Razborov' Switching Lemma for k-DNF Resolution

Ilario Bonacina*

April 23, 2018

Abstract

In this note we prove (a simplified version of) the small restrictions Switching Lemma for k-DNF resolution. That is Lemma 4.4 from [Raz15].

1 Introduction

Given the quite advanced topic of this note we do not define standard concepts such as DNF formulas, decision trees, k-DNF resolution, random restrictions etc. For the definitions of those concepts the reader could look at the introduction of [SBI04]. We denote with **bold** symbols random variables. Given a formula F and a restriction α, the restricted formula is denoted with $F\restriction\alpha$. If G is a DNF formula, $\mathrm{depth}(G)$ is the minimum depth of a decision tree (strongly) representing G. A decision tree T *(strongly) represents* a DNF G if for every leaf ℓ of T with label 0, the assignment π_ℓ corresponding to the path in T leading to ℓ is such that $t\restriction\pi_\ell = 0$ for every term $t \in G$; and for every leaf ℓ of T with label 1, the assignment π_ℓ corresponding to the path in T leading to ℓ is such that $t\restriction\pi_\ell = 1$ for at least a term $t \in G$.

Definition 1.1 ((r,p)-independent)**.** Given $r \in \mathbb{N}$ and $p \in [0,1]$, a random partial assignment $\boldsymbol{\rho}$ is (r,p)-*independent* if for any set of terms $t,\dots,t_m$ such that $X = \bigcup_{i=1}^m \mathrm{vars}(t_i)$ has size at most r then

$$\Pr\left[\bigwedge_{i\in[m]} (t_i\restriction\boldsymbol{\rho} \neq 1)\right] \leqslant \Pr\left[\bigwedge_{i\in[m]} (t_i\restriction\boldsymbol{\sigma} \neq 1)\right],$$

where $\boldsymbol{\sigma}$ is a random partial assignment such that each variable outside of X is unassigned, and for each variable in X it is set to 0 with probability $p/2$, it is set to 1 with probability $p/2$ and it is left unset with probability $1-p$. (For shortness call this distribution $\mathcal{U}_{p,X}$.)

Theorem 1.2 (Small restrictions Switching Lemma (Lemma 4.4 in [Raz15]))**.** *Let F be a k-DNF and $\boldsymbol{\rho}$ a random partial assignment that is (r,p)-independent, then for every $s \leqslant r$*

$$\Pr[\mathrm{depth}(F\restriction\boldsymbol{\rho}) > s] \leqslant \mathrm{e}^{-s(p\epsilon)^{O(k)}}, \tag{1}$$

where ϵ is a small enough absolute constant.

[*] Universitat Politècnica de Catalunya, Dept. Ciències de la Computació, Jordi Girona Salgado 31, Omega-223 08034 Barcelona, Catalonia, Spain, bonacina@cs.upc.edu

⚠ This is **a draft** that has not been peer reviewed yet: comments or bug-reports are more than welcome.

This result is proven in section 3 and the proof we give is essentially the same from [Raz15] with few minor changes here and there, in particular in the exposition. The original statement holds for *weighted* DNF formulas but this is a minor generalization of the statement (and proof) of Theorem 1.2. For sake of clarity we ignore this generalization. As consequences of Theorem 1.2 we could prove the following results.

1. Random k-CNFs with high probability require sub-exponential size refutations in $\mathcal{O}(k)$-DNF resolution. As far as we know, this result is not published in the literature but in [SBI04] it is proven that random k-CNFs require sub-exponential size refutations in $\mathcal{O}(\sqrt{k})$-DNF resolution. The exact same argument given there works using Theorem 1.2 instead of the Switching Lemma from [SBI04].

2. Resolution over $(k+1)$-DNFs is sub-exponentially separated from resolution over k-DNFs up to $k = \mathcal{O}(\log n)$ [Seg05].

3. The onto-PigeonHole Principle with cn pigeons and n holes with c constant, require sub-exponential size refutations in k-DNF resolution up to $k = \mathcal{O}(\log n/ \log\log n)$ [Raz15]. Previously this was known for k up to $\mathcal{O}(\sqrt{\log n/ \log\log n})$ [SBI04].

4. Neither resolution over $(\log n)$-DNFs nor polynomial calculus + resolution (over a field of characteristic $\neq 2$) have polynomial-size proofs of $\mathbf{NP} \subseteq \mathbf{P}/\mathrm{poly}$ [Raz15]. Actually, to prove this result Theorem 1.2 is not enough: we need the more general result for weighted DNFs as in [Raz15, Lemma 4.4].

We conclude this introduction with two open problems arising naturally in this context.

Open Problem 1.3. Random 3-XOR formulas with high probability require sub-exponential size refutation in k-DNF resolution up to $k = \mathcal{O}(\sqrt{\log n})$ [Ale11]. Prove the same result for k up to $\mathcal{O}(\log n)$. (Maybe with a modification of Theorem 1.2 proved here.)

Open Problem 1.4. Prove that polynomial calculus + resolution over a field of characteristic 2 doesn't have polynomial-size proofs of $\mathbf{NP} \subseteq \mathbf{P}/\mathrm{poly}$.

Acknowledgements. The author thanks Susanna Susanna F. De Rezende (KTH, Stockholm), Michal Garlík (UPC, Barcelona) and Massimo Lauria (Sapienza Universitá di Roma) for valuable discussions and spotting typos/errors in previous drafts of this note. The author was partially funded by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013) / ERC grant agreement no. 279611 and by the European Union's Horizon 2020 Research and Innovation Programme / ERC grant agreement no. 648276 AUTAR.

2 Preliminaries

The first ingredient we need to prove Theorem 1.2 is *Janson's inequality*. (This result is standard but anyway a proof is given in Appendix A.)

