

# Kalman Filtered Compressive Sensing Using Pseudo-Measurements

Haibin Zhao<sup>1</sup>, Christopher Funk<sup>2</sup>, Benjamin Noack<sup>2</sup>, Uwe Hanebeck<sup>3</sup>, and Michael Beigl<sup>1</sup>

**Abstract**— In this paper, we combine the Kalman filter and compressive sensing using pseudo-measurements in order to reduce the number of measurements usually required by the Kalman filter. To overcome the non-sparsity of the measurement vectors, we make use of the change of their coefficients when represented in a certain basis, reduce the dimensionality of the coefficients, and learn a sparse basis for the measurement vectors. We further improve our proposed method by introducing dynamic weighting of the pseudo-measurements, by aiding compressive measurement reconstruction with Kalman filter estimates and by employing iterative versions of this process. Simulations show that our approach achieves a 37% improvement with respect to the mean-square error compared to the traditional Kalman filter with the same number of measurements. Our approach yields better results when the measurement noise is relatively large compared to the system noise, and it significantly improves the accuracy of state estimation in sensor networks with low sensor precision.

**Index Terms**— Kalman Filter, Compressive Sensing, Kalman Filtered Compressive Sensing, Pseudo Measurements

## I. INTRODUCTION

State estimation for dynamic systems, i.e., the inference of the system state from a series of incomplete and noisy measurements, is one of the most persistent research topics in the field of signal processing. The Kalman filter (KF) [1] is an efficient recursive algorithm to solve this problem. Based on the KF, there are many extensions such as the extended Kalman filter [2], the unscented Kalman filter [3], and the distributed Kalman filter [4]. In some cases, the number of measurements required by the Kalman filter is large, e.g., when the system models are derived from partial differential equations using the finite difference method (FDM) [5], [6], [7].

Compressive Sensing (CS) [8] has emerged as a new framework for signal acquisition, with which the signal of interest can be recovered from fewer measurements than required by the Nyquist-Shannon theorem [9]. There are many applications of CS, e.g., a single-pixel camera [10], a conventional camera with a coded aperture [11], a new method for Magnetic Resonance Imaging [12], and a new

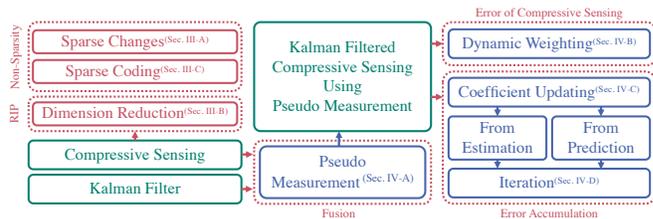


Fig. 1. Structure of this paper. Red frames describe the problems; blue frames express the solutions to the problems, while green frames represent relevant theories.

DNA microarray [13]. However, the prerequisite for using CS is that the signal must be sparse or compressible [14] when resolved in a certain basis, and the sensing matrix must satisfy the restricted isometry property (RIP) [15].

Guided by an example, this paper proposes a method to fuse the KF and CS to alleviate these disadvantages. We solve three problems that arise when applying CS to non-compressible signals by focusing on the changes of the signals, reducing the dimension of the coefficients, and using sparse coding to represent the signal sparsely. The KF and CS are then combined to Kalman filtered compressive sensing (KFCS) using pseudo-measurements (PMs). We handle the uncertainty of the PMs (i.e., the error of CS reconstruction) and the error-accumulation resulting from working with the coefficient changes by introducing dynamic weighting, a coefficient update, and an iterated update.

Fig. 1 shows the organization of this paper: Section I-A discusses previous works and their relationship to our own; Section II introduces the notation, the KF, CS, and the main problems faced; Section III proposes solutions to the main problems, and Section IV introduces our proposal, the KFCS using PMs. Section V gives a simulation and discusses the effectiveness of our approach. In Section VI, we summarize our work and discuss future issues.

### A. Related Work

There are some existing works on the combination of CS with other methods [16]. In [17] the concept of a homotopy continuation was introduced into CS to find a more accurate solution faster and the resulting method was applied to streaming signals such as videos [18]. In [19], Bayesian compressive sensing was proposed, where measurements are adaptively selected, and the variance of the errors is estimated simultaneously, which leads to significantly lower errors. Distributed compressive sensing was introduced in [20], which aims to estimate states based on multiple signals from distributed coding algorithms in

<sup>1</sup>Chair for Pervasive Computing Systems, Institute for Telematics, Department of Informatics, Karlsruhe Institute of Technology (KIT), Germany.

<sup>2</sup>Institute for Intelligent Cooperating Systems, Otto von Guericke University Magdeburg, Germany.

<sup>3</sup>Intelligent Sensor-Actuator-Systems Laboratory, Institute for Anthropomatics and Robotics, Department of Informatics, Karlsruhe Institute of Technology (KIT), Germany.

haibin.zhao@kit.edu, christopher.funk@ovgu.de,  
benjamin.noack@ieee.org, uwe.hanebeck@ieee.org,  
michael.beigl@kit.edu

sensor networks. Vaswani [21] combined CS with least squares (LS) by replacing CS on the observation with CS on the LS-residual. The LS-residual is computed using a previous estimate of the support set. In [22] a new technique for efficiently acquiring and reconstructing signals based on convolution with a fixed FIR filter having random taps was proposed. In [23], Vaswani proposed Kalman filtered compressive sensing (KFCS), which combines the KF and CS for the first time. After the KF innovation, the filter error norm (FEN) is calculated to detect if there is an addition to the support set. If there is, the new addition is estimated using CS. In [24], the same author proposed the modified KFCS. The author utilized a so-called temporary Kalman prediction and an update to reduce the computational cost. The method works particularly well for slowly changing signals.

Building on their works, we focus on combining the advantages of the KF and CS, making CS more applicable to non-compressible signals and reducing the number of measurements required for state estimation.

