Convergence properties of a data-assimilation method based on a Gauss-Newton iteration

Nazanin Abedini, Svetlana Dubinkina

Vrije Universiteit Amsterdam

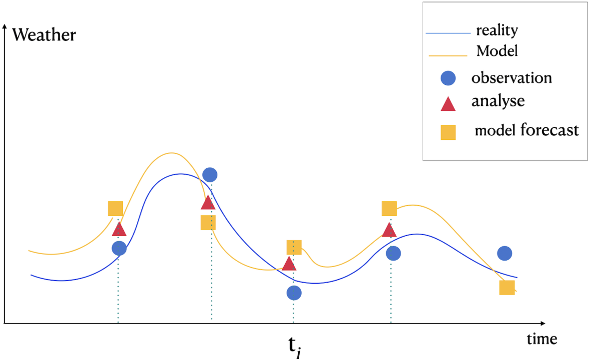
31-05-2022

Introduction

Outline

- Data assimilation framework
- Observations and data measurements
- Variational methods
- Main theorems
- Numerical experiments

Data Assimilation Framework



$$J(\blacktriangle) = \frac{1}{2}\left\| \blacktriangle - \blacksquare \right\|_B^2 + \frac{1}{2}\left\| \bullet - H(\blacktriangle) \right\|_R^2$$

Problem Formulation

Consider the following nonlinear problem

$$\frac{dx}{dt} = F(x), \quad x(t) \in \mathbb{R}^n, \; t \in [0, T].$$

Problem Formulation

Consider the following nonlinear problem

$$\frac{dx}{dt} = F(x), \quad x(t) \in \mathbb{R}^n, \ t \in [0, T].$$

Let $0 = t_0 < t_1 < \cdots < t_{N-1} = T$ be an equidistant partition of $I = [0, T]$ with $t_j = j\Delta t$. Here we consider the Euler time-discretization:

$$x_{j+1} = F(x_j), \quad x_j \in \mathbb{R}^n, \quad j = 0, \ldots, N-1,$$

where $x_{j+1} = x(t_{j+1})$.

Problem Formulation (continued)

- Let the sequence $\mathbf{x} := \{x_0, \ldots, x_N\}$ be a true solution of the dynamical model and presumed to be unknown.
- Suppose we are given a sequence of observations y_j related to x_j

$$y_j = Hx_j + \eta_j, \quad y_j \in \mathbb{R}^b, \quad j = 0, \ldots, N,$$

where $H : \mathbb{R}^n \to \mathbb{R}^b$, $b \leq n$ is the observation operator and $\eta_j \sim \mathcal{N}(0, \Gamma) = (0, \gamma^2 I)$.

Problem Formulation (continued)

- Let the sequence $\mathbf{x} := \{x_0, \ldots, x_N\}$ be a true solution of the dynamical model and presumed to be unknown.
- Suppose we are given a sequence of observations y_j related to x_j

$$y_j = Hx_j + \eta_j, \quad y_j \in \mathbb{R}^b, \quad j = 0, \ldots, N,$$

where $H : \mathbb{R}^n \to \mathbb{R}^b$, $b \leq n$ is the observation operator and $\eta_j \sim \mathcal{N}(0, \Gamma) = (0, \gamma^2 I)$.

The goal is looking for the approximation solution of dynamical model, $\mathbf{u} = \{u_0, \ldots, u_N\}$ such that

$$\|\mathbf{u} - \mathbf{x}\| \to 0$$

Variational Methods

Variational data assimilation methods lead to the minimization of a cost function involving quadratic forms based on the both the background and observation covariance matrices.

Variational Methods

Variational data assimilation methods lead to the minimization of a cost function involving quadratic forms based on the both the background and observation covariance matrices.

- 3D-Var

$$J(x) = \frac{1}{2}(x^b - x)^T B^{-1}(x^b - x) + \frac{1}{2}(y - Hx)^T R^{-1}(y - Hx) \quad (1)$$

Variational Methods

Variational data assimilation methods lead to the minimization of a cost function involving quadratic forms based on the both the background and observation covariance matrices.

