Computing Expected Shapley Values for Deterministic and Decomposable Circuits in Polynomial Time

Let V be a set of variables, 2^V its powerset. For $Y \subseteq V$ and $x \in Y$ we write $Y\backslash x$ for $Y\backslash\{x\}$. A *Boolean function on* V is a mapping $f : 2^V \to \{0,1\}$. For $x \in V$ let $p_x \in [0,1]$ be a probability value. For $Z \subseteq V' \subseteq V$ define

$$\Pi_{V'}(Z) \overset{\text{def}}{=} \left(\prod_{x \in Z} p_x\right) \times \left(\prod_{x \in V'\backslash Z} (1-p_x)\right).$$

(We note that the p-values do not appear in the notation $\Pi_{V'}(Z)$ but it should be clear from context in what follows.)

Definition 0.1. *Let $\varphi : 2^V \to \{0,1\}$ and probability values as above and $x \in V$. Define the* expected Shapley value of x for φ *as*

$$\mathsf{EShapley}(\varphi, x) \overset{\text{def}}{=} \sum_{\substack{Z \subseteq V \\ x \in Z}} \Pi_V(Z) \sum_{E \subseteq Z\backslash x} \frac{|E|!(|Z|-|E|-1)!}{|Z|!}(\varphi(E \cup \{x\}) - \varphi(E)).$$

The goal of this note is to show the following.

Theorem 0.2. *Given as input a deterministic and decomposable circuit (d-D) C on variables V and $x \in V$ and probability values we can compute* $\mathsf{EShapley}(C, x)$ *in PTIME.*

To prove this, we will first define some quantities, then show that computing the expected Shapley values can be reduced in PTIME to computing these quantities, and finally show that these quantities can be computed in PTIME. The proof is similar to that of [Arenas et al., 2023, Theorem 2], but different and more involved because we need to do a bidimensional parameterization.

Definition 0.3. *Let $\varphi : 2^V \to \{0,1\}$ and $0 \leqslant \ell \leqslant k \leqslant |V|$. Define*

$$\alpha^{\varphi}_{k,l} \overset{\text{def}}{=} \sum_{\substack{Z \subseteq V \\ |Z|=k}} \sum_{\substack{E \subseteq Z \\ |E|=\ell}} \Pi_V(Z)\varphi(E).$$

Reducing $\mathsf{EShapley}$ **to the quantities** $\alpha^g_{k,\ell}$**.** For $\varphi : 2^V \to \{0,1\}$ and $x \in V$ we denote by φ_{+x} (resp., φ_{-x}) the Boolean function on $V\backslash x$ that maps $Z \subseteq V\backslash x$ to $\varphi_{+x}(Z) \overset{\text{def}}{=} \varphi(Z \cup \{x\})$ (resp., $\varphi_{-x}(Z) \overset{\text{def}}{=} \varphi(Z)$). We explain the reduction next.

We have $\mathsf{EShapley}(\varphi, x) = A - B$ where

$$
\begin{aligned}
A &= \sum_{\substack{Z\subseteq V\\ x\in Z}} \Pi_V(Z) \sum_{E\subseteq Z\backslash x} \frac{|E|!(|Z|-|E|-1)!}{|Z|!}\varphi(E\cup\{x\})\\
B &= \sum_{\substack{Z\subseteq V\\ x\in Z}} \Pi_V(Z) \sum_{E\subseteq Z\backslash x} \frac{|E|!(|Z|-|E|-1)!}{|Z|!}\varphi(E).
\end{aligned}
$$

Let us focus on A. Letting $V' \overset{\text{def}}{=} V\backslash x$, notice that these are the variables over which φ_{+x} is defined. Letting $n \overset{\text{def}}{=} |V'|$, we have

