Supplementary material to
“ Detection of Planted Solutions for Flat Satisfiability Problems”

Quentin Berthet and Jordan Ellenberg

Appendix A: Technical proofs

Proof [of Lemma 2] It holds that

$$Z = \sum_{x \in \mathbb{F}_2^n} \prod_{i=1}^m \mathbf{1}\{x \notin V_j\}\,.$$

By linearity, symmetry of the distribution, and independence of the V_j, we have for any $x_0 \in \mathbb{F}_2^n$

$$\mathbf{E}[Z] = 2^n (\mathbf{P}_{\text{unif}}(x_0 \notin V_1))^m\,.$$

Furthermore, for each k-flat of $\mathbb{F}_2^n$, $|V_1| = 2^{n-k}$, which yields the desired result. ■

Proof [of Lemma 3] We derive the second moment of Z

$$\begin{aligned} Z^2 &= \sum_{x,x' \in \mathbb{F}_2^n} \mathbf{1}\{x \in \mathcal{S}(V)\}\mathbf{1}\{x' \in \mathcal{S}(V)\} \\ &= \sum_{x} \mathbf{1}\{x \in \mathcal{S}(V)\} + \sum_{x \neq x'} \mathbf{1}\{x \in \mathcal{S}(V)\}\mathbf{1}\{x' \in \mathcal{S}(V)\}\,. \end{aligned}$$

Taking expectation yields

$$\mathbf{E}[Z^2] = \mathbf{E}[Z] + \sum_{x \neq x'} \mathbf{P}_{\text{unif}}\big(\{x \in \mathcal{S}(V)\} \cap \{x' \in \mathcal{S}(V)\}\big)\,.$$

The uniform distribution is invariant under the action of the affine group G, which is doubly transitive on $\mathbb{F}_2^n$. Therefore, the term $\mathbf{P}_{\text{unif}}\big(\{x \in \mathcal{S}(V)\} \cap \{x' \in \mathcal{S}(V)\}\big)$ is constant for all couples of distinct elements (x, x') of $\mathbb{F}_2^n$. To compute this distribution, it thus suffices to consider that x and x' are uniformly randomly chosen among the set of pairs of distinct elements. For all $j \in [m]$, this yields

$$\mathbf{P}_{\text{unif}}\big(\{x \notin V_j\} \cap \{x' \notin V_j\}\big) = \frac{2^n - 2^{n-k}}{2^n} \cdot \frac{2^n - (2^{n-k} - 1)}{2^n - 1} = (1 - 2^{-k})\Big(1 - 2^{-k} + \frac{2 - 2^{-k}}{2^n - 1}\Big)\,.$$

Using this in the derivation of the second moment, we have

$$\begin{aligned} \mathbf{E}[Z^2] &= \mathbf{E}[Z] + (2^{2n} - 2^n)(1 - 2^{-k})^m \Big(1 - 2^{-k} + \frac{2 - 2^{-k}}{2^n - 1}\Big)^m \\ &\leq \mathbf{E}[Z] + 2^{2n}(1 - 2^{-k})^{2m}\Big(1 + \frac{2 - 2^{-k}}{1 - 2^{-k}}\frac{1}{2^n - 1}\Big)^m \\ &\leq \mathbf{E}[Z] + \mathbf{E}[Z]^2\Big(1 + \frac{2 - 2^{-k}}{1 - 2^{-k}}\frac{1}{2^n - 1}\Big)^{\Delta n}\,. \end{aligned}$$

Note that the last term is a $1 + o(1)$. ■

Proof [of Theorem 4] We first note that $2(1 - 2^{-k})^{\Delta_k} = 1$, so that $\mathbf{E}[Z] = [2(1 - 2^{-k})^{\Delta}]^n$ is exponentially large when $\Delta < \Delta_k$, and exponentially small when $\Delta > \Delta_k$.

