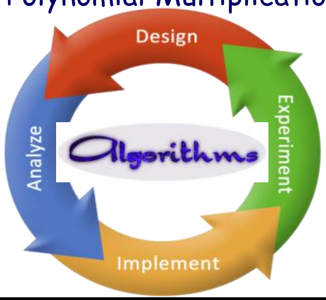
Algorithm Design and Analysis

Victor Adamchik — CS 15-451 — Spring 2015

Lecture 2 — Jan 14, 2015 — Carnegie Mellon University

# Strassen's Algorithm
# Polynomial Multiplication




## Outline

1) Strassen's Algorithm
2) Karatsuba for Polynomials
3) Legendre's Interpolation
2) Vandemonde Matrix

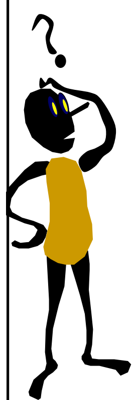
## Matrix Multiplication

A B C = AxB

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11}+a_{12}b_{21} & a_{11}b_{12}+a_{12}b_{22} \\ a_{21}b_{11}+a_{22}b_{21} & a_{21}b_{12}+a_{22}b_{22} \end{pmatrix}$$

Let A and B be nxn matrices, then their product C=A*B is

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$



$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

What is the complexity of this algorithm (in terms of multiplications)?

$O(n^3)$

## Divide and Conquer

The idea is to divide the size of the problem in half. This corresponds to dividing each of the matrices into quarters, each n/2 x n/2 size, and multiply those quarters.

## Algorithm

Let $n = 2^k$ and M(A,B) denote the matrix product

1. if A is 1x1 matrix, return $a_{11} * b_{11}$.
2. write $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ $B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$

where $A_{ij}$ and $B_{ij}$ are n/2 x n/2 matrices.

3. Compute $C_{ij} = M(A_{i1},B_{1j}) + M(A_{i2},B_{2j})$
4. Return $\begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}$

## Correctness

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \begin{pmatrix} \boxed{A_{11}B_{11} + A_{12}B_{21}} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{pmatrix}$$

This basically says that if the entries of A and B are themselves matrices, the usual matrix multiplication works by substituting the blocks into the formula.

Let us prove it for the left upper entry.

We know $$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

## Correctness

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$$

Since $1 \le i,j \le n/2$.

$$c_{ij} = \sum_{k=1}^{n/2} a_{ik} b_{kj} + \sum_{k=n/2+1}^{n} a_{ik} b_{kj}$$

$$= \sum_{k=1}^{n/2} A_{11}(i,k) B_{11}(k,j) + \sum_{k=n/2+1}^{n} A_{12}(i,k) B_{12}(k,j)$$

$$= (A_{11} \cdot B_{11})(i,j) + (A_{12} \cdot B_{21})(i,j)$$

$$= (A_{11} \cdot B_{11} + A_{12} \cdot B_{21})(i,j)$$

Similar proof for the other blocks.

## Worst-case complexity

$C_{ij} = M(A_{i1}, B_{1j}) + M(A_{i2}, B_{2j})$ $\qquad \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}$

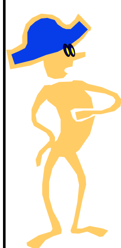
On each step we compute 4 matrices $C_{ij}$, each requires two recursive calls.

Let T(n) denote the number of multiplications, then

$$T(n) = 8T(n/2) + \underbrace{O(n^2)}_{\text{Matrix addition}}$$
$$T(1) = 1$$

The Master Theorem gives $\Theta(n^3)$.

## Strassen's Algorithm (1968)



Do we need all 8 multiplications or can we find a clever way of doing it with fewer?

