博 士 学 位 論 文
Doctoral Thesis

論文題目

Thesis Title

Reduced-order Reconstruction of Flow Field with Optimized Sparse Sampling and Data-driven Model

東 北 大 学 大 学 院 工 学 研 究 科

Graduate School of Engineering,

TOHOKU UNIVERSITY

専攻/Department: Department of Aerospace Engineering

学籍番号/ ID No: C1TD1615

氏 名 /Name: Keigo YAMADA

Tohoku University

Graduate School of Engineering

Reduced-order Reconstruction of Flow Field with Optimized Sparse Sampling and Data-driven Model

（最適化されたスパースサンプリングとデータ駆動型モデルによる流れ場の低次元再構成）

A dissertation submitted for the degree of Doctor of Philosophy (Engineering)

Department of Aerospace Engineering

by

Keigo Yamada

February 14, 2024

Reduced-order Reconstruction of Flow Field with Optimized Sparse Sampling and Data-driven Model

Keigo YAMADA

Abstract

A systematic method is developed in this research to reconstruct a flow field from only a few point measurements. The presented formulations give an estimate of flow fields instead of spatially resolved measurements of the entire field, which is typically not feasible in real-world applications. Existing frameworks for state estimation and model construction from a sparse sensing system are extended to achieve more accurate and efficient reconstruction of flow fields.

Main contributions posed in the series of work are threefold: First, the reconstruction of a flow field is conducted with a given linear reduced-order model and spatially partial measurements. The upshot is that these sparse measurements are effectively designed by efficient algorithms based on the model property, the observability Gramian. The observability Gramian is used as an optimization objective that denotes the error covariance of the state estimation provided the linear development of the state. A scalar evaluation is obtained by the matrix determinant of the Gram matrix, and an ideal sparse measurement is sought by maximizing this quantity. Here, the model is obtained beforehand from measurement data of the entire field, by means of such as a visualization of the planar distribution of a flow field and its singular value decomposition. The proposed two algorithms improved the expression of computational cost compared with some state-of-the-art methods, a simple greedy selection and a continuous relaxation with semidefinite constraints. Estimation of the state of a dynamical system is improved by selecting appropriate measurement positions based on the state-space representation.

In the second perspective, sensing methods are investigated in the following chapters to construct a flow model from spatially limited measurements. It is sometimes infeasible to obtain or process sufficient data for model construction of flow field, despite the fact that existing modeling methods assume this procedure as employed in the optimization framework. Therefore, a linear dynamical representation of a flow field is reconstructed from sparse measurements using compressive sensing techniques to address the lack of spatially resolved field measurements, previously known as Compressed Sensing Dynamic Mode Decomposition (CS-DMD). Some techniques are integrated with the previous plain method of CS-DMD to incorporate domain knowledge of fluid

dynamics. The proposed method leverages structured sparsity based on some principal features of fluid flow, where the inference of compressive sensing is carried out within subspaces relative to a flow field. The ability to reconstruct flow structures is enhanced compared to some earlier results by combining the use of reduced basis vectors and the promotion of a cluster distribution in the parameter space.

The final contribution is provoked by incorporating measurement optimization into model acquisition from sparse measurements. The former approach of CS-DMD is based on the randomly distributed sparse measurements, which is a usual assumption for compressive sensing theories. However, this conflicts with the proposed inventions for CS-DMD in that unnecessarily fine-scale measurements are considered by comprising distributed sensing positions. This chapter deals with two distinctive types of measurement optimization for the preceding method of CS-DMD, namely the predefined scheme and the adaptive one. Information from the selected sensors strengthens the reconstruction of a spatial distribution with respect to the selected basis vectors introduced in the preceding chapter. Measurement optimization techniques are first smoothly integrated to construct a measurement system that focuses on a small portion of the basis vectors. An approximate Bayesian formulation then gives a posterior evaluation of the inference and the measurement system itself. This leads to the evaluation of additional measurements to mitigate uncertainty in the model estimation.

Acknowledgments

I wish to express my sincere gratitude to those who supported the completion of this research project.

First of all, I am immensely grateful to my advisor, Prof. Taku Nonomura, who led me to this tremendously fantastic research field. It was an honor to be part of his research team and to have worked diligently with Dr. Takayuki Nagata, Dr. Yuji Saito, Dr. Yuta Ozawa, Dr. Yasuo Sasaki, Dr. Kumi Nakai, Prof. Shun Takahashi and Dr. Masahito Watanabe. All the insightful feedback and discussions raised in the team have turned my research intriguing and meaningful work.

Kudos to the lab members for their relentless hard work and passionate research activities. All the achievements of the members have enormously inspired my motivation. I would also wish to thank Prof. Keisuke Asai, Prof. Naofumi Ohnishi and Ms. Mariko Shiroto for their continuous support and encouragement to us.

I would also like to thank the committee members, Prof. Yuji Hattori and Prof. Mitsuhiro Hayashibe, for their thoughtful comments and valuable suggestions.

Finally, my deepest gratitude is dedicated to my family and friends, who let their compassion and help wherever I am. Words cannot describe how much I want to thank them for their support that uplifted me.

February 14, 2024

Keigo Yamada

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter begins with a brief introduction of the key considerations for research, such as "modeling of natural phenomena" and "actual measurements." In response to several issues raised there, the essence of research is explained in Section 1.2.

1.1 Background

1.1.1 Description of flow by modeling

Completion of a phenomenon requires analyzing and predicting the behavior of the target. Therefore, it has been regarded essential to gather information from measurements and to construct a model that reasonably explains the obtained measurements. These actions, measurement and modeling, can be considered complementary because we build models from the pooled measurements and interpret measurements by reference to models.

Many physical processes have dynamical mechanisms that produce temporal evolution with certain spatial patterns. Mathematical models have been used to encompass the quantitative description by numerically explaining the relationship between quantities. Models for such phenomena are often expressed by differential equations, while eliminating many factors that should be respected in practice. Nonlinearity adds significantly to the complexity of such models. Fluid flows are one of the most influential, yet overwhelmingly complex, physical phenomena. We are surrounded by many types of fluid flows, ranging from winds that pass through the woods, blood in the veins, and convection in the mantle far below the ground. Many insights have been given

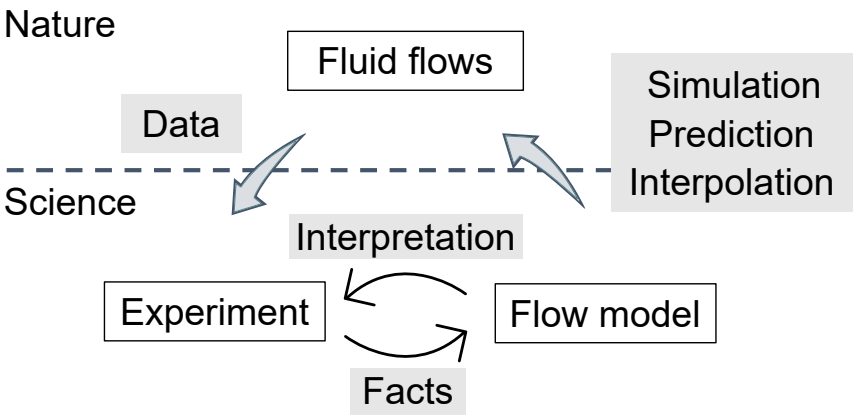


FIGURE 1.1: Relationship between nature and science

into these familiar phenomena since centuries ago, as found in famous sketches by Leonard da Vinci (Marusic and Broomhall, 2021). Air flows play a crucial role in numerous engineering fields and have long been investigated. The atmosphere is almost transparent, and the motion of air flow is difficult to observe directly. The nature of such systems is intensely dictated by means of visualization experiments and numerical simulations. Fluid models have been greatly developed by combining these means, as illustrated in Fig. 1.1.

Even though mathematical models can represent intrinsic physics, the benefit of such models is sometimes limited due to computational complexity or insufficient predictability. Data-driven modeling methods have become one of the fascinating approaches that provide a tractable reduced-order model from data obtained by measurements or simulations (Brunton et al., 2020; Glauser, 2017; Mendez et al., 2023, January). It has been even more important to efficiently acquire parsimonious models, which is represented by a superposition of modal expressions. This is sometimes related with the Koopman operator theory (Mauroy et al., 2020). Among the growing interest in computational fluid dynamics, data-driven methods of order reduction for the differential equation were developed, which are represented by proper orthogonal decomposition and related Galerkin projection. The resulting models capture the principal behavior of the phenomena of interest as low-dimensional models (Asztalos et al., 2023; Berkooz et al., 1993; Ramezanian et al., 2021; C. W. Rowley and S. T. Dawson, 2017; C. W. Rowley et al., 2003,

2004). For analyzing unknown dynamics in the data, the method of dynamic mode decomposition (DMD) (Kutz et al., 2016b, November; Schmid, 2010) is a powerful technique that extracts low-dimensional structures of spatiotemporal coherence by simple computation of the data matrix (a brief schematic is provided as Fig. 3.1), in conjunction with order reduction by proper orthogonal decomposition (POD) (Brunton and Kutz, 2019, January). These order reduction techniques are further extended to invoke nonlinear formulations, including projection on nonlinear basis (Cenedese et al., 2022; Ponsioen et al., 2018; Skene et al., 2022) and parametric regression (Brunton et al., 2016b; Heide et al., 2022; Loiseau et al., 2018).

In accordance with the intensive recent development of machine learning methods, there have been many attempts to leverage data-driven approaches for fluid dynamics. Use of autoencoder models with neural network is one of the promising method to cope with the nonlinearity underlying fluid phenomena (Maulik et al., 2021; Murata et al., 2019; Otto et al., 2022; C. W. Rowley and S. T. Dawson, 2017). As represented by these intense studies, there is no doubt that its effectiveness has been widely recognized. Meanwhile, the problem is the enormous amount of computation required by the learning process. It is also true that, at present, the performance of these data-driven methods can hardly be guaranteed (Antonelli et al., 2022; Lenaerts and Ginis, 2022). As explainability is often crucial for the validity of the learned model, another fascinating approach is to use physics-informed learning to construct a surrogate model that is more compatible with physical models. (Cai et al., 2021; Raissi et al., 2019; Yasuda et al., 2022). A similar perspective has gained popularity as physics-informed DMD (Baddoo et al., 2023). Its formulation promotes parsimonious linear models while abiding by fundamental properties of physics such as energy conservation or several symmetries. Anyway, it should be generally accepted that there is no complete ready-made model for complex phenomena like fluids.

1.1.2 Model-based flow control

As it has been hundred years since aircrafts and automobiles became a reality, there exist a lot of machines around our lives that interact with fluids. Still, the development of these machines requires much work being done to improve the performance. Greater efficiency has become of

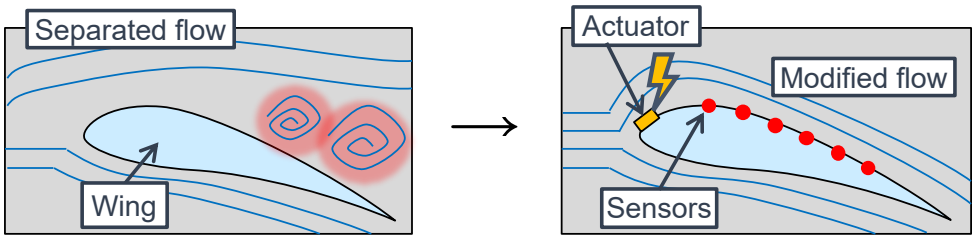


FIGURE 1.2: Schematic of flow control

intense interest due to the climate crisis in recent years in addition to improving safety and expanding its operating limits. The development of machine shapes and powering has been widely carried out, while an intriguing approach is to control the interaction between machine and fluid directly using actuators. Active feedback control is one of the developing methods in aerospace engineering that manages surrounding airflow to a desirable state. The input of momentum by actuators as Fig. 1.2 is used to achieve desirable flows states (Fujii, 2014; Kreth and Alvi, 2020).

From a technological perspective, a model transforms itself into a predictor that tells the future behavior of the target from information on the current state. As an accurate model enables the determination of the effective output of the control, the accuracy of the model plays a crucial role for greater performance of the control. Therefore, the construction of an accurate model is a fundamental task to complete the control of fluid flows. Moreover, the predictions of the model need to be obtained in a very short time, and thus, it is desirable that the computations using the model can be performed effortlessly. This is because the evolution of airflow is substantially complex and highly unsteady that changes significantly over a short period of time. This characteristic is described by nonlinear partial differential equations in the physical model of flow, as mentioned earlier. It should be mentioned that applying well-known model-based approaches to the control design tends to be prohibitive, especially for large-scale and high-speed flows. The key to a successful control scheme is to establish an efficient model that effectively anticipates the future state with less computational resource.

Accordingly, obtaining a simplified, yet meaningful, model is regarded essential for real-world applications of control. As introduced in the previous section, data-driven approaches

have established new possibilities for the description of phenomena. There are recent attempts that incorporate the perspective of control into these data-driven models. Surrogate models by those methodologies are leveraged in the analysis of the system and the construction of an effective control rule (Kaiser et al., 2021; Mauroy et al., 2020; Mendez et al., 2023, January; C. W. Rowley and S. T. Dawson, 2017). It was shown that the state of a chaotic system can be held to a certain region by means of data-driven techniques with an input-output history of a state-space model (Bai et al., 2020; Bhattacharjee et al., 2020; Kaiser et al., 2018; Korda and Mezić, 2018; Proctor et al., 2016). Those using machine learning approaches are also found to be effective (Ahmed et al., 2023; Loiseau et al., 2018; Sasaki and Tsubakino, 2020; Shimomura et al., 2020), in addition to nonlinear order reduction (Cenedese et al., 2022). Although these methods are indeed effective, it is difficult to guarantee their coverage and control limits. It strengthens the difficulty, especially in aerospace engineering, that emphasizes safety under critical conditions (Brunton et al., 2021).

In this way, accurate models lead to effective control because such models extract the fundamental relationship between quantities. A methodology of model construction increases its importance, for the model construction is not straightforward as mentioned. There are data-driven control methods, indeed, that directly learn the relationship between control input and observed output without being interpreted by models (da Silva et al., 2019; Kastsiukevich and Dmitruk, 2020; Paris et al., 2021; Van Waarde et al., 2023). These black-box models may outperform the model-based, either white- or gray-box models. However, these performances are comprehensively confirmed mainly in the linear framework, and those for nonlinear extensions are limited. The learning process also requires a lot of computations; therefore, the application to a real-world phenomenon demands an efficient methodology.

1.1.3 Efficient measurement with compressed sensing

Recent advances in data science techniques have enabled us to extract reduced-order models from a vastly large-scale measurement of complex phenomena (Brunton and Kutz, 2019, January; Kutz et al., 2016b, November). However, high-resolution data acquisition is often not available,

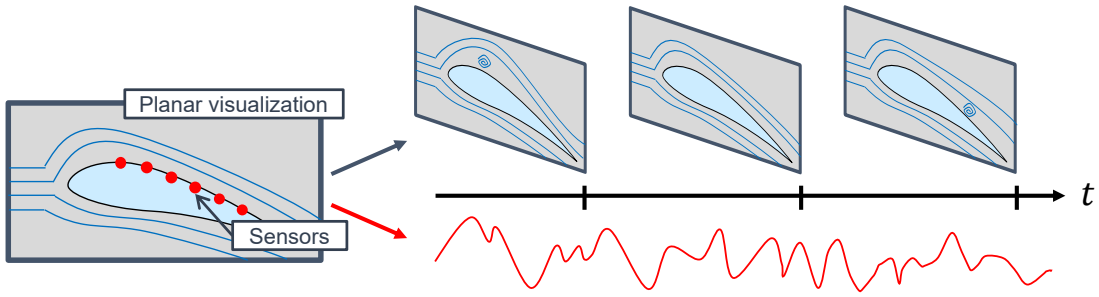


FIGURE 1.3: Schematic of the current trade-offs between spatio-temporal resolution of the flow measurement

even in an ideal environment in laboratory experiments. In general, there exist onerous trade-offs between resolutions in space and time as in Fig. 1.3. Collecting information from point measurements of electrical sensors is often the only viable approach for estimating the internal state or unobservable physical quantities. These apparatuses can improve temporal resolution but sacrifice spatial resolution instead. Increasing the degree of freedom in space is engaged by visualization experiments (I. Grant, 1997; Sugioka et al., 2019), which are often utilized for the generation of learning data for reduced-order modeling (Loiseau et al., 2018; Nonomura et al., 2021c). Still, the increase in information is strictly limited by the hardware requirements, where acquisition is often limited to some planar measurements. A similar problem is posed in the data processing of numerical simulations, where the computational cost of the enormous data for model construction is excessive in addition to that of the high-fidelity simulations for various conditions. The framework of data-driven model construction becomes even more unrealistic in control applications, where it is desirable to evenly sample the complex behavior excited by control input.

Therefore, the use of spatially compressed measurement is one of the versatile strategies to cope with the following troublesome situations.

- Too vast data for efficient analysis
- Insufficient data for informative representation

These situations are widely treated in the fields of signal processing and information technology within a framework called compressive sensing (CS) (R. G. Baraniuk et al., 2010; Candès et al.,

2006; Elad, 2010). The same data compression technique is found in a data-driven method for the construction of dynamical models (Brunton et al., 2014; Erichson et al., 2019; Fowler, 2009; Ohmichi et al., 2018). A model is constructed with compressed data by sketch projection; then the dimension is retrieved by an inverse operation that expands the sketched vectors to the original dimension. On the other hand, several methods aim to construct models from data with insufficient spatial or temporal resolution, where the data contain missing information and must be reconstructed (Bai et al., 2020; Brunton et al., 2015; Guéniat et al., 2015; Sankaranarayanan et al., 2013; Waters et al., 2011).

An intriguing approach for the latter situation is known as compressed sensing DMD of Brunton et al., 2015, which is called CS-DMD in this study. This approach incorporated the notion of compressive sensing into DMD modeling, where spatially insufficient measurement is only available ingredients for modeling. The DMD modes for complete data are represented by a sparse representation vector with respect to a certain representation basis. This formulation is technically clear, but there remains some redundancy in the estimation. This is based on the pure methodology of compressive sensing. Consequently, reconstruction of CS-DMD requires a considerable number of measurements to achieve sufficient quality in reconstruction. An important insight is, as introduced in Section 1.1.1, that the model-based approach may work efficiently instead of the pure data-driven assumptions. Especially, if the model of CS-DMD matters for applications such as flow control, we already have plenty of knowledge about its dynamics.

Until now, numerous methods have been presented to mitigate the abundance of spatial distribution, which reconstruct the distribution from a sparse and limited number of point measurements (Fukami et al., 2023; Joshi and Boyd, 2009; Saito et al., 2021; Wu et al., 2023; Yasuda et al., 2022). This is accomplished by a predetermined model that dictates the relationships between quantities. The values of unobserved quantities are interpolated using the sparse measurements. Measurement at optimized sensing positions is gaining its importance since this approach explicitly depends on the measurements. Several earlier studies elucidate measurement optimization, where redundancy is reduced based on the model. It is utilized in the sparse regression of compressive sensing, in both predefined and adaptive schemes (DeVore, 2007; Duarte

and Eldar, 2011; Joneidi et al., 2020; L. Yu et al., 2012). This procedure can be highlighted from another perspective, that is, the design of compressive measurement must consider the physical implementation of data acquisition. This is because a general framework of CS postulates that the measurement is resulted from random projection to satisfy several convenient properties. It is also applicable for data-driven methodologies; for example, optimization takes place in the estimation of complex phenomena from sensor measurements (Brunton et al., 2014; Loiseau et al., 2018; Manohar et al., 2018; Otto and C. W. Rowley, 2022; Saito et al., 2021), model-free machine learning (Carter et al., 2021; Paris et al., 2021), data-driven control (Herrmann et al., 2023) and correction for data assimilation (Misaka and Obayashi, 2014; Mons and Marquet, 2021; Yoshimura et al., 2020). Consequently, the actual implementation of sparse measurements requires a sophisticated strategy to arrange sensing positions.

1.2 Organization of the paper

1.2.1 Motivation and objective of the study

It is undoubtedly substantial to develop methodologies using sensor measurements for state estimation to achieve control over fluid flow. However, there are supposed to be several developments that need to be made, as listed below.

- An estimation method tailored for the reconstruction of fluid phenomena by exerting compressive sensing
- A systematic method to construct measurement systems, such as sensor arrays mounted on the surface of aircrafts

Spatial information from sensor measurements is often more limited compared to temporal information. One may rely on to state estimation to infer unobserved quantities based on sensor measurements and an assumed model. Since the quality of inference is explicitly dependent on the model and the measurements, it is important to develop a sophisticated method for this type of state estimation. The first key question is how to determine the sensing positions to obtain information from the measurements. Several optimization problems have thus far been proposed

based on the models employed to identify beneficial positions to take measurements. From a consideration to the flow control, a linear dynamical form is one of the realistic options to deal with a large-scale system such as fluid flows. These optimization tasks for dynamical models are still limited to relatively small systems. Therefore, an optimization method may be needed that is applicable to a large-scale model that represents fluid flows. The model itself is, however, sometimes not available in advance, which is due to difficulties in collecting sufficient data to construct an informative model. The other question posed in this study is how to efficiently construct a dynamical model from limited sensor measurements. Numerous modeling methods have been developed to do this using limited measurements based on compressive sensing, and an earlier method using DMD is a promising approach. Still, as introduced in the following chapters, there is room for improvement for this method to be implemented for real-world applications.

The common attitude under the study is to extend the ability of limited sensor measurements to obtain a field of distributed physical quantities, which is represented by a field of fluid flows. The main objective of this study is thus proposed to cope with the above claims:

> To develop effective method of flow field reconstruction using sparse sampling and data-driven reduced-order model

The methods introduced in the succeeding chapters are model-based approaches, which reflect either quantitative or qualitative models. The method for the measurement optimization that will be introduced first emphasizes a systematic evaluation of measurement points based on the given model. On the other hand, the methods of model construction in the subsequent chapters are rather considered in the latter perspective, where the methods incorporate the abstract characteristics of the target phenomenon. Several elements of essential fluid motion are characterized and treated as seen in the series of studies by Da Vinci. The proposed techniques gather existing methods in separated research areas and provoke a novel point of view for them, as briefly mentioned in the following section.

1.2.2 Structure of the main body

This paper is organized with three research topics, each of which focuses on a different point of view of the information available for the reconstruction of flow field. They are briefly outlined with several statements about the contributions presented in each chapter.

A sparse measurement is performed in Chapter 2, **Selection of Measurement Position for State Estimation of Linear Dynamical Systems**, under a given field representation by a linear dynamical system. A classical linear inverse problem and related measurement optimization problem are enhanced to the dynamical configuration. Optimization of the positions of point measurement improves the performance of state estimation assuming the linear development of the state without noises in measurement and system evolution. The development of efficient algorithms is another contribution of this study.

As the second perspective, methods of flow field modeling are investigated where spatially limited measurements are only available material for the construction of a flow field model. This chapter and the following assume the absence of a predetermined model that was available in Chapter 2. Improved methods for sparse regression in Chapter 3, **Construction of Linear Flow Model from Compressive Measurement**, are based on Dynamic Mode Decomposition (DMD) from spatially sparse measurements, which is referred to as CS-DMD (compressed sensing DMD). The method imposes a sparse reconstruction using measurements at randomly arranged positions. The first approach therein is to prune in advance the basis vectors used for compressive regression. They are limited to those relevant to a representation of fluid behavior. Another proposal is to determine the regression coefficients by engaging structured sparsity in compressive estimation. These proposed techniques enhance the ability of compressive reconstruction of DMD modes, by reducing the required number of measurements compared to those of the earlier results of the original implementation of CS-DMD.

The final topic in Chapter 4, **Optimized Measurement Position for Compressive Flow Modeling**, wraps up the doctoral research by incorporating measurement optimization into model acquisition in Chapter 3. The methods prescribe a remedy for onerous claims of CS that

force the use of randomly arranged measurements for the reconstruction from limited measurements. This chapter deals with two distinctive types of measurement optimization for CS-DMD, namely the predefined scheme and the adaptive one. It is confirmed that the former approach ruled out unnecessarily fine-scale measurements by comprising distributed sensing positions. The information from the selected sensors strengthens the reconstruction of a spatial distribution with respect to the selected basis vectors introduced in the preceding chapter. The other approach is to evaluate the uncertainty in the estimation of CS, by approximating the covariance matrix without tedious marginalization of posterior distribution. Adaptively appended measurements mitigate the ill-conditioness of the CS estimation with a limited number of measurements. These refined methods that integrate all of the above techniques attain a reduced-order representation of the fluid motion with much fewer measurements than previous attempts.

Chapter 2

Selection of Measurement Position for State Estimation of Linear Dynamical Systems

It has become possible, though gradually, to understand a complex phenomenon by expressing the behavior using data-driven methods. In spite of the high-dimensionality of the expression due to the huge amount of training data, the measurement is efficiently done by utilizing only a limited amount of measurement to infer the state of target phenomenon. The studies in this chapter seek to develop a method to sense such dynamical behaviors at the position from which their state can be observed most significantly. Objective functions of such systems and algorithms are investigated to realize an effective sensing system.

2.1 Background

2.1.1 State-space models approximating dynamical systems

Differential equations often interpret the physical aspects of dynamical phenomena of interest as state-space models. For feedback control purposes, a predictor executes the computation based on the model and the information gathered from the measurements to determine the control input applied at the future time step. There are several well-established methods, such as model predictive control and optimal control, just to name a few. However, the application

of these control methods becomes intensively challenging when the model requires a relentless series of heavy online computations to predict the temporal evolution by nonlinear evaluations. Therefore, techniques have been intensively discussed to learn and simulate control outputs based on the original, computationally costly prediction. Fluid flow is one of the typical phenomena governed by nonlinear equations, therefore, it is onerous to offer a prediction in control by online numerical simulations. Several methods can be confirmed that attempt feedback control to suppress undesirable fluctuation of fluid flows using machine learning methods (C. Lee et al., 1997; Paris et al., 2021; Sasaki and Tsubakino, 2020).

A linearized form of a state-space model is another approach for feedback control that bypasses computational difficulty. By analyzing the responses of the system to impulse input, for example, the system identification process reveals its fundamental parts as a linear representation that relies on the eigensystem realization algorithm (Brunton and Kutz, 2019, January), for example. In the current study, an autonomous representation is considered that excludes actuation driven by control input or system noise.

2.1.2 Sparse sensing and measurement optimization

Recent advances in data science techniques have enabled us to systematically extract reduced-order models from a vastly large-scale measurement of complex phenomena (Baddoo et al., 2023; Berkooz et al., 1993; Brunton and Kutz, 2019, January; Iwasaki et al., 2022; Jovanović et al., 2014; Kutz et al., 2016b, November; Scherl et al., 2020; Schmid, 2010; X. Zhang et al., 2022). A simplified dynamical representation allows easygoing computations with the number of the state variable compressed much less than the original high-dimensional data by the order reduction procedure. However, the measurement equation still contains a large measurement vector, which will induce the cost of collecting and processing for feedback control or observation of the target. In recent studies, a framework is therefore sought that entails a small number of measurements of point sensing to reconstruct the whole state using the obtained reduced-order model. This "sparse sensing" restores the state in a low-dimensional representation as an inversion estimation problem, and the entire data is recovered including unobserved sites. Estimation of

sparse sensing can be interpreted as an autoencoder that has only one middle layer, and linear activation functions that mix up the states to generate latent variables. Because the quality of the estimation obviously depends on the information obtained from the measurements, it is natural that there exist urgent requests to discriminate the effective positions of sensing site to reduce the redundancy of measurements.

Systematic methods for the optimization of sensor positions have been intensively discussed in order to obtain the most representative positions to place sensors and to reduce the resulting estimation error. Many works may be found covering various research areas including dynamical or static, and deterministic and probabilistic configurations, such as when monitoring sensor networks (B. Li et al., 2021; Nomura et al., 2022; Sakiyama et al., 2019; Sun et al., 2019), fluid flows around objects (Carter et al., 2021; DeVries and Paley, 2013; Inoba et al., 2022; Inoue et al., 2023; Kanda et al., 2021, 2022; Kaneko et al., 2021; S. Li et al., 2022; Natarajan et al., 2016; Tiwari et al., 2022), plants and factories (Alonso et al., 2004; Hoseyni et al., 2022; Ren et al., 2008), infrastructures (Castro-Triguero et al., 2013; Krause et al., 2008a; E.-T. Lee and Eun, 2022), circuits (Bates et al., 1996), and biological systems (Tzoumas et al., 2018), estimating physical field (Kraft et al., 2013, July; Nagata et al., 2022a; Nakai et al., 2023; Yildirim et al., 2009), and localizing source (Doğançay and Hmam, 2009; Yeo et al., 2022).

The optimization of sensor measurement is gaining its importance for the reconstruction of complex phenomena from sensor measurements based on data-driven reduced-order models (Manohar et al., 2018; Saito et al., 2021), as well as for model-free machine learning (Fukami et al., 2021; Paris et al., 2021) and data assimilation (Misaka and Obayashi, 2014; Mons and Marquet, 2021; Yoshimura et al., 2020). In many static configurations that exclude the temporal evolution of the state of the system, the selection criteria are statistically described by the Fisher information matrix (FIM), which originated in the parameter sensitivity analysis (Fisher, 1925; Martin et al., 2001). This matrix measure refers to the expected error ellipsoid of the estimation for the state in each instance using a corresponding snapshot of measurement, which corresponds to the inverse of the Cramer–Rao bound (Kay, 1993, Chapter 8) and (Joshi and Boyd, 2009). An approach to dealing with the combinatorial problem structure of selection is to transform them

into small subproblems in a greedy formulation (Manohar et al., 2018; Nakai et al., 2021; Saito et al., 2021), of which the quality of the approximate solution is lower-bounded by the submodularity property in such optimization problems (Hashemi et al., 2020; Krause and Golovin, 2014, February; Lovász, 1983; Mirzasoleiman et al., 2015; Nemhauser et al., 1978). This property of the optimization problem incorporates other criteria, including information entropy and various kinds of distance between information (Chepuri and Leus, 2015; Krause et al., 2008b; Sun et al., 2019). Furthermore, greedy selections have been empirically confirmed to work effectively even in the absence of the submodularity (Clark et al., 2019; Nakai et al., 2021; Yamada et al., 2021). One can also confirm other approaches that seek the global optimum under the relaxed constraints on the selection variable (Joshi and Boyd, 2009; Liu et al., 2016), or that employ proximal algorithms that form proximity operators (Dhingra et al., 2014; Nagata et al., 2021, 2022b).

This study extends the selection of sensor node subsets for dynamical systems. The selection is carried out based on the estimation error covariance, whereas the measurement quality evaluation tends to be more tedious than the static configurations. The observability Gramian is the counterpart of the FIM for linear time-invariant (LTI) representations with deterministic dynamics, as known in (Wouwer et al., 2000). Note that the optimization of the controllability Gramian is the dual problem in the selection of sensor nodes, therefore these selections are provided in the same context (T. Summers and Shames, 2016; T. H. Summers et al., 2016).

It is worth highlighting that examples of random and regular networks of nodes of tens or hundreds are used in the previous analysis, which is partly due to its computational cost. For example, the previous study (T. H. Summers et al., 2016) reported the tremendous increase in the computation time of greedy selection and a relaxation method with semidefinite programming for a power grid system of such size. There are likely to be computational issues when the optimization objective is further extended to the observability for the nonlinear state-space models (DeVries et al., 2012; Montanari et al., 2022; Yoshimura et al., 2020), the error covariance of the Kalman filter (Shamaiah et al., 2010; Tzoumas et al., 2016; Ye et al., 2018; H. Zhang et al., 2017; X. Zhang et al., 2015) and the H_2 norm of an LTI system of order reduced by the balanced

truncation (Clark et al., 2020; Manohar et al., 2021; Zhou et al., 1999). Accordingly, the main interest in this study is to apply Gramian-based selection methods for a larger-scale LTI system, which is constructed using data-driven modeling methods.

2.1.3 Objective and contribution of this chapter

The objective of this chapter is to improve the performance of the sensing framework by selecting the measurement positions for a linear dynamical system. The state-space model for the representation is given by a prior analysis of the data, although this is not the case for the following chapters. This study implements optimization frameworks for sensor node selection based on a criterion of expected estimation error defined for these models. The optimization techniques used for static problems are incorporated to ameliorate the computational cost, which should become more intense than those of static models.

The main developments that contribute to the thesis are listed below.

- *Integration of optimization formulation into linear dynamical systems:* As demonstrated in Wouwer et al., 2000, parameter sensitivity in the estimation based on the linear evolution is denoted by the observability Gramian. The connection between the optimization methods of measurement positions for dynamical and static systems is clearly discussed in order to integrate the methods.

- *Novel methods for sensor node selection:* Two novel methods are proposed to accelerate selection compared to the existing methods. These methods utilize the gradient of the objective function with relaxation to the continuous function.

- *Empirical comparison of computational costs:* A comparison of selection strategies is also provided in Section 2.3 to illustrate the characteristics of each selection method in terms of execution time and the acquired optimization measure. This comparison also elaborates on the effective computational complexities of each selection method, which is not provided in the results of the previous study.

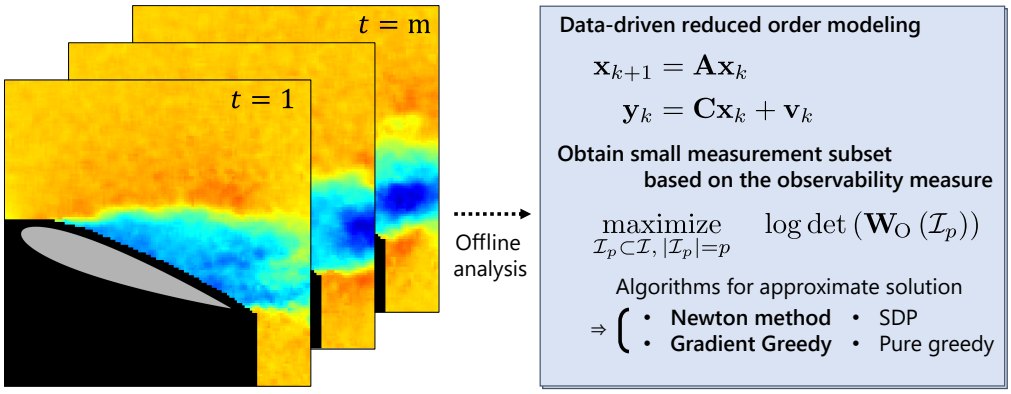


FIGURE 2.1: Brief description of this part. (Left) Representative data points are revealed from "rich" measurement data by using a data-driven model and optimization procedure. (Right) A data-driven method constructs linear reduced-order models before the optimization is conducted by approximate algorithms including our novel methods denoted by bold types.

Figure 2.1 depicts the study in this chapter. In the following sections, the basic formulation of the optimization problem is revealed in Section 2.2.2 after providing the dynamical system of interest. Sections 2.2.3 and 2.2.4 subsequently address our novel algorithms for optimization, while the previously presented approaches are briefly included. The systems are constructed in two different ways for comparisons, where the one is constructed using random numbers as a general case in Section 2.3.1, and in the other case, a data-driven method using proper orthogonal decomposition (POD) (Nankai et al., 2019) is applied as real-world datasets in Section 2.3.2.

2.2 Methods

2.2.1 Observability Gramian and inverse problem

A discrete-time LTI state-space model assumed here has $r(\in \mathbb{N})$ states and $n(\in \mathbb{N})$ measurements. Let $(\mathbf{C}, \mathbf{A})$ be respectively a measurement and a system matrix, and a model of interest is of a form,

$$\begin{cases} \mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k & \text{(2.1a)} \\ \mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{v}_k & \text{(2.1b)} \end{cases}$$

generates a trajectory of state variables $\mathbf{x}_k \in \mathbb{R}^r$ and observations $\mathbf{y}_k \in \mathbb{R}^n$. The subscript $k \in \mathbb{N}$ refers to a snapshot at time k. Assume the observation equation (2.1b) is corrupted by a Gaussian noise $\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}|\sigma^2\mathbf{I}_n)$ that is independent of the state $\mathbf{x}_k$ and has the variance of the same amplitude, $\mathrm{E}\left[\mathbf{v}_k\mathbf{v}_k^\top\right] = \sigma^2\mathbf{I}_n$, where $\mathbf{I}_n$ is the identity matrix of size n. Here, $\mathrm{E}\left[f(\xi)\right]$ takes the expectation of an arbitrary function f over a random variable ξ.

The observability of the system reflects the ability to infer the state of the system from a series of measurements $\mathbf{y}_k$. This property is quantified by a Gramian matrix, which is calles the observability Gramian

$$\overline{\mathbf{W}}_{\mathrm{O}} = \sum_{k=0}^{\infty} \left(\mathbf{A}^\top\right)^k \mathbf{C}^\top \mathbf{C}\mathbf{A}^k. \tag{2.2}$$

In the state-space representation obtained by data-driven techniques, the state vector represents a low-dimensional motion inherent in high-dimensional measurement vectors. Therefore, $n \gg r$ is assumed hereafter. One may be reminded that properties of the observability and the controllability are dual for linear systems. Therefore, the methodologies presented here are similarly adopted to controllability optimization by position selections for actuator placement.

In our formulation, this matrix is characterized by the linear least-squares estimation of the state vector that utilizes time-series measurements. The measurement equation is first concatenated from $k = 0$ to $k = l - 1$ of Eq. (2.1):

$$\begin{aligned}
\mathbf{y}_{0:l-1} &\coloneqq \left[\mathbf{y}_0^\top, \mathbf{y}_1^\top, \ldots, \mathbf{y}_{l-1}^\top\right]^\top \\
&= \begin{bmatrix} \mathbf{C} \\ \mathbf{C}\mathbf{A} \\ \vdots \\ \mathbf{C}\mathbf{A}^{l-1} \end{bmatrix} \mathbf{x}_0 + \begin{bmatrix} \mathbf{v}_0 \\ \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_{l-1} \end{bmatrix} \\
&=: \mathbf{C}_{0:l-1}\mathbf{x}_0 + \mathbf{v}_{0:l-1}.
\end{aligned} \tag{2.3}$$

Here, $\mathbf{C}_{0:l-1}$ and $\mathbf{v}_{0:l-1}$ are the stacked components in the brackets in Eq. (2.3). An estimate of $\mathbf{x}_0$ is obtained by the linear inversion as follows:

$$\tilde{\mathbf{x}}_0 = (\mathbf{C}_{0:l-1})^{\dagger}\, \mathbf{y}_{0:l-1} \tag{2.4}$$

$$= \left(\mathbf{C}_{0:l-1}^{\top}\mathbf{C}_{0:l-1}\right)^{-1} \mathbf{C}_{0:l-1}^{\top}\mathbf{y}_{0:l-1} \tag{2.5}$$

$$= \mathbf{x}_0 + \left(\mathbf{C}_{0:l-1}^{\top}\mathbf{C}_{0:l-1}\right)^{-1} \mathbf{C}_{0:l-1}^{\top}\mathbf{v}_{0:l-1}, \tag{2.6}$$

where $\tilde{\mathbf{x}}_0 \in \mathbb{R}^r$ and $\bigcirc^{\dagger}$ stands for the estimate of $\mathbf{x}_0$ and the Moore-Penrose pseudo inverse, respectively. The matrix $\mathbf{C}_{0:l-1}$ is assumed to be column full rank due to $n \gg r$. An error covariance matrix under the estimation $\tilde{\mathbf{x}}_0$ is proportional to the inverse matrix of the Gramian as shown in the following equation:

$$\mathrm{E}\left[(\mathbf{x}_0 - \tilde{\mathbf{x}}_0)\,(\mathbf{x}_0 - \tilde{\mathbf{x}}_0)^{\top}\right] \tag{2.7}$$

$$\propto \left(\mathbf{C}_{0:l-1}^{\top}\mathbf{C}_{0:l-1}\right)^{-1} \tag{2.8}$$

$$= \left(\sum_{k=0}^{l-1}\left(\mathbf{A}^{\top}\right)^{k}\mathbf{C}^{\top}\mathbf{C}\mathbf{A}^{k}\right)^{-1} \tag{2.9}$$

$$\underset{l\to\infty}{\longrightarrow} \left(\overline{\mathbf{W}}_{\mathrm{O}}\right)^{-1}. \tag{2.10}$$

2.2.2 Optimization on measurement positions

Let a small number of sensors be deployed while maintaining the quality of the state estimation. The optimization problem is formulated as a selection of a subset $\mathcal{I}_p = \{\, i_1, \ldots, i_p\}$ from all available n measurement nodes, labeled by $\mathcal{I}_n = \{1, 2, \ldots, n\}$, where a measure of the Gramian is optimized under some constraints.

A permutation matrix $\mathbf{P}(\mathcal{I}_p)$ that extracts a part of the measurement corresponding to the sensor indices $(\mathcal{I}_p)$ is defined by

$$\mathbf{P}(\mathcal{I}_p) = \begin{bmatrix} \mathbf{e}_{i_1}^\top \\ \vdots \\ \mathbf{e}_{i_p}^\top \end{bmatrix}, \tag{2.11}$$

where a unit vector $\mathbf{e}_i \in \mathbb{R}^n$ has unity in the i-th entry, and the rest is zero. The Gramian given by the selected sensors is rewritten as a subset function;

$$\mathbf{W}_{\mathrm{O}}(\mathcal{I}_p) = \sum_{k=0}^{\infty} \left(\mathbf{A}^\top\right)^k (\mathbf{PC})^\top (\mathbf{PC}) \mathbf{A}^k, \tag{2.12}$$

and therefore, $\mathbf{W}_{\mathrm{O}}(\mathcal{I}_n) = \overline{\mathbf{W}}_{\mathrm{O}}$. Here, $\mathbf{W}_{\mathrm{O}}(\mathcal{I}_p)$ can be calculated by solving the following Lyapunov equation:

$$\mathbf{A}^\top \mathbf{W}_{\mathrm{O}}(\mathcal{I}_p) \mathbf{A} - \mathbf{W}_{\mathrm{O}}(\mathcal{I}_p) + (\mathbf{PC})^\top (\mathbf{PC}) = \mathbf{0}. \tag{2.13}$$

Some measures for the optimization of $\mathbf{W}_{\mathrm{O}}(\mathcal{I}_p)$ were presented in (T. H. Summers et al., 2016), for example. A maximization of the determinant of the Gramian is configured in the current study. The maximization problem is:

$$\underset{\mathcal{I}_p \subset \mathcal{I},\, |\mathcal{I}_p| = p}{\text{maximize}} \quad \log\det\left(\mathbf{W}_{\mathrm{O}}(\mathcal{I}_p)\right). \tag{P0}$$

The logarithmic form, which is monotone, is considered for ease of calculation in algorithms. This determinant maximization strategy is commonly used in sensor placement and optimal design of interpolation methods (Drmač and Saibaba, 2018; Manohar et al., 2018).

The optimization problem (P0) is a combinatorial problem, and therefore, finding the true optimum is computationally prohibitive. This section introduces four methods as shown in table 2.1 to obtain a suboptimal but reasonable solution to (P0). Section 2.2.3 deals convex relaxation approaches to (P0), where the notations of *SDP* and *approximate convex relaxation* stand for the

selection problems based on the semidefinite programming (SDP) of (T. Summers and Shames, 2016) and an approximate smooth convex relaxation for the Newton method that is extended from (Joshi and Boyd, 2009), respectively. Subsequently, Section 2.2.4 provides formulations with regard to the *pure greedy* and *gradient greedy* methods, as a simplified greedy implementation of (T. H. Summers et al., 2016) for the matrix determinant maximization, while the latter is its linear approximation. The computational complexities are discussed in Section 2.2.5.

TABLE 2.1: Selection algorithms for each relaxed problem and expected arithmetic complexity order based on the basic matrix operations. **Bold items** refer to the selection approaches proposed in the presented study.