- For $\Delta < \Delta_k$, Markov's inequality yields

$$\mathbf{P}_{\text{unif}}(V \in \mathsf{FLAT}) = \mathbf{P}_{\text{unif}}(Z(V) \geq 1) \leq \mathbf{E}[Z] \to 0\,.$$

- For $\Delta < \Delta_k$, Paley-Zigmund's inequality and the result of Lemma 3 yields

$$\mathbf{P}_{\text{unif}}(V \in \mathsf{FLAT}) = \mathbf{P}_{\text{unif}}(Z(V) > 0) \geq \frac{\mathbf{E}[Z]^2}{\mathbf{E}[Z^2]} \to 1\,.$$

■

Proof [of Lemma 5] By definition of $\mathbf{P}_{\text{planted}}$

$$\frac{\mathbf{P}_{\text{planted}}(V)}{\mathbf{P}_{\text{unif}}(V)} = \frac{1}{2^n} \sum_{x \in \mathbb{F}_2^n} \frac{\mathbf{P}_{x^*}(V)}{\mathbf{P}_{\text{unif}}(V)}\,.$$

To compute the probabilities in the above ratios, we use the interpretation above of m drawings in $N = 2^k \mathcal{N}_k$ possible flats independently if the distribution is $\mathbf{P}_{\text{unif}}$, or otherwise in $N^* = (2^k - 1)\mathcal{N}_k$ possible choices corresponding to flats that do not contain x^*. Therefore, it holds for all V

$$\frac{\mathbf{P}_{x^*}(V)}{\mathbf{P}_{\text{unif}}(V)} = \begin{cases} 0 & \text{if } x \notin \mathcal{S}(V) \\ \left(\frac{N}{N^*}\right)^m & \text{otherwise} \end{cases}$$

Therefore, the likelihood ratio can be expressed in terms of $\mathbf{1}\{x \in \mathcal{S}(V)\}$, and $N/N^* = 1/(1 - 2^{-k})$

$$\begin{aligned} \frac{\mathbf{P}_{\text{planted}}}{\mathbf{P}_{\text{unif}}}(V) &= \frac{1}{2^n} \sum_{x \in \mathbb{F}_2^n} \left(\frac{N}{N^*}\right)^m \mathbf{1}\{x \in \mathcal{S}(V)\} \\ &= \frac{1}{\mathbf{E}[Z]} \sum_{x \in \mathbb{F}_2^n} \mathbf{1}\{x \in \mathcal{S}(V)\} = \frac{Z(V)}{\mathbf{E}[Z]}\,. \end{aligned}$$

■

Proof [of Lemma 8] Consider a fixed $Z \in \mathbb{F}_2^{\mathcal{N}_k}$ such that $Z_\emptyset = 1$. For an k-flat W described by (ℓ, α), we write $\mathcal{L}_{\alpha,\ell}(Z)$ as a function $q_{Z,\ell}$ of $\alpha \in \mathbb{F}_2^k$

$$q_{Z,\ell}(\alpha) = \sum_{\substack{S \subset [n] \\ |S| \leq k}} c_S(\ell, \alpha) Z_S\,.$$

We observe that each $c_S(\ell, \cdot)$ is a multivariate multilinear polynomial (with monomials that are squarefree), so that $q_{Z,\ell} \in \mathbb{F}_2[\alpha_1, \ldots, \alpha_k]$. Furthermore, the coefficient of the monomial $\alpha_1 \ldots \alpha_k$ is $Z_\emptyset = 1$. As the squarefree monomials are linearly independent, there exists an element of $\mathbb{F}_2^k$ such that $q_{Z,\ell}(\alpha) \neq 0$. Therefore, as α is uniformly distributed under the uniform distribution q_0, it holds that

$$\mathbf{P}_{\text{unif}}(\mathcal{L}_{\alpha,\ell}(Z) = 0) = \mathbf{P}_{\text{unif}}(q_{Z,\ell}(\alpha) = 0) \leq 1 - 2^{-k}\,.$$

As an aside, note that this bound is tight. Indeed, for all $Z \in \mathcal{V}$, the event $\mathcal{L}_{\alpha,\ell}(Z) = 0$ is equivalent to $z \notin W$, for $z = \phi^{-1}(Z)$. The probability of this event is $1 - 2^{-k}$, as seen in the proof of Lemma 2.

Let $V = (V_1, \ldots, V_m) \sim \mathbf{P}_{\text{unif}}$. By independence, we obtain directly that

$$\mathbf{P}_{\text{unif}}(\mathcal{L}_{\ell_j,\alpha_j}(X) = 0\ , \forall j \in [m]) \leq (1 - 2^{-k})^m\,.$$

By a union bound over all elements of $\mathbb{F}_2^{N_k}$, it holds that

$$\mathbf{P}_{\text{unif}}(\mathcal{L}_V \text{ has a solution}) \le 2^{N_k}(1-2^{-k})^m .$$