$$
\begin{aligned}
A &= \sum_{\substack{Z\subseteq V\\ x\in Z}} \Pi_V(Z) \sum_{E\subseteq Z\backslash x} \frac{|E|!(|Z|-|E|-1)!}{|Z|!}\varphi_{+x}(E)\\
&= p_x \sum_{Z\subseteq V'} \Pi_{V'}(Z) \sum_{E\subseteq Z} \frac{|E|!(|Z|-|E|)!}{(|Z|+1)!}\varphi_{+x}(E)\\
&= p_x \sum_{Z\subseteq V'} \sum_{E\subseteq Z} \frac{|E|!(|Z|-|E|)!}{(|Z|+1)!}\Pi_{V'}(Z)\varphi_{+x}(E)\\
&= p_x \sum_{k=0}^{n} \sum_{\substack{Z\subseteq V'\\ |Z|=k}} \sum_{\ell=0}^{k} \sum_{\substack{E\subseteq Z\\ |E|=\ell}} \frac{\ell!(k-\ell)!}{(k+1)!}\Pi_{V'}(Z)\varphi_{+x}(E)\\
&= p_x \sum_{k=0}^{n} \sum_{\ell=0}^{k} \frac{\ell!(k-\ell)!}{(k+1)!} \sum_{\substack{Z\subseteq V'\\ |Z|=k}} \sum_{\substack{E\subseteq Z\\ |E|=\ell}} \Pi_{V'}(Z)\varphi_{+x}(E)\\
&= p_x \sum_{k=0}^{n} \sum_{\ell=0}^{k} \frac{\ell!(k-\ell)!}{(k+1)!}\alpha^{\varphi_{+x}}_{k,\ell}.
\end{aligned}
$$

We can do exactly the same for B (replacing φ_{+x} by φ_{-x}), after which we obtain that

$$
\begin{aligned}
\mathsf{EShapley}(\varphi, x) &= A - B\\
&= p_x \sum_{k=0}^{n} \sum_{\ell=0}^{k} \frac{\ell!(k-\ell)!}{(k+1)!}(\alpha^{\varphi_{+x}}_{k,\ell} - \alpha^{\varphi_{-x}}_{k,\ell}).
\end{aligned}
$$

This concludes the reduction.

Computing the quantities $\alpha^{g}_{k,\ell}$. We now explain how to compute the quantities $\alpha^{\varphi}_{k,\ell}$ in PTIME for d-Ds. Let C be our input d-D circuit, on variables V. We assume that C is smooth, which is without loss of generality as this can easily be enforced in quadratic time [Shih et al., 2019]. For a gate $g \in C$ let V_g denote the set of variables that have a directed path to g in C, and denote by C_g the subcircuit rooted at g (that we identify with the Boolean function on V_g that it captures). We first explain how to compute an intermediate quantity will be needed later.

Definition 0.4. *For a gate $g \in C$ and integer $0 \leqslant k \leqslant |V_g|$, define*

$$\delta^{g}_{k} \overset{\text{def}}{=} \sum_{\substack{Z \subseteq V_g \\ |Z|=k}} \Pi_{V_g}(Z).$$

Notice that δ^{g}_{k} only depends on the "structure" of the circuit, but not on its semantics.

Lemma 0.5. *We can compute in PTIME all the quantities δ^{g}_{k}.*

Proof. We compute them by bottom-up induction on C.

Input gates. Let g be an input gate, with variable x. Then $V_g = \{x\}$, so we simply need to compute δ^{g}_{0} and δ^{g}_{1}. We have $\delta^{g}_{0} = \Pi_{V_g}(\varnothing) = 1 - p_x$ and $\delta^{g}_{1} = \Pi_{V_g}(\{x\}) = p_x$.

Negation gates. Let g be a $\neg$-gate with input g'. Notice that $V_g = V_{g'}$. So we have $\delta^{g}_{k} = \delta^{g'}_{k}$ for all $0 \leqslant k \leqslant |V_g|$ and we are done since the values $\delta^{g'}_{k}$ have already been computed inductively.

Deterministic smooth $\vee$-gates. Let g be a smooth deterministic $\vee$-gate with inputs g_1, g_2. Since g is smooth we have $V_g = V_{g_1} = V_{g_2}$. In particular we have $\delta^{g}_{k} = \delta^{g_1}_{k}$ for all $0 \leqslant k \leqslant |V_g|$ and we are done.