Problem	Algorithm	Expected complexity
Linear relaxation SDP	Path-following method	$\big(\mathcal{O}(n^4) + \mathcal{O}(n^2r^2) + \mathcal{O}(nr^3)\big)/$ iter.
Approximate convex relaxation	**BRS-Newton** [Alg. 2]	$\big(\mathcal{O}(n^3) + \mathcal{O}(n^2r^2) + \mathcal{O}(nr^3)\big)/$ iter.
Greedy	Pure greedy [Alg. 3]	$\mathcal{O}(pnr^3)$
Greedy	**Gradient greedy** [Alg. 4]	$\mathcal{O}(pnr^2) + \mathcal{O}(pr^3)$

2.2.3 Convex relaxation methods

This strategy substitutes continuous variable for the discrete selection binary. The original combinatorial Gramian $\mathbf{W}_{\mathrm{O}}(\mathcal{I}_p)$ is replaced by the weighted sum of the Gramian calculated for each sensor candidate. Let the sum of selection weight $\mathbf{s} \in [0,\, 1]^n$ be bounded to p to control the cardinality of the final result. The relaxed optimization problem of interest is:

$$\begin{aligned} \underset{\mathbf{s}}{\text{maximize}} \quad & \log\det\left(\mathbf{Q}\left(\mathbf{s}\right)\right) \\ \text{subject to} \quad & \mathbf{Q}\left(\mathbf{s}\right) = \sum_{i=1}^{n} s_i\, \mathbf{W}_{\mathrm{O}}\left(\{i\}\right), \\ & s_i \in [0,\, 1]\,, \quad \mathbf{1}^{\top}\mathbf{s} = p. \end{aligned} \tag{P1}$$

This "weight" formulation is widely used in the various types of sensor selection including linear inverse and Kalman filter estimation, (Joshi and Boyd, 2009; Liu et al., 2016) for example.

On the other hand, the "gain" formulation is employed to optimize gains to yield state variables from measurements, as can be found in the literature on sensor selection for linear inverse estimation (Nagata et al., 2021, 2022b) or based on Kalman filter (Dhingra et al., 2014; Zare et al., 2020). The group regularization to gain matrix therein distinguishes the representative measurement locations. To the best of our knowledge, no study has been conducted that applies the gain formulation to the sensor selection based on the observability Gramian. This may be because the Gramian is related to an infinite series of temporal measurements, and so are the gains for such measurements. Truncating Eq. (2.8) to a finite time horizon is obviously a possible option, but it evidently becomes infeasible to deal with a large solution vector. Defining an appropriate finite time horizon raises another difficulty of choosing a hyperparameter. Accordingly, the implementation of the gain formulation remains an interesting challenge.

Semidefinite programming based selection (SDP)

This formulation was previously introduced in (T. Summers and Shames, 2016, section II-B), and therefore, the readers should refer to the original work for more details. It should be emphasized that an optimization problem in discrete-time form is briefly revisited in our study and is included in Section 2.3 for a comparison of selection methods. An optimization problem (P1) is transformed into the following SDP representation:

$$\begin{aligned} \underset{\mathbf{s}}{\text{maximize}} \quad & \log\det\left(\mathbf{Q}\left(\mathbf{s}\right)\right) \\ \text{subject to} \quad & s_i \in [0,\,1]\,, \quad \mathbf{1}^\top \mathbf{s} = p, \quad \mathbf{Q}\left(\mathbf{s}\right) \succeq \mathbf{0}, \\ & \mathbf{A}^\top \mathbf{Q}\left(\mathbf{s}\right) \mathbf{A} - \mathbf{Q}\left(\mathbf{s}\right) + \sum_{i=1}^{n} s_i\, \mathbf{c}_i^\top \mathbf{c}_i = \mathbf{0}, \end{aligned} \tag{P2}$$

where $\mathbf{c}_i$ is the i-th row of the measurement matrix $\mathbf{C}$, and $\mathbf{0}$ is a zero matrix with appropriate dimensions. It should be noted that the Lyapunov equation imposed as a constraint in the problem can be satisfied by SDP solvers introduced later. The complexity in table 2.1 is based on a path-following method for a general primal-dual interior point method (Nesterov and Nemirovskii, 1994, January).

Newton method for approximate convex relaxation, and its customized algorithm with randomized subspace sampling (Approximate convex relaxation)

This novel method solves the convex relaxation problem by applying the Newton method and a customized randomization to Eq. (P1), with a penalty term that bounds the weight variables.

The description of the proposed method starts with the extension of the formulation of sensor selection for static systems first introduced in (Joshi and Boyd, 2009). In their approach, the Newton method solved a weight formulation of a determinant optimization for the FIM of the linear inverse problem, which returned the p-largest indices of $\mathbf{s}$ as a result of a heuristic sensor node selection. In this study, the above idea is straightforwardly extended to the Gramian. A smooth convex objective function approximates Eq. (P0) as follows:

$$\begin{aligned} \underset{\mathbf{s}}{\text{maximize}} \quad & \log\det\left(\mathbf{Q}\left(\mathbf{s}\right)\right) + \kappa \sum_{i=1}^{n} \left(\log\left(s_i\right) + \log\left(1 - s_i\right)\right) \\ \text{subject to} \quad & s_i \in \left(0,\, 1\right), \quad \mathbf{1}^{\top}\mathbf{s} = p \end{aligned} \tag{P3}$$

with $\kappa > 0$, which adjusts the smoothness of the objective function. A Newton step $\delta\mathbf{s} \in \mathbb{R}^n$ is determined by minimizing the second-order approximation of the objective function under a constraint $\delta\mathbf{s} = 0$ (Boyd et al., 2004, Section 10.2)(Joshi and Boyd, 2009), with the notation f referring to the objective function in Eq. (P3) as shown below:

$$\delta\mathbf{s} = \left(\nabla^2 f\right)^{-1} \left(-\nabla f + \frac{\mathbf{1}^{\top}\left(\nabla^2 f\right)^{-1}\nabla f}{\mathbf{1}^{\top}\left(\nabla^2 f\right)^{-1}\mathbf{1}}\mathbf{1}\right). \tag{2.14}$$

The first and second derivatives with respect to the selection weight $\mathbf{s}$ are given by

$$\left[\nabla f\right]_i = \mathbf{c}_i \left(\sum_{k=0}^{\infty} \mathbf{A}^k \mathbf{Q}^{-1}(\mathbf{s}) \left(\mathbf{A}^{\top}\right)^k\right) \mathbf{c}_i^{\top} + \frac{\kappa}{s_i} - \frac{\kappa}{1 - s_i}, \tag{2.15}$$

$$\left[\nabla^2 f\right]_{i,j} = -\mathbf{c}_i \mathbf{H}_j\left(\mathbf{s}\right) \mathbf{c}_i^{\top} - \delta_{i,j}\left(\frac{\kappa}{s_i^2} + \frac{\kappa}{\left(1 - s_i\right)^2}\right), \tag{2.16}$$

where $\delta_{i,j}$ is the Kronecker delta, and

$$\mathbf{H}_j(\mathbf{s}) := \sum_{k=0}^{\infty} \left[\mathbf{A}^k \mathbf{Q}(\mathbf{s})^{-1} \mathbf{W}_{\mathrm{O}}(\{j\}) \mathbf{Q}(\mathbf{s})^{-1} \left(\mathbf{A}^\top\right)^k \right] \tag{2.17}$$

is the solution of the following Lyapunov equation;

$$\mathbf{A}\mathbf{X}\mathbf{A}^\top - \mathbf{X} + \mathbf{Q}(\mathbf{s})^{-1} \mathbf{W}_{\mathrm{O}}(\{j\}) \mathbf{Q}(\mathbf{s})^{-1} = \mathbf{0}. \tag{2.18}$$

The procedure is carried out according to Alg. 1, where the iteration is completed when the ℓ_2 norm of the solution update falls below the threshold given.

Algorithm 1 Newton algorithm for Eq. (P3)

Input: $\mathbf{C} \in \mathbb{R}^{n\times r}$, $\mathbf{A} \in \mathbb{R}^{r\times r}$, $p \in \mathbb{N}$
Output: Indices of chosen p sensor positions $\mathcal{I}_p$
  Set an initial weight $\mathbf{s} \leftarrow \mathbf{1}p/n$
  while convergence condition not satisfied **do**
    Calculate ∇f by Eq. (2.15) and $\nabla^2 f$ by Eq. (2.16)
    Calculate $\delta\mathbf{s}$ by Eq. (2.14)
    Obtain step size t by backtracking line search
    Set $\mathbf{s} \leftarrow \mathbf{s} + t\delta\mathbf{s}$
  end while
  Return the indices of the p-largest components of $\mathbf{s}$ as $\mathcal{I}_p$

One of the most computationally demanding steps of the Newton method in Section 2.2.3 is the inverse of the Hessian in Eq. (2.14), approximately reaching $\mathcal{O}(n^3)$. The rest of this section describes an improvement of efficiency for the Newton method of Alg. 1, as the computational cost was relaxed using small sketches as in (Gower et al., 2019; Nonomura et al., 2021a). The dimension of the sketched Newton step is $\tilde{n} < n$, where the results for $\tilde{n}/n = 0.1$ in (Nonomura et al., 2021a) showed a drastic reduction in the total computation time for convergence in spite of the increase in a number of iterations. In this study, a sketching matrix $\mathbf{S}_{\tilde{n}} \in \mathcal{R}^{\tilde{n}\times n}$ is constructed using a permutation matrix for an ease of computation, as summarized in the following and in Alg. 2. The subspace referring to the permutation is randomly assigned from the n components,

which is indexed by;

$$\mathcal{I}_{\tilde{n}} = \{i'_1, \ldots, i'_\rho, \ldots i'_{\tilde{n}}\}. \tag{2.19}$$

This random selection is biased according to the selection weight $\mathbf{s}$ in the previous iteration. This heuristic leads to a reasonable acceleration of the convergence because the weights of higher weighted sensors in the first few iterations are more likely updated frequently for convergence. In this study, half of the permutation coordinates correspond to the indices of the highest $\rho = \tilde{n}/2$ values of $\mathbf{s}$. The rest half is randomly selected from remaining $n - \tilde{n}/2$ dimensions and the exploration of sensor selection is further accelerated.

Accordingly, the calculations of the gradient (the first term in Eq. (2.15)) and the Hessian (also the first term in Eq. (2.16)) are simplified to the subspace indexed by $\mathcal{I}_{\tilde{n}}$. These subsampled derivatives and the Newton step derived are denoted by $\nabla\tilde{f}$, $\nabla^2\tilde{f}$, and $\delta\tilde{\mathbf{s}}$. The criterion of the convergence is modified from that in Alg. 1 due to the randomness, where the algorithm stops if the update size is less than a given threshold in $n/\tilde{n}$ consecutive iterations.

Algorithm 2 Customized algorithm of Alg. 1 (*BRS-Newton*)

Input: $\mathbf{C} \in \mathbb{R}^{n\times r}$, $\mathbf{A} \in \mathbb{R}^{r\times r}$, $p > 0$, $\tilde{n} > 0$
Output: Indices of chosen p sensor positions $\mathcal{I}_p$
  Set $\mathbf{s} \leftarrow \mathbf{1}p/n$
  while convergence condition not satisfied **do**
    Select $\mathcal{I}_{\tilde{n}}$ [Eq. (2.19)] and set $\mathbf{S}_{\tilde{n}}$
    Calculate subsampled derivatives $\nabla\tilde{f}$ and $\nabla^2\tilde{f}$
    Calculate $\delta\tilde{\mathbf{s}}$
    Obtain step size t by backtracking line search
    Set $\mathbf{s} \leftarrow \mathbf{s} + t\left(\mathbf{S}_{\tilde{n}}^\top \delta\tilde{\mathbf{s}}\right)$
  end while
  Return the indices of the p-largest components of $\mathbf{s}$ as $\mathcal{I}_p$

2.2.4 Greedy algorithms

Selection by greedy algorithms sequentially adds a sensor that most improves the predefined objective function to the subset determined in the previous iterations. One may notice that the greedy algorithms approximate the combinatorial aspect of the original structure of the

optimization problem, yet preserve the discrete optimization structure, which is in contrast to the previous relaxation methods. However, selection by greedy algorithms is empirically known to be tremendously costly for high-dimensional systems and evaluations by complicated objective functions. The accelerating method is therefore applied (Hashemi et al., 2020; Mirzasoleiman et al., 2015) as randomization is included in the Newton method in Alg. 2. Section 2.2.4 further approximates the greedy selection by its low-order evaluation.

Greedy selection with direct evaluation of objective function (Pure greedy)

Submodularity in the objective function guarantees the quality of solutions of greedy selection (Feige et al., 2011; Lovász, 1983; Nemhauser et al., 1978). This property is derived for several metrics related to the Gramian as discussed in detail in (T. H. Summers et al., 2016). In the selection of the q-th sensor where $1 \leq q \leq p$, the Gramian $\mathbf{W}_{\mathrm{O}}(\mathcal{I}_q)$ is obtained from the algebraic Lyapunov equation (2.13). Algorithm 3 starts by calculating all the Gramian for each sensor candidate. In the subsequent selections, a candidate subset $\mathcal{I}_*$ is identified that results in the highest dimension of the observable subspace, then evaluates the determinant of the decomposed Gramian $\hat{\mathbf{W}}_{\mathrm{O}}(\mathcal{I}_{q-1} \cup \{i\})$ into the observable space by multiplying their nonzero eigenvalues.

Note that the selection starts with finding a single measurement that realizes the highest objective value, and therefore, the determinant can be zero if the observability is not obtained by any single sensor. As mentioned in (T. H. Summers et al., 2016), one may evaluate the objective function for the observable subspace on such occasions. Once the full-rank Gramian is achieved by an obtained subset, the greedy algorithm drops the computation of the rank of the candidate Gramians since the rank is monotone increasing (T. H. Summers et al., 2016).

Greedy selection with gradient approximation (Gradient greedy)

This study proposes Alg. 4 that is expected to accelerate Alg. 3 by the linear approximation of the evaluation by matrix determinant. The procedure of this approximate selection is schematically explained in Fig. 2.2. The gradient greedy algorithm selects, in the current step q, a sensor

Algorithm 3 Determinant-based greedy algorithm (*pure greedy*)

Input: $\mathbf{C} \in \mathbb{R}^{n\times r}$, $\mathbf{A} \in \mathbb{R}^{r\times r}$, $p \in \mathbb{N}$
Output: Indices of chosen p sensor positions $\mathcal{I}_p$
$\mathcal{I}_n \leftarrow \{1,\ldots,n\}$, $\mathcal{I}_0 \leftarrow \emptyset$,
for $q = 1,\ldots,p$ **do**
$\mathcal{I}_* \leftarrow \left\{ i : i \in \underset{i\in\mathcal{I}_n\setminus\mathcal{I}_{q-1}}{\operatorname{argmax}} \operatorname{rank} \mathbf{W}_{\mathrm{O}}\left(\mathcal{I}_{q-1}\cup\{i\}\right) \right\}$
$i_q \leftarrow \underset{i\in\mathcal{I}_*}{\operatorname{argmax}} \log\det \hat{\mathbf{W}}_{\mathrm{O}}\left(\mathcal{I}_{q-1}\cup\{i\}\right)$
$\mathcal{I}_q \leftarrow \mathcal{I}_{q-1}\cup\{i_q\}$
end for

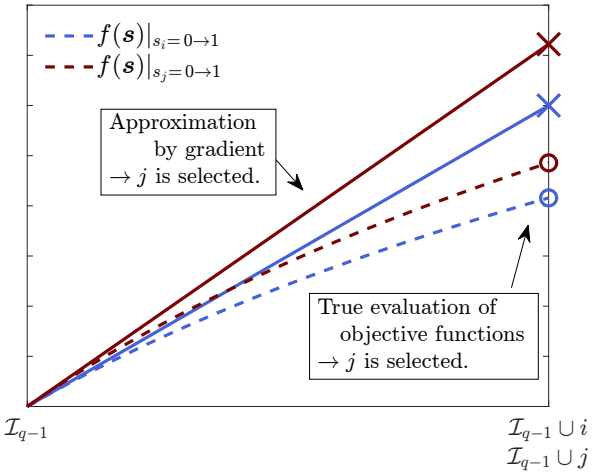


FIGURE 2.2: Schematic of the linear approximation exploited in the gradient greedy algorithm for a convex function $f(\mathbf{s})$.

corresponding to the element of the highest gradient of the objective function $f(\mathbf{s})$ in the prior step. Recall the gradient with respect to the i-th sensor candidate of Eq. (2.15);

$$\left[\nabla \log\det\left(\mathbf{Q}\left(\mathbf{s}\right)\right)\right]_i = \mathbf{c}_i \left(\sum_{k=0}^{\infty} \mathbf{A}^k \mathbf{Q}^{-1}(\mathbf{s}) \left(\mathbf{A}^\top\right)^k\right) \mathbf{c}_i^\top,$$

where $\mathbf{c}_i$ is the same notation as used in (P3). In the selection of the q-th sensor, the i-th component of $\nabla \log\det\left(\mathbf{Q}\left(\mathbf{s}\right)\right)$ is given by replacing $\mathbf{Q}\left(\mathbf{s}\right) \to \mathbf{W}_{\mathrm{O}}\left(\mathcal{I}_{q-1}\right)$ as follows:

$$\lim_{\mathbf{Q}(\mathbf{s})\to\mathbf{W}_{\mathrm{O}}(\mathcal{I}_{q-1})} \left[\nabla \log\det\left(\mathbf{Q}\left(\mathbf{s}\right)\right)\right]_i = \mathbf{c}_i \left(\sum_{k=0}^{\infty} \mathbf{A}^k \mathbf{W}_{\mathrm{O}}\left(\mathcal{I}_{q-1}\right)^{-1} \left(\mathbf{A}^\top\right)^k\right) \mathbf{c}_i^\top. \tag{2.20}$$

This approximation drops the amount of evaluations of the Lyapunov equation and the matrix determinant from the greedy selection, and instead computes inner products of vectors with

weighting matrix including the Gramian determined in the previous search. Note that the inverse in Eq. (2.20) is singular when the system is unobservable. A small regularization term $\mathrm{diag}\left[\delta, \delta, \ldots, \delta\right]$ is added to $\mathbf{W}_{\mathrm{O}}\left(\mathcal{I}_{q-1}\right)$ to ensure its regularity.

Algorithm 4 Determinant-based gradient greedy algorithm (*Gradient greedy*)

Input: $\mathbf{C} \in \mathbb{R}^{n\times r}$, $\mathbf{A} \in \mathbb{R}^{r\times r}$, $p \in \mathbb{N}$, $\delta > 0$
Output: Indices of chosen p sensor positions $\mathcal{I}_p$
$\mathcal{I}_n \leftarrow \{1, \ldots, n\}$, $\mathcal{I}_0 \leftarrow \emptyset$,
for $q = 1, \ldots, p$ **do**
  $\mathbf{W}_{\mathrm{O}}\left(\mathcal{I}_{q-1}\right) \leftarrow \mathbf{W}_{\mathrm{O}}\left(\mathcal{I}_{q-1}\right) + \mathrm{diag}\left[\delta, \delta, \ldots, \delta\right] \in \mathbb{R}^{r\times r}$
  Find $\mathbf{M}$ *s.t.* $\mathbf{AMA}^{\top} - \mathbf{M} + \mathbf{W}_{\mathrm{O}}\left(\mathcal{I}_{q-1}\right)^{-1} = \mathbf{0}$
  $i_q \leftarrow \underset{i \in \mathcal{I}_n \setminus \mathcal{I}_{q-1}}{\mathrm{argmax}}\ \mathbf{c}_i \mathbf{M} \mathbf{c}_i^{\top}$
  $\mathcal{I}_q \leftarrow \mathcal{I}_{q-1} \cup \{i_q\}$
  Calculate $\mathbf{W}_{\mathrm{O}}\left(\mathcal{I}_q\right)$
end for

2.2.5 Expected computational complexity

This section describes the expected computational complexity of selection methods based on textbook linear algebra. The leading terms of the matrix operations are summarized in table 2.1 with respect to the parameters of system dimension n, p, r. The expected overhead term of run time for each algorithm is therein pointed out, while in practice more compact expressions can be obtained depending on the libraries for computation and the considered problem structures. Readers should also note that the notation of complexity $\mathcal{O}(\bigcirc)$ conventionally omits the constant factors for simplicity. Lower-order terms might overwhelm the others due to the large constants in actual computations.

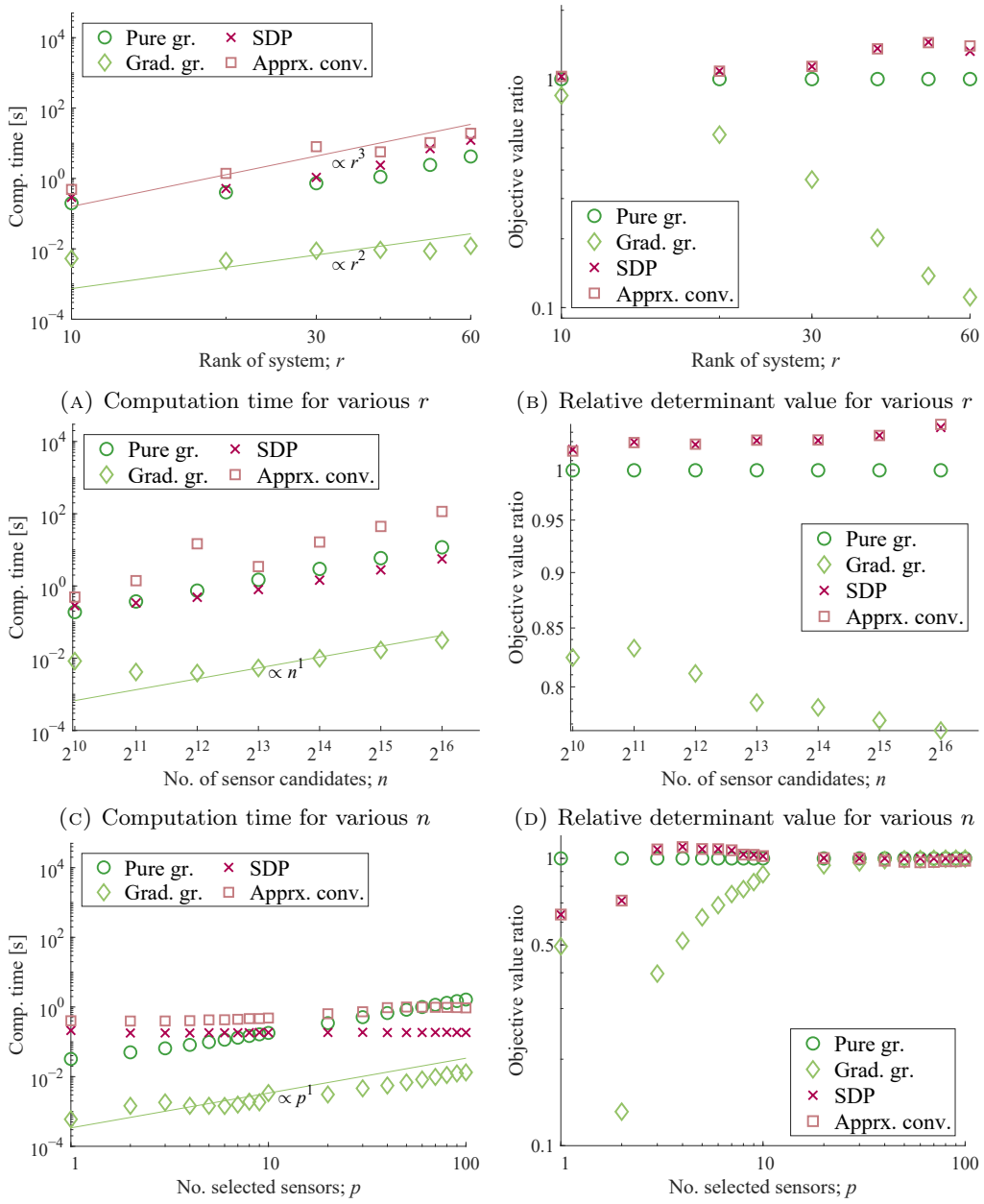
Although the optimal subset $\mathcal{I}_p$ for Eq. (P0) can be found by calculating the objective function for all subsets whose member size is p, the brute-force search will require $\mathcal{O}\left(n^p r^3\right)$ computations which will be barely accessible. As for existing approaches, a naive implementation in Eq. (P2) of the linear convex relaxation method with semidefinite problem (SDP) structure is a simplified form of (T. Summers and Shames, 2016). The interior point method and path-following iterations should require $(\mathcal{O}(n^4) + \mathcal{O}(n^2 r^2) + \mathcal{O}(nr^3))$ per iteration to construct the Newton direction. This is due to the constraints of linear matrix inequality (LMI), such as the

Gramian being semidefinite or the selection variable being bounded, while it is known that certain problem structures ameliorate the overall cost (Boyd et al., 2004, Section 11.8). A simple greedy algorithm (*pure greedy*) requires solving the Lyapunov equation and calculating the determinant of the observability Gramian, of which the cost is $\mathcal{O}\left(r^3\right)$, for all n sensor candidates and for p sensor increment iterations. The overall computational complexity is, therefore, $\mathcal{O}\left(pnr^3\right)$.

The proposed methods will accelerate both existing approaches so far by comparing the theoretical complexities. The proposed convex relaxation method (denoted by *approximate convex relaxation*) utilizes the Newton method iterations for the sensor selection. The algorithm requires $\mathcal{O}\left(n^3\right)$ computations of the inversion of Hessian. Furthermore, Eq. (2.18) will be solved for every $j \in \{1, 2, \ldots, n\}$ for Eq. (2.16), which requires computational complexities of $\mathcal{O}\left(nr^3\right)$ to solve the algebraic Lyapunov equation (2.18) and obtaining the matrix multiplication inside, respectively. The first term in Eq. (2.16) requires $\mathcal{O}\left(n^2r^2\right)$, therefore, the leading terms will be the sum $\left(\mathcal{O}\left(n^3\right) + \mathcal{O}\left(n^2r^2\right) + \mathcal{O}\left(nr^3\right)\right)$ per iteration. The sketching matrix, in Alg. 2, compresses the dimension of the Newton system to $n \to \tilde{n}$, and reduces the costs of the above mentioned computations by ratio. However, its computational complexity notation does not change. The proposed gradient greedy algorithm (*gradient greedy*) replaces the evaluations of the Lyapunov equation over n sensor candidates with inner products of vectors weighted by a $r \times r$ matrix. The overall computational complexity will be $\mathcal{O}\left(pr^3\right) + \mathcal{O}\left(pnr^2\right)$, where an evaluation of the algebraic Lyapunov equation is taken only p times.

2.3 Results and discussions

Numerical examples illustrate the efficiency of the algorithms considered in Section 2.3. The tested systems, which are represented by $(\mathbf{C}, \mathbf{A})$, are generated by synthetic data set of random numbers and real-world counterpart of large-scale measurements. As used in Section 2.2, the systems of interest are in discrete-time LTI forms. The entire computation is performed on MATLAB R2022a, and CVX 2.2 (CVX Research, 2012, August; M. Grant and Boyd, 2008) with the Mosek solver is used for the SDP-based selection, as noted in appendix A.1. The `dlyap` function is used for the solutions of discrete-time Lyapunov equations such as Eqs. (2.13)



(A) Computation time for various r

(B) Relative determinant value for various r

(C) Computation time for various n

(D) Relative determinant value for various n

(E) Computation time for various p

(F) Relative determinant value for various p

FIGURE 2.3: Computation time and optimization results for randomly generated systems: 2.3a and 2.3b for varying the number of state variable r, *i.e.* the size of $\mathbf{A}$ matrix (average of 100 times trial, $n = 1024,\ p = 10$); 2.3c and 2.3d for varying the number of sensor candidate n, *i.e.* the rows of $\mathbf{C}$ matrix (average of 20 times trial, $p = 10,\ r = 10$); 2.3e and 2.3f for varying the number of sensor selected p, (average of 100 times trial, $r = 10,\ n = 1024$).

and (2.18), which adopts the subroutine libraries, Subroutine Library in Systems and Control Theory (SLICOT) (Barraud, 1977; Benner et al., 1999; Hammarling, 1982; Van Huffel et al.,

2004). It should also be noted that there is another approach for solving the equation like (Kitagawa, 1977), for instance. The MATLAB codes are available through the GitHub repository of the present authors (Yamada, 2023).

2.3.1 Results for randomly generated system

The characteristics of the sensor selection methods are investigated for different sets of system dimension parameters (n, r, p) by varying one parameter while fixing the others. The abscissa of Figures 2.3a and 2.3b is the rank $r \in \{10, 20, \ldots, 60\}$, which is the dimension of the vector of the reduced state variable and thus of the Gramian. That of Figures 2.3c and 2.3d is $n \in \{2^{10}, 2^{11}, \ldots, 2^{16}\}$ as the size of the node members that comprise the original measurement, and accordingly the number of sensor candidates. In Figures 2.3e and 2.3f, it is $p \in \{1, 2, \ldots, 10, 20, \ldots, 100\}$ that represents the varying number of sensors selected.

The problem setting for the random number system is found in (C. Rowley et al., 2014). First, the conjugate complex numbers of which the real parts are minus are assigned to the eigenvalues of a damping continuous-time system matrix $\hat{\mathbf{A}}$. A discrete-time system matrix can be obtained by $\mathbf{A} = \exp(\hat{\mathbf{A}}\Delta t)$ that is stable and full-rank, whereas Δt is the time step of the discrete system. The observation matrix $\mathbf{C}$ is a column-orthogonal matrix generated by the singular value decomposition for a matrix of Gaussian random numbers of appropriate dimensions. Sensors up to p are selected using the algorithms presented in the previous section, and the objective function, $\log\det \mathbf{W}_O(\mathcal{I}_p)$, is calculated for each selected subset. Figure 2.3 shows the performance of the selection methods applied to the system of random numbers.

Figures 2.3a,2.3c, and 2.3e illustrate the total computation time of each algorithm in the tests, where the gradient greedy method is the least time-consuming, followed by the almost the same results of the pure greedy and the SDP-based methods. Unfortunately, the proposed approximate convex relaxation method needed the longest time to solve the optimization in almost all conditions tested despite its acceleration owing to the customized randomization. It is also expected, however, that the computation time of the SDP-based method and the pure greedy selection will exceed the others for even larger r and p, respectively.

The empirical orders of the computation time with regard to each parameter are analyzed herein. The growth rates of the computation time of each algorithm are evaluated by solely changing the system dimension parameters r, n and p, as summarized in table 2.2 based on the results shown in Figures 2.3a, 2.3c and 2.3e. First, the gradient greedy method ran in time proportional to n, but it is not clear regarding r. Since the number of sensor candidates n is much larger than r in the first experiment, the term with r^3 was not significant for the gradient greedy method. The increase against p was an order of unity, which is clearly reasonable. Second, the empirical orders of the pure greedy method when solely changing r, n and p were r^3, n and p, respectively, which agrees with the expected leading order. One can see that a term with r^3 becomes dominant as r grows in Fig. 2.3a. Meanwhile, the estimated orders of the SDP-based method when solely changing r and n were $r^{[4]}$ and n, while those of the approximate convex relaxation method were $r^{[3]}$ and $n^{[2]}$, respectively. Here, let the notation $[\,j\,]$ stand for, despite obvious abuse, a real number x bounded by $j-1$ and j for an arbitrary positive integer j. These noninteger orders may be due to the optimized arithmetic employed in the software, such as the Strassen-like algorithm (Strassen, 1969). The dependency on p was not clear in the SDP-based method, while a slight increase was observed for the approximate convex relaxation method.

TABLE 2.2: Practical orders of the computation time of selection methods: Investigated with respect to each parameter individually. *N/A ... *not admitted*

Selection method	r	n	p
SDP	$r^{[4]}$	n	p^0
Approx. conv. [Alg. 2]	$r^{[3]}$	$n^{[2]}$	$p^{[1]}$
Pure greedy [Alg. 3]	r^3	n	p^1
Gradient greedy [Alg. 4]	N/A	n	p^1

The interesting aspect is that the dominating order of the SDP-based method with respect to n was not n^4 that was initially expected, but approximately proportional to n. This is perhaps because the constraint posed as $0 \leq s_i \leq 1$ in appendix A.1 has been simplified in the tested implementation as a mere diagonal block of the semidefinite linear matrix inequality (LMI), and the Mosek solver should have taken advantage of its structure during the Newton step calculation, despite the large n assumed. It should also be noted that this efficacy was not observed for other solvers, such as SDPT3, although the CVX parser does not seem to change its

output. Nonetheless, the complexity of solving the SDP would be enormous as expected if the LMI included such a semidefinite relaxation (Liu et al., 2016; Luo et al., 2010) of the selection variable vector $\mathbf{s}$, such as $\mathbf{s}\mathbf{s}^\top$ being a semidefinite matrix. This agrees well with the observation from the experiment for the SDP-based method regarding r, where the LMI of dimension $r \times r$ is included.

TABLE 2.3: Numbers of iterations of the convex relaxation approaches: Investigated with respect to each parameter individually and averages are rounded to integer.

(A) Various r for randomly generated systems, average of 100 trials

r	10	20	30	40	50	60
SDP	18	13	11	12	12	12
Approx. conv.	203	293	348	373	380	388

(B) Various n for randomly generated systems, average of 20 trials

n	2^{10}	2^{11}	2^{12}	2^{13}	2^{14}	2^{15}	2^{16}
SDP	18	21	23	27	28	31	33
Approx. conv.	209	167	147	141	139	139	141

(C) Various p for randomly generated systems, average of 100 trials

p	1	2	4	8	10	20	40
SDP	14	13	15	17	18	18	17
Approx. conv.	171	168	172	198	206	273	411

In addition to the computation time per step, the number of iterations before convergence shown in table 2.3 also illustrate the computationally friendly features for high-dimensional problems. The number of iterations of SDP-based and approximate convex relaxation methods does not increase significantly as r, n and p individually increase. Interestingly, the iteration numbers of these convex relaxation methods changed in a different fashion. That of the SDP-based method slightly increases with an increase in n and p, and slightly decreases with increase in r, while that of the approximate convex relaxation showed the opposite results. This leads to similar growth rates against r between these two methods as shown in Fig. 2.3a, where the empirical computational complexity is less than $\mathcal{O}(r^4)$ for the SDP-based method and more than $\mathcal{O}(r^3)$ for the approximate convex relaxation method. Moreover, Fig. 2.3c shows that this difference leads to growth rates slightly higher than $\mathcal{O}(n)$ for the SDP-based method and lower than $\mathcal{O}(n^3)$ for the approximate convex relaxation method. This unexpected efficiency of the

SDP-based method leads to the better scalability with respect to n over the approximate convex relaxation, which is the other convex relaxation method. With regard to the increase in p, the iteration numbers of SDP-based method increased only slightly, while that of the approximate convex relaxation method showed the increase around $p = 20$, corresponding to the increase in the total run time in Fig. 2.3e.

As shown in Figures 2.3b, 2.3d and 2.3f, almost the same objective values were obtained by the SDP-based and the approximate convex relaxation methods, which implies a good agreement of the solutions of these relaxation methods. They yielded better or comparable objective function values compared with the greedy methods, which were up to twice as high except for $p = 1, 2$ cases. Those obtained by the gradient greedy method, on the other hand, give an inferior impression as a selection method compared with the other methods for $p \leq r$ cases. This is possibly due to the difficulty in ensuring the observability with a limited number of sensors especially $r > p$, which should lead to an unstable calculation of the gradient Eq. (2.20). The gradient greedy algorithm is concluded to have poor performance in achieving the observability, especially when the convex objective function is hardly approximated by its linear tangent due to large r.

The discussion above illustrates that the use of the SDP-based method seems to be the most favorable among the methods compared in this experiment. The use of the pure greedy method is also a reasonable choice due to its shorter computation time, straightforward implementation, and reasonable performance, which returns half the objective function value of the convex relaxation methods in the experiments.

2.3.2 Results for data-driven system derived from real-world experiment

An example of a practical application has been done using the experimental dataset of flow velocity distribution around an airfoil. The data used herein are found in (Nonomura et al., 2021b,c), which were previously acquired by particle image velocimetry (PIV) in a wind tunnel. A brief description of the experiment is shown in table 2.4 and refer to the original article of the authors for more specific information. The snapshots taken in the experiment quantify

velocity vectors that span the visualized plane, *i.e.*, two components on a two-dimensional grid of $n = 9353$ points as depicted in Fig. 2.4. Only the streamwise components (the direction shown

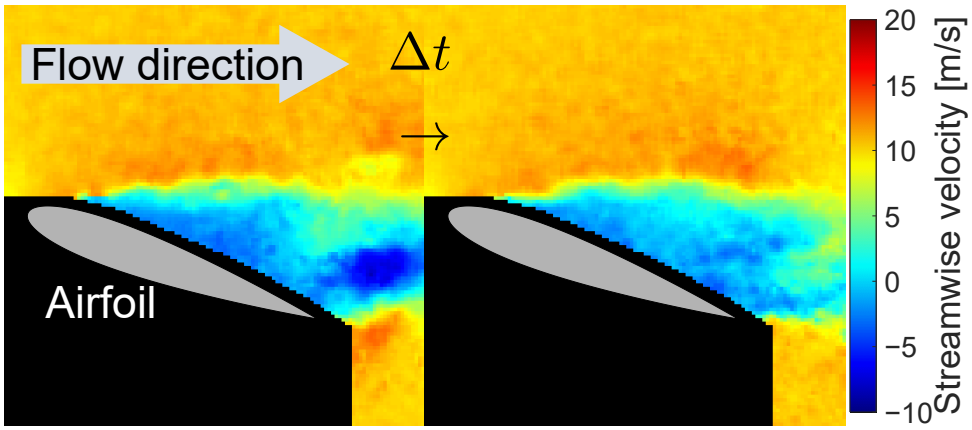


FIGURE 2.4: Visualized flow around NACA0015 airfoil (Nonomura et al., 2021b,c). Streamwise components in a time-series measurement are shown and used in the demonstration to construct the reduced-order state variables Eq. (2.21). ($\Delta t \approx 0.01$ seconds in this figure.)

Airfoil	NACA0015 (Chord length 100 mm)
Wind tunnel	Recirculating low-speed wind tunnel
Flow speed	10 m/s
Angle of attack	18°

(A) Wind tunnel experiment

Sampling rate	5,000 Hz
Spatial sample	9,353 points
Snapshot sample	2,000 snapshots × 5 sets

(B) Acquisition condition of velocity distribution

TABLE 2.4: Brief description of PIV data (Nonomura et al., 2021b,c)

by the arrow in Fig. 2.4) are used, and the ensemble averages over m snapshots are subtracted at each measurement location, that is, averages for each pixel of the calculated velocity image.

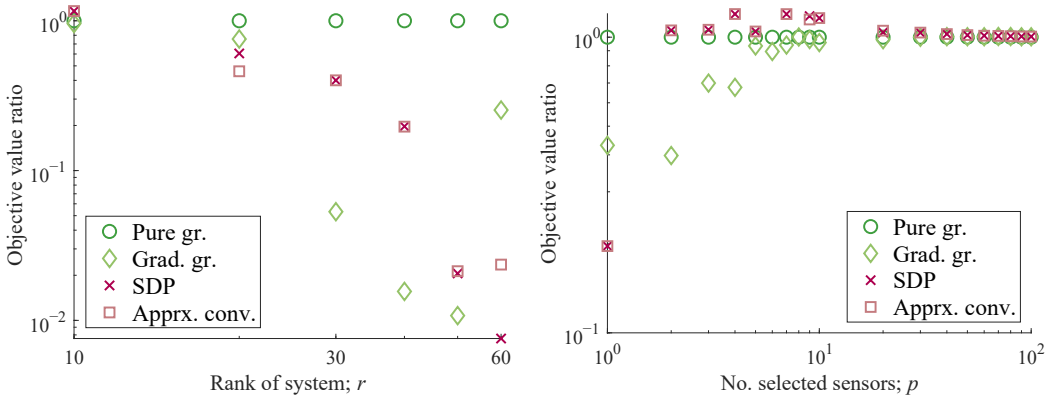
As tried in Section 2.3.1, a linear representation $(\mathbf{C}, \mathbf{A})$ is first derived. The data-driven system identification procedure is based on the modeling method of (Brunton and Kutz, 2019, January; Nankai et al., 2019). Here, the snapshots of the velocity field are reshaped to $(n \times 1)$-dimensional m vectors, $\mathbf{y}_1, \ldots, \mathbf{y}_m$, then the data matrix is defined by $\mathbf{Y} := [\mathbf{y}_1, \ldots, \mathbf{y}_m] \in \mathbb{R}^{n \times m}$. The proper orthogonal decomposition (POD) is then adopted and the data are projected onto the subspace of the leading r POD modes (Berkooz et al., 1993), resulting in

$\mathbf{Y} \approx \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ ($\mathbf{U} \in \mathbb{R}^{n\times r}, \mathbf{\Sigma} \in \mathbb{R}^{r\times r}, \mathbf{V} \in \mathbb{R}^{m\times r}$, respectively). The measurement matrix $\mathbf{C}$ is defined by $\mathbf{U}$, which consists of r left singular vectors related to the largest r singular values. These left singular vectors illustrate the dominant spatial coherent structures, while the right vectors represent temporal coherence, and the singular values are the amplitudes of these modes. The use of POD in our study is intended for a more fundamental discussion based on a linear system representation where arbitrary order low-rank systems are derived from high-dimensional measurement data.

The r-dimensional state variable vector $\mathbf{x}$ is given by

$$[\mathbf{x}_1, \ldots, \mathbf{x}_m] := \mathbf{\Sigma}\mathbf{V}^\top \in \mathbb{R}^{r\times m}. \tag{2.21}$$

After introducing $\mathbf{X}_1 = [\mathbf{x}_1, \ldots, \mathbf{x}_{m-1}]$ and $\mathbf{X}_2 = [\mathbf{x}_2, \ldots, \mathbf{x}_m]$, the system matrix is then expressed by the linear least squares via pseudo-inverse operation $\mathbf{A} = \mathbf{X}_2 (\mathbf{X}_1)^\dagger$. The system matrix $\mathbf{A}$ is manipulated to have eigenvalues less than unity, and therefore the considered systems are stable. As discussed in Section 2.3.1, the comparisons are illustrated using the objective



(A) Relative determinant value for various r ($p = 20, n = 9353$)

(B) Relative determinant value for various p ($r = 10, n = 9353$)

FIGURE 2.5: Optimization results for a system based on the experimental data of fluid dynamics.

function values obtained. The sensor candidate set corresponds to each pixel obtained by the experimental visualization, and thus, its size is fixed to $n = 9,353$. The state variable vector is set to have $r = 10, \ldots, 60$ components as a result of order reduction at different thresholds of truncation, while p is fixed to 20. Moreover, the results with respect to various p values are also

provided, where $p = 1, 2, \ldots, 10, 20, \ldots, 100$ with $r = 10$ fixed. The original dataset consisting of the $10,000$ snapshots is divided into $m = 2,000$ consecutive snapshot parts, and therefore, the following results are the five-fold average.

The values of the objective function show similar trends to Figures 2.3b and 2.3f for the gradient greedy algorithm, while a large degradation is observed in the convex relaxation methods for $r > p$. Surprisingly, those of the convex relaxation methods are 2-100 times lower than those of the pure greedy method. This trend is also partially observed for the $p = 1$ case in Fig. 2.5b, where the pure greedy selection corresponds to the true optimum for this configuration by definition. This clear degradation in the convex relaxations has not yet been explained in this study and may be due to the ill-condition of low-order approximation or the numerical error. Similar discussions have been conducted for the previous study (T. H. Summers et al., 2016). The cause should be investigated in detail for a practical application with real-world data in the future study.

