sciendo

CAIM
ISSN 2038-0909

Commun. Appl. Ind. Math. 13 (1), 2022, 10–20 DOI: 10.2478/caim-2022-0002

Review Article

Recent Trends on Nonlinear Filtering for Inverse Problems

Michael Herty[1]*, Elisa Iacomini[1], Giuseppe Visconti[2]

[1]Institute for Geometry and Applied Mathematics, RWTH Aachen University, Templergraben 55, 52064 Aachen, Germany

[2]Department of Mathematics, Sapienza University of Rome, P.le Aldo Moro 5, 00185 Rome, Italy

*Email address for correspondence: herty@igpm.rwth-aachen.de

Communicated by Lorenzo Pareschi

Received on 03 26, 2022. Accepted on 05 10, 2022.

Abstract

Among the class of nonlinear particle filtering methods, the Ensemble Kalman Filter (EnKF) has gained recent attention for its use in solving inverse problems. We review the original method and discuss recent developments in particular in view of the limit for infinitely particles and extensions towards stability analysis and multi–objective optimization. We illustrate the performance of the method by using test inverse problems from the literature.

Keywords: Ensemble Kalman inversion, nonlinear filtering methods, inverse problems, multi-objective optimization, stability analysis

AMS subject classification: 65N21, 93E11, 35Q93, 37N35

1. Introduction

This review paper focuses on the Ensemble Kalman Filter applied to general inverse problems. In this context, some literature also uses the term Ensemble Kalman Inversion (EKI). Solving inverse problems or identification problems means determining parameters of a given model in order to obtain observable data. Due to the large range of applications, several approaches have been proposed in the literature to solve inverse problems. For instance, some well-known techniques rely on Bayesian formulation [1], but they can be extremely expensive. For this reason, efficient numerical schemes to solve the Bayesian inversion have been studied [2–4].

In this paper we are interested in solving inverse problems using a classical approach, i.e. relying on an optimization viewpoint, and in the numerical solution via the so-called particle methods. These can be divided into two classes of methods: the ones coming from particle swarm optimization, e.g. see [5] and the references therein, and the ensemble Kalman methods.

The EKI method belongs indeed to the class of particle methods and it is an iterative method for solving inverse problems. The method was originally introduced in [6] for unconstrained minimization problems, and recently extended also to the presence of different types of constraints [7–9]. The original EnKF has already been introduced more than ten years ago [10–13] as a discrete time method to estimate state variables and parameters of stochastic dynamical systems. The EKI method has become popular recently, because of the fact that it does not require derivatives of the underlying model for optimization but at the same time enjoys provable convergence results. Applications have been so far, in particular, in oceanography [14], reservoir modeling [15], weather forecasting [16], milling process [17], process control [18], geophysical applications [19–21], physics [22] and also machine learning [23–25]. The literature on Kalman filtering is very rich and we can not review this in detail here, but refer to the reference for further details. Our focus is on the reformulation of the EnKF for solving inverse problems as outlined below, in Section 1.2.

1.1. *Formulation of the ensemble Kalman inversion*

In order to present the mathematical formulation of the EKI method, we denote by $\mathcal{G} : X \to Y$ the given (nonlinear) forward operator between finite dimensional Hilbert spaces $X = \mathbb{R}^d$, $d \in \mathbb{N}$, and

$Y = \mathbb{R}^K$, $K \in \mathbb{N}$. Consider the inverse problem or parameter identification problem of the type

$$\text{find } \mathbf{u} \in X \text{ s.t. } \mathcal{G}(\mathbf{u}) = \mathbf{y} + \boldsymbol{\eta} \in Y. \tag{1}$$

Throughout the paper $\mathbf{u}$ is referred to as the (unknown) control, whereas $\mathbf{y}$ represents the data measurements (that are perturbed by noise $\boldsymbol{\eta}$). In applications one typically has $d \gg K$. The perturbations due to errors in the observations is modeled by $\boldsymbol{\eta}$ whose distribution is explicitly known. We assume that the noise is normally distributed with given covariance matrix $\boldsymbol{\Gamma}^{-1} \in \mathbb{R}^{K\times K}$, namely we write $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}^{-1})$.

In order to solve the inverse problem (1), the EKI considers a number J of particles or ensemble members whose state is determined by an iterative update. The ensemble members are modeled as realizations of the control $\mathbf{u} \in \mathbb{R}^d$, in the following combined in $\mathbf{U} = \left\{\mathbf{u}^j\right\}_{j=1}^J$, with $\mathbf{u}^j \in \mathbb{R}^d$, $j = 1, \dots, J$. The iteration index is denoted by n and the collection of the ensemble members by $\mathbf{U}^n = \{\mathbf{u}^{j,n}\}_{j=1}^J$, $\forall\, n \geq 0$.

Then, at iteration $n+1$ the EKI update is given by

$$\mathbf{u}^{j,n+1} = \mathbf{u}^{j,n} + \mathbf{C}_{\mathcal{G}}(\mathbf{U}^n) \left(\mathbf{D}_{\mathcal{G}}(\mathbf{U}^n) + \frac{1}{\Delta t}\boldsymbol{\Gamma}^{-1} \right)^{-1} (\mathbf{y} - \mathcal{G}(\mathbf{u}^{j,n})) \tag{2}$$

for each $j = 1, \dots, J$, where $\Delta t \in \mathbb{R}^+$ is a parameter and where the ensemble update (2) depends on covariance matrices:

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

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

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

Then, it can be proven [6] that

$$\lim_n \overline{\mathbf{u}}^n = \arg\min_{\mathbf{u}} \frac{1}{2} \left\| \boldsymbol{\Gamma}^{\frac{1}{2}} (\mathbf{y} - \mathcal{G}(\mathbf{u})) \right\|_Y^2. \tag{3}$$

It is worth to mention that in the original formulation each observation or measurement is perturbed by additional additive noise at each iteration. The EKI satisfies the subspace property [6], i.e., the ensemble iterates stay in the subspace spanned by the initial ensemble. As consequence, the natural estimator for the solution of the inverse problem is provided by the mean of the ensemble.