Theorem 2.1 (Janson inequality (Theorem 8.1.1 in [AS16]))**.** *Given $\vec{p} = (p_1, \dots, p_n) \in [0,1]^n$, consider $\{0,1\}^n$ as a probability space setting*

$$\Pr[x] = \prod_{i\in[n]} (p_i + (1-2p_i)x_i)\,,$$

for each $x = (x_1, \ldots, x_n) \in \{0,1\}^n$. Given any m subsets $V_1, \ldots, V_m$ of $[n]$, an assignment $\alpha : \{0,1\}^n \to \{0,1\}$ and the events

$$A_i = \{(x_1, \ldots, x_n) \in \{0,1\}^n \ : \ \forall j \in V_i,\ x_j = \alpha(j)\}\,,$$

it holds that

$$\Pr\left[\bigcap_{i \in [m]} A_i^c\right] \leqslant \mathrm{e}^{-\mu + \Delta/2}\,,$$

where $\mu = \sum_{i=1}^m \Pr[A_i]$ and $\Delta = \sum_{\substack{i,j \in [m], j \neq i \\ V_i \cap V_j \neq \emptyset}} \Pr[A_i \cap A_j]$.

The second ingredient we need is [Raz15, Lemma 4.5]. This essentially will be used to reduce the proof of Theorem 1.2 to proving a similar statement but for DNF formulas with a much simpler structure.

Theorem 2.2 (Lemma 4.5 in [Raz15])**.** *Let F be a k-DNF and $\boldsymbol{G}$ a random sub-DNF of F such that*

$$\forall t \in F \ \ \Pr[t \in \boldsymbol{G}] \geqslant \epsilon\,, \tag{2}$$

where ϵ is an arbitrary parameter. Let $\boldsymbol{\rho}$ be a random partial assignment independent from $\boldsymbol{G}$ s.t.

$$\Pr_{\boldsymbol{G},\boldsymbol{\rho}}[\mathrm{depth}(\boldsymbol{G}{\restriction}\boldsymbol{\rho}) > h] \leqslant \delta\,,$$

where h, δ are some other arbitrary parameters. Then

$$\Pr\left[\mathrm{depth}(F{\restriction}\boldsymbol{\rho}) > h + k\left\lceil\frac{2h}{\epsilon}\right\rceil\right] \leqslant \frac{2\delta}{\epsilon}\,.$$

Proof. By Markov's inequality $\Pr_{\boldsymbol{\rho}}[\Pr_{\boldsymbol{G}}[\mathrm{depth}(\boldsymbol{G}{\restriction}\boldsymbol{\rho}) > h] \geqslant \frac{\epsilon}{2}] \leqslant \frac{2\delta}{\epsilon}$. So it is enough to show that for each ρ, $\Pr[\mathrm{depth}(\boldsymbol{G}{\restriction}\rho) > h] < \frac{\epsilon}{2}$ implies that $\mathrm{depth}(F{\restriction}\rho) \leqslant h + k\lceil\frac{2h}{\epsilon}\rceil$. Consider fixed ρ and suppose that $\Pr[\mathrm{depth}(\boldsymbol{G}{\restriction}\rho) > h] < \frac{\epsilon}{2}$. The main observation we need is the following claim which will be used to construct a tree T representing $F{\restriction}\rho$ and such that $\mathrm{depth}(T) \leqslant h + k\lceil\frac{2h}{\epsilon}\rceil$.

Claim 2.3. Let s be a parameter and $t_1, \ldots, t_s$ be any s terms of F, then there exists a sub-DNF G of F s.t. $\mathrm{depth}(G{\restriction}\rho) \leqslant h$ and G contains strictly more than $\frac{s\epsilon}{2}$ terms from $t_1, \ldots, t_s$.

Proof. By eq. 2 we immediately get that $\mathbb{E}[|\{t_1, \ldots, t_s\} \cap \boldsymbol{G}|] \geqslant s\epsilon$. Moreover, let I be the indicator function of the event $\mathrm{depth}(\boldsymbol{G}{\restriction}\rho) > h$,

$$\mathbb{E}[|\{t_1, \ldots, t_s\} \cap \boldsymbol{G}| \cdot I] \leqslant s \cdot \mathbb{E}[I] < \frac{s\epsilon}{2}\,.$$

Hence $\mathbb{E}[|\{t_1, \ldots, t_s\} \cap \boldsymbol{G}| \cdot (1 - I)] > \frac{s\epsilon}{2}$ and there exists a sub-DNF G of F s.t. $\mathrm{depth}(G{\restriction}\rho) \leqslant h$ and G contains strictly more than $\frac{s\epsilon}{2}$ terms from $t_1, \ldots, t_s$. ∎

Construct trees $T_0, \ldots, T_s$ as follows (the parameter s will be set later): T_0 is the trivial tree of height 0. To construct $T_{\ell+1}$ from T_ℓ just go through of all leaves of T_ℓ. If the assignment corresponding to a leaf π evaluates $F{\restriction}\rho$ to 0 or 1 then label π with the corresponding value, otherwise there is a term $t_\pi \in F{\restriction}\rho$ not evaluated to 0 or 1. Append the decision tree for that term to the leaf π and continue this process for every leaf of T_ℓ. Let $T_{\ell+1}$ be the tree we just constructed.

