Rafael V. Cacilhas - Relatório 01 (August 21, 2018)

Exercício 1

O software produzido converte um número binário em linguagem de computador para um numero decimal e converte um numero decimal para a notação padrão de binário (não em linguagem de computador).

1 a)

O numero binário que será convertido pode ser modificado editando o arquivo number.dat, bastando escrever cada dígito do número em uma linha diferente.

Foi feito um algoritmo para que o número em binário fosse lido diretamente de uma variavel integer, mas devido à grande extensão do algarismo (32 bits) não foi possivel criar uma variavel inteira tão grande e portanto o algoritmo não funciona para binários tão grandes.Ele foi deixado em comentário para quem se interesse.

Cada algarismo do numero é lido para uma posição de um array de tamanho 32 (bin) e é em seguida separado em três partes: sinal (bin(1)), expoente (bin(2) até bin(9)) e mantissa (bin(10) até bin(32)). Note que os arrays expo e mantiss invertem a ordem do array bin, por questão de conveniencia.

A saída do programa na tela é a conversão do número binário em decimal.

1 b)

A conversão de decimal para binário é feita primeiramente separando o número decimal em sua parte puramente inteira e em sua parte puramente decimal. A parte inteira é trivial, pois o algoritmo de transformação em binário é extremamente simples.

A parte decimal traz o problema de não ser possível determinar de maneira simples quando o programa deve parar de rodar, pois o algoritmo não possui um fim explicito. A detecção de uma dízima não é simples de maneira geral, pois consiste em conferir todos os N-1 termos anteriores, o que é extremamente ineficiente para N grande. Foi escolhido utilizar um numero fixos de casas decimais (variavel precisao) e analisar caso a caso em busca de dízimas. O número, convertido para binário, é escrito no arquivo bin.dat da mesma maneira que o arquivo fonte da letra a). Neste caso é 10011001011,1101110000.

O número pedido não possui uma dízima, como foi verificado até a vigésima casa decimal. Apesar de ter sido calculado até a vigésima casa foi considerado desnecessário mostrar à partir da décima, devido ao grande número de zeros presentes; o erro associado é muito pequeno.

Exercício 2

Com o objetivo de calcular o erro associado ao truncamento de uma variável foi definido o valor "exato" da subtração como sendo aquele feito com variaveis reais tipo *8.

Para truncar na quinta casa decimal foi utilizado um pequeno truque: o valor da subtração foi multiplicado por 10^3 de modo que o resultado tenha 5 casas decimais antes da vírgula. Em seguida esse valor foi salvo em uma variavel do tipo integer, de modo que as casas decimais fossem ignoradas. A seguir o valor foi transferido da variavel integer para uma variavel real*8 e dividido por 10^3 de modo a ter somente as 5 casas decimais.

As saídas de dados na tela apresentam o resultado da subtração exato, truncado, erro absoluto e relativo entre eles. Como esperado a diferença entre a subtração exata e truncada se encontra na quinta casa decimal. Os valores de erro estão um pouco abaixo desta ordem de grandeza.

Naturalmente para uma conta simples desta não faz sentido o truncamento; mas é interessante ter o valor do erro para cálculos mais complexos.

Exercício 3

O software calcula o valor do somatório dado para variaveis real*4 e real*8. Vemos que na sétima casa decimal já há divergência entre os resultados.

Em seguida é calculado o mesmo somatório, mas começando dos maiores termos. A divergência entre real*4 e real*8 ocorre aproximadamente no mesmo local, na oitava casa. No entanto é possível perceber uma diferença enorme entre o resultado obtido neste cálculo e no cálculo anterior, sendo que o cálculo atual possui quase 100% de erro relativo ao valor calculado anteriormente.

Exercício 4

Temos abaixo a equação de Van der Waals.

$$\left(p + \frac{N^2 a'}{V^2}\right)(V - Nb') = Nk_bT \qquad . \tag{1}$$

a)

Segue os gráficos para T = Tc/2, T = Tc e T = 3*Tc/2 para dois ranges de volume diferentes.

