A new data assimilation method for high-dimensional models

Guangjie Wang¹,², Xiaoqun Cao¹,²*, Xun Cai¹,², Jingzhe Sun¹,², Xiaoyong Li¹,², Heng Wang¹,²

¹ College of Meteorology and Oceanology, National University of Defense Technology, Changsha 410073, China, ² College of Computer, National University of Defense Technology, Changsha 410073, China

* caoxiaoqun@nudt.edu.cn

Abstract

In the variational data assimilation (VarDA), the typical way for gradient computation is using the adjoint method. However, the adjoint method has many problems, such as low accuracy, difficult implementation and considerable complexity, for high-dimensional models. To overcome these shortcomings, a new data assimilation method based on dual number automatic differentiation (AD) is proposed. The important advantages of the method lies in that the coding of the tangent-linear/adjoint model is no longer necessary and that the value of the cost function and its corresponding gradient vector can be obtained simultaneously through only one forward computation in dual number space. The numerical simulations for data assimilation are implemented for a typical nonlinear advection equation and a parabolic equation. The results demonstrate that the new method can reconstruct the initial conditions of the high-dimensional nonlinear dynamical system conveniently and accurately. Additionally, the estimated initial values can converge to the true values quickly, even if noise is present in the observations.

Introduction

The accuracy of numerical weather prediction (NWP) depends on exact initial and boundary conditions and perfect prediction models. In recent years, with the development of numerical models, the quality of the initial conditions has become the bottleneck of the accuracy of NWP [1]. To provide accurate and reasonable initial conditions for NWP, data assimilation is an effective method. Data assimilation derives from the objective analysis of initial values, and it has now become a novel technique that effectively uses considerable unconventional sources of information (such as satellite and radar data) [2]. By combining a mathematical model with irregularly distributed observations through a computer program, a coordinated relationship between the data and the model is constructed according to some algorithms, from which the most likely values on regular grid points can be obtained. Therefore, it has minimal errors in the analytical results and can provide initial conditions for prediction models [3]. In addition, data assimilation can not only effectively improve the accuracy of prediction models but also reduce the uncertainty of the initial conditions; therefore, data assimilation occupies an
important position in ocean and atmospheric sciences [4]. Currently, the methods of data assimilation are primarily divided into two categories: the first is sequential data assimilation methods [5], such as Kalman filter (KF), extended Kalman filter, ensemble Kalman filter (EnKF) and particle filter. The second is VarDA methods [6–10], including 3D and 4D VarDA methods (3D/4D-Var). The latter is the extension of the former in the time dimension, and its performance is considerably better. In addition, 4D-Var is the most advanced method for ocean and weather prediction, and it has achieved great success in improving the accuracy of numerical predictions. 4D-Var is essentially a large-scale optimization problem, whose constraint conditions are differential equations. When solving the problems, the value of the gradient vector for the cost function needs to be provided for optimization algorithms, which can be achieved by the reverse integration of the adjoint model [11, 12]. Since 4D-Var needs to solve the adjoint model, the amount of computation is particularly large. To reduce the computation cost, an incremental method, which requires the tangent-linear model, is often used. However, the tangent-linear and adjoint models are essentially a first-order differential mode of the nonlinear physical system equation. For complex models, the development of differential models in a manual way is a difficult task [11, 12], and the incremental method does not guarantee the convergence of the results. Moreover, the coding of the adjoint operator is a very complex and heavy task, and the parametrization of the physical process will lead to discontinuity of the cost function.

In fact, the key issue of VarDA to be solved is to calculate the partial derivative of the cost function on control variables (such as the initial value or certain parameters). Generally, there is an implicit nonlinear relationship between the cost function and the unknown parameter, both of which are typically connected by ordinary or partial differential equations. The numerical method that is commonly used to solve partial derivatives is the difference method, but the choice of the difference step is problematic. In particular, for high-dimensional models, the computation cost of the difference method is far greater than that of the adjoint method [12]. Moreover, since the initial values of multiple points are needed, it is easy to generate errors. At present, many data assimilation methods have advantages for low-dimensional nonlinear systems, but for high-dimensional nonlinear systems, their results are not as optimistic. For example, the computation cost is large, the calculation accuracy is not high and even the problems cannot be resolved in some situations. To calculate partial derivatives precisely, Lyness and Moler [13] proposed the complex variable differential method (CVD), whose main idea is to convert the calculation process of derivatives into the calculation process in the field of complex numbers. Cao et al. [14] applied the CVD method to data assimilation problems. A new data assimilation method based on dual number theory was proposed by Cao, Huang et al. [15], and its effectiveness was verified in low-dimensional nonlinear dynamical systems and data assimilation problems of non-differentiable prediction models. However, both aforementioned methods have not been applied to high-dimensional nonlinear dynamical systems and data assimilation problems. Due to the high-dimensional optimization problems whose constraint conditions are nonlinear dynamical prediction models, the gradient of the cost function is difficult to calculate, and thus, a data assimilation method based on dual number AD is proposed. In dual number space, both the function value and the gradient vector value can be accurately obtained by solving differential equations with a numerical solution and calculating the cost function. Therefore, compared to the VarDA method based on the adjoint equation, it has some obvious advantages.

