G}(\mathbf{u})$ is perturbed by noise $\boldsymbol{\eta} \in \mathbb{R}^K$. The noise is assumed independent on the control $\mathbf{u} \in \mathbb{R}^d$ and normally distributed with zero mean and known covariance matrix $\boldsymbol{\Gamma}^{-1} \in \mathbb{R}^{K\times K}$, i.e. $\boldsymbol{\eta} \sim \mathcal{N}(0, \boldsymbol{\Gamma}^{-1})$. We consider a number J of ensembles (realizations of the control) combined in $\mathbf{U} = \left\{\mathbf{u}^j\right\}_{j=1}^J$. The EnKF is originally posed as a discrete iteration on $\mathbf{U}$. The iteration index is denoted by n and the collection of the ensembles by $\mathbf{u}^{j,n} \in \mathbb{R}^d$, $\forall\, j = 1, \dots, J$ and $n \geq 0$. According to [38], the EnKF iterates each component of $\mathbf{U}^n$ at iteration $n+1$ as

$$\begin{aligned}\mathbf{u}^{j,n+1} &= \mathbf{u}^{j,n} + \mathbf{C}(\mathbf{U}^n)\left(\mathbf{D}(\mathbf{U}^n) + \frac{1}{\Delta t}\boldsymbol{\Gamma}^{-1}\right)^{-1}(\mathbf{y}^{j,n+1} - \mathcal{G}(\mathbf{u}^{j,n}))\\ \mathbf{y}^{j,n+1} &= \mathbf{y} + \boldsymbol{\xi}^{j,n+1}\end{aligned} \tag{3}$$

for each $j = 1, \dots, J$. Here, each observation or measurement $\mathbf{y}^{j,n+1} \in \mathbb{R}^K$ has been perturbed by $\boldsymbol{\xi}^{j,n+1} \sim \mathcal{N}(0, \Delta t^{-1}\boldsymbol{\Sigma})$, and $\Delta t \in \mathbb{R}^+$ is a parameter. As [38] two cases for the covariance $\boldsymbol{\Sigma}$ will be discussed: $\boldsymbol{\Sigma} = 0$ corresponding to a problem where the measurement data $\mathbf{y}$ is unperturbed and $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}$ corresponding to the case where $\boldsymbol{\xi}^{j,n+1}$ are realizations of the noise $\boldsymbol{\eta}$.

Note that the update (3) of the ensembles requires the knowledge of the operators $\mathbf{C}(\mathbf{U}^n)$ and $\mathbf{D}(\mathbf{U}^n)$ being the covariance matrices depending on the ensemble set $\mathbf{U}^n$ at iteration n and on $\mathcal{G}(\mathbf{U}^n)$, i.e. the image of $\mathbf{U}^n$ at iteration n. More precisely,

$$\begin{aligned}\mathbf{C}(\mathbf{U}^n) &= \frac{1}{J}\sum_{k=1}^{J}\left(\mathbf{u}^{k,n} - \overline{\mathbf{u}}^n\right)\otimes\left(\mathcal{G}(\mathbf{u}^{k,n}) - \overline{\mathcal{G}}^n\right) \in \mathbb{R}^{d\times K}\\ \mathbf{D}(\mathbf{U}^n) &= \frac{1}{J}\sum_{k=1}^{J}\left(\mathcal{G}(\mathbf{u}^{k,n}) - \overline{\mathcal{G}}^n\right)\otimes\left(\mathcal{G}(\mathbf{u}^{k,n}) - \overline{\mathcal{G}}^n\right) \in \mathbb{R}^{K\times K}\end{aligned} \tag{4}$$

where we have defined by $\overline{\mathbf{u}}^n$ and $\overline{\mathcal{G}}^n$ the mean of $\mathbf{U}^n$ and $\mathcal{G}(\mathbf{U}^n)$, namely

$$\overline{\mathbf{u}}^n = \frac{1}{J}\sum_{j=1}^{J}\mathbf{u}^{j,n}, \quad \overline{\mathcal{G}}^n = \frac{1}{J}\sum_{j=1}^{J}\mathcal{G}(\mathbf{u}^{j,n}).$$

In recent years, the EnKF was also studied as technique to solve classical and Bayesian inverse problems. See for instance the works [25] and [17], respectively, and the references therein. Here, we keep the attention on this type of application. The analysis of the method is proved to have a comparable accuracy with traditional least-squares approaches to inverse problems [25]. Moreover, it is known that the method provides an estimate of the unknown control $\mathbf{u}$ which lies in the subspace spanned by the initial ensemble set $\mathbf{U}^0$ [25]. We will see in this section that this property is still true at the continuous time level [38]. Concerning Bayesian inverse problems, instead, the method is proved to approximate specific Bayes linear estimators but it is able to provide only an approximation of the posterior measure by a (possibly weighted) sum of Dirac masses. For a detailed discussion we refer to [2, 17, 32].

As showed in [38], it is straightforward to compute the continuous time limit equation of the update (3) in the general case of a nonlinear model $\mathcal{G}$, even if the asymptotic analysis was performed in the easier linear setting. Consider the parameter Δt as an artificial time step for the iteration in (3), i.e. we take $\Delta t \sim N_t^{-1}$ where N_t is the maximum number of iterations. Assume then $\mathbf{U}^n \approx \mathbf{U}(n\Delta t) = \left\{\mathbf{u}^j(n\Delta t)\right\}_{j=1}^{J}$ for $n \geq 0$. Scaling by Δt and computing the limit $\Delta t \to 0^+$, the continuous time limit equation of (3) reads

$$\mathrm{d}\mathbf{u}^j = \mathbf{C}(\mathbf{U})\mathbf{\Gamma}\left(\mathbf{y} - \mathcal{G}(\mathbf{u}^j)\right)\,\mathrm{d}t + \mathbf{C}(\mathbf{U})\mathbf{\Gamma}\sqrt{\mathbf{\Sigma}}\,\mathrm{d}\mathbf{W}^j \tag{5}$$

for $j = 1, \ldots, J$, initial condition $\mathbf{U}(0) = \mathbf{U}^0$ and $\mathrm{d}\mathbf{W}^j$ are Brownian motions. Using the definition of the operator $\mathbf{C}(\mathbf{U})$, see (4), system (5) can be restated as

$$\mathrm{d}\mathbf{u}^j = \frac{1}{J}\sum_{k=1}^{J}\left\langle \mathcal{G}(\mathbf{u}^k) - \overline{\mathcal{G}}, \mathbf{y} - \mathcal{G}(\mathbf{u}^j)\right\rangle_{\mathbf{\Gamma}^{-1}}(\mathbf{u}^k - \overline{\mathbf{u}})\,\mathrm{d}t + \mathbf{C}(\mathbf{U})\mathbf{\Gamma}\sqrt{\mathbf{\Sigma}}\,\mathrm{d}\mathbf{W}^{\mathrm{j}} \tag{6}$$

for $j = 1, \ldots, J$ where $\langle \cdot, \cdot \rangle_{\mathbf{\Gamma}^{-1}} = \langle \mathbf{\Gamma}^{\frac{1}{2}} \cdot, \mathbf{\Gamma}^{\frac{1}{2}} \cdot \rangle$ and $\langle \cdot, \cdot \rangle$ is the inner-product on $\mathbb{R}^K$. From (6) it is easy to observe that the invariant subspace property holds also at the continuous time level in the case $\mathbf{\Sigma} \equiv 0$ since vector field is in the linear span of the ensemble itself.

In [38] the asymptotic behavior of the continuous time equation is analyzed in the linear setting with $\mathbf{\Sigma} \equiv 0$ so that (5) is written as gradient descent. In fact, let us consider the case of $\mathcal{G}$ linear, i.e. $\mathcal{G}(\mathbf{u}) = G\mathbf{u}$. Then the computation of the operator $\mathbf{C}(\mathbf{U})$ is $\mathbf{C}(\mathbf{U}) = \frac{1}{J}\sum_{k=1}^{J}\left(\mathbf{u}^k - \overline{\mathbf{u}}\right)\left(\mathbf{u}^k - \overline{\mathbf{u}}\right)^T G^T$. Further, note that the least squares functional (2) yields

$$\nabla_{\mathbf{u}}\Phi(\mathbf{u}, \mathbf{y}) = -G^T\mathbf{\Gamma}(\mathbf{y} - G\mathbf{u}). \tag{7}$$

Therefore, equations (5) is stated in terms of the gradient of Φ as

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{u}^j = -\frac{1}{J}\sum_{k=1}^{J}(\mathbf{u}^k - \overline{\mathbf{u}}) \otimes (\mathbf{u}^k - \overline{\mathbf{u}})\nabla_{\mathbf{u}}\Phi(\mathbf{u}^j, \mathbf{y}) \tag{8}$$

for $j = 1, \dots, J$. Equation (8) describes a preconditioned gradient descent equation for each ensemble. In fact, $\mathbf{C}(\mathbf{U})$ is positive semi-definite and hence

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi(\mathbf{u}(t), \mathbf{y}) = \frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}\left\|\mathbf{\Gamma}^{\frac{1}{2}}\left(\mathbf{y} - G\mathbf{u}\right)\right\|^2 \leq 0.$$

Observe that, although the forward operator is assumed to be linear, the gradient flow is nonlinear. For further details and properties of the gradient descent equation (8) we refer to [38]. In particular, here we recall the important result on the velocity of the collapse of the ensembles towards their mean in the large time limit.

Lemma 2.1 (Theorem 3 in [38])**.** *Let* $\mathbf{U}^0$ *be the initial set of ensembles. Then the matrix* $\mathbf{R}(t)$ *whose entries are*

$$\left(\mathbf{R}(t)\right)_{ij} = \left\langle G(\mathbf{u}^i - \overline{\mathbf{u}}), G(\mathbf{u}^j - \overline{\mathbf{u}})\right\rangle_{\mathbf{\Gamma}}$$

converges to 0 *for* $t \to \infty$ *and indeed* $\|\mathbf{R}(t)\| = O(Jt^{-1})$.

Thus, the previous Lemma states that the collapse slows down linearly as the ensemble size increases. Later, this property is also obtained in the mean-field limit for large ensemble size.

3 Mean-field limit of the Ensemble Kalman Filter

Typically, EnKF methods are applied for fixed and finite ensemble size. In fact, it is clear from (3) and (6) that the computational and memory cost of the method increases with number of ensembles. However, the analysis of the method was also studied in the large ensemble limit, see e.g. [17, 29, 31, 32]. However, to the best of our knowledge, the derivation of a kinetic equation that holds in the limit of a large number of ensembles has not yet been proposed. In this section, we derive the corresponding mean-field limit of the continuous time equation focusing on the case of a linear model G and with $\mathbf{\Sigma} = \mathbf{0}$ as in [38].

