Computers & Geosciences 106 (2017) 190–199

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Research paper

# An adjoint-free method to determine conditional nonlinear optimal perturbations

Aleid Oosterwijk[a,b], Henk A. Dijkstra[a,*], Tristan van Leeuwen[b]

[a] Department of Physics and Astronomy, Utrecht University, Utrecht, The Netherlands
[b] Department of Mathematics, Utrecht University, Utrecht, The Netherlands

ARTICLE INFO

*Keywords:*
Derivative-free nonlinear optimisation
CNOP
PCA
Ocean model application

ABSTRACT

The analysis of the growth of initial perturbations in dynamical systems is an important aspect of predictability theory because it informs on error growth. The Conditional Nonlinear Optimal Perturbation (CNOP) method is an approach where the nonlinear growth of perturbations is determined over a certain lead time. The CNOPs can be found by a nonlinear constrained optimisation problem, which is typically solved using sequential quadratic programming (SQP), a routine that requires an adjoint model. Such adjoint models are not always available and hence we here study the performance of an adjoint-free optimisation method (COBYLA), in combination with a dimension reduction technique, to determine CNOPs. The new technique is applied to a quasi-geostrophic model of the wind-driven ocean circulation. We find that COBYLA is able to find good approximations of CNOPs, albeit at a higher computational cost than conventional adjoint-based methods.

## 1. Introduction

In weather prediction and climate projection, the future state of a model is determined by evolving an ensemble of initial states forward. The spread of the states obtained from different ensemble members gives a measure of the uncertainty of the forecast (Slingo and Palmer, 2011). The European Centre for Medium-Range Weather Forecasting (ECMWF) uses linear singular vectors (LSV) for generating the ensemble members. The LSV method is based on a linearisation of the numerical weather prediction model, the tangent linear model (TLM), and provides the (linearly) fastest growing perturbations over a certain lead time.

In many applications, however, nonlinear effects are important for error growth and dominate the development of perturbations, even for short lead-times (Mu et al., 2003). Therefore, nonlinear extensions of the LSV method have been developed to determine the fastest growing initial perturbations. Examples are methods to determine nonlinear singular vectors (NSVAs) (Mu, 2000), and conditional nonlinear optimal perturbations (CNOPs) (Mu et al., 2003). The CNOP method has been applied successfully to many problems in climate dynamics, such as the wind-driven ocean circulation (Terwisscha van Scheltinga and Dijkstra, 2008), the thermohaline ocean circulation (Zu et al., 2016) and El Niño Duan et al.

A CNOP can be obtained by maximising an objective function subject to constraints and this constrained optimisation problem is usually solved by the Sequential Quadratic Programming (SQP) method (Mu et al., 2003). However, SQP is only applicable for models for which an adjoint model is available (Nocedal and Wright, 2006). This adjoint model is used to determine gradients of the objective function with respect to the initial condition (Kalnay, 2003). In this way, the initial condition which optimises the objective function can be efficiently determined. However, for many models the calculation of the adjoint model is not straightforward and derivative-free methods for optimisation are therefore needed.

Recently, particle swarm optimisation (PSO) has been successfully applied for determining CNOPs of a model of intermediate complexity (the Zebiak and Cane (ZC) model (Zebiak and Cane, 1987) of the equatorial Pacific climate to find optimal precursors of El Niño events (Mu et al., 2015). PSO is inspired by the behaviour of swarms, such as flocks of birds or schools of fish, in which each individual seeks to improve its position while responding to the behaviour of its neighbours. The computational burden of PSO rapidly increases when the dimension of the search space is increased, which is known in general as *the curse of dimensionality*. Therefore, in order to apply PSO to large-dimensional dynamical systems, dimension reduction is necessary.

There have been several studies to compute CNOPs using model reduction techniques. Sun et al. (2010) transformed the constrained optimisation problem into an unconstrained optimisation problem by reducing the dimension. A traditional unconstrained optimisation

* Corresponding author.
*E-mail address:* h.a.dijkstra@uu.nl (H.A. Dijkstra).

http://dx.doi.org/10.1016/j.cageo.2017.06.014
Received 19 October 2016; Received in revised form 19 April 2017; Accepted 19 June 2017
Available online 24 June 2017
0098-3004/ 

algorithm could then be applied to obtain the CNOP. Sun and Mu (2013) applied a differential evolution method (Storn and Price, 1997), which is an adjoint-free method, to compute the CNOP. In addition, in Chen et al. (2015) an SVD-based ensemble projection algorithm was applied to reduce the dimension of the optimisation problem to compute the CNOP. Mu et al. (2015) use a combination of PCA (principal component analysis) in conjunction with PSO. Although they show that the CPU time is shortened by reducing the dimension of the search space, this often leads to a less accurate CNOP, as compared to the SQP solution. However, for a large enough dimension of the search space, the spatial structure of the CNOPs found with PSO is comparable to the CNOPs found by SQP. For the Zebiak-Cane model used in Mu et al. (2015), the optimal size of the search space is found to be 30, whereas the original dimension of the search space is 1080. Even with this dramatic reduction of the search space dimension, the CPU time of the PCA-based PSO is approximately three times larger than that using SQP.

A disadvantage of PSO is that it is a highly heuristic method that has many free, problem-specific, parameters that need to be tuned. In this paper, we explore the use of the *Constrained Optimisation by Linear Approximation* (COBYLA) method (Powell, 1994) for the computation of CNOPs. COBYLA is based on the Nelder-Mead simplex method (Nocedal and Wright, 2006) that is adapted for constrained optimisation. This method has virtually no free parameters and therefore has application potential for a wide range of numerical weather and climate models. Here, the performance of a combination of COBYLA and a dimension-reduction method is compared to SQP, in terms of accuracy and speed, when applied to a quasi-geostrophic ocean model.

The paper is organized as follows. In Section 2, the theory behind CNOPs and the quasi-geostrophic ocean model will be briefly described. In Section 3 we will recall the algorithms behind COBYLA (Powell, 1994), as well as the dimension-reduction methodology. In Section 4, the effect of varying dimensionality on the performance of COBYLA is discussed. We will show that COBYLA is able to compute good approximations to the CNOPs, but at higher computational cost (or with lower accuracy) than with SQP. We will end with a discussion (Section 5) where the issue of the applicability of the new adjoint-free method to solving CNOPs in high-dimensional dynamical systems is addressed.

## 2. Formulation

For convenience, we shortly repeat in Section 2.1 the theory of the CNOP (Mu et al., 2003) and the equations for the ocean model (Section 2.2) (Terwisscha van Scheltinga and Dijkstra, 2008).

### 2.1. CNOP

Assume that the dynamical system can be described by the following finite dimensional system of ordinary differential equations (ODEs)

$$\frac{\partial \boldsymbol{w}}{\partial t} + F(\boldsymbol{w}) = 0, \tag{1a}$$

$$\boldsymbol{w}|_{t=0} = \boldsymbol{w}_0, \tag{1b}$$

where $\boldsymbol{w}(t)$ is the state vector, $(\boldsymbol{w}, t) \in \mathbb{R}^d \times [0, t_e]$. Furthermore, $F$ is a nonlinear operator, $\boldsymbol{w}_0$ is the initial state and $t_e \in \mathbb{R}^+$ is the lead time.

