

Multigrid-in-time for sensitivity analysis of chaotic dynamical systems

Patrick Blonigan, MIT and Qiqi Wang, MIT

December 10, 2013

Abstract

The following paper discusses the application of a multigrid-in-time scheme to Least Squares Shadowing (LSS), a novel sensitivity analysis method for chaotic dynamical systems. While traditional sensitivity analysis methods break down for chaotic dynamical systems, LSS is able to compute accurate gradients. Multigrid is used because LSS requires solving a very large Karush-Kuhn-Tucker (KKT) system constructed from the solution of the dynamical system over the entire time interval of interest.

Several different multigrid-in-time schemes are examined, and a number of factors were found to heavily influence the convergence rate of multigrid-in-time for LSS. These include the iterative method used for the smoother, how the coarse grid system is formed and how the least squares objective function at the center of LSS is weighted.

1 Introduction

Although computers are becoming more powerful at a rapid pace, the size of the problems we wish to solve are growing even faster. An example of a potentially very large problem is Least Squares Shadowing (LSS), a novel computational method that can be used for sensitivity analysis of chaotic dynamical systems.

LSS requires solving a linear system of $\mathcal{O}(mn)$ equations, where m is the number of time steps and n is the number of dimensions or degrees of freedom of the system, both of which can be very large ($\sim 10^5$ or greater) for many problems of interest in a number of fields including fluid dynamics. Fortunately this system can be written as a 2nd order boundary value problem in time with homogeneous Dirichlet boundary conditions. Therefore a multigrid-in-time scheme is attractive because of its fast convergence relative to other iterative methods for many boundary value problems in time [1].

arXiv:1305.6878v2 [math.NA] 8 Dec 2013

1.1 Motivation for Least Squares Shadowing

Sensitivity analysis of systems governed by ordinary differential equations (ODEs) and partial differential equations (PDEs) is important in many fields of science and engineering. Its goal is to compute sensitivity derivatives of key quantities of interest to parameters that influence the system. Applications of sensitivity analysis in science and engineering include design optimization, inverse problems, data assimilation, and uncertainty quantification.

Adjoint based sensitivity analysis is especially powerful in many applications, due to its efficiency when the number of parameters is large. In aircraft design, for example, the number of geometric parameters that define the aerodynamic shape is very large. As a result, the adjoint method of sensitivity analysis has proven to be very successful for aircraft design [2], [3]. Similarly, the adjoint method has been an essential tool for adaptive grid methods for solving PDEs [4], solving inverse problems in seismology, and for assimilating observation data for weather forecasting.

Sensitivity analysis for chaotic dynamical systems is important because of the prevalence of chaos in many scientific and engineering fields. One example is highly turbulent gas flow of mixing and combustion processes in jet engines. In this example, and in other applications with periodic or chaotic characteristics, statistical averaged quantities such as mean temperature and mean aerodynamic forces are of interest. Therefore, the general problem we seek to solve with sensitivity analysis is:

$$\text{Given } \frac{du}{dt} = f(u, \xi), \quad \bar{J} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T J(u, \xi) dt, \quad \text{Compute } \frac{\partial \bar{J}}{\partial \xi} \quad (1)$$

Where u is the state vector, ξ is some parameter in the governing equation, $\frac{du}{dt} = f(u, \xi)$ and $J(u, \xi)$ is some quantity of interest.

Sensitivity analysis for chaotic dynamical systems is difficult because of the high sensitivity of these systems to the initial condition, known as the "Butterfly Effect". Slightly different initial conditions will result in very different solutions, which diverge exponentially with time [5]. This also results in exponential growth of sensitivities and therefore the sensitivity of long-time averaged quantities is not equal to the long-time average sensitivities of chaotic systems [6]. Because the derivative and long-time average do not commute, the traditional adjoint method computes sensitivities that grow exponentially with time, as shown in the work done by Lea et al. [6].

The least squares shadowing (LSS) method does not encounter the exponential growth of sensitivities observed in traditional methods [7]. The method requires that the chaotic system is *ergodic*; that long time behavior of the system is independent of initial conditions. LSS finds a perturbed trajectory that does not diverge exponentially from some trajectory in phase space. This non-diverging trajectory, called a "shadow trajectory", has its existence guaranteed by the shadowing lemma [8] for a large number of chaotic systems and can be used to compute sensitivities. The shadow trajectory is found by solving a quadratic programming (QP) problem with linear constraints. It is the large size and bandwidth of the KKT system associated with the QP system that requires an efficient, iterative linear solver.

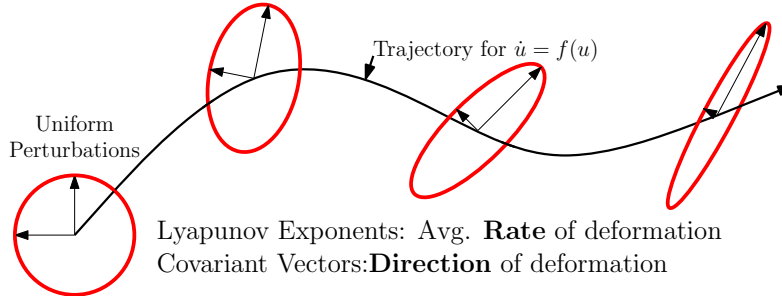


Figure 1: Schematic of Lyapunov exponents and covariant vectors

1.2 Outline

The following paper discusses the implementation of LSS using a multigrid-in-time scheme to solve the KKT system derived from the QP problem. Several schemes were considered, each resulting in different convergence rates and computational costs. The paper is structured as follows: first, LSS is presented through a discussion of non-linear dynamics and demonstrated on the Lorenz system. Next, several multigrid-in-time schemes are presented and their performance solving for sensitivities of the Lorenz system is presented and discussed. Finally, some concluding remarks and directions for future work are presented.

2 Least Squares Shadowing Method

2.1 Lyapunov Exponents and the Shadowing Lemma

Before LSS and the shadowing lemma from which it is derived is discussed, a more in-depth discussion of Lyapunov exponents and Lyapunov covariant vectors is required. For some system $\frac{du}{dt} = f(u)$, there exist Lyapunov covariant vectors $\phi_1(u), \phi_2(u), \dots, \phi_i(u)$ corresponding to each Lyapunov exponent Λ_i , which satisfy the equation [9]:

$$\frac{d}{dt}\phi_i(u(t)) = \frac{\partial f}{\partial u} \cdot \phi_i(u(t)) - \Lambda_i\phi_i(u(t))$$

To understand what Λ_i and ϕ_i represent, consider a sphere comprised of perturbations δf to a system $\frac{du}{dt} = f(u)$ at some time, as shown in the far left of figure 1. As this system evolves in time, this sphere expands in some directions, contracts in some, and remains unchanged in others. The rate at which the sphere expands or contracts corresponds to the Lyapunov exponent Λ_i and the corresponding direction of expansion or contraction is the Lyapunov covariant vector ϕ_i . It is important to note that the vectors ϕ_i are not the same as local Jacobian eigenvectors along a trajectory. The ϕ_i vectors depend on all Jacobian eigenvectors along a trajectory. Also, the ϕ_i are not necessarily orthogonal, but the number of Lyapunov covariant vectors is the same as the number of dimensions of the system.

A strange attractor, the type of attractor associated with chaotic dynamical systems, has at least one positive and one zero Lyapunov exponent [9]. To illustrate the effect of the

