SIAM J. Applied Dynamical Systems
Vol. 21, No. 2, pp. 879–902

Shadowing-Based Data Assimilation Method for Partially Observed Models*

Bart M. de Leeuw† and Svetlana Dubinkina†‡

Abstract. In this article we develop further an algorithm for data assimilation based upon a shadowing refinement technique [de Leeuw et al., *SIAM J. Appl. Dyn. Syst.*, 17 (2018), pp. 2446–2477] to take partial observations into account. Our method is based on a regularized Gauss–Newton method. We prove local convergence to the solution manifold and provide a lower bound on the algorithmic time step. We use numerical experiments with the Lorenz 63 and Lorenz 96 models to illustrate convergence of the algorithm and show that the results compare favorably with a variational technique—weak-constraint four-dimensional variational method—and a shadowing technique–pseudo-orbit data assimilation. Numerical experiments show that a preconditioner chosen based on a cost function allows the algorithm to find an orbit of the dynamical system in the vicinity of the true solution.

Key words. data assimilation, tangent space decomposition, shadowing, partial observations

AMS subject classifications. 62M20, 37C50, 65J20

DOI. 10.1137/18M1223897

1. Introduction.

Data assimilation (DA) combines measurement data (also called observations) over an interval with model simulations to construct an accurate initial condition for a chaotic dynamical system [20]. The shadowing-type data assimilation methods [22, 3, 21, 42, 13] employ (a proxy of) observations as an initial guess and, using a gradient descent algorithm, search for a trajectory of a dynamical model that is, hopefully, in a neighborhood of the true trajectory. By the *true trajectory* we mean a model trajectory from which the observations could have been generated.

When only partial observations are available, it is necessary to generate a proxy for complete observations. A proxy of observations is generated by another (not shadowing-type) DA method using partial observations and a background trajectory. The background trajectory is a model simulation that does not employ observations and thus typically has a large error with respect to the true trajectory. Such a preprocessing provides a more accurate, complete initial guess for the shadowing-type DA methods. However, it requires the availability of another DA method and consequently induces the computational costs. Needless to say, the shadowing-type DA methods are the only DA methods that require a proxy for complete observations.

*Received by the editors October 31, 2018; accepted for publication (in revised form) by C. Topaz November 10, 2021; published electronically April 11, 2022.
https://doi.org/10.1137/18M1223897
Funding: The work of the first author was partially supported by the research program Mathematics of Planet Earth 2014 EW project 657.014.001, which is financed by the Netherlands Organisation for Scientific Research (NWO).
†Centrum Wiskunde & Informatica, PO Box 94079, 1090 GB Amsterdam, Netherlands (b.m.de.leeuw@cwi.nl).
‡VU Amsterdam, Department of Mathematics, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands (s.b.dubinkina@vu.nl, https://math.vu.nl/~S.Dubinkina/).

In this paper, we propose a new shadowing-type DA method that employs partial observations and a background trajectory directly for initialization, thus without any preprocessing. The proposed shadowing-type DA method is based on the shadowing-type DA method of [9]. The shadowing-type DA method of [9] we call here the noise-reduction DA method. The noise-reduction method seeks zeros of a corresponding single-step residual function using a Gauss–Newton method and employs (a proxy of) observations as an initial guess. In order to account for partial observations without preprocessing we propose to use Levenberg–Marquardt regularization [26, 31] when seeking zeros. Since the Levenberg–Marquardt algorithm can be seen as a regularization of the Gauss–Newton method, we call the corresponding shadowing-type DA method the *regularized* shadowing DA method. A regularization parameter controls the algorithmic time step, making the Gauss–Newton method convergent to the solution manifold independently of the initial guess. We prove local convergence of the proposed regularized shadowing DA method following [3]. Note that the Levenberg–Marquardt regularization is often used in nonlinear optimization and data assimilation in particular, e.g., variational data assimilation [30], and ensemble Kalman filter [4].

Despite being convergent to the solution manifold, the regularized shadowing DA method might poorly approximate the true solution due to observations being used only as an initial guess. Therefore, in this paper we introduce a preconditioner for the corresponding gradient flow that modifies the direction of the search such that the estimate remains in the vicinity of observations. This is done in the spirit of trust region methods [33], which together with Gauss–Newton-type methods have been an inspiration for new algorithms to solve nonlinear least-squares problems (see, e.g., [10]).

First, let us review shadowing refinement that is the core of the shadowing-type DA methods. We consider a discrete deterministic model

$$x_{n+1} = F_n(x_n), \quad x_n \in \mathbb{R}^m, \quad n = 0, \dots, N-1, \tag{1}$$

where $F_n : \mathbb{R}^m \to \mathbb{R}^m$. We assume F_n to be $\mathcal{C}^3$ for all n. In many applications the model is defined by the time-discretization of an ordinary differential equation $\dot{x} = f(t, x)$, $x(t) \in \mathbb{R}^m$, which in turn may be defined as the space-discretization of a partial differential equation (or system of PDEs). We define the following single-step residual function:

$$G(\boldsymbol{u}) = \begin{pmatrix} G_0(\boldsymbol{u}) \\ G_1(\boldsymbol{u}) \\ \vdots \\ G_{N-1}(\boldsymbol{u}) \end{pmatrix}, \qquad G_n(\boldsymbol{u}) = u_{n+1} - F_n(u_n), \quad n = 0, \dots, N-1. \tag{2}$$

A sequence $\boldsymbol{u} = \{u_0, \dots, u_N\}$ is called an ε-pseudo-orbit of F if $\|G_n(\boldsymbol{u})\| < \varepsilon$ for all $n = 0, \dots, N-1$, where $\|\cdot\|$ is a norm in $\mathbb{R}^m$. Note that $\boldsymbol{u} = \{u_0, \dots, u_N\}$ is called an orbit of F if $\|G_n(\boldsymbol{u})\| = 0$ for all $n = 0, \dots, N-1$.

Assume that F admits the hyperbolic splitting into stable (contracting) and unstable (expanding) subspaces, and thus there exists a hyperbolic set for F. Suppose $\boldsymbol{u}$ is an ε-orbit in a neighborhood of a hyperbolic set for F. Then the shadowing lemma (e.g., Theorem 18.1.2 of [23]) states that for every $\delta > 0$ there exists $\varepsilon > 0$ such that $\boldsymbol{u}$ is δ-shadowed by an orbit

of F, i.e., there exists an orbit $\{x_n\}$ satisfying $x_{n+1} = F(x_n)$ such that $\|u_n - x_n\| < \delta$ for all $n = 0, \ldots, N$.

The shadowing trajectories were originally proposed to prove the existence of an exact trajectory in the vicinity of a pseudo-trajectory [35, 17, 43, 6, 5]. Shadowing refinement [17] employs a pseudo-orbit as an initial guess and iteratively refines the pseudo-orbit to obtain an improved approximation of an exact solution. The work of [43] proved convergence of the shadowing refinement algorithm to an exact orbit that is not far from the initial guess. When a dynamical system is nonuniformly hyperbolic, the hyperbolic splitting holds only at almost every point on $\mathbb{R}^m$, and then a dynamical system is shadowed for finite nontrivial lengths of time [6, 7, 8, 48]. Furthermore, for a (nonuniformly) hyperbolic system a pseudo-orbit shadows a unique exact orbit under rather strict conditions on an exponential dichotomy [35, 5].

A (nonuniformly) hyperbolic dynamical system has a fixed number of (zero) negative and positive Lyapunov exponents. The stable and nonstable subspaces correspond to Lyapunov vectors associated with negative and nonnegative Lyapunov exponents, respectively. Examples of nonuniformly hyperbolic dynamical systems are the Lorenz 63 (L63) model [28], which has one positive, one zero, and one negative Lyapunov exponent, and the Lorenz 96 (L96) model [29], which has thirteen positive Lyapunov exponents, one zero Lyapunov exponent, and all others being negative Lyapunov exponents. A nonhyperbolic dynamical system has a fluctuating number of positive Lyapunov exponents. An example of a nonhyperbolic dynamical system is a truncated quasi-geostrophic atmosphere-ocean model [49] that has two positive Lyapunov exponents, sixteen negative Lyapunov exponents, and eighteen near-zero (fluctuating) Lyapunov exponents.

In contrast to the classical shadowing refinement, shadowing-based data assimilation [22, 3, 42, 13, 9] does not search for an exact orbit of $\dot{x} = f(x)$. It rather searches for an exact orbit of a dynamical *model*—a discrete dynamical system (1) that originates from a time- and space-discretization of a system of PDE describing the behavior of the underlying dynamical system. However, it has the same spirit as shadowing refinement, since it employs a noisy orbit (proxy of observations) and searches for a shadow of an exact orbit of a dynamical model that is hopefully in a neighborhood of the true orbit. The observation noise is typically assumed to be a random Gaussian-distributed variable. However, bounded observation noise can also be reduced by a shadowing-type DA method [22].

Recent efforts to make a DA method computationally more efficient led to an approach of assimilation in the unstable subspace. This approach was first proposed for variational data assimilation [47, 46, 34, 24, 38] and later adapted for ensemble data assimilation [16, 2] and shadowing-type data assimilation [9]. The dimension of the nonstable subspace is considerably smaller than the dimension of the full space. Since the noise-reduction method showed encouraging results when the tangent space was partitioned into nonstable and unstable subspaces, we adapt the partitioning to the regularized shadowing DA method and investigate numerically the accuracy.

The rest of the paper is organized as follows. In sections 2 and 4, we briefly recall the noise-reduction DA method and the projected noise-reduction DA method of [9], respectively. In section 3, we introduce the regularized shadowing DA method for partial observations and prove its local convergence and closeness of the estimate to the observations. In section 5, we introduce the projected regularized shadowing DA method for partial observations. In

section 6, we present results for the L63 and the L96 models. Finally, we draw the conclusions in section 7.

2. The noise-reduction DA method. Let the sequence $\boldsymbol{\mathcal{X}} := \{\mathcal{X}_0, \dots, \mathcal{X}_N\}$ be a distinguished orbit of (1), referred to as the true solution of the model, and presumed to be unknown. Suppose we are given a sequence of partial noisy observations $\boldsymbol{y} := \{y_0, \dots y_N\}$ related to $\boldsymbol{\mathcal{X}}$ via

$$y_n = H_n \mathcal{X}_n + \xi_n, \qquad y_n \in \mathbb{R}^d, \quad n = 0, \dots, N,$$

where $H_n : \mathbb{R}^m \to \mathbb{R}^d$, $d \le m$, is the linear observation operator, and the noise variables ξ_n are drawn from a normal distribution $\mathcal{N}(0, E_n)$ with zero mean and known observational error covariance matrix E_n.

Data assimilation is the problem of finding a pseudo-orbit $\boldsymbol{u} = \{u_0, u_1, \dots, u_N\}$, $u_n \in \mathbb{R}^d$, of the model (1), such that the differences $\|y_n - H_n u_n\|$ and $\|u_n - F_n(u_{n-1})\|$, $n = 1, \dots, N$ are small in an appropriately defined sense. This is done with the aim of minimizing the unknown error $\|u_n - \mathcal{X}_n\|$; see, for example, [44, 25]. Well-known four-dimensional variational data assimilation aims at finding the optimal initial condition u_0 of (1) to minimize a cost function

$$C_{\text{var}}(u_0; \{y_n\}) = \sum_{n=1}^{N} (y_n - H_n u_n)^T E_n^{-1} (y_n - H_n u_n) + (u_n - F_n(u_{n-1}))^T C_n^{-1} (u_n - F_n(u_{n-1})), \tag{3}$$

where $\{C_n\}_{n=0}^N$ is the model error covariance. Here the last term with model error covariance $\{C_n\}_{n=0}^N$ is present for weak-constraint variational data assimilation (WC4DVar). WC4DVar seeks a pseudo-orbit of (1); see, e.g., [39, 27, 45, 44] and references therein. Strong-constraint variational data assimilation does not have this term in the corresponding cost function and thus seeks an exact orbit of (1). However, it suffers from a drastic increase of the number of local minima as N increases [1, 32, 37].

