University of Reading

Existence and uniqueness for variational data assimilation in continuous time

Article

Accepted Version

Bröcker, J. (2023) Existence and uniqueness for variational data assimilation in continuous time. Mathematical Control & Related Fields, 13 (1). pp. 94-115. ISSN 2156-8472 doi: 10.3934/mcrf.2021050 Available at https://centaur.reading.ac.uk/100788/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.3934/mcrf.2021050

Publisher: American Institute of Mathematical Sciences

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

University of Reading

Reading's research outputs online

EXISTENCE AND UNIQUENESS FOR VARIATIONAL DATA ASSIMILATION IN CONTINUOUS TIME

JOCHEN BRÖCKER

Abstract. A variant of the optimal control problem is considered which is nonstandard in that the performance index contains "stochastic" integrals, that is, integrals against very irregular functions. The motivation for considering such performance indices comes from dynamical estimation problems where observed time series need to be "fitted" with trajectories of dynamical models. The observations may be contaminated with white noise, which gives rise to the nonstandard performance indices. Problems of this kind appear in engineering, physics, and the geosciences where this is referred to as data assimilation. The fact that typical models in the geosciences do not satisfy linear growth nor monotonicity conditions represents an additional difficulty. Pathwise existence of minimisers is obtained, along with a maximum principle as well as preliminary results in dynamic programming. The results also extend previous work on the maximum aposteriori estimator of trajectories of diffusion processes.

1. Introduction and main results

In many branches of science, particularly in physics, the geosciences, and engineering, there appears the problem of "fitting" observed time series with trajectories of dynamical models; typically these are ordinary or partial differential equations. The aim might be to identify appropriate models or to estimate model parameters, but there is also considerable interest in the estimated trajectories themselves. These problems are the main motivation for the results in this paper. In the geosciences, these and related problems are referred to as data assimilation, and we will use this term here, although there has been relevant research in other communities which often predates research in the geosciences. We refer to [3, 18, 10, 4] and references therein for an overview over data assimilation from a geoscientist's perspective; [9, 15, 16] discuss similar problems in engineering; [13] contains many examples of data assimilation problems in biology and other applied sciences.

In the geosciences, the construction of plausible model trajectories which also fit the observational record is particularly important, for at least two reasons. Firstly, trajectories of atmospheric models that fit observations over a long temporal window might hold clues about past weather phenomena which were not directly observed. Examples are large and almost stationary pressure systems over the oceans which lead to very persistent weather patterns over the continents (blocking events) but which are not captured by historic instrumentation and thus appear only indirectly in historic weather records. Secondly, the endpoint of such fitted trajectories might be interpreted as a good guess of the then current state of the atmosphere and hence can be used as initial condition for predictions into the future.

In the context of data assimilation, "fitting" can mean different things but often it involves the optimisation of some sort of error criterion or performance index,

AMS Subj. Class.: Primary 49J55, 49K45; Secondary 86A10, 93E99, 60G35.

Keywords: Data Assimilation, Optimal Control, Young Integral, Pontryagin's maximum principle.

integrated over time, in which case data assimilation essentially becomes a variational problem. If the observations are considered as a stochastic process, the performance index and also the fitted solutions might have a probabilistic interpretation. At the same time though, observations which contain white noise give rise, as we will see, to nonstandard variational problems, because the performance index will then contain an integral against a very irregular function.

To define the class of problems we want to investigate, we consider an interval $I = [0, T]$ with $T > 0$ and a finite dimensional vector space E with norm $|.|$ and dual E'. Let $f : I \times E \to E$, $g : I \times E \to L(E, E)$ be two functions and consider the controlled initial value problem or *state equation*

$$\dot{x}(t) = f(t, x(t)) + g(t, x(t))u(t), \qquad t \in I, \tag{1}$$

with initial condition $\xi \in E$ and control function $u : I \to E$. The function f typically represents the physically relevant part of the model while g might be needed to scale the control; in geophysical applications, g is often the unit matrix. A *process* with respect to ξ is a pair (x, u) so that $u : I \to E$ is measurable, $x : I \to E$ is absolutely continuous, $x(0) = \xi$, and the state equation (1) is satisfied for almost every $t \in I$.

We aim to find controls u that strike a balance between two different aims. On the one hand, u should not be "too large", in order that the second term $g(t, x(t))u(t)$ is not "too large", whence we get that x is not "too far" from being a solution of the physically relevant part of the model, that is approximately

$$\dot{x}(t) \cong f(t, x(t)), \qquad t \in I.$$

But on the other hand, we want x to reproduce the observations, in the following sense: for some mapping $h : I \times E \to \mathbb{R}^d$ (which is part of the problem statement), we aim to find controls u so that $h(t, x(t))$ is not "too far" from the observation $y(t)$ for all $t \in I$. The observations are given and thus part of the problem statement, too. This problem might be approached by optimising an appropriate performance index with respect to the control which takes these two aims into account.

In this paper, we assume the observations to contain a white noise component, implying that referring to the observations as a function y of time is only to be understood heuristically. Any rigorous analysis will have to be based on the cumulative observations, defined as follows. Let $\{W(t), t \in I\}$ be the standard d–dimensional Wiener process on some probability space $(\Omega, \mathcal{B}, \mathbb{P})$. We can assume without loss of generality that $\Omega = C(I, \mathbb{R}^d)$, the space of continuous functions on I with values in $\mathbb{R}^d$, $\mathcal{B}$ the Borel sigma algebra generated by the supremum norm topology, $\mathbb{P}$ the standard d–dimensional Wiener measure,and $W : \Omega \times I \to \mathbb{R}^d; W(t, \omega) = \omega(t)$. Again heuristically, we define our observations as $y(t) = \dot{\zeta}(t) + r(t)$ where $\dot{\zeta}$, the desired signal, is Lebesgue integrable and r is white noise. In our analysis though we will work with the process $\eta(t) = \zeta(t) + W(t)$ for all $t \in I$, where ζ is a (probably random) absolutely continuous function. Formally, η can be considered as the observations in "integrated form", that is $\eta(t) =$"$\int_0^t y(t)\, \mathrm{d}s$". We stress however that this connection is really only formal as η will not have a classical derivative due to the presence of the Wiener process.

In terms of performance indices, quadratic functionals of the form

$$\frac{1}{2} \int_I \big\{y(t) - h(t, x(t))\big\}^{\mathrm{T}} R(t) \big\{y(t) - h(t, x(t))\big\}\, \mathrm{d}t$$

might be considered. This form is fine if the function y has the appropriate integrability properties (see e.g. [11] for such a case) but is not well defined if y contains white noise components, as is the case here. Expanding the square however, ommitting the term quadratic in y (which even if it were well defined could be ommitted

as it does not depend on x or u) and formally replacing with $y(t)\mathrm{d}t = \mathrm{d}\eta(t)$ we obtain

$$\begin{aligned}&\frac{1}{2}\int_I h(t,x(t))^{\mathrm{T}} R(t) h(t,x(t))\,\mathrm{d}t \\ &-\int_I h(t,x(t))^{\mathrm{T}} R(t)\,\mathrm{d}\eta(t) + \frac{1}{2}\int_I u(t)^{\mathrm{T}} S(t) u(t)\,\mathrm{d}t.\end{aligned} \tag{2}$$

Here, R and S are suitable functions with values in the nonnegative definite matrices. The approach is then called weakly constrained 4–dimensional variational assimilation (4D–VAR) in the geosciences, see [3, 18, 10, 4]; this is not an exhaustive list of references, and most authors use a discrete time framework. In the engineering community, the approach is known as minimum energy estimator [12, 8, 11], see also [14]. The approach has been interpreted as maximum aposteriori (MAP) estimation of diffusion trajectories (see e.g. [9]), but this cannot be justified rigorously. The correct interpretation of the functional (2) in terms of large deviations has been given by [8]. In [20, 21], it was shown that the MAP estimator is correctly defined as a minimiser of the Onsager–Machlup functional, which comprises a functional similar to (2) but with further terms added.

The performance indices we will consider in this paper encompass the minimum energy functional as well as the Onsager–Machlup functional (as in [21]); given two functions $\phi : I \times E \times E \to \mathbb{R}$ and $\psi : I \times E \to L(\mathbb{R}^d, \mathbb{R})$, consider the *cost functional*

$$A(x,u) = \int_I \phi(t,x(t),u(t))\,\mathrm{d}t + \int_I \psi(t,x(t))\,\mathrm{d}\eta(t) \tag{3}$$

with any process (x,u) so that $t \to \phi(t,x(t),u(t))$ is integrable and $t \to \psi(t,x(t))$ has finite p-variation for some $p < 2$ (this ensures that the second integral in (3) is defined as a Young integral, as we will see later). The functions ϕ and ψ will be called the *deterministic* and the *stochastic running costs*, respectively, and the first and second integral in the cost functional (3) we will call the *deterministic* and the *stochastic costs*, respectively.

We will properly define the p–variation and the Young integral in Section 2 and summarise a few properties central to our analysis in Lemmas 2-5. In particular, we obtain that for any given observation path η, the integral $\int_I \psi(t,x(t))\,\mathrm{d}\eta(t)$ is well defined and finite whenever x is a solution of the state equation (1). In particular, attempts to calculate this integral using Stratonovič or Itô style partitions will give the same result. In view of this, we can forget about the stochastic character of the observations and instead approach data assimilation pathwise for each realisation of the observations.

We also stress that the control u is *not* required to be non–anticipating. The stochastic costs (i.e. the second term in Eqn.3) are well defined nonetheless, as we will see.

The final ingredient we add to our problem is a *control set* $U \subset E$, interpreted as the set of permitted values for our control function u. A control function $u : I \to E$ will be called *feasible* with respect to ξ (the initial condition for x) if u is measurable, $u(t) \in U$ for almost all $t \in I$ and there is a function x so that (x,u) is a process with respect to ξ. In this case, the pair (x,u) will be called a *feasible process* with respect to ξ. A feasible process (x,u) with respect to ξ will be called an *admissible process* with respect to ξ if it has finite costs; a control u which is part of an admissible process will be called an *admissible control* with respect to ξ. The qualifier "with respect to ξ" might be omitted if clear from the context.

