Measurement Matrix Design for Sample-Efficient Binary Compressed Sensing

Pulak Sarangi, *Graduate Student Member, IEEE*, and Piya Pal

***Abstract*—This letter investigates the problem of recovering a binary-valued signal from compressed measurements of its convolution with a known finite impulse response filter. We show that it is possible to attain optimum sample complexity for exact recovery (in absence of noise) with a computationally efficient algorithm. We achieve this by adopting an algorithm-measurement co-design strategy where the measurement matrix is designed as a function of the filter, such that the recovery of binary signals with arbitrary sparsity is possible by using a sequential decoding algorithm. Such a filter-dependent sampler design can overcome the computational challenges associated with enforcing binary constraints, and enable us to operate in "extreme compression" regimes, where the number of measurements can be much smaller than the sparsity level.**

***Index Terms*—Binary signals, compressed sensing, extreme compression, measurement matrix design, sequential decoding.**

I. Introduction

THE objective of binary compressed sensing is to recover a binary-valued signal from compressed linear measurements [1]–[9]. In this letter, we focus on a special class of the binary compressed sensing problem, where we observe compressive measurements of the convolution of a binary signal with a known finite impulse response (FIR) filter. Recovering binary signals from such compressive convolutional measurements is of interest to several applications such as neural spike detection from fluorescence measurements [10], medical imaging [11], binary shape recovery from blurred images [12], [13], image segmentation [14], and discrete tomography [15], [16]. A concrete application is in millimeter-wave communication, where the goal is to decode binary (or finite alphabet) symbols from low-dimensional measurements obtained by a compressive spatial filtering/beamforming.

As discussed in [17], [18], there exist measurement matrices such that the linear mapping between the unknown binary vectors and the real valued compressive measurement is injective even with a single (scalar) measurement. In this case, the desired binary vector can be recovered via exhaustive search, however, it is computationally prohibitive to do so. Therefore, a major focus of binary compressed sensing has been on algorithmic developments (often via relaxations) that are computationally efficient, and at the same time, exploit additional structure such as sparsity of the binary signal [1]–[3], [5], [19], [20]. A common approach is to relax the (non-convex) binary constraints with box-constraints, and formulate various continuous valued optimization problems for recovering the binary vectors. These include l_1/l_∞-norm minimization [1], [2], [21], and semidefinite relaxation [5]. Theoretical guarantees for l_1/l_∞ minimization have been established in [1], [2], [21], but the results are mostly applicable to random measurement matrices (drawn from suitable centered distributions). Recently, the benefit of using a biased measurement matrix was established in [22], which allows recovery of binary signals simply by solving a least squares problem with box-constraints, and thereby eliminating the need for l_1/l_∞ minimization. Alternative lines of work that modify classical sparse recovery algorithms to exploit finite-valued structure include greedy orthogonal matching pursuit (OMP) algorithm [20], [23], Bayesian formulations [19], [24], graph-based decoding techniques [3], and iterative reweighting techniques [25]. A common feature of all the aforementioned approaches is that their theoretical guarantees (whenever they exist) are applicable when the number of measurements (M) is larger than the sparsity (s), similar to standard results in compressed sensing. To the best of our knowledge, exact recovery guarantees for these techniques are unavailable when $M < s < N/2$ (where N is the signal dimension).

In a recent work [26], we moved away from relaxation-based techniques and showed that it is possible to exactly recover binary signals from *uniformly downsampled* measurements of the filter output, without imposing any sparsity constraints. Specifically, we developed a new computationally efficient decoding algorithm that was inspired by successive cancellation (SC) or decision feedback decoding [24], [27] used in multiuser detection, and decoding of polar codes [28], [29]. We showed that $M > N/L$ (L being the filter length) measurements are necessary for exact recovery of any binary vector from *uniformly downsampled* convolutional measurements, and the algorithm was able to attain this under a certain decay condition on the filter.

Our contributions: We establish that by appropriately designing the measurement matrix (beyond uniform downsamplers), it is possible to achieve a sample complexity of $M \geq 1$ for the exact recovery of binary signals.[1] We achieve this by (i) developing a modified version of the sequential decoding algorithm from [26], and (ii) proposing compressive measurement design techniques that are *dependent on the filter*. This algorithm-measurement co-design strategy achieves the optimal

Manuscript received January 24, 2022; revised April 13, 2022; accepted May 14, 2022. Date of publication May 31, 2022; date of current version June 10, 2022. This work was supported in part under Grants ONR N00014-19-1-2256, DE-SC0022165, NSF 2124929 and NSF CAREER ECCS 1700506. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Xuesong Wang. *(Corresponding author: Pulak Sarangi.)*

The authors are with the Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, CA 92093 USA (e-mail: psarangi@eng.ucsd.edu; pipal@eng.ucsd.edu).

Digital Object Identifier 10.1109/LSP.2022.3179230

[1]In [5], it was noted that $M = 1$ may be achievable provided the SDP returned a rank 1 solution. However, conditions under which the SDP solution is guaranteed to be rank one with $M = 1$ measurement, are currently unavailable.

sample complexity of $M \geq 1$ (for any sparsity level), without requiring any strong decay assumptions on the filter. The measurement matrix itself can be designed in a computationally efficient manner, by solving a linear program.

