https://doi.org/10.5194/egusphere-2022-1268
Preprint. Discussion started: 21 November 2022

An Adjoint-Free Algorithm for CNOPs via Sampling

Bin Shi[1,3] and Guodong Sun[2,3]

[1]State Key Laboratory of Scientific and Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
[2]State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China
[3]University of Chinese Academy of Sciences, Beijing 100049, China

Correspondence: Bin Shi (Email: shibin@lsec.cc.ac.cn)

Abstract. In this paper, we propose a sampling algorithm based on statistical machine learning to obtain conditional nonlinear optimal perturbation (CNOP), which is different from traditional deterministic optimization methods. The new approach reduces the expensive gradient (first-order) information directly by the objective value (zeroth-order) information and does not use the adjoint technique that requires large amounts of storage and produces instability due to linearization. An intuitive analysis of the sampling algorithm is shown rigorously within the form of a concentration inequality for the approximate gradient. The numerical experiments of a theoretical model, Burgers equation with small viscosity, are implemented to obtain the CNOPs. The performance of standard spatial structures demonstrates that at the cost of losing accuracy, the sample-based method with fewer samples spends time relatively shorter than the adjoint-based method and directly from the definition. Finally, we show that the nonlinear time evolution of the CNOPs obtained by all the algorithms is nearly consistent with the quantity of norm square of perturbations, their difference and relative difference based on the definition method.

1 Introduction

The short-term behavior of a predictive model with imperfect initial data is a critical issue for weather and climate predictability. Understanding the model's sensitivity to errors in the initial data to assess subsequent errors in forecasts is important. Perhaps the simplest and most practical way is to estimate the likely uncertainty in the forecast by considering an ensemble of runs with initial data polluted by the most dangerous errors. Traditionally, based on the linear stability analysis of fluid dynamics, the well-known tool is the normal mode method (Rayleigh, 1879; Lin, 1955), and has been used to understand and analyze the observed cyclonic waves and long waves of middle and high latitudes (Eady, 1949). However, atmospheric and oceanic modes are often unstable; thus, the transient growth of perturbations can still occur in the absence of growing normal modes (Farrell and Ioannou, 1996a, b). Therefore, the normal mode theory is generally unavailable to evaluate the predictability problem generated by atmospheric and oceanic motion flows. To achieve this goal for a low-order two-layer quasi-geostrophic model in a periodic channel, Lorenz (1965) first proposes a nonnormal mode approach based on the view of linearization, which introduces the concepts of the tangent linear model, adjoint model, singular values, and singular vectors. Then, Farrell (1982) uses the linear approach to investigate the linear instability within finite time. In the last decade of the past century, such a

https://doi.org/10.5194/egusphere-2022-1268
Preprint. Discussion started: 21 November 2022

linear approach has been widely used to identify the most dangerous perturbations of atmospheric and oceanic flows, and has been extended to explore error growth and predictability, such as patterns of the general atmospheric circulations (Buizza and Palmer, 1995) and the coupled ocean-atmosphere model of the El Niño-Southern Oscillation (ENSO) (Xue et al., 1997a, b; Thompson, 1998; Samelson and Tziperman, 2001). The nonnormal approach has recently been extended to an oceanic study to investigate the predictability of the Atlantic meridional overturning circulation (Zanna et al., 2011) and the Kuroshio path variations (Fujii et al., 2008).

Both the approaches of normal and nonnormal modes are based on the assumption of linearization; thus, the perturbation must be sufficiently small such that a tangent linear model can approximately represent the evolution of the perturbation. The complex nonlinear atmospheric and oceanic processes have not still been well considered in the literature. To overcome this limitation, Mu (2000) proposed a nonlinear nonnormal mode approach, which introduces the concepts of nonlinear singular values and nonlinear singular vectors, and is then used to successfully capture the local fastest-growing perturbations for a 2D quasi-geostrophic model (Mu and Wang, 2001). However, several disadvantages still exist, such as practical inconvenience and unreasonable physics of the large norm for local fastest growing perturbations. Starting from the perspective of nonlinear programming, Mu et al. (2003) proposed an innovative approach, named conditional nonlinear optimal perturbation (CNOP), to explore the optimal perturbation that can fully consider the nonlinear effect without any linear approximation assumption. Generally, the CNOP approach captures initial perturbations with maximal nonlinear evolution given by a reasonable constraint in physics. Therefore, the CNOP approach as a powerful tool has been widely used to investigate the fastest-growing initial error in the prediction of an atmospheric and oceanic event and to reveal the related mechanisms, such as the stability of the thermohaline circulation (Mu et al., 2004; Zu et al., 2016), the predictability of ENSO (Duan et al., 2009; Duan and Hu, 2016) and the Kuroshio path variations (Wang and Mu, 2015), the parameter sensitivity of the land surface processes (Sun and Mu, 2017), and typhoon-targeted observations (Mu et al., 2009; Qin and Mu, 2012). The review paper (Wang et al., 2020) provides more details and refer (Kerswell, 2018) for more perspectives on the fields of fluid mechanics.

The primary goal of obtaining CNOPs is to implement nonlinear programming efficiently and effectively. Generally, the nonlinear optimization methods used in practice are the spectral projected gradient method (Birgin et al., 2000), sequential quadratic programming (Barclay et al., 1998) and the limited memory Broyden-Fletcher-Goldfarb-Shanno algorithm (Liu and Nocedal, 1989). In fluid mechanics, the standard gradient method is typically used to obtain the minimal finite amplitude disturbance that triggers the transition to turbulence in shear flows (Pringle and Kerswell, 2010), and the method of Lagrange multipliers is used to investigate perturbations that maximize the gain of disturbance energy in 2D isolated vortex and counter-rotating vortex-pair (Navrose et al., 2018). For the CNOPs, the objective function of the initial perturbations in a black-box model is obtained by the time evolution of a nonlinear differential equation. Thus, the essential difficulty here in computation is how to obtain the gradient information efficiently. It is not practical to obtain gradient directly from the definition-based methods, which require plenty of runs of the nonlinear model. Traditionally, the adjoint-based method is used in practice to obtain the gradient information by calculating the tangent linear model and the adjoint matrix (Kalnay, 2003). However, the adjoint-based method can only deal with smooth programming and is quite unstable for the atmospheric and oceanic models in practice. Also, a large amount of storage space to save the basic state during each iteration is a critical issue, that produces

https://doi.org/10.5194/egusphere-2022-1268
Preprint. Discussion started: 21 November 2022

drastically high-dimensional optimization problems, issues with data storage, and long computation times. The current adjoint-free computational approaches still have their faults. Considering two cases as examples, the ensemble-based methods still require colossal memory and repeated calculations (Wang and Tan, 2010; Chen et al., 2015), and the intelligent optimization methods do not guarantee finding an optimal solution (Zheng et al., 2017; Yuan et al., 2015).