In recent years, the EKI was also studied as technique to solve inverse problems in a Bayesian framework. For instance see the works [26,27] and the references therein. The analysis of the method is proven to have a comparable accuracy with traditional least–squares approaches to inverse problems [6]. The method approximates a specific Bayes linear estimators and it is able to provide an approximation of the posterior measure. For a detailed discussion we refer to [28,29]. In this work, we keep the attention on the classical approach which aims to solve the inverse problem through an optimization point–of–view, see (3). Additional properties of the EKI method, continuous–time limits [30–34], i.e., $n \to \infty$ and mean–field limits on the number of the ensemble members [27,35–37], i.e., $J \to \infty$ have been recently been developed and will be reviewed in more detail below.

1.2. *Structure of the paper*

The remainder of this paper is organized as follows. In Section 2 we review the continuous formulations of the EKI method which lead to a preconditioned gradient descent system and to a Vlasov–type partial differential equation. In Section 3 and in Section 4, instead, we present two new formulations of the EKI method for multi–objective inverse problems and for globally asymptotically convergence to the target solution, respectively. Finally, we draw conclusions and perspectives in Section 5.

2. Continuous limits of the ensemble Kalman inversion

The continuous in time limit reduces the discrete update to a coupled system of ordinary differential equations. This limit has been performed in different recent publications, starting from [33] to more recent formulations, e.g. see [38] for the hierarchical EKI. In particular, in [33] it has been shown that continuous in time limit results, in case of a linear forward model $\mathcal{G}$, to a gradient flow structure. This gradient flow provides a solution to the inverse problem (1) by minimizing the least–squares functional

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

Observe, however, that in the continuous limit [33] the regularization term originally present in (3) vanishes for certain scalings. Although typically the analysis of the continuous in time EKI focuses on linear forward models, there are recent results on the EKI formulations in nonlinear settings [39].

2.1. *Continuous–time limit*

The continuous–time limit was firstly proposed in [33]: consider the parameter Δt as an artificial time step for the discrete iteration, i.e. $\Delta t \sim N_t^{-1}$ with N_t being the maximum number of iterations and define $\mathbf{U}^n \approx \mathbf{U}(n\Delta t) = \left\{\mathbf{u}^j(n\Delta t)\right\}_{j=1}^J$ for $n \geq 0$. Computing the limit $\Delta t \to 0^+$ one obtains

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{u}^j = \mathbf{C}_{\mathcal{G}}(\mathbf{U})\boldsymbol{\Gamma}\left(\mathbf{y}-\mathcal{G}(\mathbf{u}^j)\right), \quad j=1,\dots,J \\ &\mathbf{C}_{\mathcal{G}}(\mathbf{U}) = \frac{1}{J}\sum_{k=1}^J \left(\mathbf{u}^k-\overline{\mathbf{u}}\right)\left(\mathcal{G}(\mathbf{u}^k)-\overline{\mathcal{G}}\right)^T \end{aligned} \tag{4}$$

with initial condition $\mathbf{U}(0) = \mathbf{U}^0$. Note that within this limit the noise is scaled with $\frac{1}{\Delta t}$ which allows for the continuous time limit. Further, the term $\mathbf{D}_{\mathcal{G}}$ vanishes leading to possibly unstable dynamics [37,40].

However, in the case of $\mathcal{G}$ linear, i.e., $\mathcal{G}(\mathbf{u}) = \mathbf{G}\mathbf{u}$, with $\mathbf{G} \in \mathbb{R}^{K\times d}$, the (4) can be reformulated in terms of the gradient $\nabla\Phi$ as a gradient flow:

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{u}^j = -\mathbf{C}(\mathbf{U})\nabla_{\mathbf{u}}\Phi(\mathbf{u}^j,\mathbf{y}), \quad j=1,\dots,J \\ &\mathbf{C}(\mathbf{U}) = \frac{1}{J}\sum_{k=1}^J(\mathbf{u}^k-\overline{\mathbf{u}})(\mathbf{u}^k-\overline{\mathbf{u}})^T. \end{aligned} \tag{5}$$

Since $\mathbf{C}(\mathbf{U})$ is positive semi–definite we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi(\mathbf{u}(t),\mathbf{y}) = \frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}\left\|\boldsymbol{\Gamma}^{\frac{1}{2}}\left(\mathbf{y}-\mathbf{G}\mathbf{u}\right)\right\|^2 \leq 0. \tag{6}$$

Although the forward operator is assumed to be linear, the gradient flow is nonlinear. For further details and properties of the gradient descent equation (5) we refer to [33,34]. In particular, we emphasize that the subspace property of the EKI also holds for the continuous dynamics and the following important result on the velocity of the collapse of the ensembles towards their mean in the large time limit, cf. Theorem 3 in [31,33]:

$$\left\|\Gamma^{\frac{1}{2}}G(\mathbf{u}^j(t)-\overline{\mathbf{u}}(t))\right\| = O(Jt^{-1}).$$

2.2. *Mean–field limit*

By definition, the EKI method considers a finite ensemble size $J < \infty$. The behavior of the method in the limit of infinitely many ensembles can be studied via mean–field limit in analogy with the classical mean–field derivation of multi–agent systems [41–44]. In the case of a linear foward model, the limit leads to a Vlasov–type gradient flow PDE.

$$\partial_t f(t,\mathbf{u}) - \nabla_{\mathbf{u}} \cdot (\mathbf{C}(f)\nabla_{\mathbf{u}}\Phi(\mathbf{u},\mathbf{y})f(t,\mathbf{u})) = 0, \ f(t,0) = f_0(\mathbf{u}) \tag{7}$$

for a compactly supported on $\mathbb{R}^d$ probability density f of $\mathbf{u}$ at time t denoted by

