SIAM J. APPLIED DYNAMICAL SYSTEMS
Vol. 16, No. 1, pp. 361–374

Existence and Uniqueness for Four-Dimensional Variational Data Assimilation in Discrete Time[*]

Jochen Bröcker[†]

Abstract. Variational techniques for data assimilation, i.e., estimating orbits of dynamical models from observations, are revisited. It is shown that under mild hypotheses a solution to this variational problem exists. Using ideas from optimal control theory, the problem of uniqueness is investigated and a number of results (well known from optimal control) are established in the present context. The value function is introduced as the minimal cost over all feasible trajectories starting from a given initial condition. By combining the necessary conditions with an envelope theorem, it is shown that the solution is unique if and only if the value function has a derivative at the given initial condition. Further, the value function is Lipschitz and hence has a derivative for almost all (with respect to the Lebesgue measure) initial conditions. Several examples are studied which demonstrate that points of nondifferentiability of the value function (and hence nonuniqueness of solutions) are nevertheless to be expected in practice.

Key words. data assimilation, optimal control, variational calculus, envelope theorem

AMS subject classifications. 37N35, 86A10, 86A22, 93C95

DOI. 10.1137/16M1068918

1. Introduction.

Data assimilation is a term used in atmospheric physics and geosciences and refers to methods whereby series of observations are employed to reconstruct states or orbits of relevant dynamical models. Said differently, possible states or orbits are determined which are consistent with a given dynamical model on the one hand and given observations on the other hand. Operational weather forecasting centers carry out data assimilation on a daily basis in order to find initial conditions for the simulation of future weather, and assimilation of historical weather observations into state-of-the-art weather models (known as reanalyses) is an important scientific approach to gain insight into past and long-term weather patterns. Many different approaches to data assimilation exist, based on very different philosophies and premises. For an overview of data assimilation techniques from an atmospheric physics point of view, see, e.g., [10, 12, 8].

Variational approaches have gained widespread attention both within the atmospheric physics community and in other brances of science. The basic idea of what is known as *weakly constrained four-dimensional variational assimilation*, or WC-4DVar, is to find a series of model states by minimizing a cost functional which quantifies both the deviations from the

[*] Received by the editors April 4, 2016; accepted for publication (in revised form) by A. Stuart November 3, 2016; published electronically February 8, 2017.

http://www.siam.org/journals/siads/16-1/M106891.html

Funding: This research was supported by the UK Engineering and Physical Sciences Research Council under grant agreement EP/L012669/1.

[†] School of Mathematical and Physical Sciences, University of Reading, Reading RG6 6AX, United Kingdom (j.broecker@reading.ac.uk).

observed data as well as the misfit with the given model. Depending on the cost functional and the assumptions, the solution may have a deeper probabilistic interpretation as a maximum likelihood or maximum aposteriori estimate (see also equation (1) below and the discussion thereafter). Given a probabilistic interpretation of the cost function, there is the possibility of analyzing the problem further in a statistical framework, for instance (as one referee suggested), by using the model errors to test the goodness of fit of a particular dynamical system. This is beyond the scope of this article, but see, for instance, [2] for a study along these lines.

In terms of references, an early paper on discrete time WC-4DVar in atmospheric sciences is [6]; see also [12]. In [5], the authors consider assimilation into models with infinite dimensional state space and provide a rigorous derivation of the maximum aposteriori estimator. In the engineering community, essentially the same technique has been known for much longer. In [11, Chap. 5, sect. 3], the approach is discussed for both discrete and continuous time; [14] considers both cases from a control point of view, calling it the *optimal servomechanism* (there are minor differences). The maximum a posteriori estimator for continuous time diffusions has been analyzed rigorously in [16, 17]. The present paper will deal exclusively with discrete time, the continuous time situation being considered in a forthcoming paper.

To formulate the problem in mathematical terms, fix a number $n \in \mathbb{N}$ as the length of our assimilation window. Let E, the *state space*, be a finite dimensional vector space with some norm $\|.\|$. We will use boldface letters to denote elements of E^n, like $\mathbf{x} := (x_1, \ldots, x_n)$, where $x_k \in E$ for each $k = 1, \ldots, n$. Consider a sequence of mappings $f_k : E \to E, k = 1, \ldots, n$, our time dependent *model*. A given candidate trajectory $\mathbf{x} = (x_1, \ldots, x_n)$ can always be written as

$$x_k = f_k(x_{k-1}) + u_k, \qquad k = 1, \ldots, n,$$

for certain $u_k, k = 1, \ldots, n$ which could be termed "model error" and a certain initial condition $x_0 \in E$. A common way to quantify the total model misfit of a candidate trajectory $\mathbf{x}$ is via a functional of the form $\frac{1}{2}\sum_{k=1}^{n} u_k^T B_k u_k$, where the $B_k, k = 1, \ldots, n$, are positive definite matrices.

Regarding the misfit with respect to the observations, if the state of the model at time $k = 1, \ldots, n$ is estimated to be x, then often the error of this estimate with respect to the observations y_k taken at time k has the form $(h(x) - y_k)^T G_k (h(x) - y_k)$ or similar, where $h : E \to \mathbb{R}^d$ is some function mapping the state space of the model into the space of observations, and G_k is some appropriate positive semidefinite quadratic form on $\mathbb{R}^d$, for each $k = 1, \ldots, n$.

Combining the misfit with respect to the model and with respect to the observations in the forms mentioned above, we would arrive at the following example for a cost function:

$$(1) \qquad \frac{1}{2}\sum_{k=1}^{n}(h(x_k) - y_k)^T G_k (h(x_k) - y_k) + \frac{1}{2}\sum_{k=1}^{n} u_k^T B_k u_k.$$

For an interpretation of this cost function in the context of maximum a posteriori estimation, see [11, Chap. 5, sect. 3]. We will be able to deal with more general cost functions though. Assume that for each $k = 1, \ldots, n$ we are given a function $L_k : E \times E \to \mathbb{R}_{\geq 0}$ called the *running costs* at time k. We formulate our basic problem as follows.

