SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

December 2016, Vol. 59 122309:1–122309:18
doi: 10.1007/s11432-015-0602-9

# Segment-sliding reconstruction of pulsed radar echoes with sub-Nyquist sampling

Suling ZHANG[1], Feng XI[1], Shengyao CHEN[1],
Yimin Daniel ZHANG[2] & Zhong LIU[1]*

[1]*Department of Electronic Engineering, Nanjing University of Science and Technology, Nanjing* 210094*, China;*
[2]*Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA* 19122*, USA*

Received July 27, 2015; accepted February 29, 2016; published online July 7, 2016

**Abstract** It has been shown that analog-to-information conversion (AIC) is an efficient scheme to perform sub-Nyquist sampling of pulsed radar echoes. However, it is often impractical, if not infeasible, to reconstruct full-range Nyquist samples because of huge storage and computational load requirements. Based on the analyses of AIC measurement system, this paper develops a novel segment-sliding reconstruction (SegSR) scheme to effectively reconstruct the Nyquist samples. The SegSR performs segment-by-segment reconstruction in a sliding mode and can be implemented in real time. An important characteristic that distinguishes the proposed SegSR from existing methods is that the measurement matrix in each segment satisfies the restricted isometry property (RIP) condition. Partial support in the previous segment can be incorporated into the estimation of the Nyquist samples in the current segment. The effect of interference introduced from adjacent segments is theoretically analyzed, and it is revealed that the interference consists of two interference levels with different impacts to the signal reconstruction performance. With these observations, a two-step orthogonal matching pursuit (OMP) procedure is proposed for segment reconstruction, which takes into account different interference levels and partially known support of the previous segment. The proposed SegSR scheme achieves near-optimal reconstruction performance with a significant reduction of computational loads and storage requirements. Theoretical analyses and simulations verify its effectiveness.

**Keywords** compressed sensing, analog-to-information conversion, orthogonal matching pursuit (OMP), segment-sliding reconstruction, restricted isometry property (RIP)

**Citation** Zhang S L, Xi F, Chen S Y, et al. Segment-sliding reconstruction of pulsed radar echoes with sub-Nyquist sampling. Sci China Inf Sci, 2016, 59(12): 122309, doi: 10.1007/s11432-015-0602-9

## 1 Introduction

A pulsed radar usually transmits modulated pulses during the transmit time and receives echo signals reflected from illuminated targets during the receive time. The received echoes are sampled and processed to extract target information. For wideband and ultra-wideband radars, high-rate analog-to-digital converters (ADC) have to be employed, resulting in a large volume of sampled data for storage and subsequent processing. With the development of compressed sensing (CS) theory [1–3], analog-to-information conversions (AIC) are introduced to sample wideband signals at sub-Nyquist rates. Among

* Corresponding author (email: eezliu@njust.edu.cn)

 

various AIC's, random demodulator (RD) [4], random modulator pre-integrator (RMPI) [5], Xampling [6], and quadrature compressed sensing (QuadCS) [7,8] have received particular attention for radar applications. Theoretical analyses and experimental studies have shown that these AIC systems are efficient for sub-Nyquist acquisition of sparse echo signals with a large bandwidth [9–14].

In many cases, we desire to recover the Nyquist samples of radar echoes from the sub-Nyquist outputs of AIC systems. In CS theory, the problem refers to sparse signal reconstruction, which is often expressed as

$$\min_{\boldsymbol{\sigma}} \|\boldsymbol{\sigma}\|_0 \quad \text{s.t.} \quad \boldsymbol{y} = \boldsymbol{\Phi\Psi\sigma} = \boldsymbol{A\sigma}, \tag{1}$$

where $\|\boldsymbol{\sigma}\|_0$ is the $\ell_0$-norm of $\boldsymbol{\sigma}$ and counts for the number of nonzero entries, i.e., the sparsity of the $N \times 1$ sparse coefficient vector $\boldsymbol{\sigma}$, $\boldsymbol{y}$ is an $M \times 1$ measurement vector, $\boldsymbol{\Phi}$ is an $M \times N$ observation matrix with $M \ll N$, $\boldsymbol{\Psi}$ is an $N \times N$ basis matrix, and $\boldsymbol{A} = \boldsymbol{\Phi\Psi}$ is the yielding $M \times N$ measurement matrix. It has been shown that the sparse vector $\boldsymbol{\sigma}$ with sparse level $\|\boldsymbol{\sigma}\|_0 = K$ can be exactly reconstructed by solving (1), if matrix $\boldsymbol{A}$ satisfies the restricted isometry property (RIP) condition [3]

$$(1-\delta_K)\|\boldsymbol{\sigma}\|_2^2 \leqslant \|\boldsymbol{A\sigma}\|_2^2 \leqslant (1+\delta_K)\|\boldsymbol{\sigma}\|_2^2, \tag{2}$$

where $\delta_K$ is a sufficiently small RIP constant. Unfortunately, problem (1) is an NP-complete problem and we often relax it to the following convex optimization problem,

$$\min_{\boldsymbol{\sigma}} \|\boldsymbol{\sigma}\|_1 \quad \text{s.t.} \quad \boldsymbol{y} = \boldsymbol{\Phi\Psi\sigma} = \boldsymbol{A\sigma}. \tag{3}$$

There are a wide variety of approaches to solve (3), including greedy iteration algorithms [15,16], convex optimization algorithms [17,18] and Bayesian CS [19,20] (see [21] for a review). All these techniques are suitable when the measurement matrix is of a moderate size. For the large-scale measurement matrix, on the other hand, we need to allocate a prohibitively huge size of memory space to store the data and, as such, the recovery may become impractical for the state-of-the-art computers. Consider, for example, a radar system with a signal bandwidth of 100 MHz, a pulse width of 10 μs, and a receiving time of 2490 μs. At one-fifth of the Nyquist sampling rate, we need to store about $1.24 \times 10^{10}$ elements with 49800 rows and 248000 columns, occupying about 92 GB of memory using the standard IEEE double precision. As a result, the full range reconstruction is impractical, if not infeasible, with the state-of-the-art hardware capabilities.

Motivated by such facts, this paper studies implementable, full-range reconstruction of pulsed radar echoes sampled by an RD AIC system. The results can be easily extended to other sparse sampling schemes like RMPI, QuadCS and Xampling that generate similar measurement matrices.

For the RD system, as will be discussed in Section 3, its measurement matrix $\boldsymbol{A}$ with a waveform-matched dictionary [22] is sparse with its effective elements around the diagonal and there are overlaps between the rows. With this specific matrix structure, we can decompose the large-scale reconstruction problem (3) into a series of small-scale ones by properly segmenting the measurement vector and the sparse vector. The word "properly" implies that the segmentations are made such that each segmented measurement matrix satisfies the RIP conditions. In this case, the entire sparse vector $\boldsymbol{\sigma}$ can be reconstructed segment by segment.

It is important to note that, because the matrix $\boldsymbol{A}$ is not block-diagonal, the reconstruction of a segment will be interfered by its adjacent segments. Such interference is referred to as virtual noise in this paper and is factorized into forward virtual noise and backward virtual noise. The forward virtual noise is generated by the inaccurate estimate of the sparse vector in the previous segment, whereas the backward virtual noise is formulated by the partial measurement in the subsequent segment. We theoretically analyze the effects of the virtual noise on the reconstructed positions and amplitudes of the sparse entries in current segment. It is revealed that the virtual noise has different noise levels for different measurements, and the effect of the backward virtual noise is much higher than that of the forward one.

To perform efficient estimation of the sparse echoes in each segment, we develop a two-step OMP process (TOMPP) algorithm which takes into account the effects of these two kinds of virtual noises

and partially known support in the previous segment estimate. In the proposed approach, we take the segment-sliding approach to segment both the measurement vector and the sparse vector. As a result, the overall complexity is reduced, and the full-range reconstruction of the pulsed radar echoes can be implemented in real time. Theoretical analyses and simulation results validate the effectiveness of the proposed segment estimation. When the background noise is Gaussian, the proposed scheme achieves nearly optimal reconstruction performance with significant reduction of computational loads and storage requirements.

The problem studied in this paper is, in essence, a reconstruction of sparse time-varying signals from streaming measurements, which has been discussed in [23–27]. These algorithms fully exploit the measurement systems and perform sliding estimation of the sparse coefficients. The methods in [23, 24] use a sliding window for data processing and entail recursive sampling and iterative recovery. In each sliding sampling, the measurement matrix is recursively generated by permuting the previous measurement matrix. The measurement matrix in each sampling always satisfies the RIP condition if the original one satisfies. The method described in [25] focuses on the measurement of infinite-dimensional signals using a finite-length, time-varying linear system. The compressive measurement over a finite-length window is closely associated with a segment of the infinite-dimensional signal. Signal recovery can then be implemented iteratively over sliding intervals. A similar idea is taken in [26] with special lapped orthogonal transform bases. Ref. [27] studies the large-scale reconstruction problem arising from radar imaging, and an interesting segmentation approach is presented.

There are two important issues that are not properly addressed in [25–27]. First, because the matrix $\boldsymbol{A}$ is not block-diagonal, the reconstruction of a segment will be interfered by the adjacent segments. The effect of such interference, however, has not been properly studied and analyzed. Second, the existing approaches do not guarantee the RIP conditions to be satisfied after the segmentation, even when the original problem satisfies the RIP conditions. With novel segmentation scheme and thorough theoretical analyses, the proposed approach well solves the above problems. The main contributions of this paper include: (a) we develop a new segmentation-based algorithm in which all the measurement sub-matrices satisfy the RIP condition; (b) we provide a theoretical analysis of the effect of the interference and reveal that the interference structure consists of two interference noise levels; and (c) based on these results, a new OMP algorithm, termed TOMPP, is developed for effective segment reconstruction with the impacts of different interference levels taken into account. The TOMPP algorithm achieves almost near-optimal reconstruction performance.

The remainder of the paper is organized as follows. Section 2 provides a brief description of the problem model. Section 3 analyzes the structure of the measurement matrix and explains the mechanism of sliding reconstruction. Section 4 describes the proposed segment-sliding reconstruction algorithm. An analysis of the effect of the interference from adjacent segments is provided in Section 5. The required storage capacity and computational complexity are analyzed in Section 6. Numerical results are presented in Section 7. We conclude this paper in Section 8. The proofs of the main results are given in the Appendix.

Notations. Bold letters denote the vectors or matrices. $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ represent transposition and conjugate transposition of a vector or matrix, respectively. $(\cdot)^{\dagger}$ represents the Moore-Penrose inverse of a tall, full-rank matrix. $\|\cdot\|_1$, $\|\cdot\|_2$ and $\|\cdot\|_\infty$ denote the $\ell_1$, $\ell_2$ and $\ell_\infty$ vector or matrix norms, respectively. $|\cdot|$ represents the absolute value of a number or the cardinality of a set. A vector (matrix) with a set as its subscript denotes the sub-vector (sub-matrix) containing the elements (columns) of the vector (matrix) indexed by the set.

## 2 Background materials and problem statement

### 2.1 Signal model

Consider a pulsed radar where the baseband signal $s(t)$ has a pulse width of $T_p$ and a bandwidth of $B/2$. For $K$ non-fluctuating point targets, the received echo signal at the baseband can be represented as

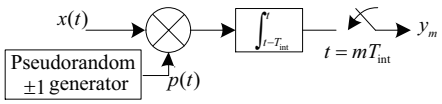


**Figure 1** Structure of an RD system.

$$x(t) = \sum_{k=0}^{K-1} \sigma_k s(t - t_k),\ t \in [0, T), \tag{4}$$

where $t_k$ and $\sigma_k$ are the delay and the gain of the $k$-th target, respectively, and $T$ refers to the receive time which is usually much larger than $T_p$, i.e., $T \gg T_p$. For notational simplicity, the background noise is not included in the model (4), but its effect on the reconstruction performance will be examined numerically in Section 7.

To apply the CS theory, $x(t)$ is assumed to be sparse in some domain. In radar applications, the transmit waveforms are known in advance and the waveform-matched dictionary [22] is often adopted. For a radar with baseband signal $s(t)$ of bandwidth $B/2$, let $\tau_0 = 1/B$ be its Nyquist sampling interval, and $N = \lceil T/\tau_0 \rceil = \lceil BT \rceil$ be the number of Nyquist samples of the received signals during the receive time $T$, where $\lceil \cdot \rceil$ denotes the ceiling operation. The waveform-matched dictionary consists of all time-shifted versions of $s(t)$ at the Nyquist-sampling grids $\{0, \tau_0, \ldots, (N-1)\tau_0\}$, i.e., $\boldsymbol{\psi}(t) = \{\psi_n(t) \,|\psi_n(t) = s(t - n\tau_0)\ , n = 0, 1, \ldots, N-1\}$. For the dictionary, the time-delay axis is discretized with resolution $\tau_0$. The discretization is reasonable, because the time resolution of the baseband signal $s(t)$ is limited to $1/B$.

Assume that the target delays are integral multiples of $\tau_0$, i.e., $t_k \in \{0, \tau_0, \ldots, (N-1)\tau_0\}$. Given the waveform-matched dictionary $\boldsymbol{\psi}(t)$, $x(t)$ in (4) can be represented as

$$x(t) = \sum_{n=0}^{N-1} \sigma_n \psi_n(t) = \boldsymbol{\psi}(t)\boldsymbol{\sigma}, \tag{5}$$

where $\boldsymbol{\sigma} = [\sigma_0, \sigma_1, \ldots, \sigma_{N-1}]^{\mathrm{T}}$ is the sparse coefficient vector to be determined. Note that there are $(N - K)$ zero coefficients in vector $\boldsymbol{\sigma}$ for $K$ targets. When $K \ll N$, $x(t)$ is said to be $K$-sparse, where the sparsity level $K$ exactly equals the number of the targets.

### 2.2 Compressive sampling of radar echoes

This sub-section takes the RD AIC as an example to illustrate the principle of compressive sampling.

Figure 1 shows the basic structure of an RD system. The input analog signal $x(t)$ is first mixed by a random-binary signal, $p(t) = \pm 1 (t \in [k/B_p, (k+1)/B_p), k = 0, 1, 2, \ldots)$, with $B_p \geqslant B$ . The mixed signal passes through a low-pass filter to prevent aliasing, and the filtered signal is then sampled. For the accumulated low-pass filter, the compressive samples are given by

$$y_m = \int_{(m-1)T_{\mathrm{int}}}^{mT_{\mathrm{int}}} x(t)p(t)\mathrm{d}t,\ m = 1, 2, \ldots, M, \tag{6}$$

where $T_{\mathrm{int}} = R\tau_0$, and $R > 1$ is an integer referred to as the down-sampling rate. During the receive time $T$, we acquire $M = \lfloor T/T_{\mathrm{int}} \rfloor$ low-rate samples, where $\lfloor \cdot \rfloor$ denotes the floor operation.