Notations: For a matrix $\mathbf{A}$, $\mathcal{N}(\mathbf{A})$ denotes its null-space, and $\mathbf{I}_N$ is the $N \times N$ identity matrix. For an integer n, define $[n] := \{1, 2, \ldots, n\}$.

II. Extreme Compression With Dense Samplers and Computationally Efficient Recovery

Consider the problem of recovering a binary valued signal $\boldsymbol{x}_0 \in \{0,1\}^N$ from compressed measurements of its convolution with a known finite impulse response filter $\boldsymbol{h} = [h_0, h_1, \ldots, h_{L-1}]^T \in \mathbb{R}^L$:

$$\boldsymbol{z} = \mathbf{\Phi}\mathbf{H}\boldsymbol{x}_0, \quad \mathbf{H} = \begin{bmatrix} h_0 & 0 & 0 & \cdots & 0 & 0 \\ h_1 & h_0 & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & h_{L-1} & h_{L-2} \\ 0 & 0 & 0 & \cdots & 0 & h_{L-1} \end{bmatrix} \tag{1}$$

Here $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_N] \in \mathbb{R}^{P \times N}$ is a Toeplitz matrix with $P = N + L - 1$, $\mathbf{\Phi} \in \mathbb{R}^{M \times P}, M < P$ is a compressive measurement matrix and $\mathbf{H}\boldsymbol{x}_0$ is the output of the filter.

It has been shown that a linear map $\mathbf{A} : \{0,1\}^N \to \mathbb{R}^M$ can be injective (over $\{0,1\}^N$) even when $M = 1$ [17], [18].[2] The linear map of interest to us has a specific structure $\mathbf{A} = \mathbf{\Phi}\mathbf{H}$. We begin by showing that for any filter $\boldsymbol{h}$, there exist infinite choices of real-valued sensing matrices $\mathbf{\Phi} \in \mathbb{R}^{M \times P}$ such that the map $\mathbf{A}$ is injective for every $M \geq 1$.

Theorem 1: Assume rank$(\mathbf{H}) = N$. Let $\mathbf{\Phi} \in \mathbb{R}^{M \times P}$ be a random matrix whose rows $\{\boldsymbol{\phi}_m\}_{m=1}^M$ are drawn independently from a distribution which is absolutely continuous with respect to the Lebesgue measure over $\mathbb{R}^P$. With probability 1, $\boldsymbol{x}_0$ is the unique binary vector that satisfies $\boldsymbol{z} = \mathbf{\Phi}\mathbf{H}\mathbf{x}$ for every $M \geq 1$, where $\boldsymbol{z}$ is given by (1).

Proof: Suppose there exist $\boldsymbol{x}, \boldsymbol{y} \in \{0,1\}^N (\boldsymbol{x} \neq \boldsymbol{y})$ such that $\mathbf{\Phi}\mathbf{H}\boldsymbol{x} = \mathbf{\Phi}\mathbf{H}\boldsymbol{y} \Rightarrow \mathbf{\Phi}\mathbf{H}(\boldsymbol{x} - \boldsymbol{y}) = \mathbf{0}$. This means that there is a non-zero ternary vector $\boldsymbol{x} - \boldsymbol{y} \in \mathbb{S}^N$, $\mathbb{S} := \{-1, 0, 1\}$, that belongs to the null space of $\mathbf{\Phi}\mathbf{H}$. We will show that this will happen with zero probability. Notice that the cardinality of $\mathbb{S}^N$ is 3^N. We denote each vector in $\mathbb{S}^N$ as $\{\boldsymbol{v}_k\}_{k=0}^{3^N-1}$, with the convention $\boldsymbol{v}_0 = \mathbf{0}$ for notational ease. Let $\mathcal{E} = \{\mathbf{\Phi} \mid \mathcal{N}(\mathbf{\Phi}\mathbf{H}) \bigcap \mathbb{S}^N \neq \{\mathbf{0}\}\}$. Then,

$$\begin{aligned} \mathbb{P}\left(\mathbf{\Phi} \in \mathcal{E}\right) &= \mathbb{P}\left(\exists\, \boldsymbol{v} \in \mathbb{S}^N \backslash \{\mathbf{0}\}, \text{ s.t. } \mathbf{\Phi}\mathbf{H}\boldsymbol{v} = \mathbf{0}\right) \\ &= \mathbb{P}\left(\bigcup_{k=1}^{3^N-1} \{\mathbf{\Phi}\mathbf{H}\boldsymbol{v}_k = \mathbf{0}\}\right) \\ &\underset{(a)}{=} \mathbb{P}\left(\bigcup_{k=1}^{3^N-1} \bigcap_{i=1}^{M} \{\boldsymbol{\phi}_i \in \mathcal{N}(\boldsymbol{v}_k^T \mathbf{H}^T)\}\right) \\ &\underset{(b)}{\leq} \sum_{k=1}^{3^N-1} \prod_{i=1}^{M} \mathbb{P}\left(\boldsymbol{\phi}_i \in \mathcal{N}(\boldsymbol{v}_k^T \mathbf{H}^T)\right) \end{aligned} \tag{2}$$