- 3D-Var

$$J(x) = \frac{1}{2}(x^b - x)^T B^{-1}(x^b - x) + \frac{1}{2}(y - Hx)^T R^{-1}(y - Hx) \quad (1)$$

- 4D-Var

$$J(x_0) = \frac{1}{2}(x^b - x_0)^T B^{-1}(x^b - x_0) + \frac{1}{2}\sum_{j=0}^{N-1}(y_j - Hx_j)^T R^{-1}(y_j - Hx_j)$$

s.t.

$$x_j = F_{0,j}(x_0) \quad (2)$$

Variational Methods

Variational data assimilation methods lead to the minimization of a cost function involving quadratic forms based on the both the background and observation covariance matrices.

- 3D-Var

$$J(x) = \frac{1}{2}(x^b - x)^T B^{-1}(x^b - x) + \frac{1}{2}(y - Hx)^T R^{-1}(y - Hx) \quad (1)$$

- 4D-Var

$$J(x_0) = \frac{1}{2}(x^b - x_0)^T B^{-1}(x^b - x_0) + \frac{1}{2}\sum_{j=0}^{N-1}(y_j - Hx_j)^T R^{-1}(y_j - Hx_j)$$

s.t.

$$x_j = F_{0,j}(x_0) \quad (2)$$

J: Penalty (Fit to background+ Fit to observations)
x^b: Background solution
B: Background error covariance
R: Observation error covariance

Variational Methods (continued)

If the dynamical model is imperfect, the cost function contains an extra term which is corresponding to the model error.

$$J(u_0; y_j) = \frac{1}{2}(u_0 - x^b)^T B^{-1}(u_0 - x^b) + \frac{1}{2}\sum_{j=1}^{N}(y_j - Hu_j)^T R^{-1}(y_j - Hu_j)$$

$$+ \frac{1}{2}\sum_{j=1}^{N}(u_j - F_j(u_{j-1}))^T Q^{-1}(u_j - F_j(u_{j-1})).$$

This cost function is weak-constraint, meaning that u does not have to be exactly a model-trajectory.

Least square problem and minimization

The nonlinear least squares problem

$$\min_{x} \Phi(x) = \frac{1}{2}\|f(x)\|_2^2, \tag{3}$$

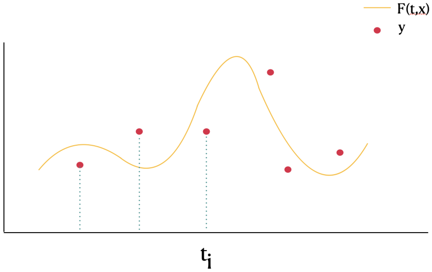
where x is an n-dimensional real vector.

Least square problem and minimization

The nonlinear least squares problem

$$\min_x \Phi(x) = \frac{1}{2}\|f(x)\|_2^2, \tag{3}$$

where x is an n-dimensional real vector.



If we define the ith component of $f(x)$ to be $f_i(x) = F(t_i, x) - y_i$, the solution of (3) gives the best model fit to the data in the sense of minimum sum of square errors.

Least square problem and minimization (continued)

Assume that:
1. $f : \mathbb{R}^n \to \mathbb{R}^m$ is a nonlinear twice continuously Frechet differentiable function. If we denote the Jacobian of the function f by $J(x) := f'(x)$. Then the gradient of $\Phi(x)$ is given by

$$\nabla\Phi(x) = J(x)^T f(x),$$

2. There exist $x^* \in \mathbb{R}^n$ such that $J(x^*)^T f(x^*) = 0$

3. The Jacobian matrix $J(x^*)$ at x^* has full rank n.

Then, finding the stationary point of ϕ is equivalent to solving the gradient equation

$$\nabla\Phi(x) = J(x)^T f(x) = 0.$$

Least square problem and minimization (continued)