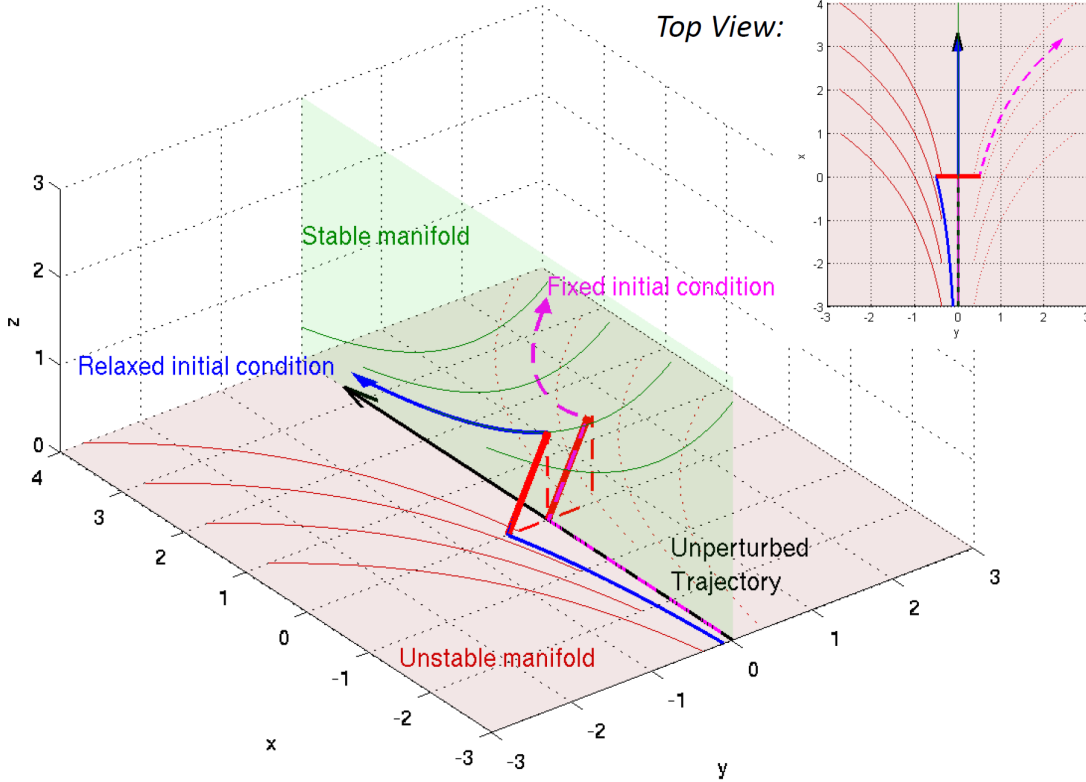


Figure 2: Phase space trajectory of a chaotic dynamical system. The unstable manifold, in red, is the space of all Lyapunov covariant vectors corresponding to positive exponents. The stable manifold, in green, corresponds to the space of all covariant vectors associated with negative exponents. A perturbation to the system (in red) has components in both manifolds, and the unstable component causes the perturbed trajectory (pink) to diverge exponentially from the unperturbed trajectory (in black). LSS chooses a perturbed trajectory with a different initial condition (in blue) that does not diverge from the unperturbed trajectory.

positive exponent, we consider figure 2. We see that if the perturbed trajectory has the same initial condition as the unperturbed trajectory, the two trajectories diverge exponentially, leading to the issues with traditional sensitivity analysis discussed in the introduction.

However, the assumption of ergodicity means that it is not necessary to compare a perturbed and an unperturbed trajectory with the same initial condition. Therefore, an initial condition can be chosen such that the perturbed and unperturbed trajectories do not diverge, resulting in the blue trajectory in figure 2. The existence of this trajectory, called a “shadow trajectory”, follows from the shadowing lemma [8] : *Consider a reference solution u_r to*

$$\frac{du}{dt} = f(u, \xi)$$

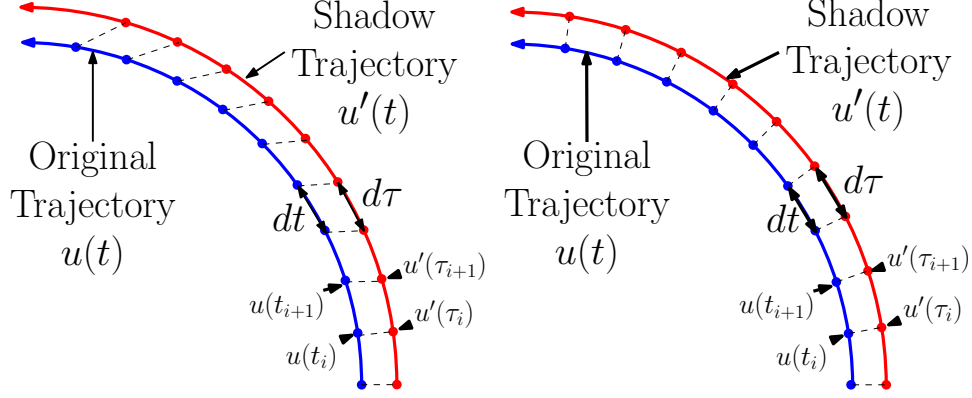


Figure 3: LEFT: Original and shadow phase space trajectories without any time transformation ($d\tau/dt = 1$). RIGHT: Original and shadow phase space trajectories with a time transformation $d\tau/dt = 1 + \eta$ that minimizes the distance between the two trajectories in phase space for all time. Note that η is a “time stretching factor” and that $-1.0 < \eta < 1.0$.

If this system has a Hyperbolic strange attractor and if some system parameter ξ is slightly perturbed:

For any $\delta > 0$ there exists $\varepsilon > 0$, such that for every u_r that satisfies $\|du_r/dt - f(u_r)\| < \varepsilon$, $0 \leq t \leq T$, there exists a true solution u_s and a time transformation $\tau(t)$, such that $\|u_s(\tau(t)) - u_r(t)\| < \delta$, $|1 - d\tau/dt| < \delta$ and $du_s/d\tau - f(u_s) = 0$, $0 \leq \tau \leq \mathcal{T}$.

Note that $\|\cdot\|$ refers to distance in phase space

Therefore, relaxing the initial condition allows us to find the shadow trajectory $u_s(\tau)$. The key assumption of the shadowing lemma is that the attractor associated with the system of interest is *hyperbolic*. The key property of hyperbolic attractors for the shadowing lemma is that tangent space can be decomposed into stable, neutrally stable and unstable components everywhere on the attractor [10]. Another way to state this property is that the Lyapunov covariant vectors make up a basis for phase space at all points on the attractor. Although this not the case for many attractors, including the Lorenz attractor, there is a *Chaotic Hypothesis* which states that many high-dimensional chaotic systems will behave as if they were hyperbolic [11]. For example, since the single point on the Lorenz attractor that is not hyperbolic is the unstable fixed point at the origin, most phase space trajectories do not pass through it and the shadowing lemma holds.

The time transformation alluded to in the shadowing lemma is required to deal with the zero (neutrally stable) Lyapunov exponent, whose covariant vector is simply $f(u)$. This time transformation, referred to as “time dilation” in this paper and other LSS literature, is required to keep a phase space trajectory and its shadow trajectory close (in phase space) for all time as demonstrated in figure 3.

2.2 Computing the Shadow Trajectory

Although LSS could be implemented by computing Lyapunov exponents and covariant vectors (as in [7]), this is not necessary. To find the shadow trajectory, an optimization problem is solved, where the objective function is the L2 norm of the tangent solution. That is, for some system of equations $\frac{du}{dt} = f(u, \xi)$, the tangent equations are solved, where $v = \frac{\partial u}{\partial \xi}$:

$$\min_{v, \eta} \frac{1}{2} \int_0^T v^2 + \alpha^2 \eta^2 dt, \quad s.t. \quad \frac{dv}{dt} = \frac{\partial f}{\partial u} v + \frac{\partial f}{\partial \xi} + \eta f, \quad 0 < t < T, \quad (2)$$

where η is the time dilation term, corresponding to the time transformation from the shadowing lemma discussed in the previous section. For a detailed derivation of equation (2), see the appendices of ‘‘Sensitivity computation of chaotic limit cycle oscillations.’’ by Q. Wang, R. Hui and P. Blonigan, available at arXiv:1204.0159.

This optimization problem is a linearly constrained least-squares problem, with the following KKT equations, derived using calculus of variations (see appendix A.1):

$$\frac{\partial w}{\partial t} = - \left(\frac{\partial f}{\partial u} \right)^* w - v \quad w(0) = w(T) = 0 \quad (3)$$

$$\alpha^2 \eta = - \langle f, w \rangle \quad (4)$$

$$\frac{dv}{dt} = \frac{\partial f}{\partial u} v + \frac{\partial f}{\partial s} + \eta f \quad (5)$$

Equations (3), (4) and (5) can be combined to form a single second order equation for the Lagrange multiplier w , as in [12]:

$$-\frac{d^2 w}{dt^2} - \left(\frac{d}{dt} \left(\frac{\partial f}{\partial u} \right)^* - \frac{\partial f}{\partial u} \frac{d}{dt} \right) w + \left(\frac{\partial f}{\partial u} \left(\frac{\partial f}{\partial u} \right)^* + \frac{1}{\alpha^2} f f^* \right) w = \frac{\partial f}{\partial \xi} \quad (6)$$

$$w(0) = w(T) = 0$$

Equation (6) shows that the LSS method has changed the tangent equation from an initial value problem in time to a boundary value problem in time. The LSS solution can be used to compute gradients for some quantity of interest J (i.e. Drag):

$$\frac{\partial \bar{J}}{\partial \xi} = \left\langle \frac{\partial J}{\partial u}, v \right\rangle + \bar{\eta} \bar{J} - \bar{\eta} \bar{J} \quad (7)$$

Where $\bar{x} \equiv \frac{1}{T} \int_0^T x dt$. See appendix A.2 for a derivation of equation (7).

