Fast Greedy Optimization of Sensor Selection in Measurement with Correlated Noise

Keigo Yamada, Yuji Saito, Koki Nankai, Taku Nonomura, Keisuke Asai, Daisuke Tsubakino

November 2019

Abstract

The noise-robust greedy sparse-sensor-selection method based on the determinant maximization is proposed. The sensor selection by maximizing the determinant of the matrix which corresponds to the inverse matrix appearing in the Bayesian estimation operator is proposed. The Bayesian estimation of state variables using the optimized sensors and the prior information robustly works even in the presence of the correlated noises. In addition, computational efficiency are improved by considering the algorithms. Then, the proposed algorithms are applied to various problems. Although the proposed method needs longer time to select the sensor location, the estimation results become more accurate than those of the previously-presented determinant-based greedy method. It should be noted that the locations of sensors determined by the proposed method are scattered compared to those of the previous method when the proposed algorithm is applied to practical dataset. This is because the proposed method avoid the very close sensors that are often contaminated by the correlated noise.

1 Introduction

Reduced-order modeling and sparse sensing are effective for calculating high-dimensional, but low-rank structured data[1]. These ideas enable us to obtain the characteristics of the data. In the fluid dynamics fields, proper orthogonal decomposition (POD), (or principal component analysis) is often adopted for the
order reduction. This method leads to the reduced order modeling of the dynamics using the Galerkin projection and dynamic mode decomposition, and many researchers try to use those technique for cost-effective fluid measurement or flow control. Although those advanced techniques are becoming available, the flow field reconstruction based on the original POD is focused in the present study. Manohar et al. illustrated that the high-dimensional, but low-rank structured data can be reconstructed by the (sub)optimized sparse sensors in the context of the least square problem of the mode reconstructions [2]. They showed that the importance of the sensor placements. Because their framework is based on the least square solver, the optimization of the sensor placement in the least square problem should be considered. With regard to the sensor placement problem for the generalized the least square estimation, Joshi and Boyd proposed the convex relaxation of the combination problem [3]. Their implementation works well, but the computational cost is significantly high because its complexity is $O(n^3)$ where $n$ is a number of the sensor candidate and it is almost impossible to apply this method to high dimensional problems of $n = 10^9$ which is recently becoming feasible in the fluid dynamics simulation. On the other hand, the computationally cheaper greedy method has recently been proposed. Manohar et al. proposed QR-based greedy method [2]. This method is related to the discrete empirical interpolation method [4] and the QR-based discrete empirical interpolation method [5] in the framework of the Galerkin projection [6] for reduced-order modeling. Recently, Saito et al. successfully extended this method to vector sensor measurements, considering the application of sparse processing of particle image velocimetry in fluid dynamic fields [7]. Furthermore, Saito et al. reconsidered the formulation of the greedy method by dividing the problem into two cases: a number of sensors is less than or equal to that of the state variables and a number of sensors is greater than that of the state variables [8]. They derived appropriate formulations for each and proved that the only former algorithm is corresponding to the previous QR method. The greedy algorithm proposed by Saito et al. is able to choose appropriate sensor (location) as well as the convex relaxation algorithm which usually works better than the previous QR algorithm, while the computational cost of the greedy method proposed by Saito et al. is much lower than those of the convex relaxation and previous QR algorithms. Although there were some studies conducted for effective sensor selection methods that compress both calculation cost and reconstruction error, all of them have advantages and disadvantages. For instance, the greedy algorithm recently developed by Saito et al. sometimes does not work well for noisy dataset as in the fluid dynamic measurements [8]. This research aims improvement of the greedy method recently presented by Saito et
by considering correlated-sensor-noise information.

Firstly, basics of SVD-based reduced order modeling and sparse sensing are briefly revisited. Algorithms of proposed in the previous and present studies are given in Section 2.1 and Section 2.2, respectively, and then the superiority of the proposed method for the noisy data set is shown by applying it to randomly generated data matrices and to NOAA-SST data matrix\[9\]in Section 3. Finally, Section 4 concludes the paper.

2 Formulations and Algorithms

2.1 Reduced-order modeling, sparse sensing and previous greedy optimization of sparse sensors

First, the reduced-order modeling is conducted by POD applied to training data, and this procedure is exactly the same as the singular value decomposition if the spatial and temporal desensitization of data are uniform. The training data has a large-size data matrix \(X\) that consists of spatial \(n\) variables and each variable has \(m\) time series, where usually \(n > m\) in the fluid dynamics. Here, \(X\) can be decomposed by economy SVD as:

\[
X = USV^T
\]

(1)

\[
= \begin{pmatrix} x(1) & x(2) & \ldots & x(m) \end{pmatrix}
\]

(2)

Here, \(U \in \mathbb{R}^{n \times m}\), \(S \in \mathbb{R}^{m \times m}\), \(V \in \mathbb{R}^{m \times m}\) and \(x(t) \in \mathbb{R}^n (t \in [1, m])\) are the spatial modes, the diagonal matrix of the singular values, the temporal modes and \(t\)-th column vector of \(X\), respectively. Here, mode in \(U\) and \(V\) is orthonormal to each other, respectively, and diagonal components of \(S\) are sorted in descending order. Here, \(U_{1:r_1}, V_{1:r_1}\) and \(S_{1:r_1}\) are defined to be the leading \(r_1\) columns of \(U\) and \(V\), and first \(r_1 \times r_1\) components of \(S\). Moreover, \(U_{(r_1+1):m}, V_{(r_1+1):m}\) and \(S_{(r_1+1):m}\) are defined to be the remainder matrix of \(U\), \(V\) and \(S\) as shown below

\[
U = \begin{bmatrix} U_{1:r_1} & U_{(r_1+1):m} \end{bmatrix}, V = \begin{bmatrix} V_{1:r_1} & V_{(r_1+1):m} \end{bmatrix}, S = \begin{bmatrix} S_{1:r_1} & 0 \\ 0 & S_{(r_1+1):m} \end{bmatrix}
\]

