Data-Driven Sensor Selection Method Based on Proximal Optimization for High-Dimensional Data With Correlated Measurement Noise

Takayuki Nagata, *nonmember, IEEE*, Keigo Yamada, *nonmember, IEEE*, Taku Nonomura, *nonmember, IEEE*, Kumi Nakai, *nonmember, IEEE*, Yuji Saito, *nonmember, IEEE*, Shunsuke Ono, *member, IEEE*

Abstract—The present paper proposes a data-driven sensor selection method for a high-dimensional nondynamical system with strongly correlated measurement noise. The proposed method is based on proximal optimization and determines sensor locations by minimizing the trace of the inverse of the Fisher information matrix under a block-sparsity hard constraint. The proposed method can avoid the difficulty of sensor selection with strongly correlated measurement noise, in which the possible sensor locations must be known in advance for calculating the precision matrix for selecting sensor locations. The problem can be efficiently solved by the alternating direction method of multipliers, and the computational complexity of the proposed method is proportional to the number of potential sensor locations when it is used in combination with a low-rank expression of the measurement noise model. The advantage of the proposed method over existing sensor selection methods is demonstrated through experiments using artificial and real datasets.

Index Terms—Alternating direction method of multipliers, optimal design of experiment, sensor selection, sparse observation

I. INTRODUCTION

MEASUREMENTS of physical quantities are essential in various fields. Typically, measurements are made at specific points using discretely installed sensors. Therefore, it is necessary to maximize the information obtained with as few sensors as possible. This kind of situation can be seen in various types of measurements, such as global positioning system [1], [2], acoustic measurements [3], [4], structural health monitoring [5], [6], environment monitoring [7], brain source localization [8], etc.

The problem that optimizes sensor locations is called the sensor selection/placement problem and is formulated as a combinatorial optimization problem, which is known as an NP-hard problem. Table. I shows typical methods used in a sensor selection problem. Global optimization techniques, such as branch and bound [9], [10], can obtain an exact solution to the sensor selection problem. However, because their computational costs are high, global optimization techniques

T. Nagata is with Tohoku University, Sendai, 980-8579, Japan (e-mail: nagata@tohoku.ac.jp), K. Yamada is with Tohoku University, Sendai, 980-8579, Japan (e-mail:keigo.yamada.t5@dc.tohoku.ac.jp), T. Nonomura is with Tohoku University, Sendai, 980-8579, Japan (e-mail: nonomura@tohoku.ac.jp), K. Nakai is with Tohoku University, Sendai, 980-8579, Japan (e-mail: kumi.nakai@tohoku.ac.jp), Y. Saito is with Tohoku University, Sendai, 980-8579, Japan (e-mail: yuji.saito@tohoku.ac.jp), S. Ono is with the Department of Computer Science, Tokyo Institute of Technology, Tokyo, JAPAN (ono@c.titech.ac.jp).

can only be used for small (choosing a small number of sensors from a small number of potential sensor locations) problems. Therefore, convex relaxation and greedy methods, which can obtain a suboptimal solution with less computational time, have been studied.

Joshi and Boyd proposed [11] a sensor optimization method based on convex optimization. Their method can obtain a global optimal solution to the relaxed problem. The computational complexity of their method is proportional to the cube of the number of potential sensor locations, and thus, the computational cost is much lower than that of global optimization techniques. The accelerated randomized convex relaxation method was recently proposed [12], but the computational cost is still high for high-dimensional data, such as surface measurement or volume measurement data. A convex optimization method based on a proximal splitting algorithm is another sensor selection method. The sparsity-promoting framework was introduced by Fardad et al. [13] and Lin et al. [14]. Their framework allows us to obtain block-sparse feedback and observer gains, as well as select actuators and sensors in dynamical systems, and Dhingra et al. [15] and Zare and Jovanović [16] extended those methods. Masazade et al. [17] proposed the method for selecting sparse sensors that minimize the estimation using an extended Kalman filter. Their method obtains the sparse Kalman gain matrix using the sparsity-promoting penalty function. In addition, sensor selection methods for a field estimation on a nondynamical system based on convex or nonconvex proximal optimization are available [18], [19].

Greedy methods provide a suboptimal solution to sensor selection problems at a low computational cost. Manohar et al. [20] proposed a greedy method based on QR decomposition. Their method is related to the discrete empirical interpolation method [21] and the QR-based discrete empirical interpolation method [22] in the framework of the Galerkin projection [23]. Saito et al. [24] proposed a greedy method based on the D-optimal design of experiments. They also constructed a sensor selection framework for multicomponent measurements [25], [26]. Nakai et al. [27] formulated a greedy method based on the E- and A-optimal design of experiments and investigated the influence of the objective function on the performance of the selected sensor subset. Greedy methods are extended to a further generalized form [25]–[37]. However, it is difficult to apply the greedy method to complex objective functions and constraints.

TABLE I
CHARACTERISTICS OF PRACTICAL ALGORITHMS FOR SENSOR SELECTION PROBLEMS.

Method	Global optimization	Convex relaxation	Greedy method
Solution	Exact	Global optimum	Local optimum
Cost	Very expensive	Expensive	Cheap
Obj. func.	-	Flexible	Inflexible

TABLE II
VARIABLES AND OPERATORS

Symbol	Description
$\mathbf{C}$	Measurement matrix
$\mathbf{H}$	Sensor location matrix
$\mathbf{I}$	Identity matrix
$\mathbf{K}, \mathbf{X}$	Gain matrix
$\mathbf{R}, \mathbf{Q}$	Noise covariance matrix
$\mathbf{R}_\mathrm{d}$	Diagonal matrix with the diagonal entries of $\mathbf{R}$
$\mathbf{U}$	Sensor candidate matrix
$\mathbf{x}$	Snapshot of full state
$\mathbf{y}$	Observation vector
$\mathbf{z}$	Latent variables

Symbol	Description
$\mathbf{A}^\dagger$	Moore–Penrose pseudo inverse of $\mathbf{A}$
$\hat{\mathbf{A}}$	Normalized $\mathbf{A}$ by noise weighting term $\mathbf{R}_\mathrm{d}^{-\frac{1}{2}}$, i.e., $\hat{\mathbf{A}} = \mathbf{R}_\mathrm{d}^{-\frac{1}{2}}\mathbf{A}$
$\mathrm{tr}(\mathbf{A})$	Trace norm of $\mathbf{A}$, i.e., $\mathrm{tr}(\mathbf{A}) = \Sigma_i \mathrm{diag}_i(\mathbf{A})$
$\mathbf{A}^\mathsf{T}$	Transposition
$\mathbf{A}_{b:c}$	Truncated matrix of $\mathbf{A}$ from b-th column/row to c-th column/row
$\|\mathbf{A}\|_\mathrm{F}$	Frobenius norm of $\mathbf{A}$, i.e., $\|\mathbf{A}\|_\mathrm{F} = \sqrt{\Sigma_i \Sigma_j \|a_{ij}\|^2}$
$\|\mathbf{A}\|_0$	Group ℓ_0 pseudo-norm $\mathbf{A}$, i.e., $\|\mathbf{A}\|_{g,0} := \|(\|\mathbf{a}_1\|_2, \cdots, \|\mathbf{a}_n\|_2)\|_0$
$\|\mathbf{a}\|_1$	Manhattan (ℓ_1) norm of $\mathbf{a}$, i.e., $\|\mathbf{a}\|_1 = \Sigma_i |a_i|$
$\|\mathbf{a}\|_2$	Euclidean (ℓ_2) norm of $\mathbf{a}$, i.e., $\|\mathbf{a}\|_2 = \sqrt{\Sigma_i \mathbf{a}_i^2}$
$\mathbf{a}_i$	i-th row or column vector of $\mathbf{A}$
$\bar{\mathbf{a}}$	Estimated $\mathbf{a}$

In most cases, including the methods mentioned above, the observed noise is assumed to be white noise, but the measured data are often corrupted by correlated noise. Therefore, the development of a sensor selection method for correlated measurements is a critical task. Several researchers have studied sensor selection in the presence of correlated measurement noise. Liu et al. [38] and Yamada et al. [30] modified the formulations of the sensor selection problem by weighting the value of each sensor based on prior noise information. They employed a Bayesian estimation framework and also used prior information in the estimation. Ucinski et al. [39] and O'Connor et al. [40] employed a similar formulation, which includes a weighting term based on prior noise information. In their framework, an ordinary least-squares estimation was used. The difficulty of treating correlated noise in the sensor selection problem is that the influence of noise cannot be evaluated unless the sensor location is determined. The previous methods considering the correlated measurement noise require the calculation of the precision matrix (inverse of the noise covariance matrix) which should be constructed by the noise information over only selected sensors. Particularly, the method based on convex relaxation requires higher calculation cost or a complicated formulation with limitations. There is no published method based on the continuous optimization problem for sensor selection that can be solved by a method with a computational complexity less than a cubic order of the number of potential sensor locations. Hence, it is difficult to employ the previously proposed methods for large-scale sensor selection problems including such correlated measurement noise.

In the present study, we consider a sensor selection problem involving selecting p sensors from n potential sensors in a high-dimensional nondynamical system with correlated measurement noise. The sensor subset gives an observation vector of a linear function of latent variables superimposed with spatially correlated measurement noise. We extended the sensor selection method based on the proximal splitting algorithm [19] to a problem including correlated measurement noise. Our goal is to choose a suboptimal sensor subset based on the optimal design of experiments considering the influence of correlated measurement noise. The main contributions of the present study are as follows:

- Sensor selection based on the proximal splitting algorithm for nondynamical systems with strongly correlated measurement noise is proposed.
- The proposed method can avoid the difficulty (described in Section II-D) of sensor selection with strongly correlated measurement noise in which the sensor locations must be known in advance for selecting the sensor locations. The problem can be efficiently solved by the alternating direction method of multipliers (ADMM) [41], [42], and the computational complexity of the proposed method is $O(n)$ using a low-rank expression of the measurement noise model [30], [31].
- The proposed method is applied to artificial and practical datasets, the National Oceanic and Atmospheric Administration Optimum Interpolation Sea Surface Temperature (NOAA-OISST) dataset. The optimization results are compared with those of the greedy method.
- Although the computational time of the proposed method is longer than that of the greedy method, the proposed method is better than the previously proposed greedy methods in terms of objective value and reconstruction error.

We laid the groundwork for further extension to complex objective functions, including constraints. MATLAB code for our algorithm can be found at https://github.com/Aerodynamics-Lab/Proximal-Optimization-based-Sensor-Selection-Algorithm-with-Correlated-Measurement-Noise [43].

II. SENSOR SELECTION PROBLEM

A mathematical description of the problem and the proposed method is shown in this section. The notation table that summarizes the main mathematical symbols and their definitions is shown in Table II.