We follow the classical formal derivation to formulate a mean-field equation of a particle system, see [10, 21, 36, 41]. Let us denote by

$$f = f(t, \mathbf{u}) : \mathbb{R}^+ \times \mathbb{R}^d \to \mathbb{R}^+ \tag{9}$$

the compactly supported on $\mathbb{R}^d$ probability density in $\mathbf{u}$ at time t and introduce the first moment $\mathbf{m} \in \mathbb{R}^d$ and the second moment $\mathbf{E} \in \mathbb{R}^{d \times d}$ of f at time t, respectively, as

$$\mathbf{m}(t) = \int_{\mathbb{R}^d} \mathbf{u} f(t, \mathbf{u}) \mathrm{d}\mathbf{u}, \quad \mathbf{E}(t) = \int_{\mathbb{R}^d} \mathbf{u} \otimes \mathbf{u} f(t, \mathbf{u}) \mathrm{d}\mathbf{u}. \tag{10}$$

Since $\mathbf{u} \in \mathbb{R}^d$, the corresponding discrete measure on the ensemble set $\mathbf{U} = \left\{\mathbf{u}^j\right\}_{j=1}^J$ is therefore given by the empirical measure

$$f(t, \mathbf{u}) = \frac{1}{J}\sum_{j=1}^J \delta(\mathbf{u}^j - \mathbf{u}) = \frac{1}{J}\sum_{j=1}^J \prod_{i=1}^d \delta(u_i^j - u_i), \tag{11}$$

where $u_i^j \in \mathbb{R}$ is the component i of the j-th ensemble. Let us define the operator

$$\boldsymbol{\mathcal{C}}(\mathbf{U}) = \frac{1}{J}\sum_{k=1}^{J}(\mathbf{u}^k - \overline{\mathbf{u}}) \otimes (\mathbf{u}^k - \overline{\mathbf{u}})$$

with the corresponding entry

$$(\boldsymbol{\mathcal{C}}(\mathbf{U}))_{\kappa,\ell} = \frac{1}{J}\sum_{k=1}^{J} u_\kappa^k u_\ell^k - \overline{u}_\kappa \frac{1}{J}\sum_{k=1}^{J} u_\ell^k - \overline{u}_\ell \frac{1}{J}\sum_{k=1}^{J} u_\kappa^k + \overline{u}_\kappa \overline{u}_\ell = \frac{1}{J}\sum_{k=1}^{J} u_\kappa^k u_\ell^k - \overline{u}_\kappa \overline{u}_\ell,$$

where $\overline{u}_i$ denotes the component i of the mean $\overline{\mathbf{u}}$ of the ensembles. This formulation allows for a mean-field limit as

$$(\boldsymbol{\mathcal{C}}(t))_{\kappa,\ell} = \int_{\mathbb{R}^d} u_\kappa u_\ell f(t,\mathbf{u})\mathrm{d}\mathbf{u} - \int_{\mathbb{R}^d} u_\kappa f(t,\mathbf{u})\mathrm{d}\mathbf{u} \int_{\mathbb{R}^d} u_\ell f(t,\mathbf{u})\mathrm{d}\mathbf{u}$$

and therefore $\boldsymbol{\mathcal{C}}(\mathbf{U})$ can be written in terms of the moments (10) of the empirical measure only as

$$\boldsymbol{\mathcal{C}}(t) = \mathbf{E}(t) - \mathbf{m}(t) \otimes \mathbf{m}(t). \tag{12}$$

Let us denote $\varphi(\mathbf{u}) \in C_0^1(\mathbb{R}^d)$ a sufficiently smooth test function. We compute

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\langle f, \varphi\rangle &= \frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}^d} \frac{1}{J}\sum_{j=1}^{J}\delta(\mathbf{u}-\mathbf{u}^j)\varphi(\mathbf{u})\mathrm{d}\mathbf{u} = -\frac{1}{J}\sum_{j=1}^{J}\nabla_{\mathbf{u}}\varphi(\mathbf{u}^j)\cdot\boldsymbol{\mathcal{C}}(t)\nabla_{\mathbf{u}}\Phi(\mathbf{u}^j,\mathbf{y}) \\ &= -\int_{\mathbb{R}^d}\nabla_{\mathbf{u}}\varphi(\mathbf{u})\cdot\boldsymbol{\mathcal{C}}(t)\nabla_{\mathbf{u}}\Phi(\mathbf{u},\mathbf{y})f(t,\mathbf{u})\mathrm{d}\mathbf{u}.\end{aligned}$$

which finally leads to the following strong form of the mean-field kinetic equation to the gradient descent equation (8):

$$\partial_t f(t,\mathbf{u}) - \nabla_{\mathbf{u}}\cdot(\boldsymbol{\mathcal{C}}(t))\nabla_{\mathbf{u}}\Phi(\mathbf{u},\mathbf{y})f(t,\mathbf{u})) = 0. \tag{13}$$

Equation (13) provides a closed formula for the evolution in time of the distribution f of the unknown control $\mathbf{u}$ when the observations $\mathbf{y}$ and the linear model G are given and when endowed with an initial guess $f^0(\mathbf{u}) = f(t=0,\mathbf{u})$ for the control.

3.1 Moments and approximation of a Bayesian estimator

As discussed in Section 2, the EnKF computes a solution to the inverse problem as mean of the ensembles in the large time behavior. Since the kinetic equation (13) formally holds in the limit of a large number of ensembles, we expect that the first moment $\mathbf{m}(t)$, defined in (10), approaches to the solution of the inverse problems as $t \to \infty$.

Due to definition (10), multiplying (13) by $\mathbf{u}$, integrating over $\mathbb{R}^d$ and integrating by parts the second term, we get the following evolution equation for the first moment:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{m}(t) + \int_{\mathbb{R}^d}\boldsymbol{\mathcal{C}}(t)\nabla_{\mathbf{u}}\Phi(\mathbf{u},\mathbf{y})f(t,\mathbf{u})\mathrm{d}\mathbf{u} = \mathbf{0}.$$

In particular, since we are assuming the simple setting of a linear model $\mathcal{G}(\mathbf{u}) = G\mathbf{u}$, using (7), we can explicitly compute the integral and obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{m}(t) + \boldsymbol{\mathcal{C}}(t)\nabla_{\mathbf{u}}\Phi(\mathbf{m},\mathbf{y}) = \mathbf{0}. \tag{14}$$

Multiplying (13) by $\mathbf{u} \otimes \mathbf{u}$ and integrating over $\mathbb{R}^d$ we obtain the following evolution equation for the second moment:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{E}(t) + \sum_{k=1}^{d}\int_{\mathbb{R}^d}\mathbf{T}_k^{(1)}(\mathbf{u})\left(\boldsymbol{\mathcal{C}}(t)\nabla_{\mathbf{u}}\Phi(\mathbf{u},\mathbf{y})f(t,\mathbf{u})\right)_k \mathrm{d}\mathbf{u} = \mathbf{0}, \quad \mathbf{T}_k^{(1)}(\mathbf{u}) = \frac{\partial}{\partial u_k}\mathbf{u}\otimes\mathbf{u}. \tag{15}$$

Hence, equation (14) and equation (15) provide a closed system of ordinary differential equations.

Remark 3.1. *As in Bayesian approach to inverse problems, also equation* (13) *poses the problem of selecting a solution out of f which only provides a distribution for the control $\mathbf{u}$. As pointed out at the beginning of this subsection, since the kinetic equation is derived via mean-field limit we choose, accordingly to the solution provided by the EnKF, the expected value $\mathbf{m}$ as estimator of the distribution f. Observe that a steady-state $\mathbf{m}^\infty$ of equation* (14) *is given by*

$$\mathbf{m}^\infty = \arg\min_{\mathbf{u}} \Phi(\mathbf{u},\mathbf{y}),$$

corresponding to a control that minimizes the least squares functional Φ. In the case of a linear model G, the above condition can be also stated as $\mathbf{y} - G\mathbf{u} \in \ker G^T$. Neither, $\mathbf{u}$ nor $\mathbf{m}^\infty$ needs to be unique.

We show that the kinetic equation (13) provides also a link between the solution of a deterministic and a Bayesian approach to inverse problems. It was already proved that the EnKF fails in the approximation of the posterior, see [17]. However, Lemma 3.2 below formally proves that the expected value of the mean-field equation (13) approximates the mean estimator of the Bayesian posterior.

Lemma 3.2 (formal result)**.** *Let $\boldsymbol{\eta} \sim \mathcal{N}(0,\boldsymbol{\Gamma}^{-1})$ be a noise normally distributed with zero mean and known covariance matrix $\boldsymbol{\Gamma}^{-1} \in \mathbb{R}^{K\times K}$ independent on the control $\mathbf{u}$. Assume that all the probability measures involved in the Bayes formula have a density. Then equation* (14) *is an approximation of the Bayesian posterior mean estimator.*

Proof. Due to the Bayes' Theorem, the posterior distribution is given by

$$g(\mathbf{u}|\mathbf{y}) = \frac{\exp(\log(\nu(\boldsymbol{\eta})))g_0(\mathbf{u})}{\mu(\mathbf{y})} \tag{16}$$

where $g(\mathbf{u})$ is the prior distribution, $\nu(\boldsymbol{\eta})$ is the likelihood and

$$\mu(\mathbf{y}) = \int_{\mathbb{R}^d}\exp(\log(\nu(\boldsymbol{\eta})))g_0(\mathbf{u})\mathrm{d}\mathbf{u}$$

is the distribution of the observation $\mathbf{y}$ treated as random variable. It is easy to show that under the hypothesis $\boldsymbol{\eta} \sim \mathcal{N}(0, \boldsymbol{\Gamma}^{-1})$ the negative log likelihood coincides with the covariance weighted model-data misfit least squares functional Φ given by (2), namely

$$\Phi(\mathbf{u}, \mathbf{y}) \equiv -\log(\nu(\boldsymbol{\eta})).$$

Hence, using Bayes' formula (16) we obtain

$$g(\mathbf{u}|\mathbf{y}) - g_0(\mathbf{u}) = \frac{1}{\mu(\mathbf{y})} \left(\exp(-\Phi(\mathbf{u}, \mathbf{y})) g_0(\mathbf{u}) \mathrm{d}\mathbf{u} - g_0(\mathbf{u}) \int_{\mathbb{R}^d} \exp(-\Phi(\mathbf{v}, \mathbf{y})) g_0(\mathbf{v}) \mathrm{d}\mathbf{v} \right)$$

and then we multiply by $\mathbf{u}$ and integrate over $\mathbb{R}^d$ to obtain

$$\int_{\mathbb{R}^d} \mathbf{u} g(\mathbf{u}|\mathbf{y}) \mathrm{d}\mathbf{u} - \mathbf{m}_0 = \frac{1}{\mu(\mathbf{y})} \left(\int_{\mathbb{R}^d} \exp(-\Phi(\mathbf{u}, \mathbf{y})) g_0(\mathbf{u}) (\mathbf{u} - \mathbf{m}_0) \mathrm{d}\mathbf{u} \right). \tag{17}$$