Suppose the initial value problem (1) is well-posed and $\mathcal{M}$ is the nonlinear propagator from 0 to time $t_e$, so $\boldsymbol{w}(t_e) = \mathcal{M}(\boldsymbol{w}_0)(t_e)$ is well-defined. Let $\overline{\boldsymbol{w}}(t)$ and $\widetilde{\boldsymbol{w}}(t)$ be two solutions of the system, with initial conditions $\overline{\boldsymbol{w}}_0$ and $\overline{\boldsymbol{w}}_0 + \boldsymbol{w}_0$, respectively. Choose $\overline{\boldsymbol{w}}_0$ a specific (in our case, steady state) solution to the system and $\boldsymbol{w}_0$ a small initial perturbation to this state. Integrating the state and the perturbed state is done by applying the nonlinear propagator, i.e. $\mathcal{M}(\overline{\boldsymbol{w}}_0)(t_e)$ and $\mathcal{M}(\overline{\boldsymbol{w}}_0 + \boldsymbol{w}_0)(t_e)$.

The goal is to find an initial perturbation which causes the largest growth after time $t_e$, as measured in a certain norm. The conditional nonlinear optimal perturbation (CNOP) is formally defined as the perturbation $\boldsymbol{w}_0^\delta$ at which the objective function

$$J(\boldsymbol{w}_0) = \| \mathcal{M}(\overline{\boldsymbol{w}}_0 + \boldsymbol{w}_0)(t_e) - \mathcal{M}(\overline{\boldsymbol{w}}_0)(t_e) \|_2^2, \tag{2}$$

attains its maximum value under the constraint that $\| \boldsymbol{w}_0 \| \leq \delta$. Here, $\|\cdot\|_2$ denotes the standard Euclidian norm and $\|\cdot\|$ is an appropriate, problem dependent, norm. To find the CNOPs of a dynamical system we now need to solve the following constrained optimisation problem

$$\max_{\boldsymbol{w}_0} J(\boldsymbol{w}_0) \quad \text{subject to} \quad \| \boldsymbol{w}_0 \| \leq \delta. \tag{3}$$

### 2.2. Quasi-geostrophic model

As a test problem, we use a quasi-geostrophic model of the midlatitude wind-driven ocean circulation. The quasi-geostrophic model is a cornerstone model not only for the ocean circulation but also for the midlatitude atmospheric circulation. Here, the model describes the flow in a rectangular ocean basin of size $L \times L$ and depth $D$. This basin is situated on a midlatitude $\beta$-plane with $\theta_0 = 45°N$, $f_0 = 2\Omega \sin\theta_0$ where $\Omega$ is the angular frequency of the Earth's rotation. The meridional gradient of the Coriolis parameter is denoted by $\beta_0$ and $x$, $y$ are local Cartesian (zonal and meridional) coordinates. We assume a constant density $\rho_0$ of the ocean water and a wind-forcing $\boldsymbol{\tau} = \tau_0[\tau^x(x, y), \tau^y(x, y)]$, where $\tau_0$ is a typical amplitude.

The governing equations are made dimensionless by using the horizontal length scale $L$, the vertical length scale $D$, a horizontal velocity scale $U$ and an advective timescale $L/U$. The dimensionless barotropic quasi-geostrophic model of the flow for the vorticity $\zeta$ and the geostrophic stream function $\psi$ is

$$\left[\frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y}\right][\zeta + \beta y] = Re^{-1}\nabla^2\zeta + \alpha_\tau\left(\frac{\partial \tau^y}{\partial x} - \frac{\partial \tau^x}{\partial y}\right) \tag{4a}$$

$$\zeta = \nabla^2\psi. \tag{4b}$$

The horizontal velocities are given by

$$u = -\frac{\partial \psi}{\partial y}, \quad v = \frac{\partial \psi}{\partial x}. \tag{5}$$

The parameters in the governing equations are given by the Reynolds number $Re$, the planetary vorticity gradient parameter $\beta$, and the wind stress forcing strength, $\alpha_\tau$. These are defined as

$$Re = \frac{UL}{A_H}, \quad \beta = \frac{\beta_0 L^2}{U}, \quad \alpha_\tau = \frac{\tau_0 L}{\rho D U^2}.$$

Here $g$ is the gravitational acceleration, $A_H$ is the lateral friction coefficient. We choose $U = \tau_0/(\rho D \beta_0 L)$ (the so-called Sverdrup scaling) from which it follows that $\alpha_\tau = \beta$. All parameter values are chosen equal to the values in Terwisscha van Scheltinga and Dijkstra (2008).

Finally, the boundary conditions are given by

$$\psi = \frac{\partial \psi}{\partial x} = 0 \text{ at } x = 0 \text{ and } x = 1, \tag{6a}$$

$$\psi = \zeta = 0 \text{ at } y = 0 \text{ and } y = 1. \tag{6b}$$

which represent no-slip at the continental boundaries $x = 0, 1$ and slip at $y = 0, 1$. In this paper, the (north-south) symmetric dimensionless wind-stress profile is prescribed as

$$\tau^x = -\frac{1}{2\pi}\cos(2\pi y),$$
$$\tau^y = 0.$$

### 2.3. Numerical implementation

The spatial discretisation of the QG model is done using second-order central differences and the grid that is used in the implementation is $N \times M$, with grid sizes $\Delta x = 1/N$ and $\Delta y = 1/M$. This implies that our state vector $\boldsymbol{w}$, which consists of the $\psi$ values at the grid points, has a dimension of $NM$. Time integration is done using the second-order implicit Crank-Nicolson (Terwisscha van Scheltinga and Dijkstra, 2008) scheme with a time step $\Delta t$. To apply the theory of CNOPs to this model, the nonlinear propagator $\mathcal{M}$ is required. This can be calculated by solving the governing equations for a small time step $\Delta t$. The action of the propagator on a specific initial condition now requires $t_e/\Delta t$ timesteps.

The norm $\|\cdot\|$ that is implemented is the kinetic energy norm as used in Terwisscha van Scheltinga and Dijkstra (2008), which is based on the kinetic energy $E = \frac{1}{2}\int_V (\boldsymbol{u}_0^2 + v_0^2)\mathrm{d}x\mathrm{d}y$, where $\boldsymbol{u}$ and $\boldsymbol{v}$ are the vectors of zonal and meridional velocities at every grid point. For the discretised quasi-geostropic model with state vector $\boldsymbol{w}$ the kinetic energy norm is calculated as follows. Let $\boldsymbol{K}$ be a linear operator that maps $\boldsymbol{w}$ (consisting of values of the streamfunction) to the velocity vector according to the discretized equations (5), which express how the velocity is calculated from the streamfunction. The kinetic energy norm[1] is then given by

$$\| \boldsymbol{w} \| = \frac{1}{2}\Delta x \Delta y \, \| \boldsymbol{K}\boldsymbol{w} \|_2^2. \tag{7}$$

## 3. Adjoint-free approach

In Terwisscha van Scheltinga and Dijkstra (2008), the CNOPs of the quasi-geostrophic model, using the same norms as above, were determined using the SQP method (Nocedal and Wright, 2006). Indeed, for this model, an adjoint can easily be determined by transposing the Jacobian matrix of the model (which is needed for the Crank-Nicholson scheme). In this section, we will present the new adjoint-free method to determined the CNOP of this model.

