Adaptive Group Testing as Channel Coding with Feedback

Matthew Aldridge
Heilbronn Institute of Mathematical Research
School of Mathematics
University of Bristol, UK
m.aldridge@bristol.ac.uk

***Abstract*—Group testing is the combinatorial problem of identifying the defective items in a population by grouping items into test pools. Recently, nonadaptive group testing – where all the test pools must be decided on at the start – has been studied from an information theory point of view. Using techniques from channel coding, upper and lower bounds have been given on the number of tests required to accurately recover the defective set, even when the test outcomes can be noisy.**

In this paper, we give the first information theoretic result on adaptive group testing – where the outcome of previous tests can influence the makeup of future tests. We show that adaptive testing does not help much, as the number of tests required obeys the same lower bound as nonadaptive testing. Our proof uses similar techniques to the proof that feedback does not improve channel capacity.

I. Introduction

The problem of *group testing* concerns detecting the defective members of a set of items through the means of pooled tests. Group testing as a subject dates back to the work of Dorfman [1] in 1940s studying practical ways of testing soldiers' blood for syphilis, and has received much attention from combinatorialists and probabilists since.

The setup is as follows: Suppose we have a number of *items*, of which some are *defective*. To identify the defective items we could test each of the items individually for defectiveness. However, when the proportion of defective items is small, most of the tests will give negative results. A less wasteful method is to test *pools* of many items together at the same time. In the noise-free model, a pool gives a negative test outcome if it contains no defective items, and gives a positive outcome if it contains at least one defective item. (In Section II of this paper we consider models with noise.) After a number T of such pooled tests, it should be possible to deduce which items were defective.

Traditionally, group testing has been seen as a combinatorial problem. One aims to find a pooling strategy such that each possible defective set gives a different sequence of outcomes. This gives a zero error probability, and one is interested in how small T can be made. (See, for example, the textbook of Du and Hwang [2] for more details on the combinatorial approach to group testing.)

Group testing splits into two main types:

- **Nonadaptive group testing**, where the entire pooling strategy is decided on beforehand;
- **Adaptive (or sequential) group testing**, where the outcomes of previous tests can be used to influence the makeup of future pools.

Recently, new results on nonadaptive group testing with arbitrarily small probability of error have been derived using information theoretic techniques. A recent paper of Atia and Saligrama [3] proves bounds on T using techniques similar to the proof of Shannon's channel coding theorem [4].

In this paper, we study adaptive group testing using information theoretic techniques. Clearly adaptive group testing cannot be more difficult than nonadaptive testing. We show that it is not much easier either.

Specifically, Theorem 2 shows that the number of tests required for adaptive group testing is no more than that required for nonadaptive testing, but is still greater than the Atia–Saligrama lower bound. The result is obtained by techniques similar to Shannon's proof that feedback does not improve capacity for channel coding [5]. As far as we are aware, this is the first information theoretic result for adaptive group testing.

In combinatorial zero-error group testing using the noise-free model, adaptive testing certainly is an improvement. Only $O(K \log N)$ are needed for adaptive testing, whereas at least $\Omega(K^2 \log N/ \log K)$ are required for nonadaptive testing [2]. We note that this is similar to the case of zero-error channel coding, where feedback may improve the zero-error capacity [5].

The structure of this paper is as follows. In Section II we outline the information theoretic approach to nonadaptive group testing, fixing notation, and reviewing the work of Atia and Saligrama [3] and others. In Section III we briefly review Shannon's result on channel coding with feedback before stating and proving our main theorem (Theorem 2). We conclude with Section IV.