Consider now T_s. For every leaf π of T_s not labeled with $0/1$ we use Claim 2.3 applied to the s terms $t_1, \ldots, t_s$ of F used to define the path leading to π and hence we get a DNF G_π such that $\mathrm{depth}(G_\pi{\upharpoonright}\rho) \leqslant h$ and G_π contains strictly more than $\frac{s\epsilon}{2}$ terms from $t_1, \ldots, t_s$. Append G_π to the leaf π (possibly identifying variables already queried in the path leading to π). We apply Claim 2.3 to every leaf of T_s not labeled with $0/1$. Let T be the tree we just constructed. We need to show that T represents $F{\upharpoonright}\rho$ and to do so we need to impose that $\frac{s\epsilon}{2} \geqslant h$. That is let $s = \lceil \frac{2h}{\epsilon} \rceil$. Clearly then $\mathrm{depth}(T) \leqslant h + ks = h + k\lceil \frac{2h}{\epsilon} \rceil$. Moreover all the leaves in T that were not leaves in T_s have label 1. Otherwise the tree T_π appended to an undecided leaf π of T_s has depth $\leqslant h$ but it must set to 0 at least $h+1$ disjoint sets of variables corresponding to at least $h+1$ terms of G_π. And this is clearly impossible. The tree T is then representing $F{\upharpoonright}\rho$. ■

3 Proof of Theorem 1.2

We have now all ingredients needed to prove Theorem 1.2. Whenever in the proof we make some $\boxed{\text{hypotheses}}$ with the intent to check them later we highlight them. (This is somewhat **(Hyp. 0)** not customary in standard mathematical writing but it might help the reader in the process of making sense of some choices of parameters.)

A k-DNF G is *pseudo-monotone* if here exists a total assignment that satisfies all terms of G. (Or equivalently that all variables appear in G only with one polarity, either positively or negatively.) Suppose that we are able to prove that for every pseudo-monotone k-DNF F and every $s \leqslant r$ that

$$\Pr[\mathrm{depth}(F{\upharpoonright}\boldsymbol{\rho}) > s] \leqslant \mathrm{e}^{-s\gamma^{2k}}, \tag{3}$$

where $\gamma = p\epsilon$ and ϵ is a small enough absolute constant.Then from eq. 3 we get immediately eq. 1. Consider $\boldsymbol{G}$ the sub-DNF of F consisting of all the terms of F satisfied by a uniformly random total assignment. Clearly $\Pr[t \in \boldsymbol{G}] \geqslant 2^{-k}$ and, by construction, $\boldsymbol{G}$ is pseudo-monotone so by eq. 3 and Theorem 2.2 we get:

$$\Pr[\mathrm{depth}(F{\upharpoonright}\boldsymbol{\rho}) > s] \leqslant 2^{k+1}\mathrm{e}^{-s(1-k+k2^{k+1})^{-1}(p\epsilon)^{2k}} = \mathrm{e}^{-s(p\epsilon)^{O(k)}}.$$

So we just need to prove eq. 3 for pseudo-monotone k-DNFs and we do it by induction on k. If $k = 0$ then F is a constant and $\mathrm{depth}(F{\upharpoonright}\boldsymbol{\rho}) = 0$ with probability 1 so eq. 1 clearly holds. We split the inductive step into two cases: Janson's inequality (Theorem 2.1) takes care of one case, the inductive hypothesis takes care of the other case.

Easy case Since all variables appear always with the same polarity in F we abuse notations a bit and we identify terms and their underlying set of literals. For instance we write $t_i \cap t_j$ meaning $\mathrm{lit}(t_i) \cap \mathrm{lit}(t_j)$. We want to find H terms $t_1, \ldots, t_H$ in F with H a parameter yet to define and, since we want to use on them the (r, p)-independent assumption, we better require that $\boxed{Hk \leqslant r}$. For every possible choice of such terms we have that **(Hyp. 1)**

$$\Pr[\mathrm{depth}(F{\upharpoonright}\boldsymbol{\rho}) > s] \leqslant \Pr\left[\bigwedge_{i\in[H]} (t_i{\upharpoonright}\boldsymbol{\rho} \neq 1)\right] \leqslant \Pr\left[\bigwedge_{i\in[H]} (t_i{\upharpoonright}\boldsymbol{\sigma} \neq 1)\right], \tag{4}$$

where in the last inequality we used the fact that $\boldsymbol{\rho}$ is (r, p)-independent and $\boldsymbol{\sigma}$ is distributed according to $\mathcal{U}_{p,X}$ with $X = \bigcup_{i\in[H]} \mathrm{vars}(t_i)$. Now, if a term $t \in F$ has strictly less than k literals we add $k - |t|$ *new* literals to it so that the term actually has exactly k literals. This operation makes $\Pr[\bigwedge_{i\in[H]}(t_i{\restriction}\boldsymbol{\sigma} \neq 1)]$ increase so it is safe as long as we are able to upper bound the new quantity. Suppose then that all terms $t_1,\dots,t_H$ have exactly k literals.

We want to use Janson's inequality (Theorem 2.1) but $\boldsymbol{\sigma}$ is not a total assignment so we extend it to a total assignment $\boldsymbol{\sigma}'$ setting all variables x_i not in X to 1 with probability 1 if the literal $\neg x_i$ appears in F, otherwise if x_i appears as a literal we set it to 0. Moreover, since variables appear in F always with the same polarity, $t_i{\restriction}\boldsymbol{\sigma} = 1$ if and only if $t_i{\restriction}\boldsymbol{\sigma}' = 1$, so let A_i denote the event $t_i{\restriction}\boldsymbol{\sigma}' = 1$. Let's write for simplicity $i \sim j$ if both $i, j \in H$, $i \neq j$ and $t_i \cap t_j \neq \emptyset$. We need to calculate $\Pr[A_i]$ and $\Pr[A_i \cap A_j]$ whenever $i \sim j$. By the assumptions on $\boldsymbol{\sigma}'$ we clearly have that $\Pr[A_i] \geqslant (p/2)^k$ so

$$\mu = \sum_{i\in[H]} \Pr[A_i] \geqslant H(p/2)^k .$$

Moreover for $i \sim j$, $\Pr[A_i \cap A_j] = (p/2)^{|t_i \cup t_j|}$, since all the variables in the terms appear with the same polarity, so

$$\Pr[A_i \cap A_j] = (p/2)^{2k}\,\delta(t_i \cap t_j)\,,$$

where $\delta(t_i \cap t_j) = (2/p)^{|t_i \cap t_j|}$ and $\delta(\emptyset) = 0$. Hence, if $\boxed{\sum_{i\neq j}\delta(t_i \cap t_j) \leqslant H(2/p)^k}$, **(Hyp. 2)**

$$\Delta = \sum_{i\sim j}\Pr[A_i \cap A_j] = (p/2)^{2k}\sum_{i\sim j}\delta(t_i \cap t_j) \leqslant H(p/2)^k\,. \tag{5}$$

