05 Curve Fitting

05 Curve Fitting

Introduction

- Data are often given for discrete values along a continuum. However, we may require estimates at points between the discrete values.
- This Chapter describes techniques to fit curves to such data to obtain intermediate estimates.

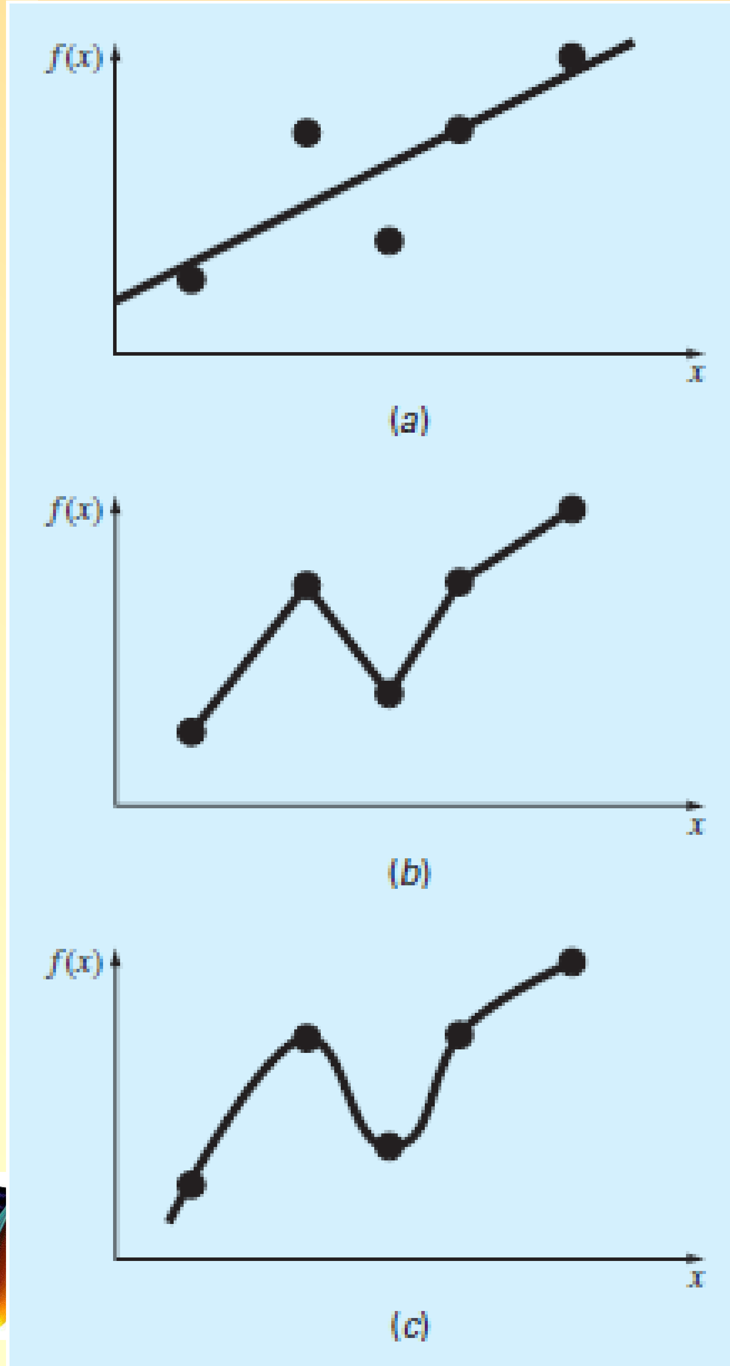
- There are two general approaches for curve fitting that are distinguished from each other on the basis of the amount of error associated with the data.
  - Regression
  - Interpolation

Regression

- The data exhibit a **significant degree of error**, the strategy is to derive a single curve that represents **the general trend** of the data.
- Because any individual data point may be incorrect, we make no effort to intersect every point. Rather, the curve is designed to follow the pattern of the points taken as a group. One approach of this nature is called *least-squares regression*

Interpolation

- The data are known to be very precise, the basic approach is to fit a curve or a series of curves that pass directly through each of the points.
- Such data usually originate from tables. Examples are values for the density of water or for the heat capacity of gases as a function of temperature.
- The estimation of values between well-known discrete points is called *interpolation*



Three attempts to fit a "best" curve through five data points:

(*a*) least-squares regression,

(*b*) linear interpolation,

(*c*) curvilinear interpolation.