For the waveform-matched dictionary, passing its elements through the RD system yields

$$a_{m,n} = \int_{(m-1)T_{\mathrm{int}}}^{mT_{\mathrm{int}}} \psi_n(t)p(t)\mathrm{d}t, m = 1, 2, \ldots, M. \tag{7}$$

Combining (5) and (6) with (7), we have

$$y_m = \sum_{n=0}^{N-1} \sigma_n a_{m,n},\ m = 1, 2, \ldots, M. \tag{8}$$

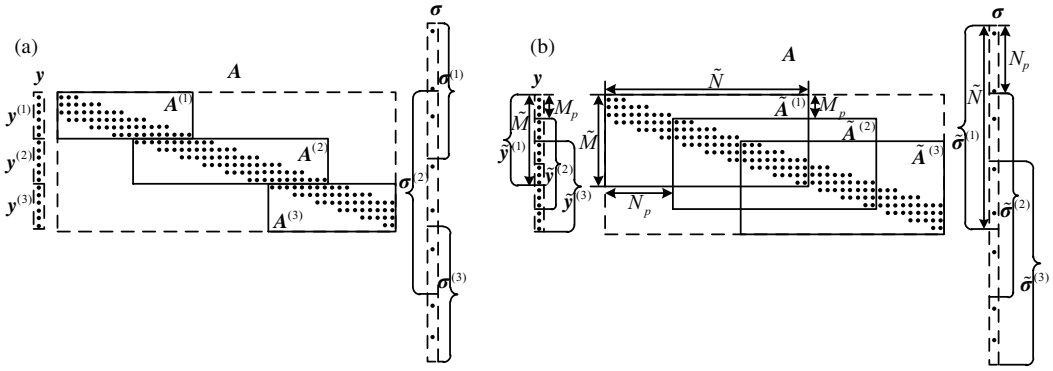


**Figure 2** (a) Segmenting in traditional processing; (b) segmenting proposed in this paper.

Define $\boldsymbol{y} = [y_1, y_2, \ldots, y_M]^{\mathrm{T}}$, $\boldsymbol{a}_n = [a_{1,n}, a_{2,n}, \ldots, a_{M,n}]^{\mathrm{T}}$ and $\boldsymbol{A} = [\boldsymbol{a}_0, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_{N-1}]$. Then, we can express (8) as

$$\boldsymbol{y} = \sum_{n=0}^{N-1} \sigma_n \boldsymbol{a}_n = \boldsymbol{A}\boldsymbol{\sigma}. \tag{9}$$

Vector $\boldsymbol{y}$ and matrix $\boldsymbol{A}$ are referred to as the measurement vector and the measurement matrix, respectively. Because $R > 1$, it is clear that $M < N$ and Eq. (9) is thus an underdetermined equation.

The reconstruction of the radar echo signal $x(t)$ is equivalent to the recovery of the sparse vector $\boldsymbol{\sigma}$. Note that the RD AIC is a special case of the QuadCS system with a zero intermediate frequency, and we have proved in [8] that the QuadCS system satisfies the RIP condition for radar signals with flat spectra. Then, the sparse vector $\boldsymbol{\sigma}$ can be exactly reconstructed by solving (1) or (3). As discussed in Section 1, the current hardware and computational capacity are infeasible to implement such a large-scale reconstruction. This paper provides an implementable reconstruction technique through the exploitation of the structure characteristic of the measurement matrix.

## 3 Structure of measurement matrix and sliding reconstruction

The structure of the measurement matrix results from the combination of the RD system and the waveform-matched dictionary. Because the radar waveform $s(t)$ has a finite duration of $T_p$, i.e., $s(t) = 0$ for $t \notin [0, T_p)$, $\psi_n(t) = s(t - n\tau_0)$ of $\boldsymbol{\psi}(t)$ also has a finite duration, i.e., $\psi_n(t) = 0$ for $t \notin [n\tau_0, T_p + n\tau_0)$, $n = 0, 1, \ldots, N-1$. Then, we learn from (7) that the $n$-th column of the measurement matrix $\boldsymbol{A}$ takes nonzero values only in finite indexes, i.e., $a_{m,n} = 0$ for $m \leqslant \lfloor n/R \rfloor$ or $m \geqslant \lceil (T_p + n\tau_0)/T_{\mathrm{int}} \rceil + 1$. That is, the matrix $\boldsymbol{A}$ is a banded matrix with non-zero elements positioned around the diagonal and there are overlaps between the rows. An example with $B = 10$ MHz, $T_p = 0.9$ μs, $T = 5.4$ μs and $R = 3$ is shown in Figure 2, where $M = 18$ and $N = 45$.

With such a special structure of the matrix $\boldsymbol{A}$, a straightforward approach, as shown in Figure 2(a), is to divide the measurement vector $\boldsymbol{y}$ into $L$ non-overlapping segments, $\boldsymbol{y}^{(l)}$, $l = 1, 2, \ldots, L$, whereas the sparse solution vector $\boldsymbol{\sigma}$ is partitioned into overlapping segments, $\boldsymbol{\sigma}^{(l)}$, $l = 1, 2, \ldots, L$, such that a segment contains all entries that are responsible for a segmented measurement vector. Let $\boldsymbol{A}^{(l)}$ be the measurement sub-matrix which is formulated by extracting the rows of $\boldsymbol{A}$ that contribute to $\boldsymbol{y}^{(l)}$. Then, the large-scale sparse reconstruction problem (3) can be decomposed into $L$ small-scale ones,

$$\min_{\boldsymbol{\sigma}^{(l)}} \left\| \boldsymbol{\sigma}^{(l)} \right\|_1 \quad \text{s.t.} \quad \boldsymbol{y}^{(l)} = \boldsymbol{A}^{(l)} \boldsymbol{\sigma}^{(l)}, \tag{10}$$

for $l = 1, 2, \ldots, L$. As the dimension of (10) can be set in an implementable range, the overall complexity and storage requirement can be significantly reduced. Once we obtain the solution of (10) for $l = 1, 2, \ldots, L$, we are able to recover the Nyquist samples of the radar echo signals.

One important problem in (10) is that the sub-matrix $\boldsymbol{A}^{(l)}$ cannot be guaranteed to satisfy the RIP condition, even if the original problem in (3) does. With the generation of $\boldsymbol{A}^{(l)}$, we can find that the

last few columns of $\boldsymbol{A}^{(l)}$ cannot contain all compressive samples of the corresponding elements. There is a high dependency between the last few columns of $\boldsymbol{A}^{(l)}$, even completely dependent. Therefore, $\boldsymbol{A}^{(l)}$ cannot satisfy the RIP condition.

In the next section, we develop a new segment-sliding reconstruction scheme which guarantees the RIP conditions to be satisfied as long as it is satisfied in the original problem of (3).

## 4 Segment-sliding reconstruction scheme

For simplicity, let the receive time $T$ be $P$ times the radar pulse width $T_p$, i.e., $T = PT_p$. Denote $N_p = N/P$ and $M_p = N/(RP)$ as the number of Nyquist samples and that of the compressive samples in a pulse width, respectively.

Segment $\boldsymbol{\sigma}$, $\boldsymbol{y}$ and $\boldsymbol{A}$ as described in Figure 2(b). For the vector $\boldsymbol{\sigma}$, we formulate the overlapping sub-vectors $\tilde{\boldsymbol{\sigma}}^{(l)}$, $l = 1, 2, \ldots, L$. Assume that the sub-vector $\tilde{\boldsymbol{\sigma}}^{(l)}$ includes $S$ radar pulses ($1 < S < P$), where $\tilde{\boldsymbol{\sigma}}^{(l)}$ is defined as

$$\tilde{\boldsymbol{\sigma}}^{(l)} = \boldsymbol{\sigma}(\tilde{n} : \tilde{n} + \tilde{N} - 1),\ l = 1, 2, \ldots, L, \tag{11}$$

$\tilde{n} = (l-1)N_p$, and $\tilde{N} = SN_p$ is the length of the sub-vector $\tilde{\boldsymbol{\sigma}}^{(l)}$. There are $L = P - S$ sparse sub-vectors. Moreover, let us express $\tilde{\boldsymbol{\sigma}}^{(l)}$ as $\tilde{\boldsymbol{\sigma}}^{(l)} = [(\tilde{\boldsymbol{\sigma}}_1^{(l)})^{\mathrm{T}}, (\tilde{\boldsymbol{\sigma}}_2^{(l)})^{\mathrm{T}}, \ldots, (\tilde{\boldsymbol{\sigma}}_S^{(l)})^{\mathrm{T}}]^{\mathrm{T}}$ where $\tilde{\boldsymbol{\sigma}}_s^{(l)} = \tilde{\boldsymbol{\sigma}}^{(l)}((s-1)N_p : sN_p - 1)$ is of length $N_p$. The selection of $S$ depends on the computational capability and sparsity in the sub-vector $\tilde{\boldsymbol{\sigma}}^{(l)}$. In the following, we assume that sub-vector $\tilde{\boldsymbol{\sigma}}^{(l)}$ is sparse for all $l = 1, 2, \ldots, L$.

Similarly, we segment the measurement vector $\boldsymbol{y}$ into $L$ overlapping measurement sub-vectors $\tilde{\boldsymbol{y}}^{(l)}$,

$$\tilde{\boldsymbol{y}}^{(l)} = \boldsymbol{y}(\tilde{m} : \tilde{m} + \tilde{M} - 1),\ l = 1, 2, \ldots, L, \tag{12}$$

where $\tilde{m} = (l-1)M_p + 1$, and $\tilde{M} = (S+1)M_p$ is the length of the sub-vector $\tilde{\boldsymbol{y}}^{(l)}$. Thus, the sub-vector $\tilde{\boldsymbol{y}}^{(l)}$ is downward shifted by $M_p$ compressive samples when compared with the sub-vector $\tilde{\boldsymbol{y}}^{(l-1)}$.

By extracting the columns and rows of $\boldsymbol{A}$ corresponding to $\tilde{\boldsymbol{\sigma}}^{(l)}$ and $\tilde{\boldsymbol{y}}^{(l)}$, we formulate an $\tilde{M} \times \tilde{N}$ measurement sub-matrix $\tilde{\boldsymbol{A}}^{(l)}$ as

$$\tilde{\boldsymbol{A}}^{(l)} = \boldsymbol{A}(\tilde{m} : \tilde{m} + \tilde{M} - 1, \tilde{n} : \tilde{n} + \tilde{N} - 1), l = 1, 2, \ldots, L. \tag{13}$$

Similar to the partition of $\tilde{\boldsymbol{\sigma}}^{(l)}$, we can express $\tilde{\boldsymbol{A}}^{(l)}$ as $\tilde{\boldsymbol{A}}^{(l)} = [\tilde{\boldsymbol{A}}_1^{(l)}, \tilde{\boldsymbol{A}}_2^{(l)}, \ldots, \tilde{\boldsymbol{A}}_S^{(l)}]$ where $\tilde{\boldsymbol{A}}_s^{(l)} = \tilde{\boldsymbol{A}}^{(l)}(1 : \tilde{M}, (s-1)N_p : sN_p - 1)$ for $s = 1, 2, \ldots, S$.

In practice, we can downward shift multiple $N_p$ coefficient elements to formulate the sub-vector $\tilde{\boldsymbol{\sigma}}^{(l)}$. Then, the sub-vector $\tilde{\boldsymbol{y}}^{(l)}$ and the measurement sub-matrix $\tilde{\boldsymbol{A}}^{(l)}$ are adjusted correspondingly. For simplicity, we assume that $N_p$ coefficient elements are shifted in this paper.

As can be observed in Figure 2(b), $\tilde{\boldsymbol{y}}^{(l)}$ is related to three sub-vectors $\tilde{\boldsymbol{\sigma}}_1^{(l-1)}$, $\tilde{\boldsymbol{\sigma}}^{(l)}$ and $\tilde{\boldsymbol{\sigma}}_S^{(l+1)}$. Define

$$\underline{\tilde{\boldsymbol{A}}}^{(l-1)} = \begin{bmatrix} \tilde{\boldsymbol{A}}_1^{(l-1)}(M_p + 1 : \tilde{M}, 0 : N_p - 1) \\ \boldsymbol{0}_{M_p \times N_p} \end{bmatrix} \quad \text{and} \quad \underset{\sim}{\tilde{\boldsymbol{A}}}^{(l+1)} = \begin{bmatrix} \boldsymbol{0}_{M_p \times N_p} \\ \tilde{\boldsymbol{A}}_S^{(l+1)}(1 : \tilde{M} - M_p, 0 : N_p - 1) \end{bmatrix}. \tag{14}$$

Then, $\underline{\tilde{\boldsymbol{A}}}^{(l-1)}\tilde{\boldsymbol{\sigma}}_1^{(l-1)}$ and $\underset{\sim}{\tilde{\boldsymbol{A}}}^{(l+1)}\tilde{\boldsymbol{\sigma}}_S^{(l+1)}$ are the contributions to $\tilde{\boldsymbol{y}}^{(l)}$ from $\tilde{\boldsymbol{\sigma}}_1^{(l-1)}$ and $\tilde{\boldsymbol{\sigma}}_S^{(l+1)}$, respectively. The $l$-th measurement sub-vector $\tilde{\boldsymbol{y}}^{(l)}$ is given by

$$\tilde{\boldsymbol{y}}^{(l)} = \begin{cases} \tilde{\boldsymbol{A}}^{(l)}\tilde{\boldsymbol{\sigma}}^{(l)} + \underset{\sim}{\tilde{\boldsymbol{A}}}^{(l+1)}\tilde{\boldsymbol{\sigma}}_S^{(l+1)}, & l = 1, \\ \underline{\tilde{\boldsymbol{A}}}^{(l-1)}\tilde{\boldsymbol{\sigma}}_1^{(l-1)} + \tilde{\boldsymbol{A}}^{(l)}\tilde{\boldsymbol{\sigma}}^{(l)} + \underset{\sim}{\tilde{\boldsymbol{A}}}^{(l+1)}\tilde{\boldsymbol{\sigma}}_S^{(l+1)}, & l = 2, 3, \ldots, L-1, \\ \underline{\tilde{\boldsymbol{A}}}^{(l-1)}\tilde{\boldsymbol{\sigma}}_1^{(l-1)} + \tilde{\boldsymbol{A}}^{(l)}\tilde{\boldsymbol{\sigma}}^{(l)}, & l = L. \end{cases} \tag{15}$$

Define

$$\underline{\tilde{\boldsymbol{y}}}^{(l)} = \begin{cases} \tilde{\boldsymbol{y}}^{(l)}, & l = 1, \\ \tilde{\boldsymbol{y}}^{(l)} - \underline{\tilde{\boldsymbol{A}}}^{(l-1)}\tilde{\boldsymbol{\sigma}}_1^{(l-1)}, & l = 2, 3, \ldots, L, \end{cases} \quad \text{and} \quad \underline{\tilde{\boldsymbol{n}}}^{(l)} = \begin{cases} \underset{\sim}{\tilde{\boldsymbol{A}}}^{(l+1)}\tilde{\boldsymbol{\sigma}}_S^{(l+1)}, & l = 1, 2, \ldots, L-1, \\ \boldsymbol{0}_{M_p \times N_p}, & l = L. \end{cases} \tag{16}$$

Then, Eq. (15) can be represented as

$$\tilde{\underline{\boldsymbol{y}}}^{(l)} = \tilde{\boldsymbol{A}}^{(l)}\tilde{\boldsymbol{\sigma}}^{(l)} + \tilde{\underline{\boldsymbol{n}}}^{(l)}, \quad l = 1, 2, \ldots, L. \tag{17}$$

This is a general form of compressive reconstruction in the presence of noise.

It should be noted that, different from the formulation in Figure 2(a), the sub-matrix $\tilde{\boldsymbol{A}}^{(l)}$ in the above expression satisfies the RIP condition as long as $\boldsymbol{A}$ does. This is because each column of $\tilde{\boldsymbol{A}}^{(l)}$ completely contains all nonzero elements of the corresponding column of $\boldsymbol{A}$.