The above results, together with those in the previous subsection, illustrate that the pure greedy method shows a more stable performance, while the convex relaxation methods show wide variations in their performance. Therefore, the pure greedy method can be considered as the most appropriate choice of method, taking into account the computation time. It should be noted that the present study also shows the prominent performance of the SDP-based method in terms of the objective function and the computation time in the randomly generated general problem settings. As long as the additional computation time is acceptable, the present authors recommend to try both the pure greedy and the SDP-based methods and to use a better sensor set when selecting the sensors based on the observability Gramian.

2.4 Conclusions

This research investigates approximate but efficient optimization of the observability Gramian of a discrete-time linear dynamical system to realize effective state estimation and control. This study offers two novel approaches to sensor selection that reduce the order of computation cost with respect to the dimension of the system compared to those in the literature. First, a convex relaxation with an approximate barrier function is solved by a customized Newton method. An

efficient randomization technique accelerates the optimization of the relaxed objective function. The other approach employs a linear approximation of a pure greedy evaluation of the matrix determinant as the objective function. The maximum of the gradient of the objective function is sought in greedy iterations.

These proposed selection methods are characterized using examples of a synthetic linear system and a data-driven representation of a state-space model obtained from a real-world dataset. A comparison covers previously presented methods in the literature in addition to the two proposed methods, which are a convex relaxation method denoted as semidefinite programming and a greedy selection with iterative naive evaluation of the objective function. An empirical comparison of the computation time is provided with respect to the system dimension parameters, such as the sizes of a state vector, available measurements, and those selected. The different trends in the increase of each selection method elaborate the effectiveness of each selection method in parallel with the comparison of the optimization performance comparisons. The proposed gradient greedy method obtains a moderate solution with orders of magnitude speedups compared with the other three methods, although the resulting sparse measurements are prone to unobservability. This is obvious when the number of selected sensors does not exceed the dimension of state variable of a system. A similar aspect is found for the convex relaxation methods, including the SDP-based and the approximate convex relaxation methods. They achieved better solutions than the greedy methods in a synthetic data, whereas the real-world example discloses their unstable optimization results especially for a large scale system. In such less observable situations by the sparse measurements, a naive implementation of the pure greedy method is the most reasonable choice in terms of the optimization metric obtained, yet the computation time is supposed to exceed the others.

Chapter 3

Construction of Linear Flow Model from Compressive Measurement

It is not straightforward to construct a model of fluid flow, as typified by nasty turbulence. For an aircraft, the dynamical properties of the flow change drastically depending on flow conditions such as its angle of attack or speed. The high spatio-temporal resolution of the data is also essential for the construction of informative approximate models. It will require a large number of numerical computations or experimental acquisitions. Therefore, from the perspective of this study, it is regarded as virtually infeasible to obtain a reliable flow model in advance. This chapter will introduce the method of flow model construction from sparse measurements as a realistic extension of the methodologies in the previous chapter.

3.1 Background

Section 3.1.1 reviews the data-driven description of complex phenomena by using dynamic mode decomposition (DMD), which will be focused on throughout this study. Section 3.1.2 also revisits methods that extend the methodology of DMD to measurements with resolutions too insufficient to delineate the nature of the physical phenomena of interest, which is common in real-world applications. Section 3.1.3 introduces several topics related to machine learning methods, especially focusing on hybrid methods of model-based and data-driven techniques.

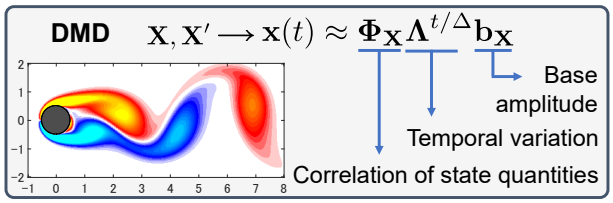


FIGURE 3.1: DMD: Analysis of obtained time-series data to extract modal structures

3.1.1 Dynamic mode decomposition and data-driven model construction of fluids

Dynamical systems have been translated into less complicated reduced-order models as introduced in the preceding chapter. Dynamic mode decomposition (DMD) is one of the data-driven methods for constructing reduced-order models (Kutz et al., 2016b, November; Schmid, 2010), which decomposes the data obtained by measurements or simulations into a series of dynamic modes. The method of DMD was originally presented to characterize the behavior of an apparently complex dynamical system. Simplified models in linear form describe the dynamical behavior in low-dimensional invariant subspaces. The structure of DMD is qualitatively shown in Fig. 3.1, where these mathematical notations are given in the following Section 3.2.1. It is often discussed in connection with the Koopman operator theory (Brunton et al., 2016a, 2022; Colbrook et al., 2023; Korda and Mezić, 2018; Mezić, 2013). Koopman operator stipulates the linear evolution of the observables of a dynamical system, which are obtained by measuring the system. These observables can take any form of the variables obtained by the measurement; therefore, the Koopman operator is infinite dimensional. This entails an infinite number of eigenfunctions, which makes it prohibitive to analyze the behavior of target phenomena with the operator. Therefore, finding the most influential low-order subspaces, or perhaps manifolds, has been the most essential step of the approach. The dynamic representation which is defined on the low-order structure transforms the original governing system of interest into an approximate reduced-order model.

It would be worthwhile to introduce some research that has exploited the technique of DMD in the analysis of complicated phenomena. The methods of DMD originated in the studies of fluid

dynamics. Therefore, it has widely facilitated the analysis of complex fluid behaviors as found in, although it is almost impossible to conduct an exhaustive survey, the analysis of bifurcation of flow regimes (Kramer et al., 2017), supersonic flows by a large-scale simulation (Ohmichi et al., 2018), in the prediction of unsteady flows (X. Zhang et al., 2022) and in the reconstruction of the density field from the visualized flow (Shigeta et al., 2023). Furthermore, applications covered in interdisciplinary fields can be found in diverse areas such as analysis of finance (Mann and Kutz, 2016), multiplayer sports (Fujii et al., 2017), epidemiology (Proctor and Eckhoff, 2015), and transportation (Y. Yu et al., 2021). The modal representations obtained through DMD are also applied to the control of a system (Korda and Mezić, 2018; Proctor et al., 2016).

3.1.2 Application of DMD on spatially-sparse flow measurement

Practical applications occasionally emphasize the use of spatially sparse measurements for the effective construction of low-rank models of physical phenomena. A recognizable example is flight operations of aircrafts, where operators grasp the state of the surrounding atmosphere using measurement instruments to manage the safety of flight. The matter is the state of atmospheric flow, but these actual measurements accompany only a limited number of point sensors mounted on the surface, which collect unsteady fluctuations in physical observables of the surroundings.

Similar situations are induced even in flow measurements with a laboratory apparatus. Two scenarios are noted in the practical acquisition and processing of the high-resolution data, which handles so many entries to represent detailed information in both space and time:

1. It is often encountered in practical and scientific experiments that the measurement apparatus cannot afford the spatial and temporal resolution simultaneously. For experiments of fluid dynamics, visualization techniques capture fluid motions under the trade-off between the resolutions of time and space.

2. The data often tend to be devastatingly high dimensional, even if the data rarely contain information about the phenomenon. This often exhausts computational resources and acquisition costs.

The data obtained in such situations often miss spatial coverage or temporal resolution. This situation is in contrast to the usual DMD-based analysis, which assumes the use of data that have sufficient spatio-temporal resolution. This enlightens the difficulty in the reduced-order modeling from insufficient measurements according to known procedures because the construction of the model becomes an inference as an underdetermined, ill-conditioned inverse problem.

Therefore, efficient approaches for DMD have been implemented to amend these discords, which incorporate ideas of compressive sensing into the DMD procedures. By including the recovery process from temporally compressed measurements, DMD-based models represent unsteady phenomena that change rapidly from data with insufficient temporal resolution (Guéniat et al., 2015; Hemati et al., 2014; Tu et al., 2014a), which are leveraged in the analysis of high-speed phenomena such as impinging jet (Ohmizu et al., 2022). (Brunton et al., 2014; Erichson and Donovan, 2016; Erichson et al., 2019; Ohmichi et al., 2018) Meanwhile, other methods have also been advocated to compress vastly large data and make the following computation easier (Bai et al., 2020; Brunton et al., 2015; Kramer et al., 2017; Manohar et al., 2019). Finally, several methods have also been proposed to compensate for spatially compressive measurements, which is considered the most difficult, but realistic case.

Compressed Sensing DMD (shortened to CS-DMD in this study) is one of the methods proposed earlier by Brunton et al., 2015. The same perspective is extended to construct the control strategy (Bai et al., 2020), introducing compressive sensing to DMD with control of Proctor et al., 2016. The procedure of CS-DMD is briefly introduced by Fig. 3.2, where the notations therein are provided later in Section 3.2.3. Although this method actually works in the absence of spatial resolution in the data, its applicability to real-world scenarios remains skeptical. The reason of this concern is that, in short, the methodology is completely model-free and too robust to be facilitated for specific applications. One should draw attention to some observations on this earlier study, which are discussed in detail in Section 3.2.4. First, it still requires a large number of observations for the reconstruction of moderately simple flow structures. The other thing is that random projections are often used for signal reconstruction from compressed measurements, which is encouraged in CS theory.

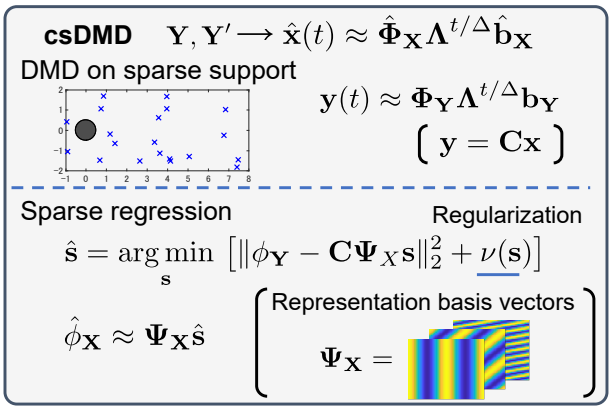


FIGURE 3.2: CS-DMD: Acquiring modal representations for entire field from sparse measurements

These aspects possibly require too stringent conditions on us, although they offer fascinating performance guarantees from the theory.

3.1.3 Model-based methods

The model-based formulation has received wide attention in numerous fields. It is also applicable to this study, where the modified implementation is performed for a better estimation with CS-DMD. In a broader context of machine learning, several established methods are known to have improved the effect of models by provoking model-based schemes. For example, the pioneering work of Physics Informed Neural Networks (PINNs) can be found in a seminal study of Raissi and Karniadakis, 2018; Raissi et al., 2019, 2020, which is also employed in the field of fluid mechanics. By training a neural model to minimize the residual of the imposed physical equations, the output of the model becomes more relevant to the physics underlying the training data. Various applications are found in the construction of a surrogate model for thermal analysis (Tanaka and Nagai, 2023), electrophysiological mapping of cardiac activation, approximation of high-speed compressible flows (Mao et al., 2020), and state inference from partial observation of flow (Cai et al., 2021).

Another approach can be found that is extended to the model-based DMD, which is more deductive compared with the earlier examples of PINNs. In the process of DMD, it is demonstrated by Baddoo et al., 2023 that several basic principles of physics can be embraced by the

obtained model made of dynamical modes by promoting "procrustes" optimization problem. The matrix representations are sought in the subspace that satisfies these properties. This is indeed model-based, yet this imposes general properties of physical laws that can be applicable to many instances without the specification of physical equations.

Prior to these machine learning methods, there are also known studies of model-based estimation in the context of compressive sensing, which introduced a deductive aspect to its inverse inference. R. G. Baraniuk et al., 2010 developed a new notion of structured sparsity, which is in contrast to general sparsity. The inference step of compressive sensing is conducted in a smaller subspace of sparse signal according to the predefined structure. The similar approach is converted to the probabilistic one in L. Yu et al., 2012, by imposing an additive term on the sparsity-promoting component in the prior distribution. This structured inference is supposed to be effective in our focus; therefore, the appropriate structure and the modified procedure of CS-DMD are sought in this chapter.

3.1.4 Objective and contribution

For obtaining better estimation results than the original CS-DMD implementation of Brunton et al., 2015, several tailored techniques are incorporated to construct a flow model with spatially sparse measurements. Here, the following requirements should be emphasized to meet the current problem settings and actual applications of CS-DMD;

- No prior training data that consists of spatially resolved measurements,
- Suppression of a noisy reconstruction for a reduced order representation of physical phenomena,
- Reproduction of structures from a smaller number of measurements than presented in the previous general formulations of CS-DMD.

The absence of prior training data is assumed first. This is advocated in the application of complex fluid phenomena, where spatially sparse measurements can be the only feasible means of data acquisition. The second aspect is often encountered in the processing of experimental

data, where it reduces the contamination of measurement noise by several filtering methods. The last claim is one of the most fascinating topics in the machine learning community in recent years, as introduced in Section 3.1.3.

Therefore, the objective of this study is to reinforce the previously presented general form of CS-DMD, which reconstructs the global distribution from compressive sampling. The specific manifesto posed in this chapter is;

- Reliable estimation excluding representation basis vectors of higher wavenumber,
- Structural CS inference, which can be posed incrementally and additively according to the different levels of postulations of underlying phenomena, ranging from general laws to specific observations that are defined by the flow configurations.

These notions are illustrated in Fig. 3.3.

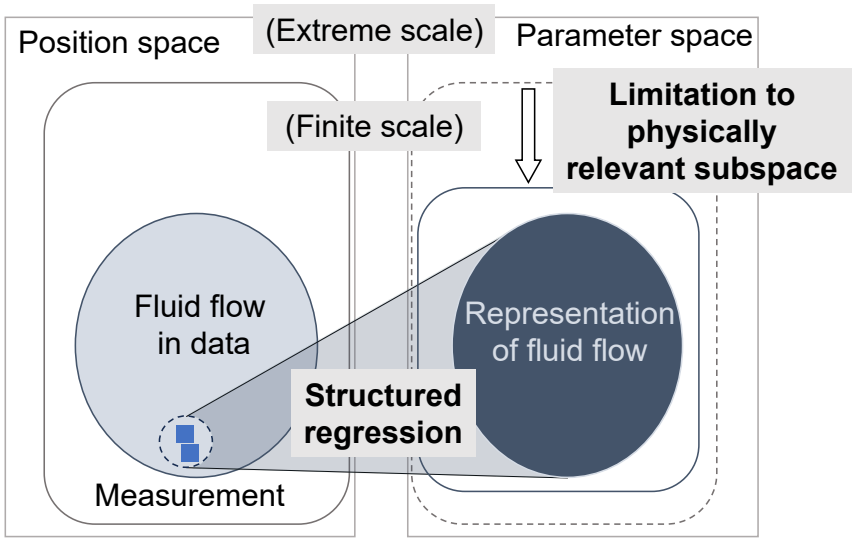


FIGURE 3.3: Schematic of the applied CS-DMD in this study

3.2 Methods

Section 3.2.1 reviews the data-driven description of complex phenomena by use of dynamic mode decomposition (DMD) to elaborate the interest in DMD which is focused in the rest of this study. Section 3.2.2 provides fundamental materials from studies of compressive sensing to illustrate the

method of CS-DMD, which was proposed in Brunton et al., 2015 and its application as in Bai et al., 2020. The proposed approaches of the study are presented descriptively in Section 3.2.4 using some illustrative examples. The main two ideas are defined separately in the subsequent sections, as shown in Section 3.2.5 and Section 3.2.6.

3.2.1 Dynamic mode decomposition

This section delivers the process and assumptions of DMD using general notations, which is followed by the specific variants of DMD where the measurement is only limited to some spatially sparse supports. Note that it is sufficient to focus on the modal representation that is obtained by the variants of DMD. Therefore, one may refer to the existing literature remarked hereafter for more detailed descriptions of the procedure.

First, T samples of measurement vectors are concatenated as a pair of data matrices $\mathbf{X}$ and $\mathbf{X}'$, where the shifted matrix $\mathbf{X}'$ reflects the development from $\mathbf{X}$ after a certain time interval Δ;

$$\mathbf{X} \coloneqq [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T], \quad \mathbf{X}' \coloneqq [\mathbf{x}'_1, \mathbf{x}'_2, \ldots, \mathbf{x}'_T], \tag{3.1}$$

where T is the sample size and $\mathbf{x}_t$, $\mathbf{x}'_t \in \mathbb{R}^N$ stand for measurements at instance t and its shifted one, respectively. The only requirement here is that the interval Δ is constant and sufficiently small, while each $\mathbf{x}_t$ is allowed to be recorded at an arbitrary time. The two-dimensional (or further high-dimensional) measurements are to be allocated as tall vectors.

DMD boils down the specific temporal development of the coherent structures between variables of those matrices, which are characterized by the eigenvalues and eigenvectors of $\mathbf{A}$. Each eigenvalue $\lambda^{(j)} \in \mathbb{C}$ ($j \in \{1, 2, \ldots, r\}$) reflects a periodic motion of the amplitude with a certain rate of dump/growth for the correlated structure, often referred to as the spatial mode,

$\phi_{\mathbf{X}}^{(j)} \in \mathbb{C}^N$. Finally, the reconstructed snapshot at arbitrary time $t \in \mathbb{R}$ is expressed by all modes,

$$\mathbf{x}(t) \approx \mathbf{\Phi}_{\mathbf{X}} \mathbf{\Lambda}^{t/\Delta} \mathbf{b}_{\mathbf{X}}, \tag{3.2}$$

$$\mathbf{\Phi}_{\mathbf{X}} = \begin{bmatrix} \phi_{\mathbf{X}}^{(1)} & \cdots & \phi_{\mathbf{X}}^{(r)} \end{bmatrix}, \tag{3.3}$$

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda^{(1)} & & \\ & \ddots & \mathbf{0} \\ \mathbf{0} & & \\ & & \lambda^{(r)} \end{bmatrix}. \tag{3.4}$$

There are some convincing methods to obtain spatial modes and corresponding eigenvalues, such known as Exact DMD (Tu et al., 2014b), Optimized DMD (K. K. Chen et al., 2012), Sparsity-promoting DMD (Jovanović et al., 2014) and Total Least-squares DMD (S. T. M. Dawson et al., 2016). Each of these methods has different approaches to improve the representation while maintaining a parsimonious model. Users can carefully choose these examples based on the assumptions and computational resources available. The initial amplitudes for each mode $\mathbf{b}_{\mathbf{X}} \in \mathbb{R}^r$ in Eq. (3.2) are arbitrarily determined to represent the dominance and phase of periodic fluctuations of each mode. One simple approach is to assign them by pseudo-inverse operation using $\mathbf{\Phi}_{\mathbf{X}} \in \mathbb{R}^{N \times r}$ and the initial state vector $\mathbf{x}$ corresponding to $t = 0$, as $\mathbf{b}_{\mathbf{X}} = (\mathbf{\Phi}_{\mathbf{X}})^{\dagger} \mathbf{x}_0$. Also note that the modal representation includes conjugate modes, as the measurement data $\mathbf{X}$, $\mathbf{X}'$ are assumed to be real matrices.

3.2.2 Compressive sensing theories

This section provides several fundamental elements of compressive sensing theories to make the following sections more comprehensive. The most remarkable aspect here (and only the thing to be remembered) is that one may be able to leverage the techniques for the recovery of the signal of great dimension while using a small number of measurements, mentioned as one of the objectives of this study. It is briefly summarized in this section to introduce related notions to present such as remarked by the following items:

- A general formulation for measurement and recovery of signals as an inverse inference

- Ideas of sparsity and compressibility
- Recovery algorithms
- Requirements for recovery with guaranteed accuracy

It is encouraged to seek some fruitful notes in R. Baraniuk, 2007; Elad, 2010; Patel and Chellappa, 2013; Sankaranarayanan et al., 2016 for abundant literature on this topic.

A linear measurement equation is here revisited using the same notations of matrices in Chapter 2 while dropping a measurement noise term;

$$\mathbf{y} = \mathbf{C}\mathbf{x}, \tag{3.5}$$

with $\mathbf{C}$ representing arbitrary linear measurement process. The Shannon–Nyquist sampling theorem clarifies the necessary conditions for the complete recovery of signal $\mathbf{x}$ from a given finite-length measurement $\mathbf{y}$. It can be considered as a fundamental theorem that indicates the upper limit of the amount of information obtained from the measurement process Eq. (3.5). It is qualitatively described as the following: For restoration of the given signal, measurements must be made at a sampling rate that is at least twice the frequency of the signal. The minimum frequency that satisfies the condition is referred to as "Shannon-Nyquist rate", and this often makes the equipment and data storage of many data acquisitions intensely costly, such as visualization of high-speed flows. It can be interpreted in the same way for spatial sampling; therefore, it further intensifies the overall cost of measurement.

As a remedy for the obstacles in signal recovery, ideas of compressive sensing provide a way to recover certain signals from fewer measurements than specified by the Shannon–Nyquist rate. First, consider a "sparse" signal $\mathbf{x} \in \mathbb{R}^N$ that is represented by a small number of representation basis vectors $\psi_i \in \mathbb{C}^N \, (\forall \, i \in \{1, \ldots, N\})$, where the number is expressed by $K \ll N \in \mathbb{N}$. Without loss of generality, a sparse signal can be expressed as

$$\mathbf{x} = \sum_{i=1}^{K} s_i \psi_i, \tag{3.6}$$

where s_i is the weighting coefficients with respect to the i-th vector. It should be noted that $K \leq N$ has to be ensured for the general class of signals and given basis vectors, for completeness. $s_i\,(\forall\, i \in \{K+1, \ldots, N\})$ takes zero due to Eq. (3.6). Therefore, if the signal x is sparse with respect to basis vectors ψ_i, the degree of freedom is mere K, which is much less than the original dimension N.

In compressive sensing, it is designed to recover the reduced representation of sparse signals from the measurements and the subsequent inference methods. Sparse representation of signal Eq. (3.6) substitutes for Eq. (3.5), which is referred to as a *sketch*, as;

$$\mathbf{y} = \sum_{i=1}^{K} \mathbf{C}\psi_i s_i = \mathbf{C}\mathbf{\Psi}\mathbf{s}. \tag{3.7}$$

Here, $\mathbf{s} \in \mathbb{C}^N$ is a sparse coefficient vector with at most $K \ll N$ nonzero entries and $\mathbf{\Psi}$ is a library matrix with orthogonal representation basis vectors.

One of the main approaches to determine the sparse components of $\mathbf{s}$ from the measurements is to construct an optimization problem by balancing the mean squared error and the belief in the sparsity of the signal. One can set an objective function as;

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \left[\|\mathbf{C}\mathbf{\Psi}\mathbf{s} - \mathbf{y}\|_2^2 + \nu\,(\mathbf{s}) \right] \tag{3.8}$$

$$\hat{\mathbf{x}} = \mathbf{\Psi}\hat{\mathbf{s}}, \tag{3.9}$$

where the second term is a function of $\mathbf{s}$ to penalize its structure. One example is to construct a convex objective function seen as the least absolute shrinkage and selection operator (Lasso);

$$\nu\,(\mathbf{s}) = c\|\mathbf{s}\|_1, \tag{3.10}$$

with a positive constant c, or to control the ℓ_0 norm by use of restrictions of an indicator function,

$$\nu\,(\mathbf{s}) = \iota_{\|\mathbf{s}\|_0 \leq s_{\max}}\,(\mathbf{s}) = \begin{cases} 0 & (\|\mathbf{s}\|_0 \leq s_{\max}) \\ \infty & (\|\mathbf{s}\|_0 > s_{\max}) \end{cases} \tag{3.11}$$

where $\iota_C(\mathbf{s})$ regulates $\mathbf{s}$ to be bounded in some subspace C, and here it confines the number of the nonzero entries in the coefficient vector $\mathbf{s}$ to $s_{\max}$. Some algorithms solve convex optimization Eq. (3.10) effortlessly. These methods of continuous relaxation such as Eq. (3.10) require a rounding procedure to make the solution back to the sparse domain, which is called "polishing" of the solution in Nagata et al., 2021, 2022b. First, the sparse support of the solution vector is determined by an arbitrary threshold with respect to each entry, of which the process is denoted as c. Indices of nonzero entries, which is denoted as the support $\mathrm{supp}(\hat{\mathbf{s}})$, are identified by this criterion, since a relaxed solution usually contains some nonzero values in all entries. The sparse solution is finally recalculated for the inferred sparse support. In this study, the least-squares estimation is placed after identifying the sparse support as previously engaged in Nagata et al., 2021. By relying on the intuitive description, this process is summarized as;

$$\hat{\mathbf{s}} \leftarrow \mathcal{P}(\hat{\mathbf{s}}; \mathbf{y}, \mathbf{\Psi}, c) = \begin{cases} \text{Identify sparse support } \mathrm{supp}(\hat{\mathbf{s}}) \text{ with arbitrary criterion } c, \\ \text{Obtain reduced solution: } \hat{\mathbf{s}}' = \left(\mathbf{\Psi}\left(\mathbf{P}\left(\mathrm{supp}(\hat{\mathbf{s}})\right)\right)^\top\right)^\dagger \mathbf{y}, \\ \text{Retrieve solution vector: } \hat{\mathbf{s}} = \left(\mathbf{P}\left(\mathrm{supp}(\hat{\mathbf{s}})\right)\right)^\top \hat{\mathbf{s}}'. \end{cases} \tag{3.12}$$

Here, a permutation $\mathbf{s}' = \mathbf{P}(\mathcal{I})\mathbf{s}$ extracts the entries of $\mathbf{s}$ labeled by the index subset $\mathcal{I}$. Its inverse operation is given by $\mathbf{P}(\mathcal{I})^\top \mathbf{s}$, where the entries in $\mathbf{s}$ are placed in the zero vector of the same dimension as $\mathbf{s}$. The overall operation for polishing is denoted as $\mathcal{P} : \mathbb{C}^N \to \mathbb{C}^N$. It should be noted that the sparse regression methods employed in the following sections implicitly assume the use of the polishing of Eq. (3.12).

These techniques can be extended to a broader class of signals where the signal is not completely sparse, as is often the case in real-world applications. An example is that the measurement Eq. (3.5) is corrupted by white noise, which results in a bandwidth broader than the pure signal, while still some remarkable peaks can be observed because of the sparsity of the underlying signal. The "compressible signal" is here denoted as a vector with a few terms of large coefficients,

while the rest as rather small;

$$\mathbf{x} = \sum_{i=1}^{N} s_i \psi_i, \tag{3.13}$$

where s_K is arranged in descending order and $\|s_K\| \leq CK^{-a}$ is observed for a given positive constant C and $a > 1$. This power-law decrease of the coefficients reflects the dominance of components with respect to the small group of representation basis vectors. A representation of $\mathbf{x}$ by the largest K terms can keep its approximation error within the range that obeys a power law.

It should be noted that the accuracy of signal reconstruction can be guaranteed within a certain degree with a high probability if the sketch obtained by the CS matrix satisfies several characteristics and conditions. The feature most commonly pronounced in publications is the restricted isometry property *(RIP)* (Candès et al., 2006), which is invoked by $\mathbf{C}$ and $\mathbf{\Psi}$;

$$(1 - \delta_K)\|\mathbf{s}\|_2^2 \leq \|\mathbf{C}\mathbf{\Psi}\mathbf{s}\|_2^2 \leq (1 + \delta_K)\|\mathbf{s}\|_2^2, \tag{3.14}$$

with small constant δ_K related to the sparsity K. This condition (for ℓ_2 norms) is a sufficient condition for the successful recovery of K sparse signals with $\mathbf{C}$ and $\mathbf{\Psi}$. Compressive measurements that maintain this property ensure the best estimate whose estimation error is bounded by a small constant related to δ_K. It can also be described that the norm of the K-sparse signal is preserved through measurements and thus preserves the information of the underlying signal as mentioned in Candès, 2008 and Elad, 2010, Section 5.2.3).

Separately, one can focus on the conditions posed on the CS matrix to meet the conditions for successful recovery by CS. For a smaller δ_K of the RIP under probabilistic description, it suffices to construct $\mathbf{C}$ as a random matrix with p rows to project the signal to compressive measurements (Berinde et al., 2008; Gilbert and Indyk, 2010), respectively. Designing measurements that satisfy this property can be defined, for example, by independent and identically distributed random variables drawn from a zero-mean Gaussian distribution with a variance of $1/N$. The condition of RIP is satisfied with high probability if the number of rows p is greater

than a well-known criterion;

$$p \geq cK \log(N/K), \tag{3.15}$$

with a small constant c, the sparsity K and the dimension N of the original vector (Candès et al., 2006; DeVore, 2007). If the signal $\mathbf{x}$ is termed as K-sparse and the measurement of size $p \geq cK \log(N/K)$ is contaminated by measurement noise at variance σ^2, the error resulting from the reconstruction using algorithms can be bounded as;

$$\|\mathbf{s} - \hat{\mathbf{s}}\|_2 \leq C\sigma, \tag{3.16}$$

with positive constant C. For the K-compressible signal, the mean squared error is bounded as

$$\|\mathbf{s} - \hat{\mathbf{s}}\|_2 \leq C_1\sigma + C_2\|\mathbf{s} - \mathbf{s}_K\|_2 + \frac{C_3}{\sqrt{K}}\|\mathbf{s} - \mathbf{s}_K\|_1, \tag{3.17}$$

where the K-term approximate of $\mathbf{s}$ is denoted as $\mathbf{s}_K$ to invoke the truncation error of compressible, yet non-sparse signals.

3.2.3 Compressive estimation of sparse coefficients for reconstruction of spatial distribution of DMD mode

In the absence of spatial resolution, one may tackle reconstructing the spatially resolved DMD representation using data of spatially sparse measurement(denoted as "sparse" measurements, hereafter). Here, the sparse measurements are related to the original, spatially resolved data ("dense" data, in contrast) according to the predefined measurement matrix $\mathbf{P} \in \mathbb{R}^{p \times N}$. The permutation matrix $\mathbf{P}$ explicitly indicates the point measurement extracted from the dense data.

This notation is valid throughout this manuscript instead of using $\mathbf{C}$, while the argument specifying the indices is omitted here for brevity. Sparse measurements are extracted as;

$$\mathbf{Y} \coloneqq \mathbf{P}\mathbf{X}, \quad \mathbf{Y}' \coloneqq \mathbf{P}\mathbf{X}' \tag{3.18}$$

$$\left(\mathbf{P} \coloneqq \begin{bmatrix} \mathbf{e}_{i_1}, & \ldots, & \mathbf{e}_{i_p} \end{bmatrix}^\top\right), \tag{3.19}$$

where $\mathbf{e}_i \in \mathbb{R}^N$ is the unit vector with unity in the i-th entry indicating the measurement at location i. The sparse observation data $\mathbf{Y}$, $\mathbf{Y}'$ is processed by DMD exactly in the same manner of Eq. (3.4) to differentiate spatiotemporal coherent structures on its support, resulting with the DMD representation;

$$\mathbf{y}(t) \approx \mathbf{\Phi}_\mathbf{Y} \mathbf{\Lambda}^{t/\Delta} \mathbf{b}_\mathbf{Y} \tag{3.20}$$

$$\mathbf{\Phi}_\mathbf{Y} = \begin{bmatrix} \phi_\mathbf{Y}^{(1)} & \ldots & \phi_\mathbf{Y}^{(r)} \end{bmatrix}. \tag{3.21}$$

One foundation for the argument in Brunton et al., 2015 is an assumption that can be expressed that the DMD mode for dense measurements indexed by j, $\phi_\mathbf{X}^{(j)}$, can be expanded with a small number of basis vectors in a transformation matrix $\mathbf{\Psi}_\mathbf{X} \in \mathbb{C}^{N \times N}$ defined for discretized Cartesian coordinate;

$$\phi_\mathbf{X}^{(j)} \approx \mathbf{\Psi}_\mathbf{X} \mathbf{s}^{(j)}, \tag{3.22}$$

where $\mathbf{s}^{(j)} \in \mathbb{R}^N$ stands for a sparse coefficient vector with at most $K \ll N$ nonzero entries.

Another lemma is also treated in Brunton et al., 2015 that the correspondence can be confirmed between the several dominating DMD modes on the sparse measurement support ($\mathbf{\Lambda}_\mathbf{Y}$, $\mathbf{\Phi}_\mathbf{Y}$) and that extracted from the dense measurements ($\mathbf{\Lambda}_\mathbf{X}$, $\mathbf{P}\mathbf{\Phi}_\mathbf{X}$), while allowing the permutation of the order;

$$\phi_\mathbf{Y}^{(j)} = \mathbf{P}\phi_X^{(j)} \approx \mathbf{P}\mathbf{\Psi}_\mathbf{X}\mathbf{s}^{(j)}. \tag{3.23}$$

In other words, it is supposed to preserve key features through sparse measurements $\mathbf{P}$. This is based on the assumption that the temporal fluctuation in the measurement is common, regardless of whether the measurement is complete or sparse. It should be noted that this assumption is occasionally violated for actual applications, though, since fluctuations due to the phenomenon of interest may be spatially local and, in addition, the inference of these structures is prone to measurement noise (S. T. M. Dawson et al., 2016).

Taking into account Eq. (3.22), each spatial correlation structure is reconstructed through the estimated coefficients $\hat{\mathbf{s}}^{(j)}$.

$$\hat{\mathbf{s}}^{(j)} = \arg\min_{\mathbf{s}} \left[\|\mathbf{P}\mathbf{\Psi}_{\mathbf{X}}\mathbf{s} - \phi_{\mathbf{Y}}^{(j)}\|_2^2 + \nu(\mathbf{s}) \right] \tag{3.24}$$

$$\hat{\phi}_{\mathbf{X}}^{(j)} \approx \mathbf{\Psi}_{\mathbf{X}}\hat{\mathbf{s}}^{(j)}, \tag{3.25}$$

where $\nu(\mathbf{s})$ is a regularization term that enforces sparsity in $\mathbf{s}$. Equation (3.24) is specifically rewritten to the use of indication function to the number of nonzero entries in the solution;

$$\begin{aligned} &[\mathrm{C}\ell_0\,(\mathrm{u})] \\ &\min_{\mathbf{s}\in\mathbb{C}^N} \left[\|\mathbf{P}\mathbf{\Psi}_{\mathbf{X}}\mathbf{s} - \phi_{\mathbf{Y}}^{(j)}\|_2 + \iota_{\|\mathbf{s}\|_0 \le s_{\max}}(\mathbf{s}) \right]. \end{aligned} \tag{3.26}$$

This minimization problem is denoted as "$\mathrm{C}\ell_0\,(\mathrm{u})$" to clarify the specific form used in this study. $\mathrm{C}\ell_0$ stands for the regression form with a "componentwise" sparsity with respect to $\|\mathbf{s}\|_0$ as Eq. (3.11), while (u), or (uniform), reveals the uniform use of the entire library matrix in contrast to the subsequent proposed methods in Section 3.2.5.

The discrete Fourier basis is assigned in Brunton et al., 2015 and our procedure follows this representation basis for the convenience of implementation with basic computation libraries, while the choice of other types of basis is also allowed without loss of generality, such as Wavelet as Krishnan et al., 2023, or some polynomial functions. One note here is that the incoherence between the measurement and the representation basis is satisfied, which is deeply related to the successful reconstruction. If one uses the Fourier basis library for the sparse representation, this condition is automatically satisfied if one combines the point measurement distributing spatially.

The process above renders the global distribution by extrapolating using the DMD on sparse supports and the representation by basis vectors. The algorithmic description is provided in Alg. 5.

Algorithm 5 Procedure of CS-DMD (Brunton et al., 2015)

Input: $\mathbf{Y}$, $\mathbf{Y'} \in \mathbb{R}^{p\times T}$, $\mathbf{P} \in \mathbb{R}^{p\times N}$, $\mathbf{\Psi_X} \in \mathbb{R}^{N\times N}$
Output: Reconstructed DMD representation ($\lambda_{\mathbf{Y}}$, $\hat{\phi}_{\mathbf{X}}^{(j)}$)
(Eq. (3.20)) Obtain DMD representation on sparse measurements: $\mathbf{Y}$, $\mathbf{Y'} \rightarrow (\lambda_{\mathbf{Y}}, \mathbf{\Phi_Y})$
(Eq. (3.23)) Find sparse coefficient vector: $\mathbf{P\Psi_X}$, $\phi_{\mathbf{Y}}^{(j)}$ by solving sparse regression problem Eq. (3.24),
(Eq. (3.25)) Reconstruct original DMD mode: $\phi_{\mathbf{X}}^{(j)} \approx \mathbf{\Psi_X s}^{(j)}$

The original work in Brunton et al., 2015 and its review in Kutz et al., 2016b, November revealed the results of the reconstruction using CS-DMD. For some toy examples, such as only a few Fourier modes configuring the data, the sparse solution is completely obtained by the inference of CS-DMD. Meanwhile, when applied to the reconstruction of fluid flows in Kutz et al., 2016b, November, Chapter 9 as treated in Section 3.3, the performance of the reconstruction is not as great as for the toy problems. Sufficient construction of the spatial distribution of DMD mode requires a lot of measurements up to $\mathcal{O}(10^3)$, although the compression rate was about 1%. Mounting such numerous sensors is technically infeasible in many applications, therefore, there is room for improvement, as mentioned by Bai et al., 2020.

3.2.4 Qualitative description of prior information for model-based estimation

This section provides key ideas to be incorporated into the original method of CS-DMD introduced in the preceding section. The description herein is intended to clarify the notions to be leveraged in the compressive reconstruction of the spatial distribution of DMD modes. Stricter definitions and tailored implementations in algorithms are deferred to the following sections.

One observation is, as pointed out in the preceding section and repeatedly, that the methods of CS-DMD in previous studies postulate compressive sensing theories in a strictly general sense. Therefore, there is still room to import methods for the compressive reconstruction of DMD modes tailored for fluid flow models. Aspects of the subject to be modified in this study are summarized as follows:

- Complete representation basis vectors that cover every form of spatial distributions in assumed spatial resolution, including fine fluctuations that are unrealistic for physical process
- General methods of compressive reconstruction that assume the sparsity in a general sense

The following statements describe the state of the art and the way of modification in the present study to slack the above feature. These features are reflected in the model obtained in a bit primitive way compared to the physics-informed approach introduced Section 3.1.3. Meanwhile, this direct incorporation of natural laws offers its effectiveness in model construction.

In short, the features considered so far to be represented by CS-DMD are summarized hierarchically;

Structural features incorporated to compressive sensing of CS-DMD

```
Characteristics of DMD representation in wavenumber domain
├─ (In-common features for physical phenomena)
│   ├─ Finite scale structure ... Nonuniform use of basis vectors
│   └─ Real-valued measurements ... Conjugate DMD modes
└─ Fluid flows
    ├─ Laminar vortex shedding
    │   ├─ Dominating advection ... Clustered nonzero coefficients
    │   └─ Symmetric flow ... Symmetric cluster at a certain scale
    ├─ Turbulent ... Scales in accordance to energy distribution
    │   └─ Broadband frequency spectra
    ├─ Shock waves ... Localized structure
    └─ (Other characteristics)
```

The features implemented in this study are highlighted by red and are technically denoted in the following sections. The weighting for sparse regression is first described in Section 3.2.5 for the refinement of the representation basis vectors, which are especially defined by selective weights of $0-1$ in our proposed method. The other idea of structured compressive sensing is discussed in Section 3.2.6 with some algorithmic descriptions.

Selective use of representation basis vectors

The first hypothesis is that the basis vectors representing spatial components do not need to be treated equally. The present work incorporates this idea to enjoy the efficient reconstruction

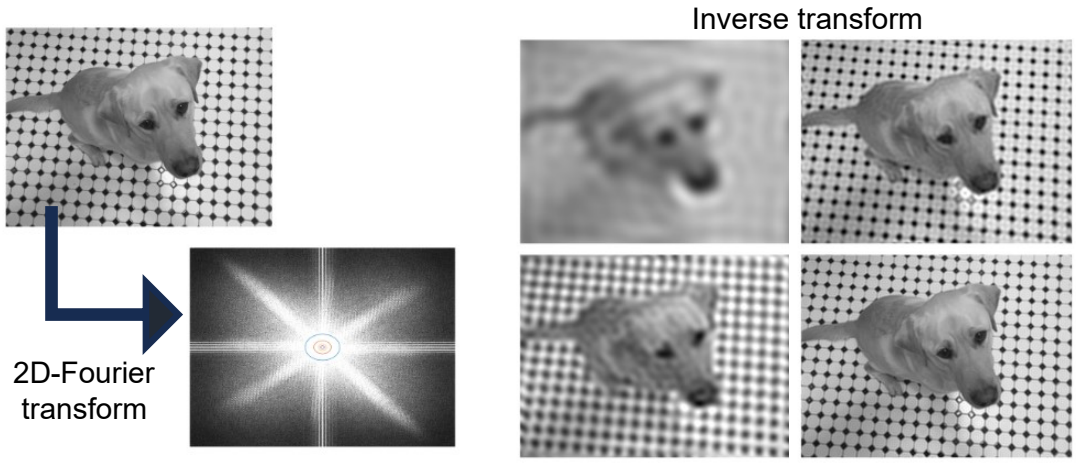


FIGURE 3.4: Reconstruction from Fourier coefficients of kobi.png (copylight 2015, Alex Taylor. All rights reserved).

by CS algorithms, which alleviates a severe ill-condition due to the insufficient amount of measurement. The model-based strategy contributes to the construction of linear models based on DMD, which focuses on the measurements and those temporal developments caused by a physical process. First, from a general perspective, one can assume several characteristics in common with physical phenomena.

It is often the case that one employs a complete system for compressive reconstruction to represent full features appearing in a field of state quantities of interest. In a general basis like Fourier or Wavelet, each element is associated with a certain spatial scale or feature, ranging from the global mean to fine fluctuations of pixel-wise scales. In the general formulation of CS-DMD in Brunton et al., 2015, one begins with setting the two-dimensional region of interest and the internal grids for spatial reconstruction. By setting the region and grids, one automatically defines the number of Fourier basis vectors and the spatial scales corresponding to each vector. If the target image for the reconstruction represents physical processes, almost all the elements of compressible coefficients introduced in Section 3.2.2 are centralized in the wavenumber space representing large-scale structures. Figure 3.4 shows the general example of such a localized distribution of coefficients in the two-dimensional wavenumber space, which can be found as image compression in signal processing texts.

The presence of multiscale physics is widely accepted in the field of fluid dynamics, as found

in the observation of vortex scales and discussion of energy distributions in turbulent flows. It can be drawn as a general remark that one can find a relationship between the scale of spatial structure and the scale of temporal fluctuation. Therefore, as long as one sticks to the low-dimensional representation that can be directed by DMD, it takes less needs that all the basis vectors representing from large- to fine-scale structures to be equally involved in the reconstruction by compressive sensing. Here, the first approach for more reliable and efficient flow models can be mentioned to prioritize the basis vectors more likely to be involved in the reconstruction. By defining the priority denoted by weightings to each basis vector, the recovery algorithm refrains from activating vectors corresponding to fine-scale, which is less seemingly relevant to fluid structures, which easily leads to overfitting. The weights can be arbitrarily taken from $[0, 1]$, while prescribing the weight to each vector is inevitably costly due to the huge number of hyperparameters. It is thus encouraged to consider a simple hard threshold to the vectors according to their corresponding spatial scales, by ignoring finer-scale structures than the threshold. This process of selective use of representation vectors promotes the pruning of the subspace under search in a realistic way, although not spontaneously invoked by the regression. In addition, one can control the number of unknown variables to be determined by compressive sensing techniques, which was originally attained by defining the dimensions of the reconstructed region. The size of the problem is astonishingly reduced as a side effect of reliable estimation by selective use of vectors.