Instead of minimizing a cost function, the noise-reduction DA method [9] searches for a zero of the single-step residual function (2) using a contractive iteration started from (a proxy of) complete, noisy observations. Just as with strong-constraint four-dimensional variational data assimilation, the noise-reduction method attempts to find an exact orbit of (1) consistent with the observations. However, instead of solving directly for the initial condition, the noise-reduction method solves for the whole orbit at once. It was shown that shadowing-type DA methods do not have the issue of multiple local minima; see, e.g., [22] and [9] for the noise-reduction method.

The noise-reduction method seeks an update $\boldsymbol{D}^{(k)}$ by approximately solving

$$G\left(\boldsymbol{u}^{(k)} + \boldsymbol{D}^{(k)}\right) = 0. \tag{4}$$

Here k denotes the index of the Gauss–Newton iteration and the solution to (4) is approximated using the right pseudoinverse of G',

$$\boldsymbol{u}^{(k+1)} = \boldsymbol{u}^{(k)} + \boldsymbol{D}^{(k)}, \quad \boldsymbol{D}^{(k)} = -G'(\boldsymbol{u}^{(k)})^{\dagger} G(\boldsymbol{u}^{(k)}) = -G'^T (G'\, G'^T)^{-1} G$$

with $\boldsymbol{u}^{(0)} = \boldsymbol{\mathcal{X}} + \boldsymbol{\xi}$. Without loss of generality, we can assume that observation operator H is the identity matrix for a proxy of complete observations. The function $G(\boldsymbol{u})$ has a zero for every orbit of the model. The Jacobian of G has an $m(N-1) \times mN$ block structure:

$$G'(\boldsymbol{u}) = \begin{bmatrix} -F_0'(u_0) & I & & & \\ & -F_1'(u_1) & I & & \\ & & \ddots & \ddots & \\ & & & -F_{N-1}'(u_{N-1}) & I \end{bmatrix} \tag{5}$$

with full row-rank. Since the Jacobian appears only when acting on a given vector (unit vector), it could be efficiently approximated by finite differences $F'(u)v \approx 1/\varepsilon(F(u+\varepsilon v) - F(u))$.

3. Regularized shadowing DA method. In this section, we introduce a new shadowing DA method that is based on Levenberg–Marquardt regularization of the noise-reduction method described in section 2. The noise-reduction method uses the Gauss–Newton method to find zeros of (2), which requires a good initial guess. When such an initial guess is not available, the noise-reduction method fails to converge. Therefore, we regularize the Gauss–Newton method following [26, 31] to have the shadowing DA method convergent to the solution manifold $G(\boldsymbol{u}) = 0$ for an initial guess that is not necessarily *good*.

Furthermore, the proposed regularized shadowing DA method takes into account the observation covariance error $\{E_n\}_{n=0}^N$ unlike existing shadowing-type DA methods. It has a tuning matrix W that keeps descent steps in the direction of observed variables small compared to descent steps in the direction of nonobserved variables. As the iteration proceeds, observed variables get denoised as well and the algorithm finds a (pseudo-)orbit compatible with observations.

Since we are interested in assimilating partial observations, we need to complete the partial observations to have an initial guess for the regularized shadowing DA method. Instead of using a proxy of observations as in the noise-reduction method, we use a combination of partial observations $\boldsymbol{y}$ and a background trajectory $\boldsymbol{x}^{\mathrm{b}}$. The background trajectory $\boldsymbol{x}^{\mathrm{b}}$ is a solution of (1). Typically the background trajectory is obtained by performing data assimilation over a previous time window, thus by employing past observations. However, we assume that the background trajectory is a solution of (1) with an arbitrary initial condition. We assume that an initial guess for the shadowing DA method is

$$\boldsymbol{u}^{(0)} = H^T \boldsymbol{y} + H^{\perp} \boldsymbol{x}^{\mathrm{b}}, \tag{6}$$

where H is the observation operator, which has a $dN \times mN$ block diagonal structure $H = \mathrm{blockdiag}\,(H_1, \ldots, H_N)$, and the operator $H^{\perp}$ is defined as

$$H^{\perp} = (I - H^T H). \tag{7}$$

We assume that for any $n = 0, \ldots, N$, H_n has only one nonzero element in a row and that element is equal to 1. This is a standard assumption about the observation operator, which corresponds to observing separate variables rather than a combination of variables. Then we have $HH^{\perp} = 0$ and from (6) it follows that $H\boldsymbol{u}^{(0)} = \boldsymbol{y}$ and $H^{\perp}\boldsymbol{u}^{(0)} = H^{\perp}\boldsymbol{x}^{\mathrm{b}}$. Thus H projects

the initial guess $\boldsymbol{u}^{(0)}$ to the observation space $\mathbb{R}^d$, while $H^{\perp}$ sets the observed components of the initial guess $\boldsymbol{u}^{(0)}$ to zero.

We seek an update $\boldsymbol{\Pi}^{(k)}$ by approximately solving

$$G\left(\boldsymbol{u}^{(k)}+\boldsymbol{\Pi}^{(k)}\right)=0 \tag{8}$$

using the Levenberg–Marquardt regularization

$$\boldsymbol{u}^{(k+1)}=\boldsymbol{u}^{(k)}+\boldsymbol{\Pi}^{(k)},\quad \boldsymbol{\Pi}^{(k)}=-\Sigma G'^{T}\left(G'\Sigma G'^{T}+\alpha^{(k)}C\right)^{-1}G. \tag{9}$$

Here k is the iteration index, $G=G(\boldsymbol{u}^{(k)})$ is defined in (2), and $G'=G'(\boldsymbol{u}^{(k)})$ is defined in (5) as in the noise-reduction method. The matrix C is a given positive definite matrix and it has an $m(N-1)\times m(N-1)$ block diagonal structure $C=\text{blockdiag}\,(C_1,\ldots,C_N)$, where $\{C_n\}_{n=0}^{N}$ plays a role of the model error covariance. The parameter $\alpha^{(k)}>0$ is called the regularization parameter and it needs to be computed at each iteration k using a specific criterion that we will introduce later. The matrix Σ is a given positive definite matrix that has the following form:

$$\Sigma:=H^{T}EH+H^{\perp}WH^{\perp}, \tag{10}$$

where the operator $H^{\perp}$ is defined in (7). The observation error covariance matrix E has a $dN\times dN$ block diagonal structure $E=\text{blockdiag}\,(E_1,\ldots,E_N)$ and the positive definite matrix W has an $mN\times mN$ block diagonal structure $W=\text{blockdiag}\,(W_1,\ldots,W_N)$. The matrices $\{W_n\}_{n=0}^{N}$ are defined in terms of a tuning parameter. Later we will show that they depend on the background trajectory error.

In order to observe why the matrix C plays a role of the model error covariance in (g), we rewrite the solution $\boldsymbol{u}^{(k+1)}$ to (9) as a minimizer of a cost function, namely,

$$\boldsymbol{u}^{(k+1)}=\min_{\boldsymbol{u}}\left[\frac{1}{2}G\left(\boldsymbol{u}\right)^{T}C^{-1}G\left(\boldsymbol{u}\right)+\frac{\alpha^{(k)}}{2}\left(\boldsymbol{u}-\boldsymbol{u}^{(k)}\right)^{T}\Sigma^{-1}\left(\boldsymbol{u}-\boldsymbol{u}^{(k)}\right)\right], \tag{11}$$

where we assumed matrices C and Σ are invertible. Here we observe that the L2-norm $\|G(\boldsymbol{u})\|$ is minimized up to the model error covariance C and $\|\boldsymbol{u}-\boldsymbol{u}^{(k)}\|$ is minimized up to the initial error Σ. Comparing the cost function (11) to the WC4DVar cost function (3), we observe that the term $\|G(\boldsymbol{u})\|_C^2$ is identical in both cost functions. However, WC4DVar minimizes the L2-norm with respect to observations, while the regularized shadowing DA method minimizes the L2-norm with respect to the previous iteration.

3.1. Local convergence. In this section we provide a necessary condition for an estimate obtained by the regularized shadowing DA method to be close to observations and prove local convergence of the algorithm to the solution manifold. Being close to observation means that the estimate remains in a ball centered at observations with a radius that depends on the observation error. Note that if the observation noise is small, then this result combined with the shadowing property guaranties that the estimate is in the vicinity of the true trajectory.

Let us rewrite the regularized shadowing DA method in the limit of a continuous algorithmic time step. This is done in the spirit of analysis of the ensemble Kalman filter based

on the optimization viewpoint [40, 41]. Assume we can set $\alpha^{(k)} = \alpha^{(0)}$ for all k. Then we introduce notation $h = 1/\alpha$ and rewrite (9) in terms of h

$$\boldsymbol{u}^{(k+1)} = \boldsymbol{u}^{(k)} + h\boldsymbol{\Psi}^{(k)}, \quad \boldsymbol{\Psi}^{(k)} = -\Sigma G'^{T}\left(hG'\Sigma G'^{T} + C\right)^{-1} G.$$

Then taking the limit of $h \to 0$, we get on $\tau \in [0, 1]$

$$\frac{d\boldsymbol{u}}{d\tau} = \psi(\boldsymbol{u}), \quad \text{with} \quad \psi(\boldsymbol{u}) = -\Sigma G'^{T}(\boldsymbol{u})C^{-1}G(\boldsymbol{u}), \quad \text{and} \quad \boldsymbol{u}(0) = \boldsymbol{u}^{0}. \tag{12}$$

Stationary points of (12) only occur if $G(\boldsymbol{u}) = 0$ or if $G'^{T}(\boldsymbol{u})$ is not full rank. Defining $\Phi(\boldsymbol{u}) = \|C^{-1/2}G(\boldsymbol{u})\|^2/2$, the ODE (12) becomes

$$\frac{d\boldsymbol{u}}{d\tau} = -\Sigma\nabla\Phi(\boldsymbol{u}). \tag{13}$$

This is a preconditioned gradient descent for $\Phi(\cdot)$ with a preconditioner Σ. We recall that Σ incorporates dependence upon the observation covariance matrix E and weighting matrix W (10).

Assuming Σ is invertible, it holds that

$$\frac{d}{d\tau}\frac{1}{2}\|C^{-1/2}G(\boldsymbol{u})\|^2 = \frac{d}{d\tau}\Phi(\boldsymbol{u}(\tau)) = \left(\frac{d\Phi}{d\boldsymbol{u}}\right)^{T}\frac{d\boldsymbol{u}}{d\tau} = -\left\|\Sigma^{-1/2}\frac{d\boldsymbol{u}}{d\tau}\right\|^2 \le 0,$$

which gives an a priori bound on $\|C^{-1/2}G(\boldsymbol{u})\|$ but does not give global existence of a solution $\|\boldsymbol{u}\|$.

Lemma 3.1. *Suppose* $\|\nabla\Phi(\boldsymbol{u})\| < 1$. *Furthermore, suppose* $\|H^{\perp}\boldsymbol{u}(1) - H^{\perp}\boldsymbol{x}^{\mathrm{b}}\|_W^2 = \varepsilon^2$ *for* $\varepsilon < 1$. *Then* $\|H\boldsymbol{u}(1) - \boldsymbol{y}\|_E^2 < 1 - \varepsilon^2$.