II. The information theoretic approach to nonadaptive group testing

First we fix some notation. We have N items, of which a subset $\mathcal{K}$ of size K is defective. We wish to accurately estimate the defective set from T tests. A pooling strategy can be defined by a *testing matrix* $\mathsf{X} = (x_{it}) \in \{\mathtt{0}, \mathtt{1}\}^{N \times T}$, where $x_{it} = \mathtt{1}$ denotes that item i is in the pool for test t, and $x_{it} = \mathtt{0}$ denotes that it is not. Test t gives an output Y_t in some output alphabet $\mathcal{Y}$ (which is usually $\{\mathtt{0}, \mathtt{1}\}$ also). Then, given

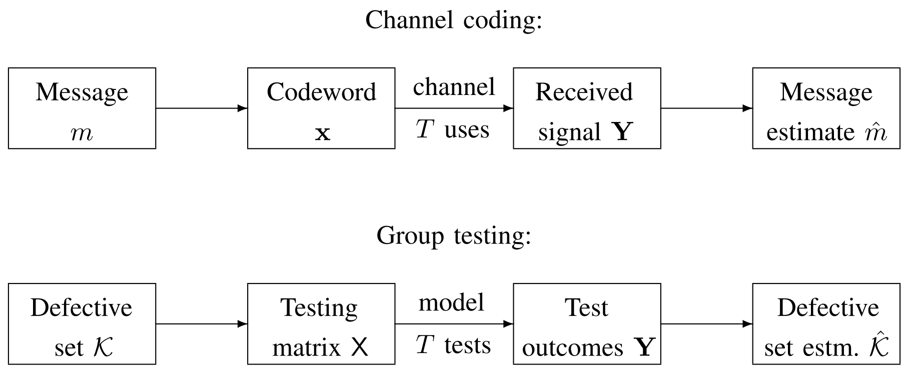

Fig. 1. A diagram illustrating the similarities between channel coding and group testing.

the test outcomes $\mathbf{Y} = (Y_t) \in \mathcal{Y}^T$, we make an estimate $\hat{\mathcal{K}} = \hat{\mathcal{K}}(\mathbf{Y})$ of the defective set. The average probability of error is ϵ.

Let $k_t = |\{i \in \mathcal{K} : x_{it} = \mathtt{1}\}|$ denote the number of defective items in test t. In the main *noise-free* case, $Y_t = \mathtt{0}$ (denoting a negative test outcome) if $k_t = 0$, and $Y_t = \mathtt{1}$ (denoting a positive test outcome) if $k_t \geq 1$.

We can also consider group testing with noise. Atia and Saligrama [3] consider two noise models:

- **Addition model**, where false positives occur with probability q. That is,

$$\text{if } k_t = 0, \qquad Y_t = \begin{cases} \mathtt{0} & \text{with probability } 1-q, \\ \mathtt{1} & \text{with probability } q; \end{cases}$$
$$\text{if } k_t \geq 1, \qquad Y_t = \mathtt{1}.$$

- **Dilution model**, where false negatives occur with probability u^{k_t}. That is,

$$\text{if } k_t = 0, \qquad Y_t = \mathtt{0};$$
$$\text{if } k_t \geq 1, \qquad Y_t = \begin{cases} \mathtt{0} & \text{with probability } u^{k_t}, \\ \mathtt{1} & \text{with probability } 1-u^{k_t}. \end{cases}$$

Sejdinovic and Johnson [6] considered a model where both addition and dilution errors can occur. Aldridge [7, Chapter 6] considered a class of models where only defects matter, in that the distribution of Y_t depends only on k_t (and not on how many nondefective items are in a test pool).

Group testing can be considered as being similar to channel coding. Here, the defective set takes the place of the message, the testing matrix is like the codebook, the test outcomes like the received signal. Then, like channel coding, we want to estimate the message/defective set using as few channel uses/tests as possible while keeping the error probability low. Figure 1 illustrates this.

Atia and Saligrama's main result was the following bounds on the number of tests required to accurately detect the defective set [3].

Theorem 1: Consider a group testing model where only defects matter. Let $T^*_{\text{NA}} = T^*_{\text{NA}}(N, K, \epsilon)$ be the minimum number of tests necessary to identify K defects among N items with error probability at most $\epsilon \neq 0, 1$. Then

$$\underline{T} + o(\log N) \leq T^*_{\text{NA}} \leq \overline{T} + o(\log N)$$

as $N \to \infty$, where

$$\overline{T} = \min_p \max_{\mathcal{L} \subset \mathcal{K}} \frac{\log \binom{N-K}{|\mathcal{L}|}\binom{K}{|\mathcal{L}|}}{I(\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}} : \mathbf{X}_{\mathcal{L}}, Y)}, \tag{1}$$

$$\underline{T} = \min_p \max_{\mathcal{L} \subset \mathcal{K}} \frac{\log \binom{N-|\mathcal{L}|}{K-|\mathcal{L}|}}{I(\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}} : \mathbf{X}_{\mathcal{L}}, Y)}. \tag{2}$$

Here, the X_i are IID $\text{Bern}(p)$, Y is related to $\mathbf{X}$ through the channel model, and I denotes mutual information. We have used the notation $\mathbf{X}_{\mathcal{L}} := (X_i : i \in \mathcal{L})$ and similar.

