Real-valued Group Testing for Quantitative Molecular Assays

Seyran Saeedi[1,5], Myrna Serrano[2,6], Dennis G. Yang[3], J. Paul Brooks[4], Gregory A. Buck[2,6], and Tomasz Arodz[1,6]

[1] Department of Computer Science, College of Engineering, Virginia Commonwealth University, Richmond, VA, USA
[2] Department of Microbiology and Immunology, School of Medicine, Virginia Commonwealth University, Richmond, VA, USA
[3] Department of Mathematics, College of Arts and Sciences, Drexel University, Philadelphia, PA, USA
[4] Department of Information Systems, School of Business, Virginia Commonwealth University, Richmond, VA, USA
[5] Department of Electrical and Computer Engineering, University of California, Santa Barbara, Santa Barbara, CA, USA
[6] Center for Microbiome Engineering and Data Analysis, Virginia Commonwealth University, Richmond, VA, USA

Abstract. Combinatorial group testing and compressed sensing both focus on recovering a sparse vector of dimensionality n from a much smaller number $m < n$ of measurements. In the first approach, the problem is defined over the Boolean field – the goal is to recover a Boolean vector and measurements are Boolean; in the second approach, the unknown vector and the measurements are over the reals. Here, we focus on real-valued group testing setting that more closely fits modern testing protocols relying on quantitative measurements, such as qPCR, where the goal is recovery of a sparse, Boolean vector and the pooling matrix needs to be Boolean and sparse, but the unknown input signal vector and the measurement outcomes are nonnegative reals, and the matrix algebra implied in the test protocol is over the reals. With the recent renewed interest in group testing, focus has been on quantitative measurements resulting from qPCR, but the method proposed for sample pooling were based on matrices designed with Boolean measurements in mind. Here, we investigate constructing pooling matrices dedicated for the real-valued group testing. We provide conditions for pooling matrices to guarantee unambiguous decoding of positives in this setting. We also show a deterministic algorithm for constructing matrices meeting the proposed condition, for small matrix sizes that can be implemented using a laboratory robot. Using simulated data, we show that the proposed approach leads to matrices that can be applied for higher positivity rates than combinatorial group testing matrices considered for viral testing previously. We also validate the approach through wet lab experiments involving SARS-CoV-2 nasopharyngeal swab samples.

Keywords: group testing · compressed sensing · qPCR · SARS-CoV-2 testing.

1 Introduction

Widely-available, fast-turnover molecular testing for the presence of highly contagious infectious diseases is considered a key tool in limiting their spread [36]. For newly emerging viral diseases, the gold-standard detection approach involves molecular assays based on polymerase chain reaction (PCR) [10], which can be quickly designed once the genetic sequence of the virus becomes available. However, rapid scaling of testing to cover affected communities may face obstacles, leading to interest in pooling strategies that allow for using m tests to screen many more than m samples, out of which k are positive [13,30,33,29,34]. Traditional, adaptive pooling approaches, such as Dorfman pooling [13,1] and its improved variants [30,21], combine biological material from multiple individuals into pools, each tested using one test. This allows for quickly eliminating a large fraction of virus-negative pools of samples. However, follow-up testing is required to confirm which individual samples in the pools that tested positive are positive, introducing delays and requiring protocols for storing and retrieving previously tested samples for re-testing.

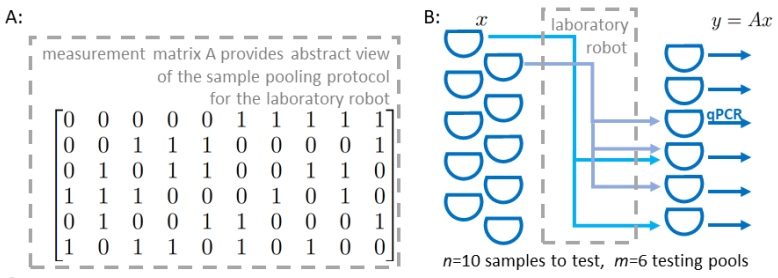
Non-adaptive, single-step protocols in which tests do not depend on each other and can be done in parallel have been studied under the umbrella of combinatorial group testing [14,2] or Boolean compressed sensing [3,27], and also in information retrieval [23]. The key challenge in non-adaptive group testing protocols is the design of a binary measurement matrix A (see Figure 1), which prescribes that sample j should be assigned to pool i if $A_{ij} = 1$. In many applications, the matrix should be sparse, and several authors considered combinatorial group testing with sparse matrices [17,18]. The matrix needs to guarantee that the identity of the positive samples can be decoded from measurements of the sample pools. Probabilistic group testing relaxes that requirement to allow the decoding guarantee to fail with some low probability [7]. These approaches all share the underlying Boolean algebra – the measurement result for each pool is binary, and only provides information whether the tested pool is all-negative or whether it contains at least one positive sample. Combinatorial quantitative group testing [37] extends this approach to measurements that provide the number of positive samples in the pool. The focus on binary or integer measurements puts a limit on how small the number of pools, m, can be for a given number of tests, n.

Contemporary molecular assays often provide more than a binary readout – for example, cycle threshold (C_t) values in qPCR can be used to provide an estimate of the quantity of the measured molecule – but the availability of this quantitative information is underutilized in pooling matrix design. Compressed sensing [6,11,4,9] approaches generalize group testing to quantitative, real-valued measurements resulting from real-valued linear algebra involving A and the unknown vector x^*, similar to how qPCR would provide an estimate of the combined abundance of the molecule in the pool. However, matrices designed for compressed sensing typically involves real-valued elements, for example sampled from a normal distribution, and thus are not feasible to implement for laboratory pooling of samples. Compressed sensing also focuses on the quality of approximating the real-valued unknown signal vector, instead of just its pattern of nonzeros that indicates which samples are positive, and on applications with high-dimensional signals, such as 3D imaging [26]. Nonnegative compressed sensing [12,5,24,25] is more aligned with molecular testing by focusing on unknown vectors that, like molecular abundance vectors, involve nonnegative reals, but is still optimizing the matrix design towards approximating the vector x^* instead of recovering its pattern of nonzeros.

Combination group testing and compressed sensing have gained renewed interest recently in the context of SARS-CoV-2 testing. It has been argued that combinatorial, nonadaptive pooling designs become advantageous compared to multi-step, adaptive pooling and to testing individual samples as the positivity rate, the fraction of samples in the tested group, increases [8]. One recent nonadaptive pooling method, P-BEST, achieves 8-fold reduction in the number of tests for groups of samples with positivity rate of around 1% or less. Another method, Tapestry [19], achieves 2.3-fold reduction for 1.9% positivity rates, extending to 10-fold reduction for 0.2% positivity rates. Both P-BEST and Tapestry utilize quantitative measurements and employ decoding techniques from compressed sensing domain, but both use matrix construction strategies designed for Boolean-measurement combinatorial group testing: P-BEST relies on Reed-Solomon codes, and Tapestry utilizes Kirkman Triple Systems resulting in 2-disjunct matrices. Neither of these approaches considered whether the matrix design can be improved if measurements are assumed to be real-valued.