Before we proceed to the proof, let us note that the assumption $\|H^{\perp}\boldsymbol{u}(1) - H^{\perp}\boldsymbol{x}^{\mathrm{b}}\|_W^2 = \varepsilon^2$ for $\varepsilon < 1$ can be fulfilled by an a posteriori choice of W.

Proof. By multiplying (13) with either H or $H^{\perp}$, integrating from 0 to 1, and then taking the L2-norm, we have

$$\|H\boldsymbol{u}(1) - \boldsymbol{y}\| = \|E\int_0^1 H\nabla\Phi(\boldsymbol{u})d\tau\|, \quad \text{and} \quad \|H^{\perp}\boldsymbol{u}(1) - H^{\perp}\boldsymbol{x}^{\mathrm{b}}\| = \|W\int_0^1 H^{\perp}\nabla\Phi(\boldsymbol{u})d\tau\|.$$

Due to the assumption $\|H^{\perp}\boldsymbol{u}(1) - H^{\perp}\boldsymbol{x}^{\mathrm{b}}\|_W^2 = \varepsilon^2$, we have $\|\int_0^1 H^{\perp}\nabla\Phi(\boldsymbol{u})d\tau\|^2 = \varepsilon^2$.

Since $HH^{\perp} = 0$, by the Pythagorean theorem we have

$$\|\int_0^1 \nabla\Phi(\boldsymbol{u})d\tau\|^2 = \|\int_0^1 H\nabla\Phi(\boldsymbol{u})d\tau\|^2 + \|\int_0^1 H^{\perp}\nabla\Phi(\boldsymbol{u})d\tau\|^2.$$

Since $\|\nabla\Phi(\boldsymbol{u})\| < 1$ by the assumption and $\|\int_0^1 H^{\perp}\nabla\Phi(\boldsymbol{u})d\tau\|^2 = \varepsilon^2$, from the above equality it follows that

$$\|\int_0^1 H\nabla\Phi(\boldsymbol{u})d\tau\|^2 < 1 - \varepsilon^2.$$

This implies that $\|H\boldsymbol{u}(1) - \boldsymbol{y}\|_E^2 < 1 - \varepsilon^2$, and the estimate $H\boldsymbol{u}$ consequently remains in a ball centered at $\boldsymbol{y}$ with radius $\|E(1 - \varepsilon^2)\|$. ∎

Now we prove the local convergence of the regularized shadowing DA method (9) when a model is fully observed. Lemma 3.2 and Theorem 3.3 provide conditions for local convergence and are proven in [3] as if for a class of general iterative schemes. Lemma 3.4 is a new result that holds for fully observed dynamical systems. Let $\mathcal{M} = \{\boldsymbol{u} : \ G(\boldsymbol{u}) = 0\}$ and assume that G' is surjective on $\mathcal{M}$, and then $\mathcal{M}$ is a manifold. We define ϕ as

$$\phi(\boldsymbol{u}) = \boldsymbol{u} - \Sigma G'^T \left(G'\Sigma G'^T + \alpha C\right)^{-1} G,$$

where we drop the iteration notation. We note that

$$D\phi = I - \Sigma G'^T \left(G'\Sigma G'^T + \alpha C\right)^{-1} G' \quad \text{for} \quad \boldsymbol{u} \in \mathcal{M}. \tag{14}$$

Since $\mathcal{M}$ is a manifold, we define the tangent and normal spaces of $\mathcal{M}$ at u as $\mathcal{T}_u\mathcal{M}$ and $\mathcal{N}_u\mathcal{M}$, respectively. We have $\mathcal{T}_u\mathcal{M} \perp \mathcal{N}_u\mathcal{M}$ and $\mathcal{T}_u\mathcal{M} = \ker\left(\Sigma G'^T \left(G'\Sigma G'^T + \alpha C\right)^{-1} G'\right)$ for $\boldsymbol{u} \in \mathcal{M}$.

Lemma 3.2. *$\mathcal{M}$ is a set of fixed points for ϕ and there are no further fixed points near $\mathcal{M}$.*

Theorem 3.3. *Suppose $\mathcal{M}$ is compact and contained in an open set $\mathcal{U}$. Furthermore, suppose $D\phi$ is continuous in $\mathcal{U}$ and $\|D\phi|_{\mathcal{N}_u\mathcal{M}}\| < 1$ for all $\boldsymbol{u} \in \mathcal{M}$. Then the sequence $\boldsymbol{u}^{(k)} = \phi^k(\boldsymbol{u}^{(0)})$ converges for $k \to \infty$ to a point on $\mathcal{M}$ if $\boldsymbol{u}^{(0)}$ is sufficiently near to $\mathcal{M}$.*

Now we prove that $\|D\phi|_{\mathcal{N}_u\mathcal{M}}\| < 1$ for the regularized shadowing DA method for all $\boldsymbol{u} \in \mathcal{M}$ in a specific case when $\Sigma G'^T \left(G'\Sigma G'^T + \alpha C\right)^{-1} G'$ is a positive semidefinite matrix. This is the case when the system is fully observed and the observation covariance matrix is a scalar times the identity. For a partially observed model unless the matrix W is set to ϵI, where $E = \epsilon I$, the matrix $\Sigma G'^T \left(G'\Sigma G'^T + \alpha C\right)^{-1} G'$ is not positive semidefinite. Letting $W = \epsilon I$ contradicts the condition of Lemma 3.1 since W is chosen such that $\|H^{\perp}\boldsymbol{u}(1) - H^{\perp}\boldsymbol{x}^{\mathrm{b}}\|_W^2 = \varepsilon^2$ for $\varepsilon < 1$.

Lemma 3.4. *Suppose $\Sigma\Omega[\alpha I + \Sigma\Omega]^{-1}$ is a positive semidefinite matrix, where $\Omega = G'^T C^{-1} G'$. Furthermore, suppose a positive α satisfies $\alpha > \lambda_{\max}(\Sigma\Omega|_{\mathcal{N}_u\mathcal{M}})/2 - \lambda_{\min}(\Sigma\Omega|_{\mathcal{N}_u\mathcal{M}})$. Then $\|D\phi|_{\mathcal{N}_u\mathcal{M}}\| < 1$ for all $\boldsymbol{u} \in \mathcal{M}$.*

Proof. Using the Sherman–Morrison–Woodbury matrix inversion formula [15] and assuming that $\alpha \neq 0$, we can rewrite (14) as

$$D\phi = I - \Sigma\Omega[\alpha I + \Sigma\Omega]^{-1}. \tag{15}$$

Since $\Sigma\Omega[\alpha I + \Sigma\Omega]^{-1}$ is a positive semidefinite matrix, then $D\phi$ is symmetric. For symmetric matrices the L2-norm is equal to the spectral radius. Thus $\|D\phi\| = \rho(D\phi)$, where the spectral radius $\rho(D\phi)$ is defined as the largest absolute value of eigenvalues of $D\phi$.

The largest absolute value of eigenvalues of $D\phi$ is

$$\rho(D\phi) = \max\{|1 - \lambda_{\max}(\Sigma\Omega[\alpha I + \Sigma\Omega]^{-1})|, |1 - \lambda_{\min}(\Sigma\Omega[\alpha I + \Sigma\Omega]^{-1})|\},$$

where $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ denote the largest and the smallest eigenvalues of a matrix A, respectively. Moreover,

$$0 \leq \lambda_{\max}(\Sigma\Omega[\alpha I + \Sigma\Omega]^{-1}) \leq \lambda_{\max}(\Sigma\Omega)\lambda_{\max}([\alpha I + \Sigma\Omega]^{-1}) = \frac{\lambda_{\max}(\Sigma\Omega)}{\lambda_{\min}(\alpha I + \Sigma\Omega)} = \frac{\lambda_{\max}(\Sigma\Omega)}{\alpha + \lambda_{\min}(\Sigma\Omega)}.$$

By choosing α such that

$$\frac{\lambda_{\max}(\Sigma\Omega)}{\alpha + \lambda_{\min}(\Sigma\Omega)} < 2,$$

we have $|1 - \lambda_{\max}(\Sigma\Omega[\alpha I + \Sigma\Omega]^{-1})| < 1$ for $\lambda_{\max}(\Sigma\Omega[\alpha I + \Sigma\Omega]^{-1}) > 0$.

Furthermore,

$$0 \le \lambda_{\min}(\Sigma\Omega[\alpha I + \Sigma\Omega]^{-1}) \le \lambda_{\max}(\Sigma\Omega[\alpha I + \Sigma\Omega]^{-1}) < 2.$$

Thus we have $|1 - \lambda_{\min}(\Sigma\Omega[\alpha I + \Sigma\Omega]^{-1})| < 1$ for $\lambda_{\min}(\Sigma\Omega[\alpha I + \Sigma\Omega]^{-1}) > 0$.

From (15) it follows that $\mathcal{T}_u\mathcal{M} = \ker\left(\Sigma\Omega[\alpha I + \Sigma\Omega]^{-1}\right)$ for $\boldsymbol{u} \in \mathcal{M}$. Since $\mathcal{T}_u\mathcal{M} \perp \mathcal{N}_u\mathcal{M}$, we have

$$\lambda\left(\Sigma\Omega[\alpha I + \Sigma\Omega]^{-1}|_{\mathcal{N}_u\mathcal{M}}\right) > 0.$$

Therefore, by choosing $\alpha > \lambda_{\max}(\Sigma\Omega|_{\mathcal{N}_u\mathcal{M}})/2 - \lambda_{\min}(\Sigma\Omega|_{\mathcal{N}_u\mathcal{M}})$, we have $\|D\phi|_{\mathcal{N}_u\mathcal{M}}\| < 1$ for all $\boldsymbol{u} \in \mathcal{M}$. ∎

Corollary 3.5. *The sequence $\boldsymbol{u}^{(k)} = \phi^k(\boldsymbol{u}^{(0)})$ defined in (9) converges for $k \to \infty$ to a point on $\mathcal{M}$ if $\boldsymbol{u}^{(0)}$ is sufficiently near to $\mathcal{M}$.*

Proof. The proof directly follows from Theorem 3.3 and Lemma 3.4. ∎

Corollary 3.6. *Suppose G has only one zero. Then for the sequence defined in (9) and a final iteration K, $\boldsymbol{u}^{(K)} = \boldsymbol{\mathcal{X}}$.*

This rather trivial corollary shows that the regularized shadowing DA method converges to the true solution for linear models or convex G^TG. The existence of several zeros of G is equivalent to the problem of several minima of G^TG.

As we have stated earlier the conditions of Lemma 3.4 are not satisfied for partially observed models. Numerical experiments conducted with the L63 and L96 models show that choosing α as

$$\alpha = \Delta t^2 \lambda_{\max}(\Sigma\Omega)/2 \tag{16}$$

provides convergence to the manifold $\mathcal{M}$. However, we do not have a rigorous answer why such α leads to $\|D\phi|_{\mathcal{N}_u\mathcal{M}}\| < 1$ for the regularized shadowing DA method for all $\boldsymbol{u} \in \mathcal{M}$. When computing α, we split the eigenvalue problem over one window length N in N eigenvalue problems over N windows length 1. Then in (16) we use the maximum eigenvalue over N windows. Moreover, to save computational costs we compute α for an initial guess $\boldsymbol{u}^{(0)}$ only and fix the same α throughout the iteration.