Decomposable $\wedge$-gates. Let g be a decomposable $\wedge$-gate with inputs g_1, g_2. Notice that $V_g = V_{g_1} \cup V_{g_2}$ with the union being disjoint. We can then decompose Z into a "left" part $Z_1 \subseteq V_{g_1}$ of size $k_1 \in \{0, \ldots, k\}$ and a "right" part $Z_2 \subseteq V_{g_2}$ of size $k - k_1$. For the summation to make sense we need $k_1 \leqslant \min(k, |V_{g_1}|)$ as well as $k - k_1 \leqslant |V_{g_2}|$ (and $k_1 \geqslant 0$). We then have:

$$\begin{aligned}
\delta^g_k &= \sum_{\substack{Z \subseteq V_g \\ |Z|=k}} \Pi_{V_g}(Z) \\
&= \sum_{k_1=\max(0,k-|V_{g_2}|)}^{\min(k,|V_{g_1}|)} \sum_{\substack{Z_1 \subseteq V_{g_1} \\ |Z_1|=k_1}} \sum_{\substack{Z_2 \subseteq V_{g_2} \\ |Z_2|=k-k_1}} \Pi_{V_{g_1}}(Z_1)\Pi_{V_{g_2}}(Z_2) \\
&= \sum_{k_1=\max(0,k-|V_{g_2}|)}^{\min(k,|V_{g_1}|)} \sum_{\substack{Z_1 \subseteq V_{g_1} \\ |Z_1|=k_1}} \Pi_{V_{g_1}}(Z_1) \sum_{\substack{Z_2 \subseteq V_{g_2} \\ |Z_2|=k-k_1}} \Pi_{V_{g_2}}(Z_2) \\
&= \sum_{k_1=\max(0,k-|V_{g_2}|)}^{\min(k,|V_{g_1}|)} \sum_{\substack{Z_1 \subseteq V_{g_1} \\ |Z_1|=k_1}} \Pi_{V_{g_1}}(Z_1)\delta^{g_2}_{k-k_1} \\
&= \sum_{k_1=\max(0,k-|V_{g_2}|)}^{\min(k,|V_{g_1}|)} \delta^{g_2}_{k-k_1} \sum_{\substack{Z_1 \subseteq V_{g_1} \\ |Z_1|=k_1}} \Pi_{V_{g_1}}(Z_1) \\
&= \sum_{k_1=\max(0,k-|V_{g_2}|)}^{\min(k,|V_{g_1}|)} \delta^{g_1}_{k_1}\delta^{g_2}_{k-k_1},
\end{aligned}$$

and we are done.

This concludes the proof of Lemma 0.5. □

We next define α-quantities for all gates of the circuit C.

Definition 0.6. *For a gate $g \in C$ and $0 \leqslant \ell \leqslant k \leqslant |V_g|$, define*

$$\alpha^g_{k,l} \overset{\text{def}}{=} \sum_{\substack{Z \subseteq V_g \\ |Z|=k}} \sum_{\substack{E \subseteq Z \\ |E|=\ell}} \Pi_{V_g}(Z)C_g(E).$$

If we can show that we can compute all quantities $\alpha^g_{k,l}$ then we are done: indeed, we can then take g to be the output gate of C, which gives us the quantities $\alpha^C_{k,l}$ that we wanted. We show just that in the next lemma.

Lemma 0.7. *We can compute in PTIME all the quantities $\alpha^g_{k,l}$.*

Proof. This is again done by bottom-up induction on C.

Input gates. Let g be an input gate, with variable x. Then $V_g = \{x\}$, so we simply need to compute $\alpha^g_{0,0}$, $\alpha^g_{1,0}$ and $\alpha^g_{1,1}$. One can easily check that $\alpha^g_{0,0} = \alpha^g_{1,0} = 0$ and $\alpha^g_{1,1} = p_x$.

Negation gates. Let g be a $\neg$-gate with input g'. Notice that $V_g = V_{g'}$ and that $C_g(E) = 1 - C_{g'}(E)$ for any $E \subseteq V_g$. We have

$$\begin{aligned}\alpha^g_{k,l} &= \sum_{\substack{Z \subseteq V_g \\ |Z|=k}} \sum_{\substack{E \subseteq Z \\ |E|=\ell}} \Pi_{V_g}(Z)(1 - C_{g'}(E)) \\ &= \left[\binom{k}{\ell} \sum_{\substack{Z \subseteq V_g \\ |Z|=k}} \Pi_{V_g}(Z)\right] - \alpha^{g'}_{k,\ell} \\ &= \binom{k}{\ell}\delta^g_k - \alpha^{g'}_{k,\ell},\end{aligned}$$

and we are done thanks to Lemma 0.5.