2.3 Solving the KKT system numerically

Equations (3), (4) and (5) are discretized using finite differences and combined to form the following symmetric system:

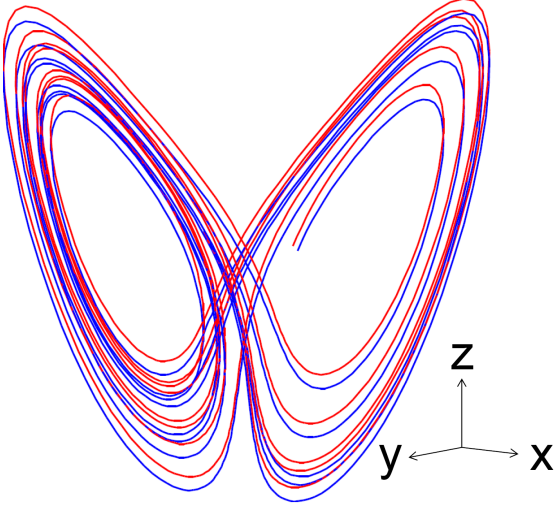


Figure 4: Lorenz equation phase space trajectory ($u(t)$) for $r = 28$ (blue) and a corresponding approximate shadow trajectory ($u(t) + v(t)$) (red). Integration time was $T = 20$ in dimensionless time units.

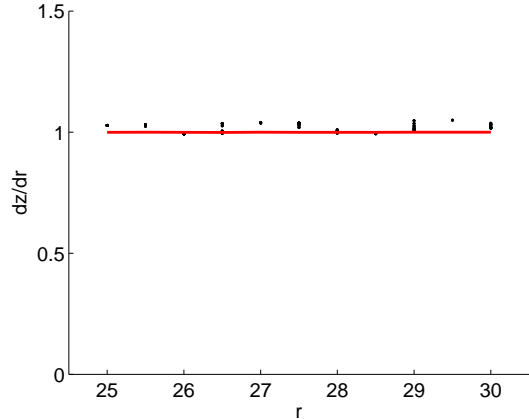


Figure 5: Gradient of long-time averaged z with respect to the parameter r . Gradients computed with trajectory length $T = 20$ are shown as black diamonds. Those computed using $T = 1000$ are shown as a red line.

Although rearranging the system to a banded system and solving with a direct works well for smaller systems like the Lorenz system, it would not be very efficient for LSS with larger systems.

Firstly, the KKT system is quite large, with $2mn + n + 1$ by $2mn + n + 1$ elements for m time steps. For a discretization with a stencil of five elements, the matrix would have approximately $23mn$ non-zero elements. Consider a CFD simulation with 1×10^5 nodes and 16000 time steps. For a 2D compressible flow solver, there are four unknowns at each node, therefore, the number of degrees of freedom, n is 4×10^5 . For this simulation, the KKT matrix would be 1.28×10^{10} by 1.28×10^{10} with 1.47×10^{11} non-zero elements.

Also, the bandwidth of the rearranged KKT system scales with the size of the system n , which means that direct methods, whose operation count scales with the matrix bandwidth squared times the rank, do not scale well for larger problems like CFD simulations. On the other hand, an iterative method scales with the number of non-zero elements, not the product of bandwidth and rank. Therefore, the operation count of an iterative method scales with mn , not mn^3 like a direct method. Because of this, iterative methods, such as multigrid, should be used to apply LSS to CFD simulations and other larger dynamical systems.

3 Multigrid-in-time for the Least Squares Shadowing Method

As discussed in section 2.3, the KKT system is very large for many problems of interest, therefore we consider iterative methods to solve the KKT system. As equation (6) is a boundary value problem in time, a multigrid-in-time scheme is attractive because of its fast convergence relative to other iterative methods for many boundary value problems [1].

Several multigrid schemes were used to solve the LSS KKT system, with vastly varying convergence rates and computational costs. The first scheme considered, referred to as “classic” multigrid, is based upon the geometric multigrid scheme outlined by Briggs et al. [1]. This scheme leads to very slow convergence of the KKT system’s residual. The second scheme considered was cyclic reduction. This scheme converges in one cycle if the system is solved exactly on the coarsest grid. However, the implementation of this scheme would use too much memory or require too many floating point operations to be viable for solving large scale systems with the computational resources that are currently available to most engineers and scientists. Finally, it was found that using a Krylov subspace solver as a smoother and higher order averaging between the KKT systems on the fine and coarse grids led to textbook multigrid convergence rates. This was found to be the case for schemes which satisfied the variational condition and those that did not, although schemes that did not satisfy the variational condition converged slightly slower than those that did.

This section is comprised of a detailed discussion of the above results. Each subsection will outline each scheme. Also, each subsection will present the performance of each scheme when used to apply the LSS method to the Lorenz equations. Finally, the implications of these results for the computational efficiency of each scheme will be discussed.

3.1 KKT system Schur complement

Rather than directly discretizing equation (6), we solve the Schur complement of the KKT system shown in section 2.3. The KKT system can be written as:

$$\begin{pmatrix} I & 0 & B^T \\ 0 & \alpha^2 I & C^T \\ B & C & 0 \end{pmatrix} \begin{pmatrix} \underline{v} \\ \underline{\eta} \\ \underline{w} \end{pmatrix} = - \begin{pmatrix} 0 \\ 0 \\ \underline{b} \end{pmatrix}$$

where B is a $mn \times (m + 1)n$ matrix and C is a $mn \times m$ matrix. Conducting block Gaussian elimination, the Schur complement is found to be:

$$\underbrace{(BB^T + \frac{1}{\alpha^2}CC^T)}_A \underline{w} = \underline{b}$$

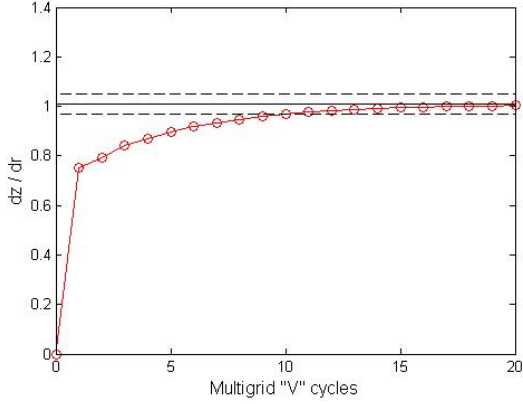


Figure 6: Convergence of the gradient of time-averaged z with respect to r , as computed using multigrid in time with 10 relaxation iterations before restriction and after prolongation on each level.

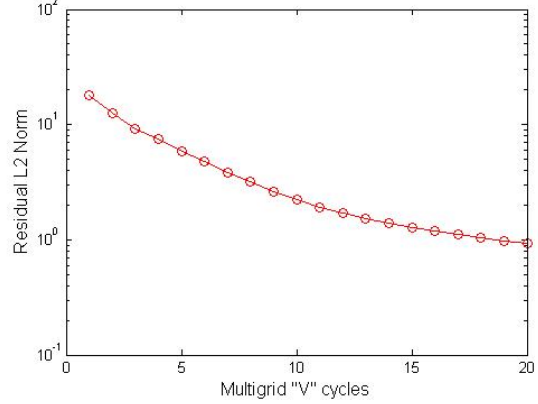


Figure 7: L2 norm of the residual while solving for the gradient of time-averaged z with respect to r using multigrid in time. Similar behavior was observed when computing other gradients

3.3 Cyclic Reduction

In classic multigrid, linear interpolation is used for prolongation, but it is not the best method when the coefficients of the equation being solved are highly oscillatory or discontinuous [15]. If this is the case, the coarse grid correction is a poor approximation for the low order error, and textbook multigrid convergence is not achieved. In this case, interpolation to some fine grid point with index i should be carried out by solving (6) with $\frac{\partial f}{\partial \xi} = 0$ between the two nearest coarse grid points, $i - 1$ and $i + 1$, with the boundary conditions $w(t_{i-1}) = w_{i-1}$ and $w(t_{i+1}) = w_{i+1}$. It can be proved that this interpolation scheme leads to a multigrid scheme that converges independently of grid size for 1D problems [15]. Furthermore, it can be shown that this scheme is equivalent to cyclic reduction [15]. Cyclic reduction is carried out as follows: defining the lower, main and upper diagonal blocks in equation (9) as L_i , D_i and U_i , elimination is conducted as follows for three given rows:

$$L_{i-1}w_{i-2} + D_{i-1}w_{i-1} + U_{i-1}w_i = b_{i-1} \quad (11)$$

$$L_iw_{i-1} + D_iw_i + U_iw_{i+1} = b_i \quad (12)$$

$$L_{i+1}w_i + D_{i+1}w_{i+1} + U_{i+1}w_{i+2} = b_{i+1} \quad (13)$$

Equations (11) and (13) give expressions for w_{i-1} and w_{i+1} , which can be substituted into equation (12):

$$L_I w_{i-2} + D_I w_i + U_I w_{i+2} = b_I$$

Where:

$$\begin{aligned}
L_I &= -L_i D_{i-1}^{-1} L_{i-1} \\
D_I &= -L_i D_{i-1}^{-1} U_{i-1} + D_i - U_i D_{i+1}^{-1} L_{i+1} \\
U_I &= -U_i D_{i+1}^{-1} U_{i+1} \\
f_I &= -L_i D_{i-1}^{-1} f_{i-1} + f_i - U_i D_{i+1}^{-1} f_{i+1} \\
b_I &= -L_i D_{i-1}^{-1} b_{i-1} + b_i - U_i D_{i+1}^{-1} b_{i+1}
\end{aligned} \tag{14}$$

If the system is solved exactly on the coarsest grid, cyclic reduction will converge in one cycle. Also, the algorithm can be implemented in parallel, as each coarse grid equation only depends on three adjacent fine grid equations.