5.1. Regression
a. Linear Regression

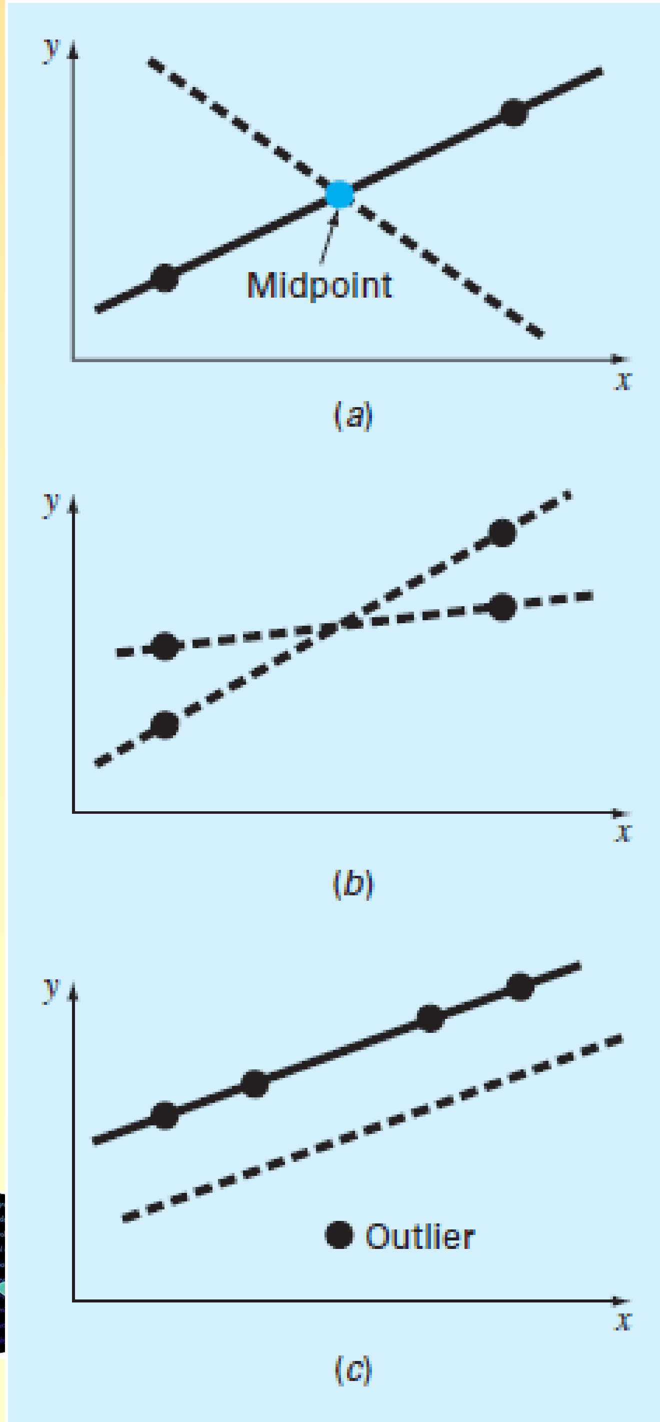
- The primary objective of this method is to how least-squares regression can be used to fit a straight line to measured data.

The mathematical expression for the straight line is,

$$y = a_0 + a_1 x$$

- One strategy for fitting a "best" line through the data would be to minimize the sum of the residual errors for all the available data, as in

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)$$



Examples of some criteria for “best fit” that are inadequate for regression:

(*a*) minimizes the sum of the residuals,

(*b*) minimizes the sum of the absolute values of the residuals, and

(*c*) minimizes the maximum error of any individual point.