To overcome the limitations of the adjoint-based method described above, we start from the perspective of stochastic optimization methods, which have been the algorithms that have powered recent developments in statistical machine learning (Bottou et al., 2018). In this paper, we use the derivative-free method proposed by Nemirovski and Yudin (1983, Section 9.3.2) that is based on the simple high-dimensional divergence theorem (i.e., Stokes' theorem). Following the popular and natural method, the derivative-free method is a stochastic approximation-type method, which imitates a stochastic oracle of the first order by the available stochastic order of the zeroth order. Then, based on the law of large numbers, we use the derivative-free method by sampling and propose the concentration estimate by the general Hoeffding inequality. The basic description of the CNOP settings and the proposed sample-based algorithm are given in Section 2 and Section 3, respectively. We then perform a preliminary numerical test for the simple Burgers equation with small viscosity in Section 4. A summary and discussion are included in Section 5.

2 The Basic CNOP Settings

In this section, we provide a brief description of the CNOP approach. Currently, the CNOP approach has been extended to investigate the influences of parameter and boundary condition errors on atmospheric and oceanic models by exploring the impact of initial errors (Mu and Wang, 2017). In this study, we only focus on the initial perturbations. The atmospheric and oceanic model in a region $x \in \Omega \subseteq \mathbb{R}^d$ with $\partial\Omega$ as its boundary is given as

$$\begin{cases} \dfrac{\partial U}{\partial t} = F(U, P) \\ U|_{t=0} = U_0 \\ U|_{\partial\Omega} = G, \end{cases} \tag{1}$$

where U is the reference state in the configuration space, P is the set of model parameters, F is a nonlinear operator, and U_0 and G are the initial reference state and boundary condition, respectively. Without loss of generality, we note $g^t(\cdot)$ to be the reference state $U(t;\cdot)$ in the configuration space evolving with time. Thus, given the initial condition U_0, we can obtain that the reference state at time T is $g^T(U_0) = U(T;U_0)$. If we consider the initial state $U_0 + u_0$ as the perturbation of U_0, then the reference state at time T is given by $g^T(U_0 + u_0) = U(T;U_0 + u_0)$.

With both the reference states $g^T(U_0)$ and $g^T(U_0 + u_0)$, the objective function of the initial perturbation u_0 based on the initial condition U_0 is

$$J(u_0;U_0) = \left\| g^T(U_0 + u_0) - g^T(U_0) \right\|^2, \tag{2}$$

https://doi.org/10.5194/egusphere-2022-1268
Preprint. Discussion started: 21 November 2022

and then the CNOP formulated as the constrained optimization problem is

$$\max_{\|u_0\| \le \delta} J(u_0; U_0). \tag{3}$$

Both the objective function (2) and the optimization problem (3) come directly from the model (1), which is a theoretical model and hence infinite-dimensional. Furthermore, when numerical computation is implemented, the optimization problem of infinite dimension is reduced to finite dimension. Without loss of generality, we shorten $J(u_0; U_0)$ as $J(u_0)$ afterward for convenience.

3 Sample-based algorithm

In this section, we briefly describe the sample-based algorithm and conclude with a formal theorem for the concentration estimate of the approximate gradient. The detailed proof is shown in Appendix A. The sample-based algorithm consists of two steps. First, for the population case, we transform the expectation of the gradient on the unit ball to that of objective values on the unit sphere by the high-dimensional Stokes' theorem. Then, we provide an intuitive analysis of the concentration in probability for the samples with the law of large numbers.

Let $\mathbb{B}^d$ be the unit ball in $\mathbb{R}^d$ and $v_0 \sim \text{Unif}(\mathbb{B}^d)$, a random variable following the uniform distribution in $\mathbb{B}^d$. Given a small real $\epsilon > 0$, we can define the expectation of J in the unit ball centering on u_0 as

$$\hat{J}(u_0) = \mathbb{E}_{v_0 \in \mathbb{B}^d} \left[J(u_0 + \epsilon v_0) \right]. \tag{4}$$

In other words, the objective function J is required to define in the ball $B(0; \delta + \epsilon) = \{u_0 \in \mathbb{R}^d : \|u_0\| \le \epsilon + \delta\}$[1]. Also, we find that $\hat{J}(u_0)$ is approximate to $J(u_0)$, thus, $\hat{J}(u_0) \approx J(u_0)$. If the gradient ∇J exists in the ball $B(0; \delta + \epsilon)$, the fact that the expectation of v_0 is zero tells us that the error estimate for the objective value is

$$\|J(u_0) - \hat{J}(u_0)\| = O(\epsilon^2).$$

With the representation of $\hat{J}(u_0)$ in (4), we can obtain the gradient $\nabla \hat{J}(u_0)$ directly from the function value J by the high-dimensional Stokes' theorem as

$$\nabla \hat{J}(u_0) = \mathbb{E}_{v_0 \in \mathbb{B}^d} \left[\nabla J(u_0 + \epsilon v_0) \right] = \frac{d}{\epsilon} \cdot \mathbb{E}_{v_0 \in \mathbb{S}^{d-1}} \left[J(u_0 + \epsilon v_0) v_0 \right], \tag{5}$$

where the random variable v_0 follows the uniform distribution on the unit sphere $\mathbb{S}^{d-1} = \partial \mathbb{B}^d$ in the last equality. Similarly, $\nabla \hat{J}(u_0)$ is approximate to $\nabla J(u_0)$, thus, $\nabla \hat{J}(u_0) \approx \nabla J(u_0)$. If the gradient ∇J exists in the ball $B(0; \delta + \epsilon)$, we can show that the error estimate of the gradient is

$$\|\nabla \hat{J}(u_0) - \nabla J(u_0)\| = O(d\epsilon). \tag{6}$$

[1] Throughout the paper, the norm $\|\cdot\|$ is defined as the Euclidean norm $\|v\| = \sqrt{\sum_{i=1}^{d} v_i^2}$.

https://doi.org/10.5194/egusphere-2022-1268
Preprint. Discussion started: 21 November 2022

The rigorous description and proof are shown in Lemma A.1 with its proof in Appendix A.

