Forty-Eighth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
September 29 - October 1, 2010

Note on Noisy Group Testing: Asymptotic Bounds and Belief Propagation Reconstruction

Dino Sejdinovic, Oliver Johnson
School of Mathematics
University of Bristol
University Walk, Bristol BS8 1TW, UK
Email: {d.sejdinovic, o.johnson}@bristol.ac.uk

***Abstract*—An information theoretic perspective on group testing problems has recently been proposed by Atia and Saligrama, in order to characterise the optimal number of tests. Their results hold in the noiseless case, where only false positives occur, and where only false negatives occur. We extend their results to a model containing both false positives and false negatives, developing simple information theoretic bounds on the number of tests required. Based on these bounds, we obtain an improved order of convergence in the case of false negatives only. Since these results are based on (computationally infeasible) joint typicality decoding, we propose a belief propagation algorithm for the detection of defective items and compare its actual performance to the theoretical bounds.**

I. Introduction and Problem Outline

The idea of group testing was introduced during World War II in order to reduce the cost of large scale blood tests by pooling blood samples together [5]. Since then, it emerged as a promising approach in various applications, including multiple access communications and DNA clone library screening (cf. [6] and references therein).

The advent of compressed sensing (CS) has revived interest in group testing [3], as both problems involve the detection of a sparse high-dimensional signal via a small number of random measurements. However, the compressed sensing literature has mostly focussed on problems with measurement matrices with entries taken from distributions with densities. Group testing naturally belongs in a broader framework of discrete compressed sensing, where the entries are random integers, often just 0s and 1s. This framework of discrete compressed sensing includes wider applications such as genotyping [7]. An extension of the group testing problem to the scenario where pools must conform to the constraints imposed by a graph has also been studied recently [4].

An information theoretic approach to a noisy version of group testing was recently developed by Atia and Saligrama [1], [2]. We adopt much of their model and notation, which we will first briefly review. We will use group testing to identify K defective items within a larger collection of N items, by testing a pool of items at a time. Each test reveals whether the pool contains any defective items, i.e., the test result is *positive*, or 1, if at least one of the items in the pool is defective, and it is otherwise *negative*, or 0. However, we will allow two types of errors to occur in the testing.

1) **False positives**, where the test result is positive with probability q when the pool does not contain any defective items. In other words, the result of the test is ORed with Bernoulli(q) random variable.
2) **False negatives** - the indicator whether an item is defective is "diluted" with probability u. In other words, the result of the test will only be positive if the indicator of some defective item passes through a Z-channel successfully.

Note that the false negatives make the analysis significantly more complicated than for standard coding theoretic problems. This is because this model makes the noise dependent on the input, because a pool with more defective items will be less likely to return a false negative. Interestingly, our results here indicate that false negatives are, in a certain sense, easier to deal with than false positives, and we obtain an improved order of convergence on the number of achievable tests in the case of false negatives only.

Another way to describe the false negative process is that the test will be positive if the sum of the indicators, thinned in the sense of Rényi, is positive. In future work we hope to explore whether the bounds on entropy under thinning proved in [8] can improve or generalize the results of this paper.

We now formally describe the model which incorporates the presence of these two kinds of testing errors.

Definition 1. Let $\beta \in \{0,1\}^N$ be a column vector of indicators corresponding to the overall set of items, i.e., $\beta_i = 1$ iff item i is defective. We consider the case $w(\beta) = K \ll N$, where $w(\cdot)$ represents the Hamming weight. Furthermore, $\mathbf{X} = (x_{ti}) \in \{0,1\}^{T\times N}$ will denote the measurement matrix, s.t. $x_{ti} = 1$ iff item i is pooled in test t. We will restrict our attention to the case where $\mathbf{X}$ is composed of i.i.d. Bern(p) entries. A set of test results is a vector $y \in \{0,1\}^T$, where $y_t = 1$ means test t is positive. The outcome y_t of test t is given by (symbol $\wedge$ stands for Boolean matrix product):

$$y_t = (x_t \wedge \mathbf{D}_t \wedge \beta) \vee z_t,\ t \in \{1, 2, \ldots, T\}. \qquad (1)$$

Here x_t denotes the t-th row of $\mathbf{X}$, $\mathbf{D}_t \in \{0,1\}^{N\times N}$ is a diagonal matrix with i.i.d. Bern($1-u$) entries on the diagonal, independent of β and $\mathbf{X}$, and z_t is a $Bern(q)$ random variable, independent of all others.