Strassen a German mathematician born in 1936

## Strassen's Algorithm

For 2 x 2 matrices

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} s_1 + s_2 - s_4 + s_6 & s_4 - s_5 \\ s_6 + s_7 & s_2 - s_3 + s_5 - s_7 \end{pmatrix}$$

$s_1 = (a_{12}-a_{22})(b_{21}+b_{22})$
$s_2 = (a_{11}+a_{22})(b_{11}+b_{22})$
$s_3 = (a_{11}-a_{21})(b_{11}+b_{12})$
$s_4 = (a_{11}+a_{12})\, b_{22}$
$s_5 = a_{11}\,(b_{12}-b_{22})$
$s_6 = a_{22}\,(b_{21}-b_{11})$
$s_7 = (a_{21}+a_{22})\, b_{11}$

It takes 7 multiplications

## Correctness

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} s_1 + s_2 - s_4 + s_6 & s_4 - s_5 \\ \boxed{s_6 + s_7} & s_2 - s_3 + s_5 - s_7 \end{pmatrix}$$

$\boxed{\begin{array}{l} s_6 = a_{22}\,(b_{21}-b_{11}) \\ s_7 = (a_{21}+a_{22})\, b_{11} \end{array}}$

Proof for a lower left entry:

$s_6+s_7 = \boxed{a_{21}\, b_{11} + a_{22}\, b_{21}}$

on the other hand:

$s_6+s_7 = a_{22}\,(b_{21}-b_{11}) + (a_{21}+a_{22})\, b_{11} =$
$a_{22}\, b_{21} - a_{22}\, b_{11} + a_{21}\, b_{11} + a_{22}\, b_{11} = \boxed{a_{21}\, b_{11} + a_{22}\, b_{21}}$

## Strassen's Algorithm

This holds for a block matrix multiplication

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \begin{pmatrix} S_1 + S_2 - S_4 + S_6 & S_4 - S_5 \\ S_6 + S_7 & S_2 - S_3 + S_5 - S_7 \end{pmatrix}$$

where $A_{ij}$ and $B_{ij}$ are n/2 x n/2 matrices and matrices $S_1$, …., $S_7$ are defined on the previous slide.

## Worst-case complexity

Let T(n) denote the number of multiplications, then

$$T(n) = 7\,T(n/2) + O(n^2)$$
$$T(1) = 1$$

Matrix addition

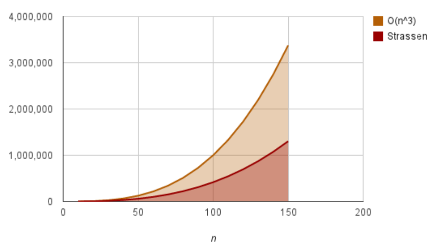
The Master Theorem gives $\Theta(n^{\log 7}) = \Theta(n^{2.807})$.

If we count additions, then

$$T(n) = 7\,T(n/2) + 18(n/2)^2$$
$$T(1) = 1$$

why-?

## Time complexity




## Space Complexity

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \begin{pmatrix} S_1 + S_2 - S_4 + S_6 & S_4 - S_5 \\ S_6 + S_7 & S_2 - S_3 + S_5 - S_7 \end{pmatrix}$$

$S_1 = (A_{12}-A_{22})(B_{21}+B_{22})$
$S_2 = (A_{11}+A_{22})(B_{11}+B_{22})$
$S_3 = (A_{11}-A_{21})(B_{11}+B_{12})$
$S_4 = (A_{11}+A_{12})\,B_{22}$
$S_5 = A_{11}\,(B_{12}-B_{22})$
$S_6 = A_{22}\,(B_{21}-B_{11})$
$S_7 = (A_{21}+A_{22})\,B_{11}$

We need to compute and then store matrices $S_1$, …, $S_7$

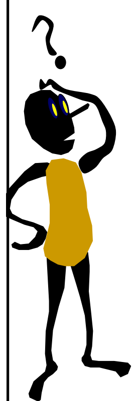
To compute them we need two scratch arrays, so 9 total.