- A strategy that overcomes the shortcomings of the aforementioned approaches is to minimize the sum of the squares of the residuals:

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

- To determine values for $a0$ and $a1$, Eq. above is differentiated with respect to each unknown coefficient:

$$\frac{\partial S_r}{\partial a_0} = -2 \sum (y_i - a_0 - a_1 x_i)$$

$$\frac{\partial S_r}{\partial a_1} = -2 \sum [(y_i - a_0 - a_1 x_i) x_i]$$

$$0 = \sum y_i - \sum a_0 - \sum a_1 x_i$$

$$0 = \sum x_i y_i - \sum a_0 x_i - \sum a_1 x_i^2$$

$$n \quad a_0 + \left(\sum x_i\right) a_1 = \sum y_i$$

$$\left(\sum x_i\right) a_0 + \left(\sum x_i^2\right) a_1 = \sum x_i y_i$$

- These are called the *normal equations*. They can be solved simultaneously for

$$a_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - \left(\sum x_i\right)^2}$$

$$a_0 = \bar{y} - a_1 \bar{x}$$

Problem

- Fit a straight line to the values in Table 14.1.

TABLE 14.1 Experimental data for force (N) and velocity (m/s) from a wind tunnel experiment.

v, m/s	10	20	30	40	50	60	70	80
F, N	25	70	380	550	610	1220	830	1450

LINEARIZATION OF NONLINEAR RELATIONSHIPS

- Linear regression provides a powerful technique for fitting a best line to data. However, it is predicated on the fact that the relationship between the dependent and independent variables is linear.
- This is not always the case, and the first step in any regression analysis should be to plot and visually inspect the data to ascertain whether a linear model applies.

- In some cases, techniques such as polynomial regression, which is described in next sub Chap., are appropriate. For others, transformations can be used to express the data in a form that is compatible with linear regression.

Contoh linierisasi hubungan yang tidak linier

1. $y = a\,x^{b}$
2. $y = a\,e^{bx}$
3. $y = a + \dfrac{b}{x}$
4. $y = \dfrac{x}{a + bx}$
5. $y = \dfrac{ax}{b + x}$

b. Regresi Polinomial

Jika pada regresi linier, data-data didekati dengan persamaan garis $y = a_1x + a_0$, maka pada regresi polinomial, data-data didekati dengan persamaan polinomial, yaitu :

$$y = a_0 + a_1x + a_2x^2 + ... + a_mx^m$$

sehingga

$$SSE = \sum_{n=1}^{n} \left(y_{data} - y_{pers}\right)^2$$

$$a_0 n + a_1\sum x_i + a_2\sum x_i^2 + \dots + a_m\sum x_i^m = \sum y_i$$

$$a_0\sum x_i + a_1\sum x_i^2 + a_2\sum x_i^3 + \dots + a_m\sum x_i^{m+1} = \sum x_i y_i$$

$$a_0\sum x_i^2 + a_1\sum x_i^3 + a_3\sum x_i^4 + \dots + a_m\sum x_i^{m+2} = \sum x_i^2 y_i$$

.

.

.

.

$$a_0\sum x_i^m + a_1\sum x_i^{m+1} + a_2\sum x_i^{m+2} + \dots + a_m\sum x_i^{2m} = \sum x_i^m y_i$$

$$\begin{bmatrix} n & \sum x_i & \sum x_i^2 & \dots & \sum x_i^m \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & & \sum x_i^{m+1} \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 & & \sum x_i^{m+2} \\ \dots & & & & \\ \sum x_i^m & \sum x_i^{m+1} & \sum x_i^{m+2} & \dots & \sum x_i^{2m} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \dots \\ a_m \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \\ \dots \\ \sum x_i^m y_i \end{bmatrix}$$

Regresi dg Matlab
Fungsi Polyfit

Dalam Matlab, fungsi `polyfit` menyelesaikan masalah pencocokan kurva untuk kuadrat terkecil.

Penggunaan fungsi `polyfit` akan menghasilkan suatu persamaan polinomial yang paling mendekati data.