Next, we provide a simple but intuitive analysis of the convergence in probability for the samples in practice. With the representation of $\nabla\hat{J}(u_0)$ in (5), the weak law of large numbers states that the sample average converges in probability toward the expected value. Thus, for any $t > 0$, we have

$$\mathbf{Pr}\left(\left\|\frac{d}{n\epsilon}\sum_{i=1}^{n}J(u_0+\epsilon v_{0,i})v_{0,i}-\nabla\hat{J}(u_0)\right\|\geq t\right)\to 0,\qquad \text{with}\quad n\to\infty.$$

Combined with the error estimate of gradient (6), if t is assumed to be larger than $\Omega(d\epsilon)$ (i.e.,there exists a constant $\tau > 0$ such that $t > \tau d\epsilon$), then the probability that the sample average approximates to $\nabla J(u_0)$ satisfies

$$\mathbf{Pr}\left(\left\|\frac{d}{n\epsilon}\sum_{i=1}^{n}J(u_0+\epsilon v_{0,i})v_{0,i}-\nabla J(u_0)\right\|\geq t-O(d\epsilon)\right)\to 0,\qquad \text{with}\quad n\to\infty.$$

Finally, we conclude the section with the rigorous Chernoff-type bound in probability for the simple but intuitive analysis above with the following theorem. The rigorous proof is shown in Appendix A with Lemma A.2 and Lemma A.3 proposed.

Theorem 1. If J is continuously differentiable and satisfies the gradient Lipschitz condition (i.e., there exists a constant $L > 0$ such that for any $u_{0,1}, u_{0,2} \in B(0,\delta)$), then we have

$$\|\nabla J(u_{0,1})-\nabla J(u_{0,2})\|\leq L\|u_{0,1}-u_{0,2}\|.$$

For any $t > Ld\epsilon/2$, there exists a constant $C > 0$ such that the concentration inequality for samples is satisfied:

$$\mathbf{Pr}\left(\left\|\frac{d}{n\epsilon}\sum_{i=1}^{n}J(u_0+\epsilon v_{0,i})v_{0,i}-\nabla J(u_0)\right\|\geq t-\frac{Ld\epsilon}{2}\right)\leq 2\exp\left(-Cnt^2\right).$$

4 Model and numerical experiments

In this section, we perform several numerical experiments to compare the proposed sample-based algorithms with the baseline algorithms for a theoretical model, the Burgers equation with small viscosity. After the concept of CNOPs was proposed in (Mu et al., 2003), there have been several methods, adjoint-based or adjoint-free, proposed to obtain the CNOPs (Wang and Tan, 2010; Chen et al., 2015; Zheng et al., 2017; Yuan et al., 2015). However, essential difficulties, such as a massive storage space to save the basic state and instability when running an atmospheric and oceanic model, have still not been overcome. In this study, different from traditional deterministic optimization methods above, we obtain the approximate gradient by sampling objective values introduced in Section 3. Then, we use the second spectral projected gradient method (SPG2) proposed in (Birgin et al., 2000) to obtain the CNOPs.

https://doi.org/10.5194/egusphere-2022-1268
Preprint. Discussion started: 21 November 2022

We consider the simple theoretical model, the Burgers equation with small viscosity under the Dirichlet condition that describes the nonlinear time evolution of the reference state U as

$$\begin{cases} \dfrac{\partial U}{\partial t} + U\dfrac{\partial U}{\partial x} = \gamma \dfrac{\partial^2 U}{\partial x^2}, & (x,t) \in [0,L] \times [0,T] \\ U(0,t) = U(L,t) = 0, & t \in [0,T] \\ U(x,0) = \sin\left(\dfrac{2\pi x}{L}\right), & x \in [0,T] \end{cases} \tag{7}$$

where $\gamma = 0.005 m^2/s$ and $L = 100m$. We use the leapfrog/DuFort-Frankel scheme (i.e., the central finite difference scheme in both the temporal and spatial directions) to numerically solve the viscous Burgers equation above (7), with $\Delta x = 1m$ as the spatial grid size ($d = 101$) and $\Delta t = 1s$ as the time step. The objective function $J(u_0)$ used for optimization (2) can be rewritten in the form of the norm square of perturbation as

$$J(u_0) = \|u(T)\|^2 = \sum_{i=1}^{d} u_i(T)^2.$$

The constraint for the initial condition is set to be $\|u_0\| \leq \delta = 8 \times 10^{-4} m/s$. With $\epsilon = 10^{-8}$ for the definition of $\hat{J}(u_0)$ in (4), we perform numerical experiments to calculate the CNOPs directly from the definition, by the adjoint-based method and the sample-based method with $n = 5$ and $n = 15$. The prediction time is set for two cases, $T = 30s$ and $T = 60s$, respectively. The CNOPs computed by the four algorithms are shown in Figure 1, and the computation times are shown in Table 1.

