IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 14 January 2024, accepted 29 January 2024, date of publication 5 February 2024, date of current version 13 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3361941

METHODS

Flexible and Comprehensive Framework of Element Selection Based on Nonconvex Sparse Optimization

TAIGA KAWAMURA, NATSUKI UENO, (Member, IEEE), AND NOBUTAKA ONO, (Senior Member, IEEE)
Graduate School of Systems Design, Tokyo Metropolitan University, Tokyo 191-0065, Japan
Corresponding author: Taiga Kawamura (kawamura-taiga@ed.tmu.ac.jp)

This work was supported by the Japan Science and Technology Agency (JST) CREST, Japan, under Grant JPMJCR19A3.

ABSTRACT We propose an element selection method for high-dimensional data that is applicable to a wide range of optimization criteria in a unifying manner. Element selection is a fundamental technique for reducing dimensionality of high-dimensional data by simple operations without the use of scalar multiplication. Restorability is one of the commonly used criteria in element selection, and the element selection problem based on restorability is formulated as a minimization problem of a loss function representing the restoration error between the original data and the restored data. However, conventional methods are applicable only to a limited class of loss functions such as ℓ_2 norm loss. To enable the use of a wide variety of criteria, we reformulate the element selection problem as a nonconvex sparse optimization problem and derive the optimization algorithm based on Douglas–Rachford splitting method. The proposed algorithm is applicable to any loss function as long as its proximal operator is available, e.g., ℓ_1 norm loss and ℓ_∞ norm loss as well as ℓ_2 norm loss. We conducted numerical experiments using artificial and real data, and their results indicate that the above loss functions are successfully minimized by the proposed algorithm.

INDEX TERMS Dimensionality reduction, element selection, sparse optimization, proximal operator, Douglas–Rachford splitting method.

I. INTRODUCTION

The recent development of machine learning technology has led to an increasing number of situations where we deal with data having not only a large number of samples but also a large number of dimensions (elements) per sample. Even though high-dimensional data are preferred in that they have rich information, they often cause increased computational costs in signal processing applications or a lack of interpretability. Dimensionality reduction is one promising technique for solving such problems [1], [2]. This technique aims to reduce the dimension of the target data with as little loss of information as possible. Among a wide variety of dimensionality reduction techniques proposed in the literature, including the well-known principal component analysis (PCA) [3], [4], [5], we focus on the element selection method in this paper as one of the most simple approaches. Unlike other methods such as the PCA, element selection does not require scalar multiplications. Therefore, this method is effective for signal processing applications in devices with limited computational resources, such as hearing aids [6]. In addition, the element selection method is useful in extracting a small number of representative elements having critical information of the whole data. The optimal placement problems of multiple sensors [7], [8], [9] are also often formulated as element selection problems.

The element selection method requires an optimization criterion that defines how to determine which elements to

The associate editor coordinating the review of this manuscript and approving it for publication was Dominik Strzalka.

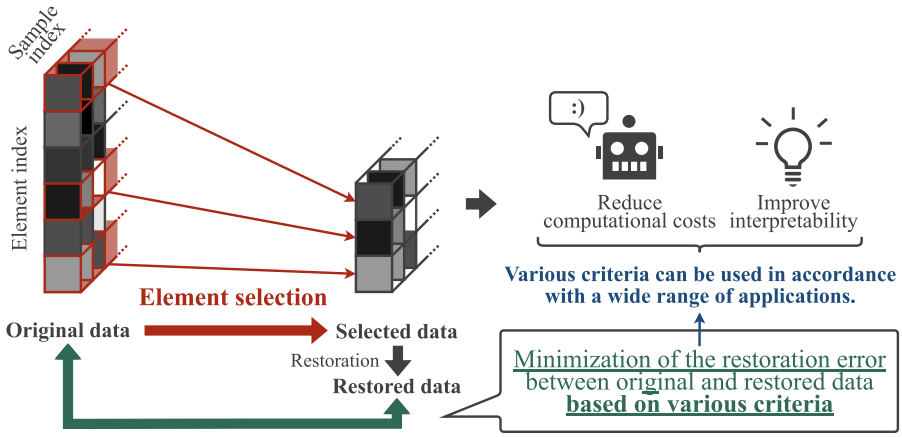

FIGURE 1. Outline of the proposed method.

select. Restorability is a commonly used criterion, based on which the elements are selected so that the original data can be restored well from the selected elements. We also refer to information-theoretic frameworks [10], [11] as other versatile approaches, which are however beyond the scope of this study. The element selection problem based on restorability is generally formulated as a combinatorial optimization problem that minimizes a loss function representing the restoration error between original and restored data. Since this optimization problem is computationally difficult, or almost impossible, to solve by a full-search algorithm in cases of high-dimensional data, it is commonly solved approximately by the greedy method [7], [9] or the convex relaxation method [9], [11]. These methods require the optimal restoration obtained in closed form for any given selected subset of elements and are therefore applicable only to specific loss functions such as the ℓ_2 norm loss function. However, there are many situations where other norms are preferable to the ℓ_2 norm for the loss function, as well as in other problems of machine learning or signal processing. For example, the ℓ_1 norm loss and Huber loss [12] are used to induce robustness against outliers [13], [14], [15]. On the other hand, the ℓ_∞ norm loss is used to decrease the worst-case error [16], [17], [18]. Therefore, there is a need for a flexible formulation of the element selection problem that is applicable to various loss functions in accordance with the purpose of applications and the structure of the target data.

To achieve this goal, we propose an element selection algorithm based on nonconvex sparse optimization that can be applied to a wide class of loss functions in a unified manner. The outline of the proposed element selection method is shown in Fig. 1. The restorability-based element selection problem is formulated as a minimization problem with respect to two matrices representing the operations of the selection and restoration of the data. Then, this minimization problem is reformulated equivalently as a minimization problem with respect to a single matrix corresponding to the product of the above two matrices, instead of dealing with them separately. Because this matrix has nonzero entries only in the specific columns corresponding to the selected elements, the reformulated optimization problem can be regarded as a nonconvex sparse optimization problem [19], [20], [21], [22], [23], [24]. Therefore, well-established algorithms using proximal operators are available in the proposed formulation. We derive an iterative algorithm based on the Douglas–Rachford splitting method [25], [26]. The proposed algorithm is applicable to a wide class of loss functions including ℓ_1, ℓ_2, and ℓ_∞ norm loss functions, as long as their proximal operators are available in closed form. In other words, the proposed method extends the scope of applicable loss functions in the element selection problem, as compared with the conventional methods. We also evaluated the proposed method and compared it with the conventional greedy method (only available for ℓ_2 norm loss) by numerical experiments using artificial and real data [27], [28]. The results show that the proposed algorithm improved the restorability of the data with respect to the corresponding optimization criterion.

Part of this paper is based on our previous conference paper [29], but this paper extends the previous work in several ways. First, we introduce the data normalization in the proposed algorithm to guarantee the invariance under uniform scaling of data, which contributes significantly to the robustness to variations in parameter values and the scaling of the data. In addition, we extend the variety of loss functions and target data in the experiments to evaluate the general validity of the proposed method, whereas only limited conditions were investigated in our previous work. We also modify slightly the experimental conditions to improve fairness in the comparison of these methods, which is described in Section IV in detail.

The rest of this paper is organized as follows. In Section II, the formulation of the element selection problem based on restorability is presented, and the difficulty in applying conventional methods to various loss functions other than the ℓ_2 norm loss is discussed. In Section III, the proposed method

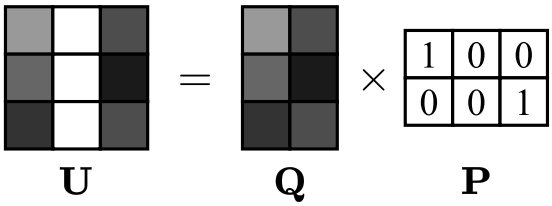

FIGURE 2. Example of selection matrix, restoration matrix, and their product.

is described with examples for several commonly used loss functions. In Section IV, results of numerical experiments are reported. Finally, Section V concludes this paper.