## II. BACKGROUND

### A. Notation

Lower case letters  $x \in \mathbb{R}$  denote scalar variables, and additional underlining  $\underline{x} \in \mathbb{R}^N$  indicates  $N$ -dimensional vectors. An additional superscript  $\hat{x}$  indicates the estimated values of  $x$ . Bold upper case letters  $\mathbf{X} \in \mathbb{R}^{M \times N}$  denote  $M \times N$ -dimensional matrices. Moreover, we utilize the letter  $n \in \mathbb{N}$  as subscript to denote the  $n$ -th time step, i.e.,  $n$  denotes the time step  $t = t_n = n \cdot \Delta t$ , where  $\Delta t$  is the length of a time step. Matrix  $\mathbf{C}^x$  denotes the covariance matrix of vector  $\underline{x}$ , i.e.,  $\mathbf{C}^x = \mathbb{E}\{(\hat{\underline{x}} - \underline{x})(\hat{\underline{x}} - \underline{x})^T\}$ , where  $\mathbb{E}\{\underline{x}\}$  indicates the expected value of  $\underline{x}$ . To distinguish the signal, measurements, and compressed measurements, we let  $\underline{f} \in \mathbb{R}^{\mathfrak{N}}$  denote the value of the signal at each discrete spatial point, where  $\mathfrak{N}$  indicates the total number of discrete spatial points. The letters  $\underline{y} \in \mathbb{R}^S$  and  $\tilde{\underline{y}} \in \mathbb{R}^M$  denote the measurements and compressed measurements respectively, where  $S$  refers to the number of sensors, and  $M$  is the number of compressed measurements.

### B. Temperature Distribution

Throughout this paper, we use the monitoring of the one-dimensional temperature distribution of a beam as an example. The mathematical model of heat conduction is given by the diffusion equation [25],

$$\underbrace{\frac{\partial f}{\partial t}(x, t) = k \frac{\partial^2 f}{\partial x^2}(x, t)}_{\text{diffusion equation}} + \underbrace{u(x, t)}_{\text{stimulus}}, \quad (1)$$

where  $x$  denotes the spatial coordinate,  $t$  denotes time, and  $f(x, t)$  indicates the temperature. We solve this partial differential equation by utilizing the finite difference method (FDM) [5]. FDM is a numerical method for calculating approximate solutions of (partial) differential equations by approximating the derivative terms using finite differences, such as

$$\frac{\partial f}{\partial t}(x, t + \Delta t) \approx \frac{f(x, t + \Delta t) - f(x, t)}{\Delta t}.$$

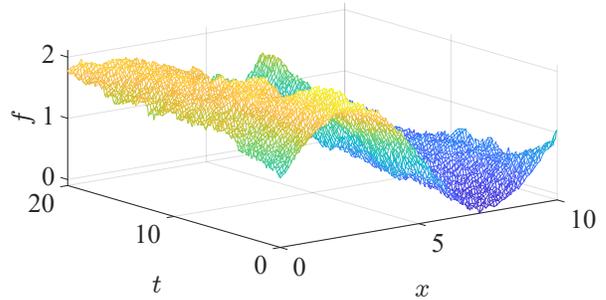


Fig. 2. Synthetic temperature distribution.

In the following, we choose the time resolution  $\Delta t = 0.1$  s and spatial resolution  $\Delta x = \frac{10}{2^{10}}$  cm. The total length is  $L = 10$  cm, which means  $\mathfrak{N} = L/\Delta x = 1024$ , and the diffusivity is  $k = 0.1$  cm<sup>2</sup>/s. As shown in Fig. 2, the initial temperature distribution is set to be a sine-shaped function and the stimulus is

$$u(x, t) = 0.1 \sin\left(t - \frac{\pi}{4}\right) \cdot \delta(x - 3) \\ - 0.2 \sin(t) \cdot \delta(x - 5) + 0.01t \cdot \delta(x - 7).$$

In addition, we add zero-mean Gaussian system noise with variance  $\sigma_w^2 = 0.005$ , and measurement noise with variance  $\sigma_v^2 = 0.025$  to each node. The time span of the simulation is  $t \in [0, 20]$  seconds, i.e.,  $n \in [0, 1, \dots, \mathfrak{T}]$ , where  $\mathfrak{T}$  represents the maximal  $n$ , corresponding to  $t = 20$ s. After the finite difference approximation, we can transform (1) into the form

$$\underline{f}_{n+1} = \mathfrak{A} \cdot \underline{f}_n + \mathfrak{B} \cdot \underline{u}_n + \underline{w}_n, \quad (2)$$

where

$$\mathfrak{A} = \begin{bmatrix} 1 + P & 0 & -P \\ -P & Q & -P \\ & \ddots & \ddots & \ddots \\ & & -P & Q & -P \\ & & & -P & 0 & 1 + P \end{bmatrix}^{-1}, \quad \mathfrak{B} = \Delta t \cdot \mathfrak{A} \cdot \underline{u}_n,$$

with  $P = k\Delta t/\Delta x^2$  and  $Q = 1 + 2P$ . Finally, we obtain the temperature distribution as shown in Fig. 2.

### C. Estimation Methods

1) *Kalman Filtering*: In the following, we will introduce the KF briefly and apply it to the example from Section II-B. We use  $\mathbf{A}_n \in \mathbb{R}^{N \times N}$  to denote the state transition matrix and  $\underline{X}_n \in \mathbb{R}^N$  to represent the state variables.  $\underline{U}_n \in \mathbb{R}^N$  describes the input to the system. Moreover,  $\mathbf{\Psi}_n \in \mathbb{R}^{S \times N}$  indicates the observation matrix. The system noise is denoted by  $\underline{w}_n \in \mathbb{R}^N$  and the uncertainty of measurements by  $\underline{v}_n \in \mathbb{R}^S$ . Both follow zero-mean Gaussian distributions with covariance matrices  $\mathbf{C}_n^w$  and  $\mathbf{C}_n^v$ , respectively. Using this notation, the mathematical description of a general linear system is

$$\underline{X}_n = \mathbf{A}_n \cdot \underline{X}_{n-1} + \underline{U}_{n-1} + \underline{w}_{n-1}, \\ \underline{y}_n = \mathbf{\Psi}_n \cdot \underline{X}_n + \underline{v}_n.$$