Deterministic smooth $\vee$-gates. Let g be a smooth deterministic $\vee$-gate with inputs g_1, g_2. Since g is smooth we have $V_g = V_{g_1} = V_{g_2}$, and since it is deterministic we have $C_g(E) = C_{g_1}(E) + C_{g_1}(E)$ for any $E \subseteq V_g$. Therefore we obtain $\alpha^g_{k,l} = \alpha^{g_1}_{k,l} + \alpha^{g_2}_{k,l}$ and we are done.

Decomposable $\wedge$-gates. Let g be a decomposable $\wedge$-gate with inputs g_1, g_2. Notice that $V_g = V_{g_1} \cup V_{g_2}$ with the union being disjoint, and that $C_g(E) = C_{g_1}(E \cap V_{g_1}) \times C_{g_2}(E \cap V_{g_2})$ and $\Pi_{V_g}(Z) = \Pi_{V_{g_1}}(Z \cap V_{g_1}) \times \Pi_{V_{g_2}}(Z \cap V_{g_2})$ for any $Z, E \subseteq V_g$. We decompose the summations over Z and E as we did in the proof of Lemma 0.5 for $\wedge$-gates. For readability we use colors to point out which parts of the expressions are modified or moved around.

$$
\begin{aligned}
\alpha^{g}_{k,l} &= \sum_{\substack{Z\subseteq V_g\\ |Z|=k}} \sum_{\substack{E\subseteq Z\\ |E|=\ell}} \Pi_{V_g}(Z)C_g(E)\\
&= \sum_{k_1=\max(0,k-|V_{g_2}|)}^{\min(k,|V_{g_1}|)} \sum_{\substack{Z_1\subseteq V_{g_1}\\ |Z_1|=k_1}} \sum_{\substack{Z_2\subseteq V_{g_2}\\ |Z_2|=k-k_1}} \sum_{\substack{E\subseteq Z_1\cup Z_2\\ |E|=\ell}} \Pi_{V_{g_1}}(Z_1)\Pi_{V_{g_2}}(Z_2)C_g(E)\\
&= \sum_{k_1=\max(0,k-|V_{g_2}|)}^{\min(k,|V_{g_1}|)} \sum_{\substack{Z_1\subseteq V_{g_1}\\ |Z_1|=k_1}} \sum_{\substack{Z_2\subseteq V_{g_2}\\ |Z_2|=k-k_1}} \sum_{\ell_1=\max(0,\ell-k+k_1)}^{\min(k_1,\ell)} \sum_{\substack{E_1\subseteq Z_1\\ |E_1|=\ell_1}} \sum_{\substack{E_2\subseteq Z_2\\ |E_2|=\ell-\ell_1}}\\
&\qquad \Pi_{V_{g_1}}(Z_1)\Pi_{V_{g_2}}(Z_2)C_{g_1}(E_1)C_{g_2}(E_2)\\
&= \sum_{k_1=\max(0,k-|V_{g_2}|)}^{\min(k,|V_{g_1}|)} \sum_{\ell_1=\max(0,\ell-k+k_1)}^{\min(k_1,\ell)} \sum_{\substack{Z_1\subseteq V_{g_1}\\ |Z_1|=k_1}} \sum_{\substack{Z_2\subseteq V_{g_2}\\ |Z_2|=k-k_1}} \sum_{\substack{E_1\subseteq Z_1\\ |E_1|=\ell_1}} \sum_{\substack{E_2\subseteq Z_2\\ |E_2|=\ell-\ell_1}}\\
&\qquad \Pi_{V_{g_1}}(Z_1)\Pi_{V_{g_2}}(Z_2)C_{g_1}(E_1)C_{g_2}(E_2)\\
&= \sum_{k_1=\max(0,k-|V_{g_2}|)}^{\min(k,|V_{g_1}|)} \sum_{\ell_1=\max(0,\ell-k+k_1)}^{\min(k_1,\ell)} \sum_{\substack{Z_1\subseteq V_{g_1}\\ |Z_1|=k_1}} \sum_{\substack{E_1\subseteq Z_1\\ |E_1|=\ell_1}} \sum_{\substack{Z_2\subseteq V_{g_2}\\ |Z_2|=k-k_1}} \sum_{\substack{E_2\subseteq Z_2\\ |E_2|=\ell-\ell_1}}\\
