# Supplementary Materials for Variable Elimination in the Fourier Domain


**Yexiang Xue** YEXIANG@CS.CORNELL.EDU
Cornell University, Ithaca, NY, 14853, USA

**Stefano Ermon** ERMON@CS.STANFORD.EDU
Stanford University, Stanford, CA, 94305, USA

**Ronan Le Bras, Carla P. Gomes, Bart Selman** {LEBRAS,GOMES,SELMAN}@CS.CORNELL.EDU
Cornell University, Ithaca, NY, 14853, USA


## 1. Proof of Lemma 1

Let $R_n^l$ be the collection of restrictions on $n$ Boolean variables $x_1, \dots, x_n$. Each restriction in $R_n^l$ leaves a set of $l$ variables $J = \{x_{i_1}, \dots, x_{i_l}\}$ open, while it fixes all other variables $x_i \notin J$ to either -1 or 1. It is easy to see that the size of $R_n^l$ is given by:

$$|R_n^l| = \binom{n}{l} \cdot 2^{n-l}. \tag{1}$$

For a restriction $J|\mathbf{z} \in R_n^l$, call $J|\mathbf{z}$ *bad* if and only if for all decision tree $h$ with depth $t$, there exists at least one input $\mathbf{x}_J$, such that $|h(\mathbf{x}_J) - f|_{J|\mathbf{z}}(\mathbf{x}_J)| > \gamma$. Let $B_n^l$ be the set of all bad restrictions, ie: $B_n^l = \{J|\mathbf{z} \in R_n^l : J|\mathbf{z} \text{ is bad}\}$. To prove the lemma, it is sufficient to prove that

$$\frac{|B_n^l|}{|R_n^l|} \leq \frac{1}{2}\left(\frac{l}{n-l} 8uw\right)^t. \tag{2}$$

In the proof that follows, for every bad restriction $\rho \in B_n^l$, we establish a bijection between $\rho$ and $(\xi, s)$, in which $\xi$ is a restriction in $R_n^{l-t}$ and $s$ is a certificate from a witness set $A$. In this case, the number of distinct $\rho$'s is bounded by the number of $(\xi, s)$ pairs:

$$|B_n^l| \leq |R_n^{l-t}| \cdot |A|. \tag{3}$$

For a restriction $\rho$, we form the *canonical decision tree* for $f|_\rho$ under precision $\gamma$ as follows:

1. We start with a fixed order for the variables and another fixed order for the factors.

2. If $f|_\rho$ is a constant function, or $||f|_\rho||_\infty \leq \gamma$, stop.

3. Otherwise, under restriction $\rho$, some factors evaluate to fixed values (all variables in these factors are fixed or there are free variables, but all assignments to these free variables lead to value 1), while other factors do not. Examine the factors according to the fixed factor order until reaching the first factor that still does not evaluate to a fixed value.

4. Expand the open variables of this factor, under the fixed variable order specified in step 1. The result will be a tree (The root branch is for the first open variable. The branches in the next level is for the second open variable, etc).

5. Each leaf of this tree corresponds to $f|_{\rho\pi_1}$, in which $\pi_1$ is a value restriction for all open variables of the factor. Recursively apply step 2 to 5 for function $f|_{\rho\pi_1}$, until the condition in step 2 holds. Then attach the resulting tree to this leaf.

Figure 1 provides a graphical demonstration of a canonical decision tree.