3.2. Existing shadowing-type DA methods. In this section we point out differences between the regularized shadowing DA method and the existing shadowing-type DA methods of [3, 13] and of [9]. We write down the methods in terms of function ϕ:

$$\begin{aligned}
\phi^{[3]} &:= u - G'^T\Lambda^{-1}G,\\
\phi^{[13]} &:= u - \gamma G'^T G,\\
\phi^{[9]} &:= u - G'^T(G'G'^T)^{-1}G,\\
\phi^{\star} &:= u - \Sigma G'^T(G'\Sigma G'^T + \alpha C)^{-1}G,
\end{aligned}$$

where $\phi^\star$ stands for ϕ of the Levenberg–Marquardt regularized shadowing DA method introduced in this paper. In $\phi^{[3]}$, Λ is chosen to be the Laplace operator. It is stated that the choice of Λ has great influence on the convergence, though without a rigorous statement whether the Laplace operator is a good choice. Local convergence is proven for the method as if for a class of general iterative schemes. In $\phi^{[13]}$, an algorithmic time step γ is chosen by tuning. For sufficiently small γ convergence of the damped Gauss–Newton method is guaranteed but the convergence rate may be only linear [15]. In $\phi^{[9]}$, the convergence rate is quadratic due to the Gauss–Newton method but the method requires a good initial guess for convergence. In $\phi^\star$, on the one hand the lower bound on α guarantees local convergence, but on the other hand the preconditioner Σ (10) may deteriorate the convergence rate. A good estimation of the true solution depends on the preconditioner Σ.

4. Projected noise-reduction DA method. The projected noise-reduction method [9] uses the Gauss–Newton iterate on the nonstable subspace and synchronization in the stable subspace. In this section we briefly review the tangential subspace decomposition, synchronization, and the projected noise-reduction method.

Let $\{x_n; n = 0, \dots, N\}$ denote an orbit of (1). Then the fundamental matrix equation associated with $\{x_n\}$

$$X_{n+1} = F_n'(x_n)X_n, \quad X_n \in \mathbb{R}^{m\times m}, \quad n = 0, \dots, N-1, \tag{17}$$

can be solved with a time-discrete QR factorization using the modified Gram–Schmidt process [11, 12]

$$Q_{n+1}R_{n+1} = F_n'(x_n)Q_n \quad \text{for} \quad n = 0, \dots, N-1, \tag{18}$$

where $X_0 = Q_0R_0$, $X_1 = F_0'(x_0)Q_0R_0 = Q_1R_1R_0$, $X_2 = F_1'(x_1)X_1 = F_1'(x_1)Q_1R_1R_0 = Q_2R_2R_1R_0$, etc. This procedure is well defined for $Q_n \in \mathbb{R}^{m\times p}$ for $p \le m$ provided $F_n'(x_n)Q_n$ is full rank for all n. The Gram–Schmidt process preserves the ordering of the columns of the Q_n and ensures that the upper triangular $R_n \in \mathbb{R}^{p\times p}$ has positive diagonal elements. The (local) p largest Lyapunov exponents of the orbit $\{x_n\}$ are extracted from the time average of the logarithm of the diagonal of R_n [14]:

$$\lambda_i = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \ln R_n^{(i,i)}, \quad i = 1, \dots, p.$$

The method of construction ensures $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_p$. If $\lambda_p \ge 0 > \lambda_{p+1}$, then the matrix $Q_n^{\mathrm{u}} = (Q_n^{(1)}, \dots Q_n^{(p)})$ provides an orthonormal basis for the nonstrongly stable tangent space at X_n. The computed factors $Q_n^{\mathrm{u}} \in \mathbb{R}^{m\times p}$ determine projections $P_n = Q_n^{\mathrm{u}}Q_n^{\mathrm{u}T}$ onto the nonstable subspace. The Gauss–Newton corrections of the projected noise-reduction DA method are applied only in the nonstable subspace.

The contraction in the stable subspace, on the other hand, is exploited by means of synchronization, e.g., [36, 19, 16, 50]. When partial observations are sufficient to constrain

the unstable subspace, an orbit of the dynamical system can be made to converge exponentially in time to a different, driving orbit. The following coupled driver-response is solved:

$$x_{n+1} = F_n(x_n), \tag{19a}$$

$$z_{n+1} = P_n x_{n+1} + (I - P_n)F_n(z_n). \tag{19b}$$

The manifold $\mathcal{S} = \{(x, z) \in \mathbb{R}^m \times \mathbb{R}^m : x = z\}$ is invariant under these dynamics. When $\mathcal{S}$ attracts a neighborhood of itself, z_n synchronizes with x_n.

5. Tangent space splitting of regularized shadowing DA method. Following [9] we propose tangent space splitting of the regularized shadowing DA method. We start by decomposing the relation (8) into the equivalent system

$$\mathcal{P}G(\boldsymbol{u}^{(k)} + \widehat{\mathcal{P}}\boldsymbol{\Pi}^{(k)} + (I - \widehat{\mathcal{P}})\boldsymbol{\Pi}^{(k)}) = 0,$$
$$(I - \mathcal{P})G(\boldsymbol{u}^{(k)} + \widehat{\mathcal{P}}\boldsymbol{\Pi}^{(k)} + (I - \widehat{\mathcal{P}})\boldsymbol{\Pi}^{(k)}) = 0.$$

Here, $\mathcal{P}$ and $\widehat{\mathcal{P}}$ are block diagonal projection matrices $\mathcal{P} = \text{blockdiag}\,(P_1, \cdots, P_N)$ and $\widehat{\mathcal{P}} = \text{blockdiag}\,(P_0, \ldots, P_N)$, where $P_0, P_1, ..., P_N \in \mathbb{R}^{m\times m}$ are projection matrices onto the non-strongly stable subspace at time levels $n = 1, \ldots, N$, respectively.

Instead of computing the update $\boldsymbol{\Pi}^{(k)}$ by simultaneously solving the above system, we split the iterate into updates in the range and complement of $\widehat{\mathcal{P}}$. We also allow the projection operators $\mathcal{P}$ and $\widehat{\mathcal{P}}$ to be updated in each iteration. In the kth iteration, we first approximate the update in the range of $\widehat{\mathcal{P}}^{(k)}$, neglecting the term $(I-\widehat{\mathcal{P}}^{(k)})\boldsymbol{\Pi}^{(k)}$ in the first equation above and solving

$$\mathcal{P}^{(k)}G(\boldsymbol{u}^{(k)} + \widehat{\mathcal{P}}^{(k)}\boldsymbol{\Pi}^{(k)}) = 0 \tag{20}$$

for ${\boldsymbol{\Pi}_{||}}^{(k)} = \widehat{\mathcal{P}}^{(k)}\boldsymbol{\Pi}^{(k)}$. Next we approximate the update in the complement of $\widehat{\mathcal{P}}^{(k)}$ by solving

$$(I - \mathcal{P}^{(k)})G(\boldsymbol{u}^{(k)} + \widehat{\mathcal{P}}^{(k)}\boldsymbol{\Pi}^{(k)} + (I - \widehat{\mathcal{P}}^{(k)})\boldsymbol{\Pi}^{(k)}) = 0 \tag{21}$$

for ${\boldsymbol{\Pi}_{\perp}}^{(k)} = (I - \widehat{\mathcal{P}}^{(k)})\boldsymbol{\Pi}^{(k)}$. Then the update is computed as $\boldsymbol{u}^{(k+1)} = \boldsymbol{u}^{(k)} + {\boldsymbol{\Pi}_{||}}^{(k)} + {\boldsymbol{\Pi}_{\perp}}^{(k)}$. Expressions (20) and (21) are solved approximately for the components ${\boldsymbol{\Pi}_{||}}^{(k)}$ and ${\boldsymbol{\Pi}_{\perp}}^{(k)}$ as described below.

5.1. Regularized Gauss–Newton step on the nonstable space. Linearization of (20) yields a projected linear system for the update ${\boldsymbol{\Pi}_{||}}^{(k)} = \widehat{\mathcal{P}}^{(k)}\boldsymbol{\Pi}^{(k)}$:

$$\mathcal{P}^{(k)}G'(\boldsymbol{u}^{(k)})\widehat{\mathcal{P}}^{(k)}\boldsymbol{\Pi}^{(k)} = -\mathcal{P}^{(k)}G(\boldsymbol{u}^{(k)}).$$

Supressing the iteration index k for the moment, we define block diagonal matrices $\mathcal{Q} = \text{blockdiag}\,(Q_1, \ldots, Q_N)$, and $\widehat{\mathcal{Q}} = \text{blockdiag}\,(Q_0, \ldots, Q_N)$ and note the relations $\mathcal{Q}\mathcal{Q}^T = \mathcal{P}$, $\mathcal{Q}^T\mathcal{Q} = I$ with analogous expressions for $\widehat{\mathcal{Q}}$. Let $\boldsymbol{\mu} = \widehat{\mathcal{Q}}^T\boldsymbol{\Pi} = \widehat{\mathcal{Q}}^T\widehat{\mathcal{P}}\boldsymbol{\Pi}$, $\widetilde{G}' = \mathcal{Q}^T G'(\boldsymbol{u})\widehat{\mathcal{Q}}$ and $\boldsymbol{b} = \mathcal{Q}^T G(\boldsymbol{u})$. Then the linear system for the update $\boldsymbol{\mu}$ may be written as

$$\widetilde{G}'\boldsymbol{\mu} = -\boldsymbol{b}, \tag{22}$$

where the matrix $\widetilde{G}'$ has the block structure

$$\widetilde{G}' = \begin{bmatrix} -R_0 & I & & & \\ & -R_1 & I & & \\ & & \ddots & \ddots & \\ & & & -R_{N-1} & I \end{bmatrix},$$

and consequently, $\widetilde{G}' \in \mathbb{R}^{Np\times(N+1)p}$. We solve (22) using the Levenberg–Marquardt regularization and define the intermediate update

$$\bar{\boldsymbol{u}}^{(k)} = \boldsymbol{u}^{(k)} + \widehat{\mathcal{Q}}^{(k)}\boldsymbol{\mu}^{(k)}, \quad \boldsymbol{\mu}^{(k)} = -\widetilde{G}'^{T}\left(\widetilde{G}'\widetilde{G}'^{T} + \widetilde{\alpha}^{(k)}\mathcal{Q}^{T}\Sigma^{T/2}C\Sigma^{1/2}\mathcal{Q}\right)^{-1}\mathcal{Q}^{T}G. \tag{23}$$

Denoting by $\widetilde{\Sigma} = \mathcal{Q}^{T}\Sigma\mathcal{Q}$ and $\widetilde{\Omega} = \widetilde{G}'^{T}(\mathcal{Q}^{T}C^{-1}\mathcal{Q})\widetilde{G}'$, we have analogously to (16) $\widetilde{\alpha}^{(k)} = \widetilde{\alpha}^{(0)} = \widetilde{\alpha}$ for all iterations k

$$\widetilde{\alpha} = \Delta t^2 \lambda_{\max}(\widetilde{\Sigma}\widetilde{\Omega})/2. \tag{24}$$

5.2. Synchronization step in the stable space. Given the update $\bar{\boldsymbol{u}}^{(k)}$, corrected in the nonstrongly stable subspace (23), the correction $\boldsymbol{\Pi}_{\perp}{}^{(k)}$ to the stable subspace can be implemented through a forward synchronization integration. This gives the following state update componentwise over the time index n,

$$u_{n+1}^{(k+1)} = P_{n+1}\bar{u}_{n+1}^{(k)} + S(I - P_{n+1})F_n(u_n^{(k+1)}), \quad n = 0, \ldots, N-1, \tag{25}$$

where $S = I$ if $n = 0$ and $S = W$ otherwise. The form of the iteration (25) is identical to that of the receiver equation (19b) in the synchronization process except for the term $S = W$ when $n > 0$. We impose continuity in the stable directions during the full assimilation, also across window boundaries, by setting $u_0^{(k+1)} = v_T$, where v_T is the converged iterate u at the terminal time on the previous assimilation time window. Therefore, we assume that $u_0^{(k+1)}$ is a denoised estimate. In turn, the iteration (25) is a noise-free forward synchronization integration at $n = 0$, meaning that $S = I$. However, when $n > 0$ the initial guess is contaminated with error and thus we need to choose S different from I. We set $S = W$, where W is a matrix consisting of tuning parameters.