The remainder of this paper is organized as follows. First, the dual number theories and algorithm rules are introduced. Second, a new data assimilation algorithm for high-dimensional numerical models is developed by combining accurate gradient information from dual number AD with the classical optimization algorithm. Third, numerical simulations for data
assimilation are implemented for a typical nonlinear advection equation and for a parabolic partial differential equation. Finally, a short conclusion is given.

**Data assimilation based on dual number automatic differentiation**

### Introduction of dual numbers

The dual number is defined as:

$$\tilde{x} = x + \varepsilon x'$$

(1)

where $x$ is the real part and $x'$ is the dual part. Both $x$ and $x'$ are real numbers. The dual label $x'$ does not indicate any specific value but rather the nature of $\varepsilon \neq 0$ and $\varepsilon^n = 0 (n > 1)$ [16, 17]. For convenient expression, Eq (1) can be written in the form such that $\tilde{x} = < x, \varepsilon x' >$. The dual number concept and its theory have been proposed for more than 100 years, but until the 1980s, dual number theory was applied to robot kinematics, spatial structures and other kinematic and dynamic problems [18, 19]. In recent years, dual number theory has opened up a new way for calculating the derivative information accurately and efficiently [20, 21].

### Dual number automatic differentiation algorithm

Using the properties of the dual number, the algebraic operations in the real number field can be extended to dual number space. Assume that two dual numbers are $\tilde{m} = m + \varepsilon m'$ and $\tilde{n} = n + \varepsilon n'$ and that their sum, difference and multiplication will be as follows [21]:

$$ (m + \varepsilon m') \pm (n + \varepsilon n') = (m \pm n) + (m' \pm n')\varepsilon $$

(2)

$$ (m + \varepsilon m') \times (n + \varepsilon n') = mn + mn'\varepsilon + nm'\varepsilon + m'n\varepsilon^2 = mn + (mn' + nm')\varepsilon $$

(3)

In Eq (3), $m'n\varepsilon^2$ is omitted, not because $m'$ and $n'$ are too small but because the square of $\varepsilon$ is 0 [17]. The dual number multiplication represented by Eq (3) is exactly equal, and there are no truncation errors. In the same way, all of the following for dual number algorithms are strictly correct, and there are no approximations or assumptions. Next, considering the dual number polynomial operation, Eq (4) provides a polynomial representation in real number space:

$$ P(x) = p_0 + p_1 x + p_2 x^2 + \ldots + p_n x^n $$

(4)

where $p_0, p_1, p_2, \ldots, p_n$ are multinomial coefficients. Replacing the real number $x$ with the dual number $\tilde{x} = < x, x' > = x + x'\varepsilon$ in Eq (4) and simultaneously using the rules of dual number...
addition and multiplication, it is easy to prove the formula as follows:

\[
P(\hat{x}) = p_0 + p_1(x + x'\epsilon) + p_2(x + x'\epsilon)^2 + \cdots + p_n(x + x'\epsilon)^n
\]

\[
= p_0 + p_1x + p_2x^2 + \cdots + p_nx^n + p_1x'\epsilon + 2p_2xx'\epsilon + \cdots + np_nx^{n-1}x'\epsilon
\]

\[
= P(x) + P^{(1)}(x)x'\epsilon
\]

\[
= <P(x), P^{(1)}(x)x'>
\]

where \(P^{(1)}\) represents the first-order derivative of the real polynomial \(P(x)\) on a real variable \(x\). \(x'\) is the seed number, which can take any value. Clearly, if \(x' = 1\), then the dual part is the exact value of the first-order derivative \(dP(x)/dx\). The other basic algebras and new algorithms of standard functions in dual number space are detailed in the literature [21].

For other basic functions, their dual number operation relation can be given by a similar derivation method, and the general two-variable basic function \(F\) is as follows:

\[
F(\hat{x}, \hat{y}) = F(<x, x'>, <y, y'>) = < F(x, y), F_x(x, y)x' + F_y(x, y)y' >
\]