As revealed by $\tilde{\underline{\boldsymbol{n}}}^{(l)}$ in (16), we treat the partial measurements in the $(l+1)$-th segment as the noise in the $l$-th segment. Thus, there will be an estimation error in solving (17). Meantime, the model (17) is formulated under the assumption that $\tilde{\boldsymbol{\sigma}}^{(l-1)}$ is exactly estimated. Considering an inaccurate estimate of $\tilde{\boldsymbol{\sigma}}^{(l-1)}$, we refine the model (17). Let $\hat{\tilde{\boldsymbol{\sigma}}}_1^{(l-1)}$ be the estimate of $\tilde{\boldsymbol{\sigma}}^{(l-1)}$ and $\Delta\tilde{\boldsymbol{\sigma}}_1^{(l-1)} = \tilde{\boldsymbol{\sigma}}_1^{(l-1)} - \hat{\tilde{\boldsymbol{\sigma}}}_1^{(l-1)}$ be the error between $\tilde{\boldsymbol{\sigma}}^{(l-1)}$ and $\hat{\tilde{\boldsymbol{\sigma}}}_1^{(l-1)}$. Then, Eq. (15) becomes

$$\tilde{\underline{\boldsymbol{y}}}^{(l)} = \begin{cases} \tilde{\boldsymbol{A}}^{(l)}\tilde{\boldsymbol{\sigma}}^{(l)} + \tilde{\underline{\underline{\boldsymbol{A}}}}^{(l+1)}\tilde{\boldsymbol{\sigma}}_S^{(l+1)}, & l = 1, \\ \tilde{\underline{\boldsymbol{A}}}^{(l-1)}\hat{\tilde{\boldsymbol{\sigma}}}_1^{(l-1)} + \tilde{\boldsymbol{A}}^{(l)}\tilde{\boldsymbol{\sigma}}^{(l)} + \tilde{\underline{\boldsymbol{A}}}^{(l-1)}\Delta\tilde{\boldsymbol{\sigma}}_1^{(l-1)} + \tilde{\underline{\underline{\boldsymbol{A}}}}^{(l+1)}\tilde{\boldsymbol{\sigma}}_S^{(l+1)}, & l = 2, 3, \ldots, L-1, \\ \tilde{\underline{\boldsymbol{A}}}^{(l-1)}\hat{\tilde{\boldsymbol{\sigma}}}_1^{(l-1)} + \tilde{\boldsymbol{A}}^{(l)}\tilde{\boldsymbol{\sigma}}^{(l)} + \tilde{\underline{\boldsymbol{A}}}^{(l-1)}\Delta\tilde{\boldsymbol{\sigma}}_1^{(l-1)}, & l = L. \end{cases} \tag{18}$$

Define

$$\tilde{\underline{\boldsymbol{y}}}^{(l)} = \begin{cases} \tilde{\boldsymbol{y}}^{(l)}, & l = 1, \\ \tilde{\boldsymbol{y}}^{(l)} - \tilde{\underline{\boldsymbol{A}}}^{(l-1)}\hat{\tilde{\boldsymbol{\sigma}}}_1^{(l-1)}, & l = 2, \ldots, L, \end{cases} \text{ and } \tilde{\underline{\boldsymbol{n}}}^{(l)} = \begin{cases} \tilde{\underline{\underline{\boldsymbol{A}}}}^{(l+1)}\tilde{\boldsymbol{\sigma}}_S^{(l+1)}, & l = 1, \\ \tilde{\underline{\boldsymbol{A}}}^{(l-1)}\Delta\tilde{\boldsymbol{\sigma}}_1^{(l-1)} + \tilde{\underline{\underline{\boldsymbol{A}}}}^{(l+1)}\tilde{\boldsymbol{\sigma}}_S^{(l+1)}, & l = 2, \ldots, L-1, \\ \tilde{\underline{\boldsymbol{A}}}^{(l-1)}\Delta\tilde{\boldsymbol{\sigma}}_1^{(l-1)}, & l = L. \end{cases} \tag{19}$$

Then, we have

$$\tilde{\underline{\boldsymbol{y}}}^{(l)} = \tilde{\boldsymbol{A}}^{(l)}\tilde{\boldsymbol{\sigma}}^{(l)} + \tilde{\underline{\boldsymbol{n}}}^{(l)}, \ l = 1, 2, \ldots, L. \tag{20}$$

When $\tilde{\boldsymbol{\sigma}}^{(l)}$ is sparse and $\|\tilde{\underline{\boldsymbol{n}}}^{(l)}\|_2 \leqslant \eta^{(l)}$, we can obtain the estimate $\hat{\tilde{\boldsymbol{\sigma}}}^{(l)}$ of $\tilde{\boldsymbol{\sigma}}^{(l)}$ by solving

$$\min_{\tilde{\boldsymbol{\sigma}}^{(l)}} \left\|\tilde{\boldsymbol{\sigma}}^{(l)}\right\|_1 \quad \text{s.t.} \quad \left\|\tilde{\underline{\boldsymbol{y}}}^{(l)} - \tilde{\boldsymbol{A}}^{(l)}\tilde{\boldsymbol{\sigma}}^{(l)}\right\|_2 \leqslant \eta^{(l)}. \tag{21}$$

The "noise" $\tilde{\underline{\boldsymbol{n}}}^{(l)}$ is generated during estimation process that integrates the interferences from the estimation error in the previous segment and the partial measurement in the subsequent segment. For convenience, we refer to $\tilde{\underline{\boldsymbol{n}}}^{(l)}$ as the virtual noise sub-vector. Similarly, we name $\tilde{\underline{\boldsymbol{y}}}^{(l)}$ as the virtual measurement sub-vector. It is clear from (19) that the virtual noise $\tilde{\underline{\boldsymbol{n}}}^{(l)}$ consists of $\tilde{\underline{\boldsymbol{A}}}^{(l-1)}\tilde{\boldsymbol{\sigma}}_1^{(l-1)}$ and $\tilde{\underline{\underline{\boldsymbol{A}}}}^{(l+1)}\tilde{\boldsymbol{\sigma}}_S^{(l+1)}$ generated by the inaccurate estimate in the $(l-1)$-th segment and the partial measurement in the $(l+1)$-th segment. For convenience, we call them forward and backward virtual noises, respectively.

The virtual noise $\tilde{\underline{\boldsymbol{n}}}^{(l)}$ is closely related to measurement matrix. From the definitions of $\tilde{\underline{\boldsymbol{A}}}^{(l-1)}$ and $\tilde{\underline{\underline{\boldsymbol{A}}}}^{(l+1)}$ in (14), we observe that the forward and backward virtual noise exist only in the range of $[1 : M_p]$ and $[\tilde{M} - M_p + 1 : \tilde{M}]$, respectively. As to be discussed in the next section, the backward virtual noise level is in general much larger than that of the forward virtual noise.

It is seen from (20) that we can formulate a series of sparse reconstruction problems by properly segmenting the measurement vector and the coefficient vector. Then a large-scale reconstruction problem is decomposed into a series of small-scale reconstruction ones. We call the method as segment-sliding sparse reconstruction (SegSR). Algorithm 1 describes the SegSR process in detail.

**Algorithm 1** SegSR scheme

**Input:** $S$, $M_p$, $N_p$, $\tilde{M}$, $\tilde{N}$, and $L$;
**Output:** Estimated sparse vector $\hat{\boldsymbol{\sigma}}$.

1: Initialize $l = 1$.
2: Extract the $\tilde{M} \times 1$ measurement sub-vector $\tilde{\boldsymbol{y}}^{(l)}$ and the $\tilde{M} \times \tilde{N}$ measurement sub-matrix $\tilde{\boldsymbol{A}}^{(l)}$.
3: Calculate the virtual measurement sub-vector $\underline{\tilde{\boldsymbol{y}}}^{(l)}$ by (19).
4: Solve (21) and obtain an estimate $\hat{\tilde{\boldsymbol{\sigma}}}^{(l)}$ of $\tilde{\boldsymbol{\sigma}}^{(l)}$.
5: Let $l = l + 1$. If $l > L$, go to Step 6; otherwise, generate the $\tilde{M} \times \tilde{N}$ sub-matrix $\underline{\tilde{\boldsymbol{A}}}^{(l-1)}$ by (14), extract the sub-vector $\hat{\tilde{\boldsymbol{\sigma}}}_1^{(l-1)}$ from the estimated $\hat{\tilde{\boldsymbol{\sigma}}}^{(l-1)}$ and go to Step 2.
6: Formulate the estimated sparse vector $\hat{\boldsymbol{\sigma}}$ of $\boldsymbol{\sigma}$ as

$$\begin{cases} \hat{\boldsymbol{\sigma}}((l-1)N_p : lN_p - 1) = \hat{\tilde{\boldsymbol{\sigma}}}_1^{(l)}, l = 1, \ldots, L-1, \\ \hat{\boldsymbol{\sigma}}((l-1)N_p : N-1) = \hat{\tilde{\boldsymbol{\sigma}}}^{(L)}, \ l = L. \end{cases}$$

Solving (21) is one of the major steps in the implementation of the proposed technique. Theoretically, any sparse estimation algorithm can be used. However, the formulation (20) has three distinct characteristics: (a) the sparsity level of each segment is in general unknown; (b) there are some overlaps between two consecutive segments of the sparse vector $\boldsymbol{\sigma}$; and (c) the distributions of the virtual noise are different for each virtual measurement. These characteristics can be incorporated into the estimate of $\tilde{\boldsymbol{\sigma}}^{(l)}$ to improve the algorithm efficiency. Toward this objective, we make some modifications on the OMP algorithm with partially known support (OMP-PKS) [28] to solve (21). Algorithm 2 presents the computational flowchart of the algorithm, termed as the two-step OMP-process (TOMPP). TOMPP consists of three main processes. One is the initialization to take into account the partially known support from the previous segment. The second one is the OMP process with a threshold $\zeta_1$ to obtain a rough estimate (Steps 2–4) derived for the whole virtual noise. The last process (Steps 5–7) is the refinement OMP process with threshold $\zeta_2$ for the sparse estimate in sparse range $[0 : \tilde{N} - N_p - 1]$ because of the virtual noise structure. The effectiveness of the TOMPP is validated by theoretical analyses and simulation experiments in the next three sections.

**Algorithm 2** TOMPP

**Input:** $\underline{\tilde{\boldsymbol{y}}}^{(l)}$, $\tilde{\boldsymbol{A}}^{(l)}$, $N_p$, $\tilde{N}$, $\hat{\tilde{\boldsymbol{\sigma}}}^{(l-1)}$, threshold parameters $\zeta_1$ and $\zeta_2$ ($\zeta_2 < \zeta_1$);

**Output:** Estimated sparse vector $\hat{\tilde{\boldsymbol{\sigma}}}^{(l)}$.

1: Initialize the index set $\Lambda^{[0]}$ as the known support in the $l$-th segment estimate, $\Lambda^{[0]} = \text{support}(\hat{\tilde{\boldsymbol{\sigma}}}^{(l-1)}(N_p : \tilde{N} - 1))$; the residual $\tilde{\boldsymbol{r}}^{[0]} = (\boldsymbol{I} - \boldsymbol{P}_0)\underline{\tilde{\boldsymbol{y}}}^{(l)}$, where $\boldsymbol{P}_0 = \tilde{\boldsymbol{A}}^{(l)}_{\Lambda^{[0]}}(\tilde{\boldsymbol{A}}^{(l)}_{\Lambda^{[0]}})^{\dagger}$ denotes the projection onto the linear space spanned by the columns of $\tilde{\boldsymbol{A}}^{(l)}_{\Lambda^{[0]}}$. Let $i = 1$.
2: Find the column of $\tilde{\boldsymbol{A}}^{(l)}$ that is most correlated with the residual $\tilde{\boldsymbol{r}}^{[i-1]}$, i.e., $\lambda^{[i]} = \arg\max_{j=0,1,\ldots,\tilde{N}-1} |\langle \tilde{\boldsymbol{a}}_j^{(l)}, \tilde{\boldsymbol{r}}^{[i-1]} \rangle|$, and update the index set $\Lambda^{[i]} = \Lambda^{[i-1]} \cup \{\lambda^{[i]}\}$.
3: Update the residual $\tilde{\boldsymbol{r}}^{[i]} = (\boldsymbol{I} - \boldsymbol{P}_i)\underline{\tilde{\boldsymbol{y}}}^{(l)}$.
4: If $\|\tilde{\boldsymbol{r}}^{[i-1]}\|_2 - \|\tilde{\boldsymbol{r}}^{[i]}\|_2 \leqslant \zeta_1$, let $i = i + 1$ and go to Step 5; otherwise, let $i = i + 1$ and return to Step 2.
5: For columns $0 \sim \tilde{N} - N_p - 1$ of $\tilde{\boldsymbol{A}}^{(l)}$, find the column that is most correlated with the residual $\tilde{\boldsymbol{r}}^{[i-1]}$, i.e., $\lambda^{[i]} = \arg\max_{j=0,1,\ldots,\tilde{N}-N_p-1} |\langle \tilde{\boldsymbol{a}}_j^{(l)}, \tilde{\boldsymbol{r}}^{[i-1]} \rangle|$, and update index set $\Lambda^{[i]} = \Lambda^{[i-1]} \cup \{\lambda^{[i]}\}$.
6: Update the residual $\tilde{\boldsymbol{r}}^{[i]} = (\boldsymbol{I} - \boldsymbol{P}_i)\underline{\tilde{\boldsymbol{y}}}^{(l)}$.
7: If $\|\tilde{\boldsymbol{r}}^{[i-1]}\|_2 - \|\tilde{\boldsymbol{r}}^{[i]}\|_2 \leqslant \zeta_2$, go to Step 8; otherwise, let $i = i + 1$ and return to Step 5.
8: Calculate the estimate $\hat{\tilde{\boldsymbol{\sigma}}}^{(l)}$ by $\hat{\tilde{\boldsymbol{\sigma}}}^{(l)}_{\Lambda^{[i]}} = (\tilde{\boldsymbol{A}}^{(l)}_{\Lambda^{[i]}})^{\dagger}\underline{\tilde{\boldsymbol{y}}}^{(l)}$.

## 5 Effects of errors on reconstruction performance

From the SegSR flowchart, we find that the estimation performance in each segment will be affected by the virtual noise. This section takes the OMP algorithm as an example to analyze the effect of the virtual noise on the estimation performance in terms of the sparse entry positions and sparse coefficients.

### 5.1 Effect of the virtual noise on position estimation

Refs. [29–31] conduct thorough analyses on the recovery performance of sparse positions in $\ell_2$-bounded noise and $\ell_\infty$-bounded noise. These analyses indicate that sparse positions can be estimated exactly by OMP for the nonzero sparse elements when their magnitudes are greater than some value. For the underlying problem considered in this paper, we first show that the virtual noise in (19) is $\ell_2$-bounded and $\ell_\infty$-bounded. We then derive the conditions for correct sparse position estimation.

Before proceeding, several notations are defined. Let $\Gamma^{(l)}$ and $\Gamma_s^{(l)}(s=1,\ldots,S)$ denote the support sets of $\tilde{\boldsymbol{\sigma}}^{(l)}$ and $\tilde{\boldsymbol{\sigma}}_s^{(l)}(s=1,\ldots,S)$, respectively. Let $\Delta\Gamma_1^{(l-1)}$ denote the support sets of $\Delta\tilde{\boldsymbol{\sigma}}_1^{(l-1)}$, and $\tilde{\delta}_k^{(l)}$ be the $k$-th order RIP constant of matrix $\tilde{\boldsymbol{A}}^{(l)}$. Then, for the virtual noise in (20), we have the following theorem.

**Theorem 1.** For $l=1,2,\ldots,L$, the virtual noise $\tilde{\underline{\boldsymbol{n}}}^{(l)}$ satisfies

$$\left\|\tilde{\underline{\boldsymbol{n}}}^{(l)}\right\|_2 \leqslant \varepsilon_2^{(l)} = \begin{cases} b_2^{(l+1)}, & l=1, \\ a_2^{(l-1)} + b_2^{(l+1)}, & l=2,3,\ldots,L-1, \\ a_2^{(l-1)}, & l=L, \end{cases} \tag{22}$$

and

$$\left\|\left(\tilde{\boldsymbol{A}}^{(l)}\right)^{\mathrm{H}}\tilde{\underline{\boldsymbol{n}}}^{(l)}\right\|_\infty \leqslant \varepsilon_\infty^{(l)} = \begin{cases} b_\infty^{(l+1)}, & l=1, \\ a_\infty^{(l-1)} + b_\infty^{(l+1)}, & l=2,3,\ldots,L-1, \\ a_\infty^{(l-1)}, & l=L, \end{cases} \tag{23}$$

where $a_2^{(l-1)} = \sqrt{1+\tilde{\delta}^{(l-1)}_{|\Delta\Gamma_1^{(l-1)}|}}\|\Delta\tilde{\boldsymbol{\sigma}}_1^{(l-1)}\|_2$ and $a_\infty^{(l-1)} = \tilde{\delta}^{(l-1)}_{|\Gamma_1^{(l-1)}|+1}\|\Delta\tilde{\boldsymbol{\sigma}}_1^{(l-1)}\|_2$ for $l=2,3,\ldots,L$, and $b_2^{(l+1)} = \sqrt{1+\tilde{\delta}^{(l+1)}_{|\Delta\Gamma_S^{(l+1)}|}}\|\tilde{\boldsymbol{\sigma}}_S^{(l+1)}\|_2$ and $b_\infty^{(l+1)} = \tilde{\delta}^{(l+1)}_{|\Gamma_S^{(l+1)}|+1}\|\tilde{\boldsymbol{\sigma}}_S^{(l+1)}\|_2$ for $l=1,2,\ldots,L-1$.