As in the projected noise-reduction DA method, here better results are obtained by switching after each projected regularized Gauss–Newton iterate to the synchronization step as opposed to switching to the synchronization after the projected regularized Gauss–Newton has converged to tolerance. We attribute this to the variation in the projections that are produced.

6. Numerical experiments. We perform numerical experiments with the L63 model [28] and the L96 model [29]. We compare the (projected) regularized shadowing DA method to WC4DVar and PDA. PDA is initialized at an initial guess $\boldsymbol{u}^{(0)}$ and an algorithmic time step is chosen as in [13], namely $\gamma = 0.1$. The maximum number of iterations for all the DA methods considered here is 100. WC4DVar is initialized at a background trajectory $\boldsymbol{x}^{\mathrm{b}}$. The minimization of a cost function is done by a MATLAB built-in Levenberg–Marquardt algorithm and stopping when the relative change in the cost function compared to the initial

value is less than 10^{-6} unless the maximum number of iterations is reached. Model error covariance for WC4DVar is assumed to be $10^{-2}I$, and the background covariance matrix is the identity.

Both the (projected) regularized shadowing DA method and PDA provide an estimation at observation times only. Therefore we use an estimation at observation times as an initial condition for forward model propagation to have an estimation at every time step of numerical discretization. For the (projected) regularized shadowing DA method, model error covariance is assumed to be $C = 10^{-3}I$. Other values such as 10^{-2} and 10^{-4} provide equivalent results to 10^{-3}. We define the weighting matrix $W = w^2 I$ in the preconditioner Σ (10) and perform sensitivity analysis in terms of w. In order to check robustness of the results, we perform 100 numerical experiments with different realizations of truth $\boldsymbol{\mathcal{X}}$, observations $\boldsymbol{y}$, and background trajectory $\boldsymbol{x}^{\mathrm{b}}$.

To analyze the (projected) regularized shadowing DA method and compare it to other methods, we compute the model error, a cost function, and the error with respect to the truth. We define mean over time of G-error as

$$\mathbb{E}^G = \frac{1}{N}\sum_{n=0}^{N-1} G_n^T G_n. \tag{26}$$

Mean over time of estimation error with respect to the truth of observed variables and nonobserved variables is

$$\mathbb{E}^{\mathrm{O}} = \frac{1}{N}\sum_{n=0}^{N-1} \mathbb{E}_n^{\mathrm{O}}, \tag{27a}$$

$$\mathbb{E}^{\mathrm{N}} = \frac{1}{N}\sum_{n=0}^{N-1} \mathbb{E}_n^{\mathrm{N}}, \tag{27b}$$

respectively. Here the errors $\mathbb{E}_n^{\mathrm{O}}$ (of observed variables) and $\mathbb{E}_n^{\mathrm{N}}$ (of nonobserved variables) are defined as

$$\mathbb{E}_n^{\mathrm{O}} = \frac{1}{\mathrm{rank}(H)}(Hu_n - H\mathcal{X}_n)^T(Hu_n - H\mathcal{X}_n), \tag{28a}$$

$$\mathbb{E}_n^{\mathrm{N}} = \frac{1}{\mathrm{rank}(I - H^T H)}[(I - H^T H)(u_n - \mathcal{X}_n)]^T[(I - H^T H)(u_n - \mathcal{X}_n)], \tag{28b}$$

respectively, and n is the numerical time step index. A cost function with respect to the observations is

$$L = \frac{1}{(k_2 - k_1 + 1)\mathrm{rank}(H)}\sum_{k=k_1}^{k_2}(Hu_k - y_k)^T(Hu_k - y_k), \tag{29}$$

where k is the index of the observation time step, $k_1 \geq 0$, and $k_2 \leq N - 1$.

6.1. Application of the regularized shadowing DA method to the Lorenz 63 model. The well-known Lorenz attractor [28] is a chaotic dynamical system commonly used as a test problem for data assimilation algorithms. The L63 model is

$$\dot{x}_1 = \sigma(x_2 - x_1), \quad \dot{x}_2 = x_1(\rho - x_3) - x_2, \quad \dot{x}_3 = x_1 x_2 - \beta x_3, \tag{30}$$

where $\sigma = 10$, $\beta = \frac{8}{3}$, and $\rho = 28$. The differential equations are discretized with a forward Euler scheme with time step $\Delta t = 0.005$. (We have also considered Runge–Kutta fourth order, but since it gives similar results, it is omitted in the paper.) We generate a set of observations by computing a trajectory of L63 on $t \in [0, 100]$, with a spin-up of $[-25, 0]$. Observations are obtained by perturbing a reference (true) trajectory with random Gaussian independent and identically distributed noise with zero mean and covariance $E = 8I$. The observations of the x_1 variable only are drawn every $\Delta t_{\text{obs}} = 0.05$. Then the map F_n (1) corresponds to 10 forward Euler steps. This map is used to define G and the derivatives of this map are needed for the shadowing iteration. The assimilation window is $\Delta t_{\text{ass}} = 5$.

In Figure 1 we display G-error (26) on the left and error with respect to the truth of non-observed variables (27b) on the right as a function of iteration. We remark that small $w = 100$ gives quicker convergence to the manifold $\mathcal{M}$, while large $w = 1000$ requires more iterations to reach the same error on average. However, error with respect to the truth of nonobserved variables is decreasing during iteration for large $w = 1000$, while increasing for small $w = 100$. In Figure 2 we plot error with respect to the truth of observed variables (27a) on the left and cost function (29) on the right as a function of iteration number, where the solid black line is the observation error. We observe again that large $w = 1000$ gives a better estimation of observed variables than small $w = 100$.

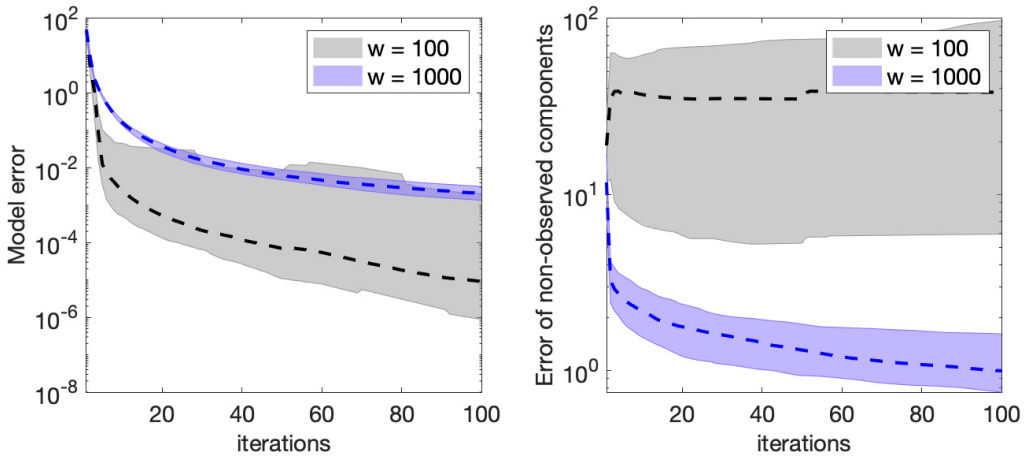

Figure 1. *Application to L63. Error of the regularized shadowing DA method as a function of iterations: median (dashed line) +/− one standard deviation (shadowed area) over* 100 *simulations. In grey error is shown for weighting matrix* $w = 100$*, in blue for* $w = 1000$*. On the left: mean over time of G-error. On the right: mean over time of error with respect to the truth of nonobserved variables.*

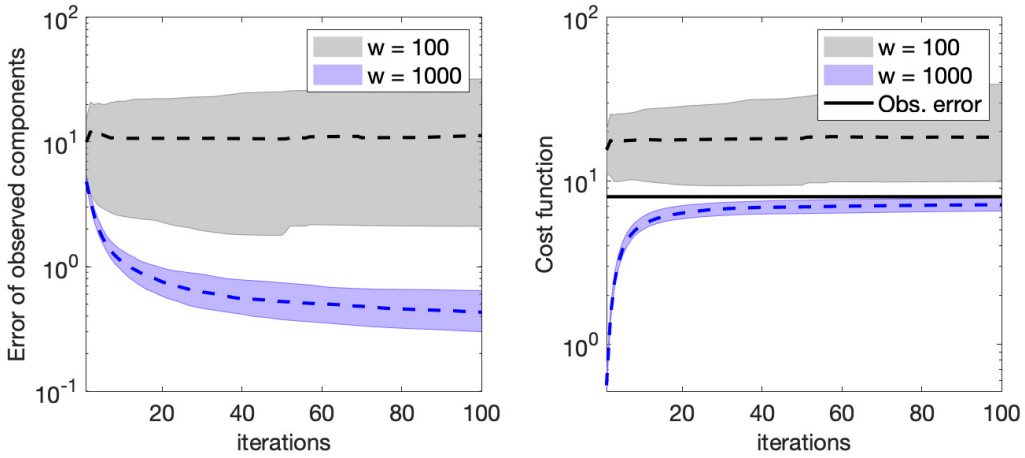

Figure 2. *Application to L63. Error of the regularized shadowing DA method as a function of iterations: median (dashed line) +/− one standard deviation (shadowed area) over* 100 *simulations. In grey error is shown for weighting matrix* $w = 100$*, in blue for* $w = 1000$*. On the left: mean over time of error with respect to the truth of observed variables. On the right: mean over time of cost function of observed variables.*

When analyzing the cost function, we observe that for small $w = 100$ the cost function quickly underestimates the observation error. In inverse problems this phenomenon is often referred to as observation overfitting, though in that context a cost function is decreasing not increasing and the observation error is overestimated not underestimated (see, e.g., [18]). For the regularized shadowing DA method the cost function (29) is zero at the first iteration, because the algorithm is initialized at $\boldsymbol{u}^{(0)}$ (6). The cost function increases during iteration due to a search for a noise-free orbit. When the cost function is larger than the observation error $\|E\|$, an estimate is not in a ball of radius $\|E\|$ centered at the observations, resulting in a larger error with respect to the truth. Therefore, we need to prevent the cost function from becoming larger than $\|E\|$. A classical approach in inverse problems is to stop the iteration when this occurs. In the regularized shadowing DA method this approach is questionable due to the cost function increasing over the course of iteration. Instead, we propose to tune the preconditioner Σ (10), namely, the weighting matrix W, to obtain the correct behavior of the cost function. We observe that the large value of $w = 1000$ results in the cost function approaching the observation error from below. This is an indication of correctly tuned w. Thus the role of preconditioner Σ is to keep descent steps in the direction of observed variables small compared to descent steps in the direction of nonobserved variables. As the iteration proceeds, observed variables get denoised as well and the algorithm finds a (pseudo-)orbit compatible with observations. We would like to stress that the cost function (29) depends only on observations, not the truth.