1.1 Problem Statement and Contribution

We introduce real-valued group testing: an approach that exploits the quantitative nature of molecular assays and aims at recovering the nonzero patterns in sparse vector. Consider an unknown nonnegative vector $x^* \in \mathbb{R}^n_+$ of molecular abundances in samples from n individuals, with up to k positives; we call such vectors k-sparse. Equal amounts of biological sample from individual i are placed by a laboratory robot into d_i distinct testing pools, with pool j having material from p_j samples. The initial amount of sample, and the time it takes for the laboratory robot to perform the work, puts limits on d_i and p_j. The assignment of samples to pools is given by a binary matrix A, where $A_{ji} = 1$ indicates that a portion of sample i is placed in pool j. The total abundances in the pools are then equal to $y^* = Ax^*$. The abundances are quantified by a molecular assay, such as qPCR, leading to observed measurement vector $\hat{y} \in \mathbb{R}^m$. Readouts are estimated, from C_t values, on a logarithmic scale, leading to noise that is approximately multiplicative, that is, $|\hat{y}_j - y^*_j| \propto y^*_j$. The goal is to design matrix A that will allow us to uncover $\mathrm{supp}(x^*)$, the support of x^*, from $\hat{y}$, that is, find which samples are positive. Matrices designed specifically for real-valued group testing allows for reducing the



C:

Variants of the sparse recovery problem: *design an $m \times n$ measurement/pooling matrix A, such that for any unknown, sparse n-dimensional signal x^*, from m-dimensional vector $\hat{y}$ of noisy measurements of y^*, where $y^* = Ax^*$, we can recover $\hat{x}$ similar to x^*.*

Problem	A_{ij}	x_j	Ax arithmetic	y_i	goal for decoding
Real-valued group testing	$\mathbb{B}$	$\mathbb{R}_+$	$\mathbb{R}$	$\mathbb{R}_+$	$\text{supp}(\hat{x}) = \text{supp}(x^*)$
Combinatorial group testing	$\mathbb{B}$	$\mathbb{B}$	$\mathbb{B}$	$\mathbb{B}$	$\hat{x} = x^*$
Comb. quantitative group testing	$\mathbb{B}$	$\mathbb{B}$	$\mathbb{R}$	$\mathbb{Z}_+$	$\hat{x} = x^*$
Compressed sensing	$\mathbb{R}$	$\mathbb{R}$	$\mathbb{R}$	$\mathbb{R}$	$\hat{x} = x^*$, or $\|\hat{x} - x^*\|$ small
Nonnegative compressed sensing	$\mathbb{R}$	$\mathbb{R}_+$	$\mathbb{R}$	$\mathbb{R}$	$\hat{x} = x^*$, or $\|\hat{x} - x^*\|$ small

Fig. 1. Conceptual illustration of the proposed real-valued group testing. **A:** According to matrix A, genetic material from sample no. 1 (left-most column) goes into testing pool 4 and 6, and testing pool no. 1 (top row) will contain genetic material from samples $6 - 10$. **B:** A robot programmed according to matrix A will distribute samples to pools, then qPCR assays will provide quantitative readout of the amount of viral genetic material in each pool. A decoding algorithm will resolve which cases are positive. **C:** Comparison of real-valued group testing with existing approaches. Note that in our approach, we construct a binary matrix but we column-normalize it to have unit column norms prior to use in testing; this ensures samples have equal contribution to the measurement vector.

number of pools for a given number of samples compared to matrices designed for Boolean combinatorial group testing. For example, for the matrix in Figure 1, quantitative measurements allow for distinguishing between the scenario with one positive at the third column and a scenario with two positives, at first and third columns. The matrix would not be appropriate for Boolean measurements, which would not be able to distinguish between these two scenarios.

We show a new necessary and sufficient condition for binary matrices to guarantee unambiguous recovery of support of k-sparse nonnegative signals. We also provide a deterministic method for constructing matrices meeting the proposed condition, for small values of k, m, n that are relevant for the viral testing setting. The approach is validated using simulated data as well as limited laboratory experiments.

2 Methods

2.1 Notation

Let $[n] = \{1, ..., n\}$. Define support of a vector x by $\text{supp}(x) = \{i \in [n] : x_i \neq 0\}$. The L_0 pseudo-norm is $||x||_0 = |\text{supp}(x)|$. A vector is k-sparse if $||x||_0 \leq k$, that is, if it has up to k non-zeros. For n-dimensional vector x, for $S \subset [n]$, we use x_S to denote a vector of dimensionality equal to x, with entries x_i of vector x where $i \in S$, and with null entries elsewhere.

By $\mathbb{B}^{m \times n}$ we denote the space of $m \times n$ binary matrices normalized to have unit sum of each column; we use $\mathbb{B}$ for brevity even though these matrices are not binary, but instead have entries $1/||A_i||_0$ for each column A_i of matrix A. By $\ker A$ we denote the nullspace of A, a set of solutions to $Ax = 0$. For $m \times n$ matrix A, given $S \subset [n]$, A_S is a submatrix of A formed by its columns $A_i = A(\cdot, i)$, where $i \in S$. Similarly, for $T \subset [m]$, A_T is a submatrix of A formed by its rows $A_j = A(j, \cdot)$, where $j \in T$. By $\bar{S}$ we will denote the complement of S in $[n]$ or $[m]$. It will be clear from the context whether we are considering a rows or columns.

2.2 Overview of the Matrix Design and Decoding Algorithms

Our approach to constructing pooling matrices is based on three observations. First, in Sections 2.3 and 2.3, we prove conditions that a matrix must meet to allow for unambiguous decoding of the positive samples. Next, in Section 2.3, we show that a matrix meeting the conditions can be obtained by starting with a wide initial matrix and removing some columns. Finally, in Section 2.3, we show that the computational cost of finding which columns to remove can be substantially reduced by focusing on a series of submatrices instead of on the initial wide matrix, and that all the submatrices are essentially the same up to permutation of rows and columns, so in fact only one submatrix needs to be considered.

These observations together allow us to formulate a deterministic matrix design method (Algorithm 1) and a corresponding algorithm for decoding positive samples from noisy measurements (Algorithm 2).