Problem 1 (WC-4DVAR). *Minimize the* objective function $A : E^n \times E^n \to \mathbb{R}_{\geq 0}$ *given by*

$$A(\mathbf{x}, \mathbf{u}) = \sum_{k=1}^{n} L_k(x_k, u_k) \tag{2}$$

over elements $(\mathbf{x}, \mathbf{u}) \in E^n \times E^n$ *satisfying the constraint*

$$x_k = f_k(x_{k-1}) + u_k \qquad \text{for } k = 1, \ldots, n \tag{3}$$

for some fixed initial condition $x_0 = \xi \in E$.

The $x_1, \ldots, x_n$ and $u_1, \ldots, u_n$ will be referred to, respectively, as the *states* and *controls* from now on. Note that the dependence on the observations is only implicit in the time dependence of the running costs. Although our analysis will focus mainly on the problem as stated in Problem 1, we shall mention a variant which will briefly be discussed at the end of the paper:

Problem 2 (WC-4DVAR with background error). *Minimize the* objective function $A : E^{n+1} \times E^{n+1} \times E \to \mathbb{R}_{\geq 0}$ *given by*

$$A(\mathbf{x}, \mathbf{u}, \xi) = \sum_{k=1}^{n} L_k(x_k, u_k) + \psi(u_0 - \xi) \tag{4}$$

over elements $(\mathbf{x}, \mathbf{u}) \in E^{n+1} \times E^{n+1}$ *satisfying the constraint*

$$\begin{aligned} x_k &= f_k(x_{k-1}) + u_k \qquad \text{for } k = 1, \ldots, n, \\ x_0 &= u_0. \end{aligned} \tag{5}$$

The additional term $\psi(u_0 - \xi)$ in the objective function is referred to as *background error*; here $\psi : E \to \mathbb{R}_{\geq 0}$ is a nonnegative function and ξ is a fixed element in E known as the *background state*.

Our main results regarding these two variants of the problem are exactly the same (for precise hypotheses and statements see the respective sections). In section 2, we will state our main hypotheses and prove that Problem 1 has global minimizers for every $\xi \in E$ (Proposition 2.2). In order to analyze uniqueness, we consider the value function (Definition 3.2), defined as $V(z) := \inf A(\mathbf{x}, \mathbf{u})$, where the infimum is taken over all $(\mathbf{x}, \mathbf{u})$ which satisfy the constraints with initial condition z. Theorem 3.3 shows that a global minimizer for initial condition $\xi \in E$ is unique if the value function is differentiable at ξ. We can show that the set of points where the value function fails to be differentiable has Lebesgue measure zero (see Corollary 3.6 and the discussion preceeding it); we cannot, however, altogether exclude the existence of such points. We expand on this by showing that uniqueness of global minimizers and differentiability of the value function is in fact an equivalence (Theorem 3.4). Further, we discuss two examples in section 4, a simple function minimization and a small data assimilation problem), where corners in the value function (and hence nonuniqueness of global minimizers) do occur for certain initial conditions $\xi \in E$. In section 5 we discuss the necessary modifications in our analysis to obtain the same results for Problem 2, that is, with background error.

2. Existence of solutions. In addition to the conditions mentioned in the introduction, we impose the following hypotheses.

Hypothesis 1. *The mappings $f_k : E \to E, k = 1, \ldots, n$, have continuous first derivatives $Df_k(x)$ for every $k = 1, \ldots, n$ and $x \in E$ which are nonsingular in x.*

Hypothesis 2. *The running costs $L_k : E \times E \to \mathbb{R}_{\geq 0}, k = 1, \ldots, n$, have continuous first partial derivatives.*

Hypothesis 3. *For each $k = 1, \ldots, n$ and $x \in E$ and $\lambda \in E$ the equation*

$$0 = \mathrm{D}_2 L_k(x, u) - \lambda^T$$

has a unique solution $u_k(x, \lambda)$.

Here and in the following, we use the symbol D_n to denote the partial derivative of a mapping with respect to the nth argument. Similarly, D denotes the total derivative.

Hypothesis 4. *For all $k = 1, \ldots, n$ and $x \in E$, we have the estimate*

$$L_k(x, u) \geq \phi(u),$$

where $\phi : E \to \mathbb{R}$ is bounded below and has bounded level sets (an example would be $\phi(u) = a\|u\|^2 - b$ for some $a, b > 0$).

Hypothesis 5. *The function $\psi : E \to \mathbb{R}_{\geq 0}$ has continuous first derivatives and we have the estimate*

$$\psi(u) \geq \phi(u),$$

where ϕ is as in Hypothesis 4.

Hypothesis 6. *For each $v, \xi \in E$ the equation $v = \mathrm{D}\psi(x - \xi)$ has a unique solution $x = x(v, \xi)$.*

We note that the objective function displayed in (1) has running costs

$$L_k(x, u) = \frac{1}{2}(h(x) - y_k)^T G_k(h(x) - y_k) + \frac{1}{2} u^T B_k u$$

which satisfy Hypotheses 2, 3, and 4 if the matrices $B_k, k = 1, \ldots, n$, are positive definite, the matrices $G_k, k = 1, \ldots, n$, are nonnegative definite, and h has continuous first derivatives. We stress, however, that Hypotheses 2, 3, and 4 do not imply that the running costs L_k are in any way "nearly quadratic." For instance, running costs of the form $L_k(x, u) = a_k(x) + \|u\|^r$ for some $r > 1$, with suitable functions a_k, would also satisfy these hypotheses.