We are now ready to formulate our problem statement:

Problem (Variational Data Assimilation—VAR)**.** *Minimise the cost functional*

$$A(x,u) = \int_I \phi(t,x(t),u(t))\,\mathrm{d}t + \int_I \psi(t,x(t))\,\mathrm{d}\eta(t)$$

subject to

$$\dot{x}(t) = f(t,x(t)) + g(t,x(t))u(t) \qquad \text{for a.a. } t \in I \tag{4}$$

$$u(t) \in U \qquad \text{for a.a. } t \in I \tag{5}$$

$$x(0) = \xi, \tag{6}$$

that is, over all admissible processes with respect to ξ*.*

The remainder of this section will be devoted to presenting our main hypotheses and results. (In special cases, it is possible to obtain part of these results using standard methods of optimal control, albeit under substantially stronger conditions that preclude important applications in the geosciences; see the end of the present section for more details.) For functions on I, we use the norms

$$\|u\|_r := \left(\int_I |u(s)|^r\,\mathrm{d}s\right)^{1/r} \qquad \text{and} \qquad \|u\|_\infty := \sup_{s\in I} |u(s)|$$

For $r > 1$, we define $\mathcal{U}_r$ as the set of all measurable control functions $u : I \to E$ with $u(t) \in U$ almost surely and $\|u\|_r < \infty$.

Hypothesis 1. *(a) f, g are continuous.*

(b) For all $R \geq 0$ the stochastic running costs ψ are Hölder on $I \times \{x \in E; |x| \leq R\}$ with constant K_R and exponent $\kappa > \frac{1}{2}$. Further

$$K_R \leq C_\psi(1 + R^\alpha) \tag{7}$$

for some nonegative constants C_ψ, α.

(c) U is closed and convex.

(d) The deterministic running costs ϕ are continuous on $I \times E \times U$, convex in u and satisfies a lower bound of the form $\phi(t,x,u) \geq C_\phi(|u|^r - |x|^\delta)$ for constants $C_\phi > 0$, $\delta > 0$ and $r > 1$.

(e) There exists $\xi \in E$ and nonegative constants C_s, C_e, β, γ so that for any control $u \in \mathcal{U}_r$ and any local solution x for the state equation (1) *with control u and initial condition ξ we have*

$$\|\dot{x}\|_r \leq C_s(1 + \|u\|_r^\beta) \tag{8}$$

$$\|x\|_\infty \leq C_e(1 + \|u\|_r^\gamma). \tag{9}$$

Further, $\gamma(\alpha + \kappa - \frac{1}{2}) + \frac{\beta}{2} < r$, and $\gamma\delta < r$.

Hypothesis 1(b) represents growth and regularity conditions on the stochastic costs; the estimate (8) quantifies the nonlinearity in the state equation, while estimate (9) might represent an energy estimate. By local solution in Hypothesis 1(e) we mean that x is a solution of the state equation (1) but possibly only over a smaller interval $[0, T_1]$ with $T_1 \leq T$. Note that we can assume without loss of generality that $\gamma \leq \beta$, since the estimate (8) always implies an energy estimate (9) with $\gamma = \beta$. We stress that these conditions might be satisfied even if f fails to exhibit linear growth. To illustrate this, we present a simple

Example (Lorenz'63 system)*.* This example belongs to a wider class of problems, to be discussed in Section 6 and motivated by data assimilation in geophysical fluid dynamics, in which the state equation has a characterisic quadratic nonlinearity, so $\beta = 2$, but $\gamma = 1$ due to energy conservation, and this will turn out to be crucial.

Our example is the Lorenz'63 system in the form given in [17]. The vector field has the form $f = f_1 + f_2$, where

$$\begin{aligned} f_1(x,y,z) &= \begin{pmatrix} -\sigma x + \sigma y \\ -\sigma x - y \\ -bz - b(r+\sigma) \end{pmatrix}, \\ f_2(x,y,z) &= \begin{pmatrix} 0 \\ -xz \\ xy \end{pmatrix}, \end{aligned} \tag{10}$$

with σ, r, b positive parameters (see [17] for their interpretation). Typical values for these parameters are $\sigma = 10, r = 28$, and $b = 8/3$. Note that f_2 is quadratic while f_1 is linear and stable. We complement this by setting $g = \mathbb{1}$ and putting no further constraints on the control, that is we set $U = \mathbb{R}^3$. Further, we consider a cost functional of the form (2) with $h(x,y,z) = x, R = 1$, and $S = \mathbb{1}$ so that $\phi(x,y,z,u) = \frac{1}{2}x^2 + \frac{1}{2}u^2$ and $\psi(x,y,z) = -x$. With regards to Hypothesis 1, only item (e) is not obvious. However, since $(x,y,z)^t f_2(x,y,z) = 0$ and f_1 is stable, the energy estimate (9) in Hypothesis 1(e) follows with $\gamma = 1$ from an application of the Bellmann–Grönwall Lemma (see [17] and Sec. 6 for more details). Using this bound on $\|x\|_\infty$ directly in the state equation, we obtain the nonlinearity estimate (8) in Hypothesis 1(e) with $\beta = 2$, due to f being quadratic in leading order, finishing the discussion of this Example.

As another note on Hypothesis 1, it might seem at first sight that Hypothesis 1(b) as well as the lower bound on r in Hypothesis 1(e) both become *less* restrictive for smaller κ. It has to be kept in mind though that a smaller Hölder exponent κ in general implies a larger Hölder constant K_R and, in our context, a larger α in the estimate (7).

Our first result ensures that under Hypothesis 1, Problem VAR is well defined.

Lemma 1. *Under Hypothesis 1, the following holds:*

- *(a) For any $u \in \mathcal{U}_r$, the state equation has a (not necessarily unique) solution x with $x(0) = \xi$.*
- *(b) If (x,u) is feasible, the stochastic part of the cost functional is well defined and finite.*
- *(c) If (x,u) is feasible, the function $t \to \phi(t, x(t), u(t))$ is measurable and bounded below by an integrable function; hence the deterministic costs are well defined as an element of $\mathbb{R} \cup \{\infty\}$.*
- *(d) If (x,u) is admissible, then $\|u\|_r < \infty$.*

In view of Lemma 1(a), every element $u \in \mathcal{U}_r$ is feasible with respect to ξ, and by item (d), every admissible u must lie in $\mathcal{U}_r$. By items (b, c), a feasible process is admissible if and only if the deterministic costs are finite. In particular, admissibility does not depend on the particular observation path. The following is our main existence result:

Theorem 1. *Suppose that Hypothesis 1 is in force and that there is an admissible process (x_0, u_0) with repect to ξ specified in Hypothesis 1. Then for any observation path $\{\eta_t, t \in I\}$, Problem VAR admits a solution (x_*, u_*). Further, $\|u_*\|_r < \infty$.*

Both Lemma 1 and Theorem 1 will be proved in Section 2.

In Section 3, we will prove a Maximum Principle (Thm. 2). For the Maximum Principle, we need to strengthen our Hypothesis. (In the following, partial derivatives with respect to the k'th argument will be written as D_k, so $\mathrm{D}_2\phi$ for instance is the partial derivative of ϕ with respect to the second argument. Further, $\mathbb{R}_{\geq 0}$ denotes the nonnegative real numbers.)

Hypothesis 2. *Hypothesis 1 and*

(a) *f, g have continuous partial derivatives with respect to the second argument.*
(b) *ϕ has a continuous partial derivative with respect to the second argument.*
(c) *There exists constants $R > 0$, $c \geq 0$ and an integrable function $d : I \to \mathbb{R}_{\geq 0}$ so that whenever (x, u) is an admissible process and $|y - x(t)| \leq R$, then $|\mathrm{D}_2\phi(t, y, u(t))| \leq c|\phi(t, y, u(t))| + d(t)$ almost surely.*
(d) *ψ has a partial derivative with respect to the second variable which is locally Hölder in (t, x) with exponent larger than $\frac{1}{2}$.*

Theorem 2 (Maximum Principle)**.** *Assume Hypothesis 2 (including Hypothesis 1 for a certain $r > 1$), and fix an initial condition ξ as well as an observation η. For these data, let (x, u) be a global minimiser for Problem VAR. Then for almost all $t \in I$ the control u satisfies the condition*

$$H(t, x(t), \lambda(t), u(t)) = \inf_{v \in U} H(t, x(t), \lambda(t), v) \tag{11}$$

where the function $\lambda : I \to E'$ is the unique solution to the integral equation

$$\lambda(t) = \int_t^T \lambda(s)\big\{\mathrm{D}_2 f(s, x(s)) + \mathrm{D}_2 g(s, x(s))u(s)\big\}\,\mathrm{d}s + \int_t^T \mathrm{D}_2\phi(s, x(s), u(s))\,\mathrm{d}s + \int_t^T \mathrm{D}_2\psi(s, x(s))\,\mathrm{d}\eta_s \tag{12}$$

and $H : I \times E \times E' \times U \to \mathbb{R}$ is defined through

$$H(t, x, \lambda, v) := \phi(t, x, \lambda, v) + \lambda\big(f(t, x) + g(t, x)v\big)$$

The function λ is referred to as the *costate*, and E' denotes the dual space of E. The third integral in Equation (12) is to be interpreted as the element of E' given by the linear function

$$E \to \mathbb{R}; z \to \int_t^T \mathrm{D}_2\psi(s, x(s))z\,\mathrm{d}\eta_s.$$

If (x, u) is an optimal pair for problem VAR with respect to initial condition ξ and if λ is a corresponding costate (in the sense of Thm. 2), then we will refer to (x, u, λ) as an *optimal triple* with respect to initial condition ξ.

Note that Hypothesis 2(a) implies uniqueness to solutions of the state equation. The condition (c) in Hypothesis 2 ensures that the second integral in the definition of λ is well defined. In fact, for the Maximum Principle it would be sufficient to know that this condition holds along optimal processes (x, u).