Jika derajat fungsi `polyfit` dipilih n = 1, maka akan dihasilkan persamaan garis lurus yaitu regresi linier.

```
kons = polyfit(x,y,1)
```

Soal 1

Bennett & Meyers memberikan data-data konduktivitas termal aseton. Hubungan ln (k) dan ln (T) mendekati persamaan garis lurus, dengan T dalam R (R = ^{O}F + 460). Tentukan nilai k pada 300 ^{O}F !

k [BTU/jam.ft.^{O}F]	T (^{O}F)
0,0057	32
0,0074	115
0,0099	212
0,0147	363

Soal 2

Tekanan uap toluena pada persamaan Antoine

Data tekanan uap toluena dalam ^{O}C dan Torr adalah

T (^{O}C)	-26,7	-4,4	6,4	18,4	31,8	40,3
p (torr)	1	5	10	20	40	60

Tentukan konstanta-konstanta persamaan Antoine yang menyatakan hubungan T dan p sebagai berikut

$$p = \exp\left(a + \frac{b}{T + 273{,}2} \right)$$

lanjutan

$$p = \exp\left(a + \frac{b}{T + 273{,}2}\right)$$

$$\ln(p) = a + \frac{b}{T + 273{,}2}$$

$$y = a + b\,x$$

$$y = \ln(p) \text{ dan } x = \frac{1}{T + 273{,}2}$$

lanjutan

```
% Data - data
p = [ 1 5 10 20 40 60];                    % C
T = [-26.7 -4.4 6.4 18.4 31.8 40.3]; % Torr

% Linierisasi
y = log(p);
x = 1./(T+273.15);

kons = polyfit(x,y,1);
a = kons(2)
b = kons(1)
```

Regresi Polinomial

data-data didekati dengan persamaan polinomial, yaitu :

$$y = a_0 + a_1 x + a_2 x^2 + \ldots\ldots + a_m x^m$$

lanjutan

Fungsi polyfit dari Matlab dapat digunakan untuk derajat n = 2, yang disebut polinomial kuadratis. Hasil fungsi polyfit adalah vektor baris yang berisi koefisien-koefisien polinomial.

Selain fungsi polyfit, Matlab juga mempunyai fungsi polyval untuk mengevaluasi polinomial pada tiap titik yang ingin diketahui nilainya.

Soal 2

Persamaan Fit untuk Data Tekanan Uap Benzena

Data tekanan uap murni benzena pada berbagai temperatur ditunjukkan pada tabel.

Temperatur (^{O}C)	**Tekanan (mmHg)**	**Temperatur (^{O}C)**	**Tekanan (mmHg)**
-36,7	1	15,4	60
-19,6	5	26,1	100
-11,5	10	42,2	200
-2,6	20	60,6	400
7,6	40	80,1	760

Hubungan tekanan sebagai fungsi temperatur dapat dinyatakan sebagai persamaan empiris polinomial sederhana sebagai berikut :

$$P = a_0 + a_1T + a_2T^2 + a_3T^3 + \ldots + a_nT^n$$

dengan a_0, a_1, …, a_n adalah parameter yang ditentukan dengan regresi dan n adalah orde polinomial.

lanjutan

```
vp = [ 1 5 10 20 40 60 100 200 400 760];
T =[-36.7 -19.6 -11.5 -2.6 7.6 15.4 26.1
  42.2 60.6 80.1];
% Penggunaan fungsi polyfit
m = 3;                      %orde polinomial
k = polyfit(T,vp,m);
% Evaluasi persamaan polinomial
z = polyval(k,55)
```

lanjutan

```
vp = [ 1 5 10 20 40 60 100 200 400 760];
T =[-36.7 -19.6 -11.5 -2.6 7.6 15.4 26.1 42.2 60.6 80.1];
% Penggunaan fungsi polyfit
m = 3;                        %orde polinomial
k = polyfit(T,vp,m);
% Evaluasi persamaan polinomial
z = polyval(k,T);
% Cetak hasil
plot(T,z,'k-',T,vp,'ko','linewidth',2)
xlabel('T(C)')
ylabel('vp(mmHg)')
```

Next meeting

05 Curve Fitting

- There are two general approaches for curve fitting that are distinguished from each other on the basis of the amount of error associated with the data.