In Eq (6), \(F_x\) is the partial derivative of the function \(F\) on the independent variable \(x\). \(F_y\) is the partial derivative of the function \(F\) on the independent variable \(y\). Eq (6) can be further extended to the case where the independent variable is the dual number vector \(\hat{x} = x + \epsilon x'\):

\[
F(\hat{x}) = F(x + \epsilon x') = < F(x), \nabla F(x) \cdot x' >
\]

where \(x\) and \(x' \in R^n\) are all multi-dimensional vectors. When the above basic arithmetic operations and functions act on mixed independent variables, such as the dual number \(<x, x'>\) and real number \(c\), the first step is to rewrite the real number \(c\) into the dual number \(<c, 0>\), and the second is to calculate it according to the above algorithms. The derivative of any function \(F(x)\) at point \(x_0\) can be obtained by calculating \(F(<x_0, 1>)\) at \(<x_0, 1>\) in dual number space directly, and the result is \(<F(x_0), F(x_0)>\). Likewise, as for the function \(F(x)\), whose independent variable is vector \(x \in R^n\), the directional derivative in direction \(x' \in R^n\) at point \(x_0\) can be derived by calculating function \(F(<x_0, x'>)\) in dual number space directly, and the result is \(<F(x_0), \nabla F(x_0) \cdot x'>\). In conclusion, the derivative value of independent variables can be obtained simultaneously when calculating the function value in dual number space, which achieves the function of automatic differentiation. At the same time, truncation errors will not be introduced because it avoids the difference operation, which results in machine accuracy when it is implemented on a computer.

A new data assimilation method

In the derivative computing method based on dual number automatic differentiation, for any cost function \(J(x)\) that takes real variable \(x\) as the independent variable, a dual variable is first constructed by taking \(x\) as the real part and taking \(\epsilon x'\) (the \(\epsilon\) is the dual label and take \(x' = 1\)) as the dual part. Second, the dual variable is substituted into the cost function. Thus, the real variable function \(J(x)\) can be transformed into the dual function \(\hat{J}(\hat{x})\), which takes the dual number \(\hat{x}\) as the independent variable. Finally, the dual function \(\hat{J}(\hat{x})\) is expanded as a Taylor series:

\[
\hat{J}(\hat{x}) = J(x) + \epsilon x' \frac{J'(x)}{1!} + \epsilon^2 x'^2 \frac{J''(x)}{2!} + \cdots + (\epsilon x')^n \frac{J^n(x)}{n!} + \cdots
\]

\[
=< J(x), \nabla J >
\]
Due to $\varepsilon^n = 0 (n > 1)$, high-order remainders that are greater than second order on the right-hand side of Eq (8) can be omitted, and the real part and the dual part on both sides of the equation are strictly equal. It can be shown as follows:

$$J(x) = \text{Re}[\hat{J}(\hat{x})], \nabla_x J = \text{Du}[\hat{J}(\hat{x})]$$

In Eq (9), Re[] and Du[] represent the real part and the dual part of the dual function, respectively. From Eq (9), if we want to calculate the first-order derivative of the cost function on the control variable, we only need to replace the independent variables in the cost function of the real number with the corresponding dual variables. After calculating the value of the function, which is in dual number form, the first-order derivative of the cost function on this independent variable can be obtained by taking its dual part. If there are no other error sources, the accuracy of the first-order derivative for the cost function based on dual number automatic differentiation is not affected by the accuracy of the computer. Therefore, it can be considered that the derivative of the function in Eq (9) is the accurate numerical solution [20, 21].

For the 2-D control vector $\mathbf{x} = (\hat{x}_1, \hat{x}_2) = (< x_1, x'_1 >, < x_2, x'_2 >)$, the dual function $\hat{J}(\hat{x})$ is expanded as a Taylor series:

$$\hat{J}(\hat{x}) = J(x) + \varepsilon x'_1 \frac{\partial J}{\partial x'}(x) + \frac{\varepsilon^2}{2!} \left( x'_1^2 \frac{\partial^2 J}{\partial x'_1^2} + x'_1 x'_2 \frac{\partial^2 J}{\partial x'_1 \partial x'_2} + x'_2^2 \frac{\partial^2 J}{\partial x'_2^2} \right) + \cdots$$

High-order remainders that are greater than second order on the right-hand side of Eq (10) can be omitted by using the property of $\varepsilon^n = 0 (n > 1)$. At the same time, the value of the seed number is $x'_1 = x'_2 = 1$; thus, the real part and the dual part on both sides of the equation are strictly equal. It can be shown as follows:

$$J(x) = \text{Re}[\hat{J}(\hat{x})], \nabla_x J = \text{Re}[\hat{J}(\hat{x})]$$

For the multi-dimensional control vector, an expression that is similar to Eq (11) can be obtained.

Dual number differentiation can solve the derivative problem of strongly nonlinear and implicit functions that cannot be solved by the general analytical method, which only requires one forward computation in dual number space. Moreover, for high-dimensional situations, the computation cost and errors are considerably less than in the conventional difference method. After obtaining the cost function gradient of the unknown initial state vector, a suitable descent algorithm is selected to optimize the parameters of each unknown parameter according to the following Eq (12):

$$x^{i+1}_j = x^i_j - \nabla_x J|_{x^i} \cdot \rho^{i+1}_j, j = 1, 2, \cdots, n$$

$\rho^{i+1}_j (j = 1, 2, \cdots, n)$ represents the $i$th iteration step, whose size is determined by the descent algorithm. When the value of $i$ is 0, $x^1$ represents the initial guess of the unknown initial state vector. The initial state vector of the high-dimensional nonlinear dynamic prediction model can finally be determined by using Eq (12). The specific process of data assimilation based on dual number automatic differentiation is shown in Fig 1, and the specific processing steps are as follows:
Step 1: First, a difference scheme (such as the upwind scheme) is used to achieve the numerical procedure for solving differential equations. Second, based on dual number rules in dual number space, the relevant procedures (including the prediction model, the observation operator, the cost function and so forth) are modified, which can be suitable for the numerical calculation of dual number space.