Assume that:
1. $f : \mathbb{R}^n \to \mathbb{R}^m$ is a nonlinear twice continuously Frechet differentiable function. If we denote the Jacobian of the function f by $J(x) := f'(x)$. Then the gradient of $\Phi(x)$ is given by

$$\nabla\Phi(x) = J(x)^T f(x),$$

2. There exist $x^* \in \mathbb{R}^n$ such that $J(x^*)^T f(x^*) = 0$

3. The Jacobian matrix $J(x^*)$ at x^* has full rank n.

Then, finding the stationary point of ϕ is equivalent to solving the gradient equation

$$\nabla\Phi(x) = J(x)^T f(x) = 0.$$

Gauss-Newton Algorithm:

Step0. Choose and initial $x_0 \in \mathbb{R}^n$.
Step1. Repeat until convergence:
Step1.1. Solve $J(x^{(k)})^T J(x^{(k)}) s^{(k)} = -J(x^{(k)})^T f(x^{(k)})$.
Step1.2. Set $x^{(k+1)} = x^{(k)} + s^{(k)}$, $k := k+1$ and go to step 1.

Minimization Problem

We are looking for the approximation solution of (4), $\mathbf{u} = \{u_0, \dots, u_N\}$ such that

$$\|u_{j+1} - F_j(u_j)\| \to 0, \quad \text{and} \quad \|Hu_j - y_j\| \to 0.$$

Therefore, the aim is to solve the following minimization problem:

$$\min_{\mathbf{u}\in D(F)} \frac{1}{2}\{\|G(\mathbf{u})\|_2^2 + \alpha\|H\mathbf{u} - \mathbf{y}\|_2^2\}, \tag{4}$$

where $G_j(\mathbf{u}) = u_{j+1} - F_j(u_j)$ and α is a parameter which plays a significant role in the convergence of the Gauss-Newton iteration and has to be chosen carefully in order to achieve either convergence or boundedness of the solution.

Minimization Problem (continued)

The method proceeds by the following Gauss-Newton iterations, starting from some initial guess $u^{(0)}$:

$$\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} - \left(G_k'^T G_k' + \alpha H^T H \right)^{-1} \left(G_k'^T G_k + \alpha H^T (H\mathbf{u}^{(k)} - \mathbf{y}) \right),$$

where k denotes the index of the Gauss-Newton's iteration.

Minimization Problem (continued)

The method proceeds by the following Gauss-Newton iterations, starting from some initial guess $u^{(0)}$:

$$\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} - \left(G_k'^T G_k' + \alpha H^T H \right)^{-1} \left(G_k'^T G_k + \alpha H^T (H\mathbf{u}^{(k)} - \mathbf{y}) \right),$$

where k denotes the index of the Gauss-Newton's iteration.

$$G(\mathbf{u}) = (G_0(\mathbf{u}), G_1(\mathbf{u}), \cdots, G_{N-1}(\mathbf{u}))^T, \quad G_j(\mathbf{u}) = u_{j+1} - F_j(u_j),$$

for $j = 0, \cdots, N-1$ and $u_j \in \mathbb{R}^n$.
Therefore G', the Jacobian of G, has an $n(N-1) \times nN$ block structure:

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

Noise-free observations

Here, we want to look at the convergence of the Gauss-Newton iterations to the true solution.

Assumption

Consider that we have a sequence of noise free observations, i.e., $\eta \equiv 0$. Therefore the data y_{j+1} used in Gauss-Newton algorithm is found from observing a true solution x_j given by

$$x_{j+1} = F(x_j), \quad y_{j+1} = Hx_{j+1}, \quad j = 0, \dots, N-1,$$

in which $H \neq I$.

Noise-free observations

Here, we want to look at the convergence of the Gauss-Newton iterations to the true solution.

