Physica D 300 (2015) 34–40

Contents lists available at ScienceDirect

Physica D

journal homepage: www.elsevier.com/locate/physd

(Non)Uniqueness of critical points in variational data assimilation

Graham Cox

Department of Mathematics, UNC Chapel Hill, Phillips Hall CB #3250, Chapel Hill, NC 27599, United States

HIGHLIGHTS

- Bayesian formulation of variational assimilation for quasilinear equations.
- Uniqueness of minimizers for small observational times.
- Uniqueness of minimizers for small prior covariance.
- Existence of critical points with large Morse index.

ARTICLE INFO

Article history:
Received 30 September 2013
Received in revised form
28 November 2014
Accepted 18 January 2015
Available online 24 January 2015
Communicated by E.S. Titi

Keywords:
Variational data assimilation
Inverse problems
Quasilinear evolution equations

ABSTRACT

In this paper we apply the 4D-Var data assimilation scheme to the initialization problem for a family of quasilinear evolution equations. The resulting variational problem is non-convex, so it need not have a unique minimizer. We comment on the implications of non-uniqueness for numerical applications, then prove uniqueness results in the following situations: (1) the observational times are sufficiently small; (2) the prior covariance is sufficiently small. We also give an example of a data set where the cost functional has a critical point of arbitrarily large Morse index, thus demonstrating that the geometry can be highly nonconvex even for a relatively mild nonlinearity.

1. Introduction

An important problem in data assimilation is to estimate the initial state of a physical system given only noisy, incomplete observations of the state at later times. To make this precise, suppose $y(t)$ solves an evolution equation $y_t = F(y)$ in some function space V and the observations of the state are given by a bounded linear operator $H: V \to \mathbb{R}^q$. Given observations $z_1, \ldots, z_N \in \mathbb{R}^q$ at times $t_1 < \cdots < t_N$, one would like to find the initial condition $u = y(0)$ that best matches the empirical data.

Of course it is important to carefully formulate what is meant by the "best" initial condition, to ensure that the problem is well-posed and has a physically meaningful solution. A common approach, which we adopt in this paper, is to minimize the cost functional

$$J(u) := \frac{1}{2}\sum_{i=1}^{N}\left|R^{-1/2}\left(Hy(t_i) - z_i\right)\right|^2 + \frac{1}{2\sigma^2}\|u - u_0\|_V^2 \qquad (1)$$

for some fixed $u_0 \in V$ and $\sigma > 0$. It can be shown through standard variational methods that J admits a minimizer (see [1,2] for details). However, if the forward problem is nonlinear J may fail to be convex and so uniqueness is not guaranteed.

It is common practice (see, for instance, [3–6]) to solve a suitable discretization of the regularized variational problem using a gradient-based algorithm. Implicit in the implementation of such an algorithm is the assumption of a unique minimizer for the variational problem—gradient descent methods are of course local and do not have the ability to distinguish between local and global minima. However, while widely recognized as a difficulty inherent in the nonlinear case (cf. [7,8]), the problem of uniqueness has so far received little attention in the literature.

A short-time uniqueness result for Burgers' equation appeared in [2] under the assumption of continuous-in-time observations, using the cost functional

$$\int_0^T |Hy(t) - z(t)|^2\, dt + \frac{1}{2\sigma^2}\|u - u_0\|_V^2.$$

There it was proved that the variational problem admits a unique minimizer when the maximal observation time T is sufficiently small, by showing that critical points are fixed points of a map (cf. (13)) that is a contraction when T is small. This turns out to be quite different from the case of discrete-time observations considered in the current paper—in the continuous case the data term becomes

E-mail address: ghcox@email.unc.edu.

http://dx.doi.org/10.1016/j.physd.2015.01.003
0167-2789/

negligible as $T \to 0$, whereas in the discrete case the data will always have a nontrivial contribution to the regularized cost function, no matter how small the observational times.

The discrete-time problem was investigated numerically in [9], where a unique minimizer was observed as long as $\sigma > 0$. For the non-regularized ($\sigma = 0$) case—corresponding to an improper prior in the Bayesian formulation—multiple minimizers were found numerically.

An interesting approach to uniqueness appeared in [10], in which a so-called curvature $\times$ size condition was used to establish uniqueness for the class of "weakly nonlinear inverse problems" (which includes the initialization problem for semilinear evolution equations). A current summary of these geometric methods can be found in [11]. It was shown there that the initialization problem admits a unique solution provided one is given an H^1 observation of the solution (in both space in time). While geometrically appealing, these methods seem less suited to situations where only partial information is available.

The variational problem for (1) can be viewed as a Tikhonov regularization of the log-likelihood functional

$$u \mapsto \sum_{i=1}^{N} \left| R^{-1/2} \left(Hy(t_i) - z_i \right) \right|^2 ,$$

where R is the observational covariance matrix and $y(t_i)$ is the solution to the evolution equation with initial condition u, evaluated at time t_i. Such regularizations for linear problems are well-established; some classical sources are [12–15] and a modern overview can be found in [16].