Although equation (14) involves inverting the main diagonal matrices D , which contains products of Jacobians $\frac{\partial f}{\partial u}$, in practice these inversions do not need to be carried out (see appendix B.1). However, the operation count of this scheme does not scale very well with the size of the KKT system. It can be shown (see appendix B.2) that the number of floating point operations, N , required for matrix multiplication with the coarse grid matrix scales approximately as:

$$N \sim \mathcal{O}(2p(2q)^l)$$

Where p is the number of operations required to multiply a vector by an F_i or G_i matrix, q is the number of iterations needed by an iterative solver to solve any system involving D_i^{-1} , and l is the number of levels, assuming the coarsest grid is $n \times n$ (one time step). This is very large relative to the approximate operation count for a single Jacobi iteration on the LSS KKT system:

$$N \sim \mathcal{O}(2^{(l+2)}p)$$

To summarize, the advantages of cyclic reduction are in its memory efficiency and potential for parallel implementation, not its operation count.

3.4 Higher Order Averaging/Krylov Subspace Scheme

Taking into account the performance of classic multigrid and cyclic reduction, a new multigrid scheme was designed for solving the KKT system associated with LSS. The slow convergence of classic multigrid and the rapid convergence of cyclic reduction indicate that the accuracy of the coarse grid correction has a large effect on the convergence rate of the scheme. To obtain a more accurate coarse grid correction, higher order averaging was used to coarsen the KKT system. Higher order averaging ensures that the coarse grid non-linear solution $u(t)$ from which the KKT system is constructed is smooth, which leads to better multigrid performance [15].

Higher order averaging schemes are formed as follows: consider a two point (first order) average of some value x at time step i :

$$\bar{x}_i = \frac{1}{2}x_{i-1/2} + \frac{1}{2}x_{i+1/2}$$

A first order scheme is formed by setting $x_{i-1/2} = x_{i-1}$ and $x_{i+1/2} = x_{i+1}$. For a second order scheme, set $x_{i-1/2} = \frac{1}{2}x_{i-1} + \frac{1}{2}x_i$ and $x_{i+1/2} = \frac{1}{2}x_i + \frac{1}{2}x_{i+1}$:

$$\bar{x}_i = \frac{1}{4}x_{i-1} + \frac{1}{2}x_i + \frac{1}{4}x_{i+1}$$

The third order scheme can be derived by substituting first order averages into the second order scheme, and so on:

$$\begin{aligned} \bar{x}_{i+1/2} &= \frac{1}{8}x_{i-1} + \frac{3}{8}x_i + \frac{3}{8}x_{i+1} + \frac{1}{8}x_{i+2} && \text{3rd Order} \\ \bar{x}_i &= \frac{1}{16}x_{i-2} + \frac{1}{4}x_{i-1} + \frac{3}{8}x_i + \frac{1}{4}x_{i+1} + \frac{1}{16}x_{i+2} && \text{4th Order} \\ \bar{x}_{i+1/2} &= \frac{1}{32}x_{i-2} + \frac{5}{32}x_{i-1} + \frac{10}{32}x_i + \frac{10}{32}x_{i+1} + \frac{5}{32}x_{i+2} + \frac{1}{32}x_{i+3} && \text{5th Order} \end{aligned}$$

Higher order averaging was applied to multigrid-in-time in two ways. *Matrix restriction multigrid* uses higher order averaging on the KKT matrix itself, using McCormick et al's variational conditions [16]. *Solution restriction multigrid* uses higher order averaging on the non-linear solution $u(t)$. The KKT system is formed on the coarse grid from the restricted solution $u(t)$.

In addition, it was observed that stationary iterative methods such as the Block Gauss Seidel method used in section 3.2 converge very slowly for the KKT system. Krylov subspace methods such as conjugate gradient and MINRES¹ were examined as smoothers as an alternative to the stationary methods used in classic multigrid. Other parameters of the KKT system and the multigrid solver were also examined and some were found to have a considerable effect on convergence rates, especially the parameter α^2 from equation (6), the weighting of the time dilation term in the minimization statement.

The following sections discuss matrix restriction multigrid and its application to LSS for the Lorenz equations, followed by a discussion of solution restriction multigrid. The computational cost of both multigrid methods is compared to MINRES.

3.4.1 Matrix restriction multigrid

Matrix restriction multigrid was designed to satisfy the variational conditions [16]:

$$I_h^{2h} = c^h(I_{2h}^h)^T, \quad A^{2h} = I_h^{2h} A^h I_{2h}^h$$

Where A^h is the fine grid matrix, A^h is the coarse grid matrix, I_h^{2h} is the restriction matrix, I_{2h}^h is the prolongation matrix and c^h is some constant that could depend on the grid dimension.

Satisfying the variational condition ensures that the error of the solution will decrease monotonically, assuming the smoother decreases or does not change the magnitude of the error on all grids [16].

¹These methods were chosen because the KKT system is Symmetric Positive Definite.

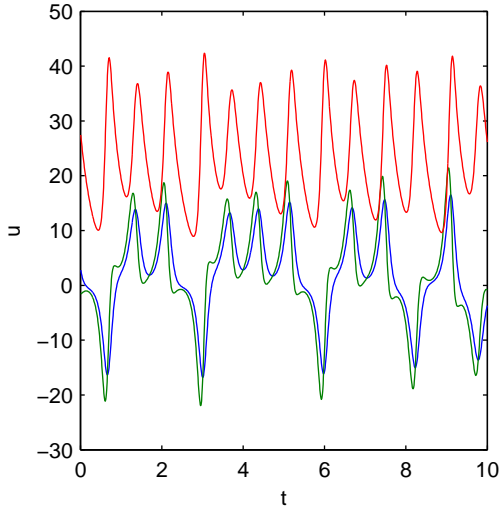


Figure 8: Lorenz equation solution; x, y, z are blue, green and red respectively.

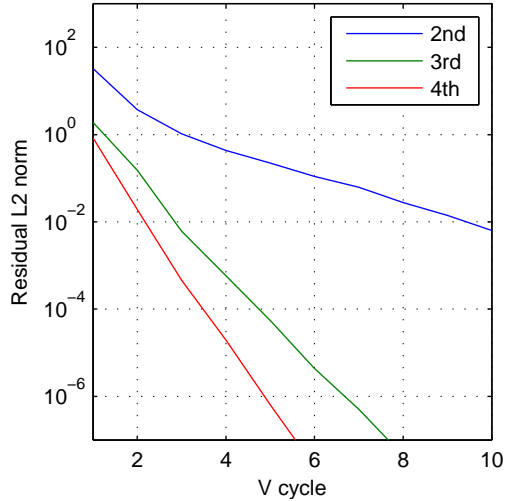


Figure 9: Convergence of matrix restriction multigrid with $dt_f = 0.01$ for different orders of averaging.

Matrix restriction multigrid has been demonstrated on LSS for the Lorenz equations, in particular the solution shown in figure 8. Unless otherwise stated, the results presented correspond to a multigrid scheme with conjugate gradient (CG) smoothing with $\nu_1 = 30$ presmoothing iterations, $\nu_2 = 30$ postsmoothing iterations, 4th order averaging, $dt = dt_f = 0.0012$ on the finest grid and $dt = 0.08$ on the coarsest grid. Although 30 is a large amount of smoothing iterations, it will be shown in section 3.4.2 that it results in a multigrid method that requires less operations than a Krylov subspace method on its own.