### 3.1. COBYLA

Constrained optimisation by linear approximation, COBYLA, is an optimisation algorithm that iteratively defines a simplex on which the objective function is linearised, by linear interpolation between the values at the vertices, and optimised. After this a new simplex is constructed by expanding, contracting or reflecting the current simplex. The new simplex is chosen in order to improve the objective value or the shape of the simplex. The size of the simplex edges is eventually reduced during the process, starting from size $\rho_{start}$. The process terminates when the edges have size $\rho_{end}$. An illustration of the COBYLA process in two-dimensions is shown in Fig. 1. The constraint is implemented using a penalty function, which is adapted during the iteration process. COBYLA can handle very general nonlinear constraint conditions, as within each iteration, these are approximated by linear constraints. For example, for a constraint $c(x) \leq u$ with $c\colon \mathbb{R}^n \to \mathbb{R}^m$ and $u \in \mathbb{R}^m$, at every iteration $x_k$, the constraint is approximated through $\nabla c(x_k)^T(x - x_k) \leq u$.

The method is initialised from a position in the search space, randomly chosen from a uniform distribution. One of the basic methods underlying COBYLA is the Nelder-Mead (Nocedal and Wright, 2006) method. This convergence of this method has been proved for strictly convex functions in one and two dimensions (Lagarias et al., 1998). This clearly does not apply to our objective function, but Nelder-Mead is widely used and generally shows good convergence results (Nocedal and Wright, 2006). Unfortunately, there is no general rule for choosing $\rho_{end}$ as the size of the simplex is measured as an absolute Euclidian distance between two solution vectors. In practice the Euclidian distance may not be a natural way to measure the distance between two solutions, as this disregards any physical interpretation the solution may have. The best one can do is to investigate several solutions as the iterations proceed and decide when to stop (and which one to use) based on expert knowledge of the application. A more detailed description of the COBYLA method can be found in Powell (1994) and its application here to computing the CNOP is provided in Appendix B3.

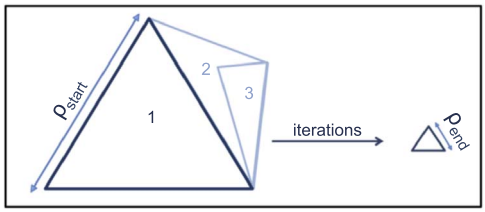



**Fig. 1.** Illustration of COBYLA algorithm in two dimensions. The initial simplex with edges of size $\rho_{start}$ is denoted by 1, subsequent simplexes in light blue are denoted by 2 and 3. The final simplex has edge length $\rho_{end}$.

### 3.2. Dimension reduction strategy

Because in general derivative-free optimisation algorithms do not perform very well on large-dimensional search spaces, dimension reduction is applied before using COBYLA. As argued in Osborne and Pastorello (1993), in a forced-dissipative system like the quasi-geostrophic ocean model, equilibrium states can often be embedded into low-dimensional subspaces. Properties of this space can, for example, be determined from the analysis of the covariances of a long transient simulation of the model (Osborne and Pastorello, 1993). Principle Component Analysis (PCA) is a widely used and effective dimension reduction method Hannachi (2004).

From the matrix $\boldsymbol{W} = [\boldsymbol{w}(t_0), \boldsymbol{w}(t_1), \ldots, \boldsymbol{w}(t_K)]$ consisting of states of a long time series, the principal components (PCs) are obtained using an eigenvalue decomposition of the matrix $\boldsymbol{W}\boldsymbol{W}^T$, i.e.,

$$\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{a} = \lambda\boldsymbol{a}$$

The eigenvectors $\boldsymbol{a}_i$, $i = 1, \ldots, d$ obtained this way are called Empirical Orthogonal Functions (EOFs). The EOFs corresponding to the largest $k$ eigenvalues account for a fraction of the variance in the data equal to the fraction of the sum of the corresponding eigenvalues and the sum of all eigenvalues. So PCA is a reduction method that captures as much variance as possible. The principal component $\boldsymbol{c}_i$ corresponding to the EOF $\boldsymbol{a}_i$ is the projection of $\boldsymbol{W}$ onto this EOF:

$$\boldsymbol{c}_i = \boldsymbol{W}^T\boldsymbol{a}_i.$$

Hence, any state $\boldsymbol{w}$ can be approximated as

$$\boldsymbol{w} \approx \boldsymbol{w}_R = \sum_{i=1}^{k} \alpha_i \boldsymbol{a}_i,$$

where $\boldsymbol{w}_R$ is the reduced state, $\boldsymbol{a}_i$ is the $i$th EOF and $\alpha_i$ is the contribution of $\boldsymbol{a}_i$ to $\boldsymbol{w}_R$. When $k = d$, the coordinate transformation is exact and $\boldsymbol{w}_R = \boldsymbol{w}$.

A step-by-step guide for the implementation of COBYLA in combination with PCA can be found in Appendix B.

## 4. Results

In this section we will compare the performance of the adjoint-free method (presented in Section 3) to the adjoint-based method SQP as applied to computing the CNOPs of the quasi-geostrophic ocean model (Section 2).

[1] Note that this is actually a 2-seminorm. However, it can still be used to measure the constraint and objective value.

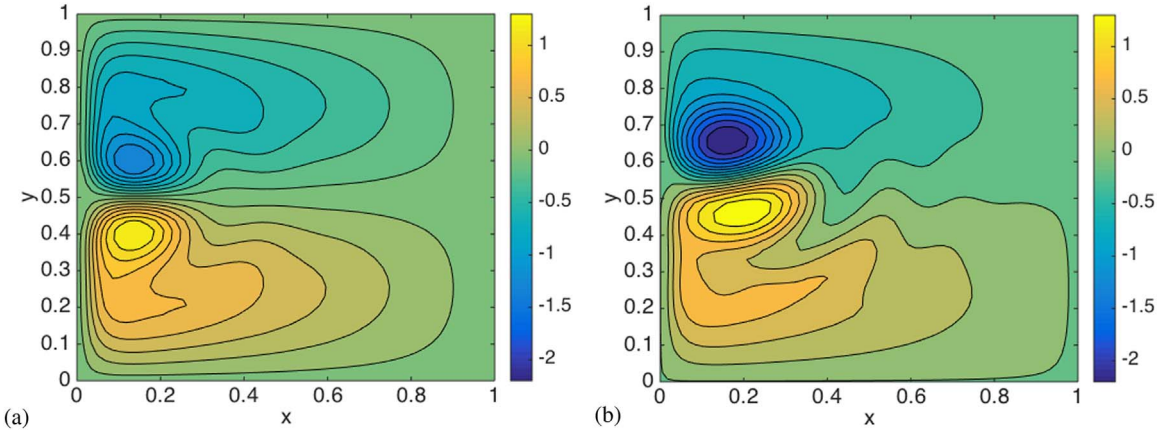



**Fig. 2.** Stream function of steady states of the quasi-geostrophic model. (a) The anti-symmetric state at $Re = 25$. (b) Jet-up state asymmetric state at $Re = 50$.