Step 2: Provide the initial guess $\hat{u}^0$ of the initial state vector.

Step 3: Using the provided guess value, the cost function and the gradient vector are calculated by using the automatic differentiation of the dual number theory, and the specific steps are as follows:

(3a) For each component of the initial state vector, the seed is 1, and the real number is transformed into the dual number. Moreover, the initial state vector is constructed in the
form of $\mathbf{u}' = (\langle u'_1, e_1 \rangle, \langle u'_2, e_2 \rangle, \ldots, \langle u'_n, e_n \rangle)$, where $e_j$ represents the multi-dimensional unit vector whose $j$th component is 1.

(3b) The dual number vector $\mathbf{u}'$ is used as the initial condition of the high-dimensional nonlinear dynamic prediction model, and the procedure implemented in the first step is used to perform a forward integration in dual number space. In this way, the evolution track of system state $\mathbf{x}(t)$ in dual number form will be obtained, where the real part and the dual part of $\mathbf{x}(t)$ represent the value of the state vector $\mathbf{x}(t)$ and the derivative of $\mathbf{x}(t)$ on the initial state vector $\mathbf{u}$ at different time steps, respectively.

(3c) The value of the cost function in dual number space will be calculated by using $\mathbf{x}(t)$ and observation data $\mathbf{y}^{\text{obs}}(t)$. Then, the value of the cost function will be calculated according to Eq (11), whose gradient of the initial state vector component is calculated in the same way.

Step 4: Using the gradient information of the cost function and combined with the typical gradient method, the step size $\rho_{j+1}$ is obtained. According to Eq (12), the initial state vector of each component is iterated, and the new estimated value $\mathbf{u}_{j+1}$ is derived. If the termination conditions (such as reaching the convergence accuracy or maximum iteration number that is set in advance) of the procedure are met, the procedure will stop, and the estimated value of the initial state vector will be derived at the same time. If the termination conditions are not met, using the value of the new initial state vector $\mathbf{u}_{j+1}$, a new iterative loop will start from the third step. To reduce the computation cost, this study uses the usual convergence criterion: the difference of the cost function gradient norm between two consecutive iterations is less than the pre-specified threshold $\|\nabla_{\mathbf{u}} J\| - \|\nabla_{\mathbf{u}_{j+1}} J\| < 10^{-6}$. In such situations, the calculation error is less than the machine error.

Results

Data assimilation for the nonlinear advection equation

To illustrate the usefulness of the new data assimilation method based on dual number automatic differentiation for high-dimensional models, numerical experiments are implemented for the nonlinear advection equation as an example. The nonlinear advection equation [22, 23] is shown as follows:

$$
\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \frac{\partial \mathbf{u}}{\partial x} = 0 \quad (0 \leq x \leq L, \quad 0 \leq t \leq T) \quad (13)
$$

$$
\mathbf{u}_{1x=0} = -3 \sin (t \pi / 6)
$$

In Eq (13), $\mathbf{u}$ is a physical quantity, $x$ is the horizontal space, and $t$ is the time. $L$ and $T$ are the space and time range of the nonlinear advection equation, respectively. The second equation is the boundary condition. The specific physical significance of this formula can be found in the literature [22, 23]. The standard explicit difference scheme, namely, the upwind scheme, is used to discretize the equations:

$$
\mathbf{u}_{j+1}^k = \mathbf{u}_j^k - \mathbf{u} \Delta t (\mathbf{u}_j^k - \mathbf{u}_{j-1}^k) / \Delta x
$$

$$
\mathbf{u}_1^k = -3 \sin (k \Delta t \pi / 6) \quad (14)
$$

where the space index of $x$ is $j = 1, 2, \ldots, J$, $k = 0, 1, \ldots, M$ is the time index. The variable $\mathbf{u}_j^k$ is used to approximate the physical quantity $u(k \Delta t, j \Delta x)$ of discrete points. Moreover, there are $\Delta t = (T/M)$ and $\Delta x = (L/J)$. The state vector is defined as $\mathbf{u} = (u_{1x}, u_{2x}, \ldots, u_{(N-1)x}, u_{Nx})^T$. The
The goal of data assimilation is to estimate the unknown initial state $u_0 = (u_0, \Delta x, u_0, 2\Delta x, \ldots, u_0, (N-1)\Delta x, u_0, N\Delta x)^T$ from a certain number of observation data $u_{\text{obs}} = (u_{\text{obs}}, \Delta x, u_{\text{obs}}, 2\Delta x, \ldots, u_{\text{obs}}, (N-1)\Delta x, u_{\text{obs}}, N\Delta x)^T$ for a discrete distribution. In this experiment, the value of $N$ is $N = 12$.