II. PROBLEM STATEMENT

Element selection is the operation of extracting specific elements from a given high-dimensional data, which can be expressed as

$$\mathbf{y} = \mathbf{P}\mathbf{x}, \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^N$ is the original data having N-dimensional features, $\mathbf{y} \in \mathbb{R}^K$ ($K < N$) is the selected data, and $\mathbf{P} \in \{0,1\}^{K\times N}$ is a row-selection matrix (hereafter referred to simply as a selection matrix). It should be noted that the element selection process requires no scalar multiplications, unlike other linear dimensionality reduction methods including the PCA, where a dense real-valued matrix $\mathbf{P} \in \mathbb{R}^{K\times N}$ is used in (1).

The objective of the element selection problem is to obtain the optimal selection matrix $\mathbf{P}$ on the basis of some criterion. In this study, we focus on the restorability, one of the widely used criteria. Let $\mathbf{X} \in \mathbb{R}^{N\times M}$ denote the collection of the data used to determine the optimal selection matrix, each of whose column vector denotes a single N-dimensional data with M being a number of samples. Then, the collection of the selected data $\mathbf{Y} \in \mathbb{R}^{K\times M}$ is given by

$$\mathbf{Y} = \mathbf{P}\mathbf{X}. \tag{2}$$

The restorability of $\mathbf{Y}$ is evaluated by the loss function $D(\hat{\mathbf{X}}, \mathbf{X})$ representing the restoration error between the original data $\mathbf{X}$ and the restored data $\hat{\mathbf{X}} \in \mathbb{R}^{N\times M}$ from $\mathbf{Y}$. Here, the restored data $\hat{\mathbf{X}}$ are assumed to be obtained by a linear transformation of $\mathbf{Y}$ as $\hat{\mathbf{X}} = \hat{\mathbf{Q}}\mathbf{Y}$, where $\hat{\mathbf{Q}} \in \mathbb{R}^{N\times K}$ is a restoration matrix defined as the optimal solution of

$$\min_{\mathbf{Q}\in\mathbb{R}^{N\times K}} \quad D(\mathbf{Q}\mathbf{Y}, \mathbf{X}). \tag{3}$$

By summarizing (2) and (3), we can formulate the element selection problem as the following optimization problem:

$$\min_{\mathbf{P}\in\mathcal{S},\mathbf{Q}\in\mathbb{R}^{N\times K}} \quad D(\mathbf{Q}\mathbf{P}\mathbf{X}, \mathbf{X}), \tag{4}$$

where $\mathcal{S}$ denotes the set of $K \times N$ selection matrices.

The element selection problem (4) is a combinatorial optimization, and the full search is computationally difficult for high-dimensional data because $\mathcal{S}$ has $\binom{N}{K} = \frac{N!}{K!(N-K)!}$ elements. Therefore, the greedy method and the convex relaxation method have been widely studied as efficient approaches to obtain an approximate solution. The greedy method solves the subproblem K times, where the elements are selected one by one, resulting in an approximate solution of (4). The convex relaxation method obtains an approximate solution of (4) by constraining $\mathbf{P} \in [0,1]^{K\times N}$ s.t. $\sum_{n=1}^{N} p_{k,n} = 1$ instead of constraining $\mathbf{P} \in \mathcal{S}$. However, these conventional methods require that $D(\cdot,\cdot)$ in (4) be a specific loss function for which (3) can be solved in closed form. For example, when the loss function is the ℓ_2 norm loss, i.e., $D(\hat{\mathbf{X}}, \mathbf{X}) = \frac{1}{M}\|\hat{\mathbf{X}} - \mathbf{X}\|_2^2$, where $\|\cdot\|_2$ denotes the ℓ_2 norm (Frobenius norm) of the matrix, (3) can be solved in closed form for any fixed selection matrix $\mathbf{P}$ as

$$\begin{aligned}\hat{\mathbf{Q}} &= \mathbf{X}\mathbf{Y}^\mathsf{T}\left(\mathbf{Y}\mathbf{Y}^\mathsf{T}\right)^{-1} \\ &= \mathbf{X}\mathbf{X}^\mathsf{T}\mathbf{P}^\mathsf{T}\left(\mathbf{P}\mathbf{X}\mathbf{X}^\mathsf{T}\mathbf{P}^\mathsf{T}\right)^{-1},\end{aligned} \tag{5}$$

where $(\cdot)^\mathsf{T}$ and $(\cdot)^{-1}$ denote the transpose and inverse of the matrix, respectively. However, for most other loss functions, e.g., the ℓ_1 and ℓ_∞ norm loss functions, (3) cannot be solved in closed form as in (5). In such cases, the greedy method requires the optimization of (3) for as many times as the number of candidate elements to add one element, and the convex relaxation method is even not applicable because its formulation is essentially based on the substitution of (5) to (4). Therefore, a more flexible framework is needed for a wider application range of loss functions.

III. PROPOSED METHOD

To overcome the limitations of the conventional methods described in Section II, we propose a comprehensive framework of element selection that is applicable to a wide class of loss functions, including those other than the ℓ_2 norm loss.

A. REFORMULATION OF ELEMENT SELECTION PROBLEM

The key idea is to regard (4) as an optimization problem of a single matrix $\mathbf{U} \in \mathbb{R}^{N\times N}$ defined as

$$\mathbf{U} = \mathbf{Q}\mathbf{P}. \tag{6}$$

An example of the selection matrix $\mathbf{P}$, the restoration matrix $\mathbf{Q}$, and the matrix $\mathbf{U}$ when $N = 3$ and $K = 2$ is shown in Fig. 2. As seen in Fig. 2, $\mathbf{U}$ has nonzero elements only for K columns corresponding to the indices of the selected elements, i.e., $\mathbf{U}$ is a sparse matrix with respect to the column vectors. Note that $\mathbf{U}$ and $(\mathbf{P}, \mathbf{Q})$ have a one-to-one correspondence except for the permutation of the selected elements. Taking these properties of $\mathbf{U}$ into account, the optimization problem in (4) can be reformulated equivalently as

$$\min_{\mathbf{U}\in\mathbb{R}^{N\times N}} \quad D(\mathbf{U}\mathbf{X}, \mathbf{X}) \quad \text{s.t.} \quad \Phi(\mathbf{U}) \le K, \tag{7}$$

Algorithm 1 Proposed Algorithm

Input: $\mathbf{X}, \rho, K, J$
Output: $\mathbf{P}, \mathbf{Q}$
1: Initialize $\mathbf{V}, \mathbf{Z}$
2: $\mathbf{X} \leftarrow \frac{1}{\|\mathbf{X}\|_2}\mathbf{X}$
3: **for** $k = 1, 2, \cdots, K$ **do**
4: **repeat** J times
5: $\mathbf{U} \leftarrow \mathcal{H}_k(\mathbf{V})$
6: $\mathbf{W} \leftarrow \mathrm{prox}_{\rho f}(\mathbf{Z})$
7: $[\mathbf{V}\ \mathbf{Z}] \leftarrow [\mathbf{V}\ \mathbf{Z}] + \Pi_{\mathbf{X}}(2[\mathbf{U}\ \mathbf{W}] - [\mathbf{V}\ \mathbf{Z}]) - [\mathbf{U}\ \mathbf{W}]$
8: **end repeat**
9: **end for**
10: Calculate $\mathbf{P}$ from the nonzero column indices of $\mathbf{U}$
11: $\mathbf{Q} \leftarrow \mathbf{U}\mathbf{P}^{\mathsf{T}}$

where $\Phi(\cdot)$ denotes the number of nonzero column vectors in the matrix.