### 4.1. Standard case

The bifurcation diagram of the quasi-geostrophic model versus the Reynolds number ($Re$) was presented in Terwisscha van Scheltinga and Dijkstra (2008) for a grid resolution of $N = 60$ and $M = 40$ (which provides sufficiently accurate results). Up to the first pitchfork bifurcation (at about $Re_P = 30$) there is a unique anti-symmetric steady state solution. For $Re > Re_P$, there are at least two stable steady asymmetric solutions and one unstable anti-symmetric solution. We choose a solution from both the unique ($Re = 25$) and from the multiple equilibrium ($Re = 50$) regime. The steady states of the model at these values of $Re$ are shown in the Figs. 2a and b, respectively.

In the transient calculations with the model, we choose a time step $\Delta t = 0.001$, which corresponds to about 1.65 days. The CNOPs are calculated with a constraint condition $\delta = 0.1$ and $t_e = 6.6$ days (4 time steps). The CNOPs for these settings, as calculated by SQP, are shown in Fig. 3. In Fig. 4 the value of the objective function between 0 and $t_e$ shows the growth of the CNOP in time. The SQP algorithm needs approximately 600 objective function calls to find the CNOP for $Re = 25$, and 200 objective function calls to find the CNOP for $Re = 50$. The solution of the SQP algorithm is reliable, as the algorithm performs very consistently. In Terwisscha van Scheltinga and Dijkstra (2008) the solution is assumed to be the global maximum; in this study this assumption is made as well.

### 4.2. Dimension reduction

To use PCA, a training set of stream function fields is produced by running the model for 10 years with an additional noise on the wind-stress, starting from each steady state. The dimensionless zonal wind stress field used is $\tau^x = -1/(2\pi)\cos(2\pi y) + r_t\zeta \sin(2\pi y)$ where $r_t = 0.1$ can be considered as the amplitude of the noise and $\zeta$ is a random variable from a Gaussian distribution with mean 0 and variance 1. Each column of the training set is a state vector at a certain time. All rows of the training set are centered (by subtracting the temporal mean of the state at each grid point), to obtain the centered training set $\boldsymbol{W} = [\boldsymbol{w}(t_0), \boldsymbol{w}(t_1), \ldots, \boldsymbol{w}(t_K)]$.

In both cases $Re = 25$ and $Re = 50$, a training set of length $K = 500$ is used. The CNOPs as calculated with SQP are projected onto the EOFs

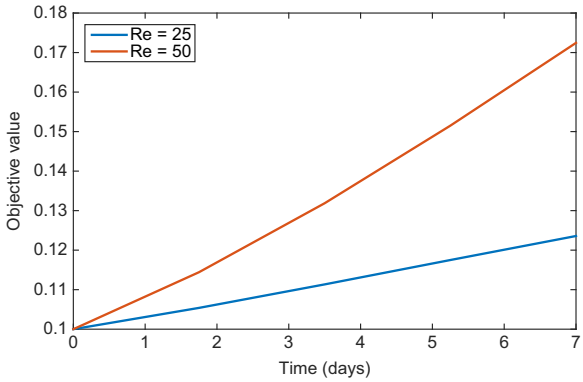



**Fig. 4.** The growth of the objective function value in time (in days) for the CNOPs as plotted in Fig. 3.

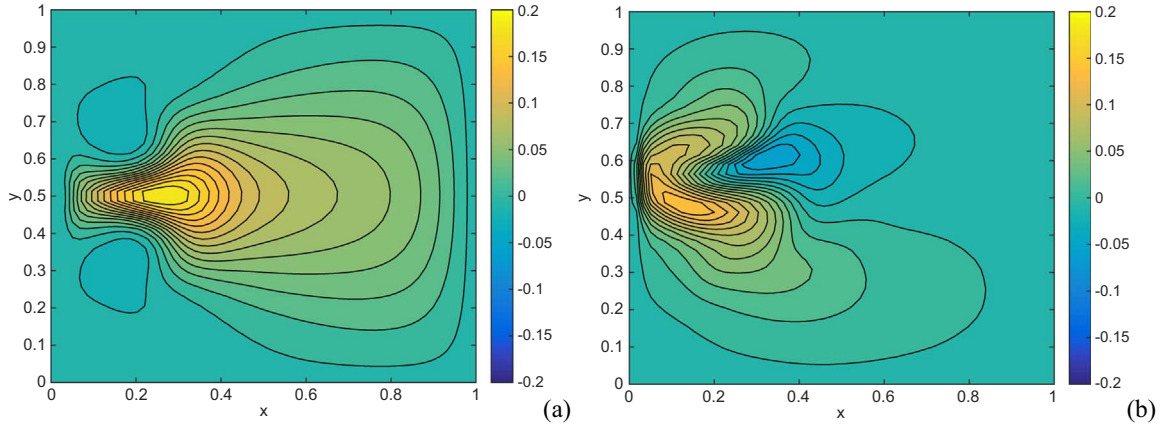



**Fig. 3.** CNOP with $\delta = 0.1$ and $t_e = 6.6$ days for the steady states as in Fig. 2. The colors indicate the dimensionless stream function values at the grid points and $x, y$ are the dimensionless spatial coordinates. (a) $Re = 25$. (b) $Re = 50$.

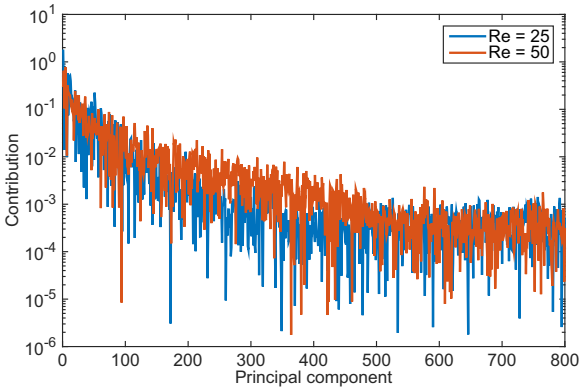



**Fig. 5.** Contribution $\alpha_i$ of the first 800 principle components to the CNOP calculated by SQP, when projected onto the EOFs for $Re$ = 25 (blue) and $Re$ = 50 (red), respectively.

derived from these data and the contribution of the principal components to the CNOP are shown in Fig. 5. The contribution $\alpha_i$ of EOF $\boldsymbol{a}_i$ is calculated as $\alpha_i = \boldsymbol{w}_0^{\delta} \cdot \boldsymbol{a}_i$, where $\boldsymbol{w}_0^{\delta}$ indicates the CNOP. Both for $Re$ = 25 and for $Re$ = 50 we observe a steep decrease of the $\alpha_i$ for the first 50 principal components. The values of $\alpha_i$ for $i$ 200 and 400 is larger for $Re$ = 50 than for $Re$ = 25. From 500 principal components onward the values of $\alpha_i$ seem comparable again for both values of $Re$. However, we are not able to draw strong conclusions about the differences in the projections for both Reynolds numbers. It should be noted that the variance of the training sets is small. Only as few as 10 PCs are needed to account for more than 90% of the total variance in the data sets (for both $Re$ = 25 and $Re$ = 50). The number of PCs actually used in the application of COBYLA will be indicated below by $P$.

