How to determine the number of terms of a polynomial.

Michael Monagan
Simon Fraser University

What is $\sum_{i=1}^{n}(2i - 1)$?

Is $\mathrm{rank}(AB) = \min(\mathrm{rank}(A), \mathrm{rank}(B))$?

What is the sign of the permutation (2 3 1) ?

Let $\mathbb{F}$ be a field and let $f = \sum_{j=1}^{t} c_j M_j(x_1, ..., x_n)$ be a polynomial where the coefficients c_j are non-zero in $\mathbb{F}$ and the monomials M_j are distinct. So t is the number of terms of f.

Let $\mathbf{B} : \mathbb{F}^n \to \mathbb{F}$ be a black box for f.
The monomials M_j, the coefficients c_j, and the number of terms t are unknown.
Problem: How can we determine t?

Erich Kaltofen and Wen-shin Lee.
Early termination in sparse interpolation algorithms.
J. Symb. Cmpt. **36**:365–400, 2003.

Let $\alpha \in \mathbb{F}^n$ and let $a_i = f(\alpha_1^i, \alpha_2^i, \ldots, \alpha_n^i)$. Define the $s \times s$ Hankel matrix

$$H_s = \begin{bmatrix} a_0 & a_1 & \cdots & a_{s-1} \\ a_1 & a_2 & \cdots & a_s \\ \vdots & \vdots & & \vdots \\ a_{s-1} & a_s & \cdots & a_{2s-2} \end{bmatrix}.$$

Let $m_j = M_j(\alpha_1, \ldots, \alpha_n)$.
Because $M_j(\alpha_1^i, \ldots, \alpha_n^i) = m_j^i$ we have the following factorization: $H_s = V_s^T D_t V_s$ where

$$H_s = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ m_1 & m_2 & \cdots & m_t \\ m_1^2 & m_2^2 & \cdots & m_t^2 \\ \vdots & \vdots & & \vdots \\ m_1^{s-1} & m_2^{s-1} & \cdots & m_t^{s-1} \end{bmatrix} \begin{bmatrix} c_1 & 0 & \cdots & 0 \\ 0 & c_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & c_t \end{bmatrix} \begin{bmatrix} 1 & m_1 & m_1^2 & \cdots & m_1^{s-1} \\ 1 & m_2 & m_2^2 & \cdots & m_2^{s-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & m_t & m_t^2 & \cdots & m_t^{s-1} \end{bmatrix}$$

Case $s = t$: if $m_i \neq m_j$ then $\det(V_t) = \prod_{1 \leq i < j \leq t} (m_j - m_i) \neq 0$.
Since $\det(V_t^T) = \det(V_t)$ and $\det(D_t) \neq 0$, we have $\det(H_t) \neq 0$.

Case $s > t$: if $m_i \neq m_j$ then $\operatorname{rank} V_s = \min(s,t) = t$, so

$$\operatorname{rank}(H_s) = \operatorname{rank}(V_s^T D_t V_s) \leq \min(\operatorname{rank}(V_s^T), \operatorname{rank}(D_t), \operatorname{rank}(V_s)) = \min(t,t,t) = t$$

we must have $\det(H_s) = 0$ for $s > t$.

Case $s < t$: $\det(H_s) = 0$ is possible. Kaltofen and Lee propose we compute

$$\det(H_1), \quad \det(H_2), \quad \det(H_3) \ \ldots, \quad \det(H_t) \neq 0, \quad \det(H_{t+1}) = 0,$$

stop when $\det(H_s) = 0$ and return $t = s - 1$. This fails for $f(x_1, x_2) = x_1^2 - x_2$ with $t = 2$ because

$$H_1 = [f(\alpha^0)] = [f(1,1)] = [0].$$

Kaltofen and Lee suggest either

1 Pick $c \in \mathbb{F}$ at random.
Set $g(x_1, \ldots, x_n) = f(x_1, \ldots, x_n) + c$ and determine t for g.
Build H_t for f and **return** $\operatorname{rank}(H_t)$.

