Data-Driven Reduced Order Control for Partially Observed Fluid Systems

Palash Sashittal*, Daniel J. Bodony†

University of Illinois at Urbana-Champaign, Urbana, IL, 61801, USA

We propose a novel method for non-intrusive reduced order control of partially observed flow systems. We formulate a rank-constrained matrix optimization problem for the maximum likelihood estimation of the reduced order model. An adjoint-based method is used for the gradient extraction and Riemannian optimization is performed for efficient convergence to the optimal solution. The resulting reduced order model is then used to design a Linear-Quadratic-Gaussian (LQG) controller. We demonstrate the performance of the proposed reduced order control method on the flow past an inclined flat plate at a high angle of attack and successfully prevent vortex shedding in the wake of the flat plate.

I. Introduction

Fluid flows of engineering interest are generally complex, multi-scale and require many degrees of freedom. This leads to a computational bottleneck for flow control and design optimization problems. One way to reduce the computational burden is through reduced order modeling of fluid flow systems. Model reduction methods that can leverage numerical and experimental data for any complex flow system with minimal knowledge about the system dynamics are desirable. Consequently, data-driven methods for reduced order modeling and flow control are an active field of research in the fluid mechanics community.

Dynamic Mode Decomposition (DMD) [1] is a powerful tool for system identification and mode extraction for unsteady fluid flows from numerical or experimental data. There are many variants of DMD that add features to the original framework by promoting sparsity [2], performing a more general optimization [3, 4, 5] and performing a total least-square regression [6] to de-bias the solution from noise present in the data. One disadvantage of these methods is that they were not intended to be used for partially observed systems which is often the case when dealing with experimental data.

An alternative way of looking at DMD is as a likelihood-based parameter estimation problem. This is the route we take in this study. We consider the observations as a linear projection of an underlying higher dimensional Hidden Markov Model (HMM) [7] with linear dynamics governed by unknown model parameters. The projection matrix (observation matrix) for the full order system is known. Our goal is to estimate these parameters given the observations over a certain time period. This requires the computation of filtered estimates of the observations followed by a Maximum Likelihood Estimate (MLE) [8, 9] of the unknown parameters.

We consider reduced order dynamics of the hidden variables by imposing a rank constraint on the state transition matrix of the full order system. The optimization for computing the MLE of the unknown parameters is performed on Riemannian manifolds for efficient convergence and the gradients are extracted using adjoint methods to further reduce the computational costs. The objective function is chosen to be the error between the filtered estimates of the measurements and the actual measurements. Unlike similar work done in [10] for system identification, we use the Kalman filter observer for the filtered estimates of the hidden state and measurements. This allows us to incorporate different levels of noise in the plant and the measurements, and to use finite time horizon Kalman filter with a time varying observer. Since the input-output relation of the linear dynamical system is invariant to a rotation of coordinates, the optimization is

*Graduate Research Assistant, Department of Aerospace Engineering, University of Illinois at Urbana-Champaign, 306 Talbot Lab, 104 S. Wright St., Urbana, 61801, IL, USA, AIAA Student Member

†Associate Professor, Department of Aerospace Engineering, University of Illinois at Urbana-Champaign

performed on a quotient manifold. We refer to this method of obtaining a reduced order model as mlDMD ('maximum likelihood dynamic mode decomposition').

II. Preliminaries

Consider the following linear discrete time-invariant (LTI) dynamical system

$$\begin{aligned} x_{k+1} &= Ax_k + w_k \\ y_k &= Cx_k + v_k \end{aligned} \tag{1}$$

with time indexed by k. The output of the system $y_k \in \mathbb{R}^p$ is a linear function of the state variable $x_k \in \mathbb{R}^m$. The observation matrix is $C \in \mathbb{R}^{p\times m}$. w_k and v_k are normally distributed zero mean random variables with covariance matrices $W \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{p\times p}$, respectively, with appropriate dimensions. The LTI system can be represented by the tuple (A, C, W, V). Two different tuples of appropriately sized matrices can give equivalent input-output systems. This fact can be used to reduced the dimensions of the reduced order model parameter space.

Only the output is observed; the state and the noise is hidden. Hence, this system is called partially observed. We denote the set of observations of the output y from iteration k to iteration ℓ by

$$\{y_k, y_{k+1}, \cdots y_\ell\} \coloneqq y_k^\ell.$$

For the sake of brevity, we will denote y_0^k as y^k. We will first look at some results from linear estimation theory that would give us the maximum likelihood estimate of the hidden variable x_k given the observations y^k. We will then use these results to formulate the reduced order modeling problem problem.

III. Linear Filtered Estimation

In this section we collect the needed results from linear estimation theory [9]. Let the estimate of x_k given observations y^ℓ be denoted by $x_{k|\ell}$. The estimate for the observation y_k will be similarly denoted as $y_{k|\ell}$. The standard estimation problem for linear state space models is determining the best estimate of x_k given y^k, denoted by $x_{k|k}$, in the sense that

$$E[(x_k - x_{k|k})^2] \leq E[(x_k - \tilde{x}_{k|k})^2],$$

for any causal estimate $\tilde{x}_{k|k}$. Using $x_{k|l} = E(x_k|y_1^l)$ and $\Sigma_{k|l} = \operatorname{var}(x_k|y_1^l)$, we get the following Kalman filter [11] forward recursions

$$\begin{aligned} x_{0|0} &= E[x_0], \\ x_{k+1|k} &= Ax_{k|k}, \\ x_{k+1|k+1} &= x_{k+1|k} + K_{k+1}(y_{k+1} - Cx_{k+1|k}), \end{aligned}$$

with the following recursive offline computations

$$\begin{aligned} \Sigma_{0|0} &= var(x_0), \\ \Sigma_{k+1|k} &= A\Sigma_{k|k}A^T + W, \\ \Sigma_{k+1|k+1} &= \Sigma_{k+1|k} - \Sigma_{k+1|k}C^T(C\Sigma_{k+1|k}C^T + V)^{-1}C\Sigma_{k+1|k}, \\ K_{k+1} &= \Sigma_{k+1|k+1}C^T V^{-1}. \end{aligned}$$

Here $K_k \in \mathbb{R}^{m\times p}$ is the time varying Kalman gain matrix and $\Sigma_k \in \mathbb{R}^{m\times m}$ is the covariance matrix of x_k. The recursion for $\Sigma_{k|k-1}$ and $\Sigma_{k|k}$ is called the Riccati equation. The offline computations can be performed before any observations of the output are made. A slightly modified formulation of the Kalman filter is also used in literature [12] where $K_{k+1} = \Sigma_{k+1|k}C^T(C\Sigma_{k+1|k}C^T + V)^{-1}$. In Appendix A, we show that both these formulations are equivalent.

The predicted observation is obtained by the predicted state estimate by

$$y_{k|k-1} = Cx_{k|k-1},$$

and the estimated error covariance matrix $S_k \in \mathbb{R}^{p\times p}$ is given by

$$\begin{aligned} S_k &= E((y_k - y_{k|k-1})(y_k - y_{k|k-1})^T), \\ &= C\Sigma_{k|k-1}C^T + V. \end{aligned}$$

Since we can compute $\Sigma_{k|k-1}$ during the offline computations, we can calculate the estimate of the error covariance before we make observations.