Here, $\mathbf{m}_0 = \int_{\mathbb{R}^d} \mathbf{u} g_0(\mathbf{u}) \mathrm{d}\mathbf{u}$ is the first moment of the prior distribution. We establish the link to the evolution equation (14) of the first moment of the kinetic distribution by approximating the right-hand side of (17) by first-order Taylor expansion at $\mathbf{m}_0$. Thus, since

$$\exp(-\Phi(\mathbf{u}, \mathbf{y})) \approx \exp(-\Phi(\mathbf{m}_0, \mathbf{y})) - \exp(-\Phi(\mathbf{m}_0, \mathbf{y})) (\mathbf{u} - \mathbf{m}_0)^T \nabla_{\mathbf{u}} \Phi(\mathbf{m}_0, \mathbf{y})$$

we compute

$$\mu(\mathbf{y}) = \int_{\mathbb{R}^d} \exp(-\Phi(\mathbf{u}, \mathbf{y})) g_0(\mathbf{u}) \mathrm{d}\mathbf{u} \approx \exp(-\Phi(\mathbf{m}_0, \mathbf{y})) \int_{\mathbb{R}^d} g_0(\mathbf{u}) \mathrm{d}\mathbf{u}$$

and

$$\begin{aligned} \int_{\mathbb{R}^d} \exp(-\Phi(\mathbf{u}, \mathbf{y})) g_0(\mathbf{u}) (\mathbf{u} - \mathbf{m}_0) \mathrm{d}\mathbf{u} &\approx -\exp(-\Phi(\mathbf{m}_0, \mathbf{y})) \int_{\mathbb{R}^d} (\mathbf{u} - \mathbf{m}_0) \otimes (\mathbf{u} - \mathbf{m}_0) g_0(\mathbf{u}) \mathrm{d}\mathbf{u} \, \nabla_{\mathbf{u}} \Phi(\mathbf{m}_0, \mathbf{y}) \\ &= -\exp(-\Phi(\mathbf{m}_0, \mathbf{y})) \left(\mathbf{E}_0 - \mathbf{m}_0 \otimes \mathbf{m}_0 \right) \nabla_{\mathbf{u}} \Phi(\mathbf{m}_0, \mathbf{y}), \end{aligned}$$

where we have defined $\mathbf{E}_0 = \int_{\mathbb{R}^d} \mathbf{u} \otimes \mathbf{u} g_0(\mathbf{u}) \mathrm{d}\mathbf{u}$ the second moment of the prior distribution. Plugging the expansions into (17) we obtain

$$\int_{\mathbb{R}^d} \mathbf{u} g(\mathbf{u}|\mathbf{y}) \mathrm{d}\mathbf{u} - \mathbf{m}_0 \approx -\left(\mathbf{E}_0 - \mathbf{m}_0 \otimes \mathbf{m}_0 \right) \nabla_{\mathbf{u}} \Phi(\mathbf{m}_0, \mathbf{y}).$$

We define now $\mathbf{m}_1 = \int_{\mathbb{R}^d} \mathbf{u} g(\mathbf{u}|\mathbf{y}) \mathrm{d}\mathbf{u}$ the mean estimator of the posterior distribution and scale $\boldsymbol{\Gamma}^{-1}$ as $\Delta t \boldsymbol{\Gamma}^{-1}$ to obtain $\nabla_{\mathbf{u}} \Phi(\mathbf{u}, \mathbf{y}) = \Delta t \nabla_{\mathbf{u}} \Phi(\mathbf{u}, \mathbf{y})$. Then,

$$\mathbf{m}_1 = \mathbf{m}_0 - \Delta t \left(\mathbf{E}_0 - \mathbf{m}_0 \otimes \mathbf{m}_0 \right) \nabla_{\mathbf{u}} \Phi(\mathbf{m}_0, \mathbf{y})$$

which gives a first-order approximation in time of (14). □

3.2 One-dimensional control: linear stability analysis

The expected value $\mathbf{m}$ depends on $\mathbf{E}$ since equation (14) and (15) give rise to a coupled system of ordinary differential equations. In the following, we employ a stability analysis for these equations in the simpler case of a one-dimensional control in order to analyze the stability of the estimator $\mathbf{m}$.

First, we observe that in the case of a scalar control the system of the moment equations reduces to

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}m(t) &= G(E(t)-m^2(t))(y-Gm(t)) \\ \frac{\mathrm{d}}{\mathrm{d}t}E(t) &= 2G(E(t)-m^2(t))(ym(t)-GE(t)) \end{aligned} \tag{18}$$

with $y \in \mathbb{R}$ and $G \in \mathbb{R} \setminus \{0\}$. The nullclines of the system of ODEs (18) are given by

$$m = \frac{y}{G}, \quad E = \frac{y}{G}m, \quad E = m^2.$$

The equilibrium or fixed points of (18) are the intersections of the nullclines and therefore we have the following three sets of points:

$$F_0 = (0,0), \quad F_1 = (\frac{y}{G}, \frac{y^2}{G^2}), \quad F_k = (k, k^2),\ k \in \mathbb{R},$$

i.e. all the fixed points are on the parabola $E = m^2$ in the phase plane (m, E). Given the Jacobian $\mathbf{J} \in \mathbb{R}^{2\times 2}$ of the ODE system (18)

$$\mathbf{J}(m,E) = \begin{bmatrix} 3G^2m^2 - 2Gym - G^2E & -G^2m + Gy \\ 2GyE + 4G^2mE - 6Gym^2 & -4G^2E + 2Gym + 2G^2m^2 \end{bmatrix} \tag{19}$$

we can compute that $\mathbf{J}(F_k)$ has eigenvalues $\mu_1 = \mu_2 = 0$. Clearly, the same holds for F_0 and F_1, since they are points of the type F_k. Therefore all the fixed points are non-hyperbolic and the stability must be analyzed directly. More precisely, since $\mu_1 = \mu_2 = 0$, the fixed points are Bogdanov-Takens-type equilibria and hence unstable as we indeed show in the following analysis.

The vector field of the system (18) can be easily analyzed on nullclines and on the m- and E-axis of the phase plane. For the sake of simplicity, let us assume that $\frac{y}{G} > 0$. The analysis is equivalent in the opposite case. Let $m(0) = \frac{y}{G}$ so that $\frac{\mathrm{d}}{\mathrm{d}t}m = 0$ for all t. We have that

$$\frac{\mathrm{d}}{\mathrm{d}t}E = -\frac{2}{G^2}(y^2 - G^2E)^2 < 0$$

and therefore E is decreasing in time on the nullcline $m = \frac{y}{G}$ which in turn means that F_1 is an attractor only if $E(0) > \frac{y}{G^2}$. Let now $E(0) = \frac{y}{G}m(0)$, for some $m(0)$. Then we have $\frac{\mathrm{d}}{\mathrm{d}t}E = 0$ and

$$\frac{\mathrm{d}}{\mathrm{d}t}m = m(y - Gm)^2 = \begin{cases} > 0, & \text{if } m(0) > 0, \\ < 0, & \text{otherwise.} \end{cases}$$

Thus, since $m(0) > 0$ is the only acceptable initial condition in order to guarantee that $E(0) > 0$, the trajectories are moving on the right side of the phase plane on the nullcline $E = \frac{y}{G}m$. Obviously,

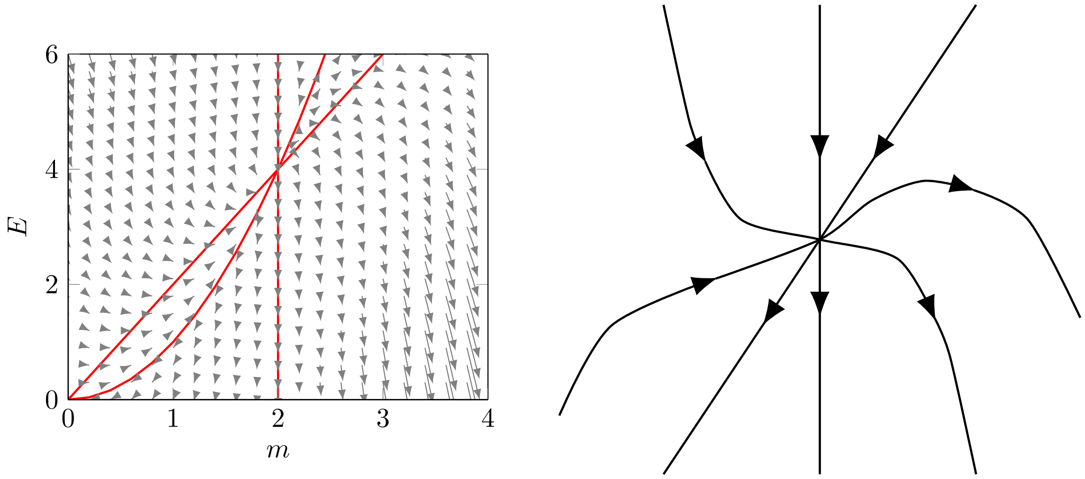

Figure 1: Left: vector field of the ODE system (18) with $(y, G) = (2, 1)$. Red lines are the nullclines. Right: trajectory behavior around the equilibrium $(\frac{y}{G}, \frac{y^2}{G^2}) = (2, 4)$.

each trajectory is still in time on the nullcline $E = m^2$ since $\frac{\mathrm{d}}{\mathrm{d}t}m = \frac{\mathrm{d}}{\mathrm{d}t}E = 0$. The nullclines and the complete vector field for the case $(y, G) = (2, 1)$ is shown in the left panel of Figure 1. We immediately observe that the behavior around the equilibrium point F_1 is unstable as showed also in the right panel of Figure 1.

The previous considerations can be also derived by looking at the solutions of (18). Assuming that the initial conditions are such that $E(0) \neq m(0)^2$, we get then the following pairs of analytical solutions:

$$m(t) = \frac{y}{G}, \quad E(t) = \frac{y^2}{G^2} + \frac{1}{2G^2(C + t)}$$

$$m(t) = \frac{y}{G} \pm \frac{1}{G\sqrt{-2C_1Gt - 2C_2G}}, \quad E(t) = m^2 + \frac{\frac{\mathrm{d}}{\mathrm{d}t}m}{G(y - Gm)}$$

with $C, C_1, C_2 \in \mathbb{R}$ constants uniquely prescribed by the initial conditions. In particular, the first set of solutions is found by assuming that $m(0) = \frac{y}{G}$ and solving the following Riccati's equation with constant coefficients

$$\frac{\mathrm{d}}{\mathrm{d}t}E(t) = -2G^2E^2 + 4Ey - 2\frac{y^4}{G^2}.$$

In this case, let $E(0) = E_0$, the constant is given by $C = \frac{1}{2}(G^2E_0^2 - y^2)$ which is positive when $E_0 > \frac{y^2}{G^2}$ and negative otherwise. In this latter case we also observe that there exists a time t in which the trajectory $E(t)$ has a vertical asymptote. For the above discussion we know that $E(t)$ is also decreasing. It is also simple to observe that in the second pair of solutions $E(t)$ can blow up driving $m(t)$ away from the equilibrium.

4 Extension of the mean-field EnKF method

The analysis of Section 3.2 shows that, at least in a one-dimensional setting, the system of moment equations (18) could lead to unconditionally unstable equilibria. This is due to the possible decay of the energy which drives the expected value far from the equilibrium value. In the general case of a d-dimensional control $\mathbf{u}$, the situation may be even more complex.

The instability of fixed points of (18) can be related to the loss of an $O(\Delta t)$ term in the derivation of the continuous time limit equation (5). In fact, instability could occur also for (8) but it is possible to show that the discrete equation (3) has stable equilibria.

Next, we stabilize the system of the moment equations (18) by introducing additional uncertainty to the microscopic interactions. This leads to a diffusive term in the kinetic equation avoiding the decay of kinetic energy and the appearance of unstable equilibria. First, we write binary microscopic interactions corresponding to the mean-field kinetic equation (13). Then, we introduce noise in these interactions and we derive a Fokker-Planck-type equation. Finally, we study the stability of the resulting moment system.