(3)

Then, the mode higher than given \(r_1\) \((r_1 \ll m)\) are truncated for the reduced order modeling, and \(X\) are approximated to be:

\[
X \approx X_{1:r_1} = U_{1:r_1}S_{1:r_1}V_{1:r_1}^T
\]

(4)
If this approximation works, the number of parameters can be reduced from \( m \) to \( r_1 \). It means that complicated data became just a superposition of only \( r_1 \) spatial modes.

Once this approximation is assumed, the original data are estimated by the sparse sensor information. When \( x \) of Eq. 2 can be effectively described by \( U_{1;r_1} z \), where \( z \in \mathbb{R}^{r_1} \) is the latent mode amplitude, the amplitude of each mode can be calculated by signal information obtained from a limited number of sensors:

\[
y = Hx \\
\approx H(U_{1;r_1} z) \\
\equiv Cz
\]

Here, \( H \in \mathbb{R}^{p \times n} \), \( C \in \mathbb{R}^{p \times r_1} \) and \( Z \in \mathbb{R}^{r_1 \times m} \) are the sensor location matrix for \( p \) sensors, the observation matrix, and the coefficient matrix of each mode, respectively. Each row vector of \( H \) is a unit vector and the position of unity in the unit vectors indicate which point among \( n \) sensor candidates is chosen. The sparse sensor gathers the signal generated by each mode. In this formulation, \( z \) seems to be the latent state variables and \( C \) the observation matrix.

Once \( H \) matrix is created, the estimated amplitudes \( \tilde{z} \) can be obtained by the pseudo-inverse operation. The notation of tilde above any variables or matrices like \( \tilde{z} \), refers they are estimated.

\[
\tilde{z} = C^+ y \\
= \begin{cases} 
C^T(CC^T)^{-1}y , & p \leq r , \\
(C^TC)^{-1}C^Ty , & p > r . 
\end{cases}
\]

and the full state can be recovered by

\[
\tilde{x} = U_{1;r_1} \tilde{z} .
\]

The variance of the estimation error is focused on because the estimation error could be minimized. Joshi and Boyd clearly showed the objective function for evaluation[3]. The covariance matrix \( \Sigma \) is calculated from the difference between estimated and original state using \( C \) in Eq. 7 and \( z \in \mathbb{R}^{r_1} \) as a vector of amplitudes of each mode:

\[
\Sigma \propto (C^TC)^{-1}
\]
Here, $\Sigma$ is minimized, and sensors maximizing the objective function below are selected:

$$\text{maximize } \det(C^T C)$$  \hspace{1cm} (12)

All the combination of $p$ point sensors out of an $n$ sensor candidates should be searched by brute-force algorithm for the optimized solution, but it may take very long time ($= \frac{n!}{(n-p)!p!}$). Instead of the brute-force algorithm, the greedy method for the suboptimized solution is adopted by adding a sensor step by step which improves our objective function. This approach effectively works, but sometimes does not as demonstrated later in the test problem. This is because the prior information of the data is not fully utilized in the estimation. More concretely, the previous method equally expects the same level of the mode amplitudes for low- to high-order modes, though the mode amplitude can actually be estimated with the training data. In addition, additional artificial noises are not considered, but the data includes at least the noises from the truncated higher POD modes. The noises from the truncated higher modes should have correlation with the different spatial variables each other. These points are not considered at all in the previous straightforward implementation. The previous determinant-based greedy method is designed in the framework of the sparse sensing of the POD modes above[8]. The observation matrix $C_k$ using first-to-$k$th sensors is defined as follows:

$$C_k = \begin{bmatrix} u_{i_1}^T & u_{i_2}^T & \ldots & u_{i_{k-1}}^T & u_{i_k}^T \end{bmatrix}^T,$$  \hspace{1cm} (13)

where $i_k$ and $u_{i_k}$ are an index of the $k$th selected sensor and the corresponding row vector of the sensor-candidate matrix, respectively. The following algorithm can be derived straightforwardly by considering pseudo inverse operation of this observation matrix. The greedy sensor-selection problem is defined as the sensor locations of $(k-1)$ sensor selected in the previous steps are known. In this algorithm, the $k$th sensor is greedily selected to maximize $\det(C^T C)$ until a number of sensors reaches $r_1$, because $(C^T C)$ is not full rank and its determinant is always zero. Saito et al. showed that this corresponds to search of a row vector of $U_{1:r_1}$ which has largest orthogonal components to the $(k-1)$ sensor vector space, which is exactly the same as the previous QR-pivoting method. After the number reaches $r_1$, a sensor which maximize $\det(C^T C)$ is simply added for each steps so that the error in the estimation by the pseudo inverse operation could be reduced. This algorithm is shown in Alg. [1]. Saito et al. also presented this objective function can be written in scholar calculation and this quickens the computation. [8]
Algorithm 1 Overview of Determinant-based greedy algorithm for sparse sensor placement