In the numerical experiment, the upwind scheme is first used to solve the nonlinear advection equation. Second, the numerical procedure for calculating the cost function (shown in Fig 2) is achieved. Third, its procedure is modified (shown in Fig 3), which is based on dual number rules in dual number space. The acquisition process of the observation data is as follows: the initial state vector is $u_0 = (u_0, \Delta x, u_0, 2\Delta x, \ldots, u_0, (N-1)\Delta x, u_0, N\Delta x)^T$. The true initial value is $u_0^{\text{obs}} = 3 \sin (j\Delta x/\pi) (j = 1, 2, \ldots, N)$. The time interval is $[0, 0.5]$, and the time step is $\Delta t = 0.001$. The space grid distance is $\Delta x = 1.0$. The state value of $u$ in discrete time series is derived by numerical integration, and then Gaussian observation noise $N(0, \sigma_0)$ are superimposed to form the observation data. The mean value and standard deviation of the noise are 0 and $\sigma_0$, respectively. If the background information is not considered, the cost function in discrete form is defined as follows:

$$
\hat{J}(\hat{u}) = \frac{1}{2} \sum_{i=1}^{N} \| u_i - u_i^{\text{obs}} \|^2 
$$

In Eq (15), $N$ represents the number of observations, and $\| \cdot \|^2$ represents the Euclidean norm. The initial guess of the unknown initial state vector is $u_0 = (1.0, 1.0, \ldots, 1.0)$. In the process of the iterative estimation for the unknown initial state vector, to update the unknown vector $u$ every time, the dual number vector $\hat{u}_i = (u_{i,1}, e_1, <u_{i,1}, e_1>, <u_{i,1}, e_2>, \ldots, <u_{i,1}, e_{12}>)$ has to first be constructed, where $e_j (j = 1, 2, \ldots, 12)$ represents the 12-D unit vector whose $j$th component is 1, and it is input into the numerical model of the nonlinear advection equation to make a forward ($0 \rightarrow T$) integration. Second, the cost function $\hat{J}(\hat{u})$ is calculated in dual number space. Third, the real part and the dual part are taken as the values of the cost function.
\[ A_{1}(:, :) = 0.0 \]
\[
\text{do i=1, mx}
\hspace{1em} A_{1}(i, i) = 1.0 
\text{enddo}
\]
\[
\text{do i=1, mx}
\hspace{1em} u(i, 0) = \text{DUAL}_\text{NUM}(\text{xyz}(i), A_{1}(i, :)) \text{! u0 is the initial guess}
\text{end do}
\]
\text{! Boundary conditions}
\[
\text{do j=0, mt}
\hspace{1em} u(0, j) = \text{DUAL}_\text{NUM}(-3*\sin(\pi*\text{dt}*j/6), 0)
\text{end do}
\]
\[
\text{do j=0, mt-1}
\hspace{1em} \text{do i=1, mx}
\hspace{2em} u(i, j+1) = u(i, j) - (u(i, j) - u(i-1, j)) * \text{dt}/\text{dx}*1
\hspace{2em} \text{end do}
\hspace{1em} \text{end do}
\text{end do}
\]

Fig 3. The model procedure of the nonlinear advection equation in dual number space.

https://doi.org/10.1371/journal.pone.0191714.g003

\( J \) and the gradient vector \( \nabla u_{j} \), respectively. Finally, the conjugate gradient algorithm is used to update the value of the initial state vector [14]. To verify the effectiveness of the new method, the state variables from the 4th time step to the 15th time step are selected as the observation data in the numerical experiment.

Figs 4 and 5 present the experimental results of data assimilation for the nonlinear advection equation without observation noise. Fig 4a and 4b present the change of the value for

![Fig 4](https://doi.org/10.1371/journal.pone.0191714.g004)

**Fig 4. The change of initial state \( u_{0} \) with iteration number.** (a) Initial state \( u_{0, \Delta x} \); (b) Initial state \( u_{0, 4\Delta x} \); (c) Initial state \( u_{0, 7\Delta x} \); (d) Initial state \( u_{0, 10\Delta x} \).
unknown initial state quantities $u_{0, \Delta x}$, $u_{0.4\Delta x}$, $u_{0.7\Delta x}$ and $u_{0.10\Delta x}$ in the iterative estimation process, respectively. It can be observed from the figures that the iterative value converges to the true value, and the other eight points have similar results. To avoid redundancy, the results of the other eight points are not presented. Fig 5a shows the change of the cost function $f(u)$ with iteration number, and Fig 5b shows the change of the gradient norm $\| \nabla u \|$ with iteration number. It can be observed from the figure that the cost function achieves convergence after the 10th iteration. The value of the unknown initial state vector $u$ is rounded to the 6th decimal place. The results are presented in Table 1. As shown in Table 1, the estimated value is very close to the true value. Theoretically, when the gradient norm of the cost function on control variables is 0, the unknown initial state vector can take the optimal estimate. By analysing the experimental results, the following conclusions are obtained: in the case without background information and only using minimal observation information, the data assimilation method based on dual number automatic differentiation can accurately estimate the initial state of the nonlinear advection equation, and its estimation accuracy reaches $O(10^{-6})$.