Taking $\Delta > \Delta_k$ yields the desired result. ■

Proof [of Lemma 10] For all $x \in \mathbb{F}_2^n$, we observe that under the null hypothesis, the variable $s(x,V)$ has distribution $\mathcal{B}(m, 1-2^{-k})$. Therefore, by Hoeffding's inequality,

$$\mathbf{P}_{\text{unif}}\big(s(x,V) > [(1-2^{-k}) + \alpha]m\big) \le \exp(-2\alpha^2 m) .$$

A union bound on $\mathbb{F}_2^n$ yields

$$\mathbf{P}_{\text{unif}}\big(\sigma(V) > [(1-2^{-k}) + \alpha]m\big) \le 2^n \exp(-2\alpha^2 m) \le \exp\big(-\big[2\alpha^2\Delta - \log(2)\big]n\big) .$$

Under $\mathbf{P}_{x^*}$ the variable $s(x^*,V)$ has distribution $\mathcal{B}\big(m, (1-2^{-k}) + \pi 2^{-k}\big)$. By Hoeffding's inequality,

$$\mathbf{P}_{x^*,\pi}\big(s(x^*,V) < [(1-2^{-k}) + \pi 2^{-k} - \alpha]m\big) \le \exp(-2\alpha^2 m) .$$

By definition of $\mathbf{P}_{\text{planted},\pi}$ and $\sigma(V) \ge s(x,V)$ for all $x \in \mathbb{F}_2^n$, we obtain the desired result. ■

Proof [of Theorem 11] For $\Delta > \tilde{\Delta}_{k,\pi}$, taking $\alpha = \pi 2^{-(k+1)}$ in the results of Lemma 10 yields the desired upper bound, as $2\alpha^2\Delta - \log(2) > 0$.

For $\Delta < \Delta_{k,\pi}$, we derive a bound on the total variation distance $d_{\mathsf{TV}}(\mathbf{P}_{\text{unif}}, \mathbf{P}_{\text{planted},\pi})$, through the inequality

$$d_{\mathsf{TV}}(\mathbf{P}_{\text{unif}}, \mathbf{P}_{\text{planted},\pi}) = \frac{1}{2}\mathbf{E}\Big[\Big|\frac{\mathbf{P}_{\text{planted},\pi}}{\mathbf{P}_{\text{unif}}}(V) - 1\Big|\Big] \le \frac{1}{2}\sqrt{\mathbf{E}\Big[\Big(\frac{\mathbf{P}_{\text{planted},\pi}}{\mathbf{P}_{\text{unif}}}(V) - 1\Big)^2\Big]} .$$

The term inside the square root being equal to the chi-square divergence $\chi^2(\mathbf{P}_{\text{planted},\pi}, \mathbf{P}_{\text{unif}})$ between the two distributions. We write $\mathbf{P}_{x,\pi} = q_{x,\pi}^{\otimes m}$ and $\mathbf{P}_{\text{unif}} = q_0^{\otimes m}$ as products of the distribution of each independent V_j. Writing out $\mathbf{P}_{\text{planted},\pi}$ as a uniform mixture of the $\mathbf{P}_{x,\pi}$ yields

$$\begin{aligned}
\chi^2(\mathbf{P}_{\text{planted},\pi}, \mathbf{P}_{\text{unif}}) &= \frac{1}{2^{2n}} \sum_{x,x' \in \mathbb{F}_2^n} \mathbf{E}\Big[\frac{\mathbf{P}_{x,\pi}}{\mathbf{P}_{\text{unif}}}\frac{\mathbf{P}_{x',\pi}}{\mathbf{P}_{\text{unif}}}(V)\Big] - 1 \\
&= \frac{1}{2^{2n}} \sum_{x,x' \in \mathbb{F}_2^n} \mathbf{E}\Big[\frac{q_{x,\pi}}{q_0}\frac{q_{x',\pi}}{q_0}(V_1)\Big]^m - 1 \\
&= \frac{1}{2^{2n}} \sum_{x \in \mathbb{F}_2^n} \mathbf{E}\Big[\Big(\frac{q_{x,\pi}}{q_0}(V_1)\Big)^2\Big]^n + \frac{1}{2^{2n}} \sum_{x \neq x'} \mathbf{E}\Big[\frac{q_{x,\pi}}{q_0}\frac{q_{x',\pi}}{q_0}(V_1)\Big]^m - 1 .
\end{aligned}$$