2 Build H_s using $a_1, a_2, \ldots, a_{2s-1}$ instead.

Let $a_i = f(\alpha_1^i, \ldots, \alpha_n^i)$ and $b_i = f(x_1^i, \ldots, x_n^i)$. Define the $s \times s$ Hankel matrices

$$H_s = \begin{bmatrix} a_1 & a_2 & \cdots & a_s \\ a_2 & a_3 & \cdots & a_{s+1} \\ \vdots & \vdots & & \vdots \\ a_s & a_{s+1} & \cdots & a_{2s-1} \end{bmatrix} \quad \text{and} \quad \widehat{H}_s = \begin{bmatrix} b_1 & b_2 & \cdots & b_s \\ b_2 & b_3 & \cdots & b_{s+1} \\ \vdots & \vdots & & \vdots \\ b_s & b_{s+1} & \cdots & b_{2s-1} \end{bmatrix}.$$

Notice that $H_s = \widehat{H}_s(\alpha)$. To use the Schwartz-Zippel lemma Kaltofen and Lee first prove that $\det(\widehat{H}_s) \neq 0$ for $1 \leq s \leq t$ and then claim without proof that $\deg(\det(\widehat{H}_s)) \leq s^2 \deg(f)$.

Let S be a large finite subset of $\mathbb{F}$. By Schwartz-Zippel, if α is chosen at random from S^n

$$\mathrm{Prob}(\det(H_s) = 0) = \mathrm{Prob}(\det(\widehat{H}_s)(\alpha) = 0) \leq \frac{\deg(\det(\widehat{H}_s))}{|S|} \leq \frac{s^2 \deg f}{|S|}.$$

Hence the probability that $\det(H_s) = 0$ for any $1 \leq s \leq t$ is

$$\leq \frac{\sum_{s=1}^t s^2 \deg f}{|S|} = \frac{\frac{1}{6} t(2t+1)(t+1) \deg f}{|S|}.$$

$$\frac{\frac{1}{6}t(2t+1)(t+1)\deg f}{|S|}.$$

This bound is cubic in t! Suppose $\mathbb{F} = \mathbb{Z}_p$ and $S = \mathbb{Z}_p$ with $p = 2^{31} - 1$. Suppose f has $t = 1000$ terms and $n = 5$ variables with $\deg f = 10$. We have

$$\frac{\frac{1}{6}t(2t+1)(t+1)\deg f}{|S| = 2^{31-1}} = \frac{3,338,335,000}{2,147,483,647}$$

In practice, this algorithm does not fail even when f has many thousands of terms.
The failure bound is a worst case bound. It assumes that the polynomials $\det(\widehat{H}_s)$ for $1 \leq s \leq t$ have the maximum possible number of roots in $\mathbb{F}$.

If f is a non-zero polynomial in $\mathbb{F}[x_1, \ldots, x_n]$ with $d = \deg f$, the Schwartz-Zippel lemma says f can have at most $d|S|^{n-1}$ roots in $\mathbb{F}$. Hence if we pick α from S^n at random,

$$\mathrm{Prob}(f(\alpha) = 0) \leq \frac{d|S|^{n-1}}{|S|^n} = \frac{d}{|S|}.$$

But for $\mathbb{F} = \mathbb{Z}_p$ where p is prime, the average number of roots in $\mathbb{Z}_p$ is p^{n-1}. Hence for random f of degree d, since there are p^n choices for α, $\mathrm{Prob}(f(\alpha) = 0) = 1/p$.

Thus if the polynomials $\det(\widehat{H}_s)$ for $1 \leq s \leq t$ behave randomly, and $\mathbb{F} = \mathbb{Z}_p$ and α is chosen randomly from $\mathbb{Z}_p^n$, $\mathrm{Prob}(\det(\widehat{H}_s(\alpha) = 0) = 1/p$ and

$$\mathrm{Prob}(\prod_{s=1}^{t} \det(\widehat{H}_s(\alpha) = 0) = t/p$$

For $t = 1000$, $n = 5$ and $p = 2^{31} - 1$, $t/p < 10^{-6}$ so the algorithm will determine t with good probability.

To reduce the probability of failure Kaltofen and Lee suggest we compute

$$\mathrm{rank}(H_1),\ \mathrm{rank}(H_2),\ \mathrm{rank}(H_3),\ \dots\ \mathrm{rank}(H_s),\ \dots$$

and stop when $\mathrm{rank}(H_s) \leq s - 2$ and use $\mathrm{rank}(H_s)$ to estimate t. This approximately squares the probability of failure since it fails only if two consecutive Hankel matrices H_{s-1} and H_s are singular for some $2 \leq s \leq t$.

