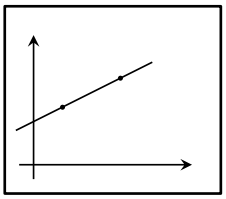
Construção e Análise de Algoritmos

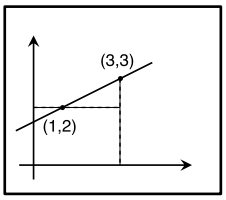
aula 13: A transformada rápida de Fourier

1 Introdução

É um fato bem conhecido que por dois pontos passa exatamente uma reta



De fato, a partir das coordenadas dos pontos não é difícil obter a equação da reta



- primeiro, nós calculamos o coeficiente que determina a sua inclinação

$$a \;=\; \frac{3-2}{3-1} \;=\; \frac{1}{2}$$

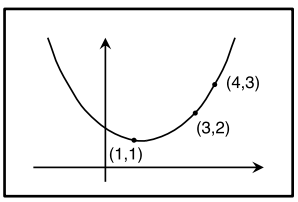
- e, a seguir, nós calculamos o ponto em que a reta cruza o eixo Y:

$$b \;=\; 2 \,-\, 1 \cdot a \;=\; 2 \,-\, 1 \cdot \frac{1}{2} \;=\; \frac{3}{2}$$

Pronto, a equação da reta acima é

$$y \;=\; \frac{x}{2} \,+\, \frac{3}{2}$$

O que algumas pessoas talvez não saibam é que por 3 pontos (não colineares) passa exatamente uma curva quadrática. Por exemplo,



Como no caso da reta, não é difícil obter a equação da curva a partir das coordenadas dos pontos.

Nós sabemos que essa equação tem a forma

$$y \;=\; ax^2 \;+\; bx \;+\; c$$

Se a curva deve passar pelos 3 pontos, então nós devemos ter

$$\begin{aligned} 1 &= a \cdot 1^2 + b \cdot 1 + c \\ 2 &= a \cdot 3^2 + b \cdot 3 + c \\ 3 &= a \cdot 4^2 + b \cdot 4 + c \end{aligned}$$

E agora, basta resolver esse pequeno sistema para obter

$$a \;=\; \frac{1}{6} \qquad\qquad b \;=\; -\frac{1}{6} \qquad\qquad c \;=\; 1$$

o que nos dá a equação

$$y \;=\; \frac{x^2}{6} \;-\; \frac{x}{6} \;+\; 1$$

De fato, no caso geral, por $n+1$ pontos (adequadamente escolhidos*) passa exatamente uma curva de grau n.

2 Multiplicação de polinômios

As observações acima implicam que qualquer polinômio de grau n

$$A(x) \;=\; a_0 \;+\; a_1 x \;+\; a_2 x^2 \;+\; \ldots \;+\; a_n x^n \qquad\qquad [\texttt{R1}]$$

pode ser representado de maneira única como uma coleção de $n+1$ pontos

$$A(x) \;:\; \Big\{(x_0, y_0),\ (x_1, y_1),\ (x_2, y_2),\ \ldots,\ (x_n, y_n)\Big\} \qquad\qquad [\texttt{R2}]$$

Isso é relevante porque algumas operações entre polinômios podem ser implementadas de maneira mais eficiente utilizando a representação `R2` do que utilizando a representação `R1`.

Por exemplo, considere o problema de multiplicar dois polinômios

$$\begin{aligned} A(x) &= a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n \\ B(x) &= b_0 + b_1 x + b_2 x^2 + \ldots + b_n x^n \end{aligned}$$

*i.e., que não pertencem a uma curva de grau menor que n

Utilizando o procedimento padrão que aprendemos na escola, a solução do problema leva tempo $\Theta(n^2)$.

E, utilizando as ideias da aula passada, o problema é resolvido em tempo $\Theta(n^{\log_2 3})$.