A. Problem Formulation

A snapshot measurement of the full state $\mathbf{x} \in \mathbb{R}^n$ including a noise signal $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}) \in \mathbb{R}^p$ through sparse sensors can be expressed as follows:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v}, \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^p$ and $\mathbf{H} \in \mathbb{R}^{p \times n}$ are the observation vector and the sensor location matrix, respectively. Here, n and p are the numbers of potential sensors and activated sensors, respectively. The sensor location matrix $\mathbf{H}$ is a sparse matrix, and each row vector of $\mathbf{H}$ is a unit vector. Unity locations in the row vectors correspond to the activated sensor locations chosen from n potential sensor locations. Equation (1) can be rewritten as follows using the sensor candidate matrix $\mathbf{U} \in \mathbb{R}^{n \times r}$ and the latent variables $\mathbf{z} \in \mathbb{R}^r$:

$$\mathbf{y} \approx \mathbf{HUz} + \mathbf{v}$$
$$= \mathbf{Cz} + \mathbf{v}. \qquad (2)$$

This system represents the problem of choosing p sensors for the observation of signals generated by r latent variables. The matrix $\mathbf{C} \in \mathbb{R}^{p \times r}$ is the measurement matrix, which is the product of the sensor location matrix and the sensor candidate matrix. Because n corresponds to the degrees of freedom in the spatial direction of the full state, the sensor candidate matrix $\mathbf{U}$ is usually a tall-and-skinny matrix (i.e., $n \gg r$) in a practical problem. The latent variables at $p > r$ can be estimated using the least-squares method using a pseudo-inverse operation,

$$\tilde{\mathbf{z}} = \mathbf{C}^\dagger \mathbf{y}$$
$$= \left(\mathbf{C}^\mathsf{T}\mathbf{C}\right)^{-1} \mathbf{C}^\mathsf{T}\mathbf{y}, \qquad (3)$$

where $\tilde{\mathbf{z}}$ is the set of estimated latent variables. The formulation in $p < r$ is not introduced due to the constraint used in the present formulation. Descriptions of the estimation of latent variables in $p < r$ are presented in [31].

In the case of measured signal including correlated measurement noise $\mathrm{E}\left[\mathbf{v}_\mathrm{c}\mathbf{v}_\mathrm{c}^\mathsf{T}\right] = \mathbf{R}_p$, "whitening" of the measurement $\mathbf{y}$ is required before conducting the least-squares estimation as follows [31]:

$$\mathbf{R}_p^{-\frac{1}{2}}\mathbf{y} = \mathbf{R}_p^{-\frac{1}{2}}\mathbf{Cz} + \mathbf{R}_p^{-\frac{1}{2}}\mathbf{v}_\mathrm{c}$$
$$\Rightarrow \quad \mathbf{y}_\mathrm{n} = \mathbf{C}_\mathrm{n}\mathbf{z} + \mathbf{v}_\mathrm{n}, \qquad (4)$$

where subscript $\circ_\mathrm{n}$ indicates the values corrected by the noise weighting term of the activated sensors $\mathbf{R}_p^{-\frac{1}{2}} \in \mathbb{R}^{p \times p}$, and $\mathbf{v}_\mathrm{n}$ is white noise. Therefore, $\mathrm{E}\left[\mathbf{v}_\mathrm{n}\mathbf{v}_\mathrm{n}^\mathsf{T}\right] = \sigma$ is a constant, and thus, the least-squares estimation of $\tilde{\mathbf{z}}$ with the whitened measurement vector $\mathbf{y}_\mathrm{n}$ at $p > r$ is obtained as follows:

$$\tilde{\mathbf{z}} = \mathbf{C}_\mathrm{n}^\dagger \mathbf{y}_\mathrm{n} = \left(\mathbf{C}_\mathrm{n}^\mathsf{T}\mathbf{C}_\mathrm{n}\right)^{-1} \mathbf{C}_\mathrm{n}^\mathsf{T}\mathbf{y}_\mathrm{n}$$
$$= \left(\mathbf{C}^\mathsf{T}\mathbf{R}_p^{-1}\mathbf{C}\right)^{-1} \mathbf{C}^\mathsf{T}\mathbf{R}_p^{-1}\mathbf{y}. \qquad (5)$$

B. A-optimality Criterion

A-optimality criterion is the trace of the inverse of the Fisher information matrix (FIM). The A-optimal design minimizes the mean squared error in the estimation of the latent variables $\tilde{\mathbf{z}}$. Here, we consider the correlated measurement noise $\mathbf{v}_\mathrm{c}$ in the observed signal of $\mathbf{x}$.

$$\tilde{\mathbf{z}} = \left(\mathbf{C}^\mathsf{T}\mathbf{R}_p^{-1}\mathbf{C}\right)^{-1} \mathbf{C}^\mathsf{T}\mathbf{R}_p^{-1}\mathbf{y}$$
$$= \left(\mathbf{C}^\mathsf{T}\mathbf{R}_p^{-1}\mathbf{C}\right)^{-1} \mathbf{C}^\mathsf{T}\mathbf{R}_p^{-1}\mathbf{Cz} + \left(\mathbf{C}^\mathsf{T}\mathbf{R}_p^{-1}\mathbf{C}\right)^{-1} \mathbf{C}^\mathsf{T}\mathbf{R}_p^{-1}\mathbf{v}_\mathrm{c} \qquad (6)$$

Therefore, the estimation error can be computed as follows:

$$\tilde{\mathbf{z}} - \mathbf{z} = \left(\mathbf{C}^\mathsf{T}\mathbf{R}_p^{-1}\mathbf{C}\right)^{-1} \mathbf{C}^\mathsf{T}\mathbf{R}_p^{-1}\mathbf{v}_\mathrm{c}. \qquad (7)$$

To consider the average variance of the estimation, the trace norm of the error covariance matrix is calculated as follows:

$$\mathrm{tr}\left(\mathrm{E}\left[\left(\tilde{\mathbf{z}} - \mathbf{z}\right)\left(\tilde{\mathbf{z}} - \mathbf{z}\right)^\mathsf{T}\right]\right)$$
$$= \mathrm{tr}\left(\mathrm{E}\left[\left(\mathbf{C}^\mathsf{T}\mathbf{R}_p^{-1}\mathbf{C}\right)^{-1} \mathbf{C}^\mathsf{T}\mathbf{R}_p^{-1}\mathbf{v}_\mathrm{c}\mathbf{v}_\mathrm{c}^\mathsf{T}\mathbf{R}_p^{-1}\mathbf{C}\left(\mathbf{C}^\mathsf{T}\mathbf{R}_p^{-1}\mathbf{C}\right)^{-1}\right]\right). \qquad (8)$$

Thus, the sensor selection problem based on the A-optimality criterion can be expressed as a minimization problem of a following objective function:

$$f = \mathrm{tr}\left(\left(\mathbf{C}^\mathsf{T}\mathbf{R}_p^{-1}\mathbf{C}\right)^{-1}\right). \qquad (p \geq r) \qquad (9)$$

C. Noise Covariance

There is a situation in which the data include not only Gaussian noise but also correlated measurement noise. In data-driven sensor selection, the reduced-order modeling of a data matrix, which is given as the r-rank approximation [44] obtained by the singular value decomposition (SVD) $\mathbf{X}_\mathrm{data} \approx \mathbf{U}_{1:r_1}\mathbf{S}_{1:r_1}\mathbf{V}_{1:r_1}^\mathsf{T}$, is used. Here, $\mathbf{X}_\mathrm{data} \in \mathbb{R}^{n \times m}$ is training data matrix containing snapshots $\mathbf{X}_\mathrm{data} = [\mathbf{x_1} \cdots \mathbf{x_m}]$. The measurement noise consists of the truncated modes of from $(r_1 + 1)$th to mth modes because those components are included in the measurement but not in the reduced-order model. In the present study, the spatial covariance derived from the truncated SVD modes of the training data matrix is used as the noise model, as in previous work based on the greedy method of Yamada et al. [30], [31]. The snapshot with full-state observation $\mathbf{x}$ and noise covariance matrix $\mathbf{R} \in \mathbb{R}^{n \times n}$ become

$$\mathbf{x} = \mathbf{Uz} + \mathbf{w}, \qquad (10)$$

where $\mathbf{w}$ is the correlated noise on the full-state observation and its covarionce is given by $\mathbf{R}$ as follows:

$$\mathrm{E}\left[\mathbf{ww}^\mathsf{T}\right] =: \mathbf{R}. \qquad (11)$$

The observations through the sensors and noise covariance are

$$\mathbf{y} = \mathbf{HUz} + \mathbf{Hw}, \qquad (12)$$

and

$$\mathrm{E}\left[\mathbf{Hww}^\mathsf{T}\mathbf{H}^\mathsf{T}\right] = \mathbf{H}\mathrm{E}\left[\mathbf{ww}^\mathsf{T}\right]\mathbf{H}^\mathsf{T}$$
$$= \mathbf{HRH}^\mathsf{T} =: \mathbf{R}_p, \qquad (13)$$

respectively. The matrix $\mathbf{R}_p \in \mathbb{R}^{p \times p}$ is the noise covariance matrix for selected sensors. The noise covariance is assumed to be represented by the truncated SVD modes from $r_1 + 1$ to m of the training data matrix as follows:

$$\mathbf{R} = \mathrm{E}\left(\mathbf{ww}^\mathsf{T}\right) \qquad (14)$$
$$= \mathrm{E}\left(\left(\mathbf{x} - \mathbf{U}_{1:r_1}\mathbf{z}\right)\left(\mathbf{x} - \mathbf{U}_{1:r_1}\mathbf{z}\right)^\mathsf{T}\right)$$
$$\approx \left(\mathbf{USV}^\mathsf{T} - \mathbf{U}_{1:r_1}\mathbf{S}_{1:r_1}\mathbf{V}_{1:r_1}^\mathsf{T}\right)\left(\mathbf{USV}^\mathsf{T} - \mathbf{U}_{1:r_1}\mathbf{S}_{1:r_1}\mathbf{V}_{1:r_1}^\mathsf{T}\right)^\mathsf{T} \qquad (15)$$
$$= \left(\mathbf{U}_{(r_1+1):m}\mathbf{S}_{(r_1+1):m}\mathbf{V}_{(r_1+1):m}^\mathsf{T}\right)\left(\mathbf{U}_{(r_1+1):m}\mathbf{S}_{(r_1+1):m}\mathbf{V}_{(r_1+1):m}^\mathsf{T}\right)^\mathsf{T}$$
$$= \mathbf{U}_{(r_1+1):m}\mathbf{S}_{(r_1+1):m}^2\mathbf{U}_{(r_1+1):m}^\mathsf{T}. \qquad (16)$$

When higher-order modes can be ignored because their contributions are small, (16) can be truncated, and a further low-dimensional form of the noise covariance matrix can be obtained as

$$\mathbf{R} \approx \mathbf{U}_{(r_1+1):r_2} \mathbf{S}^2_{(r_1+1):r_2} \mathbf{U}^\mathsf{T}_{(r_1+1):r_2}, \tag{17}$$

where $r_2 \leq m$. In the present study, the diagonal matrix with the difference between the original noise covariance matrix and the low-rank noise covariance matrix $\Delta \mathbf{S} \in \mathbb{R}^{n \times n}$ is computed as follows:

$$\Delta \mathbf{S} = \mathrm{diag}\left(\mathrm{diag}\left(\mathbf{R}\right) - \mathrm{diag}\left(\mathbf{U}_{(r_1+1):r_2} \mathbf{S}^2_{(r_1+1):r_2} \mathbf{U}^\mathsf{T}_{(r_1+1):r_2}\right)\right). \tag{18}$$

By adding this diagonal matrix, the approximation accuracy of the noise covariance matrix was improved.

$$\mathbf{R} \approx \mathbf{U}_{(r_1+1):r_2} \mathbf{S}^2_{(r_1+1):r_2} \mathbf{U}^\mathsf{T}_{(r_1+1):r_2} + \Delta \mathbf{S} \tag{19}$$

D. Difficulty in Treatment of Correlated Noise in Sensor Selection

The difficulty of treating correlated noise in sensor selection problems is that the influence of noise cannot be evaluated unless the sensor location is determined. The previous methods considering the correlated measurement noise require the calculation of the precision matrix (inverse of the noise covariance matrix) which should be constructed by the noise information over only selected sensors. As shown in (9) and (13), the matrix $\mathbf{R}_p$ constructed as $\mathbf{HRH}^\mathsf{T}$ is required for evaluation of the objective function when selecting the sensor locations $\mathbf{H}$. Thus, the sensor location matrix $\mathbf{H}$ is required for determining the sensor locations. Therefore, consideration of the correlated measurement noise in the conventional convex relaxation method [11] is difficult. Previous studies [45]–[47] considered weakly correlated noise and approximated the FIM in the following form:

$$\mathbf{C}^\mathsf{T} \mathbf{R}_p^{-1} \mathbf{C} \approx \mathbf{F} = \mathbf{U}^\mathsf{T} \left(\mathbf{ss}^\mathsf{T} \circ \mathbf{R}^{-1}\right) \mathbf{U}, \tag{20}$$

where the variable $\mathbf{s} \in \mathbb{R}^n$ is the solution vector for the convex relaxation method with constraints $\mathbf{1}^\mathsf{T}\mathbf{s} = p$ and $s_i \in (0, 1)$, which correspond to the weights for the sensor candidates. The problem is solved by semidefinite programming. The inverse of the noise covariance matrix (the precision matrix for observations) in this FIM is clearly affected by the noise information for the sensors that are not selected when taking the inverse of $\mathbf{R}$. This is because all the noise information can be observed, even though the signal can only be observed by the activated sensors in the system corresponding to this FIM. Therefore, this formulation can only be used for weakly correlated noise, as mentioned in [38].