Definition 2.1. *An element $(\mathbf{x}, \mathbf{u}) \in E^n \times E^n$ satisfying the constraint with some initial condition $\xi \in E$ will be referred to as* admissible with respect to ξ *(or simply* admissible *if ξ is specified). An element $(\mathbf{x}^*, \mathbf{u}^*) \in E^n \times E^n$ which is admissible with respect to some $\xi \in E$ will be referred to as a* minimizer of Problem 1 with respect to ξ *if $A(\mathbf{x}^*, \mathbf{u}^*) \leq A(\mathbf{x}, \mathbf{u})$ for any other $(\mathbf{x}, \mathbf{u}) \in E^n \times E^n$ which is admissible with respect to ξ.*

It is clear that for any given $\mathbf{u} \in E^n$ and $\xi \in E$ we can use the constraints (3) to generate a (uniquely defined) $\mathbf{x} \in E^n$ so that $(\mathbf{x}, \mathbf{u})$ is admissible with respect to ξ. That is, there exists a mapping $\mathbf{X} : E \times E^n \to E^n$ so that $\mathbf{x} = \mathbf{X}(\xi, \mathbf{u})$. Further, we can define a function $J : E \times E^n \to \mathbb{R}_{\geq 0}$ by

$$J(\xi, \mathbf{u}) := A(\mathbf{X}(\xi, \mathbf{u}), \mathbf{u}),$$

that is, we use the mapping $\mathbf{X}$ to eliminate $\mathbf{x}$ from the objective function.

Proposition 2.2. *For any $\xi \in E$, there exists a minimizer of Problem 1 with respect to ξ.*

Proof. It is clear that finding a minimizer of Problem 1 with respect to ξ is equivalent to solving the unconstrained problem of minimizing $J(\xi, \mathbf{u})$ over all $\mathbf{u} \in E^n$. Indeed, if $\mathbf{u}^*$ is a minimizer of $J(\xi, \mathbf{u})$ for some ξ, then $(\mathbf{x}^*, \mathbf{u}^*)$ with $\mathbf{x}^* := \mathbf{X}(\xi, \mathbf{u}^*)$ is a minimizer of Problem 1 with respect to ξ. To minimize $J(\xi, \mathbf{u})$ over $\mathbf{u}$, we need to consider only those $\mathbf{u}$ for which $J(\xi, 0) \geq J(\xi, \mathbf{u})$. Since J is continuous, those $\mathbf{u}$ form a closed set U_ξ. On the other hand, $J(\xi, \mathbf{u}) \geq \sum_{k=1}^n \phi(u_k)$ by Hypothesis 4, which means for $\mathbf{u} \in U_\xi$ that

$$J(\xi, 0) \geq \sum_{k=1}^n \phi(u_k).$$

By the properties of ϕ we obtain that U_ξ is bounded and hence compact, but a continuous function on a compact set attains its minimum. Although this is a standard fact, we will give a proof which puts Corollary 2.3 into context. A *minimizing sequence* is a sequence $\{(\mathbf{u}^{(n)}, n \in \mathbb{N}\}$ in U_ξ so that $J(\xi, \mathbf{u}^{(n)})$ is monotone decreasing in n and converges to $\inf J(\xi, \mathbf{u})$. Such a sequence clearly exists (although its actual construction might be nontrivial). Since U_ξ is compact, there is a subsequence (which we do not relabel) converging to some $\mathbf{u}^*$. It then follows from continuity that

$$J(\xi, \mathbf{u}^*) = J\left(\xi, \lim_{n\to\infty} \mathbf{u}^{(n)}\right) = \lim_{n\to\infty} J\left(\xi, \mathbf{u}^{(n)}\right) = \inf J(\xi, \mathbf{u}). \qquad \blacksquare$$

We remark that the proof uses merely Hypothesis 4 as well as the continuity of the f_k's and L_k's (which is of course implied by Hypotheses 1 and 2). The following corollary underlines the importance of the uniqueness of minimizers.

Corollary 2.3. *If a minimizer $(\mathbf{x}^*, \mathbf{u}^*)$ of Problem 1 with respect to ξ is unique, then any minimizing sequence converges to it.*

Proof. Any subsequence of your favorite minimizing sequence is still a minimizing sequence and hence (by the proof of Proposition 2.2) has a subsubsequence which converges to a minimizer which must be $(\mathbf{x}^*, \mathbf{u}^*)$. This implies that your favorite minimizing sequence also converges to it. ■

3. Uniqueness of solutions. We now investigate the uniqueness of minimizers. A very important concept in this analysis will be the *value function*, which plays a prominent role in the calculus of variations [7, 4] and optimal control theory [9, 1]. We will basically recover well-known results for the value function in the special setup considered in this paper. As part of our analysis, we need to consider the classical necessary conditions for a minimizer, involving the method of Lagrange multipliers.

Proposition 3.1. *If $(\mathbf{x}^*, \mathbf{u}^*) \in E^n \times E^n$ is a minimizer of Problem 1 with respect to ξ, then there exists $\boldsymbol{\lambda}^* \in E^n$ so that the triple $(\mathbf{x}^*, \mathbf{u}^*, \boldsymbol{\lambda}^*)$ satisfies the conditions*

$$\lambda_k^T = \lambda_{k+1}^T Df(x_k) - \mathrm{D}_1 L_k(x_k, u_k) \qquad \textit{for } k = 1, \ldots, n, \tag{6}$$

$$x_k = f_k(x_{k-1}) + u_k \qquad \textit{for } k = 1, \ldots, n, \tag{7}$$

$$0 = \mathrm{D}_2 L_k(x_k, u_k) - \lambda_k^T \qquad \textit{for } k = 1, \ldots, n, \tag{8}$$

$$\lambda_{n+1} = 0, \qquad x_0 = \xi. \tag{9}$$

Proof. The differentiability conditions on the L_k and f imply that the method of Lagrange multipliers can be applied (see, e.g., [3]), implying that there exists a $\boldsymbol{\lambda}^* \in E^n$ so that the function

$$L(\mathbf{x}, \mathbf{u}, \boldsymbol{\lambda}) := A(\mathbf{x}, \mathbf{u}) + \sum_{k=1}^{n} \lambda_k^T \left(x_k - f_k(x_{k-1}) - u_k \right),$$

where $x_0 = \xi$, has a critical point at $(\mathbf{x}^*, \mathbf{u}^*, \boldsymbol{\lambda}^*)$. A simple calculation then gives (6), (7), (8), (9). ■

The necessary conditions (6), (7), (9) provide $2n$ coupled equations for the $2n$ (times the dimension of E) unknowns $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ if we regard $\mathbf{u}^*$ as an auxilliary variable defined by relation (8) in view of Hypothesis 3. These equations cannot be solved by simply regarding them as difference equations and iterating them forward (or backward), since the initial conditions are not fully specified. We know from section 2 though that there exists at least one solution to the necessary conditions which corresponds to a global minimizer. We will see that under certain conditions, the property of being a global minimizer implies that λ_1 must have a specific value (which depends on ξ). But thanks to Hypothesis 1 there can be *at most* one solution to the necessary conditions with $x_0 = \xi$ and λ_1 being equal to some specified element in E, and it follows that such a minimizer must be unique.