Agora, veja o que acontece quando temos os dois polinômios representados no formato `R2`:

$$A(x) \ : \ \Big\{(x_0, y_0),\ (x_1, y_1),\ (x_2, y_2),\ \ldots,\ (x_n, y_n)\Big\}$$
$$B(x) \ : \ \Big\{(x_0, z_0),\ (x_1, z_1),\ (x_2, z_2),\ \ldots,\ (x_n, z_n)\Big\}$$

No ponto x_0, o polinômio $A(x)$ assume o valor y_0, e o polinômio $B(x)$ assume o valor z_0.

Mas, isso significa que, no ponto x_0 o polinômio $\big(A \cdot B\big)(x)$ deve assumir o valor $y_0 \cdot z_0$.

Levando em conta todos os pontos, nós temos que o polinômio $\big(A \cdot B\big)(x)$ é representado por

$$\big(A \cdot B\big)(x) \ : \ \Big\{(x_0, y_0 z_0),\ (x_1, y_1 z_1),\ (x_2, y_2 z_2),\ \ldots,\ (x_n, y_n z_n)\Big\}$$

e o nosso problema está resolvido!

Quase ...

Note que $\big(A \cdot B\big)(x)$ é um polinômio de grau $2n$, e isso significa que a sua representação `R2` precisa conter $2n + 1$ pontos.

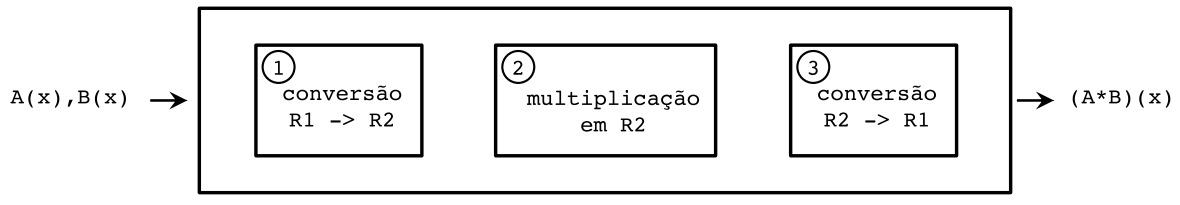
Mas, isso é um problema fácil de resolver: basta começar com representações para $A(x)$ e $B(x)$ contendo $2n + 1$ pontos.

A boa notícia é que a multiplicação baseada na representação `R2` leva tempo $\Theta(n)$.

A má notícia é que os nossos polinômios em geral não estão nesse formato ...

3 Uma estratégia de decomposição esperta

A figura abaixo ilustra o método para multiplicação de polinômios a que chegamos no final da seção anterior.



Nós já sabemos que a etapa 2 executa em tempo $\Theta(n)$.

O desafio agora é encontrar uma maneira eficiente de implementar as etapas 1 e 3.

E a solução consiste em escolher as coordenadas de maneira esperta — para reaproveitar trabalho.

Nós começamos com uma observação simples.

Suponha que nós queremos calcular o valor do polinômio $A(x)$ nos pontos 1 e -1.

Observando que $1^2 = (-1)^2$, faz sentido separar os termos com potências pares e ímpares de $A(x)$:

$$A(x) = \underbrace{a_0 + a_2x^2 + \ldots + a_nx^n}_{A_2(x)} + \underbrace{a_1x + a_3x^3 + \ldots + a_{n-1}x^{n-1}}_{A_1(x)}$$

de modo que

$$\begin{aligned} A(1) &= A_2(1) + A_1(1) \\ A(-1) &= A_2(1) + A_1(-1) \end{aligned}$$

Ou seja, nós não precisamos calcular $A_2(-1)$.

De fato, nós podemos ser um pouquinho mais espertos, escrevendo $A(x)$ como

$$A(x) = \underbrace{a_0 + a_2x^2 + \ldots + a_nx^n}_{A_2(x)} + \Big[\underbrace{a_1 + a_3x^2 + \ldots + a_{n-1}x^{n-2}}_{A_1(x)}\Big] \cdot x$$

Agora, tanto $A_2(x)$ como $A_1(x)$ só possuem potências pares, de modo que $A_2(-1) = A_2(1)$ e $A_1(-1) = A_1(1)$.