To reduce the probability of failure we compute

$$\mathrm{rank}(H_2),\ \mathrm{rank}(H_4),\ \mathrm{rank}(H_8),\ \mathrm{rank}(H_{16}),\ \dots\ , \mathrm{rank}(H_{s=2^i}),\ \dots$$

instead and stop when $\det(H_s) = 0$ and use $\mathrm{rank}(H_s)$ to estimate t. This may double the number of probes to the black box but it reduces the probability of failure to

$$\sum_{i=1}^{\lfloor \log_2 t \rfloor} \frac{(2^i)^2 d}{|S|} < \frac{\frac{4}{3} t^2 d}{|S|}$$

which is quadratic in t.

$$\widehat{H}_s = \begin{bmatrix} b_1 & b_2 & \cdots & b_s \\ b_2 & b_3 & \cdots & b_{s+1} \\ \vdots & \vdots & & \vdots \\ b_s & b_{s+1} & \cdots & b_{2s-1} \end{bmatrix} = \begin{bmatrix} f(x_1, \ldots, x_n) & \cdots & f(x_1^s, \cdots, x_n^s) \\ \vdots & & \vdots \\ f(x_1^s, \ldots, x_n^s) & \cdots & f(x_1^{2s-1}, \cdots, x_n^{2s-1}) \end{bmatrix}$$

To use Schwartz-Zippel we need to bound $\deg(\det(H_s))$. Since $\deg(b_i) = i \deg(f) = id$ we have

$$[\deg(H_{sij})] = \begin{bmatrix} d & 2d & \cdots & sd \\ 2d & 3d & \cdots & (s+1)d \\ \vdots & \vdots & & \vdots \\ sd & (s+1)d & \cdots & (2s-1)d \end{bmatrix}$$

Using the bottom row to bound $\deg(\det(H_s))$ we have

$$\deg(\det(H_s)) \leq \sum_{i=0}^{s-1} (s+i)d = (\tfrac{3}{2}s^2 - \tfrac{1}{2}s) \deg f.$$

Kaltofen and Lee state the tighter bound

$$\deg(\det(H_s)) \leq s^2 \deg f.$$

Recall that if A is an n by n matrix

$$\det A = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) \prod_{i=1}^{n} A_{i,\sigma(i)}$$

where S_n is the symmetric group on n elements, for example

$$S_3 = \{(1,2,3),(1,3,2),(2,1,3),(2,3,1),(3,1,2),(3,2,1)\}.$$

Notice

$$\det\left(\begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}\right) = \begin{matrix} A_{11}(A_{22}A_{33} - A_{23}A_{32}) \\ -A_{21}(A_{12}A_{33} - A_{13}A_{32}) \\ +A_{31}(A_{12}A_{23} - A_{22}A_{13}) \end{matrix} = \begin{matrix} A_{11}A_{22}A_{33} - A_{11}A_{23}A_{32} \\ -A_{12}A_{21}A_{33} + A_{13}A_{21}A_{32} \\ +A_{12}A_{23}A_{31} - A_{13}A_{22}A_{31} \end{matrix} \qquad \begin{matrix} (1\ 2\ 3) & (1\ 3\ 2) \\ (2\ 1\ 3) & (3\ 1\ 2) \\ (2\ 3\ 1) & (3\ 2\ 1) \end{matrix}$$

Notice for any $\sigma \in S_3$, then elements of σ sum to 6. In general for $\sigma \in S_n$ we have

$$\sum_{i=1}^{n} \sigma(i) = \sum_{i=1}^{n} i.$$

We have

$$
\begin{aligned}
\deg(\det(\widehat{H}_s)) &= \deg\left(\sum_{\sigma\in S_s} \operatorname{sign}(\sigma) \prod_{i=1}^{s} \widehat{H}s_{i,\sigma(i)}\right) \\
&\le \max_{\sigma\in S_s} \deg\left(\prod_{i=1}^{s} \widehat{H}s_{i,\sigma(i)}\right) \\
&\le \max_{\sigma\in S_s} \sum_{i=1}^{s} \deg(\widehat{H}s_{i,\sigma(i)}) \\
&= \max_{\sigma\in S_s} \sum_{i=1}^{s} (i+\sigma(i)-1)d \\
&= \max_{\sigma\in S_s} \sum_{i=1}^{s} (i+i-1)d = s^2 \deg f.
\end{aligned}
$$