This compact notation captures both the false positive test results which occur when $z_t = 1$, and the false negative test results which occur in the event that all the diagonal entries of $\mathbf{D}_t$ corresponding to the defective items in pool t equal zero.

In [1], Atia and Saligrama showed how group testing can be viewed analogously to channel coding by considering a set of K channels, with input $X_{(i)}$ and the pair $(X_{(K-i)}, Y)$ as their output, $i \in \{1, 2, \dots, K\}$. Here, $X_{(i)}$ stands for the i entries in the row of the measurement matrix corresponding to (any) i defective items, and Y is the test outcome (viewed as a random variable). Atia and Saligrama prove the following result:

Theorem 2. ([1], Theorem 3.2) *Consider the joint typicality decoder in the model of Definition 1, in the case where one or both of $q = 0$ or $u = 0$ (the model is noisefree, or allows false positives or false negatives, but not both). An achievable number of tests T_{typ} which allows perfect detection is given by:*

$$T_{typ} = \max_i \frac{\log_2 \binom{N-K}{i}\binom{K}{i}}{I(X_{(i)}; X_{(K-i)}, Y)}. \tag{2}$$

We now describe the structure of the remainder of this paper. We will consider the model of Definition 1, in the case where both q and u can be non-zero (both false positives and false negatives are allowed). For reasons of space, we will assume that an analogue of Equation (2) holds in the case $q > 0$ and $u > 0$. (To verify this requires a somewhat lengthy analysis of the probability that X and Y are jointly typical, as performed in the Appendix of [1]). This means that the key quantity of interest is the mutual information $I(X_{(i)}; X_{(K-i)}, Y)$. We will analyse this quantity in Section II, as in [1] deducing asymptotic results of the form $T_{typ} = \mathcal{O}(K \log(K(N-K))$. We also deduce that in the case where only false negatives occur, the number of tests required can be reduced by a factor of $\log K$.

In Section III, we propose a belief propagation algorithm for the detection of the defective items in noisy group testing. The analysis of Theorem 2 is based on the use of a joint typicality decoder, which is infeasible in practice, having prohibitive computational complexity in the limit of large K and N. Belief propagation offers a practically implementable alternative. Belief propagation has previously been used in the statistical physics community to address the problem of noiseless group testing and its relationship to the hitting set problem [9].

II. Asymptotic Bounds

In this section, we will derive sharp bounds on the mutual information of Equation (2):

Lemma 3. *The mutual information $I(X_{(i)}; X_{(K-i)}, Y)$ can be expressed in closed form as $I_1 + I_2$, where the "lead term" is:*

$$I_1 = i(1-q)(1-p+pu)^K \cdot \left(\frac{pu}{1-p+pu} \log_2 u - \log_2(1-p+pu) \right). \tag{3}$$

and the "error term" is:

$$I_2 = \frac{1}{\log 2} \sum_{j=2}^{\infty} \left[\frac{(1-q)^j}{j(j-1)} \cdot (1-p+pu^j)^K \left(1 - \left(\frac{(1-p+pu)^j}{1-p+pu^j} \right)^i \right) \right] \tag{4}$$