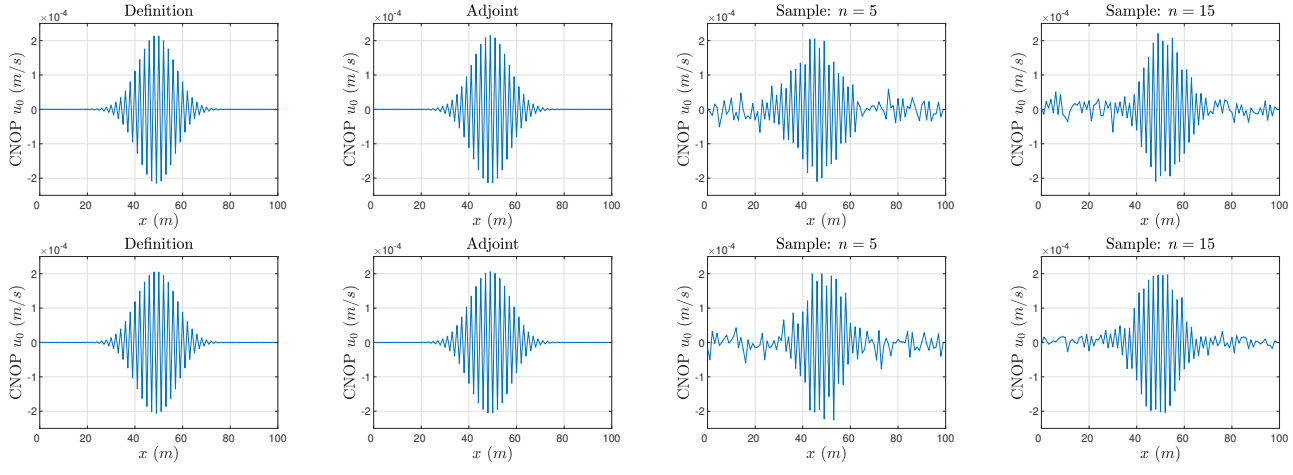

Figure 1. Spatial distributions of CNOPs (m/s). Prediction time: on the top is $T = 30s$, and on the bottom is $T = 60s$. From left to right: Definition method, Adjoint method, Sample method ($n = 5$) and Sample method ($n = 15$).

Figure 1 shows that the CNOP obtained by the adjoint method is nearly identical to that computed directly from the definition, when the numerical gradient with the spatial grid is set as $\alpha = 10^{-8}$ for both the two cases, $T = 30s$ and $T = 60s$. The computation time for the adjoint method is known to be far less than that directly computed from the definition, which is tested

https://doi.org/10.5194/egusphere-2022-1268
Preprint. Discussion started: 21 November 2022

for the Burgers equation with small viscosity in Table 1. For the sample-based method, some fluctuating errors of the CNOPs are shown in the spatial distributions (the right two columns of Figure 1) due to noise. However, the basic spatial pattern of the CNOPs can be obtained, even in the case with fewer samples $n = 5$. Also, the computation time to obtain the CNOP by taking the samples $n = 15$ is similar with that of the adjoint method. However, the computation time of the sample-based method decreases by more than half when we reduce the number of samples from $n = 15$ to $n = 5$.

Time \ Methods	Definition	Adjoint	Sample ($n = 5$)	Sample ($n = 15$)
$T = 30s$	$3.2788s$	$1.0066s$	$\mathbf{0.3836s}$	$0.8889s$
$T = 60s$	$6.3106s$	$1.4932s$	$\mathbf{0.6464s}$	$1.4845s$

Table 1. Comparison of computation time within the four algorithms for the prediction time, $T = 30s$ and $T = 60s$.

The nonlinear time evolutions of all CNOPs in terms of norm squares $\|u(t)\|^2$, which are obtained by the four algorithms above, are nearly identical in Figure 2 for both the two cases, $T = 30s$ and $T = 60s$. With $T = 30s$, the nonlinear time evolution of the CNOP in terms of norm square $\|u(t)\|^2$ is small before the perturbations start to proliferate at approximately time $t = 20s$, and then its growth changes sharply. Similarly, the nonlinear time evolution tendency is nearly identical to that with $T = 60s$ with the rapid growth of the perturbations at approximately $t = 50s$. Considering the top two figures of Figure 3, we find that the most important difference in the nonlinear time evolution of norm square of perturbations $\Delta\|u(t)\|^2$ is between the definition method and the sample-based method with $n = 5$ samples, which is less than $6 \times 10^{-7} m^2/s^2$ for the prediction time $T = 30s$ and $0.12 m^2/s^2$ for the prediction time $T = 60s$, respectively. Considering the bottom two figures of Figure 3, this phenomenon can also be observed — the relative difference within the nonlinear time evolution $\Delta\|u(t)\|^2/\|u(t)\|^2$ is nearly zero, except for a large ratio at the time $t = 11s$.

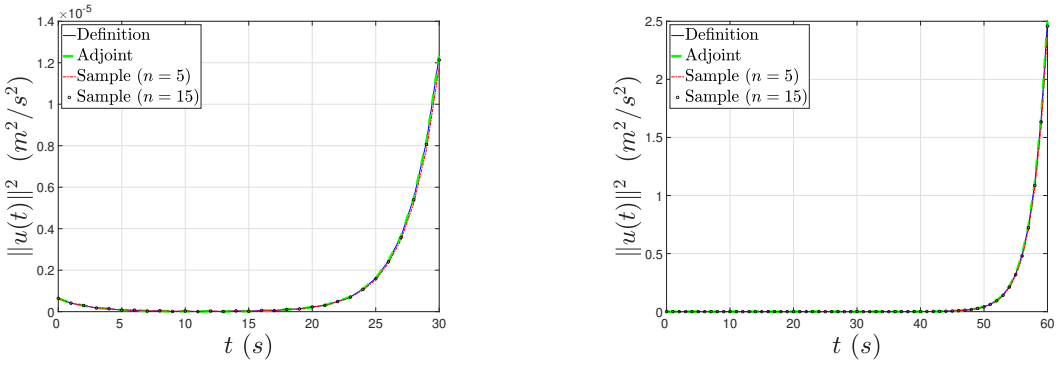

Figure 2. Nonlinear time evolution of CNOP in terms of the norm square (m^2/s^2). Left: the prediction time is $T = 30s$; Right: the prediction time is $T = 60s$, respectively.

https://doi.org/10.5194/egusphere-2022-1268
Preprint. Discussion started: 21 November 2022