Liu et al. [38] avoided this difficulty by splitting the noise covariance matrix into a diagonal matrix $a\mathbf{I}$ and a matrix $\mathbf{B}$ as follows:

$$\mathbf{R} = a\mathbf{I} + \mathbf{B}, \tag{21}$$

where a is chosen such that the matrix $\mathbf{B}$ is positive definite. In this case, the noise covariance matrix for selected sensors $\mathbf{R}_p$ can be obtained as follows:

$$\mathbf{R}_p = \mathbf{H}\left(a\mathbf{I} + \mathbf{B}\right)\mathbf{H}^\mathsf{T}$$
$$= a\mathbf{I} + \mathbf{HBH}^\mathsf{T}, \tag{22}$$

and the approximated FIM $\mathbf{F}$ (20) becomes

$$\mathbf{F} = \mathbf{U}^\mathsf{T} \mathbf{H}^\mathsf{T} \mathbf{R}^{-1} \mathbf{H} \mathbf{U}$$
$$= \mathbf{U}^\mathsf{T} \mathbf{B}^{-1} \mathbf{U} - \mathbf{U}^\mathsf{T} \mathbf{B}^{-1} \left(\mathbf{B}^{-1} + a^{-1}\mathrm{diag}\left(\mathbf{s}\right)\right)^{-1} \mathbf{B}^{-1} \mathbf{U}, \tag{23}$$

where again the variable $\mathbf{s} \in \mathbb{R}^n$ is the solution vector for the convex relaxation method with constraints $\mathbf{1}^\mathsf{T}\mathbf{s} = p$ and $s_i \in (0, 1)$, which corresponds to the weights of the sensor candidates. This operation of separating nondiagonal components in $\mathbf{R}$ results in an objective function whose covariance in measurement noise is properly evaluated. The objective function based on (23) can be solved by semidefinite programming. However, there is no published method to formulate the continuous optimization problem for sensor selection that can be solved by a method with a computational complexity less than $O\left(n^3\right)$. We will introduce a formulation that can consider correlated noise without computing the precision matrix and that can solve the problem with a computational cost of $O\left(n\right)$ in the following section.

III. PROPOSED METHOD

A. Mathematical Formulation

The mathematics of the present method is based on [19]. The difference between the previous and present methods is whether to consider independent or correlated noise. Consider a matrix $\mathbf{K}_p \in \mathbb{R}^{r_1 \times p}$ that recovers $\mathbf{z}$ from the observation $\mathbf{y}$ in (3)

$$\tilde{\mathbf{z}} = \mathbf{K}_p \mathbf{y}$$
$$= \mathbf{KUz} + \mathbf{Kw}. \tag{24}$$

The gain matrix $\mathbf{K} \in \mathbb{R}^{r_1 \times n}$ is a sparse matrix, which has p nonzero column vectors. The locations of activated sensors are indicated by that of nonzero column vectors in the gain matrix $\mathbf{K}$. As in the previous study [19], the gain matrix $\mathbf{K}$ is only used to construct the sensor matrix $\mathbf{H}$ in the polishing step (see Section III-E) and is not used directly for the estimation of $\tilde{\mathbf{z}}$. In addition, $\tilde{\mathbf{z}}$ is assumed to be an unbiased approximation of $\mathbf{z}$. In this case,

$$\mathbf{KU} = \mathbf{I} \tag{25}$$

should be satisfied. This constraint is important for the present algorithm, and the optimization problem of the objective function with several constraints should be solved. Although the faster algorithm of the proximal gradient method, such as the fast iterative shrinkage thresholding algorithm [48] potentially seems to be a candidate, it cannot be applied to problems with a nondifferentiable objective function with constraints. Thus, ADMM that can handle the constraint with proximal operator is employed, as described in III-C. The average error in $\tilde{\mathbf{z}}$ can be explained in the same way as described in (6).

$$\mathrm{tr}\left(\mathrm{E}\left[\left(\tilde{\mathbf{z}} - \mathbf{z}\right)\left(\tilde{\mathbf{z}} - \mathbf{z}\right)^\mathsf{T}\right]\right)$$
$$= \mathrm{tr}\left(\mathrm{E}\left[\mathbf{Kww}^\mathsf{T}\mathbf{K}^\mathsf{T}\right]\right) = \mathrm{tr}\left(\mathbf{KRK}^\mathsf{T}\right) \tag{26}$$

To obtain the sparse gain matrix $\mathbf{K}$, the group-sparsity paradigm [49] is introduced, and (26) is augmented with a sparsity-promoting term. When the group ℓ_1-norm penalty is

used as a sparsity-promoting term, the objective function of the sensor selection problem can be obtained as follows:

$$\underset{\mathbf{K}}{\text{minimize}} \ \text{tr}\left(\mathbf{KRK}^\mathsf{T}\right) + \lambda \sum_{i=1}^{n} \|\mathbf{k}_i\|_2 \quad \text{subject to } \mathbf{KU} = \mathbf{I}, \quad (27)$$

where $\mathbf{k}_i \in \mathbb{R}^n$ is the ith column vector of $\mathbf{K}$, and $\lambda > 0$ is the regularization parameter that adjusts the sparsity of the solution. The sparsity-promoting term penalizes a greater number of nonzero columns in $\mathbf{K}$, which corresponds to the number of selected sensors. The advantage of the present method is the flexibility of the objective function. In the case of the conventional method considering correlated noise, the covariance matrix constructed based on selected sensors, as discussed in Section II-D, is required for evaluating the objective function for sensor selection. In contrast, the present method can minimize the objective function (27) directly.

However, when sensor selection is performed using the objective function of (27), the number of sensors to be selected cannot be determined in advance due to the nature of the sparsity-promoting term. In the present study, therefore, a group ℓ_0 pseudo-norm constraint is used to obtain p sensors that determined by users in advance:

$$\underset{\mathbf{K}}{\text{minimize}} \ \text{tr}\left(\mathbf{KRK}^\mathsf{T}\right)$$
$$\text{subject to } \|(\|\mathbf{k}_1\|_2, \cdots, \|\mathbf{k}_n\|_2)\|_0 \leq p \text{ and } \mathbf{KU} = \mathbf{I}. \quad (28)$$

The operator $\|\cdot\|_0$ indicates the ℓ_0 pseudo-norm, which returns the total number of nonzero entries, and the operator $\|\cdot\|_2$ indicates the ℓ_2 norm of a vector. Although it does not lead to essential changes in the equations, a simple expression consistent with the standard problems adopted in ADMM can be obtained by taking the transpose of this formulation and the variable transformation:

$$\begin{cases} \mathbf{X} &= & \mathbf{K}^\mathsf{T} \\ \mathbf{A} &= & \mathbf{U}_{1:r_1}^\mathsf{T} \\ \Delta\mathbf{S_Q} &= & \Delta\mathbf{S} \\ \mathbf{U_Q} &= & \mathbf{U}_{(r_1+1):r_2} \\ \mathbf{S_Q} &= & \mathbf{S}_{(r_1+1):r_2} \\ \mathbf{Q} &= & \mathbf{U_Q}\mathbf{S_Q}\mathbf{U_Q}^\mathsf{T} + \Delta\mathbf{S_Q}. \end{cases} \quad (29)$$

Then, the objective function can be written as follows:

$$\underset{\mathbf{X}}{\text{minimize}} \ \text{tr}\left(\mathbf{X}^\mathsf{T}\mathbf{QX}\right)$$
$$\text{subject to } \|(\|\mathbf{x}_1\|_2, \cdots, \|\mathbf{x}_n\|_2)\|_0 \leq p \text{ and } \mathbf{AX} = \mathbf{I}, \quad (30)$$

where $\mathbf{x}_i \in \mathbb{R}^n$ is the ith row vector of $\mathbf{X}$. The matrix $\mathbf{Q}$ is the noise covariance matrix described by a low-rank expression, as given in (19).

B. Consideration of Noise Intensity in Thresholding Operations

In the present study, the sensor candidate matrix $\mathbf{U}$ was normalized according to the noise weighting term $\mathbf{R}_d^{-\frac{1}{2}}$ where $\mathbf{R}_d \in \mathbb{R}^{n \times n}$ is a diagonal matrix with diagonal entries of the noise covariance matrix $\mathbf{R}$. The normalized sensor candidate matrix $\hat{\mathbf{U}}$ is generated as follows:

$$\hat{\mathbf{U}} = \mathbf{R}_d^{-\frac{1}{2}}\mathbf{U}. \quad (31)$$

Therefore, the gain matrix $\mathbf{K}$ is

$$\hat{\mathbf{K}} = \mathbf{KR}_d^{\frac{1}{2}}, \quad (32)$$

and the objective function of (28) becomes

$$\underset{\hat{\mathbf{K}}}{\text{minimize}} \ \text{tr}\left(\hat{\mathbf{K}}\mathbf{R}_d^{-\frac{1}{2}}\mathbf{RR}_d^{-\frac{1}{2}}\hat{\mathbf{K}}^\mathsf{T}\right)$$
$$\text{subject to } \left\|\left(\|\hat{\mathbf{k}}_1\|_2, \cdots, \|\hat{\mathbf{k}}_n\|_2\right)\right\|_0 \leq p \text{ and } \hat{\mathbf{K}}\hat{\mathbf{U}} = \mathbf{I}. \quad (33)$$

The objective function, which has the same form as (30), can be obtained using the following variables:

$$\begin{cases} \mathbf{Q} &= & \Delta\mathbf{S} + \mathbf{U}_{(r_1+1):r_2}\mathbf{S}_{(r_1+1):r_2}\mathbf{U}_{(r_1+1):r_2}^\mathsf{T} \\ \mathbf{R}_d &= & \text{diag}(\mathbf{Q}) \\ \Delta\mathbf{S_Q} &= & \mathbf{R}_d^{-\frac{1}{2}}\Delta\mathbf{S}\mathbf{R}_d^{-\frac{1}{2}} \\ \mathbf{X} &= & \hat{\mathbf{K}}^\mathsf{T}(= \mathbf{KR}_d^{\frac{1}{2}}) \\ \mathbf{A} &= & \hat{\mathbf{U}}_{1:r_1}^\mathsf{T}(= \mathbf{R}_d^{-\frac{1}{2}}\mathbf{U}) \\ \mathbf{U_Q} &= & \mathbf{R}_d^{-\frac{1}{2}}\mathbf{U}_{(r_1+1):r_2} \\ \mathbf{S_Q} &= & \mathbf{S}_{(r_1+1):r_2}. \end{cases} \quad (34)$$

Then, the problem can be solved by the same procedure as described in Section III-D. It should be noted that the solution of the problem, the gain matrix $\mathbf{K}$, is obtained by denormalizing the solution matrix $\mathbf{X}$ as follows:

$$\mathbf{K} = \mathbf{X}^\mathsf{T}\mathbf{R}_d^{-\frac{1}{2}}. \quad (35)$$

Normalization is required to consider the intensity of the noise in the proximal operator used in ADMM. In the case of the objective function (28), the noise covariance matrix is considered in the first term of the objective function, but it is not considered in the sparsity-promoting term. This is because the effect of the correlated noise cannot be accurately evaluated unless the sensor locations are determined, as discussed in Section II-D. Therefore, it is difficult to consider the correlated measurement noise in the sparsity-prompting term. In the present method, the sensor candidate matrix $\mathbf{U}$ was normalized based on the diagonal entries of the noise covariance matrix, but off-diagonal entries cannot be considered. By normalization of the sensor candidate matrix based on the noise weighting term, the diagonal entries of the noise covariance matrix are considered in the computation of the sparsity-promoting term, even though the influence of the off-diagonal entries is still not considered. It should be noted that the noise considered in the present study has a structure of a certain size because the noise model consists of truncated modes. Hence, not only uncorrelated noise but also a part of the component of correlated noise will be considered by normalization using the noise weighting term, even if only the diagonal component of the noise covariance matrix is considered.