Logo,

$$\begin{aligned} A(1) &= A_2(1) + A_1(1) \\ A(-1) &= A_2(1) + A_1(1) \cdot (-1) \end{aligned}$$

e apenas $A_2(1)$ e $A_1(1)$ precisam ser calculados.

Essa ideia é muito legal!

Mas, como é que nós podemos fazê-la funcionar para uma coleção maior de pontos?

A resposta é: trabalhando com números complexos.

Suponha que nós queremos calcular o valor do polinômio $A(x)$ nos pontos $1, -1, i, -i$.

Observando que $(-1)^2 = 1^2$ e $(-i)^2 = i^2$, nós escrevemos o polinômio $A(x)$ como

$$A(x) = \underbrace{a_0 + a_2x^2 + \ldots + a_nx^n}_{A_2(x)} + \Big[\underbrace{a_1 + a_3x^2 + \ldots + a_{n-1}x^{n-2}}_{A_1(x)}\Big] \cdot x$$

de modo que só é preciso calcular $A_2(1), A_1(1), A_2(i), A_1(i)$.

A seguir, nós observamos que $1^4 = i^4$.

Logo, faz sentido separar também as potências múltiplas de 4 das demais potências de $A_2(x)$

(e $A_1(x)$ também):

$$A_2(x) \;=\; \underbrace{a_0 \;+\; a_4x^4 \;+\; \ldots \;+\; a_nx^n}_{A_{22}(x)} \;+\; \underbrace{a_2x^2 \;+\; a_6x^6 \;+\; \ldots \;+\; a_{n-2}x^{n-2}}_{A_{21}(x)}$$

de modo que

$$\begin{aligned} A_2(1) &= A_{22}(1) + A_{21}(1) \\ A_2(i) &= A_{22}(1) + A_{21}(i) \end{aligned}$$

Ou seja, nós não precisamos calcular $A_{22}(i)$.

E, como antes, nós podemos ser um pouquinho mais espertos escrevendo

$$A_2(x) \;=\; \underbrace{a_0 \;+\; a_4x^4 \;+\; \ldots \;+\; a_nx^n}_{A_{22}(x)} \;+\; \underbrace{\left[a_2 \;+\; a_6x^4 \;+\; \ldots \;+\; a_{n-2}x^{n-4}\right]}_{A_{21}(x)} \cdot x^2$$

Agora, tanto $A_{22}(x)$ como $A_{21}(x)$ só possuem potências múltiplas de 4, de modo que

$$A_{22}(1) \;=\; A_{22}(i) \qquad\qquad A_{21}(1) \;=\; A_{21}(i)$$

Logo, nós podemos calcular

$$\begin{aligned} A_2(1) &= A_{22}(1) + A_{21}(1) \\ A_2(i) &= A_{22}(1) + A_{21}(1) \cdot i^2 \end{aligned}$$

A mesma coisa, é claro, também pode ser feita com $A_1(x)$.

Não é difícil ver que nós fizems a mesma coisa duas vezes.

E ver isso é a chave para generalizar a coisa ...

4 Generalização

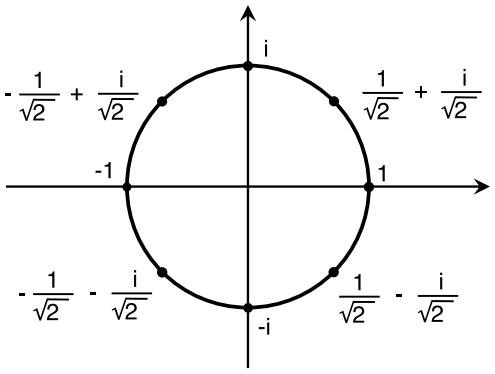
Uma propriedade importante da coleção de pontos com que trabalhamos na seção anterior é que

$$1 \;=\; 1^4 \;=\; (-1)^4 \;=\; i^4 \;=\; (-i)^4$$

Logo, para fazer a coisa funcionar, digamos, com 8 pontos, nós precisamos encontrar uma coleção com a seguinte propriedade:

$$1 \;=\; p_1^8 \;=\; p_2^8 \;=\; \ldots \;=\; p_8^8$$

Mas, isso não é uma tarefa difícil no plano complexo: basta pegar 8 pontos equidistantes no círculo unitário, começando com o 1.