Proof: As in [1], we decompose

$$I(X_{(i)}; X_{(K-i)}, Y) = H(Y|X_{(K-i)}) - H(Y|X_{(K)}), \tag{5}$$

and consider the two terms separately. First, we set $V = X \wedge \mathbf{D} \wedge \beta$, and notice that

$$\mathbb{P}(Y=0|w(X_{(K)})=j) = \mathbb{P}(Z=0)\mathbb{P}(V=0|w(X_{(K)})=j) = (1-q)u^j, \tag{6}$$

which means that, writing $h(\cdot)$ for the binary entropy function,

$$H(Y|X_{(K)}) = \sum_{j=0}^{K} \binom{K}{j} p^j (1-p)^{K-j} h\left[(1-q)u^j\right]. \tag{7}$$

Similarly,

$$\begin{aligned}
&\mathbb{P}(Y=0|w(X_{(K-i)})=l) = \\
&\quad \mathbb{P}(Z=0)\mathbb{P}(V=0|w(X_{(K-i)})=l) = \\
&\quad (1-q)\sum_{j=l}^{l+i} \Big[\mathbb{P}(V=0|w(X_{(K)})=j) \cdot \\
&\quad \mathbb{P}(w(X_{(K)})=j|w(X_{(K-i)})=l)\Big] = \\
&\quad (1-q)\sum_{j=l}^{l+i} u^j \mathbb{P}(w(X_{(i)})=j-l) = \\
&\quad (1-q)u^l \sum_{j=0}^{i} u^j \binom{i}{j} p^j (1-p)^{i-j} = \\
&\quad (1-q)u^l(1-p+pu)^i,
\end{aligned} \tag{8}$$

whereby we obtain:

$$H(Y|X_{(K-i)}) = \sum_{l=0}^{K-i} \left[\binom{K-i}{l} p^l (1-p)^{K-i-l} \cdot h\left((1-q)u^l(1-p+pu)^i\right) \right]. \tag{9}$$

We substitute the expressions (7) and (9) into Equation (5), and analyse the resulting sum. We use an expansion of binary entropy as

$$h(\theta) = (-\theta \log_2 \theta) + \frac{1}{\log 2}\left(\theta - \sum_{j=2}^{\infty} \frac{\theta^j}{j(j-1)}\right), \quad (10)$$

with the first bracketed term becoming (3), and the remaining expression becoming (18). ■

Observe that for any $j \geq 2$, the function $g(p) = (1-p+pu^j) - (1-p+pu)^j \geq 0$. This means that the bracketed term in (18) is positive, and so as in [1], we could simply use the lower bound $I(X_{(i)}; X_{(K-i)}, Y) \geq I_1$ in Theorem 2. However, in many cases I_2 turns out to play a significant role, and so by including it in our analysis we obtain better bounds.

Lemma 4. *Choosing $p = (1-u)^{-1}/K$, we obtain that for constant i, q and u*

$$I_1 = \frac{i(1-u)(1-q)(u \log u - u + 1)}{Ke \log 2} + \mathcal{O}\left(\frac{1}{K^2}\right). \quad (11)$$

Proof: By setting $p = \alpha/K$, and expanding Equation (3) in powers of $1/K$, we obtain

$$I_1 = \frac{\alpha i (1-q) e^{\alpha(u-1)}(u \log u - u + 1)}{Ke \log 2} + \mathcal{O}\left(\frac{1}{K^2}\right). \quad (12)$$

We can optimize this expression over α by taking $\alpha = 1/(1-u)$, which justifies the heuristic choice of $p = 1/K$ to define the measurement matrices in [1]. ■

Note that in the cases $u = q = 0$ and $u = 0$ respectively we recover $i/(Ke\log 2)$ from (15) of [1] and $i(1-q)/(Ke \log 2)$ from (29) of [1]. In the case $q = 0$, this optimal choice of α gives us a lower bound of $i(1-u)/(2Ke\log 2)$, a slight improvement on (37) of [1].