C. Alternating Direction Method of Multipliers

In the present study, the optimization problem is solved by ADMM. When the sparsity-promoting term is the simple group ℓ_1-norm penalty, (27) has the following form:

$$\underset{\mathbf{X},\mathbf{Z}}{\text{minimize}} \ g(\mathbf{X}) + h(\mathbf{Z}) \quad \text{subject to} \quad \mathbf{Z} = \mathbf{GX}. \quad (36)$$

For arbitrarily chosen $\mathbf{Z}^{(0)}$, $\mathbf{Y}^{(0)}$, and $\gamma > 0$, ADMM iterates the following steps:

$$
\begin{cases}
\mathbf{X}^{(k+1)} & = \underset{\mathbf{X}}{\arg\min} \; g\left(\mathbf{X}\right) + \frac{1}{2\gamma} \left\| \mathbf{Z}^{(k)} - \mathbf{GX} - \mathbf{Y}^{(k)} \right\|_2^2 \\
\mathbf{Z}^{(k+1)} & = \underset{\mathbf{Z}}{\arg\min} \; h\left(\mathbf{Z}\right) + \frac{1}{2\gamma} \left\| \mathbf{Z} - \mathbf{GX}^{(k+1)} - \mathbf{Y}^{(k)} \right\|_2^2 \\
\mathbf{Y}^{(k+1)} & = \mathbf{Y} + \mathbf{GX}^{(k+1)} - \mathbf{Z}^{(k+1)}.
\end{cases}
\tag{37}
$$

In the present formulation with the simple group ℓ_1-norm penalty, $g(\mathbf{X})$, $h(\mathbf{Z})$, $\mathbf{Z}$, and $\mathbf{G}$ correspond to

$$
g\left(\mathbf{X}\right) := \mathrm{tr}\left(\mathbf{X}^\mathsf{T}\mathbf{QX}\right), \; h\left(\mathbf{Z}\right) := \lambda \sum_i \|\mathbf{x}_i\|_2 , \tag{38}
$$

$$
\mathbf{Z} = \left[\begin{array}{c} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{array} \right], \text{ and } \mathbf{G} = \left[\begin{array}{c} \mathbf{I} \\ \mathbf{A} \end{array} \right],
$$

respectively.

D. Sensor Selection With Group ℓ_0 Pseudo-Norm Constraint

1) Reformulation of Objective Function (30): The first constraint in (30), which is the sparsity-promoting term, was computed by an ℓ_0-constrained block hard thresholding (L_0BHT) operator. This operator can determine the group ℓ_0 pseudo-norm of the solution matrix beforehand that corresponds to the number of sensors to be activated. To establish the algorithm for solving the problem (30), we first reformulate the objective function (30) into an ADMM-applicable form defined in (36). We also define the group ℓ_0 pseudo-norm used in L_0BHT as follows:

Definition 1 (Group ℓ_0 pseudo-norm): The activated sensor locations correspond to the locations of nonzero row vectors of $\mathbf{X}$, and thus, the group ℓ_0 pseudo-norm of $\mathbf{X}$ used in (30) as the constraint is defined as

$$
\|\mathbf{X}\|_{g,0} := \| \left(\|\mathbf{x}_1\|_2, \cdots, \|\mathbf{x}_n\|_2 \right) \|_0. \tag{39}
$$

Then, the objective function (30) becomes

$$
\underset{\mathbf{X}}{\text{minimize}} \; \mathrm{tr}\left(\mathbf{X}^\mathsf{T}\mathbf{QX}\right) \quad \text{subject to } \|\mathbf{X}\|_{g,0} \leq p \text{ and } \mathbf{AX} = \mathbf{I}. \tag{40}
$$

In addition, we define the indicator function of the inequality constraint on the group ℓ_0 pseudo-norm:

$$
l_{\{\|\cdot\|_{g,0} \leq p\}}\left(\mathbf{X}\right) := \begin{cases} 0, & \|\mathbf{X}\|_{g,0} \leq p, \\ \infty, & \text{otherwise.} \end{cases} \tag{41}
$$

Then, the objective function (40) can be reformulated as

$$
\underset{\mathbf{X}}{\text{minimize}} \; \mathrm{tr}\left(\mathbf{X}^\mathsf{T}\mathbf{QX}\right) + l_{\{\|\cdot\|_{g,0} \leq p\}}\left(\mathbf{X}\right) \quad \text{subject to } \mathbf{AX} = \mathbf{I}. \tag{42}
$$

2) Algorithm: The form of (42) is the same as that of (36)

$$
g\left(\mathbf{X}\right) := \mathrm{tr}\left(\mathbf{X}^\mathsf{T}\mathbf{QX}\right), \; h\left(\mathbf{Z}\right) := l_{\|\cdot\|_{g,0} \leq p}\left(\mathbf{Z}\right), \tag{43}
$$

$$
\mathbf{Z} = \left[\begin{array}{c} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{array} \right] \text{ and } \mathbf{G} = \left[\begin{array}{c} \mathbf{I} \\ \mathbf{A} \end{array} \right].
$$

This can be solved by ADMM as described in (37). For arbitrarily chosen $\mathbf{Z}^{(0)}$, $\mathbf{Y}^{(0)}$, and $\gamma > 0$, the proposed algorithm iterates following subproblems:

$$
\begin{cases}
\mathbf{X}^{(k+1)} & = \underset{\mathbf{X}}{\arg\min} \; \mathrm{tr}\left(\mathbf{X}^\mathsf{T}\mathbf{QX}\right) + \frac{1}{2\gamma} \left\| \mathbf{Z}^{(k)} - \mathbf{GX} - \mathbf{Y}^{(k)} \right\|_2^2 \\
\mathbf{Z}^{(k+1)} & = \underset{\mathbf{Z}}{\arg\min} \; l_{\{\|\cdot\|_{g,0} \leq p\}}\left(\mathbf{Z}\right) + \frac{1}{2\gamma} \left\| \mathbf{Z} - \mathbf{GX}^{(k+1)} - \mathbf{Y}^{(k)} \right\|_2^2 \\
\mathbf{Y}^{(k+1)} & = \mathbf{Y} + \mathbf{GX}^{(k+1)} - \mathbf{Z}^{(k+1)}.
\end{cases}
\tag{44}
$$

The first subproblem is differentiable and convex, and thus, the solution can be characterized as follows:

$$
\frac{\mathrm{d}}{\mathrm{d}\mathbf{X}} \left\{ \mathrm{tr}\left(\mathbf{X}^\mathsf{T}\mathbf{QX}\right) + \frac{1}{2\gamma} \left\| \mathbf{Z}^{(k)} - \mathbf{GX} - \mathbf{Y}^{(k)} \right\|_2^2 \right\} = 0
$$

$$
2\mathbf{QX} + \frac{1}{\gamma}\mathbf{G}^\mathsf{T}\mathbf{GX} + \frac{1}{\gamma}\mathbf{G}^\mathsf{T}\left(\mathbf{Y}^{(k)} - \mathbf{Z}^{(k)}\right) = 0, \tag{45}
$$

and the solution of the first subproblem can be obtained.

$$
\mathbf{X} = \left(2\mathbf{Q} + \frac{1}{\gamma}\mathbf{G}^\mathsf{T}\mathbf{G}\right)^{-1} \frac{1}{\gamma}\mathbf{G}^\mathsf{T}\left(\mathbf{Z}^{(k)} - \mathbf{Y}^{(k)}\right)
$$

$$
= \left(2\mathbf{Q} + \frac{1}{\gamma}\mathbf{I} + \frac{1}{\gamma}\mathbf{A}^\mathsf{T}\mathbf{A}\right)^{-1} \frac{1}{\gamma}\left\{(\mathbf{Z}_1 - \mathbf{Y}_1) + \mathbf{A}^\mathsf{T}(\mathbf{Z}_1 - \mathbf{Y}_1)\right\} \tag{46}
$$

Here, the covariance matrix $\mathbf{Q}$ can be approximated using a low-rank form as described in (19). In addition, the computational cost of the least-squares solution of the first subproblem can be reduced by adopting the inverse matrix lemma given below, when the number of latent variables is small:

$$
\left(2\mathbf{Q} + \frac{1}{\gamma}\mathbf{I} + \frac{1}{\gamma}\mathbf{A}^\mathsf{T}\mathbf{A}\right)^{-1} = \left(2\mathbf{U_Q}\mathbf{S_Q^2}\mathbf{U_Q^\mathsf{T}} + \underbrace{2\Delta\mathbf{S} + \frac{1}{\gamma}\mathbf{I} + \frac{1}{\gamma}\mathbf{A}^\mathsf{T}\mathbf{A}}_{\mathbf{J}} \right)^{-1}
$$

$$
= \mathbf{J}^{-1} - \mathbf{J}^{-1}\mathbf{U}\left\{\left(2\mathbf{S}^2\right)^{-1} + \mathbf{U}^\mathsf{T}\mathbf{J}^{-1}\mathbf{U}\right\}^{-1} \mathbf{U}^\mathsf{T}\mathbf{J}^{-1}, \tag{47}
$$

where

$$
\mathbf{J}^{-1} = \left(2\Delta\mathbf{S} + \frac{1}{\gamma}\mathbf{I} + \frac{1}{\gamma}\mathbf{A}^\mathsf{T}\mathbf{A}\right)^{-1}
$$

$$
= \left(2\Delta\mathbf{S} + \tfrac{1}{\gamma}\mathbf{I}\right)^{-1} - \left(2\Delta\mathbf{S} + \tfrac{1}{\gamma}\mathbf{I}\right)^{-1}\tfrac{1}{\gamma}\mathbf{A}^\mathsf{T}\left(\mathbf{I} + \mathbf{A}\left(2\Delta\mathbf{S} + \tfrac{1}{\gamma}\mathbf{I}\right)^{-1}\tfrac{1}{\gamma}\mathbf{A}^\mathsf{T}\right)^{-1}\mathbf{A}\left(2\Delta\mathbf{S} + \tfrac{1}{\gamma}\mathbf{I}\right)^{-1}. \tag{48}
$$

The present method reduces the computational cost using a low-rank expression of the correlated measurement noise [31]. Therefore, when a more general noise model in which the correlated measurement noise is inversely proportional to the distance between sensors is used, the computational cost becomes large.