In Sections 4 and 5, two applications of the Maximum Principle will be discussed. A precise formulation of the results will be given in those sections. The first result, Theorem 3, shows that under the conditions of the Maximum Principle and if $\phi(t, x, u)$ is strongly convex in u, an optimal control has finite p–variation for any $p > 2$. It turns out that the regularity of the controls is limited by the regularity of the observations, which cannot be improved beyond p–variation with $p > 2$. In particular, we do not expect optimal controls to be Lipschitz in general, as is typically the case in classical optimal control problems.

The second application (in Sec. 5) concerns the value function and its relation to uniqueness of optimal controls. The costs of any feasible control can be regarded as a function $J(\xi, u)$ on $E \times \mathcal{U}_r$, and the *value function* is defined as

$$V(\xi) = \inf_{u \in \mathcal{U}_r} J(\xi, u)$$

Theorem 5 then shows that (under suitable conditions) an optimal control u is unique for Problem VAR if and only if the value function has a derivative at ξ.

Finally, section 6 discusses an application to a typical problem in geophysical data assimilation. For this example, both a minimum energy cost function (known as weakly constrained 4d-VAR costfunction in the atmospheric sciences) as well as an Onsager–Machlup type cost function will be considered. For the latter case, we obtain the existence of minimisers under substantially weaker assumptions than in [21].

As has been mentioned already, special cases of Problem VAR can be approached using classical methods of optimal control, although stronger conditions are required in order to obtain results comparable to those presented here. Under Hypothesis 1 and if ψ has a continuous derivative with respect to (t, x), the stochastic costs can be integrated by parts and one obtains a standard control problem with running costs given by

$$\begin{aligned}\tilde{\phi}(t,x,u) &= \phi(t,x,u) \\ &\quad - \big\{ \mathrm{D}_1\psi(t,x) + \mathrm{D}_2\psi(t,x)\,(f(t,x)+g(t,x)u)\,\big\}\eta(t).\end{aligned} \tag{13}$$

This approach has been taken in [21, 22]. But in addition to the existence of the derivatives of ψ, some coercivity property of $\tilde{\phi}$ is required to obtain weakly compact level sets, and even if Hypothesis 1(d) is imposed, further assumptions such as linear growth of f are needed to make sure that for large u, the second term in Equation (13) does not override the first one. The approach presented here however does not require linear growth of f, which is in fact not present in typical geophysical models (Lorenz'63 being an example; see Sec. 6 for further discussion). Our approach instead relies on a more careful analysis of the stochastic costs, using the theory of p–variation and Young integrals.

2. Proof of existence result

Before proving Lemma 1 and the existence result Theorem 1, we will discuss the notions of p–variation and Young integrals. We will not give proofs as these are standard facts or easy modifications thereof, and references will be provided below.

Definition 1. Let $I \subset \mathbb{R}$ be an interval and V, W be finite dimensional vector spaces, with $|.|$ denoting a generic norm.

(a) Let $x \in C(I, V)$ and $p \geq 1$. Then the *p–variation* of x is defined as

$$[x]_p := \left(\sup \sum_i |x(t_{i+1}) - x(t_i)|^p \right)^{1/p}$$

where the sup is taken over all finite dissections $D = \{0 = t_0 < t_1 < \ldots < t_n = T\}$ of I.

(b) Let $x \in C(I, L(V, W))$ and $y \in C(I, V)$. The Young integral of x against y is given by

$$\int_I x(s)\,\mathrm{d}y(s) := \lim \sum_i x(\tau_i)(y(t_{i+1}) - y(t_i))$$

where the sum runs over a finite dissection of I and $\tau_i \in [t_i, t_{i+1}]$ for all i, and the limit is with respect to the resolution $|D| := \sup_i |t_{i+1} - t_i|$ of the dissections going to zero. It is required that the existence of the limit and its value does not dependent on the choice of the dissections nor on the precise placement of the τ_i.

Lemma 2. *Let $x \in C(I, L(V, W))$ and $y \in C(I, V)$, and suppose there are $p, q \geq 1$ with $\theta := \frac{1}{p} + \frac{1}{q} > 1$ and so that both $[x]_p$ and $[y]_q$ are finite. Then*

(a) the Young integral of x against y exists and

$$|\int_I x(s)\,\mathrm{d}y(s) - x(0)(y(T)-y(0))| \leq \frac{1}{1-2^{1-\theta}}\,[x]_p\,[y]_q\,,$$

(b) the Young integral permits integration by parts:

$$\int_I x(s)\,\mathrm{d}y(s) = x(T)y(T) - x(0)y(0) - \int_I y(s)\,\mathrm{d}x(s).$$

For the first and second part of this Lemma, see [7], Theorem 6.8 and Exercise 6.14, respectively. In order to make sense of the integration by parts formula, we note that every element y of V gives rise to a linear mapping $\hat{y} : L(V,W) \to W, A \to Ay$, and the integral on the right hand side of the integration by parts formula is to be understood as $\int_I \hat{y}(s)\,\mathrm{d}x(s)$. With slight abuse of notation however, we will continue to write y instead of $\hat{y}$.

Lemma 3. *Let $x \in C(I, L(V,W))$ and $y \in C(I,V)$*

(a) Suppose y is absolutely continuous, then $[y]_p \leq \|\dot{y}\|_1$ for all $p \geq 1$.

(b) If $p > q$, then $[y]_p \leq [y]_q$.

(c) (Interpolation inequality) If $p > q$, then

$$[y]_p \leq \Big([y]_q\Big)^{\frac{q}{p}} \left(\sup_{t,s\in I}|y(t)-y(s)|\right)^{1-\frac{q}{p}}.$$

(d) (Product rule)

$$[xy]_p \leq [x]_p\,\|y\|_\infty + [y]_p\,\|x\|_\infty$$

(e) (Chain rule) Suppose that the function $\psi : I\times V \to W$ is Hölder on $I\times\{x \in V; |x| \leq \|y\|_\infty\}$ with exponent κ and constant K, and that $\kappa p \geq 1$, then

$$[\psi(.,y(.))]_p \leq K([y]^\kappa_{\kappa p} + |I|^\kappa)$$

The first item follows directly from the definition of p–variation; for a proof of items (b,c), see [7], Proposition 5.3, Proposition 5.5, respectively; items (d,e) are easy consequences of the definition of p–variation.

Lemma 4. *Let $y \in C(I,V)$ and x_n in $C(I,L(V,W))$ for all $n \in \mathbb{N}$, with $\sup_n [x_n]_p$ as well as $[y]_q$ being finite and $\frac{1}{p}+\frac{1}{q} > 1$. Then $x_n \to x$ uniformly implies*

$$\int_0^t x_n(s)\,\mathrm{d}y(s) \to \int_0^t x(s)\,\mathrm{d}y(s)$$

uniformly as a function of the upper limit t.

This is a consequence of Proposition 6.13 in [7].

Lemma 5. *For any $p > 2$, the p–variation of the observations $[\eta]_p$ is a measurable and almost surely finite random variable.*

Proof. We are assuming $\eta = \zeta + W$, and since ζ is absolutely continuous, it has finite variation of any order by Lemma 3(a). It is well known that almost any path of the Wiener process is $\frac{1}{p}$–Hölder for any $p > 2$ on compact intervals, which implies finite p–variation for $p > 2$. It follows directly from the definition and Minkowski's inequality that p–variation is subadditive, hence $[\eta]_p$ is almost surely finite. □

In view of Lemma 5, we can find a set $\Omega_0 \subset \Omega$ with $\mathbb{P}(\Omega_0) = 1$ so that whenever $\omega \in \Omega_0$, we have $[\eta]_p < \infty$ for any $p > 2$. Without loss of generality, we will henceforth assume $\Omega = \Omega_0$. In other words, we can simply assume the observations to be a continuous function η having finite p-variation for any $p > 2$. This allows us to forget about the stochastic character of the observations and instead approach

data assimilation pathwise for each realisation of the observations. In particular, we obtain from Lemma 2 that for any given observation path, the integral $\int_I y(t)\,\mathrm{d}\eta(t)$ is well defined and finite for all functions y which have finite p-variation with $p < 2$.

Proof of Lemma 1. Item (a) is a basic result in the theory of ODE's. To prove item (b) use the remark made just prior to this proof and the fact that $t \to \psi(t, x(t))$ has bounded $\frac{1}{\kappa}$–variation with $\kappa < \frac{1}{2}$ (as we will show in the proof of Theorem 1). Item (c) follows from Hypothesis 1(d), the fact that x is continuous and hence bounded on I, and the fact that a continuous function of measurable functions is measurable. Item (d) follows because if (x, u) is feasible with $\|u\|_r = \infty$, the lower bound on ϕ in Hypothesis 1(d) yields that the deterministic costs are infinite. □

The following Lemma will be essential in the proof of Theorem 1, ensuring the coercivity of the stochastic costs:

Lemma 6. *Suppose that there are nonegative constants $r, q, \gamma_1, \gamma_2, C_1, C_2$ with $r \geq 1$ and $1 \leq q < 2$ so that for any feasible process (x, u) we have*

$$\|\psi(., x(.))\|_\infty \leq C_1(1 + \|u\|_r^{\gamma_1}) \tag{14}$$

$$[\psi(., x(.))]_q \leq C_2(1 + \|u\|_r^{\gamma_2}). \tag{15}$$

If we let

$$\theta_1 := \min\{\gamma_1(1 - \frac{q}{2}) + \gamma_2\frac{q}{2}, \gamma_2\}$$

$$\theta_2 := \max\{\gamma_1(1 - \frac{q}{2}) + \gamma_2\frac{q}{2}, \gamma_2\}$$

then for any θ with $\theta_1 < \theta < \theta_2$ there is a random variable $D_\theta \geq 0$ depending on θ and the observations η only so that

$$|\int_I \psi(t, x(t))\,\mathrm{d}\eta_t| \leq D_\theta\left(1 + \|u\|_r^\theta\right).$$