Now suppose restriction $\rho$ is bad. By definition, for any decision tree of depth $t$, there exists at least one input $\mathbf{x}$, such that $|h(\mathbf{x}) - f|_\rho(\mathbf{x})| > \gamma$. The canonical decision tree is no exception. Therefore, there must be a path $l$ in the canonical decision tree of $f|_\rho$, which has more than $t$ variables. Furthermore, these $t$ variables can be split into $k$ ($1 \leq k \leq t$) segments, each of which corresponds to one factor. Let $f_i$ ($i \in \{1, \dots, k\}$) be these factors, and let $\pi_i$ be the assignments of the free variables for $f_i$ in path $l$. Now for each factor $f_i$, by the definition of the canonical decision tree, under the restriction $\rho\pi_1 \dots \pi_{i-1}$, $f_i|\rho\pi_1 \dots \pi_{i-1}$ must have a branch whose value is no greater than $\eta$ (otherwise $f_i|\rho\pi_1 \dots \pi_{i-1}$ all evaluates to 1). We call this branch the "compressing" branch for factor $f_i|\rho\pi_1 \dots \pi_{i-1}$. Let the variable assignment which leads to this compressing branch for $f_i|\rho\pi_1 \dots \pi_{i-1}$ be $\sigma_i$. Let $\sigma = \sigma_1 \dots \sigma_k$. Then we map the bad restriction $\rho$ to $\rho\sigma$ and an auxiliary advice string that we are going to describe.

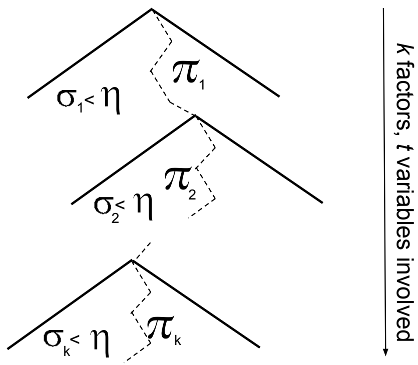


*Figure 1.* A graphical illustration of a canonical decision tree.

It is self-explanatory that we can map from any bad restriction $\rho$ to $\rho\sigma$. The auxiliary advice is used to establish the backward mapping, i.e. the mapping from $\rho\sigma$ to $\rho$. When we look at the result of $f|\rho\sigma$, we will notice that at least one factor is set to its compressing branch (because we set $f_1$ to its compressing branch in the forward mapping). Now there could be other factors set at their compressing branches (because of $\rho$), but an important observation is that: *the number of factors at their compressing branches cannot exceed* $u = \lceil \log_\eta \gamma \rceil + 1$, because otherwise, the other $u-1$ factors already render $||f|\rho||_\infty \leq \gamma$, and the canonical decision tree should have stopped on expanding this branch. We therefore could record the index number of $f_1$ out of all the factors that are fixed at their compressing branches in the auxiliary advice string, so we can find $f_1$ in the backward mapping. Notice that this index number will be between 1 and $u$, so it takes $\log u$ bits to store it.

Now with the auxiliary information, we can identify which factor is $f_1$. The next task is to identify which variables in $f_1$ are fixed by $\rho$, and which are fixed by $\sigma_1$. Moreover, if one variable is fixed by $\sigma_1$, we would like to know its correct values in $\pi_1$. To do this, we introduce additional auxiliary information: for each factor $f_i$, suppose it has $r_i$ free variables under restriction $f_i|\rho\pi_1\ldots\pi_{i-1}$, we use $r_i$ integers to mark the indices of these free variables. Because each $f_i$ is of width at most $w$, every integer of this type is between 1 and $w$ (therefore can be stored in $\log w$ bits). Also, it requires $t$ integers of this type in total to keep this information, because we have $t$ free variables in total for $f_1, \ldots, f_k$.

Notice that it is not sufficient to keep these integers. We further need $k-1$ separators, which tell which integer belongs to which factor $f_i$. Aligning these integers in a line, we need $k-1$ separators to break the line into $k$ segments. These separators can be represented by $t-1$ bits, in which the $i$-th bit is 1 if and only if there is a separator between the $i$-th and ($i$+1)-th integer (we have $t$ integers at most). With these two pieces of information, we are able to know the locations of free variables set by $\sigma_i$ for each factor $f_i$.