## Space Complexity

Let W(n) be the space complexity

$$W(n) = W(n/2) + 9(n/2)^2$$
$$W(1) = 1$$

Solving this gives $W(n) = 3\,n^2$.



How would you extend Strassen's algorithm to matrix dimensions differ from $2^k$?

Pad the matrices with zeros.

## Polynomial Multiplication

How would you multiply two polynomials?

## Polynomial Multiplication

$$A(x) = \sum_{k=0}^{n} a_k x^k \qquad B(x) = \sum_{k=0}^{n} b_k x^k$$

$$C(x) = A(x)B(x) = \sum_{j=0}^{n} \sum_{k=0}^{n} a_j b_k x^{j+k}$$

This has $O(n^2)$ complexity. We can do much better!

## Karatsuba Revisited

$$A(x) = \sum_{k=0}^{n} a_k x^k \qquad B(x) = \sum_{k=0}^{n} b_k x^k$$

$$A(x) = A_1\, x^{n/2} + A_0 \qquad \text{Same for } B(x)$$

For example, $1 + 3x + x^2 + 7x^3 = (1 + 3x) + x^2\,(1+7x)$

$$A(x)B(x) = C_2\, x^n + C_1 x^{n/2} + C_0$$

where

$$C_2 = A_1 B_1$$
$$C_1 = (A_0 + A_1)(B_0 + B_1) - A_0 B_0 - A_1 B_1$$
$$C_0 = A_0 B_0$$

## Polynomial Multiplication

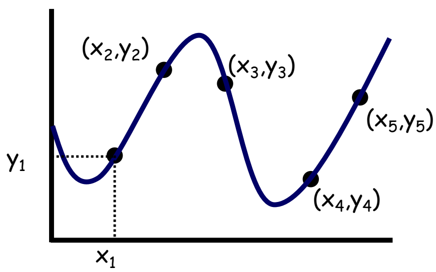
Karatsuba's polynomial multiplication can be done with at most $O(n^{\log 3})$ operations.

Observe, the algorithm in fact multiplies only <u>linear</u> polynomials (2 terms) with three scalar multiplications.

In the next slides we outline a slightly different approach that is based on interpolation.

## Interpolation

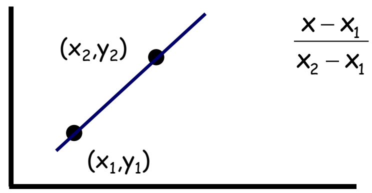
Say you're given a bunch of "data points"




Can you find a (low-degree) polynomial which "fits the data"?

## Interpolation

There is a unique linear polynomial going through 2 points




$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1}$$

Correspondence between a set of 2 points and a line (a polynomial of the first order)

## Uniqueness

Correspondence between a set of distinct points

$(x_1,y_1), (x_2,y_2), \ldots, (x_{n+1},y_{n+1})$

and a polynomial over a field F.

Theorem:

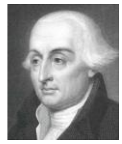
There is exactly one polynomial P(x) of degree at most n such that $P(x_k) = y_k$ for all k = 1, … , n+1.

This theorem was proved in 15-251, so review that lecture.

## Lagrange's Interpolation Formula

The method for constructing the polynomial is called Lagrange Interpolation.

$$\boxed{y(x) = \sum_{j=0}^{n} y_j \prod_{\substack{k=0 \\ k \neq j}}^{n} \frac{x - x_k}{x_j - x_k}}$$



Discovered in 1795 by J.-L. Lagrange.