The recursion for $\Sigma_{k|k-1}$, converges to a steady-state value Σ provided that (C, A) is observable and (A, W) is controllable. In that case, Σ satisfies the following algebraic Riccati equation (ARE),

$$\Sigma = A\Sigma A^T + W - A\Sigma C^T(C\Sigma C^T + V)^{-1}C\Sigma A^T.$$

The resulting steady-state Kalman gain matrix is

$$K = \Sigma C^T(C\Sigma C^T + V)^{-1}.$$

The estimate of the hidden variable using the steady-state Kalman gain matrix is generally used when the time horizon for the control is large relative to the effective convergence time of the Riccati equation.

IV. Problem Formulation for Model Learning

In real flow experiments, the flow state variables are high dimensional and the number of sensors is generally very small in comparison, *i.e.* $p \ll m$. However, many fluid flows exhibit low dimensional flow dynamics, especially in the asymptotic limit after the transient in the flow has died out. The reduced order dynamics can be captured with a low rank state transition matrix A. Therefore, in this work we will consider a rank-constrained state transition matrix and an infinite time horizon.

In the infinite time horizon, for a given reduced order model for the hidden dynamics, we can find the steady-state Kalman filtered estimate of the observations using the method shown in Section III. Let Θ be the model parameters of the reduced order model that describes the dynamics of the hidden variables as well as the observation matrix. In Section VI we present the model parameterization Θ used in this study. Our goal is to find the optimal Θ that minimizes the reconstruction error of the predicted observation given by,

$$J(y^k, \Theta) = \frac{1}{2}\sum_{k=0}^{n}(y_k - y_{k|k-1})^T(y_k - y_{k|k-1}). \tag{2}$$

We present the model learning formulation for a single sequence of observations y^n, but the formulation can easily be extended to multiple sequences. The usefulness of using multiple observation sequences for model learning will become evident in Section VIII.

Remark. *In this study we have used the reconstruction error as the objective function. Another choice of objective function is the negative log-likelihood function for the observations given the model parameters* Θ*,*

$$J(y^k, \Theta) = \sum_{k=0}^{n}\frac{1}{2}(\log\det(S) + (y_k - y_{k|k-1})^T S^{-1}(y_k - y_{k|k-1}))$$

with $S = (C\Sigma C^T + V)^{-1}$.

V. Adjoint Equations

Here we present the adjoint equations for the steady-state Kalman filter with the reconstruction error objective function (Eq. (2)). We formulate the Lagrangian for the optimization problem as follows,

$$
\begin{aligned}
\mathcal{L} = J - \eta_{0|0}^T(x_{0|0} - E[x_0]) - \sum_{k=0}^{n-1} \eta_{k+1|k}^T(x_{k+1|k} - Ax_{k|k}) \\
- \sum_{k=0}^{n-2} \eta_{k+1|k+1}^T(x_{k+1|k+1} - x_{k+1|k} - K(y_{k+1} - y_{k+1|k})) \\
- \sum_{k=0}^{n-1} z_{k+1|k}^T(y_{k+1|k} - Cx_{k+1|k}) \\
- \operatorname{tr}(\Lambda^T(\Sigma - A\Sigma A^T - W + A\Sigma C^T S^{-1} C\Sigma A^T)) \\
- \operatorname{tr}(H^T(K - \Sigma C^T S^{-1})) - \operatorname{tr}(T^T(S - C\Sigma C^T - V)).
\end{aligned}
$$

Here, $\eta_{k|k} \in \mathbb{R}^m, \eta_{k+1|k} \in \mathbb{R}^m, z_{k+1|k} \in \mathbb{R}^p, \Lambda \in \mathbb{R}^{m\times m}, H \in \mathbb{R}^{m\times p}$ and $T \in \mathbb{R}^{p\times p}$ are the adjoint variables. Taking variations of the Lagrangian with respect to the Kalman filter variables and setting them to zero, we get the following adjoint equations,

$$
\begin{aligned}
z_{n|n-1} &= -(y_n - y_{n|n-1}), \\
\eta_{n|n-1} &= C^T z_{n|n-1}, \\
z_{k|k-1} &= -(y_k - y_{k|k-1}) - K^T \eta_{k|k}, \\
\eta_{k|k-1} &= \eta_{k|k} + C^T z_{k|k-1}, \\
\eta_{k|k} &= A^T \eta_{k+1|k}, \\
H &= \sum_{k=1}^{n-1} \eta_{k|k}(y_k - y_{k|k-1})^T, \\
W &= S^{-T} C\Sigma^T A^T \Lambda A \Sigma^T C^T S^{-T} - S^{-T} C\Sigma^T H S^{-T}, \\
\Lambda &= A^T \Lambda A - A^T \Lambda A \Sigma^T C^T S^{-T} C - C^T S^{-T} C \Sigma^T A^T \Lambda A \\
&\quad + H S^{-T} C + C^T W C.
\end{aligned}
$$

Even though the Kalman filter recursions are nonlinear, because of the Riccati equation, the adjoint equations are linear. The equation for Λ and W are in the form of coupled generalized Sylvester equations. Reference [13] presents an efficient method to solve such equations. Adjoint equations for the finite time horizon Kalman filter equations are shown in Appendix B.

VI. Model Parameterization

As mentioned in section IV, we assume that the dynamics of the hidden state x is confined to a $r-$dimensional subspace in $\mathbb{R}^m$. Let $L \in \mathbb{R}^{m\times r}$ be a orthogonal basis of this subspace containing the trajectories of x such that $L^T L = I_r$, where I_r is the $r \times r$ identity matrix. We can now reduce the system by the following transformation

$$x_k = La_k,$$

where $a_k \in \mathbb{R}^r$ is the representation of x_k on the space spanned by the columns of L. This leads to a model with $O(mr + r^2)$ parameters. The system of equations now become,

$$
\begin{aligned}
a_{k+1} &= A_r a_k + L^T w_k, \\
y_k &= CLa_k + v_k.
\end{aligned}
\tag{3}
$$

where $A_r \in \mathbb{R}^{r\times r}$ which is much smaller than A and can now be fully-parameterized. Note that the noise covariance due to $L^T w_k$ for the transformed system will be $L^T W L$ instead of W.

We solve for the orthogonal matrix L and the entries of the matrix A_r as well as the mean of the initial condition $E(a_0)$ which we denote by $\bar{a}_0$. With this parameterization, input-output LTI model representation is $(A_r, CL, L^T WL, V)$ with the model parameters $\Theta = (\bar{a}_0, A_r, L)$ with the constraint that $L^T L = I_r$. The space of orthogonal matrices of fixed rank is known as the Stiefel manifold [14]. Stiefel manifold for r-ranked orthogonal matrices in m-dimensional space is denoted as $\mathcal{S}_{m,r}$. Table 1 shows the model parameterization with the resulting input-output equivalence relations.

Governing Equations	Reduced Order Model Parameters	Equivalence Class
$a_{k+1} = A_r a_k + L^T w_k$ $y_k = CLa_k + v_k$	$\Theta = (\bar{a}_0, A_r, L)$ $\bar{a}_0 \in \mathbb{R}^r, A_r \in \mathbb{R}^{r\times r}, L \in \mathcal{S}_{m,r}$	$\Theta_U = (U^T \bar{a}_0, U^T A_r U, LU)$ $U \in \mathcal{O}_r,\ U^T U = I_r$

Table 1: This table shows model parameterization with the corresponding equivalence relations.