Set sensor-candidate matrix $C_k = [u_{i_1}^T \ u_{i_2}^T \ \ldots \ u_{i_{k-1}}^T \ u_{i_k}^T]_T$

for $k = 1, \ldots, r, \ldots, p$ do
  if $k \leq r$ then
    $i_k = \arg \max_i \det(C_k C_k^T)$
    $= \arg \max_i \det \left( \begin{bmatrix} C_{k-1} \\ u_i \end{bmatrix} \begin{bmatrix} C_{k-1}^T \\ u_i^T \end{bmatrix} \right)$
    $= \arg \max_i u_i \left( I - C_{k-1}^T (C_{k-1} C_{k-1}^T)^{-1} C_{k-1} \right) u_i^T$
  else
    $i_k = \arg \max_i \det(C_k^T C_k)$
    $= \arg \max_i \det \left( \begin{bmatrix} C_{k-1}^T \\ u_i^T \end{bmatrix} \begin{bmatrix} C_{k-1} \\ u_i \end{bmatrix} \right)$
    $= \arg \max_i \left( 1 + u_i \left( C_{k-1}^T C_{k-1} \right)^{-1} u_i^T \right)$
  end if
end for
2.2 Bayesian Estimation using Sparse Sensor

Two more conditions based on prior information are involved for more robust estimation in the present study instead of the straightforward application of the pseudo inverse operation as in the previous subsection. One is the expected POD mode amplitude, and the other is the covariance of the sensor noise. Both information can be estimated by the training data matrix. The covariance of the POD mode amplitudes can be expected from the training data:

\[ E(zz^T) \equiv Q \]
\[ \approx \frac{1}{m} S_{1:r_1} V_{1:r_1}^T V_{1:r_1} S_{1:r_1} \]
\[ = \frac{1}{m} S_{1:r_1}^2, \quad (14) \]

where \( E(\theta) \) is the expectation value of variable \( \theta \) and diagonal matrix \( Q \in \mathbb{R}^{r_1 \times r_1} \) is approximated by square of singular value matrix, which is also diagonal matrix. In addition, if there is sensor noise of \( w \), the full-state observation becomes

\[ x = U_{1:r_1} z + w, \quad (15) \]
\[ E(ww^T) \equiv R \quad (16) \]

and the sparse observation becomes

\[ y = HU_{1:r_1} z + Hw, \quad (17) \]
\[ E(Hww^T H^T) = HE(ww^T)H^T \]
\[ = HRH^T \]
\[ = R. \quad (20) \]

where \( R \in \mathbb{R}^{p \times p} \) represents the covariance matrix of the noises in signals that \( p \) sensors capture. Here, the sensor noise is assumed to be estimated from the truncated high-order modes in the present study. Therefore,

\[ R = E(ww^T) \]
\[ = E((x - U_{1:r_1} z)(x - U_{1:r_1} z)^T) \]
\[ \approx (USV^T - U_{1:r_1} S_{1:r_1} V_{1:r_1}^T)(USV^T - U_{1:r_1} S_{1:r_1} V_{1:r_1}^T)^T \]
\[ = (U_{(r_1+1):m} S_{(r_1+1):m} V_{(r_1+1):m}^T)(U_{(r_1+1):m} S_{(r_1+1):m} V_{(r_1+1):m}^T)^T \]
\[ = U_{(r_1+1):m} S_{(r_1+1):m} U_{(r_1+1):m}^T, \quad (21) \]

where \( U \) and \( S \) are the left and right singular vectors and \( V \) are the singular values of the training data matrix.
and

\[ R = H\Phi H^T \]
\[ \approx H(U_{(r+1):m} S_{(r+1):m}^2 U_{(r+1):m}^T) H^T. \]  

(22)  
(23)

In the present prior information, the Bayesian estimation is derived. Here, a priori probability density function of POD mode amplitude becomes:

\[ p(z) \propto \exp(-z^T Q^{-1} z). \]

(24)

and the conditional probability density function of \( y \) under given \( z \) is as follows:

\[ p(y|z) \propto \exp(-(y - Cz)^T R^{-1} (y - Cz)). \]

(25)

These relation gives us the a posteriori probability density function:

\[ p(z|y) \propto p(y|z)p(z) \]
\[ \propto \exp(-(y - Cz)^T R^{-1} (y - Cz)) \exp(-z^T Q^{-1} z) \]
\[ = \exp(-(y - Cz)^T R^{-1} (y - Cz) - z^T Q^{-1} z). \]

(26)

This equation gives us the maximum likelihood estimation:

\[ \hat{z} = (C^T R^{-1} C + Q^{-1})^{-1} C^T R^{-1} y. \]

(27)

Here, it should be noted that normalization term \( Q \) enables inverse operation in Eq. [27] for both \( p \leq r \) and \( p > r \) conditions, unlike the previous greedy algorithm. In this estimation, the covariance matrix is estimated to be

\[ \hat{\Sigma} = (C^T R^{-1} C + Q^{-1})^{-1}. \]

(28)

Here, \( \hat{\Sigma} \) is to be minimized, and the sensors maximizing the objective function below are selected:

\[ \max \text{det}(C^T R^{-1} C + Q^{-1}). \]

(29)

It should be noted that the convex optimization of sensor selection for the Bayesian estimation has already derived under the condition in which \( R \) is diagonal matrix by Joshi and Boyd[3]. In the present study, \( R \) includes nondiagonal components which presents the correlated noise as is often the case in the reduced order modeling which truncates the high-order modes. The sensor selection is
considered in the framework of the Bayesian estimation. Because of the regularization by $Q$, the case classification is not required and the straightforward greedy optimization of $\det(C^T R^{-1} C + Q^{-1})$ can be available unlike the previous determinant optimization\cite{8}. This Bayesian-determinant-based greedy (BDG) algorithm is presented in Alg. 2. The previous method (Alg. 1) tends to select largely fluctuated points regardless of the similarity and the amount of noise in the signals. On the other hand, the proposed method avoids those points by considering correlation matrix of noises. As shown in the previous presented paper\cite{8}, the greedy algorithm reduces the computational cost by alternating determinant calculation into scalar one by utilizing matrix determinant lemma (see Appendix B). This lemma is applied to the proposed method and the algorithm is accelerated as shown in the next subsection.