But even in this situation, there might exist further solutions to the necessary conditions corresponding, for instance, to local minima or other nonminimal critical points. The existence of such solutions cannot be ruled out by our approach.

The following function will be important in our analysis.

Definition 3.2. *The* value function $V : E \to \mathbb{R}_{\geq 0}$ *is defined as*

$$V(z) := \inf A(\mathbf{x}, \mathbf{u}),$$

where the inf *is over all elements $(\mathbf{x}, \mathbf{u}) \in E^n \times E^n$ which are admissible with respect to initial condition $x_0 = z \in E$.*

As we consider V as a function of the initial condition, we use the symbol z for generic initial conditions in E. Further, we fix an arbitrary point $\xi \in E$ as a specific initial condition we want to investigate the value function at.

From the very definition of the value function, it follows that

$$V(z) \leq A(\mathbf{x}, \mathbf{u}) \tag{10}$$

for any $(\mathbf{x}, \mathbf{u})$ admissible with respect to z, while we have equality in (10) if and only if any $(\mathbf{x}, \mathbf{u})$ is a minimizer with respect to z. Now assume that $(\mathbf{x}^*, \mathbf{u}^*)$ is a minimizer with respect to ξ. We modify this minimizer by considering an element $(\mathbf{x}', \mathbf{u}') \in E^n \times E^n$ with $x'_1 = x^*_1, \ldots, x'_n = x^*_n$ and also $u'_2 = u^*_2, \ldots, u'_n = u^*_n$, but $u'_1 = u^*_1 + f(\xi) - f(z)$, where $z \in E$ is a generic initial condition. Thus, we can consider $(\mathbf{x}', \mathbf{u}')$ and also $A(\mathbf{x}', \mathbf{u}')$ as a function of z. More explicitly, we have

$$A(\mathbf{x}', \mathbf{u}') = A(\mathbf{x}^*, \mathbf{u}^*) - L_1(x^*_1, u^*_1) + L_1(x^*_1, u^*_1 + f(\xi) - f(z)) =: \alpha(z).$$

Since $(\mathbf{x}', \mathbf{u}')$ is still admissible with respect to z, but not necessarily a minimizer, it follows from (10) and the discussion that

$$V(z) \leq A(\mathbf{x}', \mathbf{u}') = \alpha(z)$$

for all $z \in E$, with equality if $z = \xi$. This implies that if V has a derivative at ξ, it must be equal to $\mathrm{D}\alpha(\xi) = -\mathrm{D}_2 L(x^*_1, u^*_1)\mathrm{D}f(\xi) = -\lambda^{*T}_1 \mathrm{D}f(\xi)$. We arrive at the following conclusion (which is basically the envelope theorem; see, e.g., [13]).

Theorem 3.3. *If $(\mathbf{x}, \mathbf{u})$ is a minimizer with respect to the initial condition $x_0 = \xi \in E$ and the value function V has a derivative at ξ, then this minimizer is unique, and*

$$\lambda^{*T}_1 = -\mathrm{D}V(\xi)\mathrm{D}f(\xi)^{-1}. \tag{11}$$

We emphasize again that even uniqueness of the minimizer does not rule out the existence of further solutions to the necessary conditions. We have just proved that only one of them can correspond to a minimizer.

Now that we have solved one problem, the next one immediately presents itself: When is the value function differentiable? The rest of the paper will essentially be concerned with this question. First, we show that differentiability is not only sufficient but actually necessary for uniqueness of minimizers.

Theorem 3.4. *If there is only one minimizer $(\mathbf{x}^*, \mathbf{u}^*)$ of Problem 1 with respect to ξ, then V is differentiable at ξ and in fact $\mathrm{D}_1 J(\xi, \mathbf{u}^*) = \mathrm{D}V(\xi)$.*

This follows very easily from Danskin's theorem [4, p. 204], but the proof requires some nonsmooth analysis machinery. Here is a simplified proof which will serve our much more humble needs.

Proof. Take a sequence $x_k \in E, k \in \mathbb{N}$, with $x_k \to \xi$, and for every k, let $\mathbf{u}^{(k)}$ be a minimizer of $J(x_k, \mathbf{u})$ with respect to $\mathbf{u}$. As x_k is convergent, it is bounded (by R, say). As discussed in the proof of Proposition 2.2 we can assume that each $\mathbf{u}^{(k)}$ satisfies the bound $J(x_k, 0) \geq \sum_{l=1}^n \phi(u_l^{(k)})$. On the other hand, $J(z, 0)$ is continuous on the compact set $\{z \in E, \|z\| \leq R\}$ and hence bounded by some $S > 0$ depending on R. Hence $\sum_{l=1}^n \phi(u_l^{(k)}) \leq S$, so we can assume that $\mathbf{u}^{(k)}$ is bounded as well. We recall that $\mathbf{u}^*$ is the unique minimizer of $J(\xi, \mathbf{u})$, and assert that

$$\mathbf{u}^{(k)} \to \mathbf{u}^*. \tag{12}$$

Indeed, taking any subsequence of $(x_k, \mathbf{u}^{(k)})$ (which we do not relabel), we still have that $x_k \to \xi$ and $\mathbf{u}^{(k)}$ minimizes $J(x_k, \mathbf{u})$. As $\mathbf{u}^{(k)}$ is bounded, there is a subsubsequence converging to some $\mathbf{u}'$. Because V and J are continuous, we can now take limits in the equality $V(x_k) = J(x_k, \mathbf{u}^{(k)})$, obtaining $V(\xi) = J(\xi, \mathbf{u}')$. This shows that $\mathbf{u}'$ minimizes $J(\xi, \mathbf{u})$ and hence $\mathbf{u}' = \mathbf{u}^*$. This proves our claim (12).