Remark. *In case of finite time horizon, the reduced order model parameters will also include the variance of the initial condition of the hidden variable* $\Sigma_{0|0}$.

VII. Riemannian Optimization

In this seciton we describe procedure to perform Riemannian optimization on the model parameter space. The reader is referred to [15] for a detailed review of Riemannian optimization methods. Let $\bar{\mathcal{M}}$ denote the space of the model parameters $\Theta = (\bar{a}_0, A_r, L)$ *i.e.*

$$\bar{\mathcal{M}} := \mathbb{R}^r \times \mathbb{R}^{r\times r} \times \mathcal{S}_{m,r}.$$

The space $\bar{\mathcal{M}}$ is viewed as an embedded manifold in the ambient space $\mathcal{M}_a := \mathbb{R}^r \times \mathbb{R}^{r\times r} \times \mathbb{R}^{m\times r}$ The tangent space at any point $\Theta \in \bar{\mathcal{M}}$, is denoted by $\mathcal{T}_\Theta \bar{\mathcal{M}}$. At any given iteration k, the update rule for the current point $\Theta_k \in \bar{\mathcal{M}}$ is given by the retraction operator $\mathrm{Retr} : \mathcal{M}_a \to \bar{\mathcal{M}}$ so that,

$$\Theta_{k+1} = \mathrm{Retr}(\Theta_k + d\Theta_k), \tag{4}$$

where $d\Theta_k \in \mathcal{T}_{\Theta_k}\bar{\mathcal{M}}$ in the direction of descent with magnitude given by the step size. The action of the retraction operator on any point $(a', A', L') \in \mathcal{M}_a$ is given by

$$\mathrm{Retr}(a', A', L') = (a', A', Q'_L),$$

where Q'_L is the orthogonal matrix given by the QR factorization of the matrix L such that $L' = Q'_L R'_L$. In case of steepest descent, the direction $d\Theta_k$ will be given by the negative of the Riemannian gradient for the objective function defined on $\bar{\mathcal{M}}$. For conjugate gradient, the search direction depends on the previous search direction and the gradient of the objective function at the current iteration.

Computation of the Euclidean derivative using the adjoint variables is shown in Appendix C. Appendix D shows the procedure to get the Riemannian gradient from the Euclidean derivative and Appendix E shows how to perform conjugate gradient based optimization in the model parameter space.

VIII. Learning the Control Matrix

The formulation that is presented so far does not contain a control matrix in the full order hidden dynamical system. We can still infer a reduced order description of the control matrix if we have impulse response data of the dynamical system. Let $B \in \mathbb{R}^{m\times d}$ be the control matrix for the full-order system with d independent actuators. We assume that the system is controllable. The flow system equations are

$$x_{k+1} = Ax_k + Bu_k + w_k,$$
$$y_k = Cx_k + v_k,$$

where x_k, y_k, w_k, v_k, A and C are same as described in Equation (1) and u_k is the control whose action on the state dynamics is realized by the control matrix B.

We need to learn a reduced order description of the flow dynamics as well as the control matrix B. To that end, we first collect impulse responses $\{y_{i,1}, y_{i,2}, \cdots, y_{i,n_i}\}$ of the full-order system of length n_i, for each of actuator $i \in \{1, \cdots, d\}$. We then learn the reduced order model parameters with the given d sequences of observations by minimizing the following cost function

$$J_B(\Theta) = \sum_{i=1}^{d} \left(\frac{1}{2} \sum_{k=0}^{n} (y_{i,k} - y_{i,k|k-1})^T (y_{i,k} - y_{i,k|k-1}) \right).$$

where $\Theta = (\bar{a}_{1,0}, \cdots, \bar{a}_{d,0}, A_r, L)$ are the model parameters. The reduced order description of the control matrix is given by

$$B_r \coloneqq (\ \bar{a}_{1,0} \mid \ \cdots \ \mid \bar{a}_{d,0} \ ).$$

Therefore, by using an objective function that sums over multiple sequences of observations and augmenting the model parameters with initial conditions of the hidden variables for each sequence, we can estimate the reduced order description of the control matrix along with the dynamical matrix A_r and the reduced order subspace L.

IX. Initialization using DMD with Delay Embedding

The optimization problem described in Section IV is non-convex which makes finding the global minima or a solution close to the global minima very challenging using gradient based methods. One way to tackle this problem is to use multiple random initializations and take the best converged solution. However, when the parameter space is huge, a large number of initializations is required to get a good estimate of the global minimum. Another method that is often applied, in particular for combinatorial optimization problems, is to relax some of the constraints of the optimization problem and then massage the solution to get a reasonable solution to the problem that also satisfies all the constraints [16].

In this work we employ a similar approach to relaxation of non-convex constraints. We simplify the optimization problem by assuming that the reduced order hidden states can be represented as linear combination of the delay-coordinates of the observations. More formally, we assume that if a_k is the $r-$dimensional hidden state variable at iteration k, there exists $\tau \in \mathbb{N}$ matrices $L_i \in \mathbb{R}^{p\times r}$, where $i \in [1, \cdots, \tau]$ such that

$$a_k = \sum_{j=1}^{\tau} L_j^T y_{k+j-1},$$

where $\{y_k, \cdots, y_{k+\tau-1}\}$ are τ consecutive observations of the system. Let us define the matrix $L \in \mathbb{R}^{m\times r}$ such that $L^T = [L_1^T | \cdots | L_\tau^T]$ where $m = p\tau$. We can impose orthogonality constraint on the matrix L such that $L^T L = I_r$ without loss of expressivity.

Under this assumption, we can use delay-DMD [17, 18, 19] (also referred to as Hankel-DMD) to get a good initial condition for the model learning optimization problem. Say we have only one sequence of observations y^n from one impulse response of the system. We generate two data matrices $X_1, X_2 \in \mathbb{R}^{m\times(n-\tau)}$ with observations as follows,

$$X_1 = \begin{bmatrix} y_0 & y_1 & \cdots & y_{n-\tau-1} \\ y_1 & y_2 & \cdots & y_{n-\tau} \\ \vdots & \vdots & \ddots & \vdots \\ y_{\tau-1} & y_\tau & \cdots & y_{n-1} \end{bmatrix} \quad \text{and} \quad X_2 = \begin{bmatrix} y_1 & y_2 & \cdots & y_{n-\tau} \\ y_2 & y_3 & \cdots & y_{n-\tau+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_\tau & y_{\tau+1} & \cdots & y_n \end{bmatrix}.$$

Applying a $r-$ranked truncated singular value decomposition of X_1 we get

$$X_1 \approx U_X \Sigma_X V_X^T,$$

where $U_X, V_X \in \mathbb{R}^{m\times r}$ and $\Sigma_X \in \mathbb{R}^{r\times r}$. The initial conditions for the model learning optimization algorithm are given by

$$\bar{a}_0 = U_X^T \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{\tau-1} \end{bmatrix}, \quad A_r = U_X^T X_2 V \Sigma^{-1}, \quad \text{and} \quad L = U_X.$$

