O Teorema de Karp-Lipton

Nicollas Sdroievski

10 de Junho de 2019

1 Introdução

O teorema de Karp-Lipton garante que, caso $\mathsf{NP} \subset \mathsf{P/poly}$, então a hierarquia polinomial colapsa para o segundo nível, ou seja $\mathsf{PH} = \Sigma_2^p$. De forma parecida com outros teoremas deste tipo, a ideia é mostrar que a hipótese $\mathsf{NP} \subset \mathsf{P/poly}$ é pouco provável.

Um ingrediente fundamental da demonstração é o fato de que o problema SAT (e todos os problemas NP-completos) é "*downward self-reducible*". Essa propriedade nos permite encontrar uma valoração que satisfaz uma fórmula (satisfazível) dado um oráculo para o problema de decisão. Ilustramos essa propriedade na seguinte afirmação, que será útil na demosntração do teorema principal.

Afirmação 1. *Se* $\mathsf{SAT} \in \mathsf{P/poly}$*, então para todo* $n \in \mathbb{N}$ *existe um circuito* C'_n *de tamanho polinomial que, ao receber como entrada uma fórmula satisfazível* φ*, tem como saída uma valoração* v *tal que* $\varphi(v) = 1$*.*

Demonstração. Assumindo $\mathsf{SAT} \in \mathsf{P/poly}$, existe uma família de circuitos $\{C_n\}_{n\in\mathbb{N}}$ tal que cada circuito C_n possui $|C_n| = O(n^c)$ para alguma constante c fixa, e decide instâncias de tamanho n de SAT.

Usamos o circuito C_n para construir uma MT M_n que ao receber uma fórmula satisfazível tem como saída uma valoração que a satisfaz. A ideia principal é usar a seguinte propriedade do problema SAT:

$$\varphi \in \mathsf{SAT} \iff (\varphi|x = 1) \in \mathsf{SAT} \vee (\varphi|x = 0) \in \mathsf{SAT}.$$

Onde x é uma variável qualquer de φ e $\varphi|x = b$ indica que a variável x é trocada pela constante b.

O comportamento de M_n é o seguinte: na entrada φ com $|\varphi| = n$ e variáveis $x_1, x_2, \ldots, x_m$, faça $\psi = \varphi$ e para i de 1 até m:

1. Se $C_n(\psi|x_i = 1) = 1$, então marque x_i como 1. Faça $\psi = (\psi|x_i = 1)$
2. Senão, marque x_i como 0. Faça $\psi = (\psi|x_i = 0)$.

Ao fim da execução retorne os valores de $x_1, x_2, \ldots, x_n$.

Perceba que a cada execução do laço a seguinte invariante é preservada: se a fórmula φ é satisfazível, então a fórmula ψ é satisfazível. Isso garante que ao fim da execução, caso φ seja satisfazível, a valoração encontrada realmente satisfaz φ. Além disso, como o tamanho de C_n é polinomial e avaliamos $C_n(\psi)$ no máximo n vezes, temos que o tempo de execução de M_n é $\mathsf{poly}(n)$.

Para obter o circuito C'_n, basta aplicar a transformação padrão de MT para circuito na maquina M_n, que resultará em um circuito C'_n de tamanho polinomial com as características desejadas. □

2 Teorema e Demonstração

Teorema 1. *Se* $\mathsf{NP} \subset \mathsf{P/poly}$*, então* $\mathsf{PH} = \Sigma_2^p$.

Demonstração. Assuma $\mathsf{NP} \subset \mathsf{P/poly}$. É suficiente mostrar que $\Pi_2^p \subseteq \Sigma_2^p$ pois $\Pi_2^p = \mathsf{co}\Sigma_2^p$. Para isso basta mostrar que $\Pi_2\mathsf{SAT} \in \Sigma_2^p$. Seja φ uma instância de $\Pi_2\mathsf{SAT}$, logo:

$$\varphi \in \Pi_2\mathsf{SAT} \iff \forall u \in \{0,1\}^m \exists v \in \{0,1\}^m \mid \varphi(u,v) = 1. \tag{1}$$

Fixando uma valoração u obtemos uma instância $\varphi(u)$ de SAT, assim:

$$\varphi(u) \in \mathsf{SAT} \iff \exists v \in \{0,1\}^m \mid \varphi(u,v) = 1. \tag{2}$$

Seja $n = |\varphi(u)| \leq |\varphi|$. Pela hipótese do Teorema e pela Afirmação 1, temos que existe um circuito C'_n com $|C'_n| = \mathsf{poly}(n)$ tal que, caso exista uma valoração v que satisfaz $\varphi(u)$, a saída de $C'_n(\varphi(u))$ é justamente v.

Certamente a hipótese $\mathsf{NP} \subset \mathsf{P/poly}$ garante apenas a *existência* do circuito C'_n e não nos mostra como encontrá-lo. A ideia principal do Teorema de Karp-Lipton é que podemos adivinhar esse circuito usando um quantificador $\exists$ e então, como esse circuito retorna uma valoração, verificar se sua resposta é correta. Perceba que o circuito C'_n pode ser representado por uma string

de tamanho $p(n)$ para um polinômio em n. Neste caso, considere o seguinte predicado:

$$\exists w \in \{0,1\}^{p(n)} \forall u \in \{0,1\}^m \mid w \text{ codifica um circuito } C \text{ e } \varphi(u, C(\varphi(u)) = 1. \tag{3}$$

Afirmamos que (1) é verdade se e somente se (3) é verdade. Caso (1) seja verdade, então *para todo* u, *existe* um valor de v que satisfaz a fórmula $\varphi(u)$. Como visto o circuito C'_n é capaz de encontrar essa valoração, e logo (3) é verdadeira. Por outro lado, caso (1) seja falsa, então *existe* um valor de u tal que *nenhum* v satisfaz $\varphi(u)$, logo (3) também é falsa.

Como (3) pode ser computada em Σ_2^p, temos que $\Pi_2\mathsf{SAT} \in \Sigma_2^p$ e logo $\mathsf{PH} = \Sigma_2^p$. □