**Algorithm 2** Determinant-based greedy algorithm considering correlation between sensors

Set sensor-candidate matrix $C_k = [ \mathbf{u}_{1}^T \mathbf{u}_{2}^T \ldots \mathbf{u}_{i_k}^T \mathbf{u}_{i_{k-1}}^T \mathbf{u}_{i_k}^T ]^T$

Set amplitudes variance matrix $Q = \text{diag}(\sigma_1^2 \sigma_2^2 \ldots \sigma_{(r_1-1)}^2 \sigma_{r_1}^2 )$

\begin{verbatim}
for $k = 1, \ldots, r, \ldots, p$ do
    $\hat{R}_k = H_k R H_k^T$
    $i_k = \arg \max_{i} \det \left( C_k^T \hat{R}_k^{-1} C_k + Q^{-1} \right)$
    $= \arg \max_{i} \det \left( \begin{bmatrix} C_{k-1}^T \mathbf{u}_i^T \end{bmatrix} \hat{R}_k^{-1} \begin{bmatrix} C_{k-1} \mathbf{u}_i \end{bmatrix} + Q^{-1} \right)$
    $R_k \leftarrow \hat{R}_k$
    $k \leftarrow k + 1$
end for
\end{verbatim}

| Name          | Method Implementation                        |
|---------------|---------------------------------------------|
| DG (previous) | $p \leq r : \arg \max \det(C C^T)$ Alg. 1 |
|               | $p > r : \arg \max \det(C C^T)$            |
| BDG (present) | $\arg \max \det(C^T R^{-1} C + Q^{-1}) $ Alg. 2 & Alg. 3 |

Table 1: Greedy sensor selection methods investigated in the present study
2.3 Fast algorithm

The fast implementation is considered based on the algorithm shown in previous subsection. Given \( R_{k-1} \) and \( C_{k-1} \), the objective function is reformulated. First, \( R_k \) and its inverse are considered for \( k \)th sensor candidate. The covariance matrix \( R_k \) is formulated:

\[
R_k = \begin{pmatrix} R_{k-1} & p_k^T \\ p_k & q_k \end{pmatrix},
\]

(30)

where

\[
p_k = \frac{1}{m} \mathbb{E}(h_iww^TH_{k-1}^T) \]
\[
= \frac{1}{m} h_i\mathbb{E}(ww^T)H_{k-1}^T \]
\[
= \frac{1}{m} h_iRH_{k-1}^T \]
\[
\approx \frac{1}{m} h_i(U(r_1+1):mS_{(r_1+1):m}U^T_{(r_1+1):m})H_{k-1}^T
\]

(31)

and

\[
q_k = \frac{1}{m} \mathbb{E}(h_iww^T h_i^T) \]
\[
= \frac{1}{m} h_i\mathbb{E}(ww^T)h_i^T \]
\[
= \frac{1}{m} h_iRH_i^T \]
\[
\approx \frac{1}{m} h_i(U(r_1+1):mS_{(r_1+1):m}U^T_{(r_1+1):m})h_i^T.
\]

(32)

(33)

(34)

(35)

(36)

Here, \( H_{k-1} \) is a first-to-\((k-1)\)th-sensor selection matrix, and \( h_i \) is a sensor selection row vector for \( i \)th sensor, of which the \( i \)th component is unity and the others are zero. In addition, \( R_{k-1} \) is obtained in the previous \((k-1)\)th step. Once, \( q_k \) and \( p_k \) are defined for every sensor candidate \( i \) in the \( k \)th sensor selection, \( R_{k-1}^{-1} \) is obtained as follows:

\[
R_{k-1}^{-1} \equiv \begin{pmatrix} \alpha_k & \beta_k^T \\ \beta_k & \delta_k \end{pmatrix}
\]

(37)
where

\[ \alpha_k = R_k^{-1} + \frac{1}{q_k - p_k R_k^{-1} p_k^T} R_k^{-1} p_k R_k^{-1}, \]

\[ \beta_k = -\frac{p_k R_k^{-1}}{q_k - p_k R_k^{-1} p_k^T}, \]

\[ \delta_k = \frac{1}{q_k - p_k R_k^{-1} p_k^T}. \]

The objective function is now considered based on the expression above.