Remark. *In case of multiple actuators, say we have observation sequences $\{y_{i,1}, y_{i,2}, \cdots, y_{i,n_i}\}$ of length n_i, for each of actuator $i \in \{1, \cdots, d\}$. The data matrices can be augmented so that*

$$X_1 = \begin{bmatrix} X_{1,1} \mid \cdots \mid X_{1,d} \end{bmatrix}, \quad \textit{and} \quad X_2 = \begin{bmatrix} X_{2,1} \mid \cdots \mid X_{2,d} \end{bmatrix}.$$

where

$$X_1 = \begin{bmatrix} y_{i,0} & y_{i,1} & \cdots & y_{i,n-\tau-1} \\ y_{i,1} & y_{i,2} & \cdots & y_{i,n-\tau} \\ \vdots & \vdots & \ddots & \vdots \\ y_{i,\tau-1} & y_{i,\tau} & \cdots & y_{i,n-1} \end{bmatrix} \quad \textit{and} \quad X_2 = \begin{bmatrix} y_{i,1} & y_{i,2} & \cdots & y_{i,n-\tau} \\ y_{i,2} & y_{i,3} & \cdots & y_{i,n-\tau+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_{i,\tau} & y_{i,\tau+1} & \cdots & y_{i,n} \end{bmatrix},$$

for each $i \in \{1, \cdots, d\}$. The initial condition for the model parameters A_r and L will be the same as before, and for each $i \in \{1, \cdots, d\}$ we have

$$\bar{a}_{i,0} = U_X^T \begin{bmatrix} y_{i,0} \\ y_{i,1} \\ \vdots \\ y_{i,\tau-1}. \end{bmatrix}$$

where U_X is the matrix whose columns are the r leading left singular vectors of the data matrix X_1.

X. Linear Quadratic Gaussian Control

Our goal is to construct an observer-based feedback control law that minimizes the expectation of the quadratic cost given by

$$J_\infty = E\left[\sum_{k=0}^{\infty}(x_k^T Q x_k + u_k^T R u_k)\right].$$

where $Q \in \mathbb{R}^{m\times m}$ and $R \in \mathbb{R}^{p\times p}$ are positive semi-definite matrices.

We can now use the Linear-Quadratic-Gaussian control framework for construct the observer-based feedback control law. For the given reduced order linear system in Equation 3, we can construct the following observer equations

$$\begin{aligned} \hat{a}_{k+1} &= A_r \hat{a}_k + B_r u_k + K_r(y_k - CL(A\hat{a}_k + B_r u_k)) \\ \hat{y}_k &= CL\hat{a}_k, \quad u_k = -G_r \hat{a}_k \end{aligned}$$

where $B_r \in \mathbb{R}^{r\times d}$ is the control matrix, $K_r \in \mathbb{R}^{r\times p}$ is the Kalman filter gain matrix and $G_r \in \mathbb{R}^{d\times r}$ is the control feedback grain matrix. The matrices K_r and G_r are computed by first solving two separate discrete-time algebraic Riccati equations,

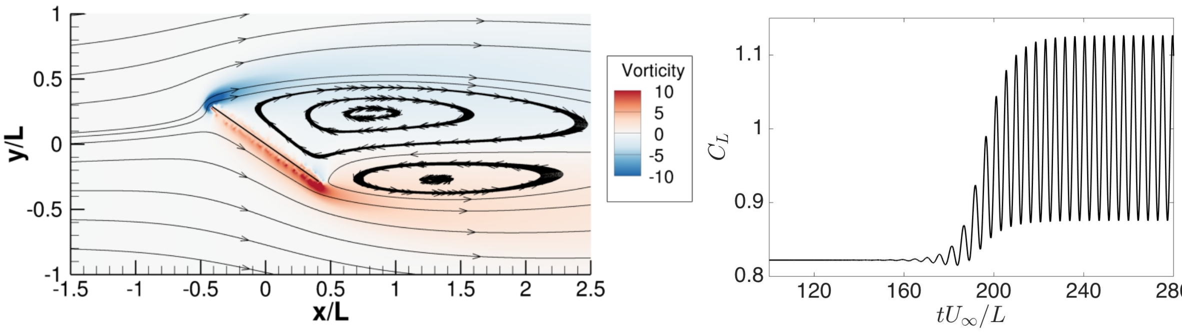
$$\begin{aligned} \Sigma_r &= A_r \Sigma_r A_r^T + L^T W L - A_r \Sigma_r (CL)^T (CL\Sigma_r (CL)^T + V)^{-1} (CL) \Sigma_r A_r^T \\ \Pi_r &= A_r^T \Pi A_r + L^T Q L + - A_r^T \Pi_r B_r (R + B_r^T \Pi B_r)^{-1} B_r^T \Pi A_r. \end{aligned}$$

for $\Sigma_r \in \mathbb{R}^{r\times r}$ and $G_r \in \mathbb{R}^{r\times r}$. K_r and G_r are then given by

$$K_r = \Sigma_r (CL)^T (CL\Sigma_r (CL)^T + V)^{-1} \quad \text{and} \quad G_r = (R + B_r^T \Pi B_r)^{-1} B_r^T \Pi A_r.$$

XI. Results

In this section we control the flow over an inclined flat plate and suppress the vortex shedding in the wake. The freestream flow is at a low Reynolds number of 100 and the flat plate is inclined at an angle of 35°. At these conditions, the flow has been shown to have an unstable steady state [20]. The goal of our data-driven reduced order controller is to keep the system at steady state and prevent periodic vortex shedding in the flow.



(a) (b)

Figure 1: (a) Vorticity contours and velocity streamlines for steady flow over 35° inclined flat plate (b) C_L vs. time for 35° inclined flat plate with the unstable steady state as the initial condition

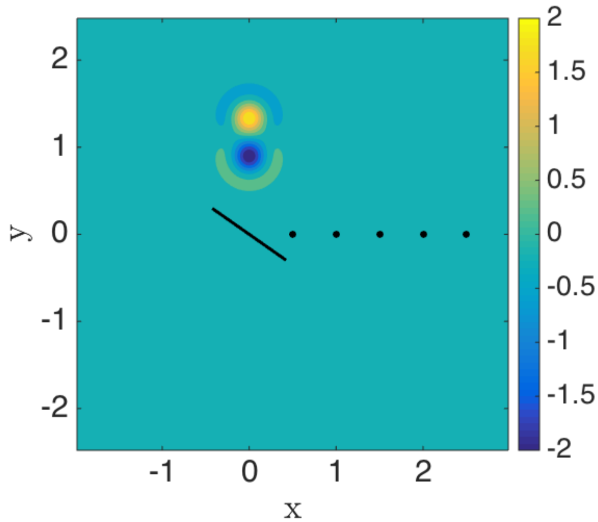


Figure 2: Vorticity contour plot of the impulse response of the actuator. Actuator is modeled as a localized body force placed near the leading edge of the flat plate. The 5 sensor locations are highlighted with black dots. The flat plate is shown by a thick black line.

