11th International Symposium on Turbulence and Shear Flow Phenomena (TSFP11)
Southampton, UK, July 30 to August 2, 2019

A PRECONDITIONED MULTIPLE SHOOTING SHADOWING ALGORITHM FOR THE SENSITIVITY ANALYSIS OF CHAOTIC SYSTEMS

Karim Shawki
Department of Aeronautics
Imperial College
Exhibition Road, London
karim.shawki14@imperial.ac.uk

George Papadakis
Department of Aeronautics
Imperial College
Exhibition Road, London
g.papadakis@imperial.ac.uk

ABSTRACT

We propose a preconditioner that significantly accelerates the rate of convergence of the Multiple Shooting Shadowing (MSS) method (Blonigan and Wang, 2018) used to compute derivatives of time-averaged objectives to system parameter(s) for chaotic systems. Obtaining such derivatives (also known as sensitivities) accurately and quickly, is useful for example in optimal control and uncertainty quantification applications. MSS currently suffers from slow convergence. We propose a block diagonal preconditioner, which is based on a partial singular value decomposition of the MSS constraint matrix. Two chaotic systems are considered; the Lorenz system and the 1D Kuramoto Sivashinsky equation. The preconditioner is matrix-free and can be applied using existing time-steppers for the forward and adjoint equations.

INTRODUCTION

A dynamical system can be written as a set of ordinary differential equations, $^{d\mathbf{u}}/_{dt} = \mathbf{f}(\mathbf{u}, s)$, where $\mathbf{u}$ is a vector of state variables (length N), $\mathbf{f}$ is a vector of non-linear equations, and s is one (or more) system or control parameters. We define a time-averaged objective as

$$\bar{J} = \lim_{T\to\infty} \frac{1}{T} \int_0^T J(\mathbf{u}(t,s),s)\, dt \tag{1}$$

and seek the sensitivity $^{d\bar{J}}/_{ds}$. The sensitivity can be used to compute a search direction to iteratively find the optimal controls that minimise or maximise $\bar{J}$. For aeronautical engineering applications, $\bar{J}$ could be, for example, the averaged drag force on an airfoil, and s, a set of geometrical or actuation controls. Finite difference methods for computing $^{d\bar{J}}/_{ds}$ are usually not used, since the computational cost grows with the number of controls. The conventional method for computing $^{d\bar{J}}/_{ds}$ requires integration of the linearised tangent evolution equation

$$\frac{d\mathbf{v}}{dt} = \frac{\partial f}{\partial u}\mathbf{v} + \frac{\partial \mathbf{f}}{\partial s} \tag{2}$$

where $\mathbf{v}(t,s) = {}^{du(t,s)}/_{ds}$ and $^{\partial f}/_{\partial u}$ is a Jacobian matrix. The chain rule is applied to give the sensitivity

$$\frac{d\bar{J}}{ds} = \frac{1}{T}\int_0^T \left(\frac{\partial J(\mathbf{u}(t,s),s)}{\partial u}\mathbf{v}(t,s) + \frac{\partial J}{\partial s} \right) dt \tag{3}$$

The linearised solution $\mathbf{v}(t)$ for chaotic systems grows with $O(e^{\lambda_{max} t})$, where λ_{max} is the largest system Lyapunov exponent, rendering $^{d\bar{J}}/_{ds}$ meaningless for long T. 'Chaos' here refers to systems that are sensitive to the control s, meaning that small perturbations to s lead to exponentially diverging trajectories. For chaotic systems, (2) is 'ill-conditioned', and a number of methods have been proposed to counter this issue (Wang et al., 2014; Lasagna, 2017; Lea et al., 2002)

Least Squares Shadowing (LSS) (Wang et al., 2014) was proposed to compute bounded solutions $\mathbf{v}(t)$ for chaotic systems, and hence to compute accurate sensitivities. LSS requires the solution of a large linear matrix system, making it prohibitively expensive to be applied to high DOF systems. Its variant, Multiple Shooting Shadowing (MSS) (Blonigan and Wang, 2018) is a promising alternative, but suffers from slow convergence. In this paper, we propose a preconditioner to pave the way forward for MSS to be applied to high DOF systems. We show that convergence can be made almost independent of the number of DOF and trajectory lengths.