(Verifique isso!)

Em geral, para formar uma coleção de tamanho 2^m, nós pegamos 2^m pontos equidistantes no círculo unitário (começando no 1).

Nós vamos chamar essa coleção de U_m.[†]

- para todo ponto p em U_m, o ponto $-p$ também está em U_m — pois, p e $-p$ ficam em extremos opostos no círculo
- para todo ponto p em U_m, o ponto p^2 também está em U_m — (porque?)

Essas duas propriedades tem uma consequência interessante:

- quando nós elevamos todos os elementos de U_m ao quadrado — denote isso por $\big(U_m\big)^2$ — nós obtemos um subconjunto de U_m com a metade do tamanho (i.e., 2^{m-1} pontos).

E o mais interessante de tudo é que

$$\big(U_m\big)^2 \quad = \quad U_{m-1}$$

Isso já estava acontecendo na seção anterior, veja só

$$U_1 \quad = \quad \{1,\, -1\} \qquad\qquad U_2 \quad = \quad \{1,\, -1,\, i,\, -i\}$$

e

$$\big(U_2\big)^2 \quad = \quad U_1$$

Agora nós estamos prontos para fazer as coisas funcionarem recursivamente.

O nosso problema consiste em calcular o valor do polinômio $A(x)$ em todos os pontos da coleção U_m.

[†]Esse nome é duplamente apropriado: de um lado, U e m se referem ao fato de que são 2^m pontos no círculo unitário, e de outro U_m é parecido com Um.

Como já sabemos, a ideia consiste em separar as potências pares e ímpares de $A(x)$

$$A(x) \;=\; \underbrace{a_0 \;+\; a_2x^2 \;+\; \ldots \;+\; a_nx^n}_{\text{par}} \;+\; \underbrace{a_1x \;+\; a_3x^3 \;+\; \ldots \;+\; a_{n-1}x^{n-1}}_{\text{ímpar}}$$

e, a seguir, nós fatoramos o termo x na parte ímpar, para obter

$$A(x) \;=\; \underbrace{a_0 \;+\; a_2x^2 \;+\; \ldots \;+\; a_nx^n}_{A_2(x)} \;+\; \underbrace{\Big[a_1 \;+\; a_3x^2 \;+\; \ldots \;+\; a_{n-1}x^{n-2}\Big]}_{A_1(x)} \cdot x$$

de modo que

$$A(x) \;=\; A_2(x) \;+\; A_1(x) \cdot x$$

Essa decomposição permite que os termos $A_2(x)$ e $A_1(x)$ sejam calculados apenas sobre a metade dos pontos — digamos, a metade de cima do círculo.

Isto é, nós estamos calculando valores de polinômios com a metade do tamanho de $A(x)$ (i.e., a metade dos termos), para uma coleção de pontos com a metade do tamanho de U_m.

No entanto, a recursão ainda não está perfeita porque $A_2(x)$ e $A_1(x)$ não são polinômios comuns, mas polinômios que só possuem potências pares.

Mas, isso é um problema fácil de resolver.