where (a) follows from the fact that $\mathbf{\Phi}\mathbf{H}\boldsymbol{v}_k = \mathbf{0}$ if and only if $\boldsymbol{\phi}_i \in \mathcal{N}(\boldsymbol{v}_k^T \mathbf{H}^T)$ for all $i \in [M]$. The inequality (b) follows from union bound, and the independence assumption on the rows of $\mathbf{\Phi}$. Rank$(\mathbf{H}) = N$ implies that $\boldsymbol{v}_k^T \mathbf{H}^T \neq \mathbf{0}$ for non-zero $\boldsymbol{v}_k$, and therefore $\mathcal{N}(\boldsymbol{v}_k^T \mathbf{H}^T)$ is a $P - 1$ dimensional subspace of $\mathbb{R}^P$ whose Lebesgue measure is zero. Since $\boldsymbol{\phi}_i$ is generated from a distribution that is absolutely continuous with respect to the Lebesgue measure over $\mathbb{R}^P$, we have $\mathbb{P}(\boldsymbol{\phi}_i \in \mathcal{N}(\boldsymbol{v}_k^T \mathbf{H}^T)) = 0$. Using (2), we can conclude that $\mathbb{P}(\mathbf{\Phi} \in \mathcal{E}) = 0$. ∎

Theorem 1 suggests that for almost all choices of $\mathbf{\Phi}$, one can uniquely identify a binary $\boldsymbol{x}_0$ from $\boldsymbol{z}$, with only $M = \Omega(1)$ measurements (independent of the sparsity-level), possibly via exhaustive search. Relaxation-based techniques succeed in a regime where M is larger than the sparsity of $\boldsymbol{x}_0$, and exact recovery may not be possible with $M = \Omega(1)$ measurements. We now present a simple and computationally efficient algorithm that sequentially decodes the binary entries of $\boldsymbol{x}_0$. We further show that by using the idea of *filter-dependent* sampler design, it is possible to achieve $M = \Omega(1)$ with this algorithm.

Algorithm 1: Sequential Block-wise Decoding Algorithm.

Input: Measurement $\boldsymbol{z}$, Sensing matrix $\mathbf{\Phi} \in \mathbb{R}^{M \times P}$, Filter $\mathbf{H}$, Tolerance $\epsilon \geq 0$.
Output: $\hat{\boldsymbol{x}} \in \{0,1\}^N$ //Estimate of $\boldsymbol{x}_0$
$\mathbf{A} \leftarrow \mathbf{\Phi}\mathbf{H}, b \leftarrow \lceil N/M \rceil, m \leftarrow 1, \hat{\boldsymbol{x}} \leftarrow \mathbf{0}$ //Initialization
Repeat
  $l \leftarrow 1, r \leftarrow z_m$//Reset Residual
  Repeat
    if $r/A_{m,(m-1)b+l} \geq 1 - \epsilon$ //Detection threshold
      $\hat{x}_{(m-1)b+l} \leftarrow 1$
    end
    $r \leftarrow r - \hat{x}_{(m-1)b+l} A_{m,(m-1)b+l}, l \leftarrow l + 1$ //Update residual
  until $l \leq b$ **or** $(m-1)b + l \leq N$
  $m \leftarrow m + 1$
until $m \leq M$

A. Sequential Block-Wise Decoding and Performance Guarantees

The proposed Sequential Block-wise Decoding Algorithm is summarized in Table 1. The main idea is to partition the entries of $\boldsymbol{x}_0$ into $b = \lceil N/M \rceil$ disjoint blocks, one corresponding to each scalar measurement z_m, and decode the entries of a block sequentially. For the m^{th} block, suppose that the first $k < b$ indices within the block, denoted by the set $\mathcal{J}_{m,k} = \{(m-1)b + i\}_{i=1}^k$ have already been decoded. The sequential decoding algorithm computes a residual $r = z_m - \sum_{i \in \mathcal{J}_{m,k}} A_{m,i}\hat{x}_i$, and compares it against a suitable threshold determined by $\epsilon (\geq 0)$, in order to estimate the $(k+1)^{\text{th}}$ element. The estimate is given as $\hat{x}_{(m-1)b+k+1} = \mathbb{1}_{\{r/A_{m,(m-1)b+k+1} \geq (1-\epsilon)\}}$ where $\mathbb{1}_{\{x \geq \gamma\}} : \mathbb{R} \to \{0,1\}$ denotes an indicator function defined as $\mathbb{1}_{\{x \geq \gamma\}} = \begin{cases} 1, & \text{if } x \geq \gamma \\ 0, & \text{else} \end{cases}$. It is important to note that the residual computation subtracts only the elements that have been decoded within the *current block*. The previous blocks that have already been decoded, are not subtracted out. This algorithmic choice has been made to avoid error propagation between blocks. Such disjoint decoding can be especially beneficial in presence of

[2] In [18], it is shown that $\mathbf{A}$ can be linearly dependent over $\mathbb{R}$, but linearly independent over $\{0,1\}$, and [17] shows the existence of such a rational $\mathbf{A}$.

noise. Decoding each block requires $O(b)$ operations, resulting in a total computational complexity of $O(Mb) = O(N)$ for decoding all M blocks. Given integers $m \in [M]$, and $l \in [b]$ we define $\eta(m,l) := (m-1)b + l$. For ease of exposition, we assume that N/M is an integer. The following theorem specifies sufficient conditions on $\mathbf{A} = \mathbf{\Phi H}$ under which Algorithm 1 exactly recovers $\boldsymbol{x}_0$ from noiseless measurements (1).