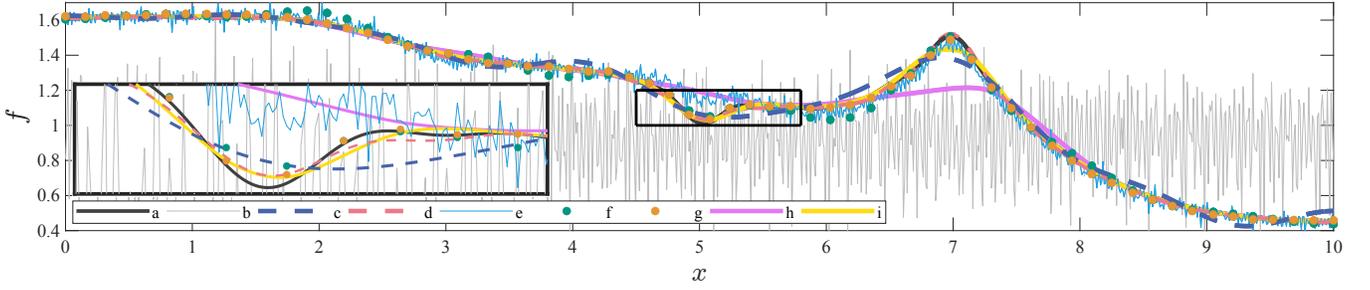


Fig. 3. a: the temperature distribution  $f_{\mathcal{T}}$ ; b: the signal recovered using CS with  $Z = 1024$  and DCT basis; c: the best  $k$ -term approximation for the signal with  $Z = 1024$  and DCT basis; d: the best  $k$ -term approximation for the changes of the signal with  $Z = 1024$  and DCT basis; e: the signal recovered by using CS to estimate the changes of the coefficients with  $Z = 1024$  and DCT basis; f: the signal recovered by using CS to estimate the changes of the coefficients with  $Z = 64$  and DCT basis; g: the signal recovered by using CS to estimate the changes of the coefficients with  $Z = 64$  and SCR basis; h: the signal recovered by Kalman Filter with 12 measurements; i: the signal recovered by Kalman Filter with 64 measurements. The box on the left side in the Figure is a partial enlargement of the rectangular box.

In the KF, the state  $\underline{X}_n$  is the variable to be estimated. At each time step the KF performs two steps: The prediction step

$$\begin{aligned}\underline{X}_{n+1}^p &= \mathbf{A}_n \cdot \hat{\underline{X}}_n + \underline{U}_n, \\ \mathbf{C}_{n+1}^{X^p} &= \mathbf{A}_n \cdot \mathbf{C}_n^{\hat{X}} \cdot \mathbf{A}_n^T + \mathbf{C}_n^w,\end{aligned}$$

and the update step

$$\begin{aligned}\hat{\underline{X}}_{n+1} &= (\mathbf{I} - \mathbf{K}_{n+1} \cdot \Psi_{n+1}) \cdot \underline{X}_{n+1}^p + \mathbf{K}_{n+1} \cdot \underline{y}_{n+1}, \\ \mathbf{C}_{n+1}^{\hat{X}} &= (\mathbf{I} - \mathbf{K}_{n+1} \cdot \Psi_{n+1}) \cdot \mathbf{C}_{n+1}^{X^p},\end{aligned}$$

with the Kalman gain  $\mathbf{K}_{n+1}$  given by

$$\mathbf{K}_{n+1} = \mathbf{C}_{n+1}^{X^p} \cdot \Psi_{n+1}^T \cdot \left( \Psi_{n+1} \cdot \mathbf{C}_{n+1}^{X^p} \cdot \Psi_{n+1}^T + \mathbf{C}_{n+1}^v \right)^{-1}.$$

To apply the KF to the specific case of the example from Section II-B, we assume to have  $S$  equidistant sensors with 10 Hz sampling rate, i.e., the time-step length  $\Delta t = 0.1s$ . Then, we have  $\mathbf{A}_n = \mathfrak{A}_n$ ,  $\underline{X}_n = \underline{f}_n$ , and  $\underline{U}_n = \Delta t \cdot \mathfrak{A}_n \cdot \underline{u}_n$ . As the state variables  $\underline{X}$  are exactly the temperature  $\underline{f}$  at each node, and as the purpose of the observation matrix  $\Psi_n$  is to select some coefficients of  $\underline{f}$  that correspond to positions where there are sensors, the matrix  $\Psi_n$  has the following form:

$$\Psi_n^{i,j} = \begin{cases} 1, & i\text{-th sensor is at } j\text{-th node of the beam,} \\ 0, & \text{otherwise.} \end{cases}$$

For the KF, the estimated temperature distributions from the considered example are shown in Fig. 3 for varying numbers of sensors as violet ( $S = 12$ ) and yellow ( $S = 64$ ) curves (curve h and i).

2) *Compressive Sensing*: In compressive sensing, instead of using a system model, we utilize a basis  $\Theta \in \mathbb{R}^{Z \times Z}$  and associated coefficients  $\underline{z}_n \in \mathbb{R}^Z$  to represent the signals  $\underline{s}_n = \Theta \cdot \underline{z}_n \in \mathbb{R}^Z$ , where  $Z$  is the dimension of the coefficient vectors  $\underline{z}_n$  as well as the number of discrete points of each signal. We recover the signal  $\underline{s}_n$  by estimating the coefficients  $\underline{z}_n$  from the compressed measurements  $\tilde{\underline{y}}_n = \tilde{\Theta}_n \cdot \underline{z}_n$ , where  $\tilde{\Theta}_n \in \mathbb{R}^{M \times Z}$  denotes the sensing matrix. However, in CS,  $M$  is in generally much smaller than  $Z$ , which means that solving  $\tilde{\underline{y}}_n = \tilde{\Theta}_n \cdot \underline{z}_n$  is an

underdetermined problem. Therefore, to obtain a unique solution, additional constraints must be imposed. In the case of CS, the signal must be sparse or compressible, and the sensing matrix must satisfy the restricted isometry property (RIP). To clarify the definition of RIP, we must explain *sparsity* first: A signal is sparse, if  $\|\underline{z}\|_0 \ll \dim(\underline{z})$ . A weaker condition than sparsity is compressibility. A signal is compressible if  $|z_s| \leq C_1 \cdot s^{-q}$  holds for  $s = 1 \dots Z$  and certain constants  $C_1, q > 0$  [26], where  $z_s$  denotes the  $s$ -th-largest coefficient of  $\underline{z}$  in terms of absolute value. The sensing matrix  $\tilde{\Theta}$  is said to satisfy the RIP( $\delta_K, K$ ) if

$$(1 - \delta_K) \|\underline{z}\|_2^2 \leq \left\| \tilde{\Theta} \cdot \underline{z} \right\|_2^2 \leq (1 + \delta_K) \|\underline{z}\|_2^2, \quad (3)$$

holds for all  $\underline{z}$  with sparsity  $K$ , i.e.,  $\|\underline{z}\|_0 = K$ . The estimation process is then described by the optimization problem

$$\begin{aligned}\hat{\underline{z}}_n &= \arg \min_{\underline{z}_n} \|\underline{z}_n\|_0, \\ \text{s.t. } & \left\| \tilde{\underline{y}}_n - \tilde{\Theta}_n \cdot \hat{\underline{z}}_n \right\|_2 \leq \varepsilon,\end{aligned} \quad (4)$$

where  $\varepsilon$  describes the recovery error. It's worth noting that the  $\ell_0$ -norm means the number of non-zero elements in the vector. The  $\ell_0$ -norm optimization problem is non-convex [27], but it can be replaced by an  $\ell_1$ -norm optimization problem [28].