*Proof.* See Appendix A.

From Theorem 1, it is clear that the virtual noise $\tilde{\underline{\boldsymbol{n}}}^{(l)}$ is both $\ell_2$-bounded and $\ell_\infty$-bounded and is limited by $\varepsilon_2^{(l)}$ and $\varepsilon_\infty^{(l)}$, respectively. The upper limits consist of two parts, $a_2^{(l-1)}$ ($a_\infty^{(l-1)}$) and $b_2^{(l+1)}$ ($b_\infty^{(l+1)}$), which are related to the estimate error $\Delta\tilde{\boldsymbol{\sigma}}_1^{(l-1)}$ in the previous segment and the sparse sub-vector $\tilde{\boldsymbol{\sigma}}_S^{(l+1)}$ in the subsequent segment, respectively. The noise limits $\varepsilon_2^{(l)}$ and $\varepsilon_\infty^{(l)}$ are usually dominated by $b_2^{(l+1)}$ and $b_\infty^{(l+1)}$, respectively.

For the $\ell_2$-bounded and $\ell_\infty$-bounded virtual noise, we have the following theorem to find the correct sparse positions.

**Theorem 2.** Suppose that $\|\tilde{\underline{\boldsymbol{n}}}^{(l)}\|_2 \leqslant \varepsilon_2^{(l)}(\|(\tilde{\boldsymbol{A}}^{(l)})^{\mathrm{H}}\tilde{\underline{\boldsymbol{n}}}^{(l)}\|_\infty \leqslant \varepsilon_\infty^{(l)})$ and $\tilde{\boldsymbol{A}}^{(l)}$ satisfies condition $\tilde{\delta}^{(l)}_{|\Gamma^{(l)}|+1} < 1/(\sqrt{|\Gamma^{(l)}|}+1)$. Then OMP with the stopping rule $\|\tilde{\boldsymbol{r}}^{[i]}\|_2 \leqslant \varepsilon_2^{(l)}$ ($\|(\tilde{\boldsymbol{A}}^{(l)})^{\mathrm{H}}\tilde{\boldsymbol{r}}^{[i]}\|_\infty \leqslant \varepsilon_\infty^{(l)}$) exactly recovers the support $\Gamma^{(l)}$ of $\tilde{\boldsymbol{\sigma}}^{(l)}$, if all the nonzero coefficients $\tilde{\sigma}_j^{(l)}$ ($j \in \Gamma^{(l)}$) satisfy

$$\left|\tilde{\sigma}_j^{(l)}\right| \geqslant \frac{\left(\sqrt{1+\tilde{\delta}^{(l)}_{|\Gamma^{(l)}|+1}}+1\right)\varepsilon_2^{(l)}}{1-\left(\sqrt{|\Gamma^{(l)}|}+1\right)\tilde{\delta}^{(l)}_{|\Gamma^{(l)}|+1}} \quad \left(\left|\tilde{\sigma}_j^{(l)}\right| \geqslant \frac{\left(\sqrt{1+\tilde{\delta}^{(l)}_{|\Gamma^{(l)}|+1}}+1\right)\sqrt{|\Gamma^{(l)}|}\varepsilon_\infty^{(l)}}{1-\left(\sqrt{|\Gamma^{(l)}|}+1\right)\tilde{\delta}^{(l)}_{|\Gamma^{(l)}|+1}}\right). \tag{24}$$

The proof of Theorem 2 closely follows that in [30, 31] and thus is omitted here.

Equipped with Theorem 2, we can find the correct sparse entry positions by using the OMP algorithm if the nonzero coefficients satisfy (24). However, it should be noted that the conditions are derived with overall effects of the virtual noise and the resulting minimum nonzero magnitude may be high. In this case, we may miss some nonzero elements of the sparse vector.

As discussed in the previous section, the virtual noise $\tilde{\underline{\boldsymbol{n}}}^{(l)}$ exists only in the ranges $[1:M_p]$ and $[\tilde{M}-M_p+1:\tilde{M}]$. Then, the columns of $\tilde{\boldsymbol{A}}_s^{(l)}(s=2,\ldots,S)$ are orthogonal to the forward virtual noise

$\tilde{\underline{\boldsymbol{A}}}^{(l-1)}\Delta\tilde{\boldsymbol{\sigma}}_1^{(l-1)}$ and the columns of $\tilde{\boldsymbol{A}}_s^{(l)}(s=1,\ldots,S-1)$ are orthogonal to the backward virtual noise $\tilde{\underline{\underline{\boldsymbol{A}}}}^{(l+1)}\tilde{\boldsymbol{\sigma}}_S^{(l+1)}$. In particular, when $S\geqslant 3$, the columns of $\tilde{\boldsymbol{A}}_s^{(l)}(s=2,\ldots,S-1)$ are orthogonal to both forward and backward virtual noise. By the principle of the OMP, the estimation of the sparse positions of $\tilde{\boldsymbol{\sigma}}_1^{(l)}$ is mainly affected by the forward virtual noise and, similarly, that of $\tilde{\boldsymbol{\sigma}}_S^{(l)}$ is mainly affected by the backward virtual noise. In general, the noise level of the backward virtual noise is higher than that of the forward virtual noise. Therefore, we can use a refinement OMP process in Algorithm 2 with a small threshold $\zeta_2$ to improve the estimate of the sparse vector $\tilde{\boldsymbol{\sigma}}^{(l)}$ within the range $[0:\tilde{N}-N_p-1]$.

### 5.2 Effect of the virtual noise on amplitude estimation

We now analyze the effect of the virtual noise on the estimated amplitude of sub-vectors $\tilde{\boldsymbol{\sigma}}_s^{(l)}$, $s=1,2,\ldots,S$, of $\tilde{\boldsymbol{\sigma}}^{(l)}$. To simplify the analysis, we assume that the support $\Gamma^{(l)}$ of $\tilde{\boldsymbol{\sigma}}^{(l)}$ is correctly recovered.

With the support $\Gamma^{(l)}$ of $\tilde{\boldsymbol{\sigma}}^{(l)}$, we obtain the estimate $\hat{\tilde{\boldsymbol{\sigma}}}^{(l)}_{\Gamma^{(l)}}$ of $\tilde{\boldsymbol{\sigma}}^{(l)}_{\Gamma^{(l)}}$ through the least squares solution of the following expression:

$$\tilde{\underline{\boldsymbol{y}}}^{(l)}=\tilde{\boldsymbol{A}}^{(l)}_{\Gamma^{(l)}}\tilde{\boldsymbol{\sigma}}^{(l)}_{\Gamma^{(l)}}+\tilde{\underline{\boldsymbol{n}}}^{(l)}=\sum_{s=1}^{S}\tilde{\boldsymbol{A}}^{(l)}_{\Gamma_s^{(l)}}\tilde{\boldsymbol{\sigma}}^{(l)}_{\Gamma_s^{(l)}}+\tilde{\underline{\boldsymbol{n}}}^{(l)},\ l=1,2,\ldots,L, \tag{25}$$

where

$$\tilde{\underline{\boldsymbol{n}}}^{(l)}=\begin{cases}\tilde{\underline{\underline{\boldsymbol{A}}}}^{(l+1)}_{\Gamma_S^{(l+1)}}\tilde{\boldsymbol{\sigma}}^{(l+1)}_{\Gamma_S^{(l+1)}}, & l=1,\\ \tilde{\underline{\boldsymbol{A}}}^{(l-1)}_{\Gamma_1^{(l-1)}}\Delta\tilde{\boldsymbol{\sigma}}^{(l-1)}_{\Gamma_1^{(l-1)}}+\tilde{\underline{\underline{\boldsymbol{A}}}}^{(l+1)}_{\Gamma_S^{(l+1)}}\tilde{\boldsymbol{\sigma}}^{(l+1)}_{\Gamma_S^{(l+1)}}, & l=2,3,\ldots,L-1,\\ \tilde{\underline{\boldsymbol{A}}}^{(l-1)}_{\Gamma_1^{(l-1)}}\Delta\tilde{\boldsymbol{\sigma}}^{(l-1)}_{\Gamma_1^{(l-1)}}, & l=L.\end{cases} \tag{26}$$

The estimated $\hat{\tilde{\boldsymbol{\sigma}}}^{(l)}_{\Gamma^{(l)}}$ is given by

$$\hat{\tilde{\boldsymbol{\sigma}}}^{(l)}_{\Gamma^{(l)}}=\left(\tilde{\mathbf{A}}^{(l)}_{\Gamma^{(l)}}\right)^{\dagger}\tilde{\underline{\boldsymbol{y}}}^{(l)}. \tag{27}$$

Theorem 3 describes the upper bound of the estimated error of $\hat{\tilde{\boldsymbol{\sigma}}}^{(l)}_{\Gamma^{(l)}}$.

**Theorem 3.** For $l=1,2,\ldots,L$, the least squares estimate error of (25), $\|\Delta\tilde{\boldsymbol{\sigma}}^{(l)}_{\Gamma_s^{(l)}}\|_2=\|\tilde{\boldsymbol{\sigma}}^{(l)}_{\Gamma_s^{(l)}}-\hat{\tilde{\boldsymbol{\sigma}}}^{(l)}_{\Gamma_s^{(l)}}\|_2$ $(s=1,\ldots,S)$, is upper bounded by

$$\left\|\Delta\tilde{\boldsymbol{\sigma}}^{(l)}_{\Gamma_s^{(l)}}\right\|_2\leqslant\begin{cases}\alpha^{S-s+1}\left\|\tilde{\boldsymbol{\sigma}}^{(l+1)}_{\Gamma_S^{(l+1)}}\right\|_2, & l=1,\\ \beta\left\|\Delta\tilde{\boldsymbol{\sigma}}^{(l-1)}_{\Gamma_1^{(l-1)}}\right\|_2+\alpha^{S-s+1}\left\|\tilde{\boldsymbol{\sigma}}^{(l+1)}_{\Gamma_S^{(l+1)}}\right\|_2, & l=2,3,\ldots,L-1,\\ \beta\left\|\Delta\tilde{\boldsymbol{\sigma}}^{(l-1)}_{\Gamma_1^{(l-1)}}\right\|_2, & l=L,\end{cases} \tag{28}$$

where $\beta=\alpha^s(1-\alpha^{2(S-s+1)})/(1-\alpha^2)$ and $\alpha=\bar{\delta}^{(l)}_{\bar{K}^{(l)}}/(1-\bar{\delta}^{(l)}_{\bar{K}^{(l)}})$, with $\bar{K}^{(l)}$ and $\bar{\delta}^{(l)}_k$ respectively defined as $\bar{K}^{(l)}=\max\{|\Gamma^{(l-1)}|,|\Gamma^{(l)}|,|\Gamma^{(l+1)}|\}$ and $\bar{\delta}^{(l)}_k=\max\{\tilde{\delta}^{(l-1)}_k,\tilde{\delta}^{(l)}_k,\tilde{\delta}^{(l+1)}_k\}$.

*Proof.* See Appendix B.

The error bound (28) not only gives the least squares bound but also shows the bound distribution in each segment. Note that, when $\bar{\delta}^{(l)}_{\bar{K}^{(l)}}<1/3$, we have $\alpha<1/2$ and $\beta<1$. From the relation $\beta=\alpha^s\left(1-\alpha^{2(S-s+1)}\right)/\left(1-\alpha^2\right)$, we know that $\beta$ decreases as $s$ increases. Therefore, the estimation error due to the forward virtual noise decreases as $s$ increases. However, the error introduced by the backward virtual noise increases with $s$ because of the $\alpha^{S-s+1}$ term. Similar to the analyses on the noise bounds, the error bound of the current segment is usually dominated by the error introduced in the subsequent segment. With the effect behavior of the subsequent segment, we can deduce that the sub-vector $\tilde{\boldsymbol{\sigma}}_1^{(l)}$ of the sparse vector $\tilde{\boldsymbol{\sigma}}^{(l)}$ can be more accurately estimated than other sub-vectors $\tilde{\boldsymbol{\sigma}}_s^{(l)}(s=2,\ldots,S)$. Therefore, we downward slide one pulse width in Algorithm 1.

**Table 1** Storage requirements

| Reconstruction by (3) | SegSR scheme | | |
|---|---|---|---|
| | $S=2$ | $S=3$ | $S=4$ |
| 92.02 GB | 9.16 MB | 18.31 MB | 30.52 MB |

## 6 Storage requirements and computational complexity

### 6.1 Storage requirements

For the reconstruction problem (3), we need to store $M \times N$ matrix $\boldsymbol{A}$ which occupies $8MN = 8P(P-1)$ $M_pN_p$ bytes using the standard IEEE double precision. For the SegSR scheme, we only need to store $\tilde{M} \times \tilde{N}$ sub-matrix $\tilde{\boldsymbol{A}}^{(l)}$ involved in the sub-problem (21). Therefore, the required storage space is $8\tilde{M}\tilde{N} = 8S\,(S+1)\,M_pN_p$ bytes. Table 1 shows the storage requirements for an example with $P = 249$, $M_p = 200$ and $N_p = 1000$. It is clear that the storage space is significantly reduced when the SegSR scheme is used.

### 6.2 Computational complexity

The computational complexity of solving the problem (3) by the OMP algorithm is known to be $O(KMN)$ [15]. For the proposed SegSR scheme, the sub-problem (21) is solved by the TOMPP. Although the TOMPP contains two OMP processes, each OMP process estimates different sparse elements in the sparse sub-vector and, therefore, the complexity of the TOMPP is $O(|\Gamma^{(l)}|\tilde{M}\tilde{N})$. To solve the problem (3), the SegSR scheme requires a computational complexity of $O(\tilde{M}\tilde{N}\sum_{i=1}^{L}|\Gamma^{(l)}|)$. When the non-zeros positions in the sparse vector are uniformly distributed, it can be shown that $\tilde{M}\tilde{N}\sum_{i=1}^{L}|\Gamma^{(l)}| = (P-S)KS^2(S+1)M_pN_p/P$. It is noted that $KMN = KP(P-1)M_pN_p$ and $P \gg S$. Therefore, we obtain $\tilde{M}\tilde{N}\sum_{i=1}^{L}|\Gamma^{(l)}| \ll KMN$, which means that the proposed SegSR scheme has a much lower complexity to solve the problem (3) as compared to conventional OMP.

## 7 Simulation results

In this section, we evaluate the performance of the proposed SegSR scheme through several simulation experiments. Linear frequency modulated (LFM) pulse signals with a chirp rate of $\gamma = B/T_p$, expressed as $s(t) = \cos(\gamma\pi(t - T_p/2)^2)$, $0 \leqslant t \leqslant T_p$, is taken as an example. Unless otherwise specified, 500 realizations are conducted and the averaged results are presented.

To compare with direct reconstruction, we set the measurement size of the signal to be moderate. $B = 100$ MHz, $T_p = 10$ μs and $T = 100$ μs are used in the simulations. For the RD AIC, we set $B_p = 100$ MHz and $T_{\text{int}} = 0.05$ μs, which imply that the sampling rate of the ADC is 20 MHz and the down-sampling rate is $R = 5$. With these parameters, we can obtain $N = 10000$ Nyquist samples and $M = 2000$ measurements in a receive interval, and $N_p = 1000$ Nyquist samples and $M_p = 200$ measurements in a pulse width.

For received echo model (4), we assume that the echo amplitudes follow a uniform distribution in $(0, 1]$, and the time delays are randomly chosen in resolution grids. Furthermore, we assume that each element of coefficient vector $\boldsymbol{\sigma}$ takes a nonzero value with probability $p$, and zero value with probability $(1-p)$. Then the mean sparsity of coefficient vector $\boldsymbol{\sigma}$ is equal to $p$ times the Nyquist samples over the receive time.