Assumption

Consider that we have a sequence of noise free observations, i.e., $\eta \equiv 0$. *Therefore the data* y_{j+1} *used in Gauss-Newton algorithm is found from observing a true solution* x_j *given by*

$$x_{j+1} = F(x_j), \quad y_{j+1} = Hx_{j+1}, \quad j = 0, \dots, N-1,$$

in which $H \neq I$.

Theorem

Let assume that the observations are noise-free and $\|e_0\|_2 := \|\mathbf{u}^{(0)} - \mathbf{x}\|_2 < 1$, *where* $\mathbf{u}^{(0)}$ *is the initial guess for the iteration and* $\mathbf{x}$ *is the true solution. Furthermore, assume that there exists* $L2 > 0$ *and* $0 < \lambda < 2/L2$ *such that*

$$\|G'(u) - G'(v)\|_2 \leq L_2 \|u - v\|_2, \quad \forall u, v \in \mathbb{R}^n, \tag{5}$$

$$\|(G'(x)^T G'(x) + \alpha H^T H)^{-1} G'^T(x)\|_2 \leq \lambda, \ \forall x \in \mathbb{R}^n. \tag{6}$$

Then the sequence $\mathbf{u}^{(k)}$ *converges to the true solution* $\mathbf{x}$.

Proof

Let $e_k := \mathbf{u}^{(k)} - \mathbf{x}$. By using Guass-Newton iteration, we have

$$\begin{aligned} e_{k+1} &= \mathbf{u}^{(k+1)} - \mathbf{x} \\ &= \mathbf{u}^{(k)} - (G'^T G' + \alpha H^T H)^{-1} \left(G'^T G - \alpha H^T (\mathbf{y} - H\mathbf{u}^{(k)}) \right) - \mathbf{x}. \end{aligned}$$

Using the following steps

- $H\mathbf{x} = \mathbf{y}$ and $G(\mathbf{x}) = 0$
- Mean value theorem and Lipschitz condition on G'
- $\|(G'(x)^T G'(x) + \alpha H^T H)^{-1} G'^T(x)\|_2 \leq \lambda$

Proof

Let $e_k := \mathbf{u}^{(k)} - \mathbf{x}$. By using Guass-Newton iteration, we have

$$\begin{aligned} e_{k+1} &= \mathbf{u}^{(k+1)} - \mathbf{x} \\ &= \mathbf{u}^{(k)} - (G'^T G' + \alpha H^T H)^{-1} \left(G'^T G - \alpha H^T (\mathbf{y} - H\mathbf{u}^{(k)}) \right) - \mathbf{x}. \end{aligned}$$

Using the following steps

- $H\mathbf{x} = \mathbf{y}$ and $G(\mathbf{x}) = 0$
- Mean value theorem and Lipschitz condition on G'
- $\|(G'(x)^T G'(x) + \alpha H^T H)^{-1} G'^T(x)\|_2 \leq \lambda$

$$\|e_k\|_2 \leq (\frac{L_2 \lambda}{2})^{2^k - 1} \|e_0\|_2^{2^k}.$$

Noisy Observations

Since the noise-free observations is not a real assumption. We would like to consider the problem with noisy observations. In this section, we assume that we have a sparse and noisy observations, i.e, $H \neq I$ and $y_{j+1} = Hx_{j+1} + \eta$, $\eta \sim \mathcal{N}(0, \Gamma) = (0, \gamma^2 I)$.

Noisy Observations

Since the noise-free observations is not a real assumption. We would like to consider the problem with noisy observations. In this section, we assume that we have a sparse and noisy observations, i.e, $H \neq I$ and $y_{j+1} = Hx_{j+1} + \eta$, $\eta \sim \mathcal{N}(0, \Gamma) = (0, \gamma^2 I)$.