In Figure 3 we compare the regularized shadowing DA method with $w = 1000$ to WC4DVar and PDA, where we plot error with respect to the truth over time of observed variables (28a) and of nonobserved variables (28b) on the left and right, respectively. We observe that the regularized shadowing DA method with tuned w outperforms both WC4DVar and PDA.

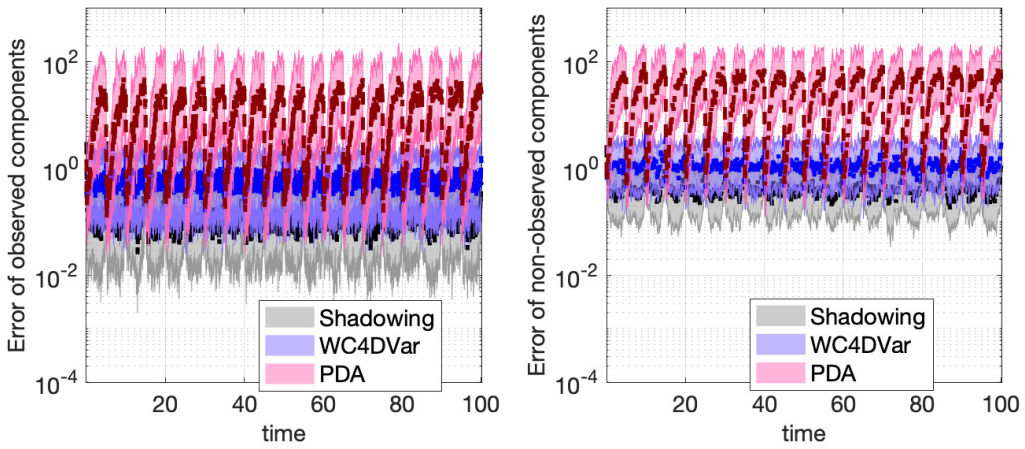

Figure 3. *Application to L63. Error as a function of time: median (dashed line) +/− one standard deviation (shadowed area) over* 100 *simulations. On the left: error with respect to the truth of observed variables. On the right: error with respect to the truth of nonobserved variables. The regularized shadowing DA method with* $w = 1000$ *in grey, WC4DVar in blue, and PDA in pink.*

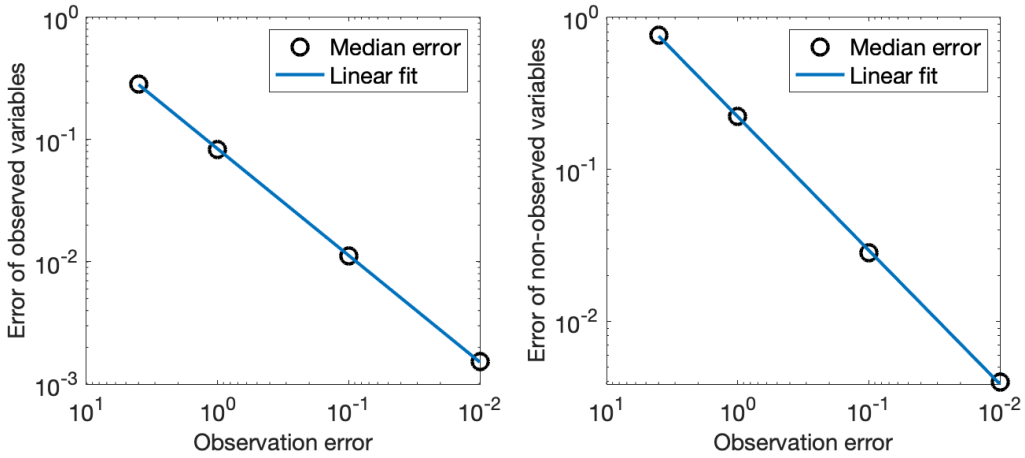

Figure 4. *Regularized shadowing DA method applied to L63. Median error over* 100 *simulations as a function of observation error. The median error is depicted as circles and the linear fit as a blue curve. On the left: error with respect to the truth of observed variables with the linear fit coefficient* 0.87. *On the right: error with respect to the truth of nonobserved variables with the linear fit coefficient* 0.88.

Finally, we investigate the convergence of the algorithm in the small noise limit. We perform the numerical experiments with E equal to $4I$, I, $0.1I$, and $0.01I$. On the left in Figure 4, we plot the median over 100 simulations of the error with respect to observed components (28a) as a function of the observation error and compute the linear fit. The algorithm converges and shows the order of convergence 0.87. We also investigate the performance of the

algorithm in terms of the median error of nonobserved components (28b) shown on the right in Figure 4. Even though we do not expect to see a clear convergence here in terms of the observation noise, since the observation noise enters implicitly in the estimation of the nonobserved components, the median error of nonobserved components shows clear convergence of the algorithm with the order of convergence 0.88.

6.2. Application of the regularized shadowing DA method to the Lorenz 96 model. Lorenz [29] proposed the following model as an example of a simple one-dimensional model with features of the atmosphere. The L96 model is

$$\dot{x_l} = -x_{l-2}x_{l-1} + x_{l-1}x_{l+1} - x_l + \mathcal{F}, \qquad (l = 1, ..., d), \tag{31}$$

where the dimension d and forcing $\mathcal{F}$ are parameters. Cyclic boundary conditions are imposed. We implement the L96 model with the parameter choices $d = 36$ and $\mathcal{F} = 8$. The differential equations are discretized with a forward Euler scheme with time step $\Delta t = 0.005$. (We have also considered Runge–Kutta fourth order but since it gives similar results, it is omitted in the paper.) We generate a set of observations computing a trajectory of L96 on $t \in [0, 100]$, with a spin-up of $[-25, 0]$ for a true trajectory to reside on the attractor. Observations are obtained by perturbing a reference (true) trajectory with random Gaussian independent and identically distributed noise with zero mean and covariance $E = 8I$. The observations of every second variable are drawn every $\Delta t_{\text{obs}} = 0.05$. Then the map F_n (1) corresponds to 10 forward Euler steps. This map is used to define G and the derivatives of this map are needed for the shadowing iteration. The assimilation window is $\Delta t_{\text{ass}} = 5$.

In Figure 5 we display G-error (26) on the left and error with respect to the truth of nonobserved variables (27b) on the right as a function of iteration. As for L63 displayed in Figure 1, large $w = 1000$ requires more iterations to reach the same G-error than small $w = 100$. Error with respect to the truth of nonobserved variables decreases over iterations for large $w = 1000$ while it increases for small $w = 100$.

In Figure 6, we plot error with respect to the truth of observed variables (27a) and cost function (29) as a function of iteration number on the left and on the right, respectively. A better estimation of observed variables is obtained with large $w = 1000$ than with small $w = 100$, as was the case for L63 displayed in Figure 2. Moreover, small $w = 100$ gives a considerable increase in the error. The cost function is underestimated with small $w = 100$ and well estimated with large $w = 1000$.

In Figure 7 we compare the regularized shadowing DA method with $w = 1000$ to WC4DVar and PDA, where we plot error with respect to the truth over time of observed variables (28a) and of nonobserved variables (28b) on the left and right, respectively. Here we observe that the regularized shadowing DA method with correctly chosen preconditioner Σ outperforms both WC4DVar and PDA.

Finally, we investigate the convergence of the algorithm in the small noise limit. We perform the numerical experiments with E equal to $4I$, I, $0.1I$, and $0.01I$. On the left in Figure 8, we plot median over 100 simulations of the error with respect to observed components (28a) as a function of the observation error and compute the linear fit. The algorithm appears to converge and shows the order of convergence 0.74. We also investigate performance of the

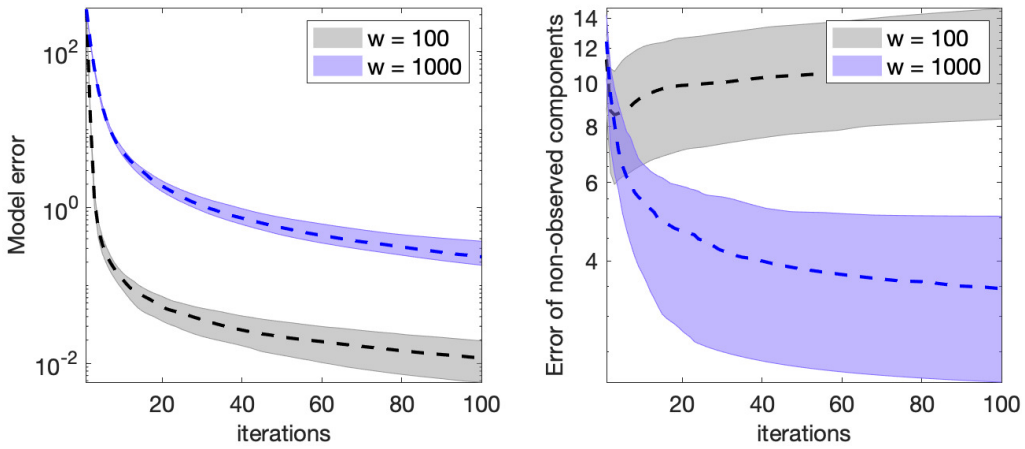

Figure 5. *Application to L96. Error of the regularized shadowing DA method as a function of iterations: median (dashed line) +/− one standard deviation (shadowed area) over* 100 *simulations. In grey error is shown for weighting matrix* $w = 100$*, in blue for* $w = 1000$*. On the left: mean over time of G-error. On the right: mean over time of error with respect to the truth of nonobserved variables.*

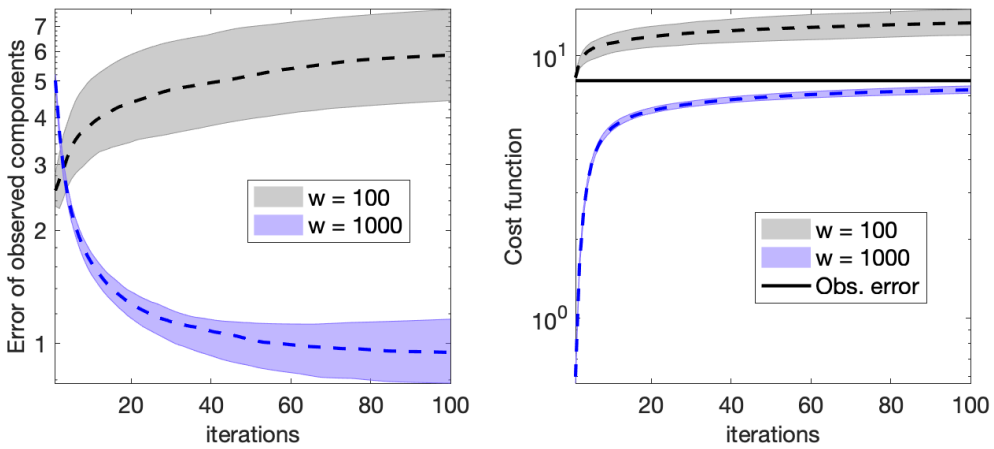

Figure 6. *Application to L96. Error of the regularized shadowing DA method as a function of iterations: median (dashed line) +/− one standard deviation (shadowed area) over* 100 *simulations. In grey error is shown for weighting matrix* $w = 100$*, in blue for* $w = 1000$*. On the left: mean over time of error with respect to the truth of observed variables. On the right: mean over time of cost function of observed variables.*

algorithm in terms of the median error of nonobserved components (28b) shown on the right in Figure 8. Here we observe the error decrease as well, though the error stalls for observation errors smaller than 10^{-1}.