### 7.1 Properties of virtual noise

We first present the existence range of the virtual noise. Note that, when $M_p = 200$ and $S = 3$, the length of the virtual measurement in a segment is $\tilde{M} = 800$. In this case, the forward virtual noise exists in $[1 : 200]$ and the backward virtual noise exists in $[601 : 800]$. Figure 3(a) shows a realization of the virtual noise with respect to the measurement time in the second segment. The result is obtained with

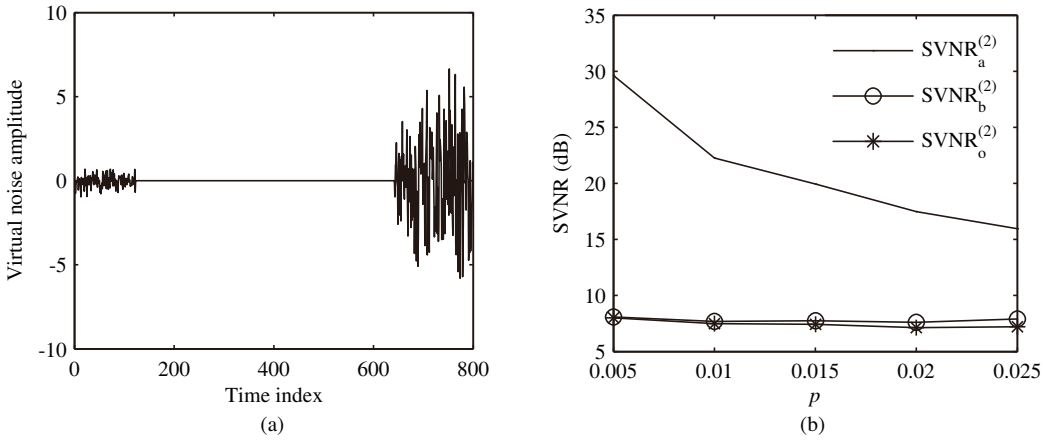


**Figure 3** (a) Variations of virtual noise amplitude; (b) signal to virtual noise ratio.

$p = 0.005$ and $S = 3$, and Eq. (21) is solved by the OMP-PKS. The result presented in Figure 3(a) is consistent with the theoretical analysis.

Next, we measure the virtual noise level. We define the following three signal-to-virtual-noise ratio (SVNR) metrics:

$$\mathrm{SVNR}_o^{(l)} = \frac{\left\|\tilde{\boldsymbol{A}}^{(l)}\tilde{\boldsymbol{\sigma}}^{(l)}\right\|_2^2}{E\left[\left\|\tilde{\underline{\boldsymbol{n}}}^{(l)}\right\|_2^2\right]}, \quad \mathrm{SVNR}_a^{(l)} = \frac{\left\|\tilde{\boldsymbol{A}}^{(l)}\tilde{\boldsymbol{\sigma}}^{(l)}\right\|_2^2}{E\left[\left\|\tilde{\underline{\boldsymbol{A}}}^{(l-1)}\Delta\tilde{\boldsymbol{\sigma}}_1^{(l-1)}\right\|_2^2\right]}, \quad \mathrm{SVNR}_b^{(l)} = \frac{\left\|\tilde{\boldsymbol{A}}^{(l)}\tilde{\boldsymbol{\sigma}}^{(l)}\right\|_2^2}{E\left[\left\|\tilde{\underset{\sim}{\boldsymbol{A}}}^{(l+1)}\tilde{\boldsymbol{\sigma}}_S^{(l+1)}\right\|_2^2\right]},$$

to describe the levels of total virtual noise, forward virtual noise, and backward virtual noise in the $l$-th segment, respectively. Figure 3(b) shows the results of the second segment for different values of $p$, where $S = 3$, and the results are obtained using the OMP-PKS. It is clear that, because the backward noise level is much larger than that of the forward virtual noise, the virtual noise is dominated by the backward virtual noise.

### 7.2 Signal reconstruction in the noise-free case

We now compare the reconstruction error and the running time of the proposed SegSR scheme with the OMP-PKS and the TOMPP. For simplification, the SegSR scheme with OMP-PKS and TOMPP is simply denoted as OMP-PKS and TOMPP, and the direct reconstruction by OMP is denoted as OMP.

(1) Reconstruction error. We take the relative reconstruction error $E_r = \|\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}\|_2/\|\boldsymbol{\sigma}\|_2$ as the performance criterion. Figure 4(a) shows the relative reconstruction errors versus $p$ for different number of pulses $S$ in a segment. The reconstruction errors by the OMP are not shown here because they derive much smaller errors in the noise-free case. It is seen that the relative reconstruction errors by the TOMPP are much lower than by the OMP-PKS. This is because the TOMPP can find sparse coefficients with smaller amplitudes than the OMP-PKS can do, as we discussed in Section 5. Figure 4(b) further validates the observation through the correct discovery rate (CDR), which is defined as the ratio of the number of coefficients correctly declared as nonzero to the total number of nonzero coefficients.

In addition, we notice that the relative reconstruction errors $E_r$ by the OMP-PKS remain unchanged for different values of $S$. However, the relative reconstruction error $E_r$ by the TOMPP substantially decreases when $S$ is changed from 2 to 3. When $S = 4$, the relative reconstruction error $E_r$ does not significantly change from the $S = 3$ case. This is because the virtual noise has different effects on the estimate error $\|\Delta\tilde{\boldsymbol{\sigma}}_s^{(l)}\|_2$ for $s = 1, \ldots, S$. To elaborate this, we show in Figure 5 the variation of $\|\Delta\tilde{\boldsymbol{\sigma}}_s^{(l)}\|_2$ versus $s$ in the second segment with $p = 0.01$. It is seen that $\|\Delta\tilde{\boldsymbol{\sigma}}_s^{(l)}\|_2$ for $s = 1$ is the smallest among all $\|\Delta\tilde{\boldsymbol{\sigma}}_s^{(l)}\|_2$, $s = 1, \ldots, S$. Meanwhile, $\|\Delta\tilde{\boldsymbol{\sigma}}_s^{(l)}\|_2$ for $s = 1$ and $S = 3$ is almost same as for $s = 1$ and $S = 4$. It is seen that the TOMPP achieves the optimal performance in terms of reconstruction error when we assign the segment length as 4 times the pulse width.

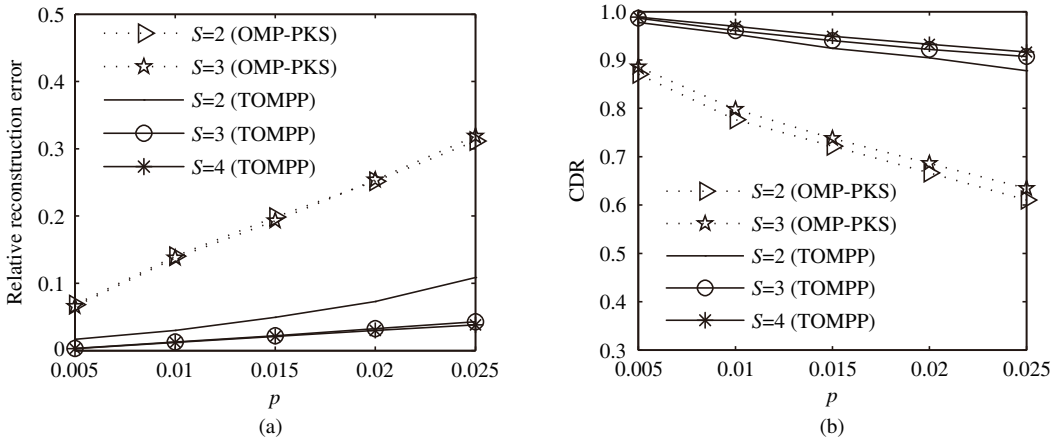


**Figure 4** (a) Relative reconstruction error; (b) correct discovery rate.

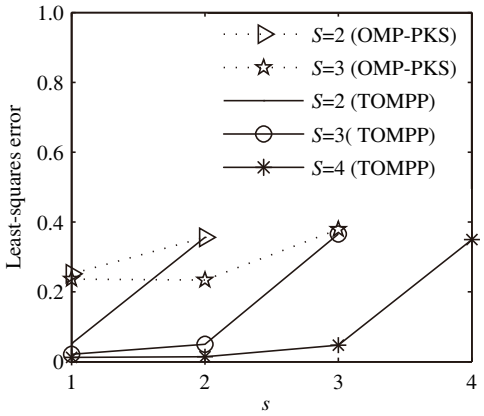


**Figure 5** $\left\|\Delta\tilde{\boldsymbol{\sigma}}_s^{(2)}\right\|_2$ versus $s$.

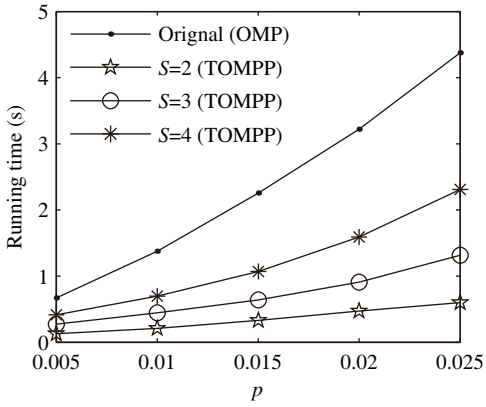


**Figure 6** Running time.

(2) Running time. We use CPU time to demonstrate the running time. The simulation is performed in the MATLAB 2011b environment on a PC with 3.1 GHz Intel core i5-2400 processor and 4 GB RAM. Figure 6 shows that the TOMPP is faster than the OMP. It should be noted that the simulations are only illustrative for a moderate size example. However, the TOMPP solves large-scale reconstruction problems which cannot be solved by the OMP.

### 7.3 Signal reconstruction in the noise case

We assume that the received baseband signal is corrupted by band-limited, additive, and white Gaussian noise $n(t)$ with bandwidth $B/2$ and power spectral density $N_0/2$. We use the reconstruction signal-to-noise ratio (RSNR), RSNR $= \|\boldsymbol{\Psi\sigma}\|_2^2/E[\|\boldsymbol{\Psi}(\hat{\boldsymbol{\sigma}}-\boldsymbol{\sigma})\|_2^2]$, to evaluate the reconstruction performance.

The variations of RSNR versus $p$ obtained from the TOMPP and the OMP-PKS are shown in Figure 7(a) for a given input signal-to-noise ratio (ISNR), defined as ISNR $= 2\int_0^{\mathrm{T}}|x(t)|^2\mathrm{d}t/(TN_0B)$. Figure 7(b) shows the variations of the RSNR versus the ISNR for the TOMPP and the OMP-PKS, where $p = 0.01$ is assumed. For comparison, the RSNR results obtained from the OMP are aslo depicted. It is seen that the RSNRs obtained by both the TOMPP and the OMP-PKS are smaller than those by the OMP. However, the RSNRs offered by the TOMPP are close to those by the OMP and are much higher than the OMP-PKS. When the ISNR is low, e.g., ISNR $\leqslant 0$ dB, the OMP, OMP-PKS, and TOMPP all perform poor. This is because that they can only find large nonzero coefficients. For high ISNR values, the TOMPP yields much better results than the OMP-PKS, thus further validating the ability of the TOMPP of finding small nonzero coefficients. The simulation results in a noisy case further confirm that

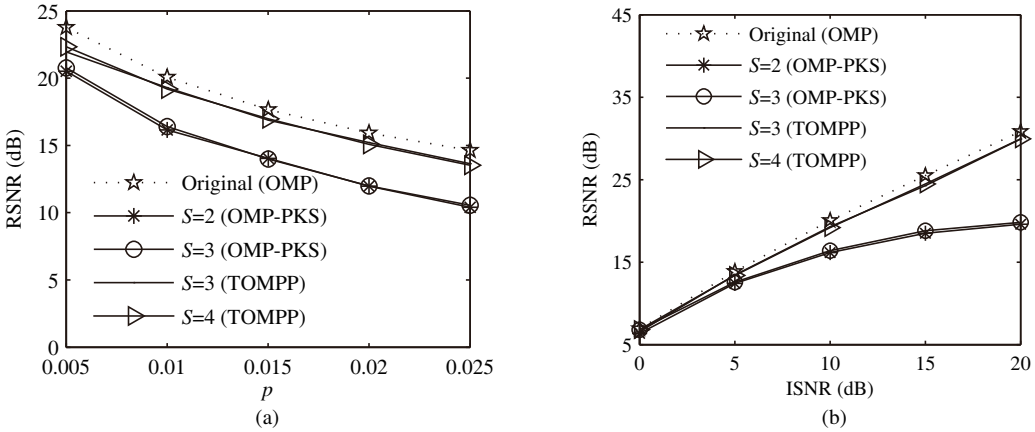


**Figure 7** (a) RSNR versus $p$ for ISNR = 10 dB; (b) RSNR versus ISNR for $p = 0.01$.

the segment length should be set to 4 for optimal segmentation.

## 8 Conclusion

In this paper, we have developed a SegSR scheme for implementable recovery of full-range pulsed radar echoes from AIC-based sub-Nyquist samples. In summary, the works conducted include (a) developing a novel segmenting approach such that, for the first time, the RIP condition is satisfied in each segment, (b) providing a rigorous analysis of the virtual noise caused by the segmenting, (c) proposing a two-step OMP algorithm that performs segment-sliding reconstruction with virtual noise elimination capabilities, and (d) providing extensive simulation results that clearly support our analytical results.

The proposed SegSR scheme decomposes a large-scale reconstruction problem into a series of small-scale reconstruction ones, thereby significantly reducing the storage and computational load. The SegSR is implemented in a sliding mode and thus is suitable for real-time applications.

This paper takes the RD AIC system as an example to introduce the proposed SegSR scheme. We maintain that it can be applied to other AIC systems in which the measurement matrices have a similar structure.

**Acknowledgements** This work was supported by National Natural Science Foundation of China (Grant Nos. 61171166, 61401210, 61571228), and China Postdoctoral Science Foundation (Grant No. 2014M551597).