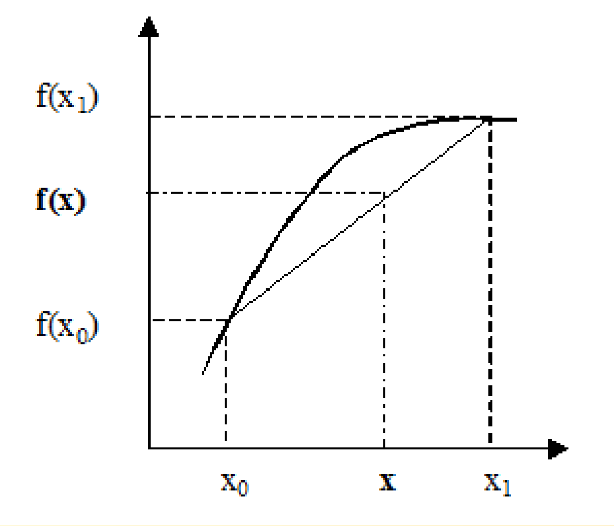
  - Regression
  - Interpolation

6.2. Interpolation

- Interpolation is used to estimate intermediate values between precise data points.

a. Linear Interpolation

- The simplest form of interpolation is to connect two data points with a straight line. This technique, called *linear interpolation*, is depicted graphically in Fig.



$$\frac{f(x)-f(x_0)}{x-x_0}=\frac{f(x_1)-f(x_0)}{x_1-x_0}$$

$$f(x)=f(x_0)+\frac{f(x_1)-f(x_0)}{(x_1-x_0)}(x-x_0)$$

Newton linear-interpolation formula.

- Find ρ, μ, and v at T = 27 °C

T, °C	ρ, kg/m³	μ, N · s/m²	v, m²/s
−40	1.52	1.51×10^{-5}	0.99×10^{-5}
0	1.29	1.71×10^{-5}	1.33×10^{-5}
20	1.20	1.80×10^{-5}	1.50×10^{-5}
50	1.09	1.95×10^{-5}	1.79×10^{-5}
100	0.946	2.17×10^{-5}	2.30×10^{-5}
150	0.835	2.38×10^{-5}	2.85×10^{-5}
200	0.746	2.57×10^{-5}	3.45×10^{-5}
250	0.675	2.75×10^{-5}	4.08×10^{-5}
300	0.616	2.93×10^{-5}	4.75×10^{-5}
400	0.525	3.25×10^{-5}	6.20×10^{-5}
500	0.457	3.55×10^{-5}	7.77×10^{-5}

$$f(x) = f(x_0) + \frac{f(x_1) - f(x_0)}{(x_1 - x_0)}(x - x_0)$$

b. Quadratic Interpolation

- A strategy for improving the estimate is to introduce some curvature into the line connecting the points. If three data points are available, this can be accomplished with a second-order polynomial (also called a quadratic polynomial or a parabola).

Bentuk umum

$$f(x) = b_0 + b_1(x-x_0) + b_2(x-x_0)(x-x_1)$$

dengan

$$b_0 = f(x_0)$$

$$b_1 = \frac{f(x_1)-f(x_0)}{(x_1-x_0)}$$

$$b_2 = \frac{\dfrac{f(x_2)-f(x_1)}{(x_2-x_1)} - \dfrac{f(x_1)-f(x_0)}{(x_1-x_0)}}{(x_2-x_0)}$$

- Find ρ, μ, and v at T = 27 ^oC

T, °C	ρ, kg/m³	μ, N · s/m²	υ, m²/s
−40	1.52	1.51×10^{-5}	0.99×10^{-5}
0	1.29	1.71×10^{-5}	1.33×10^{-5}
20	1.20	1.80×10^{-5}	1.50×10^{-5}
50	1.09	1.95×10^{-5}	1.79×10^{-5}
100	0.946	2.17×10^{-5}	2.30×10^{-5}
150	0.835	2.38×10^{-5}	2.85×10^{-5}
200	0.746	2.57×10^{-5}	3.45×10^{-5}
250	0.675	2.75×10^{-5}	4.08×10^{-5}
300	0.616	2.93×10^{-5}	4.75×10^{-5}
400	0.525	3.25×10^{-5}	6.20×10^{-5}
500	0.457	3.55×10^{-5}	7.77×10^{-5}

Bentuk umum

$$f(x) = b_0 + b_1(x-x_0) + b_2(x-x_0)(x-x_1)$$

dengan

$$b_0 = f(x_0)$$

$$b_1 = \frac{f(x_1)-f(x_0)}{(x_1-x_0)}$$

$$b_2 = \frac{\frac{f(x_2)-f(x_1)}{(x_2-x_1)} - \frac{f(x_1)-f(x_0)}{(x_1-x_0)}}{(x_2-x_0)}$$

General Form of Newton's Interpolating Polynomials

Bentuk umum interpolasi polinomial Newton

$$f(x) = b_0 + b_1(x-x_0) + b_2(x-x_0)(x-x_1) + \dots + b_n(x-x_0)(x-x_1)(x-x_2)\dots(x-x_n)$$

dengan

$$b_0 = f(x_0)$$

$$b_1 = f[x_1,x_0]$$

$$b_2 = f[x_2,x_1,x_0]$$

.