Similarly, it can be shown that with $p = \alpha/K$,

$$\lim_{K\to\infty} K I_2 = \frac{\alpha i}{\log 2} \sum_{j=2}^{\infty}\left[\frac{(1-q)^j}{j(j-1)} \cdot e^{\alpha u^j - \alpha}(u^j + j - ju - 1)\right]. \quad (13)$$

It is easy to see that the series in (13) is converging for $q \neq 0$. Furthermore, by repeatedly using the sum

$$\sum_{j=2}^{\infty} \frac{\theta^j}{j(j-1)} = \theta + (1-\theta)\log(1-\theta), \quad (14)$$

and the obvious inequality $0 \leq u^j \leq u^2$, for $j \geq 2$, we obtain:

$$\frac{\alpha e^{-\alpha} i}{\log 2} C_{q,u} \leq \lim_{K\to\infty} K I_2 \leq \frac{\alpha e^{-\alpha} i}{\log 2} e^{\alpha u^2} C_{q,u}, \quad (15)$$

where

$$C_{q,u} = q - (1-u+qu)(1 + \log q - \log(1-u+qu)). \quad (16)$$

Notice that $C_{q,u} = \infty$ when $q = 0$. This suggests that I_2 is of a larger order in this case. Indeed, the following Lemma holds:

Lemma 5. *In case $q = 0$, $I_2 = \mathcal{O}(\frac{\log K}{K})$. In particular,*

$$\frac{\alpha e^{-\alpha} i}{\log 2}(1-u) \leq \lim_{K\to\infty} \frac{K}{\log K} I_2 \leq \frac{\alpha e^{-\alpha} i}{\log 2} e^{\alpha u^2}(1-u). \quad (17)$$

Proof: In

$$I_2 = \frac{1}{\log 2}\sum_{j=2}^{\infty}\left[\frac{(1-p+pu^j)^K}{j(j-1)} \cdot \left(1 - \left(\frac{(1-p+pu)^j}{1-p+pu^j}\right)^i\right)\right] \quad (18)$$

we notice that $(1-p+pu^j)^K \uparrow e^{-\alpha(1-u^j)} \leq e^{-\alpha(1-u^2)}$, as $K \to \infty$, whereas $(1-p+pu^j)^K \geq (1-p)^K$. Therefore, by applying (14) with $\theta = (1-p+pu)^i$, we obtain that $\forall i, K$,

$$\frac{(1-p)^K}{\log 2}\left[1 - \frac{(1-p+pu)^i}{(1-p)^i} + \frac{(1-(1-p+pu)^i)\log(1-(1-p+pu)^i)}{(1-p)^i}\right] \leq$$
$$\leq I_2 \leq$$
$$\leq \frac{e^{-\alpha(1-u^2)}}{\log 2}\left[1 - \frac{(1-p+pu)^i}{(1-p+pu^2)^i} + \frac{(1-(1-p+pu)^i)\log(1-(1-p+pu)^i)}{(1-p+pu^2)^i}\right]. \quad (19)$$

Now, by developing both sides in powers of $1/K$,

$$\frac{\alpha e^{-\alpha} i}{\log 2} \cdot \left(\frac{(1-u)\left[\log K - \log(\alpha i - \alpha i u)\right] - u}{K}\right) + \mathcal{O}\left(\frac{1}{K^2}\right) \leq$$
$$\leq I_2 \leq$$
$$\leq \frac{\alpha e^{-\alpha(1-u^2)} i}{\log 2} \cdot \left(\frac{(1-u)\left[\log K - \log(\alpha i - \alpha i u)\right] - u + u^2}{K}\right) + \mathcal{O}\left(\frac{1}{K^2}\right), \quad (20)$$

which proves the claim. ■

Theorem 6. *Assuming that an equivalent of Theorem 2 holds in the general case, using Lemmas 3, 4 and 5, we deduce:*

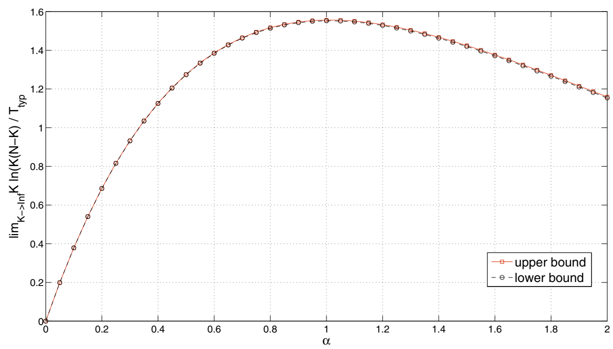

Figure 1. The bounds on the constant in the asymptotic estimate $T_{typ} = \mathcal{O}(K\log(K(N-K))$ for $u = 0.05$, $q = 0.01$.