Since the typical Lorenz equation solution in figure 8 is oscillatory, a higher order averaging scheme can be used to improve the coarse grid correction by ensuring that the components of the block matrices (which depend on $u(t)$) that make up KKT matrix vary smoothly even on very coarse grids. Because of this, the use of higher order averaging drastically improves the convergence rate of multigrid-in-time, as shown by figure 9.

Block Gauss-Seidel and other stationary solvers that are used for smoothing in classic multigrid schemes were found to converge very slowly. Also, the convergence of these solvers worsen as the grid is coarsened, as shown in figure 10. This is because the largest eigenvalue of the KKT system, which can be shown to set the convergence rate of a stationary solver [17], decreases at a slower rate as the KKT system is coarsened, as shown in figure 11. This is in contrast to the behavior of the finite difference (or finite element) matrix of the Poisson equation, whose largest eigenvalue decreases exponentially as the grid is coarsened, resulting in much faster convergence of Jacobi or Gauss-Seidel solvers on the coarser grids [17].

To accelerate convergence of multigrid-in-time, conjugate gradient (CG) is used as a smoother. With a CG smoother, the rate of convergence is bounded by $((\kappa - 1)/\kappa)^N$, where κ is the condition number of the KKT matrix A_h and N is the number of smoothing iterations

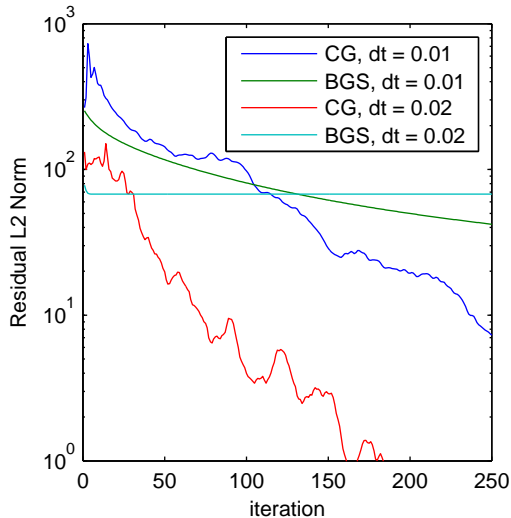


Figure 10: Convergence of Block Gauss-Seidel and Conjugate Gradient on the fine grid ($dt=0.01, \alpha^2 = 40$) and a coarsened grid ($dt=0.02, \alpha^2 = 40$).

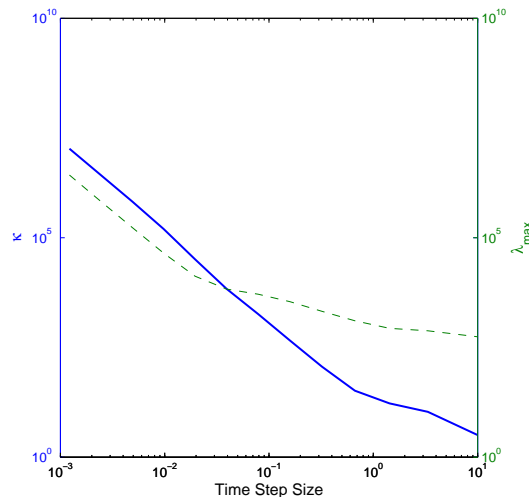


Figure 11: Condition Number κ (solid line) and maximum eigenvalue λ_{max} (dashed line) versus time step dt for coarsened grids corresponding to a fine grid with $dt = 0.001$ for the LSS system associated with the Lorenz equations

[18]. The condition number decreases quickly as the grid is coarsened (see figure 11), leading to faster CG convergence on coarser grids, as seen in figure 10.

A matrix restriction scheme with a CG smoother and 4th order averaging has been observed to converge independently of the number of fine grid points, as shown in figure 12.

In addition, a number of tuning parameters were observed to affect this convergence rate. The parameter α^2 from equation (6), the weighting of the time dilation term in the minimization statement, has a strong effect on multigrid convergence, as shown in figure 13. There is an optimal α^2 , found to be equal to around 40 in the case of our sample problem, but this number is most likely specific to the Lorenz system. α^2 has an effect on convergence because it affects the condition number and the eigenvalue spectrum of A on all grids.

The coarsening threshold dt_c is defined as the time step size below which multigrid is not called recursively. Figure 14 shows that there is an optimal value for dt_c . Below this, the coarse grid correction actually slows convergence in some cases, because it is a poor approximation for low order errors.

To determine the relative efficiency of matrix restriction multigrid we compare its computational cost to that of using MINRES to solve the fine grid solution. Figure 20 shows that the solution of the KKT system converges after about 4700 iterations, and the gradient converges after about 1800 iterations. For a Krylov subspace method applied to a sparse matrix, the number of floating point operations for a single iteration, p_{MINRES} , is $p_{MINRES} \sim \mathcal{O}(mn)$, m is the number of time steps, n is the number of dimensions of the dy-

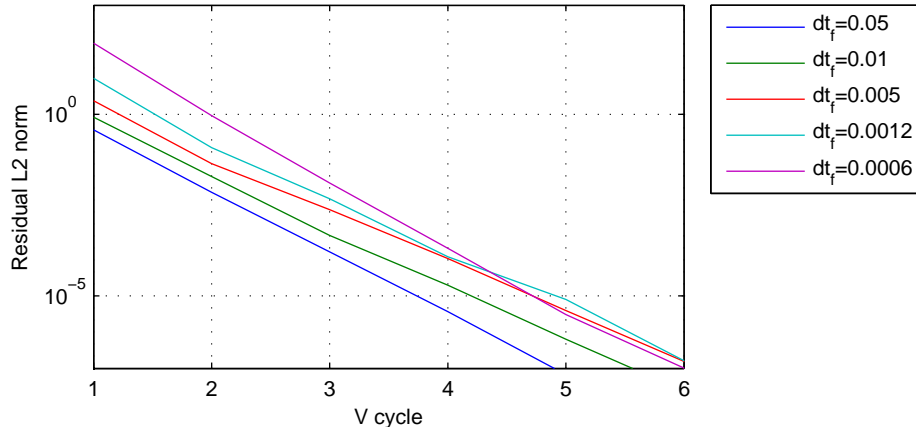


Figure 12: Convergence of matrix restriction multigrid for different fine grid time steps dt_f .

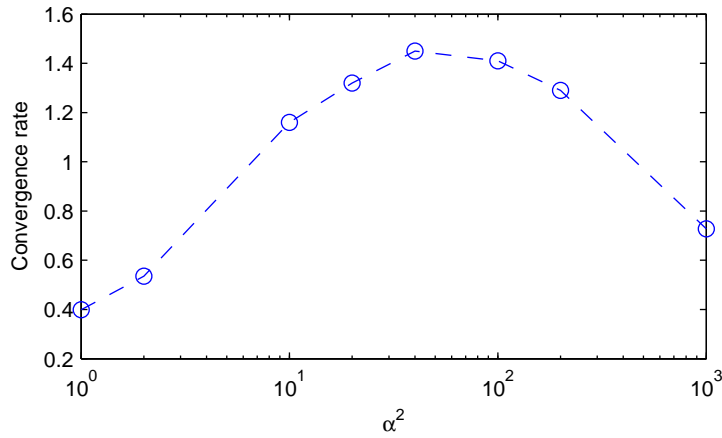


Figure 13: Convergence rate of matrix restriction multigrid versus α^2 . The convergence rate is $-\gamma$, from the curve fit $\log_{10} \|r\|_{L2} = \gamma \log_{10} N_V + \log_{10} C$ of the residual L2 norm $\|r\|_{L2}$ versus the number of V-cycles N_V .

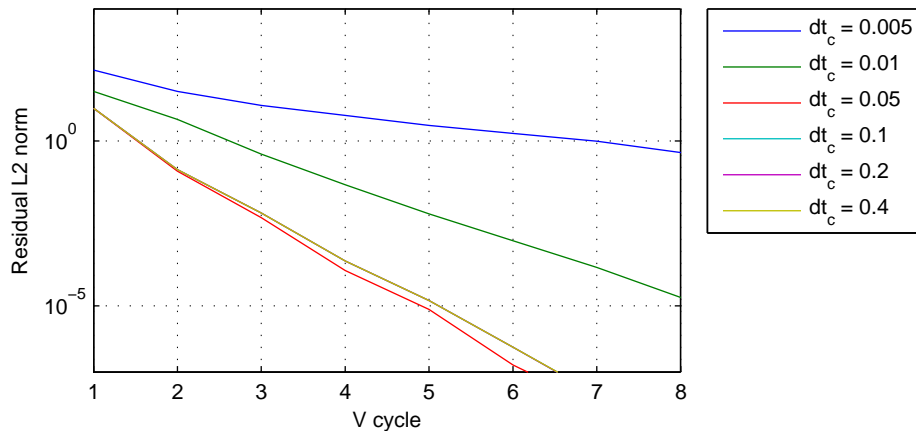


Figure 14: Convergence of matrix restriction multigrid for different coarsening thresholds dt_c .