\[ \det \left( C_k^T R_k^{-1} C_k + Q^{-1} \right) = \det \left[ \begin{bmatrix} C_k^T & u_i^T \end{bmatrix} \begin{bmatrix} \alpha_k & \beta_k^T \\ \beta_k & \delta_k \end{bmatrix} \begin{bmatrix} C_k^{-1} \\ u_i \end{bmatrix} + Q^{-1} \right] \]

\[ = \det \left( C_k^{-1} \alpha_k C_k^{-1} + u_i^T \beta_k C_k^{-1} + C_k^{-1} \beta_k^T u_i + \delta_k u_i^T u_i + Q^{-1} \right) \]

\[ = \det \left( C_k^{-1} + Q^{-1} + \frac{(C_k^{-1} R_k^{-1} p_k^T - u_i^T) (p_k R_k^{-1} C_k^{-1} - u_i)}{q_k - p_k R_k^{-1} p_k^T} \right) \]

\[ = \left( 1 + \frac{(p_k R_k^{-1} C_k^{-1} - u_i) (C_k^{-1} R_k^{-1} C_k^{-1} + Q^{-1})^{-1} (C_k^{-1} R_k^{-1} p_k^T - u_i)}{q_k - p_k R_k^{-1} p_k^T} \right) \times \det \left( C_k^{-1} R_k^{-1} C_k^{-1} + Q^{-1} \right). \]

Because \( C_k^{-1} R_k^{-1} C_k^{-1} \) has been already obtained, the sensor can be selected by the following scholar evaluation:

\[ \arg \max \det \left( C_k^T R_k^{-1} C_k + Q^{-1} \right) \]

\[ = \arg \max \frac{(p_k R_k^{-1} C_k^{-1} - u_i) (C_k^{-1} R_k^{-1} C_k^{-1} + Q^{-1})^{-1} (C_k^{-1} R_k^{-1} p_k^T - u_i)}{q_k - p_k R_k^{-1} p_k^T}. \]

Once the sensor is selected, \( R \) and \((CRC)^{-1}\) are updated. This algorithm is described in Alg.\[5\] including another improvement for computational efficiency introduced in the next subsection.

### 2.4 Memory Efficient Implementation

As was already introduced in Eq.\[21\] the covariance of noise for every pair of observation points can be estimated by multiplying \( U_{(r_1+1)\cdot m} \) and \( S_{(r_1+1)\cdot m} \) and stored.
to be $\mathcal{R}$. In Alg. 2, $p_k$ and $q_k$ are constructed by taking the corresponding parts of $\mathcal{R}$. It takes the same calculation time in every step to obtain $p_k$ and $q_k$ of the all sensor candidate $i$. Although this could be a straightforward way to calculate $p_k$ and $q_k$ with less complexity, it often runs over the memory to store $\mathcal{R}$ since $n^2$, the size of $\mathcal{R}$, can often reach billions or more in the fluid dynamic application. Therefore, the following implementation is proposed: only $q_k$ is constructed and stored as an $n$-components vector which corresponds to diagonal components of $\mathcal{R}$, and $p_k$ is approximated by using the dominant first $(r_2 - r_1)$ columns of $U_{(r_1+1):m}$ in Eq. 21. This modification is reasonable because nondiagonal components are inherently small compared to diagonal ones of $\mathcal{R}$ thereby higher modes have less effect on the $R_k$ and its inverse as the mode number increases. Here, $q_k$ can be estimated and directly stored by

$$q_k \approx \frac{1}{m} \| \xi_k - \bar{\xi}_k \|^2,$$  (43)

where $\xi_k$ and $\bar{\xi}_k$ are $k$th row vectors of $X$ and $X_{1:r_1}$, respectively. (Here, it should be noted that $x_k$ is defined to be the column vector of $X$). In addition, nondiagonal components can be approximated as follows:

$$\tilde{p}_k \approx h_i(U_{(r_1+1):r_2}S_{(r_1+1):r_2}^2U_{(r_1+1):r_2}^T)H_{k-1}^T,$$  (44)

where, $U_{(r_1+1):r_2}$ and $S_{(r_1+1):r_2}$ are the leading $(r_2 - r_1)$ columns of remainder spatial modes matrix $U_{(r_1+1):m}$ and first $(r_2 - r_1) \times (r_2 - r_1)$ components of a remainder singular value matrix $S_{(r_1+1):m}$. If the number $p$ of sensors selected is much smaller than $n$, the covariance of noises between all sensor candidates $i$ and the already chosen $k$ sensors can be computed at the end of $k$th step and stored as $\mathcal{R}_k \in \mathbb{R}^{k \times m}$ for the next step. These processes are called the $r_2$ truncation. This approximation simply reduces amount of memory stored or computation complexity. Furthermore, $q_k$ can be also approximated using the POD modes instead of application of Eq. 43 as follows:

$$\tilde{q}_k \approx h_i(U_{(r_1+1):r_2}S_{(r_1+1):r_2}^2U_{(r_1+1):r_2}^T)h_i^T.$$  (45)

Those modification in Sections 2.3 and 2.4 are integrated into the Alg. 3. This formulation is for the case that $q$ is computed with all the $(r_1 + 1 \approx m)$ components as in Eq. 36 and stored as a vector $d = \text{diag}(\mathcal{R})$, and $p$ is approximated by truncating modes higher than $r_2$ as in Eq. 44.