Note that $q_{x,\pi} = (1-\pi)q_0 + \pi q_x$, where q_x is the uniform distribution on k-flats that do not contain x (the planting distribution), so that

$$\frac{q_{x,\pi}}{q_0} = 1 + \pi\Big[\frac{q_x}{q_0} - 1\Big] .$$

Substituting this in the above yields

$$\begin{aligned}
\chi^2(\mathbf{P}_{\text{planted},\pi}, \mathbf{P}_{\text{unif}}) &= \frac{1}{2^{2n}} \sum_{x \in \mathbb{F}_2^n} \Big(1 + \pi^2\Big[\mathbf{E}\Big[\Big(\frac{q_x}{q_0}(V_1)\Big)^2\Big] - 1\Big]\Big)^m \\
&\quad + \frac{1}{2^{2n}} \sum_{x \neq x'} \Big(1 + \pi^2\Big[\mathbf{E}\Big[\frac{q_x}{q_0}\frac{q_{x'}}{q_0}(V_1)\Big] - 1\Big]\Big)^m - 1 .
\end{aligned}$$

Furthermore, for any k-flat V_1, it holds that $q_x/q_0(V_1) = (N/N_k)\mathbf{1}\{x \notin V_1\}$. We give the following upper bound the last two terms of this equation's RHS,

$$\begin{aligned}
\frac{1}{2^{2n}} \sum_{x \neq x'} \Big(1 + \pi^2 \Big[\mathbf{E}\Big[\frac{q_x}{q_0}\frac{q_{x'}}{q_0}(V_1)\Big] - 1\Big]\Big)^m - 1 \quad &\leq \quad \frac{1}{2^{2n}} 2^n \Big(1 - \pi^2 + \pi^2 \frac{\mathbf{P}_{\text{unif}}(x, x' \notin V_1)}{(1-2^{-k})^2}\Big)^m - 1 \\
&\leq \quad \Big(\frac{1-\pi^2}{2}\Big)^n \Big(1 + \frac{\pi^2}{1-\pi^2}\frac{2-2^{-k}}{(1-2^{-k}}\frac{1}{2^n - 1}\Big)^{\Delta n} - 1 \\
&\leq \quad \Big(1 + \frac{c_k \pi^2}{2^n - 1}\Big)^{c_k n/\pi^2} - 1\,,
\end{aligned}$$

for some constant $c_k > 0$ (independent of n and π), by the formula for $\mathbf{P}_{\text{unif}}(x, x' \notin V_1)$ derived in the proof of Lemma **??**. The last term converges to 0 when $n \to +\infty$. We bound as well the first term of the main equation's RHS

$$\begin{aligned}
\frac{1}{2^{2n}} \sum_{x \in \mathbb{F}_2^n} \Big(1 + \pi^2 \Big[\mathbf{E}\Big[\Big(\frac{q_x}{q_0}(V_1)\Big)^2\Big] - 1\Big]\Big)^m \quad &\leq \quad \frac{1}{2^{2n}} 2^n (1 + \pi^2(\mathbf{P}_{\text{unif}}(x \notin V_1) - 1))^m \\
&\leq \quad \frac{1}{2^n}\Big(1 + \frac{\pi^2}{2^k - 1}\Big)^{\Delta n}.
\end{aligned}$$

Taking $\Delta < \Delta_{k,\pi} = 2^k \log(2)/\pi^2$ yields $1/2(1 + \pi^2/(2^k - 1))^{\Delta} < 1$, and all the terms of $\chi^2(\mathbf{P}_{\text{planted},\pi}, \mathbf{P}_{\text{unif}})$ go to 0 when $n \to +\infty$.

■

Proof [of Lemma 12] In all cases, the k-flats are independent, and the m sets of k linear forms are uniformly distributed. If (A, b) is uniformly random, so are the b_j, and as a consequence, the ε_j. This yields the desired $V \sim \mathbf{P}_{\text{unif}}$. However, if there is a secret x, $\phi_j(x) = 1 - b_j$ with probability η. The distribution of $1 - b_j - \phi_j(x)$ is therefore is a mixture of the uniform distribution on $\mathbb{F}_2$ (with weight $1 - \pi$) and of the unit mass at 1 (with weight π). The distribution of $\varepsilon_j - \ell_j(x)$ is thus the mixture of the uniform distribution on $\mathbb{F}_2^n$ (with weight $1 - \pi$) and of the the distribution on $\mathbb{F}_2^k \setminus \{0\}$ generated by placing a 1 in one of the coefficients of $\varepsilon_j - \ell_j(x)$, and letting the others be independent and uniform. As shown in Remark 1, the flat V_j has distribution $q_{x,\pi}$ and $V \sim \mathbf{P}_{x,\pi}$, as desired. ■