B. OPTIMIZATION ALGORITHM

The optimization problem of (7) is nonconvex due to the constraint $\Phi(\mathbf{U}) \leq K$. Although there have been many studies [30], [31], [32], [33], [34] dealing with algorithms for nonconvex problem with guaranteed convergence, several conditions on the constraint (or the regularizing function) are required therein, and it is still difficult to directly apply these theories to the constraint $\Phi(\mathbf{U}) \leq K$ in (7) without relaxation. For a similar (but slightly different) optimization problem to (7), on the other hand, an efficient algorithm [21], [22], [23], also known as the SPADE algorithm, is proposed as a state-of-the-art audio declipping method and showed high performances although the convergence is not guaranteed from a theoretical perspective. Motivated by the algorithms in these studies, we propose the optimization algorithm for solving (7) based on the Douglas–Rachford splitting method and the SPADE algorithm. The proposed algorithm is summarized in Algorithm 1, and its derivation is provided in Appendix A. Here, $\mathrm{prox}_{\rho f}(\cdot)$ is the proximal operator [35] of $f(\cdot)$ given by

$$\mathrm{prox}_{\rho f}(\mathbf{Z}) = \arg\min_{\mathbf{A}} \ f(\mathbf{A}) + \frac{1}{2\rho}\|\mathbf{A} - \mathbf{Z}\|_2^2, \tag{8}$$

where f is defined as

$$f(\mathbf{Z}) = D(\mathbf{Z}, \mathbf{X}), \tag{9}$$

and $\rho \in (0, \infty)$ is a step size parameter. Several examples of loss functions $D(\cdot, \cdot)$ and their corresponding proximal operators $\mathrm{prox}_{\rho f}(\cdot)$ are given in Section III-C. On the other hand, $\mathcal{H}_k(\cdot)$ and $\Pi_{\mathbf{X}}(\cdot)$ are given independently of the loss functions $D(\cdot, \cdot)$, and correspond to the projection onto the sets $\{\mathbf{U} \in \mathbb{R}^{N\times N} \mid \Phi(\mathbf{U}) \leq k\}$ and $\{[\mathbf{U}\ \mathbf{W}] \in \mathbb{R}^{N\times(N+M)} \mid \mathbf{U}\mathbf{X} = \mathbf{W}\}$, respectively. In particular, $\mathcal{H}_k(\cdot)$ is an operation that sets all but the k largest column vectors in the ℓ_2 norm of the input matrix to zero, and $\Pi_{\mathbf{X}}(\cdot)$ is given by

$$\begin{aligned}\Pi_{\mathbf{X}}([\mathbf{U}\ \mathbf{W}]) &= [\mathbf{U}\ \mathbf{W}] - [\mathbf{U}\ \mathbf{W}]\begin{bmatrix}\mathbf{X}\\ -\mathbf{I}\end{bmatrix}\left(\begin{bmatrix}\mathbf{X}\\ -\mathbf{I}\end{bmatrix}\right)^{\dagger} \\ &= [\mathbf{U}\ \mathbf{W}] - [\mathbf{U}\ \mathbf{W}]\begin{bmatrix}\mathbf{X}\\ -\mathbf{I}\end{bmatrix}(\mathbf{X}^{\mathsf{T}}\mathbf{X}+\mathbf{I})^{-1}\begin{bmatrix}\mathbf{X}\\ -\mathbf{I}\end{bmatrix}^{\mathsf{T}},\end{aligned} \tag{10}$$

where $\mathbf{I}$ is the $M \times M$ identity matrix and $(\cdot)^{\dagger}$ denotes the pseudo-inverse matrix.

The fourth to eighth lines in Algorithm 1 correspond to the optimization of the problem

$$\min_{\mathbf{U}\in\mathbb{R}^{N\times N}, \mathbf{W}\in\mathbb{R}^{N\times M}} \ D(\mathbf{W}, \mathbf{X}) \quad \text{s.t.} \ \ \Phi(\mathbf{U}) \leq k, \ \mathbf{U}\mathbf{X} = \mathbf{W}, \tag{11}$$

i.e., the optimization problem (7) with K replaced by k. These lines are iterated for increasing $k = 1, \ldots, K$ instead of setting the desired number of selected elements, K, from the beginning. This strategy is incorporated for better convergence and is motivated by the SPADE algorithm [21], [22], [23]. The performance improvement by the use of increasing k was also demonstrated in our previous study [29]. Although convergence of the algorithm is not guaranteed in a strict sense due to the fifth line (i.e., projection to a nonconvex set), we did not observe divergence when the parameter ρ is appropriately set in the experiments in Section IV; further theoretical analysis remains as future work.

The normalization in the second line in Algorithm 1 is introduced to guarantee the invariance of the algorithm against the scalar multiple of $\mathbf{X}$ and to facilitate the tuning of the parameter ρ. As seen from (8), the proximal operator is generally not a homogeneous function. Therefore, even if we have found the good parameter ρ for $\mathbf{X}$, it does not necessarily work well for $a\mathbf{X}$ with general $a \in \mathbb{R}$. To avoid such dependence of the optimal parameter ρ on the scale of the data, $\mathbf{X}$ is first normalized by its ℓ_2 norm in the proposed algorithm. This is the main difference between the proposed algorithm and the algorithm in our previous work [29], and its validity is demonstrated by the preliminary experiments in Section IV-B.

C. LOSS FUNCTIONS AND PROXIMAL OPERATORS

As long as $\mathrm{prox}_{\rho f}(\cdot)$ corresponding to $D(\cdot, \cdot)$ can be calculated in closed form, various loss functions can be used for element selection in a unified manner in Algorithm 1. By comparing (8) and (3), we find that (8) includes no linear terms with respect to the optimization variable, although it has an additional quadratic term. Thus, a wide class of useful loss functions, including the ℓ_1, ℓ_2, and ℓ_∞ norm loss functions, have a closed-form solution for (8), even though many of them do not have a closed-form solution for (3), e.g., the ℓ_1 and ℓ_∞ norm loss functions. This means that the proposed framework allows a flexible formulation of the element selection problem that cannot be

TABLE 1. Examples of loss functions $D(\cdot,\cdot)$ and corresponding proximal operators $\mathrm{prox}_{\rho f}$.

$D(\mathbf{Z},\mathbf{X})$	$[\mathrm{prox}_{\rho f}(\mathbf{Z})]_{n,m}$
$\frac{1}{M}\|\mathbf{Z}-\mathbf{X}\|_1$	$\max(0, 1-\frac{\rho}{M}\lvert[\mathbf{Z}-\mathbf{X}]_{n,m}\rvert^{-1})\cdot[\mathbf{Z}-\mathbf{X}]_{n,m}+[\mathbf{X}]_{n,m}$
$\frac{1}{M}\|\mathbf{Z}-\mathbf{X}\|_2^2$	$(2\frac{\rho}{M}+1)^{-1}(2\frac{\rho}{M}[\mathbf{X}]_{n,m}+[\mathbf{Z}]_{n,m})$
$\|\mathbf{Z}-\mathbf{X}\|_\infty$	$\mathrm{sign}([\mathbf{Z}-\mathbf{X}]_{n,m})\cdot\max(\lvert[\mathbf{Z}-\mathbf{X}]_{n,m}\rvert, s)+[\mathbf{X}]_{n,m}$ where $s\in\mathbb{R}$ is such that $\sum_{n=1}^{N}\sum_{m=1}^{M}\max(0,[\mathbf{Z}-\mathbf{X}]_{n,m}-s)=\rho$

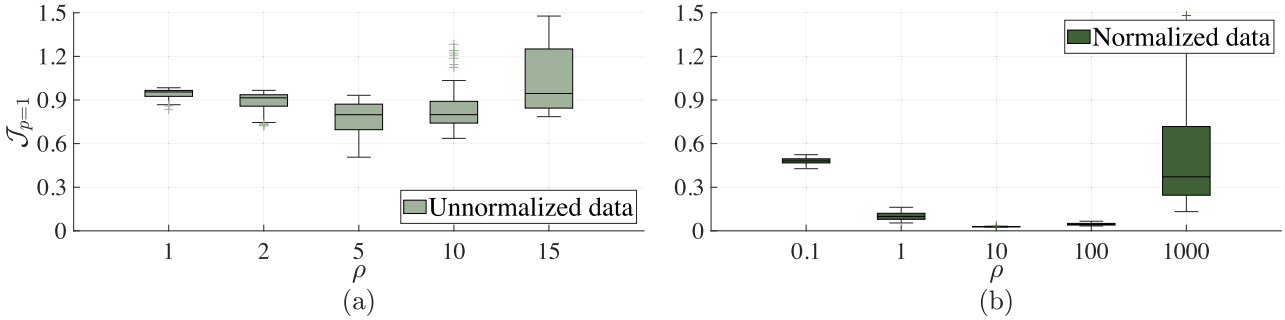

FIGURE 3. Box plot of restoration error (optimization criterion) $\mathcal{J}_{p=1}$ obtained using Proposed ℓ_1 (a) with and (b) without normalization in the preliminary experiment.