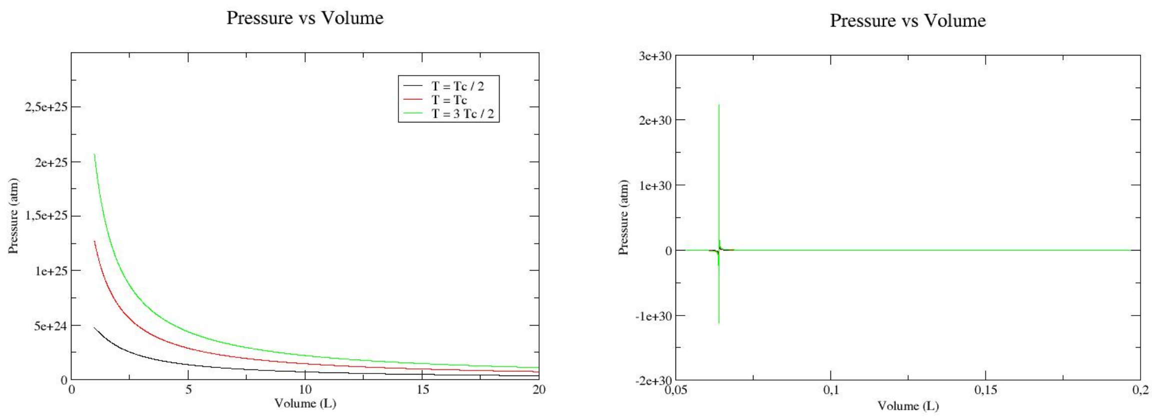

Figure 1: Pressão em função do volume para dois ranges de volume.

b)

Definindo $T' = \frac{T}{T_c}$,$p' = \frac{p}{p_c}$ e $V' = \frac{V}{V_c}$, onde $T_c = \frac{27k_b b'}{8a'}$, $p_c = \frac{a'}{27b'^2}$ e $V_c = 3Nb'$ e multiplicando os dois lados da equação por $\frac{27k_b b'}{8a'}$, podemos reescrever a equação como:

$$\left(p' + \frac{3}{8V'^2}\right)\left(V' - \frac{1}{3}\right) = 3T' \qquad . \tag{2}$$

Abaixo seguem os graficos de P' em função de V'.

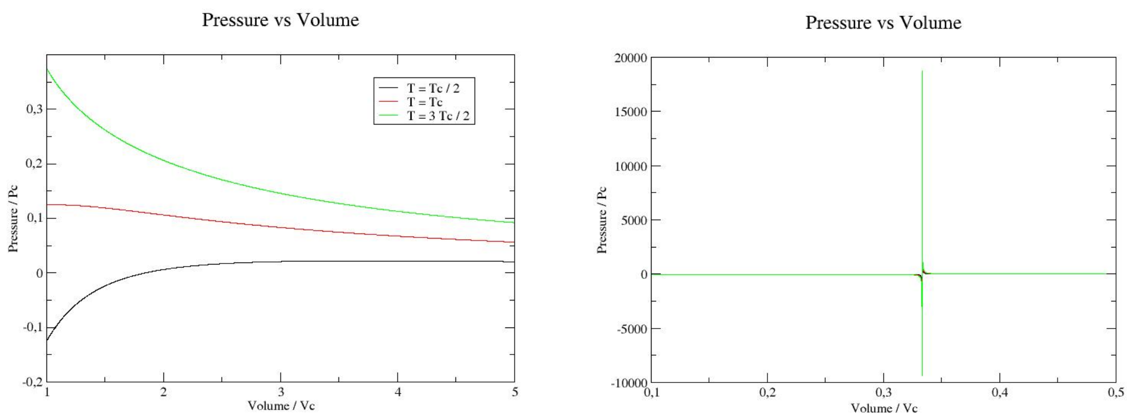

Figure 2: Pressão em função do volume para dois ranges de volume.

Como podemos ver na Figura 2 o uso das variaveis auxiliares faz com que as grandezas P' e V' tenham aproximadamente a mesma ordem de grandeza, fato que auxilia na compreensão do gráfico, além de fazer com que P não tenha valores extremamente altos, o que ajuda no cálculo computacional. Além disso, escrever a pressão como P' = P / Pc faz com que o comportamento das curvas em temperaturas diferentes seja drásticamente diferente, onde vemos que para T = Tc/2, abaixo de V' = 2, temos a separação das três curvas, o que não ocorre no caso da Figura

1. Por fim, a separação também traz facilidade no entendimento da expressão analítica, onde agora fica claro que a função deve divergir perto de V' = 1/3, fato que não é tão explícito (ao menos não em termos do valor numérico) na equação 1.