Theorem

Let assume that $H \neq I$, $\eta \neq 0$ and $\|e_0\|_2 := \|\mathbf{u}^{(0)} - \mathbf{x}\|_2 < 1$, where $\mathbf{u}^{(0)}$ is the initial guess for the iteration and $\mathbf{x}$ is the true solution. Moreover, assume that there exists constants $L2 > 0$, $0 < \alpha < 1$, $\lambda \leq \frac{1}{2L_2}$ and $\beta \leq \frac{1}{2\|H^T\eta\|}$ such that,

$$
\begin{aligned}
&\|G'(u) - G'(v)\|_2 \leq L_2 \|u - v\|_2, \quad \forall u, v \in \mathbb{R}^n, \\
&\|(G'(x)^T G'(x) + \alpha H^T H)^{-1}\|_2 \leq \beta, \quad \forall x \in \mathbb{R}^n, \\
&\|(G'(x)^T G'(x) + \alpha H^T H)^{-1} G'^T(x)\|_2 \leq \lambda, \quad \forall x \in \mathbb{R}^n.
\end{aligned}
$$

Then $\|\mathbf{u}^{(k)} - \mathbf{x}\|_2 \leq c_k$, where $c_k := \sum_{i=0}^{k-1} (2\alpha\lambda L_2)^{2^i}$.

Numerical Experiment

The Lorenz 63 model is a chaotic dynamical system commonly used as a test problem for data assimilation algorithms. The model consists of three nonlinear, ordinary differential equations given as

$$\frac{dx_1}{dt} = \sigma(x_2 - x_1), \ \frac{dx_2}{dt} = x_1(\rho - x_3) - x_2, \ \frac{dx_3}{dt} = x_1 x_2 - b x_3,$$

where $\sigma = 10$, $b = \frac{8}{3}$ and $\rho = 28$.
In the following numerical experiments, a forward Euler method is used to discretize the model equations using a time step $\Delta t = 0.005$.

Numerical Experiments (continued)

Noise-Free Observations:

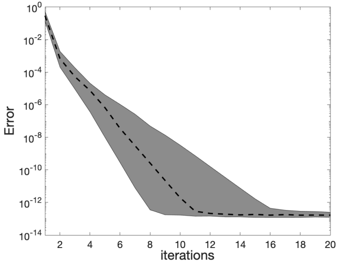
For the experiments in this section we generate a set of observations computing a trajectory of Lorenz 63 on $t \in [0, 20]$, using $\Delta t = 0.005$ and $\gamma = 0$. The observations of x^1-variable only are drawn every $\Delta t = 0.05$ which means the dynamical model F_j corresponds to 10 forward Euler steps.

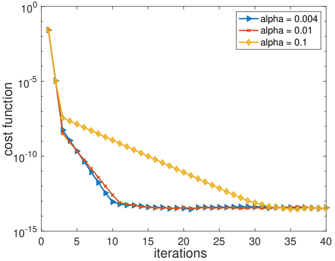
Numerical Experiments (continued)

Noise-Free Observations:

For the experiments in this section we generate a set of observations computing a trajectory of Lorenz 63 on $t \in [0, 20]$, using $\Delta t = 0.005$ and $\gamma = 0$. The observations of x^1-variable only are drawn every $\Delta t = 0.05$ which means the dynamical model F_j corresponds to 10 forward Euler steps.

Error : $\|\mathbf{x} - \mathbf{u}\|_2$, cost function : $\frac{1}{2}\|G(\mathbf{u})\|_2^2 + \frac{\alpha}{2}\|H\mathbf{u} - \mathbf{y}\|_2^2$.

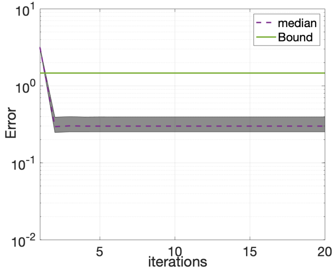




Numerical Experiments (continued)

Noisy Observations:

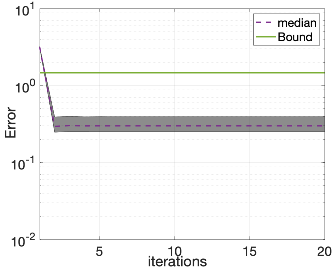
Here, we generate a set of noisy observations with $\gamma^2 = 0.1$. The observation matrix is, H=[1,0,0], which means the first variable x^1 is observed. In the following figures, we display error with respect to the truth on the left, the green line in this figure is the bound which we obtained in theorem 2. On the right, we display error for different values of η. We can see that error goes to zero as $\gamma \to 0$.