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

1 Donoho D L. Compressed sensing. IEEE Trans Inf Theory, 2006, 52: 1289–1306
2 Candès E J, Romberg J K, Tao T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans Inf Theory, 2006, 52: 489–509
3 Candès E J, Tao T. Decoding by linear programming. IEEE Trans Inf Theory, 2005, 51: 4203–4215
4 Tropp J A, Laska J N, Duarte M F, et al. Beyond Nyquist: efficient sampling of sparse bandlimited signals. IEEE Trans Inf Theory, 2010, 56: 520–544
5 Becker S. Practical compressed sensing: modern data acquisition and signal processing. Dissertation for Ph.D. Degree. Pasadena: California Institute of Technology, 2011
6 Mishali M, Eldar Y C, Elron A J. Xampling: signal acquisition and processing in union of subspaces. IEEE Trans Signal Process, 2011, 59: 4719–4734
7 Xi F, Chen S, Liu Z. Quadrature compressive sampling for radar echo signals. In: Proceedings of the International Conference on Wireless Communications and Signal Processing (WCSP), Nanjing, 2011. 1–5
8 Xi F, Chen S, Liu Z. Quadrature compressive sampling for radar signals. IEEE Trans Signal Process, 2014, 62: 2787–2802
9 Bu H X, Bai X, Tao R. Compressed sensing SAR imaging based on sparse representation in fractional Fourier domain. Sci China Inf Sci, 2012, 55: 1789–1800

10 Hou Q K, Liu Y, Fan L J, et al. Compressed sensing digital receiver and orthogonal reconstructing algorithm for wideband ISAR radar. Sci China Inf Sci, 2015, 58: 020302
11 Zeng J S, Fang J, Xu Z B. Sparse SAR imaging based on L1/2 regularization. Sci China Inf Sci, 2012, 55: 1755–1775
12 Bar-Ilan O, Eldar Y C. Sub-Nyquist radar via Doppler focusing. IEEE Trans Signal Process, 2014, 62: 1796–1811
13 Yoo J, Turnes C, Nakamura E B, et al. A compressed sensing parameter extraction platform for radar pulse signal acquisition. IEEE J Emerg Sel Topics Circ Syst, 2012, 2: 626–638
14 Liu C, Xi F, Chen S, et al. Pulse-doppler signal processing with quadrature compressive sampling. IEEE Trans Aerosp Electron Syst, 2015, 51: 1217–1230
15 Tropp J A, Gilbert A C. Signal recovery from random measurements via orthogonal matching pursuit. IEEE Trans Inf Theory, 2007, 53: 4655–4666
16 Needell D, Tropp J A. CoSaMP: iterative signal recovery from incomplete and inaccurate samples. Appl Comput Harmon Anal, 2009, 26: 301–321
17 Yin W, Osher S, Goldfarb D, et al. Bregman iterative algorithm for $\ell_1$-minimization with applications to compressive sensing. SIAM J Imaging Sci, 2008, 1: 143–168
18 Becker S R, Candès E J, Grant M C. Templates for convex cone problems with applications to sparse signal recovery. Math Program Comput, 2011, 3: 165–218
19 Ji S, Dunson D, Carin L. Multitask compressive sensing. IEEE Trans Signal Process, 2009, 57: 92–106
20 Wu Q, Zhang Y D, Amin M G, et al. Complex multitask Bayesian compressive sensing. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Florence, 2014. 3375–3379
21 Fornasier M. Numerical methods for sparse recovery. Radon Series Comp Appl Math, 2010, 9: 1–110
22 Shi G, Lin J, Chen X, et al. UWB echo signal detection with ultra-low rate sampling based on compressed sensing. IEEE Trans Circ Syst Express Briefs, 2008, 55: 379–383
23 Freris N M, Öçal O, Vetterli M. Compressed sensing of streaming data. In: Proceedings of the 51st Annual Allerton Conference on Communication, Control, and Computing, Monticello, 2013. 1242–1249
24 Freris N M, Öçal O, Vetterli M. Recursive compressed sensing. ArXiv:1312.4895, 2013
25 Petros T B, Asif M S. Compressive sensing for streaming signals using the streaming greedy pursuit. In: Proceedings of the Military Communications Conference (MILCOM), San Jose, 2010. 1205–1210
26 Asif M S, Romberg J. Sparse recovery of streaming signals using $\ell_1$-homotopy. IEEE Trans Signal Process, 2014, 62: 4209–4223
27 Qin S, Zhang Y D, Wu Q, et al. Large-scale sparse reconstruction through partitioned compressive sensing. In: Proceedings of the 19th International Conference on Digital Signal Processing, Hong Kong, 2014. 837–840
28 Carrillo R E, Polania L F, Barner K E. Iterative algorithms for compressed sensing with partially known support. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Dallas, 2010. 3654–3657
29 Cai T T, Wang L. Orthogonal matching pursuit for sparse signal recovery with noise. IEEE Trans Inf Theory, 2011, 57: 4680–4688
30 Wu R, Huang W, Chen D. The exact support recovery of sparse signals with noise via orthogonal matching pursuit. IEEE Signal Process Lett, 2013, 20: 403–406
31 Dan W, Wang R. Robustness of orthogonal matching pursuit under restricted isometry property. Sci China Math, 2014, 57: 627–634


## Appendix A Proof of Theorem 1

To prove Theorem 1, we introduce some properties of restricted isometry constant of matrix $\boldsymbol{A}$. Such properties have appeared in different publications. The following Lemma 1 gives a summary of these results. One can refer to [3,16] for more details.

**Lemma 1.** Assume that the matrix $\boldsymbol{A}$ has a restricted isometry constant $\delta_K$. Denote $\Gamma$ and $\Gamma'$ as disjoint support sets of the sparse vectors $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$, respectively, where $|\Gamma| \leqslant k$, $|\Gamma'| \leqslant k'$ and $k + k' \leqslant K$. Then,

(i) $|\langle \boldsymbol{A\sigma}, \boldsymbol{A\sigma}' \rangle| \leqslant \delta_{k+k'} \|\boldsymbol{\sigma}\|_2 \|\boldsymbol{\sigma}'\|_2$;

(ii) $\left\| \boldsymbol{A}_{\Gamma}^{\mathrm{H}} \boldsymbol{A}_{\Gamma'} \right\|_2 \leqslant \delta_K$;

(iii) $\|(\boldsymbol{A}_{\Gamma}^{\mathrm{H}} \boldsymbol{A}_{\Gamma})^{-1}\|_2 \leqslant \frac{1}{1-\delta_K}$;

(iv) for any positive integer $k \leqslant K$, $\delta_k \leqslant \delta_K$.

**Proof of Theorem 1**: First, we show that the virtual noise $\tilde{\underline{\boldsymbol{n}}}^{(l)}$ is $\ell_2$-bounded. With the definition (19) of $\tilde{\underline{\boldsymbol{n}}}^{(l)}$, for $l = 2, 3, \ldots, L-1$, we have

$$\begin{aligned} \left\| \tilde{\underline{\boldsymbol{n}}}^{(l)} \right\|_2 &\leqslant \left\| \underline{\tilde{\boldsymbol{A}}}^{(l-1)} \Delta \tilde{\boldsymbol{\sigma}}_1^{(l-1)} \right\|_2 + \left\| \underline{\tilde{\boldsymbol{A}}}^{(l+1)} \tilde{\boldsymbol{\sigma}}_S^{(l+1)} \right\|_2 \leqslant \left\| \tilde{\boldsymbol{A}}^{(l-1)} \Delta \tilde{\boldsymbol{\sigma}}_1^{(l-1)} \right\|_2 + \left\| \tilde{\boldsymbol{A}}^{(l+1)} \tilde{\boldsymbol{\sigma}}_S^{(l+1)} \right\|_2 \\ &\leqslant \sqrt{1 + \tilde{\delta}^{(l-1)}_{|\Delta\Gamma_1^{(l-1)}|}} \left\| \Delta \tilde{\boldsymbol{\sigma}}_1^{(l-1)} \right\|_2 + \sqrt{1 + \tilde{\delta}^{(l+1)}_{|\Gamma_S^{(l+1)}|}} \left\| \tilde{\boldsymbol{\sigma}}_S^{(l+1)} \right\|_2. \end{aligned} \tag{A1}$$

The last inequality is from the RIP (2). Define $a_2^{(l-1)} = \sqrt{1 + \tilde{\delta}^{(l-1)}_{|\Delta\Gamma_1^{(l-1)}|}} \|\Delta \tilde{\boldsymbol{\sigma}}_1^{(l-1)}\|_2$ and $b_2^{(l+1)} = \sqrt{1 + \tilde{\delta}^{(l+1)}_{|\Gamma_S^{(l+1)}|}} \|\tilde{\boldsymbol{\sigma}}_S^{(l+1)}\|_2$. Then (A1) can be simplified as

$$\left\| \tilde{\underline{\boldsymbol{n}}}^{(l)} \right\|_2 \leqslant a_2^{(l-1)} + b_2^{(l+1)}, \ l = 2, 3, \ldots, L-1. \tag{A2}$$

For $l=1$, $\|\tilde{\underline{\boldsymbol{n}}}^{(1)}\|_2=\|\tilde{\underline{\boldsymbol{A}}}^{(2)}\tilde{\boldsymbol{\sigma}}_S^{(2)}\|_2\leqslant b_2^{(2)}$. For $l=L$, $\|\tilde{\underline{\boldsymbol{n}}}^{(L)}\|_2=\|\tilde{\underline{\boldsymbol{A}}}^{(L-1)}\Delta\tilde{\boldsymbol{\sigma}}_1^{(L-1)}\|_2\leqslant a_2^{(L-1)}$.

Next, we show that the virtual noise $\tilde{\underline{\boldsymbol{n}}}^{(l)}$ is $\ell_\infty$-bounded. Similarly, for $l=2,3,\ldots,L-1$, we have

$$\left\|\left(\tilde{\boldsymbol{A}}^{(l)}\right)^{\mathrm{H}}\tilde{\underline{\boldsymbol{n}}}^{(l)}\right\|_\infty\leqslant\left\|\left(\tilde{\boldsymbol{A}}^{(l)}\right)^{\mathrm{H}}\tilde{\underline{\boldsymbol{A}}}^{(l-1)}\Delta\tilde{\boldsymbol{\sigma}}_1^{(l-1)}\right\|_\infty+\left\|\left(\tilde{\boldsymbol{A}}^{(l)}\right)^{\mathrm{H}}\tilde{\underline{\boldsymbol{A}}}^{(l+1)}\tilde{\boldsymbol{\sigma}}_S^{(l+1)}\right\|_\infty.\tag{A3}$$

Note that $\tilde{\underline{\boldsymbol{A}}}^{(l-1)}$ is orthogonal to $\tilde{\boldsymbol{A}}_s^{(l)}$ for $s=2,3,\ldots,S$, and $\tilde{\underline{\boldsymbol{A}}}^{(l+1)}$ is orthogonal to $\tilde{\boldsymbol{A}}_s^{(l)}$ for $s=1,2,\ldots,S-1$. Then, we have

$$\left\|\left(\tilde{\boldsymbol{A}}^{(l)}\right)^{\mathrm{H}}\tilde{\underline{\boldsymbol{n}}}^{(l)}\right\|_\infty\leqslant\left\|\left(\tilde{\boldsymbol{A}}_1^{(l)}\right)^{\mathrm{H}}\tilde{\underline{\boldsymbol{A}}}^{(l-1)}\Delta\tilde{\boldsymbol{\sigma}}_1^{(l-1)}\right\|_\infty+\left\|\left(\tilde{\boldsymbol{A}}_S^{(l)}\right)^{\mathrm{H}}\tilde{\underline{\boldsymbol{A}}}^{(l+1)}\tilde{\boldsymbol{\sigma}}_S^{(l+1)}\right\|_\infty.\tag{A4}$$

We discuss the two terms in the right-hand side of (A4) separately. For the first term, note that $(\tilde{\boldsymbol{A}}_1^{(l)})^{\mathrm{H}}\tilde{\underline{\boldsymbol{A}}}^{(l-1)}=(\tilde{\boldsymbol{A}}_2^{(l-1)})^{\mathrm{H}}\tilde{\boldsymbol{A}}_1^{(l-1)}$. Then, by Property (i) of Lemma 1, we have

$$\begin{aligned}\left\|\left(\tilde{\boldsymbol{A}}_1^{(l)}\right)^{\mathrm{H}}\tilde{\underline{\boldsymbol{A}}}^{(l-1)}\Delta\tilde{\boldsymbol{\sigma}}_1^{(l-1)}\right\|_\infty&=\left\|\left(\tilde{\boldsymbol{A}}_2^{(l-1)}\right)^{\mathrm{H}}\tilde{\boldsymbol{A}}_1^{(l-1)}\Delta\tilde{\boldsymbol{\sigma}}_1^{(l-1)}\right\|_\infty\\&=\max_{j\in\{N_p,N_p+1,\ldots,2N_p-1\}}\left|\left\langle\tilde{\boldsymbol{A}}_j^{(l-1)},\tilde{\boldsymbol{A}}_1^{(l-1)}\Delta\tilde{\boldsymbol{\sigma}}_1^{(l-1)}\right\rangle\right|\leqslant\tilde{\delta}^{(l-1)}_{\left|\Delta\Gamma_1^{(l-1)}\right|+1}\left\|\Delta\tilde{\boldsymbol{\sigma}}_1^{(l-1)}\right\|_2.\end{aligned}\tag{A5}$$

Similarly, for the second term, we have

$$\left\|\left(\tilde{\boldsymbol{A}}_S^{(l)}\right)^{\mathrm{H}}\tilde{\underline{\boldsymbol{A}}}^{(l+1)}\tilde{\boldsymbol{\sigma}}_S^{(l+1)}\right\|_\infty=\left\|\left(\tilde{\boldsymbol{A}}_{S-1}^{(l+1)}\right)^{\mathrm{H}}\tilde{\boldsymbol{A}}_S^{(l+1)}\tilde{\boldsymbol{\sigma}}_S^{(l+1)}\right\|_\infty\leqslant\tilde{\delta}^{(l+1)}_{\left|\Gamma_S^{(l+1)}\right|+1}\left\|\tilde{\boldsymbol{\sigma}}_S^{(l+1)}\right\|_2.\tag{A6}$$

Substituting (A5) and (A6) into (A4) results in

$$\left\|\left(\tilde{\boldsymbol{A}}^{(l)}\right)^{\mathrm{H}}\tilde{\underline{\boldsymbol{n}}}^{(l)}\right\|_\infty\leqslant\tilde{\delta}^{(l-1)}_{|\Delta\Gamma_1^{(l-1)}|+1}\left\|\Delta\tilde{\boldsymbol{\sigma}}_1^{(l-1)}\right\|_2+\tilde{\delta}^{(l+1)}_{|\Gamma_S^{(l+1)}|+1}\left\|\tilde{\boldsymbol{\sigma}}_S^{(l+1)}\right\|_2.\tag{A7}$$

Define $a_\infty^{(l-1)}=\tilde{\delta}^{(l-1)}_{|\Gamma_1^{(l-1)}|+1}\|\Delta\tilde{\boldsymbol{\sigma}}_1^{(l-1)}\|_2$ and $b_\infty^{(l+1)}=\tilde{\delta}^{(l+1)}_{|\Gamma_S^{(l+1)}|+1}\|\tilde{\boldsymbol{\sigma}}_S^{(l+1)}\|_2$. We have

$$\left\|\left(\tilde{\boldsymbol{A}}^{(l)}\right)^{\mathrm{H}}\tilde{\underline{\boldsymbol{n}}}^{(l)}\right\|_\infty\leqslant a_\infty^{(l-1)}+b_\infty^{(l+1)},l=2,3,\ldots,L-1.\tag{A8}$$

Specifically, for $l=1$, we have $\|(\tilde{\boldsymbol{A}}^{(1)})^{\mathrm{H}}\tilde{\underline{\boldsymbol{n}}}^{(1)}\|_\infty=\|(\tilde{\boldsymbol{A}}^{(1)})^{\mathrm{H}}\tilde{\underline{\boldsymbol{A}}}^{(2)}\tilde{\boldsymbol{\sigma}}_S^{(2)}\|_\infty\leqslant b_\infty^{(2)}$; for $l=L$, $\|(\tilde{\boldsymbol{A}}^{(L)})^{\mathrm{H}}\tilde{\underline{\boldsymbol{n}}}^{(L)}\|_\infty=\|(\tilde{\boldsymbol{A}}^{(L)})^{\mathrm{H}}\tilde{\underline{\boldsymbol{A}}}^{(L-1)}\Delta\tilde{\boldsymbol{\sigma}}_1^{(L-1)}\|_\infty\leqslant a_\infty^{(L-1)}$. Then Theorem 1 is proved.

## Appendix B Proof of Theorem 3

Before proceeding, we derive a lemma which is used in the proof of Theorem 3.

**Lemma 2.** Let matrix $\boldsymbol{U}$ be of full column rank and be partitioned as $\boldsymbol{U}=[\boldsymbol{U}_1,\boldsymbol{U}_2,\ldots,\boldsymbol{U}_S]$. For $S\geqslant 3$, assume that $\boldsymbol{U}_s(s=1,\ldots,S-2)$ is orthogonal to $\boldsymbol{U}_{s+2}(s=1,\ldots,S-2)$. Then

$$\boldsymbol{U}^\dagger=\begin{bmatrix}\boldsymbol{U}_1^\dagger+\ldots+(-1)^{j-1}\boldsymbol{U}_1^\dagger\prod\limits_{i=2}^{j}\boldsymbol{U}_i\boldsymbol{C}_i^\dagger+\ldots+(-1)^{S-1}\boldsymbol{U}_1^\dagger\prod\limits_{i=2}^{S}\boldsymbol{U}_i\boldsymbol{C}_i^\dagger\\ \boldsymbol{C}_2^\dagger+\ldots+(-1)^{j-1}\boldsymbol{C}_2^\dagger\prod\limits_{i=3}^{j+1}\boldsymbol{U}_i\boldsymbol{C}_i^\dagger+\ldots+(-1)^{S-2}\boldsymbol{C}_2^\dagger\prod\limits_{i=3}^{S}\boldsymbol{U}_i\boldsymbol{C}_i^\dagger\\ \vdots\\ \boldsymbol{C}_s^\dagger+\ldots+(-1)^{j-1}\boldsymbol{C}_s^\dagger\prod\limits_{i=s+1}^{j+s-1}\boldsymbol{U}_i\boldsymbol{C}_i^\dagger+\ldots+(-1)^{S-s}\boldsymbol{C}_s^\dagger\prod\limits_{i=s+1}^{S}\boldsymbol{U}_i\boldsymbol{C}_i^\dagger\\ \vdots\\ \boldsymbol{C}_S^\dagger\end{bmatrix},\tag{B1}$$

where $\boldsymbol{C}_i=(\boldsymbol{I}-[\boldsymbol{U}_1,\boldsymbol{U}_2,\ldots,\boldsymbol{U}_{i-1}][\boldsymbol{U}_1,\boldsymbol{U}_2,\ldots,\boldsymbol{U}_{i-1}]^\dagger)\boldsymbol{U}_i,\ i=2,\ldots,S$.