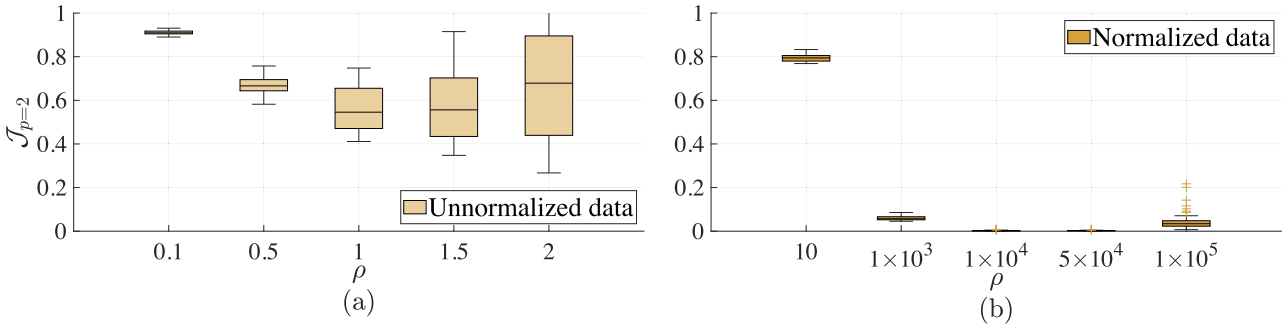

FIGURE 4. Box plot of restoration error (optimization criterion) $\mathcal{J}_{p=2}$ obtained using Proposed ℓ_2 (a) with and (b) without normalization in the preliminary experiment.

dealt with in conventional methods. Various useful functions and their proximal operators are reviewed in [35], [36], and [37].

Three examples of the loss functions, i.e., ℓ_1, ℓ_2, and ℓ_∞ norm loss functions, and their proximal operators are listed in Table 1. Here, $\|\cdot\|_1$ denotes the ℓ_1 norm of the matrix given by

$$\|\mathbf{X}\|_1=\sum_{n=1}^{N}\sum_{m=1}^{M}|x_{n,m}|, \tag{12}$$

where $x_{n,m}$ (and also $[\mathbf{X}]_{n,m}$ in Table 1) denote the (n,m)th entry of $\mathbf{X}$ and $|\cdot|$ is the absolute value, and $\|\cdot\|_\infty$ denotes the ℓ_∞ norm of the matrix given by

$$\|\mathbf{X}\|_\infty=\max_{1\le n\le N,1\le m\le M}|x_{n,m}|. \tag{13}$$

All loss functions in Table 1 are scaled so that they are invariant against the repetition of data, i.e., $D(\mathbf{Z},\mathbf{X})=D([\mathbf{Z}\ \mathbf{Z}],[\mathbf{X}\ \mathbf{X}])$.

IV. NUMERICAL EXPERIMENTS

We conducted numerical experiments to evaluate the proposed method. First, preliminary experiments were conducted using artificially generated data to evaluate the sensitivity for the parameters in the proposed method and demonstrate validity of the normalization process. Then, experiments using real data of sea surface temperature and clothing images were conducted for more practical evaluation and compared with the conventional greedy method.

A. CONDITIONS

The proposed methods for three different criteria shown in Table 1 were evaluated, which are denoted by Proposed ℓ_1, Proposed ℓ_2, and Proposed ℓ_∞ in the order from above in Table 1. In addition, the greedy method (Greedy) and the random method (Random) were evaluated for comparison. In the proposed methods, the number of iterations J was set to 500 for Proposed ℓ_1, 100 for Proposed ℓ_2, and 1000 for Proposed ℓ_∞. The step size parameter ρ was set on the basis of the preliminary experiment in Section IV-B. The

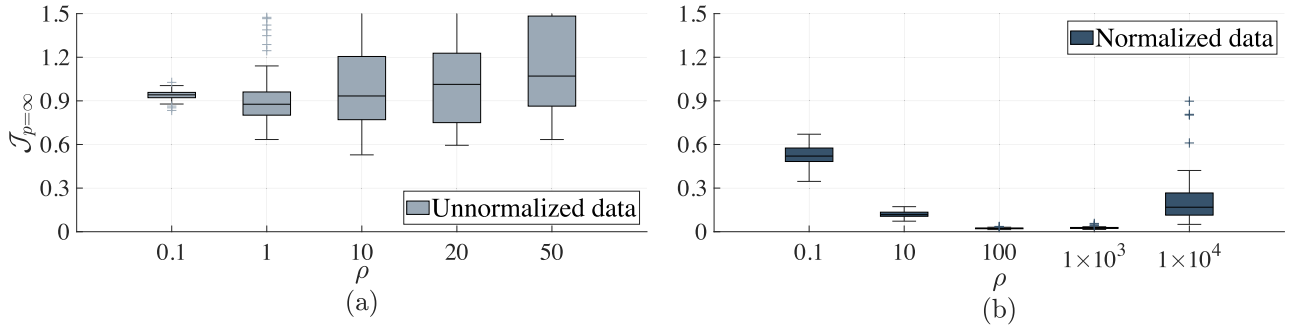

FIGURE 5. Box plot of restoration error (optimization criterion) $\mathcal{J}_{p=\infty}$ obtained using Proposed ℓ_∞ (a) with and (b) without normalization in the preliminary experiment.

initializations for the first line in Algorithm 1 were all set to 0. In Greedy, K elements were selected one by one to minimize the ℓ_2 norm loss, starting from an empty set. Here, $\mathbf{X}\mathbf{X}^\mathsf{T}$ in (5) were replaced by $\mathbf{X}\mathbf{X}^\mathsf{T} + \lambda$ for numerical stability, where λ is the largest eigenvalue of $\mathbf{X}$ multiplied by 10^{-8}. In Random, all elements were selected uniformly at random.

As the evaluation criteria, we used the restoration error (optimization criterion) $\mathcal{J}_p$ defined as

$$\mathcal{J}_p(\mathbf{P}, \mathbf{Q}) = \begin{cases} \dfrac{\|\mathbf{QPX} - \mathbf{X}\|_p^p}{\|\mathbf{X}\|_p^p} & (p = 1, 2) \\ \dfrac{\|\mathbf{QPX} - \mathbf{X}\|_p}{\|\mathbf{X}\|_p} & (p = \infty) \end{cases} \tag{14}$$

to evaluate the behavior of the optimization algorithms and the restoration error (evaluation criterion) $\mathcal{E}_p$ defined as

$$\mathcal{E}_p(\mathbf{P}) = \begin{cases} \dfrac{\|\hat{\mathbf{Q}}\mathbf{PX} - \mathbf{X}\|_p^p}{\|\mathbf{X}\|_p^p} & (p = 1, 2) \\ \dfrac{\|\hat{\mathbf{Q}}\mathbf{PX} - \mathbf{X}\|_p}{\|\mathbf{X}\|_p} & (p = \infty) \end{cases} \tag{15}$$

to evaluate the restorability of the selected elements. The restoration error (optimization criterion) $\mathcal{J}_p$ was calculated using $\mathbf{P}$ and $\mathbf{Q}$ optimized on the basis of the *optimization criteria* in the iterative algorithm, i.e., the output variables in Algorithm 1 for Proposed and $\mathbf{P}$ and $\mathbf{Q}$ obtained by (5) for Greedy. On the other hand, $\mathcal{E}_p$ was calculated for the selection matrix $\mathbf{P}$ using the $\hat{\mathbf{Q}}$ optimized on the basis of the *evaluation criteria*, which minimizes the right-hand side of (15) with fixed $\mathbf{P}$. When calculating $\mathcal{E}_{p=2}$, $\hat{\mathbf{Q}}$ was obtained in closed form given by (5). For $\mathcal{E}_{p=1}$ and $\mathcal{E}_{p=\infty}$, $\hat{\mathbf{Q}}$ was optimized using Algorithm 2 described in Appendix B. To mitigate bias, the average of the restoration errors (evaluation criterion) $\mathcal{E}_p$ over 50 runs was used in Random.