Atia and Saligrama [3] proved the theorem for the noise-free, addition and dilution models. Aldridge [7, Chapter 6.4] pointed out that their analysis extends to any model where only defects matter. Atia and Saligrama [3] also extended their result to the $K = o(N)$ asymptotic regime.

The proof of the upper bound is similar to Gallager's proof [8] of the direct part Shannon's channel coding theorem [4]. Test pools designed at random, with $X_{it} = \mathtt{1}$ with probability p and $X_{it} = \mathtt{0}$ with probability $1-p$, IID over i and t. Estimation of the defective set is done on a maximum likelihood basis, in that $\hat{K}$ is chosen to maximise the probabilty of the outcome $\mathbf{Y}$ given the testing matrix X.

The proof of the lower bound resembles the converse part of Shannon's theorem (see for example [9, Section 7.9]), where Fano's inequality bounds the error probability. Unfortunately, unlike in Shannon's theorem, we are not so lucky that the upper and lower bounds asymptotically coincide, although they are close up to a logarithmic factor in N.

There has been other recent work on nonadaptive information theoretic group testing. Sejdinovic and Johnson [6] gave accurate asymptotic expressions for $\overline{T}$ for the noise-free, addition and dilution models. Cheraghchi et al [10] considered group testing when the makeup of the pools is constrained by a graphical structure. Numerous authors [6], [11], [12], [13] have used modern decoding algorithms on nonadaptive group testing simulations.

Some similar work has occured in the compressed sensing community; see the survey of Malyutov [14].

III. Adaptive group testing

In adaptive group testing, the makeup of a testing pool can depend on the outcomes of earlier tests, so

$$x_{it} = x_{it}(Y_1, \dots, Y_{t-1}).$$

This is similar to channel coding with feedback, where future inputs to the channel can depend on past outputs. Shannon proved that (perhaps surprisingly) feedback does not improve the capacity of a single-user channel [5]. Since a transmitter could choose not to use the feedback, it's clear that the capacity with feedback is at least as high as the capacity without. However by being more careful with Fano's inequality in the proof of the converse, it can apply to the case of feedback also. See [9, Section 7.12], for example, for a detailed proof.

Our result proceeds similarly. Due to the non-tightness of the bounds on testing in the nonadaptive case, we will not be able to show that adaptive group testing requires the *same* number of tests as nonadaptive testing, but we will be able to show that it obeys the same lower bound and requires no more tests than the nonadaptive case.

The lack of much improvement due to adaptive testing may initially seem surprising. However, the analogy with Shannon's feedback result explains why we should in fact expect this.

We emphasise that our theorem holds not only for the noise-free model, but also for the dilution and addition models, and any model where only defects matter.

Theorem 2: Consider a group testing model where only defects matter. Let T^*_{NA} and T^*_{A} (dependent on N, K and ϵ) be the minimum number of tests necessary to identify K defects among N items with error probability at most $\epsilon \neq 0, 1$ for nonadaptive and adaptive group testing respectively. Then, as $N \to \infty$, we have the inequalities

$$\underline{T} - o(\log N) \leq T^*_{\text{A}} \leq T^*_{\text{NA}} \leq \overline{T} + o(\log N)$$

where $\underline{T}$ and $\overline{T}$ are as in (1) and (2).

Proof: The third inequality is part of Theorem 1. The second inequality is trivial, as nonadaptive group testing is merely a special case of adaptive group testing where the tester chooses to ignore the information of previous test results.