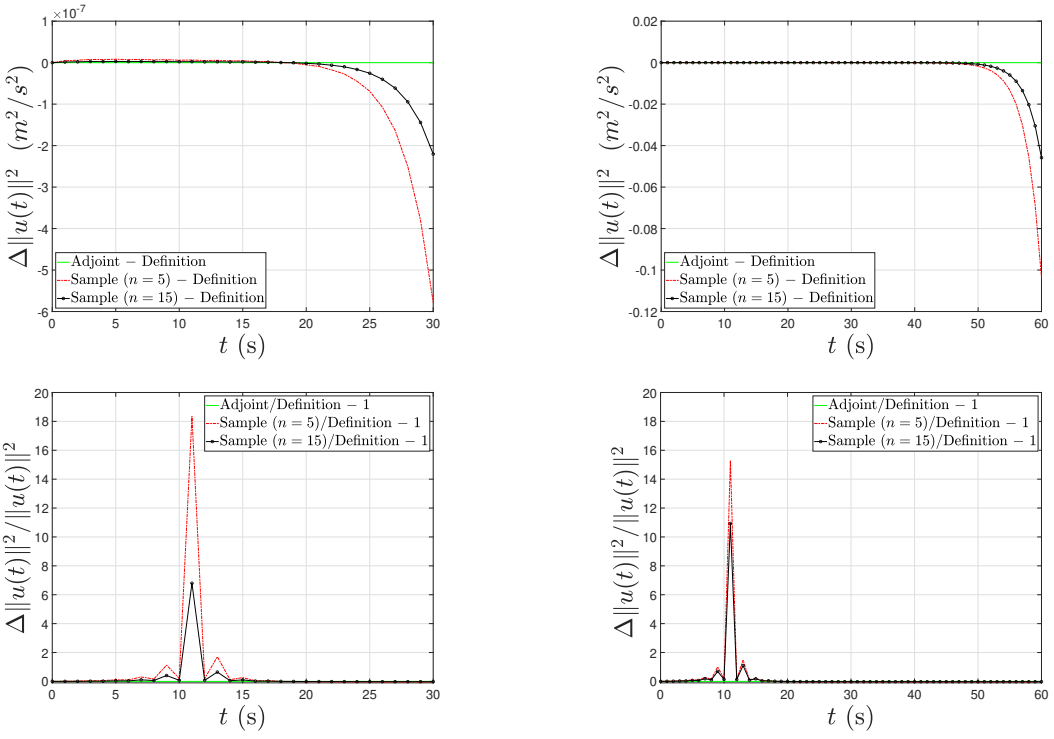

Figure 3. Nonlinear evolution of CNOPs in terms of the difference (m^2/s^2) and relative difference of the norm square of perturbations. The top two figures show the difference and the bottom two show the relative difference. Left: the prediction time is $T = 30s$; Right: the prediction time is $T = 60s$.

Based on all the results in the numerical experiments above, for Figures 1 - 3 and Table 1, we use the sample-based method with $n = 5$ to obtain the CNOP within a shorter time at the cost of losing accuracy while overcoming the disadvantage of the massive storage space for the basic state and the running instability based on the linearization.

5 Summary and discussion

In this paper, we introduced a sample-based algorithm to obtain the CNOPs, that is based on the high-dimensional Stokes' theorem and the law of large numbers. We have also provided a rigorous concentration estimate for the exact gradient by averaging the samples and compared the performance between the sample-based and baseline algorithms by performing several numerical experiments. Compared with the classical adjoint-based method, this approach is easier to implement and reduces the amount of required storage for the basic state. When we reduce the number of samples to some extent, it reduces the computation markedly more when using the sample-based method, which can guarantee that the CNOP obtained is nearly consistent with some minor fluctuating errors oscillating in spatial distribution. Currently, the CNOP method has been widely applied to the predictability of atmospheric and oceanic motions. However, for an earth system model that is more realistic,

https://doi.org/10.5194/egusphere-2022-1268
Preprint. Discussion started: 21 November 2022

such as the Community Earth System Model (CESM), many difficulties still exist in obtaining the CNOP (Wang et al., 2020), or even for a high–regional resolution model, such as the Weather Research and Forecasting (WRF) Model, which is used widely in the operational forecasting (Yu et al., 2017). Based on increasingly reliable models developed in atmospheric science and oceanography, we will now comment on some extensions of the sample-based method to investigate them using more complex models, whether theoretical or practical, to obtain the CNOPs. First, to test the validity of the sampling algorithm to calculate the CNOPs, we start from an idealized ocean-atmosphere coupling model with its adjoint model, the Zebiak-Cane (ZC) model (Zebiak and Cane, 1987), which might characterize the oscillatory behavior of ENSO in amplitude and period based on oceanic wave dynamics. Mu et al. (2007) obtained the CNOP and studied the spring predictability barrier for El Niño events by applying the ZC model and its adjoint model. In addition, Mu et al. (2009) also used the PSU/NCAR mesoscale model (i.e.,the MM5 model) with its adjoint model to explore the predictability of tropical cyclones by computing its CNOP. We implement the sample-based methods to obtain the CNOPs on the above models and compare their performances with the adjoint-based methods. Second, based on the fact about the validity of the sampling algorithm, the CNOPs will be investigated for an earth system model or atmosphere-ocean general circulation models (AOGCMs) without its adjoint model. We take the sample-based algorithms to investigate further the nonlinear instability and predictability in a popular atmospheric blocking model, the nonlinear multiscale interaction model (NMI) (Luo et al., 2014), which also successfully implements an eddy-blocking matching mechanism.

Appendix A: Proof of Theorem 1

Lemma A.1. If $\hat{J}(u_0)$ is defined in (4), then the equation (5) is satisfied. Also, under the same assumption of Theorem 1, the estimates for the objective value and gradient difference can be described by

$$\|\hat{J}(u_0) - J(u_0)\| \leq \frac{L\epsilon^2}{2}, \tag{A1}$$

$$\|\nabla\hat{J}(u_0) - \nabla J(u_0)\| \leq \frac{Ld\epsilon}{2}. \tag{A2}$$

Proof of Lemma A.1. First, with the definition of $\hat{J}(u_0)$, we show the proof about the computation of gradient $\nabla\hat{J}(u_0)$ in (5).