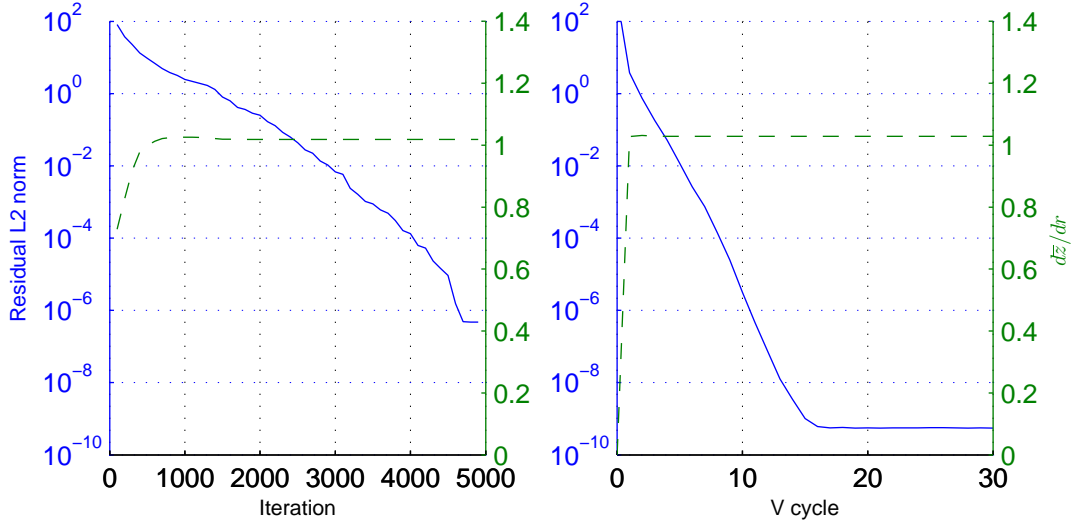


Figure 15: LEFT: Convergence of MINRES for an LSS system for the Lorenz equations with $dt_f = 0.004$ and $\alpha^2 = 40$. RIGHT: Convergence of matrix restriction multigrid for an LSS system for the Lorenz equations. The dashed lines shows the gradient computed at a given iteration, which should be roughly 1.01 ± 0.04 [7]

namical system, and mn is the size of the Schur complement of the KKT system in equation (9) [17]. Therefore, for the solution in figure 15 the total number of operations, P_{MINRES} , is roughly

$$P_{MINRES} \sim 4700\mathcal{O}(mn)$$

Using the variational condition to form the coarse grid matrix makes the matrix restriction multigrid more expensive than classic multigrid. To conduct matrix multiplication on a coarse grid, the coarse grid solution is prolonged to the fine grid, multiplied by the fine grid matrix A_h and then restricted to the coarse grid. Ignoring the cost of restriction and prolongation, for a fixed number of smoothing iterations $N = \nu_1 + \nu_2$ and 10 grids, the cost of one V-cycle of matrix restriction multigrid, p_{MMG} , is:

$$p_{MMG} \sim 10\mathcal{O}(mn(\nu_1 + \nu_2)) = 600\mathcal{O}(mn) \quad (15)$$

Figure 15 shows that the solution of the KKT system converges after about 17 cycles with matrix restriction multigrid, therefore:

$$P_{MMG} \sim 10200\mathcal{O}(mn) \approx 2P_{MINRES}$$

Matrix restriction multigrid requires roughly twice as many operations as MINRES. However, matrix restriction multigrid computes the correct gradient after 2 cycles, which requires $1200\mathcal{O}(mn)$ operations, slightly less than the roughly $1800\mathcal{O}(mn)$ operations required by MINRES.

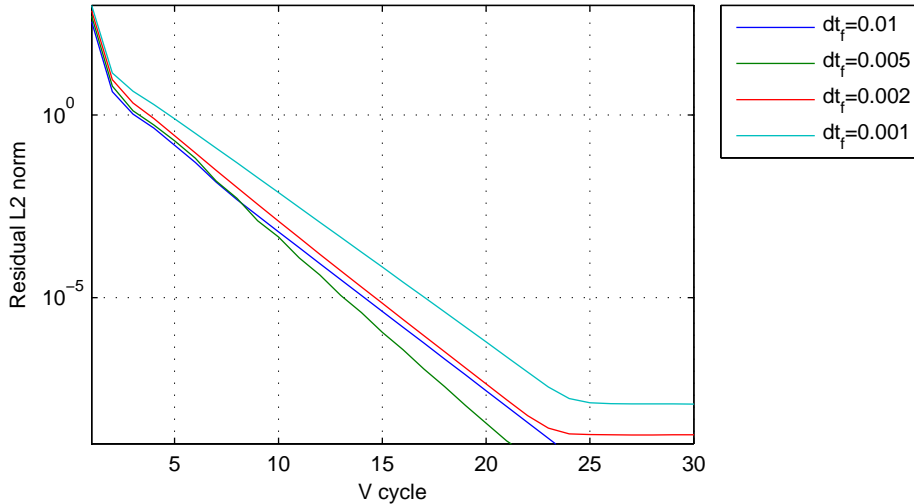


Figure 16: Convergence of solution restriction multigrid for different values of dt_f .

3.4.2 Solution restriction multigrid

Although matrix restriction multigrid performs very well, it costs much more than MINRES. We can reduce computational costs of multigrid by using solution restriction instead of matrix restriction. Solution restriction only satisfies part of the variational condition, the restriction and prolongation operators are transposes of one another:

$$I_h^{2h} = c^h (I_{2h}^h)^T$$

By restricting the non-linear solution $u(t)$ instead of the matrix A_h , solution restriction multigrid results in reduced smoothing costs on the coarse grid, because the KKT system formed on the coarse grid from the restricted $u(t)$ is half the size of the fine grid system. As shown by figure 16, the solution restriction scheme also leads to convergence rates with little dependence on dt_f for a given dt_c for the Lorenz equations ².

As observed for matrix restriction multigrid, the value of the parameter α^2 and the order of averaging both affect the convergence rate, as shown in figures 17 and 18. However, higher order averaging is only beneficial to a certain degree, as the 5th order scheme leads to slower convergence than the 3rd order one (figure 18). This is because high order averaging could smooth $u(t)$ too much. If this is the case the course grid solution is a poor approximation of the errors on the fine grid. Also, α^2 has a much greater effect than the averaging scheme on the convergence rate of solution restriction multigrid.

There is a slight trade-off for the lower costs of solution restriction multigrid. Figure 19 shows that solution restriction multigrid converges slightly slower than matrix restriction multigrid. However, this slower convergence does not outweigh the benefits of the lower cost

²Parameter values used for this plots in this section (unless otherwise stated): $dt_f = 0.004$ or $m = 4096$, $\nu_1 = \nu_2 = 30$, $dt_c = 0.2$, $\alpha^2 = 40$, MINRES smoothing, and 3rd order averaging

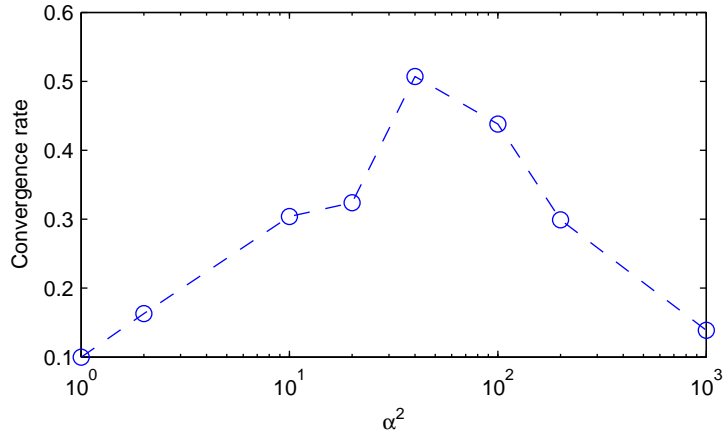


Figure 17: Convergence rate of solution restriction multigrid versus α^2 . The convergence rate is $-\gamma$, from the curve fit $\log_{10} \|r\|_{L2} = \gamma \log_{10} N_V + \log_{10} C$ of the residual L2 norm $\|r\|_{L2}$ versus the number of V-cycles N_V .