To prove the first inequality, we adapt Atia and Saligrama's proof of converse part of Theorem 1 [3], and Shannon's proof that feedback fails to improve channel capacity [5], as exposited by Cover and Thomas [9, Theorem 7.12.1].

Choose a set of items $\mathcal{L}$ of size $|\mathcal{L}|$ uniformly at random from $\{1, 2, \dots, N\}$, and choose $\mathcal{K}$ of size K uniformly at random from sets containing $\mathcal{L}$.

Suppose a genie reveals to us the $|\mathcal{L}|$ defective items $\mathcal{L} \subset \mathcal{K}$, leaving us to work out the remaining $K - |\mathcal{L}|$ defective items. Given $\mathcal{L}$, there are $\binom{N-|\mathcal{L}|}{K-|\mathcal{L}|}$ equally likely choices of the random $\mathcal{K}$, so

$$H(\mathcal{K} \mid \mathcal{L}) = \log \binom{N - |\mathcal{L}|}{K - |\mathcal{L}|}. \tag{3}$$

Using a standard identity we can rewrite (3) as

$$\log \binom{N - |\mathcal{L}|}{K - |\mathcal{L}|} = H(\mathcal{K} \mid \hat{\mathcal{K}}, \mathcal{L}) + I(\mathcal{K} : \hat{\mathcal{K}} \mid \mathcal{L}). \tag{4}$$

We can now use Fano's inequality (see for example [9, Theorem 2.10.1]) to bound the conditional entropy term in (4) in terms of the error probability ϵ. Specifically, we have

$$H(\mathcal{K} \mid \hat{\mathcal{K}}, \mathcal{L}) \leq 1 + \epsilon \log \binom{N - |\mathcal{L}|}{K - |\mathcal{L}|}, \tag{5}$$

since there are again $\binom{N-|\mathcal{L}|}{K-|\mathcal{L}|}$ choices for $\mathcal{K}$. Substituting (5) into (4) gives

$$\log \binom{N - |\mathcal{L}|}{K - |\mathcal{L}|} \leq 1 + \epsilon \log \binom{N - |\mathcal{L}|}{K - |\mathcal{L}|} + I(\mathcal{K} : \hat{\mathcal{K}} \mid \mathcal{L}). \tag{6}$$

A series of standard information theory inequalities and identities show that the mutual information term in (6) can be bounded by

$$I(\mathcal{K} : \hat{\mathcal{K}} \mid \mathcal{L}) \leq T I(\mathbf{X}_{\mathcal{K} \setminus \mathcal{L}} : \mathbf{X}_{\mathcal{L}}, Y). \tag{7}$$

We relegate the elementary (but slightly long-winded) verification of (7) to the Appendix. Substituting (7) into (6) gives

$$\log \binom{N - |\mathcal{L}|}{K - |\mathcal{L}|} \leq 1 + \epsilon \log \binom{N - |\mathcal{L}|}{K - |\mathcal{L}|} + T I(\mathbf{X}_{\mathcal{K} \setminus \mathcal{L}} : \mathbf{X}_{\mathcal{L}}, Y). \tag{8}$$

Rearranging (8) to make ϵ the subject gives

$$\epsilon \geq 1 - T \frac{I(\mathbf{X}_{\mathcal{K} \setminus \mathcal{L}} : \mathbf{X}_{\mathcal{L}}, Y)}{\log \binom{N-|\mathcal{L}|}{K-|\mathcal{L}|}} - \frac{1}{\log \binom{N-|\mathcal{L}|}{K-|\mathcal{L}|}}. \tag{9}$$

Sending $N \to \infty$ in (9), it is clear that we require

$$\begin{aligned} T &\geq \frac{\log \binom{N-|\mathcal{L}|}{K-|\mathcal{L}|}}{I(\mathbf{X}_{\mathcal{K} \setminus \mathcal{L}} : \mathbf{X}_{\mathcal{L}}, Y)} \big(1 + o(1)\big) \\ &= \frac{\log \binom{N-|\mathcal{L}|}{K-|\mathcal{L}|}}{I(\mathbf{X}_{\mathcal{K} \setminus \mathcal{L}} : \mathbf{X}_{\mathcal{L}}, Y)} + o(\log N) \end{aligned} \tag{10}$$

to force the error probability to be arbitrarily small.