In the context of the example considered in this paper, the dimensionality  $Z$  equals  $\mathfrak{N}$ , as the signal  $\underline{s}_n$  is the temperature distribution  $\underline{f}_n$  and the compressed measurements are generated by randomly selecting elements from  $\underline{y}_n \in \mathbb{R}^S$ , where  $\underline{y}_n$  indicates the  $S$  measurements from  $S$  sensors. This can be implemented by randomly activating the sensors. We utilize the matrix  $\Phi_n \in \mathbb{R}^{M \times S}$  to realize this approach, where

$$\Phi_n^{i,j} = \begin{cases} 1, & j = \underline{\Lambda}_n^i, \\ 0, & \text{otherwise.} \end{cases}$$

In the above equation,  $\underline{\Lambda}_n \in \mathbb{N}^M$  represents the set of indices of the elements selected from  $\underline{y}_n$ . With this definition we

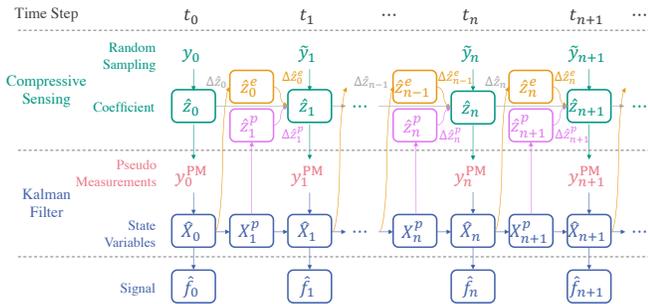


Fig. 4. Structure of KFCS using Pseudo-Measurements.

have the following relationship:

$$\begin{aligned} \tilde{y}_n &= \Phi_n \cdot y_n, \\ \tilde{\Theta}_n &= \Phi_n \cdot \Theta. \end{aligned}$$

The estimated temperature distribution in the example from Section II-B – using CS with  $M = 12$ ,  $Z = \mathfrak{N} = 1024$ , and the discrete cosine transform (DCT) basis – is presented in Fig. 3 as grey curve. This result is unacceptable, i.e., we cannot use the CS directly to estimate this signal. For a quantitative evaluation of the errors, we also calculate the MSE of the recovered signal (see Tab. I), the definition of which can be found in (7). In the following subsection, we will discuss the main problems of combining KF with CS and the problems of CS being used for non-compressible signals.

TABLE I  
MSE FOR CURVES IN FIG. 3

Curve	b	c	d	e
MSE	272	0.067	1.196	1.918
Curve	f	g	h	i
MSE	0.843	0.141	4.311	0.150

#### D. Main Problems

It can be seen from the violet and yellow curves in Fig. 3 that the number of measurements required by the KF to obtain an accurate estimate is large. To overcome this drawback, we will combine the KF with CS. However, due to the limitations of CS, it cannot be directly used to estimate arbitrary signals. An example of this has been presented in Section II-C.2 where the temperature distribution estimated by CS, which is shown as a grey curve in Fig. 3, bears no resemblance to the ground truth temperature distribution (the black curve). Thus, to fuse the KF and the CS, we must solve the following problems: The signal is not sparse or compressible with respect to the DCT basis, and the sensing matrix  $\tilde{\Theta}$  may not satisfy the RIP with sufficient probability.

In the next section, we will consider three approaches to solve the problems mentioned above. After that, we will propose pseudo-measurements (PM) in Section IV to combine the KF and the CS. Subsequently, we will face two further problems: the uncertainty of the PMs (i.e., the error of the CS reconstruction) and the error accumulation caused

by considering coefficient changes. To solve these problems, we will suggest three techniques.

### III. SOLUTIONS TO MAIN PROBLEMS

We solve the aforementioned problems of CS in three steps. First, we focus on the changes of the signal rather than the signal itself, since the presence of the system model implies that the signal changes not arbitrarily but regularly, thus, a basis might exist with which the changes of the signal can be sparsely described. Secondly, we reduce the dimension of the coefficient vector to increase the probability that the sensing matrix meets the RIP. Thirdly, we find a basis that is better suited to the considered signals than the DCT basis by applying sparse coding. In the new basis the considered signals become sparser.

#### A. Sparse Changes

It can be seen from the example, that we can choose the initial temperature and stimulus arbitrarily, but once they are fixed, the temperature cannot evolve arbitrarily, as the temperature must vary following the rule of the system model. I.e., the presence of a system model suggests that the signal changes in a regular way, which indicates that the changes of the signal are more likely to be sparse with respect to a specific basis than the signal itself. With the best  $k$ -term-approximation [29], we can determine the compressibility of signals. The best  $k$ -term-approximation is a method to compress a signal, by which the  $k$  largest terms of a signal's basis expansion remain unchanged and the rest of its terms are set to zero. The less  $k$  could be, the more compressible the signal is, and then, we can recover the signal with less measurements. We find  $k = 10$  is a suitable value, and we can see from the Fig. 3 that the compressibility of the changes of the signal (red dashed line) is better than that of the signal itself (dark blue dashed line). Thus, we perform CS at each time step, not to estimate the coefficients, but to estimate the changes of the coefficients:

$$\begin{aligned} \Delta \hat{z}_{n-1} &= \underset{\Delta z_{n-1}}{\operatorname{argmin}} \|\Delta z_{n-1}\|_1, \\ \text{s.t.} \|\tilde{y}_n - \tilde{\Theta}_n \cdot \underbrace{(\hat{z}_{n-1} + \Delta \hat{z}_{n-1})}_{=\hat{z}_n}\|_2 &\leq \varepsilon. \end{aligned} \quad (5)$$

The use of the coefficient changes is shown in grey in Fig. 4. To find the changes of the coefficients, the initial coefficients must be known. Therefore, all sensors are utilized to measure  $y_0$  to obtain more accurate initial coefficients at the first time step  $t_0$ . Interestingly, we find in the simulation that our algorithm converges even with a zero-initialization, i.e.,  $z_0 = 0$ .