Table 1 shows the estimated value of the initial condition, the final value of the cost function and the iteration number for the nonlinear advection equation under different levels of observation noise. As shown in Table 1, when there are no observation errors, the estimation accuracy of the unknown initial states is the highest. Moreover, each initial state quantity can be accurate to the 6th decimal place, which verifies the effectiveness of the data assimilation method based on dual number automatic differentiation in estimating the unknown initial state for high-dimensional nonlinear physical systems. In addition, with increasing observation errors, the accuracy of data assimilation decreases, but the iterative estimation results still converge to the true value. When the standard deviation of the observation errors is $\sigma_o = 0.2$, the results of data assimilation are still close to the true value, and it can be accurate to the first

| $\sigma_o$ | $u_{0, \Delta x}$ | $u_{0.2\Delta x}$ | $u_{0.3\Delta x}$ | $u_{0.4\Delta x}$ | $u_{0.5\Delta x}$ | $u_{0.6\Delta x}$ | $f$ |
|-----------|------------------|------------------|------------------|------------------|------------------|------------------|-----|
| 0.0       | 1.50000000      | 2.5980763        | 3.00000000      | 2.5980762        | 1.5000002       | -0.0000002       | 5.3×10^{-12} |
| 0.1       | 1.5129420       | 2.5967695        | 3.0089850       | 2.5938779        | 1.4849668       | 0.0041435        | 0.6717583    |
| 0.2       | 1.4535858       | 2.6274088        | 2.9734110       | 2.5785377        | 1.4725619       | -0.014866        | 2.8869303    |

| $\sigma_o$ | $u_{0.7\Delta x}$ | $u_{0.8\Delta x}$ | $u_{0.9\Delta x}$ | $u_{0.10\Delta x}$ | $u_{0.11\Delta x}$ | $u_{0.12\Delta x}$ | $f$ |
|-----------|------------------|------------------|------------------|------------------|------------------|------------------|-----|
| 0.0       | -1.50000000      | -2.5980763       | -3.00000000      | -2.5980762       | -1.4999993       | -0.0000005       | 5.3×10^{-12} |
| 0.1       | -1.5000002       | -2.5980764       | -3.00000000      | -2.5980756       | -1.4999992       | 0.0000005        | 0.6717583    |
| 0.2       | -1.5067966       | -2.6231179       | -3.0218678       | -2.5734933       | -1.5258349       | -0.0271324       | 2.8869303    |

https://doi.org/10.1371/journal.pone.0191714.t001
decimal place, which demonstrates that the new data assimilation method based on dual number automatic differentiation is capable of removing noise in the observations when working with high-dimensional models.

It is noteworthy that Figs 4 and 5 show a spike in the iteration process. This is because the descent algorithm obtains the iteration step according to the gradient information of the cost function every iteration, the iteration step is not a constant. If the iteration step is not appropriate in the iteration process, there may be a spike. In the next iteration, the descent algorithm will adjust the iteration step according to the gradient information of the cost function until the difference of the cost function gradient norm between two consecutive iterations is less than the pre-specified threshold. The spike is due to the descent algorithm and not due to the dual number approach, and it will not influence the final convergence result.

Data assimilation for the parabolic equation
To further illustrate the usefulness of the new data assimilation method for high-dimensional models, this part takes the heat conduction equation as an example. This equation [24] is shown as follows:

\[
\frac{\partial u}{\partial t} = \sigma \frac{\partial^2 u}{\partial x^2}, \quad (0 < x < L, \ 0 < t < T)
\]

\[
u_{|x=0} = 0, \quad u_{|x=L} = 0
\]

In Eq (16), \(u\) is the state variable of temperature and \(\sigma\) is the heat conduction coefficient, which is a constant. The specific physical significance of this formula can be found in the literature [24]. The discrete heat conduction equation is as follows:

\[
u_{j}^{k+1} = \nu_{j}^{k} + \sigma \Delta t (\nu_{j-1}^{k} - 2\nu_{j}^{k} + \nu_{j+1}^{k})/\Delta x^2
\]

\[
u_{0}^{k} = 0, \ \nu_{N}^{k} = 0
\]

where the variable \(\nu_{j}^{k}\) is used to approximate the state variable of the temperature \(u(k\Delta t, j\Delta x)\) of discrete points. The state vector and the goal of data assimilation are the same as the first experiment, but the value of \(N\) is \(N = 6\) in this experiment.