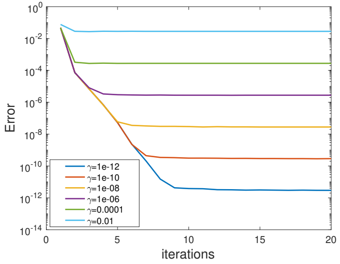


Numerical Experiments (continued)

Noisy Observations:

Here, we generate a set of noisy observations with $\gamma^2 = 0.1$. The observation matrix is, H=[1,0,0], which means the first variable x^1 is observed. In the following figures, we display error with respect to the truth on the left, the green line in this figure is the bound which we obtained in theorem 2. On the right, we display error for different values of η. We can see that error goes to zero as $\gamma \to 0$.





Comparison with WC4DVar

The cost function for weak constraint 4-D Var is:

$$C(u_0; y_j) = \frac{1}{2}\sum_{j=1}^{N}(y_j - Hu_j)^T R^{-1}(y_j - Hu_j) + \frac{1}{2}\sum_{j=1}^{N}(u_j - F_j(u_{j-1}))^T Q^{-1}(u_j - F_j(u_{j-1})). \tag{7}$$

(8)

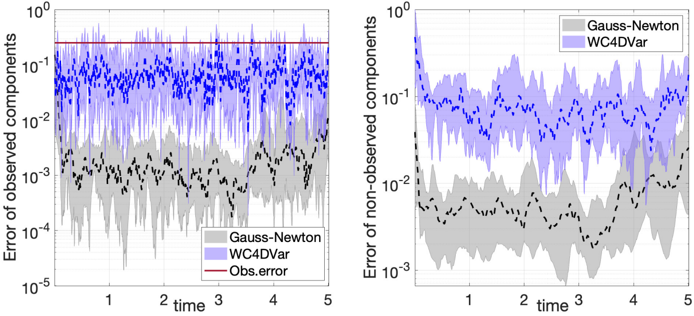


Figure 1: Application to L63. Error as a function of time: median (dashed line) +/- one standard deviation over 20 simulations with length of assimilation window, 5 (25 days). On the left: error with respect to the truth of observed variables. On the right: error with respect to the truth of non-observed variables. The Gauss-Newton method in grey, WC4DVar method in blue, and the observational error is in red.

Lorenz 96 Model

The Lorenz 96 model is a dynamical system with dimension $d \in \mathbb{N}$, formulated by Edward Lorenz in 1996, which is completely determined by the equation for the l-th variable as follows

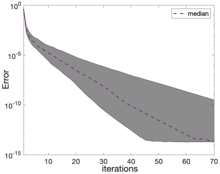
$$\dot{x}_l = -x_{l-2}x_{l-1} + x_{l-1}x_{l+1} - x_l + \mathcal{F}, \quad l = 1, \dots, d, \tag{9}$$

where the dimension d and forcing $\mathcal{F}$ are parameters. Cyclic boundary conditions are imposed. We implement the L96 model with the parameter choices $d = 40$ and $\mathcal{F} = 8$.

Lorenz 96 Model (continued)

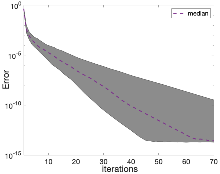
Noise-free Observations:

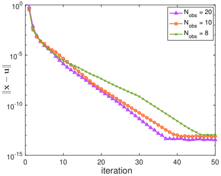
As we have noise free observations, they are obtained by a reference (true) trajectory with partially observation operator, it means $H \neq I$.