SHADOWING METHODS

Least Squares Shadowing (LSS) finds a 'shadow' trajectory satisfying $^{d\mathbf{u}'}/_{dt} = \mathbf{f}(\mathbf{u}', s+\delta s)$, that stays close in

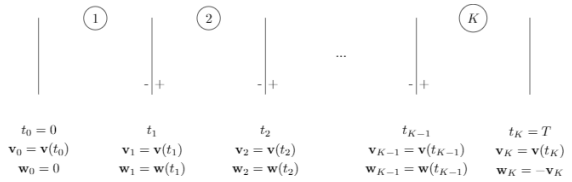


Figure 1: A sketch illustrating the segmenting approach of MSS. The constraint equations are propagated forward in time in all K segments, such that $\mathbf{v}(t_i^+) = \mathbf{v}(t_i^-)$ is satisfied for $i = 1, 2, ..., K-1$.

phase space to a reference trajectory $\mathbf{u}_{ref}$ (satisfying $d\mathbf{u}/dt = \mathbf{f}(\mathbf{u}, s)$) at all time instances. It achieves this by relaxing the initial condition for the shadowing trajectory. This regularises the problem leading to meaningful sensitivities $d\bar{J}/ds$ for chaotic systems that are sensitive to small changes in s. The existence of the shadowing trajectory is guaranteed for hyperbolic systems from the Shadowing Lemma (Pilyugin, 1999). The linear form of LSS can be written as follows

$$\begin{aligned} &\underset{\mathbf{v},\eta}{\text{Minimise}} \ \frac{1}{T}\int_0^T \|\mathbf{v}\|^2 + \alpha^2\eta^2\, dt, \\ &\text{subject to} \ \frac{d\mathbf{v}}{dt} = \frac{\partial f}{\partial u}\mathbf{v} + \frac{\partial \mathbf{f}}{\partial s} + \eta\mathbf{f}(u_{ref}, s) \end{aligned} \tag{4}$$

where $\mathbf{v}(t) = d(\mathbf{u}' - \mathbf{u}_{ref})/ds$, $\eta(t) = d(d\tau/dt - 1)/ds$ (time dilatation term) and α is a penalisation parameter. Solving (4) is too expensive for systems with a large number of degrees of freedom.

The recently proposed Multiple Shooting Shadowing (MSS) method (Blonigan and Wang, 2018) reduces costs by minimising the Euclidean norm of $\mathbf{v}(t)$ at K discreet checkpoints in time only (instead of at all time instances as in LSS). This is illustrated in Figure (1).

By using state transition theory, the MSS optimisation problem can be written as

$$\underset{\mathbf{v}_i}{\text{Minimise}} \ \frac{1}{2}\sum_{i=0}^{K} \|\mathbf{v}_i\|_2^2 \tag{5a}$$

$$\text{subject to} \ \mathbf{v}_{i+1} = \Phi_{i+1}\mathbf{v}_i + \mathbf{b}_{i+1} \tag{5b}$$

where Φ_i are $N \times N$ state transition matrices, and $\mathbf{b}_i$ are zero-state response term vectors. (5b) enforces continuity between segments and satisfies (4b) at any given time checkpoint i. (5) is equivalent to (4) when $\alpha = 0$ and as $K \to \infty$. For more details, refer to (Blonigan and Wang, 2018).