Proof. Fix a θ as stated and set τ so that $\theta = \gamma_1(1 - \tau) + \gamma_2\tau$. Then $\frac{q}{2} < \tau < 1$, and if we define $p = \frac{q}{\tau}$, we obtain $q < p < 2$, so we can assume that $p = \frac{2}{1+2\epsilon}$ for some $\epsilon > 0$. Put $\hat{p} = \frac{2}{1-\epsilon} > 2$ and observe that $\frac{1}{\hat{p}} + \frac{1}{p} = \frac{1-\epsilon}{2} + \frac{1+2\epsilon}{2} = 1 + \epsilon > 1$. In view of Lemma 5, we can invoke Lemma 2 for $[\psi(., x(.))]_p$ and $[\eta]_{\hat{p}}$ which gives that

$$|\int_I \psi(t, x(t))\,\mathrm{d}\eta_t| \leq C_p\left(1 + [\psi(., x(.))]_p\right)$$

for some random variable C_p which depends on the observations η and also on the choice of p. Next, the interpolation inequality (Lemma 3.c) yields

$$|\int_I \psi(t, x(t))\,\mathrm{d}\eta_t| \leq C_p\left(1 + [\psi(., x(.))]_q^{\frac{q}{p}}\,\|\psi(., x(.))\|_\infty^{1-\frac{q}{p}}\right).$$

Now use that $p = \frac{q}{\tau}$ as well as the estimates (14,15) to complete the proof. □

Note that bounds on the stochastic costs could also be obtained through integration by parts, as already mentioned in the introduction. This requires ψ to have a derivative with respect to x to begin with, but even if this derivative is bounded, then under the assumed conditions the stochastic term is bounded by the nonlinearity estimate (8), which can be much worse than the estimates given by Lemma 6. We are now ready to prove Theorem 1.

Proof of Theorem 1. Let $(x_n, u_n), n = 0, 1, \ldots$ be a minimising sequence with $\|u_n\|_r < \infty$ for all n and (x_0, u_0) being the admissible process which exists by assumption.

We know that the costs $A(x_n, u_n)$ are bounded above (by $A(x_0, u_0) < \infty$). Hypothesis 1.(d) implies that for any process

$$A(x,u) \geq C_\phi \int_I \left(|u(t)|^r - |x(t)|^\delta \right) \, \mathrm{d}t + \int_I \psi(t, x(t)) \, \mathrm{d}\eta_t, \tag{16}$$

and using the energy estimate (9) and taking into account that $\gamma\delta < r$ (Hyp. 1,(e)), we obtain for the first integral

$$\int_I \left(|u(t)|^r - |x(t)|^\delta \right) \, \mathrm{d}t \geq C \|u(t)\|^r - b \tag{17}$$

for some $C > 0$ and $b \in \mathbb{R}$.

We aim to control the second integral in Equation (16) using Lemma 6 with $q = 1/\kappa$. By Hypothesis 1,(b) and the chain rule in Lemma 3 we obtain

$$[\psi(., x(.))]_q \leq K_{\|x\|_\infty} C'(1 + [x]_1^\kappa)$$

for some constant C'. Invoking Lemma 3,(a) as well as the estimates (7,8,9) we obtain

$$[\psi(., x(.))]_q \leq C''(1 + \|u\|_r^{\beta\kappa + \gamma\alpha}). \tag{18}$$

Again by Hypothesis 1,(b) and the estimate (7) we get

$$\|\psi(., x(.))\|_\infty \leq C'''(1 + \|x\|_\infty^{\kappa+\alpha}),$$

and invoking the energy estimate (9) once again this becomes

$$\|\psi(., x(.))\|_\infty \leq C''''(1 + \|u\|_r^{\gamma(\kappa+\alpha)}). \tag{19}$$

In equations (18,19), we can use Lemma 6 with $q := 1/\kappa$, $\gamma_1 := \gamma(\kappa + \alpha)$ and $\gamma_2 := \beta\kappa + \gamma\alpha$. We obtain $\theta_1 = \gamma(\alpha + \kappa - \frac{1}{2}) + \frac{\beta}{2}$ and $\theta_2 = \beta\kappa + \gamma\alpha$ (note that indeed $\theta_1 \leq \theta_2$ because $\gamma \leq \beta$ as discussed just after Hypothesis 1). Now according to the lower bound on r in Hypothesis 1,(e), we can find θ so that $\theta_1 < \theta = r - \epsilon$ for some $\epsilon > 0$ and therefore by Lemma 6 there exists a constant D only depending on $r - \epsilon$ and the observations so that

$$\left| \int_I \psi(t, x(t)) \, \mathrm{d}\eta_t \right| \leq D \left(1 + \|u\|_r^{r-\epsilon} \right).$$

Using this as well as the estimate (17) in Equation (16), we are finally able to find a constant C_A so that

$$A(x,u) \geq C_A(1 + \|u\|_r^r). \tag{20}$$

Since $A(x_n, u_n)$ is bounded, we obtain from Equation (20) that $\|u_n\|_r$ must be bounded. Hypothesis 1.(e) then implies that $\|x_n\|_\infty$ and $\|\dot{x}_n\|_r$ are bounded.

Since L_r is reflexive, by weak compactness, we can assume (taking subsequences which we do not relabel) that $u_n \to u_*$ and $\dot{x}_n \to v_*$. The boundedness of $\|\dot{x}_n\|_r$ implies that x_n is equicontinuous, so applying Arzela–Ascoli (and again taking subsequences), we can assume $x_n \to x_*$ uniformly as well. (Note that a subsequence of a minimising sequence is still a minimising sequence.) Further

$$x_*(t) = \xi + \int_0^t v_*(s) \, \mathrm{d}s$$

so $v_* = \dot{x}_*$.

Since the functions x_n are uniformly bounded, we obtain from Hypothesis 1(d) that the functions $t \to \phi(t, x_n(t), u_n(t))$ admit a uniform lower bound ϕ_0. We can therefore replace ϕ with $\max\{\phi, \phi_0\}$ on $I \times E \times U$ without affecting continuity or convexity. We can now apply the Integral semicontinuity theorem of [2], (Thm. 6.38) to conclude that

$$\int_I \phi(t, x_*(t), u_*(t)) \, \mathrm{d}t \leq \liminf \int_I \phi(t, x_n(t), u_n(t)) \, \mathrm{d}t \tag{21}$$

The continuity of ψ together with the uniform convergence of x_n implies that $\psi(.,x_n(.)) \to \psi(.,x_*(.))$ uniformly on I. The boundedness of $[\psi(.,x_n(.))]_q$ follows from the boundedness of $\|u_n\|_r$ and the estimate (18). It then follows from Lemma 4 that

$$\int_I \psi(t,x_n(t))\,\mathrm{d}\eta(t) \to \int_I \psi(t,x_*(t))\,\mathrm{d}\eta(t)$$

This fact together with (21) implies

$$\begin{aligned}
&\int_I \phi(t,x_*(t),u_*(t))\,\mathrm{d}t - \int_I \psi(t,x_*(t))\,\mathrm{d}\eta(t)\\
&\leq \liminf \int_I \phi(t,x_n(t),u_n(t))\,\mathrm{d}t - \lim \int_I \psi(t,x_n(t))\,\mathrm{d}\eta(t)\\
&\leq \liminf \left\{ \int_I \phi(t,x_n(t),u_n(t))\,\mathrm{d}t - \int_I \psi(t,x_n(t))\,\mathrm{d}\eta(t) \right\}\\
&= \inf A(x,u)
\end{aligned}$$

where the last inf is over all admissible trajectories with $\|u\|_r < \infty$ and the last equality follows because (x_n,u_n) is a minimising sequence. It remains to show that (x_*,u_*) is a feasible process. The proof is the same as in [2], Theorem 23.11. □

3. A Maximum Principle

The aim of this section is to prove that under suitable conditions, global minimisers of Problem VAR satisfy a set of necessary conditions akin to the celebrated Pontryagin Maximum Principle in optimal control, see for instance [5], Sec.I.6. The classical Pontryagin Maximum Principle requires the problem to exhibit certain regularity, in particular f and g need to be more regular than we have assumed so far. Recently, there has been considerable progress in relaxing these conditions, at the price of having to use more general notions of derivatives as well as some heavy nonsmooth analysis machinery, see [1], Sec.III.3.4. and in particular [2], Theorem 22.26.

Taking one step at a time, we will prove the Maximum Principle under essentially classical regularity assumptions, but admitting systems bearing the essential features of the example in Section 6. Our main problem will again be dealing with the stochastic part of the costs in an appropriate manner. We note that here again, this can be attempted by integrating the stochastic costs by parts, which is now permitted in view of Hypothesis 2(d). This results in a classical optimal control problem, but applying the standard Maximum Principle then requires (among other things) that the running costs of this problem have one further derivative with respect to x, which is not guaranteed by our assumptions. The following proof however shows that the stochastic costs can be dealt with directly.