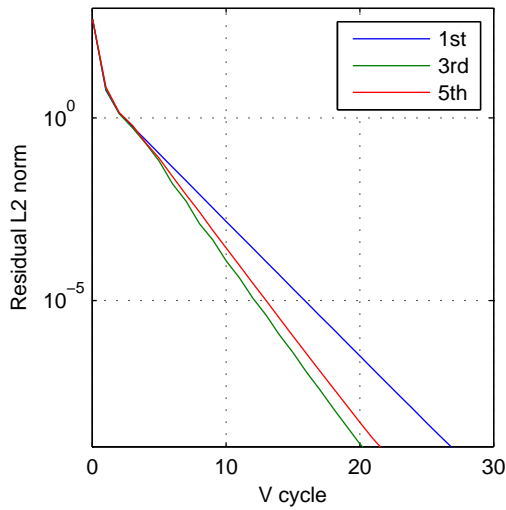


Figure 18: Convergence of solution restriction multigrid with 1st, 3rd and 5th order averaging.

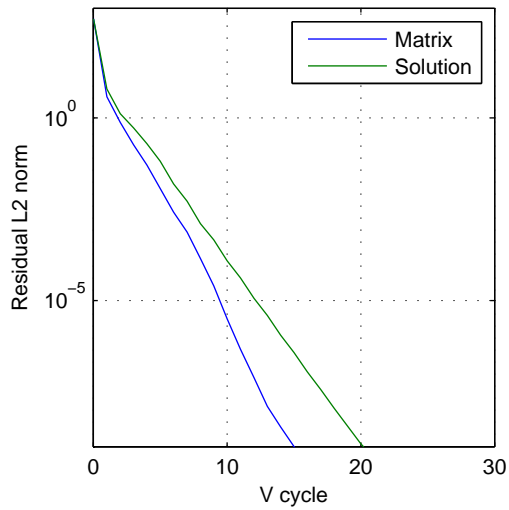


Figure 19: Convergence plots for matrix restriction multigrid with MINRES smoothing, $dt_c = 5$, and 4th order averaging and solution restriction multigrid.

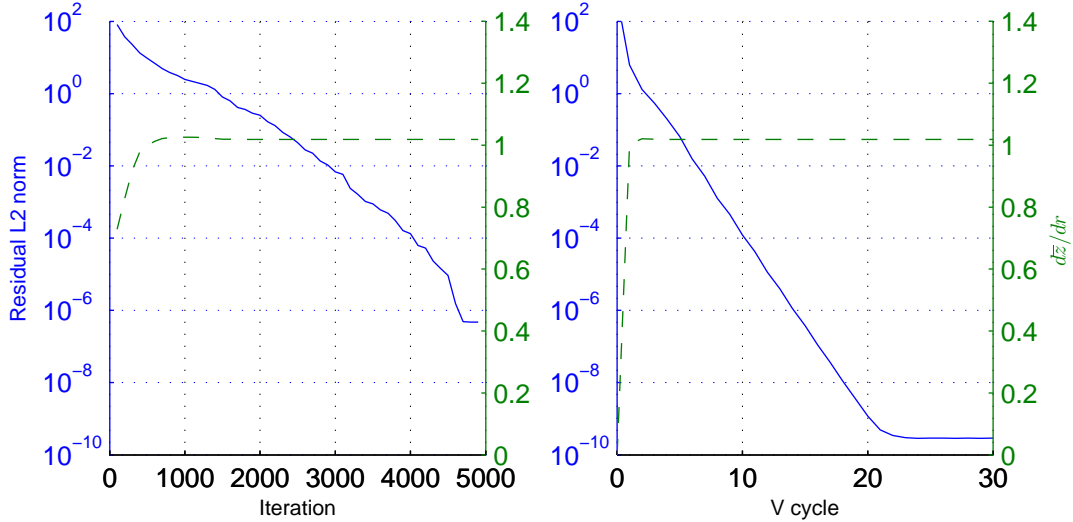


Figure 20: LEFT: Convergence of MINRES for an LSS system for the Lorenz equations with $dt_f = 0.004$ and $\alpha^2 = 40$. RIGHT: Convergence of solution restriction multigrid for an LSS system for the Lorenz equations. The dashed lines shows the gradient computed at a given iteration, which should be roughly 1.01 ± 0.04 [7]

of solution restriction multigrid.

These benefits can be seen by comparing the cost of solution restriction multigrid to that of matrix restriction multigrid and MINRES. For a fixed number of iterations $N = \nu_1 + \nu_2$, the cost of smoothing is halved for one coarsening as mn is halved when the system is coarsened. When a large number of grids are used for a V-cycle, the cost of one V-cycle of solution restriction multigrid, p_{SMG} , is:

$$\begin{aligned}
 p_{SMG} &\sim \mathcal{O}(mn(\nu_1 + \nu_2)) + \frac{1}{2}\mathcal{O}(mn(\nu_1 + \nu_2)) + \frac{1}{4}\mathcal{O}(mn(\nu_1 + \nu_2)) + \dots \\
 &\approx 2\mathcal{O}(mn(\nu_1 + \nu_2)) = 120\mathcal{O}(mn)
 \end{aligned}$$

Figure 20 shows that the solution of the KKT system converges after about 20 cycles with solution restriction multigrid, therefore:

$$P_{SMG} \sim 2400\mathcal{O}(mn) \approx \frac{1}{2}P_{MINRES} \approx \frac{1}{4}P_{MMG}$$

In addition to requiring only half of the floating point operations of MINRES and a quarter of the operations of matrix restriction multigrid, solution restriction multigrid computes the correct gradient after 2 cycles, which requires $240\mathcal{O}(mn)$ operations, an order of magnitude less than the roughly $1800\mathcal{O}(mn)$ operations required by MINRES and a quarter of the cost of matrix restriction multigrid.

4 Conclusion

In conclusion, a multigrid-in-time scheme could be used to implement the least squares shadowing (LSS) method in a relatively computationally efficient manner. To solve the large KKT system associated with LSS, several multigrid-in-time schemes have been investigated. Classic geometric multigrid with a Gauss-Seidel smoother was found to converge very slowly. Cyclic reduction converged in one cycle and can be run in parallel but it used a very large number of floating point operations to solve the system if the number of dimensions of the system, n , was very large. A higher order averaging multigrid scheme with a Krylov subspace smoother was found to give textbook multigrid convergence. Matrix restriction multigrid converges quickly, but smoothing on the coarse grid requires a similar amount of operations to smoothing on the fine grid. Solution restriction multigrid converges slightly slower than matrix restriction multigrid, but is considerably less expensive. It was also found that the parameter α^2 , the weighting of the time dilation term in equation (2), had a large effect on the rate of convergence of matrix restriction and solution restriction multigrid.

Because of its lower cost, solution restriction multigrid is currently the most promising numerical method for implementing LSS for large scale systems. Therefore, future work will start with further investigation of solution restriction multigrid-in-time. In particular, a method to determine the α^2 value corresponding to the fastest rate of convergence needs to be found.

Once a robust, scalable numerical method for solving the KKT equation is found, LSS will be extensively tested on chaotic and turbulent fluid flows, such as homogeneous, isotropic turbulence and turbulent channel flow. Eventually, LSS could be used to investigate more complicated flows such as flow around a lifting body, the fuel injection system in a jet engine or scramjet combustor, or an internal flow in a rocket engine.

Acknowledgements

The authors would like to acknowledge AFSOR Award F11B-T06-0007 under Dr. Fariba Fahroo, NASA Award NNH11ZEA001N under Dr. Harold Atkins as well as financial support from the NDSEG fellowship.

A Least Squares Shadowing Sensitivity Analysis

A.1 Deriving the KKT System

First form the Lagrangian function for equation (2):

$$L = \int_0^T \frac{v^2 + \alpha^2 \eta^2}{2} + \left\langle w, \left(-\frac{dv}{dt} + \frac{\partial f}{\partial u} v + \frac{\partial f}{\partial \xi} + \eta f \right) \right\rangle dt$$

Now consider the first variation, which can be rearranged and transformed by integration by parts:

$$\delta L = \int_0^T \left\langle \delta v, \left(v + \frac{dw}{dt} + \frac{\partial f^*}{\partial u} w \right) \right\rangle dt + \int_0^T (\alpha^2 \eta + \langle f, w \rangle) \delta \eta dt + \langle \delta v, w \rangle \Big|_0^T$$

For first order optimality $\delta L = 0$ for all δv and $\delta \eta$, therefore:

$$\boxed{v + \frac{dw}{dt} + \frac{\partial f^*}{\partial u} w = 0, \quad w(0) = w(T) = 0} \quad (16)$$

and

$$\boxed{\alpha^2 \eta + \langle f, w \rangle = 0} \quad (17)$$

Equations (16) and (17), along with the tangent equation, make up the KKT system.