By noticing the definition of the indicator function (41), the second subproblem, the update of $\mathbf{Z}$ can be rewritten as follows:

$$
\mathbf{Z}^{(k+1)} = \underset{\mathbf{Z}}{\arg\min} \; \frac{1}{2\gamma} \left\| \mathbf{Z} - \mathbf{GX}^{(k+1)} - \mathbf{Y}^{(k)} \right\|_2^2
$$

$$
\text{subject to } \|\mathbf{Z}\|_{g,0} \leq p, \tag{49}
$$

and this update can be rewritten by the proximal operator for the group ℓ_0 pseudo-norm constraint,

$$
\mathbf{Z}^{(k+1)} = \mathrm{prox}_{l_{\{\|\cdot\|_{g,0} \leq p\}}}\left(\mathbf{GX}^{(k+1)} + \mathbf{Y}^{(k)}\right). \tag{50}
$$

The proximal operator of the indicator function $l_{\{\|\cdot\|_{g,0} \le p\}}$ corresponds to metric projection onto the set satisfying the constraint $\|\mathbf{Z}\|_{g,0} \le p$ [50],

$$\text{prox}_{l_{\{\|\cdot\|_{g,0} \le p\}}}(\mathbf{V}) = \underset{\mathbf{W}}{\arg\min} \sum_{i=1}^{n} \|\mathbf{v}_i - \mathbf{w}_i\|_2^2 \tag{51}$$

$$\text{subject to } \|\mathbf{W}\|_{g,0} \le p$$

$$=: P_{\{\|\cdot\|_{g,0} \le p\}}(\mathbf{V}), \tag{52}$$

where vectors $\mathbf{v}_i$ and $\mathbf{w}_i$ are row vectors of the matrices $\mathbf{V}$ and $\mathbf{W}$, respectively. The operation for this projection is almost the same as the block hard thresholding, but the threshold is $\left(\|\mathbf{v}_j\|_2\right)_p$, which is the pth largest value of $\|\mathbf{v}_i\|_2$:

$$P_{\{\|\cdot\|_{g,0} \le p\}}(\mathbf{V}) := \begin{cases} \mathbf{v}_i, & \|\mathbf{v}_i\|_2 \ge \left(\|\mathbf{v}_j\|_2\right)_p, \\ 0, & \|\mathbf{v}_i\|_2 < \left(\|\mathbf{v}_j\|_2\right)_p. \end{cases} \tag{53}$$

The procedures of the proposed algorithm are summarized in Algorithm 1. Because the problem is nonconvex when the hard thresholding operator is used, convergence is not guaranteed. ADMM may converge to a just local optimum, but it is expected that it will possibly have better convergence properties (e.g., faster convergence or convergence to a point with better objective value) than other local optimization methods. There is ample experimental evidence in the literature that supports the empirical convergence of ADMM, especially when the nonconvex problem at hand exhibits a "favorable" structure [51]. The present problem consists of the convex function and the nonconvex function. In general, this is hard to compute, but the nonconvex function of the present problem is the cardinality constraint, which keeps the p largest magnitude elements and zeros out the rest. This is one of the special cases that can be exactly carried out. In addition, a decreasing γ strategy was employed and ADMM was stabilized for nonconvex optimization. A scalar η is introduced to gradually decrease the value of γ. Similar strategies were adopted in existing methods using ℓ_0-type norms [50], [52]–[55]. This technique is supported by recent convergence analyses of ADMM for nonconvex cases (e.g., [56], [57]). In these studies, the sequence generated by ADMM under appropriate conditions with sufficiently small γ converges to a stationary point. Although these theoretical analyses employ overly strict assumptions and cover the problem formulations in many applications, including the present study, they provide insight into parameter settings.

The convergence and error bound of the nonconvex problem in the present study might be provided based on the Kurdyka-Lojasiewicz (KL) property, etc. The convergence of ADMM in nonconvex problems with certain properties has been analyzed by several researchers [58], [59]. However, we were unable to provide proof that our nonconvex problem satisfies the conditions that are required for convergence at present. The objective of the present paper is to provide the result that better sensor selection can be performed practically by the present ADMM-based method than the previous methods. Therefore, the theoretical analysis of convergence and the error bound are left for future researches.

Algorithm 1 Proposed Algorithm for Sensor Selection With Group ℓ_0 Pseudo-Norm Constraint (MATLAB code available online [43])

Input: $\mathbf{X}^{(0)} = \mathbf{A}^{-1}$, $\mathbf{Y}^{(0)} = \mathbf{Z}^{(0)} = 0$, $p > 0$, $\gamma > 0$, and $0 < \eta < 1$
Output: $\mathbf{X}^{(k)}$
  while $\|\mathbf{X}^{(k)} - \mathbf{X}^{(k-1)}\|_F > \epsilon$ **do**
    $\mathbf{X}^{(k+1)} =$
    $\left(\mathbf{J}^{-1} - \mathbf{J}^{-1}\mathbf{U}\left\{\left(2\mathbf{S}^2\right)^{-1} + \mathbf{U}^\mathsf{T}\mathbf{J}^{-1}\mathbf{U}\right\}^{-1}\mathbf{U}^\mathsf{T}\mathbf{J}^{-1}\right) \frac{1}{\gamma}\left\{(\mathbf{Z}_1 - \mathbf{Y}_1) + \mathbf{A}^\mathsf{T}(\mathbf{Z}_1 - \mathbf{Y}_1)\right\}$;
    $\mathbf{Z}^{(k+1)} = \text{prox}_{l_{\{\|\cdot\|_{g,0} \le p\}}}\left(\mathbf{GX}^{(k+1)} + \mathbf{Y}^{(k)}\right)$;
    $\mathbf{Y}^{(k+1)} = \mathbf{Y} + \mathbf{GX}^{(k+1)} - \mathbf{Z}^{(k+1)}$;
    $\gamma \leftarrow \eta\gamma$;
    $k \leftarrow k + 1$;
  end while

E. Polishing Step

Although latent variables $\mathbf{z}$ can be estimated directly by the gain matrix $\mathbf{K}$, the accuracy of the estimation is degraded to some extent by the sparsity-promoting terms. Therefore, the sensor location matrix $\mathbf{H}$ is constructed based on the gain matrix $\mathbf{K}$, and the gain is recalculated using (5) rather than the gain matrix $\mathbf{K}$ directly. The corresponding entry in $\mathbf{H}$ is set to unity (which activates the corresponding sensor) when the ℓ_2 norm of the corresponding column vector of $\mathbf{K}$ is nonzero (e.g., greater than 10^{-4}). The latent variables $\tilde{\mathbf{z}}$ is estimated using (5) after constructing $\mathbf{H}$.

IV. Results and Discussion

The performance of the proposed method was evaluated by applying it to an artificial dataset used in a previous study [30]. The obtained results were compared to those obtained by the greedy methods considering white noise or correlated noise and the ADMM-based method considering white noise. Random data matrices, $\mathbf{X}_{\text{data}} = \mathbf{USV}^\mathsf{T}$ ($\mathbf{X}_{\text{data}} \in \mathbb{R}^{n \times m}$), were generated, where $\mathbf{U} \in \mathbb{R}^{n \times m}$ and $\mathbf{V} \in \mathbb{R}^{m \times m}$ consist of m-orthogonal column vectors that were generated by QR decomposition of normally distributed [$\mathcal{N}(0,1)$] random matrix and diagonal entries of the diagonal matrix $\mathbf{S} \in \mathbb{R}^{m \times m}$ were $\text{diag}(\mathbf{S}) = \left[1, 1/\sqrt{2}, 1/\sqrt{3}, ..., 1/\sqrt{m}\right]$, respectively.

The leading-$r_1(= 10)$ modes were retained for the construction of a reduced-order model of $\mathbf{X}_{\text{data}}$, and a certain number of subsequent modes up to the r_2th mode (from the 11th to the 40th modes in the present study) were retained for the construction of the noise covariance matrix. The objective function is the A-optimality criterion for all methods. The greedy method considering white noise [Greedy(WN)] [27], the greedy method considering correlated noise [Greedy(CN)], the ADMM-based method considering white noise [ADMM(WN)] [19], and the ADMM-based method considering correlated noise [ADMM(CN)] were tested. Here, Greedy(CN) selects sensor locations by minimizing the objective function (9) in each single-sensor subproblem. This method is the same as the reference [31], except for the point that the objective function is based on the D-optimality criterion. In the case of Greedy(WN) [27],
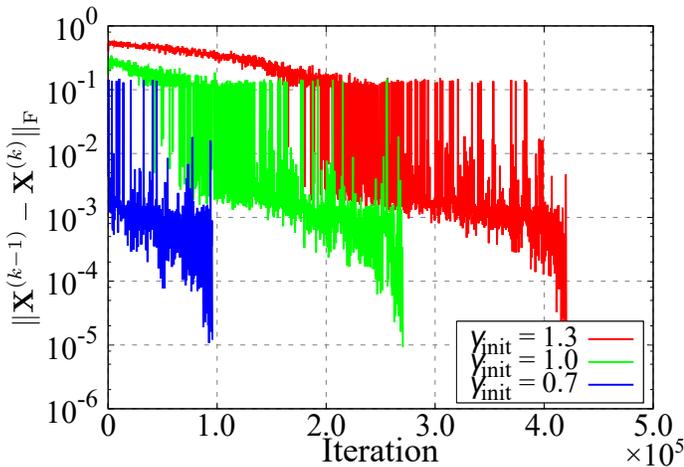
Fig. 1. Iteration behavior for a certain case of ADMM(CN) for different γ_{init} for $p = 30$.

it is the same as Greedy(CN), except for the noise covariance matrix $\mathbf{R}_{\text{d}}$ in the objective function (9) is the identity matrix. In the case of ADMM(WN) [19], it is the same as ADMM(CN), except for the noise covariance matrix $\mathbf{Q}$ in the objective function of ADMM(CN) is the identity matrix.

In addition, ADMM(CNw/oN), which is the ADMM-based method without normalization by the noise weighting term described in Section III-B, was also tested to investigate the effectiveness of normalization. The difference between ADMM(CN) and ADMM(CNw/oN) is described by the difference in their objective functions (33) and (28). The computations using the ADMM-based methods were conducted with the ℓ_0-constrained block hard thresholding (L_0BHT). Because the problem is nonconvex when the hard thresholding operator is used, convergence is not guaranteed. Therefore, the decreasing γ strategy was used. The step size γ is initially set to be $\gamma = \gamma_{\text{init}}$, and it was gradually reduced by multiplying by $\eta = 0.99$ [50] every 5000 iterations. The iteration behavior of the present method [ADMM(CN)] for a certain case is shown in Figure 1. The size of dataset was $\mathbf{X}_{\text{data}} \in \mathbb{R}^{10,000 \times 100}$. The objective values obtained for $\gamma_{\text{init}} = 0.7, 1.0$, and 1.3 were $\text{tr}\left(\mathbf{X}^{\mathsf{T}} \mathbf{R} \mathbf{X}\right) = 0.1526, 0.1526$, and 0.1517, respectively.

In the following experiments, the value of γ for ADMM-based methods was initially set at $\gamma_{\text{init}} = 1.0$. Each calculation in the numerical experiments was carried out 100 times with different datasets, and the computational time, the averaged objective value, and the reconstruction error were calculated.

A. Computational Complexity and Computational Time

The computational complexities of the previously proposed methods and the present method are summarized in Table. III. Here, SDP(CN) indicates the method based on semidefinite programming (SDP) proposed by Liu et al. [38]. This method selects sensors while considering the influence of correlated measurement noise. Although prior knowledge (the matrix $\mathbf{\Sigma}$ in reference [38]) can be considered when selecting sensor locations, it was not included in the present numerical experiment for a fair comparison.