If the sensing matrix  $\tilde{\Theta}$  satisfies the RIP, CS will yield a similar result as the best  $k$ -term approximation, as the two previously mentioned prerequisites (see Section II-C.2) of CS are met. However, CS yields the cyan curve in Fig. 3, which significantly differs from the red dashed line. This difference suggests that the sensing matrix does not satisfy the RIP since the number of columns is much larger than the number of rows.

### B. Dimension Reduction

In [26] Baraniuk et al. proved the following relationship between the number of rows  $M$  and columns  $Z$  of a sensing matrix satisfying the RIP:

$$M \geq \kappa \cdot K \cdot \log \left( \frac{Z}{K} \right),$$

where  $\kappa$  is a variable related to the probability of  $\tilde{\Theta}$  satisfying the RIP. The higher the probability we require, the greater  $M$  should be.

From above, it is clear that reducing the dimension  $Z$  of the coefficients can increase the probability of the sensing matrix satisfying the RIP. Thus, we do not calculate the value at each node by CS, but only the values at some key points. These points can be selected manually or randomly. In this paper, we choose the sensor positions, i.e., the points where the  $\underline{y}_n$  are measured, as key points. The computed values will later be utilized as the measurements of the KF. Therefore, we call them **pseudo-measurements**.

In our example, we choose the key points at the fixed sensors' positions, which means that the number of estimated values reduces from  $Z = \mathfrak{N} = 1024$  to  $Z = S = 64$ . As we can see from Fig. 3, the result from CS after the dimension reduction (the green points) fits the actual signal (black curve) better than the results before (the cyan curve). By comparing the grey curve with the green points in Fig. 3, we have significantly improved the efficacy of CS applied to non-compressible signals.

By scrutinizing formula (4), we find that we can improve the reconstruction result of CS not only by reducing the number of coefficients  $Z$ , but also by modifying the sensing matrix  $\Psi$ . Thus, we will no longer use the DCT-basis and employ sparse coding [30] instead to find a sparsifying basis for the signal to further improve CS estimation.

### C. Sparse Coding

Sparse Coding was first proposed by Olshausen [31] to explain visual processing in the brain and has been an essential topic in machine learning in recent years [32]. The goal of sparse coding is to represent most or all of the original signal as a linear combination of few basic signals. We can utilize sparse coding to find a basis with respect to which the signal can be sparsely represented. One approach to sparse coding is to solve the following optimization problem:

$$\begin{aligned} \min_{\underline{\alpha}_n, \Theta} \sum_{n=1}^K \left\{ \left\| \underline{y}_n - \Theta \cdot \underline{\alpha}_n \right\|_2^2 + \lambda_n \|\underline{\alpha}_n\|_0 \right\}, \\ \text{s.t. } \Theta^T \Theta = \mathbf{I}. \end{aligned}$$

The first term is the approximation error that arises from reconstructing the signal  $\underline{y}_n$  using the basis  $\Theta$  and the coefficients  $\underline{\alpha}_n$  at each time step. In the second term,  $\lambda_n$  is a large constant, which performs as a penalty to ensure the sparsity of the coefficients  $\underline{\alpha}_n$  at each time step, and the constraint guarantees the orthogonality of the basis  $\Theta$  [33]. When the matrix  $\Theta$  is orthogonal, the sensing matrix

$\tilde{\Theta}_n = \Phi_n \cdot \Theta$  has a high probability of satisfying the RIP [34].

Using the basis obtained from sparse coding, which in the following we will call the sparse coding representation (SCR), we can see that the estimated signal in our example (Fig. 3, yellow points) shows a further improvement compared to that obtained using the DCT basis (green points).

## IV. KALMAN FILTERED COMPRESSIVE SENSING USING PSEUDO-MEASUREMENTS

In the previous section, we made CS more applicable to non-compressible signals in three steps. This section will focus on the fusion of the KF and CS, new problems that arise due to the fusion, and their corresponding solutions.

### A. Pseudo-Measurements

As stated in Section III-B, after the dimensionality reduction, we obtain values of the signal (i.e., the temperature distribution) at key points (i.e., the sensors' positions)

$$\underline{y}_n^{\text{PM}} = \Theta \cdot \hat{\underline{z}}_n.$$

As mentioned in Section II-C.2, the dimensionality of  $\underline{y}_n^{\text{PM}}$  is usually much larger than the dimension of the compressed measurements  $M$ , so the number of the values recovered by CS (i.e.,  $\underline{y}_n^{\text{PM}}$ ) is often sufficient for the KF to estimate the state variables with satisfactory accuracy. Therefore, we use these values as measurements for the KF (i.e., as pseudo-measurements).

By now, we have successfully combined the KF with CS through PMs. We coin the resulting method Kalman filtered compressive sensing (KFCS). In this method, CS plays the role to enlarge the number of measurements available to the KF.

After establishing the KFCS, two problems arise: the estimation error of the CS reconstruction and the accumulation of said error over time. In our proposal, the values estimated by CS are used as measurements in the KF. However, due to the reconstruction error, the uncertainty of the pseudo-measurements is greater than that of the sensor measurements. Moreover, as can be seen from the black part in the block diagram in Fig. 6, there is no feedback in the system yet, which means that the accumulation of the reconstruction error cannot be suppressed. To address these issues, we will quantify the uncertainty of the PMs and introduce feedback into the system in the following subsections.