In the numerical experiment, the procedure, the cost function, the acquisition process of the observations and the process of the iterative estimation for the unknown initial state vector are the same as the first experiment. But the true initial value is \(u_{0|Ax}^{0} = \sin (j\Delta \pi) \ (j = 1, 2, \ldots, N)\). The value of the heat conduction coefficient \(\sigma\) is 1. The time interval is \([0, 0.1]\), and the time step and the space grid are \(\Delta t = 0.001\) and \(\Delta x = 0.25\), respectively. Additionally, the initial guess of the initial state vector is \(u_{0} = (2.0, 2.0, \ldots, 2.0)\). To verify the effectiveness, the state variables from the second time step to the 7th time step are selected as the observation data in the numerical experiment.

Figs 6 and 7 present the experimental results of data assimilation for the heat conduction equation without observation noise. Fig 6a and 6b present the change of the value for unknown initial state quantities \(u_{0,2\Delta x}\) and \(u_{0,4\Delta x}\) in the iterative estimation process, respectively. It can be observed from the figures that the iterative value converges to the true value, and the other four points have similar results. Fig 7a and 7b show the change of the cost function \(J(u)\) and the gradient norm \(\|\nabla J\|\) with iteration number, respectively. It can be seen from the figure that the cost function achieves convergence after the 12th iteration. It is noteworthy that Figs 6 and 7 also show a spike caused by the same reason as the first experiment. The value of the unknown initial state vector \(u\) is rounded to the 7th decimal place. The results are shown in Table 2, from which it can be observed that the estimated value is very close to the
true value. By analysing the experimental results, the following conclusions are obtained: in the case of no background information and only using minimal observation information, the new method can accurately estimate the initial state of the heat conduction equation, and its estimation accuracy reaches $O(10^{-7})$, which basically can be used as true values. In addition, with increasing observation errors, the accuracy of data assimilation decreases, but the iterative estimation results still converge to the true value. When the standard deviation of observation errors is $\sigma_o = 0.2$, the data assimilation results are still close to the true value, which can be accurate to the first decimal place. Therefore, this experiment further verifies the effectiveness and the ability of removing noise in the observations of the new data assimilation method when working with high-dimensional models.

**Conclusion**

To overcome the shortcomings of the gradient computation for high-dimensional prediction models in variational data assimilation when using the adjoint method, a new data assimilation method based on dual number automatic differentiation is proposed. By using the dual number automatic differentiation, the process of gradient analysis is transformed to the computation of
the cost function in dual number space, which can obtain the value of the gradient vector simply, efficiently and accurately. The important advantages are that the coding of the adjoint model and the reverse integration are no longer necessary and that the values of the cost function and its corresponding gradient vector can be obtained simultaneously by only one forward computation in dual number space. Compared to the direct difference method, the calculation precision is higher, because the new method is not affected by the truncation errors and the cancellation errors. Compared to the adjoint method, the computation cost is less, because the new method not only does not need to develop the first order differential model, but also does not need to reverse integration of the adjoint equation. The results show that the new method can effectively estimate the initial conditions for high-dimensional numerical prediction models and is capable of removing noise. Therefore, the method proposed in this study is a new data assimilation method with strong adaptability.

However, because different optimization algorithms, such as Quasi-Newton method and conjugate gradient method (used in this paper), may have a slight influence on running time, iteration number and storage space, we will do more experiments to choose the most efficient optimization algorithms according to different cases in the future. Moreover, we expect that using more advanced optimization algorithms in this method can improve its efficiency. In addition, the data assimilation effects and advantages of this new method aiming at more high-dimensional and practical numerical prediction models will be carried out in the subsequent research work.

Supporting information

S1 File. The assimilation result for the nonlinear advection equation. (TXT)

S2 File. The result of the cost function for the nonlinear advection equation. (TXT)

S3 File. The result of the gradient norm for the nonlinear advection equation. (TXT)

S4 File. The assimilation result for the parabolic equation. (TXT)

S5 File. The result of the cost function for the parabolic equation. (TXT)

S6 File. The result of the gradient norm for the parabolic equation. (TXT)

S7 File. All the data used in the manuscript. This file includes all the result and figures used in the manuscript. (ZIP)

Author Contributions

Conceptualization: Guangjie Wang, Xiaoqun Cao, Jingzhe Sun.

Formal analysis: Guangjie Wang, Xun Cai, Jingzhe Sun.

Funding acquisition: Xiaoqun Cao, Xiaoyong Li.

Investigation: Guangjie Wang, Xun Cai, Heng Wang.
Resources: Guangjie Wang, Xiaoqun Cao, Heng Wang.
Software: Guangjie Wang, Xiaoqun Cao, Heng Wang.
Visualization: Guangjie Wang, Xun Cai, Heng Wang.
Writing – original draft: Jingzhe Sun, Heng Wang.
Writing – review & editing: Xiaoyong Li, Heng Wang.