(5) is an under-determined problem that can be written in matrix form as

$$\underset{\mathbf{v}_i}{\text{Minimize}} \ \frac{1}{2}\sum_{i=0}^{K} \|\mathbf{v}_i\|_2^2 \tag{6a}$$

$$\text{subject to} \ A\underline{\mathbf{v}} = \underline{\mathbf{b}} \tag{6b}$$

where

$$A = \begin{bmatrix} -\Phi_1 & I & & & \\ & -\Phi_2 & I & & \\ & & \ddots & \ddots & \\ & & & -\Phi_K & I \end{bmatrix} \quad \underline{\mathbf{v}} = \begin{bmatrix} \mathbf{v}_0 \\ \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_K \end{bmatrix} \quad \underline{\mathbf{b}} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_K \end{bmatrix} \tag{7}$$

and A is a $NK \times N(K+1)$ matrix, $\underline{\mathbf{v}}$ and $\underline{\mathbf{b}}$ are vectors of length $N(K+1)$ and NK respectively, and $\mathbf{b}_i$ are the zero-state response terms. An optimality system can be derived by introducing a set of discrete adjoint variables $\underline{\mathbf{w}}$, and differentiating the resulting Lagrangian, giving

$$\begin{bmatrix} -I & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \underline{\mathbf{v}} \\ \underline{\mathbf{w}} \end{bmatrix} = \begin{bmatrix} 0 \\ \underline{\mathbf{b}} \end{bmatrix} \tag{8}$$

with a Schur complement

$$S\underline{\mathbf{w}} = \begin{bmatrix} \Phi_1\Phi_1^T + I & -\Phi_2^T & & & \\ -\Phi_2 & \Phi_2\Phi_2^T + I & -\Phi_3^T & & \\ & \ddots & \ddots & \ddots & \\ & & & -\Phi_K & \Phi_K\Phi_K^T + I \end{bmatrix} \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_K \end{bmatrix} = \underline{\mathbf{b}} \tag{9}$$

that can be solved for the discreet adjoint variables $\underline{\mathbf{w}}$. The solution $\underline{\mathbf{w}}$ can be used to find $\underline{\mathbf{v}}$ and then $d\bar{J}/ds$ by evaluating the following integral:

$$\begin{aligned} \frac{d\bar{J}}{ds} = \frac{1}{T}\sum_{i=0}^{K-1}\int_{t_i}^{t_{i+1}} \left\langle \left.\frac{\partial J}{\partial u}\right|_t, \mathbf{v}' \right\rangle dt + \\ \frac{1}{T}\sum_{i=0}^{K-1} \frac{\langle \mathbf{f}_{i+1}, \mathbf{v}'(t_{i+1})\rangle}{\|\mathbf{f}_{i+1}\|_2^2}(\bar{J} - J_{i+1}) + \frac{\partial \bar{J}}{\partial s} \end{aligned} \tag{10}$$

Matrix S has size $NK \times NK$. Equation (9) can be solved iteratively in a matrix-free fashion by supplying matrix-vector products $S\underline{\mathbf{w}}^{(j)}$ to a Krylov-type solver, such as Conjugate Gradient or GMRES. $\underline{\mathbf{w}}^{(j)}$ is a sequence of solver generated vectors. Products involving Φ_i require integration of the homogeneous form of (2), and those involving Φ_i^T require backwards in time integration of the homogeneous adjoint form of (2). There is an adjoint version of (9) for computing simultaneously the sensitivities to multiple control inputs. For more details, refer to (Blonigan and Wang, 2018).

A BLOCK DIAGONAL PRECONDITIONER FOR THE MSS SCHUR COMPLEMENT

The convergence rate of iterative Krylov subspace solvers for symmetric, positive definite matrices (like S) depends on the distribution of the matrix eigenvalues (Saad, 2003). For such systems, the eigenvalues are all positive and real. If all of them are tightly clustered around a few single points away from the origin, then one would expect fast convergence. On the other hand, widely spread eigenvalues without tight clustering can lead to slow convergence. The objective of a preconditioner is to reduce the spread of the eigenvalues, and thereby reduce the condition number, $\kappa(S) = \frac{\mu_{max}(S)}{\mu_{min}(S)}$, where $\mu_{max}(S)$ and $\mu_{min}(S)$ are the maximum and minimum eigenvalues of S, respectively.