(i) For any $q > 0$, $u \geq 0$, as $K \to \infty$,

$$T_{typ} = \mathcal{O}(K\log(K(N-K)). \quad (21)$$

In particular,

$$\alpha e^{-\alpha} C_{q,u} \leq \lim_{K\to\infty}\left(\frac{K\log(K(N-K))}{T_{typ}} - \alpha\left[e^{-\alpha(1-u)}(1-q)(u\log u - u + 1)\right]\right) \leq \alpha e^{-\alpha(1-u^2)} C_{q,u}. \quad (22)$$

(ii) For $q = 0$, and any $u \geq 0$, as $K \to \infty$,

$$T_{typ} = \mathcal{O}\left(K(1 + \frac{\log(N-K)}{\log K})\right). \quad (23)$$

In particular,

$$\alpha e^{-\alpha}(1-u) \leq \lim_{K\to\infty}\frac{K\log(K(N-K))}{T_{typ}\log K} \leq \alpha e^{-\alpha(1-u^2)}(1-u). \quad (24)$$

Notice that in the case of noiseless group testing, i.e., $u = 0$, $q = 0$, we arrive at the exact asymptotic expressions for T_{typ}:

$$T_{typ} = eK(1 + \frac{\log(N-K)}{\log K}). \quad (25)$$

In the noisy case, the derived bounds are sharp. Fig. 1 depicts the quantities bounding $\lim_{K\to\infty}\frac{K\log(K(N-K))}{T_{typ}}$ as a function of α for $u = 0.05$ and $q = 0.01$. The bounds coincide in the first two decimal places.

III. Belief Propagation Reconstruction

The joint typicality decoder analysed in Section II has prohibitive computational complexity in the limit of large K and N. In this section, we compare the theoretical performance with the performance of belief propagation (BP) decoder, which performs an approximate bitwise MAP (maximum a posteriori) detection of defective items by solving:

$$\hat{\beta}_i^{(MAP)} = \arg\max_{\beta_i\in\{0,1\}} \mathbb{P}(\beta_i|y),\ i \in \{1,2,\dots,N\}. \quad (26)$$

The above can be transformed into:

$$\begin{aligned}\hat{\beta}_i^{(MAP)} &= \arg\max_{\beta_i\in\{0,1\}}\sum_{\sim\beta_i}\left[\prod_{t=1}^{T}\mathbb{P}(y_t|\beta_{supp(y_t)})\prod_{j=1}^{N}\mathbb{P}(\beta_j)\right] \\ &= \arg\max_{\beta_i\in\{0,1\}}\sum_{\sim\beta_i}\left[\prod_{t=1}^{T}\mathbb{P}(y_t|w(\beta_{supp(y_t)}))\cdot\right. \\ &\left.\prod_{j=1}^{N}\left(\lambda\delta_{\beta_j}(1) + (1-\lambda)\delta_{\beta_j}(0)\right)\right], \quad (27)\end{aligned}$$

where $\lambda = K/N$. Therefore, MAP detection amounts to the marginalisation of a function which permits a sparse factorisation, and as such can be performed efficiently via message passing on a factor graph corresponding to the measurement matrix $\mathbf{X}$.