- For $d = 1$, the gradient $\hat{J}(u_0)$ about u_0 can be computed as

$$\frac{d\hat{J}(u_0)}{du_0} = \frac{d}{du_0}\left(\frac{1}{2}\int_{-1}^{1} J(u_0 + \epsilon v_0)dv_0\right) = \frac{1}{2}\int_{-1}^{1}\frac{dJ(u_0 + \epsilon v_0)}{\epsilon dv_0}dv_0 = \frac{J(u_0 + \epsilon) - J(u_0 - \epsilon)}{2\epsilon}.$$

https://doi.org/10.5194/egusphere-2022-1268
Preprint. Discussion started: 21 November 2022

– For the case of $d \geq 2$, we assume that $\mathbf{a} \in \mathbb{R}^d$ is an arbitrary vector. Then, the gradient $\nabla \hat{J}(u_0)$ satisfies the following equality as

$$\begin{aligned}
\boldsymbol{a} \cdot \nabla \hat{J}(u_0) &= \int_{v_0 \in \mathbb{B}^d} \boldsymbol{a} \cdot \nabla_{u_0} J(u_0 + \epsilon v_0) dV \\
&= \frac{1}{\epsilon} \int_{v_0 \in \mathbb{B}^d} \nabla_{v_0} \cdot (J(u_0 + \epsilon v_0)\boldsymbol{a}) dV \\
&= \frac{1}{\epsilon} \int_{v_0 \in \mathbb{S}^{d-1}} J(u_{0,k} + \epsilon v_0)\boldsymbol{a} \cdot v_0 dS \\
&= \boldsymbol{a} \cdot \frac{1}{\epsilon} \int_{v_0 \in \mathbb{S}^{d-1}} J(u_0 + \epsilon v_0) v_0 dS.
\end{aligned}$$

Because the vector $\boldsymbol{a}$ is arbitrary, we can obtain the following equality:

$$\nabla \int_{v_0 \in \mathbb{B}^d} J(u_0 + \epsilon v_0) dV = \frac{1}{\epsilon} \int_{v_0 \in \mathbb{S}^{d-1}} J(u_0 + \epsilon v_0) v_0 dS.$$

Then, with d being the ratio of the surface area and the volume of $\mathbb{B}^d$, the representation of the gradient in (5) is satisfied.

If J is continuously differentiable and satisfies the gradient Lipschitz condition, we can obtain the following inequality for J:

$$\left| J(u_0 + \epsilon v_0) - J(u_0) - \epsilon \left\langle \nabla J(u_0), v_0 \right\rangle \right| \leq \frac{L\epsilon^2}{2} \|v_0\|^2.$$

Because $\int_{v_0 \in \mathbb{B}^d} \langle \nabla J(u_0), v_0 \rangle \, dV = 0$, the estimate (A1) can be obtained directly. For any $i \neq j \in \{1, \ldots, d\}$, $v_{0,i}$ and $v_{0,j}$ are uncorrelated, thus

$$\int_{v_0 \in \mathbb{S}^{d-1}} v_{0,i} v_{0,j} dS = 0;$$

for the case $i = j \in \{1, \ldots, d\}$, we have

$$\int_{v_0 \in \mathbb{S}^{d-1}} v_{0,i}^2 dS = \frac{1}{d} \int_{v_0 \in \mathbb{S}^{d-1}} \left(\sum_{i=1}^{d} v_{0,i}^2 \right) dS = \frac{1}{d} \int_{v_0 \in \mathbb{S}^{d-1}} dS.$$

Then, with the row vector representation of v_0, we can obtain the following equality:

$$\mathbb{E}_{v_0 \in \mathbb{S}^{d-1}} [v_0^T v_0] = \frac{1}{d} \cdot \mathbf{I}.$$

Also, we can obtain the equivalent representation of the gradient $\nabla J(u_0)$ as

$$\nabla J(u_0) = \frac{d}{\epsilon} \cdot \mathbb{E}_{v_0 \in \mathbb{S}^{d-1}} \left[\epsilon \langle \nabla J(u_0), v_0 \rangle v_0 \right].$$

https://doi.org/10.5194/egusphere-2022-1268
Preprint. Discussion started: 21 November 2022

Finally, with $\mathbb{E}_{v_0\in\mathbb{S}^{d-1}}[v_0]=0$, the norm of the gradient difference is estimated as

$$\begin{aligned}\|\nabla\hat{J}(u_0)-\nabla J(u_0)\| &\le \left\|\frac{d}{\epsilon}\cdot\mathbb{E}_{v_0\in\mathbb{S}^{d-1}}\left[\left(J(u_0+\epsilon v_0)-J(u_0)\right)v_0\right]-\frac{d}{\epsilon}\cdot\mathbb{E}_{v_0\in\mathbb{S}^{d-1}}\left[\epsilon\langle\nabla J(u_0),v_0\rangle v_0\right]\right\| \\ &\le \frac{d}{\epsilon}\cdot\mathbb{E}_{v_0\in\mathbb{S}^{d-1}}\left[\left\|J(u_0+\epsilon v_0)-J(u_0)-\epsilon\langle\nabla J(u_0),v_0\rangle\right\|\cdot\left\|v_0\right\|\right] \\ &\le \frac{Ld\epsilon}{2},\end{aligned}$$

where the last inequality follows the gradient Lipschitz condition.

□

Considering any $\epsilon>0$, to proceed with the concentration inequality, we must still know that the random variable $J(u_0+\epsilon v_0)$ for $v_0\sim\text{Unif}(\mathbb{S}^{d-1})$ is sub-Gaussian. Thus, we first introduce the following lemma.

Lemma A.2 (Proposition 2.5.2 in Vershynin (2018))**.** Let X be a random variable. If there exist two constants $K_1,K_2>0$ such that the moment generating function of X^2 is bounded:

$$\mathbb{E}\left[\exp\left(\frac{X^2}{K_1^2}\right)\right]\le K_2,$$

then the random variable X is sub-Gaussian.

Because $J(u_0+\epsilon v_0)$ is bounded on $\mathbb{S}^{d-1}$, $\exp((J(u_0+\epsilon v_0)^2/K_1^2)$ is integrable on $\mathbb{S}^{d-1}$ for any $K_1>0$, i.e., there exists a constant $K_2>0$ such that

$$\mathbb{E}_{v_0\in\mathbb{S}^{d-1}}\left[\exp\left(\frac{J(u_0+\epsilon v_0)^2}{K_1^2}\right)\right]\le K_2.$$

With Lemma A.2, the random variable $J(u_0+\epsilon v_0)$ is sub-Gaussian. Therefore, for any fixed vector $v_0'\in\mathbb{S}^{d-1}$, we know the random variable $J(u_0+\epsilon v_0)\langle v_0,v_0'\rangle$ is sub-Gaussian. We will now introduce the following lemma to proceed with the concentration inequality.