Theorem 2: Let $\mathbf{A} := \mathbf{\Phi H}$ and $\epsilon = 0$. For any $M \geq 1$, Algorithm 1 recovers $\boldsymbol{x}_0$, if for each $m \in [M]$, the following holds

$$A_{m,l} > 0, \forall\, l \in [N],\ A_{m,\eta(m,l)} > \sum_{\substack{k=1\\ k \notin \mathcal{J}_{m,l}}}^{N} A_{m,k}, \forall\, l \in [b-1] \quad (3)$$

Proof: Condition (3) implies that for any binary $\boldsymbol{x} \in \{0,1\}^N$ and $l \in [b-1]$,

$$x_{\eta(m,l)} \leq \sum_{k \notin \mathcal{J}_{m,l}} x_k \frac{A_{m,k}}{A_{m,\eta(m,l)}} + x_{\eta(m,l)} < 1 + x_{\eta(m,l)} \quad (4)$$

We now show that for every m, we can decode the indices of $\boldsymbol{x}_0$ given by $\{\eta(m,l)\}_{l=1}^{b}$. Fix m. Our proof proceeds via induction on l. For $l = 1$, we have $r = z_m = \sum_{k \notin \mathcal{J}_{m,1}} x_k A_{m,k} + x_{\eta(m,1)} A_{m,\eta(m,1)}$ (Line 5 of Algorithm 1). Hence from (4) we have $x_{\eta(m,1)} \leq r/A_{m,\eta(m,1)} < x_{\eta(m,1)} + 1$. Since $\hat{x}_{\eta(m,1)} = \mathbb{1}_{\{r/A_{m,\eta(m,1)} \geq 1\}}$, it follows that $\hat{x}_{\eta(m,1)} = 1$ if $x_{\eta(m,1)} = 1$, and 0 otherwise, implying $\hat{x}_{\eta(m,1)} = x_{\eta(m,1)}$. Next assume that after $l-1 < b$ iterations we have correctly decoded $\{x_{\eta(m,k)}\}_{k=1}^{l-1}$. The residual satisfies:

$$r/A_{m,\eta(m,l)} = \left(z_m - \sum_{k \in \mathcal{J}_{m,l-1}} \hat{x}_k A_{m,k} \right) / A_{m,\eta(m,l)}$$
$$\overset{(a)}{=} \sum_{k \notin \mathcal{J}_{m,l}} x_k \frac{A_{m,k}}{A_{m,\eta(m,l)}} + x_{\eta(m,l)} \quad (5)$$

where (a) holds due to the induction hypothesis. Using a similar argument as $l = 1$, from (4), (5) we can again show that $\hat{x}_{\eta(m,l)} = \mathbb{1}_{\{r/A_{m,\eta(m,l)} \geq 1\}} = x_{\eta(m,l)}$, which concludes the proof. ■

The success of Algorithm 1 therefore depends on condition (3), which reveals the dependence of the sampler $\mathbf{\Phi}$ on the filter $\boldsymbol{h}$, and implicitly governs the sample complexity. If the entries of $\mathbf{\Phi}$ are drawn randomly, agnostic to $\mathbf{h}$, the condition (3) may not be satisfied with high probability. Therefore, it becomes essential to explicitly tune the design of sampler $\mathbf{\Phi}$ to the structure of the filter $\boldsymbol{h}$.

B. Filter-Dependent Sampler Design Via Linear Program

It can be verified that condition in (3) is satisfied if and only if for every $m \in [M]$, $\boldsymbol{\phi}_m$ belongs to the following set

$$\mathcal{F}_{\boldsymbol{h}}^{(m)} = \Bigg\{ \boldsymbol{\phi} \in \mathbb{R}^P \mid \boldsymbol{\phi}^T \mathbf{h}_i > 0, \quad i = 1, 2, \ldots, N,$$
$$\boldsymbol{\phi}^T \left(\mathbf{h}_{(m-1)b+j} - \sum_{k=1 k \notin \mathcal{J}_{m,j}}^{N} \mathbf{h}_k \right) > 0, 1 \leq j \leq b-1 \Bigg\}$$

Notice that $\mathcal{F}_{\boldsymbol{h}}^{(m)}$ is a polyhedral set, whose geometry depends on the choice of the filter $\boldsymbol{h}$. Hence the sampling operators $\boldsymbol{\phi}_m$ can be designed to satisfy (3), by solving the following linear program for every m:

$$\text{find} \quad \boldsymbol{\phi}_m \quad \text{subject to} \quad \boldsymbol{\phi}_m \in \mathcal{F}_{\boldsymbol{h}}^{(m)} \qquad \text{(LPH)}$$

For any $\boldsymbol{\phi} \in \mathcal{F}_{\boldsymbol{h}}^{(m)}$, the scaled vector $\alpha\boldsymbol{\phi}$ for any $\alpha > 0$ is also a valid solution, i.e., $\alpha\boldsymbol{\phi} \in \mathcal{F}_{\boldsymbol{h}}^{(m)}$. Therefore, the sensing matrix can always be scaled to avoid solutions close to $\mathbf{0}$, as well as meet any desired power constraint. The following lemma, whose proof is in the Appendix, ensures that $\mathcal{F}_{\boldsymbol{h}}^{(m)}$ is non-empty under mild conditions on $\boldsymbol{h}$.