We will now show that $\mathrm{D}_1 J(\xi, \mathbf{u}^*) = \mathrm{D}V(\xi)$. Because $V(\xi) \leq J(\xi, \mathbf{u}^{(k)})$ we have that

$$V(x_k) - V(\xi) \geq J\left(x_k, \mathbf{u}^{(k)}\right) - J\left(\xi, \mathbf{u}^{(k)}\right) = \mathrm{D}_1 J\left(z_k, \mathbf{u}^{(k)}\right)(x_k - \xi)$$

for some z_k on the line connecting x_k with ξ. This gives

$$V(x_k) - V(\xi) - \mathrm{D}_1 J\left(\xi, \mathbf{u}^{(*)}\right)(x_k - \xi) \geq \left(\mathrm{D}_1 J\left(z_k, \mathbf{u}^{(k)}\right) - \mathrm{D}_1 J\left(\xi, \mathbf{u}^{(*)}\right)\right)(x_k - \xi)$$

and hence

$$\liminf_{k\to\infty} \left(V(x_k) - V(\xi) - \mathrm{D}_1 J\left(\xi, \mathbf{u}^{(*)}\right)(x_k - \xi)\right) \Big/ \|(x_k - \xi)\| \geq 0.$$

On the other hand, because $V(x_k) \leq J(x_k, \mathbf{u}^*)$ we have

$$V(x_k) - V(\xi) \leq J(x_k, \mathbf{u}^*) - J(\xi, \mathbf{u}^*) = \mathrm{D}_1 J(z_k, \mathbf{u}^*)(x_k - \xi)$$

for some z_k on the line connecting x_k with ξ. This gives

$$V(x_k) - V(\xi) - \mathrm{D}_1 J\left(\xi, \mathbf{u}^{(*)}\right)(x_k - \xi) \leq \left(\mathrm{D}_1 J\left(z_k, \mathbf{u}^*\right) - \mathrm{D}_1 J\left(\xi, \mathbf{u}^{(*)}\right)\right)(x_k - \xi)$$

and hence

$$\limsup_{k\to\infty} \left(V(x_k) - V(\xi) - \mathrm{D}_1 J\left(\xi, \mathbf{u}^{(*)}\right)(x_k - \xi)\right) \Big/ \|(x_k - \xi)\| \leq 0. \tag{13}$$ ■

With this result and Theorem 3.3 in mind, the question of differentiability of the value function becomes even more important. We always hope that real-world problems "tend to be" smooth, but value functions are one of the many important exceptions. In the remainder of the present section, we will demonstrate that V fails to be differentiable only on a set of volume zero. This makes precise our previous statement that minimizers are unique for "most" initial conditions.

Proposition 3.5. *Let R be a positive constant. Then for any z_1, z_2 in E with $\|z_1\| \leq R$, $\|z_2\| \leq R$ we have*

$$|V(z_1) - V(z_2)| \leq L\|z_1 - z_2\|,$$

where L is a constant depending only on R.

Proof. The function J has a continuous derivative with respect to the first argument and hence

$$|J(z_1, \mathbf{u}) - J(z_2, \mathbf{u})| \leq \|\mathrm{D}_1 J(\zeta, \mathbf{u})\|\|z_1 - z_2\| \tag{14}$$

for some point $\zeta \in E$ with $\|\zeta\| \leq R$ by the mean value theorem. Considering for the moment only $\mathbf{u}$'s so that $\sum_{k=1}^{n} \phi(u_k) \leq S$ with some $S > 0$, we know from Hypothesis 4 that this is a bounded set and hence $\|\mathrm{D}_1 J(\zeta, \mathbf{u})\| \leq L$ for some L depending on R and S only. Using this in (14), we obtain

$$|J(z_1, \mathbf{u}) - J(z_2, \mathbf{u})| \leq L\|z_1 - z_2\| \tag{15}$$

with L depending on R and S.

It is clear that $V(z) = \inf_{\mathbf{u} \in E^n} J(z, \mathbf{u})$, and we will now prove that it is sufficient to consider the inf only over those $\mathbf{u}$ for which $\sum_{k=1}^{n} \phi(u_k) \leq S$, where S is a sufficiently large constant depending on R. A similar argument was used in the proof of Theorem 3.4; for a given $z \in E$, it suffices to minimize over those $\mathbf{u}$ that satisfy $\sum_{k=1}^{n} \phi(u_k) \leq J(z, 0)$, while $J(z, 0)$ is continuous on the compact set $\{z \in E, \|z\| \leq R\}$ and hence bounded by some $S > 0$ depending on R. We have established that V is the infimum of functions that satisfy relation (15) with L depending on R only, and it is easy to see that V must satisfy this relation as well. ■

Using Rademacher's theorem [7], we obtain the following.

Corollary 3.6. *The points where the value function V fails to be differentiable have Lebesgue measure zero in E.*

Notwithstanding this fact, our examples in the next section will demonstrate that value functions for perfectly smooth problems can exhibit points of nondifferentiability, and it seems that the existence of such points cannot be ruled out without much more stringent structural assumptions. Having said this, it is fairly easy to see that Problem 1 has a unique solution if the function $J(x_0, \mathbf{u})$ could somehow be shown to be strictly convex in $\mathbf{u}$. For then, if $\mathbf{u}^{(1)}$ and $\mathbf{u}^{(2)}$ were two distinct global minimizers with value, say, J_0, any $t \in]0, 1[$ would give rise to the contradiction

$$J\left(x_0, t \cdot \mathbf{u}^{(1)} + (1 - t) \cdot \mathbf{u}^{(2)}\right) < t \cdot J\left(x_0, \mathbf{u}^{(1)}\right) + (1 - t) \cdot J\left(x_0, \mathbf{u}^{(2)}\right) = J_0.$$

But even if A is convex in $(\mathbf{x}, \mathbf{u})$, the dependence of $\mathbf{x}$ on $\mathbf{u}$ for any admissible pair will involve the dynamics f, rendering J a nonconvex function of $\mathbf{u}$ in general.