Algorithm 1 Pooling Matrix Design Algorithm

Input: m - number of rows; k - sparsity; d - maximum number of nonzeros per matrix column
Output: A - pooling matrix

1: PoolingMatrix(m, k, d)
2: $D_m = \text{WidestBinaryMatrix}(m, d)$
3: $D_{dk} = \text{WidestBinaryMatrix}(dk, d)$
4: $V_{dk} = \text{ViolatingColumnSubsets}(D_{dk})$
5: $V_m = \emptyset$
6: **for all** S : dk-row subset of D_m **do**
7: $V'_{dk} = \text{MapColumnIDs}_{D_{dk} \to D_m}(V_{dk}, S)$
8: $V_m = V_m \cup V'_{dk}$
9: $H = \text{HittingSet}(\text{columns of } D_m, \text{sets } V_m)$
10: $A =$ remove columns H from D_m
11: **return** column-normalized A

Algorithm 2 COMP-NNLS Decoding Algorithm

Input: $\hat{y}$ - measurements; A - pooling matrix
Output: P - set of positives

1: COMP-NNLS$(\hat{y}, A)$
2: **for all** i **do**
3: **if** $\hat{y}_i = 0$ **then**
4: remove columns j from A if $A_{ij} = 1$
5: remove row i from A, $\hat{y}$
6: $\hat{x} = \arg\min_x ||Ax - \hat{y}||_2$ s.t. $x \geq 0$
7: $P = \text{supp}(\hat{x})$
8: **return** P

Pooling Matrix Design Algorithm The pooling matrix construction proceeds in three broad steps. First, (ln. 2 in Alg. 1), an initial wide binary m-row matrix is constructed by concatenating all possible binary columns containing between 2 and d nonzeros. This matrix will lead to many ambiguities if used in group testing. Second (ln. 3-8), we analyze which sets of columns lead to ambiguities – we perform the analysis on a smaller, dk-row binary matrix (ln. 3-4), and translate the results to the initial matrix (ln. 5-8). These steps rely on our main technical contribution, Theorems 2 and 4 and Lemmas 1 and 2 described below in Sections 2.3–2.3. Each of these column sets needs to be broken apart by removing at least one column from each set from the initial matrix, which can be achieved (ln. 9) by solving an instance of the hitting set problem, as described in Section 2.3.

Decoding Algorithm The decoding algorithm (Alg. 2) has two steps. First, in a manner similar to combinatorial orthogonal matching pursuit (COMP) [7], to simplify computations, we eliminate from A all columns that have nonzero entry in rows with null measured y_j, and we also eliminate these rows from A and y (Alg. 2, ln. 2–6). This leaves us with a smaller set of possible positives. For pooling matrices constructed by the proposed approach, in the noise-free case Theorem 2 shows that the set of positives can be unambiguously decoded from measurements y^* by solving $y^* = Ax$ s.t. $x \geq 0$ and taking the support of the solution vector. When noisy measurements $\hat{y}$ are available instead of y^*, this constrained linear system may have empty set of sparse solutions. Indeed, the space of possible measurement noise vectors $\hat{y} - y^*$ is m-dimensional, while the set of k-sparse nonnegative solutions to $y = Ax$ s.t. ≥ 0 is a union of a finite number of k-dimensional sets, with m typically much higher than k. Thus, to find a sparse x we use (Alg. 2, ln. 7) nonnegative least squares (NNLS), that is, we find the signal with the smallest, in the L_2 sense, change in $\hat{y}$ needed to find a nonnegative solution. This approach is aligned with the decoding guarantee for the noisy case provided by Theorem 4.

2.3 Constructing Matrices for Real-valued Group Testing

Necessary and Sufficient Condition in Noise-free Case We will show that any matrix with the following property allows for recovering the pattern of non-zeros of any unknown nonnegative signal with sparsity k.

Definition 1 **k-balanced nullspace property.** *Matrix $A \in \mathbb{B}^{m\times n}$ has k-balanced nullspace property if for all $\eta \in \ker A \setminus \{0\}$, at least $k+1$ entries η_i are positive and at least $k+1$ entries η_j are negative.*

Intuitively (see Fig. 2), the property precludes having $\eta = x' - x''$ with a k-sparse, nonnegative x' and another nonnegative, nonzero x'' with $\text{supp}(x') \neq \text{supp}(x'')$ that could serve as an alternative set of positive cases. If we recover a set of up to k positive cases from the measurement vector y^*, we are guaranteed that there is no other set of positive cases, not just with up to k cases but of any cardinality, consistent with y^*. Thus, the guarantee is stronger than in group testing based on k-disjunct matrices, where the guarantee is limited to alternative sets of positives with cardinality up to k. The intuition is formalized as follows.

Theorem 2. *A matrix $A \in \mathbb{B}^{m\times n}$ allows for decoding* $\text{supp}(x^*)$ *from y^* given by $y^* = Ax^*$ if and only if it meets the k-balanced nullspace property.*

Proof. ($\Longleftarrow$): Let x^* be a k-sparse vector with support S, let $y^* = Ax^*$, and let $x \neq x^*$ be another nonnegative vector with different pattern of non-zeros (i.e., $\text{supp}(x^*) \neq \text{supp}(x)$) such that $y^* = Ax$. Take $\eta = x^* - x$; then $\eta \in \ker A \setminus \{0\}$. Vector η constructed this way will have at most k positive entries, since $-x$ only contributes negative or null entries and x^* is k-sparse nonnegative; k-balanced nullspace property will not hold. Thus, the property implies no such x exists. The unknown pattern of non-zeros in x^* can be recovered from the unique solution to the constrained linear system of equations $y^* = Ax$, $x \geq 0$.
($\Longrightarrow$): Assume k-balanced nullspace property is violated. We will show that recovering $\text{supp}(x^*)$ unambiguously is not possible for some nonnegative x^*, that is, k-balanced nullspace property is necessary for real-valued group testing. For A with only positive entries, any $\eta \in \ker A \setminus \{0\}$ must have at least one positive and one negative entry. Let S be the support of the positive entries in η, and $-S$ the support of the negative entries; both are nonempty. Both η_S and $-\eta_{-S}$ are nonnegative, at least one of them is k-sparse, and $\text{supp}(\eta_S) \neq \text{supp}(-\eta_{-S})$. We have $A\eta = A\eta_S + A\eta_{-S} = 0$. Both η_S and $-\eta_{-S}$ lead to the same measurement vector $y = A\eta_S = A(-\eta_{-S})$; given y and A, S and $-S$ resulting from η provide two different, equally possible sets of positive cases.

Sufficient Condition for the Multiplicative Noise Case The measurements of $y = Ax$ are expected to be noisy, with the observed $\hat{y} \neq y$. For nonnegative x and A, the true y is also nonnegative, and in qPCR and similar settings, $\hat{y}$ is also constrained to be nonnegative by the nature of the measurement process. Often, with high probability, null y_j leads to null $\hat{y}_j$. To fit these characteristics, we focus on multiplicative noise.

Consider a k-sparse solution $x \in \mathbb{R}^n_+$ resulting in noise-free measurement $y \in \mathbb{R}^m$. The observed results of the measurements is $\hat{y}$, with noise of magnitude limited by $\delta > 0$ in the sense $|\hat{y}_i - y_i| \leq \delta y_i$. This noise bound is not symmetric, that is, bound on $\hat{y}_i/y_i$ is $1+\delta$ while the bound on $y_i/\hat{y}_i$ is not $1/(1+\delta)$; for small values of δ it approximates the symmetric multiplicative noise model that captures qPCR noise well.

The bound on the magnitude of noise can help us establish how much the measurement noise can impact the decoding of the positive cases from $\hat{y}$.

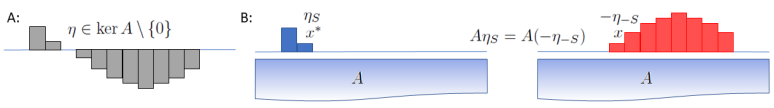


Fig. 2. Illustration of k-balanced nullspace property for $k = 2$. A: Vector η that violates it by having only $k = 2$ positive entries. B: The violation allows for two alternative solutions, $x^* = \eta_S$ and $x = -\eta_{-S}$, with the same measurement outcome.

Definition 3 k, l**-sparse** δ**-robustness property.** *Matrix* $A \in \mathbb{B}^{m \times n}$ *has* k, l*-sparse* δ*-robustness property if for all* k*-sparse* $x' \in \mathbb{R}_+^n$ *with* $||x'||_1 = 1$*, for all* l*-sparse* $x'' \in \mathbb{R}_+^n$ *with* $\text{supp}(x') \neq \text{supp}(x'')$*, we have* $||Ax' - Ax''||_2 > 2\delta$*.*

The δ-robustness property is an extension of the k-balanced property, which is a special case for $\delta = 0$ and $l = n - 1$. The property guarantees, via Theorem 4, that for two different sets of positives, the corresponding noise-free measurements are so different that even in the presence of measurement noise, the two sets of positives can be distinguished by the decoding algorithm (see Fig. 3).

Theorem 4. *Consider* $A \in \mathbb{B}^{m \times n}$*, and noisy measurements such that for every* $x \in \mathbb{R}_+^n$ *we observe* $\hat{y}$ *instead of* $y = Ax$*, with multiplicative error* $|\hat{y}_i - y_i| \leq \delta y_i$ *of magnitude limited by a constant* $\delta \in \mathbb{R}_+$*. Consider a* k*-sparse* $x^* \in \mathbb{R}_+^n$ *and a set* $X \subset \mathbb{R}_+^n$ *of alternative,* l*-sparse solutions with different support than* x^**. If* A *has* k, l*-sparse* δ*-robustness property, then* $\text{supp}(\arg\min_x ||Ax - \hat{y}||_2 \text{ s.t. } x \geq 0)$ *will correctly decode* $\text{supp}(x^*)$*.*

Proof. For any $A \in \mathbb{B}^{m \times n}$, we have $||y||_1 = ||Ax||_1 = ||x||_1$, since columns of $A \in \mathbb{B}^{m \times n}$ add up to one. Noise of the form $|\hat{y}_i - y_i| \leq \delta y_i$ guarantees that $||\hat{y} - y||_2 \leq \delta ||y||_2 \leq \delta ||y||_1 = \delta ||x||_1$. Consider any $x' \in X$ and let $y' = Ax'$; also, let $y^* = Ax^*$. For x' to possibly be a solution $x' = \arg\min_x ||Ax - \hat{y}||_2$ s.t. $x \geq 0$ instead of x^*, we need $||y' - \hat{y}||_2 \leq ||y^* - \hat{y}||_2$. This implies $||y' - \hat{y}||_2 \leq \delta ||x||_1$. From triangle inequality, it further implies $||y' - y^*||_2 \leq ||y^* - \hat{y}||_2 + ||y' - \hat{y}||_2 \leq 2\delta ||x^*||_1$, that is, $||Ax' - Ax^*||_2 \leq 2\delta ||x^*||_1$. If A has k, l-sparse δ-robustness property, we have a contradiction.

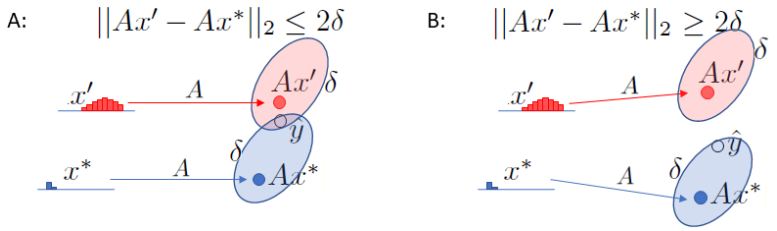


Fig. 3. Illustration of k, l-sparse δ-robustness property. A: Violation of the property may lead to ambiguity in decoding noisy measurement $\hat{y}$. B: If the property holds, x^* can be decoded from $\hat{y}$.

Property-violating Column Sets To arrive with a matrix meeting the k-balanced nullspace property and the k, l-sparse δ-robustness property, we can start with a matrix that does not meet the property, and then trim the matrix, as shown by the following result.

Definition 5 Violating set. *A* violating set *of a matrix* $A \in \mathbb{B}^{m \times n}$ *is a subset* V *of columns of* A *such that submatrix* A_V *does not meet* k*-balanced nullspace property, or the* k, l*-order* δ*-robustness property. A* minimal violating set *is a violating set that contains no smaller violating sets.*

Lemma 1. *For a matrix* D*, a submatrix* $A \in \mathbb{B}^{m \times n}$*, formed by eliminating columns from* D*, meets the* k*-balanced nullspace property and the* k, l*-order* δ*-robustness property if one column has been removed from all minimal violating sets of* D*.*

Proof. Removing columns from a matrix does not introduce new violating sets, but removing even one column from a minimal violating set removes the violating set. Thus, removing one column from each minimal violating set eliminates all violating sets.

All minimal violating sets for the k-balanced nullspace property can be obtained by enumerating all minimal linearly dependent sets of the matrix and filtering them according to the number of positive and

negative linear weights. Finding minimal dependent sets is a special case of enumerating all circuits in a matroid [22]. For our experiments, we use the implementation in the function `circuits` of `4ti2` [35]. For the δ-robustness property, quadratic programming iterated over sets of up to $k + l$ columns can produce all minimal violating sets.

Reducing the Search for Property-violating Sets to a Small Base Problem Detecting all minimal dependent sets for a large $m \times n$ matrix in $\mathbb{B}^{m\times n}$ has combinatorial complexity and is computationally infeasible. For matrices $A \in \mathbb{B}^{m\times n}$ with up to d non-zeros in each column, $y = Ax$ will be sparse for nonnegative, k-sparse x, and we have the following decomposition that can make the search for testing matrices with m rows more efficient, by allowing us to focus on a small submatrices with dk rows, irrespective of how large m is.

Lemma 2. Decomposition into Base Problem. *Consider $A \in \mathbb{B}^{m\times n}$ with $||A_i||_0 \leq d$ for each column i. The matrix meets the k-balanced nullspace property or the k, l-sparse δ-robustness property if and only if for every dk-sized subset $T \in [m]$, the submatrix A_T formed by taking only rows in T, and then eliminating columns that have support extending beyond T, has those properties.*

Proof. Whether a set of matrix columns violates either of the properties is not affected by removing rows that have all-null values in these columns. If the property is violated for A_T, it is violated by some specific columns of A_T. Extending these columns by bringing back the remaining rows of A, which have all-null values, does not remove the violation involving the columns, and A does not meet the property. Conversely, if A does not meet the property, then some k-sparse vector x that leads to a measurement with at most dk nonzeros has an alternative x' that also has to lead to a measurement with up to dk nonzeros, in the same set T, and the property that guaranties lack of ambiguity must be violated for submatrix A_T.

It is not necessary to detect all minimal dependent sets in all the submatrices A_T, we only need to do it once, for a small matrix. Up to permutation of columns and rows, all these submatrices are identical to, or a submatrix of, of a single small matrix D_{dk} with dk rows and $\sum_{i=2}^{d} \binom{dk}{i}$ columns that results from concatenating all possible binary columns with two ones, with three ones, and up to dk ones. Given all minimal violating column sets in D_{dk}, a simple, efficient procedure can compute column indices in A from indices in D_{dk}. We can iteratively translate minimal violating sets of D_{dk} to minimal violating sets of each kd-row subset of A, which together form all the minimal violating sets of A.

Constructing the Pooling Matrix based on Property-violating Column Sets For a given k and d, finding matrix A with the largest n for a given m by starting from some initial matrix D_m and then removing smallest number of columns that break up all minimal violating sets corresponds to the hitting set problem. The hitting set problem involves a universe U, and a set Σ of subsets of U. The goal is to select the smallest number of items from U such that each set from Σ contains at least one of those items. Here, the initial matrix D_m is an m-row matrix that has all $N = \binom{m}{2} + ... + \binom{m}{d}$ possible binary vectors with $2, ..., d$ ones as columns. Universe U is the set of columns of the initial matrix D_m, and Σ are all the minimal violating sets of D_m. We aim to break each minimal violating set by removing as few columns from D_m as possible. The problem, through equivalence with set cover problem, is NP-hard. The trivial approach of checking every subset of columns of D_m matrix has complexity of $|\Sigma|2^N$, and algorithms that require instead $N|\Sigma|2^{|\Sigma|}$ also have been formulated [16], but in our case $|\Sigma| > N$. While an approximation algorithm can be used for large set cover problems, for small values of m, k, and d considered here, integer programming is effective. We used GUROBI [20] solver with the number of threads limited to 32, limiting the time of calculations for each m to 500,000 seconds. All calculations were performed on a machine with 4 Intel Xeon E7-8894V4 2.4GHz CPUs, with 24 cores per CPU, and 6TB RAM.

3 Results

3.1 Comparison of Matrix Properties with Existing Approaches

We used the proposed approach to construct two $k = 2$, $d = 3$ testing matrices, for $m = 12$ and $m = 16$. These two numbers of pools correspond well to the standard dimensions of a $384 = 16x24$ well plate. A robot with a multichannel pipette can work on multiple columns (for $m = 16$) or half-rows (for $m = 12$) in parallel. We used the noise limit of $\delta = 0.125$ for $m = 12$ and $\delta = 0.15$ for $m = 16$. The properties of the two pooling matrices are presented in Table 1. For comparison, we also summarized the properties of matrices of similar sizes resulting from existing approaches for combinatorial group testing or nonnegative compressed sensing.

In combinatorial group testing, unambiguous decoding of k-sparse signals from Boolean measurements is guaranteed for a binary pooling matrix if the matrix is k-disjunct [15], that is, if for any set of k columns, every other column has at least one non-zero element in a row where all the k columns are null. Reed-Solomon codes [32] with Kautz-Singleton construction [23] are a popular way of constructing k-disjunct matrices. The three RS/KS matrices with smallest m, comparable to o m in our matrices, are 12×16 (RS code: $[3, 2]_4$), 15×25 ($[3, 2]_5$), and 21×49 ($[3, 2]_7$); these have lower compression rates than our matrices.

RS/KS construction is not optimal [31]. To find optimal codes for a given m, we used integer programming to find 2-disjunct matrices with highest possible n for values of $m = 12, 16$ used in our matrices. The resulting widest-possible 2-disjunct matrices are 12×20 and 16×37. These have much lower compression rates n/m than matrices constructed using our approach that exploits the fact that the matrices will be used with real-valued measurements instead of Boolean measurements.

RS codes are better suited for larger m and n, and have been used recently in SARS-CoV-2 testing for settings that involve lower positivity rates. P-BEST [33], a method based on RS codes, involves a 48×384 matrix designed for positivity rates up to 1%, much lower rate than our matrices; it also has much higher sample dilution rate. However, it offers two-fold higher compression rate than our matrices, and is a better choice in low-positivity scenarios.

Table 1. Properties of pooling matrices resulting from our approach for $k = 2$, $d = 3$, compared with existing methods for matrix construction. For each $m \times n$ matrix, we provide the maximum number of non-zeros per column (d_{max}) and per row (p_{max}), the maximum sample dilution dil_{max} resulting from d_{max} and p_{max}, maximum sparsity assumed in matrix design, k_{max}, the maximum positivity rate, k_{max}/n the matrix is designed for, and the testing compression rate, n/m.

matrix design method	m	n	d_{max}	p_{max}	dil_{max}	k_{max}	k_{max}/n	n/m
proposed method	16	66	3	14	42	2	3.03%	4.13:1
proposed method	12	36	3	10	30	2	5.55%	3.00:1
Combinatorial Group Testing Matrices								
Tapestry [19]	45	105	3	8	24	2	1.90%	2.33:1
P-BEST [33]	48	384	6	48	288	4	1.04%	8.00:1
$[3, 2]_7$ Reed-Solomon/Kautz-Singleton	21	49	3	7	21	2	4.08%	2.33:1
$[3, 2]_5$ RS/KS [32,23]	15	25	3	5	15	2	8.00%	1.67:1
$[3, 2]_4$ RS/KS [32,23]	12	16	3	4	12	2	12.50%	1.33:1
Optimal 2-disjunct [15]	16	37	3	7	21	2	5.40%	2.32:1
Optimal 2-disjunct [15]	12	20	3	5	15	2	10.00%	1.67:1
Nonnegative Compressed Sensing Matrices								
Random-Permutation [24]	16	66	3	15	45	–	–	4.13:1
Random-Permutation [24]	12	36	3	12	36	–	–	3.00:1
Random-Binomial [25]	16	66	7	17	119	–	–	4.13:1
Random-Binomial [25]	12	36	5	13	65	–	–	3.00:1
One-sided Coherence [5]	16	20	4	5	20	2	10.00%	1.25:1
One-sided Coherence [5]	12	9	4	3	12	2	22.20%	0.75:1

Another method proposed for SARS-CoV-2 testing is Tapestry [19], which relies on Kirkman Triple Systems [28] for constructing 2-disjunct matrices. The 45×105 Tapestry matrix, while having lower compression rate than our 16×66 matrix, also only has unambiguous decoding guarantees for positivity of up to 1.9%.

In addition to matrices used in combinatorial group testing, we also evaluated matrix design approaches proposed in the nonnegative compressed sensing domain which, unlike our matrices, are designed with the goal of recovering the vector, not only its support. One-side coherence [5] has been proposed as a computationally-efficient sufficient condition for sparse recovery of nonnegative signals. It uses a condition $\rho/(1+\rho) < 1/2k$ with $\rho = \max_{i \neq j} |A_i^T A_j| / ||A_i||_2^2$; for k=2, one needs $\rho < 1/3$. The condition was initially used for random $[0,1]$-uniform matrices, which have low $A_i^T A_j$, but which are not feasible to use in laboratory setting. For binary matrices, $A_i^T A_j$ reduces to overlap and $||A_i||_2^2$ to $||A_i||_0$; to have a chance of finding A with $\rho < 1/3$ we need $||A_i||_0 \geq 4$, that is, columns with at least four non-zeros. Maximum independent set on the graph of all possible $||A_i||_0 \geq 4$ columns, with edge if $A_i^T A_j > 1$, is tractable for small m, but highest-n-for-m matrices are very poor: 12×9, and 16×20.

Other approaches [25,24] for creating matrices for nonnegative compressed sensing use random binary matrices. We tested matrices with entries sampled from Bernoulli distribution, an approach that has been used previously to construct sensing matrices that meet sparse recovery guarantees with high probability, including in the nonnegative case [25]. In order to increase the quality of the random matrices we used as comparison to our approach, instead of single random matrix, we sampled 100 Bernoulli matrices, and we picked the matrix that has highest mean of sensitivity and specificity when tested on 100 random 2-sparse inputs x. Bernoulli matrices may have highly varying number of ones between columns; as an alternative, we used a permutation approach that guarantees that the number of elements in each column is between 2 and 3, as in matrices resulting from our method. A similar approach was used previously in nonnegative sparse recovery [24]. Again, we picked the best of 100 random matrices for comparison with our approach. Random matrices only provide probability guarantees that hold with high probability for large n, large m. The small random matrices used here do not come with guarantees for unambiguous recovery for any sparsity value k, which is indicated by a dash in Table 1.

3.2 Effectiveness on Simulated Data

We have used synthetic, simulated data to evaluate the sensitivity and specificity of decoding the positive cases from real-valued measurements using the matrices in Table 1. For each $m \times n$ matrix, we simulated 10,000 random input vectors of dimensionality n and with given number of nonzeros. The nonzero elements of each vector were sampled uniformly at random from $[1.0, 1000.0]$ range, that is, from a range spanning around 10 qPCR C_t cycles. True, noise-free measurements were obtained by setting $y = Ax$. To simulate noisy qPCR measurements $\hat{y}$, we used a realistic model of qPCR noise [19], of the form $\hat{y}_i = (1+q)^{\mathcal{N}(0,\sigma^2)} y_i$,

Table 2. Sensitivity and specificity of pooling matrices resulting from our approach for varying positivity rates in simulated noise-free and noisy measurements scenarios, for 10,000 simulated experiments. For each matrix, we provide its dimensions and its compression rate. Maximum positivity rate for which the matrix was designed is marked with an asterisk.

	positivity	Noise-free measurements		Measurements with simulated noise	
matrix	rate	sensit. [%]	specif. [%]	sensit. [%]	specif. [%]
proposed 16×66, $4.13:1$	1.52%	100.0	100.0	100.00	99.99
proposed 16×66, $4.13:1$	3.03% *	100.0	100.0	99.48	99.18
proposed 16×66, $4.13:1$	4.55%	100.0	99.25	98.06	96.50
proposed 12×36, $3:1$	2.78%	100.0	100.0	100.00	99.91
proposed 12×36, $3:1$	5.56% *	100.0	100.0	99.26	97.88
proposed 12×36, $3:1$	8.33%	100.0	97.39	97.72	92.36

with the hyperparameters set to $q = 0.95$ and $\sigma = 0.1$. Given A and $\hat{y}$, we used COMP-NNLS decoding algorithm (Alg. 2) to identify the positive samples.

First, we focused on the matrices resulting from the proposed method, and analyzed how they behave when the set of samples have varying positivity rates. We used three positivity rates: two that are within the rate range for which the matrix is designed for, and one that exceeds the rate by half. Results in Table 2 show that within the designed range of positivity rates, the matrices have high sensitivity, above 99%, and high specificity, above 99% for the 16×66 matrix and above 97% for the 14×36 matrix. When the maximum number of positives that the matrix was designed for is exceeded by half, the sensitivity drops by about 1.5% and remains above 97%, while the specificity drops to between 92% and 96%, indicating a small increase in the number of false negatives and a moderate increase in the number of false positives.

Next, we compared the proposed matrices with matrices from existing approaches in experiments with $k = 2$ positives, that is, the highest number of positives the matrices are designed for (only P-BEST is designed for higher number of positives, $k_{max} = 4$). As Table 3 show, in the scenario with noise-free measurements, all the matrices, with the exception of some of the random matrices, exhibit perfect accuracy.

As expected, matrices originating from Boolean combinatorial group testing, which are designed to work without quantitative information available for the decoding process, are not significantly affected by the presence of multiplicative measurement noise. This comes at the cost of having relatively low compression rates, (small m, n matrices) or being limited to low positivity rates (large m, n P-BEST and Tapestry).

Random matrices originating from nonnegative compressed sensing approaches are by design of the same shape as matrices from our approach, with the same compression rates. Permutation-based matrices exhibit slightly lower sensitivity than our deterministically-designed matrices, and show specificity lower by 0.78% for the 16×66 matrix, and by 1.44% for the 12×36 matrix – in both cases, this drop translates to almost doubling of the number of false positives. Binomial random matrices have lower specificity and sensitivity. Matrices based on one-sides coherence, while having high sensitivity and specificity, achieve that by offering very low compression rate.

Table 3. Sensitivity and specificity pooling matrices resulting from our approach compared to existing approaches in simulated noise-free and noisy measurements scenarios, for 10,000 simulated experiments with $k = 2$ positive samples. For each matrix, we provide its dimensions and its compression rate.

	Noise-free measurements		Measurements with simul. noise	
matrix	sensit. [%]	specif. [%]	sensit. [%]	specif. [%]
proposed method 16×66, $4.13 : 1$	100.0	100.0	99.48	99.18
proposed method 12×36, $3 : 1$	100.0	100.0	99.26	97.88
Tapestry 45×105, $2.33 : 1$	100.0	100.0	99.93	100.0
P-BEST 48×384, $8 : 1$	100.0	100.0	99.92	100.0
$[3, 2]_7$ RS/KS 21×49, $2.33 : 1$	100.0	100.0	99.87	100.0
$[3, 2]_5$ RS/KS 15×25, $1.67 : 1$	100.0	100.0	99.81	100.0
$[3, 2]_4$ RS/KS 12×16, $1.33 : 1$	100.0	100.0	99.84	100.0
Opt. 2-disj. 16×37, $2.31 : 1$	100.0	100.0	99.85	100.0
Opt. 2-disj. 12×20, $1.67 : 1$	100.0	100.0	99.79	100.0
Rand-Perm 16×66, $4.13 : 1$	100.0	99.77	99.10	98.40
Rand-Perm 12×36, $3 : 1$	100.0	99.19	98.87	96.44
Rand-Binom 16×66, $4.13 : 1$	95.39	98.18	94.50	95.30
Rand-Binom 12×36, $3 : 1$	100.0	97.57	99.00	93.09
1-sided coh. 16×20, $1.25 : 1$	100.0	100.0	99.77	100.0
1-sided coh. 12×9, $0.75 : 1$	100.0	100.0	99.76	100.0

3.3 Effectiveness in Wet Lab

To validate the proposed approach in a laboratory setting, we performed limited testing of biological samples using qPCR assay. We focused on the 12×36 pooling matrix that offers higher maximum positivity rate, $k/n = 5.5\%$, than the 16×66 matrix. We used one set of 36 human samples: two samples were previously confirmed to be positive for SARS-CoV-2, and the remaining 34 were confirmed negatives. Briefly, nasopharyngeal swabs were collected and immediately transferred to 3 ml of PrimeStore MTM medium to deactivate the virus. Nucleic acids were purified using ThermoFisher MagMAX Viral/Pathogen Nucleic Acid Isolation Kit. Nucleic Acid-extracts were stored at -80°C. Virus detection and viral loads were performed by qPCR using IDT 2019-nCoV CDC qPCR Probe Assay targeting the SARS-CoV-2 nucleocapsid gene (N1 and N2). The human RNase P gene was used as a control for sample integrity. Amplification was performed in our ViiA7 Real-Time PCR System using the TaqMan Fast Virus 1-Step Master Mix.

We randomly assigned the two positive samples to two columns of the pooling matrix. Samples were thawed in ice and spun down for 5 seconds. We then prepared 12 master wells. For each $A_{ij} = 1$ in the sensing matrix A, we pipetted 3 μl of biological material from sample j into master well i. We then pipetted 5 μl of volume from each master well into a separate testing well. The qPCR assay was performed in a 20 μl volume containing 5 μl of 4×TaqMan Fast Virus 1-Step Master Mix (Thermo Fisher Scientific), 1.5 μl of primers/probe set, and 8.5 μl DEPC-treated water. The qPCR was performed using ViiA7 Real-Time PCR System (Thermo Fisher Scientific) with the following cycling conditions: reverse transcription at 50°C for 15 min. and 95°C for 2 min., followed by 40 cycles of PCR at 95°C for 3 sec. and 55°C for 30 seconds. We ran 12 individual qPCR assays, one per testing well. Total viral load in each testing well was estimated from the qPCR C_t value. Briefly, a standard curve was obtained by amplification of known amounts of SARS-CoV-2 (IDT 2019-CoV Plasmid Controls). Five consecutive dilutions (dilution factor 1:10) were prepared containing from 104 to 1 copies/reaction. The amounts of SARS-CoV-2 in samples were obtained by plotting C_t values onto the standard curve.

A second round of 12 qPCR assays was carried out using 5 μl volume from each master wells, leading to a second set of measurements, with the same underlying viral loads but differing due to technical variability of the qPCR assay. Finally, we performed the same experiment on the same pooled master wells, by diluting the master well 1:5, then pipetting 5 μl of volume into a testing well, and performing qPCR. This set of tests was again performed in duplicate, resulting in two additional data sets differing due to technical variability. For all four data sets, we used NNLS to recover the viral loads and, subsequently, the sparsity pattern of the 36 samples. In all four data sets, the method correctly identified both positive samples, and correctly labeled the remaining 34 samples as negative.

4 Conclusion

We provided a theoretical and empirical exploration of real-valued group testing, a setting that is relevant for improving efficiency of community testing for viral diseases. The proposed approach is focused on small-to-medium sized matrices that are convenient in a laboratory setting. The resulting matrices are useful for positivity rates below 5%, where they offer higher reduction in the number of tests than matrices designed for Boolean combinatorial group testing while maintaining high sensitivity and specificity. For higher positivity rates, the approach is not recommended. For much lower positivity rates, up to about 1%, existing approaches such as P-BEST offer higher reduction in the number of tests.

Acknowledgements

T.A., G.A.B, and M.S. are funded by NSF grant CBET-2034995.