## How does this apply to multiplication?

Given two polynomials

$A(x) = 1 + x + x^2$ $B(x) = 1 + 2x + 3x^2$

Compute their values at x = -2, -1, 0, 1, 2

{A(-2), A(-1), A(0), A(1), A(2)} = {3, 1, 1, 3, 7}

{B(-2), B(-1), B(0), B(1), B(2)} = {9, 2, 1, 6, 17}

Pointwise multiplication:

{C(-2), C(-1), C(0), C(1), C(2)} = {27, 2, 1, 18, 119}

## How does this apply to multiplication?

Points to fit

(-2,27), (-1,2), (0, 1), (1, 18), (2, 119)

This yields (with n=5)

$$y(x) = \sum_{j=0}^{n} y_j \prod_{\substack{k=0 \\ k \neq j}}^{n} \frac{x - x_k}{x_j - x_k}$$

$$y(x) = 1 + 3x + 6x^2 + 5x^3 + 3x^4$$

## Karatsuba again…

Recall that in Karatsuba we split a polynomial in half

Thus, after some proper renaming, on each iteration we multiply linear polynomials

Let us apply the idea of multiplication by interpolation to Karatsuba's divide and conquer

## Multiplication by Interpolation

Let us multiply polynomials of degree one

$A(x) = a_0 + a_1 x$, $B(x) = b_0 + b_1 x$

Suggested points for evaluation: 0, 1, $\infty$

$A(0) = a_0$, $A(1) = a_0 + a_1$, $A(\infty) = a_1$

$B(0) = b_0$, $B(1) = b_0 + b_1$, $B(\infty) = b_1$

$\infty$ means, take the leading coefficient

Compute the pointwise product:

$c_0 = a_0 b_0$, $c_1 = (a_0 + a_1)(b_0 + b_1)$, $c_2 = a_1 b_1$

We restore the polynomial with no use of Langange's formula

## Multiplication by Interpolation

Find a polynomial passing through these points

$(0, a_0 b_0)$, $(1, (a_0 + a_1)(b_0 + b_1))$, $(\infty, a_1 b_1)$

Clearly it must be a quadratic polynomial

$c_0 + c_1 x + c_2 x^2$

Setting $x = 0$, gives that $c_0 = a_0 b_0$

Setting $x = \infty$, gives that $c_2 = a_1 b_1$

Setting $x = 1$, gives that $c_0 + c_1 + c_2 = (a_0 + a_1)(b_0 + b_1)$

It follows, $c_1 = (a_0 + a_1)(b_0 + b_1) - a_0 b_0 - a_1 b_1$

Wow, exactly like in Karatsuba's algorithm

## Toward the Fast Fourier Transform

To compute the polynomial product $A(x) B(x)$,

1) evaluate $A(x)$ and $B(x)$ at some points $x_k$,
2) multiply $A(x_k) B(x_k)$ pointwise,
3) find the polynomial which passes through these points.

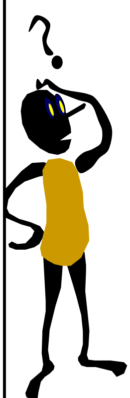
## The worst-case complexity

To compute the polynomial product $A(x) B(x)$,

1) evaluate $A(x)$ and $B(x)$ at some points $x_k$,
   - what's the complexity of $A(x_k)$-?
   - what's the complexity of $A(x)$ at n points?
2) multiply $A(x_k) B(x_k)$ pointwise,
3) find the polynomial which passes through these points.
   - what is it complexity?

## Vandermonde Matrix

We will rewrite Lagrange's formula in a matrix form



What is the runtime complexity of Lagrange's interpolation?

$O(n^3)$, if we expand

## Matrix Form

$$y(x) = \sum_{j=0}^{n} y_j \prod_{\substack{k=0 \\ k \neq j}}^{n} \frac{x - x_k}{x_j - x_k}$$

Consider a case of two points

$$y = y_0 \frac{x - x_1}{x_0 - x_1} + y_1 \frac{x - x_0}{x_1 - x_0} = \underbrace{\frac{x_0 y_1 - x_1 y_0}{x_0 - x_1}}_{a_0} + x \underbrace{\frac{y_0 - y_1}{x_0 - x_1}}_{a_1}$$

$$y = a_0 + a_1 x$$

where new coefficients $a_0$ and $a_1$ can be found by solving the following system

$$\begin{pmatrix} 1 & x_0 \\ 1 & x_1 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \end{pmatrix}$$

This could be generalized...