References
1. Guan YH, Zhou GQ, Lu WS, Chen JP. Theory development and application of data assimilation methods. Meteorology and Disaster Reduction Research. 2007; 30(4):938–950.
2. Li H, Xu JP. Development of data assimilation and its application in ocean science. Marine Science Bulletin. 2011; 30(4):463–472.
3. Zhu GF. Understanding of the fundamental concept of atmospheric data assimilation. Meteorological Monthly. 2015; 41(4):456–463.
4. She ZQ, Tang YM, Gao YQ. The theoretical framework of the ensemble-based data assimilation method and its prospect in oceanic data assimilation. Haiyang Xuebao. 2016; 0(3):1–14.
5. Evensen G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. Journal of Geophysical Research Atmospheres. 1994; 99(C5):10143–10162. https://doi.org/10.1029/94JC00572
6. Thepaut JN, Courtier P. Four-dimensional data assimilation using the adjoint of a multiple primitive-equation mode. Quarterly Journal of the Royal Meteorological Society. 1991; 117(302):1225–1254. https://doi.org/10.1002/qj.49711750206
7. Rabier F, Mahlouf JF, Jarvinen H, Klinker E, Simmons A. The ECMWF operational implementation of four-dimensional variational assimilation. II: Experimental results with improved physics. Quarterly Journal of the Royal Meteorological Society. 2000; 126(564):1171–1190. https://doi.org/10.1002/qj.49712656416
8. Cao XQ, Huang SX, Du Hd. The new method of modeling horizontal error functions in variational assimilation with orthogonal wavelet. Acta Physica Sinica. 2008; 57(3):1984–1989.
9. Cao XQ, Song JQ, Zhang WM, Zhao YL, Liu BN. A new data assimilation method using complex-variable differentiation. Acta Physica Sinica. 2013; 62(17):170504–170504.
10. Zhang WM, Cao XQ, Song JQ. Design and implementation of four-dimensional variational data assimilation system constrained by the global spectral model. Acta Physica Sinica. 2012; 61(24):249202–249202.
11. Giering R, Kaminski T. Recipes for adjoint code construction. ACM Transactions on Mathematical Software. 1998; 24(4):437–474. https://doi.org/10.1145/293686.293695
12. Cheng Q, Zhang HB, Wang B. Algorithms of automatic differentiation. Mathematical Numerica Sinica. 2009; 31(1):15–36.
13. Lyness JN; Moler CB. Numerical differentiation of analytic functions. SIAM Journal on Numerical Analysis. 1967; 4(2):202–210. https://doi.org/10.1137/0704019
14. Cao XQ, Song JQ, Zhang WM, Zhao YL, Liu BN. A new data assimilation method using complex-variable differentiation. Acta Physica Sinica. 2013; 62(17):170504–170504.
15. Cao XQ, Huang QB, Liu BN, Zhu MB, Yu Y. A new data assimilation method based on dual-number theory. Acta Physica Sinica. 2015; 64(13):8–19.
16. He JH; Lee EWM. A constrained variational principle for heat conduction. Physics Letters A. 2009; 373(31):2614–2615. https://doi.org/10.1016/j.physleta.2009.05.039
17. He JH; Wu XH. Variational iteration method: New development and applications. Computers & Mathematics with Applications. 2007; 54(7-8):881–894. https://doi.org/10.1016/j.camwa.2006.12.083
18. Brodsky V; Shoham M. Dual numbers representation of rigid body dynamics. Mechanism & Machine Theory. 1999; 34(5):693–718. https://doi.org/10.1016/S0094-114X(98)00049-4
19. Wang JY, Liang HC, Sun ZW. Dual number-based relative coupled dynamics control. Journal of Astronautics. 2010; 31(7):1711–1717.
20. Spall RE; Yu W. Imbedded dual-number automatic differentiation for Computational Fluid Dynamics sensitivity analysis. Journal of Fluids Engineering. 2012; 135(1):279–284.
21. Yu W; Blair M. DNAD, a simple tool for automatic differentiation of Fortran codes using dual numbers. Computer Physics Communications. 2013; 184(5):1446–1452. https://doi.org/10.1016/j.cpc.2012.12.025

22. Yang XZ, Zhang H, Lu P, Ji ZZ. Stability analysis for difference schemes of nonlinear advection equations. Journal of China Electric Power University. 1999; 26(4):84–89.

23. Zhou BB. A kind of explicit finite difference schemes with better stability for advective equations. Scientia Atmospherica Sinica. 1995; 19(2):252–256.

24. Isakov V; Kindermann S. Identification of the diffusion coefficient in a one-dimensional parabolic equation. Inverse Problems. 2000; 16(3):665–68010. https://doi.org/10.1088/0266-5611/16/3/309