But (10) has to be true for all $\mathcal{L} \subset \mathcal{K}$, and we can optimise over the test inclusion parameter p. This gives the result. ■

IV. Conclusion

In conclusion, we have considered adaptive group testing for models where only defects matter with arbitrarily low probability of error. We have shown that adaptive testing requires no more tests than nonadaptive and, since it still obeys the Atia–Saligrama lower bound, cannot reduce the number of tests very much.

It remains an open question whether or not $T_{\text{A}} = T_{\text{NA}}$ (either exactly or in an asymptotic sense), or whether, as with zero-error testing for the noise-free model, there is a gap between T_{A} and T_{NA}.

A 'halfway house' between adaptive and nonadaptive testing is *S-stage testing*, where S test pools are decided on at a time. Clearly the number of tests required for S-stage testing lies between T^*_{NA} and T^*_{A} and is nondecreasing in S. We are not aware that this has received any attention from an information theoretic point of view.

Appendix. An inequality about mutual information

In this appendix we verify the claim (7), that

$$I(\mathcal{K} : \hat{\mathcal{K}} \mid \mathcal{L}) \leq T I(\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}} : \mathbf{X}_{\mathcal{L}}, Y),$$

which is required in the proof of Theorem 2.

We use the data processing inequality left-hand side of (7), to write

$$I(\mathcal{K} : \hat{\mathcal{K}} \mid \mathcal{L}) \leq I(\mathcal{K} : \mathbf{Y} \mid \mathcal{L}) = H(\mathbf{Y} \mid \mathcal{L}) - H(\mathbf{Y} \mid \mathcal{K}) \quad (11)$$

where the second equality in (11) is standard identity and we have used that $\mathcal{L} \cup \mathcal{K} = \mathcal{K}$.

We now unwrap the conditional entropy terms in (11) using the chain rule for entropy (see for example [9, Theorem 2.5.1]) and standard identities and inequalities. This gives

$$\begin{aligned}
I(\mathcal{K} : \hat{\mathcal{K}} \mid \mathcal{L}) &\leq \sum_{t=1}^{T} \big(H(Y_t \mid Y_1, \ldots, Y_{t-1}, \mathcal{L}) \\
&\qquad - H(Y_t \mid Y_1, \ldots, Y_{t-1}, \mathcal{K}) \big) &(12)\\
&= \sum_{t=1}^{T} \big(H(Y_t \mid Y_1, \ldots, Y_{t-1}, \mathcal{L}, \mathbf{X}_{\mathcal{L}t}) \\
&\qquad - H(Y_t \mid Y_1, \ldots, Y_{t-1}, \mathcal{K}, \mathbf{X}_{\mathcal{K}t}) \big) &(13)\\
&\leq \sum_{t=1}^{T} \big(H(Y_t \mid \mathbf{X}_{\mathcal{L}t}) \\
&\qquad - H(Y_t \mid Y_1, \ldots, Y_{t-1}, \mathcal{K}, \mathbf{X}_{\mathcal{K}t}) \big) &(14)\\
&= \sum_{t=1}^{T} \big(H(Y_t \mid \mathbf{X}_{\mathcal{L}t}) - H(Y_t \mid \mathbf{X}_{\mathcal{K}t}) \big), &(15)
\end{aligned}$$

where we have used the notation $\mathbf{X}_{\mathcal{L}t} := (X_{it} : i \in \mathcal{L})$ for fixed t and similar. We justify the above steps as follows:

(12) is from applying the chain rule to the right hand side of (11);
(13) is because $\mathbf{X}_{\mathcal{L}t}$ is a function of $Y_1, \ldots, Y_{t-1}$ and $\mathcal{L}$, and the same for $\mathcal{K}$;
(14) is because conditioning reduces entropy, so removing conditioning increases it;
(15) is because, conditional on $\mathbf{X}_{\mathcal{K}t}$, we know Y_t is independent of the previous outcomes $Y_1, \ldots, Y_{t-1}$ and the defective set $\mathcal{K}$.