.

.

$$b_n = f[x_n,x_{n-1}, \dots,x_1,x_0]$$

Differensiasi terbagi pertama

$$f[x_i,x_j] = \frac{f(x_i)-f(x_j)}{(x_i-x_j)}$$

Differensiasi terbagi kedua

$$f[x_i,x_j,x_k] = \frac{f[x_i, x_j] - f[x_j, x_k]}{x_i - x_j}$$

Interpolasi polinomial order 3

i	x_i	$f(x_i)$	I	II	III
0	x_0	$f(x_i)$			
			$f[x_1,x_0]$		
1	x_1	$f(x_1)$		$f[x_2,x_1,x_0]$	
			$f[x_2,x_1]$		$f[x_3,x_2,x_1,x_0]$
2	x_2	$f(x_2)$		$f[x_3,x_2,x_1]$	
			$f[x_3,x_2]$		
3	x_3	$f(x_3)$			

From data of acetic acid density at 25 ^{O}C, find the density of acetic acid at 65 % !

Data densitas asam asetat pada 25 ^{O}C

Konsentrasi (%)	Densitas (g/cm^3)
0	0,997
10	1,011
20	1,024
30	1,035
40	1,045
50	1,053
60	1,06
70	1,064
80	1,065
90	1,061
100	1,044

Interpolation with Matlab

Function

interp1 for linier interpolation.

YI = interp1 (X,Y,XI,'the methode')

For find Y YI if X = X1, with data Y vs X

The methode

- 'nearest'
- 'linier'
- 'spline'
- 'cubic'

Konduktivitas termal aseton

Bennett & Meyers memberikan data-data konduktivitas termal aseton. Hubungan ln (k) dan ln (T) mendekati persamaan garis lurus, dengan T dalam R (R = ^{O}F + 460). Tentukan nilai k pada 300 ^{O}F !

k [BTU/jam.ft.^{O}F]	**T (^{O}F)**
0,0057	32
0,0074	115
0,0099	212
0,0147	363

lanjutan

```
k = [0.0057 0.0074 0.0099 0.0147];
T = [32 115 212 363]+460;
  % Hubungan k dan T linier
ln_k = log(k);
ln_T = log(T);
  % Interpolasi Linier
T300 = log(300+460);
k300 = interp1(ln_T,ln_k,T300,'linier');
  % Hasil
k_pd_300 = exp(k300)
```

b. Splines Interpolation

- n[th]-order polynomials were used to interpolate between *(n+1)* data points.
- For example, for eight points, we can derive a perfect seventh-order polynomial. This curve would capture all the meanderings (at least up to and including seventh derivatives) suggested by the points.
- However, there are cases where these functions can lead to erroneous results because of round-off error and oscillations.
- An alternative approach is to apply lower-order polynomials in a piecewise fashion to subsets of data points. Such connecting polynomials are called *spline functions*.

- For example, third-order curves employed to connect each pair of data points are called *cubic splines*.
- These functions can be constructed so that the connections between adjacent cubic equations are visually smooth.