B. PRELIMINARY EXPERIMENT

The preliminary experiment was conducted using the artificially generated data to confirm the behavior with respect to the variation of ρ and the effectiveness of the normalization in Algorithm 1. The artificially generated data $\mathbf{X}$ was given by $\mathbf{X} = a(\mathbf{C} + \mathbf{N})$, where $\mathbf{C} \in \mathbb{R}^{N\times M}$ is a K-rank matrix and $\mathbf{N} \in \mathbb{R}^{N\times M}$ is the random noise, each of whose elements independently follows the normal distribution $\mathcal{N}(0, 0.1)$, and

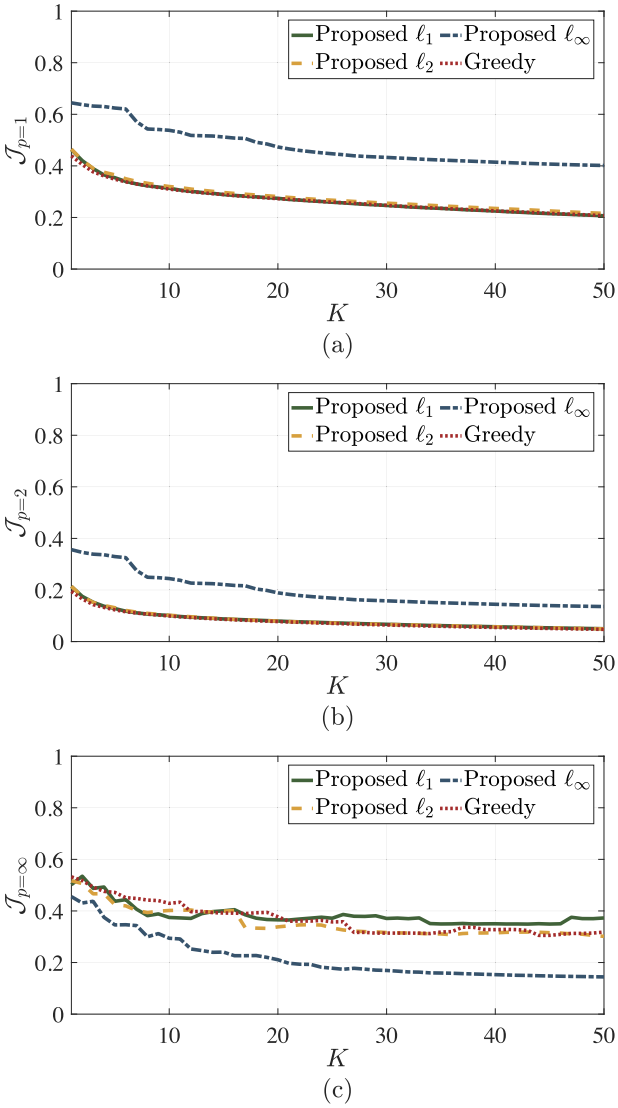

FIGURE 6. Restoration errors (optimization criteria) (a) $\mathcal{J}_{p=1}$, (b) $\mathcal{J}_{p=2}$, and (c) $\mathcal{J}_{p=\infty}$ for each K when training using sea surface temperature data.

$a \in [1, 10]$ is a parameter that controls the scale of $\mathbf{X}$, which follows the uniform distribution $U(1, 10)$. The matrix $\mathbf{C}$ was defined as the product of $\mathbf{A} \in \mathbb{R}^{K\times M}$, each of whose elements independently follows the Laplace distribution Laplace(0, 1)

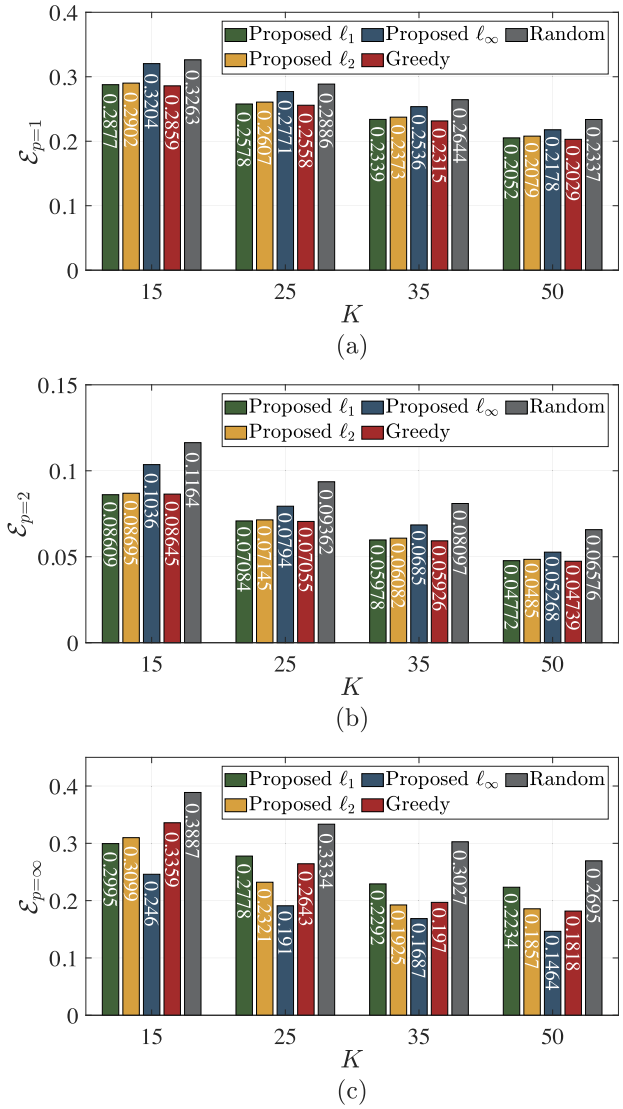

FIGURE 7. Restoration errors (evaluation criteria) (a) $\mathcal{E}_{p=1}$, (b) $\mathcal{E}_{p=2}$, and (c) $\mathcal{E}_{p=\infty}$ at $K = 15$, 25, 35, and 50 using sea surface temperature data.

and $\mathbf{B} \in \mathbb{R}^{N\times K}$, each of whose elements independently follows $\mathcal{N}(0, 1)$, i.e., $\mathbf{C} = \mathbf{BA}$. Here, we set $N = 50$, $M = 4000$, and $K = 10$.

Each of the three proposed methods (Proposed ℓ_1, Proposed ℓ_2 and Proposed ℓ_∞) was evaluated for five different values of ρ with and without the normalization in the second line of Algorithm 1. The results of Proposed ℓ_1, Proposed ℓ_2 and Proposed ℓ_∞ are shown in Figs. 3, 4 and 5 as box plots of the restoration error (optimization criterion) $\mathcal{J}_{p=1,2 \text{ and } \infty}$ for 50 generated $\mathbf{X}$, respectively. By comparing (a) without normalization and (b) with normalization in these figures, we found that the normalization results in smaller restoration errors (optimization criteria) for many ρ and less variation within a single ρ. These results indicate that the normalization eliminates the need for detailed tuning of ρ depending on each data set. From these results, the following experiments were conducted using $\rho = 10$ for Proposed ℓ_1, $\rho = 10000$ for Proposed ℓ_2, and $\rho = 100$ for Proposed ℓ_∞.

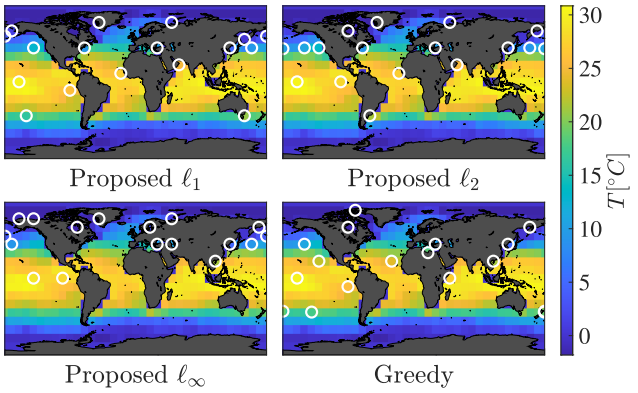

FIGURE 8. Virtual observation points selected at $K = 15$ using sea surface temperature data.