### 4.3. Performance of COBYLA

The COBYLA routine has been implemented using $\rho_{start} = 0.1$ and $\rho_{end} = 10^{-4}$. All results are obtained by applying the optimisation method 10 times to the problem and by reporting the average and best solution. For one specific case, $Re$ = 25 and $P$ = 100, the optimisiation is repeated 100 times. The results are not significantly different than those for 10 times and therefore using 10 simulations is considered sufficient to address the accuracy of the approximation of the CNOP. Apart from the objective value and number of objective function calls also an 'error norm', which is a measure for the error in the CNOP solution obtained with respect to the SQP solution for that problem, is determined. The error $E$ is defined as

$$\mathrm{E} = \frac{\| \boldsymbol{w}_C - \boldsymbol{w}_{SQP} \|_2^2}{\| \boldsymbol{w}_{SQP} \|_2^2},$$

where $\boldsymbol{w}_C$ is the CNOP solution found by COBYLA and $\boldsymbol{w}_{SQP}$ is the CNOP solution as found by SQP. Actually, the CNOPs come in pairs $\{\boldsymbol{w}_0^{\delta}, -\boldsymbol{w}_0^{\delta}\}$ which both have the same objective function value. As COBYLA could find either of these solutions, the error $E$ is calculated with respect to the closest of one of these solutions.

The error $E$ is included because in this case our interest does not focus on the best objective value (in which case the objective value would be sufficient to measure the accuracy of the methods) but in the CNOP (initial perturbation) that yields this objective value. This means that if the objective value is nearly as good as the global maximum value, but the value is found at a local maximum far from the global maximum, the solution will not resemble the CNOP at all. Indeed, the performance of the method can be low even though the objective value is nearly perfect. The error $E$ determines how close the solution is to the CNOP obtained by the SQP method. The graphical results that are plotted below are also included as Tables in Appendix A.

The results for $Re$ = 25 are shown in Fig. 7 using different values of $P$. The minimum number of principal components needed to find a solution at the global optimum is about $P$ = 80. However, using $P$ = 100 leads to better results, and 150 principal components or more is almost a guarantee for finding the global optimum. Of course using more principal components also leads to a much more objective function calls, as more principal components lead to more dimensions in the search space. A typical example of a CNOP stream function pattern for $P$ = 100 is shown in Fig. 6a, showing quite a good agreement with the one computed using the SQP method (cf. Fig. 3a).

It is interesting that the solution for $P$ = 30 is significantly closer to the correct solution than the solution at $P$ = 50 which is probably due to a poor convergence of the EOF decomposition of the CNOP. From Fig. 7c, it is observed that for $P$ = 100 the confidence interval is relatively large as about half of the simulations converge to a local optimum instead of to the CNOP. It should also be noted that, even though not very well visible in Fig. 7c, a slight increase in the width of the confidence interval occurs for $P$ = 250 (see Table 1 in Appendix A).

The results for $Re$ = 50 are shown in Fig. 8 using $P$ = 50, 100 and 200. From Fig. 8 a and b we again observe an increase in objective function value and number of objective function calls as the number of principal components included increases. A typical example of a CNOP stream function pattern for $P$ = 200 is shown in Fig. 6b, showing quite a good agreement with the one computed using the SQP method (cf. Fig. 3b).

However, the error $E$ decreases only slightly compared the results for $Re$ = 25. This suggests that more principal components are needed to resolve the CNOP at $Re$ = 50, which is in accordance with the results in Fig. 5. However, when inspecting the solutions for $P$ = 200, the pattern of the CNOP is overall similar to the one found by SQP. This suggests that even though the error $E$ is quite large, and there are clear differences between the CNOP solutions of COBYLA and SQP, the solutions found by COBYLA might be useful if we are only interested in the global pattern of the CNOP.

We already indicated that up to $P$ = 150, some solutions approached a local optimum and not the global optimum. The increase in principal components leads to more solutions with the global optimum.

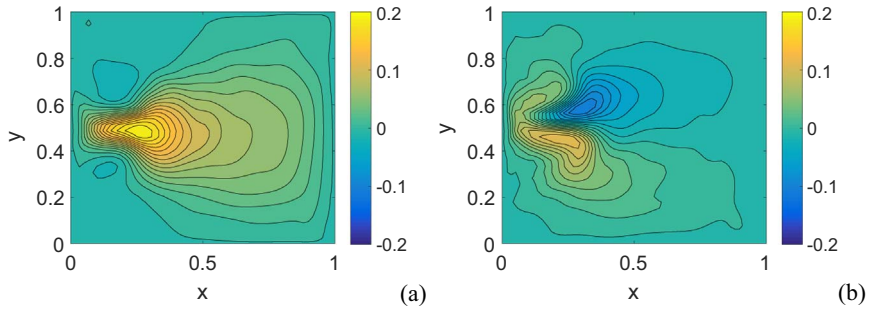
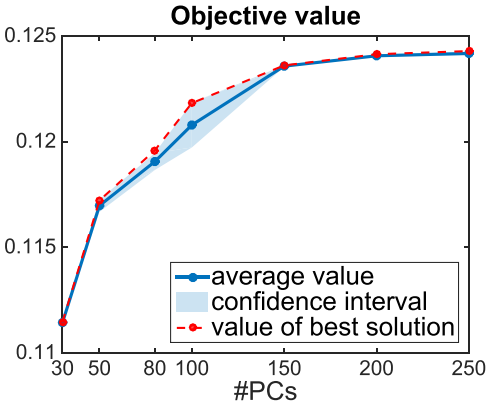
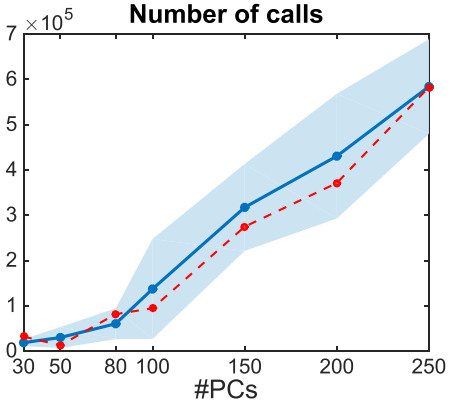


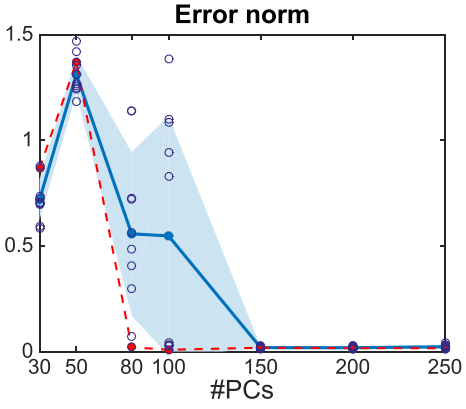

**Fig. 6.** (a) Spatial pattern of the streamfunction for the CNOP determined with COBYLA and $P$ = 100 for $Re$ = 25; for this solution, the value of the error $E$ = 0.0177. (b) Same as (a) but now for $P$ = 200 for $Re$ = 50; for this solution, the value of the error $E$ = 0.3308.