Basta escrever

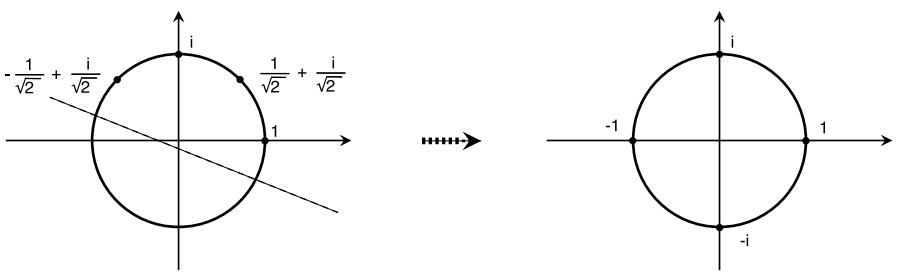
$$A_2(x) \;=\; a_0 \;+\; a_2\big(x^2\big) \;+\; a_4\big(x^2\big)^2 \;+\; a_6\big(x^2\big)^3 \;+\; \ldots \;+\; a_n\big(x^2\big)^{n/2}$$

$$A_1(x) \;=\; a_1 \;+\; a_3\big(x^2\big) \;+\; a_5\big(x^2\big)^2 \;+\; a_7\big(x^2\big)^3 \;+\; \ldots \;+\; a_{n-1}\big(x^2\big)^{(n-2)/2}$$

Isto é, nós podemos ver $A_2(x)$ e $A_1(x)$ como polinômios (comuns) de graus $n/2$ e $(n-2)/2$, sobre o parâmetro x^2.

Isso significa que nós vamos calcular os valores $A_2(x)$ e $A_1(x)$ (na versão acima) sobre os quadrados dos pontos na metade de cima do círculo.

O passo final consiste em observar o que acontece quando nós elevamos os pontos na metade de cima do círculo ao quadrado:[‡]



[‡]Na realidade, isso funciona para qualquer metade de pontos consecutivos no círculo — (porque?)

É isso mesmo! nós obtemos a coleção U_{m-1}!

Resumindo tudo o que nós acabamos de fazer:

- para calcular o valor do polinômio $A(x)$ sobre a coleção U_m
- nós decompomos $A(x)$ nos polinômios $A_2(x)$ e $A_1(x)$ (como indicado acima)
- calculamos os valores de $A_2(x)$ e A_1 sobre a coleção U_{m-1}
- e depois calculamos os valores de $A(x)$ a partir dos valores obtidos para $A_2(x)$ e $A_1(x)$

Essas observações já são suficientes para construir um algoritmo de divisão e conquista para a etapa 1:

```
Procedimento  ConversãoR1R2-DC  ( A(x), U_m )
 {
     Se ( m = 0 )   Retorna A(1)

     (A2,A1)  <--  Quebra ( A(x) )

          V2  <--  ConversãoR1R2-DC ( A2(x), U_m-1 )
          V1  <--  ConversãoR1R2-DC ( A1(x), U_m-1 )

     Para cada par de pontos (p,-p) em U_m
      {
          V(p)  <--  V2(p)  +  V1(p)
         V(-p)  <--  V2(p)  +  V1(p) * -p
      {
     Retorna (V)
 }
```

Examinando o pseudo-código desse procedimento, não é difícil chegar à seguinte equação de recorrência para o seu tempo de execução:[§]

$$T(n) \quad = \quad 2 \cdot T(n/2) \; + \; O(n)$$

Essa é uma equação de recorrência bem conhecida, que possui solução

$$T(n) \quad = \quad \Theta(n \log n)$$

[§]Aqui, nós estamos assumindo que $2^m = O(n)$, o que é o caso para a nossa aplicação de multiplicação de polinômios.

5 Conversão R2 $\rightarrow$ R1

Do ponto de vista da nossa disciplina de algoritmos, talvez a gente já pudesse parar por aqui.

Quer dizer, a solução que vimos para o problema da conversão R1 $\rightarrow$ R2 nos dá o exemplo mais interessante de divisão e conquista que nós vamos ver no curso.

Mas, do ponto de vista do problema da multiplicação de polinômios, é preciso seguir adiante e mostrar como se resolve a conversão R2 $\rightarrow$ R1.

(Nós vamos ver que essa solução é tão interessante quanto a anterior ...)