Compressive reconstruction leveraging structured sparsity

In addition to the selective use of the representation basis vectors *(what is used for the inference)*, one can promote specific sparsity in the estimation of representation coefficients *(how we use it)*. The framework presented in this thesis facilitates the idea of structured compressibility treated in R. G. Baraniuk et al., 2010; L. Yu et al., 2012 to promote the low-rank approximation of the physical phenomena. In the studies of the model-based inference in compressive sensing, one can find an idea of "structured sparsity", which is a rather specific concept compared to the general sparsity introduced in Section 3.2.2. Under the property, one can rely on a smaller number of

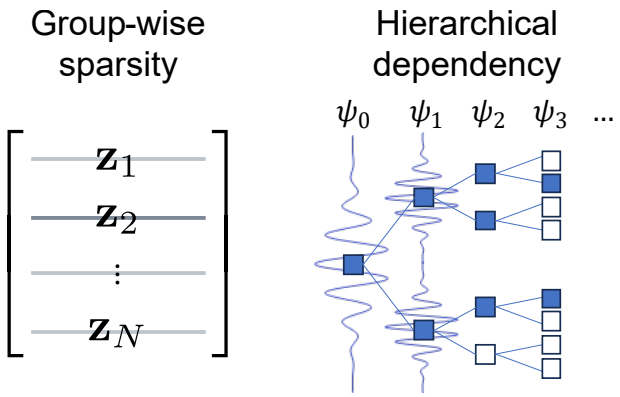


FIGURE 3.5: Examples of structured sparsity

measurements than required by the theory of compressive sensing, and thus, a more reliable reconstruction is realized than that provided by the original formulation of CS-DMD.

As mentioned using Fig. 3.4, large coefficients exist only in the small region of the wavenumber domain. It is pointed out in many compressive sensing literature, moreover, the configuration of the large coefficients in the wavenumber domain has certain kinds of characteristics due to the underlying nature. The characteristics of structured sparsity are shown in Fig. 3.5, for an illustrative example. According to R. G. Baraniuk et al., 2010, the structured sparsity is termed by the union of the subspaces, each of which is a support of the K-sparse subspaces with respect to the given representation basis;

$$\mathcal{M}_K = \bigcup_{m=1}^{M_K} \mathcal{X}_m, \tag{3.27}$$

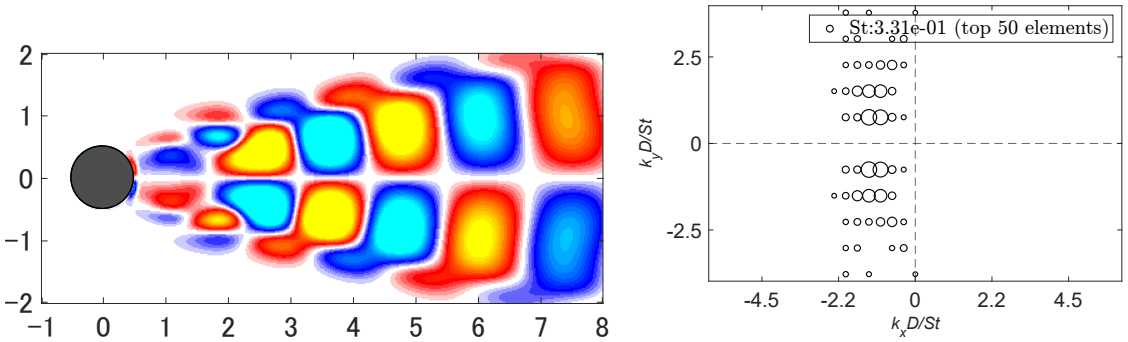
where each $\mathcal{X}_m$ reflects the subspace of the signals that contain coefficients on a K-sparse support. The original K-sparse subspace is composed of a whole $\binom{N}{K}$ ($\gg M_K$) combination of K-sparse subspaces. Therefore, structured sparsity suppresses the degree of freedom for CS compared to the original expression considered in Eq. (3.6). A general notion of RIP is also customized to RIP based on the structured sparse signals, by regulating that the measurement process retains the norm of the sparse signals in Eq. (3.27) in a certain small range. In short, the concept of structured sparsity regulates the combination of K-sparse supports that take large coefficients, assuming the relations between the sparse supports. The selective use of representation vectors introduced earlier is, therefore, a kind of structure that grants only the use of basis vectors within

a certain range of wavenumbers.

The underlying structures are exerted to estimate a sparse vector, as several studies performed compressive recovery of structured compressible/sparse signals. Deterministic methods can be found in R. G. Baraniuk et al., 2010; Cevher et al., 2009. They considered the reconstruction of specific types of signals using a tree-like hierarchical distribution and blockwise sparsity, as shown in Fig. 3.5. More involved structures are obtained by sampling strategy in L. Yu et al., 2012, by manipulating the hyperparameters in a sparse prior to realize the clustered structure in estimated nonzero entries in the sparse parameter vector.

Beneficial features of fluid flows and structured sparsity

One convincing example is that periodic flow structures occasionally occur in regions of a wake flow, which invokes periodic smooth structures in the DMD modes. In the subsequent sections of the results, where the quantities of a flow passing left-to-right are visualized, it can be confirmed that this aspect leads to smooth structures in the modal representation of the vorticity field in Fig. 3.6. It is true that this clarity is enhanced, however, by the absence of measurement noise



(A) Spatial structure of DMD mode of a two dimensional flow

(B) Amplitudes of 2D Fourier coefficient of spatial distribution of DMD mode

FIGURE 3.6: Example: Sparial distribution of DMD mode for the vorticity field of 2D-flow past a cylinder

which would blur the characteristics in the actual acquisition process. Anyway, one powerful structure may be the assumption of this smoothness, namely, the coefficients arranged locally as a cluster structure in the parameter space as shown in Fig. 3.6b.

As the concept of structured sparsity is termed as a union of subspaces, it is worth considering various characteristics of individual phenomena and taking the common parts to form more specific structures in sparsity. The taxonomy of fluid flows is the fundamental approach in many applications, since it can modify the onerous structure of the basic governing equation of flows. By capturing the characteristics of the flow and approximating the equation, one can carefully extract the properties specific to the flow under consideration. In the same manner, by focusing on reproducing the spatial structure of specific flows by CS-DMD as summarized previously, it can further narrow down the sparse subspace in compressive sensing. For example, in a moderately high-speed flow, the advection speed of flow structures and its direction give clear insight into the region where the representation coefficients reside. Furthermore, the interaction between the flow and the surrounding environment sometimes defines the flow geometry. As clearly shown in Fig. 3.6, the flow passing through the circular cylinder constructs a strong (anti-)symmetry with respect to a baseline due to the movement of Kàrmàn vortices. This can be endorsed by incorporating a symmetric distribution of the representation coefficients into the estimation. It is also important, although by no means concrete, to consider the reconstruction of turbulence or shock waves. These interesting aspects still require further investigation.

In Section 3.2.6, a method for compressive estimation is proposed to give a clustered distribution of the coefficients. This is deemed the first step to impose the various characteristics of fluid flow.

3.2.5 Selective weights on representation basis vectors and a priori reduction of library

This modification to CS-DMD is proposed as a regression with a customized penalty for each representation basis vector as a selective weighting matrix $\mathcal{W}$;

$$\mathcal{W} = \begin{bmatrix} w_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & w_N \end{bmatrix}, \tag{3.28}$$

$$\hat{\mathbf{s}}^{(j)} = \arg\min_{\mathbf{s}} \left[\|\mathbf{P}\boldsymbol{\Psi}_{\mathbf{X}}\mathbf{s} - \phi_{\mathbf{Y}}^{(j)}\|_2 + \nu \left(\mathcal{W}^{-1/2}\mathbf{s} \right) \right] \tag{3.29}$$

This can be justified because the sparse regression relies on some prior knowledge as Bayesian inference that promotes a parsimonious solution, and thus, the limitation on the reconstructed wavenumber is additional information.

Specifically, it is assumed in this manuscript that only a few Fourier basis vectors of the lower wavenumber are responsible for the representation of the spatial structure of each DMD mode. This structure is simply introduced by $0-1$ weights w_i that are selectively assigned to each basis vector, depending on whether its wavenumber number exceeds a predefined threshold $k_{\max}$ or not,

$$\boldsymbol{\Psi}_{\mathbf{X}} \in \mathbb{C}^{N\times N} \rightarrow \check{\boldsymbol{\Psi}}_{\mathbf{X}} \in \mathbb{C}^{N\times \check{N}}, \tag{3.30}$$

where the number of column is reduced to $\check{N} \ll N$. The selected vectors are labeled by entries with unity in the weight vector $\mathbf{w}$, of which the number of such entries is denoted by $\check{N}$. Therefore, a vector in the original parameter space is connected by $\mathrm{supp}(\mathbf{s})$. The sparse regression problem used in this study is transformed from Eq. (3.24) by substituting for $\boldsymbol{\Psi}_{\mathbf{X}}$ with the windowed one. This form is denoted as $\mathrm{C}\ell_0\,(\mathrm{w})$ with a notation (w), or (window), to indicate the use of

the windowed basis $\check{\mathbf{\Psi}}_{\mathbf{X}}$;

$$
\begin{aligned}
&[\mathrm{C}\ell_0\,(\mathrm{w})] \\
&\min_{\mathbf{s}\in\mathbb{C}^{\check{N}}} \left[\|\mathbf{P}\check{\mathbf{\Psi}}_{\mathbf{X}}\mathbf{s} - \phi_{\mathbf{Y}}^{(j)}\|_2 + \iota_{\|\mathbf{s}\|_0 \le s_{\max}}(\mathbf{s}) \right],
\end{aligned} \tag{3.31}
$$

where the last term is the hard-threshold in Eq. (3.11).

This threshold is assumed beforehand by committing an appropriate knowledge on the target phenomena. In this example, the scale of the spatial structures can be estimated to scales according to the size of the object in the flow. They are also supposed to be in conjunction with the corresponding eigenvalues. The assumption is a first analogy of the analysis of the boundary layer and instability therein, and the empirical observation of generated vortices behind an object. The latter of the above hypotheses reflects the nature of the modal decomposition by DMD, which is analytically obtained for the flow past a circular cylinder (Bagheri, 2013) by revealing the existence of the harmonics of the DMD modes. Meanwhile, the results in Section 3.3 adopt the appropriate window size from the analysis of the complete data of visualization of the physical quantities of the flow. It is intended for empirical analysis of the dependence of the reconstructed results on the window size. This reduction in the number of available basis vectors immediately lessens the theoretical requirement for the recovery, which is defined by RIP Eq. (3.15).

It must be mentioned for the selective use of representation basis vectors that the reconstruction of DMD modes with no temporal oscillations must rely on interpolation. This is because the selective use of the window Eq. (3.30) cannot be defined due to the zero division. It is immediately concluded by the absence of assumption on the non-oscillating modes that can contain a localized distribution. The flow promotes global correlation by advection toward the downstream direction, whereas they may stagnate in some regions of the flow field.

3.2.6 Compressive estimation as clustered structure

As the Lasso problem Eqs. (3.8) and (3.10) is found in the maximum a posteriori estimation of a mixture of the likelihood and the Laplacian, a sparsified parameter vector can be obtained by

imposing a class of prior distributions that promotes sparsity. Such prior distributions generally used are known as Slab-and-Spike, Laplacian, Horse-shoe, Student-t and other distributions (Carvalho et al., 2010). However, because of the difficulty in analytically solving for the posterior distribution, it is often obtained by sampling methods such as a Gibbs sampler. Furthermore, if one instead assumes the structured sparsity, it is often necessary to consider the correlation between each coefficient (L. Yu et al., 2012). This makes the solution for the inference even harder, despite the solution space reduced to union-of-subspace Eq. (3.27). This paper also assumes the existence of the "cluster" structure in the parameter vector, therefore the solution is not obtained straightforwardly using the previous approaches Eq. (3.8). It should be introduced here that for some cases of amenable structures, the inference for the structured parameter was demonstrated by simple and deterministic algorithms in R. G. Baraniuk et al., 2010; Cevher et al., 2008; C. Chen and Huang, 2012; L. Yuan et al., 2013. The method proposed hereafter is one of those of concise structured inference.

In this paper, a cluster structure in the parameter space is pursued to find a sparse coefficient vector to express the spatial structure of DMD modes. The structure is realized by the evaluation of the norm over the predefined groups, which is composed of a small member of nodes on a Cartesian lattice in wavenumber space. In effect, it is possible to define every cluster shape, including the second-neighbor elements and even more proximate as the same group as introduced in Cevher et al., 2008. However, only the adjacent elements are considered in this study because counting up all combinations for the multidimensional neighbors evidently imposes a problem that is far beyond affordable. Therefore, the adjacent coefficients in the parameter space are sorted into overlapping groups. This type of approximation can be related to the method presented in L. Yu et al., 2012, which introduced a type of mean-field approximation when compressing the Spike-and-Slab prior.

An optimization problem is posed as overlapping group Lasso (M. Yuan and Lin, 2005) to realize the cluster structure. This is denoted as $(\mathrm{C}\ell_1 + \mathrm{G}\ell_1\,(\mathrm{w}))$ where G is taken from the "Group" and the explicit use of the windowed basis $\check{\mathbf{\Psi}}_{\mathbf{X}}$ let the notation be combined with (w).

Its numerical expression is presented as follows;

$$
\begin{aligned}
&[\mathrm{C}\ell_1 + \mathrm{G}\ell_1\,(\mathrm{w})] \\
&\min_{\mathbf{s}\in\mathbb{C}^{\check{N}}} \left[\frac{1}{2}\|\phi_{\mathbf{Y}} - \mathbf{P}\check{\mathbf{\Psi}}_{\mathbf{X}}\mathbf{s}\|_2^2 + \lambda_{\mathrm{C}\ell_1}\|\mathbf{s}\|_1 + \lambda_{\mathrm{G}\ell_2}\sum_{i=1}^{\check{N}} \|\mathbf{s}_{\mathrm{g}(i)}\|_2 \right] && (3.32) \\
&\text{s.t.} \quad \mathbf{s}_{\mathrm{g}(i)} = \left[s_i, s_i^{(1)}, \ldots, s_i^{(2d)} \right] = \mathbf{G}_i\mathbf{s}, && (3.33)
\end{aligned}
$$

where $\mathbf{s}_{g(i)} \in \mathbb{C}^{2d+1}$ denotes a group extracted from a sparse parameter as $\mathbf{G}_i\mathbf{s}$. The i-th group is indicated herein as the adjacent group to the i-th entry of the parameter vector. Each group contains $2d+1$ members to represent $2d$ adjacent nodes that surround a central node except for the boundaries in the assumed d-dimensional wavenumber space. The examples in the study only take the reconstruction of the two-dimensional distribution of DMD modes; therefore, there are at most five nonzero elements in each group. The last term in Eq. (3.32) evaluates ℓ_2-ℓ_1 mixed norm, where one can simply add values of the ℓ_2 norm calculated for each group $g(i)$. This formula is obviously obtained by adding a sparsity-promoting term based on the group norm to the Lasso formula Eqs. (3.8) and (3.10). More general notation for the last term can be considered by other ℓ_p norms (R. G. Baraniuk et al., 2010).

Note that there are overlapping elements in the definition of groups in this study. It induces difficulty to express a proximal operator using the primal variable $\mathbf{s}$ to update the parameter vector. Therefore, this study employs the ADMM (Alternate Direction Method of Multipliers) algorithm, Alg. 6, one of the iterative methods using proximal mapping (Boyd, 2010). To be accommodated in the framework of ADMM, the expression Eq. (3.32) posed above is rewritten as shown below, by introducing auxiliary variables that are constructed by the linear transformation

of the main variable;

$$\underset{\mathbf{s}\in\mathbb{C}^{\check{N}},\,\mathbf{z}\in\mathbb{C}^{2(d+1)\check{N}}}{\arg\min}\ [g(\mathbf{s})+h(\mathbf{z})] \tag{3.34}$$

$$\text{s.t.}\quad \mathbf{z}=\begin{bmatrix}\mathbf{z}_1\\ \\ \mathbf{z}_2\\ \\ \end{bmatrix}=\begin{bmatrix}\mathbf{I}\\ \\ \mathbf{G}\\ \\ \end{bmatrix}\mathbf{s}=\begin{bmatrix}\mathbf{I}\\ \mathbf{G}_1\\ \vdots\\ \mathbf{G}_{\check{N}}\end{bmatrix}\mathbf{s} \tag{3.35}$$

$$g(\mathbf{s})=\frac{1}{2}\|\phi_{\mathbf{Y}}-\mathbf{P}\check{\mathbf{\Psi}}_{\mathbf{X}}\mathbf{s}\|_2^2 \tag{3.36}$$

$$h(\mathbf{z})=\lambda_{\mathrm{C}\ell_1}\|\mathbf{z}_1\|_1+\lambda_{\mathrm{G}\ell_2}\sum_{i=1}^{\check{N}}\|\mathbf{z}_{2,i}\|_2, \tag{3.37}$$

where $\mathbf{z}_1\in\mathbb{C}^N$ are $\mathbf{s}$ itself and groups $\mathbf{z}_{2,i}=\mathbf{s}_{g(i)}$ are aligned as $\mathbf{z}_2\in\mathbb{C}^{(2d+1)N}$. The objective function Eq. (3.34) is composed of the differentiable convex function g and the proximable function h. The latter function of the auxiliary variable enforces the sparsity over the elements of $\mathbf{s}$ and over its adjacent values in a given distributed parameter space.

Algorithm 6 Alternate direction method of multipliers

Input: g [*Eq.* (3.36)], h [*Eq.* (3.37)], $\mathbf{G}\in\mathbb{R}^{(2d+1)\check{N}\times\check{N}}$, $\lambda_{\mathrm{C}\ell_1},\lambda_{\mathrm{G}\ell_2},\gamma\in\mathbb{R}$
Output: Sparse solution vector $\check{\mathbf{s}}$
  Initialize $\check{\mathbf{s}}$, $\mathbf{y}$, $\mathbf{z}$
  while (until halting criterion satisfied) **do**
    $\check{\mathbf{s}}_{\text{new}}=\operatorname{argmin}\left[g(\check{\mathbf{s}})+\frac{1}{2\gamma}\|\mathbf{z}-\mathbf{G}\check{\mathbf{s}}-\mathbf{y}\|_2^2\right]$
    $\mathbf{z}_{\text{new}}=\operatorname{prox}_{\gamma,\,h}\left(\mathbf{G}\check{\mathbf{s}}_{\text{new}}+\mathbf{y}\right)$
    $\mathbf{y}_{\text{new}}=\mathbf{y}+\mathbf{G}\check{\mathbf{s}}_{\text{new}}-\mathbf{z}_{\text{new}}$
    $\check{\mathbf{s}}=\check{\mathbf{s}}_{\text{new}},\quad \mathbf{z}=\mathbf{z}_{\text{new}},\quad \mathbf{y}=\mathbf{y}_{\text{new}}$
  end while

From experience, the entries in the solution vector do not shrink completely to 0 since ADMM promotes sparse solutions through the update of auxiliary variables. The values are in effect shrunk to small numbers, but are yet above the computer precision. Therefore, it is handy to

set the threshold to identify nonzero elements of the obtained solution;

$$\hat{s}_j = \begin{cases} \hat{s}_j & (\hat{s}_j \geq \epsilon \|\hat{\mathbf{s}}\|_\infty) \\ 0 & (\hat{s}_j < \epsilon \|\hat{\mathbf{s}}\|_\infty) \end{cases} \tag{3.38}$$

with ϵ adjusting the threshold, which is usually set to 10^{-5} in this study. The number of the identified nonzero entries is finally set to satisfy the predefined value, if assumed as in Section 3.3, and retain the most energetic $s'_{\max}$ elements according to;

$$s'_{\max} = \min\left(s_{\max}, \|\hat{\mathbf{s}}\|_0\right). \tag{3.39}$$

The final sparsified solution is obtained by polishing procedure in the same way as done by Eq. (3.12) for the retained $s'_{\max}$ entries. As introduced in Section 3.2.2, the polishing step ultimately determines the sparse solution. By a criterion c that infers the nonzero elements using thresholds in Eqs. (3.38) and (3.39), the polished solution is obtained by Eq. (3.12).

3.3 Results using simulated 2D flow past a cylinder

The introduced methods are evaluated in this section by adopting the simple benchmark data, namely, the data of simulated two-dimensional incompressible flow passing a cylinder. Swirling vortices are easily observed and constantly generated in these computed results, as confirmed hereafter. This dataset is due to the computation method of (Taira and Colonius, 2007), and was used for the demonstration of CS-DMD in Kutz et al., 2016b, November, Section 9.4.4. It is distributed through the author website (Kutz et al., 2016a) and is easy to install. In the following sections, the empirical comparisons endorse the proposed methods by highlighting the improvement from the previous methods. The variations found in the reconstruction results are also reported depending on the hyperparameters defined in the reconstruction process.

3.3.1 Data specification

The dataset of the time-series vorticity field of the simulated two-dimensional flow was originally provided in Taira et al., 2017. The description of the data is summarized in table 3.1. The flow shows successive vortex shedding alternately from the upper and lower sides of the cylinder as a stable limit cycle. The simulation was based on the immersed boundary projection method (IBPM) of Taira and Colonius, 2007, while the post-processing scripts of the data and the dataset for the current analysis are provided in Kutz et al., 2016b, November, Chapter 1 and Kutz et al., 2016a, respectively. The Strouhal number St is used to represent the oscillations in the fluid dynamics, which is a dimensionless frequency based on the cylinder diameter D, freestream velocity U, and fluctuation frequency f, respectively.

TABLE 3.1: Simulated 2D flow past a cylinder (Taira and Colonius, 2007) (* represents dimensionless values based on the flow geometry)

Item	Value
Reynolds number	100
Temporal sampling interval	*0.2
Number of snapshots	150
Number of grids for freestream direction	449
Number of grids for vertical direction	199
Vortex shedding period (St number)	*6 (0.16)

A snapshot of 449×199 pixels Fig. 3.7 shows the instantaneous vorticity field recorded in the dataset. The number of the representation basis vectors is therefore $449 \times 199 = 89,351$, and the same number of coefficients are determined in the normal CS-DMD. The successive snapshots between 1–149 and 2–150 are assigned as $\mathbf{X}$ and $\mathbf{X}'$, respectively. The point measurements $\mathbf{Y}$ and $\mathbf{Y}'$ in the CS-DMD are also defined by the measurements of the corresponding snapshots at randomly selected pixels.

3.3.2 DMD modes and representation coefficients obtained from complete data

This section briefly introduces the results of DMD on spatially dense measurement data of fluid flow. In addition to describing what can be found by modal decomposition, it will be explained

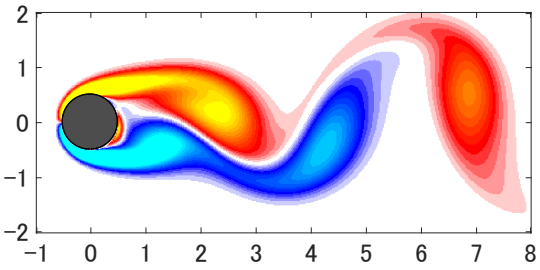
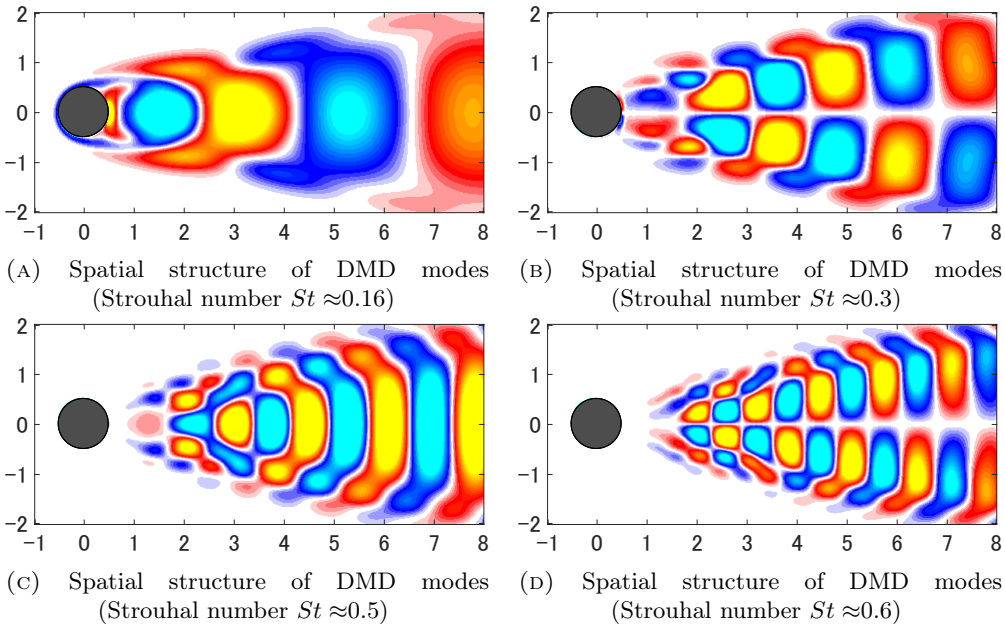


FIGURE 3.7: Instantaneous distribution of simulated vorticity of two-dimensional cylinder wake flow; flow passing left to right

below what the structure used for sparse reconstruction is like.

Structure of each DMD mode



(A) Spatial structure of DMD modes (Strouhal number $St \approx 0.16$)

(B) Spatial structure of DMD modes (Strouhal number $St \approx 0.3$)

(C) Spatial structure of DMD modes (Strouhal number $St \approx 0.5$)

(D) Spatial structure of DMD modes (Strouhal number $St \approx 0.6$)

FIGURE 3.8: Comparison of spatial structures of DMD modes as a reference obtained from spatially dense data (only real parts)

The results of normal DMD on a field measurement $\mathbf{X}/\mathbf{X}'$ are shown in Fig. 3.8, and the corresponding eigenvalues are in Fig. 3.9. Only the first four modes are shown in Fig. 3.8 labeled by the corresponding Strouhal number, of which the corresponding eigenvalues are indicated in Fig. 3.9. Since the obtained eigenvalues lie on the unit circle drawn by the dashed line, such a mode produces a periodic motion in the amplitude without amplification or dumping. It is straightforwardly confirmed from this distribution of eigenvalues, as previously discussed

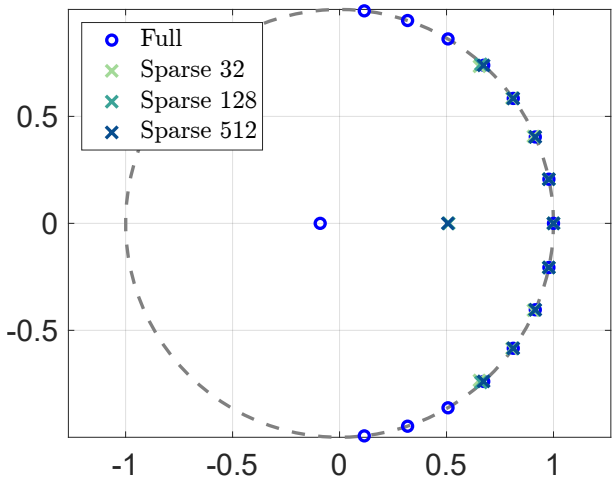
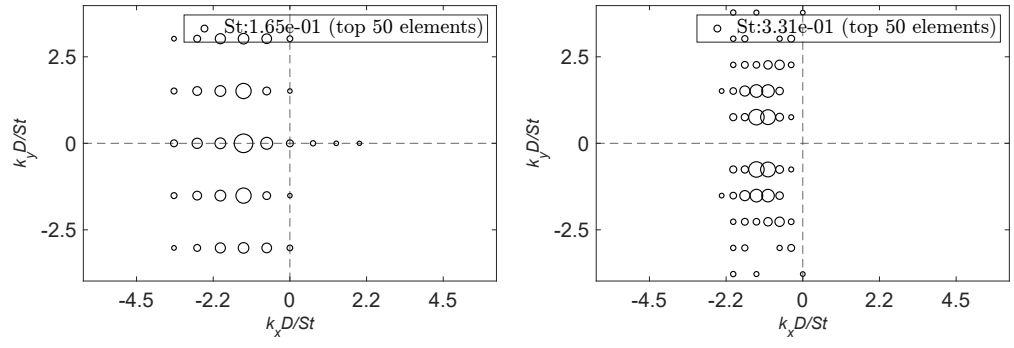


FIGURE 3.9: Obtained eigenvalues in the procedure of DMD; the circles are due to the spatially dense measurements, while the cross marks are obtained from sparse measurements at corresponding numbers .

in Bagheri, 2013; Kutz et al., 2016b, November, that the wake of a circular cylinder produces characteristic modes and their harmonics. The structures of Figures 3.8b and 3.8d are antisymmetric with respect to the horizontal axis parallel to the freestream, while those of Figures 3.8a and 3.8c show symmetric distributions. The former modes seem to represent the alternately shedding vortices, and the latter modes should be responsible for the advection of the flow structures.

Scatter plots in Fig. 3.10 depict the spatial structure of coherence found in Fig. 3.8. These spatial structures are converted into coefficients in the wavenumber space by the two-dimensional Fourier transformation, and the absolute values of each coefficient are shown therein. Both axes are normalized by the Strouhal number $St = fD/U$ to see the common tendency between these modes. The absence of circles on the $k_y = 0$ line at Figures 3.10b and 3.10d endorses the antisymmetric distribution of DMD modes of $St \approx 0.3$ and 0.6, respectively.

The distribution of power spectrum density, PSD, is illustrated for each mode by Fig. 3.8 to purify the spatial characteristics of the DMD modes obtained. The abscissa is a dimensionless wavenumber that is determined by kD/St for each mode and aspect ratio of the acquired image. The peaks are observed around $kD/St \approx 1.5$ where the maximum values of each power spectrum density are normalized for ease of display. This coincidence of the peaks is fairly reasonable



(A) Spatial structure of DMD modes (Strouhal number $St \approx 0.16$)

(B) Spatial structure of DMD modes (Strouhal number $St \approx 0.3$)

(C) Spatial structure of DMD modes (Strouhal number $St \approx 0.5$)

(D) Spatial structure of DMD modes (Strouhal number $St \approx 0.6$)

FIGURE 3.10: Comparison of distributions of Fourier coefficients for each DMD mode as a reference obtained from spatially dense data (both axes are normalized by St for each mode)

because the phenomenon behind the data is the Kàrmàn vortex shedding, and each DMD mode is responsible for a similar fluctuation of a quite close advection velocity. The threshold for windowing truncation $k_{\max}$ in Section 3.2.5 is determined depending on St assigned to each DMD mode;

$$k_{\max}(St) = k'_{\max} St/D, \tag{3.40}$$

where $k'_{\max} = 5$ is mainly used based on the discussion above, and the low-wavenumber region is considered for the following CS-DMD application. Specifically, in the reconstruction of DMD mode $St \approx 0.33$, the above criterion results in the reduction of degrees of freedom to only 309, of which the proportion is $309/449/199 \approx 3.5 \times 10^{-3}$. The number of measurements demanded by RIP is thus reduced to $(\log(309) - \log(K))/(\log(449) + \log(199) - \log(K)) < 0.5$, which means

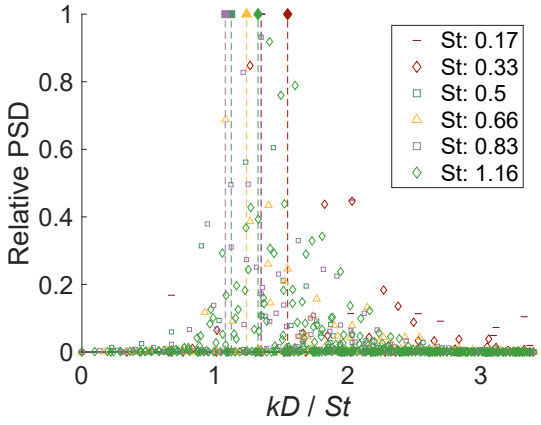


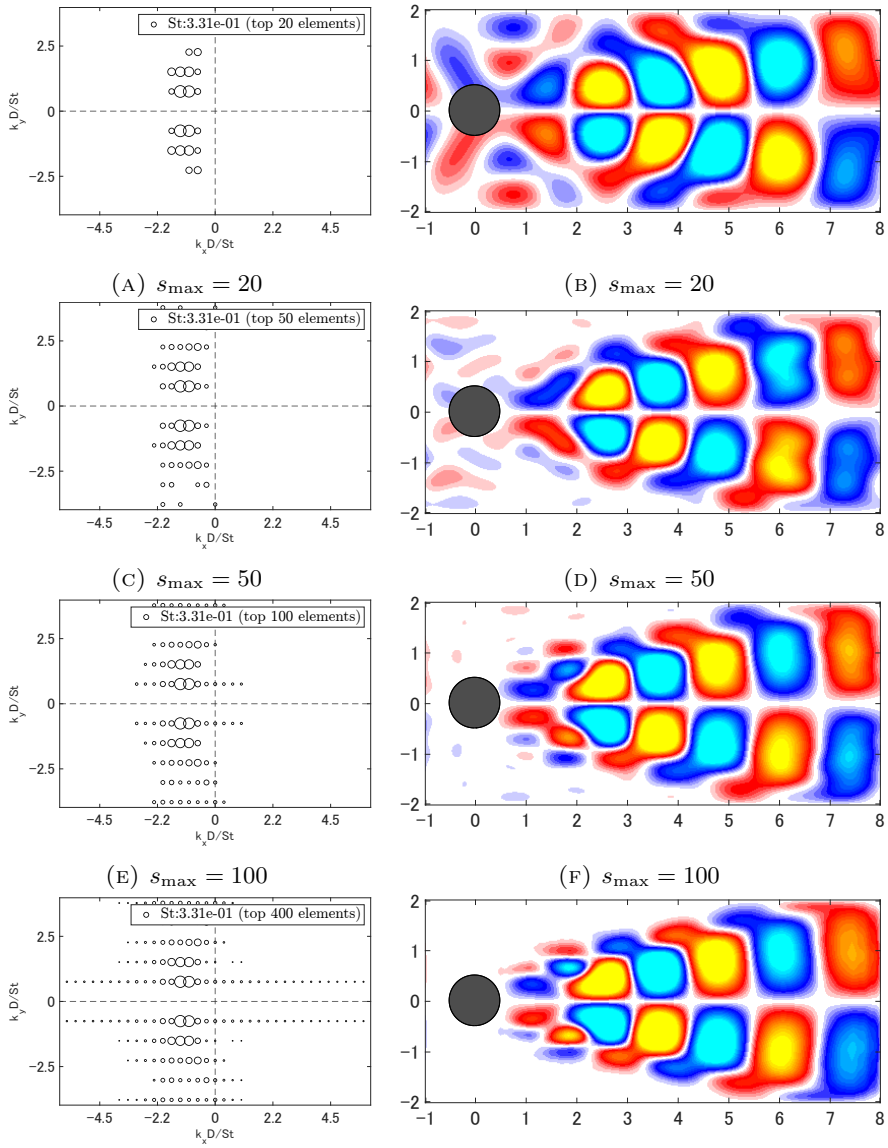
FIGURE 3.11: Structure of DMD modes in Fourier space; all of which has peak around 1.5 of dimensionless wavenumber $k' = kD/St$

that it becomes less than half of the original amount. This result holds without the assumption of structured sparsity; thus, it is supposed to be reduced by an order of magnitude in effect when applied to the current example of flow reconstruction.

Sparse approximation of DMD mode $St \approx$0.33

The compressed representations of the spatial distribution of the DMD mode are shown in **??** to give a reference of compressive solution of CS-DMD in the subsequent result sections. These are obtained from the original DMD mode corresponding to the temporal fluctuation $St \approx$0.33 of Fig. 3.8b. Each figure in the left column of Fig. 3.12 shows the truncated coefficients of the Fourier transformation of DMD mode at various numbers of thresholds for larger components, 20, 50, 100, and 400. Correspondingly, the right column in Fig. 3.12 shows the inverse transformation of the truncated coefficients.

Figures indicate nonzero coeffcients in the left column in Fig. 3.12. Nonzero values denoted by circles collocate in the left-handed plane as found in Fig. 3.10b, where larger values are in the central part with smaller values aligned on the periphery. Moreover, as pointed out previously, two separated clusters are observed to adhere to the antisymmetric structure. The basis vectors with the smaller wavenumbers in the y-direction are used for the representation if there are more nonzero components involved, as Figures 3.12d and 3.12f show that components are allocated along the abscissa. The selective use of basis vectors by window in Section 3.2.5 may only work



(A) $s_{\max} = 20$ (B) $s_{\max} = 20$

(C) $s_{\max} = 50$ (D) $s_{\max} = 50$

(E) $s_{\max} = 100$ (F) $s_{\max} = 100$

(G) $s_{\max} = 400$ (H) $s_{\max} = 400$

FIGURE 3.12: Comparison of sparse approximation of Fourier coefficients of DMD mode $St \approx 0.33$, which was previously shown in Figures 3.8b and 3.10b

for a smaller number of nonzero components due to this distribution.

The spatial distributions of DMD mode reconstructed from the sparsified coefficients are shown in the right column. The vortices shed from the cylinder are clear as in Fig. 3.12b, yet the absence of vorticity in the freestream region can not be retrieved with only 20 components. Allowing more components to be involved in the representation leads to a reduction in the artificial distribution in the freestream region in addition to strengthening the structures of the finer scale. The original structure of DMD mode in Fig. 3.8b are almost recovered with 400 components as in Fig. 3.12h.

3.3.3 Model-based estimation of DMD modes

This section provides the reconstruction results of DMD modes by compressive sensing approach of the CS-DMD. It was confirmed in (Kutz et al., 2016b, November) that the CS-DMD performed well for this example with $p = \mathcal{O}(10^3)$ measurements, and thus, compression rate was up to $\approx 1\%$ to the original data of the $\mathcal{O}(10^5)$ pixels. This number of measurements gives an impression of somewhat too many measurements to apply the CS-DMD to modeling of fluid flows. The lack of information from sparse measurements seemed to be the main reason for this, which ultimately caused the reconstruction to suffer from overfitting or aliasing for a smaller number of measurements.

Reconstructed distribution of DMD modes

One may start by confirming the assumption of the correspondence of the eigenvalues obtained from point measurements. As given in Fig. 3.9, the eigenvalues on the unit circle show significantly good accordance with those obtained by the field measurements and those based on a small number of point measurements. The data can be described as an ideal measurement of a characteristic phenomenon without noise contamination. (Especially, the main idea for the comparison is the reduction of the number of point measurements for the same level of reconstruction quality to highlight the proposed method.) This can be concluded that the data are well conditioned so that one can extract ubiquitous information of flows from any positions of

sensing, so the only matter is the lack of spatial information obtained by sparse measurements of sensors.

The comparison investigated the reconstruction of CS-DMD with respect to the parameters in table 3.2; namely, the number of point measurements p, the target Strouhal number St of DMD modes for reconstruction and the threshold value $k'_{\max}$ for the selection window of basis vectors in Eq. (3.40);

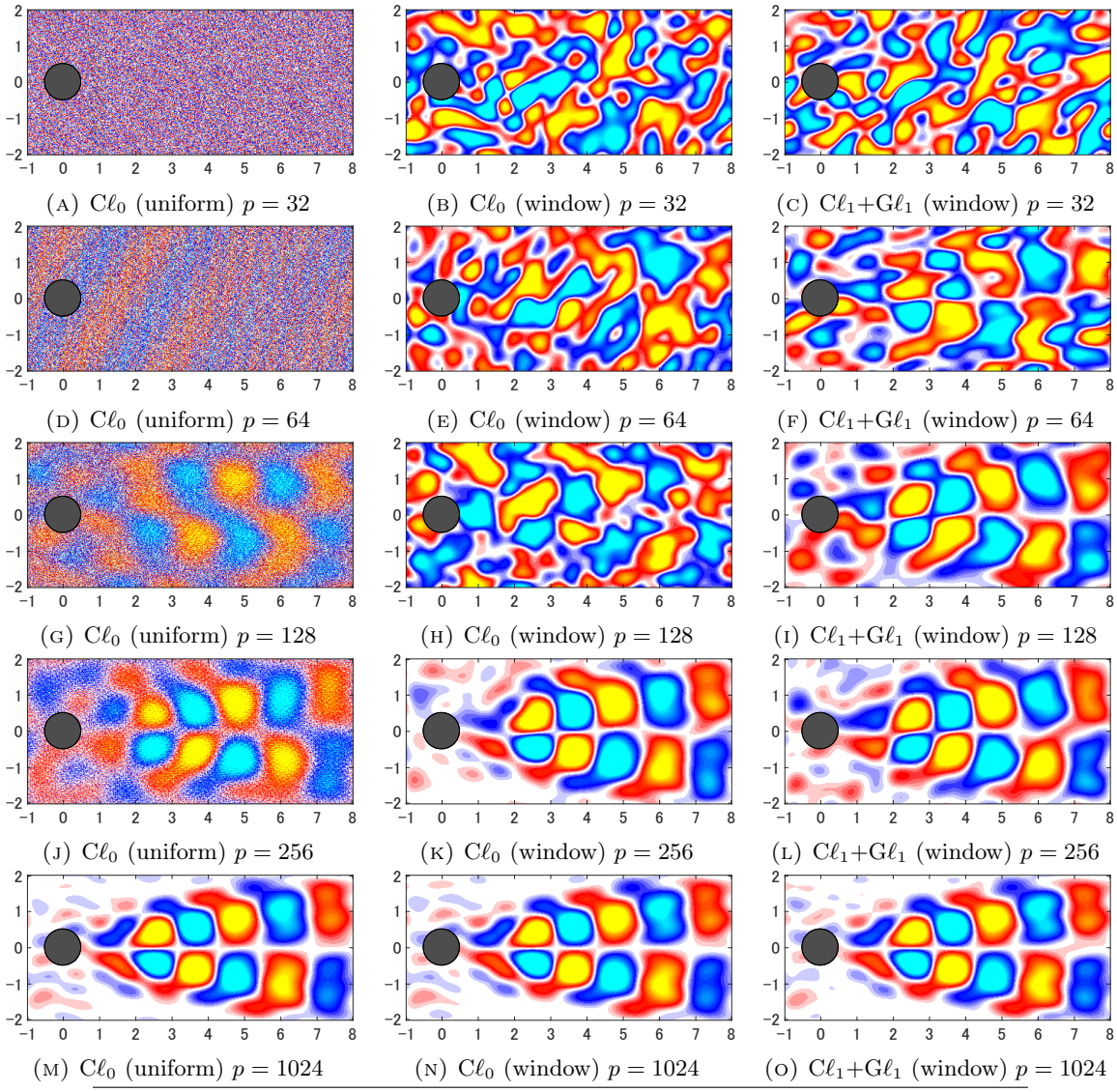
Model parameter	Value
p	2^4 to 2^{10} with 25 increments
St	0.16, 0.33, 0.50, 0.66
$k'_{\max}$	3,5,7
$s_{\max}$	30,50,100

TABLE 3.2: Parameters for comparative study of model-based reconstruction targeting the flow past a cylinder

First, the reconstructed distributions of the DMD mode are compared in Fig. 3.13, in which the DMD mode of $St \approx 0.33$ is treated. Each column is classified by the estimation methods conducted; from left to right, the original implementation Cℓ_0 (u) as a reference, Cℓ_0 (w) with a reduced number of basis vectors as introduced in Section 3.2.5, and finally, the results of Cℓ_1+Gℓ_1 (u) by ADMM encouraging the cluster structure. These figures visualize the real parts in the reconstructed DMD modes, where only one trial composed these figures with a random configuration of sensing points in common. Therefore, the results can be deviated for a different set of sensing positions, although they are expected to be the same to some extent. It should be noted that the number of Fourier modes activated is up to 50, which is according to the criterion Eq. (3.38).

As reported in Kutz et al., 2016b, November, effective reconstruction with normal implementation of the CS-DMD required almost 1,000 sparse measurements, which is consistent with current results. Three chief mentions can be drawn from Fig. 3.13 as:

1. By comparing the leftmost column with others, the proposed selection of representation basis vectors eliminates the fine-scaled, but noisy structures to be involved.