An important exception is the case of linear dynamics (i.e., the functions f_k are linear) and a cost function of the form (1) with the h_k being linear as well. As is well known, Problem 1 then has a unique solution for all $x_0 = \xi$, and the value function is in fact a quadratic function, the coefficients of which are obtained by solving a matrix valued difference equation of Riccati type *backward* in time. For details, see, for example, [14, Ex. 6.2-8], dealing with the closely related optimal regulator, or [15, Exer. 8.2.7.].

4. Examples.

In this chapter we will further illustrate the concept of the value function in the context of a very simple analytic example as well as a small data assimilation problem. In particular, it will be seen that the value function might exhibit points of nondifferentiability even in perfectly smooth situations. The notorious nonsmoothness of value functions in variational calculus and optimal control is of course a well-known phenomenon; see, e.g., [4].

Analytic example. Consider the function

$$\gamma : [0,1] \times \mathbb{R} \to \mathbb{R}; \gamma(\xi, x) = \frac{1}{4}x^4 - \frac{1}{3}x^3 \cdot (2\xi - 1) + \frac{1}{2}x^2 \cdot \xi(\xi - 1)$$

which we want to minimize over $x \in \mathbb{R}$, where ξ is a parameter. The derivative factorizes as $\mathrm{D}_2\gamma(\xi, x) = x(x - \xi + 1)(x - \xi)$, and it is easily seen that for all $\xi \in [0,1]$, the function has a local maximum at $x = 0$ and two local minima at $x = \xi - 1$ and $x = \xi$. For $\xi < \frac{1}{2}$ the former is the global minimum and for $\xi > \frac{1}{2}$ the latter. In particular, the minimizer is not unique at $\xi = \frac{1}{2}$. An elementary calculation reveals that the value function $V(\xi) = \inf_x \gamma(\xi, x)$ is given as

$$V(\xi) = \begin{cases} \dfrac{1}{12}(\xi - 1)^4 + \dfrac{1}{6}(\xi - 1)^3 & \text{if } \xi \leq \dfrac{1}{2}, \\ \dfrac{1}{12}\xi^4 - \dfrac{1}{6}\xi^3 & \text{if } \xi \geq \dfrac{1}{2}. \end{cases}$$

The value function is Lipschitz continuous but not differentiable everywhere, even though the original problem statement leaves very little to be desired in terms of regularity. Figure 1 shows the value function V as well as γ as a function of ξ (sic) for several values of x. Let us fix a value x_0 and assume that it happens to be a minimizer for γ if $\xi = \xi_0$. We can see that the function $\xi \to \gamma(\xi, x_0)$ "touches" the value function from above at $\xi = \xi_0$, since $V(\xi_0) = \gamma(\xi_0, x_0)$ but $V(\xi) \leq \gamma(\xi, x_0)$ for other values of ξ (because x_0 is not necessarily a minimizer for other ξ). In other words, the function $\xi \to \gamma(\xi, x_0) - V(\xi)$ has a minimum at $\xi = \xi_0$, which implies that $\mathrm{D}_1\gamma(\xi_0, x_0) - \mathrm{D}V(\xi_0) = 0$ or

$$\mathrm{D}_1\gamma(\xi_0, x_0) = \mathrm{D}V(\xi_0), \tag{16}$$

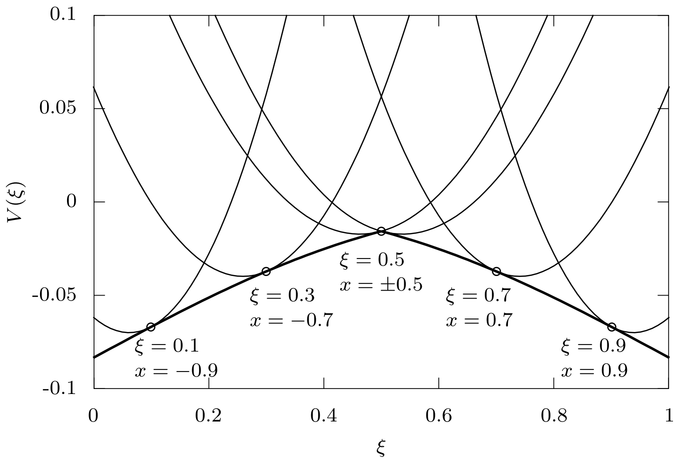

Figure 1. *The value function V as a function of ξ (thick line), showing the concave corner at $\xi = 0.5$. The function $\gamma(\cdot, x)$ is shown for several values of x (thin line; ordinates are scaled), touching the value function from above. For $\xi = 0.1, 0.3, 0.7,$ and 0.9, the minimizer is unique ($-0.9, -0.7, 0.7,$ and 0.9, respectively), and the value function is differentiable at these points. For $\xi = 0.5$, there are two minimizers $x = \pm 0.5$, and both $\gamma(\cdot, 0.5)$ and $\gamma(\cdot, -0.5)$ touch the value function from above, giving rise to the concave corner.*

provided that V has a derivative at ξ_0. Equation (16) is known as the *envelope theorem*, and it provides a necessary condition for a minimizer *in addition to* the usual condition $\mathrm{D}_2\gamma(\xi_0, x_0) = 0$. The reader will realize that this is exactly the argument behind (11).