Exercicio 5

Não consegui aplicar a soma ao problema proposto, pois o algoritmo da soma trabalha com a soma de dois números, dados em log. O cálculo da função de partição trabalha com Z, não com logZ. Se tentarmos aplicar o log, ficamos com:

$$Z = Z_1 + Z_2 + Z_3 + ... + Z_n \qquad . \tag{3}$$

$$logZ = log(Z_1 + Z_2 + Z_3 + ... + Z_n) \qquad . \tag{4}$$

Como o log da soma não é a soma dos logs não vejo como aplicar o algoritmo da soma para este caso, visto que o lado direito não é a soma de dois logs.

Exercicio 6

a)

$$S = \sum_{k=1}^{N} [a * x_k + b - y_k]^2 \qquad . \tag{5}$$

Derivando em relação a a e b, temos:

$$\frac{\partial S}{\partial a} = 2\sum_{k=1}^{N} [ax_k + b - y_k]x_k \qquad . \tag{6}$$

$$\frac{\partial S}{\partial b} = 2\sum_{k=1}^{N} [ax_k + b - y_k] \qquad . \tag{7}$$

Separando os três termos do somatório e igualando a zero (pois queremos o mínimo), temos:

$$0 = \sum_{k=1}^{N} ax_k^2 + \sum_{k=1}^{N} bx_k - \sum_{k=1}^{N} y_k x_k \qquad . \tag{8}$$

$$0 = \sum_{k=1}^{N} ax_k + \sum_{k=1}^{N} b - \sum_{k=1}^{N} y_k \tag{9}$$

Vamos definir A = , A = $\sum x$, B = $\sum y$, C = $\sum x^2$ e D = $\sum xy$. Assim, o sistema fica da forma:

$$aC + bA - D = 0 \tag{10}$$

$$aA + Nb - B = 0 \tag{11}$$

Cujas soluções são:

$$a = \frac{ND - AB}{NC - A^2} \tag{12}$$

$$b = \frac{BC - AD}{NC - A^2} \tag{13}$$

Note que se quisermos forçar b = 0, temos que $B = \frac{AD}{C}$ e temos que:

$$a_{b=0} = \frac{ND - \frac{A^2 D}{C}}{NC - A^2} \tag{14}$$

b

A saída do programa apresenta primeiramente na tela os valores do coeficiente angular a e o coeficiente linear b. A equação da reta obtida pelo Grace é $y = 1.6382 * x + 0.029209$, onde vemos que o coeficiente angular bateu até a ultima casa de precisão do Grace, mas o coeficiente linear apresentou diferença já na segunda casa decimal.

c)

A segunda linha da saída do programa é o coeficiente angular quando forçamos $b = 0$. Como o b encontrado foi maior que zero, temos que a possui um valor ligeiramente maior do que o encontrado anteriormente, como esperado.

O cálculo analítico já foi apresentado na letra a).

d)

O coeficiente angular do experimento de Millikan representa a razão entre a massa das gotículas e a carga do elétron. Como não temos a unidade da massa utilizada o valor absoluto da razão não faz sentido, sendo que podemos então ignorar o expoente e trabalhar com o valor numérico. Comparando o valor numérico (esquecendo os expoentes) do coeficiente angular com a carga do elétron, temos que o experimento apresenta um erro relativo da ordem de 2%

Exercicio 7

a)

A equação da reta encontrada pelo Grace é $y = 67.415 * x + 758.39$. A saída do programa mostra primeiramente os resultados do fit através da rotina do Exercicio 6. Como podemos ver, $a = 66.772$ e $b = 908.71$. Os resultados de a discordam do Grace na segunda casa decimal e os de b discordam até mesmo na primeira.

b)

O software foi feito seguindo a dedução analítica do Recipe e reescrevendo o algoritmo para FORTRAN90. Os resultados do calculo do coeficiente angular parecem coerentes, mas o erro calculado é bem grande, o que torna dificil saber se o algoritmo foi feito com sucesso. O cálculo e erro do coeficiente linear b estão bem fora dos valores esperados pelo fit do Grace.

Infelizmente não consigo encontrar nenhum erro no algoritmo.