Lemma 1: For any $\boldsymbol{h} \in \mathbb{R}^L$ satisfying $\text{rank}(\mathbf{H}) = N$ and $M \geq 1$, the set $\mathcal{F}_{\boldsymbol{h}}^{(m)}$ is non-empty for every $m \in [M]$.

We obtain the following exact recovery guarantee for Algorithm 1 by combining Theorem 2 and Lemma 1.

Theorem 3: Let $\mathbf{\Phi}^\star$ be a sensing matrix whose mth row is a solution to (LPH), $m \in [M]$. Consider noiseless measurements $\boldsymbol{z} \in \mathbb{R}^M$ acquired using $\mathbf{\Phi}^\star$ as $\boldsymbol{z} = \mathbf{\Phi}^* \mathbf{H} \boldsymbol{x}_0$, where $\boldsymbol{x}_0 \in \{0,1\}^N$. For every $M \geq 1$, Algorithm 1 recovers $\boldsymbol{x}_0$, regardless of its sparsity.

C. Remarks on Noise Resilience and Sampler Design

The main objective of this letter was to achieve optimum sample complexity for exact recovery of binary signals in absence of noise with a computationally efficient algorithm. In presence of noise, the threshold ϵ in Algorithm 1 should be optimized based on the noise level. Increasing M increases the number of blocks which is also important to promote noise resilience, since Algorithm 1 prevents error propagation from one block to another. Another important, but perhaps less obvious consideration is the effect of block length on the dynamic range of the sampler. By decreasing b (increasing M), measurement matrices with smaller dynamic range can be designed, which leads to better numerical stability and more reliable decoding. Determining the optimal choices of b and ϵ based on the noise level and dynamic range considerations will be of future interest. Other directions will be to design $\mathbf{\Phi}$ by using a suitable optimization criterion over the set $\mathcal{F}_{\boldsymbol{h}}^{(m)}$ (instead of a simple feasibility search), and explore adaptive filter-dependent sensing strategies.

III. Simulations

We consider binary signals of dimension $N = 100$, and FIR filters of length $L = 5$. The filter coefficients are generated independently as product of two independent random variables $h_i = s_i d_i$ where $d_i \sim \mathcal{U}[1,2]$, and s_i is a Rademacher random variable, and these coefficients are kept fixed throughout the experiments. We compare the performance of Algorithm 1 and the filter-dependent sampler design strategy (LPH) against two recent binary compressed sensing algorithms (i) SDP relaxation [5] and (ii) box-constrained least squares with biased measurement (LS-Bias) [22]. We generate noiseless measurements of the form $\boldsymbol{z} = \mathbf{A}\boldsymbol{x}_0$. For our approach, $\mathbf{A} = \mathbf{\Phi H}$ where $\mathbf{\Phi}$ is obtained by solving (LPH). For SDP relaxation, the entries of $\mathbf{A}$ are generated i.i.d as $A_{i,j} \sim \mathcal{N}(0,1)$. For LS-Bias, following [22], the entries of $\mathbf{A}$ are generated i.i.d as $A_{i,j} \sim \mathcal{N}(1,1)$, where the non-zero mean of 1 acts as the bias. In the first experiment, we study the noiseless performance of each technique as a function of sparsity s. The probability of exact recovery, i.e., number of times $\hat{\boldsymbol{x}} = \boldsymbol{x}_0$ over 100 Monte Carlo runs, is used as the performance metric. Fig. 1 shows that the proposed strategy exactly recovers $\boldsymbol{x}_0$ regardless of the sparsity level (even when

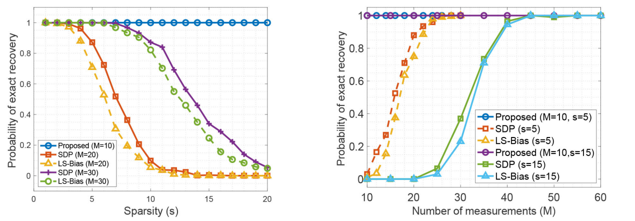

Fig. 1. Noiseless recovery performance of different methods (Left) versus s for different M, and (Right) versus M for $s = 5, 15$.