*Proof.* We prove this lemma by mathematical induction.

For $S=2$, it can be shown [1)] that

$$\boldsymbol{U}^\dagger=[\boldsymbol{U}_1,\boldsymbol{U}_2]^\dagger=\begin{bmatrix}\boldsymbol{U}_1^\dagger-\boldsymbol{U}_1^\dagger\boldsymbol{U}_2\boldsymbol{C}_2^\dagger\\ \boldsymbol{C}_2^\dagger\end{bmatrix}.\tag{B2}$$

Similarly, for $S=3$, we have

$$\boldsymbol{U}^\dagger=[\boldsymbol{U}_1,\boldsymbol{U}_2,\boldsymbol{U}_3]^\dagger=\begin{bmatrix}[\boldsymbol{U}_1,\boldsymbol{U}_2]^\dagger-[\boldsymbol{U}_1,\boldsymbol{U}_2]^\dagger\boldsymbol{U}_3\boldsymbol{C}_3^\dagger\\ \boldsymbol{C}_3^\dagger\end{bmatrix}.\tag{B3}$$

1) Cline R E. Representations for the generalized inverse of a partitioned matrix. J Soc Indust Appl Math, 1964, 12: 588–600

Substituting (B2) into (B3) results in

$$\boldsymbol{U}^{\dagger}=\begin{bmatrix}\boldsymbol{U}_1^{\dagger}-\boldsymbol{U}_1^{\dagger}\boldsymbol{U}_2\boldsymbol{C}_2^{\dagger}+\boldsymbol{U}_1^{\dagger}\boldsymbol{U}_2\boldsymbol{C}_2^{\dagger}\boldsymbol{U}_3\boldsymbol{C}_3^{\dagger}\\ \boldsymbol{C}_2^{\dagger}-\boldsymbol{C}_2^{\dagger}\boldsymbol{U}_3\boldsymbol{C}_3^{\dagger}\\ \boldsymbol{C}_3^{\dagger}\end{bmatrix}. \tag{B4}$$

Then Eq. (B1) is true when $S=3$.

Assume that Eq. (B1) is true when $S=Q$. Then for $S=Q+1$, by (B2), we can get

$$\boldsymbol{U}^{\dagger}=\begin{bmatrix}[\boldsymbol{U}_1,\boldsymbol{U}_2,\ldots,\boldsymbol{U}_Q]^{\dagger}-[\boldsymbol{U}_1,\boldsymbol{U}_2,\ldots,\boldsymbol{U}_Q]^{\dagger}\boldsymbol{U}_{Q+1}\boldsymbol{C}_{Q+1}^{\dagger}\\ \boldsymbol{C}_{Q+1}^{\dagger}\end{bmatrix}. \tag{B5}$$

Substituting (B1) with $S=Q$ into (B5), we can prove that Eq. (B1) is true for $S=Q+1$. Lemma 2 is proved.

**Proof of Theorem 3**: We fist prove the bounds (28) for $l=2,3,\ldots,L-1$. With the least-squares estimate (27), the estimate error is given by

$$\Delta\tilde{\boldsymbol{\sigma}}_{\Gamma^{(l)}}^{(l)}=\tilde{\boldsymbol{\sigma}}_{\Gamma^{(l)}}^{(l)}-\hat{\tilde{\boldsymbol{\sigma}}}_{\Gamma^{(l)}}^{(l)}=-\left(\tilde{\boldsymbol{A}}_{\Gamma^{(l)}}^{(l)}\right)^{\dagger}\tilde{\underline{\boldsymbol{n}}}^{(l)}=-\left(\tilde{\boldsymbol{A}}_{\Gamma^{(l)}}^{(l)}\right)^{\dagger}\left(\underline{\tilde{\boldsymbol{A}}}_{\Gamma_1^{(l-1)}}^{(l-1)}\Delta\tilde{\boldsymbol{\sigma}}_{\Gamma_1^{(l-1)}}^{(l-1)}+\underset{\sim}{\tilde{\boldsymbol{A}}}_{\Gamma_S^{(l+1)}}^{(l+1)}\tilde{\boldsymbol{\sigma}}_{\Gamma_S^{(l+1)}}^{(l+1)}\right). \tag{B6}$$

Note that the matrix $\tilde{\boldsymbol{A}}_{\Gamma^{(l)}}^{(l)}$ is of full column rank and can be partitioned as $\tilde{\boldsymbol{A}}_{\Gamma^{(l)}}^{(l)}=[\tilde{\boldsymbol{A}}_{\Gamma_1^{(l)}}^{(l)},\tilde{\boldsymbol{A}}_{\Gamma_2^{(l)}}^{(l)},\ldots,\tilde{\boldsymbol{A}}_{\Gamma_S^{(l)}}^{(l)}]$, and for $S\geqslant 3$, $\tilde{\boldsymbol{A}}_{\Gamma_s^{(l)}}^{(l)}(s=1,2,\ldots,S-2)$ is orthogonal to $\tilde{\boldsymbol{A}}_{\Gamma_{s+2}^{(l)}}^{(l)}(s=1,2,\ldots,S-2)$. Then with Lemma 2, we have

$$\Delta\tilde{\boldsymbol{\sigma}}_{\Gamma_s^{(l)}}^{(l)}=\tilde{\boldsymbol{\sigma}}_{\Gamma_s^{(l)}}^{(l)}-\hat{\tilde{\boldsymbol{\sigma}}}_{\Gamma_s^{(l)}}^{(l)}=-\left(\tilde{\boldsymbol{P}}_{\Gamma_s^{(l)}}^{(l)}\Delta\tilde{\boldsymbol{\sigma}}_{\Gamma_1^{(l-1)}}^{(l-1)}+\tilde{\boldsymbol{Q}}_{\Gamma_s^{(l)}}^{(l)}\tilde{\boldsymbol{\sigma}}_{\Gamma_S^{(l+1)}}^{(l+1)}\right),\ s=1,\ldots,S, \tag{B7}$$

where $\tilde{\boldsymbol{P}}_{\Gamma_s^{(l)}}^{(l)}$ and $\tilde{\boldsymbol{Q}}_{\Gamma_s^{(l)}}^{(l)}$ are defined by

$$\tilde{\boldsymbol{P}}_{\Gamma_s^{(l)}}^{(l)}=\begin{cases}\left(\tilde{\boldsymbol{A}}_{\Gamma_1^{(l)}}^{(l)}\right)^{\dagger}\underline{\tilde{\boldsymbol{A}}}_{\Gamma_1^{(l-1)}}^{(l-1)}+\displaystyle\sum_{j=2}^{S}\left((-1)^{j-1}\left(\tilde{\boldsymbol{A}}_{\Gamma_1^{(l)}}^{(l)}\right)^{\dagger}\tilde{\boldsymbol{A}}_{\Gamma_2^{(l)}}^{(l)}\left(\prod_{i=2}^{j-1}\left(\tilde{\boldsymbol{C}}_{\Gamma_i^{(l)}}^{(l)}\right)^{\dagger}\tilde{\boldsymbol{A}}_{\Gamma_{i+1}^{(l)}}^{(l)}\right)\left(\tilde{\boldsymbol{C}}_{\Gamma_j^{(l)}}^{(l)}\right)^{\dagger}\underline{\tilde{\boldsymbol{A}}}_{\Gamma_1^{(l-1)}}^{(l-1)}\right),\ s=1,\\ \left(\tilde{\boldsymbol{C}}_{\Gamma_s^{(l)}}^{(l)}\right)^{\dagger}\underline{\tilde{\boldsymbol{A}}}_{\Gamma_1^{(l-1)}}^{(l-1)}+\displaystyle\sum_{j=2}^{S-s+1}\left((-1)^{j-1}\left(\prod_{i=s}^{j+s-2}\left(\tilde{\boldsymbol{C}}_{\Gamma_i^{(l)}}^{(l)}\right)^{\dagger}\tilde{\boldsymbol{A}}_{\Gamma_{i+1}^{(l)}}^{(l)}\right)\left(\tilde{\boldsymbol{C}}_{\Gamma_{j+s-1}^{(l)}}^{(l)}\right)^{\dagger}\underline{\tilde{\boldsymbol{A}}}_{\Gamma_1^{(l-1)}}^{(l-1)}\right),\ s=2,\ldots,S-1,\\ \left(\tilde{\boldsymbol{C}}_{\Gamma_S^{(l)}}^{(l)}\right)^{\dagger}\underline{\tilde{\boldsymbol{A}}}_{\Gamma_1^{(l-1)}}^{(l-1)},\ s=S.\end{cases}$$

$$\tilde{\boldsymbol{Q}}_{\Gamma_s^{(l)}}^{(l)}=\begin{cases}(-1)^{S-1}\left(\tilde{\boldsymbol{A}}_{\Gamma_1^{(l)}}^{(l)}\right)^{\dagger}\tilde{\boldsymbol{A}}_{\Gamma_2^{(l)}}^{(l)}\left(\displaystyle\prod_{i=2}^{S-1}\left(\tilde{\boldsymbol{C}}_{\Gamma_i^{(l)}}^{(l)}\right)^{\dagger}\tilde{\boldsymbol{A}}_{\Gamma_{i+1}^{(l)}}^{(l)}\right)\left(\tilde{\boldsymbol{C}}_{\Gamma_S^{(l)}}^{(l)}\right)^{\dagger}\underset{\sim}{\tilde{\boldsymbol{A}}}_{\Gamma_S^{(l+1)}}^{(l+1)},\ s=1,\\ (-1)^{S-s}\left(\displaystyle\prod_{i=s}^{S-1}\left(\tilde{\boldsymbol{C}}_{\Gamma_i^{(l)}}^{(l)}\right)^{\dagger}\tilde{\boldsymbol{A}}_{\Gamma_{i+1}^{(l)}}^{(l)}\right)\left(\tilde{\boldsymbol{C}}_{\Gamma_S^{(l)}}^{(l)}\right)^{\dagger}\underset{\sim}{\tilde{\boldsymbol{A}}}_{\Gamma_S^{(l+1)}}^{(l+1)},\ s=2,\ldots,S-1,\\ \left(\tilde{\boldsymbol{C}}_{\Gamma_S^{(l)}}^{(l)}\right)^{\dagger}\underset{\sim}{\tilde{\boldsymbol{A}}}_{\Gamma_S^{(l+1)}}^{(l+1)},\ s=S.\end{cases}$$

with $\tilde{\boldsymbol{C}}_{\Gamma_i^{(l)}}^{(l)}=(\boldsymbol{I}-[\tilde{\boldsymbol{A}}_{\Gamma_1^{(l)}}^{(l)},\tilde{\boldsymbol{A}}_{\Gamma_2^{(l)}}^{(l)},\ldots,\tilde{\boldsymbol{A}}_{\Gamma_{i-1}^{(l)}}^{(l)}][\tilde{\boldsymbol{A}}_{\Gamma_1^{(l)}}^{(l)},\tilde{\boldsymbol{A}}_{\Gamma_2^{(l)}}^{(l)},\ldots,\tilde{\boldsymbol{A}}_{\Gamma_{i-1}^{(l)}}^{(l)}]^{\dagger})\tilde{\boldsymbol{A}}_{\Gamma_i^{(l)}}^{(l)}$ for $i=2,\ldots,S$. Then by (B7), we have

$$\left\|\Delta\tilde{\boldsymbol{\sigma}}_{\Gamma_s^{(l)}}^{(l)}\right\|_2\leqslant\left\|\tilde{\boldsymbol{P}}_{\Gamma_s^{(l)}}^{(l)}\right\|_2\left\|\Delta\tilde{\boldsymbol{\sigma}}_{\Gamma_1^{(l-1)}}^{(l-1)}\right\|_2+\left\|\tilde{\boldsymbol{Q}}_{\Gamma_s^{(l)}}^{(l)}\right\|_2\left\|\tilde{\boldsymbol{\sigma}}_{\Gamma_S^{(l+1)}}^{(l+1)}\right\|_2,\ s=1,\ldots,S. \tag{B8}$$

We now derive the upper bounds of $\|\tilde{\boldsymbol{P}}_{\Gamma_s^{(l)}}^{(l)}\|_2$ and $\|\tilde{\boldsymbol{Q}}_{\Gamma_s^{(l)}}^{(l)}\|_2$. From the definitions of $\tilde{\boldsymbol{P}}_{\Gamma_s^{(l)}}^{(l)}$ and $\tilde{\boldsymbol{Q}}_{\Gamma_s^{(l)}}^{(l)}$, we can find that $\|\tilde{\boldsymbol{P}}_{\Gamma_s^{(l)}}^{(l)}\|_2$ and $\|\tilde{\boldsymbol{Q}}_{\Gamma_s^{(l)}}^{(l)}\|_2$ are related to terms $\|(\tilde{\boldsymbol{A}}_{\Gamma_1^{(l)}}^{(l)})^{\dagger}\tilde{\boldsymbol{A}}_{\Gamma_2^{(l)}}^{(l)}\|_2$, $\|(\tilde{\boldsymbol{A}}_{\Gamma_1^{(l)}}^{(l)})^{\dagger}\underline{\tilde{\boldsymbol{A}}}_{\Gamma_1^{(l-1)}}^{(l-1)}\|_2$, $\|(\tilde{\boldsymbol{C}}_{\Gamma_i^{(l)}}^{(l)})^{\dagger}\underline{\tilde{\boldsymbol{A}}}_{\Gamma_1^{(l-1)}}^{(l-1)}\|_2(i=2,\ldots,S)$, $\|(\tilde{\boldsymbol{C}}_{\Gamma_i^{(l)}}^{(l)})^{\dagger}\tilde{\boldsymbol{A}}_{\Gamma_{i+1}^{(l)}}^{(l)}\|_2(i=2,\ldots,S-1)$ and $\|(\tilde{\boldsymbol{C}}_{\Gamma_S^{(l)}}^{(l)})^{\dagger}\underset{\sim}{\tilde{\boldsymbol{A}}}_{\Gamma_S^{(l+1)}}^{(l+1)}\|_2$. Now we discuss these terms separately.