### B. Dynamic Weighting

The PMs consist of two parts: the actual measurements from sensors and the estimated measurements from CS. Although we can get acceptable results using solely CS, following Section III (see Fig. 3), the estimated measurements still have a higher level of uncertainty. Despite there being some works that explore the distribution of

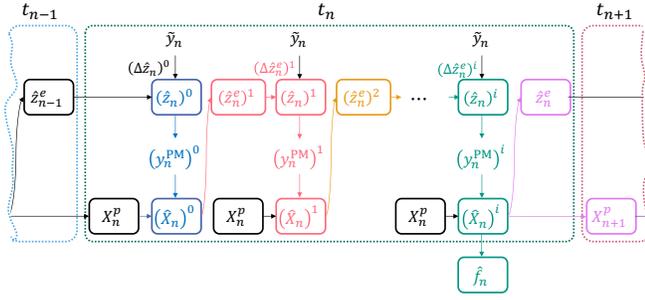


Fig. 5. Iterated update with coefficients updated from the estimation. The large boxes represent time steps and the different colours represent different iterations.

estimation error of CS [35], [36], [37], it is still an open question. What has been shown is the error bound

$$\|e\|_2 := \|\hat{z} - z\|_2 \leq C \frac{\min_{v_K \in \Sigma_K} \|z - v_K\|_1}{\sqrt{K}} =: B, \quad (6)$$

where  $\Sigma_K$  is the set of all  $K$ -sparse vectors [26]. Formula (6) shows that the estimation error lies inside a hyper-sphere around the true value  $z$ . As an approximation, we assume that CS is an unbiased estimator, i.e.,  $\mathbb{E}\{\hat{z}\} = \mathbb{E}\{z\}$ , moreover, we regard the radius of the hyper-sphere  $B$  as the  $3\sigma$  bound of the estimation error. According to the principle of maximum entropy [38], the error can then be regarded to follow Gaussian distribution with  $\mathbb{E}\{e_n^{PM}\} = \underline{0}$  and

$$\sigma_{PM} = \frac{B}{3} \approx C \frac{\|\hat{z}^{(K)}\|_1}{3\sqrt{K}},$$

where  $\sigma_{PM}$  denotes the standard deviation of the PMs and  $\hat{z}^{(K)}$  indicates the best  $K$ -term-approximation of  $\hat{z}$ . The covariance matrix  $C_n^v$  used in the KF is then dynamically adjusted to the types of measurements, i.e., pseudo measurements and actual measurements, at each time step.

### C. Coefficients Update

Since we use CS only to estimate the changes of the coefficients, the errors accumulate over time, leading to divergence of the estimation. The reason for this problem is that so far the system is an open-loop system (Fig. 6, black frames): The CS result is fed into the KF, while the KF estimate does not affect the CS result. Hence, we introduce feedback from the KF to the CS block by introducing a so-called coefficient update. With that modification a closed-loop system is constituted (red frames in Fig. 6 as well as the orange and violet parts in Fig. 4).

There are two approaches to update the CS coefficients: updating using the estimate produced by the KF's update step, (Fig. 4, yellow part), and updating using the estimate produced by the KF's prediction step (Fig. 4, violet part). The formulae for the coefficient updates are respectively

$$\begin{aligned} \hat{z}_n^e &= \Theta^{-1} \cdot \Psi \cdot \hat{X}_n, \\ \hat{z}_n^p &= \Theta^{-1} \cdot \Psi \cdot X_n^p. \end{aligned}$$

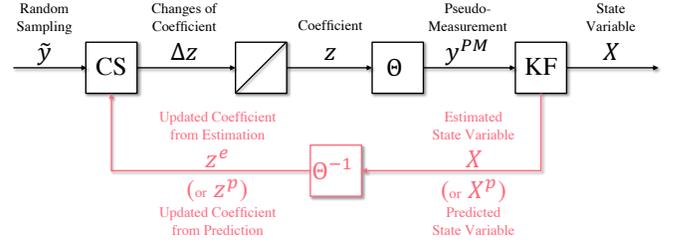


Fig. 6. System block diagram of KFCS.

After the coefficient update  $\hat{z}_n^e$  and  $\hat{z}_n^p$  can be used to estimate the change of the coefficients  $\Delta\hat{z}_n^e$  or  $\Delta\hat{z}_n^p$  as in (5). For systems with unknown input  $\underline{U}$ , such as in temperature monitoring, coefficient updating using the estimate produced by the KF's update step works better. In contrast, for systems with known inputs, such as control systems, it is better to update the coefficients using the KF's predictions.

### D. Iterated Update

Inspired by the iterated extended Kalman filter (IEKF) [39], we introduce an iterative approach in our KFCS. In each time step  $t_n$ , the coefficients  $(\hat{z}_n)^i$  and the state variables  $(\hat{T}_n)^i$  are computed iteratively, using the same measurements  $\tilde{y}_n$  and prediction  $T_n^p$ , until they converge or the maximum number of iterations is reached. Fig. 5 depicts this iterative process, with each dashed box indicating one time step. In each time step, different colors indicate different iterations.

## V. SIMULATION

To evaluate the effectiveness of our approach, we used different methods to recover the simulated temperature distribution from the example presented in Section II-B. After the signal recovery, we calculated their mean-square-error (MSE) and average MSE over time. Furthermore, we computed the average squared error over time under different system noise and measurement noise levels to explore the relationship between the error and the variance of the noises.

### A. Signal Recovery

To recover the temperature distribution, we assumed that there are  $S = 64$  equidistant sensors in total. To realize the random sampling, we activated  $M = 12$  sensors at each time step randomly and utilized the matrix  $\Phi_n$  to select these measurements. The interval between two time steps  $\Delta t$  was set to be 0.1s. To compare the effectiveness of our approach to other methods, we also recovered the temperature distribution using the traditional KF with 12, 40, and 64 equidistant sensors. We performed  $\mathfrak{R} = 1000$  simulations, for each of which, we generated temperature distributions  $f_n(\mathfrak{k})$  contaminated by system noise. Here  $n$  is the time step and  $\mathfrak{k}$  indicates the particular run of the simulation. Based on  $f_n(\mathfrak{k})$  with added measurement noise, we then computed an estimate  $\hat{f}_n(\mathfrak{k})$ . To measure the effectiveness, we use the mean-square error  $MSE_n$  at

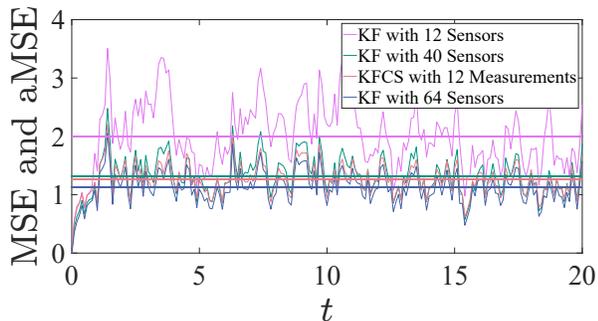


Fig. 7. MSE and aMSE of different methods, the curves indicate the MSE, while the horizontal dash-dot lines show the aMSE over time.

each time step and the average MSE over time (aMSE) as criterion. They are given by

$$\text{MSE}_n = \frac{1}{\mathfrak{K}} \sum_{\mathfrak{k}=1}^{\mathfrak{K}} \left\| \hat{\underline{f}}_n(\mathfrak{k}) - \underline{f}_n(\mathfrak{k}) \right\|_2^2, \quad (7)$$

$$\text{aMSE} = \frac{\sum_{n=1}^{\mathfrak{T}} \text{MSE}_n}{\mathfrak{T}}, \quad (8)$$

where  $\mathfrak{T}$  indicates the total number of time steps,  $\mathfrak{K}$  refers to the number of simulations.