The flow is simulated using the fast immersed boundary method developed in [21]. In order to achieve uniform flow conditions in the far field, a multi-domain approach is employed. The domain of interest is considered to be embedded in a series of domains, each twice-as-large as the preceding but with the same number of the uniform grid points. The grid size used is 250×250 and the domain of interest is given by $[-2, 3] \times [-2.5, 2.5]$ where the lengths are non-dimensionalized by the chord length of the flat plate, D. The center of the flat plate is located at the origin. Five domains, each with the same number of grid points are used for an effective computational domain that is 2^4 times larger than the domain of interest. The time-step is taken as $dt = 0.01D/U_\infty$ where U_∞ is the freestream velocity.

Figure 1a shows the vorticity contours and velocity streamlines of the steady state while Figure 1b shows the coefficient of lift as instabilities grow with the unstable steady state as the initial condition.

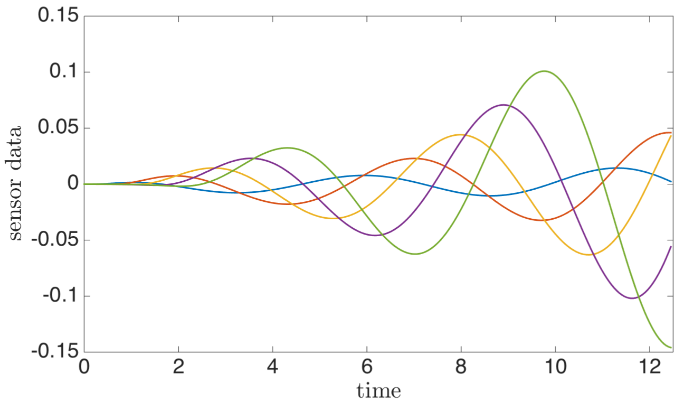


Figure 3: Sensor data from the impulse response of the flow with the steady state initial condition.

We use a single actuator for the control that is as a simple model of localized body force [20, 22] at the actuator location near the leading edge of the flat plate. The instantaneous vorticity field generated by impulse control input of the actuator is

$$B(r) = c[(1 - ar_1^2)\exp(-ar_1^2) - (1 - ar_2^2)\exp(-ar_2^2)]$$

where $r_i^2 = (x - x_{c,i})^2 + (y - y_{c,i})^2$ for $i = 1, 2$. The constants a and c determine the shape and strength of the control, respectively. The actuator parameters used in this study are $x_{c,1} = 0, y_{c,1} = 1.3423, x_{c,2} = 0, y_{c,2} = 0.89, a = 20$ and $c = 2$. Figure 2 shows the vorticity field generated by the actuator relative to the position of the flat plate in the computational domain.

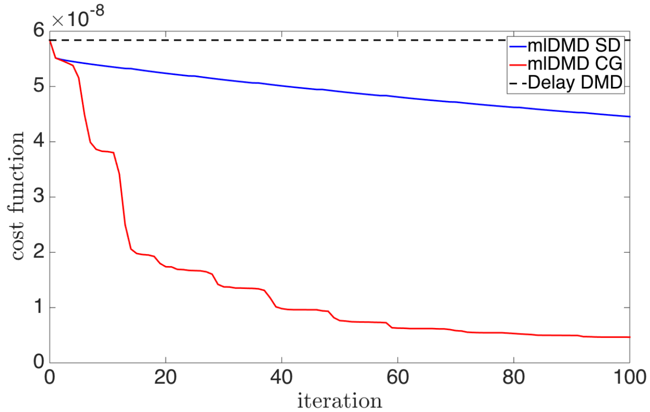
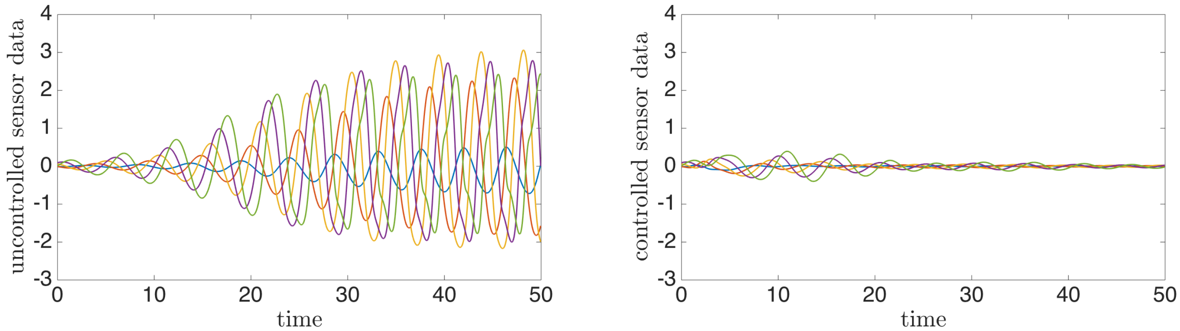


Figure 4: Convergence of mlDMDSD (mlDMD with steepest descent) and mlDMDCG (mlDMD with conjugate gradient) methods with delay-DMD initial condition for learning the reduced order model for the flow over the inclined flat plate.

In [22], we controlled the flow with a data-driven reduced order feedback controller that used full-state observations. These kind of controllers are not always feasible since flow experiment sensors are generally scarce, rendering the system partially observed. In this study, we control the flow using only 5 vorticity point-sensors in the wake of the flat plate. Figure 2 shows the location of the point-sensors relative to the flat plate. Each sensor records the deviation of the instantaneous vorticity value from the steady state value.

We collect sensor data from the impulse response of the flow with the steady state initial condition. Consecutive data points separated by a time-step of $\Delta T = 0.05$. Figure 3 shows the sensor data, which is the deviation of the instantaneous vorticity values from the steady state value of the vorticity at the sensor locations.

We choose the dimension of the hidden state $m = 250$ and rank of the reduced order model $r = 20$. The noise covariance matrices W and V are set to $0.01I_m$ and $0.01I_p$ respectively. The initial condition for the gradient based optimization method mlDMD, is given by using delay-DMD with $\tau = 50$ delay coordinates.



(a) uncontrolled simulation sensor data

(b) controlled simulation sensor data

Figure 5: The sensor data for the (a) uncontrolled and (b) controlled simulation of the flow over inclined flat plate. Figures (a) and (b) have the same scale on the y-axis.

Figure 4 shows the convergence of the mlDMD optimization problem with steepest descent (denoted by mlDMDSD) and conjugate gradient (denoted by mlDMDCG). We use an Armijo-type line search for both the methods with the Fletcher-Reeves update rule for the conjugate gradient direction update for the conjugate gradient method [23]. Steepest descent shows very slow convergence compared to conjugate gradient. We use the solution from the conjugate gradient method after 100 iterations to generate observer-based feedback control for the flow over the inclined flat plate.

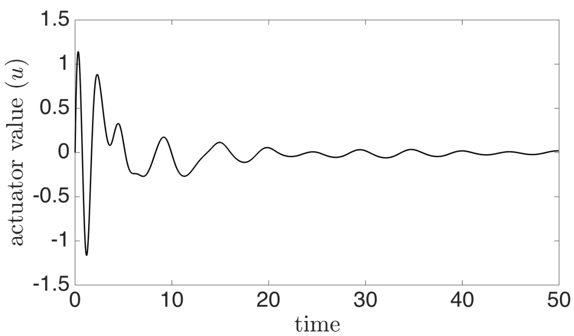


Figure 6: The actuator value for the controlled flow simulation over $t \in [0, 50]$.