Then, using Janson's inequality (Theorem 2.1), we upper bound eq. 4:

$$\begin{aligned}\Pr\left[\bigwedge_{i\in[H]}(t_i{\restriction}\boldsymbol{\sigma} \neq 1)\right] &\leqslant \mathrm{e}^{-\mu+\Delta/2} \\ &= \mathrm{e}^{-\frac{H}{2}(p/2)^k} \\ &\leqslant \mathrm{e}^{-s\gamma^{2k}}\,,\end{aligned}$$

where the last inequality holds if we set $H = 2s(2p)^k\epsilon^{2k}$. The condition on H we need to satisfy (Hyp. 1) is that $Hk \leqslant r$ but for ϵ small enough this condition holds since by hypothesis $s \leqslant r$ (and $p \leqslant 1$).

The other assumption (Hyp. 2), that is $\sum_{i\sim j}\delta(t_i \cap t_j) \leqslant H(2/p)^k$, is more problematic. First of all since we have yet to find $t_1,\dots,t_H$. We find them now using a probabilistic process. Let $\boldsymbol{t}$ be a random term in F and let $\boldsymbol{t}_1,\dots\boldsymbol{t}_H$ be H independent samples from F according to the distribution of $\boldsymbol{t}$. (Notice that $\boldsymbol{t}$ is in general *not* uniformly distributed.) Then

$$\mathbb{E}[\sum_{i\neq j}\delta(\boldsymbol{t}_i \cap \boldsymbol{t}_j)] = \frac{H(H-1)}{2}\,\mathbb{E}[\delta(\boldsymbol{t}\cap\boldsymbol{t}')] \leqslant H^2\,\mathbb{E}[\delta(\boldsymbol{t}\cap\boldsymbol{t}')]\,,$$

where $\boldsymbol{t}'$ is a random term of F with the same distribution of $\boldsymbol{t}$ independent from it. Now, the previous inequality does not depend on the distribution of $\boldsymbol{t}$ and if there exists a distribution for $\boldsymbol{t}$ such that $\mathbb{E}[\delta(\boldsymbol{t}\cap\boldsymbol{t}')] \leqslant H^{-1}(2/p)^k$ then from the previous equation we get that for such specific distribution

$$\mathbb{E}[\sum_{i\neq j}\delta(\boldsymbol{t}_i \cap \boldsymbol{t}_j)] \leqslant H(2/p)^k\,.$$

This means there exists some particular sampling $t_1,\dots,t_H$ such that

$$\sum_{i\neq j}\delta(t_i \cap t_j) \leqslant H(2/p)^k\,,$$

hence satisfying the assumption (Hyp. 2) on $\sum_{i\neq j}\delta(t_i \cap t_j)$.

Not-so-easy case We have to deal now with the case that for all possible distributions of random terms $\boldsymbol{t}$ it holds that $\mathbb{E}[\delta(\boldsymbol{t}\cap\boldsymbol{t}')] > H^{-1}(2/p)^k$, where as before $\boldsymbol{t}'$ is a random term of F with the same distribution of $\boldsymbol{t}$ independent from it. Now we cannot guarantee the existence of $t_1,\dots t_H$ in F such that eq. 5 holds so we use the inductive hypothesis to prove eq. 3.

The idea/hope now is to group together terms of F of similar importance, that is find k-
DNFs $F_1,\dots,F_k$ such that $\boxed{F=\bigvee_{\mu\in[k]}F_\mu}$ and for each of those find a set of variables V^μ *not* **(Hyp. 3)**
too big such that for every $\pi\in\{0,1\}^{V^\mu}$, $\boxed{F_\mu{\restriction}\pi \text{ is a } (k-\mu)\text{-DNF}}$. This clearly implies that **(Hyp. 4)**
for every restriction ρ, $\operatorname{depth}(F{\restriction}\rho)\leqslant\sum_{\mu\in[k]}\operatorname{depth}(F_\mu{\restriction}\rho)$ and hence

$$\Pr[\operatorname{depth}(F{\restriction}\boldsymbol{\rho})>s]\leqslant\sum_{\mu\in[k]}\Pr[\operatorname{depth}(F_\mu{\restriction}\boldsymbol{\rho})>sq_\mu]\,, \tag{6}$$

where $q_\mu\in\mathbb{N}$ is a quantity yet to decide such that $\boxed{\sum_{\mu\in[k]}q_\mu\leqslant 1}$.[1] Suppose that for each **(Hyp. 5)**
$\mu\in[k]$, $\boxed{|V^\mu|\leqslant\frac{sq_\mu}{2}}$, then **(Hyp. 6)**

$$\begin{aligned}
\Pr[\operatorname{depth}(F_\mu{\restriction}\boldsymbol{\rho})>sq_\mu] &\leqslant \Pr[\exists\pi\in\{0,1\}^{V^\mu},\ \operatorname{depth}((F_\mu{\restriction}\pi){\restriction}\boldsymbol{\rho})>\frac{sq_\mu}{2}]\\
&\leqslant \sum_{\pi\in\{0,1\}^{V^\mu}}\Pr[\operatorname{depth}((F_\mu{\restriction}\pi){\restriction}\boldsymbol{\rho})>\frac{sq_\mu}{2}]\\
&\leqslant 2^{|V^\mu|}\Pr[\operatorname{depth}((F_\mu{\restriction}\pi){\restriction}\boldsymbol{\rho})>\frac{sq_\mu}{2}]\\
&\leqslant 2^{|V^\mu|}\mathrm{e}^{-\frac{sq_\mu}{2}\gamma^{2k-2\mu}} && \text{(by the inductive hypothesis)}\\
&\leqslant \mathrm{e}^{-\frac{sq_\mu}{4}\gamma^{2k-2\mu}}\,,
\end{aligned}$$