C. EXPERIMENTS USING SEA SURFACE TEMPERATURE DATA

One application of element selection, that is, improvement of interpretability by extracting representative data, was evaluated by experiments using sea surface temperature data. The data consists of weekly mean temperatures for 1500 weeks from December 31, 1989 to September 29, 2018 on a grid of 1 degree across the globe, with 360 × 180 virtual observation points. Among them, we used 10 degrees each for 36×18 observation points with excluding 223 observation points located on land. Therefore, the number of elements, N, was 425, and the number of samples, M, was 1500. Different numbers of selected elements were evaluated up to $K = 50$.

First, a line graph of the restoration error (optimization criterion) $\mathcal{J}_p$ at each K for the three proposed methods and Greedy is shown in Fig. 6. The subplots show the restoration errors (optimization criteria) $\mathcal{J}_{p=1}$ in Fig. 6(a), $\mathcal{J}_{p=2}$ in Fig. 6(b), and $\mathcal{J}_{p=\infty}$ in Fig. 6(c). In Fig. 6(a) and Fig. 6(b), the difference between Proposed ℓ_1, Proposed ℓ_2, and Greedy was small. In Fig. 6(c), however, Proposed ℓ_∞ achieved the smallest restoration error (optimization criterion), indicating that the proposed method can select elements that reduce the value of the corresponding loss function. Next, Fig. 7 shows the restoration error (evaluation criterion) $\mathcal{E}_p$ for the selection matrix $\mathbf{P}$ obtained with each method at $K = 15$, 25, 35, and 50. The subplots show the restoration errors (evaluation criteria) $\mathcal{E}_{p=1}$ in Fig. 7(a), $\mathcal{E}_{p=2}$ in Fig. 7(b), and $\mathcal{E}_{p=\infty}$ in Fig. 7(c). By comparing Figs. 6 and 7, one can see that the differences between each method for $\mathcal{E}_p$ were smaller than that for $\mathcal{J}_p$. This is because the restoration matrix was optimized on the basis of the evaluation criteria in $\mathcal{E}_p$. Nevertheless, Proposed ℓ_∞ still achieved the smallest restoration errors (evaluation criterion) for $\mathcal{E}_{p=\infty}$. Conversely, Proposed ℓ_∞ showed larger restoration errors (evaluation criteria) than Proposed ℓ_1, Proposed ℓ_2, and Greedy for $\mathcal{E}_{p=1}$ and $\mathcal{E}_{p=2}$, although these three methods did not show significant differences. These results indicate that the choice of different optimization criteria indeed contributes to the improvement of the corresponding restorability. Finally, Fig. 8 shows the virtual observation points selected for each

TABLE 2. Restoration error (evaluation criterion) $\mathcal{E}_p$ (left, ℓ_1 norm; middle, ℓ_2 norm; right, ℓ_∞ norm) in training data.

Method	$K=25$			$K=50$			$K=75$			$K=100$		
	$\mathcal{E}_{p=1}$	$\mathcal{E}_{p=2}$	$\mathcal{E}_{p=\infty}$	$\mathcal{E}_{p=1}$	$\mathcal{E}_{p=2}$	$\mathcal{E}_{p=\infty}$	$\mathcal{E}_{p=1}$	$\mathcal{E}_{p=2}$	$\mathcal{E}_{p=\infty}$	$\mathcal{E}_{p=1}$	$\mathcal{E}_{p=2}$	$\mathcal{E}_{p=\infty}$
Proposed ℓ_1	0.3146	0.1236	0.9037	0.2529	0.0885	0.8740	0.2161	0.0701	0.8847	**0.1890**	0.0583	0.6979
Proposed ℓ_2	0.3170	0.1242	0.9480	0.2568	0.0888	0.9094	0.2186	0.0701	0.7179	0.1917	0.0580	0.7040
Proposed ℓ_∞	0.3475	0.1448	**0.7876**	0.2823	0.1045	**0.7000**	0.2379	0.0819	**0.6530**	0.2118	0.0700	**0.6376**
Greedy	**0.3122**	**0.1216**	0.9849	**0.2520**	**0.0878**	1.0047	**0.2159**	**0.0696**	0.8954	0.1899	**0.0578**	0.6467
Random	0.3587	0.1560	0.9783	0.2886	0.1113	0.9295	0.2520	0.0910	0.8961	0.2231	0.0761	0.8670

TABLE 3. Restoration error (evaluation criterion) $\mathcal{E}_p$ (left, ℓ_1 norm; middle, ℓ_2 norm; right, ℓ_∞ norm) in test data.

Method	$K=25$			$K=50$			$K=75$			$K=100$		
	$\mathcal{E}_{p=1}$	$\mathcal{E}_{p=2}$	$\mathcal{E}_{p=\infty}$	$\mathcal{E}_{p=1}$	$\mathcal{E}_{p=2}$	$\mathcal{E}_{p=\infty}$	$\mathcal{E}_{p=1}$	$\mathcal{E}_{p=2}$	$\mathcal{E}_{p=\infty}$	$\mathcal{E}_{p=1}$	$\mathcal{E}_{p=2}$	$\mathcal{E}_{p=\infty}$
Proposed ℓ_1	0.3094	0.1215	1.3403	0.2502	0.0880	1.4809	0.2139	0.0704	1.3725	**0.1875**	0.0593	1.3073
Proposed ℓ_2	0.3122	0.1223	**1.2350**	0.2533	0.0882	1.3781	0.2157	0.0703	**1.3119**	0.1890	0.0592	1.3393
Proposed ℓ_∞	0.3417	0.1423	1.2462	0.2778	0.1034	**1.2495**	0.2351	0.0825	1.3483	0.2100	0.0713	1.3978
Greedy	**0.3075**	**0.1200**	1.3180	**0.2486**	**0.0873**	1.4677	**0.2131**	**0.0697**	1.6108	0.1878	**0.0586**	**1.2943**
Random	0.3540	0.1540	2.6900	0.2857	0.1123	4.9133	0.2505	0.0926	5.7889	0.2223	0.0779	4.4831

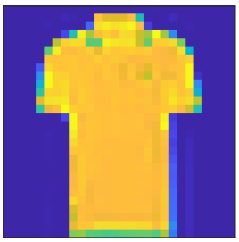

FIGURE 9. Original clothing image.

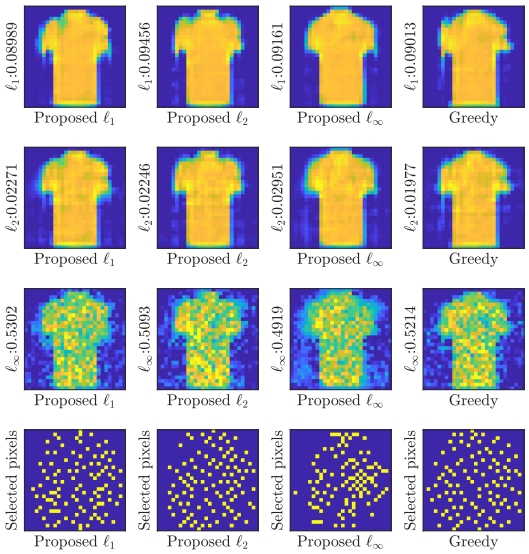

FIGURE 10. Restored clothing images at $K = 100$. Below each image is the name of the corresponding method, and the left of each image in the first to third lines is the loss function used to optimize $\hat{\mathbf{Q}}$ and the restoration error (evaluation criterion) for that image.

method. Some of the selected observation points were located at similar positions in each method, but others were located at distinct positions.