Proof of Theorem 2. We will focus on the bits relevant for the stochastic costs which are nonstandard, sketching the more standard parts of the proof. A few technical details will be left to subsequent Lemmas. As in the proof of Theorem 2.1 in [19], Chapter 2, and the proof of Theorem 6.3 in [5], Chapter I, we will use a needle variation. Let $s \in I \setminus \{T\}$ be a Lebesgue point for u. (Recall that s is a Lebesgue point for u if $\lim_{\epsilon\to 0} \frac{1}{2\epsilon}\int_{[-\epsilon,\epsilon]} u(s+t)\,\mathrm{d}t = u(s)$, and that if u is integrable over I, almost every $s \in I$ is a Lebesgue point of u.) For some $v \in U$ and $\epsilon > 0$ so that $s+\epsilon \in I$, define the interval $I_\epsilon := [s, s+\epsilon]$ and the control

$$u_\epsilon(t) = \begin{cases} v & \text{if } t \in I_\epsilon \\ u(t) & \text{else.} \end{cases}$$

Note that $u_\epsilon \in \mathcal{U}_r$. Let x_ϵ be the (unique) solution to the state equation (1) with $u = u_\epsilon$ and $x_\epsilon(0) = \zeta$. To be able to write the state equation in a more compact form in the following, we introduce the function $h(t, x, u) := f(t, x) + g(t, x)u$. Using this abbreviation we can write

$$\dot{x}_\epsilon(t) = h(t, x_\epsilon(t), u(t)) + g(t, x_\epsilon(t))(v - u(t))\mathbb{1}_{I_\epsilon}(t).$$

We now consider a Taylor expansion of the state equation and the costs to first order in the perturbation $t \to g(t, x(t))(v - u(t))\mathbb{1}_{I_\epsilon}(t)$. To this end, we define

$$\begin{aligned}\dot{\zeta}_\epsilon(t) &= \mathrm{D}_2 h(t, x(t), u(t))\zeta_\epsilon(t) + g(t, x(t))(v - u(t))\mathbb{1}_{I_\epsilon}(t),\\ \zeta_\epsilon(0) &= 0.\end{aligned}$$

We make the following claims:

$$\|x_\epsilon - x\|_\infty = O(\epsilon), \tag{22}$$

$$\|x_\epsilon - x - \zeta_\epsilon\|_\infty = o(\epsilon), \tag{23}$$

and

$$\begin{aligned}&A(x_\epsilon, u_\epsilon) - A(x, u)\\ &= \int_I \mathrm{D}_2\phi(t, x(t), u(t))\zeta_\epsilon(t)\,\mathrm{d}t\\ &\quad + \int_I \mathrm{D}_2\psi(t, x(t))\zeta_\epsilon(t)\,\mathrm{d}\eta_t\\ &\quad + \int_{I_\epsilon} \phi(t, x(t), v) - \phi(t, x(t), u(t))\,\mathrm{d}t + o(\epsilon).\end{aligned} \tag{24}$$

The proof of these claims is virtually identical to the proof of Lemma 2.2 in Chapter 2 of [19], with the exception of the stochastic cost contribution to Equation (24), which will be considered in Lemma 7.

As will be shown in Lemma 8, the costate equation (12) has a unique solution λ which has finite q–variation for any $q > 2$. Similar to the classical Maximum Principle, the functions ζ_ϵ and λ are in duality, leading to the identity

$$\begin{aligned}&\int_I \lambda(t)g(t, x(t))(v - u(t))\mathbb{1}_{I_\epsilon}(t)\,\mathrm{d}t\\ &= \int_I \mathrm{D}_2\phi(t, x(t), u(t))\zeta_\epsilon(t)\,\mathrm{d}t + \int_I \mathrm{D}_2\psi(t, x(t))\zeta_\epsilon(t)\,\mathrm{d}\eta_t.\end{aligned}$$

This will be proved in Lemma 9. We use this relation as well as the definition of the Hamiltonian H in Equation (24) and invoke the optimality of u to obtain

$$\begin{aligned}0 \le &\int_I \mathbb{1}_{I_\epsilon}(t)\left\{H(t, x(t), \lambda(t), v) - H(t, x(t), \lambda(t), u(t))\right\}\,\mathrm{d}t\\ &+ o(\epsilon).\end{aligned}$$

The result now follows if we divide by ϵ and send $\epsilon \to 0$, keeping in mind that s is a Lebesgue point of u □

Lemma 7. *With the definitions as in the proof of Theorem 2, we have*

$$\begin{aligned}&\int_I \psi(t, x_\epsilon(t))\,\mathrm{d}\eta_t - \int_I \psi(t, x(t))\,\mathrm{d}\eta_t\\ &- \int_I \mathrm{D}_2\psi(t, x(t))\zeta_\epsilon(t)\,\mathrm{d}\eta_t\\ &= o(\epsilon).\end{aligned}$$

Proof. According to Lemma 4, this follows if

$$\frac{1}{\epsilon}\big[\psi(t,x_\epsilon(t)) - \psi(t,x(t)) - \mathrm{D}_2\psi(t,x(t))\cdot\zeta_\epsilon(t)\big] \longrightarrow 0$$

uniformly in t, and for some $q < 2$ the q–variation of the left hand side stays bounded.

By Taylor's theorem, we can write the left hand side as

$$\begin{aligned}&\int_0^1 \mathrm{D}_2\psi\big(t,\tau x_\epsilon(t) + (1-\tau)x(t)\big) - \mathrm{D}_2\psi\big(t,x(t)\big)\,\mathrm{d}\tau\cdot\frac{1}{\epsilon}\big[x_\epsilon(t)-x(t)\big]\\ &+\mathrm{D}_2\psi\big(t,x(t)\big)\frac{1}{\epsilon}\big[x_\epsilon(t)-x(t)-\zeta_\epsilon(t)\big].\end{aligned} \tag{25}$$

Now using Equation (22) and the fact that $\mathrm{D}_2\psi$ is uniformly continuous (by Hypothesis 2(d)), we obtain that the first term goes to zero uniformly. Equation (23) guarantees that the second term goes to zero uniformly as well.

The proof of Equation (22) reveals, upon closer inspection, that the L_r–norm of $\frac{1}{\epsilon}\frac{\mathrm{d}}{\mathrm{d}t}[x_\epsilon - x]$ remains bounded, and the same is true for $\frac{1}{\epsilon}\frac{\mathrm{d}}{\mathrm{d}t}\zeta_\epsilon$. By Lemma 3(a) this implies that the q–variation of these functions remains bounded for any $q \geq 1$. By Lemma 3(d) this leaves to show that also the term

$$\int_0^1 \mathrm{D}_2\psi\big(t,\tau x_\epsilon(t) + (1-\tau)x(t)\big)\,\mathrm{d}\tau$$

in expression (25) has q–variation bounded in ϵ. Define the function $y_\epsilon(\tau,t) = \mathrm{D}_2\psi\big(t,\tau x_\epsilon(t) + (1-\tau)x(t)\big)$. By Hypothesis 2,(d), the q–variation of $t \to y_\epsilon(\tau,t)$ is bounded uniformly in ϵ and τ, for $q < \frac{1}{2}$. But it is easily seen that if $[y_\epsilon(\tau,.)]_q \leq C$, then $t \to \int_0^1 y_\epsilon(\tau,t)\,\mathrm{d}\tau$ has q–variation bounded by C as well. We conclude that as a function of t, the expression in Equation (25) has q–variation bounded in ϵ. □

Lemma 8. *There is a unique solution to the costate equation* (12) *which has finite q–variation for $q > 2$.*

Proof. According to Hypothesis 2(c) the function $t \to \mathrm{D}_2\phi(t,x(t),u(t))$ is integrable. Further, from Hypothesis 2(d), along with the absolute continuity of x and the chain rule in Lemma 3, we can conclude that $t \to \mathrm{D}_2\psi(t,x(t))$ has q–variation for some $q < 2$. These two facts imply that the second and third integral in the definition of λ (Eqn. 12) are well defined and continuous as a function of the lower limit. We can thus write Equation (12) in the form

$$\lambda(t) = \int_t^T \lambda(s)M(s)\,\mathrm{d}s + F(t)$$

where $M(t) := \mathrm{D}_2 f(t,x(t)) + \mathrm{D}_2 g(t,x(t))u(t)$ and F is continuous with finite q–variation for $q > 2$. Noting that $|M|$ is integrable, it follows by standard ODE arguments that λ is well defined and unique. Further, λ is the sum of an absolutely continuous part and a part with finite q–variation for $q > 2$. From this, it is easy to see that λ itself has finite q–variation for $q > 2$. □

Lemma 9 (Duality Lemma)**.** *Let $I = [0,T]$ be an interval and V a finite dimensional vector space with norm $|.|$ and dual space V'. Consider functions $a : I \to V$ and $b : I \to V'$ so that both $[a]_p$ and $[b]_q$ are finite, where $\frac{1}{p} + \frac{1}{q} > 1$. Further, consider a measurable function $M : I \to L(V,V)$ so that the function $t \to |M(t)|$ is*

integrable (here $|.|$ denotes the operator norm). Then the equations

$$\zeta(t) = \zeta(0) + \int_0^t M(s)\zeta(s)\,\mathrm{d}s + a(t),$$
$$\lambda(t) = \lambda(T) + \int_t^T \lambda(s)M(s)\,\mathrm{d}s + b(t)$$

have unique solutions, and both $[\zeta]_p$ and $[\lambda]_q$ are finite. Further, the duality relation

$$\lambda(T)\zeta(T) - \lambda(0)\zeta(0) = \int_I \zeta(t)\,\mathrm{d}b(t) + \int_I \lambda(t)\,\mathrm{d}a(t)$$

holds.

Proof. The uniqueness of ζ and λ as well as the claims about $[\zeta]_p$ and $[\lambda]_q$ follow as in the proof of Lemma 8. A direct calculation gives

$$\int_I \lambda(t)\,\mathrm{d}\zeta(t) = \int_I \lambda(t)M(t)\zeta(t)\,\mathrm{d}t + \int_I \lambda(t)\,\mathrm{d}a(t)$$

and similarly

$$\int_I \zeta(t)\,\mathrm{d}\lambda(t) = -\int_I \lambda(t)M(t)\zeta(t)\,\mathrm{d}t + \int_I \zeta(t)\,\mathrm{d}b(t).$$

Adding these two relations gives

$$\int_I \zeta(t)\,\mathrm{d}\lambda(t) + \int_I \lambda(t)\,\mathrm{d}\zeta(t) = \int_I \zeta(t)\,\mathrm{d}b(t) + \int_I \lambda(t)\,\mathrm{d}a(t),$$

and integrating by parts on the left hand side gives the duality relation. □

4. Regularity of controls

To investigate the regularity of controls, we assume that for every $(t, z, \mu) \in I \times E \times E'$ the function $H(t, z, \mu, .)$ has a unique minimiser, that is, there is a function $\upsilon : I \times E \times E' \to U$ so that

$$H(t, z, \mu, \upsilon(t, z, \mu)) = \inf_{u\in U} H(t, z, \mu, u).$$

By the characterisation (11) of optimal controls, regularity properties of υ can be translated into regularity properties of optimal controls. We observe that the regularity of υ is determined by the Hamiltonian H which neither contains the stochastic running costs nor the observations.