Note that $(\tilde{\boldsymbol{A}}_{\Gamma_1^{(l)}}^{(l)})^{\dagger}\tilde{\boldsymbol{A}}_{\Gamma_2^{(l)}}^{(l)}=((\tilde{\boldsymbol{A}}_{\Gamma_1^{(l)}}^{(l)})^{\mathrm{H}}\tilde{\boldsymbol{A}}_{\Gamma_1^{(l)}}^{(l)})^{-1}(\tilde{\boldsymbol{A}}_{\Gamma_1^{(l)}}^{(l)})^{\mathrm{H}}\tilde{\boldsymbol{A}}_{\Gamma_2^{(l)}}^{(l)}$. Then for term $\|(\tilde{\boldsymbol{A}}_{\Gamma_1^{(l)}}^{(l)})^{\dagger}\tilde{\boldsymbol{A}}_{\Gamma_2^{(l)}}^{(l)}\|_2$, with Properties (ii)–(iv) of Lemma 1, we have

$$\left\|\left(\tilde{\boldsymbol{A}}_{\Gamma_1^{(l)}}^{(l)}\right)^{\dagger}\tilde{\boldsymbol{A}}_{\Gamma_2^{(l)}}^{(l)}\right\|_2\leqslant\left\|\left(\left(\tilde{\boldsymbol{A}}_{\Gamma_1^{(l)}}^{(l)}\right)^{\mathrm{H}}\tilde{\boldsymbol{A}}_{\Gamma_1^{(l)}}^{(l)}\right)^{-1}\right\|_2\left\|\left(\tilde{\boldsymbol{A}}_{\Gamma_1^{(l)}}^{(l)}\right)^{\mathrm{H}}\tilde{\boldsymbol{A}}_{\Gamma_2^{(l)}}^{(l)}\right\|_2\leqslant\frac{1}{1-\tilde{\delta}_{|\Gamma^{(l)}|}^{(l)}}\tilde{\delta}_{|\Gamma^{(l)}|}^{(l)}\leqslant\frac{\bar{\delta}_{\bar{K}^{(l)}}^{(l)}}{1-\bar{\delta}_{\bar{K}^{(l)}}^{(l)}}. \tag{B9}$$

Similarly, for term $\|(\tilde{\boldsymbol{A}}_{\Gamma_1^{(l)}}^{(l)})^{\dagger}\underline{\tilde{\boldsymbol{A}}}_{\Gamma_1^{(l-1)}}^{(l-1)}\|_2$, we have

$$\left\|\left(\tilde{\boldsymbol{A}}_{\Gamma_1^{(l)}}^{(l)}\right)^{\dagger}\underline{\tilde{\boldsymbol{A}}}_{\Gamma_1^{(l-1)}}^{(l-1)}\right\|_2\leqslant\left\|\left(\left(\tilde{\boldsymbol{A}}_{\Gamma_2^{(l-1)}}^{(l-1)}\right)^{\mathrm{H}}\tilde{\boldsymbol{A}}_{\Gamma_2^{(l-1)}}^{(l-1)}\right)^{-1}\right\|_2\left\|\left(\tilde{\boldsymbol{A}}_{\Gamma_2^{(l-1)}}^{(l-1)}\right)^{\mathrm{H}}\tilde{\boldsymbol{A}}_{\Gamma_1^{(l-1)}}^{(l-1)}\right\|_2\leqslant\frac{\tilde{\delta}_{|\Gamma^{(l-1)}|}^{(l-1)}}{1-\tilde{\delta}_{|\Gamma^{(l-1)}|}^{(l-1)}}\leqslant\frac{\bar{\delta}_{\bar{K}^{(l)}}^{(l)}}{1-\bar{\delta}_{\bar{K}^{(l)}}^{(l)}}. \tag{B10}$$

For term $\|(\tilde{\boldsymbol{C}}^{(l)}_{\Gamma_i^{(l)}})^{\dagger}\tilde{\boldsymbol{A}}^{(l-1)}_{\Gamma_1^{(l-1)}}\|_2 (i=2,\ldots,S)$, we first derive the bound for $i=2$. Note that $\tilde{\boldsymbol{A}}^{(l-1)}$ is orthogonal to $\tilde{\boldsymbol{A}}^{(l)}_s (s=2,3,\ldots,S)$. Then

$$\left(\tilde{\boldsymbol{C}}^{(l)}_{\Gamma_2^{(l)}}\right)^{\dagger}\tilde{\boldsymbol{A}}^{(l-1)}_{\Gamma_1^{(l-1)}} = \left(\left(\tilde{\boldsymbol{C}}^{(l)}_{\Gamma_2^{(l)}}\right)^{\mathrm{H}}\tilde{\boldsymbol{C}}^{(l)}_{\Gamma_2^{(l)}}\right)^{-1}\left(\tilde{\boldsymbol{A}}^{(l)}_{\Gamma_2^{(l)}}\right)^{\mathrm{H}}\tilde{\boldsymbol{A}}^{(l)}_{\Gamma_1^{(l)}}\left(\tilde{\boldsymbol{A}}^{(l)}_{\Gamma_1^{(l)}}\right)^{\dagger}\tilde{\boldsymbol{A}}^{(l-1)}_{\Gamma_1^{(l-1)}}. \tag{B11}$$

Thus

$$\left\|\left(\tilde{\boldsymbol{C}}^{(l)}_{\Gamma_2^{(l)}}\right)^{\dagger}\tilde{\boldsymbol{A}}^{(l-1)}_{\Gamma_1^{(l-1)}}\right\|_2 \leqslant \left\|\left(\left(\tilde{\boldsymbol{C}}^{(l)}_{\Gamma_2^{(l)}}\right)^{\mathrm{H}}\tilde{\boldsymbol{C}}^{(l)}_{\Gamma_2^{(l)}}\right)^{-1}\right\|_2\left\|\left(\tilde{\boldsymbol{A}}^{(l)}_{\Gamma_2^{(l)}}\right)^{\mathrm{H}}\tilde{\boldsymbol{A}}^{(l)}_{\Gamma_1^{(l)}}\right\|_2\left\|\left(\tilde{\boldsymbol{A}}^{(l)}_{\Gamma_1^{(l)}}\right)^{\dagger}\tilde{\boldsymbol{A}}^{(l-1)}_{\Gamma_1^{(l-1)}}\right\|_2. \tag{B12}$$

With the Properties (iii) and (iv) of Lemma 1, we have

$$\left\|\left(\left(\tilde{\boldsymbol{C}}^{(l)}_{\Gamma_2^{(l)}}\right)^{\mathrm{H}}\tilde{\boldsymbol{C}}^{(l)}_{\Gamma_2^{(l)}}\right)^{-1}\right\|_2 \leqslant \left\|\left(\left[\tilde{\boldsymbol{A}}^{(l)}_{\Gamma_1^{(l)}},\tilde{\boldsymbol{A}}^{(l)}_{\Gamma_2^{(l)}}\right]^{\mathrm{H}}\left[\tilde{\boldsymbol{A}}^{(l)}_{\Gamma_1^{(l)}},\tilde{\boldsymbol{A}}^{(l)}_{\Gamma_2^{(l)}}\right]\right)^{-1}\right\|_2 \leqslant \frac{1}{1-\tilde{\delta}^{(l)}_{|\Gamma^{(l)}|}} \leqslant \frac{1}{1-\bar{\delta}^{(l)}_{\bar{K}^{(l)}}}. \tag{B13}$$

By Properties (ii) and (iv) of Lemma 1, we have

$$\left\|\left(\tilde{\boldsymbol{A}}^{(l)}_{\Gamma_2^{(l)}}\right)^{\mathrm{H}}\tilde{\boldsymbol{A}}^{(l)}_{\Gamma_1^{(l)}}\right\|_2 \leqslant \tilde{\delta}^{(l)}_{|\Gamma^{(l)}|} \leqslant \bar{\delta}^{(l)}_{\bar{K}^{(l)}}. \tag{B14}$$

Substituting (B13), (B14) and (B10) into (B12), we have

$$\left\|\left(\tilde{\boldsymbol{C}}^{(l)}_{\Gamma_2^{(l)}}\right)^{\dagger}\tilde{\boldsymbol{A}}^{(l-1)}_{\Gamma_1^{(l-1)}}\right\|_2 \leqslant \left(\frac{\bar{\delta}^{(l)}_{\bar{K}^{(l)}}}{1-\bar{\delta}^{(l)}_{\bar{K}^{(l)}}}\right)^2. \tag{B15}$$

For $i=3,\ldots,S$,

$$\left\|\left(\tilde{\boldsymbol{C}}^{(l)}_{\Gamma_i^{(l)}}\right)^{\dagger}\tilde{\boldsymbol{A}}^{(l-1)}_{\Gamma_1^{(l-1)}}\right\|_2 \leqslant \left\|\left(\left(\tilde{\boldsymbol{C}}^{(l)}_{\Gamma_i^{(l)}}\right)^{\mathrm{H}}\tilde{\boldsymbol{C}}^{(l)}_{\Gamma_i^{(l)}}\right)^{-1}\right\|_2\left\|\left(\tilde{\boldsymbol{A}}^{(l)}_{\Gamma_i^{(l)}}\right)^{\mathrm{H}}\tilde{\boldsymbol{A}}^{(l)}_{\Gamma_{i-1}^{(l)}}\right\|_2\left\|\left(\tilde{\boldsymbol{C}}^{(l)}_{\Gamma_{i-1}^{(l)}}\right)^{\dagger}\tilde{\boldsymbol{A}}^{(l-1)}_{\Gamma_1^{(l-1)}}\right\|_2. \tag{B16}$$

Similar to (B13) and (B14), we have

$$\left\|\left(\left(\tilde{\boldsymbol{C}}^{(l)}_{\Gamma_i^{(l)}}\right)^{\mathrm{H}}\tilde{\boldsymbol{C}}^{(l)}_{\Gamma_i^{(l)}}\right)^{-1}\right\|_2 \leqslant \left\|\left(\left[\tilde{\boldsymbol{A}}^{(l)}_{\Gamma_1^{(l)}},\tilde{\boldsymbol{A}}^{(l)}_{\Gamma_2^{(l)}},\ldots,\tilde{\boldsymbol{A}}^{(l)}_{\Gamma_i^{(l)}}\right]^{\mathrm{H}}\left[\tilde{\boldsymbol{A}}^{(l)}_{\Gamma_1^{(l)}},\tilde{\boldsymbol{A}}^{(l)}_{\Gamma_2^{(l)}},\ldots,\tilde{\boldsymbol{A}}^{(l)}_{\Gamma_i^{(l)}}\right]\right)^{-1}\right\|_2 \leqslant \frac{1}{1-\bar{\delta}^{(l)}_{\bar{K}^{(l)}}}, \tag{B17}$$

$$\left\|\left(\tilde{\boldsymbol{A}}^{(l)}_{\Gamma_i^{(l)}}\right)^{\mathrm{H}}\tilde{\boldsymbol{A}}^{(l)}_{\Gamma_{i-1}^{(l)}}\right\|_2 \leqslant \bar{\delta}^{(l)}_{\bar{K}^{(l)}}. \tag{B18}$$

Substituting (B17) and (B18) into (B16), we have

$$\left\|\left(\tilde{\boldsymbol{C}}^{(l)}_{\Gamma_i^{(l)}}\right)^{\dagger}\tilde{\boldsymbol{A}}^{(l-1)}_{\Gamma_1^{(l-1)}}\right\|_2 \leqslant \frac{\bar{\delta}^{(l)}_{\bar{K}^{(l)}}}{1-\bar{\delta}^{(l)}_{\bar{K}^{(l)}}}\left\|\left(\tilde{\boldsymbol{C}}^{(l)}_{\Gamma_{i-1}^{(l)}}\right)^{\dagger}\tilde{\boldsymbol{A}}^{(l-1)}_{\Gamma_1^{(l-1)}}\right\|_2 \leqslant \left(\frac{\bar{\delta}^{(l)}_{\bar{K}^{(l)}}}{1-\bar{\delta}^{(l)}_{\bar{K}^{(l)}}}\right)^{i-2}\left\|\left(\tilde{\boldsymbol{C}}^{(l)}_{\Gamma_2^{(l)}}\right)^{\dagger}\tilde{\boldsymbol{A}}^{(l-1)}_{\Gamma_1^{(l-1)}}\right\|_2, i=3,\ldots,S. \tag{B19}$$

Substituting (B15) into (B19), we have

$$\left\|\left(\tilde{\boldsymbol{C}}^{(l)}_{\Gamma_i^{(l)}}\right)^{\dagger}\tilde{\boldsymbol{A}}^{(l-1)}_{\Gamma_1^{(l-1)}}\right\|_2 \leqslant \left(\frac{\bar{\delta}^{(l)}_{\bar{K}^{(l)}}}{1-\bar{\delta}^{(l)}_{\bar{K}^{(l)}}}\right)^i, i=3,\ldots,S. \tag{B20}$$

For term $\|(\tilde{\boldsymbol{C}}^{(l)}_{\Gamma_i^{(l)}})^{\dagger}\tilde{\boldsymbol{A}}^{(l)}_{\Gamma_{i+1}^{(l)}}\|_2 (i=2,\ldots,S-1)$ and $\|(\tilde{\boldsymbol{C}}^{(l)}_{\Gamma_S^{(l)}})^{\dagger}\tilde{\boldsymbol{A}}^{(l+1)}_{\Gamma_S^{(l+1)}}\|_2$, similar to the derivation of (B20), we have

$$\left\|\left(\tilde{\boldsymbol{C}}^{(l)}_{\Gamma_i^{(l)}}\right)^{\dagger}\tilde{\boldsymbol{A}}^{(l)}_{\Gamma_{i+1}^{(l)}}\right\|_2 \leqslant \frac{\bar{\delta}^{(l)}_{\bar{K}^{(l)}}}{1-\bar{\delta}^{(l)}_{\bar{K}^{(l)}}}, i=2,\ldots,S-1, \tag{B21}$$

$$\left\|\left(\tilde{\boldsymbol{C}}^{(l)}_{\Gamma_S^{(l)}}\right)^{\dagger}\tilde{\boldsymbol{A}}^{(l+1)}_{\Gamma_S^{(l+1)}}\right\|_2 \leqslant \frac{\bar{\delta}^{(l)}_{\bar{K}^{(l)}}}{1-\bar{\delta}^{(l)}_{\bar{K}^{(l)}}}. \tag{B22}$$

With (B9), (B10) and (B20)–(B22), we have $\|\tilde{\boldsymbol{P}}^{(l)}_{\Gamma_s^{(l)}}\|_2 \leqslant \alpha^s(1-\alpha^{2(S-s+1)})/(1-\alpha^2)$ and $\|\tilde{\boldsymbol{Q}}^{(l)}_{\Gamma_s^{(l)}}\|_2 \leqslant \alpha^{S-s+1}$, $s=1,\ldots,S$, where $\alpha=\bar{\delta}^{(l)}_{\bar{K}^{(l)}}/(1-\bar{\delta}^{(l)}_{\bar{K}^{(l)}})$. Then by (B8), we have

$$\left\|\Delta\tilde{\boldsymbol{\sigma}}^{(l)}_{\Gamma_s^{(l)}}\right\|_2 \leqslant \beta\left\|\Delta\tilde{\boldsymbol{\sigma}}^{(l-1)}_{\Gamma_1^{(l-1)}}\right\|_2 + \alpha^{S-s+1}\left\|\tilde{\boldsymbol{\sigma}}^{(l+1)}_{\Gamma_S^{(l+1)}}\right\|_2, s=1,\ldots,S, \tag{B23}$$

where $\beta=\alpha^s\left(1-\alpha^{2(S-s+1)}\right)/\left(1-\alpha^2\right)$, $\alpha=\bar{\delta}^{(l)}_{\bar{K}^{(l)}}/(1-\bar{\delta}^{(l)}_{\bar{K}^{(l)}})$.

When $l=1$, $\Delta\tilde{\boldsymbol{\sigma}}^{(l)}_{\Gamma_s^{(l)}} = -\tilde{\boldsymbol{Q}}^{(l)}_{\Gamma_s^{(l)}}\tilde{\boldsymbol{\sigma}}^{(l+1)}_{\Gamma_S^{(l+1)}}$ and then $\|\Delta\tilde{\boldsymbol{\sigma}}^{(l)}_{\Gamma_s^{(l)}}\|_2 \leqslant \alpha^{S-s+1}\|\tilde{\boldsymbol{\sigma}}^{(l+1)}_{\Gamma_S^{(l+1)}}\|_2$ for $s=1,\ldots,S$. Similarly, when $l=L$, $\|\Delta\tilde{\boldsymbol{\sigma}}^{(l)}_{\Gamma_s^{(l)}}\|_2 \leqslant \beta\|\Delta\tilde{\boldsymbol{\sigma}}^{(l-1)}_{\Gamma_1^{(l-1)}}\|_2$ for $s=1,\ldots,S$. Then Theorem 3 is proved.