An extensive survey of preconditioners for saddle point problems, such as the MSS KKT system (8), is available in (Benzi et al., 2005). Preconditioners can be applied to the 2×2 block system (8) or directly to the Schur complement system (9). In either case, an easily invertible approximation of S is required. There is an additional restriction, namely that the preconditioner should be matrix free, i.e. it should rely on matrix-vector products only (computing and

storing S is out of the question for long trajectories and large N).

We consider first deflating the eigenvalues greater than unity. There are typically $O(\#\text{ Positive Lyapunov exponents} \times K)$ eigenvalues greater than one. The singular value decomposition (SVD) of a given state transition matrix Φ_i reads

$$\Phi_i = U_i \Sigma_i V_i^T \tag{11}$$

where U_i, Σ_i and V_i are $N \times N$ matrices. The columns of U_i contain the left singular vectors of Φ_i, while the right singular vectors of Φ_i make up the columns of V_i. Σ_i is a diagonal matrix containing the singular values of Φ_i in descending order, which we denote by $\sigma(\Phi_i)$.

We can form an efficient preconditioner (see Shawki and Papadakis, 2018) by ignoring the off-diagonal identities of matrix A (7), and denote this matrix by $\tilde{A}$. We can then compute an approximation to $\tilde{S}^{-1} = (\tilde{A}\tilde{A}^T)^{-1}$

$$\mathbf{M_{BD}} \approx \tilde{S}^{-1} = \begin{bmatrix} (\Phi_1\Phi_1^T)^{-1} & & & \\ & (\Phi_2\Phi_2^T)^{-1} & & \\ & & \ddots & \\ & & & (\Phi_K\Phi_K^T)^{-1} \end{bmatrix} \tag{12}$$

using a partial singular value decomposition (SVD) of the state transition matrices Φ_i. A full SVD is not required, since Φ_i can be estimated with a small number of singular modes only (those with large σ) . This reduces the cost of computing the SVD. The resulting preconditioner takes the form

$$\mathbf{M_{BD}}_{(l)}^{(q)} = \text{diag}(\mathbf{M}_{(l),1}^{(q)}, \mathbf{M}_{(l),2}^{(q)}, ..., \mathbf{M}_{(l),K}^{(q)}) \tag{13a}$$

$$\mathbf{M}_{(l),i}^{(q)} = U_i \Sigma_i^{-2} U_i^T + (I - U_i U_i^T) \tag{13b}$$

where l is the number of retained singular modes in each segment, i is the segment number and q is the number of Lanczos bidiagonalisation iterations. The superscript (q) and subscript (l) have been removed from Σ_i and U_i for clarity. We call (13) the 'Block Diagonal Preconditioner' (BDP). The MATLAB function 'svds' was used to compute (13b). The decoupling between segments as a result of neglecting the identities of the matrix A allows for a much faster and more efficient parallel-in-time evaluation of the blocks $\mathbf{M}_{(l),i}^{(q)}$. We use $\mathbf{M_{BD}}_{(l)}^{(q)}$ as a left preconditioner for the original system, i.e. we solve

$$\mathbf{M_{BD}}_{(l)}^{(q)} S \underline{\mathbf{w}} = \mathbf{M_{BD}}_{(l)}^{(q)} \underline{\mathbf{b}} \tag{14}$$

The function of $\mathbf{M_{BD}}_{(l)}^{(q)}$ is to deflate the $K \times l$ largest eigenvalues of S and bring their values closer to unity. Regularising the very small μ_{min} is essential for solution accuracy and convergence speed for non strictly hyperbolic systems. Using Tikhonov regularisation, we get:

$$(\gamma I + \mathbf{M_{BD}}_{(l)}^{(q)} S)\underline{\mathbf{w}} = \mathbf{M_{BD}}_{(l)}^{(q)} \underline{\mathbf{b}} \tag{15}$$

where the regularisation parameter $\gamma > 0$. The above form brackets eigenvalues within the narrow range $[\gamma, \gamma +$