TABLE III
COMPARISON OF COMPUTATIONAL COMPLEXITY

Method	Computational complexity
Convex relaxation method(WN) [11]	$O\left(n^3\right)$ per iteration
SDP(CN) [38]	$O\left(n^{4.5}\right)$ per iteration
Greedy(WN) [27]	$O\left(pnr_1^2\right)$
Greedy(CN)	$O\left(pn(r_1 - r_2)^2\right)$
ADMM(WN) [19]	$O\left(nr_1^2\right)$ per iteration
ADMM(CN)	$O\left(n(r_1 - r_2)^2\right)$ per iteration

The number of operations in matrix multiplication and addition was evaluated based on the size of the matrices, whereas the matrix operations were accounted based on the Matlab code implementation. The largest complexity among each term was taken as the computational complexity of the algorithms. Here, the problem considered in the present study was $r_1 \leq p \approx (r_1 - r_2) \ll n$.

The computational complexities of Greedy(WN) and ADMM(WN) are $O\left(pnr_1^2\right)$ and $O\left(nr_1^2\right)$ (per iteration), respectively, where p is the number of sensors to be activated, n is the number of potential sensor locations, and r_1 is the number of latent variables. The computational complexity increases for methods that consider correlated noise. The computational complexities of Greedy(CN) and ADMM(CN) are $O\left(pn(r_1 - r_2)^2\right)$ and $O\left(n(r_1 - r_2)^2\right)$ (per iteration), respectively. Regarding the computational complexity of each subproblem of the proposed method (Algorithm 1), the complexity of the first subproblem is dominant, and it is $O\left(n(r_1 - r_2)^2\right)$. The complexities of the second and third subproblems are $O(pn)$ and $O\left(nr_1^2\right)$, respectively. Although the computational cost of ADMM-based methods is independent of p, these methods require iterative computation, and thus, the characteristics of the dataset also have an influence on the total computation time.

Fig. 2 shows the comparison of computational time. The numerical test was carried out on a desktop personal computer (Intel Xeon W-2295 3.0 GHz CPU with 256 GB RAM). The size of dataset was $\mathbf{X}_{\text{data}} \in \mathbb{R}^{n \times 50}$. The number of potential sensor locations n was changed from 10^2 to 5×10^4, and the average computational time and objective value of 100 times computations with different datasets were compared. The number of selected sensors was fixed at $p = 30$.

The computational time for all methods increases as n increases. The computational time of Greedy(CN) is much less than that of SDP(CN) and ADMM(CN). Although the computational time of SDP(CN) is less than that of ADMM(CN) at $n = 10^2$, it rapidly increases as n increases. The computational complexity of SDP(CN) is roughly given by $O\left(n^{4.5}\right)$ [38], [60], if the number of latent variables is much less than that of potential sensor locations. Consequently, approximately 10^4 s is required even if the potential sensor location was $n = 500$ when SDP(CN) is used. In addition, approximately 200 GB of random access memory is required for the problem with 500 sensor candidates. On the other hand, an increase in the computational time with respect to n of ADMM(CN) is less than that of SDP(CN) and is at the same level as that of Greedy(CN), and problems with $n = 5 \times 10^4$ can be treated.

Fig. 2. Computational time with respect to the number of potential sensor locations for $p = 30$.
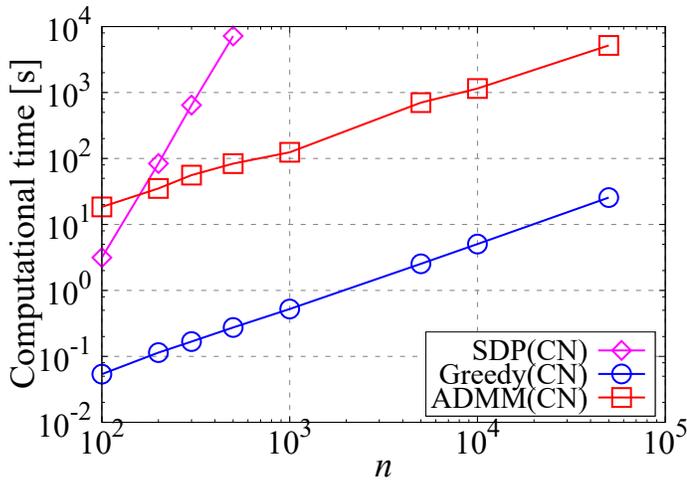
Fig. 3. Objective value with respect to the number of potential sensor locations for $p = 30$.

It should be noted that the present method assumes that the noise covariance matrix can be approximated by a low-rank form based on the framework proposed by Yamada et al. [30], [31]. Therefore, an additional computational cost might be required when using a more general noise model, e.g. an inverse-distance noise model in which the correlated measurement noise is inversely proportional to the distance between sensors.

B. A-optimality Criterion and Reconstruction Error

Fig. 3 shows the trace of the inverse of the FIM with respect to the number of potential sensor locations for $p = 30$. The size of the dataset was $\mathbf{X}_{\mathrm{data}} \in \mathbb{R}^{10,000 \times 100}$. The objective value was normalized by that obtained by Greedy(CN). Although SDP(CN) [38] is quite general, the objective value obtained by SDP(CN) is larger, thus, worse than that obtained by Greedy(CN) and ADMM(CN). In addition, obtained objective value using SDP(CN) gets worse as the number of potential sensor locations increases. The computational time of ADMM(CN) is longer than that of Greedy(CN), as illustrated in Fig. 3, but the objective value obtained is superior to that obtained by Greedy(CN) under the tested conditions.

Fig. 4 shows the trace of the inverse of the FIM with respect to the number of selected sensors. Here, the size of dataset was $\mathbf{X}_{\mathrm{data}} \in \mathbb{R}^{10,000 \times 100}$. This figure illustrates that the objective values obtained by the methods that consider correlated measurement noise are smaller than those obtained by the methods that consider only white noise, which is an expected result. The objective value obtained by ADMM(CN) is better than that obtained by Greedy(CN) and is the best among the compared methods, except for $p = 10$. The objective values obtained by Greedy(WN) and ADMM(WN) are larger because the noise covariance matrix $\mathbf{R}$ is not considered in the sensor selection. The objective value obtained by ADMM(CNw/oN) is larger than that obtained by the method that does not consider correlated noise, even though the method considers correlated noise in the sensor selection.

As described in Section III-B, the difference between ADMM(CN) and ADMM(CNw/oN) is described by the dif-

ference in their objective functions (33) and (28). In the case of ADMM(CNw/oN), the noise covariance matrix is considered in the first term of the objective function, but it is not considered in the sparsity-promoting term. In the case of ADMM(CN), on the other hand, the sensor candidate matrix $\mathbf{U}$ is normalized by the noise weighting term $\mathbf{R}_{\mathrm{d}}$, which is the diagonal matrix with the diagonal entries of the noise covariance matrix $\mathbf{R}$. Therefore, the noise intensity is considered in the computation of the sparsity-promoting term for ADMM(CN). The consideration of off-diagonal entries is more difficult because the effect of correlated noise cannot be accurately evaluated unless the sensor locations are determined, as discussed in Section II-D. Therefore, it is difficult to consider the correlated measurement noise in the sparsity-prompting term, and the off-diagonal entries are not fully considered even in ADMM(CN). It should be noted that the noise considered in the present study has a certain size structure because the noise model consists of truncated modes. Hence, not only the uncorrelated noise but also part of the component of correlated noise will be considered by normalization using the noise weighting term, even if only the diagonal component of the noise covariance matrix is considered. The result of this numerical experiment indicates that the normalization of the sensor candidate matrix by the noise weighting term is effective and suggests that consideration of the intensity of noise in the thresholding process is important for optimization using proximal methods.

Fig. 5 shows the reconstruction error, which is the difference between the reconstructed and original data, defined as follows:

$$\epsilon_{\mathrm{reconst}} = \frac{\|\mathbf{X}_{\mathrm{data}} - \mathbf{U}\tilde{\mathbf{Z}}\|_{\mathrm{F}}}{\|\mathbf{X}_{\mathrm{data}}\|_{\mathrm{F}}}, \quad (54)$$

where $\tilde{\mathbf{Z}}$ is the estimated coefficients of modes obtained by (5). The notation $\|\cdot\|_{\mathrm{F}}$ indicates the Frobenius norm of a matrix. The reconstruction error is reduced by considering the correlated measurement noise, and the smallest reconstruction error is obtained for ADMM(CN). As pointed out with regard to Fig. 4, the performance of ADMM(CNw/oN) is the worst,
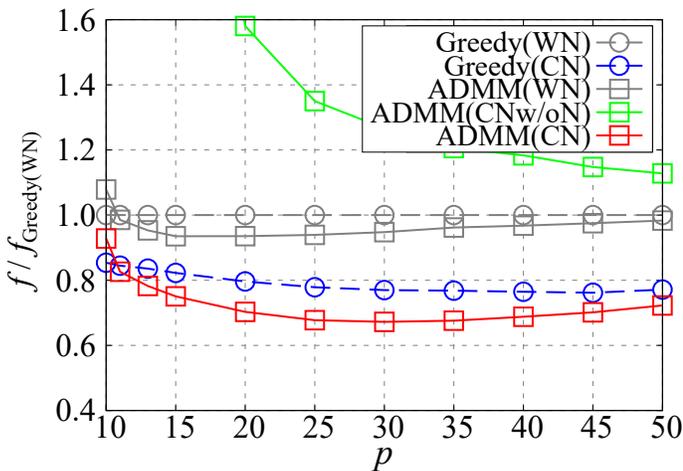
Fig. 4. Comparison of objective values obtained in the experiment using the artificial dataset.

Fig. 5. Reconstruction error for an artificial dataset.

even though correlated noise is considered. The trends of the reconstruction error are the same as those of the objective value because it is an ideal problem setting.

The greedy algorithm can only provide a sensor subset consisting of p sensors only by adding a new sensor that gives the largest increment of the objective to the sensor subset consisting of $p-1$ sensors. The influence of this constraint on greedy strategy is relatively small at $p < r$ due to the nature of the undersampling condition, but the constraint tends to degrade the performance of the sensor subset obtained under oversampling conditions. On the other hand, the ADMM-based method can obtain different suboptimal solutions for each number of sensors, similar to the convex relaxation method. Hence, the performance of the sensor subset obtained by ADMM(CN) is higher than that obtained by Greedy(CN) at larger p. However, the difference in the objective value and reconstruction error depending on the method decreases as the number of selected sensors p increases. This is because as the number of sensors p increases, the importance of location and combination of each selected sensor becomes smaller than when the number of selected sensors is small. On the other hand, the condition on $p = 10$ is on the constraint of the ADMM-based method, and thus, the optimization can be severe and the quality of the solution deteriorates. Hence, the difference in the performance between ADMM(CN) and Greedy(CN) is the largest at around $p = 25$.