Let again $f = f(t, \mathbf{u}) : \mathbb{R}^+ \times \mathbb{R}^d \to \mathbb{R}$ be the probability density of the control $\mathbf{u} \in \mathbb{R}^d$ at time $t > 0$ as defined in (9). Let $\mathbf{m} \in \mathbb{R}^d$ and $\mathbf{E} \in \mathbb{R}^{d\times d}$ be the first and the second moment of f, respectively, as given in (10). We introduce the microscopic interaction rules:

$$\begin{aligned}\mathbf{u} &= \mathbf{u}_* - \epsilon(\mathbf{E} - \mathbf{m}\otimes\mathbf{m})\nabla_{\mathbf{u}}\Phi(\mathbf{u}_*, \mathbf{y}) + \sqrt{\epsilon}\,\mathbf{K}(\mathbf{u}_*)\boldsymbol{\xi}\\ &= \mathbf{u}_* - \epsilon\,\boldsymbol{\mathcal{C}}(t)\nabla_{\mathbf{u}}\Phi(\mathbf{u}_*, \mathbf{y}) + \sqrt{\epsilon}\,\mathbf{K}(\mathbf{u}_*)\boldsymbol{\xi}\end{aligned} \tag{20}$$

where $\mathbf{u}$ is the post-interaction value of the ensemble member, $\mathbf{u}_*$ is its pre-interaction value and $\boldsymbol{\xi} \in \mathbb{R}^d$ is a random variable with given distribution $\theta(\boldsymbol{\xi})$ having zero mean and covariance matrix $\boldsymbol{\Lambda} \in \mathbb{R}^{d\times d}$. Instead, $\mathbf{K}(\mathbf{u}_*) \in \mathbb{R}^{d\times d}$ is an arbitrary function of $\mathbf{u}_*$. For $\mathbf{K} = \mathbf{C}$ we observe a similar structure as in equation (5). The quantity ϵ describes the strength of the interactions and it is a scattering rate.

Remark 4.1. *Observe that* (20) *is in fact the microscopic interaction corresponding to the mean-field equation* (13)*, that is in the case of a linear model*

$$\mathbf{u} = \left(\mathbf{1} - \epsilon\,\boldsymbol{\mathcal{C}}(t)G^T\boldsymbol{\Gamma}G\right)\mathbf{u}_* + \epsilon\,\boldsymbol{\mathcal{C}}(t)G^T\boldsymbol{\Gamma}GG^{-1}\mathbf{y}, \tag{21}$$

with an additional term representing the uncertainty in the interaction. The interaction (21) *has a probabilistic interpretation [1] provided*

$$\epsilon\,\rho\left(\boldsymbol{\mathcal{C}}(t)G^T\boldsymbol{\Gamma}G\right) \le 1,$$

where $\rho(\cdot)$ is the spectral radius.

The probability density f satisfies the following (linear) Boltzmann equation in weak form

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}^d} f(t,\mathbf{u})\varphi(\mathbf{u})\mathrm{d}\mathbf{u} = \left\langle \int_{\mathbb{R}^d} (\varphi(\mathbf{u}) - \varphi(\mathbf{u}_*))f(t,\mathbf{u})\mathrm{d}\mathbf{u} \right\rangle \tag{22}$$

where $\varphi \in C_c^\infty(\mathbb{R}^d)$ is a test function and where the operator $\langle \cdot \rangle$ denotes the mean with respect to the distribution θ, i.e. $\langle g \rangle = \int_{\mathbb{R}^d} g(\boldsymbol{\xi})\theta(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi}$. Consider the time asymptotic scaling by setting

$$\tau = t\epsilon, \quad f(t,\mathbf{u}) = \tilde{f}(\tau,\mathbf{u}) \tag{23}$$

and allow $\epsilon \to 0^+$. This corresponds to large interaction frequencies and small interaction strengths, a situation similar to the so-called grazing collision limit [14, 15, 37, 43]. We denote the scaled quantities again by f and t, respectively. A second-order Taylor expansion yields the corresponding formal Fokker-Planck equation:

$$\begin{aligned}\varphi(\mathbf{u}) - \varphi(\mathbf{u}_*) =& \nabla_{\mathbf{u}}\varphi(\mathbf{u}_*)\cdot(\mathbf{u}-\mathbf{u}_*) + \frac{1}{2}(\mathbf{u}-\mathbf{u}_*)^T\mathbf{H}(\varphi(\mathbf{u}_*))(\mathbf{u}-\mathbf{u}_*)\\ &+ \frac{1}{2}(\mathbf{u}-\mathbf{u}_*)^T\widehat{\mathbf{H}}(\varphi;\tilde{\mathbf{u}},\mathbf{u}_*)(\mathbf{u}-\mathbf{u}_*)\end{aligned}$$

with $\tilde{\mathbf{u}} = \alpha\mathbf{u}_* + (1-\alpha)\mathbf{u}$, $\alpha \in (0,1)$ and where $\mathbf{H} \in \mathbb{R}^{d\times d}$ is the Hessian matrix and $\widehat{\mathbf{H}}(\varphi;\tilde{\mathbf{u}},\mathbf{u}_*) = \mathbf{H}(\varphi(\tilde{\mathbf{u}})) - \mathbf{H}(\varphi(\mathbf{u}_*))$. Substituting this expression in equation (22) and using definition (20) of the microscopic interactions, we obtain

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}^d} f(t,\mathbf{u})\varphi(\mathbf{u})\mathrm{d}\mathbf{u} = \frac{1}{\epsilon}\left(\mathcal{A}+\mathcal{B}+\mathcal{R}\right),\\ &\mathcal{A} = -\epsilon\left\langle\int_{\mathbb{R}^d}\nabla_{\mathbf{u}}\varphi(\mathbf{u}_*)\cdot\left(\boldsymbol{\mathcal{C}}(t)\nabla_{\mathbf{u}}\Phi(\mathbf{u}_*,\mathbf{y})\right)f(t,\mathbf{u}_*)\mathrm{d}\mathbf{u}_*\right\rangle + \sqrt{\epsilon}\left\langle\int_{\mathbb{R}^d}\nabla_{\mathbf{u}}\varphi(\mathbf{u}_*)\cdot\mathbf{K}(\mathbf{u}_*)\boldsymbol{\xi}f(t,\mathbf{u}_*)\mathrm{d}\mathbf{u}_*\right\rangle,\\ &\mathcal{B} = \frac{\epsilon^2}{2}\left\langle\int_{\mathbb{R}^d}\left(\boldsymbol{\mathcal{C}}(t)\nabla_{\mathbf{u}}\Phi(\mathbf{u}_*,\mathbf{y})\right)^T\mathbf{H}(\varphi(\mathbf{u}_*))\left(\boldsymbol{\mathcal{C}}(t)\nabla_{\mathbf{u}}\Phi(\mathbf{u}_*,\mathbf{y})\right)f(t,\mathbf{u}_*)\mathrm{d}\mathbf{u}_*\right\rangle\\ &-\epsilon\sqrt{\epsilon}\left\langle\int_{\mathbb{R}^d}\left(\mathbf{K}(\mathbf{u}_*)\boldsymbol{\xi}\right)^T\mathbf{H}(\varphi(\mathbf{u}_*))\left(\boldsymbol{\mathcal{C}}(t)\nabla_{\mathbf{u}}\Phi(\mathbf{u}_*,\mathbf{y})\right)f(t,\mathbf{u}_*)\mathrm{d}\mathbf{u}_*\right\rangle\\ &+\frac{\epsilon}{2}\left\langle\int_{\mathbb{R}^d}\mathrm{Tr}\left((\boldsymbol{\xi}\otimes\boldsymbol{\xi})^T\mathbf{K}(\mathbf{u}_*)^T\mathbf{H}(\varphi(\mathbf{u}_*))\mathbf{K}(\mathbf{u}_*)\right)f(t,\mathbf{u}_*)\mathrm{d}\mathbf{u}_*\right\rangle,\end{aligned}$$

where $\mathrm{Tr}(\cdot)$ is the matrix trace and $\mathcal{R}$ is the remaining term. One can easily prove that $\epsilon^{-1}\mathcal{R}$ vanishes in the asymptotic scaling (23). In order to show this, it is sufficient the fact that φ is an enough smooth function and thus each second partial derivative is Lipschitz continuous so that $\exists\, L > 0$ such that

$$\left|\frac{\partial^2\varphi(\tilde{\mathbf{u}})}{\partial u_i u_j} - \frac{\partial^2\varphi(\mathbf{u}_*)}{\partial u_i u_j}\right| \le L|\tilde{\mathbf{u}}-\mathbf{u}_*| < L|\mathbf{u}-\mathbf{u}_*| = L|\epsilon\boldsymbol{\mathcal{C}}(t)\nabla_{\mathbf{u}}\Phi(\mathbf{u}_*,\mathbf{y}) + \sqrt{\epsilon}\mathbf{K}(\mathbf{u}_*)\boldsymbol{\xi}| \xrightarrow{\epsilon\to 0^+} 0$$

for all $i,j = 1,\dots,d$.

For $\mathbf{K}(\mathbf{u}_*)$ constant or depending on moments of the kinetic distribution f, the grazing limit in strong form is then obtained as

$$\partial_t f(t,\mathbf{u}) = \nabla_{\mathbf{u}}\cdot\left(\boldsymbol{\mathcal{C}}(t)\nabla_{\mathbf{u}}\Phi(\mathbf{u},\mathbf{y})f(t,\mathbf{u})\right) + \frac{1}{2}\nabla_{\mathbf{u}}\cdot\left(\boldsymbol{\Lambda}\mathbf{K}^T\mathbf{K}\nabla_{\mathbf{u}}f(t,\mathbf{u}_*)\right) \tag{24}$$

where we used the basic fact

$$\mathrm{Tr}\left(\boldsymbol{\Lambda}\mathbf{K}^T\mathbf{K}\mathbf{H}(f(t,\mathbf{u}_*))\right) = \nabla_{\mathbf{u}}\cdot\left(\boldsymbol{\Lambda}\mathbf{K}^T\mathbf{K}\nabla_{\mathbf{u}}f(t,\mathbf{u}_*)\right).$$

Some remarks are in order. As expected, the Fokker-Planck-type equation (24) is consistent with the kinetic equation (13) in the limit of vanishing covariance $\mathbf{\Lambda}$. The introduction of the uncertainty in (20) allows for a different interpretation of the data perturbation in (3) and (6) within the kinetic model.

4.1 Moments equations and linear stability analysis

In the setting of [38] we have $\mathcal{G}(\mathbf{u}) = G\mathbf{u}$ and, for $\mathbf{K} = \mathbf{I}$ identity matrix, a straightforward computation leads to the following moment equations based on the Fokker–Planck equation (24).

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{m}(t) = -\boldsymbol{\mathcal{C}}(t)\nabla_{\mathbf{u}}\Phi(\mathbf{m},\mathbf{y})\\ &\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{E}(t) = -\sum_{k=1}^{d}\int_{\mathbb{R}^d}\mathbf{T}_k^{(1)}(\mathbf{u})\left(\boldsymbol{\mathcal{C}}(t)\nabla_{\mathbf{u}}\Phi(\mathbf{u},\mathbf{y})f(t,\mathbf{u})\right)_k \mathrm{d}\mathbf{u} + \frac{1}{2}\sum_{i,j=1}^{d}\Lambda_{ij}^2\int_{\mathbb{R}^d}\mathbf{T}_{ij}^{(2)}(\mathbf{u})f(t,\mathbf{u})\mathrm{d}\mathbf{u},\end{aligned} \tag{25}$$

where $\mathbf{m}$ and $\mathbf{E}$ are defined as before and where we have

$$\mathbf{T}_k^{(1)}(\mathbf{u}) = \frac{\partial}{\partial u_k}\mathbf{u}\otimes\mathbf{u}, \quad \mathbf{T}_{ij}^{(2)}(\mathbf{u}) = \frac{\partial}{\partial u_i u_j}\mathbf{u}\otimes\mathbf{u}.$$

Comparing equation (25) and (14), we observe that they are equivalent. This implies that the equation for $\mathbf{m}$ is still providing a solution according to Remark 3.1 and Lemma 3.2. Instead, the equation of the second moment $\mathbf{E}$ has an additional term that stabilize the equilibria of (25).