Applying the LQG framework as described in Section X with $Q = I_m$ and $R = 10^{13} I_p$, we get a observer and a controller gain matrix for the observer-based feedback controller. The control simulation for the reduced order model inferred using delay-DMD blows up in 70 iterations. The controller for the reduced order model inferred by mlDMDCG, on the other hand, is able to successfully control the flow. Figure 5a and Fig. 5b show the sensor data from the uncontrolled and the controlled simulation of the flow over the inclined flat plate after 1000 iterations ($t \in [0, 50]$). Both the figures have the same y-axis for easy comparison. As can be seen from the comparison, the controller successfully stabilizes the flow and keeps it close to the steady-state value. Figure 6 shows the actuator value for the control simulation for the entire time horizon $t \in [0, 50]$. Even though the actuator oscillated in the beginning for $t < 20$ until it settles around 0 for the rest of the simulation.

XII. Conclusions and Future Work

In this study we propose a novel method, mlDMD, for maximum likelihood estimation based system identification with a low-rank structure applicable for fluid flow systems. Using well-known results from linear estimation theory, we infer a reduced order model that minimizes the reconstruction error for the Kalman filtered estimate of the observations. The adjoint formulation is used to efficiently extract the gradient of the objective function with respect to the model parameters. We present the forward and adjoint equations for both the finite and the infinite horizon Kalman filters. The proposed method can take observations of the impulse response of the system, from one or multiple actuators, and generate a linear reduced order description of the state transition matrix, the low-dimensional subspace that captures the hidden state dynamics and the control matrix.

The optimal controller is then constructed using the Linear-Quadratic-Guassian framework on the inferred reduced order model. The performance of the resulting observer-based feedback controller is demonstrated on the flow over an inclined flat plate at a high angle of attack (35°) and low Reynolds number ($Re = 100$). Under the same conditions, delay-DMD is unable to control the flow while mlDMD is able to suppress the vortex shedding in the wake of the flat plate and keep the flow near the steady state.

There are various avenues for future work to improve and further the capabilities of the current method. The current method requires impulse response observations for the reduced order inference of the control matrix. A fully parameterized control matrix can be included in the model parameters which will allow us to conduct model inference using observations with arbitrary control input. Additionally, the current method assumes that the resulting reduced order model is controllable and observable. In the future, we will add a matrix inequality constraints so that the reduced order model both detectable and stabilizable [24]. Further, this method can also be used to identify Koopman invariant subspaces from linear observations of a nonlinear system. These avenues will be explored in a future publication.

Acknowledgements

This work was sponsored, in part, by the Office of Naval Research (ONR) as part of the Multidisciplinary University Research Initiatives (MURI) Program, under grant number N00014-16-1-2617.

A. Alternate Formulations for the Kalman Filter

We need to show equivalence of $K_{k+1} = \Sigma_{k+1|k+1}C^TV^{-1}$ and $K_{k+1} = \Sigma_{k+1|k}C^T(C\Sigma_{k+1|k}C^T + V)^{-1}$. Using the expression for $\Sigma_{k+1|k+1}$ from the offline Kalman filter recursion we can show,

$$\begin{aligned}
K_{k+1} &= \Sigma_{k+1|k+1}C^TV^{-1} \\
&= (\Sigma_{k+1|k} - \Sigma_{k+1|k}C^T(C\Sigma_{k+1|k}C^T + V)^{-1}C\Sigma_{k+1|k})C^TV^{-1} \\
&= \Sigma_{k+1|k}C^TV^{-1} - \Sigma_{k+1|k}C^T(C\Sigma_{k+1|k}C^T + V)^{-1}C\Sigma_{k+1|k}C^TV^{-1} \\
&= \Sigma_{k+1|k}C^TV^{-1} - \Sigma_{k+1|k}C^T(C\Sigma_{k+1|k}C^T + V)^{-1}(C\Sigma_{k+1|k}C^T + V - V)V^{-1} \\
&= \Sigma_{k+1|k}C^TV^{-1} - \Sigma_{k+1|k}C^T(I - (C\Sigma_{k+1|k}C^T + V)^{-1}V)V^{-1} \\
&= \Sigma_{k+1|k}C^TV^{-1} - \Sigma_{k+1|k}C^TV^{-1} - \Sigma_{k+1|k}C^T(C\Sigma_{k+1|k}C^T + V)^{-1} \\
&= \Sigma_{k+1|k}C^T(C\Sigma_{k+1|k}C^T + V)^{-1}
\end{aligned}$$

Therefore the two formulations are equivalent.

B. Adjoint Formulation for Finite Horizon Kalman Filter

In order to compute the gradient of the cost function with respect to the parameters Θ, we use the adjoint equations. We begin by constructing the Lagrangian as follows

$$
\begin{aligned}
\mathcal{L} = J &- \eta_{0|0}^T(x_{0|0} - E[x_0]) \\
&- \sum_{k=0}^{n-1} \eta_{k+1|k}^T(x_{k+1|k} - Ax_{k|k}) \\
&- \sum_{k=0}^{n-2} \eta_{k+1|k+1}^T(x_{k+1|k+1} - x_{k+1|k} - K_{k+1}(y_{k+1} - y_{k+1|k})) \\
&- tr(\Lambda_{0|0}^T(\Sigma_{0|0} - var(x_0))) \\
&- \sum_{k=0}^{n-1} tr(\Lambda_{k+1|k}^T(\Sigma_{k+1|k} - A\Sigma_{k|k}A^T - W)) \\
&- \sum_{k=0}^{n-2} tr(\Lambda_{k+1|k+1}^T(\Sigma_{k+1|k+1} - \Sigma_{k+1|k} + \Sigma_{k+1|k}C^T(C\Sigma_{k+1|k}C^T + V)^{-1}C\Sigma_{k+1|k})) \\
&- \sum_{k=0}^{n-2} tr(H_{k+1}^T(K_{k+1} - \Sigma_{k+1|k+1}C^TV^{-1})) \\
&- \sum_{k=0}^{n-1} z_{k+1|k}^T(y_{k+1|k} - Cx_{k+1|k}) \\
&- \sum_{k=0}^{n-1} tr(T_{k+1}^T(S_{k+1} - C\Sigma_{k+1|k}C^T - V))
\end{aligned}
$$

where $\eta \in \mathbb{R}^m, \Lambda \in \mathbb{R}^{m\times m}, H \in \mathbb{R}^{m\times p}, z \in \mathbb{R}^p$ and $T \in \mathbb{R}^{p\times p}$ are adjoint variables. Taking variation of $\mathcal{L}$ with respect to the all the variables and setting it to zero gives us first order conditions for the optimal solution,