Lemma A.3 (Theorem 2.6.3 in Vershynin (2018))**.** Let $X_1,\dots,X_n$ be independent, mean zero, sub-Gaussian random variables, and $a=(a_1,\dots,a_n)\in\mathbb{R}^n$. Then, for every $t\ge0$, we have

$$\mathbf{Pr}\left(\left|\sum_{i=1}^n a_iX_i\right|\ge t\right)\le 2\exp\left(-\frac{ct^2}{K^2\|a\|^2}\right),$$

where $K=\max_{1\le i\le n}\|X_i\|_{\psi_2}$.[2]

[2]The sub-Gaussian norm of a random variable X is defined as

$$\|X\|_{\psi_2}=\inf\left\{t>0:\ \mathbb{E}\exp\left(\frac{X^2}{t^2}\right)\le 2\right\}.$$

https://doi.org/10.5194/egusphere-2022-1268
Preprint. Discussion started: 21 November 2022

With Lemma A.1 and Lemma A.3, we can obtain the concentration inequality for the samples as

$$\mathbf{Pr}\left(\left|\frac{d}{n\epsilon}\sum_{i=1}^{n}\left\langle J(u_0+\epsilon v_{0,i})v_{0,i}v_0'\right\rangle-\left\langle\nabla\hat{J}(u_0),v_0'\right\rangle\right|\geq t\right)\leq 2\exp\left(-\frac{cnt^2}{K^2}\right).$$

Because v_0' is a unit vector on $\mathbb{S}^{d-1}$, we can proceed with the concentration estimate as follows by the Cauchy-Schwarz inequality:

$$\mathbf{Pr}\left(\left\|\frac{d}{n\epsilon}\sum_{i=1}^{n}J(u_0+\epsilon v_{0,i})v_{0,i}-\nabla\hat{J}(u_0)\right\|\geq t\right)\leq 2\exp\left(-\frac{cnt^2}{K^2}\right).$$

Based on the triangle inequality, the concentration inequality can proceed with the estimate of the gradient difference (A2) as

$$\mathbf{Pr}\left(\left\|\frac{d}{n\epsilon}\sum_{i=1}^{n}J(u_0+\epsilon v_{0,i})v_{0,i}-\nabla J(u_0)\right\|\geq t-\frac{Ld\epsilon}{2}\right)\leq 2\exp\left(-\frac{cnt^2}{K^2}\right)$$

for any $t > Ld\epsilon/2$. With $C = c/K^2$, the proof of Theorem 1 is complete.

Author contributions. Bin Shi constructed the basic idea of this paper, wrote the Matlab code of the sampling method and plotted all the figures, and wrote the manuscript. Guodong Sun joined the discussions of this manuscript and provided some suggestions. All the authors contributed to the writing and reviewing of the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We are indebted to Mu Mu for seriously reading an earlier version of this paper and providing his suggestions about this theoretical study. Bin Shi would also like to thank Ya-xiang Yuan, Ping Zhang, and Yu-hong Dai for their encouragement to understand and analyze the nonlinear phenomena in nature from the perspective of optimization in the early stages of this project. This work was supported by Grant No.YSBR-034 of CAS.

https://doi.org/10.5194/egusphere-2022-1268
Preprint. Discussion started: 21 November 2022

References

Barclay, A., Gill, P. E., and Ben Rosen, J.: SQP methods and their application to numerical optimal control, in: Variational calculus, optimal control and applications, pp. 207–222, Springer, 1998.

Birgin, E. G., Martínez, J. M., and Raydan, M.: Nonmonotone spectral projected gradient methods on convex sets, SIAM Journal on Optimization, 10, 1196–1211, 2000.

Bottou, L., Curtis, F. E., and Nocedal, J.: Optimization methods for large-scale machine learning, SIAM Review, 60, 223–311, 2018.

Buizza, R. and Palmer, T. N.: The singular-vector structure of the atmospheric global circulation, J. Atmos. Sci., 52, 1434–1456, 1995.

Chen, L., Duan, W., and Xu, H.: A SVD-based ensemble projection algorithm for calculating the conditional nonlinear optimal perturbation, Science China Earth Sciences, 58, 385–394, 2015.

Duan, W. and Hu, J.: The initial errors that induce a significant "spring predictability barrier" for El Niño events and their implications for target observation: Results from an earth system model, Climate Dynamics, 46, 3599–3615, 2016.

Duan, W., Liu, X., Zhu, K., and Mu, M.: Exploring the initial errors that cause a significant "spring predictability barrier" for El Niño events, Journal of Geophysical Research: Oceans, 114, 2009.

Eady, E. T.: Long waves and cyclone waves, Tellus, 1, 33–52, 1949.

Farrell, B. F.: The initial growth of disturbances in a baroclinic flow, J. Atmos. Sci., 39, 1663–1686, 1982.

Farrell, B. F. and Ioannou, P. J.: Generalized stability theory. Part I: Autonomous operators, Journal of Atmospheric Sciences, 53, 2025–2040, 1996a.

Farrell, B. F. and Ioannou, P. J.: Generalized stability theory. Part II: Nonautonomous operators, Journal of Atmospheric Sciences, 53, 2041–2053, 1996b.

Fujii, Y., Tsujino, H., Usui, N., Nakano, H., and Kamachi, M.: Application of singular vector analysis to the Kuroshio large meander, Journal of Geophysical Research: Oceans, 113, 2008.

Kalnay, E.: Atmospheric modeling, data assimilation and predictability, Cambridge university press, 2003.

Kerswell, R. R.: Nonlinear nonmodal stability theory, Annual Review of Fluid Mechanics, 50, 319–345, 2018.

Lin, C.-C.: The theory of hydrodynamic stability, Cambridge University Press, Cambridge, UK., 1955.

Liu, D. C. and Nocedal, J.: On the limited memory BFGS method for large scale optimization, Mathematical programming, 45, 503–528, 1989.