(a) Objective value

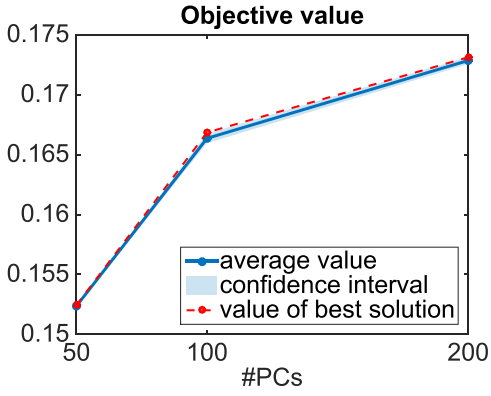



(b) Number of objective function calls made

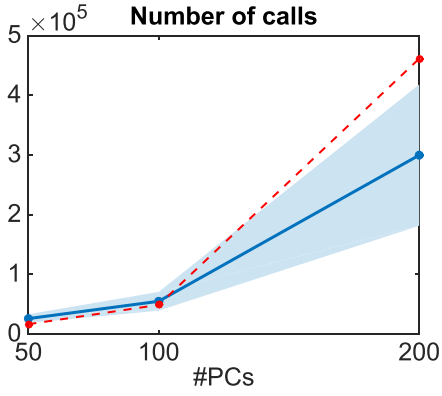



(c) Error norm with respect to SQP solution

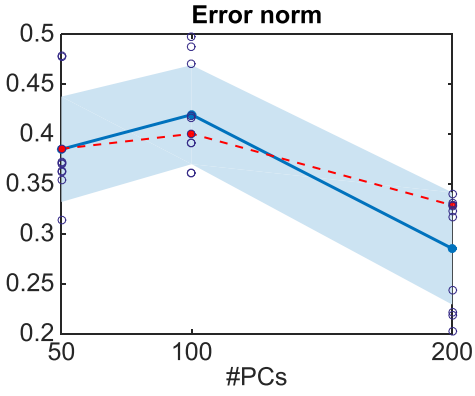
**Fig. 7.** Best and average of 10 results for COBYLA with a confidence interval of one standard deviation, using $\rho_{start} = 0.1$, $\rho_{end} = 10^{-4}$, $Re = 25$ and varying the number of PCs. The dots in (c) denote the individual runs.




(a) Objective value




(b) Number of objective funcion calls made



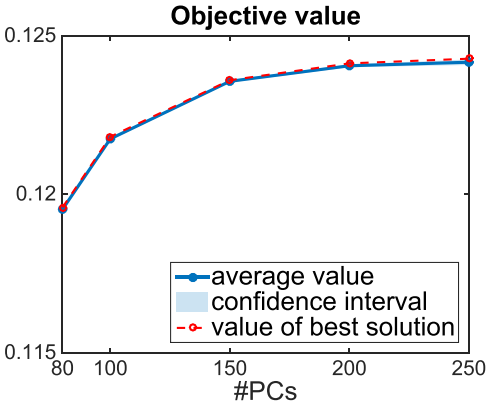

(c) Error norm with respect to SQP solution

**Fig. 8.** Best and average of 10 results for COBYLA with a confidence interval of one standard deviation, using $\rho_{start} = 0.1$, $\rho_{end} = 10^{-4}$, $Re = 50$ and varying the number of PCs. The dots in (c) denote the individual simulations.

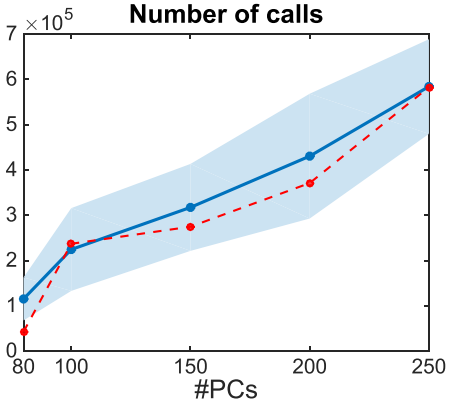
For $Re = 25$ the convergence with $P$ of only those solutions that are in the global optimum have been investigated. These solutions are selected by a subjective measure, i.e., inspecting the spatial pattern of the CNOP. For these solutions (Fig. 9) the objective function is almost converged at $P = 250$ and the number of objective function calls appears to increase linearly with $P$. The error norm of the best solution is decreasing with the number of principal components. The average value is increasing from $P = 250$, as is its confidence interval. Only including the solutions that converged to the global optimum, we find that the error norm converges to a value around 0.02 instead of zero. Because the error norm increases slightly from $P = 250$, more principle components will not necessarily lead to a more accurate solution.

### 4.4. Restarted COBYLA

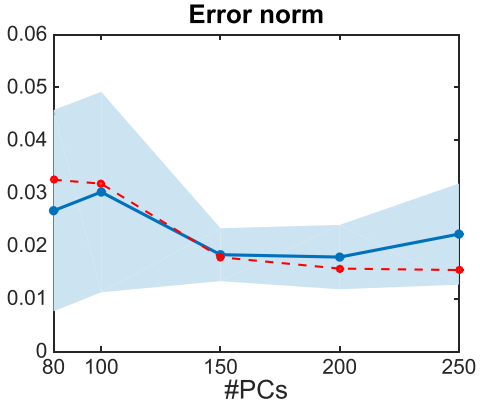
We have seen that, in general, a large dimensionality of the search space leads to more accurate solutions, while small dimensionality leads to fast convergence. This suggests that it might be efficient to first solve a low dimensional version of the problem, and restarting COBYLA with a higher dimension using the low dimensional solution as a starting point. This has been done for a value $P = 30$, using this solution as initial point for $P = 100$ calculations, and for initial dimensionality $P = 50$, using the




(a) Objective value

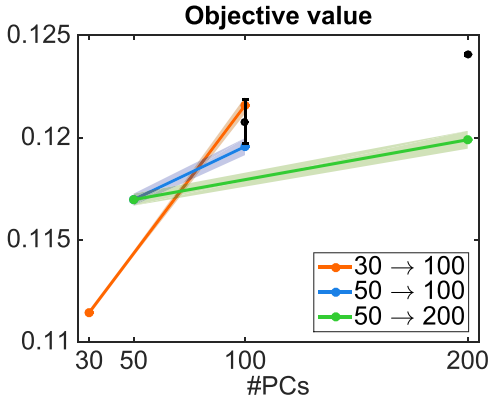



(b) Number of objective function calls made

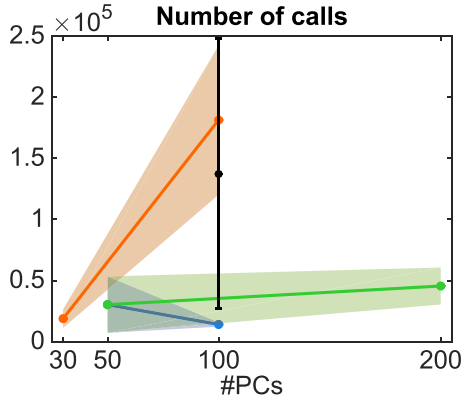



(c) Error norm with respect to SQP solution

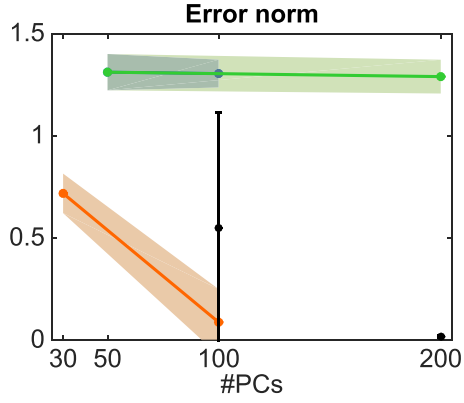
**Fig. 9.** Best and average of only the solutions that are determined with COBYLA to be in the global optimum, with confidence interval of one standard deviation, using $\rho_{start} = 0.1$, $\rho_{end} = 10^{-4}$, $Re = 25$ and varying the number of PCs.