Nós vimos na Introdução que a conversão R2 $\rightarrow$ R1 consiste basicamente na solução de um sistema de equação lineares.

$$\begin{aligned} 1 &= a \cdot 1^2 + b \cdot 1 + c \\ 2 &= a \cdot 3^2 + b \cdot 3 + c \\ 3 &= a \cdot 4^2 + b \cdot 4 + c \end{aligned}$$

Isto é, cada ponto da coleção U_m nos dá uma equação, onde a equação é obtida substituindo a variável x no polinômio pelo valor ponto, e o outro lado da equação é o valor do polinômio naquele ponto.

Abaixo nós temos o sistema de equação no caso geral (em notação matricial):

$$\begin{bmatrix} 1 & p_0 & p_0^2 & \dots & p_0^{n-1} \\ 1 & p_1 & p_1^2 & \dots & p_1^{n-1} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & p_{n-1} & p_{n-1}^2 & \dots & p_{n-1}^{n-1} \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \\ \dots \\ a_{n-1} \end{bmatrix} = \begin{bmatrix} A(p_0) \\ A(p_1) \\ \dots \\ A(p_{n-1}) \end{bmatrix}$$

Note o formato peculiar dessa matriz.

Um dia[†], em que não tinha nada de mais importante para fazer, Alexandre Vandermonde ficou curioso a respeito dessa matriz e resolveu estudá-la.

Ele descobriu que a sua inversa também tem um formato peculiar.

No nosso caso particular (onde as entradas da matriz são pontos da coleção U_m) a inversa é a seguinte:

[†]Provavelmente mais de um dia ...

$$\frac{1}{n} \cdot \begin{bmatrix} 1 & p_0^{-1} & p_0^{-2} & \dots & p_0^{-(n-1)} \\ 1 & p_1^{-1} & p_1^{-2} & \dots & p_1^{-(n-1)} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & p_{n-1}^{-1} & p_{n-1}^{-2} & \dots & p_{n-1}^{-(n-1)} \end{bmatrix}$$

Multiplicando essa inversa nos dois lados da equação acima (e trocando a ordem dos lados), nós obtemos

$$\frac{1}{n} \cdot \begin{bmatrix} 1 & p_0^{-1} & p_0^{-2} & \dots & p_0^{-(n-1)} \\ 1 & p_1^{-1} & p_1^{-2} & \dots & p_1^{-(n-1)} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & p_{n-1}^{-1} & p_{n-1}^{-2} & \dots & p_{n-1}^{-(n-1)} \end{bmatrix} \cdot \begin{bmatrix} A(p_0) \\ A(p_1) \\ \dots \\ A(p_{n-1}) \end{bmatrix} = \begin{bmatrix} a_0 \\ a_1 \\ \dots \\ a_{n-1} \end{bmatrix}$$

Ou seja, para obter os coeficientes $a_0, a_1, ldots, a_n$, basta fazer uma multiplicação de matriz por vetor.

Infelizmente, essa multiplicação leva tempo $\Theta(n^2)$ ...

Mas, isso não é necessário!

Note que essa multiplicação corresponde a calcular o valor do polinômio com coeficientes $A(p_0)$, $A(p_1)$, ..., $A(p_{n-1})$ sobre a coleção de pontos $U'_m = \{p_0^{-1}, p_1^{-1}, \dots, p_{n-1}^{-1}\}$.

A coleção de pontos U'_m possui as mesmas propriedades da coleção U_m.

Logo, a conversão `R2` $\rightarrow$ `R1` pode ser realizada da mesma forma que a conversão `R1` $\rightarrow$ `R2`.

E isso leva tempo $\Theta(n \log n)$.

Juntando todas as etapas do processo, nós tempos um algoritmo para a multiplicação de polinômios de grau n que executa em tempo

$$T(n) \quad = \quad \Theta(n \log n) \; + \; \Theta(n) \; + \; \Theta(n \log n) \quad = \quad \Theta(n \log n)$$