We analyze linear stability of (25) in the case of a one-dimensional control. In this particular case the moment equations are

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}m(t) = G(E(t) - m(t)^2)(y - Gm(t))\\ &\frac{\mathrm{d}}{\mathrm{d}t}E(t) = 2G(E(t) - m(t)^2)(ym(t) - GE(t)) + \lambda^2\end{aligned} \tag{26}$$

with $y \in \mathbb{R}$, $G \in \mathbb{R} \setminus \{0\}$ and where now $\lambda^2 \in \mathbb{R}$ represents the variance of the univariate noise ξ. Following the same analysis performed in Section 3.2, we compute the nullclines of the ODE system (26) and they are given by

$$m = \frac{y}{G}, \quad E = m^2, \quad E = \frac{m(y + Gm) \pm \sqrt{m^2(y - Gm)^2 + 2\lambda^2}}{2G}.$$

We are interested in the behavior around the equilibrium with $m = \frac{y}{G}$ which is obtained as intersection of the first and the third nullcline:

$$\tilde{F}_1^{\pm} = (\frac{y}{G}, \frac{y^2}{G^2} \pm \frac{\sqrt{2\lambda^2}}{2G}).$$

Observe that this equilibrium point is in fact the equilibrium point F_1 given in Section 3.2 when $\lambda \to 0^+$. For simplicity, in the following we consider $y, G > 0$. Similar considerations can be done in the other cases. Letting $m(0) = \frac{y}{G}$ so that $\frac{\mathrm{d}}{\mathrm{d}t}m = 0$ for all t, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}E = -2G^2\left(E - \frac{y^2}{G^2}\right)^2 + \lambda^2$$

where the right-hand side represents a parabola in E with negative leading coefficient. Therefore, using classical arguments of stability theory for ODEs, we can state that the greater root $\tilde{F}_1^+$ is the stable equilibrium and the smaller root $\tilde{F}_1^-$ is the unstable equilibrium. This result can be also obtained by looking at the eigenvalues of the Jacobian matrix of the system (26) which is equivalent to (19). In fact, computing the eigenvalues $\mu_{1,2}^{\pm}$ of $\mathbf{J}(m, E)$ evaluated in $\tilde{F}_1^{\pm}$ we have

$$\mu_1^{\pm} = \mp \frac{G}{2}\sqrt{2\lambda^2}, \quad \mu_2^{\pm} = \mp 2G\sqrt{2\lambda^2}$$

and therefore the equilibrium $\tilde{F}_1^+$ corresponding to the two negative eigenvalues is stable. Moreover, we stress the fact that the in the case of (26) the equilibria are no longer non-hyperbolic as in the case of (18). However, the variance λ^2 plays the role of a bifurcation parameter since for $\lambda^2 \to 0^+$ we recover the Bogdanov-Takens-type equilibria and thus λ^2 changes the stability of the equilibrium point. In view of this consideration we wish to avoid λ^2 going to zero and, furthermore, we can apply a control on it in order to guarantee that the unstable equilibrium $\tilde{F}_1^-$ is always negative and thus not admissible. More precisely, the standard deviation should satisfy

$$\lambda > \frac{y^2\sqrt{2}}{G}.$$

Then, the solutions of (26) are given by

$$m(t) = \frac{y}{G}, \quad E(t) = \frac{\left(\tanh(\sqrt{2\lambda^2}GC + \sqrt{2\lambda^2}Gt\right)\sqrt{2\lambda^2}G + 2y^2}{2G^2}$$

and

$$\begin{aligned} m(t) &= \frac{\pm e^{2\sqrt{2\lambda^2}Gt}C_1 y \mp C_2 y + \sqrt{\sqrt{2\lambda^2}C_1 e^{3\sqrt{2\lambda^2}Gt} - C_2\sqrt{2\lambda^2}e^{2\sqrt{2\lambda^2}Gt}}}{(C_1 e^{2\sqrt{2\lambda^2}Gt} - C_2)G}, \\ E(t) &= \frac{m(t)^3 G^2 - m(t)^2 G y - \frac{\mathrm{d}}{\mathrm{d}t}m(t)}{m(t)G^2 - Gy} \end{aligned}$$

with $C, C_1, C_2 \in \mathbb{R}$ constants uniquely prescribed by the initial conditions. We observe that, in the large time behavior, $m \to \frac{y}{G}$ unconditionally, as in the case of (18). Instead, the large time behavior of E is changed and shifted by a quantity depending on λ^2 which avoids the possibility of having a decay in the energy which drives m away from the expected equilibrium value. In Figure 2 we show the nullclines and the complete vector field for the case $(y, G) = (2, 1)$ (left panel) and the behavior around the stable equilibrium point $\tilde{F}_1^+ = (2, 8)$ (right panel).

Remark 4.2. *Lemma* (2.1) *shows that the collapse of the ensembles towards their mean slows down linearly as the number of the ensemble increases. The kinetic equation* (24) *holds in the limit of a large ensemble size and the energy* $\mathbf{E}$ *gives information on the concentration of the distribution* f *of the control* $\mathbf{u}$ *around its mean. The previous analysis shows also that, in fact, the result of Lemma* (2.1) *holds at the kinetic level since* $\mathbf{E}$ *does not decay to zero as* $t \to \infty$.

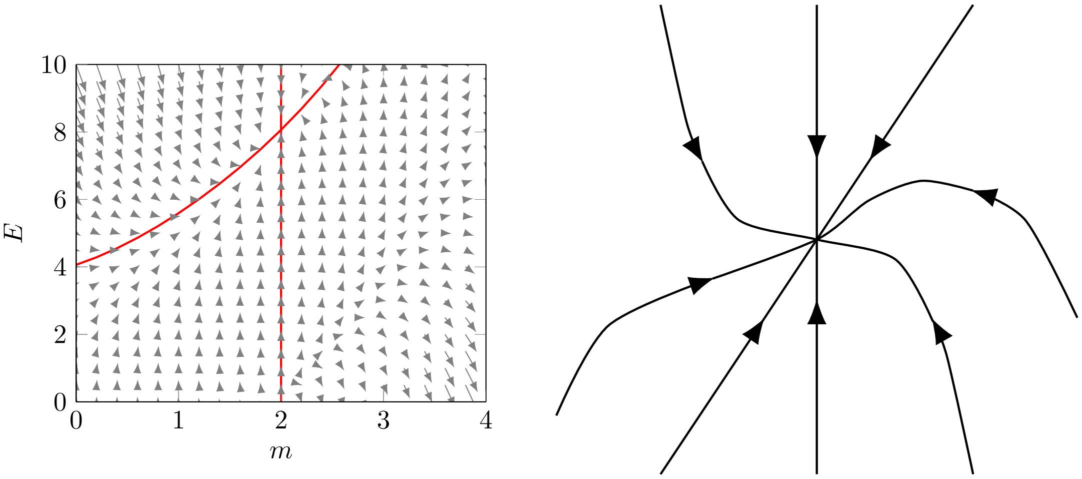

Figure 2: Left: vector field of the ODE system (26) with $(y, G) = (2, 1)$. Red lines are the nullclines. Right: trajectory behavior around the equilibrium $\tilde{F}_1^+ = (2, 8)$.

5 Numerical simulation results

The simulations are performed by using a standard Monte Carlo approach [9] to solve the kinetic equation (24). More precisely, we use a simple modification of the mean-field interaction algorithm given in [1] which is a direct simulation Monte Carlo method based on the mean-field microscopic dynamics described by (20) giving rise to the corresponding kinetic equation (24). For further details on the method we refer to [7, 19, 33, 34, 35, 36].

The algorithmic details are as follows. In each example we consider a sampling of J controls $\{\mathbf{u}^j\}_{j=1}^J$ from the prior or initial distribution $f_0(\mathbf{u})$. Then, each sample is updated according to the mean-field microscopic rule (20) by selecting $M \leq J$ interacting particles uniformly distributed without repetition. The parameter ϵ in (20) is closely related with the concept of a time step and it is taken such that stability of the discrete method is guaranteed. [1]. In particular, for the kinetic model (13) we require that

$$\epsilon \leq \frac{1}{\max_i \left(|(\Re(\mu_i)| \right)} \tag{27}$$

where the μ_i's are the eigenvalues of $\boldsymbol{\mathcal{C}}(t) G^T \boldsymbol{\Gamma} G$, cf. Remark 4.1. In the following we will use an adaptive choice of ϵ by recomputing it at each iteration. As we observe that $\boldsymbol{\mathcal{C}}(t) G^T \boldsymbol{\Gamma} G$ is characterized by large spectral radius at initial time that reduces over time, we chose an adaptive computation of ϵ.

As already pointed out in Section 4, the microscopic interactions (20) are closely related to a time discretization of the gradient descent equation (8). However, a deterministic numerical method for (8) requires $O(J^2)$ operations due to the direct evaluation of the sum for J ensembles. The numerical discretization of the kinetic equation by means of a Monte Carlo approach allows to compute the microscopic dynamics with a cost directly proportional to the number J of ensembles.

Information on the simulation results is presented in the following norms

$$\begin{aligned} v &= \frac{1}{J}\sum_{j=1}^{J} \|\mathbf{v}^j\|_2^2, \quad r = \frac{1}{J}\sum_{j=1}^{J} \|\mathbf{r}^j\|_2^2 \\ V &= \frac{1}{J}\sum_{j=1}^{J} |\mathbf{V}_{jj}|^2, \quad R = \frac{1}{J}\sum_{j=1}^{J} |\mathbf{R}_{jj}|^2 \end{aligned} \tag{28}$$

at each iteration, where

$$\begin{aligned} \mathbf{v}^j &= \mathbf{u}^j - \overline{\mathbf{u}}, \quad \mathbf{r}^j = \mathbf{u}^j - \mathbf{u}^\dagger \\ \mathbf{V}_{ij} &= \left\langle G\mathbf{v}^i, G\mathbf{v}^j \right\rangle_{\mathbf{\Gamma}^{-1}}, \quad \mathbf{R}_{ij} = \left\langle G\mathbf{r}^i, G\mathbf{r}^j \right\rangle_{\mathbf{\Gamma}^{-1}}. \end{aligned} \tag{29}$$

The quantity $\mathbf{v}^j$ measures the deviation of the j-th sample from the mean $\overline{\mathbf{u}}$ of the approximated distribution by the samples and $\mathbf{r}^j$ measures the deviation of the j-th sample from the truth solution $\mathbf{u}^\dagger$. The quantities $\mathbf{V}$ and $\mathbf{R}$ give information on the deviation of $\mathbf{v}^j$ and $\mathbf{r}^j$ under application of the model G.