(a) Objective value




(b) Number of objective function calls made




(c) Error norm with respect to SQP solution

**Fig. 10.** Best and average of 10 solutions with confidence interval of one standard deviation, using COBYLA with $\rho_{start} = 10^{-2}$, $\rho_{end} = 10^{-4}$, $Re = 25$ and using $P = 100$, with as initial state the solution for $P = 30$ or $P = 50$, and for $P = 200$ with as initial state the solution with $P = 50$. The black error bars are the solutions of COBYLA using $P = 100$ and $P = 200$ as in Fig. 7.

solution as initial point for $P = 100$ and $P = 200$ computations. For all implementations, a small value of $\rho_{start}$ of 0.01 is used, as it is to be expected that the solution is closer to the correct one than a random solution would be. The results are seen in Fig. 10.

It is interesting that using a solution obtained in 30 dimensions as initial point for the restart leads to good results, while using a solution obtained in 50 dimensions does not. This is, however, in agreement with the results in Fig. 7, where we observed that the solution with $P = 30$ resembles the correct solution better than the solution with $P = 50$. This means that even though the objective function value of this solution is worse, it is still a better restart point for the larger dimensional optimisation. We also compare the COBYLA result starting with $P = 30$ and restarting with $P = 100$ to the general COBYLA result with $P = 100$. It can be observed that the number of objective function calls has increased, the objective function value increased and the error norm decreased, all within the confidence interval of the original $P = 100$ solution. The number of objective function calls in restarted COBYLA is less than that of the general COBYLA version (cf. compare with Fig. 9b).

## 5. Summary and discussion

A new adjoint-free optimisation method for calculating CNOPs, combining COBYLA and Principal Component Analysis (PCA) dimension reduction, was presented. The performance of this method has been tested using a quasi-geostrophic (QG) ocean model using a comparison with the results obtained with the adjoint-based SQP method. We have seen that COBYLA is robust for 150 principal components or more at $Re = 25$. There is no convergence yet with 200 PCs for $Re = 50$. This shows that for $Re = 25$ and $Re = 50$ the objective landscape and/or the effect of dimension reduction is significantly different. It is clear that the number of PCs does increase with $Re$, as the flow shows more variability, but it is out of the scope of this paper to determine the precise dependence of the number of PCs on $Re$, if at all possible.

When comparing COBYLA to SQP in terms of performance, we see that for robust results COBYLA needs approximately 30,000 objective function calls, while SQP needs 200–600 calls to obtain the CNOPs to a much higher accuracy. So SQP outperforms COBYLA in terms of both accuracy and speed. However COBYLA is able to find the CNOPs without the use of the adjoint model. Restarted COBYLA is an improvement compared to general COBYLA when a good initial solution is used for the restart, for example by using a COBYLA solution determined within a lower dimensional search space.

Comparing our results to Mu et al. (2015), we conclude that even though better values of the objective function are found with COBYLA compared to SQP, the amount of dimensions required is much larger for the QG model than for the ZC model: approximately 200 dimensions for a 2400 dimensional search space for the QG model, and only 30 dimensions for a 1080 dimensional search space for the ZC model. This results in a relatively slow algorithm compared to the PCA based PSO of Mu et al. (2015). While their algorithm is only three times as slow as SQP, PCA based COBYLA is approximately 100 times as slow as SQP. The large difference in amount of principal components required to find the CNOPs in both models is remarkable. It is likely due to the fact that the ZC model is quasi linear (where only a nonlinearity in the temperature equation appears and the equatorial wave dynamics is linear) while there is a strong advective nonlinearity in the QG model.

Many other adjoint-free optimisation methods could be implemented on comparable models, to see whether they are suitable to resolve the CNOPs. Many of these models, however, suffer from the 'curse of dimensionality', which means it is hard to find a method that is known a priori to converge also on more complex high-dimensional models, such as general circulation models. This shows that the dimension reduction is essential to solve these kind of problems. We have discussed PCA as a method of dimension reduction but other methods of dimension reduction could also be considered. For the QG model, one could also project on eigenvectors instead of using PCs. The clear challenge is to find an efficient algorithm for dimension reduction that is able to resolve CNOPs of dynamical systems using COBYLA in low-dimensional spaces, even if the model dimensionality is very large.

## Acknowledgements

The third author is financially supported by the Netherlands Organisation for Scientific Research (NWO) as part of research programme 613.009.032.

## Appendix A. Tables of results

Table A.1 Table A.2 Table A.3

**Table A.1**
Best objective value, average objective value, average number of objective function calls and average error norm with respect to SQP solution for $\rho_{start} = 0.1$, $\rho_{end} = 10^{-4}$, $Re = 25$ versus the number of PCs.

| #PC | Best obj | Mean objective | | Mean #Calls | | Mean error norm | |
|---|---|---|---|---|---|---|---|
| 30 | 0.1115 | 0.11143 | ± 0.00002 | 19000 | ± 7600 | 0.7 | ± 0.1 |
| 50 | 0.1172 | 0.1170 | ± 0.0003 | 30000 | ± 23000 | 1.31 | ± 0.09 |
| 80 | 0.1196 | 0.1191 | ± 0.0004 | 61000 | ± 34000 | 0.6 | ± 0.4 |
| 100 | 0.1218 | 0.121 | ± 0.001 | 140000 | ± 110000 | 0.5 | ± 0.6 |
| 150 | 0.1236 | 0.12357 | ± 0.00003 | 320000 | ± 96000 | 0.018 | ± 0.005 |
| 200 | 0.1241 | 0.12406 | ± 0.00005 | 430000 | ± 140000 | 0.018 | ± 0.006 |
| 250 | 0.1243 | 0.12417 | ± 0.00009 | 580000 | ± 100000 | 0.02 | ± 0.01 |

**Table A.2**
Best objective function value, average objective function value, average number of objective function calls and average error norm with respect to SQP solution for $\rho_{start} = 0.1$, $\rho_{end} = 10^{-4}$, $Re = 50$ versus the number of PCs.

| #PC | Best obj | Mean objective | | Mean #Calls | | Mean error norm | |
|---|---|---|---|---|---|---|---|
| 50 | 0.1524 | 0.15236 | ± 0.00005 | 26000 | ± 7300 | 0.38 | ± 0.05 |
| 100 | 0.1668 | 0.1664 | ± 0.0004 | 55000 | ± 16000 | 0.42 | ± 0.05 |
| 200 | 0.1731 | 0.1729 | ± 0.0003 | 300000 | ± 120000 | 0.29 | ± 0.06 |

**Table A.3**
Best objective function value, average objective function value, average number of objective function calls and average error norm with respect to SQP solution for $\rho_{start} = 0.01$, $\rho_{end} = 10^{-4}$, $Re = 25$ versus the number of PCs: using the $P = 30$ solution as initial point in a $P = 100$ computation and using the$P = 50$ solution as initial point in $P = 100$ and $P = 200$ computations.

| PC-PC | Best obj | Mean objective | | Mean #Calls | | Mean error norm | |
|---|---|---|---|---|---|---|---|
| 30–100 | 0.1218 | 0.1216 | ± 0.0004 | 200000 | ± 60000 | 0.09 | ± 0.16 |
| 50–100 | 0.1199 | 0.1196 | ± 0.0004 | 10000 | ± 2000 | 1.3 | ± 0.1 |
| 50–200 | 0.1203 | 0.1199 | ± 0.0004 | 50000 | ± 20000 | 1.3 | ± 0.1 |

## Appendix B. Implementation

Here we will discuss which steps have to be taken to use COBYLA to calculate CNOPs in a dynamical system after dimension reduction. We assume that for a certain system a model integration function exists, and that we want to calculate a CNOP in a specific norm, with constraint boundary $\delta$ and integration time $t_e$, with respect to a certain background state $\boldsymbol{w}_b$.