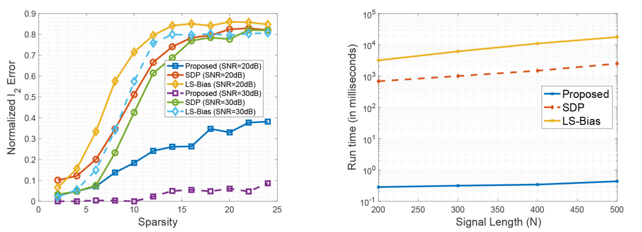

Fig. 2. (Left) Noisy Reconstruction: Normalized l_2 error vs. sparsity s, for different SNR. Here $M = 25$ and threshold is fixed at $\epsilon = 0.1$. (Right) Comparison of runtime versus N ($M = \lceil 0.2N \rceil$, $s = 5$).

$M < s$), whereas SDP relaxation and LS-Bias, the probability of exact recovery falls below 0.5 when s exceed $M/2$.

In the next experiment, we study the performance of SDP relaxation and LS-Bias as a function of M, keeping the sparsity fixed at $s = 5$ and 15. For the proposed strategy, we keep M fixed at $M = 10$ ($b = 10$). Fig. 1(b) shows SDP and LS-Bias require $M = 25$ (for $s = 5$) and $M = 45$ (for $s = 15$) to exactly recover $\boldsymbol{x}_0$ whereas the proposed strategy is able to do so with only $M = 10$ filter-dependent measurements.

Next, we evaluate the performance of the proposed strategy in presence of noise. We generate noisy measurements of the form $\mathbf{z} = \mathbf{A}\boldsymbol{x}_0 + \mathbf{n}$, where the additive noise $\mathbf{n}$ is distributed as $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma_n^2 \mathbf{I}_M)$. The signal-to-noise ratio (SNR) for each sensing strategy is defined as $10 \log_{10} \left(\frac{\|\mathbf{A}\boldsymbol{x}_0\|_2^2}{M\sigma_n^2} \right)$. To ensure a consistent SNR across different approaches, we normalize the measurement matrices (in our case, we normalize $\mathbf{\Phi}$) such that $\|\mathbf{A}\boldsymbol{x}_0\|_2$ is the same for each method, and σ_n is chosen according to the desired SNR. In Fig. 2(a), we plot the l_2 error averaged over 100 Monte Carlo runs for each sparsity level. The proposed algorithm achieves a significantly smaller error especially when the sparsity increases. We operate in the regime $s < M$ to ensure sufficient measurements for all algorithms.

Finally, in Fig. 2(b) we compare the average run-time (averaged over 10 runs) of all three algorithms as a function of N. We choose $s = 5$, and $M = \lceil 0.2N \rceil$. The run-time of Algorithm 1 is significantly smaller than the others, which were implemented using off-the-shelf convex solver (CVX) [30].

IV. Conclusion

We proposed a measurement matrix design framework for recovery of binary-valued signals from compressed convolutional measurements. The filter-dependent sensing matrix design guarantees exact recovery in absence of noise, using a computationally efficient sequential block-wise decoding algorithm. The overall strategy achieves an optimal sample complexity of $M \geq 1$. The proposed framework also paves way for several interesting future directions such as optimizing the algorithm and measurement design parameters using the knowledge of noise level, and extending the strategy to more general alphabets and different classes of filters.

Appendix
Proof of Lemma 1

Proof: We will first establish that if $\mathcal{F}_{\boldsymbol{h}}^{(1)}$ is non-empty then $\mathcal{F}_{\boldsymbol{h}}^{(m)}$ is also non-empty for every $m \in [M]$. For each m, we define a permutation matrix $\mathbf{\Pi}_m \in \mathbb{R}^{N \times N}$ as follows:

$$[\mathbf{\Pi}_m \mathbf{x}]_j = \begin{cases} x_{(m-1)b+j}, & 1 \leq j \leq b \\ x_{j-(m-1)b}, & (m-1)b+1 \leq j \leq mb \\ x_j, & \text{otherwise} \end{cases}$$

This permutation swaps the first and m^{th} block (of size b) of the vector $\mathbf{x}$. The set $\mathcal{F}_{\boldsymbol{h}}^{(m)}$ is described by $N + b - 1$ inequalities, which can be compactly represented as $\mathbf{B}\mathbf{\Pi}_m \mathbf{H}^T \boldsymbol{\phi} \succ \mathbf{0}$. Here, $\succ$ denotes element-wise inequality constraints, and $\mathbf{B} = [\mathbf{I}_N, \tilde{\mathbf{B}}^T]^T$ is a $(N + b - 1) \times N$ matrix, with $\tilde{\mathbf{B}}_{i,j} = 1$ if $i = j$, $\tilde{\mathbf{B}}_{i,j} = -1$ if $i < j$, and 0 otherwise. If $\mathcal{F}_{\boldsymbol{h}}^{(1)}$ is non-empty, then $\exists\ \boldsymbol{\phi}_1$ such that $\mathbf{B}\mathbf{H}^T \boldsymbol{\phi}_1 \succ \mathbf{0}$ (since $\mathbf{\Pi}_1 = \mathbf{I}_N$). Since $\text{rank}(\mathbf{H}) = N$, we can always find $\tilde{\boldsymbol{\phi}} \in \mathbb{R}^P$ satisfying $\mathbf{H}^T \tilde{\boldsymbol{\phi}} = \mathbf{\Pi}_m^T \mathbf{H}^T \boldsymbol{\phi}_1$. Such a vector $\tilde{\boldsymbol{\phi}}$ also satisfies $\mathbf{B}\mathbf{\Pi}_m \mathbf{H}^T \tilde{\boldsymbol{\phi}} = \mathbf{B}\mathbf{\Pi}_m \mathbf{\Pi}_m^T \mathbf{H}^T \boldsymbol{\phi}_1 \overset{(a)}{=} \mathbf{B}\mathbf{H}^T \boldsymbol{\phi}_1 \succ \mathbf{0}$, since $\mathbf{\Pi}_m$ is a permutation matrix with $\mathbf{\Pi}_m \mathbf{\Pi}_m^T = \mathbf{I}_N$. Therefore, $\tilde{\boldsymbol{\phi}} \in \mathcal{F}_{\boldsymbol{h}}^{(m)}$, whenever $\mathcal{F}_{\boldsymbol{h}}^{(1)}$ is non-empty.