where the last inequality holds if we suppose that $\boxed{|V^\mu|\leqslant\frac{sq_\mu}{4}\gamma^{2k-2\mu}}$, which is a more stringent **(Hyp. 7)**
condition than $|V^\mu|\leqslant\frac{sq_\mu}{2}$ (Hyp. 6). Under this stronger assumption, from eq. 6 we get

$$\Pr[\operatorname{depth}(F{\restriction}\boldsymbol{\rho})>s]\leqslant\sum_{\mu\in[k]}\mathrm{e}^{-\frac{sq_\mu}{4}\gamma^{2k-2\mu}}\leqslant\mathrm{e}^{-s\gamma^{2k}}\,,$$

if we make the assumption that $\boxed{\sum_{\mu\in[k]}\mathrm{e}^{-\frac{sq_\mu}{4}\gamma^{2k-2\mu}}\leqslant\mathrm{e}^{-s\gamma^{2k}}}$. This last equation is concluding **(Hyp. 8)**
the inductive step *but* we have yet to find F_μ, V^μ and q_μ satisfying the assumptions we made. Their construction is our job now.

Remember that we are in the case where for every possible distribution for $\boldsymbol{t}$ it holds that $\mathbb{E}[\delta(\boldsymbol{t}\cap\boldsymbol{t}')]>H^{-1}(2/p)^k=(2s\gamma^{2k})^{-1}$, where as before $\boldsymbol{t}'$ is a random term of F with the same distribution of $\boldsymbol{t}$ independent from it. Roughly the idea is to use this assumption to assign a weight $\mathrm{w}(V)$ to every subset of variables V and use those weights to organize terms in F_μ.

First of all let's extend the definition of $\delta(\cdot)$ to generic sets of variables:

$$\delta(V)=\begin{cases}0 & \text{if } V=\emptyset\,,\\ (2/p)^{|V|} & \text{otherwise}\,.\end{cases}$$

[1] Razborov in [Raz15, Lemma 4.4] sets $q_\mu=2^{-\mu}$. For the moment we leave q_μ as a parameter. (Later we will actually set it to a different value than Razborov's but we will still have some similar exponential decay in μ.)

Then

$$
\begin{aligned}
(2s\gamma^{2k})^{-1} = H^{-1}(2/p)^k \leqslant \mathbb{E}[\delta(\boldsymbol{t} \cap \boldsymbol{t}')] &= \sum_V \delta(V)\Pr[\boldsymbol{t} \cap \boldsymbol{t}' = V] && (7)\\
&\leqslant \sum_V \delta(V)\Pr[V \subseteq \boldsymbol{t} \wedge V \subseteq \boldsymbol{t}'] && (8)\\
&= \sum_V \delta(V)\Pr[V \subseteq \boldsymbol{t}]^2 \,. && (9)
\end{aligned}
$$

The quadratic form $\sum_V \delta(V)\Pr[V \subseteq \boldsymbol{t}]^2$ attains a minimum value ξ (since it is a continuous function and the space of all probability distributions over the terms of F is compact). From now on suppose $\boldsymbol{t}$ has a fixed distribution: one realizing this minimum ξ, (and by assumption $\xi \geqslant (2s\gamma^k)^{-1}$).

Let the weight of V be $\mathrm{w}(V) = \Pr[V \subseteq \boldsymbol{t}]$. Then eq. 9 can be rewritten as

$$\sum_V \delta(V)\,\mathrm{w}(V)^2 = \xi \tag{10}$$

Now we need an analogous expression but specialized for every term $t \in F$. We have that $\Pr[V \subseteq \boldsymbol{t}] = \sum_{t \supseteq V} \Pr[\boldsymbol{t} = t]$. That is $\mathrm{w}(V) = \sum_{t \supseteq V} y_t$ where $y_t = \Pr[\boldsymbol{t} = t]$. Taking partial derivatives on the variables y_t we claim that for each term $t \in F$

$$\sum_{V \subseteq t} \delta(V)\,\mathrm{w}(V) \geqslant \xi \,. \tag{11}$$

We prove now eq. 11. This proof is self contained and it is not relevant to understand the remaining part of the argument.

Proof of eq. 11. [2] Notice that

$$
\begin{aligned}
\sum_{t\in F} \frac{\partial(\sum_V \delta(V)\,\mathrm{w}(V)^2)}{\partial y_t} y_t &= \sum_{t\in F}\sum_V 2\,\delta(V)\,\mathrm{w}(V)\frac{\partial(\mathrm{w}(V))}{\partial y_t} y_t \\
&= 2\sum_V \delta(V)\,\mathrm{w}(V)\sum_{t\in F}\frac{\partial(\mathrm{w}(V))}{\partial y_t} y_t \\
&= 2\sum_V \delta(V)\,\mathrm{w}(V)^2 \,.
\end{aligned}
$$

If we had that $\frac{\partial(\sum_V \delta(V)\,\mathrm{w}(V)^2)}{\partial y_t} < 2\xi$ for each $t \in F$, then

$$\sum_{t\in F} \frac{\partial(\sum_V \delta(V)\,\mathrm{w}(V)^2)}{\partial y_t} y_t < 2\xi \sum_{t\in F} y_t = 2\xi \,.$$

But this together with the previous chain of inequalities will contradict eq. 10. So this means for the moment just that there exists a term $t_0 \in F$ such that

$$\frac{\partial(\sum_V \delta(V)\,\mathrm{w}(V)^2)}{\partial y_{t_0}} \geqslant 2\xi \,.$$

[2] This proof is a bit technical but somewhat standard. In Razborov's paper it is left as an exercise for the reader with some hint. The hint given there seems to be slightly misleading/wrong. Or at least we didn't manage to make it work.

