Fingerprinting

1 Deciding if a Polynomial is Equivalent to Zero

Basic version: Given a polynomial $Q(x)$, is $Q \equiv 0$ (*i.e.,* does $Q(x) = 0$ for all x)? This also lets us answer if $P \equiv Q$ (*i.e.,* if $P - Q \equiv 0$).

If $Q(x)$ is given explicitly as

$$Q(x) = \sum_{i=0}^{k} a_i x^i,$$

then $Q \equiv 0$ iff all $a_i = 0$. If $Q(x)$ is given as

$$Q(x) = \prod_{i=1}^{k} (x - b_i),$$

then $Q \not\equiv 0$. Thus, if Q is given explicitly as a sum or product, this problem is uninteresting. We will instead look at the case where Q can be evaluated efficiently on any input, but we have no explicit form (*i.e.,* we have an oracle for Q).

The idea is to try $Q(x)$ on a bunch of random inputs x; output zero iff it is zero everywhere. This may of course be wrong because we cannot query infinitely many points. In general, this will work for polynomials over any field $\mathbb{F}$, but we will use it only for $\mathbb{F} = \mathbb{Q}$ here.

Lemma 1. *If $Q : \mathbb{F} \to \mathbb{F}$ has degree d, and we choose a random input x uniformly from any set $S \subseteq \mathbb{F}$ of size $s = |S|$, then*

$$\Pr[Q(x) = 0 \mid Q \not\equiv 0] \leq \frac{d}{s}.$$

Of course, if $Q \equiv 0$, we always answer correctly.

Proof. Any polynomial of degree d that is not all zero has at most d zeros. □

A more interesting question is a generalization to multivariate polynomials.

Theorem 2. *[Schwartz-Zippel] Let $Q(x_1, \ldots, x_n)$ be a polynomial in n variables over $\mathbb{F}$ with degree d. If $x_1, \ldots, x_n$ are chosen independently and uniformly from $S \subseteq \mathbb{F}$ of size $s = |S|$, then*

$$\Pr[Q(x_1, \ldots, x_n) = 0 \mid Q \not\equiv 0] \leq \frac{d}{s}.$$

Here, the degree is the largest sum of exponents of any term.

Proof. By induction on n. The base case $n = 1$ is our earlier lemma. For the induction step, let $k \geq 0$ be the largest exponent of x_1 in any term. Write

$$Q(x_1, \ldots, x_n) = \sum_{i=0}^{k} x_1^i Q_i(x_2, \ldots, x_n).$$

Note $Q_k(x_2, \ldots, x_n)$ has degree at most $d - k$. And, $Q_k(x_2, \ldots, x_n) \not\equiv 0$ because the term x_1^k actually occurred in Q. By the induction hypothesis, the probability that $Q_k(x_2, \ldots, x_n) = 0$ (evaluates to 0 at the random points we sample) is at most $\frac{d-k}{s}$. Now, condition on $Q_k(x_2, \ldots, x_n) \neq 0$.

Define

$$p(y) \coloneqq \sum_{i=0}^{k} y^i Q_i(x_2, \ldots, x_n).$$

Here, all of the $Q_i(x_2, \ldots, x_n)$ are constants (we have sampled $x_2, \ldots, x_n$). Therefore p is a univariate degree k polynomial, and $p \neq 0$ because $Q_k(x_2, \ldots, x_n) \neq 0$. By the base case, we get that

$$\Pr[p(y) = 0] \leq \frac{k}{s}.$$

By a union bound,

$$\Pr[Q(x_1, \ldots, x_n) = 0] \leq \frac{d-k}{s} + \frac{k}{s} = \frac{d}{s}.$$

□

We can of course amplify the probability by drawing multiple independent samples.

2 Deciding if a Bipartite Graph has a Perfect Matching

2.1 Matching as Permutation

We are given a graph with $V = X \cup Y$, $|X| = |Y| = n$, and $E \subseteq X \times Y$. A matching M is a set of edges such that each vertex is the endpoint of at most one edge in M. A perfect matching is when each vertex is the endpoint of exactly one edge in M.

We represent the graph by its adjacency matrix A with

$$a_{ij} = \begin{cases} 1 & \text{if } (x_i, y_j) \in G \\ 0 & \text{o.w.} \end{cases}$$

We can think of M as a bijection/permutation

$$\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}.$$

So $\sigma(i) \neq \sigma(i')$ for $i \neq i'$, and $a_{i,\sigma(i)} = 1$ for all i.