We now establish that $\mathcal{F}_{\boldsymbol{h}}^{(1)}$ is indeed non-empty, i.e., $\exists\ \boldsymbol{\phi} \in \mathbb{R}^P$, such that:

$$\boldsymbol{\phi}^T \mathbf{h}_i > 0\ \forall\ i, \quad \boldsymbol{\phi}^T \left(\mathbf{h}_j - \sum_{k=j+1}^{N} \mathbf{h}_k \right) > 0, j \in [b-1] \quad (6)$$

Define $\mathbf{u}_j := \mathbf{h}_j - \sum_{k=j+1}^N \mathbf{h}_k, j \in [b-1]$. Let $\mathcal{S} = \{\mathbf{h}_1, \mathbf{h}_2 \cdots, \mathbf{h}_N, \mathbf{u}_1, \mathbf{u}_2 \cdots, \mathbf{u}_{b-1}\}$, and consider its convex hull $\mathcal{A}_{\mathbf{H}} := \text{conv}(\mathcal{S})$ which is a (closed) polyhedral set. Observe that there exists a $\boldsymbol{\phi} \in \mathbb{R}^P$ satisfying (6) if there exists a hyperplane $\boldsymbol{\phi}^T \mathbf{x} = c$ ($c > 0$), which strictly separates the point $\mathbf{0}$ from $\mathcal{A}_{\mathbf{H}}$, i.e., $\boldsymbol{\phi}^T \mathbf{x} > 0$ for all $\mathbf{x} \in \mathcal{A}_{\mathbf{H}}$. Since $\mathcal{A}_{\mathbf{H}}$ is a closed convex set, the strict hyperplane separation theorem will guarantee existence of the desired $\boldsymbol{\phi}$ provided $\mathbf{0} \notin \mathcal{A}_{\mathbf{H}}$ [31, Prop 1.5.3]. We show $\mathbf{0} \notin \mathcal{A}_{\mathbf{H}}$ by contradiction. Suppose $\mathbf{0} \in \mathcal{A}_{\mathbf{H}}$. Then $\exists\ \alpha_i, \beta_j \geq 0$ satisfying $\sum_{i=1}^N \alpha_i + \sum_{j=1}^{b-1} \beta_j = 1$, and $\sum_{i=1}^N \alpha_i \mathbf{h}_i + \sum_{j=1}^{b-1} \beta_j \mathbf{u}_j = \mathbf{0}$ which can rearranged as $(\alpha_1 + \beta_1)\mathbf{h}_1 + \sum_{i=2}^{b-1} \left(\alpha_i + \beta_i - \sum_{j=1}^{i-1} \beta_j \right) \mathbf{h}_i + \sum_{i=b}^N \left(\alpha_i - \sum_{j=1}^{b-1} \beta_j \right) \mathbf{h}_i = \mathbf{0}$. Since $\text{rank}(\mathbf{H}) = N$, we must have

$$\alpha_1 + \beta_1 = 0, \left(\alpha_i + \beta_i - \sum_{j=1}^{i-1} \beta_j \right) = 0, \quad 2 \leq i \leq b-1,$$
$$\left(\alpha_i - \sum_{j=1}^{b-1} \beta_j \right) = 0, \quad b \leq i \leq N. \quad (7)$$

Since α_i, β_i are also non-negative, it can be easily verified that (7) holds only if $\alpha_i = 0, \beta_j = 0$ for all i, j. This contradicts the fact that $\sum_{i=1}^N \alpha_i + \sum_{j=1}^{b-1} \beta_j = 1$. Hence $\mathbf{0} \notin \mathcal{A}_{\mathbf{H}}$, completing the proof. ∎

REFERENCES

[1] M. Stojnic, "Recovery thresholds for $\ell 1$ optimization in binary compressed sensing," in *Proc. IEEE Int. Symp. Inf. Theory*, 2010, pp. 1593–1597.

[2] S. Keiper, G. Kutyniok, D. G. Lee, and G. E. Pfander, "Compressed sensing for finite-valued signals," *Linear Algebra Appl.*, vol. 532, pp. 570–613, 2017.

[3] U. Nakarmi and N. Rahnavard, "Bcs: Compressive sensing for binary sparse signals," in *Proc. IEEE MILCOM IEEE Mil. Commun. Conf.*, 2012, pp. 1–5.