V. APPLICATION TO DATA-DRIVEN SENSOR SELECTION

The proposed method was applied to a data-driven sensor selection problem, and comparisons with the previously proposed method were conducted. The adopted dataset was the NOAA-OISST V2 dataset, which comprises weekly global sea surface temperature measurements between 1990 and 2000 [61]. The dataset used consisted of 520 snapshots with 44,219 points of a spatial grid. The temperature data for 10 years were split into training data $\mathbf{X}_{\text{train}}$ (80% of the dataset) and the test data $\mathbf{X}_{\text{test}}$ (20% of the dataset), and five-fold cross-validation was conducted.
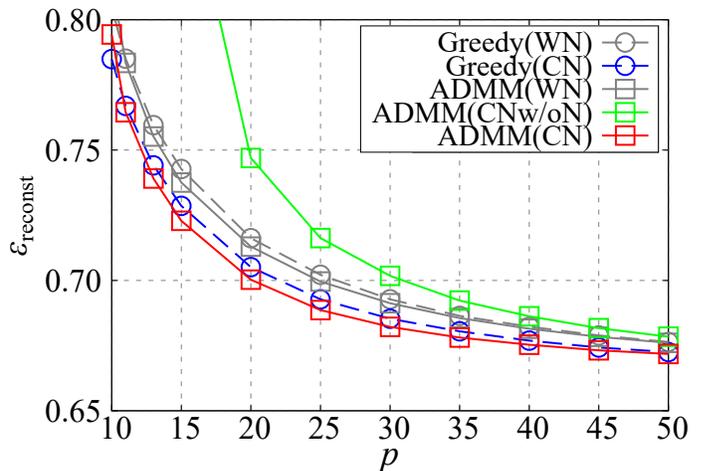
The data matrix $\mathbf{X}_{\text{data}} \in \mathbb{R}^{n \times m}$, which consisted of m snapshots with a spatial dimension of n, was decomposed into a left singular matrix $\mathbf{U} \in \mathbb{R}^{n \times m}$, which shows the spatial modes; a diagonal matrix of singular values $\mathbf{S} \in \mathbb{R}^{m \times m}$; and a right singular matrix $\mathbf{V} \in \mathbb{R}^{m \times m}$, which shows the temporal modes. Here, the spatial dimension and the number of snapshots for this dataset were $n = 44,219$ and $m = 416$, respectively. The dimensional reduction was conducted by the truncated SVD, and the rank-r reduced-order modeling of a data matrix is given as the r-rank approximation $\mathbf{X}_{\text{data}} \approx \mathbf{U}_{1:r_1} \mathbf{S}_{1:r_1} \mathbf{V}_{1:r_1}^{\mathsf{T}}$ [44]. The reduced-order model was constructed by retaining the leading-r_1 singular values and vectors, where $r_1 = 10$ in the present experiment. The noise covariance matrix was constructed using truncated modes up to $r_2 = 40$ modes based on (17). The objective value (9) and the reconstruction error (54) were evaluated. The initial values of γ_{init} for ADMM(WN) and ADMM(CN) were set at 0.7 and 1.2, respectively, and the value was multiplied by $\eta = 0.99$ every 5000 iterations, similar to the numerical experiments in Section IV. The latent variables $\mathbf{Z}_{1:r}(= \mathbf{S}_{1:r_1} \mathbf{V}_{1:r_1}^{\mathsf{T}})$ in this problem are mode coefficients of the spatial modes of the time variation in the temperature field $\mathbf{U}_{1:r}$. The mode coefficients $\mathbf{Z}_{1:r}$ indicate the time variation in the strength of each mode.

Fig. 6 illustrates the locations of the 30 sensors selected by the greedy and ADMM-based methods. The white open circles represent sensor locations. There is a clear difference in the locations of the selected sensors obtained by the methods considering white noise and correlated noise. The sensors selected by the methods assuming white noise are aggregated into several regions. In contrast, the location of sensors is dispersed by considering the correlated measurement noise. Observations using the latter sensor subset can be more informative than observations with the former subset.

The objective values are shown in Fig. 7. As discussed in Section IV, the objective values obtained by Greedy(CN) and ADMM(CN) are smaller than those obtained by Greedy(WN) and ADMM(WN). The objective value obtained by ADMM(CN) is better than that obtained by Greedy(CN) at around $15 \leq p \leq 40$ and is worse at smaller
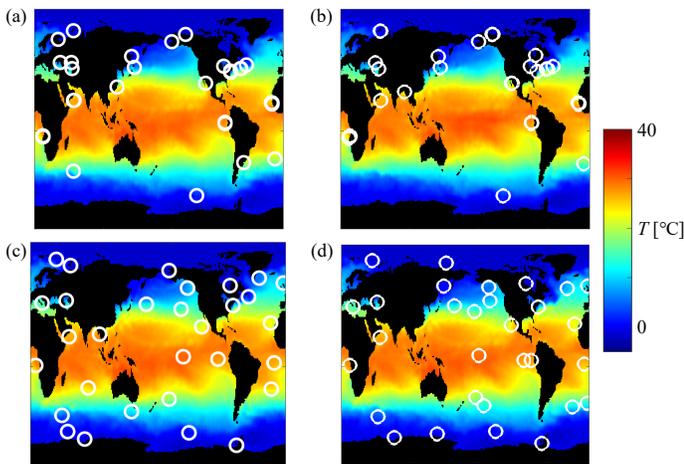
Fig. 6. Locations of selected sensors at $p = 30$. (a) Greedy(WN), (b) ADMM(WN), (c) Greedy(CN), (d) ADMM(CN). The white open circles represent the locations of activated sensor locations.

Fig. 8. Reconstruction error for the NOAA-OISST dataset.
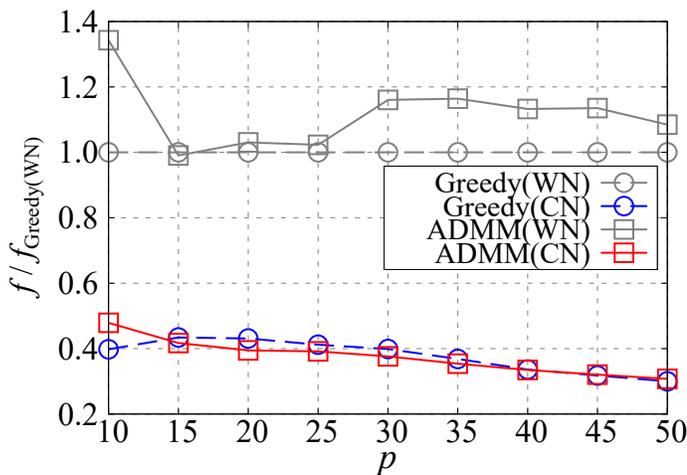
Fig. 7. Comparison of objective values obtained in experiment using practical dataset.

and larger values of p. For $p = 10$, the condition is on the constraint of the ADMM-based method, and thus, the optimization can be severe and the quality of the solution deteriorates. A similar trend for the ADMM-based method has been reported in a previous study [19] and can also be seen in the convex relaxation method [11] (see [19], [24], [27]).

Fig. 8 shows that the reconstruction error is reduced by considering the correlated measurement noise, as was seen for the numerical experiment with the artificial dataset. ADMM (CN) achieved the smallest reconstruction error in the range $15 \le p \le 40$. As discussed in the numerical experiment using the artificial dataset, the importance of the location and combination of each activated sensor gradually decreases as the number of sensors p increases. Therefore, the difference in the reconstruction error between ADMM(CN) and Greedy(CN) decreases for larger p. In addition, the performance of the ADMM-based methods deteriorates at $p = 10$ compared to greedy methods.

VI. CONCLUSIONS

The present paper proposes a data-driven sensor selection method for high-dimensional nondynamical systems with correlated measurement noise. The proposed method is based on the proximal splitting algorithm and the A-optimal design of experiments. The sensor locations are determined by minimizing the trace of the inverse of the FIM with group ℓ_0 pseudo-norm constraint. The proposed method can avoid the difficulty of sensor selection with strongly correlated measurement noise, namely, that the sensor locations must be known in advance for selecting the sensor locations, as described in Section II-D. The problem can be efficiently solved with ADMM, and the computational complexity is $O(n)$ when the method is combined with a low-rank expression of the measurement noise model [30], [31]. Therefore, the proposed method can treat large-scale problems that have more than 10,000 potential sensor locations, such as data-driven sensor selection problems.

The performance of the proposed method was compared with those of the previously proposed methods by the numerical experiments using artificial and practical datasets. The adopted practical dataset was NOAA-OISST V2, which contains weekly mean sea surface temperatures.

The results of numerical experiments using the artificial dataset and NOAA-OISST dataset showed that the performance of ADMM(CN) is the best among the compared methods in terms of the objective value and reconstruction error. The ADMM-based method without normalization by the noise weighting term [ADMM(CNw/oN)] was also tested to investigate the effectiveness of noise normalization in the thresholding process. Although ADMM(CNw/oN) considers correlated noise, the obtained objective value was larger than those obtained by methods that do not consider correlated noise. This indicates that the normalization of the sensor candidate matrix by the noise weighting term is effective and suggests that consideration of the noise intensity in the thresholding process is essential for optimization using proximal methods.

The previous methods considering the correlated measure-

ment noise require the calculation of the precision matrix (inverse of the noise covariance matrix) which should be constructed by the noise information over only selected sensors. In particularly, the method based on convex relaxation requires much larger calculation cost or a complicated formulation with limitations. There was no published method to formulate the continuous optimization problem for sensor selection that can be solved by a method with a computational complexity less than a cubic order of the number of potential sensor locations. Hence, it was difficult to employ the previously proposed methods for high-dimensional data including such correlated measurement noise. Although the computational cost becomes quite large when using a more general noise model, the proposed method can achieve low-cost nonconvex sensor optimization considering correlated measurement noise by means of a low-rank expression of correlated measurement noise [30]. Further extension of the complex objective function and its constraints is expected using the framework proposed in the present study.

REFERENCES

[1] M. Kihara and T. Okada, "A satellite selection method and accuracy for the global positioning system," *Navigation*, vol. 31, no. 1, pp. 8–20, 1984.

[2] M. S. Phatak, "Recursive method for optimum gps satellite selection," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 37, no. 2, pp. 751–754, 2001.

[3] K. Doğançay and H. Hmam, "On optimal sensor placement for time-difference-of-arrival localization utilizing uncertainty minimization," in *2009 17th European Signal Processing Conference*. IEEE, 2009, pp. 1136–1140.

[4] R. Macho-Pedroso, F. Domingo-Perez, J. Velasco, C. Losada-Gutierrez, and J. Macias-Guarasa, "Optimal microphone placement for indoor acoustic localization using evolutionary optimization," in *2016 international conference on indoor positioning and indoor navigation (IPIN)*. IEEE, 2016, pp. 1–8.

[5] K. Worden and A. Burrows, "Optimal sensor placement for fault detection," *Engineering structures*, vol. 23, no. 8, pp. 885–901, 2001.

[6] T.-H. Yi, H.-N. Li, and M. Gu, "Optimal sensor placement for structural health monitoring based on multiple optimization strategies," *The Structural Design of Tall and Special Buildings*, vol. 20, no. 7, pp. 881–900, 2011.

[7] W. Du, Z. Xing, M. Li, B. He, L. H. C. Chua, and H. Miao, "Optimal sensor placement and measurement of wind for water quality studies in urban reservoirs," in *IPSN-14 Proceedings of the 13th International Symposium on Information Processing in Sensor Networks*. IEEE, 2014, pp. 167–178.

[8] W.-J. Yeo, S. Taulu, and J. N. Kutz, "Efficient magnetometer sensor array selection for signal reconstruction and brain source localization," *arXiv preprint arXiv:2205.10925*, 2022.

[9] W. J. Welch, "Branch-and-bound search for experimental designs based on d optimality and other criteria," *Technometrics*, vol. 24, no. 1, pp. 41–48, 1982.

[10] E. L. Lawler and D. E. Wood, "Branch-and-bound methods: A survey," *Operations research*, vol. 14, no. 4, pp. 699–719, 1966.

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

[12] T. Nonomura, S. Ono, K. Nakai, and Y. Saito, "Randomized subspace newton convex method applied to data-driven sensor selection problem," *IEEE Signal Processing Letters*, vol. 28, pp. 284–288, 2021.

[13] M. Fardad, F. Lin, and M. R. Jovanović, "Sparsity-promoting optimal control for a class of distributed systems," in *Proceedings of the 2011 American Control Conference*. IEEE, 2011, pp. 2050–2055.

[14] F. Lin, M. Fardad, and M. R. Jovanović, "Design of optimal sparse feedback gains via the alternating direction method of multipliers," *IEEE Transactions on Automatic Control*, vol. 58, no. 9, pp. 2426–2431, 2013.