Another additional important quantities is given by the misfit which allows to measure the quality of the solution at each iteration. The misfit for the j-th sample is defined as

$$\boldsymbol{\vartheta}^j = G\mathbf{r}^j - \boldsymbol{\eta}. \tag{30}$$

By using (30) we finally look at

$$\vartheta = \frac{1}{J}\sum_{j=1}^{J} \|\boldsymbol{\vartheta}^j\|_{\mathbf{\Gamma}^{-1}}^2. \tag{31}$$

Driving this quantity to zero leads to over-fitting of the solution. For this reason, usually it is suitable introducing a stopping criterion which avoids this effect. In the following we will consider the discrepancy principle which check and stop the simulation when the condition $\vartheta \leq \|\boldsymbol{\eta}\|_2^2$ is satisfied.

5.1 Linear elliptic problem

A test proposed e.g. in [25, 38], is the ill-posed inverse problem of finding the force function of an elliptic equation in one spatial dimension assuming that noisy observation of the solution to the problem are available. This problem is widely used since is explicitly solvable due to the linearity of the model.

The problem is prescribed by the following one dimensional elliptic equation

$$-\frac{\mathrm{d}^2}{\mathrm{d}x^2}p(x) + p(x) = u(x), \quad x \in [0, \pi]$$

endowed with boundary conditions $p(0) = p(\pi) = 0$. The linear model is thus defined as

$$A = \left(-\frac{\mathrm{d}^2}{\mathrm{d}x^2} + 1\right)^{-1}$$

which can be discretized, for instance, by a finite difference method or by the explicit solution

$$p(x) = A\,u(x) = \exp(x)\left(C_1 - \frac{1}{2}\int_0^x \exp(y)u(y)\mathrm{d}y\right) + \exp(-x)\left(C_2 + \frac{1}{2}\int_0^x \exp(-y)u(y)\mathrm{d}y\right)$$

where the constants C_1 and C_2 can be uniquely determined by the boundary conditions. We assign a continuous control $u(x)$ and then introduce a uniform mesh consisting of $d = K = 2^8$ equidistant points in the interval $[0, \pi]$. Let $\mathbf{u}^\dagger \in \mathbb{R}^d$ be the vector of the evaluations of the control function $u(x)$ on the mesh. We simulate noisy observations $\mathbf{y} \in \mathbb{R}^K$ as

$$\mathbf{y} = \mathbf{p} + \boldsymbol{\eta} = G\mathbf{u}^\dagger + \boldsymbol{\eta},$$

where G is the finite difference discretization of the continuous operator A. For simplicity we assume that $\boldsymbol{\eta}$ is a Gaussian white noise, more precisely $\boldsymbol{\eta} \sim \mathcal{N}(0, \gamma^2\mathbf{I})$ with $\gamma \in \mathbb{R}^+$ and $\mathbf{I} \in \mathbb{R}^{d\times d}$ is the identity matrix. We are interested in recovering the control $\mathbf{u}^\dagger \in \mathbb{R}^d$ from the noisy observations $\mathbf{y} \in \mathbb{R}^K$ only.

The initial ensemble of particles is sampled by an initial distribution $f_0(\mathbf{u}) = \mathcal{N}(0, \mathbf{C}_0)$. The choice of $f_0(\mathbf{u})$ is related to the choice of the prior distribution in Bayesian problems. In this case $f_0(\mathbf{u})$ represents a Brownian bridge as in [38]

Test case 1. Let us consider $u(x) = 1$, $\forall\, x \in [0, \pi]$. We solve the inverse problem by the proposed method for different values M of the interacting samples. We observe that taking $M < J$ does not strongly influence the results of the simulation. But, $M < J$ allows to have a computational gain.

We allow for two values of the noise level $\gamma = 0.01$, see Figure 3, and $\gamma = 0.1$, see Figure 4. In both figures, the top panels show the residual (left) and misfit (right) decrease over the number of iterations. Due to the discrepancy principle, the simulation is automatically stopped when the misfit reaches $\|\boldsymbol{\eta}\|$. The final residual values are obviously larger in the case of $\gamma = 0.1$ due to the larger noise level present in the initial observations. The bottom panels show, form left to right, the initial noisy data which are spread around the exact solution $p(x)$ of the problem, the reconstruction of $p(x)$ and the reconstruction of the control $u(x)$ by using the mean of the samples as estimator of the solution. We observe that the different values of M does not give significantly different results.

In the left panel of Figure 5 we show the spectrum of $\boldsymbol{\mathcal{C}}(t)G^T\boldsymbol{\Gamma}G$ at initial time for $\gamma = 0.01$ and $\gamma = 0.1$. We observe that the ratio between the largest and smaller eigenvalues is very large, reflecting the ill-posedness of the problem and the need of using a small ϵ in (21) in order to guarantee stability. However, we consider an adaptive ϵ since the spectral radius is observed to decrease quickly over iterations. See the red lines in the right panel of Figure 5, where, instead, the blue lines show the corresponding values of ϵ which avoid the lack of stability.

Test case 2. Let us consider $u(x) = \sin(8x)$, $\forall\, x \in [0, \pi]$, and a fixed value of the noise level $\gamma = 0.01$. We show that the method provides a good performance also cases where the control function has a high-frequency profile. In Figure 6 we consider the results obtained with $J = 25$ and $J = 25 \cdot 2^9 = 12800$ sampling from the initial distribution $f_0(\mathbf{u})$. In order to measure the quality of the solution to the inverse problem, we again compare the residual r and the projected residual R

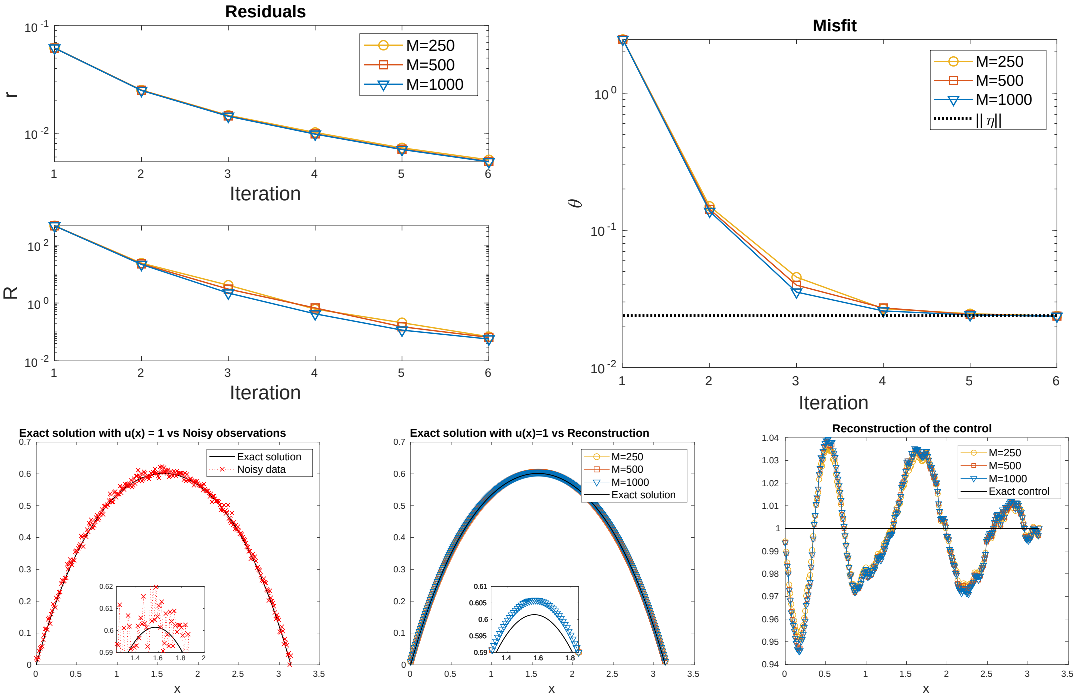

Figure 3: Elliptic problem - Test case 1 with $\gamma = 0.01$. Top row: plots of the residual r, the projected residual R and the misfit ϑ for $M = 250, 500, 1000$. Bottom row: plots of the noisy data, the reconstruction of $p(x)$ and the reconstruction of the control $u(x)$ at final iteration for $M = 250, 500, 1000$.

(top left plot) and the misfit ϑ (top right plot) for the two values of J. The misfit reaches the noise level in a very small number of iterations for both J's but the residual r for $J = 12800$ is reaching a smaller value than the residual computed with $J = 25$. This result is observable also in the middle right plot and in the bottom left plot where we compare the reconstruction of $p(x)$ and of the exact control $u(x)$ at the final iteration with the two values of J and using the mean as estimator of the solution. It is very clear that the case with $J = 12800$ is providing a better resolution. Finally, in the bottom right plot we show the relative error $\frac{\|\bar{\mathbf{u}}-\mathbf{u}\|_2^2}{\|\bar{\mathbf{u}}\|_2^2}$ as function of the increasing value of J noticing a decreasing behavior.

5.2 Nonlinear elliptic problem

The second numerical experiment is a slightly modified example proposed in [17]. We consider a one-dimensional elliptic boundary value problem given by

$$-\frac{\mathrm{d}}{\mathrm{d}x}\left(\exp(u_1)\frac{\mathrm{d}}{\mathrm{d}x}p(x)\right) = f(x), \quad x \in [0, 1]$$

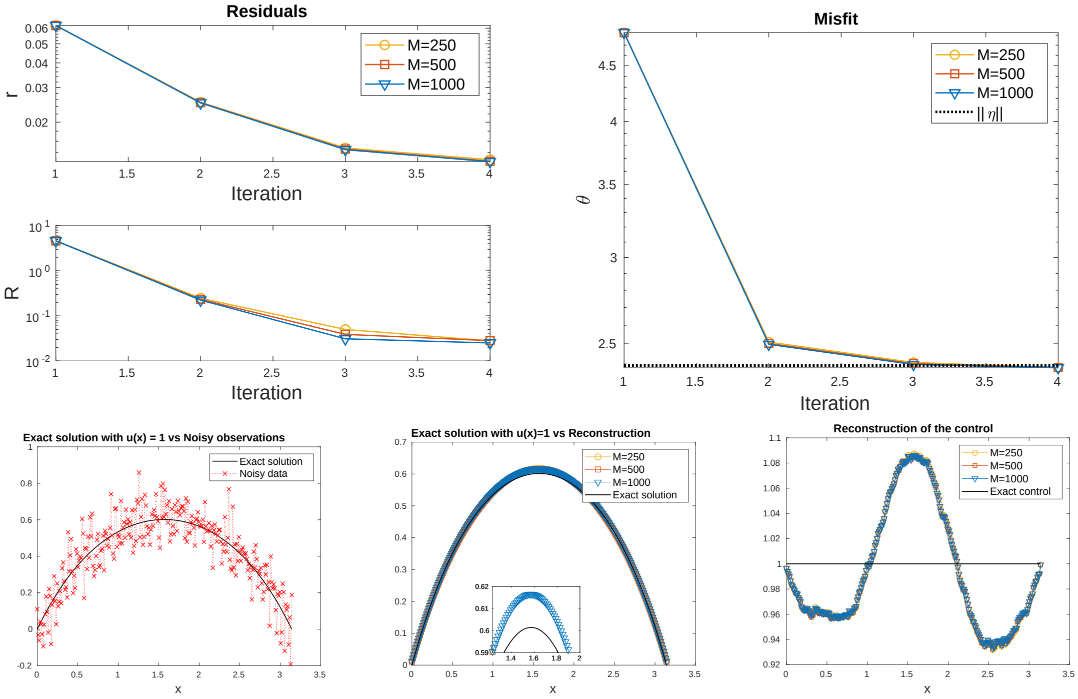

Figure 4: Elliptic problem - Test case 1 with $\gamma = 0.1$. Top row: plots of the residual r, the projected residual R and the misfit ϑ for $M = 250, 500, 1000$. Bottom row: plots of the noisy data, the reconstruction of $p(x)$ and the reconstruction of the control $u(x)$ at final iteration for $M = 250, 500, 1000$.