Noise-free Observations:

As we have noise free observations, they are obtained by a reference (true) trajectory with partially observation operator, it means $H \neq I$.





Lorenz 96 Model (continued)

Noise-free Observations:

As we have noise free observations, they are obtained by a reference (true) trajectory with partially observation operator, it means $H \neq I$.

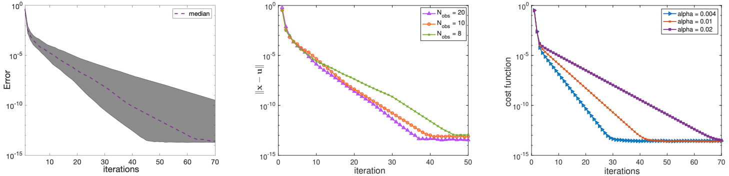


Figure 2: Application to L96 with partially and noise free observations. On the left: Error of the Gauss-Newton iteration with respect to the truth as a function of iterations: median (dashed line), +/- one standard deviation (shadowed area) over 100 simulations. On the middle: Error of the Gauss-Newton method as a function of iterations for different size of observations with $\alpha = 0.001$, $\Delta t = 0.005$. On the right:Cost function with respect to iterations for different values of α with 20 observations and $\Delta t = 0.005$.

Comparison with WC4DVar

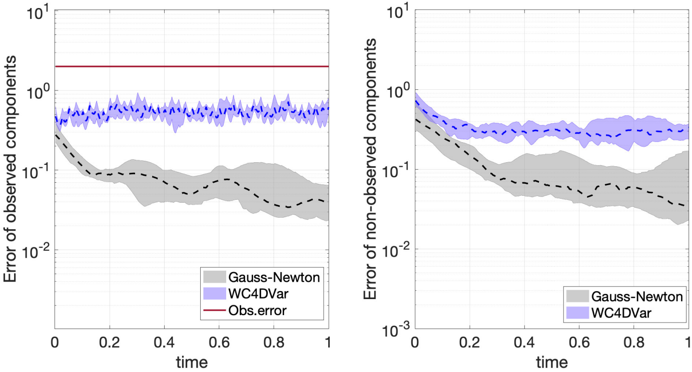


Figure 3: Application to L96. Error as a function of time: median (dashed line) +/- one standard deviation over 20 simulations with length of assimilation window, 1 (5 days). On the left: error with respect to the truth of observed variables. On the right: error with respect to the truth of non-observed variables. The Gauss-Newton method in grey, WC4DVar method in blue, and the observational error is in red.

Summary and outlook

Summary

- The Gauss-Newton method for solving the nonlinear least square problem is described.
- Deriving conditions for the convergence of this approximate method for noise-free observations.
- Obtaining the boundedness of the approximate solution for noisy observations under some conditions.

Outlook

- How much we can actually observe?

References

S. Gratton, A. S. Lawless, N. K. Nichols,
Approximate Gauss-Newton methods for nonlinear least squares problems,
SIAM Journal on Optimization (2007).

A. Lawless, S. Gratton, N. Nichols,
An investigation of incremental 4d-var using non-tangent linear models,
Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography 131 (2005) 459–476.

C. Cartis, M. H. Kaouri, A. S. Lawless, N. K. Nichols,
Convergent least-squares optimisation methods for variational data assimilation,
arXiv preprint arXiv:2107.12361 (2021).

H. Du, L. A. Smith,
Pseudo-orbit data assimilation. part i: The perfect model scenario,
Journal of the Atmospheric Sciences (2014).

B. M. de Leeuw, S. Dubinkina,
Shadowing-based data assimilation method for partially observed models,
SIAM Journal on Applied Dynamical Systems (2022).

B. De Leeuw, S. Dubinkina, J. Frank, A. Steyer, X. Tu, E. V. Vleck,
Projected shadowing-based data assimilation,
SIAM Journal on Applied Dynamical Systems 17 (2018) 244 6–2477.

Thank you!