The effects of truncating $r_2$ of $p$ are shown in subsection 3.4.
Algorithm 3 Detailed accelerated determinant-based greedy algorithm considering correlation between sensors

Set amplitudes variance matrix \( Q = \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_{(r_1-1)}^2, \sigma_{r_1}^2) \)

Set noise variance vector \( d = (\|\xi_1 - \bar{\xi}_1\|^2, \|\xi_2 - \bar{\xi}_2\|^2, \ldots, \|\xi_n - \bar{\xi}_n\|^2)/m \)

\( k = 1 \)

\( q_i = d(i) \)

\( i_1 = \arg \max_i \det \left( u_i^T q_i^{-1} u_i + Q^{-1} \right) \)

\( R \leftarrow (q_{i_1}) \)

Set the sensor-candidate matrix \( C = [u_{i_1}] \)

\textbf{for} \( k = 2, \ldots, r, \ldots, p \) \textbf{do}

Calculate and store \( (C^T R^{-1} C + Q^{-1})^{-1}, R^{-1} C, C^T R^{-1} \)

Construct the noise covariance vector \( p_i \) and the noise variance \( q_i \)

\( p_i = h_i(U_{(r_1+1):r_2} S^2_{(r_1+1):r_2} U_{(r_1+1):r_2}^T H_{k-1}^T) \)

\( q_i = d(i) \)

\( i_k = \arg \max_i \det \left( C^T \hat{R}^{-1} C + Q^{-1} \right) \)

\( = \arg \max_i \left( p_i R^{-1} C - u_i \right) (C^T R^{-1} C + Q^{-1})^{-1} \times \left( C^T R^{-1} p_i^T - u_i^T \right) / \left( q_i - p_i R^{-1} p_i^T \right) \)

\( R \leftarrow \left( \begin{array}{c} R \\ p_i \\ q_i \end{array} \right) \)

Set sensor-candidate matrix \( C = [u_{i_1}^T u_{i_2}^T \ldots u_{i_k}^T]^T \)

\textbf{end for}
3 Results and Discussions

3.1 Sensors for randomized matrix

The numerical experiments are conducted and the proposed method is validated. The random sensor candidate matrices, \( X_{\text{rand}} = USV_{\text{train}}^T \) \((X_{\text{rand}} \in \mathbb{R}^{1000 \times 500})\), were set, where \( U \) and \( V_{\text{train}} \) consist of 500 orthogonal vectors generated by QR decomposition of normally distributed random matrices and components of diagonal matrix \( S \) are \( \text{diag}(S) = [1, 1/\sqrt{2}, 1/\sqrt{3}, ..., 1/\sqrt{500}] \), respectively. These slowly decaying diagonal components of \( S \) are simulating the data of an actual flow field. Validation on the test data is also conducted by reconstructing \( X_{\text{test}} = USV_{\text{test}}^T \) with first 10 POD modes, where \( V_{\text{train}} \) is a newly generated normally distributed random matrix of \( 500 \times 500 \).

First, the reconstruction error which is used throughout the paper is defined as follows:

\[
\epsilon = \frac{\sum_{t=1}^m \|x(t) - \hat{x}(t)\|_2}{\|x(t)\|_2},
\]

(46)

where numerator and denominator are \( L_2 \) norm of difference of the original and estimated states and \( L_2 \) norm of original state, respectively. Figure illustrates the reconstruction errors defined above against the number of sensors. Moreover, the black lines are the results of reconstruction using sensors of previously presented greedy method and the red ones are those using sensors of the proposed method. The results show that the proposed method estimates the whole state better than the previous method. Especially, the reconstruction error of the proposed method does not increase at \( p = 10 \) in contrast to that of the previous method. Increase of error in the previous method is because sensors directly converts the signals to amplitudes of \( r_1 \) mode amplitudes when the numbers of sensors \( p \) and reconstruction modes \( r_1 \) are equal or close, even if they contain intense correlated noises corresponding to higher POD modes. Because sensors of the proposed method are selected and less correlated noise are contaminated, the error of the proposed method at \( p = 10 \) is smaller than that of the previous method. It should be noted that the proposed method recorded better results than the previous method with regardless of choice of the training data or the test data.

Another numerical experiment is conducted and the effectiveness of the noise-robust sensor selection and the Bayesian estimation is investigated individually. The novelty of the proposed method can be decomposed into two parts; the noise-robust sensor selection as in Eq. and the Bayesian estimation as in Eq. 27.
Either of them can be replaced by one corresponding to those of the previous method. Two more combinations of algorithms are tested in addition to the results shown in Fig. 1 for verification. In total, four different combinations of the algorithms are tested, and they are presented in Table 2. Here, LSS-LSE and NRS-BE correspond to black and red lines in Fig. 1 respectively.

Table 2: Four testing conditions of sensor selection and estimation for randomized matrices

| Name          | Sensor selection | Estimation                                                                 |
|---------------|------------------|---------------------------------------------------------------------------|
| LSS-BE        | arg max det(CCᵀ) | \[ \hat{z} = (CᵀR⁻¹C + Q⁻¹)⁻¹CᵀR⁻¹y \] [Eq. 27]                       |
|               | \[ p \leq r \]   | [Alg. 1]                                                                  |
|               | arg max det(CᵀC) | \[ p > r \]                                                                |
| NRS-LSE       | arg max det(CᵀR⁻¹C + Q⁻¹) | \[ \hat{z} = Cᵀ(CᵀC)⁻¹Cᵀy \] \[ p \leq r \] [Eq. 8]                     |
|               | \[ p \leq r \]   | [Alg. 2]                                                                  |
|               | arg max det(CᵀC) | \[ p > r \]                                                                |
| LSS-LSE (DG)  | arg max det(CCᵀ) | \[ \hat{z} = (CᵀR⁻¹C + Q⁻¹)⁻¹CᵀR⁻¹y \] [Eq. 27]                       |
|               | \[ p \leq r \]   | [Alg. 1]                                                                  |
|               | arg max det(CᵀC) | \[ p > r \]                                                                |
| NRS-BE (BDG)  | arg max det(CᵀR⁻¹C + Q⁻¹) | \[ \hat{z} = (CᵀR⁻¹C + Q⁻¹)⁻¹CᵀR⁻¹y \] [Eq. 27]                     |
|               | \[ p \leq r \]   | [Alg. 2]                                                                  |
|               | arg max det(CᵀC) | \[ p > r \]                                                                |

Figure 1: Comparison of error on randomized matrix

Figure 2 which shows the results of the reconstruction errors illustrates that
the Bayesian estimation plays an important role. Even if sensors of LSS receive some level of noise, the inverse of $R$ in Eq. 27 gives us the noise robust weights (regularization) for signals from sensors based on the prior information of the noise at sensor points.