D. EXPERIMENTS USING IMAGE DATA

We also conducted experiments using clothing images from Fashion-MNIST dataset [27] to evaluate the restoration performance for unknown test data. This data consists of 10000 images of 28×28 pixels, and we divided them into 5000 training data used for optimization of selected elements and 5000 test data for evaluation. Therefore, in element selection, the number of elements, N, corresponded to 784 pixels, the number of samples, M, corresponded to 5000 images, and the size of the training data, denoted by $\mathbf{X}_{\text{train}}$, was 784×5000. For evaluation, the size of the test data, denoted by $\mathbf{X}_{\text{test}}$, was also 784×5000. Four different values of K, i.e., $K = 25$, 50, 75, and 100 were evaluated. Restoration performance was evaluated by $\mathcal{E}_p$ for $\mathbf{X}_{\text{train}}$ and $\mathbf{X}_{\text{test}}$, but in both cases $\mathbf{P}$ and $\mathbf{Q}$ were optimized using $\mathbf{X}_{\text{train}}$.

First, the restoration errors (evaluation criteria) $\mathcal{E}_p$ obtained using train and test data are shown in Tables 2 and 3, respectively. In Table 2, Proposed ℓ_∞ performed best with $\mathcal{E}_{p=\infty}$, whereas the performance of the other methods, except Random, remained flat, similar to the results in Section IV-C. On the other hand, the results on the test data in Table 3 show that $\mathcal{E}_{p=\infty}$ was higher than 1. This is due to $\mathcal{E}_{p=\infty}$ evaluating the restoration error (evaluation criterion) of the single worst pixel in the test data. Proposed ℓ_1, Proposed ℓ_2, and Greedy show the same trend as in Table 2 and achieved lower restoration errors (evaluation criteria) than Random. Next, one sample of the training data and the restored images of that sample at $K = 100$ are shown in Fig. 9 and Fig. 10, respectively. Fig. 10 shows the results obtained using Proposed ℓ_1, Proposed ℓ_2, Proposed ℓ_∞, and Greedy, starting from the right column, and using the ℓ_1, ℓ_2, and ℓ_∞ norms as criteria for optimization of $\hat{\mathbf{Q}}$ in calculation of $\mathcal{E}_p$, starting from the top row, with the fourth row showing the selected elements. By comparing row by row, we confirmed significant differences in the behavior of restored images depending on the optimization criteria of the restoration matrix $\hat{\mathbf{Q}}$. Although differences for columns were not as significant as those for rows, the amount of blur around the object was different for the four methods; Proposed ℓ_1 shows

Algorithm 2 Algorithm for Optimization of $\mathbf{Q}$

Input: $\mathbf{Y}, \sigma, I$
Output: $\mathbf{Q}$
1: Initialize: $\mathbf{W}, \mathbf{D}$
2: **repeat** I times
3: $\quad \mathbf{Q} \leftarrow (\mathbf{W}-\mathbf{D})\mathbf{Y}^\mathsf{T}(\mathbf{Y}\mathbf{Y}^\mathsf{T})^{-1}$
4: $\quad \mathbf{W} \leftarrow \mathrm{prox}_{\sigma\psi}(\mathbf{Q}\mathbf{Y}+\mathbf{D})$
5: $\quad \mathbf{D} \leftarrow \mathbf{D}+\mathbf{Q}\mathbf{Y}-\mathbf{W}$
6: **end repeat**

the least amount of blur. One possible reason is that pixels that rarely take large values, that is, marginal pixels in this data set, were relatively ignored in the ℓ_1 norm criterion. In addition, from the fourth row, we confirmed that Proposed ℓ_∞ differs significantly from the other three methods in selected pixels. These results demonstrated the effect of the use of different optimization criteria in element selection, which are enabled by the proposed method.

V. CONCLUSION

We proposed an element selection method that can apply a wide range of loss functions in a unified manner based on nonconvex optimization. The proposed method enables element selection with loss functions that are not applicable to conventional methods, which was also confirmed by numerical experiments. Future works include the theoretical extension of our framework to criteria other than restorability and the application of the proposed element selection method to signal processing in devices with limited computational resources.

APPENDIX A
DERIVATION OF THE PROPOSED ALGORITHM

First, (11) is equivalently reformulated as

$$\min_{\mathbf{U}\in\mathbb{R}^{N\times N},\mathbf{W}\in\mathbb{R}^{N\times M}} f(\mathbf{W})+g_k(\mathbf{U})+h(\mathbf{U},\mathbf{W}), \tag{16}$$

where $f(\cdot)$, $g_K(\cdot)$, and $h(\cdot)$ are defined as

$$f(\mathbf{W}) = D(\mathbf{W},\mathbf{X}), \tag{17}$$

$$g_k(\mathbf{U}) = \begin{cases} 0 & (\Phi(\mathbf{U})\le k) \\ \infty & (\Phi(\mathbf{U})>k) \end{cases}, \tag{18}$$

and

$$h(\mathbf{U},\mathbf{W}) = \begin{cases} 0 & (\mathbf{W}=\mathbf{U}\mathbf{X}) \\ \infty & (\mathbf{W}\neq\mathbf{U}\mathbf{X}) \end{cases}, \tag{19}$$

respectively. Then, the objective function of (16) can be regarded as the sum of two functions $f(\mathbf{W})+g_k(\mathbf{U})$ and $h(\mathbf{U},\mathbf{W})$, each of which is a function with respect to $(\mathbf{U},\mathbf{W})$ whose proximal operator is available in closed form. In particular, $\mathcal{H}_k(\cdot)$ and $\Pi_{\mathbf{X}}(\cdot)$ are the proximal operators of $g_k(\cdot)$ and $h(\cdot,\cdot)$, respectively. Therefore, the Douglas–Rachford splitting method can be applied to (11), which results in the fourth to eighth lines of Algorithm 1. These update rules are applied to increasing $k=1,2,\ldots,K$ to finally solve the optimization problem (7).

APPENDIX B
CALCULATION OF THE OPTIMAL RESTORATION MATRIX

We have to solve (3) with $\mathbf{Y}=\mathbf{P}\mathbf{X}$ to obtain $\hat{\mathbf{Q}}$ in (15). Since, $\hat{\mathbf{Q}}$ cannot be obtained in closed form for $p=1$ and $p=\infty$, we used Algorithm 2 with

$$\psi(\mathbf{Z}) = \begin{cases} \dfrac{1}{M}\|\mathbf{Z}-\mathbf{X}\|_1 & (p=1) \\ \|\mathbf{Z}-\mathbf{X}\|_\infty & (p=\infty) \end{cases}. \tag{20}$$

This algorithm is derived by applying the alternating direction method of multipliers (ADMM) to the following optimization problem:

$$\min_{\mathbf{Q}\in\mathbb{R}^{N\times K},\mathbf{W}\in\mathbb{R}^{N\times M}} \psi(\mathbf{W}) \quad \text{s.t. } \mathbf{W}=\mathbf{Q}\mathbf{Y}. \tag{21}$$

Throughout the experiment, the parameters I and σ were set to 100 and 10 for $p=1$ and 1000 and 100 for $p=\infty$, respectively.

REFERENCES

[1] J. P. Cunningham and Z. Ghahramani, "Linear dimensionality reduction: Survey, insights, and generalizations," *J. Mach. Learn. Res.*, vol. 16, no. 89, pp. 2859–2900, 2015.
[2] S. Velliangiri, S. Alagumuthukrishnan, and S. I. T. Joseph, "A review of dimensionality reduction techniques for efficient computation," *Proc. Comput. Sci.*, vol. 165, pp. 104–111, 2019.
[3] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdiscipl. Rev., Comput. Statist.*, vol. 2, no. 4, pp. 433–459, 2010.
[4] K. Pearson, "On lines and planes of closest fit to systems of points in space," *London, Edinburgh, Dublin Phil. Mag. J. Sci.*, vol. 2, no. 11, pp. 559–572, 1901.
[5] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, Sep. 1936.
[6] C. Haruta and N. Ono, "A low-computational DNN-based speech enhancement for hearing aids based on element selection," in *Proc. 29th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2021, pp. 1025–1029.
[7] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies," *J. Mach. Learn. Res.*, vol. 9, no. 8, pp. 235–284, 2008.
[8] 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 J.*, vol. 21, no. 13, pp. 15248–15257, Jul. 2021.
[9] T. Nishida, N. Ueno, S. Koyama, and H. Saruwatari, "Region-restricted sensor placement based on Gaussian process for sound field estimation," *IEEE Trans. Signal Process.*, vol. 70, pp. 1718–1733, 2022.
[10] M. Chu, H. Haussecker, and F. Zhao, "Scalable information-driven sensor querying and routing for ad hoc heterogeneous sensor networks," *Int. J. High Perform. Comput. Appl.*, vol. 16, no. 3, pp. 293–313, Aug. 2002.
[11] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 451–462, Feb. 2009.
[12] P. J. Huber, "Robust estimation of a location parameter," *Ann. Math. Statist.*, vol. 35, no. 1, pp. 73–101, Mar. 1964.
[13] Q. Ke and T. Kanade, "Robust L_1 norm factorization in the presence of outliers and missing data by alternative convex programming," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2005, pp. 739–746.
[14] H. Avron, A. Sharf, C. Greif, and D. Cohen-Or, "ℓ_1-sparse reconstruction of sharp point set surfaces," *ACM Trans. Graph.*, vol. 29, no. 5, pp. 1–12, Oct. 2010.
[15] P. Indyk and M. Ruzic, "Near-optimal sparse recovery in the ℓ_1 norm," in *Proc. 49th Annu. IEEE Symp. Found. Comput. Sci.*, Oct. 2008, pp. 199–207.