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

The initial probability density distribution is denoted by f_0. The operator $\mathbf{C}(f)$ is the mean–field limit of the covariance of the ensemble and can be written in terms of moments of f as

$$\mathbf{C}(f) = \mathbf{E}(t) - \mathbf{m}(t)\mathbf{m}^T(t) \geq 0,$$

where $\mathbf{m} \in \mathbb{R}^d$ and $\mathbf{E} \in \mathbb{R}^{d\times d}$ are defined, respectively, as

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

For the rigorous mean–field derivation and analysis of the EKI we refer to [35,36]. Equation (7) is a nonlinear transport equation arising from non–linear gradient flow interactions and in [35,40] it is observed that the counterpart of (6) holds at the kinetic level. In fact, for

$$\mathcal{L}(f,\mathbf{y}) = \int_{\mathbb{R}^d} \Phi(\mathbf{u},\mathbf{y}) f(t,\mathbf{u})\mathrm{d}\mathbf{u},$$

we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{L}(f,\mathbf{y}) = \int_{\mathbb{R}^d} \Phi(\mathbf{u},\mathbf{y})\partial_t f(t,\mathbf{u})\mathrm{d}\mathbf{u} = -\int_{\mathbb{R}^d} (\nabla_{\mathbf{u}}\Phi(\mathbf{u},\mathbf{y}))^T \mathbf{C}(f)\nabla_{\mathbf{u}}\Phi(\mathbf{u},\mathbf{y})\mathrm{d}\mathbf{u} \leq 0,$$

since $\mathbf{C}(f)$ is positive semi-definite. In particular, $\mathcal{L}(f,\mathbf{y})$ is strictly decreasing unless f is a Dirac measure. Also, for $f(\mathbf{u}) = \delta(\mathbf{u}-\mathbf{u}^*)$ for $\mathbf{u}^* = \arg\min_{\mathbf{u}\in\mathbb{R}^d} \Phi(\mathbf{u},\mathbf{y})$ provides a steady solution of the continuous–limit formulation, but the converse is not necessarily true. In fact, all Dirac distributions, satisfy $\mathbf{C}(f) = 0$ and hence provide steady solutions of (7). Convergence to the distribution $f(\mathbf{u}) = \delta(\mathbf{u}-\mathbf{u}^*)$ has been proven to be linear in [35]: $\|\mathbf{C}(f)\| = O(t^{-1})$.

The mean–field interpretation of the EKI has allowed to design computationally efficient methods based on the mean–field formulation [37,45]. In particular, it is possible to use a large number of particles which guarantees significantly better reconstructions of the unknown control, cf. Section 5 in [37].

3. Multi–objective ensemble Kalman inversion

The EKI can also be extended to treat also multi-objective optimization problems within a weighted function approach. Here, a vector of controls has to be determined for competitive models $\mathcal{G}_i$ for $i = 1,\dots,l$ and given observational data:

$$\mathbf{y}_i = \mathcal{G}_i(\mathbf{u}) + \boldsymbol{\eta}_i \quad i = 1,\dots,l \tag{8}$$

for l models $\mathcal{G}_i : X \to Y$ and l observations $\mathbf{y}_1,\dots,\mathbf{y}_l \in Y$, where $\boldsymbol{\eta}_i$ is observational noise. A solution to (8) can be obtained e.g. using a multi–objective optimization [46–48]:

$$\min_{\mathbf{u}\in X} \left(\|\Gamma^{\frac{1}{2}} (\mathbf{y}_1 - \mathcal{G}_1(\mathbf{u}))\|, \dots, \|\Gamma^{\frac{1}{2}} (\mathbf{y}_l - \mathcal{G}_l(\mathbf{u}))\| \right). \tag{9}$$

Solution in this framework is related to the notion of Pareto optimality [48] that defines a concept of minimum for the vector–valued optimization problem (9).

Definition 3.1. A point $\mathbf{u}^* \in \mathbb{R}^d$ is called Pareto optimal if and only if there exists no point $\mathbf{u} \in \mathbb{R}^d$ such that $\mathcal{G}_i(\mathbf{u}) \leq \mathcal{G}_i(\mathbf{u}^*)$ for all $i = 1, 2, \ldots, l$ and $\mathcal{G}_j(\mathbf{u}) \leq \mathcal{G}_j(\mathbf{u}^*)$ for at least one $j \in \{1, 2, \ldots, l\}$.

The set $\mathcal{S}_U$ of all $\mathbf{u}^*$ fulfilling Definition 3.1 is called Pareto set, while its representation in the space of objectives $\mathcal{S}_G := \{(\mathbf{y}_i - \mathcal{G}_i(\mathbf{u}))_{i=1}^{l} : \mathbf{u} \in \mathcal{S}\}$ is called Pareto front. An approximation of $\mathcal{S}_G$ can be recovered following an approach based on weighted function method [47]. Let $\mathbf{1} = (1, \ldots, 1)^T$ and let $\lambda \in \Lambda$ be a fixed vector in the set

$$\Lambda := \{\lambda \in \mathbb{R}^l_+ : \lambda \cdot \mathbf{1} = 1\}.$$

Define the weighted objective function and the weighted observations as

$$\mathcal{G}(\mathbf{u}, \lambda) := \sum_{i=1}^{l} \lambda_i \mathcal{G}_i(\mathbf{u}) : X \times \Lambda \to Y \text{ Â and Â } \mathbf{y} = \sum_{i=1}^{l} \lambda_i \mathbf{y}_i.$$

An approximation to the Pareto front $\mathcal{S}_U$ is then obtained by $P := \{\mathbf{u}^*(\lambda) : \lambda \in \Lambda\}$, where for each λ

$$\mathbf{u}^*(\lambda) = \arg\min_{\mathbf{u} \in X} \Phi(\mathbf{u}, \lambda), \quad \Phi(\mathbf{u}, \mathbf{y}, \lambda) = \frac{1}{2} \left\| \Gamma^{\frac{1}{2}} \sum_{i=1}^{l} \lambda_i \left(\mathbf{y}_i - \mathcal{G}_i(\mathbf{u})\right) \right\|^2 \quad \forall \lambda \in \Lambda. \tag{10}$$

In case of a convex problem, $S_U = P$, see [47, Theorem 3.1.4]. In theory the previous problem (10) needs to be solved for all $\lambda \in \Lambda$.