A.2 Computing Sensitivities using a Shadow Trajectory

For a trajectory $u(t)$ and shadow trajectory $u'(t)$:

$$\begin{aligned} \delta \bar{J} &= \frac{1}{T} \int_0^T J(u'(\tau)) d\tau - \frac{1}{T} \int_0^T J(u(t)) dt \\ &= \int_0^T \frac{1}{T} J(u'(\tau(t))) \frac{d\tau}{dt} + \frac{1}{T} J(u(t)) dt \end{aligned}$$

For some perturbation to the tangent equation $\delta f = \varepsilon \frac{\partial f}{\partial \xi}$, the time transformation is $\frac{d\tau}{dt} = 1 + \varepsilon \eta$:

$$\mathcal{T} = \int_0^T \frac{d\tau}{dt} dt = T + \varepsilon \underbrace{\int_0^T \eta dt}_{\equiv H} = T + \varepsilon H$$

Therefore:

$$\delta \bar{J} = \frac{1}{T + \varepsilon H} \int_0^T J(u'(\tau(t))) - J(u(t)) + \varepsilon \eta J(u'(\tau(t))) - \frac{\varepsilon H}{T} J(u(t)) dt$$

Diving through by ε :

$$\begin{aligned}\frac{\partial \bar{J}}{\partial \xi} &= \lim_{\varepsilon \rightarrow 0} \left[\frac{1}{T + \varepsilon H} \int_0^T \frac{(J(u'(\tau(t))) - J(u(t)))}{\varepsilon} + \eta J(u'(\tau(t))) - \frac{H}{T} J(u(t)) dt \right] \\ &= \frac{1}{T} \int_0^T \left\langle \frac{\partial J}{\partial u}, v \right\rangle dt + \frac{1}{T} \int_0^T \eta J(u(t)) dt - \frac{H}{T} \frac{1}{T} \int_0^T J(u(t)) dt\end{aligned}$$

By the definition of H , and defining $\bar{x} \equiv \frac{1}{T} \int_0^T x dt$:

$$\boxed{\frac{\partial \bar{J}}{\partial \xi} = \overline{\left\langle \frac{\partial J}{\partial u}, v \right\rangle} + \bar{\eta} \bar{J} - \bar{\eta} \bar{J}}$$

B Cyclic Reduction

B.1 Conducting cyclic reduction without inverting main diagonal matrices

Consider the following system:

$$\begin{pmatrix} D_1 & U_1 & 0 \\ L_2 & D_2 & U_2 \\ 0 & L_3 & D_3 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

Applying equation (14), the following system is obtained:

$$Aw_2 = b$$

with:

$$\begin{aligned}A &= -L_2 D_1^{-1} U_1 + D_2 - U_2 D_3^{-1} L_3 \\ b &= -L_2 D_1^{-1} b_1 + b_2 - U_2 D_3^{-1} b_3\end{aligned}$$

This system can be solved iteratively, using some preconditioner P :

$$P \Delta x = b - Ax_k, \quad x_{k+1} = x_k + \Delta x$$

Where x_k is the value of w_2 after k iterations. To compute Ax_k , it is decomposed into three parts:

$$Ax_k = -L_2 D_1^{-1} U_1 x_k + D_2 x_k - U_2 D_3^{-1} L_3 x_k = \alpha + \beta + \gamma$$

α and γ include an inverted matrix, but the inversion can be avoided as follows:

$$-L_2 D_1^{-1} U_1 x_k = \alpha$$

Compute $y_k = U_1 x_k$:

$$-L_2 D_1^{-1} y_k = \alpha$$

Next, define $z_k = D_1^{-1} y_k$. Iteratively solve:

$$D_1 z_k = y_k$$

and use the result to compute α :

$$\alpha = -L_2 z_k$$

γ and the right hand side b can be computed using a similar method. This idea can be applied to a much larger system and allows cyclic reduction to be conducted without inverting any Jacobian matrices.

B.2 Estimating the operation count for cyclic reduction

The following section estimates the operation count of the algorithm of section B.1, in which block cyclic reduction is conducted without directly inverting any blocks of the matrix. First define:

- p : the number of flops required to multiply a vector by a Jacobian matrix for the system of interest.
- q : the number of iterations required to carry out multiplication by an inverse matrix.
- n : the number of states in the system

As $L_i = F_{i-1} G_{i-1}^T$ and $U_i = G_i F_i^T$, multiplication by these matrices requires $2p$ flops. As $D_i = F_{i-1} F_{i-1}^T + G_i G_i^T + f_i f_i^T$, multiplication by these matrices requires $4p + 2n$ flops. For a Jacobi solver, the number of flops for inverse matrix multiplication is qp .

An estimate of the operation count of cyclic reduction is shown for a few low order terms. These were computed using a symbolic calculator. Starting from a 1 by 1 coarse grid, the number of flops for multiplication by D , U or L were substituted into equation (14) recursively. The highest order term for a few different grids is shown in table 1, with the number of operations for a single Jacobi iteration (derived in the same way) for comparison. These estimates do not take into account fixed number of operations to backward substitute the coarse grid solution for the fine grid solution.

Time steps m	CR Flops	Jacobi Flops
3	$8pq^2$	$20p + 13n$
5	$16pq^3$	$36p + 23n$
9	$32pq^4$	$68p + 43n$
17	$64pq^5$	$132p + 83n$

Table 1: Estimate of Operation Count per iteration for cyclic reduction and for the Jacobi method for comparison. Note that the cost of cyclic reduction is similar to that of a block Gaussian elimination scheme as block cyclic reduction is like applying block Gaussian elimination in parallel to a permuted system.

References

- [1] Briggs WL, Henson VE, McCormick, SF. *A multigrid tutorial*. Society for Industrial and Applied Mathematics: Philadelphia, second edition, 2000.
- [2] Jameson A. Aerodynamic design via control theory. *Journal of Scientific Computing* 1988; **3** : 233-260.
- [3] Reuther JJ, Jameson A, Alonso JJ, Rimlinger MJ, Sanders D. Constrained multipoint aerodynamic shape optimization using an adjoint formulation and parallel computers. *Journal of Aircraft* 2001; **36** : 21-58.
- [4] Venditti D, Darmofal D. Grid adaptation for functional outputs: Application to two-dimensional inviscid flow. *Journal of Computational Physics* 2002; **176** : 40-69.
- [5] Lorenz E. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences* 1963; **20** : 130-141.
- [6] Lea D, Allen M, Haine T. Sensitivity analysis of the climate of a chaotic system. *Tellus* 2000; **52A** : 523-532.
- [7] Wang Q. Forward and adjoint sensitivity computation of chaotic dynamical systems. *Journal of Computational Physics* 2013; **235** : 1-13.
- [8] Pilyugin SY. *Shadowing in dynamical systems*. Volume 1706, Lecture Notes in Mathematics, Springer: Berlin, 1999.
- [9] Ginelli F, Poggi P, Turchi A, Chaté H, Livi R, Politi A. Characterizing dynamics with covariant lyapunov vectors. *Physical Review Letters* 2007; **99** : 130601.
- [10] Hasselblatt B. Hyperbolic dynamical systems. *Handbook of Dynamical Systems*, **2002** **1(A)**, pp. 239-319, DOI: S1874-575X(02)80005-4
- [11] Eyink G, Haine T, Lea D. Ruelle's linear response formula, ensemble adjoint schemes and lévy flights. *Nonlinearity* 2004; **17** : 1867–1889.
- [12] Wang Q, Gomez S, Blonigan P, Gregory A, Qian E. Towards scalable parallel-in-time turbulent flow simulations. *Physics of Fluids* 2013; **25** : 110818, DOI: 10.1063/1.4819390
- [13] Giles M, Pierce N. An introduction to the adjoint approach to design. *Flow, Turbulence and Combustion* 2000; **65** : 393-415.
- [14] Mineck R, Diskin B, Thomas J. Textbook multigrid efficiency for leading edge stagnation. May 2004 Web. Nov. 30, 2013. https://ia600605.us.archive.org/20/items/nasa_techdoc_20040081104/20040081104.pdf
- [15] Chan TF, Wan WL. Robust multigrid methods for nonsmooth coefficient elliptic linear systems. *Journal of Computational and Applied Mathematics* 2000; **123(1)** : 323-352.

- [16] McCormick SF, Ruge JW. Multigrid methods for variational problems. *SIAM Journal on Numerical Analysis* 1982; **19(5)** : 924-929.
- [17] Golub GH, Van Loan CF. *Matrix Computations*. The Johns Hopkins University Press: Baltimore, 1996.
- [18] Bank RE, Douglas CC. Sharp estimates for multigrid rates of convergence with general smoothing and acceleration. *SIAM journal on numerical analysis* 1985; **22(4)** : 617-633.