We further need to know the values for each variable in $\pi_i$. Therefore, we add in another $t$-bit string, each bit is either 0 or 1. 0 means the assignment of the corresponding variable in $\pi_i$ is the same as the one in $\sigma_i$, 1 means the opposite.

With all this auxiliary information, we can start from $\rho\sigma$, find the first factor $f_1$, further identify which variables are set by $\sigma_1$ in $f_1$, and set back its values in $\pi_1$. Then we start with $f|\pi_1$, we can find $\pi_2$ in the same process, and continue. Finally, we will find all variables in $\sigma$ and back up the original restriction $\rho$.

Now to count the length of the auxiliary information, the total length is $t\log u + t\log w + 2t - 1$ bits. Therefore, we can have a one-to-one mapping between elements in $B_n^l$ and $R_n^{l-t} \times A$, in which the size of $A$ is bounded by $2^{t\log u + t\log w + 2t - 1} = (uw)^t \cdot 2^{2t-1}$.

In all,

$$\frac{|B_n^l|}{|R_n^l|} \leq \frac{\binom{n}{l-t} 2^{n-l+t} (uw)^t \cdot 2^{2t-1}}{\binom{n}{l} 2^{n-l}} \tag{4}$$

$$= \frac{\binom{n}{l-t} \frac{1}{2} (8uw)^t}{\binom{n}{l}} \tag{5}$$

$$= \frac{l(l-1)\ldots(l-t+1)}{(n-l+1)\ldots(n-l+t)} \frac{1}{2} (8uw)^t \tag{6}$$

$$\leq \frac{1}{2} \left( \frac{l}{n-l} 8uw \right)^t. \tag{7}$$

## 2. Proof of Theorem 3

For each term in the Fourier expansion whose degree is less than or equal to $d$, we can treat this term as a weighted function involving less than or equal to $d$ variables. Therefore, it can be represented by a decision tree, in which each path of the tree involves no more than $d$ variables (therefore the tree is at most at the depth of $d$). Because $f$ is represented as the sum over a set of Fourier terms up to degree $d$, it can be also represented as the sum of the corresponding decision trees.

## 3. Proof of Theorem 6

Let the Fourier expansion of $f$ be: $f(\mathbf{x}) = \sum_S \hat{f}(S)\chi_S(\mathbf{x})$, we have:

$$
\begin{aligned}
& f'(\mathbf{x} \setminus x_i) \\
=& f(\mathbf{x} \setminus x_i, x_i = +1) + f(\mathbf{x} \setminus x_i, x_i = -1) \\
=& \sum_{S: i \in S} \hat{f}(S) \cdot \chi_{S \setminus i}(\mathbf{x} \setminus x_i) \cdot 1 + \\
& \sum_{S: i \notin S} \hat{f}(S) \cdot \chi_{S \setminus i}(\mathbf{x} \setminus x_i) \\
& + \sum_{S: i \in S} \hat{f}(S) \cdot \chi_{S \setminus i}(\mathbf{x} \setminus x_i) \cdot (-1) + \\
& \sum_{S: i \notin S} \hat{f}(S) \cdot \chi_{S \setminus i}(\mathbf{x} \setminus x_i) \\
=& \sum_{S: i \in S} \hat{f}(S) \cdot \chi_{S \setminus i}(\mathbf{x} \setminus x_i) \cdot 1 + \\
& \sum_{S: i \in S} \hat{f}(S) \cdot \chi_{S \setminus i}(\mathbf{x} \setminus x_i) \cdot (-1) \\
& + \sum_{S: i \notin S} \hat{f}(S) \cdot \chi_{S \setminus i}(\mathbf{x} \setminus x_i) + \\
& \sum_{S: i \notin S} \hat{f}(S) \cdot \chi_{S \setminus i}(\mathbf{x} \setminus x_i) \\
=& \sum_{S: i \notin S} 2 \cdot \hat{f}(S) \cdot \chi_{S \setminus i}(\mathbf{x} \setminus x_i).
\end{aligned}
$$