That is eq. 11 holds for t_0. Notice that for the term t_0 we found must be that $y_{t_0} > 0$. We now claim that for each term $t \neq t_0$ it holds that $\sum_{V \subseteq t} \delta(V)\,\mathrm{w}(V) \geqslant \sum_{V \subseteq t_0} \delta(V)\,\mathrm{w}(V)$ and hence eq. 11 clearly holds. Suppose, for sake of contradiction, that $\sum_{V \subseteq t} \delta(V)\,\mathrm{w}(V) < \sum_{V \subseteq t_0} \delta(V)\,\mathrm{w}(V)$ for a term $t \neq t_0$ and let

$$A = \sum_{V \subseteq t_0} \delta(V)\,\mathrm{w}(V) - \sum_{V \subseteq t} \delta(V)\,\mathrm{w}(V) > 0\,.$$

Since $t \neq t_0$ then let

$$B = \sum_{\substack{V \subseteq t_0 \\ V \not\subseteq t}} \delta(V) + \sum_{\substack{V \not\subseteq t_0 \\ V \subseteq t}} \delta(V) > 0\,.$$

Let $0 < \epsilon < \min\{\frac{2A}{B}, y_{t_0}, 1 - y_t\}$. Such ϵ exists since by assumption $y_{t_0} > 0$ and hence $y_t < 1$. Consider then the distribution $\widetilde{\boldsymbol{t}}$ equal to $\boldsymbol{t}$ except on t_0 and t where $\Pr[\widetilde{\boldsymbol{t}} = t_0] = y_{t_0} - \epsilon$ and $\Pr[\widetilde{\boldsymbol{t}} = t] = y_t + \epsilon$. This is still a valid distribution.

Now let $\widetilde{w}(V) = \Pr[V \subseteq \widetilde{\boldsymbol{t}}]$, it is immediate to see from the definition that

$$\widetilde{w}(V) = w(V) + \epsilon(I_{V \subseteq t} - I_{V \subseteq t_0})\,,$$

where I_A is the indicator function of A: 1 if A holds, 0 if not. By the minimality of the quadratic form in eq. 9 attained in $\boldsymbol{t}$ and the definition of $\widetilde{\boldsymbol{t}}$,

$$\begin{aligned} \sum_V \delta(V)\,\mathrm{w}(V)^2 &\leqslant \sum_V \delta(V) \widetilde{w}(V)^2 \\ &= \sum_V \delta(V)\left(\mathrm{w}(V) + \epsilon(I_{V \subseteq t} - I_{V \subseteq t_0})\right)^2\,. \end{aligned}$$

From the previous equation, using that $\epsilon > 0$, we immediately get that

$$\epsilon \sum_V \delta(V)(I_{V \subseteq t} - I_{V \subseteq t_0})^2 + 2 \sum_V \delta(V)\,\mathrm{w}(V)(I_{V \subseteq t} - I_{V \subseteq t_0}) \geqslant 0\,,$$

which is exactly the same as

$$\epsilon B - 2A \geqslant 0\,,$$

but this is not possible for our choice of ϵ. This concludes the proof of eq. 11. ■

From now on we just use eq. 10 and eq. 11 ignoring completely the fact that the $\mathrm{w}(V)$ are probabilities.

Order the sets of variables as $V_1, V_2, V_3, \ldots$ in such a way that $\mathrm{w}(V_1) \geqslant \mathrm{w}(V_2) \geqslant \mathrm{w}(V_3) \geqslant \cdots$ and order the variables according to this order: first put the variables in V_1 (in some order), then the ones in $V_2 \setminus V_1$ (in some order), then the ones in $V_3 \setminus (V_1 \cup V_2)$ and so on. Given a term $t \in F$ and a set of variables V, let $\mu_t(V)$ be the size of the minimal initial segment of t containing V if there is one, otherwise set $\mu_t(V) = 0$. Notice that if $V \subseteq t$ then $\mu_t(V) \geqslant |V|$. Now we have all the ingredients set to define F_μ, V^μ and q_μ. Let's start with F_μ:

$$F_\mu = \left\{ t \in F \;:\; \sum_{\substack{V \subseteq t \\ \mu_t(V) = \mu}} \delta(V)\,\mathrm{w}(V) \geqslant \xi q_\mu \right\}. \tag{12}$$

The only claim on F_μ we had is that $F = \bigvee_{\mu \in [k]} F_\mu$ (Hyp. 3). Suppose, for sake of contradiction, that there exist a term $t \in F$ such that for each $\mu \in [k]$, $\sum_{V \subseteq t,\ \mu_t(V) = \mu} \delta(V)\,\mathrm{w}(V) < \xi q_\mu$. Then

$$\sum_{V \subseteq t} \delta(V)\,\mathrm{w}(V) \leqslant \sum_{\mu \in [k]} \sum_{\substack{V \subseteq t \\ \mu_t(V) = \mu}} \delta(V)\,\mathrm{w}(V) < \xi \sum_{\mu \in [k]} q_\mu \leqslant \xi\,, \tag{13}$$

where the last inequality follows assuming that $\sum_{\mu\in[k]} q_\mu \leqslant 1$ which is Hyp. 5. Then eq. 13 would contradict eq. 11 and hence it must be that $F = \bigvee_{\mu\in[k]} F_\mu$.