with boundary conditions $p(0) = p_0$ and $p(1) = u_2$, where $\mathbf{u} = (u_1, u_2)$ is the unknown control. The exact solution of this problem is given by

$$p(x) = p_0 + (u_2 - p_0) + \exp(-u_1)\left(-S_x(F) + S_1(F)x\right)$$

where $S_x(g) = \int_0^x g(y)\mathrm{d}y$ and $F(x) = S_x(f) = \int_0^x f(y)\mathrm{d}y$. In the following example we consider $f(x) = 1$, $\forall\, x \in [0, 1]$ and $p_0 = 0$, so that the explicit solution is given by

$$p(x) = u_2 x + \exp(-u_1)\left(-\frac{x^2}{2} + \frac{x}{2}\right).$$

We assume to have noisy measurements of p at the points $x_1 = \frac{1}{4}$ and $x_2 = \frac{3}{4}$ with value $\mathbf{y} = (27.5, 79.7)$. The goal is to seek the control $\mathbf{u}$ based on the knowledge of $\mathbf{y}$, of the prior $f_0(\mathbf{u})$ and of the noise model. More precisely, we consider a prior information given by $\mathbf{u} \sim \mathcal{N}(0, 1) \otimes \mathcal{U}(90, 110)$ and a Gaussian white noise $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \gamma^2 \mathbf{I})$, with $\gamma = 0.1$ and $\mathbf{I} \in \mathbb{R}^{2\times 2}$ begin the identity matrix. Thus, as in Section 5.1, noisy observations are simulated by

$$\mathbf{y} = \mathbf{p} + \boldsymbol{\eta} = \mathcal{G}(\mathbf{u}^\dagger) + \boldsymbol{\eta}$$

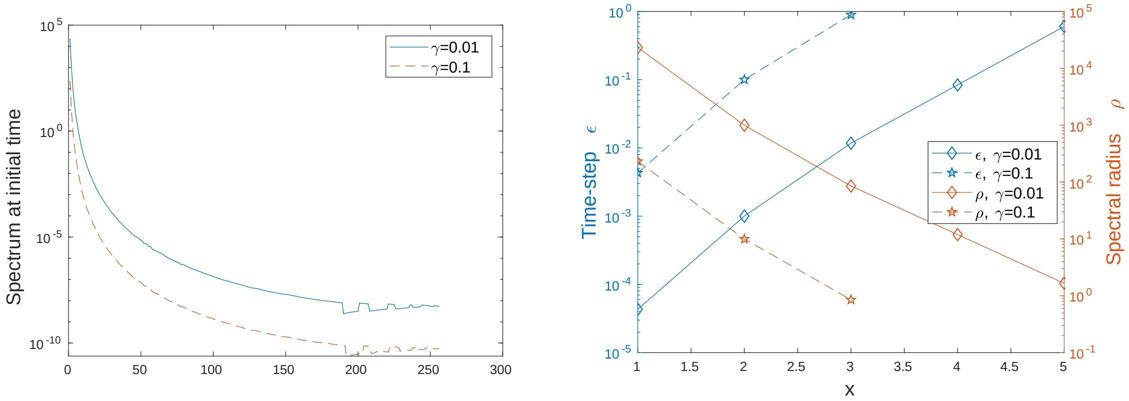

Figure 5: Elliptic problem - Test case 1. Left: spectrum of $\mathcal{C}(t)G^T\mathbf{\Gamma}G$ for the initial data with $\gamma = 0.01$ and $\gamma = 0.1$. Right: adaptive ϵ and spectral radius of $\mathcal{C}(t)G^T\mathbf{\Gamma}G$ over iterations with $\gamma = 0.01$ and $\gamma = 0.1$.

where the forward model is defined as

$$\mathcal{G}\colon \mathbf{u} \in \mathbb{R}^2 \mapsto \mathbf{p} = (p(x_1), p(x_2)) \in \mathbb{R}^2.$$

The example has $d = 2$ dimension of the control in order to make a comparison between the solution to the inverse problem provided by the kinetic method and by the Bayes' formula. In particular, it is possible to analyze the approximation of the mean estimator and of the posterior distribution computed by the kinetic equation (24). However, observe that (24) is derived by assuming a linear forward operator G but in this example the model $\mathcal{G}$ is nonlinear. Thus, inspired by (5), we consider a small modification of the microscopic interaction rule (20) with $\mathbf{K} = \mathbf{I}$ identity matrix given by

$$\mathbf{u} = \mathbf{u}_* + \epsilon\, \mathbf{C}(\mathbf{U}_*)\mathbf{\Gamma}(\mathbf{y} - \mathcal{G}(\mathbf{u}_*)) + \sqrt{\epsilon}\,\boldsymbol{\xi}$$

in order to perform the simulations for the nonlinear model.

For this example, the true posterior mean is computed in [17] thanks to Bayes' formula and it is given by $(-2.65, 104.5)$. In Figure 7 we show the results provided by the kinetic model. The top row shows the density estimation of the $J = 10^5$ sampling from the initial distribution (left plot) and the positions of the samples at the last iterations. Again, we use the discrepancy principle as stopping criterion. The middle row shows the marginals of u_1 and u_2 as relative frequency plots. The solution computed as the mean estimator of the kinetic distribution is $\mathbf{u} = (-2.56, 104.77)$ is very close to the true posterior mean, as proved also by the plot of the residuals in the bottom left panel of Figure 7.

6 Conclusions

In this paper we have introduced a kinetic model for the solution to inverse problems. The kinetic equation has been derived as mean-field limit of the Ensemble Kalman Filter method for infinitely large ensemble. The introduction of a continuous equation describing the evolution of the probability distribution of the unknown control guarantees several advantages: information on statistical quantities of the solution, implicit regularization modeled by the initial distribution, analysis of the properties of the solution and computational gain for numerical simulations.

Thanks to Lemma 3.2, we have proved that when the solution to the inverse problem is selected as the mean of the kinetic distribution, it provides an approximation of the linear mean Bayes' estimator. A linear stability analysis for the simple setting of a one dimensional control has showed that the method has only stable solutions. Numerical simulations have been performed in order to investigate the good performance of the kinetic equation in providing solutions to inverse problems.

Acknowledgments

We acknowledge support by DFG HE5386/14,15 and the Excellence Cluster of Production at RWTH Aachen University.

References

[1] G. Albi and L. Pareschi. Binary interaction algorithms for the simulation of flocking and swarming dynamics. *Multiscale Model. Simul.*, 11(1):1–29, 2013.

[2] A. Apte, M. Hairer, A. M. Stuart, and J. Voss. Sampling the posterior: An approach to non-Gaussian data assimilation. *Phys. D*, 230:50–64, 2007.

[3] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis.* Springer, 2nd edition, 1985.

[4] D. Bianchi, A. Buccini, M. Donatelli, and S. Serra-Capizzano. Iterated fractional Tikhonov regularization. *Inverse Problems*, 31(5):055005, 2015.

[5] D. Bloemker, C. Schillings, and P. Wacker. A strongly convergent numerical scheme from ensemble kalman inversion. *SIAM J. Numer. Anal.*, 56(4):2537–2562, 2018.

[6] D. Bloemker, C. Schillings, P. Wacker, and S. Weissman. Well Posedness and Convergence Analysis of the Ensemble Kalman Inversion. Preprint. arxiv:1810.08463, 2018.

[7] H. Bobovsky and H. Neunzert. On a simulation scheme for the boltzmann equation. *Math. Methods Appl. Sci.*, 8(2):223–233, 1986.

[8] M. Burger and F. Lucka. Maximum a posteriori estimates in linear inverse problems with log-concave priors are proper Bayes estimators. *Inverse Problems*, 30:114004, 2014.

[9] R. E. Caflisch. Monte Carlo and quasi-Monte Carlo methods. *Acta Numerica*, 1998:1–49, 1998.

[10] J. A. Carrillo, M. Fornasier, G. Toscani, and F. Vecil. *Mathematical Modeling of Collective Behavior in Socio-Economic and Life Sciences*, chapter Particle, kinetic, and hydrodynamic models of swarming, pages 297–336. Modeling and Simulation in Science, Engineering and Technology. Birkhäuser Boston, 2010.

[11] J. A. Carrillo, L. Pareschi, and M. Zanella. Particle based gPC methods for mean-field models of swarming with uncertainty. *Commun. Comput. Phys.*, 25:508–531, 2019.

[12] E. Cristiani, B. Piccoli, and A. Tosin. *MS&A: Modeling, Simulation and Applications*, volume 12, chapter Multiscale Modeling of Pedestrian Dynamics. Springer International Publishing, 2014.

[13] M. Dashti and A. M. Stuart. *The Bayesian Approach to Inverse Problems*, pages 311–424. Springer International Publishing, 2016.

[14] L. Desvillettes. On asymptotics of the Boltzmann equation when the collisions become grazing. *Transport Theor. Stat.*, 21(3):259–276, 1992.

[15] R. J. DiPerna and P. L. Lions. On the Fokker-Planck-Boltzmann equation. *Commun. Math. Phys.*, 120(1):1–23, 1988.

[16] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375. Springer Science and Business Media, 1996.

[17] O. G. Ernst, B. Sprungk, and H.-J. Starkloff. Analysis of the ensemble and polynomial chaos kalman filters in bayesian inverse problems. *SIAM/ASA J. Uncertain. Quantif.*, 3(1):823–851, 2015.

[18] G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *J. Geophys. Res*, 99:10143–10162, 1994.

[19] M. Fornasier, J. Haskovec, and J. Vybíral. Particle systems and kinetic equations modeling interacting agents in high dimension. *Multiscale Model. Simul.*, 9:1727–1764, 2011.

[20] C. W. Groetsch. *The theory of Tikhonov regularization for Fredholm equations of the first kind*, volume 105. Pitman Advanced Publishing Program, 1984.

[21] S.-Y. Ha and E. Tadmor. From particle to kinetic and hydrodynamic descriptions of flocking. *Kinet. Relat. Models*, 3(1):415–435, 2008.

[22] P. C. Hansen. *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion.* SIAM, 1998.

[23] M. Herty and C. Ringhofer. Averaged kinetic models for flows on unstructured networks. *Kinet. Relat. Models*, 4(4):1081–1096, 2011.

[24] M. Iglesias. Iterative regularization for ensemble data assimilation in reservoir models. *Computational Geosciences*, 19(1):177–212, 2015.

[25] M. Iglesias, K. Law, and A. M. Stuart. Analysis of the Ensamble Kalman methods for inverse problems. *Inverse Problems*, 29(4):045001, 2013.

[26] M. Iglesias, K. Law, and A. M. Stuart. Evaluation of Gaussian approximations for data assimilation in reservoir models. *Comput. Geosci.*, 17:851–885, 2013.