Using a mean–field approach as in the last section, allows for an analysis on the dependence of $u^*(\lambda)$ on λ which in turn is used as adaptive grid on Λ. The evolution of the formal sensitivity of the mean-field description f of the particle distribution with respect to λ_i is given by

$$\begin{aligned} 0 = \partial_t \partial_{\lambda_i} f(t, \mathbf{u}, \lambda) - \nabla_{\mathbf{u}} \Big(& \partial_{\lambda_i} \mathbf{C}(f) \nabla_{\mathbf{u}} \Phi(\mathbf{u}, \mathbf{y}, \lambda) f(t, \mathbf{u}, \lambda) + \\ & \mathbf{C}(f)\ \partial_{\lambda_i} (\nabla_{\mathbf{u}} \Phi(\mathbf{u}, \mathbf{y}, \lambda)) f(t, \mathbf{u}, \lambda) + \mathbf{C}(f)\ \nabla_{\mathbf{u}} \Phi(\mathbf{u}, \mathbf{y}, \lambda) \partial_{\lambda_i} f(t, \mathbf{u}, \lambda) \Big), \end{aligned} \tag{11}$$

for zero initial data. The set of equations (11) for $i = 1, \ldots, l$ is defined on the extended phase space $\mathbb{R} \times X \times \Lambda$ and therefore computationally infeasible. However, the Pareto set is given as moment of f where the first moment $\mathbf{m}$ depends additionally on λ :

$$P(t) = \left\{ \mathbf{m}(\lambda, t) := \int_X \mathbf{u} \, \mathrm{d}f(t, \mathbf{u}, \lambda) : \ \lambda \in \Lambda \right\}.$$

Similarly, moments of (11) can be defined leading to a set of ordinary differential equations for $\nabla_\lambda \mathbf{m}(\lambda, t)$ [49, Lemma 2.3]. This in turn allows to define an adaptive grid $\{\lambda^k\, k = 1, \ldots\} \subset \Lambda$: Let for a fixed $\overline{\lambda} = \lambda^{k-1}$ the corresponding optimal parameter be approximated by $\mathbf{m}(\overline{\lambda}, T)$ for some T fixed and sufficiently large. Then, consider the following Taylor expansion

$$\mathbf{m}(\overline{\lambda} + \Delta\lambda, T) \approx \mathbf{m}(\overline{\lambda}, T) + \Delta\lambda \cdot \nabla \mathbf{m}(\overline{\lambda}, T), \text{ and Â } \Delta\lambda = \lambda^{k+1} - \overline{\lambda}. \tag{12}$$

Reformulating (12) allows to obtain $\lambda^{k+1} \in \Lambda$ adaptively based on the approximation on the Pareto set $P(t)$. It also yields an estimate on the norm of the update $\Delta\lambda$ on an approximation of S_U with given tolerance $\delta > 0$ by $\|\Delta\lambda\| \|\nabla \mathbf{m}(\lambda^k)\| \leq \delta$.

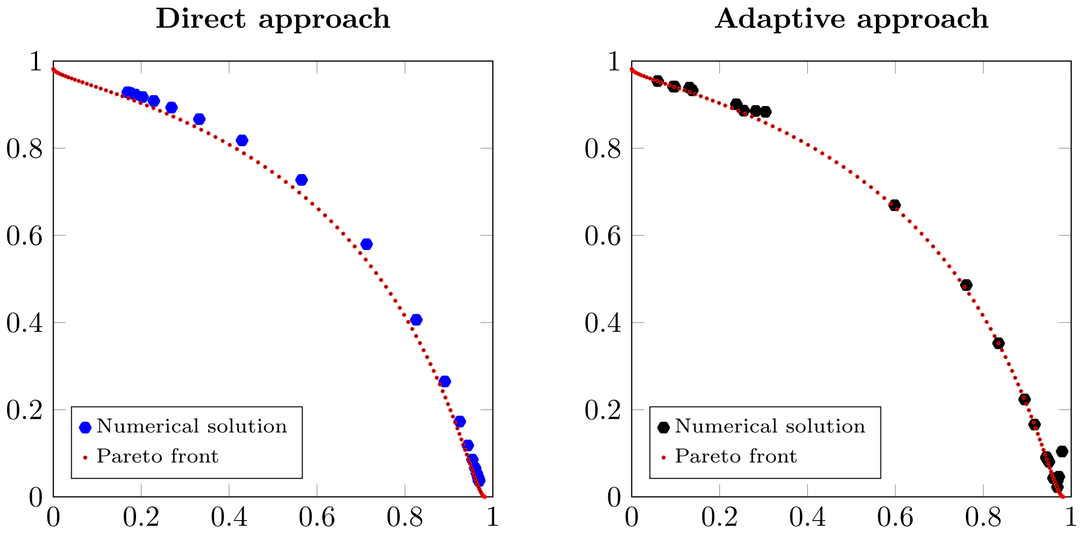

Figure 1. Approximation of the Pareto front with the direct approach (left) and the adaptive approach (right).

3.1. *Numerical experiment*

In the numerical experiment we show that the adaptive strategy leads to results that approximate the Pareto front S_G very well with only a few discretization points $\lambda^k, k = 1, \ldots, K$. We set $l = 2$ so that Λ is parameterized by a single parameter $\lambda \in [0, 1]$, i.e. $\mathcal{G} = \lambda\mathcal{G}_1 + (1 - \lambda)\mathcal{G}_2$. Then, we consider two non convex functions $\mathcal{G}_1, \mathcal{G}_2 : \mathbb{R}^2 \to \mathbb{R}^2$ as in [50]

$$\mathcal{G}_1(u_1, u_2) = 1 - e^{-\left(u_1-\frac{1}{\sqrt{2}}\right)^2-\left(u_2-\frac{1}{\sqrt{2}}\right)^2}, \quad \mathcal{G}_2(u_1, u_2) = 1 - e^{-\left(u_1+\frac{1}{\sqrt{2}}\right)^2-\left(u_2+\frac{1}{\sqrt{2}}\right)^2},$$

and $\mathbf{y}_i = 0$, for $i = 1, 2$. As further parameters we use $J = 25$ particles sampled from the uniform distribution $U_0 \sim \mathcal{U}([-2, 2]^2)$, the tolerance is set $\delta = 5 \cdot 10^{-3}$, $T_{fin} = 10$, $\Gamma = \mathbb{1}$ and $K = 22$.

Even so, the theoretical results have been proven in the linear case [49, Sec. 3], they are applied here in a nonlinear framework. We compare a naive choice for the discretization of Λ using an equidistant grid (direct approach) with the outlined adaptive strategy.

We observe that the solution obtained with the adaptive approach covers a larger part of the Pareto front, showing additionally a relatively sharper resolution compared with the direct approach, see Figure 1.

Moreover, the approximation of the Pareto set in Figure 2 shows the expected behavior. Here, the adaptive strategy yields a cloud of points relatively close to the (analytically known) Pareto set S_U compared with the direct approach.

4. Stabilized continuous limit of the ensemble Kalman inversion

In the continuous–time limit the term $\mathbf{D}_{\mathcal{G}}$ present in the discrete formulation vanish due to scaling. This consideration inspired [40], where a stability analysis of the moment system of the time–continuous EKI (5) is performed. Therein, it has been established that the system has infinitely many non–hyperbolic Bogdanov–Takens equilibria leading to several undesirable consequences. The latter are structurally unstable, i.e., sensitive to small perturbations. Since those equilibria lie on the set where the preconditioner $\mathbf{C}$ collapse to zero, low order of convergence in time holds true. Further, numerical approximations may push the trajectory in the unfeasible region of the phase space or get the method stuck in the wrong equilibrium.

These considerations led to a modified formulation of the method is globally asymptotically stable by introducing the regularization term R to the dynamics. More precisely, given $\mathbf{\Sigma} \in \mathbb{R}^{d\times d}$ symmetric and

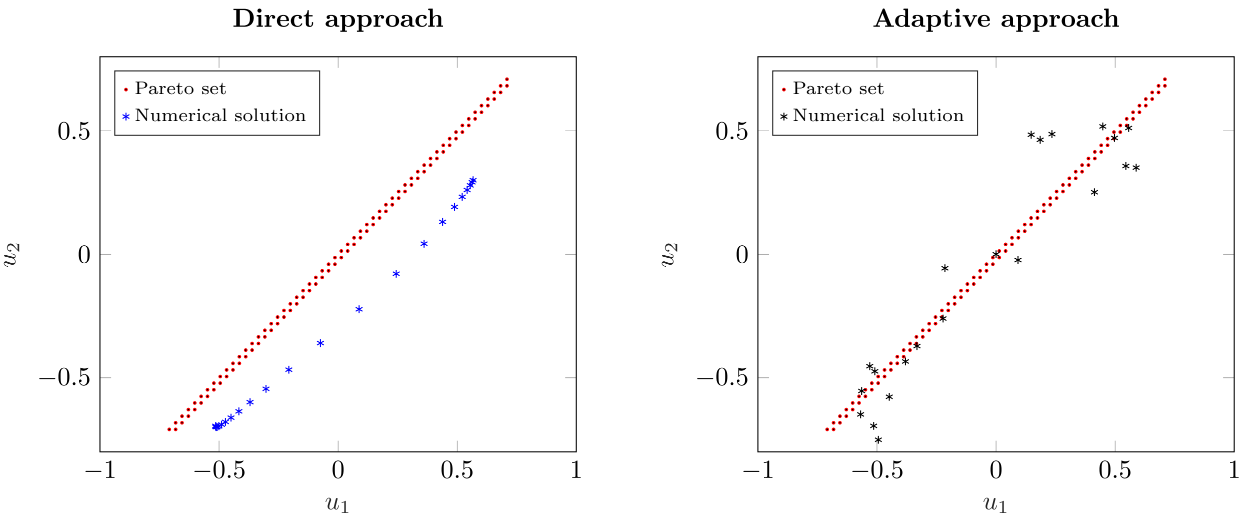

Figure 2. Approximation of the Pareto set with the direct approach (left) and the adaptive approach (right).

full–rank, in particular positive definite, in [40] it is proposed to consider the following general discrete dynamics for each ensemble member $j = 1, \dots, J$ in the case of a linear model:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{u}^j &= -\tilde{\mathbf{C}}(\mathbf{U})\nabla_{\mathbf{u}}\Phi(\mathbf{u}^j, \mathbf{y}) + R(\mathbf{U}),\\ R(\mathbf{U}) = \beta\tilde{\mathbf{C}}(\mathbf{U})(\mathbf{u}^j - \bar{\mathbf{u}}),& \quad \tilde{\mathbf{C}}(\mathbf{U}) = \mathbf{C}(\mathbf{U}) + (1-\alpha)\mathbf{\Sigma},\end{aligned} \tag{13}$$

with parameters $\alpha, \beta \in \mathbb{R}$. The choices $\alpha = 1$ and $\beta = 0$ yield the continuous–time limit (5) for the original EKI. The modified dynamics (13) differs from (5) in the formulation of the preconditioner $\tilde{\mathbf{C}}(\mathbf{U})$ and in the presence of the additive term $R(\mathbf{U})$. The new preconditioner is related to an inflation of the covariance $\mathbf{C}(\mathbf{U})$ for $\alpha < 0$. This modification allows to stabilize the phase space of the moments. The term $R(\mathbf{U})$, instead, has been shown to be an acceleration term for the convergence towards equilibrium. The modified dynamical system (13) has also a mean–field interpretation:

$$\partial_t f(t, \mathbf{u}) - \nabla_{\mathbf{u}} \cdot \Big(\tilde{\mathbf{C}}(f)\left(\nabla_{\mathbf{u}}\Phi(\mathbf{u}, \mathbf{y}) - \beta(\mathbf{u} - \mathbf{m})\right) f(t, \mathbf{u})\Big) = 0, \tag{14}$$

where $\tilde{\mathbf{C}}(f)$ is the mean–field of $\tilde{\mathbf{C}}(\mathbf{U})$ leading to $\tilde{\mathbf{C}}(f) = \mathbf{E}(t) - \mathbf{m}(t)\mathbf{m}^T(t) + (1-\alpha)\mathbf{\Sigma}$.

The stability analysis of the moment equations is performed in the simplified setting where $K = d$ and $\mathbf{\Gamma}$, $\mathbf{G}$ are identity matrices. The $d + d^2$ dynamical system of the moments of (14) is then

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{m}(t) &= \mathbf{C}(\mathbf{y} - \mathbf{m}) + (1-\alpha)\mathbf{\Sigma}(\mathbf{y} - \mathbf{m})\\ \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{C}(t) &= -2\mathbf{C}\mathbf{C} - (1-\alpha)\mathbf{\Sigma}\mathbf{C} - (1-\alpha)\mathbf{C}\mathbf{\Sigma},\end{aligned}$$

and its linearization $(\delta\mathbf{m}, \delta\mathbf{E})$ at target equilibrium $(\mathbf{m}^*, \mathbf{C}^*) = (\mathbf{y}, \mathbf{0})$ fulfills

$$\frac{\mathrm{d}}{\mathrm{d}t}\delta\mathbf{m}(t) = -(1-\alpha)\mathbf{\Sigma}\delta\mathbf{m}, \quad \frac{\mathrm{d}}{\mathrm{d}t}\delta\mathbf{C}(t) = -4(1-\alpha)\mathbf{\Sigma}\delta\mathbf{C}.$$

For $\alpha < 1$ and $\mathbf{\Sigma}$ positive definite the target equilibrium is hyperbolic. This formal presentation of the role of the parameters is mathematically rigorous in [40], where also exponentially fast convergence to the target equilibrium is proven.

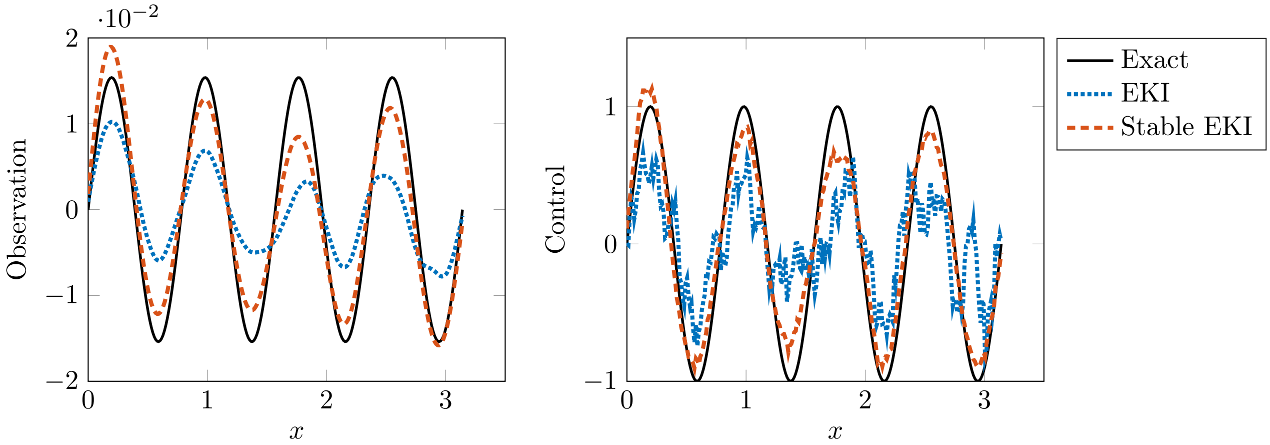

Figure 3. Reconstruction of the observation (left) and reconstruction of the control (right) for the continuous–time limit of the original EKI method and for the stabilized formulation

.

4.1. *Numerical experiments*

We consider the inverse problem of identifying the force function $u(x)$ of a linear elliptic equation in one spatial dimension assuming that noisy observation of the solution to the problem are available, e.g. see [6,33,37].

The problem is prescribed by the following one dimensional elliptic PDE for p

$$-\frac{\mathrm{d}^2}{\mathrm{d}x^2}p(x) + p(x) = u(x), \quad x \in [0,\pi] \tag{15}$$

subject to boundary conditions $p(0) = p(\pi) = 0$. To obtain measurement data we use the continuous control $u(x) = \sin(8x)$. The problem is discretized using a uniform mesh with $d = K = 2^8$ equidistant points $\{x_i\}$ on the interval $[0,\pi]$. Denote by $\mathbf{u}_i^\dagger = u(x_i)$ the evaluations of the control function $u(x)$ on the mesh $i = 0, 1 \dots, d$. Noisy observations $\mathbf{y} \in \mathbb{R}^K$ are obtained by

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

where $\mathbf{G} \in \mathbb{R}^{K\times d}$ is a first–order finite difference discretization of the PDE (15). For simplicity we assume that $\boldsymbol{\eta}$ is a Gaussian white noise, i.e., $\boldsymbol{\eta} \sim \mathcal{N}(0,\gamma^2\mathbf{I})$ with $\gamma \in \mathbb{R}^+$ and $\mathbf{I} \in \mathbb{R}^{d\times d}$ is the identity matrix. We are interested in recovering an approximation to the discrete control $\mathbf{u}^\dagger \in \mathbb{R}^d$ from the noisy observations $\mathbf{y} \in \mathbb{R}^K$.

In Figure 3 we show the solution to this problem provided by the time–continuous limit of the original EKI and the stabilized formulation proposed in [40]. Both method use $J = 20$ ensemble members and a noise level of $\gamma = 0.01$. We observe that the stable EKI produces a qualitatively improved reconstruction of the control and observation compared to the classical EKI. Moreover, as expected by the analysis, we observe that the stable EKI converges faster than the classical method, see Figure 4.

5. Conclusions

An overview on the EKI and its current developments has been provided. The analytical properties have been investigated and, in particular, the mean–field equation and its corresponding moment system has been presented. Two recent extensions of the EKI has been shown and discussed towards coupled inverse problems and towards numerically stable formulations. Further developments may involve a mixture between the two novelties presented and since many physical problems are subject to additional

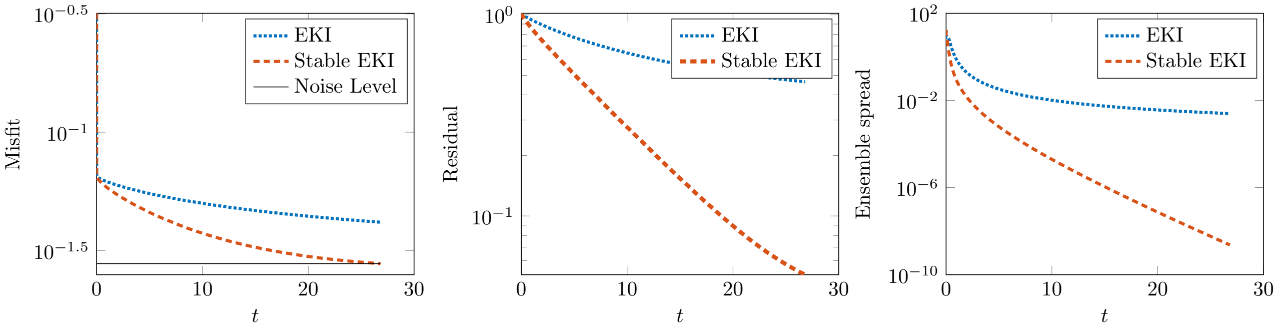

Figure 4. Behavior of the misfit, of the residual and of the ensemble spread around the mean in time, both for the continuous–time limit of the original EKI method and for the stabilized formulation in [40].

parameteric uncertainty, a suitable treatment of the then stochastic EKI might be of further interest. In case of large parameter spaces $X = \mathbb{R}^d$ with $d \gg 1$ computational issues need to be addressed as well, since e.g. $\mathbf{C}_{\mathcal{G}}$ grows quadratic in d. Furthermore, the outlined approach of time-continuous and mean-field limit is applicable to wider range of particle methods and might serve as a starting point for future investigation into nonlinear filtering from a mathematical perspective.

Acknowledgments

The authors thank the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for the financial support through 20021702/GRK2326, 333849990/IRTG-2379, HE5386/15,18-1,19-1,22-1,23-1 and under GermanyÂ's Excellence Strategy EXC-2023 Internet of Production 390621612. The funding through HIDSS-004 is acknowledged.

G.V. is member of the "National Group for Scientific Computation (GNCS-INDAM)" and acknowledges support by MUR (Ministry of University and Research) PRIN2017 project number 2017KKJP4X.

References

1. M. Dashti and A. M. Stuart, *The Bayesian Approach to Inverse Problems*, pp. 311–424. Springer International Publishing, 2016.
2. J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*. Springer, 2nd ed., 1985.
3. M. Burger and F. Lucka, Maximum a posteriori estimates in linear inverse problems with log-concave priors are proper Bayes estimators, *Inverse Problems*, vol. 30, p. 114004, 2014.
4. H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of inverse problems*, vol. 375. Springer Science and Business Media, 1996.
5. J. Carrillo, F. Hoffmann, A. Stuart, and U. Vaes, Consensus-based sampling, *Studies in Applied Mathematics*, vol. 148, no. 3, pp. 1069–1140, 2022.
6. M. Iglesias, K. Law, and A. M. Stuart, Ensemble Kalman methods for inverse problems, *Inverse Probl.*, vol. 29, no. 4, p. 045001, 2013.
7. N. K. Chada, C. Schillings, and S. Weissmann, On the incorporation of box-constraints for ensemble Kalman inversion, *Foundations of Data Science*, vol. 1, no. 2639-8001_2019_4_433, p. 433, 2019.
8. M. Herty and G. Visconti, Continuous limits for constrained ensemble Kalman filter, *Inverse Probl.*, 2020.
9. D. J. Albers, P.-A. Blancquart, M. E. Levine, E. E. Seylabi, and A. M. Stuart, Ensemble Kalman methods with constraints, *Inverse Probl.*, vol. 35, no. 9, p. 095007, 2019.
10. K. Bergemann and S. Reich, An ensemble Kalman-Bucy filter for continuous data assimilation, *Meteorologische Zeitschrift*, vol. 21, no. 3, pp. 213–219, 2012.

11. Y. Chen and D. S. Oliver, Parameterization techniques to improve mass conservation and data assimilation for ensemble Kalman filter, 2010.

12. A. A. Emerick and A. C. Reynolds, Ensemble smoother with multiple data assimilation, *Computers and Geosciences*, vol. 55, pp. 3–15, 2013.

13. G. Evensen, Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res*, vol. 99, pp. 10143–10162, 1994.

14. G. Evensen and P. J. Van Leeuwen, Assimilation of geosat altimeter data for the agulhas current using the ensemble Kalman filter with a quasi-geostrophic model, *Monthly Weather*, vol. 128, pp. 85–96, 1996.

15. S. I. Aanonsen, G. Naevdal, D. S. Oliver, A. C. Reynolds, and B. Valles, The ensemble Kalman filter in reservoir engineering–a review, *SPE J.*, vol. 14, no. 3, pp. 393–412, 2009.

16. T. Janjìc, D. McLaughlin, S. E. Cohn, and M. Verlaan, Conservation of mass and preservation of positivity with ensemble-type Kalman filter algorithms, *Monthly Weather Review*, vol. 142, no. 2, pp. 755–773, 2014.