[16] J. Tan, D. Baron, and L. Dai, "Signal estimation with low infinity-norm error by minimizing the mean p-norm error," in *Proc. 48th Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2014, pp. 1–5.
[17] J. Tan, D. Baron, and L. Dai, "Wiener filters in Gaussian mixture signal estimation with ℓ_∞-Norm error," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6626–6635, Oct. 2014.
[18] X. Jiang, J. Chen, H. C. So, and X. Liu, "Large-scale robust beamforming via ℓ_∞-minimization," *IEEE Trans. Signal Process.*, vol. 66, no. 14, pp. 3824–3837, Jul. 2018.
[19] Z. Wen, C. Yang, X. Liu, and S. Marchesini, "Alternating direction methods for classical and ptychographic phase retrieval," *Inverse Problems*, vol. 28, no. 11, Nov. 2012, Art. no. 115010.
[20] S. Ono and I. Yamada, "Color-line regularization for color artifact removal," *IEEE Trans. Comput. Imag.*, vol. 2, no. 3, pp. 204–217, Sep. 2016.
[21] C. Gaultier, S. Kitić, R. Gribonval, and N. Bertin, "Sparsity-based audio declipping methods: Selected overview, new algorithms, and large-scale evaluation," *IEEE/ACM Trans. Audio, Speech, Lang., Process.*, vol. 29, pp. 1174–1187, 2021.
[22] P. Záviška, P. Rajmic, A. Ozerov, and L. Rencker, "A survey and an extensive evaluation of popular audio declipping methods," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 1, pp. 5–24, Jan. 2021.
[23] S. Kitić, N. Bertin, and R. Gribonval, "Sparsity and cosparsity for audio declipping: A flexible non-convex approach," in *Proc. Int. Conf. Latent Variable Anal. Signal Separaiton (LVA/ICA)*, 2015, pp. 243–250.
[24] A. Themelis and P. Patrinos, "Douglas-rachford splitting and ADMM for nonconvex optimization: Tight convergence results," *SIAM J. Optim.*, vol. 30, no. 1, pp. 149–181, Jan. 2020.
[25] J. Douglas and H. H. Rachford, "On the numerical solution of heat conduction problems in two and three space variables," *Trans. Amer. Math. Soc.*, vol. 82, no. 2, pp. 421–439, 1956.
[26] J. Eckstein and D. P. Bertsekas, "On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, nos. 1–3, pp. 293–318, Apr. 1992.
[27] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms," 2017, *arXiv:1708.07747*.
[28] R. W. Reynolds, N. A. Rayner, T. M. Smith, and D. C. Stokes, "An improved in situ and satellite SST analysis for climate," *J. Climate*, vol. 15, no. 13, pp. 1609–1625, Jul. 2002.
[29] T. Kawamura, N. Ueno, and N. Ono, "Element selection with wide class of optimization criteria using non-convex sparse optimization," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2023, pp. 1–5.
[30] E. C. Marques, N. Maciel, L. Naviner, H. Cai, and J. Yang, "A review of sparse recovery algorithms," *IEEE Access*, vol. 7, pp. 1300–1322, 2019.
[31] T. Sun, H. Jiang, L. Cheng, and W. Zhu, "Iteratively linearized reweighted alternating direction method of multipliers for a class of nonconvex problems," *IEEE Trans. Signal Process.*, vol. 66, no. 20, pp. 5380–5391, Oct. 2018.
[32] T. Sun, H. Jiang, and L. Cheng, "Convergence of proximal iteratively reweighted nuclear norm algorithm for image processing," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 5632–5644, Dec. 2017.
[33] I. Selesnick, "Sparse regularization via convex analysis," *IEEE Trans. Signal Process.*, vol. 65, no. 17, pp. 4481–4494, Sep. 2017.
[34] E. J. Candes, M. B. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, pp. 877–905, Oct. 2008.
[35] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 127–239, Nov. 2014.
[36] R. Chartrand, "Shrinkage mappings and their induced penalty functions," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 1026–1029.
[37] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. New York, NY, USA: Springer, 2011, pp. 185–212.

TAIGA KAWAMURA received the B.E. degree in engineering from the Tokuyama College, National Institute of Technology, Yamaguchi, Japan, in 2022. He is currently pursuing the M.S. degree in engineering with Tokyo Metropolitan University, Tokyo, Japan. He is a Student Member of the Acoustical Society of Japan (ASJ). His research interest includes acoustic signal processing.

NATSUKI UENO (Member, IEEE) received the B.E. degree in engineering from Kyoto University, Kyoto, Japan, in 2016, and the M.S. and Ph.D. degrees in information science and technology from The University of Tokyo, Tokyo, Japan, in 2018 and 2021, respectively. He is currently a Project Assistant Professor with Tokyo Metropolitan University, Tokyo. His research interests include spatial audio and acoustic signal processing.

NOBUTAKA ONO (Senior Member, IEEE) received the B.E., M.S., and Ph.D. degrees in mathematical engineering and information physics from The University of Tokyo, Tokyo, Japan, in 1996, 1998, and 2001, respectively.

He was a Research Associate with The University of Tokyo, in 2001, and became a Lecturer, in 2005. He was also an Associate Professor with the National Institute of Informatics, Tokyo, in April 2011, and became a Professor, in 2017. In 2017, he was with Tokyo Metropolitan University, Hino, Japan. He is the author or coauthor of more than 310 articles in international journal articles and peer-reviewed conference proceedings. His research interests include acoustic signal processing, especially microphone array processing, source localization and separation, machine learning, and optimization algorithms.

Dr. Ono is a Senior Member of IEEE Signal Processing Society and a member of the Acoustical Society of Japan (ASJ), Institute of Electronics, Information and Communications Engineers (IEICE), Information Processing Society of Japan (IPSJ), and Society of Instrument and Control Engineers (SICE), Tokyo. He was a Tutorial Speaker with ISMIR 2010 and ICASSP 2018. He was a recipient of the Awaya Award from ASJ, in 2007, the Igarashi Award at the Sensor Symposium from IEEJ, in 2004, the Best Paper Award at IEEE ISIE, in 2008, the Measurement Division Best Paper Award from SICE, in 2013, the Best Paper Award in IEEE IS3C, in 2014, the Excellent Paper Award in IIHMSP, in 2014, the Unsupervised Learning ICA Pioneer Award from SPIE, DSS, in 2015, the Sato Paper Award from ASJ, in 2000 and 2018, two TAF Telecom System Technology Awards, in 2018, the Best Paper Award in APSIPA ASC, in 2018 and 2021, and the Sadaoki Furui Prize Paper Award from APSIPA, in 2021. He is also the Chair of IEEE Signal Processing Society Tokyo Joint Chapter. He was the Chair of Signal Separation Evaluation Campaign Evaluation Committee, in 2013 and 2015, the Technical Program Chair of IWAENC 2018, the General Chair of DCASE 2020 workshop, and a member of IEEE Audio and Acoustic Signal Processing Technical Committee, from 2014 to 2019. From 2012 to 2015, he was an Associate Editor of IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING.