Sampling Schemes and Recovery Algorithms for Functions of Few Coordinate Variables

Simon Foucart[*]— Texas A&M University

Abstract

When a multivariate function does not depend on all of its variables, it can be approximated from fewer point evaluations than otherwise required. This has been previously quantified e.g. in the case where the target function is Lipschitz. This note examines the same problem under other assumptions on the target function. If it is linear or quadratic, then connections to compressive sensing are exploited in order to determine the number of point evaluations needed for recovering it exactly. If it is coordinatewise increasing, then connections to group testing are exploited in order to determine the number of point evaluations needed for recovering the set of active variables. A particular emphasis is put on explicit sets of evaluation points and on practical recovery methods. The results presented here also add a new contribution to the field of group testing.

Key words and phrases: High-dimensional approximation, active variables, sparsity, adaptivity vs. nonadaptivity, group testing.

AMS classification: 65D15, 41A63, 26B99.

1 Introduction

In this note, we revisit the problem put forward in [3], namely, we examine functions F defined on the high-dimensional cube $[0,1]^N$ (or $[-1,1]^N$) that actually depend on at most s coordinate variables. These variables, called active variables, are of course unknown. We can write succinctly

$$(1) \qquad F(x) = f(x_{j_1}, \ldots, x_{j_s}), \qquad x = (x_1, \ldots, x_N) \in [0,1]^N,$$

for some function f defined on the lower-dimensional cube $[0,1]^s$. In the scenario studied here, the functions F are only available through point evaluations, i.e., users can access samples

$$(2) \qquad y_i = F(x^{(i)}), \qquad i \in [1:m],$$

[*]S. F. partially supported by NSF grants DMS-1622134 and DMS-1664803.

while being allowed to freely select the points $x^{(1)}, \ldots, x^{(m)} \in [0,1]^N$. Given an accuracy ε, the goal is to approximate the function F by a function $\widetilde{F}$ constructed from the data $(x^{(1)}, y_1), \ldots, (x^{(m)}, y_m)$ —and possibly the knowledge of s (an upper estimation for the exact number of active variables)— in such a way that

$$(3) \qquad \|F - \widetilde{F}\|_\infty = \max_{x \in [0,1]^N} |F(x) - \widetilde{F}(x)| \le \varepsilon |F|,$$

where $|\cdot|$ is a seminorm on the space $\mathcal{X}$ of functions being considered. Since one wishes to achieve this approximation goal with as few samples as possible, we informally introduce the sample complexity $m(s, N, \varepsilon, \mathcal{X})$ to denote the minimal value of m in (2) that makes (3) realizable. There should in fact be two such quantities, $m_{\text{ada}}(s, N, \varepsilon, \mathcal{X})$ and $m_{\text{nonada}}(s, N, \varepsilon, \mathcal{X})$, to distinguish between two settings, a priori different, where the samples are adaptive, i.e., each point $x^{(i)}$ may be chosen depending on the previous data $((x^{(1)}, y_1), \ldots, (x^{(i-1)}, y_{i-1}))$, and where the samples are nonadaptive, i.e., all the points $x^{(i)}$ are chosen once and for all at the beginning of the acquisition process.

To summarize the results brought to light in [3] and improved in [11], which both consider in particular the space $\mathcal{X} = \text{Lip}$ of Lipschitz functions, we state that[1]

$$(4) \qquad m_{\text{ada}}(s, N, \varepsilon, \text{Lip}) \le C(s) \varepsilon^{-s} \log(N/s), \qquad C(s) = C\, s\, e^s,$$

$$(5) \qquad m_{\text{nonada}}(s, N, \varepsilon, \text{Lip}) \le C(s) \varepsilon^{-s} \log(N/s), \qquad C(s) = C\, s^2 e^s.$$

The term ε^{-s} has to be here, because it is necessary even if the locations of the active variables were known in advance. The term $\log(N/s)$, interpreted as the price to pay for not knowing these locations in advance, also has to be here, because it must appear when F is linear (see Section 2). As for the term $C(s)$, its exact dependence on s does not really matter, so long as $C(s) \le \kappa^s$, $\kappa > 1$. Indeed, such a behavior can be absorbed in the term ε^{-s}, which then guarantees that (3) holds with ε replaced by $\kappa \varepsilon$ as soon as $m \ge \varepsilon^{-s} \log(N/s)$. Therefore, the articles [3, 11] settled the sample complexity question for the approximation of functions of few coordinate variables, revealing essentially no difference between the adaptive and nonadaptive cases.

There are, however, two important questions left unanswered by [3, 11]. The first questions concerns the sampling scheme: can one produce explicitly a set $\{x^{(1)}, \ldots, x^{(m)}\}$ of $m \asymp m(s, N, \varepsilon, \mathcal{X})$ evaluation points that makes (3) realizable? The probabilistic argument of [3, 11] 'only' shows the existence such a set. The second question concerns the recovery algorithm: can one devise an implementable construction of the approximant $\widetilde{F}$ from the data $((x^{(1)}, y_1), \ldots, (x^{(m)}, y_m))$? The ingredients coming into play in [3, 11] are not quite practical.

The purpose of this note is to answer these questions when the assumption $F \in \text{Lip}$ gives way to some other assumptions. In Sections 2 and 3, we assume that the function F is linear and quadratic, respectively, and establish connections with the theory of compressive sensing, namely, with sparse

[1] The results of [3, 11] featured $\ln(N)$ instead of $\log(N/s)$, but they do also hold as written in (4)-(5), with log standing for the logarithm in base 2 throughout this note.

recovery and with bisparse and low-rank recovery, respectively. In both cases, we determine the order of the sample complexity for exact recovery of F, we point out that nonadaptive random sampling schemes are optimal, and we highlight some efficient recovery algorithms. The substance of this note resides in subsequent sections, where we assume that the function F is coordinatewise monotone. In Section 4, we show that the problem then reduces the problem of group testing, which features known differences between the adaptive and nonadaptive settings. The former setting is briefly treated in Section 5, while Section 6 deals with the former setting: we first recall useful properties of the sampling (testing) procedure, before presenting an explicit sampling scheme (borrowed from compressive sensing and seemingly new in group testing) that satisfies the so-called disjunctiveness property, and we conclude by proposing a simple recovery method proved to succeed under the disjunctiveness property.

2 Linearity Assumption

In this section, we assume that the function F, defined on $[-1,1]^N$ here, depends on at most s active variables and is known a priori to be linear, so that

$$(6) \qquad F(x) = \langle a, x \rangle \qquad \text{where } a \in \mathbb{R}^N \text{ is } s\text{-sparse}.$$

In this situation, the function F can be recovered exactly. Indeed, the objective is now similar to recovering the s-sparse vector $a \in \mathbb{R}^N$ from samples of the type $y_i = F(x^{(i)}) = \langle a, x^{(i)} \rangle$, i.e., from the vector

$$(7) \qquad y = Xa, \qquad X = \begin{bmatrix} {x^{(1)}}^\top \\ \hline \vdots \\ \hline {x^{(m)}}^\top \end{bmatrix} \in \mathbb{R}^{m \times N}.$$

This is exactly the standard compressive sensing problem, which has been extensively studied in the last fifteen years or so, see e.g. [5] and the references therein. In a nutshell, the minimal number of samples to achieve exact recovery of a is of the order of $s \log(N/s)$ (or simply $2s$ if one is not bothered by the unstability of the recovery). Explicit sampling schemes working in this optimal regime are not known, but nonadaptive random sampling schemes do the job. For instance, if the $x^{(i)}$ are chosen independently uniformly at random in $[-1,1]$ (so they are independent zero-mean subgaussian random variables) and if $m \asymp s \log(N/s)$, then the matrix $X/\sqrt{m}$ satisfies with high probability the so-called restricted isometry property of order proportional to s. This property allows one to recover a via a number of efficient algorithms, such as ℓ_1-minimization, iterative hard thresholding (IHT), hard thresholding pursuit (HTP), orthogonal matching pursuit (OMP), compressive sampling matching pursuit (CoSaMP), among others.

3 Quadraticity Assumption

In this section, we assume that the function F, again defined on $[-1,1]$, depends on at most s active variables and is known a priori to be the sum of few quadratic terms, say

$$F(x) = \sum_{k=1}^{r} \langle a_k, x \rangle^2 = x^\top A x, \qquad \text{where } A = \sum_{k=1}^{r} a_k a_k^\top \in \mathbb{R}^{N \times N}. \tag{8}$$

The symmetric matrix A has rank $r \le s$ and is s-bisparse, in the sense that it is supported on some $S \times S$ with $|S| \le s$. Thus, the problem reduces to the recovery of jointly low-rank and bisparse matrices. This problem has recently been studied in various places, for instance in [6]. There, as an extension of the case $r = 1$ treated in [8], it was shown that $m \asymp rs\log(N/s)$ nonadaptive random samples suffice to guarantee, with high probability, the exact recovery of symmetric s-bisparse matrices $A \in \mathbb{R}^{N\times N}$ satisfying $\mathrm{rank}(A) \le r$. These random samples are factorized in nature, namely one can take $x^{(1)}, \ldots, x^{(m)} \in [-1,1]^N$ with entries $x_j^{(i)} = p^{-1}(\eta_{j,1} z_1^{(i)} + \cdots + \eta_{j,p} z_p^{(i)})$, where $p \asymp s\log(N/s)$, the $\eta_{j,k}$ are independent Rademacher random variables, and the $z_k^{(i)}$ are independent random variables uniformly distributed in $[-1,1]$. The recovery algorithm is adapted to this sampling strategy: it proceeds in two steps, both of which being practical, hence is overall efficiently implementable.

4 Monotonicity Assumption — Reduction to Group Testing

In the rest of this note, we assume that F, defined on $[0,1]^N$, is coordinatewise monotone, i.e., that the map $\xi \in [0,1] \mapsto F(x_1, \ldots, x_{j-1}, \xi, x_{j+1}, \ldots, x_N)$ is increasing or decreasing, when not constant, for any $j \in [1:N]$ and any $x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_N \in [0,1]$. Replacing the variable ξ by $1-\xi$ if necessary, we can in fact assume that F is coordinatewise increasing. Under this monotonicity assumption, with $S = \{j_1, \ldots, j_s\}$ denoting the set of active variables and with $\mathbb{1}_R$, $R \subseteq [1:N]$, denoting the vector with entries equal to one on R and to zero outside R, we make the crucial observation that

$$S \cap R \neq \emptyset \iff F(\mathbb{1}_R) > F(0). \tag{9}$$

Thus, by evaluating F at $x^{(0)} = 0$ and at $x^{(1)} = \mathbb{1}_{R_1}, \ldots, x^{(m)} = \mathbb{1}_{R_m}$ for some $R_1, \ldots, R_m \subseteq [1:N]$, we can test if each R_i contains an active variable or not. Based on this binary information, we aim at recovering the set S of active variables — once this is done, the approximation of the high-dimensional function F reduces to approximation of the low-dimensional function f. On this account, we shall only be concerned with the recovery of the set of active variables from now on.

The problem at hand is then exactly the one considered in group testing. As a matter of historical fact, the problem considered in this note shaped the initial development of the theory of group testing, see [9]. We refer the reader to [4] for a detailed account of this theory and only give a condensed description here. In group testing terminology, among a set of items, one tries to identify which ones are defective by testing the items in groups. A test, identified as the subset of items being tested, is positive if it contains at least one defective item. The traditional notation n for the number of items and d for the number of defective ones is replaced here by N and s, respectively, since an item corresponds to a variable and a defective one to an active variable. Group testing comes in two flavors: a relatively easy one where the tests are adaptive and a more delicate one where the tests are nonadaptive. These two flavors are differentiated by the number of tests required to successfully identify the defective objects — in other words, by their sample complexities. This difference between adaptive and nonadaptive settings will of course be reflected in the recovery of active variables, as we will see in the next two sections. These sections are not only concerned with sample complexities, though, but mainly with the two critical points mentioned in the introduction: explicit sampling schemes and efficient recovery algorithms. In keeping with the philosophy of making such considerations practical, we have implemented the proposed sampling schemes and recovery algorithms in a MATLAB reproducible file to accompany this note. The reproducible can be downloaded from the author's webpage.

5 Monotonicity Assumption — Adaptive Setting

When adaptive sampling is allowed, it is well-known and rather easy to see that the actice variables can be recovered with roughly $2s\log(N)$ points evaluations, which is the optimal order according to the often-invoked information-theoretic lower bound. For completeness, we briefly explain how to achieve this number of point evaluations via a simple splitting strategy.

Let $n := \lceil \log(N) \rceil$, so that $[1:N] \subseteq [1:2^n]$, and let us view F as a function of 2^n variables if necessary. In a first round, we split $R^{(0)} = [1:2^n]$ into $R^{(1)}_{\rm left} = [1:2^{n-1}]$ and $R^{(1)}_{\rm right} = [2^{n-1}+1:2^n]$ and test both $F(\mathbb{1}_{R^{(1)}_{\rm left}}) > F(0)$ and $F(\mathbb{1}_{R^{(1)}_{\rm right}}) > F(0)$. At least one of these tests is positive — let $R^{(1)}$ denote such a positive test and keep in mind that it has size 2^{n-1}. In a second round, we split $R^{(1)}$ into two sets $R^{(2)}_{\rm left}$ and $R^{(2)}_{\rm right}$ of size 2^{n-2} and we test both $F(\mathbb{1}_{R^{(2)}_{\rm left}}) > F(0)$ and $F(\mathbb{1}_{R^{(2)}_{\rm right}}) > F(0)$. At least one of these tests is positive — let $R^{(2)}$ denote such a positive test and keep in mind that it has size 2^{n-2}. We continue in this way until an nth round, which provides a positive test $R^{(n)}$ of size $2^{n-n} = 1$, i.e., an active variable that we denote by j_1. Remark that $1+2n$ point evaluations were used so far. We now repeat the whole procedure by removing the variable x_{j_1} from consideration and by performing the n rounds above to find another active variable j_2 using $2n$ more point evaluations. The procedure finishes after at most s iterations, totaling in a number of point evaluations at most $1 + s \times 2n \approx 2s\log(N)$. We note in passing that the procedure runs without having to know an upper estimate for s beforehand.

6 Monotonicity Assumption — Nonadaptive Setting

When adaptive sampling is disallowed, the order of the sample complexity is known to increase to $s^2 \log(N)$. In contrast to Sections 2 and 3, however, we can provide an explicit sampling scheme in this optimal regime of parameters, together with a practical recovery algorithm. We outline our contribution in the following statement, which could have been phrased solely in the language of group testing.

Theorem 1. The active variables of any function of few coordinate variables satisfying (1) with a coordinatewise increasing f can be recovered from $m \asymp s^2 \log^2(N)$ nonadaptive samples. The evaluation points are $x^{(0)} = 0$ and $x^{(i)} = \mathbb{1}_{R_i}$, $i \in [1:m]$, where the $R_i \subseteq [1:N]$ are identified with the rows of the test matrix explicitly given in (14). The recovery is efficiently performed by solving the linear feasibility problem (18).

We prove the three parts of this statement in the following three subsections. Precisely, we first recall in Subsection 6.1 the notion of test matrix and some properties surrounding it, in particular disjunctiveness, then we prove in Subsection 6.2 that a matrix found in the compressive sensing literature is indeed disjunct, and we finally establish in Subsection 6.3 that disjunctiveness ensures that a solution of the feasibility problem yields the correct locations of the active variables.

6.1 Test matrices and their properties

A nonadaptive testing procedure is usually summarized by a so-called test matrix $T \in \mathbb{R}^{m \times N}$ with entries in $\{0,1\}$. Each row $T_{i,:} \in \{0,1\}^N$ is identified with the subset R_i of $[1:N]$ defining the ith test, in the sense that

$$T_{i,j} = 1 \iff j \in R_i. \tag{10}$$

Each column $T_{:,j} \in \{0,1\}^m$ is identified to a subset C_j of $[1:m]$ in a similar fashion. The outcome of the ith test is negative if and only if $S \cap R_i = \emptyset$, which is equivalent to $(T\mathbb{1}_S)_i = 0$. Thus, the output of the whole testing procedure on $S \subseteq [1:N]$ is the binary vector $y = \chi(T\mathbb{1}_S)$ defined by

$$y_i = \begin{cases} 0, & \text{if } (T\mathbb{1}_S)_i = 0, \\ 1, & \text{if } (T\mathbb{1}_S)_i \geq 1. \end{cases} \tag{11}$$

Since the vector $\chi(T\mathbb{1}_S)$ can be identified with the set $\cup_{j \in S} C_j$ under the above formalism, the primary requisite of group testing, i.e., making sure that every $S \subseteq [1:N]$ of size exactly s, respectively at most s, can be deduced from $\chi(T\mathbb{1}_S)$ is seen to be equivalent to the property of

s-separability, respectively **$\bar{\text{s}}$-separability:** the sets $\cup_{j \in S} C_j$ are all distinct when S runs over all subsets of $[1:N]$ with $|S| = s$, respectively $|S| \leq s$.

Another central property of the testing procedure is that of

s-disjunctiveness: there are no $S \subseteq [1:N]$ with $|S| = s$ and $\ell \notin S$ such that $C_\ell \subseteq \cup_{j \in S} C_j$.

These two properties are closely related, since s-disjunctiveness implies $\overline{s}$-separability and conversely $\overline{s}$-separability implies $(s-1)$-disjunctiveness, according to [4, Lemma 7.2.2 and Lemma 7.2.4]. Another closely related property of the testing procedure will come into play, namely that of

s-strong selectivity: for any $S \subseteq [1:N]$ with $|S| \leq s$ and any $\ell \in S$, there exists $i \in [1:m]$ such that $S \cap R_i = \{\ell\}$.

This property appeared in [10], which announced that $(s+1)$-strong selectivity implies $\overline{s}$-separability and conversely that $\overline{s}$-separability implies s-strong selectivity. To tie all these properties together, we state the following (maybe well-known) chain of implications, which is justified in the appendix. The implication (a) will be used in the proof of Proposition 4.

Proposition 2. Separability, disjunctiveness, and strong selectivity are almost equivalent, since

$$s\text{-disjunct} \underset{(a)}{\Rightarrow} (s+1)\text{-strongly selective} \underset{(b)}{\Rightarrow} \overline{s}\text{-separable} \underset{(c)}{\Rightarrow} (s-1)\text{-disjunct} \underset{(d)}{\Rightarrow} s\text{-strongly selective.}$$

6.2 Sampling scheme

In view the previous considerations, we are now interested in s-disjunct testing procedures. We first examine the minimal number $m(s,N)$ of tests needed to achieve s-disjunctiveness. On the one hand, this number is known to be bounded from below as

$$m(s,N) \geq c \, \frac{s^2}{\log(s)} \log(N), \tag{12}$$

see [7] for a very short proof. On the other hand, it is also bounded from above as

$$m(s,N) \leq C \, s^2 \log(N), \tag{13}$$

as can be derived by a probabilistic argument presented in the appendix. Next, we wish to exhibit explicit testing procedures that achieve s-disjunctiveness with the opimal number of tests, possibly up to logarithmic factors. Several constructions exist. We highlight one of them for its simplicity. It arose from another context in [2] and seems uncommon in the group testing literature.

Proposition 3. Given a prime number $p \geq 3$ and an integer $d \leq p$, let $m := p^2$ and $N := p^d$. The $m \times N$ test matrix T with rows indexed by couples $(i,j) \in \mathbb{Z}_p \times \mathbb{Z}_p$, columns indexed by polynomials g of degree less than d over $\mathbb{Z}_p$, and explicitly defined by

$$T_{(i,j),g} = \begin{cases} 1, & \text{if } g(i) = j \pmod{p}, \\ 0, & \text{if } g(i) \neq j \pmod{p}, \end{cases} \tag{14}$$

is s-disjunct as soon as $m \geq s^2 \log^2(N)$.

Proof. It is known (and justified again in the appendix) that the matrix T has a coherence μ satisfying

$$\mu := \max_{g \neq h} \frac{|\langle T_{:,g}, T_{:,h} \rangle|}{\|T_{:,g}\|_2 \|T_{:,h}\|_2} \leq \frac{d-1}{p}, \tag{15}$$

which also reads $|\langle T_{:,g}, T_{:,h} \rangle| \leq d-1$ for all $g \neq h$, since the squared ℓ_2-norm of each column, which counts the number of 1's in that column, equals p. We now essentially reproduces the argument of [4, Lemma 7.3.2] showing that T is s-disjunct whenever $s < 1/\mu$. Suppose that T is not s-disjunct, i.e., that there exist $S \subseteq [1:N]$ with $|S| = s$ and $h \notin S$ such that $C_h \subseteq \cup_{g \in S} C_g$. This implies that $T_{:,h} \leq \sum_{g \in S} T_{:,g}$, the inequality being understood entrywise. Taking the inner product with $T_{:,h}$, we obtain

$$p = \|T_{:,h}\|_2^2 = \sum_{g \in S} \langle T_{:,g}, T_{:,h} \rangle \leq s\,(d-1). \tag{16}$$

This is evidently impossible when $s \leq p/d$. Thus, s-disjunctiveness of T is achieved as soon as $p > s\,d$, i.e., $m = p^2 > s^2 d^2 = s^2 (\log(N)/\log(p))^2$, which does occurs when $m \geq s^2 \log^2(N)$. $\square$

6.3 Recovery algorithm

We conclude this note by highlighting a practical method for recovering the set of active variables. The method simply consists in solving a linear feasibility problem. It is worth noting that knowing an upper estimate for s is not required, unlike e.g. the LiPo algorithm of [1]. Moreover, our theoretical analysis is carried out beyond random test matrices and our guarantees hold with certainty rather than with small error probability. The result stated in Proposition 4 below implies the theorem announced at the beginning of this section when combined with Proposition 3 above.

Proposition 4. Let $T \in \{0,1\}^{m \times N}$ be the test matrix of an s-disjunct procedure. Given a set $S \subseteq [1:N]$ with $|S| \leq s$ being acquired through $y = \chi(T \mathbb{1}_S) \in \{0,1\}^m$, define

$$I_0 := \{i \in [1:m] : y_i = 0\} \qquad \text{and} \qquad I_1 := \{i \in [1:m] : y_i = 1\}. \tag{17}$$

The set S is recovered as the support of a solution to

$$\text{find } x \in \mathbb{R}^N \qquad \text{subject to } x \geq 0, \quad (Tx)_{I_0} = 0, \quad (Tx)_{I_1} \geq 1. \tag{18}$$

Proof. We first note that the problem (18) is feasible since the vector $\mathbb{1}_S$ satisfies the constraints. Let now x' denote any solution to (18) and let $S' := \{j \in [1:N] : x'_j > 0\}$ denote its support. We observe that S' also has $y \in \{0,1\}^m$ as testing outcome, in view of

$$(T\mathbb{1}_{S'})_i = 0 \iff S' \cap R_i = \emptyset \iff (Tx')_i = 0 \iff i \in I_0 \iff y_i = 0. \tag{19}$$

Therefore, if we know that S' has size at most s, we can conclude that $S' = S$ using $\overline{s}$-separability, which follows from s-disjunctiveness. To justify that $|S'| \leq s$, we shall prove that $S' \subseteq S$. Suppose on the contrary that there exists $\ell \in S'$, $\ell \notin S$. By (a), the testing procedure is $(s+1)$-strongly selective, so there exists $i \in [1:m]$ such that $(S \cup \{\ell\}) \cap R_i = \{\ell\}$. This means that $\ell \in R_i$ and that $S \cap R_i = \emptyset$. The condition $\ell \in R_i$ implies that $(Tx')_i \geq x'_\ell > 0$, so $i \notin I_0$, while the condition $S \cap R_i = \emptyset$ implies that $(T\mathbb{1}_S)_i = 0$, i.e., $i \in I_0$. This contradiction shows that $S' \subseteq S$ and hence concludes the proof. $\square$

References

[1] C. L. Chan, S. Jaggi, V. Saligrama, and S. Agnihotri. *Non-adaptive group testing: explicit bounds and novel algorithms.* IEEE Trans. Info. Theory 60.5 (2014): 3019–3035.

[2] R. DeVore. *Deterministic constructions of compressed sensing matrices.* Journal of Complexity 23.4-6 (2007): 918–925.

[3] R. DeVore, G. Petrova, and P. Wojtaszczyk. *Approximation of functions of few variables in high dimensions.* Constructive Approximation 33.1 (2011): 125–143.

[4] D.-Z. Du and F. K. Hwang. Combinatorial Group Testing and its Applications. Second edition, World Scientific, 2000.

[5] S. Foucart and H. Rauhut. A Mathematical Introduction to Compressive Sensing. Birkhäuser, 2013.

[6] S. Foucart, R. Gribonval, L. Jacques, and H. Rauhut. *Jointly low-rank and bisparse recovery: questions and partial answers.* arXiv preprint arXiv:1902.04731 (2019).

[7] Z. Füredi. *On r-cover-free families.* Journal of Combinatorial Theory, Series A 73.1 (1996): 172–173.

[8] M. Iwen, A. Viswanathan, and Y. Wang. *Robust sparse phase retrieval made easy.* Applied and Computational Harmonic Analysis 42.1 (2017): 135–142.

[9] C. H. Li. *A sequential method for screening experimental variables.* Journal of the American Statistical Association 57.298 (1962): 455–477.

[10] E. Porat and A. Rothschild. *Explicit nonadaptive combinatorial group testing schemes.* IEEE Trans. Info. Theory 57.12 (2011): 7982–7989.

[11] P. Wojtaszczyk. *Complexity of approximation of functions of few variables in high dimensions.* Journal of Complexity 27.2 (2011): 141–150.

Appendix. The supplementary material added here makes this note (almost) self-contained by proving a few statements appearing in the main text but not justified there.

Proof of Proposition 2. (a), hence (d): Consider $S \subseteq [1:N]$ with $|S| \le s+1$ and $\ell \in S$. By s-disjunctiveness, $C_\ell \not\subseteq \cup_{j \in S \setminus \{\ell\}} C_j$, so there is some $i \in C_\ell$ with $i \notin \cup_{j \in S \setminus \{\ell\}} C_j$. The fist condition is similar to $T_{i,\ell} = 1$, or $\ell \in R_i$; the second condition is similar to $T_{i,j} = 0$, or $j \notin R_i$, for all $j \in S \setminus \{\ell\}$, i.e., to $(S \setminus \{\ell\}) \cap R_i = \emptyset$. Now $\ell \in R_i$ and $(S \setminus \{\ell\}) \cap R_i = \emptyset$ also read $S \cap R_i = \{\ell\}$, which proves $(s+1)$-strong selectivity.

(b): Consider distinct $S, S' \subseteq [1:N]$ with $|S|, |S'| \le s$ and suppose that $\cup_{j \in S} C_j = \cup_{j \in S'} C_j =: C$. Assume without loss of generality that $S' \setminus S \neq \emptyset$ and pick $\ell \in S' \setminus S$. By $(s+1)$-strong selectivity, there is some $i \in [1:m]$ such that $(S \cup \{\ell\}) \cap R_i = \{\ell\}$. On the one hand, we have $\ell \in R_i$, i.e., $T_{i,\ell} = 1$, so $i \in C_\ell \subseteq \cup_{j \in S'} C_j = C$. On the other hand, for any $j \in S$, we have $j \notin R_i$, i.e., $T_{i,j} = 0$, or $i \notin C_j$, so that $i \notin \cup_{j \in S} C_j = C$. This is a contradiction.

(c) [4, Lemma 7.2.4]: Consider $S \subseteq [1:N]$ with $|S| \le s-1$, $\ell \notin S$, and suppose that $C_\ell \subseteq \cup_{j \in S} C_j$. Then $\cup_{j \in S} C_j = \cup_{j \in S \cup \{\ell\}} C_j$, which contradicts $\overline{s}$-separability. □

Proof of (13). The following probabilistic argument is a variation of [4, Theorem 8.1.3] under a random model guaranteeing the same number t of 1's per column of the test matrix. So let $C_1, \ldots, C_N$ be random subsets of $[1:m]$ with size $t := 3\, s \lceil \log(N) \rceil$, where $m = 2\, s\, t = 6\, s^2 \lceil \log(N) \rceil$. Fixing $S \subseteq [1:N]$ with $|S| \le s$, let $u \le s\, t$ denote the size of $\cup_{j \in S} C_j$. Fixing also $\ell \notin S$, we have

$$\mathbb{P}(C_\ell \subseteq \cup_{j \in S} C_j) = \frac{\binom{u}{t}}{\binom{m}{t}} = \frac{u(u-1)\cdots(u-t+1)}{m(m-1)\cdots(m-t+1)} \le \left(\frac{u}{m}\right)^t \le \left(\frac{s\,t}{m}\right)^t \le \left(\frac{1}{2}\right)^t.$$

Then, by a union bound over all S and ℓ, we obtain

$$\mathbb{P}(\exists S, \exists \ell : C_\ell \subseteq \cup_{j \in S} C_j) \le \binom{N}{s}(N-s)\frac{1}{2^t} \le \frac{N^{s+1}}{2^t} \le \frac{N^{s+1}}{N^{3s}} \le \frac{1}{N^s}.$$

This proves that s-disjunctiveness is achieved with probability at least $1 - N^{-s}$. □

The coherence of DeVore's matrix. With T being the test matrix defined in (14), we have already noticed that $\|T_{g,:}\|_2^2 = p$ for any polynomial g of degree less than d over $\mathbb{Z}_p$. If h is another such polynomial, then

$$(20) \qquad |\langle T_{:,g}, T_{:,h} \rangle| = \sum_{(i,j) \in \mathbb{Z}_p \times \mathbb{Z}_p} T_{(i,j),g} T_{(i,j),h} = \sum_{(i,j) \in \mathbb{Z}_p \times \mathbb{Z}_p} \mathbb{1}_{\{g(i) = j \pmod p\}} \mathbb{1}_{\{h(i) = j \pmod p\}}$$
$$= \sum_{i \in \mathbb{Z}_p} \mathbb{1}_{\{(g-h)(i) = 0 \pmod p\}} \le d-1,$$

where the last inequality simply reflects the fact that the number of roots in $\mathbb{Z}_p$ of a nonzero polynomial is at most equal to its degree. This fully justifies (15). □