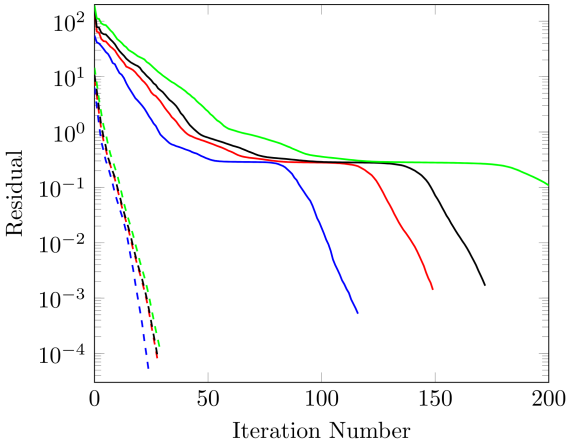


Figure 2: Residuals for the original system S (solid lines) and the regularised BDP system $\gamma I + \mathbf{M_{BD}}_{(1)} S$ (dashed lines) with $\gamma = 0.1$. The segment size is $\Delta T = 1$ and $\rho = 40$ (Lorenz system). Blue: $T = 200$, red: $T = 300$, black: $T = 500$, green: $T = 1000$.

$\mu_{max}(\mathbf{M_{BD}}_{(l)}^{(q)} S)]$. The presence of γ results in discontinuities in $\mathbf{v}(t)$. Choosing a small γ improves the conditioning without affecting much the accuracy of the computed sensitivity. A larger γ however can degrade the accuracy.

RESULTS

Two dynamical systems were considered; the Lorenz system and the Kuramoto Sivashinsky equation. The Lorenz system reads

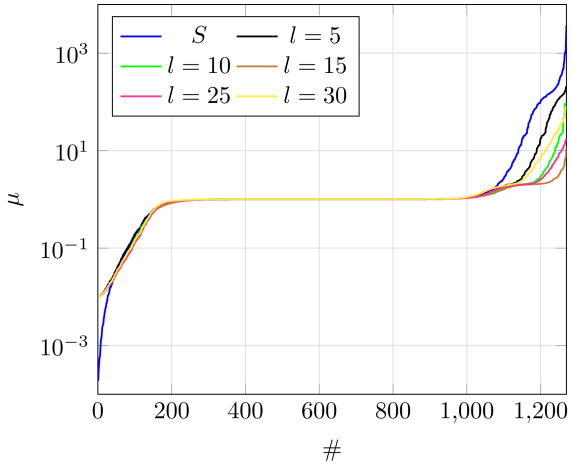
$$\frac{dx}{dt} = \sigma(y - x) \qquad \frac{dy}{dt} = x(\rho - z) - y \qquad \frac{dz}{dt} = xy - \beta z \tag{16}$$

where σ, ρ and β are system parameters. When (15) was solved using $l = 1$ (since there is one positive exponent for the Lorenz system), and $\gamma = 0.1$, the convergence was found to be almost independent on T (see Figure 2). While $\kappa(S)$ (and the number of iterations) increases with T, $\kappa(\gamma I + \mathbf{M_{BD}}_{(l)}^{(q)} S)$ stays almost constant with T, thus keeping convergence independent on T. The clustering effect of $\mu(S)$ is clearly essential for efficient and scalable use of MSS.