### *B.1. PCA*

For principal component analysis, we need the following steps.

1. Obtain a training set by running the model for a long period, at least 100 times $t_e$. The training set should be a matrix with columns consisting of states: each column is a time-sample of the integration.
2. Centre the training set: subtract from each row the average of this row. This yields $\boldsymbol{W}$, the centered training set.
3. Calculate $\boldsymbol{W}\boldsymbol{W}^T$, and its eigenvalue decomposition. For example with NAG routine F01CKF (Mark 21)[2] and F02FCF (Mark 21).
4. Write the eigenvectors in a file, to be read by the optimisation program. Preferably the eigenvectors are written in direction of decreasing corresponding eigenvalue.

### *B.2. Working in reduced dimensionality*

The search space used by COBYLA is of a reduced dimensionality. However, the model integration is done in the original dimension of the model. Therefore we need a translation from reduced space to the original space.

1. Read the eigenvectors from the file. Read as many as the desired dimensionality in which the CNOPs are to be solved, say, the first $k$.
2. Write a function that projects a state $\boldsymbol{w}$ in the reduced dimensionality to a model state in the original dimension. The model state should be a sum $\boldsymbol{w}_m = \sum_{i=1}^{k} \boldsymbol{w}(i)\boldsymbol{a}_i$ where $\boldsymbol{w}(i)$ is the $i$th component of the reduced state, $\boldsymbol{a}_i$ is the $i$th eigenvector, $\boldsymbol{w}_m$ is the model state, and $k$ is the reduced dimensionality.

### *B.3. COBYLA*

To implement COBYLA and calculate the CNOPs, we need the following steps.

1. Download the COBYLA software (for Fortran) from http://mat.uc.pt/zhang/software.html#cobyla. Example optimisation problems are included in the code.
2. Write an implementation of the norm in which the CNOP is to be calculated, which is a function of the state (in the original dimensionality) and returns the norm.
3. Write a function CALCFC of variable $\boldsymbol{w}$, to be used by COBYLA, in which the objective and constraint are calculated and returned. The constraint value $c(\boldsymbol{w})$ is calculated as $c(\boldsymbol{w}) = \| \boldsymbol{w}_m \| - \delta$ where the implementation of the norm is used, $\delta$ is the chosen constraint boundary, and $\boldsymbol{w}_m$ is the model state obtained from the reduced state. The objective value $J(\boldsymbol{w})$ is calculated by integrating the state $\boldsymbol{w}_b + \boldsymbol{w}_m$ for time $t_e$, integrating the state $\boldsymbol{w}_b$ for time $t_e$, subtracting the integrated states and calculating the norm of the result. Here $\boldsymbol{w}_m$ is again the model state obtained from the reduced state. Note that if $\boldsymbol{w}_b$ is a steady state, of course integrating $\boldsymbol{w}$ separately is not necessary.
4. Provide COBYLA with CALCFC, a value of $\rho_{start}$ and $\rho_{end}$, an initial solution, the search space dimension and a maximum number of calls to CALCFC.

## References


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

Duan, W., Liu, X., Zhu, K., Mu, M., Exploring the initial errors that cause a significant spring predictability barrier for El Nino events, J. Geophys. Res. 114 (C4).

Hannachi, A., 2004. A Primer for EOF Analysis of Climate Data. Department of Meteorology University of Reading, 1–33.

Kalnay, E., 2003. Atmospheric Modeling, Data Assimilation and Predictability. Cambridge University Press.

Lagarias, J.C., Reeds, J.A., Wright, M.H., Wright, P.E., 1998. Convergence properties of the Nelder-Mead simplex method in low dimensions. SIAM J. Optim. 9 (1), 112–147.

Mu, B., Wen, S., Yuan, S., Li, H., 2015. PPSO: PCA based particle swarm optimization for solving conditional nonlinear optimal perturbation. Comput. Geosci. 83, 65–71.

Mu, M., 2000. Nonlinear singular vectors and nonlinear singular values. Sci. China Ser. D: Earth Sci. 43 (4), 375–385. http://dx.doi.org/10.1007/BF02959448, (ISSN 1006-9313 ⟨http://dx.doi.org/10.1007/BF02959448⟩).

Mu, M., Duan, W., Wang, B., 2003. Conditional nonlinear optimal perturbation and its applications. Nonlinear Process. Geophys. 10 (6), 493–501.

Nocedal, J., Wright, S., 2006. Numerical Optimization, & Business Media. Springer Science.

Osborne, A., Pastorello, A., 1993. Simultaneous occurrence of low-dimensional chaos and colored random noise in nonlinear physical systems. Phys. Lett. A 181 (2), 159–171.

Powell, M.J., 1994. A Direct Search Optimization Method That Models the Objective and Constraint Functions by Linear Interpolation. Advances in optimization and numerical analysis, Springer, 51–67.

Slingo, J., Palmer, T., 2011. Uncertainty in weather and climate prediction. Philos. Trans. R. Soc. A-Math. Phys. Eng. Sci. 369 (1956), 4751–4767.

Storn, R., Price, K., 1997. Differential evolution a simple and efficient heuristic for global optimization over continuous spaces. J. Glob. Optim. 11, 341–359.


[2] Information on the NAG library can be found at http://www.nag.co.uk/numeric/fl/manual/xhtml/mark21.xml

Sun, G.D., Mu, M., 2013. Understanding variations and seasonal characteristics of net primary production under two types of climate change scenarios in China using the LPJ model. Clim. Change 120, 755–769.
Sun, G.D., Mu, M., Zhang, Y.L., 2010. Algorithm studies on how to obtain a conditional nonlinear optimal perturbation (CNOP). Adv. Atmos. Sci. 27, 1311–1321.
Terwisscha van Scheltinga, A.D., Dijkstra, H.A., 2008. Conditional nonlinear optimal perturbations of the double-gyre ocean circulation. Nonlinear Process. Geophys. 15 (5), 727–734. http://dx.doi.org/10.5194/npg-15-727-2008, ((http://www.nonlin-processes-geophys.net/15/727/)).
Zebiak, S.E., Cane, M.A., 1987. A model El Nino-Southern oscillation. Mon. Weather Rev. 115, 2262–2278.
Zu, Z., Mu, M., Dijkstra, H.A., 2016. Optimal initial excitations of decadal modification of the atlantic meridional overturning circulation under the prescribed heat and freshwater flux boundary conditions. J. Phys. Oceanogr. 46 (7), 2029–2047.