The results are shown in Fig. 7. It can be seen that the KFCS with 12 measurements has a 37% advantage over the traditional KF with 12 sensors and even a 5% improvement over the traditional KF with 40 sensors w.r.t. the aMSE. Of course, there is still a 11% disadvantage in our approach compared to the traditional KF with 64 sensors, which activates all sensors all the time.

### B. Relationship between Error and Noise

We chose 10 different system noise levels  $\sigma_w$  and 10 different measurement noise levels  $\sigma_v$  to explore the relationship between error and noise. We then combined them into 100 different combinations and performed  $\mathfrak{K} = 300$  simulations for each combination using the KF with 12 sensors and KFCS with 12 measurements. For each combination, we generated and estimated the signal in the same way as in the previous subsection. From the results, we calculated the aMSE(c) for each combination  $c$ . We denote the aMSE of the two considered methods by  $\text{aMSE}_{\text{KF}}$  and  $\text{aMSE}_{\text{KFCS}}$ , respectively. To show the advantages and disadvantages of the KF and KFCS more clearly, we define the ratio of aMSE (raMSE) as

$$\text{raMSE}(c) = \frac{\text{aMSE}_{\text{KFCS}}(c)}{\text{aMSE}_{\text{KF}}(c)}.$$

From the definition of raMSE we know that, if raMSE is less than 1, the KFCS performs better than the traditional KF, vice versa. Fig. 8 shows the aMSE and raMSE. We can see that the estimation error increases with growing noise. The more remarkable point is that the KFCS is better suited when the measurement noise is relatively large compared to the system noise.

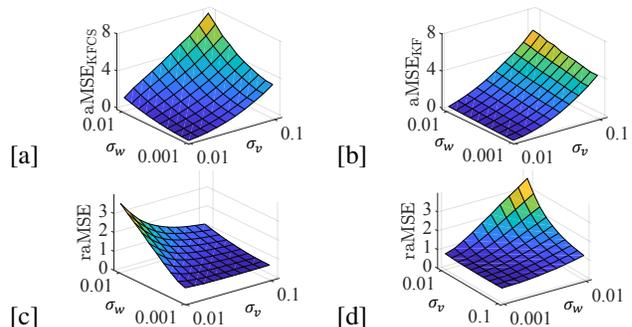


Fig. 8. [a] The aMSE(c) from KF; [b] The aMSE(c) from KFCS; [c] and [d] The raMSE(c) for different ranges of  $\sigma_w$  and  $\sigma_v$ .

### C. Discussion

In this section the temperature distribution from our example was recovered by KFCS with 12 measurements and by the KF with 12, 40, and 64 sensors. The results show that our approach using 12 measurements yields a better result than the traditional KF using 12 or even 40 sensors.

To determine the relationship between error and noise, we created 100 different noise combinations and performed 300 simulations using the KF with 12 sensors and KFCS with 12 measurements to recover the signal for each. The results demonstrate that our method has better results when the measurement noise is relatively large as compared to the system noise. The reason is that when the system noise is large, the actual signal deviates strongly from the ideal signal. Due to the strong noise, the changes of the coefficients are no longer sparse under the SCR basis which can result in wrong pseudo-measurements being given to the KF. In other words, our approach is more robust against measurement uncertainty than against system uncertainty. Our approach significantly improves the estimation in comparison to the traditional KF when one can model the system accurately and the sensors are inaccurate.

Although we assumed a known initial signal at  $t_0$  in Section III-A, we note that our approach can still converge with a zero-initialization.

## VI. CONCLUSION

In this paper, we have proposed three approaches to make CS more applicable to non-compressible signals and a method to combine the KF and CS using PMs. In addition, we have introduced dynamic weighting, a coefficient update, and an iterated update into KFCS to improve its accuracy. Finally, we have performed simulations to validate the effectiveness of the new approach. In the simulations, the KFCS has demonstrated better results than the traditional KF. Furthermore, we have found that our method is more suitable for accurate system model is given and the measurement accuracy is low.

There are some possibilities for further improvements in the future. Firstly, the random sampling in this paper is achieved by randomly selecting sensors at fixed positions. This could also be achieved by other methods, e.g., installing

sensors on mobile agents. Secondly, the positions of the PMs in this paper are fixed. In future works, adaptive PMs could be considered, which means that the PMs are set to be at the positions where the uncertainties of the estimate are the largest. Thirdly, we have assumed that the uncertainty of PMs obeys a Gaussian distribution, but it has not been rigorously proven. Although the iterative process can overcome this problem to some extent, one can still design more effective filters by taking the error distribution of the CS estimate into account. Forthly, in this paper, we have applied sparse coding to obtain a sparsifying basis for the considered signals. However, according to our idea from Section III-A, one might find a sparsifying basis for the changes of the signal directly from the system model. Moreover, we only talk about the application of the KFCS to temperature monitoring, we will apply the algorithm to other examples later.