The non-regularized problem is ill-posed, in the sense that it does not necessarily possess a minimizer in the space V. Ill-posed inverse problems are well-represented in the literature, with classic examples being the determination of a diffusion coefficient for the heat equation, or a source term for a semilinear reaction–diffusion equation. By definition these problems have one of the following properties:

(1) no solutions exist;
(2) a solution exists but is not stable with respect to perturbations of the data;
(3) multiple solutions exist.

In the literature one can find many results of existence [7,17,18]; uniqueness [19]; existence and uniqueness [20,21]; uniqueness and stability [22]. A recent survey of some analytic uniqueness results can be found in Chapter 6 of [23]. However, the question of uniqueness for regularized, nonlinear variational problems seem to be less well studied.

There is a Bayesian interpretation of (1) in which the $\|\cdot\|_V$ term corresponds to a prior distribution with covariance proportional to σ^2 and minimizers of J correspond to modes of the posterior distribution, hence the variational problem for J admits multiple minimizers precisely when the posterior distribution is multimodal. This interpretation also provides a useful interpretation of the parameter σ, and suggests how it should be chosen based on prior information.

The goal of this paper is to describe this Bayesian formalism for a family of quasilinear evolution equations (which includes reaction–diffusion equations and viscous conservation laws) and determine sufficient conditions to guarantee unimodality of the resulting posterior distribution. We emphasize that our methods only assume finite data at each observational time (for instance, projection onto the first N Fourier modes), rather than complete L^2 or H^1 knowledge of the state.

1.1. Some notation and conventions

Throughout we denote the $L^2(0, 1)$ norm and inner product by $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$, respectively. We let $V := H_0^1(0, 1)$, with norm $\|u\|_V := \|u_x\|$. This is equivalent to the standard $H^1(0, 1)$ norm, because $\pi\|u\| \le \|u_x\|$ for any $u \in H^1(0, 1)$. It is well known that $H_0^1(0, 1) \subset L^\infty(0, 1)$, with $\sup|u| \le \|u_x\|$. We will frequently make use of the inequality between the arithmetic and geometric means,

$$2ab \le \lambda a^2 + \lambda^{-1} b^2 \tag{2}$$

for any positive a, b and λ, which we refer to as the AM–GM inequality.

2. Statement of results

For the remainder of the paper we consider a quasilinear parabolic equation

$$y_t + f(y)_x = y_{xx} + r(y) \tag{3}$$

on the interval $[0, 1]$, with Dirichlet boundary conditions. We make the standing assumption that f and r are both of class C^2. This is more than sufficient to guarantee that the initial value problem for (3) is well-posed, as will be seen in Proposition 3. The additional regularity is needed in computing the first and the second variation of the cost functional. We also need to ensure that the initial value problem admits a global (in time) solution for any initial value, so that J is well-defined on all of H_0^1. This will be the case if

$$\int_{-\infty}^{0} \frac{1}{|r(y)|+1}\,dy = \int_{0}^{\infty} \frac{1}{|r(y)|+1}\,dy = \infty. \tag{4}$$

If this condition is not satisfied, there may exist initial conditions for which the solution blows up in a finite amount of time.

We also assume that the observation operator H is bounded on L^2, and hence has a bounded adjoint $H^*: \mathbb{R}^q \to L^2$.

Our first result is that the problem has a natural Bayesian formulation with respect to a Gaussian prior distribution, the significance of which will be discussed in Section 5. This requires an additional regularity assumption on r and f that will not be needed elsewhere in the paper. We let Δ_D denote the Dirichlet Laplacian on $L^2(0, 1)$—that is, the unbounded, selfadjoint operator with domain $H^2(0, 1) \cap H_0^1(0, 1)$ and bounded inverse denoted by $\Delta_D^{-1}: L^2(0, 1) \to H^2(0, 1) \cap H_0^1(0, 1)$.

Theorem 1. *Let μ_0 denote the Gaussian measure on $L^2(0, 1)$ with covariance $\mathcal{C}_0 = -\sigma^2 \Delta_D^{-1}$ and mean u_0, and suppose that $r(y)$ and $f'(y)$ are uniformly Lipschitz. Given observations $\{z_i\}_{i=1}^N$ with i.i.d. $\mathcal{N}(0, R)$ Gaussian noise, there is a well-defined posterior measure μ_z, with Radon–Nikodym derivative*

$$\frac{d\mu_z}{d\mu_0}(u) \propto \exp\left\{ -\sum_{i=1}^{N} \left| R^{-1/2}\left(Hy(t_i) - z_i \right) \right|^2 \right\}. \tag{5}$$

Moreover, the mean and covariance of the posterior distribution are continuous functions of the data, $z = \{z_i\}$.

In fact, one has that the posterior measure is Lipschitz with respect to the Hellinger metric; the reader is referred to [1] for further details. The nontriviality of this result is due to the infinite-dimensional setting of the problem. Because there is no analog of the Lebesgue measure for infinite-dimensional spaces, one cannot define the posterior measure using the exponential of J as a density, as is done in finite dimensions. Thus it is necessary to define the posterior relative to the prior distribution, and care must be taken to ensure that this density, given by (5), is in fact μ_0-integrable and hence can be normalized. This normalizability will follow from estimates on solutions to the nonlinear evolution equation.

There is thus a Bayesian formulation of the regularized variational problem, for which the MAP (Maximum *A Posteriori*) estimators are precisely the global minima of the cost functional (1).