17. M. Schwenzer, G. Visconti, M. Ay, T. Bergs, M. Herty, and D. Abel, Identifying trending coefficients with an ensemble Kalman filter, *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 2292–2298, 2020.

18. B. O. S. Teixeira, L. A. B. Târres, L. A. Aguirre, and D. S. Bernstein, On unscented Kalman filtering with state interval constraints, *J. Process Contr.*, vol. 20, no. 1, pp. 45–57, 2010.

19. J. Keller, H.-J. Franssen, and W. Nowak, Investigating the pilot point ensemble kalman filter for geostatistical inversion and data assimilation, *Adv. Water Resour.*, vol. 155, 2021.

20. J. B. Muir and V. C. Tsai, Geometric and level set tomography using ensemble Kalman inversion, *Geophysical Journal International*, vol. 220, no. 2, pp. 967–980, 2019.

21. C.-H. M. Tso, M. Iglesias, P. Wilkinson, O. Kuras, J. Chambers, and A. Binley, Efficient multiscale imaging of subsurface resistivity with uncertainty quantification using ensemble Kalman inversion, *Geophysical Journal International*, vol. 225, no. 2, pp. 887–905, 2021.

22. Z. Li, An iterative ensemble kalman method for an inverse scattering problem in acoustics, *Modern Physics Letters B*, vol. 34, no. 28, p. 2050312, 2020.

23. E. Haber, F. Lucka, and L. Ruthotto, Never look back - A modified EnKF method and its application to the training of neural networks without back propagation. Preprint arXiv:1805.08034, 2018.

24. N. B. Kovachki and A. M. Stuart, Ensemble Kalman inversion: a derivative-free technique for machine learning tasks, *Inverse Probl.*, vol. 35, no. 9, p. 095005, 2019.

25. A. Yegenoglu, S. Diaz, K. Krajsek, and M. Herty, Ensemble Kalman filter optimizing deep neural networks, in *Conference on Machine Learning, Optimization and Data Science*, vol. 12514, 2020.

26. 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.*, vol. 3, no. 1, pp. 823–851, 2015.

27. A. Garbuno-Inigo, F. Hoffmann, W. Li, and A. M. Stuart, Interacting Langevin Diffusions: Gradient Structure and Ensemble Kalman Sampler, *SIAM J. Appl. Dyn. Syst.*, vol. 19, no. 1, pp. 412–441, 2020.

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

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

30. D. Bloemker, C. Schillings, and P. Wacker, A strongly convergent numerical scheme from ensemble Kalman inversion, *SIAM J. Numer. Anal.*, vol. 56, no. 4, pp. 2537–2562, 2018.

31. D. Bloemker, C. Schillings, P. Wacker, and S. Weissman, Well Posedness and Convergence Analysis of the Ensemble Kalman Inversion, *Inverse Probl.*, vol. 35, no. 8, 2019.

32. N. K. Chada, A. M. Stuart, and X. T. Tong, Tikhonov regularization within ensemble Kalman inversion, *SIAM J. Numer. Anal.*, vol. 58, no. 2, pp. 1263–1294, 2020.

33. C. Schillings and A. M. Stuart, Analysis of the Ensamble Kalman Filter for Inverse Problems, *SIAM J. Numer. Anal.*, vol. 55, no. 3, pp. 1264–1290, 2017.

34. C. Schillings and A. M. Stuart, Convergence analysis of ensemble Kalman inversion: the linear, noisy case, *Appl. Anal.*, vol. 97, no. 1, pp. 107–123, 2018.

35. J. A. Carrillo and U. Vaes, Wasserstein stability estimates for covariance-preconditioned Fokker-Planck equations, *Nonlinearity*, vol. 34, no. 4, p. 2275, 2021.

36. Z. Ding and Q. Li, Ensemble Kalman Inversion: mean-field limit and convergence analysis, *Stat. Comput.*, vol. 31, p. 9, 2021.

37. M. Herty and G. Visconti, Kinetic methods for inverse problems, *Kinet. Relat. Models*, vol. 12, no. 5, pp. 1109–1130, 2019.

38. N. K. Chada, Limit analysis of hierarchical ensemble Kalman inversion, *J. Inverse Ill-Posed Probl.*, 2020. In press.

39. Z. Ding, Q. Li, and J. Lu, Ensemble Kalman inversion for nonlinear problems: Weights, consistency, and variance bounds, *Found. Data Sci.*, vol. 3, no. 3, pp. 371–411, 2021.

40. A. Armbruster, M. Herty, and G. Visconti, A stabilization of a continuous limit of the ensemble Kalman inversion, *SIAM J. Numer. Anal.*, 2022. Accepted. Preprint arXiv:2006.15390.

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

42. F. Golse, On the dynamics of large particle systems in the mean field limit, in *Macroscopic and large scale phenomena: coarse graining, mean field limits and ergodicity*, pp. 1–144, Springer, 2016.

43. P.-E. Jabin, A review of the mean field limits for Vlasov equations, *Kinetic & Related Models*, vol. 7, no. 4, pp. 661–711, 2014.

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

45. G. Albi and L. Pareschi, Binary interaction algorithms for the simulation of flocking and swarming dynamics, *Multiscale Model. Simul.*, vol. 11, no. 1, pp. 1–29, 2013.

46. M. Ehrgott, *Multicriteria optimization*, vol. 491. Springer Science & Business Media, 2005.

47. K. Miettinen, *Nonlinear multiobjective optimization*, vol. 12. Springer Science & Business Media, 2012.

48. P. M. Pardalos, A. Žilinskas, J. Žilinskas, *et al.*, *Non-convex multi-objective optimization.* Springer, 2017.

49. M. Herty and E. Iacomini, Filtering methods for coupled inverse problems. Preprint. arXiv:2203.09841, 2022.

50. K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE transactions on evolutionary computation*, vol. 6, no. 2, pp. 182–197, 2002.