(A) Cℓ_0 (uniform) $p = 32$ (B) Cℓ_0 (window) $p = 32$ (C) Cℓ_1+Gℓ_1 (window) $p = 32$

(D) Cℓ_0 (uniform) $p = 64$ (E) Cℓ_0 (window) $p = 64$ (F) Cℓ_1+Gℓ_1 (window) $p = 64$

(G) Cℓ_0 (uniform) $p = 128$ (H) Cℓ_0 (window) $p = 128$ (I) Cℓ_1+Gℓ_1 (window) $p = 128$

(J) Cℓ_0 (uniform) $p = 256$ (K) Cℓ_0 (window) $p = 256$ (L) Cℓ_1+Gℓ_1 (window) $p = 256$

(M) Cℓ_0 (uniform) $p = 1024$ (N) Cℓ_0 (window) $p = 1024$ (O) Cℓ_1+Gℓ_1 (window) $p = 1024$

FIGURE 3.13: Spatial distribution of reconstructed DMD modes of $St \approx 0.33$ for flow past cylinder

2. The targeted structure of the DMD mode is partially reconstructed by Cℓ_1+Gℓ_1 (window) only with $p = 64$, of which the compression rate against the original spatial resolution is $\approx 0.073\%$.

3. Structures are occasionally observed in the freestream region where the vorticity does not exist.

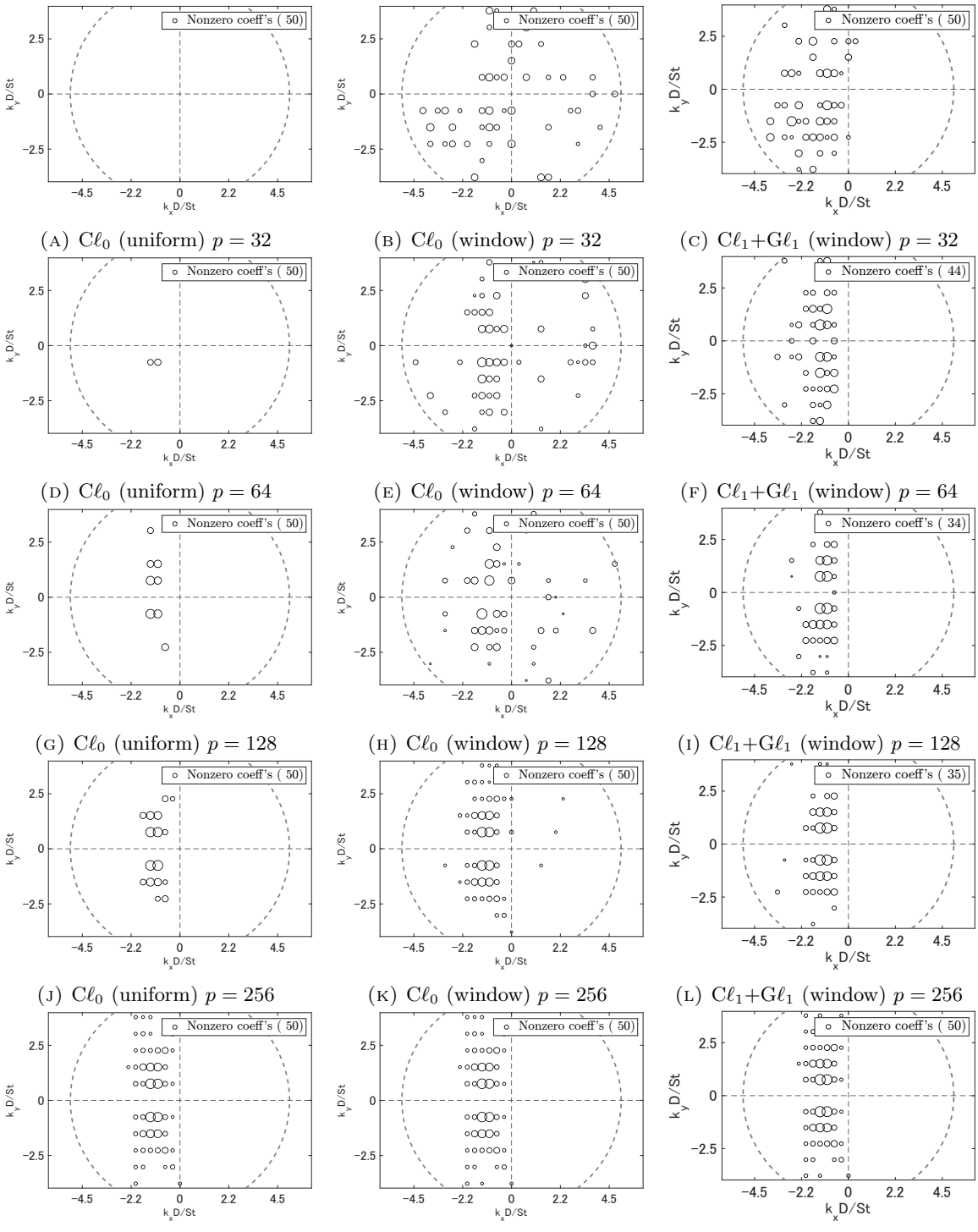
From the first perspective, the general formulation of CS-DMD is supposed to cause overfitting due to a subspace that is too redundant for the representation, as indicated in Section 3.2.4.

Moreover, the ability of faithful reconstruction is greatly improved by the modified objective function, which incorporates the structure of expected signals into the inverse inference. The vortex-like structure is, in fact, partially observed in Fig. 3.14g which alternately produces circles with opposite signs in the wake region, but is completely buried in the noisy components. The final remark can be observed in the result of Cℓ_1+Gℓ_1 (window) with the case of $p = 128$ in Fig. 3.13i. Although the obtained modal structures are qualitatively similar to original modal structure in Fig. 3.8b, the spurious structures are found in the freestream regions. This can be related to the ability to adequately sparsify the results. This is due to the too strict sparsity imposed in the reconstruction procedure, since it requires more of the basis vectors to mend these artifacts. Therefore, it is still an open avenue to automatically decide on the appropriate threshold to preserve the obtained sparse vector by considering the number of measurements and the uncertainty in the obtained model. The discussion on the sparsity imposed on the estimation is also placed in the following section of Section 3.3.4.

Inferred Fourier coefficients

It can be intriguing to see how the structural demand is satisfied by different types of reconstruction methods. The obtained Fourier coefficients are projected into the two-dimensional parameter space by circles denoting the intensity of each component in the inferred sparse vector. The arrangement of the figures completely corresponds to Fig. 3.13, with each row for the number of measurements and each column for the estimation configurations. The outer circle of the dashed line indicates the boundary of the window imposed on the estimation according to Eq. (3.40).

The numbers in each legend in the figures denote the number of nonzero elements determined by the procedure Eq. (3.38), which depend on the returned solution vectors of each estimation algorithm and configuration. Because the presented figures focus on the tiny region of interest defined by Eq. (3.40) with $k' = 5$, there will be nothing to show if the inferred entries reside outside of the region. The possible understandings of Fig. 3.14 are listed threefold:



(A) Cℓ_0 (uniform) $p = 32$ (B) Cℓ_0 (window) $p = 32$ (C) Cℓ_1+Gℓ_1 (window) $p = 32$

(D) Cℓ_0 (uniform) $p = 64$ (E) Cℓ_0 (window) $p = 64$ (F) Cℓ_1+Gℓ_1 (window) $p = 64$

(G) Cℓ_0 (uniform) $p = 128$ (H) Cℓ_0 (window) $p = 128$ (I) Cℓ_1+Gℓ_1 (window) $p = 128$

(J) Cℓ_0 (uniform) $p = 256$ (K) Cℓ_0 (window) $p = 256$ (L) Cℓ_1+Gℓ_1 (window) $p = 256$

(M) Cℓ_0 (uniform) $p = 1024$ (N) Cℓ_0 (window) $p = 1024$ (O) Cℓ_1+Gℓ_1 (window) $p = 1024$

FIGURE 3.14: Reconstructed DMD modes of $St \approx 0.33$ for flow past cylinder

1. Without the limitation of the window, the fitting of CS involves the components of much higher wavenumbers.

2. Non-trivial entries are first retrieved as the number of sensors increases, as observed for all reconstruction methods.

3. The estimated distributions by ADMM with a group sparsity term spontaneously show a clear bias to the left-hand side plane, which is desired from the original distribution.

As externally imposed by the selective use of basis vectors, reconstruction methods seek the most adequate components within the window, as shown in the middle and right columns in Fig. 3.14. The number of measurements in which the nonzero coefficients start to collocate on the left-hand side around $k_x \approx -2.2$ decreases by introducing the window threshold and is further decreased by the structural reconstruction. The second observation is, somewhat trivial, more applicable for Cℓ_0 (u), say Fig. 3.14d. Most energetic components are first identified, although other entries are placed far beyond the adequate region. This hints at the reason why the assumption of a clustered structure works effectively in this case. Therefore, the use of Eq. (3.32) without the selective use of basis vectors may be a possible choice, as the determination of the shape of the window should be more involved for the actual applications. However, this is hardly achievable from another aspect, because the computational burden of ADMM increases significantly as the size of the sparse vector increases. The use of an efficient algorithm of L. Yuan et al., 2013 is clearly a remedy for this, but is a future subject. The last remark is the most impressive for this comparison, for the only induced structure on the structured sparsity is to form the cluster distribution in the parameter space. As hierarchically listed in Section 3.2.4, this bias in the distribution should be respected when the flow of interest has a clear advection in one direction. From the observation on the group sparsity formulation, one may be capable of discovering the nature of physics underlying the phenomenon in a relatively explainable way.

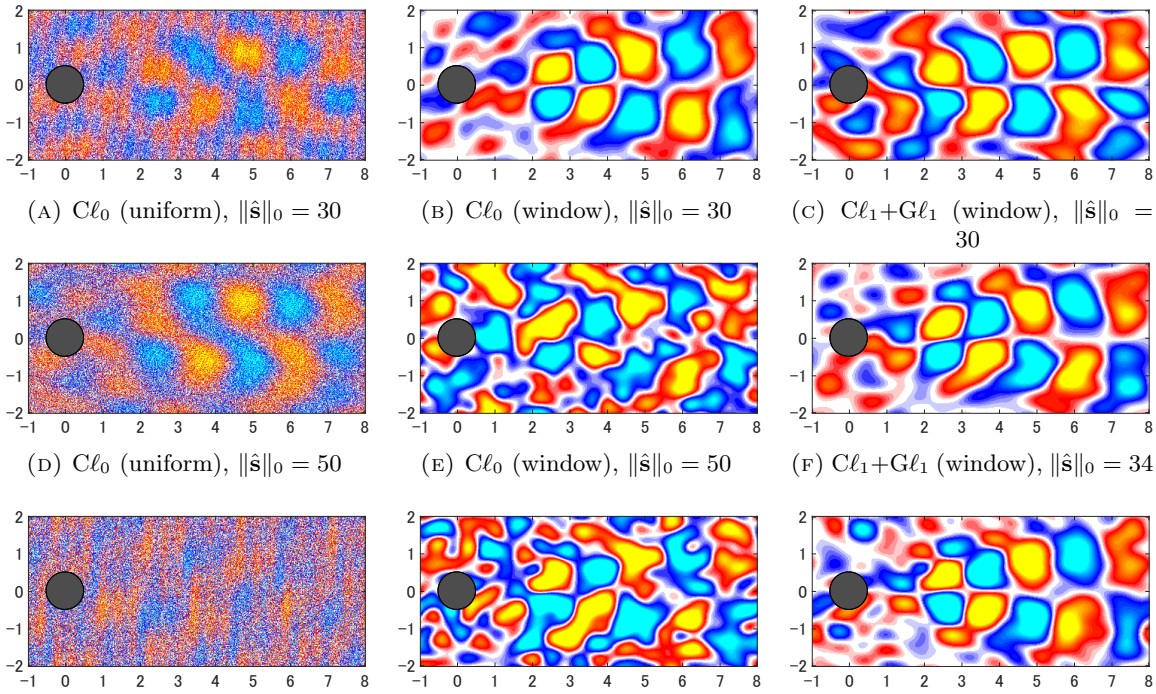
3.3.4 Variation of compressive reconstruction against model parameters

Two model parameters are investigated to show the certainty of reconstruction by CS-DMD. These parameters, that is, the sparsity of the solution and the window size used for the selection of basis vectors, are defined in advance of the compressive estimation. Therefore, these values must be set adequately to properly restrict the ability of CS.

Sparsity imposed on nonzero components in solution

This value determines the number of nonzero entries in the solution obtained. It is applied directly to solutions based on the ℓ_0 norm in Eq. (3.11), while this is accomplished by managing the regularization parameters in Eqs. (3.10) and (3.34). As described in relation to CS theories in Section 3.2.2, the number is given due to the nature of the target signal for the case of the sparse signal. The required number of measurements is composed according to the sufficient condition for complete recovery such as the RIP (Eq. (3.15)). On the other hand, one has to determine this value in advance, especially when dealing with the compressible signal. The results in the previous section Section 3.3.3 show that a suitable setting is necessary to realize a good sparse solution. It is expected that too large a value will cause overfitting due to the ill-condition of the estimation, while too small a value will degrade the representation ability.

The results shown in Figures 3.15 and 3.16 show the reconstructed distribution of the DMD modes $St \approx 0.33$ and Fourier coefficients thereof. These results are according to three different sparsity assumptions, $s_{\max} \leq 30, 50, 100$, imposed on three estimation methods with measurements of $p = 128$. It can be easily observed that, in each column of Fig. 3.15, the values differ when the estimates most closely approach the original distribution of Fig. 3.8b. The figures in the middle column, the results of $\mathrm{C}\ell_0(w)$, showed a fair resemblance to the original state when the number of nonzero entries is restricted to $s_{\max} = 30$, but gradually collapsed as $s_{\max}$ increases. In contrast, the result of the group sparse regression in the right column showed the improvement in the expression of boundaries as $s_{\max}$ increases. This is also observed from the obtained distribution of Fourier coefficients in Fig. 3.16. In the figures of the middle column, those of $\mathrm{C}\ell_0(w)$ tend to be distributed within the dashed circle, indicating the threshold to the



(A) Cℓ_0 (uniform), $\|\hat{\mathbf{s}}\|_0 = 30$ (B) Cℓ_0 (window), $\|\hat{\mathbf{s}}\|_0 = 30$ (C) Cℓ_1+Gℓ_1 (window), $\|\hat{\mathbf{s}}\|_0 = 30$

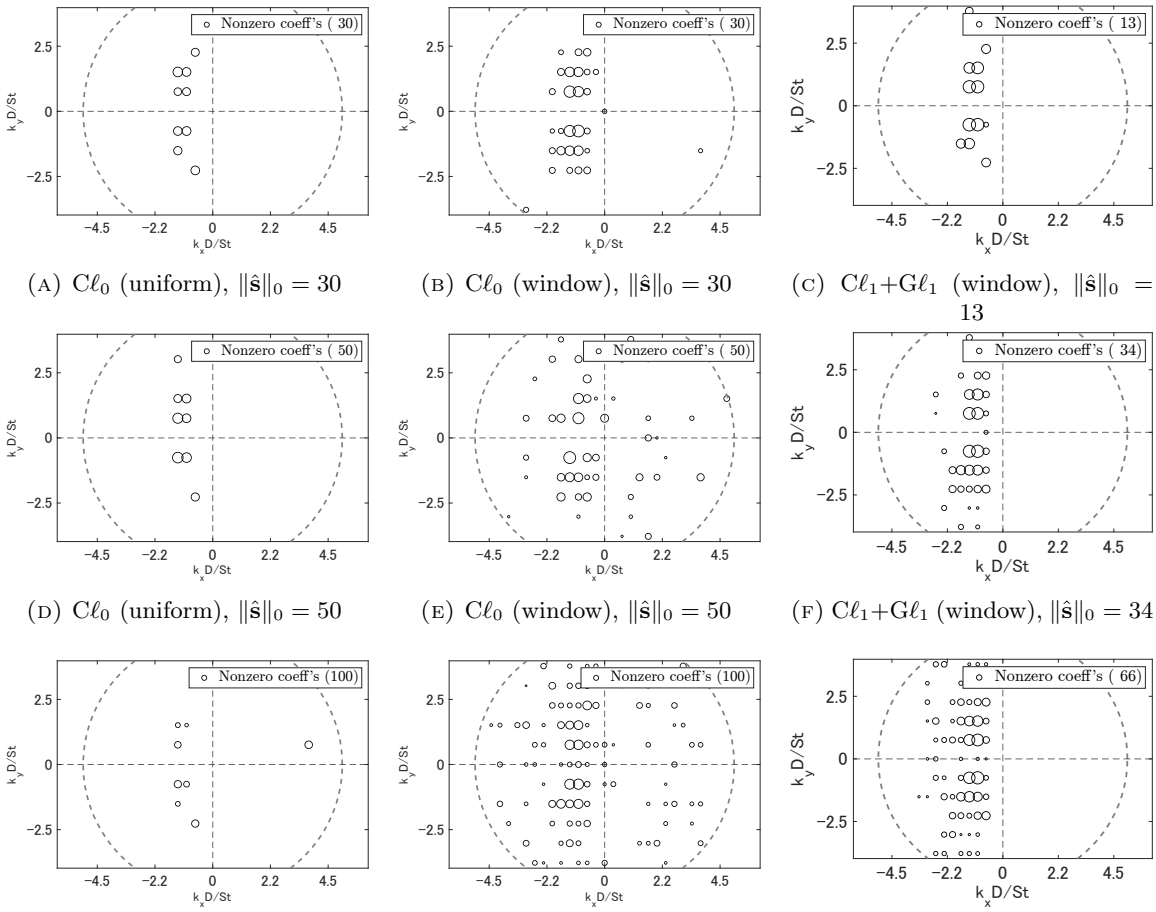
(D) Cℓ_0 (uniform), $\|\hat{\mathbf{s}}\|_0 = 50$ (E) Cℓ_0 (window), $\|\hat{\mathbf{s}}\|_0 = 50$ (F) Cℓ_1+Gℓ_1 (window), $\|\hat{\mathbf{s}}\|_0 = 34$

(G) Cℓ_0 (uniform), $\|\hat{\mathbf{s}}\|_0 = 100$ (H) Cℓ_0 (window), $\|\hat{\mathbf{s}}\|_0 = 100$ (I) Cℓ_1+Gℓ_1 (window), $\|\hat{\mathbf{s}}\|_0 = 66$

FIGURE 3.15: Spatial distribution of reconstructed DMD modes of $St \approx 0.33$ for flow past cylinder for $p = 128$; variation against $s_{\max}$

wavenumber of the basis vector. It is observed, in contrast, that those nonzero values collocate near each other as the sparsity gets relaxed, indicating the reinforcement of a cluster distribution.

From these observations, the determination of the adequate value for the sparsity is in no way straightforward. This value should surely embody the low-dimensional nature of the phenomena of interest as a sparse vector. Nevertheless, this value should reflect the conditions of measurements, such as the signal-to-noise ratio, and the performance of the inference algorithm. Therefore, as mentioned in Fig. 3.13, the automatic way to determine this value will be a precise help for the modeling by the CS-DMD. It has been explained that the sparseness of the reconstruction reflects the belief in the certainty of the existence of a phenomenon. This parameter can be mapped one by one to the parameter indicating the magnitude of the measurement noise, which influences the fidelity of the regression. The only affordable way to obtain the reliable result thus far is to employ multiple values of sparsity, which is leveraged in Chapter 4 to adaptively obtain informative measurement positions.



(A) Cℓ_0 (uniform), $\|\hat{\mathbf{s}}\|_0 = 30$ (B) Cℓ_0 (window), $\|\hat{\mathbf{s}}\|_0 = 30$ (C) Cℓ_1+Gℓ_1 (window), $\|\hat{\mathbf{s}}\|_0 = 13$

(D) Cℓ_0 (uniform), $\|\hat{\mathbf{s}}\|_0 = 50$ (E) Cℓ_0 (window), $\|\hat{\mathbf{s}}\|_0 = 50$ (F) Cℓ_1+Gℓ_1 (window), $\|\hat{\mathbf{s}}\|_0 = 34$

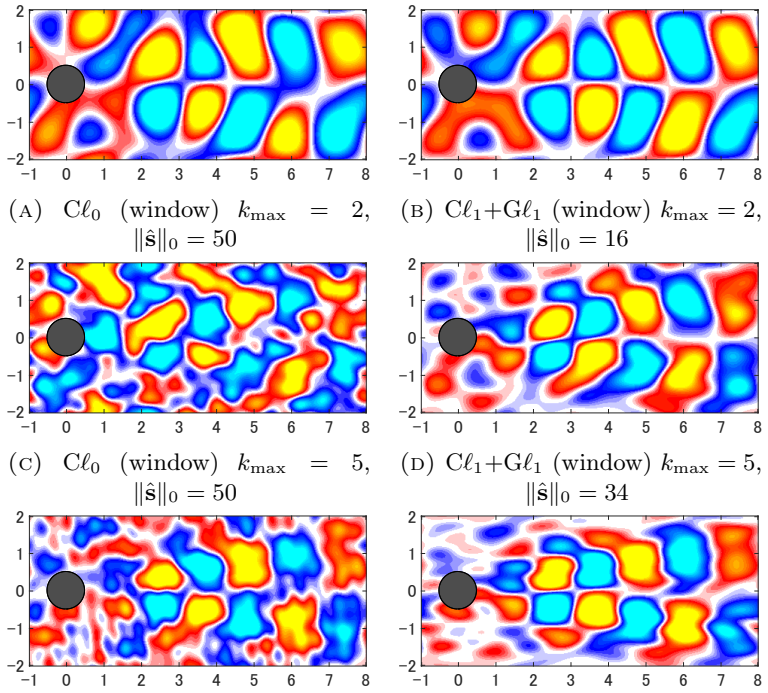
(G) Cℓ_0 (uniform), $\|\hat{\mathbf{s}}\|_0 = 100$ (H) Cℓ_0 (window), $\|\hat{\mathbf{s}}\|_0 = 100$ (I) Cℓ_1+Gℓ_1 (window), $\|\hat{\mathbf{s}}\|_0 = 66$

FIGURE 3.16: Reconstructed DMD modes of $St \approx 0.33$ for flow past cylinder for $p = 128$; variation against $s_{\max}$

Window size

This parameter is interpreted as an expectation of the spatial scale of the phenomenon. Once the window imposes a much smaller scale than the proper one, the misunderstanding on the target phenomenon will ruin the estimation by squeezing it into the subspace that never fits. It causes overfitting, in contrast, by applying too large a window size as addressed in Section 3.2.5. Moreover, the computational time will surely increase as the size of the problem increases, depending on the number of basis vectors considered. The degree of prominence of each basis vector is essentially distributed with respect to the wavenumber as shown in Fig. 3.11. An adequate value for the threshold in Eq. (3.40) must be assumed without analyzing the entire flow field. Reasonable determination entirely relies on the experience and prior knowledge of the user. It

is described in this section how the obtained results change, or deteriorate, without a proper selection of basis vector.



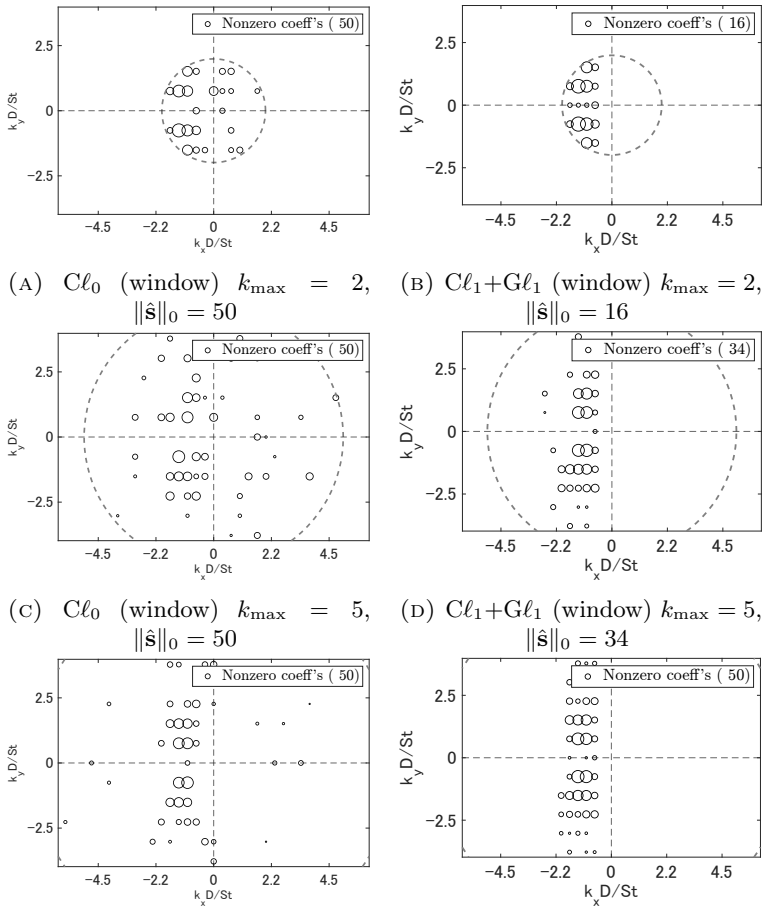
(A) Cℓ_0 (window) $k_{\max} = 2$, $\|\hat{\mathbf{s}}\|_0 = 50$ (B) Cℓ_1+Gℓ_1 (window) $k_{\max} = 2$, $\|\hat{\mathbf{s}}\|_0 = 16$

(C) Cℓ_0 (window) $k_{\max} = 5$, $\|\hat{\mathbf{s}}\|_0 = 50$ (D) Cℓ_1+Gℓ_1 (window) $k_{\max} = 5$, $\|\hat{\mathbf{s}}\|_0 = 34$

(E) Cℓ_0 (window) $k_{\max} = 7$, $\|\hat{\mathbf{s}}\|_0 = 50$ (F) Cℓ_1+Gℓ_1 (window) $k_{\max} = 7$, $\|\hat{\mathbf{s}}\|_0 = 50$

FIGURE 3.17: Spatial distribution of reconstructed DMD modes of $St \approx 0.33$ for flow past a cylinder for $p = 128$; variation against $k_{\max}$

The results shown in Figures 3.17 and 3.18 exhibit the reconstructed distribution of the DMD modes $St \approx 0.33$ and the Fourier coefficients thereof according to three different thresholds from the available basis vector, $k'_{\max} = 2, 5, 7$. The figures shown are those of the two proposed methods that incorporate the selective use of basis vectors in Section 3.2.5. It should be mentioned here that the results of C$\ell_0(w)$ tend to those with a whole library when $k'_{\max} \to \infty$. The outer circle with a dashed line indicates the boundary of the window imposed on the estimation according to Eq. (3.40) and each value of $k'_{\max}$. The first rows in Fig. 3.17 show the distribution obtained from $k'_{\max} = 2$, which corresponds to the case where the restriction on the basis vectors is excessive. This threshold abandons the vectors of wavenumber at dropping shoulder in the higher wavenumber domain, yet contains the most energetic parts. The results of both methods are very similar to the original distribution Fig. 3.8b. Meanwhile, the artifacts outside the proper



(A) Cℓ_0 (window) $k_{\max} = 2$, $\|\hat{\mathbf{s}}\|_0 = 50$

(B) Cℓ_1+Gℓ_1 (window) $k_{\max} = 2$, $\|\hat{\mathbf{s}}\|_0 = 16$

(C) Cℓ_0 (window) $k_{\max} = 5$, $\|\hat{\mathbf{s}}\|_0 = 50$

(D) Cℓ_1+Gℓ_1 (window) $k_{\max} = 5$, $\|\hat{\mathbf{s}}\|_0 = 34$

(E) Cℓ_0 (window) $k_{\max} = 7$, $\|\hat{\mathbf{s}}\|_0 = 50$

(F) Cℓ_1+Gℓ_1 (window) $k_{\max} = 7$, $\|\hat{\mathbf{s}}\|_0 = 50$

FIGURE 3.18: Reconstructed DMD modes of $St \approx 0.33$ for flow past a cylinder for $p = 128$; variation against $k_{\max}$

region are as energetic as the inferred structures. This is entirely due to the mismatch between the presumed threshold and the actual structures. On the other hand, the final row of $k'_{\max} = 7$, shows the case that the estimation is driven by a somewhat relaxed assumption. The obtained results are akin to those in the second row with a moderate threshold value of $k'_{\max} = 5$, of which the comparison was previously conducted in Section 3.3.3. Meanwhile, the structures obtained show a finer spatial deviation in Figures 3.18e and 3.18f compared to Figures 3.18c and 3.18d. This can be blamed on overfitting, but partially.

The determination of an adequate value is a remaining task, as is the sparsity value mentioned above. However, it is more convincing to presume this value than that of the sparsity, because one can exploit the knowledge of the scaling law ubiquitous in fluid mechanics. There are several

studies investigating scales of fluctuation in wake flow for various flow configurations, such as the convection velocity of the vortices and the size of interfering object inside the flow. The scaling law can be leveraged here to estimate the expected fluctuation scales, both in time and space.

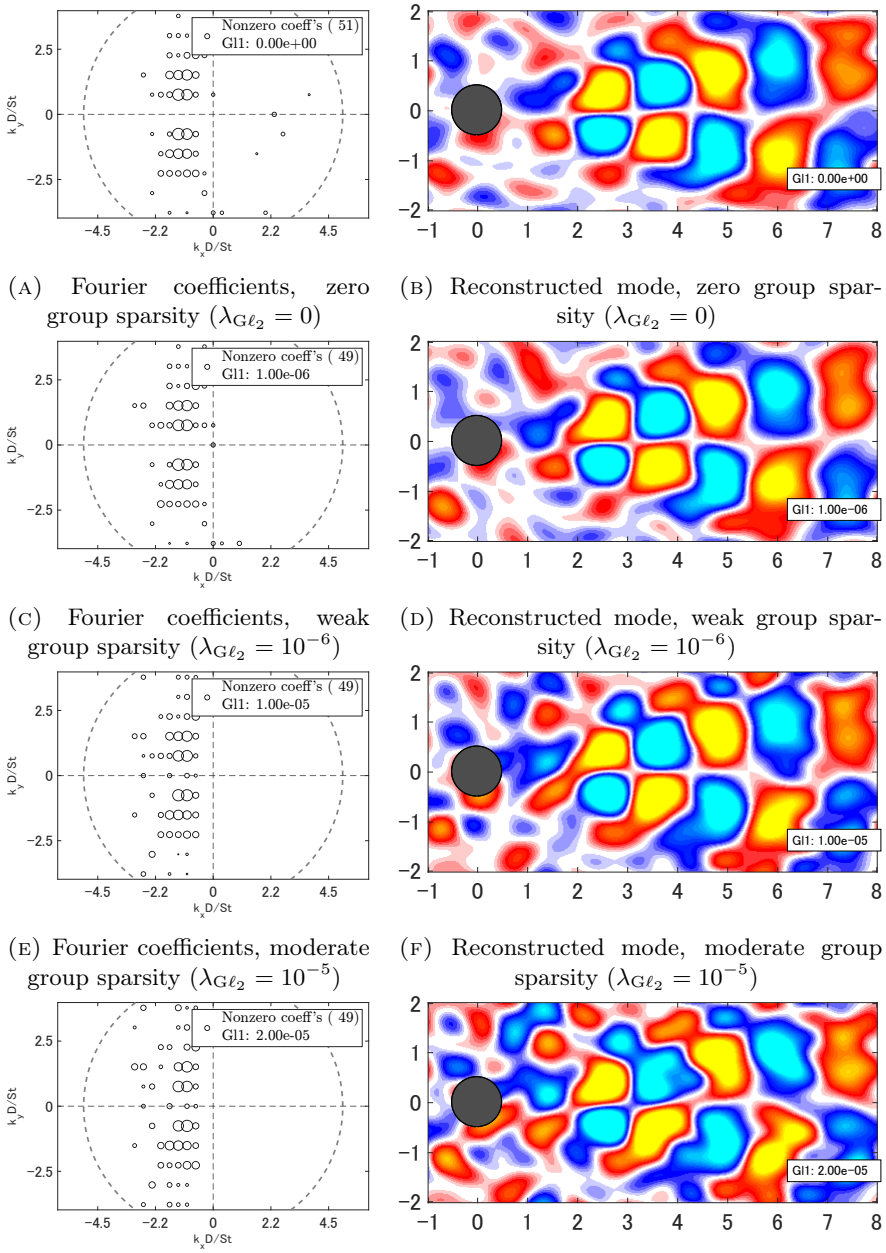
Investigation of group sparsity

Two sparsity-promoting terms distinctly affect the quality of the solution as defined in Eq. (3.34). Let us examine the compressive reconstruction resulting from various sparse parameters imposed on the group sparsity term. The goal here is to understand the behavior of the proposed CS-DMD enforcing the cluster in the solution.

The dataset used for the testing is the same as that given in Section 3.3. The results of Fig. 3.19 show that the reconstructed spatial distribution of the DMD mode of St=0.33. The results are based on the different regularization parameters for the group-sparse term, which are varied from $\lambda_{G\ell_2} = 0$ (zero group sparsity) to three small values. The estimation of zero sparsity only concerns componentwise sparsity in the sense of ℓ_1 norm, which is namely Cℓ_1 (w). The other three higher parameters are set to represent cases of weak, moderate, and somewhat stringent sparsity.

The left column in Fig. 3.19 indicates the nonzero coefficients identified by the CS-DMD with various group sparsity parameters. There are several nonzero values in the right-hand plane in Fig. 3.19a as expected due to the absence of the group sparsity. These false values disappear as the group sparsity is incorporated and the distributions of the solution tend to cluster structures in Figures 3.19e and 3.19g. Therefore, the last term in Eq. (3.32) plays a regularization role, yet it is also confirmed that there are several problematic nonzero coefficients identified on the line of $k_y = 0$ as $\lambda_{G\ell_2}$ increases. It may be worthwhile to incorporate constraints that are more related to a flow, such as the symmetry in the spatial (or in the parameter) structure.

The right column of Fig. 3.19 is the reconstructed spatial structures of the DMD mode. One may be impressed by the comparison between the reconstruction of C$\ell_1(w)$ therein and those of C$\ell_0(w)$ in Fig. 3.13h. The sparse regression based on the ℓ_1 penalization is assumed appropriate for this case, which is technically infeasible due to an astonishing computational cost of using



(A) Fourier coefficients, zero group sparsity ($\lambda_{G\ell_2} = 0$)

(B) Reconstructed mode, zero group sparsity ($\lambda_{G\ell_2} = 0$)

(C) Fourier coefficients, weak group sparsity ($\lambda_{G\ell_2} = 10^{-6}$)

(D) Reconstructed mode, weak group sparsity ($\lambda_{G\ell_2} = 10^{-6}$)

(E) Fourier coefficients, moderate group sparsity ($\lambda_{G\ell_2} = 10^{-5}$)

(F) Reconstructed mode, moderate group sparsity ($\lambda_{G\ell_2} = 10^{-5}$)

(G) Fourier coefficients, stringent group sparsity ($\lambda_{G\ell_2} = 2 \times 10^{-5}$)

(H) Reconstructed mode, stringent group sparsity ($\lambda_{G\ell_2} = 2 \times 10^{-5}$)

FIGURE 3.19: Comparison of spatial distribution for several regularization of group sparsity, using reconstructed DMD modes of $St \approx 0.33$ for flow past cylinder without additive measurement noise ($k'_{\max} = 5$, $p = 128$ and $\|\hat{\mathbf{s}}\|_0 \approx 50$)

the whole basis vector. It is finally realized by reducing the problem size due to the selective use of basis vectors in Section 3.2.5. Regarding the group sparsity, little differences were observed except for a deviation in Fig. 3.19h where a distribution of strong vorticity lies in the freestream. Although they are in fact less effective components for this experiment, it is expected that a too

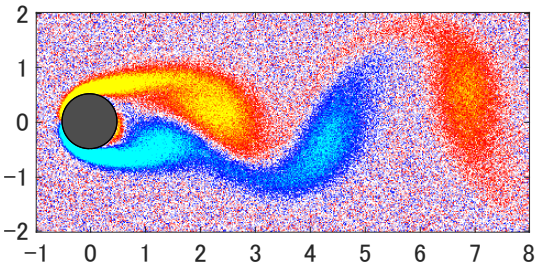


FIGURE 3.20: Instantaneous distribution of simulated vorticity data corrupted by additive measurement noise according to Eq. (3.41)

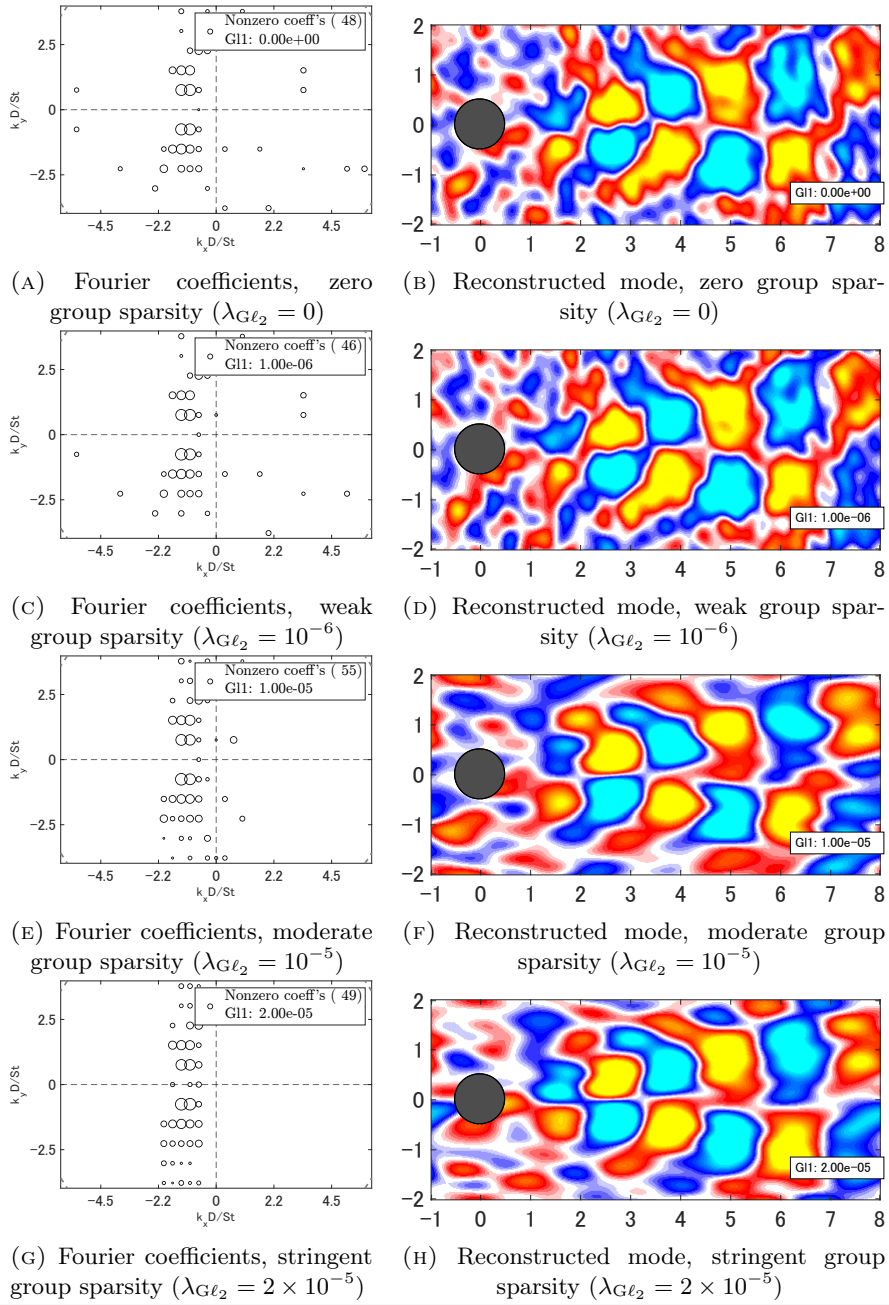
high value assigned for the group sparsity leads to an over-regularized solution. In this case, an appropriate value for $\lambda_{G\ell_2}$ is supposed to be between 10^{-6} and 10^{-5}.

One can observe a more prominent difference when the additive noise is contaminated and a sparse solution is hardly acquired by normal compressive sensing procedures. Reconstruction is carried out for the corrupted dataset, which is contaminated with independent Gaussian noise that changes with time. The intensity of the noise is proportional to the variance of the plain data σ_{d}^2 which is the maximum over the each datapoints distributed in space and is controlled by a specified constant $c = 0.2\,\sigma_{\mathrm{d}}$ as Eq. (3.41).

$$\mathbf{x}_{\mathrm{noise}} \sim c\mathcal{N}(\mathbf{0}|\mathbf{I}_N) \tag{3.41}$$

This noise term is added to $\mathbf{X}$, $\mathbf{X}'$ in Eq. (3.1). This means the standard deviation of the measurement noise is 0.2 of the maximum fluctuation of vorticity. The distributions used for the test is finally obtained as shown in Fig. 3.20. For compressive reconstruction, sparse measurements are similarly defined for the noiseless case, and the following results share the same configuration of point measurements.

The window size for the compressive reconstruction is set to $k'_{\mathrm{max}} = 7$, which is a more severe condition compared to the previous investigation. The inferred Fourier coefficients in the left column in Fig. 3.21 indicate effective reconstructions that take advantage of regularization to form a cluster. In this case, the "weak" group sparsity does not contribute to the reliable reconstruction; therefore, the reconstructed spatial distribution has finer-scale fluctuation. When a stronger group sparsity is imposed, the estimated distribution returns a more realistic spatial



(A) Fourier coefficients, zero group sparsity ($\lambda_{G\ell_2} = 0$)

(B) Reconstructed mode, zero group sparsity ($\lambda_{G\ell_2} = 0$)

(C) Fourier coefficients, weak group sparsity ($\lambda_{G\ell_2} = 10^{-6}$)

(D) Reconstructed mode, weak group sparsity ($\lambda_{G\ell_2} = 10^{-6}$)

(E) Fourier coefficients, moderate group sparsity ($\lambda_{G\ell_2} = 10^{-5}$)

(F) Reconstructed mode, moderate group sparsity ($\lambda_{G\ell_2} = 10^{-5}$)

(G) Fourier coefficients, stringent group sparsity ($\lambda_{G\ell_2} = 2 \times 10^{-5}$)

(H) Reconstructed mode, stringent group sparsity ($\lambda_{G\ell_2} = 2 \times 10^{-5}$)

FIGURE 3.21: Comparison of spatial distribution for several regularization of group sparsity, using reconstructed DMD modes of $St \approx 0.33$ for flow past cylinder with additive measurement noise ($k'_{\max} = 7$, $p = 128$ and $\|\hat{\mathbf{s}}\|_0 \approx 50$)

structure, although the reconstruction accuracy is inferior to that in the noiseless case. An appropriate $\lambda_{G\ell_2}$ is supposed to be 2×10^{-5} or higher in this case, which differs from the previous investigation without additive noise to the plain dataset.

In summary, the effect of group sparsity is investigated by varying the sparsity parameter

controlling the degree of clustered coefficients. Stipulated inference was certainly realized by the group-sparse term in the sparse regression, yet it is still challenging to presume an appropriate parameter for this term. The improvement due to the group sparsity was more clarified when the dataset was corrupted by artificial measurement noise, which means that it will be more appealing when applied to real measurement data. Accordingly, in the following results of the quantified comparison in Section 3.3.5, the data used for the reconstruction are obtained using the plain dataset without additive noise. The comparison is carried out including the proposed method denoted as Cℓ_1+Gℓ_1 (w), with $\lambda_{\mathrm{G}\ell_2} = 10^{-6}$ corresponding to the "weak" group sparsity setting in the preceding investigation.