Definition 2. The deterministic running costs are *strongly convex* in u if for every bounded subset $C \subset I \times E \times U$ there is a $c > 0$ so that whenever (t, z, w_1) and (t, z, w_2) are in C, we have

$$(\mathrm{D}_3\phi(t, z, w_1) - \mathrm{D}_3\phi(t, z, w_2))\,(w_1 - w_2) \geq c|w_1 - w_2|^2.$$

Note that H is strongly convex if ψ is convex, meaning in particular that υ is well defined in this case (see also [2]).

Theorem 3 (Regularity of controls)**.** *Suppose Hypothesis 2 is in force.*

(a) If υ is well defined, then an optimal control has a continuous modification.
(b) If ψ is strongly convex, and $\mathrm{D}_3\phi$ as well as g are locally Lipschitz in (t, x), then an optimal control has a continuous modification that has bounded q–variation for any $q > 2$.

Proof. If (x, u, λ) is an optimal triple for initial condition ξ, then by the definition of υ we must have

$$u(t) = \upsilon(t, x(t), \lambda(t))$$

for almost all $t \in I$. Hence item a follows if we can show that υ is continuous, while establishing that υ is locally Lipschitz, together with Lemmas 8 and 3(e) will prove item b. Our proof of these two properties of υ follows [2], Theorem 23.17.

With w be an arbitrary element of U, we have the estimate

$$\begin{aligned} &H(t, z, \mu, w) \geq H(t, z, \mu, \upsilon(t, z, \mu)) \\ &\geq C_\phi |\upsilon(t, z, \mu)|^r - |z|^\delta + \mu \cdot [f(t, z) + g(t, z)\upsilon(t, z, \mu)] , \end{aligned}$$

and because f, g are continuous, we can infer that υ is bounded on bounded sets. Consider a sequence $\{(t_n, z_n, \mu_n), n \in \mathbb{N}\}$ so that $(t_n, z_n, \mu_n) \to (t, z, \mu)$. Taking any subsequence, there is a subsubsequence so that $\upsilon(t_n, z_n, \mu_n) \to w$ because this sequence is bounded. Then by definition of υ

$$H(t_n, z_n, \mu_n, \upsilon(t_n, z_n, \mu_n)) \leq H(t_n, z_n, \mu_n, \upsilon(t, z, \mu))$$

and taking limits we obtain

$$H(t, z, \mu, w) \leq H(t, z, \mu, \upsilon(t, z, \mu))$$

whence $w = \upsilon(t, z, \mu)$, and we can conclude that υ is continuous.

To prove Theorem 3(b), consider (t, z_1, μ_1) and (s, z_2, μ_2) in $I \times E \times E'$ and use the shorthands $w_1 := \upsilon(t, z_1, \mu_1)$ and $w_2 := \upsilon(s, z_2, \mu_2)$. Note that due to the convexity of H in u and the defintion of the function υ we must have

$$\begin{aligned} \mathrm{D}_4 H(t, z_1, \mu_1, w_1)(w_2 - w_1) \geq 0, \\ \mathrm{D}_4 H(s, z_2, \mu_2, w_2)(w_1 - w_2) \geq 0. \end{aligned}$$

The strong convexity condition implies

$$\begin{aligned} &c|w_1 - w_2|^2 \\ &\leq (\mathrm{D}_4 H(t, z_1, \mu_1, w_1) - \mathrm{D}_4 H(t, z_1, \mu_1, w_2)) (w_1 - w_2) \\ &= (\mathrm{D}_4 H(t, z_1, \mu_1, w_1) - \mathrm{D}_4 H(s, z_2, \mu_2, w_2) \\ &\quad + \mathrm{D}_4 H(s, z_2, \mu_2, w_2) - \mathrm{D}_4 H(t, z_1, \mu_1, w_2)) (w_1 - w_2) \\ &\leq (\mathrm{D}_4 H(s, z_2, \mu_2, w_2) - \mathrm{D}_4 H(t, z_1, \mu_1, w_2)) (w_1 - w_2) \\ &\leq L (|s - t| + |z_2 - z_1| + |\mu_2 - \mu_1|) |w_1 - w_2|. \end{aligned}$$

Hence, υ is Lipschitz, and by Lemma 3(e), together with the fact that x is absolutely continuous and that λ has q–variation for $q > 2$, we can conclude that the control also has q–variation for $q > 2$. □

5. The value function

In classical optimal control, there are well known relationships between the (generalised) derivative of the value function and the costate, permitting one to study the uniqueness of optimal controls, among other things (see for instance [6]). In this section, we will present a few results in this direction. Crucial to this analysis is the fact that if (x, u, λ) is an optimal triple for some $\xi \in E$, then (x, λ) satisfy a Hamiltonian ODE (Eqn. (26) in Thm. 4 below), which is obtained by using the minimum condition (11) to eliminate the control u from the state and costate equation. Further, it is required that solutions to Equation (26) are unique with respect to the initial condition $(x(0), \lambda(0))$. Currently, we are only able to prove this under the assumption that the function $(t, z) \to \mathrm{D}_2\psi(t, z)$ has continuous partial derivatives with respect to both arguments which are Lipschitz in z. We conjecture though

that weaker conditions might be sufficient; this will be investigated in a future paper. Given our incomplete understanding of these equations at this point, we have to impose these properties as assumptions, and pending a better understanding, the results in this section have to be regarded as preliminary.

Throughout this section, we fix a $\xi_0 \in E$ and assume there is an admissible control for ξ_0. Further we impose

Hypothesis 3. *Hypothesis 2 and*

(a) The mapping υ defined in Section 4 is well defined.
(b) There is a neighbourhood $X \subset E$ of ξ_0 so that Hypothesis 1(e) is valid for all $\xi \in X$.

We first show that the necessary conditions in Theorem 2 can be recast as a Hamiltonian system with stochastic perturbation. Consider the function

$$m(t,z,\mu) := \inf_{u\in U} H(t,z,\mu,u) = H(t,z,\mu,\upsilon(t,z,\mu))$$

which is continuous. From the fact that υ is well defined (i.e. H has a unique minimiser in u) and is bounded on bounded sets, and because H has continuous derivatives with respect to z,μ, we can conclude that also m has continuous derivatives with respect to z,μ given by

$$\mathrm{D}m(t,z,\mu) = \mathrm{D}H(t,z,\mu,u)|_{u=\upsilon(t,z,\mu)}$$

where D is the partial derivative with respect to z or μ. The proof of this well known fact is omitted. We immediately obtain the following theorem:

Theorem 4 (Hamiltonian system)**.** *Let (x,u,λ) be an optimal triple for initial condition ξ. Then the state and costate x,λ satisfy the equation*

$$\begin{aligned} \dot{x}(t) &= \mathrm{D}_3 m(t,x(t),\lambda(t)) \\ \lambda(t) &= \int_t^T \mathrm{D}_2 m(s,x(s),\lambda(s))\,\mathrm{d}s + \int_t^T \mathrm{D}_2\psi(s,x(s))\,\mathrm{d}\eta_s \\ x(0) &= \xi. \end{aligned} \tag{26}$$

To define the value function, one usually extends the Problem VAR to a class of problems by asking for optimal controls on the interval $[s,T]$ with $s \geq 0$ and initial condition $x(s) = \xi$, but we will not do this here. To any $u \in \mathcal{U}_r$ and any $\xi \in X$ there corresponds a unique absolutely continuous function $x : I \to E$ so that (x,u) is a feasible process with respect to ξ. The costs of this process can be regarded as a function $J(\xi,u)$ on $X \times \mathcal{U}_r$. Note that J is well defined but might be infinite.

Definition 3. On X we define the *value function* as

$$V(\xi) = \inf_{u\in\mathcal{U}_r} J(\xi,u),$$

with $V(\xi) = \infty$ if there is no admissible control for ξ.

Regarding J and the value function, we have the following

Lemma 10. *(a) There exists a neighbourhood $X_0 \subset X$ so that $J(.,u)$ is bounded on X_0, where u is the admissible control for ξ_0.*
(b) If $\xi \in X_0$ and $u \in \mathcal{U}_r$ is admissible for ξ, then $J(.,u)$ has a partial derivative at ξ, and

$$\mathrm{D}_1 J(\xi,u) = \lambda(0)$$

where λ is given by Equation (12)*.*
(c) The value function is Lipschitz continuous on bounded subsets of X_0.

Proof. Let (x, u) be the admissible process for ξ_0. Then due to Hypothesis 3(b), there is a neighbourhood $X_0 \subset X$ of ξ_0 so that any solution (y, u) of the state equation with control u and initial condition $\xi \in X_0$ satisfies the bound $\|y-x\|_\infty \leq R$, where R is as in Hypothesis 2(c). Further,

$$\begin{aligned} J(\xi, u) = J(\xi_0, u) \\ + \int_I \phi(t, y(t), u(t)) - \phi(t, x(t), u(t)) \, \mathrm{d}t \\ + \int_I \psi(t, y(t)) - \psi(t, x(t)) \, \mathrm{d}\eta_t. \end{aligned} \tag{27}$$

From the mean value theorem, we get

$$|\phi(t, y(t), u(t)) - \phi(t, x(t), u(t))| \leq |\mathrm{D}_2\phi(t, \theta(t), u(t))|R$$

for some function $\theta : I \to E$ with $\|\theta - x\|_\infty \leq R$. It can be shown (see [2], proof of theorem 22.17) that Hypothesis 2(c) implies the apparently stronger statement that if $|z-x(t)| \leq R$, then $|\mathrm{D}_2\phi(t, z, u(t))| \leq c'|\phi(t, x(t), u(t))| + d'(t)$ almost surely, with $c' \geq 0$ and d' integrable. This shows that the first integral in Equation (27) is bounded. The stochastic costs are easily seen to be continuous with respect to varying the initial condition of the state equation, since this causes the trajectories to vary continuously in the uniform topology, while the 1-variation remains bounded. Hence, all terms on the right hand side of Equation (27) remain bounded.