[4] M. Shirvanimoghaddam, Y. Li, B. Vucetic, J. Yuan, and P. Zhang, "Binary compressive sensing via analog fountain coding," *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6540–6552, Dec. 2015.

[5] S. M. Fosson and M. Abuabiah, "Recovery of binary sparse signals from compressed linear measurements via polynomial optimization," *IEEE Signal Process. Lett.*, vol. 26, no. 7, pp. 1070–1074, Jul. 2019.

[6] V. Bioglio, G. Coluccia, and E. Magli, "Sparse image recovery using compressed sensing over finite alphabets," in *Proc. IEEE Int. Conf. Image Process.*, 2014, pp. 1287–1291.

[7] J.-H. Ahn, "Compressive sensing and recovery for binary images," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4796–4802, Oct. 2016.

[8] R. Hayakawa and K. Hayashi, "Convex optimization-based signal detection for massive overloaded MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7080–7091, Nov. 2017.

[9] A. Aissa-El-Bey, D. Pastor, S. M. A. Sbai, and Y. Fadlallah, "Sparsity-based recovery of finite alphabet solutions to underdetermined linear systems," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 2008–2018, Apr. 2015.

[10] R. Brette and A. Destexhe, *Handbook of Neural Activity Measurement*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[11] T. Zhang, J. M. Pauly, and I. R. Levesque, "Accelerating parameter mapping with a locally low rank constraint," *Magn. Reson. Med.*, vol. 73, no. 2, pp. 655–661, 2015.

[12] S. Razavikia, A. Amini, and S. Daei, "Reconstruction of binary shapes from blurred images via Hankel-structured low-rank matrix recovery," *IEEE Trans. Image Process.*, vol. 29, pp. 2452–2462, 2020.

[13] S. Razavikia, H. Zamani, and A. Amini, "Sampling and recovery of binary shapes via low-rank structures," in *Proc. 13th Int. Conf. Sampling Theory Appl.*, 2019, pp. 1–4.

[14] N. Alajlan, M. S. Kamel, and G. H. Freeman, "Geometry-based image retrieval in binary image databases," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 6, pp. 1003–1013, Jun. 2008.

[15] J. Nemeth and P. Balazs, "Restoration of blurred binary images using discrete tomography," in *Proc. Int. Conf. Adv. Concepts Intell. Vis. Syst.*, 2013, pp. 80–90.

[16] K. J. Batenburg and J. Sijbers, "DART: A practical reconstruction algorithm for discrete tomography," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2542–2553, Sep. 2011.

[17] J.-H. Lange, M. E. Pfetsch, B. M. Seib, and A. M. Tillmann, "Sparse recovery with integrality constraints," *Discrete Appl. Math.*, vol. 283, pp. 346–366, 2020.

[18] L. Fukshansky, D. Needell, and B. Sudakov, "An algebraic perspective on integer sparse recovery," *Appl. Math. Computation*, vol. 340, pp. 31–42, 2019.

[19] Z. Tian, G. Leus, and V. Lottici, "Detection of sparse signals under finite-alphabet constraints," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2009, pp. 2349–2352.

[20] A. Flinth and G. Kutyniok, "Promp: A sparse recovery approach to lattice-valued signals," *Appl. Comput. Harmon. Anal.*, vol. 45, no. 3, pp. 668–708, 2018.

[21] O. L. Mangasarian and B. Recht, "Probability of unique integer solution to a system of linear equations," *Eur. J. Oper. Res.*, vol. 214, no. 1, pp. 27–30, 2011.

[22] A. Flinth and S. Keiper, "Recovery of binary sparse signals with biased measurement matrices," *IEEE Trans. Inf. Theory*, vol. 65, no. 12, pp. 8084–8094, Dec. 2019.

[23] S. Sparrer and R. F. Fischer, "Soft-feedback omp for the recovery of discrete-valued sparse signals," in *Proc. 23rd Eur. Signal Process. Conf.*, 2015, pp. 1461–1465.

[24] H. Zhu and G. B. Giannakis, "Exploiting sparse user activity in multiuser detection," *IEEE Trans. Commun.*, vol. 59, no. 2, pp. 454–465, Feb. 2011.

[25] S. M. Fosson, "Non-convex approach to binary compressed sensing," in *Proc. IEEE 52nd Asilomar Conf. Signals, Syst., Comput.*, 2018, pp. 1959–1963.

[26] P. Sarangi and P. Pal, "No relaxation: Guaranteed recovery of finite-valued signals from undersampled measurements," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2021, pp. 5440–5444.

[27] S. Verdu *et al.*, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[28] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.

[29] A. Alamdar-Yazdi and F. R. Kschischang, "A simplified successive-cancellation decoder for polar codes," *IEEE Commun. Lett.*, vol. 15, no. 12, pp. 1378–1380, Dec. 2011.

[30] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," Mar. 2014. [Online]. Available: http://cvxr.com/cvx

[31] D. Bertsekas, *Convex Optimization Theory*, vol. 1. Belmont, MA, USA: Athena Scientific, 2009.