$$
\begin{aligned}
\eta_{k|k} &= A^T\eta_{k+1|k} \\
\eta_{k|k-1} &= \eta_{k|k} - C^TK_k^T\eta_{k|k} + C^Tz_{k|k-1} \\
\eta_{n|n-1} &= C^Tz_{n|n-1} \\
\Lambda_{n|n-1} &= C^TT_nC \\
H_k &= \eta_{k|k-1}(y_k - y_{k|k-1})^T \\
\Lambda_{k|k} &= A^T\Lambda_{k+1|k}A + H_kV^{-T}C \\
\Lambda_{0|0} &= A^T\Lambda_{1|0}A \\
\Lambda_{k|k-1} &= \Lambda_{k|k} - (\Lambda_{k|k}\Sigma_{k|k-1}^TC^TS_k^{-T}C + C^TS_k^{-T}C\Sigma_{k|k-1}^T\Lambda_{k|k}) + C^TT_kC \\
T_k &= \frac{1}{2}S_k^{-T} - \frac{1}{2}S_k^{-T}r_kr_k^TS_k^{-T} + S_k^{-T}C\Sigma_{k|k-1}^T\Lambda_{k|k}\Sigma_{k|k-1}^TC^TS_k^{-T} \\
T_n &= \frac{1}{2}(S_n^{-T} - S_n^{-T}r_nr_n^TS_n^{-T}) \\
z_{k|k-1} &= -(y_k - y_{k|k-1}) - K_k^T\eta_{k|k}
\end{aligned}
$$

where $r_k = y_k - y_{k|k-1}$. The equations can be solved anti-casually with respect to the forward Kalman filter. The only terms that need to be stored in the reverse pass of the adjoint equations are r_k, $\Sigma_{k|k-1}$ and K_k.

C. Gradient Computation using Forward and Adjoint Solutions

In this section we present the equations of the gradient of cost function with respect to the model parameters $\Theta = (\bar{a}_0, A_r, L)$ in terms of the solution of the adjoint equations. The gradients are

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \bar{a}_0} &= \eta_{0|0}, \\
\frac{\partial \mathcal{L}}{\partial A_r} &= \sum_{k=0}^{n-1} \eta_{k+1|k} x_{k|k}^T + \Lambda A \Sigma^T + \Lambda^T A \Sigma \\
&\quad - \Lambda A \Sigma^T C^T S^{-T} C \Sigma - \Lambda^T A \Sigma C^T S^{-1} C \Sigma, \\
\frac{\partial \mathcal{L}}{\partial L} &= -\sum_{k=0}^{n-2} C^T K_{k+1}^T \eta_{k+1|k+1} x_{k+1|k}^T + \sum_{k=0}^{n-1} C^T z_{k+1|k} x_{k+1|k}^T \\
&\quad + QL\Lambda^T + Q^T L \Lambda - C^T S^{-1} C L \Sigma A^T \Lambda^T A \Sigma - C^T S^{-T} C L \Sigma^T A^T \Lambda A \Sigma^T \\
&\quad + C^T S^{-1} H^T \Sigma + C^T W C L \Sigma^T + C^T W^T C L \Sigma
\end{aligned}$$

The gradient with respect to any particular model parameterization can be computed by using the above gradients.

D. Riemannian Optimization for Stiefel Manifold

We use the procedure outlined in [5] to perform the gradient descent. The reader is refered to [14] and [15] for a detailed review on Riemannian optimization. Since L is an orthogonal matrix, we perform the optimization on the Stiefel manifold $\mathcal{S}_{m,r}$ defined by

$$\mathcal{S}_{m,r} := \{L \in \mathbb{R}^{m\times r} | L^T L = I_r, r \leq m\}$$

The tangents of the Stiefel manifold can be defined by differentiating the constraint $L^T L = I_r$ to get $\Delta^T L + L^T \Delta = 0$, *i.e.* $\Delta^T L$ is skew-symmetric. Additionally, since the Stiefel manifold can be viewed as an embedded manifold in Euclidean space, we choose the standard inner product

$$g_e(dL_1, dL_2) = \mathrm{tr}\left(dL_1^T dL_2\right),$$

where dL_1 and dL_2 are tangent at the same point in the Stiefel manifold. We can define a projection operator [14] that projects any $Z \in \mathbb{R}^{m\times r}$ onto the tangent space at point $L \in \mathcal{S}_{m,r}$ as follows,

$$\pi_T(\Delta) = \Delta - \frac{1}{2} L(L^T \Delta + \Delta^T L).$$

For some function $F : \mathcal{S}_{m,r} \to \mathbb{R}$, let $\nabla F(L)$ be the Euclidean gradient,

$$(\nabla F(L))_{ij} = \frac{dF}{dL_{ij}},$$

then the Riemannian gradient of the function at L is defined as the tangent vector, $\mathrm{grad}\, F(L) \in \mathcal{T}_L \mathcal{S}_{m,r}$ such that,

$$\mathrm{tr}(\nabla F^T dL) = g_e(\mathrm{grad}\, F(L), dL) \;\forall\; dL \in \mathcal{T}_L \mathcal{S}_{m,r}$$

such that $\mathrm{grad}\, L^T L$ is skew-symmetric. This gives the following relation between Euclidean and Riemannian gradients,

$$\mathrm{grad}\, F(L) = \nabla F(L) - \frac{1}{2} L(L^T \nabla F(L) + \nabla F(L)^T L) = \pi_T(\nabla F(L)).$$

Remark. *In this work we have used the standard inner product to define the metric on the Stiefel manifold. The same procedure described here can be followed to derive the gradient for the canonical metric on the Stiefel manifold [14].*

E. Conjugate Gradient Method

Let $\bar{\mathcal{M}}$ denote the full parameter space for the model parameters $\Theta = (\bar{a}_0, A_r, L)$, *i.e.*

$$\bar{\mathcal{M}} := \mathbb{R}^r \times \mathbb{R}^{r\times r} \times \mathcal{S}_{m,r}.$$

Let $\mathcal{O}_r$ denote the space of real unitary matrices of size $r \times r$. Then for any arbitrary matrix $U \in \mathcal{O}_r$, $\Theta = (\bar{a}_0, A_r, L)$ and $\Theta_U = (U^T\bar{a}_0, U^T A_r U, LU)$ represent equivalent input-output systems. Therefore we endow $\bar{\mathcal{M}}$ with the equivalence relation $\sim$ so that $(\bar{a}_{0,1}, A_{r,1}, L_1) \sim (\bar{a}_{0,2}, A_{r,2}, L_2)$ if and only if there exists a $U \in \mathcal{O}_r$ such that

$$(\bar{a}_{0,1}, A_{r,1}, L_1) = (U^T\bar{a}_{0,2}, U^T A_{r,2} U, L_2 U).$$

Now let us define the equivalent case as follows,

$$[\Theta] = \{\Theta_1 \in \bar{\mathcal{M}} | \Theta_1 \sim \Theta\}.$$

Let $pi : \bar{\mathcal{M}} \to \mathcal{M}$ be the canonical projection such that $\pi(\Theta) = [\Theta]$ for any $\Theta \in \bar{\mathcal{M}}$. The vertical space is defined as $\mathcal{V}_\Theta := \mathcal{T}_\Theta \pi^{-1}([\Theta])$. We want to avoid moving in this space since it is tangential to the equivalent class orbit.