To prove item (b), let λ be the unique solution of Equation (12). We regard x as a function of time and of the initial condition ξ, and make this explicit by writing $x_\xi(t)$ for $t \in I$. Evidently, $x_\xi(0) = \xi$. For any $v \in E$ we have

$$\begin{aligned} \mathrm{D}_1 J(\xi, u)v = \int_0^T \mathrm{D}_2\phi(t, x_\xi(t), u(t)) \, \zeta(t) \, \mathrm{d}t \\ + \int_0^T \mathrm{D}_2\psi(t, x_\xi(t)) \, \zeta(t) \, \mathrm{d}\eta_t \end{aligned} \tag{28}$$

where

$$\frac{\mathrm{d}}{\mathrm{d}t}\zeta(t) = \mathrm{D}_2\{f(t, x_\xi(t)) + g(t, x_\xi(t))u(t)\}\zeta(t), \qquad \zeta(0) = v.$$

A proof of this runs similar to the proof of the claims (22,23,24). Applying the Duality Lemma 9 gives that the right hand side of Equation (28) is equal to $\lambda(0)^{\mathrm{T}}v$.

To prove item (c), first note that by item (a), there is a control u that is admissible for any $\xi \in X_0$, and $J(., u)$ is bounded over X_0 by J_0, say. This implies that for any $\xi \in X_0$ there exist an optimal triple $(x_\xi, u_\xi, \lambda_\xi)$, and we need to show that $\lambda_\xi(0)$ is bounded for $\xi \in X_0$. The estimate (20) shows that $\|u_\xi\|_r$ is bounded over $\xi \in X_0$. Going back to Hypothesis 3(b) we see that $\|x_\xi\|_\infty$ and $\|\dot{x}_\xi\|_r$ are also bounded. We remember that by the Maximum Principle (Thm. 2), we have

$$\begin{aligned} \lambda_\xi(0) = \int_I \lambda(s)\big\{\mathrm{D}_2 f(s, x_\xi(s)) + \mathrm{D}_2 g(s, x_\xi(s))u_\xi(s)\big\} \, \mathrm{d}s \\ + \int_I \mathrm{D}_2\phi(s, x_\xi(s), u_\xi(s)) \, \mathrm{d}s + \int_I \mathrm{D}_2\psi(s, x_\xi(s)) \, \mathrm{d}\eta_s. \end{aligned} \tag{29}$$

The function $t \to \mathrm{D}_2 f(s, x_\xi(s)) + \mathrm{D}_2 g(s, x_\xi(s))u_\xi(s)$ is bounded in $\|.\|_1$ thanks to Hypothesis 2(a) and the boundedness of $\|u_\xi\|_r$. The term $\int_I \mathrm{D}_2\psi(s, x_\xi(s)) \, \mathrm{d}\eta_s$ is bounded due to Hypothesis 2(d), the boundedness of $\|x_\xi\|_\infty$ and our estimates of

the Young integral. Hypothesis 2(c) provides the estimate

$$\begin{aligned}&\int_I \mathrm{D}_2\phi(s, x_\xi(s), u_\xi(s))\,\mathrm{d}s\\ &\leq \int_I c|\phi(s, x_\xi(s), u_\xi(s))|\,\mathrm{d}s + D\\ &\leq cJ_0 + c\,|\int_I \psi(s, x_\xi(s))\,\mathrm{d}\eta_s| + D,\end{aligned}$$

and the integral is bounded again due to Hypothesis 2(d), the boundedness of $\|x_\xi\|_\infty$ and our estimates of the Young integral. □

We will from now on restrict attention to X_0 which we rename X. Since V is locally Lipschitz, $\mathrm{D}V$ exists on a dense set in X by Rademacher's theorem. The *reachable gradient* of V at $\xi \in X$, denoted by $\partial^* V(\xi)$, is the set of all cluster points of $\mathrm{D}V(\xi_n)$ when $\xi_n \to \xi$. Note that $\partial^* V(\xi)$ is nonempty for all $\xi \in X$, and it is well known (see [6], prop. 2) that V is differentiable at ξ if and only if $\partial^* V(\xi)$ is a singleton, in which case $\{\mathrm{D}V(\xi)\} = \partial^* V(\xi)$.

Theorem 5. *(a) Let $\xi \in X$ and suppose that for any control u that is optimal with respect to ξ, the corresponding pair (x, λ) is the unique solution of the Hamiltonian system* (26) *with respect to the initial condition $(\xi, \lambda(0))$. Then if V has a derivative at ξ, the optimal control u for ξ is unique (a.s. wrt Lebesgue measure), and $\mathrm{D}V(\xi) = \lambda(0)$.*

(b) Assume that in addition to the conditions in the first item, solutions to the Hamiltonian system (26) *depend continuously on initial conditions in the uniform topology. Then if $w \in \partial^* V(\xi)$, there exists an optimal state control pair (x, u) with $\lambda(0) = w$. In particular, if u is a unique optimal control with respect to ξ, then V has a derivative at ξ.*

Proof. Let u_1 and u_2 be optimal controls, with (x_1, u_1, λ_1) and (x_2, u_2, λ_2) being the corresponding optimal triples. Note that $x_1(0) = x_2(0) = \xi$, and both (x_1, λ_1) and (x_2, λ_2) are solutions of the Hamiltonian system (26). Since $V(z) \leq J(z, u_1)$ for $z \in X$, with equality if $z = \xi$, we can conclude that $\mathrm{D}V(\xi) = \mathrm{D}_1 J(\xi, u_1)$, in case V has a derivative at ξ. Since the same is true for u_2, we see that $\lambda_1(0) = \lambda_2(0) = \mathrm{D}V(\xi)$ and therefore $(x_1, \lambda_1) = (x_2, \lambda_2)$ because the Hamiltonian system (26) has unique solutions with respect to initial conditions. Therefore $u_1(t) = \upsilon(x_1(t), \lambda_1(t)) = \upsilon(x_2(t), \lambda_2(t)) = u_2(t)$, and item (a) is proved.

To prove item (b), let $w \in \partial^* V(\xi)$. Take a sequence $\xi_k \in X, k \in \mathbb{N}$ with $\xi_k \to \xi$ and so that V is differentiable at ξ_k; letting $\lambda_k(0) = \mathrm{D}V(\xi_k)$, we also assume $\lambda_k(0) \to w$. This is possible owing to the definition of $\partial^* V(\xi)$. For each $k \in \mathbb{N}$, let (x_k, λ_k) be the solution to the Hamiltonian system (26) with initial condition $(\xi_k, \lambda_k(0))$. By assumption, $(x_k, \lambda_k) \to (x, \lambda)$ uniformly, where (x, λ) is the solution to the Hamiltonian system (26) with initial condition (ξ, w). Putting $u_k(t) = \upsilon(x_k(t), \lambda_k(t)), t \in I$, we see that u_k is the unique optimal control for ξ_k for each $k \in \mathbb{N}$. Now u_k converges uniformly to some control u, and it follows as in the proof of Theorem 1 that u is optimal for ξ. The remainder of item (b) follows from the remark just before the Theorem and the fact that if u is the only optimal control for ξ, then $\partial^* V(\xi)$ must be singleton. □

6. Motivating example from geophysical fluid dynamics

We consider a class of examples with a state equation of the form

$$\dot{x}(t) = f_1(t, x(t)) + f_2(x(t)) + g(t, x(t))u(t) \tag{30}$$

on the interval $I = [0, T]$ with initial condition $x(0) = \xi \in E$. We use $E = \mathbb{R}^n$ for some n and use the standard scalar product and norm on E. We impose the conditions

(a) f_1 is continuous and has a bounded and continuous derivative with respect to x.
(b) f_2 is a bilinear form with the property $x^{\mathrm{T}} f_2(x) = 0$.
(c) g is bounded and continuous and has a bounded and continuous derivative with respect to x.

In the context of geophysical fluid dynamics, f_1 represents viscosity, coriolis forces and other physical effects. The bilinear term f_2 is a kinematic term inherited from the advection term (or "material acceleration") in the Navier Stokes equations. The function g need not have any physical interpretation but might be present to scale the control u or enforce balance conditions. We also assume the presence of a closed and convex control set $U \subset E$ (which might be equal to E); again, this set might represent balance conditions.

The class of systems described by the conditions (a-c) above contains various conceptual weather and climate models, such as Lorenz'63, Lorenz'96, and truncation approximations to the Navier Stokes and Barotropic Vorticity models. Of course, in the original form of these models, there is no g, which is an additional component coming in with the data assimilation.

We will now demonstrate that the conditions (a-c) above (plus further conditions on the running costs discussed below) imply Hypotheses 1 and 2. By multiplying the state equation (30) with x^{T} and integrating from 0 to t we obtain

$$\begin{aligned}\frac{1}{2}x^2(t) &= \frac{1}{2}\xi^2 + \int_0^t x(s)^{\mathrm{T}} f_1(s, x(s))\,\mathrm{d}s \\ &\quad + \int_0^t x(s)^{\mathrm{T}} g(s, x) u(s)\,\mathrm{d}s \\ &\leq C_1 + C_2 \int_0^t |x(s)|^2\,\mathrm{d}s + C_3 \int_0^t |x(s)||u(s)|\,\mathrm{d}s \\ &\leq C_1' + C_2' \int_0^t |x(s)|^2\,\mathrm{d}s + C_3' \int_0^t |u(s)|^2\,\mathrm{d}s\end{aligned}$$

where we have used first the property of f_2, next the properties of f_1 and g, and finally Young's inequality. Applying Grönwall's inequality, we obtain the energy estimate (9) with $\gamma = 1$ and $r = 2$. Using the energy estimate directy in the state equation (30), we obtain the nonlinearity estimate (8) with $\beta = 2$ and $r = 2$.

We introduce an observation function $h : I \times E \to \mathbb{R}^d$ and impose the condition

(d) The function h satisfies Hypothesis 1(b) for the stochastic running costs ψ with $\alpha = 0$ and Hypothesis 2. Further, the derivative of h with respect to x is bounded.

For the costs, we use the functional (2), where

(e) R, S are continuously differentiable matrix valued functions, and $S(t) \geq s\mathbb{1}$ for some $s > 0$.

This implies

$$\phi(t, x, u) = \frac{1}{2}h(t, x)^{\mathrm{T}} R(t) h(t, x) + \frac{1}{2}u^{\mathrm{T}} S(t) u.$$

The cost functional corresponds to a minimum energy estimation or weakly constrained 4d–VAR in the geosciences. By our choice of ϕ, we can conclude that Hypothesis 1(d) is satisfied with $r = 2, \delta = 0$. Further

$$\psi(t, x) = -h(t, x)^{\mathrm{T}} R(t)$$

and hence ψ satisfies Hypothesis 1(b) with $\kappa \leq 1$ and $\alpha = 0$, and we obtain that $\gamma(\alpha + \kappa - \frac{1}{2}) + \frac{\beta}{2} \leq \frac{3}{2}$ which is smaller than $r = 2$. Theorem 1 is therefore in force and we obtain that there is an optimal solution to this data assimilation problem for every observation path, and every optimal solution will satisfy $\|u\|_2 < \infty$.

With regards to the Maximum Principle, we note that Hypothesis 2(b) is satisfied and ϕ is strongly convex in u. Further $\mathrm{D}_2\phi$ does not depend on u and is continuous in (t, x), whence Hypothesis 2(c) is satisfied. Therefore, the Maximum Principle (Thm. 2) applies, and so does Theorem 4. The Hamiltonian equations (26) in the present case read as

$$\dot{x} = f(t, x(s)) + g(s, x(s))u(s), \tag{31}$$

and (written in coordinates)

$$\begin{aligned} \lambda_m = &\sum_j \int_t^T \lambda_j(s)\{\mathrm{D}_2 f(s, x(s)) + \mathrm{D}_2 g(s, x(s))u(s)\}_{jm}\, \mathrm{d}s \\ &+ \sum_{jk} \int_t^T \partial_m h_j(s, x(s)) R_{jk}(s)\{h_k(s, x(s))\, \mathrm{d}s - \mathrm{d}\eta_k(s)\}, \end{aligned} \tag{32}$$

with $u(t) = -S^{-1} g(t, x(t))^{\mathrm{T}} \lambda(t)$ and $x(0) = \xi$. In this case, it follows directly that the control is continuous and has q–variation for any $q > 2$.

Primitive conditions that render Theorem 5 applicable will be investigated in a future paper, but we can already present the following preliminary result: Suppose that the function h is linear in x and that $\mathrm{D}_2 f, \mathrm{D}_2 g$ are locally Lipschitz continuous with respect to x. It is then easy to see that solutions to the Hamiltonian system (Eqns. 31,32 in the present case) for given initial values $(\xi, \lambda(0))$ are unique and depend continuously on these values. Thus in this special but important situation, we obtain the conclusions of Theorem 5; in particular, optimal controls are unique if the value function has a derivative at ξ.

As a final note, we discuss the variational problems encountered in [21] in connection with the maximum aposteriori (MAP) estimator of trajectories of diffusion processes. The conditions (a, b, c) on the state equation 30 are strengthened to

- (a') f_1 has bounded and continuous derivatives up to second order with respect to x.
- (b') same as (b).
- (c') g does not depend on t, and gg^{T} has bounded and continuous derivatives up to third order with respect to x. Further, $0 < c_1 \mathbb{1} \leq gg^{\mathrm{T}} \leq c_2 \mathbb{1}$.

Under these conditions, $\Gamma := (gg^{\mathrm{T}})^{-1}$ defines a Riemannian metric, and we write σ and div for the scalar curvature and the divergence, respectively, associated with this metric. The conditions on ψ and R remain the same.

In [21] it is shown that under certain conditions (which we do not verify here) a MAP estimator for trajectories of the diffusion

$$\mathrm{d}x(t) = \tilde{f}(t, x(t))\mathrm{d}t + g(x(t))\, \mathrm{d}B(t)$$

with respect to observations of the form

$$\mathrm{d}\eta(t) = h(t, x(t))\mathrm{d}t + \rho_t\, \mathrm{d}B'(t)$$

is a solution of Problem VAR with ψ

$$\begin{aligned}\psi(t,x) &= -h(t,x)^{\mathrm{T}} R(t),\\ R(t) &= (\rho(t)\rho(t)^{\mathrm{T}})^{-1},\\ \phi(t,x,u) &= \frac{1}{2}h(t,x)^{\mathrm{T}} R(t) h(t,x)\\ &\quad + \frac{1}{2}u^{\mathrm{T}}\Gamma(x)u - \operatorname{div} f(t,x) + \frac{1}{6}\sigma(x),\end{aligned}$$

and $\tilde{f}$ is related to f through an Itô–Stratonovič conversion (see [21] for details). The new terms appearing in the cost function are σ, $\operatorname{div} f_1$ and $\operatorname{div} f_2$. The first two are bounded with bounded derivatives, while the third is linear in x. Hence Hypothesis 2 is again satisfied with α, β, γ the same constants as before, only $\delta = 1$ now instead of zero as before. We can thus draw the same conclusions as for the minimum energy estimator, in particular we can conclude the existence of global minimisers. Note that in [21] this conclusion is reached, albeit under much stronger assumptions. In particular, the boundedness of $\operatorname{div} f$ is needed for that proof to work, meaning that state equations with a quadratic nonlinearity as considered here are not covered in [21].

Acknowledgements. The author was supported by the UK Engineering and Physical Sciences Research Council under grant agreement EP/L012669/1. Fruitful discussions with Horatio Boedihardjo, Tobias Kuna, and Dan Crisan are gratefully acknowledged.

References

[1] M. Bardi and I. Capuzzo-Dolcetta, *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations*, Systems & Control: Foundations & Applications, Birkhäuser Boston, Inc., Boston, MA, 1997, https://doi.org/10.1007/978-0-8176-4755-1, http://dx.doi.org/10.1007/978-0-8176-4755-1. With appendices by Maurizio Falcone and Pierpaolo Soravia.

[2] F. Clarke, *Functional analysis, calculus of variations and optimal control*, vol. 264 of Graduate Texts in Mathematics, Springer, London, 2013, https://doi.org/10.1007/978-1-4471-4820-3, http://dx.doi.org/10.1007/978-1-4471-4820-3.

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

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

[5] W. H. Fleming and H. M. Soner, *Controlled Markov processes and viscosity solutions*, vol. 25 of Stochastic Modelling and Applied Probability, Springer, New York, second ed., 2006.

[6] H. Frankowska and A. Ochal, *On singularities of value function for bolza optimal control problem*, Journal of Mathematical Analysis and Applications, 306 (2005), pp. 714 – 729, https://doi.org/http://dx.doi.org/10.1016/j.jmaa.2004.10.003, http://www.sciencedirect.com/science/article/pii/S0022247X0400825X.

[7] P. K. Friz and N. B. Victoir, *Multidimensional stochastic processes as rough paths*, vol. 120 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 2010. Theory and applications.

[8] O. Hijab, *Asymptotic bayesian estimation of a first order equation with small diffusion*, Annals of Probability, 12 (1984), pp. 890–902.

[9] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, vol. 64 of Mathematics in Science and Engineering, Academic Press, 1970.

[10] E. Kalnay, *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge University Press, first ed., 2001.

[11] A. Krener, *The convergence of the minimum energy estimator*, in New Trends in Nonlinear Dynamics and Control and their Applications, W. Kang, C. Borges, and M. Xiao, eds., vol. 295 of Lecture Notes in Control and Information Sciences, Springer Berlin / Heidelberg, 2003, pp. 187–208.

[12] R. E. Mortensen, *Maximum-likelihood recursive nonlinear filtering*, Journal of Optimization Theory and Applications, 2 (1968), pp. 386–394.
[13] G. Nakamura and R. Potthast, *Inverse Modeling*, 2053-2563, IOP Publishing, 2015, https://doi.org/10.1088/978-0-7503-1218-9, http://dx.doi.org/10.1088/978-0-7503-1218-9.
[14] L. C. G. Rogers, *Least-action filtering*, ArXiv e-prints, (2013), https://arxiv.org/abs/1301.5157.
[15] A. Sage, *Optimum Systems Control*, Englewood Cliffs, N. J.: Prentice-Hall, 1968.
[16] E. D. Sontag, *Mathematical Control Theory. Deterministic Finite-Dimensional Systems*, vol. 6 of Texts in Applied Mathematics, Springer-Verlag, New York, second ed., 1998.
[17] R. Temam, *Infinite-dimensional dynamical systems in mechanics and physics*, vol. 68 of Applied Mathematical Sciences, Springer-Verlag, New York, second ed., 1997, https://doi.org/10.1007/978-1-4612-0645-3, http://dx.doi.org/10.1007/978-1-4612-0645-3.
[18] Y. Tremolet, *Accounting for an imperfect model in 4d-var*, Quarterly Journal of the Royal Meteorological Society, 132 (2006), pp. 2483–2504.
[19] J. Yong and X. Y. Zhou, *Stochastic controls, Hamiltonian systems, and HJB equations*, vol. 43 of Applications of Mathematics, Springer-Verlag, New York, 1999, https://doi.org/10.1007/978-1-4612-1466-3, http://dx.doi.org/10.1007/978-1-4612-1466-3.
[20] O. Zeitouni and A. Dembo, *A maximum a posteriori estimator for trajectories of diffusion processes*, Stochastics, 20 (1987), p. 221.
[21] 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, https://doi.org/10.1080/17442508808833490, http://www.informaworld.com/10.1080/17442508808833490.
[22] O. Zeitouni and A. Dembo, *A change of variables formula for Stratonovich integrals and existence of solutions for two-point stochastic boundary value problems*, Probab. Theory Related Fields, 84 (1990), pp. 411–425, https://doi.org/10.1007/BF01197893, http://dx.doi.org/10.1007/BF01197893.

School of Mathematical, Physical, and Computational Sciences, University of Reading, Reading RG6 6AX, United Kingdom.
Email address: j.broecker@reading.ac.uk