&\qquad \Pi_{V_{g_1}}(Z_1)\Pi_{V_{g_2}}(Z_2)C_{g_1}(E_1)C_{g_2}(E_2)\\
&= \sum_{k_1=\max(0,k-|V_{g_2}|)}^{\min(k,|V_{g_1}|)} \sum_{\ell_1=\max(0,\ell-k+k_1)}^{\min(k_1,\ell)} \sum_{\substack{Z_1\subseteq V_{g_1}\\ |Z_1|=k_1}} \sum_{\substack{E_1\subseteq Z_1\\ |E_1|=\ell_1}} \Pi_{V_{g_1}}(Z_1)C_{g_1}(E_1)\\
&\qquad \sum_{\substack{Z_2\subseteq V_{g_2}\\ |Z_2|=k-k_1}} \sum_{\substack{E_2\subseteq Z_2\\ |E_2|=\ell-\ell_1}} \Pi_{V_{g_2}}(Z_2)C_{g_2}(E_2)\\
&= \sum_{k_1=\max(0,k-|V_{g_2}|)}^{\min(k,|V_{g_1}|)} \sum_{\ell_1=\max(0,\ell-k+k_1)}^{\min(k_1,\ell)} \sum_{\substack{Z_1\subseteq V_{g_1}\\ |Z_1|=k_1}} \sum_{\substack{E_1\subseteq Z_1\\ |E_1|=\ell_1}} \Pi_{V_{g_1}}(Z_1)C_{g_1}(E_1)\alpha^{g_2}_{k-k_1,\ell-\ell_1}\\
&= \sum_{k_1=\max(0,k-|V_{g_2}|)}^{\min(k,|V_{g_1}|)} \sum_{\ell_1=\max(0,\ell-k+k_1)}^{\min(k_1,\ell)} \alpha^{g_2}_{k-k_1,\ell-\ell_1} \sum_{\substack{Z_1\subseteq V_{g_1}\\ |Z_1|=k_1}} \sum_{\substack{E_1\subseteq Z_1\\ |E_1|=\ell_1}} \Pi_{V_{g_1}}(Z_1)C_{g_1}(E_1)\\
&= \sum_{k_1=\max(0,k-|V_{g_2}|)}^{\min(k,|V_{g_1}|)} \sum_{\ell_1=\max(0,\ell-k+k_1)}^{\min(k_1,\ell)} \alpha^{g_1}_{k_1,\ell_1} \times \alpha^{g_2}_{k-k_1,\ell-\ell_1}.
\end{aligned}
$$

and we are done.

This concludes the proof of Lemma 0.7. □

We also need to argue that the number of bits of all these quantities stays polynomial (in fact linear?), otherwise we cannot do the arithmetic operations; this should come from the fact that $\wedge$-gates are decomposable. Formally, to show this, we would need to find the magical constant K such that $\alpha^g_{k,\ell} \leqslant 2^{K|V_g|}$, i.e., the *value* of $\alpha^g_{k,\ell}$ itself is at most an exponential in $|V_g|$, so its number of bits is linear in $|V_g|$.

References

[Arenas et al., 2023] Arenas, M., Barceló, P., Bertossi, L. E., and Monet, M. (2023). On the Complexity of SHAP-Score-Based Explanations: Tractability via Knowledge Compilation and Non-Approximability Results. *J. Mach. Learn. Res.*, 24(63):1–58.

[Shih et al., 2019] Shih, A., Van den Broeck, G., Beame, P., and Amarilli, A. (2019). Smoothing structured decomposable circuits. *Advances in Neural Information Processing Systems*, 32.