Our example, however, demonstrates that V need not have a derivative for all ξ but can develop concave corners. Let X_0 be the set of all minimizers of $\gamma(\xi_0, x)$ for fixed ξ_0. If $\mathrm{D}_1\gamma(\xi_0, x)$ varies across X_0, then V cannot be differentiable as this would contradict the envelope theorem. The value function then has, roughly speaking, "several derivatives" as several functions with different derivatives touch V from above. Conversely, if V has a concave corner at ξ_0, then we can at least not exclude that several functions with different derivatives touch V from above, giving rise to several minimizers.

Numerical example. Next we will consider a numerical example, involving data assimilation into a two-dimensional nonlinear system. We use the notation of sections 1 and 2. In this example, $E = \mathbb{R}^2$, and an objective function as in (1) is used with $h(x^{(1)}, x^{(2)}) = (x^{(1)})^2$. The dynamics is given by

$$f : \mathbb{R}^2 \to \mathbb{R}^2, x \to \begin{pmatrix} x^{(2)} - a \cdot \exp\left(-\left(x^{(1)}\right)^2\right) \cdot x^{(1)} \\ b \cdot x^{(1)} \end{pmatrix} \tag{17}$$

with parameters $a = 3.5, b = 0.3$. The observations were generated using the modified dynamical system

$$\tilde{f} := f + \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \tag{18}$$

which we have not analyzed in any depth, but it seems to create a complex attractor shown in Figure 2.

An initial condition was sampled from this attractor and an orbit $(w_1, \ldots, w_n)$ of length $n = 6$ was generated. Putting $y_k = h(x_k) + s \cdot r_k$ for $k = 1, \ldots, 6$ gave the observations, where

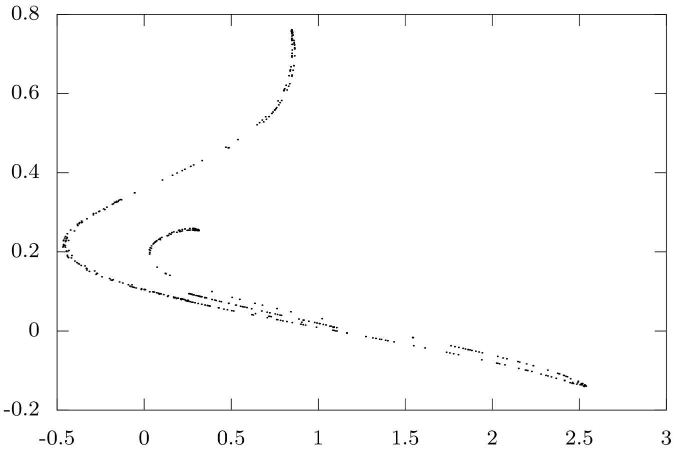

Figure 2. *The attractor for the dynamical system in* (18)*, approximated by an orbit of* 500 *points.*

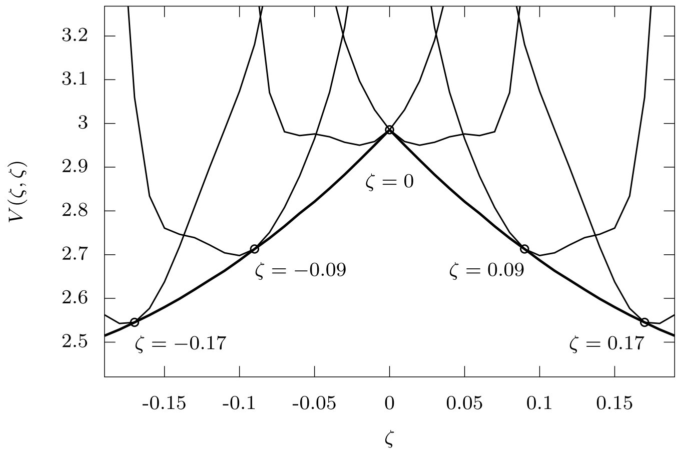

Figure 3. *The value function $V(\zeta,\zeta)$ as a function of ζ (thick line), showing the concave corner at $\zeta = 0$. The function $J(\cdot,\mathbf{u})$ is shown for several values of $\mathbf{u}$ (thin line), touching the value function from above at those ζ for which $\mathbf{u}$ is a minimizer. The minimizer is unique for $\zeta = \pm 0.17, \pm 0.09$, and the value function is differentiable at these points. For $\zeta = 0$, there are two minimizers $\mathbf{u}_1, \mathbf{u}_2$ (with $\mathbf{u}_1 = -\mathbf{u}_2$), and both $J(\cdot,\mathbf{u}_1), J(\cdot,\mathbf{u}_2)$ touch the value function from above, giving rise to the concave corner.*

the r_k are independent standard normal random variables and $s \cong 0.45$ is about 0.2 times the empirical standard deviation of $h(x_k)$. The initial condition ξ in the constraint (3) is two-dimensional, but in order to better illustrate the results, we investigated the initial conditions $\xi = (\zeta,\zeta)$ with $\zeta \in [-2,2]$. The functions J and V will be presented as functions of this parameter ζ.

Before looking at the results, however, we note that if $(\mathbf{x},\mathbf{u})$ is an admissible solution with respect to some initial condition $\xi = (\zeta,\zeta)$, then $(-\mathbf{x},-\mathbf{u})$ is an admissible solution with respect to the initial condition $-\xi = (-\zeta,-\zeta)$. Further, both solutions provide exactly the same value of the objective function, implying that if $(\mathbf{x},\mathbf{u})$ is an optimal solution with respect to $\xi = (\zeta,\zeta)$, then $(-\mathbf{x},-\mathbf{u})$ is optimal for the initial condition $-\xi = (-\zeta,-\zeta)$. In particular, if $(\mathbf{x},\mathbf{u})$ is an optimal solution with respect to $\zeta = 0$ (i.e., $\xi = 0$), then so will be $(-\mathbf{x},-\mathbf{u})$, that is, we can expect the problem to have multiple solutions for $\zeta = 0$.