## The Vandermonde Matrix

The Lagrange formula defines a polynomial $A(x)=\sum_{k=0}^{n} a_k x^k$

where coefficients $a_k$ can be found from

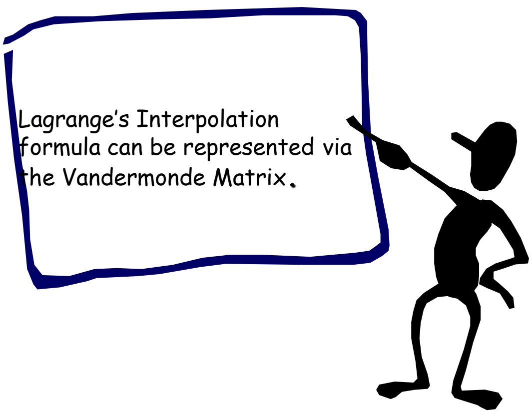
$$\begin{pmatrix} 1 & x_0 & x_0^2 & \dots & x_0^n \\ 1 & x_1 & x_1^2 & \dots & x_1^n \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^n \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \dots \\ a_n \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \\ \dots \\ y_n \end{pmatrix}$$

or in short $V \cdot a = y$ Thus, $a = V^{-1} \cdot y$

## The Vandermonde Matrix

$$V = \begin{pmatrix} 1 & x_0 & x_0^2 & \dots & x_0^n \\ 1 & x_1 & x_1^2 & \dots & x_1^n \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^n \end{pmatrix}$$

In order to inverse the matrix V, we have to prove that it's nonsingular.




## Determinant of the Vandermonde Matrix

$$\det \begin{pmatrix} 1 & x_0 & x_0^2 & \dots & x_0^n \\ 1 & x_1 & x_1^2 & \dots & x_1^n \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^n \end{pmatrix} = \prod_{k=0}^{n}\prod_{j=0}^{k-1}(x_k - x_j)$$

Since the n + 1 points are distinct, the determinant can't be zero, so the matrix V is not singular and its inverse does exist.

## Proof

$$\det \begin{pmatrix} 1 & x_0 & x_0^2 & \dots & x_0^n \\ 1 & x_1 & x_1^2 & \dots & x_1^n \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^n \end{pmatrix} = \prod_{k=0}^{n}\prod_{j=0}^{k-1}(x_k - x_j)$$

First we subtract the first row from all other rows

$$V_n = \begin{vmatrix} 1 & x_0 & x_0^2 & \dots & x_0^n \\ 1 & x_1 & x_1^2 & \dots & x_1^n \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^n \end{vmatrix} = \begin{vmatrix} 1 & x_0 & x_0^2 & \dots & x_0^n \\ 0 & x_1 - x_0 & x_1^2 - x_0^2 & \dots & x_1^n - x_0^n \\ \dots & \dots & \dots & \dots & \dots \\ 0 & x_n - x_0 & x_n^2 - x_0^2 & \dots & x_n^n - x_0^n \end{vmatrix}$$

## Proof

$$\begin{vmatrix} 1 & x_0 & \dots & x_0^{n-1} & x_0^n \\ 0 & x_1 - x_0 & \dots & x_1^{n-1} - x_0^{n-1} & x_1^n - x_0^n \\ \dots & \dots & \dots & \dots & \dots \\ 0 & x_n - x_0 & \dots & x_n^{n-1} - x_0^{n-1} & x_n^n - x_0^n \end{vmatrix}$$

Next step, multiply (n-1)-column by $x_0$ and subtract from n-column.