Given any permutation, it encodes a perfect matching iff $a_{i,\sigma(i)} = 1$ for all i (*i.e.,* if the edges exist). In other words, iff

$$\prod_{i=1}^{n} a_{i,\sigma(i)} = 1.$$

To count all perfect matchings of G, we write

$$\operatorname{perm}(A) = \sum_{\sigma} \prod_{i=1}^{n} a_{i,\sigma(i)}.$$

The determinant of A is

$$\det(A) = \sum_{\sigma} (-1)^{\operatorname{sgn}(\sigma)} \prod_{i=1}^{n} a_{i,\sigma(i)}.$$

As a reminder, $\operatorname{sgn}(\sigma)$ is the number of inversions in σ. The determinant can be computed by writing $A = LU$ where L is lower triangular and U is upper triangular. (This can be done by solving a linear system). Then, $\det(A)$ is the product of all diagonal entries of L and U. But, computing $\operatorname{perm}(A)$ is $\#P$-complete. We will see approximation algorithms for adjacency matrices of bipartite graphs in a few weeks.

2.2 Matching and Polynomials

If $\text{perm}(A) = 0$ then $\det(A) = 0$, but not vice-versa. We will use our polynomial technique to help with this. Define the matrix B via

$$b_{ij} = \begin{cases} x_{ij} & \text{if } a_{ij} = 1 \\ 0 & \text{if } a_{ij} = 0 \end{cases}$$

That is, we introduce a new variable for each edge. Let Q_B be the multivariate polynomial

$$Q_B \coloneqq \det(B).$$

Theorem 3. *[Edmonds]. G has a perfect matching iff $Q_B \not\equiv 0$.*

Proof. If G has a perfect matching $\hat{\sigma}$, we set $x_{i,\hat{\sigma}(i)} = 1$ for all i and $x_{i,j} = 0$ for all $j \neq \hat{\sigma}(i)$. Then each product

$$\prod_{i=1}^{n} a_{i,\pi(i)}$$

has at least one zero term in it, so

$$\begin{aligned} \sum_{\sigma} (-1)^{\text{sgn}(\sigma)} \prod_{i=1}^{n} a_{i,\sigma(i)} &= (-1)^{\text{sgn}(\hat{\sigma})} \prod_{i=1}^{n} a_{i,\hat{\sigma}(i)} \\ &= (-1)^{\text{sgn}(\hat{\sigma})} \\ &\neq 0. \end{aligned}$$

In particular, $Q_B \not\equiv 0$. Conversely, if G has no perfect matching, then $\text{perm}(A) = 0$ so $\det(A) = 0$, which implies $Q_B \equiv 0$. □

The resulting algorithm is:

1. Compute B
2. Pick values x_{ij} i.i.d. uniformly from $S = \{1, \ldots, 2n\}$
3. Evaluate Q_B for those x_{ij}
4. If the answer is zero, output "no perfect matching"; otherwise, we output "perfect matching"

Because the determinant has degree at most n, by Schwartz-Zippel, with probability at least $1/2$, if $Q_B \not\equiv 0$ we output "matching", and if $Q_B \equiv 0$ we always give the right answer. This gives a Monte Carlo algorithm with success probability at least $1/2$, and we can repeat k times independently to boost the success probability to $1 - 2^{-k}$.

We cannot compute $Q_B = \det(B)$ generically (with variables), but we can compute it efficiently once we assign values. Thus, we have an oracle for Q_B, but an explicit representation could have exponentially many terms.

3 Extension to Non-Bipartite Matching

If we want to find a perfect matching in a non-bipartite graph, we start from the adjacency matrix A of G. In a bipartite graph, a_{ij} referred to the edge (x_i, y_j) while a_{ji} referred to (x_j, y_i) – these are different edges. In a general undirected graph, $a_{ij} = a_{ji}$ refers to the edge (i, j), so we should use the same variable on both.

We are worried about odd cycles, and we want them to cancel out. Define the Tutte matrix T as follows: for each edge (i, j) define a variable x_{ij}. Set

$$t_{ij} = \begin{cases} x_{ij} & \text{if } a_{ij} = 1 \text{ and } i < j \\ -x_{ij} & \text{if } a_{ij} = 1 \text{ and } i > j \\ 0 & \text{o.w.} \end{cases}$$

Theorem 4. *[Tutte]. The multivariate polynomial* $\det(T) \not\equiv 0$ *iff* G *has a perfect matching.*

We again identify perfect matchings with permutations σ such that (1) $a_{i,\sigma(i)} = 1$ for all i and (2) $\sigma^2 = Id$ (to prevent cycles).

Proof. We begin with the converse. Assume G has a perfect matching σ. For all $i < j$, set $x_{ij} = 1$ iff $\sigma(i) = j$ and zero otherwise. By the same proof as for Edmonds, all other terms of $\det(T)$ are zero, so $\det(T) \in \{-1, 1\}$. Therefore $\det(T) \not\equiv 0$. □