The belief propagation message-update rules are given by:

$$m_{i\to t}^{(l+1)}(\beta_i) \propto (\lambda\delta_{\beta_i}(1) + (1-\lambda)\delta_{\beta_i}(0))\cdot\prod_{b\in\mathcal{N}(i)\setminus\{t\}}\hat{m}_{b\to i}^{(l)}(\beta_i), \quad (28)$$

$$\hat{m}_{t\to i}^{(l)}(\beta_i) \propto \sum_{\sim\beta_i}\left[\mathbb{P}(y_t|w(\beta_{supp(y_t)}))\cdot\prod_{j\in\mathcal{N}(t)\setminus\{i\}} m_{j\to t}^{(l)}(\beta_j)\right]. \quad (29)$$

The fact that $\mathbb{P}(y_t|\beta_{supp(y_t)}) = \mathbb{P}(y_t|w(\beta_{supp(y_t)}))$ greatly simplifies the message-passing update rules. In particular, since $\mathbb{P}(y_t|w(\beta_{supp(y_t)})) = (1-q)u^{w(\beta_{supp(y_t)})}$ due to the symmetry between x_t and β in (1), the above equations, by rewriting message-update rules in terms of log-ratios, i.e.,

$$L_{i\to t}^{(l)} = \log\frac{m_{i\to t}^{(l)}(1)}{m_{i\to t}^{(l)}(0)}\quad,\quad \hat{L}_{t\to i}^{(l)} = \log\frac{\hat{m}_{t\to i}^{(l)}(1)}{\hat{m}_{t\to i}^{(l)}(0)}. \quad (30)$$

simplify to:

$$L_{i\to t}^{(l)} = \begin{cases}\log\frac{\lambda}{1-\lambda}, & l = 0,\\ \log\frac{\lambda}{1-\lambda} + \sum_{b\in\mathcal{N}(i)\setminus\{t\}}\hat{L}_{b\to i}^{(l)}, & l \geq 1,\end{cases} \quad (31)$$

and

$$\hat{L}^{(l)}_{t\to i} =$$

$$\log\left(u + \frac{1-u}{1-(1-q)\prod_{j\in\mathcal{N}(t)\setminus\{i\}}\left(u+\frac{1-u}{1+\exp(L^{(l)}_{j\to t})}\right)}\right), \tag{32}$$

in the case of a positive t-th test, i.e., when $y_t = 1$, and simply

$$\hat{L}^{(l)}_{t\to i} = \log u, \tag{33}$$

for $y_t = 0$.

In a preliminary assessment of the belief propagation reconstruction, we simulated BP decoder for noisy group testing in the case where $N = 5000$, $K = 50$, $u = 0.05$, $q = 0.01$, and for various values of parameter p. We performed at least 200 trials at the various numbers of tests. The number of iterations was fixed to 50.

As illustrated in Fig. 2, the detected probability of perfect reconstruction increases with p, and is about 99% when the number of tests was $T \approx 1600$ for $p = 0.02$. The value of p which performs best here is $1/K$, suggesting that the same heuristics concerning the optimal p discussed for typical set decoding also apply for belief propagation.

In Fig. 3, we illustrated the number of detection errors per size of the support as a function of the number of tests. This figure illustrates that even though a large probability of perfect reconstruction is achieved only at the relatively large number of tests, the BP decoder typically diagnoses only a few items incorrectly at the number of tests as small as $T \approx 900$. These results are still far from the estimate arising from the asymptotic analysis of joint typicality decoder in the previous section, which is $T_{typ} \approx 400$, but nonetheless confirm the utility of the belief propagation algorithm in noisy group testing, even though no design of the measurement matrix that complies well with belief propagation algorithm has been taken into consideration. It may also be possible to achieve further improvements in performance by using belief propagation with decimation as in [9].

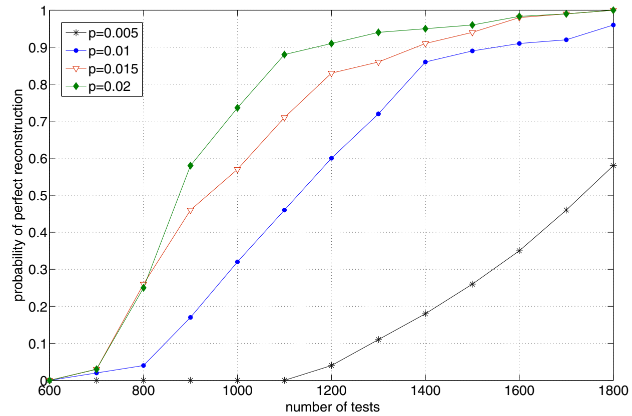

Figure 2. Probability of perfect reconstruction with BP at $N = 5000$, $K = 50$.