The objective function was minimized using a standard Nelder–Mead Simplex algorithm as implemented in the Octave (or MATLAB) `fminsearch` function, which was able to find the minimum in all cases with minimal problems. Figure 3 shows the function $J(\cdot,\mathbf{u})$ for several values of $\mathbf{u}$ (thin line), touching the value function from above at those ζ for which $\mathbf{u}$ is a minimizer ($\zeta = \pm 0.17, \pm 0.09$, and $\zeta = 0$ are shown as examples). The minimizer is unique for $\zeta \neq 0$, and the value function is differentiable at these points. For $\zeta = 0$, there are two minimizers $\mathbf{u}_1, \mathbf{u}_2$ (with $\mathbf{u}_1 = -\mathbf{u}_2$), and both $J(\cdot,\mathbf{u}_1), J(\cdot,\mathbf{u}_2)$ touch the value function from above, giving rise to the concave corner.

5. Discussion of 4DVar with background error.

In this section, we will briefly sketch the modifications necessary to prove that our main conclusions hold for Problem 2 as well. It is evident that for any given $\mathbf{u} \in E^n$ we can use the constraints (5) to define a mapping

$\mathbf{X} : E^n \to E^n$ so that $\mathbf{x} = \mathbf{X}(\mathbf{u})$ jointly with $\mathbf{u} \in E^n$ satisfies the constraints (we will now call the pair $(\mathbf{x}, \mathbf{u})$ admissible). We define a function $J : E \times E^n \to \mathbb{R}_{\geq 0}$ by

$$J(\xi, \mathbf{u}) := A(\mathbf{X}(\mathbf{u}), \mathbf{u}, \xi),$$

that is, we use the mapping $\mathbf{X}$ to eliminate $\mathbf{x}$ from the objective function. For the proof of Proposition 2.2, note that we again only need to consider only those $\mathbf{u}$ for which $J(\xi, 0) \geq J(\xi, \mathbf{u})$, but $J(\xi, \mathbf{u}) \geq \sum_{k=1}^{n} \phi(u_k) + \phi(u_0 - \xi)$ by Hypotheses 4 and 5, which means $J(\xi, 0) \geq \phi(u_k)$ for all $k = 1, \dots, n$ and also $J(\xi, 0) \geq \phi(u_0 - \xi)$. Since ϕ has bounded level sets, there exists $S > 0$ so that $S \geq \|u_k\|$ for all $k = 1, \dots, n$ and also $S \geq \|u_0 - \xi\|$ or $S + \|\xi\| \geq \|u_0\|$. We again obtain that U_ξ is bounded and hence compact. The necessary conditions remain the same, apart from the boundary conditions (9) which read as

$$\lambda_{n+1} = 0, \qquad 0 = -\lambda_1^T \mathrm{D} f(x_0) + \mathrm{D}\psi(x_0 - \xi). \tag{19}$$

The value function is simply defined as the infimum

$$V(z) = \inf A(\mathbf{x}, \mathbf{u}, \xi)$$

over all admissible pairs $(\mathbf{x}, \mathbf{u})$. If $(\mathbf{x}^* \mathbf{u}^*)$ is a minimizer with respect to $\xi \in E$ and the value function has a derivative at $z = \xi$, the envelope theorem gives

$$\mathrm{D}V(\xi) = -\mathrm{D}\psi(x_0^* - \xi).$$

Using the necessary conditions (with boundary conditions (19)) and Hypothesis 6, we can conclude as before that the minimizer must be unique. All other results of the paper apply with the same proofs.

Acknowledgments. Fruitful discussions with Tobias Kuna, Nikos Katzourakis, Peter Jan van Leeuwen, Dan Crisan, and Andrew Stuart are gratefully acknowledged.

REFERENCES

[1] M. Bardi and I. Capuzzo-Dolcetta, *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations*, Systems Control Found. Appl., Birkhäuser Boston, Boston, MA, 1997.

[2] J. Bröcker and I. G. Szendro, *Sensitivity and out–of–sample error in continuous time data assimilation*, Quart. J. the Roy. Meteorological Soc., 138 (2012), pp. 785–801.

[3] T. Bröcker, *Analysis.* II. Bibliographisches Institut, Mannheim, 1992.

[4] F. Clarke, *Functional Analysis, Calculus of Variations and Optimal Control*, Grad. Texts in Math. 264, Springer, London, 2013.

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

[6] J. C. Derber, *A variational continuous assimilation technique*, Monthly Weather Rev., 117 (1989), pp. 2437–2446.

[7] L. C. Evans, *Partial Differential Equations*, 2nd ed., Grad. Stud. Math. 19, AMS, Providence, RI, 2010.

[8] G. Evensen, *Data Assimilation. The Ensemble Kalman Filter.* Springer-Verlag, New York, 2007.

[9] W. H. Fleming and H. M. Soner, *Controlled Markov Processes and Viscosity Solutions*, 2nd ed., Stoch. Model. Appl. Probab. 25, Springer, New York, 2006.

[10] K. Ide, P. Courtier, M. Ghil, and A. C. Lorenc, *Unified notation for data assimilation: Operational, sequential and variational*, J. Meteorologcial Soc. Japan, 75 (1997), pp. 181–189.

[11] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Math. Sci. Engrg. 64, Academic Press, New York, 1970.

[12] E. Kalnay, *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge University Press, Cambridge, UK, 2001.

[13] Y. S. Ledyaev, *On generic existence and uniqueness in nonconvex optimal control problems*, Set-Valued Anal., 12 (2004), pp. 147–162.

[14] A. Sage, *Optimum Systems Control*, Prentice-Hall, Englewood Cliffs, NJ, 1968.

[15] E. D. Sontag, *Mathematical Control Theory. Deterministic Finite-Dimensional Systems*, 2nd ed., Texts in Appl. Math. 6, Springer-Verlag, New York, 1998.

[16] O. Zeitouni and A. Dembo, *A maximum a posteriori estimator for trajectories of diffusion processes*, Stochastics, 20 (1987), p. 221.

[17] O. Zeitouni and A. Dembo, *An existence theorem and some properties of maximum a posteriori estimators of trajectories of diffusions*, Stochastics, 23 (1988), p. 197.