$$V_n = \begin{vmatrix} 1 & x_0 & \dots & x_0^{n-1} & 0 \\ 0 & x_1 - x_0 & \dots & x_1^{n-1} - x_0^{n-1} & x_1^{n-1}(x_1 - x_0) \\ \dots & \dots & \dots & \dots & \dots \\ 0 & x_n - x_0 & \dots & x_n^{n-1} - x_0^{n-1} & x_n^{n-1}(x_n - x_0) \end{vmatrix}$$

## Proof

$$\begin{vmatrix} 1 & x_0 & \dots & x_0^{n-1} & 0 \\ 0 & x_1 - x_0 & \dots & x_1^{n-1} - x_0^{n-1} & x_1^{n-1}(x_1 - x_0) \\ \dots & \dots & \dots & \dots & \dots \\ 0 & x_n - x_0 & \dots & x_n^{n-1} - x_0^{n-1} & x_n^{n-1}(x_n - x_0) \end{vmatrix}$$

Next step, multiply (n-2)-column by $x_0$ and subtract from (n-1)-column.

$$V_n = \begin{vmatrix} 1 & x_0 & \dots & 0 & 0 \\ 0 & x_1 - x_0 & \dots & x_1^{n-2}(x_1 - x_0) & x_1^{n-1}(x_1 - x_0) \\ \dots & \dots & \dots & \dots & \dots \\ 0 & x_n - x_0 & \dots & x_n^{n-2}(x_n - x_0) & x_n^{n-1}(x_n - x_0) \end{vmatrix}$$

Finally, multiply 1st-column by $x_0$ and subtract from 2nd

## Proof

$$V_n = \begin{vmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & x_1 - x_0 & \dots & x_1^{n-2}(x_1 - x_0) & x_1^{n-1}(x_1 - x_0) \\ \dots & \dots & \dots & \dots & \dots \\ 0 & x_n - x_0 & \dots & x_n^{n-2}(x_n - x_0) & x_n^{n-1}(x_n - x_0) \end{vmatrix}$$

Remove a constant factor $(x_k - x_0)$, $k = 1, \dots, n$ from each row

$$V_n = \prod_{k=1}^{n}(x_k - x_0) \begin{vmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & x_1 & \dots & x_1^{n-2} & x_1^{n-1} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 1 & x_n & \dots & x_n^{n-2} & x_n^{n-1} \end{vmatrix}$$

## Proof

$$V_n = \prod_{k=1}^{n}(x_k - x_0) \begin{vmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & x_1 & \dots & x_1^{n-2} & x_1^{n-1} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 1 & x_n & \dots & x_n^{n-2} & x_n^{n-1} \end{vmatrix}$$

This leads to

$$V_n = \prod_{k=1}^{n}(x_k - x_0)\, V_{n-1}$$

Repeating the above steps for $V_{n-1}$, we get

$$V_{n-1} = \prod_{k=2}^{n}(x_k - x_1)\, V_{n-2}$$

And finally

$$V_1 = \begin{vmatrix} 1 & x_{n-1} \\ 1 & x_n \end{vmatrix} = x_n - x_{n-1}$$

QED.

## Complexity of Interpolation

$$\begin{pmatrix} a_0 \\ a_1 \\ \dots \\ a_n \end{pmatrix} = \begin{pmatrix} 1 & x_0 & x_0^2 & \dots & x_0^n \\ 1 & x_1 & x_1^2 & \dots & x_1^n \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^n \end{pmatrix}^{-1} \begin{pmatrix} y_0 \\ y_1 \\ \dots \\ y_n \end{pmatrix}$$

It follows that the complexity of interpolation depends on how fast can we inverse the Vandermonde matrix.

The success depends on the values of $x_k$, k=0, …,n

## Toward the Fast Fourier Transform

To compute the polynomial product A(x) B(x),

1) evaluate A(x) and B(x) at some points $x_k$,
2) multiply $A(x_k)$ $B(x_k)$ pointwise,
3) find the polynomial which passes through these points , by using Vandermonde's matrix