[15] N. K. Dhingra, M. R. Jovanović, and Z.-Q. Luo, "An ADMM algorithm for optimal sensor and actuator selection," in *53rd IEEE Conference on Decision and Control*. IEEE, 2014, pp. 4039–4044.

[16] A. Zare and M. R. Jovanović, "Optimal sensor selection via proximal optimization algorithms," in *2018 IEEE Conference on Decision and Control (CDC)*. IEEE, 2018, pp. 6514–6518.

[17] E. Masazade, M. Fardad, and P. K. Varshney, "Sparsity-promoting extended kalman filtering for target tracking in wireless sensor networks," *IEEE Signal Processing Letters*, vol. 19, no. 12, pp. 845–848, 2012.

[18] S. Liu, E. Masazade, M. Fardad, and P. K. Varshney, "Sparsity-aware field estimation via ordinary kriging," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 3948–3952.

[19] T. Nagata, T. Nonomura, K. Nakai, K. Yamada, Y. Saito, and S. Ono, "Data-driven sparse sensor selection based on a-optimal design of experiment with ADMM," *IEEE Sensors Journal*, no. 13, pp. 15 248–15 257, 2021.

[20] 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, 2018.

[21] 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.

[22] 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.

[23] 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.

[24] Y. Saito, T. Nonomura, K. Nankai, K. Yamada, K. Nakai, T. Nagata, K. Asai, Y. Sasaki, and D. Tsubakino, "Determinant-based fast greedy sensor selection algorithm," *IEEE Access*, vol. 9, pp. 68 535–68 551, 2021.

[25] Y. Saito, T. Nonomura, K. Nankai, K. Yamada, K. Asai, Y. Sasaki, and D. Tsubakino, "Data-driven vector-measurement-sensor selection based on greedy algorithm," *IEEE Sensors Letters*, vol. 4, 2020.

[26] Y. Saito, K. Yamada, N. Kanda, K. Nakai, T. Nagata, T. Nonomura, and K. Asai, "Data-driven determinant-based greedy under/oversampling vector sensor placement," *CMES-Computer Modeling in Engineering & Sciences*, vol. 129, no. 1, pp. 1–30, 2021.

[27] K. Nakai, K. Yamada, T. Nagata, Y. Saito, and T. Nonomura, "Effect of objective function on data-driven greedy sparse sensor optimization," *IEEE Access*, vol. 9, pp. 46 731–46 743, 2021.

[28] E. Clark, T. Askham, S. L. Brunton, and J. N. Kutz, "Greedy sensor placement with cost constraints," *IEEE Sensors Journal*, vol. 19, no. 7, pp. 2642–2656, 2018.

[29] K. Manohar, J. N. Kutz, and S. L. Brunton, "Optimal sensor and actuator selection using balanced model reduction," *IEEE Transactions on Automatic Control*, vol. 67, no. 4, pp. 2108–2115, 2021.

[30] K. Yamada, Y. Saito, K. Nankai, T. Nonomura, K. Asai, and D. Tsubakino, "Fast greedy optimization of sensor selection in measurement with correlated noise," *Mechanical Systems and Signal Processing*, vol. 158, p. 107619, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0888327021000145

[31] K. Yamada, Y. Saito, T. Nonomura, and K. Asai, "Greedy sensor selection for weighted linear least squares estimation under correlated noise," *IEEE Access*, accepted.

[32] C. Jiang, Z. Chen, R. Su, and Y. C. Soh, "Group greedy method for sensor placement," *IEEE Transactions on Signal Processing*, vol. 67, no. 9, pp. 2249–2262, 2019.

[33] E. Clark, S. L. Brunton, and J. N. Kutz, "Multi-fidelity sensor selection: Greedy algorithms to place cheap and expensive sensors with cost constraints," *IEEE Sensors Journal*, vol. 21, no. 1, pp. 600–611, 2020.

[34] E. Clark, J. N. Kutz, and S. L. Brunton, "Sensor selection with cost constraints for dynamically relevant bases," *IEEE Sensors Journal*, vol. 20, no. 19, pp. 11 674–11 687, 2020.

[35] B. Li, H. Liu, and R. Wang, "Efficient sensor placement for signal reconstruction based on recursive methods," *IEEE Transactions on Signal Processing*, vol. 69, pp. 1885–1898, 2021.

[36] K. Nakai, T. Nagata, K. Yamada, Y. Saito, and T. Nonomura, "Nondominated-solution-based multiobjective-greedy sensor selection for optimal design of experiments," *arXiv preprint arXiv:2204.12695e*, 2022.

[37] T. Nagata, K. Yamada, K. Nakai, Y. Saito, and T. Nonomura, "Randomized group-greedy method for data-driven sensor selection," *arXiv preprint arXiv:2205.04161*, 2022.

[38] S. Liu, S. P. Chepuri, M. Fardad, E. Maşazade, G. Leus, and P. K. Varshney, "Sensor selection for estimation with correlated measurement noise," *IEEE Transactions on Signal Processing*, vol. 64, no. 13, pp. 3509–3522, 2016.

[39] D. Uciński, "D-optimal sensor selection in the presence of correlated measurement noise," *Measurement*, vol. 164, p. 107873, 2020.

[40] M. O'Connor, W. B. Kleijn, and T. Abhayapala, "Distributed sparse mvdr beamforming using the bi-alternating direction method of multipliers," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 106–110.

[41] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Computers & mathematics with applications*, vol. 2, no. 1, pp. 17–40, 1976.

[42] J. Eckstein and D. P. Bertsekas, "On the douglas–rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, vol. 55, no. 1-3, pp. 293–318, 1992.

[43] T. Nagata, "Proximal-optimization-based-sensor-selection-algorithm-with-correlated-measurement-noise," Available online, 2022, https://github.com/Aerodynamics-Lab/Proximal-Optimization-based-Sensor-Selection-Algorithm-with-Correlated-Measurement-Noise.

[44] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, 1936.

[45] E. Rigtorp, "Sensor selection with correlated noise," Master's thesis, KTH Royal Institute of Tehcnology, 2010.

[46] H. Jamali-Rad, A. Simonetto, G. Leus, and X. Ma, "Sparsity-aware sensor selection for correlated noise," in *17th International Conference on Information Fusion (FUSION)*. IEEE, 2014, pp. 1–7.

[47] X. Shen and P. K. Varshney, "Sensor selection based on generalized information gain for target tracking in large sensor networks," *IEEE Transactions on Signal Processing*, vol. 62, no. 2, pp. 363–375, 2014.

[48] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[49] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.

[50] S. Ono, "L_0 gradient projection," *IEEE Transactions on Image Processing*, vol. 26, no. 4, pp. 1554–1564, 2017.

[51] S. Boyd, N. Parikh, and E. Chu, *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc, 2011.

[52] X. Cheng, M. Zeng, and X. Liu, "Feature-preserving filtering with l0 gradient minimization," *Computers & Graphics*, vol. 38, pp. 150–157, 2014.

[53] M. Storath, A. Weinmann, and L. Demaret, "Jump-sparse and sparse recovery using potts functionals," *IEEE Transactions on Signal Processing*, vol. 62, no. 14, pp. 3654–3666, 2014.

[54] R. M. Nguyen and M. S. Brown, "Fast and effective l_0 gradient minimization by region fusion," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 208–216.

[55] R. Matsuoka, S. Kyochi, S. Ono, and M. Okuda, "Joint sparsity and order optimization based on ADMM with non-uniform group hard thresholding," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 65, no. 5, pp. 1602–1613, 2017.

[56] G. Li and T. K. Pong, "Global convergence of splitting methods for nonconvex composite optimization," *SIAM Journal on Optimization*, vol. 25, no. 4, pp. 2434–2460, 2015.

[57] M. Hong, Z.-Q. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," *SIAM Journal on Optimization*, vol. 26, no. 1, pp. 337–364, 2016.

[58] L. Yang, T. K. Pong, and X. Chen, "Alternating direction method of multipliers for a class of nonconvex and nonsmooth problems with applications to background/foreground extraction," *SIAM Journal on Imaging Sciences*, vol. 10, no. 1, pp. 74–110, 2017.

[59] Y. Wang, W. Yin, and J. Zeng, "Global convergence of ADMM in nonconvex nonsmooth optimization," *Journal of Scientific Computing*, vol. 78, no. 1, pp. 29–63, 2019.

[60] A. Nemirovski, "Interior point polynomial time methods in convex programming," 2004. [Online]. Available: http://www2.isye.gatech.edu/nemirovs/Lect_IPM.pdf

[61] NOAA/OAR/ESRL, "Noaa optimal interpolation (oi) sea surface temperature (sst) v2." [Online]. Available: https://www.esrl.noaa.gov/psd/data/gridded/data.noaa.oisst.v2.html

Takayuki Nagata received the B.S. and M.S. degrees in mechanical and aerospace engineering from Tokai University, Hiratsuka, Japan, in 2015 and 2017, respectively. He received the Ph.D. degree in aerospace engineering from Tohoku University, Sendai, Japan, in 2020. From 2018 to 2020, he was a Research Fellow of the Japan Society for the Promotion of Science (JSPS) at Tohoku University, Japan. He is currently a Project Assistant Professor at Tohoku University, Sendai, Japan.

Keigo Yamada received the B.S. degree in physics and M.S. degree in aerospace engineering from Tohoku University, Sendai, Japan, in 2019. He is currently a Ph.D. student with the Department of Aerospace Engineering with Tohoku University, Sendai, Japan.

Taku Nonomura received the B.S. degree in mechanical and aerospace engineering from Nagoya University, Nagoya, Japan, in 2003, and the Ph.D. degree in aerospace engineering from the University of Tokyo, Japan in 2008. He is currently an Associate Professor with the Department of Aerospace Engineering at Tohoku University, Sendai, Japan.

Kumi Nakai received the Ph.D. degree in mechanical systems engineering from Tokyo University of Agriculture and Technology, Tokyo, Japan, in 2020. From 2017 to 2020, she was a Research Fellow of the Japan Society for the Promotion of Science (JSPS) at Tokyo University of Agriculture and Technology, Tokyo, Japan. She is currently a researcher with National Institute of Advanced Industrial Science and Technology, Japan.

Yuji Saito received the B.S. degree in mechanical engineering, and the Ph.D. degree in mechanical space engineering from Hokkaido University, Japan, in 2018. He is currently an Assistant Professor with the Department of Aerospace Engineering at Tohoku University, Sendai, Japan.

Shunsuke Ono (S'11–M'15) received a B.E. degree in Computer Science in 2010 and M.E. and Ph.D. degrees in Communications and Computer Engineering in 2012 and 2014 from the Tokyo Institute of Technology, respectively.

From April 2012 to September 2014, he was a Research Fellow (DC1) of the Japan Society for the Promotion of Science (JSPS). He is currently an Associate Professor in the Department of Computer Science, School of Computing, Tokyo Institute of Technology. From October 2016 to March 2020 and from October 2021 to present, he was/is a Researcher of Precursory Research for Embryonic Science and Technology (PRESTO), Japan Science and Technology Corporation (JST), Tokyo, Japan. His research interests include signal processing, image analysis, remote sensing, mathematical optimization, and data science.

Dr. Ono received the Young Researchers' Award and the Excellent Paper Award from the IEICE in 2013 and 2014, respectively, the Outstanding Student Journal Paper Award and the Young Author Best Paper Award from the IEEE SPS Japan Chapter in 2014 and 2020, respectively, the Funai Research Award from the Funai Foundation in 2017, the Ando Incentive Prize from the Foundation of Ando Laboratory in 2021, and the Young Scientists' Award from MEXT in 2022. He has been an Associate Editor of IEEE TRANSACTIONS ON SIGNAL AND INFORMATION PROCESSING OVER NETWORKS since 2019.