Lorenz, E. N.: A study of the predictability of a 28-variable atmospheric model, Tellus, 17, 321–333, 1965.

Luo, D., Cha, J., Zhong, L., and Dai, A.: A nonlinear multiscale interaction model for atmospheric blocking: The eddy-blocking matching mechanism, Quarterly Journal of the Royal Meteorological Society, 140, 1785–1808, 2014.

Mu, M.: Nonlinear singular vectors and nonlinear singular values, Science in China Series D: Earth Sciences, 43, 375–385, 2000.

Mu, M. and Wang, J.: Nonlinear fastest growing perturbation and the first kind of predictability, Science in China Series D: Earth Sciences, 44, 1128–1139, 2001.

Mu, M. and Wang, Q.: Applications of nonlinear optimization approach to atmospheric and oceanic sciences, SCIENTIA SINICA Mathematica, 47, 1207–1222, 2017.

Mu, M., Duan, W. S., and Wang, B.: Conditional nonlinear optimal perturbation and its applications, Nonlinear Processes in Geophysics, 10, 493–501, 2003.

https://doi.org/10.5194/egusphere-2022-1268
Preprint. Discussion started: 21 November 2022

Mu, M., Sun, L., and Dijkstra, H. A.: The sensitivity and stability of the ocean's thermohaline circulation to finite-amplitude perturbations, Journal of Physical Oceanography, 34, 2305–2315, 2004.

Mu, M., Xu, H., and Duan, W.: A kind of initial errors related to "spring predictability barrier" for El Niño events in Zebiak-Cane model, Geophysical Research Letters, 34, 2007.

Mu, M., Zhou, F., and Wang, H.: A method for identifying the sensitive areas in targeted observations for tropical cyclone prediction: Conditional nonlinear optimal perturbation, Monthly Weather Review, 137, 1623–1639, 2009.

Navrose, Johnson, H. G., Brion, V., Jacquin, L., and Robinet, J.-C.: Optimal perturbation for two-dimensional vortex systems: route to non-axisymmetric state, Journal of Fluid Mechanics, 855, 922–952, 2018.

Nemirovski, A. S. and Yudin, D. B.: Problem complexity and method efficiency in optimization, 1983.

Pringle, C. C. T. and Kerswell, R. R.: Using nonlinear transient growth to construct the minimal seed for shear flow turbulence, Physical review letters, 105, 154 502, 2010.

Qin, X. and Mu, M.: Influence of conditional nonlinear optimal perturbations sensitivity on typhoon track forecasts, Quarterly Journal of the Royal Meteorological Society, 138, 185–197, 2012.

Rayleigh, L.: On the stability, or instability, of certain fluid motions, Proceedings of the London Mathematical Society, 1, 57–72, 1879.

Samelson, R. M. and Tziperman, E.: Instability of the chaotic ENSO: The growth-phase predictability barrier, J. Atmos. Sci., 58, 3613–3625, 2001.

Sun, G. and Mu, M.: A new approach to identify the sensitivity and importance of physical parameters combination within numerical models using the Lund–Potsdam–Jena (LPJ) model as an example, Theoretical and Applied Climatology, 128, 587–601, 2017.

Thompson, C. J.: Initial conditions for optimal growth in a coupled ocean–atmosphere model of ENSO, J. Atmos. Sci., 55, 537–557, 1998.

Vershynin, R.: High-dimensional probability: An introduction with applications in data science, vol. 47, Cambridge university press, 2018.

Wang, B. and Tan, X.: Conditional nonlinear optimal perturbations: Adjoint-free calculation method and preliminary test, Monthly Weather Review, 138, 1043–1049, 2010.

Wang, Q. and Mu, M.: A new application of conditional nonlinear optimal perturbation approach to boundary condition uncertainty, Journal of Geophysical Research: Oceans, 120, 7979–7996, 2015.

Wang, Q., Mu, M., and Sun, G.: A useful approach to sensitivity and predictability studies in geophysical fluid dynamics: conditional nonlinear optimal perturbation, National Science Review, 7, 214–223, 2020.

Xue, Y., Cane, M. A., and Zebiak, S. E.: Predictability of a coupled model of ENSO using singular vector analysis. Part I: Optimal growth in seasonal background and ENSO cycles, Mon. Wea. Rev., 125, 2043–2056, 1997a.

Xue, Y., Cane, M. A., Zebiak, S. E., and Palmer, T. N.: Predictability of a coupled model of ENSO using singular vector analysis. Part II: Optimal growth and forecast skill, Mon. Wea. Rev., 125, 2057–2073, 1997b.

Yu, H., Wang, H., Meng, Z., Mu, M., Huang, X.-Y., and Zhang, X.: A WRF-based tool for forecast sensitivity to the initial perturbation: The conditional nonlinear optimal perturbations versus the first singular vector method and comparison to MM5, Journal of Atmospheric and Oceanic Technology, 34, 187–206, 2017.

Yuan, S., Zhao, L., and Mu, B.: Parallel cooperative co-evolution based particle swarm optimization algorithm for solving conditional nonlinear optimal perturbation, in: International Conference on Neural Information Processing, pp. 87–95, Springer, 2015.

Zanna, L., Heimbach, P., Moore, A. M., and Tziperman, E.: Optimal excitation of interannual Atlantic meridional overturning circulation variability, Journal of Climate, 24, 413–427, 2011.

Zebiak, S. E. and Cane, M. A.: A model El Niño–Southern Oscillation, Monthly Weather Review, 115, 2262–2278, 1987.

https://doi.org/10.5194/egusphere-2022-1268
Preprint. Discussion started: 21 November 2022

335 Zheng, Q., Yang, Z., Sha, J., and Yan, J.: Conditional nonlinear optimal perturbations based on the particle swarm optimization and their applications to the predictability problems, Nonlinear Processes in Geophysics, 24, 101–112, 2017.

Zu, Z., Mu, M., and Dijkstra, H. A.: Optimal initial excitations of decadal modification of the Atlantic Meridional Overturning Circulation under the prescribed heat and freshwater flux boundary conditions, Journal of Physical Oceanography, 46, 2029–2047, 2016.