But the term in the summand of (15) is precisely the mutual information

$$H(Y_t \mid \mathbf{X}_{\mathcal{L}t}) - H(Y_t \mid \mathbf{X}_{\mathcal{K}t}) = I(\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}\ t} : Y_t \mid \mathbf{X}_{\mathcal{L}t}), \quad (16)$$

and this is independent of t. Hence substituting (16) into (15) gives

$$\begin{aligned}
I(\mathcal{K} : \hat{\mathcal{K}} \mid \mathcal{L}) &\leq \sum_{t=1}^{T} I(\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}\ t} : Y_t \mid \mathbf{X}_{\mathcal{L}t}) \\
&= T I(\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}} : Y \mid \mathbf{X}_{\mathcal{L}}). &(17)
\end{aligned}$$

The mutual information term in (17) can alternatively be written as

$$\begin{aligned}
I(\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}} : Y \mid \mathbf{X}_{\mathcal{L}}) &= I(\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}} : \mathbf{X}_{\mathcal{L}}, Y) - I(\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}} : \mathbf{X}_{\mathcal{L}}) \\
&= I(\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}} : \mathbf{X}_{\mathcal{L}}, Y), &(18)
\end{aligned}$$

since $\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}}$ and $\mathbf{X}_{\mathcal{L}}$ are independent.

Substituting (18) into (17) gives

$$I(\mathcal{K} : \hat{\mathcal{K}} \mid \mathcal{L}) \leq T I(\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}} : \mathbf{X}_{\mathcal{L}}, Y),$$

thus verifying the claim (7).

Acknowledgments

The author was supported by the Heilbronn Institute for Mathematical Research and by the Engineering and Physical Sciences Research Council, via the University of Bristol Bridging the Gaps cross-disciplinary feasibility account (EP/H024786/1). The author thanks Oliver Johnson and Dino Sejdinovic for helpful discussions.

References

[1] R. Dorfman, "The detection of defective members of large populations," *Ann. Math. Statist.*, vol. 14, no. 4 pp. 436–440, 1943.
[2] D.-Z. Du and F. K. Hwang, *Combinatorial Group Testing and Its Applications*, 2nd edition, Series on Applied Mathematics, vol. 18, World Scientific, 2000.
[3] G. Atia and V. Saligrama, "Boolean compressed sensing and noisy group testing," submitted to *IEEE Trans. Inform. Theory*, arXiv:0907.1061v4 [cs.IT], 2010.
[4] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379-423 and 623–656, 1948.
[5] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inform. Theory*, vol. 2, no. 3, pp. 8–19, 1956.
[6] D. Sejdinovic and O. Johnson, "Note on noisy group testing: asymptotic bounds and belief propagation reconstruction," *Proc. Allerton Conf. on Commun., Control and Computing*, pp. 998–1003, 2010.
[7] M. Aldridge, "Interference mitigation in large random wireless networks," PhD thesis, University of Bristol, UK, 2011.
[8] R. Gallager, "A simple derivation of the coding theorem and applications," *IEEE Trans. Inform. Theory*, vol. 11, no. 1, pp. 3–18, 1965.
[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, second edition, Wiley–Interscience, 2006.
[10] M. Cheraghchi, A. Karbasi, S. Mohajer and V. Saligrama, "Graph-constrained group testing," *IEEE Trans. Inform. Theory*, vol. 58, no. 1, pp. 248–262, 2012.
[11] E. Porat and A. Rothschild, "Explicit nonadaptive combinatorial group testing schemes," *IEEE Trans. Inform. Theory*, vol. 57, no. 12, pp. 7982–7989, 2011.
[12] M. Cheraghchi, A. Hormati, A. Karbasi and M. Vetterli, "Group testing With probabilistic tests: theory, design and application," *IEEE Trans. Inform. Theory*, vol. 57, no. 10, pp. 7057–7067, 2011.
[13] C. L. Chan, S. Jaggi, V. Saligrama and S. Agnihotri, "Non-adaptive group testing: explicit bounds and novel algorithms," arXiv:1202.0206v1 [cs.IT], 2012.
[14] M. Malyutov, "Recovery of sparse active inputs in general systems: a review," *IEEE Region 8 SIBIRCON-2010*, pp. 15–22, 2010.