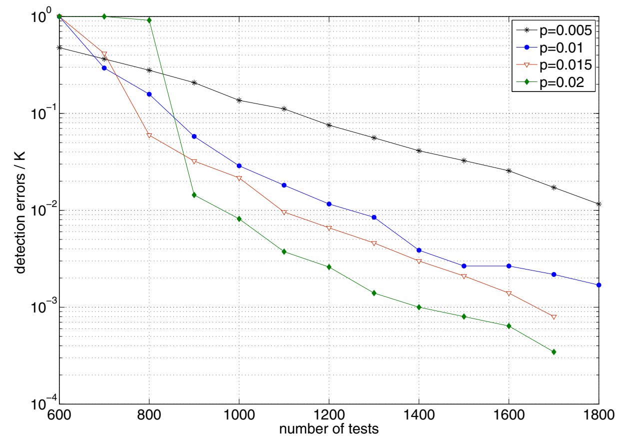

Figure 3. The number of detection errors per size of the support with BP at $N = 5000$, $K = 50$.

IV. Conclusions

This extended abstract studies the information theoretic bounds arising in the problem of noisy group testing and proposes an efficient algorithm for noisy group testing based on belief propagation. We develop a sharp estimate on the constants arising in the asymptotic approximation of the number of tests sufficient for the perfect detection via a joint typicality decoder, as a function of the noise parameters. We show how the presence of the false positives in the noisy group testing changes the order of the achievable number of tests. These result allows us to benchmark the performance of a belief propagation algorithm. We restrict our attention here to the case where the measurement matrix is composed of i.i.d. Bernoulli entries. More general measurement matrices can be studied in a similar manner, in particular those with row weights generated according to a pre-optimised degree distribution. A judicious choice of degree distributions may further improve the performance of the belief propagation algorithm, in analogy with well known results in sparse graph coding.

References

[1] G. Atia and V. Saligrama. Boolean compressed sensing and noisy group testing. arXiv:0907.1061v2, 2009.

[2] G. Atia and V. Saligrama. Noisy group testing: An information theoretic perspective. In *Proc. of the 47th Annual Allerton Conf. on Communication, Control and Computing*, Monticello, IL, 2009.

[3] M. Cheraghchi, A. Hormati, A. Karbasi, and M. Vetterli. Compressed sensing with probabilistic measurements: A group testing solution. In *Proc. of the 47th Annual Allerton Conf. on Communication, Control and Computing*, Monticello, IL, 2009.

[4] M. Cheraghchi, A. Karbasi, S. Mohajer, and V. Saligrama. Graph-constrained group testing. arXiv:1001.1445v1, 2010.

[5] R. Dorfman. The detection of defective members of large populations. *Ann. Math. Stat.*, 14(4):436–440, 1943.

[6] D.-Z. Du and F. K. Hwang. *Combinatorial group testing and its applications*. World Scientific, 2nd edition, 2000.
[7] Y. Erlich, A. Gordon, M. Brand, G. J. Hannon, and P. P. Mitra. Compressed genotyping. *IEEE Transactions on Information Theory*, 56(2):706–723, Feb 2010.
[8] O. Johnson and Y. Yu. Monotonicity, thinning and discrete versions of the entropy power inequality. arXiv:0909.0641v1, to appear in IEEE Transactions on Information Theory.
[9] M. Mezard, M. Tarzia, and C. Toninelli. Statistical physics of group testing. In *Proc. of the International Workshop on Statistical-Mechanical Informatics*, 2007.