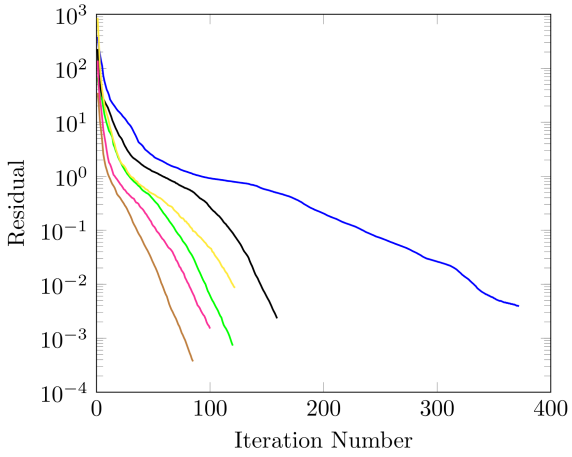
MSS was then applied to the Kuramoto Sivashinsky equation

$$\begin{aligned} &\frac{\partial u}{\partial t} = -(u + c)\frac{\partial u}{\partial x} - \frac{\partial^2 u}{\partial x^2} - \frac{\partial^4 u}{\partial x^4} \\ &x \in [0, L] \\ &u(0,t) = u(L,t) = 0 \\ &\left.\frac{\partial u}{\partial x}\right|_{x=0} = \left.\frac{\partial u}{\partial x}\right|_{x=L} = 0 \end{aligned} \tag{17}$$

where $L = 128$ and $N \geq 127$ (# grid points). We first study the combined effect of the number of preconditioner retained singular modes, l and of γ on the eigenvalue spectrum the convergence rate. For this, we solve (15) for different l and using $\gamma = 0.01$. We see from Figure (3) that increasing l up to $l = 15$ (number of positive Lyapunov exponents) reduces μ_{max} by up to at least two orders of magnitude and

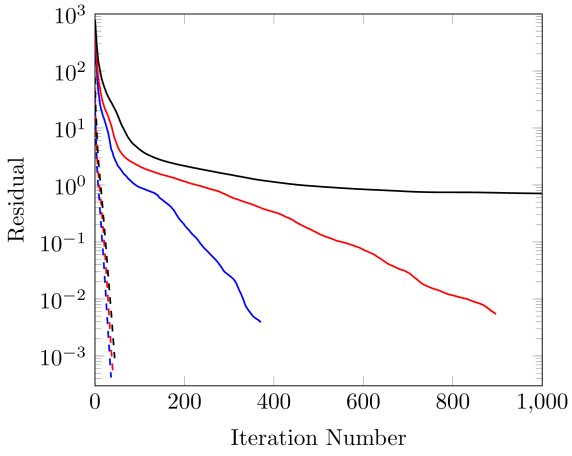


(a) Eigenvalues

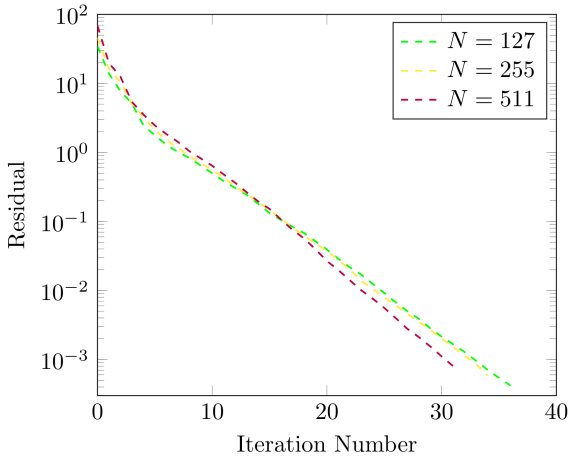


(b) Residuals

Figure 3: Eigenvalues and residuals of the original system S (blue line), and of the BDP system $(\gamma I + \mathbf{M_{BD}}^{(2)}_{(l)} S)$ for $N = 127$, $T = 100$ and $c = 0.8$. The preconditioners were constructed for different l, and a regularisation value $\gamma = 0.01$ was used.