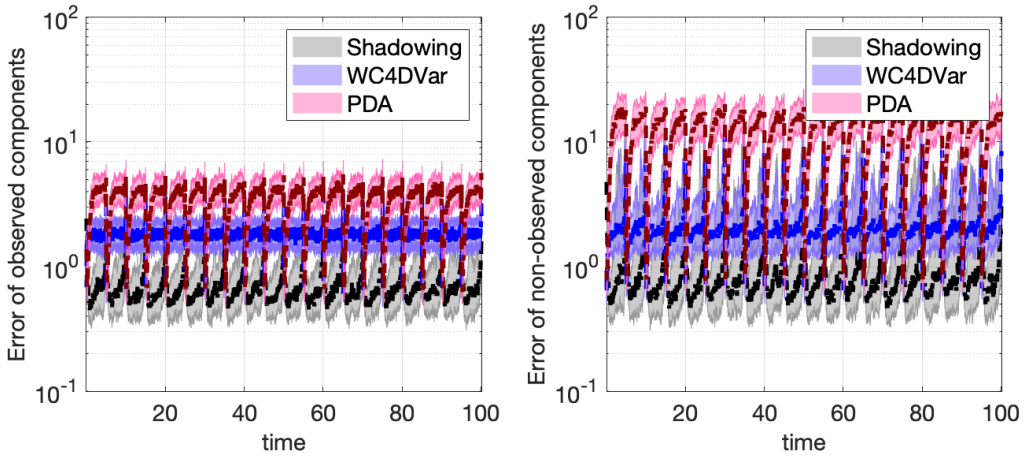

Figure 7. *Application to L96. Error as a function of time: median (dashed line) +/− one standard deviation (shadowed area) over* 100 *simulations. On the left: error with respect to the truth of observed variables. On the right: error with respect to the truth of nonobserved variables. The regularized shadowing DA method with* $w = 1000$ *in grey, WC4DVar in blue, and PDA in pink.*

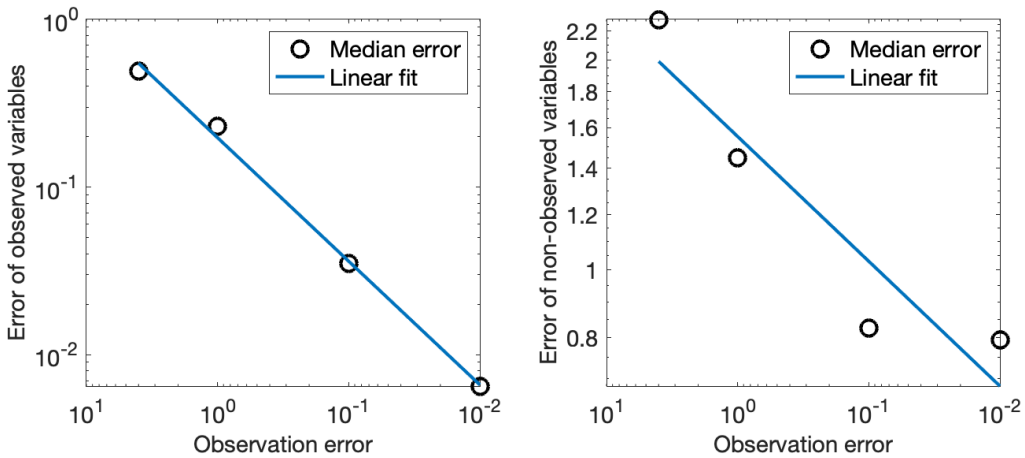

Figure 8. *Regularized shadowing DA method applied to L96. Median error over* 100 *simulations as a function of observation error. The median error is depicted as circles and the linear fit as blue curve. On the left: error with respect to the truth of observed variables with the linear fit coefficient* 0.74. *On the right: error with respect to the truth of nonobserved variables.*

6.3. Sensitivity of W with respect to the background trajectory. Let us investigate the relation between the background trajectory $\boldsymbol{x}^{\mathrm{b}}$ and W in Σ (10). One can consider W as a confidence in the background trajectory $\boldsymbol{x}^{\mathrm{b}}$. Indeed, let us construct the background trajectory $\boldsymbol{x}^{\mathrm{b}}$ by perturbing the true trajectory with the observation noise. Thus we have a fully observed system with observation error $E = 8I$. Since $W = w^2 I$, we set $w = \sqrt{8}$ to

have $W = 8I$. We perform numerical experiments with the regularized shadowing method with both the L63 model and the L96 model. We observe that the cost function becomes larger than the observation noise level at some iteration. If we set $w = 100$, then the cost function is below the noise level and the estimate has a smaller error with respect to the truth compared to the case when $w = \sqrt{8}$ (not shown). We attribute this to an estimation being more accurate when not all variables get denoised at once.

We note that when the background trajectory is a trajectory propagated forward in time from an arbitrary initial condition, then on average the error (28b) for $N = 1000$ is equal to 60 and 20 for L63 and L96, respectively. We recall that in that case $w = 1000$ led to the most accurate estimate. If we decrease the assimilation window by 10, then on average the error (28b) for $N = 100$ is equal to 1 and 3 for L63 and L96, respectively. Numerical experiments indicated that in this case $w = \sqrt{10}$ leads to the most accurate results. Therefore, W can be chosen smaller for a more accurate background trajectory. However, w still remains a tuning parameter. On the other hand, to tune it we only require the cost function (29) that does not depend on the unknown truth.

We note that the background trajectory can also be chosen as a trajectory propagated forward in time not from an arbitrary initial condition but from the past estimate. This will decrease error in the background trajectory and lead to smaller values of w.

6.4. Application of the projected regularized shadowing DA method. Next we perform numerical experiments with the projected regularized shadowing DA method. In [9], it was argued that the dimension of the nonstable subspace on which the Gauss–Newton iteration is performed needs to be larger than the dimension of the unstable and neutral subspaces combined. The projection dimensions $p = 2$ and $p = 25$ for L63 and L96, respectively, gave satisfactory results in numerical experiments in [9]. Therefore, we use the same projection dimensions p for the projected regularized shadowing DA method. We compare the projected regularized shadowing DA method to the noise-reduction method of [9]. We note that the assimilation window needs to be decreased by 10 for both methods to converge. For the L63 model, such an assimilation window corresponds to a solution being on one wing of the attractor for the majority of the time. For the L96 model, such an assimilation window is equivalent to 60 hours. Furthermore, the numerical experiments showed that $w = \sqrt{0.1}$ leads to the cost function below the noise level. We attribute the smaller value of $w = \sqrt{0.1}$ compared to $w = \sqrt{10}$ for the nonprojected regularized shadowing DA method to smaller errors due to tangential splitting.

In Figures 9 and 10, we plot error with respect to the truth over time for L63 and L96, respectively. On the left is the error of observed variables (28a) and on the right is the error of nonobserved variables (28b). First, we observe that the projected regularized shadowing DA method depicted in grey and the noise-reduction method depicted in pink perform comparably for both models. This also holds for PDA (not shown). Next, we note that the shadowing-type methods outperform WC4DVar. For the observed components the improvement is more significant than for the nonobserved ones, as well as for the L96 model than for the L63 model. We can conclude that the projection operators $\mathcal{P}$ and $\widehat{\mathcal{P}}$ in (20) and (21), respectively, are sufficiently accurate when computed based on partial observations and a background trajectory.

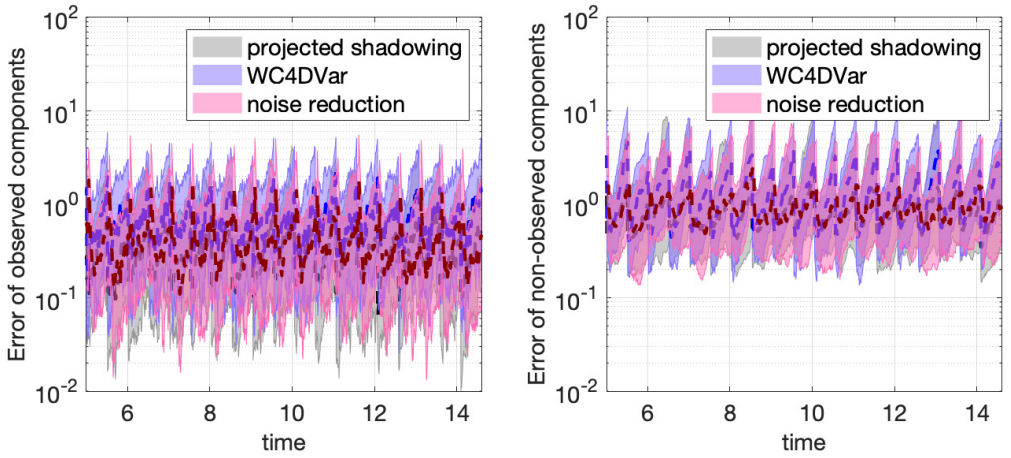

Figure 9. *Application to L63 with the assimilation window* $\Delta t_{ass} = 0.5$. *Error as a function of time: median (dashed line) +/− one standard deviation (shadowed area) over* 100 *simulations. On the left: error with respect to the truth of observed variables. On the right: error with respect to the truth of nonobserved variables. The projected regularized shadowing DA method with* $p = 2$ *in grey, WC4DVar in blue, and the noise-reduction DA method in pink.*

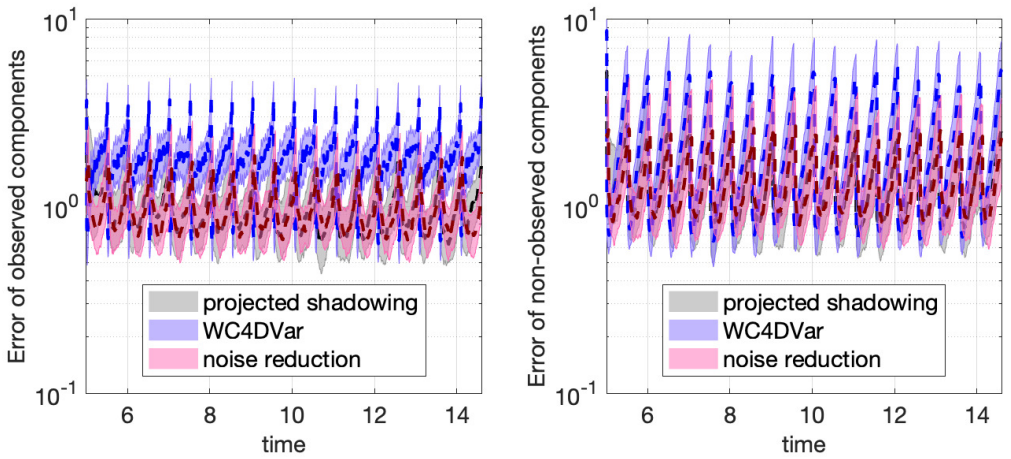

Figure 10. *Application to L96 with the assimilation window* $\Delta t_{ass} = 0.5$. *Error as a function of time: median (dashed line) +/− one standard deviation (shadowed area) over* 100 *simulations. On the left: error with respect to the truth of observed variables. On the right: error with respect to the truth of nonobserved variables. The projected regularized shadowing DA method with* $p = 25$ *in grey, WC4DVar in blue, and the noise-reduction DA method in pink.*

7. Conclusions. We introduced a shadowing-type data assimilation method for partial observations. This method does not require any preprocessing to obtain a proxy of observations, unlike the existing shadowing-type data assimilation methods. The method is derived from the Levenberg–Marquardt regularization of a shadowing-type DA method of [9]. We prove its local convergence and obtain a lower bound for the algorithmic time step required

for the method to converge to the manifold $G(\boldsymbol{u}) = 0$. The regularized shadowing DA method incorporates a preconditioner. The preconditioner scales the descent steps such that they are larger for nonobserved variables compared to the observed ones. This allows the algorithm to find a solution $\boldsymbol{u}$ of $G(\boldsymbol{u}) = 0$ in the vicinity of the truth. Numerical experiments with the L63 and L96 models show encouraging results: the regularized shadowing DA method outperforms both WC4Var and PDA. However, the regularized shadowing DA method is more expensive than these methods, since it requires finding eigenvalues at the first iteration and forming large matrices and inverting them.