Let's construct now the desired V^μ for each F_μ. Consider fixed then F_μ and let $t \in F_\mu$. We have at most 2^μ possible $V \subseteq t$ such that $\mu_t(V) = \mu$ so, by the definition of F_μ in eq. 12, there must exist some $V^{(t)} \subseteq t$ such that $\mu_t(V^{(t)}) = \mu$ and

$$\mathrm{w}(V^{(t)}) \geqslant \xi q_\mu\, \delta(V^{(t)})^{-1} 2^{-\mu} \geqslant \xi q_\mu (2/p)^{-|Vt|} 2^{-\mu} \geqslant \xi q_\mu (p/4)^\mu\,.$$

Let this $V^{(t)}$ be the ℓ_t-th in the global ordering of all the V_i:s. Now, using eq. 10 and the previous equation let's upper bound the size of $\bigcup_{i\leqslant\ell_t} V_i$:

$$\begin{aligned}\xi \geqslant \sum_{i\leqslant\ell_t} \delta(V_i)\, \mathrm{w}(V_i)^2 \geqslant \mathrm{w}(V_{\ell_t})^2 &\sum_{i\leqslant\ell_t} \delta(V_i)\\ \geqslant \mathrm{w}(V_{\ell_t})^2 &\sum_{i\leqslant\ell_t} |V_i|\\ \geqslant \mathrm{w}(V_{\ell_t})^2 &\Big|\bigcup_{i\leqslant\ell_t} V_i\Big|\\ \geqslant \xi^2 q_\mu^2 (p/4)^{2\mu} &\Big|\bigcup_{i\leqslant\ell_t} V_i\Big|\,.\end{aligned}$$

So we get that

$$\Big|\bigcup_{i\leqslant\ell_t} V_i\Big| \leqslant \xi^{-1} q_\mu^{-2} (4/p)^{2\mu} \leqslant 2s\gamma^{2k} q_\mu^{-2} (4/p)^{2\mu}\,, \tag{14}$$

and notice how this expression does not depend on the particular $t \in F_\mu$ we chose. So take as V^μ the largest union $\bigcup_{i\leqslant\ell_t} V_i$ for all possible terms $t \in F_\mu$. Clearly the upper bound in eq. 14 still holds and by construction for each term $t \in F_\mu$, $|V^\mu \cap t| \geqslant \mu$. So, however we set the variables in V^μ by some π then $F_\mu\restriction\pi$ is a $(k-\mu)$-DNF. The remaining bits are just to check that we can define q_μ in such a way that Hyp. 5, Hyp. 7 and Hyp. 8 hold (for ϵ a small enough absolute constant). That is, for ϵ small enough, we want to satisfy the following inequalities

$$\sum_{\mu\in[2k]} q_\mu \leqslant 1\,, \tag{15}$$

$$|V^\mu| \leqslant \frac{sq_\mu}{4}\gamma^{2k-2\mu}\,, \tag{16}$$

$$\sum_{\mu\in[k]} \mathrm{e}^{-\frac{sq_\mu}{4}\gamma^{2k-2\mu}} \leqslant \mathrm{e}^{-s\gamma^{2k}}\,. \tag{17}$$

From eq. 14 we get that in order for eq. 16 to hold it is enough to set $q_\mu = 4(4\epsilon)^{2\mu/3}$. With this choice it is clear that for ϵ small enough eq. 15 holds. To check eq. 17 we need to be slightly more careful. For simplicity we use just the fact that $q_\mu \geqslant 4\gamma^{2\mu/3}$. To prove eq. 17 it is enough to have that

$$\begin{aligned}\sum_{\mu\in[k]} \mathrm{e}^{-s\gamma^{2k-\frac{4\mu}{3}}} &= \sum_{\mu\in[k]} \left(\mathrm{e}^{-s\gamma^{2k}}\right)^{\gamma^{-\frac{4\mu}{3}}}\\ &\leqslant \sum_{\mu\in[k]} \left(\mathrm{e}^{-s\gamma^{2k}}\right)^{\epsilon^{-\frac{4\mu}{3}}}\\ &\leqslant \mathrm{e}^{-s\gamma^{2k}}\,,\end{aligned}$$

where the last inequality holds for small enough ϵ. ■

References

[Ale11] Michael Alekhnovich. Lower bounds for k-DNF resolution on random 3-CNFs. *Computational Complexity*, 20(4):597–614, 2011.

[AS16] Noga Alon and Joel H. Spencer. *The Probabilistic Method.* Wiley Publishing, 4th edition, 2016.

[Raz15] Alexander A. Razborov. Pseudorandom generators hard for k-DNF resolution and polynomial calculus resolution. *Ann. Math. (2)*, 181(2):415–472, 2015.

[SBI04] Nathan Segerlind, Sam Buss, and Russell Impagliazzo. A switching lemma for small restrictions and lower bounds for k-dnf resolution. *SIAM J. Comput.*, 33(5):1171–1200, May 2004.

[Seg05] Nathan Segerlind. Exponential separation between res(k) and res($k+1$) for $k \leqslant \epsilon \log n$. *Inf. Process. Lett.*, 93(4):185–190, February 2005.

A Proof of Janson's Inequality (Theorem 2.1)

A set $A \subseteq \{0,1\}^n$ is *up-closed* if for every $i \in [n]$, if $(x_1, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_n) \in A$ then $(x_1, \dots, x_{i-1}, 1, x_{i+1}, \dots, x_n) \in A$. Similarly $A \subseteq \{0,1\}^n$ is *down-closed* if its complement A^c is up-closed.

Theorem A.1 (Harris-Kleitman inequality)**.** *Consider $\{0,1\}^n$ as a probability space as in Theorem 2.1. If A and B are two events that are both up-closed (or down closed) then*

$$\Pr[A \mid B] \geqslant \Pr[A]\,.$$

If A is up-closed and B is down-closed then

$$\Pr[A \mid B] \leqslant \Pr[A]\,.$$

Proof. For simplicity let's consider only the case when A and B are both up-closed since the other cases are equivalent. We prove by induction on n that $\Pr[A \cap B] \geqslant \Pr[A]\Pr[B]$. From this we can immediately derive the desired inequality recalling that $\Pr[A \cap B] = \Pr[A \mid B]\Pr[B]$.