To characterize the vertical space, consider a curve $U(t) \in \mathcal{O}_r$ defined for $t \in \mathbb{R}$ and $U(0) = I_r$. All the points $\Theta(t) = (U(t)^T\bar{a}_0, U(t)^T A_r U(t), LU(t))$ are equivalent. Differentiating on both sides with respect to t, we get

$$\dot{\Theta}(t) = (\dot{U}(t)^T\bar{a}_0, \dot{U}(t)^T A U(t) + U(t)^T A\dot{U}(t), L\dot{U}(t)).$$

where $\dot{U}(t) \in \mathcal{T}_{U(t)}\mathcal{O}_r$. This means that $\dot{U}(t)^T U(t) =$ skew-symmetric at each t. Evaluating at $t = 0$, we get the vertical space defined at Θ as

$$\mathcal{V}_\Theta = \{(U'^T\bar{a}_0, U'^T A + AU', LU') | U' \in \mathbb{R}^{r\times r} \text{ and } U' \text{ is skew-symmetric}\}.$$

We need a projection operator that, for any given point Θ, can act on any point in $\bar{\mathcal{M}}$ and project it onto the orthogonal complement of the vertical space $\mathcal{V}_\Theta$. The orthogonal complement space of the vertical space is called the horizontal space and is denoted by $\mathcal{H}_\Theta$. Let (a', A', L') be a point in $\mathcal{H}_\Theta$, then orthogonality constraint requires,

$$\begin{aligned} &\operatorname{tr}(a'^T U'^T\bar{a}_0 + A'^T(U'^T A + AU') + L'^T(LU')) = 0 \\ \implies &\operatorname{tr}(U'(a'a_0^T + A'^T A + A'A^T + L'^T L)) = 0 \end{aligned}$$

for all skew-symmetric matrices $U' \in \mathbb{R}^{r\times r}$. From this we can conclude that,

$$a'a_0^T + A'^T A + A'A^T + L'^T L = \text{symmetric matrix.}$$

We can now construct the operator $\pi_H : \bar{\mathcal{M}} \to \mathcal{H}_\Theta$ such that for $\Theta = (\bar{a}_0, A, L)$ we have

$$\pi_H((a', A', L')) = (a' - U^T\bar{a}_0, A' - U^T A - AU, L' - LU)$$

where U is given by the solution of the following equation,

$$\Omega + U\Xi + \Xi U - 2\Gamma(U) = 0$$

where

$$\begin{aligned} \Omega &= a'\bar{a}_0^T - \bar{a}_0 a'^T + A'^T A + A'A^T - A^T A' - AA'^T + L'^T L - L^T L' \\ \Xi &= A^T A + AA^T + \bar{a}_0\bar{a}_0^T + I_r \\ \Gamma(U) &= A^T U A + AUA^T \end{aligned}$$

This equation for U is linear and in the generalized Sylvester equation form. The conjugate gradient update for the search direction $d\Theta_k$ at iteration k is given by

$$d\Theta_k = -\alpha_k \operatorname{grad} F(\Theta_k) + \beta_k \pi_H(d\Theta_{k-1}).$$

where $F : \mathcal{M} \to \mathbb{R}$ is the objective function, α_k is the step-size and β_k is computed using the Fletcher-Reeves formula. Equation 4 can then be used to move towards the minima of the objective function.

References

[1] Peter J Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5–28, 2010.

[2] Mihailo R Jovanović, Peter J Schmid, and Joseph W Nichols. Sparsity-promoting dynamic mode decomposition. *Physics of Fluids*, 26(2):024103, 2014.

[3] Paul J Goulart, Andrew Wynn, and David Pearson. Optimal mode decomposition for high dimensional systems. In *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, pages 4965–4970. IEEE, 2012.

[4] Kevin K Chen, Jonathan H Tu, and Clarence W Rowley. Variants of dynamic mode decomposition: boundary conditions, Koopman, and Fourier analyses. 2011.

[5] Palash Sashittal and Daniel J Bodony. Low-rank dynamic mode decomposition using riemannian manifold optimization. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 2265–2270. IEEE, 2018.

[6] Maziar S Hemati, Clarence W Rowley, Eric A Deem, and Louis N Cattafesta. De-biasing the dynamic mode decomposition for applied koopman spectral analysis. *arXiv preprint arXiv:1502.03854*, 2015.

[7] Lawrence R Rabiner and Biing-Hwang Juang. An introduction to hidden markov models. *ieee assp magazine*, 3(1):4–16, 1986.

[8] Khaled Ibrahim Yared and Nils Richard Sandell. Maximum likelihood identification of state space models for linear dynamic systems. 1978.

[9] Herbert E Rauch, CT Striebel, and F Tung. Maximum likelihood estimates of linear dynamic systems. *AIAA journal*, 3(8):1445–1450, 1965.

[10] Hiroyuki Sato and Kazuhiro Sato. Riemannian optimal system identification algorithm for linear mimo systems. *IEEE control systems letters*, 1(2):376–381, 2017.

[11] Greg Welch, Gary Bishop, et al. An introduction to the kalman filter. 1995.

[12] Robert Grover Brown, Patrick YC Hwang, et al. *Introduction to random signals and applied Kalman filtering*, volume 3. Wiley New York, 1992.

[13] Feng Ding and Tongwen Chen. On iterative solutions of general coupled matrix equations. *SIAM Journal on Control and Optimization*, 44(6):2269–2284, 2006.

[14] Alan Edelman, Tomás A Arias, and Steven T Smith. The geometry of algorithms with orthogonality constraints. *SIAM journal on Matrix Analysis and Applications*, 20(2):303–353, 1998.

[15] P-A Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.

[16] Tong Zhang. Analysis of multi-stage convex relaxation for sparse regularization. *Journal of Machine Learning Research*, 11(Mar):1081–1107, 2010.

[17] Jonathan H Tu, Clarence W Rowley, Dirk M Luchtenburg, Steven L Brunton, and J Nathan Kutz. On dynamic mode decomposition: theory and applications. *arXiv preprint arXiv:1312.0041*, 2013.

[18] Hassan Arbabi and Igor Mezic. Ergodic theory, dynamic mode decomposition, and computation of spectral properties of the koopman operator. *SIAM Journal on Applied Dynamical Systems*, 16(4):2096–2126, 2017.

[19] Kathleen P Champion, Steven L Brunton, and J Nathan Kutz. Discovery of nonlinear multiscale systems: Sampling strategies and embeddings. *SIAM Journal on Applied Dynamical Systems*, 18(1):312–333, 2019.

[20] Sunil Ahuja and Clarence W Rowley. Feedback control of unstable steady states of flow past a flat plate using reduced-order estimators. *Journal of Fluid Mechanics*, 645:447–478, 2010.

[21] Tim Colonius and Kunihiko Taira. A fast immersed boundary method using a nullspace approach and multi-domain far-field boundary conditions. *Computer Methods in Applied Mechanics and Engineering*, 197(25-28):2131–2146, 2008.

[22] Palash Sashittal and Daniel J Bodony. Reduced-order control using low-rank dynamic mode decomposition. *Theoretical and Computational Fluid Dynamics*, pages 1–21, 2018.

[23] Li Zhang, Weijun Zhou, and Donghui Li. Global convergence of a modified Fletcher–Reeves conjugate gradient method with armijo-type line search. *Numerische Mathematik*, 104(4):561–572, 2006.

[24] Geir E Dullerud and Fernando Paganini. *A course in robust control theory: a convex approach*, volume 36. Springer Science & Business Media, 2013.