[27] R. E. Kalman. A new approach to linear filtering and prediction problems. *J. Basic Eng.-T. ASME*, 1960.

[28] E. Klann and Ramlau R. Regularization by fractional filter methods and data smoothing. *Inverse Problems*, 24(2):0125018, 2008.

[29] E. Kwiatkowski and J. Mandel. Convergence of the square root ensemble Kalman filter in the large ensemble limit. *SIAM/ASA J. Uncertain. Quantif.*, 3(1):1–17, 2015.

[30] K. J. H. Law and A. M. Stuart. Evaluating data assimilation algorithms. *Mon. Weather Rev.*, 140:3757–3782, 2012.

[31] K. J. H. Law, H. Tembine, and R. Tempone. Deterministic mean-field ensemble kalman filtering. *SIAM J. Sci. Comput.*, 38(3), 2016.

[32] F. Le Gland, V. Monbet, and V.-D. Tran. Large sample asymptotics for the ensemble Kalman filter. Research Report RR-7014, INRIA, 2009.

[33] M. Lemou. Multipole expansions for the Fokker-Planck equation. *Numer. Math.*, 78(4):597–618, 1998.

[34] C. Mouhot and L. Pareschi. Fast algorithms for computing the Boltzmann collision operator. *Math. Comput.*, 75:1833–1852, 2006.

[35] L. Pareschi and G. Russo. An introduction to Monte Carlo methods for the Boltzmann equation. In *CEMRACS 1999 (Orsay), ESAIM Proc.*, volume 10, Paris, 1999. Soc. Math. Appl. Indust.

[36] L. Pareschi and G. Toscani. *Interacting Multiagent Systems. Kinetic equations and Monte Carlo methods.* Oxford University Press, 2013.

[37] L. Pareschi, G. Toscani, and C. Villani. Spectral methods for the non cut-off Boltzmann equation and numerical grazing collision limit. *Numer. Math.*, 93(3):527–548, 2003.

[38] C. Schillings and A. M. Stuart. Analysis of the Ensamble Kalman Filter for Inverse Problems. *SIAM J. Numer. Anal.*, 55(3):1264–1290, 2017.

[39] C. Schillings and A. M. Stuart. Convergence analysis of ensemble Kalman inversion: the linear, noisy case. *Applicable Analysis*, 97(1):107–123, 2018.

[40] A. M. Stuart. Inverse problems: a Bayesian perspective. *Acta Numer.*, 19:451–559, 2010.

[41] G. Toscani. Kinetic models of opinion formation. *Commun. Math. Sci.*, 4(3):481–496, 2006.

[42] T. Trimborn, L. Pareschi, and M. Frank. Portfolio Optimization and Model Predictive Control: A Kinetic Approach. Preprint. arxiv:1711.03291, 2018.

[43] C. Villani. Conservative forms of Boltzmann's collision operator: Landau revisited. *ESAIM Math. Model. Numer. Anal.*, 33(1):209–227, 1999.

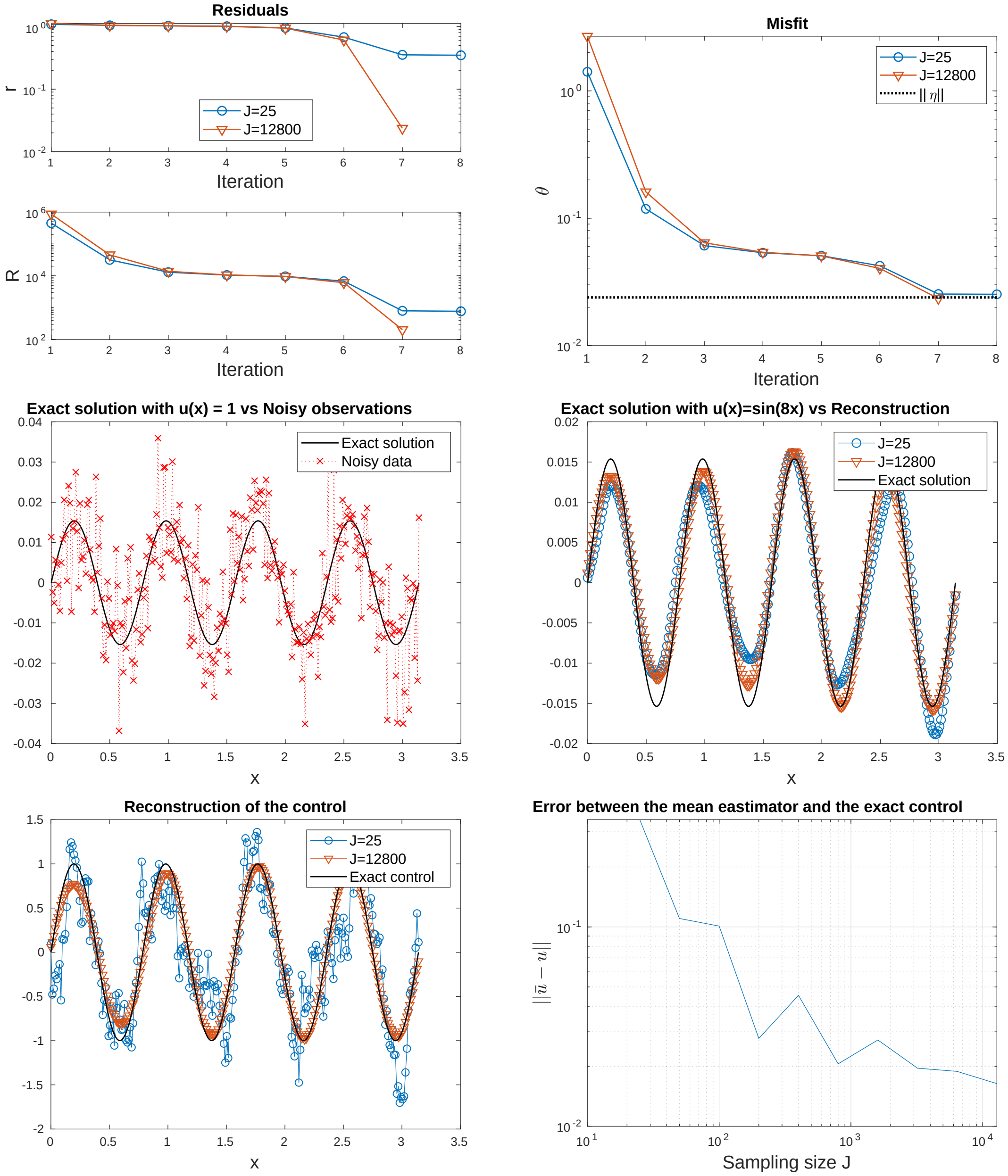

Figure 6: Elliptic problem - Test case 2 with $\gamma = 0.01$. Top row: plots of the residual r, the projected residual R and the misfit ϑ for $J = 25, 25 \cdot 2^9$. Middle row: plots of the noisy data and of the reconstruction of $p(x)$ at final iteration for $J = 25, 25 \cdot 2^9$. Bottom row: plots of the reconstruction of the control $u(x)$ at final iteration for $J = 25, 25 \cdot 2^9$ and behavior of the relative error $\frac{\|\bar{\mathbf{u}}-\mathbf{u}\|_2^2}{\|\bar{\mathbf{u}}\|_2^2}$.

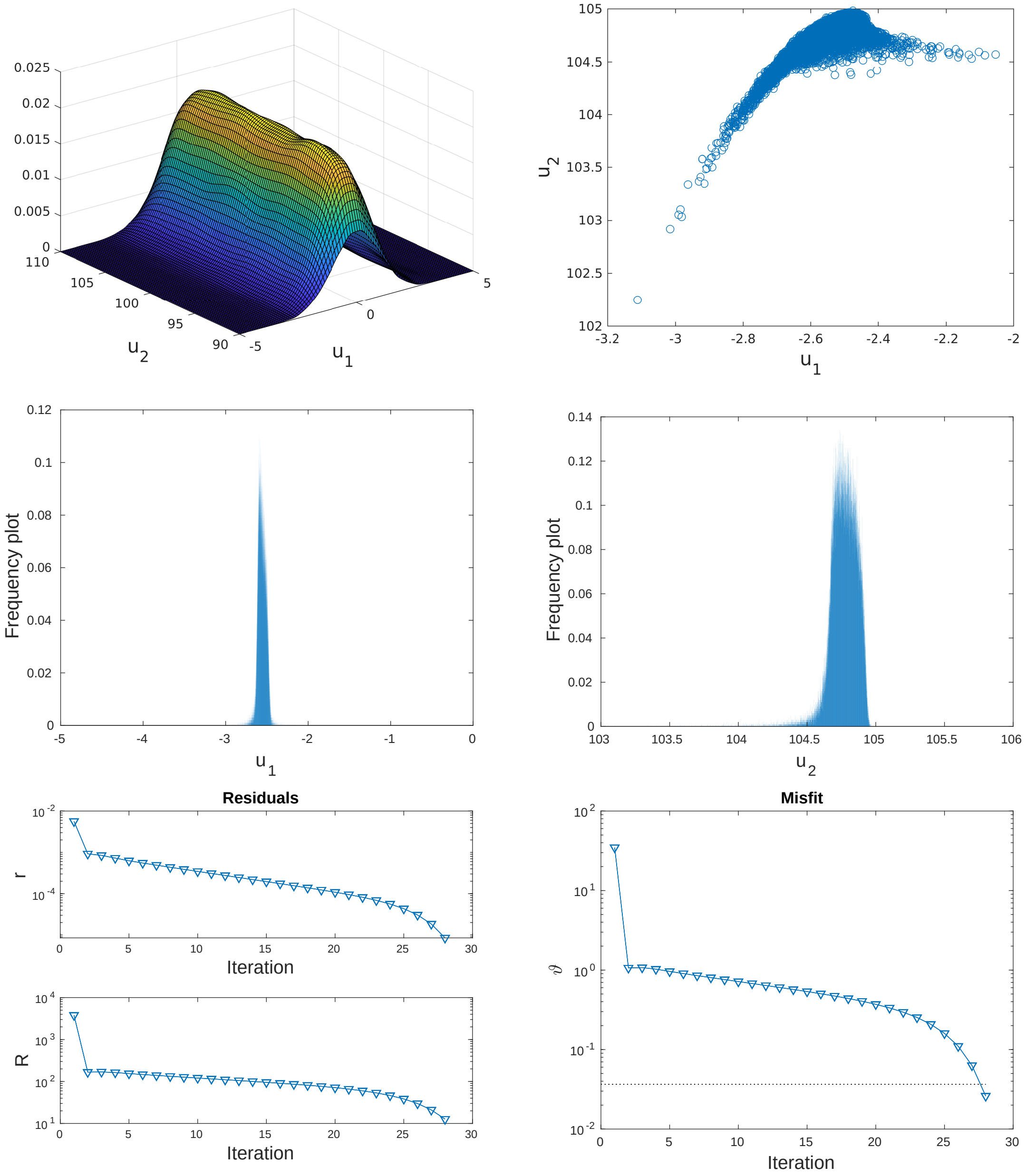

Figure 7: Nonlinear problem. Top row: plots of the density estimation of the initial samples (left) and position of the samples at final iteration (right). Middle row: Marginals of u_1 (left) and u_2 (right) as relative frequency plot. Bottom row: residual errors r and R (left) and misfit error (right).