For $n = 1$ the result is clear. For $n > 1$ define

$$\begin{aligned} A_0 &= \{(x_1, \dots, x_{n-1}) \,:\, (x_1, \dots, x_n) \in A \wedge x_n = 0\}\,, \\ A_1 &= \{(x_1, \dots, x_{n-1}) \,:\, (x_1, \dots, x_n) \in A \wedge x_n = 1\}\,. \end{aligned}$$

Similarly define B_0 and B_1. Since A and B are up-closed, then $A_0 \subseteq A_1$ and $B_0 \subseteq B_1$ and both A_0, A_1, B_0, B_1 are up-closed. For sake of conciseness let $a_0 = \Pr[A_0]$, $a_1 = \Pr[A_1]$ and analogously define b_0, b_1. So, in particular

$$(a_1 - a_0)(b_1 - b_0) \geqslant 0\,,$$

and expanding this expression we get

$$a_1 b_1 + a_0 b_0 \geqslant a_1 b_0 + a_0 b_1\,. \tag{18}$$

Let for shortness $q_n = 1 - p_n$. We then have the following chain of inequalities:

$$\begin{aligned}
\Pr[A \cap B] &= p_n \Pr[A_0 \cap B_0] + q_n \Pr[A_1 \cap B_1] && (19)\\
&\geqslant (p_n a_0 b_0 + q_n a_1 b_1)(p_n + q_n) && (20)\\
&= p_n^2 a_0 b_0 + q_n^2 a_1 b_1 + p_n q_n (a_1 b_1 + a_0 b_0) && (21)\\
&\geqslant p_n^2 a_0 b_0 + q_n^2 a_1 b_1 + p_n q_n (a_1 b_0 + a_0 b_1) && (22)\\
&= (p_n a_0 + q_n a_1)(p_n b_0 + q_n b_1) && (23)\\
&= \Pr[A] \Pr[B]\,. && (24)
\end{aligned}$$

In eq. 20 we used the inductive hypothesis on A_0, B_0 and A_1, B_1 and the trivial fact that $p_n + q_n = 1$. In eq. 22 we just used eq. 18. The rest are just algebraic manipulations or the definition of $\Pr[A]$ and $\Pr[B]$. ■

Proof of Theorem 2.1 (Janson's inequality). Notice that for each i, A_i is up-closed and

$$\begin{aligned}
\Pr[\bigcap_{i\in[m]} A_i^c] &= \prod_{i\in[m]} \Pr[A_i^c \mid \bigcap_{j=1}^{i-1} A_j^c] && (25)\\
&= \prod_{i\in[m]} \left(1 - \Pr[A_i \mid \bigcap_{j=1}^{i-1} A_j^c]\right) && (26)\\
&\leqslant \exp\left(-\sum_{i\in[m]} \Pr[A_i \mid \bigcap_{j=1}^{i-1} A_j^c]\right)\,, && (27)
\end{aligned}$$

where we used that $1 - x \leqslant e^{-x}$ and so we just need to lower bound $\Pr[A_i \mid \bigcap_{j=1}^{i-1} A_j^c]$.[3] We will apply at some point the Harris-Kleitman inequality but first we need to massage $\Pr[A_i \mid \bigcap_{j=1}^{i-1} A_j^c]$. Fix i, we split $\bigcap_{j=1}^{i-1} A_j^c$ as follows

$$\bigcap_{j=1}^{i-1} A_j^c = \left(\bigcap_{j\leqslant i-1 : V_i \cap V_j = \emptyset} A_j^c\right) \cap \left(\bigcap_{j\leqslant i-1 : V_i \cap V_j \neq \emptyset} A_j^c\right)\,.$$

Call the first intersection B and denote with $j \sim i$ the fact that $j \leqslant i - 1$ and $V_i \cap V_j \neq \emptyset$. Now, by the definition of the probability space $\{0,1\}^n$, all coordinates are set independently

[3] In general A_j^c is down-closed and any arbitrary intersection of down-closed is down-closed, but the Harris-Kleitman inequality gives us $\Pr[A_i \mid \bigcap_{j=1}^{i-1} A_j^c] \leqslant \Pr[A_i]$ which is useless since we need a lower bound.

and hence $\Pr[A_i \mid B] = \Pr[A_i]$. We then have the following chain of inequalities:

$$\begin{aligned}
\Pr[A_i \mid \bigcap_{j=1}^{i-1} A_j^c] &= \Pr[A_i \mid B \cap \bigcap_{j\sim i} A_j^c] && (28)\\
&\geqslant \Pr[A_i \cap \bigcap_{j\sim i} A_j^c \mid B] && (29)\\
&= \Pr[A_i \mid B]\Pr[\bigcap_{j\sim i} A_j^c \mid A_i \cap B] && (30)\\
&\geqslant \Pr[A_i \mid B](1 - \sum_{j\sim i} \Pr[A_j \mid A_i \cap B]) && (31)\\
&= \Pr[A_i \mid B] - \sum_{j\sim i} \Pr[A_i \mid B] \cdot \Pr[A_j \mid A_i \cap B] && (32)\\
&= \Pr[A_i] - \sum_{j\sim i} \Pr[A_i \cap A_j \mid B] && (33)\\
&\geqslant \Pr[A_i] - \sum_{j\sim i} \Pr[A_i \cap A_j]\,. && (34)
\end{aligned}$$

In eq. 29 we used the trivial fact, following from Bayes theorem, that for every triple of events E_1, E_2, E_3, $\Pr[E_1 \mid E_2 \cap E_3] \geqslant \Pr[E_1 \cap E_2 \mid E_3]$. In eq. 33 we used that $\Pr[A_i \mid B] = \Pr[A_i]$ and the definition of conditional probability. In eq. 34 we used the Harris-Kleitman inequality applied to $A_i \cap A_j$, which is up-closed, and B, which is down closed.

Substituting the bound from eq. 34 into eq. 27 we get immediately the statement of the theorem noticing that $\Delta = 2\sum_{i=1}^{m}\sum_{j\sim i}\Pr[A_i \cap A_j]$. ■