Larutan metil dietanol amin (MDEA) banyak digunakan untuk absorpsi gas seperti H_2S dan CO_2. Dalam suatu menara absorpsi, perlu diketahui viskositas 43 % larutan metil dietanol amin (MDEA) dalam air pada 45°C. Untuk itu, ingin dipergunakan persamaan berikut:

$$\ln \mu = B_1 + \frac{B_2}{T} + B_3 T$$

Dengan μ dalam poise, dan T dalam K. B_i adalah parameter dengan ketentuan sebagai berikut:

Fraksi berat MDEA	B_1	$B_2.10^{-3}$	$B_3.10^2$
0	-19,52	3,913	2,112
0,1	-22,14	4,475	2,470
0,2	-25,16	5,157	2,859
0,3	-28,38	5,908	3,255
0,4	-31,52	6,678	3,634
0,5	-34,51	7,417	3,972

Tentukan viskositas larutan pada 43 % larutan dan 45°C.

```
% Data parameter viskositas sebagai fungsi x
x = [0 0.1 0.2 0.3 0.4 0.5];
B1 = [-19.52 -22.14 -25.16 -28.38 -31.52 -34.51];
B2 = [3.913 4.475 5.157 5.908 6.678 7.417]*10^3;
B3 = [2.112 2.470 2.859 3.255 3.634 3.972]*10^-2;

% Data x dan T yang ingin dicari viskositasnya
x_i = 0.43;                          % Konsentrasi
T_i = 45+273.15;                     % Temperatur, K

% Interpolasi hubungan fraksi berat vs konstanta
% Konstanta B1, B2, dan B3 pada 0.43 ditentukan terlebih dahulu
% Interpolasi nilai B pada konsentrasi 0.43
B1_i = interp1(x,B1,x_i,'spline');
B2_i = interp1(x,B2,x_i,'spline');
B3_i = interp1(x,B3,x_i,'spline');
% Viskositas dihitung
viscA = exp(B1_i + B2_i/T_i + B3_i*T_i)

% Interpolasi hubungan fraksi berat vs viskositas
% Viscositas pada 45 C ditentukan terlebih dahulu
% Viskositas pada 45 C dihitung
visc = exp(B1 + B2./T_i + B3.*T_i)
% Interpolasi viskositas pada konsentrasi 0.43
viscB = interp1(x,visc,x_i,'spline')
```

c. Two dimension Interpolation

Interpolasi dua dimensi mempunyai prinsip yang sama dengan interpolasi satu dimensi. Perbedaanya adalah interpolasi dua dimensi menginterpolasikan fungsi dua variabel $z = f(x,y)$.

ZI = interp2 (X,Y,Z,XI,YI) adalah interpolasi untuk menentukan nilai ZI, dari fungsi Z pada titik XI dan YI dari matriks X dan Y. Metode yang ditampilkan pada interpolasi 1 dimensi dapat juga digunakan pada interpolasi 2 dimensi.

Tentukan berapakah ΔH (beda entalpi) dari uap yang didinginkan menjadi 480 ^{0}F dan 52 psia, jika kondisi mula-mula adalah 640 ^{0}F dan 92 psia !

Entalphi, H dalam Btu/lb				
p (psia)	T (^{0}F)			
	400	500	600	700
50	1258,7	1282,6	-	-
55	1258,2	1282,2	-	-
90	-	-	1328,7	1378,1
95	-	-	1328,4	1377,8

```
% Kondisi awal
To = [600 700];                          % Temperatur, oF
po = [90 95];                            % Tekanan, psia
Ho = [1328.7 1378.1
      1328.4 1377.8];                    % Entalphi, BTU/lb
H1 = interp2(To,po,Ho,640,92)            % Interpolasi 2 dimensi

% Kondisi akhir
Tn = [400 500];                          % Temperatur, oF
pn = [50 55];                            % Tekanan, psia
Hn = [1258.7 1282.6
      1258.2 1282.2];                    % Entalphi, BTU/lb
H2 = interp2(Tn,pn,Hn,480,52)            % Interpolasi 2 dimensi

% Perhitungan beda entalpi
del_H = H2 - H1
```

Use the portion of the given steam table for superheated H_2O at 200 MPa to find the corresponding entropy s for a specific volume v of 0.108 m^3/kg with linear interpolation

v (m^3/kg)	0.10377	0.11144	0.1254
s (kJ/kg·K)	6.4147	6.5453	6.7664

The following data defines the sea-level concentration of dissolved oxygen for fresh water as a function of temperature:

T, °C	0	8	16	24	32	40
o, mg/L	14.621	11.843	9.870	8.418	7.305	6.413

Estimate $o(27)$ using **(a)** linear interpolation, **(b)** Newton's interpolating polynomial, and **(c)** cubic splines. Note that the exact result is 7.986 mg/L..