3.3.5 Quantitative comparison of recovery

The reconstruction quality of the CS-DMD Eq. (3.25) is evaluated using the discrepancy between the representation coefficients of the DMD modes obtained from field measurements and those from the compressive measurements. The first metric is the difference between the two normalized complex vectors regarding the ℓ_2 norm,

$$\epsilon\left(\mathbf{a}\,,\,\mathbf{b}\right) = \min_{\theta} \|\mathrm{R}(\theta)\mathbf{a} - \mathbf{b}\|_2, \tag{3.42}$$

where the rotation $\mathrm{R}(\theta)\,(\|\mathrm{R}(\theta)\|_2 = 1)$ compensates the phase difference between two complex vectors. This is determined by minimizing the difference, and specifically, the pseudo inverse of either vector is taken. Another metric is the detection rate, which denotes how much the nonzero entries in the original vector have been spotted;

$$\delta\left(\mathbf{a}\,,\,\mathbf{b}\right) = \|\mathrm{supp}(\mathbf{a}) \cap \mathrm{supp}(\mathbf{b})\|_0 / \|\mathrm{supp}(\mathbf{a})\|_0, \tag{3.43}$$

where the set of indices for the nonzero entries in $\mathbf{a}$ is called its support and it is denoted by $\mathrm{supp}(\mathbf{a})$. The numerator identifies the intersection of each support, and the ratio is taken by counting the member. The vector in the denominator, a, is treated as a reference, and two vectors that are close to each other tend to return unity, while incoherent estimates of CS-DMD

return nearly zero.

The error plots in Fig. 3.22 quantify the performance of the three different implementations of CS-DMD for the first four DMD modes obtained. These figures compare the general form of CS-DMD in Brunton et al., 2015 against that using the selective basis vectors (Section 3.2.5) and further combined with the structured inference of ADMM using the Cℓ_1+Gℓ_1 term (Section 3.2.6). The reconstruction error is calculated for the reconstruction of the first 4 DMD modes, $St \approx 0.16, 0.3, 0.5, 0.6$ which are shown in Fig. 3.8. The results shown are averaged from 36 different trials of measurements at randomly selected points, which are common for each method. Each symbol represents the median values of the tests, while the error bars indicate the maximum (*i.e.* worst-case) and the minimum (*i.e.* best-case) values, respectively. These results are obtained by the threshold Eq. (3.38) for $k' = 5$. Each reconstruction is made with at most, $\|\hat{\mathbf{s}}\|_0 = 50$ components of the Fourier basis, except for the results of ADMM that employs a Cℓ_1+Gℓ_1 penalty term. It should be noticed here that the number of nonzero components is assumed 50, therefore these estimation methods left an inferior impression for the results of smaller measurement size, $p < 50$. This can be observed in the decrease in the range of error bars of Cℓ_1+Gℓ_1 around $p \geq 2^6 = 64$ in Fig. 3.22a. It is also reflected in the change in the trends of Cℓ_1+Gℓ_1 around $p \geq 2^6 = 64$ in Figures 3.22b to 3.22d.

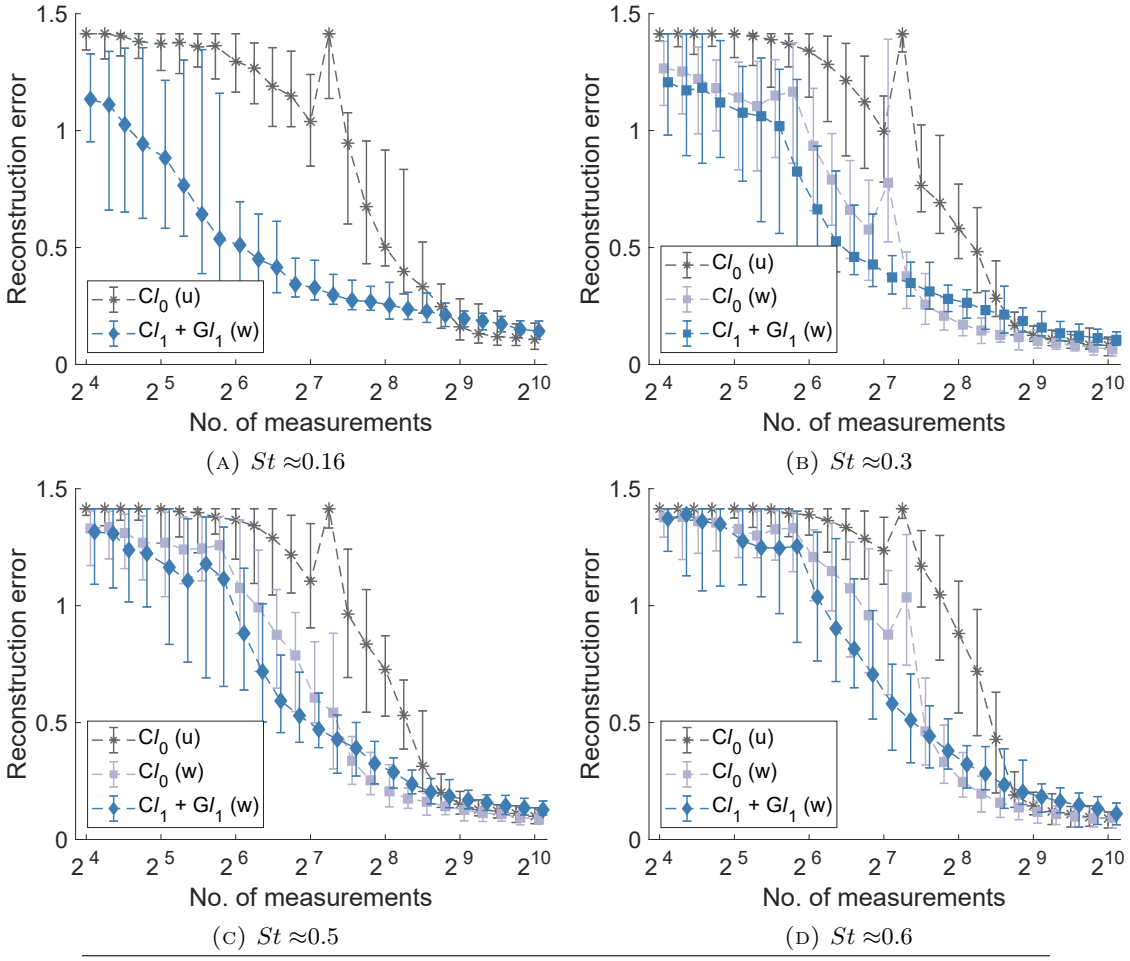
From these results in Fig. 3.22 except for $p < 2^6$ for the above reason, several remarks are found as summarized in the following list:

1. Lower values of reconstruction error are achieved by the reduced representation basis vectors by window-selection.

2. The results of Cℓ_1+Gℓ_1 are better than those of Cℓ_0 for $2^6 \leq p < 2^8$.

3. On the contrary, the results of Cℓ_1+Gℓ_1 are worse than those of Cℓ_0 for a larger number of measurements.

One can confirm the positive effect of introduced selective basis vectors from the first observation above for smaller p, yet the margin gradually decreases as p increases. The latter decrease is because the number of obtained measurements is too sufficient for the inference of coefficient



(A) $St \approx 0.16$ (B) $St \approx 0.3$ (C) $St \approx 0.5$ (D) $St \approx 0.6$

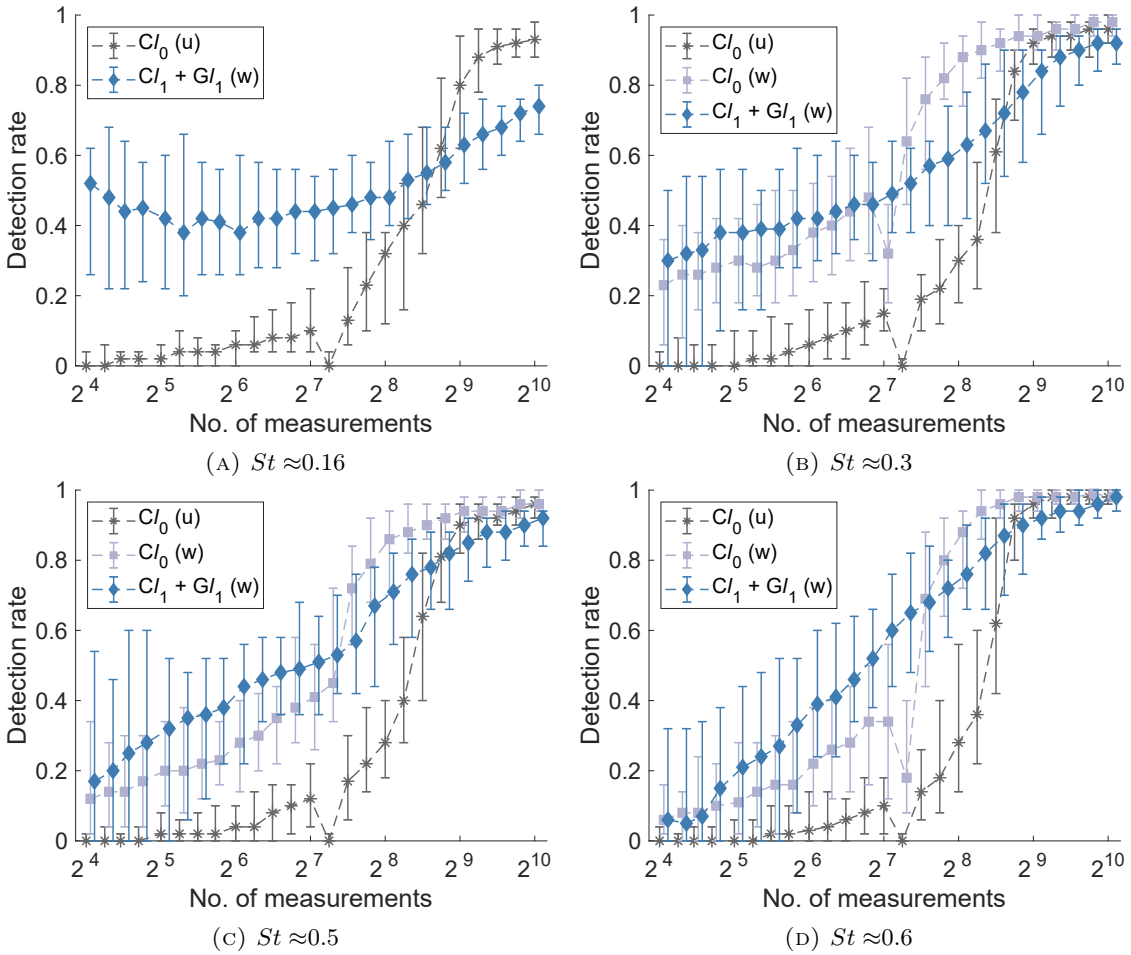
FIGURE 3.22: Reconstruction error of DMD modes for flow past cylinder

values of at most 50 components. Moreover, for every graph in Fig. 3.22, the performance of the regression with the group sparsity, Cℓ_1+Gℓ_1, is dominant in the region of $2^6 \leq p < 2^8$. This immediately shows a significant improvement by the presented method. These observations are in good agreement with the previous qualitative evaluation of the reconstructed distribution of DMD modes of Figures 3.13 and 3.14. However, results in a region of $p \geq 2^8$ show different trends. The results of component-wise sparsity continue to decrease while the reconstruction error for Cℓ_1+Gℓ_1 is slightly worse than the others. This is not anticipated from the results of reconstructed distributions in Fig. 3.13.

Concluding the performance from the above comparison leaves several questions that have not been clarified. The key aspect to be taken into account is the number of nonzero entries $\|\hat{\mathbf{s}}\|_0$ based on the Lasso regression of Cℓ_1+Gℓ_1. The number of nonzero entries identified by the

ADMM algorithm and the subsequent polishing step is indicated in each picture of the Fourier coefficients in Fig. 3.14. Although almost all coefficients reside in the left-hand plane in the figures of the rightmost column, $\|\hat{\mathbf{s}}\|_0$ is around 30–40 for $p = 64, 128, 256$. This might work positively for inference suppressing overfitting when the measurement size is relatively small, whereas it limits the representation for a larger number of measurements. This is confirmed by the following comparison for different measurement sizes $\|\hat{\mathbf{s}}\|_0$ in Fig. 3.15.

The correctness of the inferred support of the solution vector is subsequently compared using the results of the detection rate for each DMD mode. It should be noted that a higher value means better recovery by sparse reconstruction. Figure 3.23 shows that similar trends are observed as



(A) $St \approx 0.16$ (B) $St \approx 0.3$

(C) $St \approx 0.5$ (D) $St \approx 0.6$

FIGURE 3.23: Detection rate of 50 nonzero elements for DMD modes $St \approx 0.33$

shown in the previous comparison of reconstruction error in Fig. 3.13. At first, the results due to the group sparsity Cℓ_1+Gℓ_1 show better performance when $p < 2^8$. However, the other plots

gradually overtake as p increases. The turnaround between Cℓ_1+G$\ell_1(w)$ and C$\ell_0(w)$ occurs around $p \approx 2^7$, which is the same circumstance in Fig. 3.13. It is supposed that the Lasso inference including a group sparsity requires careful choosing of the sparsity parameter.

Furthermore, the variance of the results is relatively high at first, yet diminishes as p increases. This may be due to the random arrangement of the sensing positions for these results.

3.4 Conclusions

In this chapter, the advanced implementation of Compressed Sensing DMD (CS-DMD) is developed to reconstruct the spatial distribution of each DMD mode from spatially sparse measurements. In contrast to the general formulation based on compressive sensing theories, the modified approach of CS-DMD leverages the notion of structured sparsity to recognize a more physical distribution hidden in the partially observed data. The presumed structures in fluid flows are listed in Section 3.2.4, especially with a focus on reconstruction by means of the Fourier transformation. Several observations on the structure are provided to show the hierarchical properties underlying the fluid behavior.

The introduced modifications are the following;

- Selective use of Fourier basis vector for compressive representation to suppress the unrealistic fine-scale reconstruction,
- Structural inference for recovery by CS, especially by promoting the clustered distribution of representation coefficients in the wavenumber domain.

These considerations are expressed in algorithms to acquire a more reliable estimation of DMD modes. First, a library matrix for the compressive representation was reduced in advance according to the selection rule using a threshold based on the temporal oscillation frequency of the target DMD mode. This procedure immediately reduces the dimension of the solution space; therefore, overfitting due to the exceeding ill-condition can be suppressed for a smaller number of measurements. In addition, the second modification was properly triggered by the group-wise sparsity-promoting term in the inference step. The cluster structure in the parameter space is

relaxed to the correlation in the adjacent entries, which encourages the nonzero entries to reside next to each other. Accordingly, overlapping groups are invoked by arranging the first adjacent components to each entry into a corresponding group. The optimization problem for the inference is therefore posed by the minimization of the sum of the residual and metrics of entry-wise and group-wise sparsity, which is solved by ADMM seamlessly.

The demonstration is carried out using data of the two-dimensional CFD flow past a circular cylinder in Section 3.3. By analyzing the DMD modes obtained from the complete data, the coincidence in the peak wavenumber is utilized to determine the threshold for the available wavenumber for the reconstruction. The results of reconstruction with CS-DMD based on different regression types show the effectiveness of the modifications introduced as intended. The reconstructed distributions of the DMD modes are significantly improved for a smaller number of measurements, where the original implementation suffers from the ill-condition of the inference. It became even more obvious when comparing the reconstruction error Fig. 3.22 and the detection rate of the appropriate nonzero entries Fig. 3.23. The dependence of reconstruction results on the model parameters is also confirmed in this study, by varying the sparsity during the compressive inference of the threshold for the window in the wavenumber domain.

Chapter 4

Optimized Measurement Position for Compressive Flow Modeling

The randomness significantly contributes to the basic theories of compressive sensing, as mentioned in Section 3.2.2. Several prerequisites are sought for exploiting the notion of sparsity, where only a small portion of signals can contain nonzero coefficient values;

- The signal has to be sparse or compressible with respect to the predefined basis vectors. Here, one usually assumes randomness in the positions of nonzero entries or their values.
- The success of recovery depends largely on the measurement obtained to realize the sketch without loss of information on the signal. This is generally satisfied if the measurement matrix is constructed using random variables according to the sub-Gaussian distribution (R. Baraniuk et al., 2008).

These necessities will surely impose on us the burden of detecting the signal in an unnecessarily broader subspace using extra measurements. A remedy for this perspective is partially settled for the application of CS-DMD for fluid flows in the previous chapter, by means of the selective use of basis vectors and the structured estimation. This chapter concerns the optimization of the measurement matrix for CS by selecting appropriate positions to arrange the sensing.

4.1 Background

4.1.1 Measurement optimization applied to compressive sensing

The restricted isometry property (RIP) is known to be one of the sufficient conditions for the complete recovery of a signal from compressed measurements. As developed in R. Baraniuk et al., 2008; Candès et al., 2006, random matrices generated from distributions of the sub-Gaussian class have the RIP with high probability. It has been pointed out that a deterministic method is sometimes necessary when considering physically possible observation systems. This randomness of observations is occasionally regarded as problematic, since that is hard to fulfill for the sensing system of distributed physical quantities, as pointed out for the example of a converter of infinite length continuous signals (Duarte and Eldar, 2011) and the imaging using compressive sensing (Boyer et al., 2019), to name a few examples. Their claims are summarized in the following bullet list,

- *Hardware limitation:* It is hard to apply the randomized measurement matrix to the measurement apparatus because the measurement relies much on the physical process.
- *Implementation:* The number of measurements is limited due to cost budget and hardware considerations.
- *Too strict conditions in effect:* It works well even though the actual implementations violate the condition demanded by the theory.

Therefore, it is worthwhile to facilitate the structured acquisition of structured signals. In normal problem settings of compressive sensing, one cannot assume in advance where the K nonzero elements exist in a large-dimensional vector of parameters. As denoted in one of the interpretations for the RIP, it is implied that information from K-sparse parameter vectors must be projected through the measurement process. Stable reconstruction in a general sense obliges the sensing system to cover evenly all basis vectors defined on the discretized grid, including those less likely to be centric for the reconstruction. However, in Chapter 3, the efficient reconstruction

of CS-DMD is successfully performed at the cost of the robustness of compressive reconstruction. Thus, in the same sense, the number of required measurements is subject to reduction by concentrating on a small part of the informative subspaces.

4.1.2 Adaptive measurement optimization

Due to the great number of basis vectors, although reduced as demonstrated in the preceding chapter Chapter 3, the measurement system is essentially underdetermined to a high degree. Therefore, this section provides a different approach that adaptively appends sensing positions by incorporating information on each basis vector, which is acquired by the improved estimation in Sections 3.2.4 and 3.2.5. Although mounting additional sensors to the object is not straightforward, adding a measurement can be easily implemented in the reconstruction of a flow using sparse visualization methods such as Kanda et al., 2022. It should also be noted that a similar methodology can be applied to other scientific and industrial areas. The importance of effective measurement positions is demonstrated by the methods in this section, which aim to append measurement positions for better, more reliable estimation.

Some methods to select sensing positions extend the optimization regarding the Fisher information matrix introduced in Chapter 2 to the Bayesian formalism. The Bayesian counterpart conditions an uncertainty in the estimation and leverages for the evaluation of upcoming measurements. Examples are found in signal processing applications, known as sensor networks (Liu et al., 2016; Shamaiah et al., 2010; Tzoumas et al., 2016; Uciński, 2020). The information from distributed sensors is gathered at the fusion center, and the state of the target is estimated from the gathered measurements. The estimation is often delivered by use of the Kalman Filter, which updates the covariance matrices of the state and measurement of a state-space model. These updated matrices enable tracking of the state variables of the target that change over time, by constantly optimizing the measurement system. In short, these formulations assume known dynamical systems as denoted in Chapter 2, while there is no assumption on the signal vector, such as sparsity. The evaluation criterion utilizes scalar metrics of the error covariance matrix resulting from the maximum a posteriori estimation of Kalman filters. In this study, the use

of the greedy method is encouraged because of the simple expressions during the evaluations of sensing positions. This will ease the burden of dealing with a large number of potential measurement positions, which are distributed as the region of interest. It is confirmed in the discussion of Jawaid and Smith, 2015; Shamaiah et al., 2010, a certain kind of evaluation metric is proved to have the property of submodularity utilized for the sequential selection of the measurement position by a greedy method, which is also ensured in the observability Gramian Chapter 2.

As a complementary method to those treated in the preceding review in Section 4.1.1, some techniques have been provided to amend the unfavorable underdetermined measurement by adaptive selection of measurement positions based on a Bayesian formalism. If the prior information sufficiently elucidates the uncertainty of each parameter, it can be used in advance to allocate sensing positions based on what should be scrutinized by the compressive sensing. Furthermore, if one obtains "error bars" of the estimation of CS in a Bayesian formalism (Ji et al., 2008; Kilic et al., 2022; Wang and B. Zhang, 2012), they can be used as a measure of confidence in the current result of the parameter estimation. It may be used as posterior information to adaptively reinforce the sensing for a more targeted measurement system. For example, in the Bayesian Compressive Sensing (BCS) of Ji et al., 2008, the sparse parameter vector is obtained by iterative substitutions using likelihood functions, such as performed in Expectation-Maximization (EM) algorithms. This framework postulates a hierarchical structure onto the sparse parameter vector, and maximum likelihood processes illustrate the distribution of the target sparse vector. The variance in the distribution enables one to update the sensing system by quantifying the uncertain elements. This information of error bars will be suitable for dictating the measurement optimization of this study. The Bayesian approach can be employed by integrating this information into the selection criterion in Section 4.2.2.

A covariance matrix of given variables expresses how much variation the variables can have. Especially in Bayesian CS, this should first be assumed as hyperparameters in advance based on the knowledge for the target phenomenon. This portion reflects the a priori information on the distribution of intensities of the parameters. It simultaneously dictates the uncertainty in the current estimation by introducing the result of Bayesian inference, resulting from the

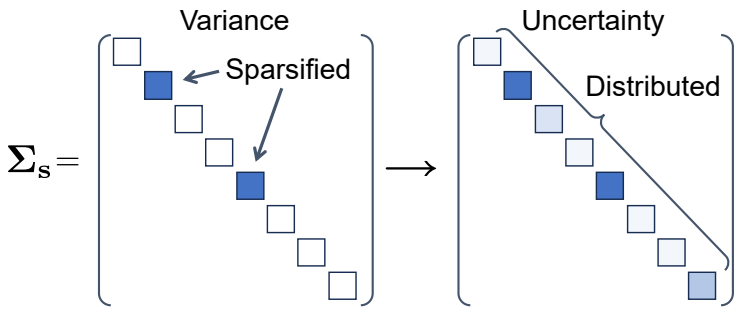


FIGURE 4.1: Parameter covariance matrix leveraged in the earlier and the current formulations of measurement optimization

current set of the measurement vectors and the assumption on sparsity. One should note here that the earlier attempt of measurement optimization of Bayesian CS only matters the variance of the parameter with respect to a sparsified subspace, while mentions on this are deferred to the following sections providing the proposed method Section 4.2.3. In this study, a method of measurement optimization is extended to utilize this covariance matrix, not as a variance of the inferred sparse parameter but as an uncertainty of the inference process, as shown in Fig. 4.1.

4.1.3 Objective and contribution

Concrete guidelines for measurement design are irreplaceable for actual implementations of CS-DMD, as several studies pointed out for general cases of CS Section 4.1.1. In Chapter 3, the tailored estimation reduces the redundancy of the basis vectors for compressive reconstruction. The results therein show that the reconstruction of the spatial distribution of DMD modes is substantially improved by reducing the freedom of degrees and by casting certain structures in the coefficient vectors. This hints at the nonuniform distribution of larger values in the coefficient vector representing the DMD mode with respect to the Fourier basis vectors. Therefore, it is supposed to be desirable to somehow establish a sensing system focusing on a subspace tailored for the "biased" components among all available basis vectors, although it does not explicitly obey the performance guarantee of compressive sensing.

Accordingly, the objective of the current study is to develop techniques of measurement optimization in order to further compress the dimension of measurement by structural measurement

design and improve reconstruction performance. Sensing systems are combined with the estimation methods from the previous chapter to construct a unified system that more efficiently executes the mode recovery process by CS-DMD. The main claims of the current chapter are described as;

- Demonstrated the capability of applying the previous selection schemes of measurement positions based on the state estimation in a linear system to the current consideration
- Established an adaptive method for measurement position selection with a *a posteriori* covariance matrix of uncertainty of estimated sparse parameters
- Improved reconstruction accuracy of DMD modes in the CS-DMD with structurally obtained sparse measurements

4.2 Methods

This section describes the optimization methods presented in this study, which obtain effective measurement positions to promote reliable compressive sensing with a smaller number of measurements. First, Section 4.2.1 comes with short reviews of methods for measurement optimization in compressive sensing. The formulation of the optimization problem is then provided in both a predetermined Section 4.2.2 and an adaptive manner Section 4.2.3 of the measurement design. It should be noted that selection methods for sensing position, such as those treated in Chapter 2, can be immediately adopted for this purpose.

4.2.1 Design of measurement system in compressive sensing literature

Several studies have thus far investigated how to design observations via a deterministic pathway to develop a structured sensing system. Some examples can be found in Duarte and Eldar, 2011 and other intensive works with more realistic applications of CS in each target (Y. Chen et al., 2023; Jiang et al., 2022). They have been established to encompass the structural way of measurement design, instead of the robustness and performance guarantee provided by the configuration under the compressive sensing theory. In this section, some studies will be reviewed to

share some methods that aim to optimize measurements. The relationship with the optimization method for the ordinary, noncompressive inverse problems will be provided.

In a common approach, the entries in the measurement matrix are determined by sampling from a given distribution to satisfy the notion of RIP. This process ensures that the difference between two elements is preserved by the transform by the measurement matrix, and it guarantees the uniqueness of the reconstruction (R. Baraniuk et al., 2008). Meanwhile, there have been several studies to embed structural construction strategies in measurement design. Some studies focused on how information is preserved by various types of deterministic measurement matrices. In the earlier attempt in DeVore, 2007, a structural strategy was presented by approximately bounding the spectral norm. The lack of efficiency in signal recovery was noted in (Berinde et al., 2008; Gilbert and Indyk, 2010) as the bound for the inequalities gives a poorer impression than that due to random matrices. Nevertheless, a deterministic construction of the CS matrix performs well, at least empirically, in some applications. The importance of considering the physical measurement process was stressed in Duarte and Eldar, 2011, as measurement vectors are chosen from a predetermined library that reflects the physical constraints of acquisition, such as magnetic resonance imaging (MRI) and analog-to-digital converters. Boyer et al., 2019 further embedded the notion of blocked measurement, which also considers the process of MRI. Joneidi et al., 2020 focused on reducing a high-dimensional representation matrix that originally represented large-scale data. The measurement matrix is constructed to minimize the constant used in the RIP inequality Eq. (3.14). This can be interpreted as a deviated optimization problem in the first chapter Eq. (P0), where the expected error is considered on the unit sphere. Instead of optimizing the matrix determinant of the Fisher information matrix known as D-optimality, the formulation in Joneidi et al., 2020 formulated the worst error minimization for the union of K-sparse vectors. It was a double-fold combinatorial optimization of subset selection for an arbitrary K-sparse vector, and the selection was finally carried out using a relaxed problem.

Similar approaches are presented in this chapter by applying an optimization method based on a general linear inverse problem to a compressive stage, where the observable subspace is gradually sought by measurement. The general form considering non-sparse vectors is first

presented in Section 4.2.2, which is followed by an adaptive selection method of measurement positions that identifies important subspaces provoked by estimation uncertainty in Section 4.2.3.

4.2.2 A priori optimization of measurement position for reduced redundancy

The process of compressive sensing is considerably akin to the state estimation of sparse sensing, the latter of which is previously described in Chapter 2. Based on a linear measurement model, $\mathbf{y} = \mathbf{C}\mathbf{s}$, the parameter vector $\mathbf{s}$ is reconstructed as an estimate by inverse inference from the measurement vector $\mathbf{y}$. The clear difference is in the number of measurements and in the number of parameters to be estimated. In the framework of sparse sensing, the inference is performed based on the reduced-order model, which has only a few degrees of freedom underlying. A small number of parameters are estimated from observation by means of maximum likelihood or maximum a posteriori schemes. On the contrary, in compressive sensing, the dimension of the parameter vector is overwhelmingly high as compared to that of the measurement vector, although the sparsity corroborates that only a small portion of the coefficients has nonzero values. One may utilize a bit of specialized tools to detect the index for the nonzero values by enforcing the sparsity in the estimate solution. Former frameworks of sparse sensing can be interpreted as the ultimate case in which we know which elements in the parameter vector take nonzero values.

Measurement optimization has been studied to identify a meaningful trial or measurements to capture sufficient information about a system, which is known as the design of experiments (Bates et al., 1996; Kincaid and Padula, 2002; Udwadia, 1994). The measurement sensitivities to each parameter are stored for configurable positions in a candidate matrix, where the optimization is carried out by the subset selection of rows in the candidate matrix. This sensitivity matrix is known as the Fisher information matrix (Joshi and Boyd, 2009; Manohar et al., 2018; Nakai et al., 2021; Saito et al., 2021), or the observability Gramian in the dynamical standpoint (DeVries et al., 2012; Wouwer et al., 2000; Yamada et al., 2023). The size of the measurement is overwhelmed by the enormous number of latent degrees of freedom, which results in the optimization objective for the so-called underdetermined case. The optimization objective is posed as follows, with a

bit of abuse of notations from Chapter 2;

$$\mathcal{I}_p = \{ i_1, \ldots, i_p \} = \underset{|\mathcal{I}|=p}{\arg\max} \quad f\left(\left(\mathbf{P}(\mathcal{I}) \check{\mathbf{\Psi}}_{\mathbf{X}} \right) \left(\mathbf{P}(\mathcal{I}) \check{\mathbf{\Psi}}_{\mathbf{X}} \right)^* \right), \tag{4.1}$$

where $\mathbf{P}(\mathcal{I})$ extracts rows of $\check{\mathbf{\Psi}}_{\mathbf{X}}$ indicated by indices $\mathcal{I}$ and f is a function of matrix of an appropriate dimension $f : \mathbb{R}^{p \times p} \to \mathbb{R}$ such as the matrix determinant or trace. $\check{\mathbf{\Psi}}_{\mathbf{X}}$ denotes the selectively adopted representation basis matrix defined in Eq. (3.30).

Qualitatively, optimization in an underdetermined system is described as related to maximizing the information obtained from observation. The determined subset of sensing sites achieves the best optimization measure along a given p-dimensional subspace subsampled from the original N-dimensional space. The determined measurement systems preserve information about their corresponding p-dimensions, thus, the complete inverse estimation is guaranteed if the signal has supports on the subspace. The interpretation of this is guided by Nakai et al., 2021 by relating with the estimation based on the minimum norm. Another preferable aspect is that this strategy for the determining sensing position does not indeed need a priori field measurements that contain a number of spatial information, as assumed in the previous attempts of selection mentioned earlier. However, this method is certainly thought of as a kind of ad-hoc approach due to some concerns regarding the use under compressive sensing;

- The optimization is based on an assumption of minimum-norm estimation in the previous formalism of sensing optimization. This is not consistent to the sparsity-promoting estimation of compressive sensing, which is, in fact, in contrast to the RIP-based optimization in Section 4.2.1.

- The selection result is not completely deterministic for a certain kind of basis, such as a complex Fourier basis. Specifically, the first choice of position is not unique because of the uniform norm of each basis vector.

- The basis matrix has to be reduced in advance to some extent for meaningful selection. This is a technique for suppressing redundancy of measurement for a basis matrix that is reduced or coarse-grained in parameter space.

The second perspective is because the determination of the measurement positions only certifies the orthogonality between measurements, and thus, gives positions mere relatively. Combining the third aspect with this, there could be little difference from random placement without a considerable number of measurements compared with the number of basis vectors. One should also note that reducing the robustness leads to aliasing if the modeling has a mismatch in the selection of the basis vector introduced in the preceding section. That said, the application in the CS-DMD is performed to investigate the outcome of delimiting the target subspace of sensing.

Finally, a simple form of the greedy method is proposed based on the logarithm of matrix determinant to select a subset of measurement positions, which is denoted as the D-optimality criterion (Udwadia, 1994). This method introduced from the design of experiments is not completely consistent with the current problem setting, yet it is still worth considering because of its stiff theoretical background. The greedy selection method of (Saito et al., 2021) is introduced to select the measurement positions that is sensitive to the tailored representation matrix $\check{\mathbf{\Psi}}_{\mathrm{X}}$ of Eq. (3.30). The greedy selection evaluates the matrix determinant of the sensitivity matrix, which is constructed from the selected rows of the candidate matrix. It is proposed in the present study that measurement positions are selected based on the windowed representation basis, $\check{\mathbf{\Psi}}_{\mathbf{X}}$;

$$\text{(Repeat until } q = p) \tag{4.2}$$

$$\begin{cases} i_q = \underset{i \in \mathcal{I} \setminus \mathcal{I}_{q-1}}{\arg\max} \quad \log\det\left(\left(\mathbf{P}(\mathcal{I}_{q-1} \cup \{i\})\check{\mathbf{\Psi}}_{\mathbf{X}}\right)\left(\mathbf{P}(\mathcal{I}_{q-1} \cup \{i\})\check{\mathbf{\Psi}}_{\mathbf{X}}\right)^*\right), \\ \mathcal{I}_q = \mathcal{I}_{q-1} \cup i_q, \end{cases} \tag{4.3}$$

where $\mathbf{P}$ extracts the rows of $\check{\mathbf{\Psi}}_{\mathbf{X}}$ corresponding to the index subset of selected positions. The total selection procedure is noted in Alg. 7. This form of measurement optimization is referred to as "DG" selection, which is according to the earlier methods treating the D-optimality criterion (Saito et al., 2021).

Algorithm 7 A priori optimization of measurement position [DG]

Input: $\check{\mathbf{\Psi}}_{\mathbf{X}} \in \mathbb{C}^{N \times \check{N}}$, $p \in \mathbb{N}$
Output: Indices of chosen p sensing positions $\mathcal{I}_p$
$\mathcal{I}_n \leftarrow \{1, \ldots, n\}$, $\mathcal{I}_0 \leftarrow \emptyset$,
for $q = 1, \ldots, p$ **do**
$\mathcal{I}^* \leftarrow \left\{ i : i \in \underset{i \in \mathcal{I}_n \setminus \mathcal{I}_{q-1}}{\operatorname{argmax}} \operatorname{rank} \left[\left(\mathbf{P}\left(\mathcal{I}_{q-1} \cup i\right) \check{\mathbf{\Psi}}_{\mathbf{X}} \right) \left(\mathbf{P}\left(\mathcal{I}_{q-1} \cup i\right) \check{\mathbf{\Psi}}_{\mathbf{X}} \right)^* \right] \right\}$
$i_q \leftarrow \underset{i \in \mathcal{I}^*}{\operatorname{argmax}} \log\det \left[\left(\mathbf{P}\left(\mathcal{I}_{q-1} \cup i\right) \check{\mathbf{\Psi}}_{\mathbf{X}} \right) \left(\mathbf{P}\left(\mathcal{I}_{q-1} \cup i\right) \check{\mathbf{\Psi}}_{\mathbf{X}} \right)^* \right]$
$\mathcal{I}_q \leftarrow \mathcal{I}_{q-1} \cup i_q$
end for

4.2.3 Measurement optimization conscious of uncertainty in CS

The selection in Section 4.2.2 reduces redundancy, in fact, in measurement by considering the orthogonal projections contained in the measurements using a small portion of the basis vectors. However, it does not necessarily guarantee complete signal recovery, because the subspace spanned by the measurement is chosen uniformly by the selection method based on Eqs. (4.1) and (4.3). Therefore, this section explores the methodology to focus on the identified subspace for reconstruction, which utilizes the improved estimation scheme in Chapter 3. Some expressions are first borrowed from the previous work to illustrate the aspects in common with our intentions, while the new framework integrated in this study is located later.

Again, each basis vector is assumed to have different levels of importance, as utilized in Chapter 3. The selection of sensing positions now reflects the bias for each basis vector, according to the weighting matrix. One can leverage some existing selection algorithms from the sensor networks literature, despite the absence in the assumption of the sparsity in the parameter. In the works of Jawaid and Smith, 2015; Shamaiah et al., 2010, the selection method extends to the estimation derived for the estimation by Kalman filtering. A state-space representation of dynamical models is first assumed as in Shamaiah et al., 2010, the same form of Chapter 2, repeated as:

$$\mathbf{y}_k = \mathbf{P}(\mathcal{I})\check{\mathbf{\Psi}}_{\mathbf{X}}\mathbf{s}_k + \mathbf{v}_k, \tag{4.4}$$

where some notations are defined the same as those in Eq. (3.18) besides the parameter $\mathbf{s}_k$ is

statistically represented by the covariance matrix $\mathbf{\Sigma_s} \in \mathbb{R}^{r \times r}$. The measurement matrix for the selected sensing sites $\mathbf{P}(\mathcal{I})\check{\mathbf{\Psi}}_\mathbf{X}$ is indicated by a permutation matrix Eq. (2.11) and a subsampled library matrix in Eq. (3.30). Selection is carried out on the scalar metric with a slight modification to Eq. (4.1);

[Shamaiah et al., 2010, Eq. (4),]

$$\mathcal{I}_p = \{ i_1, \ldots, i_p \} = \underset{|\mathcal{I}|=p}{\arg\max} \quad f\left(\left(\mathbf{P}(\mathcal{I})\check{\mathbf{\Psi}}_\mathbf{X} \right)^* \left(\mathbf{P}(\mathcal{I})\check{\mathbf{\Psi}}_\mathbf{X} \right) + \mathbf{\Sigma}_\mathbf{s}^{-1} \right). \tag{4.5}$$

The greedy selection is conducted in the same manner as Eq. (4.3), whose property of submodularity for certain types of metric guarantees optimization results bounded by a fraction of the true optimum. A similar formulation can be found that the selection criterion is equipped with the covariance matrix due to the measurement noise. It immediately loses the property of submodularity, yet greedy selection is empirically shown to have the capability of selecting the appropriate positions for sensing (Uciński, 2020; Yamada et al., 2021). This formulation sounds appealing, but one has to admit the chief dissimilarity between the error covariance above and that in compressive sensing (see Section 3.2.4); in the latter, the sparsity in the parameter vector must be privileged. Accordingly, correct evaluation of the covariance matrix is irreplaceable to enjoy the efficient selection procedure.

There should also be several ideas remarked from compressive sensing studies, which also proposed measurement optimization criteria similar to those introduced above. It will suffice to introduce measurement optimization in the Bayesian context in the earlier work presented by Ji et al., 2008 and several variants found in Jiang et al., 2022; Kilic et al., 2022; Wang and B. Zhang, 2012. As previously summarized in Section 3.2.4, the compressive estimation of BCS hypothesizes a sparse prior that promotes estimation as a sparse vector. By imposing a hierarchical distribution, nonzero entries are detected to be in a few portions of the parameter vector. This procedure is of pure Bayesian formalism; thus, the posterior distribution is invoked by the measurement and the appropriate prior. The adaptive determination of a measurement vector is proposed in Ji et al., 2008, Section IV-A by exerting the error covariance matrix of

the Gaussian distribution, which represents "error bars" of the resulting estimation. It enhances the ability for compressive estimation by evaluating this posterior distribution as a quality of the parameter estimate. After the error covariance is determined by the process above, the measurement matrix $\mathbf{\Psi_X}$ appends a new row $\mathbf{r_X}$ of which projection reduces the differential entropy the most $h(\hat{\mathbf{s}})$;

[Ji et al., 2008, Eq. (17)]

$$h(\hat{\mathbf{s}}) = -\int p(\hat{\mathbf{s}}) \log p(\hat{\mathbf{s}}) d\hat{\mathbf{s}} = -\frac{1}{2} \log\det\left(\mathbf{\Psi}_\mathbf{X}^* \mathbf{\Psi_X} + \mathbf{r}_\mathbf{X}^* \mathbf{r_X} + \mathbf{\Sigma_s}\right), \tag{4.6}$$

where the notations are suitably modified to ours while several scripts are dropped for brevity. This metric is found in information theories and statistics contexts, just as the Fisher information matrix. Not surprisingly, the expression of differential entropy derived from the Gaussian distribution (Ji et al., 2008, Section IV-A) is equivalent to that evaluates the Fisher information matrix by its matrix determinant Eq. (4.5). It increments its resemblance even more when the "selection" approach is assumed, and the indication matrix $\mathbf{P}$ is introduced to select from the predefined library $\mathbf{\Psi_X}$ as Eq. (4.5).

It must be clarified here that the covariance matrix $\mathbf{\Sigma_s}$ used in Ji et al., 2008 only represents the variance with respect to the sparsified parameters. This approach of choosing the next sampling based on the inferred sparse vector is referred to as "adaptive sampling" in the field of signal processing and information theory, which is sometimes deemed to have limited utility (Arias-Castro et al., 2013). This is certainly because the procedure of posterior inference usually squashes the prior distribution to obtain a sparse representation. Consequently, the constitution of the projection vector for the complement subspace in Ji et al., 2008 resorts to randomly determined values because there is no information on how it observes the complement components by CS measurement. This persists even in the selection approach to append measurements, since the evaluation of the metric ignores the components corresponding to zero variances in the sparse covariance matrix. An intuitive thought is that the covariance matrix of the estimation $\mathbf{\Sigma_s}$ should dictate the uncertainty in the estimation, not the variance of the

sparsified parameters. The estimation in the current consideration explicitly depends on the location of nonzero entries in the parameter vector or the positions of point measurements of compressive measurements. This aspect sheds light on the importance of the new approach which embodies the uncertainty $\mathbf{\Sigma_s}$ in the compressive estimation of the sparse parameter vector. including the determination of a number or positions of nonzero entries. In the next section, a specific method is presented to construct an informative covariance matrix for the selection of measurement positions.

4.2.4 Approximated uncertainty in the parameter estimation

The estimation is established upon the uncertainty due to several assumptions necessary for the estimation. Estimation is settled by using measurements from a predefined projection, and the obtained measurements should embed unpredictable measurement noise and transition. Moreover, the validity against the postulated sparse model should be taken into consideration, since the estimation procedure sets a criterion of compressive estimation using the model. A more reliable compressive measurement scheme can be constructed by minimizing these uncertainties. This study introduces a method to mitigate such uncertainty in the CS estimation by appending the measurements according to a criterion involving an uncertainty measure. The aim of these adaptive procedures is intended as:

- Reducing the uncertainty in the CS estimation by reinforcing the measurements,
- Enhancing representability of CS-DMD by targeting more basis vectors.

As pointed out in the previous section, the evaluation including the unidentified subspace shown in Fig. 4.1 is of special interest in the current study to scrutinize the effective measurement positions for the CS-DMD of fluid flows. These quantities of uncertainty are, however, rarely obtained in many approaches of the CS. An approximate approach implemented in this study involves simple arithmetic to avoid costly marginalization for posterior inference of the distribution.

Representation of uncertainties in the estimation usually requires a probabilistic generative model. In many existing approaches of CS using the probabilistic model, the estimation is

realized according to the maximum a posteriori (MAP) or the maximum likelihood (ML) solution resulting from the predefined model and the measurement. The predefined sparse priors, such as Horse-shoe, Student-t, and Spike-and-Slab distributions, denote the sparse distribution of the parameters. The estimation process manipulates the hyperparameters of the prior through posterior inference and promotes a large fraction of the parameters to take zero values.

Indeed, the hypothesis of the sparsity, that is, the belief that only a small fraction of the parameters should have nonzero entries, focuses our attention only on the nonzero parts of the parameters. However, to give a more reliable estimate, it is natural to be suspicious of the presence of nonzero values in the parameter, such as their positions and the number of elements thereof. Regarding the width of the distribution used for the MAP/ML estimation, there is a chance to evaluate the steadiness of the resulting estimation and reinforce the measurement to make the estimation more solid, as operated in Ji et al., 2008.