## REFERENCES

- [1] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME - Journal of Basic Engineering*, vol. 82, pp. 35–45, 1960.
- [2] T. Lefebvre, H. Bruyninckx, and J. De Schutter, "Kalman Filters for Nonlinear Systems," in *Nonlinear Kalman Filtering for Force-Controlled Robot Tasks*. Springer, 4, pp. 51–76.
- [3] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte, "A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators," *IEEE Transactions on Automatic Control*, vol. 45, no. 3, pp. 477–482, 2000.
- [4] S. Radtke, B. Noack, and U. D. Hanebeck, "Distributed Estimation using Square Root Decompositions of Dependent Information," in *Proceedings of the 22nd International Conference on Information Fusion (Fusion 2019)*, Ottawa, Canada, Jul. 2019.
- [5] W. F. Ames, *Numerical Methods for Partial Differential Equations*. Academic Press, 2014.
- [6] J. Sijs, U. D. Hanebeck, and B. Noack, "An Empirical Method to Fuse Partially Overlapping State Vectors for Distributed State Estimation," in *Proceedings of the 2013 European Control Conference (ECC 2013)*, Zürich, Switzerland, Jul. 2013.
- [7] B. Noack, J. Sijs, and U. D. Hanebeck, "Fusion Strategies for Unequal State Vectors in Distributed Kalman Filtering," in *Proceedings of the 19th IFAC World Congress (IFAC 2014)*, Cape Town, South Africa, Aug. 2014.
- [8] M. A. Davenport, P. T. Boufounos, M. B. Wakin, and R. G. Baraniuk, "Signal Processing with Compressive Measurements," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 445–460, 2010.
- [9] C. E. Shannon, "Communication in the Presence of Noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, 1949.
- [10] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-Pixel Imaging via Compressive Sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, 2008.
- [11] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and Depth from a Conventional Camera with a Coded Aperture," *ACM Transactions on Graphics (TOG)*, vol. 26, no. 3, pp. 70–es, 2007.
- [12] A. C. Zelinski, L. L. Wald, K. Setsompop, V. K. Goyal, and E. Adalsteinsson, "Sparsity-Enforced Slice-Selective MRI RF Excitation Pulse Design," *IEEE Transactions on Medical Imaging*, vol. 27, no. 9, pp. 1213–1229, 2008.
- [13] W. Dai, M. A. Sheikh, O. Milenkovic, and R. G. Baraniuk, "Compressive Sensing DNA Microarrays," *EURASIP Journal on Bioinformatics and Systems Biology*, vol. 2009, pp. 1–12, 2008.
- [14] D. L. Donoho, "Compressed Sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [15] E. J. Candes and T. Tao, "Decoding by Linear Programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [16] A. Y. Carmi, L. S. Mihaylova, and S. J. Godsill, "Introduction to Compressed Sensing and Sparse Filtering," in *Compressed Sensing & Sparse Filtering*. Springer, 2014, pp. 1–23.
- [17] M. Salman Asif and J. Romberg, "Dynamic Updating for  $\ell_1$  Minimization," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 421–434, 2010.
- [18] M. S. Asif, "Dynamic Compressive Sensing: Sparse Recovery Algorithms for Streaming Signals and Video," Ph.D. dissertation, Georgia Institute of Technology, 2013.
- [19] S. Ji, Y. Xue, and L. Carin, "Bayesian Compressive Sensing," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2346–2356, 2008.
- [20] D. Baron, M. F. Duarte, M. B. Wakin, S. Sarvotham, and R. G. Baraniuk, "Distributed Compressive Sensing," *arXiv e-prints*, p. arXiv:0901.3403, Jan. 2009.
- [21] N. Vaswani, "LS-CS-residual (LS-CS): Compressive Sensing on Least Squares Residual," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4108–4120, 2010.
- [22] J. A. Tropp, M. B. Wakin, M. F. Duarte, D. Baron, and R. G. Baraniuk, "Random Filters for Compressive Sampling and Reconstruction," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, vol. 3. IEEE, 2006, pp. III–III.
- [23] N. Vaswani, "Kalman Filtered Compressed Sensing," in *2008 15th IEEE International Conference on Image Processing*. IEEE, 2008, pp. 893–896.
- [24] N. Vaswani, "Analyzing Least Squares and Kalman Filtered Compressed Sensing," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2009, pp. 3013–3016.
- [25] J. Crank and P. Nicolson, "A Practical Method for Numerical Evaluation of Solutions of Partial Differential Equations of the Heat-Conduction Type," in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 43, no. 1. Cambridge University Press, 1947, pp. 50–67.
- [26] R. Baraniuk, M. A. Davenport, M. F. Duarte, C. Hegde *et al.*, "An Introduction to Compressive Sensing," *Connexions E-Textbook*, pp. 24–76, 2011.
- [27] Y. Tsaig and D. L. Donoho, "Extensions of Compressed Sensing," *Signal Processing*, vol. 86, no. 3, pp. 549–571, 2006.
- [28] D. L. Donoho and M. Elad, "Optimally Sparse Representation in General (Nonorthogonal) Dictionaries via  $\ell_1$  Minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.
- [29] A. Cohen, W. Dahmen, and R. DeVore, "Compressed Sensing and Best  $k$ -term Approximation," *Journal of the American Mathematical Society*, vol. 22, no. 1, pp. 211–231, Jan. 2009.
- [30] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online Dictionary Learning for Sparse Coding," in *Proceedings of the 26th Annual International Conference on Machine Learning*, 2009, pp. 689–696.
- [31] B. A. Olshausen and D. J. Field, "Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1?" *Vision Research*, vol. 37, no. 23, pp. 3311–3325, 1997.
- [32] D. Cai, H. Bao, and X. He, "Sparse Concept Coding for Visual Analysis," in *CVPR 2011*. IEEE, 2011, pp. 2905–2910.
- [33] C. Bao, J.-F. Cai, and H. Ji, "Fast sparsity-based orthogonal dictionary learning for image restoration," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 3384–3391.
- [34] H. Rauhut, "Compressive Sensing and Structured Random Matrices," *Theoretical Foundations and Numerical Methods for Sparse Recovery*, vol. 9, pp. 1–92, 2010.
- [35] B. Wang, L. Hu, J. An, G. Liu, and J. Cao, "Recovery Error Analysis of Noisy Measurement in Compressed Sensing," *Circuits, Systems, and Signal Processing*, vol. 36, no. 1, pp. 137–155, 2017.
- [36] A. Jung, Z. Ben-Haim, F. Hlawatsch, and Y. C. Eldar, "On Unbiased Estimation of Sparse Vectors Corrupted by Gaussian Noise," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2010, pp. 3990–3993.
- [37] A. Jung, Z. Ben-Haim, F. Hlawatsch, and Y. C. Eldar, "Unbiased Estimation of a Sparse Vector in White Gaussian Noise," *IEEE Transactions on Information Theory*, vol. 57, no. 12, pp. 7856–7876, 2011.
- [38] E. T. Jaynes, "Information Theory and Statistical Mechanics," *Physical Review*, vol. 106, no. 4, p. 620, 1957.
- [39] Z. Zhang, "Parameter Estimation Techniques: A Tutorial with Application to Conic Fitting," *Image and Vision Computing*, vol. 15, no. 1, pp. 59–76, 1997.