To decrease the computational costs of the regularized shadowing DA method, we propose the projected regularized shadowing DA method following [9]. The method is based on tangential splitting into nonstable and stable subspaces. It consists of the regularized shadowing DA method on the nonstable subspace and the synchronization step on the stable subspace. Numerical experiments with the L63 and L96 models show that the projected regularized shadowing DA method outperforms WC4Var as well. However, the assimilation window needs to be smaller for the method to converge. Then on such an assimilation window both PDA and the noise-reduction DA method of [9] provide results similar to the projected regularized shadowing DA method, without any preprocessing.

REFERENCES

[1] M. Berliner, *Likelihood and Bayesian prediction for chaotic systems*, J. Amer. Staist. Assoc., 86 (1991), pp. 938–952.

[2] M. Bocquet and A. Carrassi, *Four-dimensional ensemble variational data assimilation and the unstable subspace*, Tellus A, 69 (2017), 1304504, https://doi.org/10.1080/16000870.2017.1304504.

[3] J. Bröcker and U. Parlitz, *Efficient noncausal noise reduction for deterministic time series*, Chaos, 11 (2001), pp. 319–326, https://doi.org/10.1063/1.1357454.

[4] Y. Chen and D. Oliver, *Levenberg–Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification*, Comput Geosci, 17 (2013), pp. 689–703, https://doi.org/10.1007/s10596-013-9351-5.

[5] S.-N. Chow and E. S. Van Vleck, *A shadowing lemma for random diffeomorphisms*, Random Comput. Dyn., 1 (1992), pp. 197–218.

[6] B. A. Coomes, H. Koçak, and K. J. Palmer, *Shadowing orbits of ordinary differential equations*, J. Comput. Appl. Math., 52 (1994), pp. 35–43.

[7] B. A. Coomes, H. Koçak, and K. J. Palmer, *Rigorous computational shadowing of orbits of ordinary differential equations*, Numer. Math., 69 (1995), pp. 401–421.

[8] B. A. Coomes, H. Koçak, and K. J. Palmer, *A shadowing theorem for ordinary differential equations*, Z. Angew. Math. Phys., 46 (1995), pp. 85–106.

[9] B. de Leeuw, S. Dubinkina, J. Frank, A. Steyer, X. Tu, and E. V. Vleck, *Projected shadowing-based data assimilation*, SIAM J. Appl. Dyn. Syst., 17 (2018), pp. 2446–2477, https://doi.org/10.1137/17M1141163.

[10] E. de Sturler and M. E. Kilmer, *A regularized Gauss–Newton trust region approach to imaging in diffuse optical tomography*, SIAM J. Sci. Comput., 33 (2011), pp. 3057–3086, https://doi.org/10.1137/100798181.

[11] L. Dieci and E. S. Van Vleck, *Lyapunov and Sacker–Sell spectral intervals*, J. Dynam. Differential Equations, 19 (2007), pp. 265–293, https://doi.org/10.1007/s10884-006-9030-5.

[12] L. Dieci and E. S. Van Vleck, *Lyapunov exponents: Computation*, in Encyclopedia of Applied and Computational Mathematics, B. Engquist, ed., Springer, New York, 2015, pp. 834–838, https://doi.org/10.1007/978-3-540-70529-1_421.

[13] H. Du and L. A. Smith, *Pseudo-orbit data assimilation. Part* I: The perfect model scenario, J. Atmos. Sci., 71 (2014), pp. 469–482.

[14] S. V. Ershov and A. B. Potapov, *On the concept of stationary Lyapunov basis*, Phys. D, 118 (1998), pp. 167–198, https://doi.org/10.1016/S0167-2789(98)00013-X.
[15] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, MD, 1996.
[16] C. González-Tokman and B. R. Hunt, *Ensemble data assimilation for hyperbolic systems*, Phys. D, 243 (2013), pp. 128–142.
[17] C. Grebogi, S. Hammel, J. Yorke, and T. Sauer, *Shadowing of physical trajectories in chaotic dynamics: Containment and refinement*, Phys. Rev. Lett., 65 (1990), pp. 1527–1530.
[18] M. Hanke, *A regularizing Levenberg-Marquardt scheme, with applications to inverse groundwater filtration Problems*, Inverse Problems, 13 (1997), 79.
[19] K. Hayden, E. Olson, and E. S. Titi, *Discrete data assimilation in the Lorenz and 2D Navier–Stokes equations*, Phys. D, 240 (2011), pp. 1416–1425, https://doi.org/10.1016/j.physd.2011.04.021.
[20] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Math. Sci. Eng., Academic Press, New York, 1970.
[21] K. Judd, C. A. Reynolds, T. E. Rosmond, and L. A. Smith, *The geometry of model error*, J. Atmos. Sci., 65 (2008), pp. 1749–1772, https://doi.org/10.1175/2007JAS2327.1.
[22] K. Judd and L. Smith, *Indistinguishable states* I. *Perfect model scenario*, Phys. D, 151 (2001), pp. 125–141.
[23] A. Katok and B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems*, Encyclopedia Math. Appl., 54, Cambridge University Press, Cambridge, 1995.
[24] K. Law, D. Sanz-Alonso, A. Shukla, and A. Stuart, *Controlling Unpredictability with Observations in the Partially Observed Lorenz '96 Model*, https://arxiv.org/abs/1411.3113, 2014.
[25] K. Law, A. Stuart, and K. Zygalakis, *Data assimilation: A mathematical introduction*, Texts in Appl. Math., Springer, New York, 2015, https://doi.org/10.1007/978-3-319-20325-6.
[26] K. Levenberg, *A method for the solution of certain non-linear problems in least squares*, Quart. Appl. Math., 2 (1944), pp. 164–168, https://doi.org/10.1090/qam/10666.
[27] J. Lewis and J. Derber, *The use of adjoint equations to solve a variational adjustment problem with advective constraint*, Tellus A, 37 (1985), pp. 309–322.
[28] E. N. Lorenz, *Deterministic Nonperiodic Flow*, J. Atmos. Sci., 20 (1963), pp. 130–148, https://doi.org/10.1175/1520-0469(1963)020⟨0130:DNF⟩2.0.CO;2.
[29] E. N. Lorenz, *Predictability—A problem partly solved*, in Proceedings of Seminar on Predictability, T. Palmer and R. Hagedorn, eds., Reading, UK, Cambridge University Press, Cambridge, 1996, pp. 1–18.
[30] J. Mandel, E. Bergou, S. Gürol, S. Gratton, and I. Kasanický, *Hybrid Levenberg–Marquardt and weak-constraint ensemble Kalman smoother method*, Nonlinear Process. Geophys., 23 (2016), pp. 59–73, https://doi.org/10.5194/npg-23-59-2016.
[31] D. W. Marquardt, *An algorithm for least-squares estimation of nonlinear parameters*, J. Soc. Indust. Appl. Math., 11 (1963), pp. 431–441, https://doi.org/10.1137/0111030.
[32] R. N. Miller, M. Ghil, and F. Gauthiez, *Advanced data assimilation in strongly nonlinear dynamical systems*, J. Atmos. Sci., 51 (1994), pp. 1037–1056.
[33] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, New York, 2006.
[34] L. Palatella, A. Carrassi, and A. Trevisan, *Lyapunov vectors and assimilation in the unstable subspace: Theory and applications*, J. Phys. A, 46 (2013), 254020.
[35] K. J. Palmer, *Exponential dichotomies, the shadowing lemma and transversal homoclinic points*, in Dynamics Reported, U. Kirchgraber and H.-O. Walther, eds., Vieweg+Teubner Verlag, Wiesbaden, 1988, pp. 265–306, https://doi.org/10.1007/978-3-322-96656-8_5.
[36] L. M. Pecora and T. L. Carroll, *Synchronization in chaotic systems*, Phys. Rev. Lett., 64 (1990), pp. 821–824.
[37] C. Pires, R. Vautard, and O. Talagrand, *On extending the limits of variational assimilation in nonlinear chaotic systems*, Tellus A, 48 (1996), pp. 96–121.
[38] D. Sanz-Alonso and A. M. Stuart, *Long-time asymptotics of the filtering distribution for partially observed chaotic dynamical systems*, SIAM/ASA J. Uncertain. Quantific., 3 (2015), pp. 1200–1220, https://doi.org/10.1137/140997336.
[39] Y. Sasaki, *Some basic formalisms in numerical variational analysis*, Monthly Wea. Rev., 98 (1970), pp. 875–883.

[40] C. Schillings and A. M. Stuart, *Analysis of the ensemble Kalman filter for inverse problems*, SIAM J. Numer. Anal., 55 (2017), pp. 1264–1290, https://doi.org/10.1137/16M105959X.
[41] C. Schillings and A. M. Stuart, *Convergence analysis of ensemble Kalman inversion: The linear, noisy case*, Appl. Anal., 97 (2018), pp. 107–123, https://doi.org/10.1080/00036811.2017.1386784.
[42] T. Stemler and K. Judd, *A guide to using shadowing filters for forecasting and state estimation*, Phys. D, 238 (2009), pp. 1260–1273, https://doi.org/10.1016/j.physd.2009.04.008.
[43] J. Y. T. Sauer, *Rigorous verification of trajectories for the computer simulation of dynamical systems*, Nonlinearity, 4 (1991), pp. 961–979.
[44] O. Talagrand, *Assimilation of observations, an introduction*, J. Meteorol. Soc. Japan, 75 (1997), pp. 191–209.
[45] O. Talagrand and P. Courtier, *Variational assimilation of meteorological observations with the adjoint vorticity equation*, Quart. J. Roy. Meteorol. Soc., 113 (1987).
[46] A. Trevisan, M. D'Isidoro, and O. Talagrand, *Four-dimensional variational assimilation in the unstable subspace and the optimal subspace dimension*, Quart. J. Roy. Meteorol. Soc., 136 (2010), pp. 487–496, https://doi.org/10.1002/qj.571.
[47] A. Trevisan and F. Uboldi, *Assimilation of standard and targeted observations within the unstable subspace of the observation–analysis–forecast cycle system*, J. Atmos. Sci., 61 (2004), pp. 103–113, https://doi.org/10.1175/1520-0469(2004)061⟨0103:AOSATO⟩2.0.CO;2.
[48] E. S. Van Vleck, *Numerical shadowing near hyperbolic trajectories*, SIAM J. Sci. Comput., 16 (1995), pp. 1177–1189, https://doi.org/10.1137/0916068.
[49] S. Vannitsem and V. Lucarini, *Statistical and dynamical properties of covariant Lyapunov vectors in a coupled atmosphere-ocean model—Multiscale effects, geometric degeneracy, and error dynamics*, J. Phys. A, 49 (2016), 224001, https://doi.org/10.1088/1751-8113/49/22/224001.
[50] S. Zhuk and J. Frank, *A detectability criterion and data assimilation for non-linear differential equations*, Nonlinearity, 31 (2018), 5235, https://doi.org/10.1088/1361-6544/aaddcb.