On the other hand, the error of LSE increases around $p = 10 = r_1$ regardless of choice of the LSS or NRS. This corresponds to the fact that the error becomes larger when the problem becomes an overdetermined problem from an underdetermined problem with increasing $p$. The components in the smallest-sensitivity direction should also be estimated which includes the larger error due to the noise when $p \approx r$ as noted before, while the components in the smallest-sensitivity direction are assumed to be zero in the pseudo-inverse operation and such an error is suppressed when $p < r$.

The error of NRS-LSE is the largest in the range of $6 < p < 18$ because the sensors of NRS are chosen with assuming the regularization of the Bayesian estimation, and therefore the smallest-sensitivity direction components of least square estimation are not accurately predicted in NRS-LSE than that in LSS-LSE. However, NRS-LSE works better than LSS-LSE in the condition of $p > 18$. This is because NRS-LSE choose the sensor which is not contaminated by correlated noise and such sensors work better even in the least square estimation when the
Finally, a combination of NRS-BE always works better than that of LSS-BE. This illustrates that NRS appropriately choose the sensor positions for the Bayesian estimation.

### 3.2 Sensors for flow around airfoil

The particle image velocimetry (PIV) for acquiring time-resolved data of velocity fields around an airfoil was conducted previously [10]. The effectiveness of the present method for the PIV data is demonstrated hereafter. The test conditions are listed in Table 3 and the fluctuation of the freestream direction velocity obtained by PIV is only employed, unlike the previous study in which the two components are simultaneously treated [7]. Figure 3 shows clearly that the present method selects different points as sensors from those selected by the previous method.

| Laser                  | Double pulse lasers                  |
|------------------------|--------------------------------------|
| Time between pulse     | 100 μs                               |
| Sampling rate          | 5000 Hz                              |
| Particle image resolution | 1024 × 1024 pixels                |
| Total number of image pairs | 9700                              |

The sensors of the previous method tend to be concentrated in small parts of highly fluctuated regions, whereas those of the presented method are located evenly in the recirculation region. Comparison of the reconstruction error illustrates that the proposed method is effective and its error approaches to the lower limit of error estimated by full observation. Figure 5 shows that the sum of squared singular value of the data matrix. The behaviour of the line slowly approaches to unity represents the data characteristic that a large part of higher POD components is regarded as noise. In spite of large amount of correlated noise of modes higher than \( r_1 \), spread sensors in Fig. 3 robustly work for the reconstruction.

### 3.3 Sensors for sea surface temperature distribution

The sea surface temperature (SST) data are given at the NOAA website [9]. The data formatted by Manohar et al. is adopted in the present study. The details of the data are listed in Table 4.

---

17
Figure 3: Sensors for flow around airfoil

Figure 4: Comparison of reconstruction error on PIV data

Figure 5: Power of singular values of PIV data
Figure 6: Sensors for sea surface temperature distribution

Figure 7: Comparison of reconstruction error

Figure 8: Power of singular value of SST data
Table 4: SST data conditions

| Brief Description | NOAA Optimum Interpolation (OI) SST V2 |
|-------------------|----------------------------------------|
| Temporal Coverage | Weekly means from 1989/12/31 to 2005/12/04 |
| Spatial Coverage  | 1.0 degree latitude x 1.0 degree longitude global grid |

Figure 6 shows the difference in sensor positions as cross mark and backgrounds are root mean squares (RMS) of data used in the procedure. As was conducted in the Section 3.2, ten modes are used for the reconstruction of state, and its results are shown in Fig. 7. Those figures show that the spread sensors are so scattered that effects of the correlated noises due to truncated POD modes could be minimized. The quick increase in Fig. 8 which shows that the sum of squared singular value of the data matrix illustrates that the first few modes are dominant in the whole state. This represents that first ten modes dominates the large part of SST data, and, therefore, the noise of higher modes are relatively small. This is the reason why the superiority of the proposed method is not very significant in this problem compared with that in the PIV data problem.

3.4 Results of \( r_2 \)-truncation

The effect of \( r_2 \) truncation for an efficient memory implementation is illustrated in this section. This idea is investigated by applying the algorithm to the SST data introduced in Sec. 3.3 because \( U \) in this data set has enormous size and it requires considerable time to calculate the covariance of \( R \) if all the components are used. The variation of the reconstruction error and the calculation time are shown in Figs. 9 and 10. Those errors are calculated for five cases, \( r_2 = 10, 20, 50, 200, 832 \). Especially, \( r_2 = 10 \) and \( r_2 = 832 \) correspond to the case ignoring all the non-diagonal terms of \( R_k \), and the case using all the components of \( U_{(r_1+1):m} \) and \( S_{(r_1+1):m} \) for the calculation of the covariance, respectively. As expected, these plots show that more accurate estimation is realized when higher modes are used for calculation of covariance, but it indeed needs longer calculation time. A reasonable \( r_2 \) which leads to both less calculation time and less reconstruction error appears to be around \( r_2 = 50 \), only 6% of the original modes. It seems reasonable to ignore modes higher than 40 in this problem because the curve of Fig. 8 stops increasing there. The criteria of reasonable \( r_2 \) parameter in the generalized problem should be addressed in the future research.
4 Conclusions