One may need to marginalize the source of the uncertainty to evaluate it. The current problem setting has two different sources of uncertainty as;

- Sampling of compressive measurements
- Sparse model assumed for CS inference

Specifically, the information obtained from the measurements depends on projection vectors in a measurement matrix. They are usually obtained by randomly chosen projections in a general CS formulation (Candès et al., 2006). Moreover, it is indispensable to consider the timing of sampling or its duration when dealing with an unsteady, transient phenomenon. It usually requires an increase in the number of measurements in either space or time to mitigate the uncertainty in the measurements. However, this process could pose a combinatorial problem for taking all possible samplings. The other uncertainty may also be concerned, as this study focuses on, with adequate parameters in the sparse model, such as the Lasso model or the assumption of structured sparsity. One may iteratively generate CS estimates to check the deviation in the obtained solution while varying each parameter according to a prior distribution as conducted in Cevher et al., 2009. However, this will be infeasible for several reasons; the computation will easily become intractable, and a reasonable assumption on the prior is also required.

This section presents a method to evaluate estimation uncertainties. A formulation is also proposed to enhance the sensed information by adding spatial sampling after the evaluation of measurement positions, which is based on a contrasting method from general adaptive CS sampling. The quantity of uncertainty should be essentially obtained from the marginalization over the hyperparameters, as mentioned, yet this study approximates the uncertainty due to the compressive measurement for ease of computation and application. The uncertainty is herein defined by the sparsified solutions obtained by compressive estimation with different sparsity parameters;

$$\boldsymbol{\Sigma}_{\mathbf{s}} = \mathrm{Cov}\left[\mathbf{s} - \mu(\mathbf{s}),\, \mathbf{s} - \mu(\mathbf{s})\right] \tag{4.7}$$

$$\approx \mathrm{diag}\left[(\mathbf{s}_{\mathrm{K}'} - \mathbf{s}_{\mathrm{K}}) \circ (\mathbf{s}_{\mathrm{K}'} - \mathbf{s}_{\mathrm{K}})\right], \tag{4.8}$$

where $\circ$ is an elementwise production, and $\bar{\mathbf{s}}$ is its conjugate, while Cov and μ take the covariance and expected value of the argument, respectively. Subscriptions to parameter vectors denote the number of nonzero elements. In other words, this means that a parameter vector with nonzeros entries of a certain number K is regarded as the expectation of inference, while a change in this model parameter is reflected in the variation in a sparse solution. One can conclude the solidity of the solution if Eq. (4.8) is small in some measures. Otherwise, this information can be leveraged for the additional arrangement of measurements. Here, the deviation of the sparsity parameter is assumed to lead to an increase in the number of nonzero elements of the parameter vector $\mathbf{s}_{\mathrm{K}'}$. This will guide appended measurements to appropriately capture the components that may improve the representability of the model.

Algorithmic descriptions are summarized in Algs. 8 and 9 individually, while they are used together in effect. Adaptive optimization based on Alg. 9 is referred to as "Bayesian-A-optimality-criterion-based Greedy algorithm" in this study, and is shortened as BAG hereafter.

Algorithm 8 Approximation of uncertainty

Input: $\mathbf{X}, \mathbf{X}' \in \mathbb{R}^{N\times T}$, $\check{\mathbf{\Psi}}_{\mathbf{X}} \in \mathbb{C}^{N\times \check{N}}$, $\mathcal{I}_q \in \mathbb{N}^q$, $\lambda_1, \lambda_2 \in \mathbb{R}$
Output: Uncertainty in the estimation with respect to the sparsity, $\mathbf{\Sigma_s}$
$\mathbf{Y} = \mathbf{P}(\mathcal{I}_q)\mathbf{X}$, $\mathbf{Y}' = \mathbf{P}(\mathcal{I}_q)\mathbf{X}$
$\mathbf{s}_1 \leftarrow$ (CS-DMD [Alg. 5] with $\mathbf{Y}, \mathbf{Y}'$ and regularizaton parameter $\lambda = \lambda_1$)
$\mathbf{s}_2 \leftarrow$ (CS-DMD [Alg. 5] with $\mathbf{Y}, \mathbf{Y}'$ and regularizaton parameter $\lambda = \lambda_2$)
$\mathbf{\Sigma_s} \leftarrow \mathrm{diag}\left[(\mathbf{s}_1 - \mathbf{s}_2) \circ (\bar{\mathbf{s}}_1 - \bar{\mathbf{s}}_2)\right]$ ($\circ$ is a elementwise production and $\bar{\mathbf{s}}$ is its conjugate)

Algorithm 9 Append of sensing positions through Bayesian formulation [BAG]

Input: $\check{\mathbf{\Psi}}_{\mathbf{X}} \in \mathbb{C}^{N\times \check{N}}$, $p > p' \in \mathbb{N}$, $\mathcal{I}_{p'} \in \mathbb{N}^{p'}$
Output: Indices of chosen p sensing positions $\mathcal{I}_p$
$\mathcal{I}_n \leftarrow \{1, \ldots, n\}$, $\mathcal{I}_0 \leftarrow \emptyset$, $\mathbf{\Sigma_s} \leftarrow$ (Evaluate obtained measurement set $\mathcal{I}_{p'}$ by Alg. 8)
for $q = p' + 1, \ldots, p$ **do**
$i_q \leftarrow \underset{i \in \mathcal{I}_n \setminus \mathcal{I}_{q-1}}{\mathrm{argmax}} \ \mathrm{trace}\left[\left(\mathbf{P}\left(\mathcal{I}_{q-1} \cup i\right)\check{\mathbf{\Psi}}_{\mathbf{X}}\right)^* \left(\mathbf{P}((\mathcal{I}_{q-1} \cup i))\check{\mathbf{\Psi}}_{\mathbf{X}}\right) + \mathbf{\Sigma}_{\mathbf{s}}^{-1}\right]^{-1}$
$\mathcal{I}_q \leftarrow \mathcal{I}_{q-1} \cup i_q$
$\mathbf{\Sigma_s} \leftarrow$ (Evaluate obtained measurement set $\mathcal{I}_q$ by Alg. 8)
end for

4.3 Results and discussions

The results and discussions hereafter concern the reconstruction of the DMD mode by CS-DMD using the data of visualized flow past a cylinder as provided in Section 3.3. The target DMD mode for the sensing optimization is the mode of $St \approx 0.33$ (Fig. 3.8b). It is applicable to other modes in Fig. 3.8, while a result can be different due to a current consideration on $\check{\mathbf{\Psi}}_{\mathbf{X}}$ depending on St as Eq. (3.40).

4.3.1 Obtained sensing positions and reconstruction performance

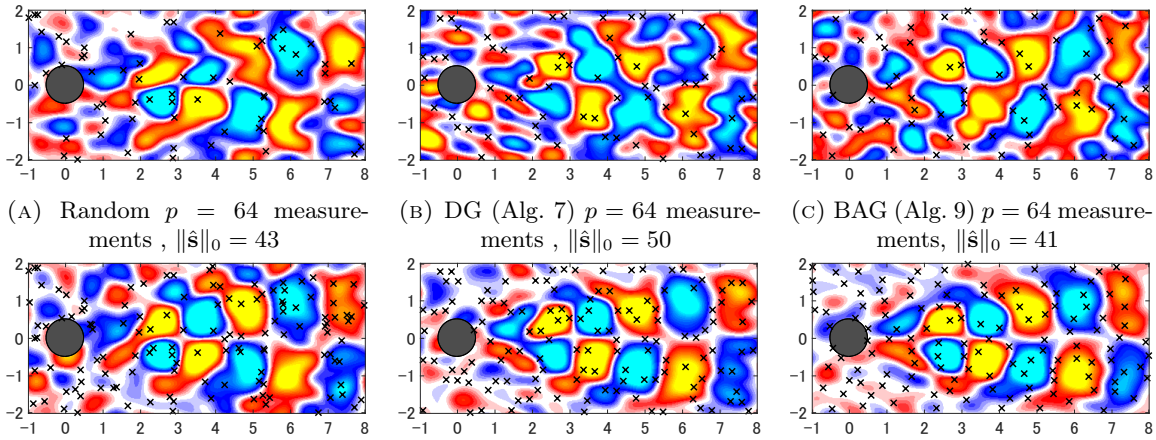
Qualitative comparisons of reconstruction are first placed in this section. The positions for sensing determined by each algorithm, namely, random selection, DG [Alg. 7] and BAG [Alg. 9], are illustrated in the following figures Fig. 4.4. Reconstruction was performed according to the model parameters given in the table below, which are almost the same as those in table 3.2 except for the number of DMD modes.

Model parameter	Value
p	2^6 to 2^{10} with 17 increments
St	0.33
$k'_{\max}$	3,5,7
$s_{\max}$	30,50,100

TABLE 4.1: Parameters for comparative study of measurement optimization targeting the flow past a cylinder

Sensing positions and reconstructed distribution due to selection methods

Comparison of sensing positions are illustrated in Fig. 4.2 with corresponding reconstructed DMD mode in backgrounds. The optimization objective used for each reconstruction is Cℓ_1+Gℓ_1 (w). The numbers of nonzero entries determined are denoted in each caption. The results shown in the left column are due to randomly chosen measurements, while those obtained by DG (Alg. 7) and BAG (Alg. 9) are in the middle and right columns, respectively. For adaptive selection of BAG, the first 16 measurements are obtained by the DG algorithm. Additional measurements are appended by iterating sparse reconstructions according to Alg. 9. The increments of the number of measurements are given as $(16, 19, 22, 26, 32, 38, 45, 53, 64, 76, 90, 107, 128)$, which is according to the powers of two.



(A) Random $p = 64$ measurements , $\|\hat{\mathbf{s}}\|_0 = 43$

(B) DG (Alg. 7) $p = 64$ measurements , $\|\hat{\mathbf{s}}\|_0 = 50$

(C) BAG (Alg. 9) $p = 64$ measurements, $\|\hat{\mathbf{s}}\|_0 = 41$

(D) Random $p = 128$ measurements , $\|\hat{\mathbf{s}}\|_0 = 53$

(E) DG (Alg. 7) $p = 128$ measurements, $\|\hat{\mathbf{s}}\|_0 = 60$

(F) BAG (Alg. 9) $p = 128$ measurements, $\|\hat{\mathbf{s}}\|_0 = 49$

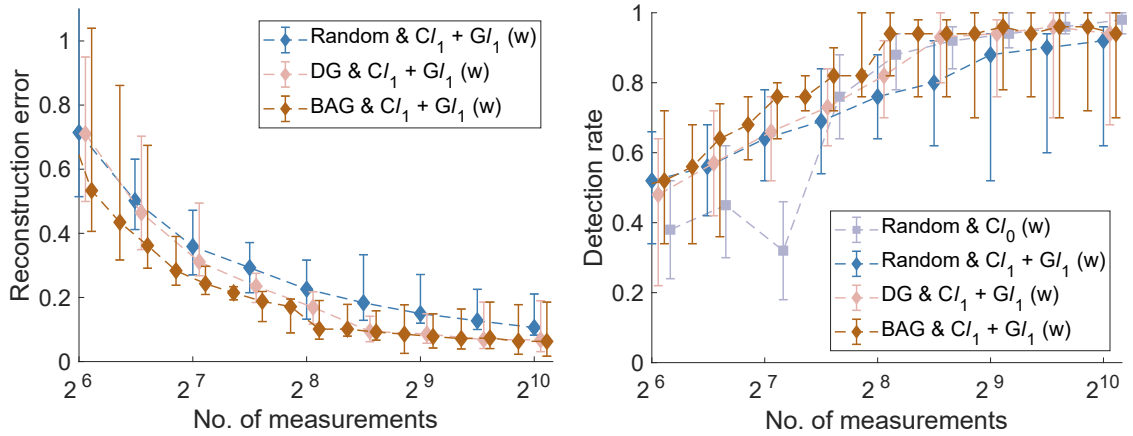
FIGURE 4.2: Comparison of reconstructed distribution of DMD mode $St \approx 0.3$ with positions of measurements determined by each selection method

Measurement points obtained by the proposed selection methods are distributed with larger

relative distances compared with the random selection. Intuitively, the constructed measurements focus on the larger structures invoked by the basis vectors selected as $\check{\mathbf{\Psi}}_{\mathbf{X}}$ according to Eq. (3.40). The differences in the reconstructed distributions are not quite significant in Figures 4.2a to 4.2c, yet the result in Fig. 4.2b is slightly worse than the others. However, the differences shown in Figures 4.2d to 4.2f are more pronounced. The reconstruction in Fig. 4.2d is slightly improved compared with the result of fewer random measurements Fig. 4.2a, which is in contrast to the improvements by two selection methods in Figures 4.2e and 4.2f compared with Figures 4.2b and 4.2c, respectively.

Reconstruction comparison against random distribution

The differences in results due to the various measurements are characterized by the reconstruction error and detection rate of nonzero entries, as shown in the previous chapter, Section 3.3.3. The graphs display the reconstruction results of the 50 most dominant Fourier coefficients for the spatial structure of the DMD mode $St \approx 0.3$ (Fig. 3.8b). These values are calculated using 36 tests of different measurement configurations for each selection method, with median, maximum, and minimum values shown. Therefore, lower values mean better results in the reconstruction error, whereas higher values mean better results for the detection rate. Data points are trimmed except for those of (BAG & Cℓ_1+G$\ell_1(w)$) in Fig. 4.3, for convenience of comparison. As found



(A) Reconstruction error (B) Detection rate

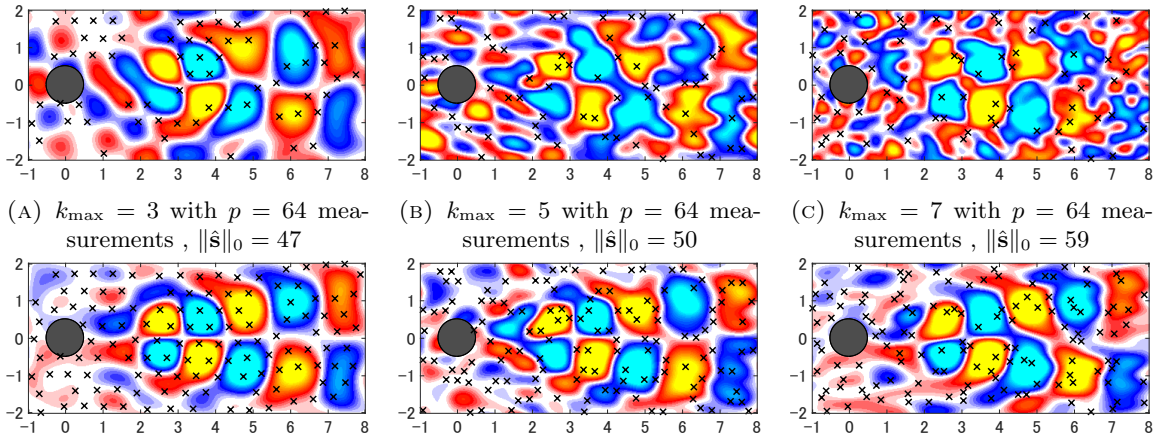
FIGURE 4.3: Error and detection rate of reconstruction of DMD modes for flow past a cylinder with systematic measurement

in Section 3.3.3, the results of estimation based on (Random & Cℓ_1+G$\ell_1(w)$), (Eq. (3.32)) with randomly chosen measurements, are better than those of C$\ell_0(w)$ for a small number of measurements, yet the superiority is lost as there are more measurements used. This is applicable for the detection rate, as also observed previously. The results of (DG & Cℓ_1+Gℓ_1 (w)) are similar to those of (Random & Cℓ_1+Gℓ_1 (w)), except for the slightly lower values for $p > 2^8$. Interestingly, the results of the other selection methods proposed, (BAG & Cℓ_1+Gℓ_1 (w)) differ from this trend. Specifically, the reconstruction errors obtained are the least for $p > 2^6$, although the difference is statistically small. A similar trend is observed in the detection rate, where the superiority of (BAG & Cℓ_1+Gℓ_1 (w)) maintains for $p > 2^8$, which is in contrast to those of the other selection methods, (Random & Cℓ_1+Gℓ_1 (w)).

4.3.2 Variation of sensing positions due to parameters

It should be noted that from Eqs. (4.1) and (4.5), the selection results depend on the reduced representation matrix $\check{\mathbf{\Psi}}_{\mathbf{X}}$, which is defined by Eq. (3.40). Therefore, the dependency on the representation matrix is investigated here by varying the threshold parameter $k_{\max}$ of Eq. (3.40). The reconstructed distributions are illustrated in these figures, where each estimate is made based on Eq. (3.32), thus, they are all denoted as (DG & Cℓ_1+Gℓ_1 (w)). Since these estimations explicitly depend on the representation matrix, each reconstruction is performed based on the different basis matrix as regression materials. The parameter $k_{\max}$ is set to 3, 5, and 7, respectively; therefore, the middle column shows the same figures as those in Fig. 4.2.

The relative distances between measurement points obviously show different trends for each $k_{\max}$. The representation matrix for $k_{\max}$ is composed of the basis vectors of the lowest wavenumbers; therefore, measurement positions are arranged in distance with each other. The distances between the measurements tend to decrease as $k_{\max}$ increases, as expected. These results suggest the possibility of an ad hoc method if one can add measurements adaptively, where the window size gradually increases with the number of measurements.



(A) $k_{\max} = 3$ with $p = 64$ measurements , $\|\hat{\mathbf{s}}\|_0 = 47$

(B) $k_{\max} = 5$ with $p = 64$ measurements , $\|\hat{\mathbf{s}}\|_0 = 50$

(C) $k_{\max} = 7$ with $p = 64$ measurements , $\|\hat{\mathbf{s}}\|_0 = 59$

(D) $k_{\max} = 3$ with $p = 128$ measurements , $\|\hat{\mathbf{s}}\|_0 = 39$

(E) $k_{\max} = 5$ with $p = 128$ measurements , $\|\hat{\mathbf{s}}\|_0 = 60$

(F) $k_{\max} = 7$ with $p = 128$ measurements , $\|\hat{\mathbf{s}}\|_0 = 31$

FIGURE 4.4: Comparison of reconstructed distribution of DMD mode $St \approx 0.3$ for several window sizes of selection method of DG [Alg. 7]

4.4 Conclusions

In this chapter, the measurement optimization schemes for advanced CS-DMD in the previous chapter are introduced. The theory of compressive sensing ensures the recovery of sparse signals based on a plethora of randomly configured measurements. However, the previous results in Chapter 3 hint at a strong bias in the representation coefficients that are responsible for the representation of spatial structures of the DMD modes. It also shows a structure in the Fourier domain as pointed out. The randomness in the measurements is also troublesome in the sense of actual applications, where hardware limitations are clear obstacles for such measurement implementation. Accordingly, systematic selection strategies for the measurement design are proposed in this chapter as summarized in the following three perspectives:

- Selection of measurements based on the existing methods of designs of experiments
- Evaluation of sparse inference of CS using the current set of measurements with approximate marginalization over sparsity parameter, hence, evaluation of the uncertainty in the inference
- Adaptive selection of measurements leveraging the estimated uncertainty for more reliable recovery of sparse vectors

The first method is an extension of existing measurement optimization frameworks. Each measurement position is selected based on the predefined reduced library matrix, which is the reduced Fourier basis vectors introduced in Chapter 3. Contrary to the assumption of random measurements in the CS theories, these measurements enhance the sensitivity with respect to a small portion of subspaces spanned by the reduced library. An efficient implementation is also provided as a greedy formulation. This approach is further extended to the CS problems, where the estimate yields a sparse vector. What has not been well discussed in the optimization of compressive measurement is how to deal with uncertainty in compressive inference. Therefore, the other selection approach is proposed to mitigate the uncertainty in the inference combined with an approximate method for the posterior quantification of the uncertainty.

The comparison of the reconstruction ability of the above methods is performed by using the same dataset as in the previous chapter, that is, the two-dimensional simulated flow past a circular cylinder. The proposed selection methods are concluded to improve the sparse reconstruction of CS-DMD to some extent, by means of both qualitative and quantitative comparison. In particular, the adaptive approach reduced the reconstruction error regarding the group-sparsity for all number of measurement points when the number of nonzero elements was fixed. This is in contrast to the previous results of randomly configured point measurements, where the reconstruction performance has plateaued despite the increase in measurements.

Chapter 5

Concluding remarks

The main results of the preceding studies are summarized in the following statements. Remaining and newly recognized issues conclude the present paper as a future outlook.

5.1 Summary of each chapter

In the series of studies in this dissertation, novel methodologies are presented for the reconstruction of flow state from compressive measurements. In particular, the application for the control of fluid flow is considered, where efficient state estimation using the sparse measurements and the construction of a flow model are critical tasks. Several research themes are presented to focus on these aspects, as summarized below. Quantitative comparisons revealed the superiority of the methods compared in each chapter.

Selection of Measurement Position for State Estimation of Linear Dynamical Systems

Efficient algorithms are developed in Chapter 2 to identify beneficial positions to configure the point measurement for a linear dynamical system. The optimization criterion is the logarithm of the matrix determinant of the observability Gramian of a discrete-tie linear model. After revealing the connection of the criterion to that of the nondynamical systems, several existing methods of measurement optimization for a linear static system were integrated with the current dynamical formulation. Two distinctive methods are proposed to mitigate computations for the

optimization; one that approximates the original binary optimization by relaxing the constraint on the variable. This novel convex relaxation method utilizes the second-order approximation by a Newton method, while the problem dimension is reduced by a randomization technique as previously introduced for static systems. Therefore, it was supposed to be more efficient than the previously presented relaxation method based on semidefinite programming. The other proposed method is an approximate greedy evaluation method, where the first-order approximation of the increment of the criterion caused by an additional measurement is appended to the determined sensing subset.

The proposed gradient greedy method obtains a moderate solution with orders of magnitude speedups compared with the other methods when the dimension of the state variable of a system does not exceed the number of point measurements deployed. Meanwhile, this method is found to be incapable of ensuring the observability for larger dimensions of the state variable. The convex relaxation methods, including the SDP-based and the approximate convex relaxation methods, achieved better solutions than the greedy methods in synthetic data, while the real-world example discloses their unstable optimization results, especially for a large-scale system. In such less observable situations, the pure greedy selection is supposed to be the most reliable choice in terms of the optimization of the matrix determinant of the observability Gramian.

Construction of Linear Flow Model from Compressive Measurement

This chapter features several improvements to the existing methods for constructing a reduced-order linear dynamical model. This is based on a completely different postulate than that of the previous chapter, since there is no training data for field measurements. The modification for efficient inference of the model is the protagonist in this chapter. Several intuitive yet reasonable observations are provided with respect to an existing data-driven method that incorporates the use of compressive sensing techniques for model construction by Dynamic Mode Decomposition (DMD). This method, called CS-DMD, has strong generality; therefore, the modified approach of CS-DMD takes advantage of the notion of structured sparsity to recognize a more physical distribution hidden in the partially observed data.

The introduced in this study is the selectively retained Fourier basis vectors and the structural CS inference for the recovery of spatial distribution from sparse measurements. The former is used for compressive representation to obtain a more reliable and physically relevant reconstruction. This reduction compresses the dimension of the inference to a great extent as a side effect. The latter structural inference assists in reconstructing the presumed characteristics of the target. Several basic structures of fluid flows are listed, and the demonstration of the structured inference that promotes a clustered distribution of representation coefficients in the Fourier domain. This is meant to reflect the smooth advection that transports flow structures at a certain scale. These considerations are handled straightforwardly in algorithms. First, a library matrix for the compressive representation was thinned in advance according to the selection rule using a threshold that depends on the temporal oscillation frequency of the target DMD mode. The second modification is incorporated into CS-DMD by invoking the group-sparsity used as a penalty term in the sparse inference of CS-DMD. Each parameter in the Fourier domain is put together with its nearest adjacent to compose overlapping groups. This is a relaxed form of the combinatorial problem that controls every form of cluster. This optimization problem is solved by the Alternating Direction Method of Multipliers (ADMM).

The results obtained from the above processes made a great improvement in the reconstruction by CS-DMD. The CFD data of two-dimensional flow past a cylinder are used for the demonstration, where the periodic motion represented by each DMD mode is quite characteristic. The reconstructed distributions of the DMD modes are significantly improved for a smaller number of measurements, where the original implementation suffers from the ill-condition of the inference. It became even more obvious when comparing the mean squared error of the reconstruction and the detection rate of the appropriate nonzero entries.

Optimized Measurement Position for Compressive Flow Modeling

Finally, this chapter integrates the optimization of measurement into CS-DMD to further compress the size of measurements. It is in contrast to the CS theory that encourages randomly arranged measurements, yet the results in the previous chapter suggest that a small fraction

of basis vectors are frequently used. Therefore, the redundancy in the point measurements is reduced as much as possible by arranging the measurements to maximize the sensitivity with respect to the selected basis vectors. This optimization is extended from previous studies of measurement optimization based on the linear least-squares estimation; thus, it does not take into account the sparsity necessary for the inference in CS.

The other method implemented in this study is a sort of Bayesian optimization, which utilizes the posterior evaluation of the parameter covariance matrix. The key insight of this approach is to elaborate on the uncertainty in the sparse solution of CS, where the previous approaches of adaptive measurement omit the ignored zero-entries. A simple heuristic for the evaluation of uncertainty is also tested and involved in the measurement optimization. Iterative append of measurements based on the reduction of uncertainty is expected to lead to a more reliable estimation of CS-DMD.

The results show that the mean square errors of the reconstruction are reduced, although not spectacularly. The comparison is performed with the same dataset as used in the previous chapter, that is, the two-dimensional flow around a cylinder by CFD. The results using a smaller number of measurements are improved compared with the previous general implementation of CS-DMD. By this result, a systematic strategy of measurement design is valid for the application of CS-DMD.

5.2 Future outlook

This study showed that using sparse measurements is a meaningful procedure for building a linear dynamical model. However, at the same time, there are some issues that have not been fully resolved or fully investigated. After introducing these aspects, this section shows how the author thinks this series of research should develop.

What is not fully addressed in this thesis

First of all, the robustness to real-world applications should be investigated in forthcoming researches. The results and subsequent discussions in this study did not fully resolve the actual

applications, where the model obtained from data analysis may be corrupted by some "noise". In Chapter 2, the optimization results showed a clear change depending on whether the model is created synthetically or from real-world measurements. The reason for this has not been investigated in this study. In the following reports considering CS-DMD, the demonstrations were based on the ideal data, which are known to be well approximated by linear models such as DMD. The data are not corrupted by measurement noise and can generate trustworthy reduced-order models. The assumed dynamical models are of reduced-order representations, therefore, interactions from these less dominant materials should be considered as (Otto et al., 2022; Yamada et al., 2021).

The dependence on the model assumption comes in second place. It is assumed in Chapter 2 that the algorithms presented are applied to low-dimensional linear models obtained in advance. The state-space models therein were constructed by POD and the linear least squares. The limitation of the presented formulation is not considered completely, such as the conditions of the applied model to ensure the observability. In Chapters 3 and 4, there should be a dependency on how the DMD models are obtained from sensor measurements. The data for the demonstrations are considerably concise, as mentioned earlier, where one can obtain eigenvalues from only a few dozen measurements that are the same as those obtained with complete data. However, this is not the case for many data obtained from experiments. Considerable attention should be paid to the type of DMD used, as well as to the uncertainty in sparse measurements. Furthermore, the assumptions in a sparse model should be robust since the influence due to a sparsity parameter was also confirmed. A novel method of (semi-) automatic selection is desirable for determining these hyperparameters.

It is also noted that the development of efficient algorithms for CS-DMD has not been presented. In particular, the implementation of ADMM is based on a naive formulation of overlapping group Lasso instead of some developed forms such as L. Yuan et al., 2013. The development of efficient algorithms is considered an important topic to be addressed.

How this research should develop

The results of the presented paper suggest several directions for future research. Since there are too diverse thoughts evoked, two main potential perspectives are provided here.

The most rewarding topic is to extend the methods to the reconstruction of higher-dimensional flow fields. The current configuration is completely based on the reconstruction of a two-dimensional field of one component (often shortened as 2D1C). Therefore, it is an undoubted demand to extend the presented methodology to 3D2C (or more) fields for actual applications of flow measurements.

Another insight that should be explored is the uncertainty with respect to the temporal sampling of CS-DMD. The presented methods focused on constructing a linear model from limited spatial measurements of finite duration and discrete sampling. It is meaningful to consider the uncertainty with respect to time, since the proposed method relies on such measurements to illustrate the nonlinear nature of phenomena.

Appendix A

Implementations focusing on program

A.1 CVX implementation for Eq. (P2)

Thanks to the parser CVX CVX Research, 2012, August; M. Grant and Boyd, 2008, the SDP formulation defined by Eq. (P2) is easily implemented in MATLAB and solved by use of solver Mosek in T. Summers and Shames, 2016 and in our case. The actual implementation is shown below, and MATLAB codes are provided in Yamada, 2023.

```
[r,~] = size(A);
[n,~] = size(C);
cvx_solver('Mosek')
cvx_begin sdp
   variable z(n) nonnegative
   variable X(r,r) symmetric semidefinite
   maximize( det_rootn(X) )
   subject to
       % matrix inequality
       A'*X*A-X+C'*(repmat(z,1,r).*C) >= zeros(r,r);
       % bounds selection variables
       z <= 1;
       sum(z) == p;
cvx_end
```

References

Ahmed, A., del Rio-Chanona, E. A., & Mercangöz, M. (2023). Linearizing nonlinear dynamics using deep learning. *Computers &amp Chemical Engineering*, *170*, 108104. https://doi.org/10.1016/j.compchemeng.2022.108104 (cit. on p. 5).

Alonso, A. A., Frouzakis, C. E., & Kevrekidis, I. G. (2004). Optimal sensor placement for state reconstruction of distributed process systems. *AIChE Journal*, *50*(7), 1438–1452. https://doi.org/https://doi.org/10.1002/aic.10121 (cit. on p. 14).

Antonelli, G., Chiaverini, S., & Lillo, P. D. (2022). On data-driven identification: Is automatically discovering equations of motion from data a chimera? *Nonlinear Dynamics*, *111*(7), 6487–6498. https://doi.org/10.1007/s11071-022-08192-x (cit. on p. 3).

Arias-Castro, E., Candes, E. J., & Davenport, M. A. (2013). On the fundamental limits of adaptive sensing. *IEEE Transactions on Information Theory*, *59*(1), 472–481. https://doi.org/10.1109/tit.2012.2215837 (cit. on p. 107).

Asztalos, K. J., Almashjary, A., & Dawson, S. T. M. (2023). Galerkin spectral estimation of vortex-dominated wake flows. *Theoretical and Computational Fluid Dynamics*. https://doi.org/10.1007/s00162-023-00670-1 (cit. on p. 2).

Baddoo, P. J., Herrmann, B., McKeon, B. J., Kutz, J. N., & Brunton, S. L. (2023). Physics-informed dynamic mode decomposition. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *479*(2271). https://doi.org/10.1098/rspa.2022.0576 (cit. on pp. 3, 13, 43).

Bagheri, S. (2013). Koopman-mode decomposition of the cylinder wake. *Journal of Fluid Mechanics*, *726*, 596–623. https://doi.org/10.1017/jfm.2013.249 (cit. on pp. 63, 70).

Bai, Z., Kaiser, E., Proctor, J. L., Kutz, J. N., & Brunton, S. L. (2020). Dynamic mode decomposition for compressive system identification. *AIAA Journal*, *58*(2), 561–574. https://doi.org/10.2514/1.j057870 (cit. on pp. 5, 7, 42, 46, 55).

Baraniuk, R. (2007). Compressive sensing [lecture notes]. *IEEE Signal Processing Magazine*, *24*(4), 118–121. https://doi.org/10.1109/msp.2007.4286571 (cit. on p. 48).

Baraniuk, R., Davenport, M., DeVore, R., & Wakin, M. (2008). A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, *28*(3), 253–263. https://doi.org/10.1007/s00365-007-9003-x (cit. on pp. 95, 96, 101).

Baraniuk, R. G., Cevher, V., Duarte, M. F., & Hegde, C. (2010). Model-based compressive sensing. *IEEE Transactions on Information Theory*, *56*(4), 1982–2001. https://doi.org/10.1109/tit.2010.2040894 (cit. on pp. 6, 44, 58–60, 64, 65).

Barraud, A. (1977). A numerical algorithm to solve $AXA - X = Q$. *IEEE Transactions on Automatic Control*, *22*(5), 883–885. https://doi.org/10.1109/tac.1977.1101604 (cit. on p. 30).

Bates, R., Buck, R., Riccomagno, E., & Wynn, H. (1996). Experimental design and observation for large systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 77–94. https://doi.org/10.1111/j.2517-6161.1996.tb02068.x (cit. on pp. 14, 102).

Benner, P., Mehrmann, V., Sima, V., Van Huffel, S., & Varga, A. (1999). Slicot—a subroutine library in systems and control theory. In *Applied and computational control, signals, and*

circuits (pp. 499–539). Birkhäuser Boston. https://doi.org/10.1007/978-1-4612-0571-5_10 (cit. on p. 30).

Berinde, R., Gilbert, A. C., Indyk, P., Karloff, H., & Strauss, M. J. (2008). Combining geometry and combinatorics: A unified approach to sparse signal recovery. *2008 46th Annual Allerton Conference on Communication, Control, and Computing*. https://doi.org/10.1109/allerton.2008.4797639 (cit. on pp. 51, 101).

Berkooz, G., Holmes, P., & Lumley, J. L. (1993). The proper orthogonal decomposition in the analysis of turbulent flows. *Annual review of fluid mechanics*, *25*(1), 539–575. https://doi.org/10.1146/annurev.fl.25.010193.002543 (cit. on pp. 2, 13, 35).

Bhattacharjee, D., Klose, B., Jacobs, G. B., & Hemati, M. S. (2020). Data-driven selection of actuators for optimal control of airfoil separation. *Theoretical and Computational Fluid Dynamics*, *34*(4), 557–575. https://doi.org/10.1007/s00162-020-00526-y (cit. on p. 5).

Boyd, S. (2010). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, *3*(1), 1–122. https://doi.org/10.1561/2200000016 (cit. on p. 65).

Boyd, S., Boyd, S. P., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press. (Cit. on pp. 23, 29).

Boyer, C., Bigot, J., & Weiss, P. (2019). Compressed sensing with structured sparsity and structured acquisition. *Applied and Computational Harmonic Analysis*, *46*(2), 312–350. https://doi.org/10.1016/j.acha.2017.05.005 (cit. on pp. 96, 101).

Brunton, S. L., Brunton, B. W., Proctor, J. L., & Kutz, J. N. (2016a). Koopman invariant subspaces and finite linear representations of nonlinear dynamical systems for control (H. A. Kestler, Ed.). *PLOS ONE*, *11*(2), e0150171. https://doi.org/10.1371/journal.pone.0150171 (cit. on p. 40).

Brunton, S. L., Budišić, M., Kaiser, E., & Kutz, J. N. (2022). Modern koopman theory for dynamical systems. *SIAM Review*, *64*(2), 229–340. https://doi.org/10.1137/21m1401243 (cit. on p. 40).

Brunton, S. L., & Kutz, J. N. (2019, January). *Data-driven science and engineering*. Cambridge University Press. https://doi.org/10.1017/9781108380690 (cit. on pp. 3, 5, 13, 35).

Brunton, S. L., Kutz, J. N., Manohar, K., Aravkin, A. Y., Morgansen, K., Klemisch, J., Goebel, N., Buttrick, J., Poskin, J., Blom-Schieber, A. W., Hogan, T., & McDonald, D. (2021). Data-driven aerospace engineering: Reframing the industry with machine learning. *AIAA Journal*, 1–26. https://doi.org/10.2514/1.j060131 (cit. on p. 5).

Brunton, S. L., Noack, B. R., & Koumoutsakos, P. (2020). Machine learning for fluid mechanics. *Annual Review of Fluid Mechanics*, *52*(1), 477–508. https://doi.org/10.1146/annurev-fluid-010719-060214 (cit. on p. 2).

Brunton, S. L., Proctor, J. L., & Kutz, J. N. (2016b). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the national academy of sciences*, *113*(15), 3932–3937. https://doi.org/10.1073/pnas.1517384113 (cit. on p. 3).

Brunton, S. L., Proctor, J. L., Tu, J. H., & Kutz, J. N. (2015). Compressed sensing and dynamic mode decomposition. *Journal of Computational Dynamics*, *2*(2), 165. https://doi.org/10.3934/jcd.2015002 (cit. on pp. 7, 42, 44, 46, 53–55, 57, 90).

Brunton, S. L., Tu, J. H., Bright, I., & Kutz, J. N. (2014). Compressive sensing and low-rank libraries for classification of bifurcation regimes in nonlinear dynamical systems. *SIAM Journal on Applied Dynamical Systems*, *13*(4), 1716–1732. https://doi.org/10.1137/130949282 (cit. on pp. 7, 8, 42).

Cai, S., Wang, Z., Fuest, F., Jeon, Y. J., Gray, C., & Karniadakis, G. E. (2021). Flow over an espresso cup: Inferring 3-D velocity and pressure fields from tomographic background oriented schlieren via physics-informed neural networks. *Journal of Fluid Mechanics*, *915*. https://doi.org/10.1017/jfm.2021.135 (cit. on pp. 3, 43).

Candès, E. J. (2008). The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique*, *346*(9–10), 589–592. https://doi.org/10.1016/j.crma.2008.03.014 (cit. on p. 51).

Candès, E. J., Romberg, J. K., & Tao, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, *59*(8), 1207–1223. https://doi.org/10.1002/cpa.20124 (cit. on pp. 6, 51, 52, 96, 109).

Carter, D. W., Voogt, F. D., Soares, R., & Ganapathisubramani, B. (2021). Data-driven sparse reconstruction of flow over a stalled aerofoil using experimental data. *Data-Centric Engineering*, *2*. https://doi.org/10.1017/dce.2021.5 (cit. on pp. 8, 14).

Carvalho, C. M., Polson, N. G., & Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, *97*(2), 465–480. https://doi.org/10.1093/biomet/asq017 (cit. on p. 64).

Castro-Triguero, R., Murugan, S., Gallego, R., & Friswell, M. I. (2013). Robustness of optimal sensor placement under parametric uncertainty. *Mechanical Systems and Signal Processing*, *41*(1-2), 268–287. https://doi.org/10.1016/j.ymssp.2013.06.022 (cit. on p. 14).

Cenedese, M., Axås, J., Bäuerlein, B., Avila, K., & Haller, G. (2022). Data-driven modeling and prediction of non-linearizable dynamics via spectral submanifolds. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-28518-y (cit. on pp. 3, 5).

Cevher, V., Duarte, M., Hegde, C., & Baraniuk, R. (2008). Sparse signal recovery using markov random fields. In D. Koller, D. Schuurmans, Y. Bengio, & L. Bottou (Eds.), *Advances in neural information processing systems* (Vol. 21). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2008/file/287e03db1d99e0ec2edb90d079e142f3-Paper.pdf (cit. on p. 64).

Cevher, V., Indyk, P., Hegde, C., & Baraniuk, R. (2009). Recovery of clustered sparse signals from compressive measurements. *International conference on Sampling Theory and Applications(SAMPTA)*. http://infoscience.epfl.ch/record/151473,%20http://dsp.rice.edu/software/ (cit. on pp. 60, 109).

Chen, C., & Huang, J. (2012). Compressive sensing MRI with wavelet tree sparsity. In F. Pereira, C. Burges, L. Bottou, & K. Weinberger (Eds.), *Advances in neural information processing systems* (Vol. 25). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2012/file/65658fde58ab3c2b6e5132a39fae7cb9-Paper.pdf (cit. on p. 64).

Chen, K. K., Tu, J. H., & Rowley, C. W. (2012). Variants of dynamic mode decomposition: Boundary condition, Koopman, and Fourier analyses. *Journal of Nonlinear Science*, *22*(6), 887–915. https://doi.org/10.1007/s00332-012-9130-9 (cit. on p. 47).

Chen, Y., Wang, S., Luo, Y., & Liu, H. (2023). Measurement matrix optimization based on target prior information for radar imaging. *IEEE Sensors Journal*, *23*(9), 9808–9819. https://doi.org/10.1109/jsen.2023.3263591 (cit. on p. 100).

Chepuri, S. P., & Leus, G. (2015). Sparse sensing for distributed detection. *IEEE Transactions on Signal Processing*, *64*(6), 1446–1460 (cit. on p. 15).

Clark, E., Askham, T., Brunton, S. L., & Kutz, J. N. (2019). Greedy sensor placement with cost constraints. *IEEE Sensors Journal*, *19*(7), 2642–2656. https://doi.org/10.1109/jsen.2018.2887044 (cit. on p. 15).

Clark, E., Kutz, J. N., & Brunton, S. L. (2020). Sensor selection with cost constraints for dynamically relevant bases. *IEEE Sensors Journal*, *20*(19), 11674–11687. https://doi.org/10.1109/jsen.2020.2997298 (cit. on p. 16).

Colbrook, M. J., Ayton, L. J., & Szőke, M. (2023). Residual dynamic mode decomposition: Robust and verified Koopmanism. *Journal of Fluid Mechanics*, *955*. https://doi.org/10.1017/jfm.2022.1052 (cit. on p. 40).

CVX Research, I. (2012, August). Cvx: Matlab software for disciplined convex programming, version 2.0. (Cit. on pp. 29, 123).

da Silva, G. R. G., Bazanella, A. S., Lorenzini, C., & Campestrini, L. (2019). Data-driven LQR control design. *IEEE Control Systems Letters*, *3*(1), 180–185. https://doi.org/10.1109/lcsys.2018.2868183 (cit. on p. 5).

Dawson, S. T. M., Hemati, M. S., Williams, M. O., & Rowley, C. W. (2016). Characterizing and correcting for the effect of sensor noise in the dynamic mode decomposition. *Experiments in Fluids*, *57*(3). https://doi.org/10.1007/s00348-016-2127-7 (cit. on pp. 47, 54).

DeVore, R. A. (2007). Deterministic constructions of compressed sensing matrices. *Journal of Complexity*, *23*(4-6), 918–925. https://doi.org/10.1016/j.jco.2007.04.002 (cit. on pp. 7, 52, 101).

DeVries, L., Majumdar, S. J., & Paley, D. A. (2012). Observability-based optimization of coordinated sampling trajectories for recursive estimation of a strong, spatially varying flowfield. *Journal of Intelligent & Robotic Systems*, *70*(1–4), 527–544. https://doi.org/10.1007/s10846-012-9718-1 (cit. on pp. 15, 102).

DeVries, L., & Paley, D. A. (2013). Observability-based optimization for flow sensing and control of an underwater vehicle in a uniform flowfield. *2013 American Control Conference*, 1386–1391. https://doi.org/10.1109/ACC.2013.6580030 (cit. on p. 14).

Dhingra, N. K., Jovanović, M. R., & Luo, Z.-Q. (2014). An ADMM algorithm for optimal sensor and actuator selection. *53rd IEEE Conference on Decision and Control*, 4039–4044. https://doi.org/10.1109/CDC.2014.7040017 (cit. on pp. 15, 22).

Doğançay, K., & Hmam, H. (2009). On optimal sensor placement for time-difference-of-arrival localization utilizing uncertainty minimization. *2009 17th European Signal Processing Conference*, 1136–1140 (cit. on p. 14).

Drmač, Z., & Saibaba, A. K. (2018). The discrete empirical interpolation method: Canonical structure and formulation in weighted inner product spaces. *SIAM Journal on Matrix Analysis and Applications*, *39*(3), 1152–1180. https://doi.org/10.1137/17m1129635 (cit. on p. 20).

Duarte, M. F., & Eldar, Y. C. (2011). Structured compressed sensing: From theory to applications. *IEEE Transactions on Signal Processing*, *59*(9), 4053–4085. https://doi.org/10.1109/tsp.2011.2161982 (cit. on pp. 7, 96, 100, 101).