(a) Varying T with $N = 127$



(b) Varying N with $T = 100$

Figure 4: Residuals for the original system S (solid lines) and for $\gamma I + \mathbf{M_{BD}}^{(q)}_{(l)} S$ with $\gamma = 0.09$, $q = 2$ and $l = 15$ (dashed lines). The segment size is $\Delta T = 10$ for all cases. In the top figure (a), Blue: $T = 100$, red: $T = 200$, black: $T = 500$

leads to faster convergence rates (panel b). Interestingly, a further increase to $l = 25$ or $l = 30$ increases the condition number and slows down the convergence (Figure 3b). This indicates that after a certain value of l, adding more singular modes to the individual preconditioner blocks $\mathbf{M}^{(q)}_{(l),i}$ starts to provide unreliable information to the approximation $(\Phi_1 \Phi_1^T)^{-1}$, and hence to the preconditioner $\mathbf{M_{BD}}^{(q)}_{(l)}$. This value can be closely linked to the number of positive exponents.

When equation (15) was solved using $l = 15$, convergence was found to be almost independent on both T and N (see Figure 4). For $T = 500$ (Figure 4a), the total cost (in terms of the number of constraint and adjoint integrations) of constructing $\mathbf{M_{BD}}^{(q)}_{(l)}$ and solving (15) was reduced by a factor of 35 (compared to the cost of solving without preconditioning).

CONCLUSIONS

We proposed a block diagonal preconditioner to accelerate the convergence rate for the solution of the linear system arising from the application of the Multiple Shooting Shadowing algorithm. The preconditioner is based on the partial singular value decomposition of the diagonal blocks of the Schur complement. It was applied to the Lorenz system and the Kuramoto Sivashinsky equation.

When the preconditioner was combined with a regularisation method, the condition number was significantly suppressed, and the convergence rate was found to be weakly dependent on the number of degrees of freedom and the length of the trajectory. The total number of operations was significantly reduced as a result. The results show that a faster implementation of MSS for the sensitivity analysis of higher DOF systems is possible.

The number of singular modes to retain in the partial SVD, l, is case dependent. A well-chosen value is required for fast convergence. If the number of positive Lyapunov exponents is known, it can be used to inform the choice of l. Strictly speaking however, this is not necessary. A self-adaptive algorithm is currently being investigated to estimate l that requires no prior knowledge of the number of positive Lyapunov exponents.

Apart from the value of l mentioned above, the question on how to choose an appropriate value of the regularisation parameter γ that can provide an adequate balance

between solution accuracy and rate of convergence is still open. The application of MSS for the control of the KS equation is currently being investigated.

REFERENCES

Benzi, Michele, and Golub, G.H., and Liesen, Jrg, 2005, Numerical solution of saddle point problems, Acta Numerica, Vol. 14, pp. 1-137.

Blonigan, Patrick, and Wang, Qiqi, 2018, Multiple shooting shadowing for sensitivity analysis of chaotic dynamical systems, Journal of Computational Physics, Vol. 354, pp. 447-475.

Lasagna, Davide, Sensitivity Analysis of Chaotic Systems using Unstable Periodic Orbits, SIAM Journal on Applied Dynamical Systems, Vol. 17, pp. 547-580.

Lea, D.J., and Allen M.R., and Haine, T.W.N, Sensitivity analysis of the climate of a chaotic system, Quarterly Journal of the Royal Meteorological Society, Vol. 128, pp. 25872605.

Pilyugin, S.Y., *Shadowing in Dynamical Systems*, 1999, Springer-Verlag Berlin Heidelberg.

Saad, Yousef, *Iterative Methods for Sparse Linear Systems (Book)*, 2003, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2nd edition.

Shawki, Karim, and Papadakis, George, 2018, *A Preconditioned Multiple Shooting Shadowing Algorithm for the Sensitivity Analysis of Chaotic Systems*, https://arxiv.org/abs/1810.12222.

Wang, Qiqi, and Hu, Rui, and Blonigan, Patrick, 2014, Least Squares Shadowing sensitivity analysis of chaotic limit cycle oscillations, Journal of Computational Physics, Vol. 267, pp. 210-224.