The noise-robust greedy sparse-sensor-selection method based on the determinant maximization is proposed. The proposed method selects sensors by maximizing the determinant of the matrix which corresponds to the inverse matrix appearing in the Bayesian estimation operator, then estimates the state with prior information such as the covariance of noise and that of the POD amplitudes. Those two procedures result in the noise robust estimation against the noise of truncated POD modes. In addition, some improvements for the computational efficiency are introduced. The effectiveness of this algorithm on data-sets related to randomized matrix, global climate dataset, visualized velocity field around airfoil is demonstrated by comparing with the results by the other algorithms in this study. Although the proposed method takes longer time to select the sensor location, the estimation results become much stable than those of the previously-presented determinant-based greedy method. It should be noted that the locations of sensors determined by the proposed method are scattered compared to those of the previous method when the proposed algorithm is applied to the latter two cases. This is because the proposed method avoid the very close sensors that are often contaminated by the correlated noise due to the truncated high-order POD modes.
Acknowledgements

The fourth author T.N. is grateful for support of the grant JPMJPR1678 of JST Presto.

References

[1] K. Taira, S. L. Brunton, S. T. Dawson, C. W. Rowley, T. Colonius, B. J. McK-eon, O. T. Schmidt, S. Gordeyev, V. Theofilis, and L. S. Ukeiley, “Modal analysis of fluid flows: An overview,” *AIAA Journal*, pp. 4013–4041, 2017.

[2] K. Manohar, B. W. Brunton, J. N. Kutz, and S. L. Brunton, “Data-driven sparse sensor placement for reconstruction: Demonstrating the benefits of exploiting known patterns,” *IEEE Control Systems Magazine*, vol. 38, no. 3, pp. 63–86, June 2018.

[3] S. Joshi and S. Boyd, “Sensor selection via convex optimization,” *IEEE Transactions on Signal Processing*, vol. 57, no. 2, pp. 451–462, 2009.

[4] S. Chaturantabut and D. C. Sorensen, “Nonlinear model reduction via discrete empirical interpolation,” *SIAM Journal on Scientific Computing*, vol. 32, no. 5, pp. 2737–2764, 2010.

[5] Z. Drmac and S. Gugercin, “A new selection operator for the discrete empirical interpolation method—improved a priori error bound and extensions,” *SIAM Journal on Scientific Computing*, vol. 38, no. 2, pp. A631–A648, 2016.

[6] C. W. Rowley, T. Colonius, and R. M. Murray, “Model reduction for compressible flows using pod and galerkin projection,” *Physica D: Nonlinear Phenomena*, vol. 189, no. 1-2, pp. 115–129, 2004.

[7] Y. Saito, T. Nonomura, K. Nankai, K. Asai, Y. Sasaki, and D. Tsubakino, “Data-driven vector-measurement-sensor selection based on greedy algorithm for particle-image-velocimetry measurement,” *arXiv preprint arXiv:1906.00778*, 2019.

[8] Y. Saito, K. Yamada, T. Nonomura, K. Asai, D. Tsubakino, and Y. Sasaki, “Determinant based fast greedy optimization on sparse sensor selection,” *arXiv preprint arXiv:1911.08757*, 2019.
[9] NOAA/OAR/ESRL, “Noaa optimal interpolation (oi) sea surface temperature (sst) v2,” Boulder CO, July 2019. [Online]. Available: https://www.esrl.noaa.gov/psd/

[10] K. Nankai, Y. Ozawa, T. Nonomura, and K. Asai, “Linear reduced-order model based on piv data of flow field around airfoil,” TRANSACTIONS OF THE JAPAN SOCIETY FOR AERONAUTICAL AND SPACE SCIENCES, vol. 62, no. 4, pp. 227–235, 2019.
Appendix

A Woodbury’s identity

Given appropriate matrices \( A, B, C \) and \( D \), Woodbury’s identity is defined as below:

\[
(A + BDC)^{-1} = A^{-1} - A^{-1}B(D^{-1} + CA^{-1}B)^{-1}CA^{-1}
\]  \( (47) \)

If \( A \) and \( D \) are selected to be identity matrix, the identity transforms:

\[
(I + BC)^{-1} = I - B(I + CB)^{-1}C
\]  \( (48) \)

Further, if \( B \) is chosen, \( C \) to be column vector and row vector respectively, the simpler result is obtained because \( (I + CB)^{-1} \) becomes scalar value.

B Matrix determinant lemma

Suppose \( A \) to be regular matrix and \( u,v \) to be column vectors, then

\[
\det(A + uv^T) = (1 + v^TA^{-1}u) \det A
\]  \( (49) \)

Generalization is formulated for matrix \( A, U, V \) and \( W \), each of which is \( n \) by \( n \), \( n \) by \( m \), \( n \) by \( m \) and \( m \) by \( m \) matrix:

\[
\det(A + UWV^T) = \det(W^{-1} + V^TA^{-1}U) \det W \det A
\]  \( (50) \)

Of course, set \( W \) to be identity matrix of \( m \) by \( m \), formulation above is written much simpler.