Elad, M. (2010). *Sparse and redundant representations*. Springer New York. https://doi.org/10.1007/978-1-4419-7011-4 (cit. on pp. 7, 48, 51).

Erichson, N. B., & Donovan, C. (2016). Randomized low-rank dynamic mode decomposition for motion detection. *Computer Vision and Image Understanding*, *146*, 40–50. https://doi.org/10.1016/j.cviu.2016.02.005 (cit. on p. 42).

Erichson, N. B., Mathelin, L., Kutz, J. N., & Brunton, S. L. (2019). Randomized dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems*, *18*(4), 1867–1891. https://doi.org/10.1137/18m1215013 (cit. on pp. 7, 42).

Feige, U., Mirrokni, V. S., & Vondrák, J. (2011). Maximizing non-monotone submodular functions. *SIAM Journal on Computing*, *40*(4), 1133–1153. https://doi.org/https://doi.org/10.1137/090779346 (cit. on p. 26).

Fisher, R. A. (1925). Theory of statistical estimation. *Mathematical Proceedings of the Cambridge Philosophical Society*, *22*(5), 700–725. https://doi.org/10.1017/s0305004100009580 (cit. on p. 14).

Fowler, J. (2009). Compressive-projection principal component analysis. *IEEE Transactions on Image Processing*, *18*(10), 2230–2242. https://doi.org/10.1109/tip.2009.2025089 (cit. on p. 7).

Fujii, K., Inaba, Y., & Kawahara, Y. (2017). Koopman spectral kernels for comparing complex dynamics: Application to multiagent sport plays. In Y. Altun, K. Das, T. Mielikäinen, D. Malerba, J. Stefanowski, J. Read, M. Žitnik, M. Ceci, & S. Džeroski (Eds.), *Machine learning and knowledge discovery in databases* (pp. 127–139). Springer International Publishing. https://doi.org/10.1007/978-3-319-71273-4_11 (cit. on p. 41).

Fujii, K. (2014). High-performance computing-based exploration of flow control with micro devices. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *372*(2022), 20130326. https://doi.org/10.1098/rsta.2013.0326 (cit. on p. 4).

Fukami, K., Fukagata, K., & Taira, K. (2023). Super-resolution analysis via machine learning: A survey for fluid flows. *Theoretical and Computational Fluid Dynamics*, *37*(4), 421–444. https://doi.org/10.1007/s00162-023-00663-0 (cit. on p. 7).

Fukami, K., Maulik, R., Ramachandra, N., Fukagata, K., & Taira, K. (2021). Global field reconstruction from sparse sensors with voronoi tessellation-assisted deep learning. *Nature Machine Intelligence*, *3*(11), 945–951. https://doi.org/10.1038/s42256-021-00402-2 (cit. on p. 14).

Gilbert, A., & Indyk, P. (2010). Sparse recovery using sparse matrices. *Proceedings of the IEEE*, *98*(6), 937–947. https://doi.org/10.1109/jproc.2010.2045092 (cit. on pp. 51, 101).

Glauser, A. P. L. C. L. D. M. (Ed.). (2017). *Whither turbulence and big data in the 21st century?* Springer International Publishing. https://doi.org/10.1007/978-3-319-41217-7 (cit. on p. 2).

Gower, R., Koralev, D., Lieder, F., & Richtarik, P. (2019). Rsn: Randomized subspace newton. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in neural information processing systems 32* (pp. 616–625). Curran Associates, Inc. http://papers.nips.cc/paper/8351-rsn-randomized-subspace-newton.pdf (cit. on p. 24).

Grant, I. (1997). Particle image velocimetry: A review. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, *211*(1), 55–76. https://doi.org/10.1243/0954406971521665 (cit. on p. 6).

Grant, M., & Boyd, S. (2008). Graph implementations for nonsmooth convex programs [http://stanford.edu/~boyd/graph_dcp.html]. In V. Blondel, S. Boyd, & H. Kimura (Eds.), *Recent advances in learning and control* (pp. 95–110). Springer-Verlag Limited. (Cit. on pp. 29, 123).

Guéniat, F., Mathelin, L., & Pastur, L. R. (2015). A dynamic mode decomposition approach for large and arbitrarily sampled systems. *Physics of Fluids*, *27*(2), 025113. https://doi.org/10.1063/1.4908073 (cit. on pp. 7, 42).

Hammarling, S. J. (1982). Numerical solution of the stable, non-negative definite lyapunov equation. *IMA Journal of Numerical Analysis*, *2*(3), 303–323. https://doi.org/10.1093/imanum/2.3.303 (cit. on p. 30).

Hashemi, A., Ghasemi, M., Vikalo, H., & Topcu, U. (2020). Randomized greedy sensor selection: Leveraging weak submodularity. *IEEE Transactions on Automatic Control*, *66*(1), 199–212 (cit. on pp. 15, 26).

Heide, A. L., Asztalos, K. J., Dawson, S. T., & Hemati, M. (2022). A low-order nonlinear model of a stalled airfoil from data: Exploiting sparse regression with physical constraints. *AIAA AVIATION 2022 Forum*. https://doi.org/10.2514/6.2022-3976 (cit. on p. 3).

Hemati, M. S., Williams, M. O., & Rowley, C. W. (2014). Dynamic mode decomposition for large and streaming datasets. *Physics of Fluids*, *26*(11). https://doi.org/10.1063/1.4901016 (cit. on p. 42).

Herrmann, B., Baddoo, P. J., Dawson, S. T., Semaan, R., Brunton, S. L., & McKeon, B. J. (2023). Interpolatory input and output projections for flow control. *Journal of Fluid Mechanics*, *971*. https://doi.org/10.1017/jfm.2023.680 (cit. on p. 8).

Hoseyni, S. M., Maio, F. D., & Zio, E. (2022). Subset simulation for optimal sensors positioning based on value of information. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, *0*(0), 1748006X2211184. https://doi.org/10.1177/1748006x221118432 (cit. on p. 14).

Inoba, R., Uchida, K., Iwasaki, Y., Nagata, T., Ozawa, Y., Saito, Y., Nonomura, T., & Asai, K. (2022). Optimization of sparse sensor placement for estimation of wind direction and surface pressure distribution using time-averaged pressure-sensitive paint data on automobile model. *Journal of Wind Engineering and Industrial Aerodynamics*, *227*, 105043. https://doi.org/10.1016/j.jweia.2022.105043 (cit. on p. 14).

Inoue, T., Ikami, T., Egami, Y., Nagai, H., Naganuma, Y., Kimura, K., & Matsuda, Y. (2023). Data-driven optimal sensor placement for high-dimensional system using annealing machine. *Mechanical Systems and Signal Processing*, *188*, 109957. https://doi.org/10.1016/j.ymssp.2022.109957 (cit. on p. 14).

Iwasaki, Y., Nonomura, T., Nakai, K., Nagata, T., Saito, Y., & Asai, K. (2022). Evaluation of optimization algorithms and noise robustness of sparsity-promoting dynamic mode decomposition. *IEEE Access*, *10*, 80748–80763. https://doi.org/10.1109/access.2022.3193157 (cit. on p. 13).

Jawaid, S. T., & Smith, S. L. (2015). Submodularity and greedy algorithms in sensor scheduling for linear dynamical systems. *Automatica*, *61*, 282–288. https://doi.org/10.1016/j.automatica.2015.08.022 (cit. on pp. 98, 105).

Ji, S., Xue, Y., & Carin, L. (2008). Bayesian compressive sensing. *IEEE Transactions on Signal Processing*, *56*(6), 2346–2356. https://doi.org/10.1109/TSP.2007.914345 (cit. on pp. 98, 106, 107, 109).

Jiang, X., Li, N., Guo, Y., Liu, J., & Wang, C. (2022). Sensing matrix optimization for multi-target localization using compressed sensing in wireless sensor network. *China Communications*, *19*(3), 230–244. https://doi.org/10.23919/jcc.2022.03.017 (cit. on pp. 100, 106).

Joneidi, M., Zaeemzadeh, A., Shahrasbi, B., Qi, G.-J., & Rahnavard, N. (2020). E-optimal sensor selection for compressive sensing-based purposes. *IEEE Transactions on Big Data*, *6*(1), 51–65. https://doi.org/10.1109/tbdata.2018.2868120 (cit. on pp. 8, 101).

Joshi, S., & Boyd, S. (2009). Sensor selection via convex optimization. *IEEE Transactions on Signal Processing*, *57*(2), 451–462. https://doi.org/10.1109/TSP.2008.2007095 (cit. on pp. 7, 14, 15, 21, 23, 102).

Jovanović, M. R., Schmid, P. J., & Nichols, J. W. (2014). Sparsity-promoting dynamic mode decomposition. *Physics of Fluids*, *26*(2), 1–22. https://doi.org/10.1063/1.4863670 (cit. on pp. 13, 47).

Kaiser, E., Kutz, J. N., & Brunton, S. L. (2021). Data-driven discovery of Koopman eigenfunctions for control. *Machine Learning: Science and Technology*, *2*(3), 035023. https://doi.org/10.1088/2632-2153/abf0f5 (cit. on p. 5).

Kaiser, E., Morzyński, M., Daviller, G., Kutz, J. N., Brunton, B. W., & Brunton, S. L. (2018). Sparsity enabled cluster reduced-order models for control. *Journal of Computational Physics*, *352*, 388–409. https://doi.org/10.1016/j.jcp.2017.09.057 (cit. on p. 5).

Kanda, N., Abe, C., Goto, S., Yamada, K., Nakai, K., Saito, Y., Asai, K., & Nonomura, T. (2022). Proof-of-concept study of sparse processing particle image velocimetry for real time flow observation. *Experiments in Fluids*, *63*(9). https://doi.org/10.1007/s00348-022-03471-0 (cit. on pp. 14, 97).

Kanda, N., Nakai, K., Saito, Y., Nonomura, T., & Asai, K. (2021). Feasibility study on real-time observation of flow velocity field using sparse processing particle image velocimetry. *Transactions of the Japan Society for Aeronautical and Space Sciences*, *64*(4), 242–245. https://doi.org/10.2322/tjsass.64.242 (cit. on p. 14).

Kaneko, S., Ozawa, Y., Nakai, K., Saito, Y., Nonomura, T., Asai, K., & Ura, H. (2021). Data-driven sparse sampling for reconstruction of acoustic-wave characteristics used in aeroacoustic beamforming. *Applied Sciences*, *11*(9), 4216. https://doi.org/10.3390/app11094216 (cit. on p. 14).

Kastsiukevich, D., & Dmitruk, N. (2020). Data-driven optimal control of linear time-invariant systems. *IFAC-PapersOnLine*, *53*(2), 7191–7196. https://doi.org/10.1016/j.ifacol.2020.12.549 (cit. on p. 5).

Kay, S. M. (1993). *Fundamentals of statistical signal processing: Estimation theory*. Prentice-Hall, Inc. (Cit. on p. 14).

Kilic, B., Gungor, A., Kalfa, M., & Arkan, O. (2022). Adaptive measurement matrix design in direction of arrival estimation. *IEEE Transactions on Signal Processing*, *70*, 4742–4756. https://doi.org/10.1109/tsp.2022.3209880 (cit. on pp. 98, 106).

Kincaid, R. K., & Padula, S. L. (2002). D-optimal designs for sensor and actuator locations. *Computers & Operations Research*, *29*(6), 701–713. https://doi.org/https://doi.org/10.1016/S0305-0548(01)00048-X (cit. on p. 102).

Kitagawa, G. (1977). An algorithm for solving the matrix equation $X = FXF^{\top} + S$. *International Journal of Control*, *25*(5), 745–753. https://doi.org/10.1080/00207177708922266 (cit. on p. 31).

Korda, M., & Mezić, I. (2018). Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control. *Automatica*, *93*, 149–160. https://doi.org/10.1016/j.automatica.2018.03.046 (cit. on pp. 5, 40, 41).

Kraft, T., Mignan, A., & Giardini, D. (2013, July). *Optimization of a large-scale microseismic monitoring network in northern Switzerland* (tech. rep. No. 1). Oxford University Press (OUP). https://doi.org/10.1093/gji/ggt225 (cit. on p. 14).

Kramer, B., Grover, P., Boufounos, P., Nabi, S., & Benosman, M. (2017). Sparse sensing and DMD-based identification of flow regimes and bifurcations in complex flows. *SIAM Journal on Applied Dynamical Systems*, *16*(2), 1164–1196. https://doi.org/10.1137/15m104565x (cit. on pp. 41, 42).

Krause, A., & Golovin, D. (2014, February). Submodular function maximization. In *Tractability* (pp. 71–104). Cambridge University Press. https://doi.org/10.1017/cbo9781139177801.004 (cit. on p. 15).

Krause, A., Leskovec, J., Guestrin, C., VanBriesen, J., & Faloutsos, C. (2008a). Efficient sensor placement optimization for securing large water distribution networks. *Journal of Water Resources Planning and Management*, *134*(6), 516–526. https://doi.org/10.1061/(ASCE)0733-9496(2008)134:6(516) (cit. on p. 14).

Krause, A., Singh, A., & Guestrin, C. (2008b). Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, *9*(Feb), 235–284 (cit. on p. 15).

Kreth, P. A., & Alvi, F. S. (2020). Using high-frequency pulsed supersonic microjets to control resonant high-speed cavity flows. *AIAA Journal*, *58*(8), 3378–3392. https://doi.org/10.2514/1.j058912 (cit. on p. 4).

Krishnan, M., Gugercin, S., & Tarazaga, P. A. (2023). A wavelet-based dynamic mode decomposition for modeling mechanical systems from partial observations. *Mechanical Systems and Signal Processing*, *187*, 109919. https://doi.org/10.1016/j.ymssp.2022.109919 (cit. on p. 54).

Kutz, J. N., Brunton, S. L., Brunton, B. W., & Proctor, J. L. (2016a). Dynamic mode decomposition (website). http://www.dmdbook.com/ (cit. on pp. 67, 68).

Kutz, J. N., Brunton, S. L., Brunton, B. W., & Proctor, J. L. (2016b, November). *Dynamic mode decomposition: Data-driven modeling of complex systems* (Vol. 149). Society for Industrial; Applied Mathematics. https://doi.org/https://doi.org/10.1137/1.9781611974508 (cit. on pp. 3, 5, 13, 40, 55, 67, 68, 70, 74, 75).

Lee, C., Kim, J., Babcock, D., & Goodman, R. (1997). Application of neural networks to turbulence control for drag reduction. *Physics of Fluids*, *9*(6), 1740–1747. https://doi.org/10.1063/1.869290 (cit. on p. 13).

Lee, E.-T., & Eun, H.-C. (2022). Optimal sensor placement in reduced-order models using modal constraint conditions. *Sensors*, *22*(2), 589. https://doi.org/10.3390/s22020589 (cit. on p. 14).

Lenaerts, J., & Ginis, V. (2022). Questioning the question: Exploring how physical degrees of freedom are retrieved with neural networks. *Physical Review Research*, *4*(2), 023206. https://doi.org/10.1103/physrevresearch.4.023206 (cit. on p. 3).

Li, B., Liu, H., & Wang, R. (2021). Efficient sensor placement for signal reconstruction based on recursive methods. *IEEE Transactions on Signal Processing*, *69*, 1885–1898. https://doi.org/10.1109/tsp.2021.3063495 (cit. on p. 14).

Li, S., Li, W., & Noack, B. R. (2022). Least-order representation of control-oriented flow estimation exemplified for the fluidic pinball. *Journal of Physics: Conference Series*, *2367*(1), 012024. https://doi.org/10.1088/1742-6596/2367/1/012024 (cit. on p. 14).

Liu, S., Chepuri, S. P., Fardad, M., Maşazade, E., Leus, G., & Varshney, P. K. (2016). Sensor selection for estimation with correlated measurement noise. *IEEE Transactions on Signal Processing*, *64*(13), 3509–3522. https://doi.org/10.1109/TSP.2016.2550005 (cit. on pp. 15, 21, 33, 97).

Loiseau, J.-C., Noack, B. R., & Brunton, S. L. (2018). Sparse reduced-order modelling: Sensor-based dynamics to full-state estimation. *Journal of Fluid Mechanics*, *844*, 459–490. https://doi.org/10.1017/jfm.2018.147 (cit. on pp. 3, 5, 6, 8).

Lovász, L. (1983). Submodular functions and convexity. In *Mathematical programming the state of the art* (pp. 235–257). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-68874-4_10 (cit. on pp. 15, 26).

Luo, Z.-q., Ma, W.-k., So, A., Ye, Y., & Zhang, S. (2010). Semidefinite relaxation of quadratic optimization problems. *IEEE Signal Processing Magazine*, *27*(3), 20–34. https://doi.org/10.1109/msp.2010.936019 (cit. on p. 33).

Mann, J., & Kutz, J. N. (2016). Dynamic mode decomposition for financial trading strategies. *Quantitative Finance*, *16*(11), 1643–1655. https://doi.org/10.1080/14697688.2016.1170194 (cit. on p. 41).

Manohar, K., Brunton, B. W., Kutz, J. N., & Brunton, S. L. (2018). Data-driven sparse sensor placement for reconstruction: Demonstrating the benefits of exploiting known patterns. *IEEE Control Systems Magazine*, *38*(3), 63–86. https://doi.org/10.1109/MCS.2018.2810460 (cit. on pp. 8, 14, 15, 20, 102).

Manohar, K., Kaiser, E., Brunton, S. L., & Kutz, J. N. (2019). Optimized sampling for multiscale dynamics. *Multiscale Modeling & Simulation*, *17*(1), 117–136. https://doi.org/https://doi.org/10.1137/17M1162366 (cit. on p. 42).

Manohar, K., Kutz, J. N., & Brunton, S. L. (2021). Optimal sensor and actuator selection using balanced model reduction. *IEEE Transactions on Automatic Control*, *67*(4), 2108–2115. https://doi.org/10.1109/tac.2021.3082502 (cit. on p. 16).

Mao, Z., Jagtap, A. D., & Karniadakis, G. E. (2020). Physics-informed neural networks for high-speed flows. *Computer Methods in Applied Mechanics and Engineering*, *360*, 112789. https://doi.org/10.1016/j.cma.2019.112789 (cit. on p. 43).

Martin, M., Perez, J., & Plastino, A. (2001). Fisher information and nonlinear dynamics. *Physica A: Statistical Mechanics and its Applications*, *291*(1-4), 523–532. https://doi.org/10.1016/s0378-4371(00)00531-8 (cit. on p. 14).

Marusic, I., & Broomhall, S. (2021). Leonardo da Vinci and fluid mechanics. *Annual Review of Fluid Mechanics*, *53*(1), 1–25. https://doi.org/10.1146/annurev-fluid-022620-122816 (cit. on p. 2).

Maulik, R., Lusch, B., & Balaprakash, P. (2021). Reduced-order modeling of advection-dominated systems with recurrent neural networks and convolutional autoencoders. *Physics of Fluids*, *33*(3), 037106. https://doi.org/10.1063/5.0039986 (cit. on p. 3).

Mauroy, A., Mezić, I., & Susuki, Y. (Eds.). (2020). *The koopman operator in systems and control*. Springer International Publishing. https://doi.org/10.1007/978-3-030-35713-9 (cit. on pp. 2, 5).

Mendez, M. A., Ianiro, A., Noack, B. R., & Brunton, S. L. (Eds.). (2023, January). *Data-driven fluid mechanics*. Cambridge University Press. https://doi.org/10.1017/9781108896214 (cit. on pp. 2, 5).

Mezić, I. (2013). Analysis of fluid flows via spectral properties of the Koopman operator. *Annual Review of Fluid Mechanics*, *45*(1), 357–378. https://doi.org/10.1146/annurev-fluid-011212-140652 (cit. on p. 40).

Mirzasoleiman, B., Badanidiyuru, A., Karbasi, A., Vondrak, J., & Krause, A. (2015). Lazier than lazy greedy. *Proceedings of the AAAI Conference on Artificial Intelligence*, *29*(1). https://doi.org/10.1609/aaai.v29i1.9486 (cit. on pp. 15, 26).

Misaka, T., & Obayashi, S. (2014). Sensitivity analysis of unsteady flow fields and impact of measurement strategy. *Mathematical Problems in Engineering*, *2014*, 1–12. https://doi.org/10.1155/2014/359606 (cit. on pp. 8, 14).

Mons, V., & Marquet, O. (2021). Linear and nonlinear sensor placement strategies for mean-flow reconstruction via data assimilation. *Journal of Fluid Mechanics*, *923*. https://doi.org/10.1017/jfm.2021.488 (cit. on pp. 8, 14).

Montanari, A. N., Freitas, L., Proverbio, D., & Gonçalves, J. (2022). Functional observability and subspace reconstruction in nonlinear systems. *Physical Review Research*, *4*(4), 043195. https://doi.org/10.1103/physrevresearch.4.043195 (cit. on p. 15).

Murata, T., Fukami, K., & Fukagata, K. (2019). Nonlinear mode decomposition with convolutional neural networks for fluid dynamics. *Journal of Fluid Mechanics*, *882*. https://doi.org/10.1017/jfm.2019.822 (cit. on p. 3).

Nagata, T., Nakai, K., Yamada, K., Saito, Y., Nonomura, T., Kano, M., Ito, S.-i., & Nagao, H. (2022a). Seismic wavefield reconstruction based on compressed sensing using data-driven reduced-order model. *Geophysical Journal International*, *322*(1), 33–50. https://doi.org/10.1093/gji/ggac443/6824439 (cit. on p. 14).

Nagata, T., Nonomura, T., Nakai, K., Yamada, K., Saito, Y., & Ono, S. (2021). Data-driven sparse sensor selection based on A-optimal design of experiment with ADMM. *IEEE*

Sensors Journal, *21*(13), 15248–15257. https://doi.org/10.1109/jsen.2021.3073978 (cit. on pp. 15, 22, 50).

Nagata, T., Yamada, K., Nonomura, T., Nakai, K., Saito, Y., & Ono, S. (2022b). Data-driven sensor selection method based on proximal optimization for high-dimensional data with correlated measurement noise. *IEEE Transactions on Signal Processing*, *70*, 5251–5264. https://doi.org/10.1109/tsp.2022.3212150 (cit. on pp. 15, 22, 50).

Nakai, K., Nagata, T., Yamada, K., Saito, Y., Nonomura, T., Kano, M., Ito, S., & Nagao, H. (2023). Observation site selection for physical model parameter estimation towards process-driven seismic wavefield reconstruction. *Geophysical Journal International*, *234*(3), 1786–1805. https://doi.org/10.1093/gji/ggad165 (cit. on p. 14).

Nakai, K., Yamada, K., Nagata, T., Saito, Y., & Nonomura, T. (2021). Effect of objective function on data-driven greedy sparse sensor optimization. *IEEE Access*, *9*, 46731–46743. https://doi.org/10.1109/ACCESS.2021.3067712 (cit. on pp. 15, 102, 103).

Nankai, K., Ozawa, Y., Nonomura, T., & Asai, K. (2019). Linear reduced-order model based on piv data of flow field around airfoil. *Transactions of the Japan Society for Aeronautical and Space Sciences*, *62*(4), 227–235. https://doi.org/10.2322/tjsass.62.227 (cit. on pp. 17, 35).

Natarajan, M., Freund, J. B., & Bodony, D. J. (2016). Actuator selection and placement for localized feedback flow control. *Journal of Fluid Mechanics*, *809*, 775–792. https://doi.org/10.1017/jfm.2016.700 (cit. on p. 14).

Nemhauser, G. L., Wolsey, L. A., & Fisher, M. L. (1978). An analysis of approximations for maximizing submodular set functions—i. *Mathematical Programming*, *14*(1), 265–294. https://doi.org/10.1007/bf01588971 (cit. on pp. 15, 26).

Nesterov, Y., & Nemirovskii, A. (1994, January). *Interior-point polynomial algorithms in convex programming*. Society for Industrial; Applied Mathematics. https://doi.org/10.1137/1.9781611970791 (cit. on p. 22).

Nomura, S., Hara, J., & Tanaka, Y. (2022). Dynamic sensor placement based on graph sampling theory. https://doi.org/10.48550/ARXIV.2211.04019 (cit. on p. 14).

Nonomura, T., Ono, S., Nakai, K., & Saito, Y. (2021a). Randomized subspace Newton convex method applied to data-driven sensor selection problem. *IEEE Signal Processing Letters*, *28*, 284–288. https://doi.org/10.1109/LSP.2021.3050708 (cit. on p. 24).

Nonomura, T., Nankai, K., Iwasaki, Y., Komuro, A., & Asai, K. (2021b). Airfoil PIV data for linear ROM. https://github.com/Aerodynamics-Lab/Airfoil-PIV-data-for-linear-ROM (cit. on pp. 34, 35).

Nonomura, T., Nankai, K., Iwasaki, Y., Komuro, A., & Asai, K. (2021c). Quantitative evaluation of predictability of linear reduced-order model based on particle-image-velocimetry data of separated flow field around airfoil. *Experiments in Fluids*, *62*(5). https://doi.org/10.1007/s00348-021-03205-8 (cit. on pp. 6, 34, 35).

Ohmichi, Y., Ishida, T., & Hashimoto, A. (2018). Modal decomposition analysis of three-dimensional transonic buffet phenomenon on a swept wing. *AIAA Journal*, *56*(10), 3938–3950. https://doi.org/https://doi.org/10.2514/1.J056855 (cit. on pp. 7, 41, 42).

Ohmizu, K., Ozawa, Y., Nagata, T., Nonomura, T., & Asai, K. (2022). Demonstration and verification of exact DMD analysis applied to double-pulsed schlieren image of supersonic impinging jet. *Journal of Visualization*. https://doi.org/10.1007/s12650-022-00836-9 (cit. on p. 42).

Otto, S. E., Padovan, A., & Rowley, C. W. (2022). Model reduction for nonlinear systems by balanced truncation of state and gradient covariance. https://doi.org/10.48550/ARXIV.2207.14387 (cit. on pp. 3, 121).

Otto, S. E., & Rowley, C. W. (2022). Inadequacy of linear methods for minimal sensor placement and feature selection in nonlinear systems: A new approach using secants. *Journal of Nonlinear Science*, *32*(5). https://doi.org/10.1007/s00332-022-09806-9 (cit. on p. 8).

Paris, R., Beneddine, S., & Dandois, J. (2021). Robust flow control and optimal sensor placement using deep reinforcement learning. *Journal of Fluid Mechanics*, *913*(A25). https://doi.org/10.1017/jfm.2020.1170 (cit. on pp. 5, 8, 13, 14).

Patel, V. M., & Chellappa, R. (2013). *Sparse representations and compressive sensing for imaging and vision*. Springer New York. https://doi.org/10.1007/978-1-4614-6381-8 (cit. on p. 48).

Ponsioen, S., Pedergnana, T., & Haller, G. (2018). Automated computation of autonomous spectral submanifolds for nonlinear modal analysis. *Journal of Sound and Vibration*, *420*, 269–295. https://doi.org/10.1016/j.jsv.2018.01.048 (cit. on p. 3).

Proctor, J. L., Brunton, S. L., & Kutz, J. N. (2016). Dynamic mode decomposition with control. *SIAM Journal on Applied Dynamical Systems*, *15*(1), 142–161. https://doi.org/10.1137/15m1013857 (cit. on pp. 5, 41, 42).

Proctor, J. L., & Eckhoff, P. A. (2015). Discovering dynamic patterns from infectious disease data using dynamic mode decomposition. *International Health*, *7*(2), 139–145. https://doi.org/10.1093/inthealth/ihv009 (cit. on p. 41).

Raissi, M., & Karniadakis, G. E. (2018). Hidden physics models: Machine learning of nonlinear partial differential equations. *Journal of Computational Physics*, *357*, 125–141. https://doi.org/10.1016/j.jcp.2017.11.039 (cit. on p. 43).

Raissi, M., Perdikaris, P., & Karniadakis, G. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, *378*, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045 (cit. on pp. 3, 43).

Raissi, M., Yazdani, A., & Karniadakis, G. E. (2020). Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations. *Science*, *367*(6481), 1026–1030. https://doi.org/10.1126/science.aaw4741 (cit. on p. 43).

Ramezanian, D., Nouri, A. G., & Babaee, H. (2021). On-the-fly reduced order modeling of passive and reactive species via time-dependent manifolds. *Computer Methods in Applied Mechanics and Engineering*, *382*, 113882. https://doi.org/10.1016/j.cma.2021.113882 (cit. on p. 2).

Ren, Y., Ding, Y., & Liang, F. (2008). Adaptive evolutionary Monte Carlo algorithm for optimization with applications to sensor placement problems. *Statistics and Computing*, *18*(4), 375. https://doi.org/https://doi.org/10.1007/s11222-008-9079-6 (cit. on p. 14).

Rowley, C. W., Colonius, T., & Murray, R. M. (2004). Model reduction for compressible flows using POD and Galerkin projection. *Physica D: Nonlinear Phenomena*, *189*(1-2), 115–129. https://doi.org/10.1016/j.physd.2003.03.001 (cit. on p. 2).

Rowley, C. W., & Dawson, S. T. (2017). Model reduction for flow analysis and control. *Annual Review of Fluid Mechanics*, *49*(1), 387–417. https://doi.org/10.1146/annurev-fluid-010816-060042 (cit. on pp. 2, 3, 5).

Rowley, C. W., Kevrekidis, I. G., Marsden, J. E., & Lust, K. (2003). Reduction and reconstruction for self-similar dynamical systems. *Nonlinearity*, *16*(4), 1257–1275. https://doi.org/10.1088/0951-7715/16/4/304 (cit. on p. 2).

Rowley, C., Williams, M., & Kevrekidis, I. (2014). Dynamic mode decomposition and the Koopman operator: Algorithms and applications. *IPAM, UCLA*, 47 (cit. on p. 31).

Saito, Y., Nonomura, T., Yamada, K., Nakai, K., Nagata, T., Asai, K., Sasaki, Y., & Tsubakino, D. (2021). Determinant-based fast greedy sensor selection algorithm. *IEEE Access*, *9*,

68535–68551. https://doi.org/10.1109/access.2021.3076186 (cit. on pp. 7, 8, 14, 15, 102, 104).

Sakiyama, A., Tanaka, Y., Tanaka, T., & Ortega, A. (2019). Eigendecomposition-free sampling set selection for graph signals. *IEEE Transactions on Signal Processing*, *67*(10), 2679–2692. https://doi.org/10.1109/tsp.2019.2908129 (cit. on p. 14).

Sankaranarayanan, A. C., Turaga, P., Herman, M. A., & Kelly, K. F. (2016). Enhanced compressive imaging using model-based acquisition: Smarter sampling by incorporating domain knowledge. *IEEE Signal Processing Magazine*, *33*(5), 81–94. https://doi.org/10.1109/msp.2016.2581846 (cit. on p. 48).

Sankaranarayanan, A. C., Turaga, P. K., Chellappa, R., & Baraniuk, R. G. (2013). Compressive acquisition of linear dynamical systems. *SIAM Journal on Imaging Sciences*, *6*(4), 2109–2133. https://doi.org/10.1137/120863307 (cit. on p. 7).

Sasaki, Y., & Tsubakino, D. (2020). Designs of feedback controllers for fluid flows based on model predictive control and regression analysis. *Energies*, *13*(6), 1325. https://doi.org/10.3390/en13061325 (cit. on pp. 5, 13).

Scherl, I., Strom, B., Shang, J. K., Williams, O., Polagye, B. L., & Brunton, S. L. (2020). Robust principal component analysis for modal decomposition of corrupt fluid flows. *Physical Review Fluids*, *5*(5), 054401. https://doi.org/10.1103/PhysRevFluids.5.054401 (cit. on p. 13).

Schmid, P. J. (2010). Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, *656*(July 2010), 5–28. https://doi.org/10.1017/S0022112010001217 (cit. on pp. 3, 13, 40).

Shamaiah, M., Banerjee, S., & Vikalo, H. (2010). Greedy sensor selection: Leveraging submodularity. *49th IEEE conference on decision and control (CDC)*, 2572–2577. https://doi.org/https://doi.org/10.1109/CDC.2010.5717225 (cit. on pp. 15, 97, 98, 105, 106).

Shigeta, T., Nagata, T., & Nonomura, T. (2023). Density field reconstruction from time-series schlieren images via extended phase-consistent dynamic mode decomposition. *Experiments in Fluids*, *64*(7). https://doi.org/10.1007/s00348-023-03668-x (cit. on p. 41).

Shimomura, S., Sekimoto, S., Oyama, A., Fujii, K., & Nishida, H. (2020). Closed-loop flow separation control using the deep q network over airfoil. *AIAA Journal*, *58*(10), 4260–4270. https://doi.org/10.2514/1.j059447 (cit. on p. 5).

Skene, C. S., Yeh, C.-A., Schmid, P. J., & Taira, K. (2022). Sparsifying the resolvent forcing mode via gradient-based optimisation. *Journal of Fluid Mechanics*, *944*. https://doi.org/10.1017/jfm.2022.519 (cit. on p. 3).

Strassen, V. (1969). Gaussian elimination is not optimal. *Numerische Mathematik*, *13*(4), 354–356. https://doi.org/10.1007/bf02165411 (cit. on p. 32).

Sugioka, Y., Sato, H., Nakakita, K., Nakajima, T., Nonomura, T., & Asai, K. (2019). In-flight visualization of shock wave on a jet aircraft wing using lifetime-based pressure-sensitive paint technique. *AIAA Scitech 2019 Forum*. https://doi.org/10.2514/6.2019-0024 (cit. on p. 6).

Summers, T., & Shames, I. (2016). Convex relaxations and gramian rank constraints for sensor and actuator selection in networks. *2016 IEEE International Symposium on Intelligent Control (ISIC)*. https://doi.org/10.1109/isic.2016.7579985 (cit. on pp. 15, 21, 22, 28, 123).

Summers, T. H., Cortesi, F. L., & Lygeros, J. (2016). On submodularity and controllability in complex dynamical networks. *IEEE Transactions on Control of Network Systems*, *3*(1), 91–101. https://doi.org/https://doi.org/10.1109/TCNS.2015.2453711 (cit. on pp. 15, 20, 21, 26, 37).

Sun, C., Yu, Y., Li, V. O. K., & Lam, J. C. K. (2019). Multi-type sensor placements in gaussian spatial fields for environmental monitoring. *Sensors*, *19*(1), 189. https://doi.org/10.3390/s19010189 (cit. on pp. 14, 15).

Taira, K., Brunton, S. L., Dawson, S. T., Rowley, C. W., Colonius, T., McKeon, B. J., Schmidt, O. T., Gordeyev, S., Theofilis, V., & Ukeiley, L. S. (2017). Modal analysis of fluid flows: An overview. *AIAA Journal*, 4013–4041. https://doi.org/10.2514/1.J056060 (cit. on p. 68).

Taira, K., & Colonius, T. (2007). The immersed boundary method: A projection approach. *Journal of Computational Physics*, *225*(2), 2118–2137. https://doi.org/10.1016/j.jcp.2007.03.005 (cit. on pp. 67, 68).

Tanaka, H., & Nagai, H. (2023). Data-driven thermal state estimation for in-orbit systems via physics-informed machine learning. *Acta Astronautica*, *212*, 316–328. https://doi.org/10.1016/j.actaastro.2023.07.039 (cit. on p. 43).

Tiwari, N., Uchida, K., Inoba, R., Saito, Y., Asai, K., & Nonomura, T. (2022). Simultaneous measurement of pressure and temperature on the same surface by sensitive paints using the sensor selection method. *Experiments in Fluids*, *63*(11). https://doi.org/10.1007/s00348-022-03501-x (cit. on p. 14).

Tu, J. H., Rowley, C. W., Kutz, J. N., & Shang, J. K. (2014a). Spectral analysis of fluid flows using sub-nyquist-rate piv data. *Experiments in Fluids*, *55*(9), 1–13. https://doi.org/https://doi.org/10.1007/s00348-014-1805-6 (cit. on p. 42).

Tu, J. H., Rowley, C. W., Luchtenburg, D. M., Brunton, S. L., & Kutz, J. N. (2014b). On dynamic mode decomposition: Theory and applications. *Journal of Computational Dynamics*, *1*(2), 391–421. https://doi.org/10.3934/jcd.2014.1.391 (cit. on p. 47).

Tzoumas, V., Jadbabaie, A., & Pappas, G. J. (2016). Sensor placement for optimal kalman filtering: Fundamental limits, submodularity, and algorithms. *2016 American Control Conference (ACC)*. https://doi.org/10.1109/acc.2016.7524914 (cit. on pp. 15, 97).

Tzoumas, V., Xue, Y., Pequito, S., Bogdan, P., & Pappas, G. J. (2018). Selecting sensors in biological fractional-order systems. *IEEE Transactions on Control of Network Systems*, *5*(2), 709–721. https://doi.org/10.1109/tcns.2018.2809959 (cit. on p. 14).

Uciński, D. (2020). D-optimal sensor selection in the presence of correlated measurement noise. *Measurement*, *164*, 107873. https://doi.org/10.1016/j.measurement.2020.107873 (cit. on pp. 97, 106).

Udwadia, F. E. (1994). Methodology for optimum sensor locations for parameter identification in dynamic systems. *Journal of engineering mechanics*, *120*(2), 368–390. https://doi.org/10.1061/(ASCE)0733-9399(1994)120:2(368) (cit. on pp. 102, 104).

Van Huffel, S., Sima, V., Varga, A., Hammarling, S., & Delebecque, F. (2004). High-performance numerical software for control. *IEEE Control Systems Magazine*, *24*(1), 60–76. https://doi.org/10.1109/MCS.2004.1272746 (cit. on p. 30).

Van Waarde, H. J., Eising, J., Camlibel, M. K., & Trentelman, H. L. (2023). The informativity approach: To data-driven analysis and control. *IEEE Control Systems*, *43*(6), 32–66. https://doi.org/10.1109/mcs.2023.3310305 (cit. on p. 5).

Wang, W., & Zhang, B. (2012). Adaptive sampling with Bayesian compressive sensing in radar sensor networks and image. *EURASIP Journal on Wireless Communications and Networking*, *2012*(1). https://doi.org/10.1186/1687-1499-2012-257 (cit. on pp. 98, 106).

Waters, A., Sankaranarayanan, A., & Baraniuk, R. (2011). SpaRCS: Recovering low-rank and sparse matrices from compressive measurements. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, & K. Weinberger (Eds.), *Advances in neural information processing systems*

(Vol. 24). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2011/file/0ff8033cf9437c213ee13937b1c4c455-Paper.pdf (cit. on p. 7).

Wouwer, A. V., Point, N., Porteman, S., & Remy, M. (2000). An approach to the selection of optimal sensor locations in distributed parameter systems. *Journal of process control*, *10*(4), 291–300. https://doi.org/https://doi.org/10.1016/S0959-1524(99)00048-7 (cit. on pp. 15, 16, 102).

Wu, J., Xiao, D., & Luo, M. (2023). Deep-learning assisted reduced order model for high-dimensional flow prediction from sparse data. *Physics of Fluids*, *35*(10). https://doi.org/10.1063/5.0166114 (cit. on p. 7).

Yamada, K. (2023). Sensor placement based on observability gramian. https://github.com/Aerodynamics-Lab/Sensor-Placement-Based-on-Observability-Gramian (cit. on pp. 31, 123).

Yamada, K., Saito, Y., Nankai, K., Nonomura, T., Asai, K., & Tsubakino, D. (2021). Fast greedy optimization of sensor selection in measurement with correlated noise. *Mechanical Systems and Signal Processing*, *158*, 107619. https://doi.org/https://doi.org/10.1016/j.ymssp.2021.107619 (cit. on pp. 15, 106, 121).

Yamada, K., Sasaki, Y., Nagata, T., Nakai, K., Tsubakino, D., & Nonomura, T. (2023). Efficient sensor node selection for observability gramian optimization. *Sensors*, *23*(13), 5961. https://doi.org/10.3390/s23135961 (cit. on p. 102).

Yasuda, Y., Onishi, R., Hirokawa, Y., Kolomenskiy, D., & Sugiyama, D. (2022). Super-resolution of near-surface temperature utilizing physical quantities for real-time prediction of urban micrometeorology. *Building and Environment*, *209*, 108597. https://doi.org/10.1016/j.buildenv.2021.108597 (cit. on pp. 3, 7).

Ye, L., Roy, S., & Sundaram, S. (2018). On the complexity and approximability of optimal sensor selection for kalman filtering. *2018 Annual American Control Conference (ACC)*, 5049–5054. https://doi.org/10.23919/ACC.2018.8431016 (cit. on p. 15).

Yeo, W.-J., Taulu, S., & Kutz, J. N. (2022). Efficient magnetometer sensor array selection for signal reconstruction and brain source localization. *arXiv preprint arXiv:2205.10925* (cit. on p. 14).

Yildirim, B., Chryssostomidis, C., & Karniadakis, G. (2009). Efficient sensor placement for ocean measurements using low-dimensional concepts. *Ocean Modelling*, *27*(3-4), 160–173. https://doi.org/10.1016/j.ocemod.2009.01.001 (cit. on p. 14).

Yoshimura, R., Yakeno, A., Misaka, T., & Obayashi, S. (2020). Application of observability Gramian to targeted observation in WRF data assimilation. *Tellus A: Dynamic Meteorology and Oceanography*, *72*(1), 1–11. https://doi.org/10.1080/16000870.2019.1697602 (cit. on pp. 8, 14, 15).

Yu, L., Sun, H., Barbot, J., & Zheng, G. (2012). Bayesian compressive sensing for cluster structured sparse signals. *Signal Processing*, *92*(1), 259–269. https://doi.org/10.1016/j.sigpro.2011.07.015 (cit. on pp. 8, 44, 58, 60, 64).

Yu, Y., Zhang, Y., Qian, S., Wang, S., Hu, Y., & Yin, B. (2021). A low rank dynamic mode decomposition model for short-term traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems*, *22*(10), 6547–6560. https://doi.org/10.1109/tits.2020.2994910 (cit. on p. 41).

Yuan, L., Liu, J., & Ye, J. (2013). Efficient methods for overlapping group Lasso. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(9), 2104–2116. https://doi.org/10.1109/tpami.2013.17 (cit. on pp. 64, 79, 121).

Yuan, M., & Lin, Y. (2005). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *68*(1), 49–67. https://doi.org/10.1111/j.1467-9868.2005.00532.x (cit. on p. 64).

Zare, A., Mohammadi, H., Dhingra, N. K., Georgiou, T. T., & Jovanovic, M. R. (2020). Proximal algorithms for large-scale statistical modeling and sensor/actuator selection. *IEEE Transactions on Automatic Control*, *65*(8), 3441–3456. https://doi.org/10.1109/tac.2019.2948268 (cit. on p. 22).

Zhang, H., Ayoub, R., & Sundaram, S. (2017). Sensor selection for Kalman filtering of linear dynamical systems: Complexity, limitations and greedy algorithms. *Automatica*, *78*, 202–210. https://doi.org/10.1016/j.automatica.2016.12.025 (cit. on p. 15).

Zhang, X., Tian, Y., Cheng, R., & Jin, Y. (2015). An efficient approach to nondominated sorting for evolutionary multiobjective optimization. *IEEE Transactions on Evolutionary Computation*, *19*(2), 201–213. https://doi.org/10.1109/TEVC.2014.2308305 (cit. on p. 15).

Zhang, X., Ji, T., Xie, F., Zheng, H., & Zheng, Y. (2022). Unsteady flow prediction from sparse measurements by compressed sensing reduced order modeling. *Computer Methods in Applied Mechanics and Engineering*, *393*, 114800. https://doi.org/10.1016/j.cma.2022.114800 (cit. on pp. 13, 41).

Zhou, K., Salomon, G., & Wu, E. (1999). Balanced realization and model reduction for unstable systems. *International Journal of Robust and Nonlinear Control*, *9*(3), 183–198. https://doi.org/10.1002/(sici)1099-1239(199903)9:3<183::aid-rnc399>3.0.co;2-e (cit. on p. 16).