References

1. Abdalhamid, B., Bilder, C.R., McCutchen, E.L., Hinrichs, S.H., Koepsell, S.A., Iwen, P.C.: Assessment of specimen pooling to conserve SARS CoV-2 testing resources. American Journal of Clinical Pathology **153**(6), 715–718 (2020)
2. Aldridge, M., Johnson, O., Scarlett, J., et al.: Group testing: An information theory perspective. Foundations and Trends® in Communications and Information Theory **15**(3-4), 196–392 (2019)
3. Atia, G.K., Saligrama, V.: Boolean compressed sensing and noisy group testing. IEEE Transactions on Information Theory **58**(3), 1880–1901 (2012)
4. Berinde, R., Gilbert, A.C., Indyk, P., Karloff, H., Strauss, M.J.: Combining geometry and combinatorics: A unified approach to sparse signal recovery. In: 46th Annual Allerton Conference on Communication, Control, and Computing. pp. 798–805. IEEE (2008)
5. Bruckstein, A.M., Elad, M., Zibulevsky, M.: On the uniqueness of non-negative sparse & redundant representations. In: 2008 IEEE International Conference on Acoustics, Speech and Signal Processing. pp. 5145–5148. IEEE (2008)
6. Candes, E.J., Romberg, J.K., Tao, T.: Stable signal recovery from incomplete and inaccurate measurements. Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences **59**(8), 1207–1223 (2006)
7. Chan, C.L., Che, P.H., Jaggi, S., Saligrama, V.: Non-adaptive probabilistic group testing with noisy measurements: Near-optimal bounds with efficient algorithms. In: 2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton). pp. 1832–1839. IEEE (2011)
8. Cleary, B., Hay, J.A., Blumenstiel, B., Harden, M., Cipicchio, M., Bezney, J., Simonton, B., Hong, D., Senghore, M., Sesay, A.K., Gabriel, S., Aviv, R., Mina, M.J.: Using viral load and epidemic dynamics to optimize pooled testing in resource-constrained settings. Science translational medicine **13**(589) (2021)
9. Cohen, A., Dahmen, W., DeVore, R.: Compressed sensing and best k-term approximation. Journal of the American Mathematical Society **22**(1), 211–231 (2009)
10. Corman, V.M., Landt, O., Kaiser, M., Molenkamp, R., Meijer, A., Chu, D.K., Bleicker, T., Brünink, S., Schneider, J., Schmidt, M.L., et al.: Detection of 2019 novel coronavirus (2019-ncov) by real-time RT-PCR. Eurosurveillance **25**(3), 2000045 (2020)
11. Donoho, D.L.: Compressed sensing. IEEE Transactions on Information Theory **52**(4), 1289–1306 (2006)
12. Donoho, D.L., Tanner, J.: Sparse nonnegative solution of underdetermined linear equations by linear programming. Proceedings of the National Academy of Sciences **102**(27), 9446–9451 (2005)
13. Dorfman, R.: The detection of defective members of large populations. The Annals of Mathematical Statistics **14**(4), 436–440 (1943)
14. Du, D., Hwang, F.K., Hwang, F.: Combinatorial group testing and its applications. World Scientific (2000)
15. D'yachkov, A.G., Rykov, V.V.: Bounds on the length of disjunctive codes. Problemy Peredachi Informatsii **18**(3), 7–13 (1982)
16. Fomin, F.V., Kratsch, D.: Exact Exponential Algorithms. Springer-Verlag, Berlin, Heidelberg (2010)
17. Gandikota, V., Grigorescu, E., Jaggi, S., Zhou, S.: Nearly optimal sparse group testing. IEEE Transactions on Information Theory **65**(5), 2760–2773 (2019)
18. Gebhard, O., Hahn-Klimroth, M., Parczyk, O., Penschuck, M., Rolvien, M.: Near optimal sparsity-constrained group testing: improved bounds and algorithms. arXiv:2004.11860 (2020)
19. Ghosh, S., Agarwal, R., Rehan, M.A., Pathak, S., Agarwal, P., Gupta, Y., Consul, S., Gupta, N., Ritika, R., Goenka, R., Rajwade, A., Gopalkrishnan, M.: A compressed sensing approach to pooled rt-pcr testing for covid-19 detection. IEEE Open Journal of Signal Processing (2021)
20. Gurobi Optimization, L.: Gurobi optimizer reference manual (2020), `http://www.gurobi.com`
21. Heidarzadeh, A., Narayanan, K.R.: Two-stage adaptive pooling with RT-qPCR for COVID-19 screening. arXiv:2007.02695 (2020)
22. Hemmecke, R.: On the positive sum property and the computation of Graver test sets. Mathematical Programming, Series B **96**, 247–269 (2003)
23. Kautz, W., Singleton, R.: Nonrandom binary superimposed codes. IEEE Transactions on Information Theory **10**(4), 363–377 (1964)
24. Khajehnejad, M.A., Dimakis, A.G., Xu, W., Hassibi, B.: Sparse recovery of nonnegative signals with minimal expansion. IEEE Transactions on Signal Processing **59**(1), 196–208 (2010)
25. Kueng, R., Jung, P.: Robust nonnegative sparse recovery and the nullspace property of 0/1 measurements. IEEE Transactions on Information Theory **64**(2), 689–703 (2017)
26. Lustig, M., Donoho, D.L., Santos, J.M., Pauly, J.M.: Compressed sensing MRI. IEEE Signal Processing Magazine **25**(2), 72–82 (2008)

27. Malioutov, D., Malyutov, M.: Boolean compressed sensing: LP relaxation for group testing. In: 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 3305–3308. IEEE (2012)
28. Ngo, H.Q., Du, D.Z.: A survey on combinatorial group testing algorithms with applications to dna library screening. Discrete mathematical problems with medical applications **55**, 171–182 (2000)
29. Perchetti, G.A., Sullivan, K.W., Pepper, G., Huang, M.L., Breit, N., Matthias, P., Jerome, K.R., Greninger, A.L.: Pooling of SARS-CoV-2 samples to increase molecular testing throughput. Journal of Clinical Virology p. 104570 (2020)
30. Pilcher, C.D., Westreich, D., Hudgens, M.G.: Group testing for severe acute respiratory syndrome–coronavirus 2 to enable rapid scale-up of testing and real-time surveillance of incidence. The Journal of Infectious Diseases **222**(6), 903–909 (2020)
31. Porat, E., Rothschild, A.: Explicit nonadaptive combinatorial group testing schemes. IEEE Transactions on Information Theory **57**(12), 7982–7989 (2011)
32. Reed, I.S., Solomon, G.: Polynomial codes over certain finite fields. Journal of the Society for Industrial and Applied Mathematics **8**(2), 300–304 (1960)
33. Shental, N., Levy, S., Wuvshet, V., Skorniakov, S., Shalem, B., Ottolenghi, A., Greenshpan, Y., Steinberg, R., Edri, A., Gillis, R., et al.: Efficient high-throughput SARS-CoV-2 testing to detect asymptomatic carriers. Science Advances p. eabc5961 (2020)
34. Täufer, M.: Rapid, large-scale, and effective detection of COVID-19 via non-adaptive testing. Journal of Theoretical Biology **506**, 110450 (2020)
35. 4ti2 team: 4ti2—a software package for algebraic, geometric and combinatorial problems on linear spaces (2020), `https://4ti2.github.io`
36. Walensky, R.P., Del Rio, C.: From mitigation to containment of the COVID-19 pandemic: Putting the SARS-CoV-2 genie back in the bottle. JAMA–Journal of the American Medical Association **323**, 1889–1890 (2020)
37. Wang, C., Zhao, Q., Chuah, C.N.: Optimal nested test plan for combinatorial quantitative group testing. IEEE Transactions on Signal Processing **66**(4), 992–1006 (2017)