1. Physical quantities

Pretty much all natural sciences deal with physical quantities. A *quantity* in the general sense is a property ascribed to phenomena, bodies, or substances that can be quantified for, or assigned to, a particular phenomenon, body, or substance [1]. Examples are mass and electric charge. A *physical quantity* is a quantity that can be used in the mathematical equations of science and technology. A *unit* is a particular physical quantity, defined and adopted by convention, with which other particular quantities of the same kind are compared to express their value. The value of a physical quantity is the quantitative expression of a particular physical quantity *as the product of a number and a unit*,

$$\text{physical quantity} = \text{numerical value} \times \text{unit} \qquad (1)$$

the number being its numerical value. Thus, the numerical value of a particular physical quantity depends on the unit in which it is expressed.

For example, the value of the height h_W of the Washington Monument is $h_W = 169$ m = 555 ft. Here h_W is the physical quantity, its value expressed in the unit "meter," unit symbol m, is 169 m, and its numerical value when expressed in meters is 169. However, the value of h_W expressed in the unit "foot," symbol ft, is 555 ft, and its numerical value when expressed in feet is 555.

Units are treated like algebraic quantities, and may be multiplied, divided, and cancelled [2]. Thus, the expression (physical quantity)/unit is simply the numerical value of the measurement in the specified units, and hence is a dimensionless quantity.

The system of units used by scientists and engineers around the world is commonly called "the metric system," but since 1960 it has been known officially as the International System, or SI (the abbreviation for its French name, *Systeme International*) [3]. The SI is founded on seven SI base units for seven base quantities assumed to be mutually independent, as given in Table 1 [1].

Note that physical quantities are denoted by italic or (sloping) Greek letters (as in *m* mass and *t* for time). Units are denoted by Roman letters (as in m for meter) [2].

All other physical quantities, called *derived quantities*, are defined and expressed in terms of combinations of the seven base quantities via a system of quantity equations [1]. The *SI derived units* for these derived quantities are obtained from these equations and the seven SI base units. Examples of such SI derived units are given in Table 2.

A number of derived units have special names and symbols [2]. The names of units derived from names of people are lower case (as in joule, pascal, coulomb, and watt), but their symbols are upper case (as in J, Pa, C, and W). Among the most important for our purposes are those listed in Table 3.

In all cases (both for base and derived quantities), the units may be modified by a prefix that denotes a factor of a power of 10. In a perfect world, Greek prefixes of units are upright (as in μm) and sloping for physical properties (as in μ for chemical potential), but available typefaces are not always so obliging. The 20 SI prefixes used to form decimal multiples and submultiples of SI units are given in Table 4.

Examples of the use of these prefixes are

$$1 \text{ nm} = 10^{-9} \text{ m} \qquad 1 \text{ ps} = 10^{-12} \text{ s} \qquad 1 \text{ μK} = 10^{-6} \text{ K} \qquad (2)$$

Table 1. SI base units and their definitions [1].

Base quantity		SI base unit		
Name	Common symbol	Name	Symbol	Definition
length	l	meter	m	1 m is the length of the path travelled by light in vacuum during a time interval of 1/299 792 458 of a second.
mass	m	kilogram	kg	1 kg is equal to the mass of the international prototype of the kilogram.
time	t	second	s	1 s is the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium 133 atom.
electric current	I	ampere	A	1 A is that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross-section, and placed 1 meter apart in vacuum, would produce between these conductors a force equal to $2 \times 10\text{-}7$ newton per meter of length.
thermodynamic temperature	T	kelvin	K	1 K is the fraction 1/273.16 of the thermodynamic temperature of the triple point of water.
amount of substance	n	mole	mol	1 mol is the amount of substance of a system which contains as many elementary entities as there are atoms in 0.012 kilogram of carbon 12. When the mole is used, the elementary entities must be specified and may be atoms, molecules, ions, electrons, other particles, or specified groups of such particles.
luminous intensity	I_v	candela	cd	1 cd is the luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency 540×10^{12} hertz and that has a radiant intensity in that direction of 1/683 watt per steradian (squared radian)

The kilogram (kg) is anomalous: although it is a base unit, it is interpreted as 10^3 g, and prefixes are attached to the gram (as in 1 mg = 10^{-3} g). Powers of units apply to the prefix as well as the unit they modify [2]:

$$1 \text{ cm}^3 = 1 \text{ (cm)}^3 = 1 \text{ } (10^{-2} \text{ m})^3 = 10^{-6} \text{ m}^3 \tag{3}$$

Note that 1 cm^3 does not mean 1 c(m^3). When carrying out numerical calculations, it is usually safest to write out the numerical value of an observable as powers of 10.

Table 2. Examples of SI derived quantities and units [1].

Derived quantity		SI derived unit	
Name	Common symbol	Name	Symbol
area	A	square meter	m^2
volume	V	cubic meter	m^3
speed, velocity	v	meter per second	$m\ s^{-1}$
acceleration	a	meter per second squared	$m\ s^{-2}$
wave number	$\tilde{v}$	reciprocal meter	m^{-1}
mass density (density)	ρ	kilogram per cubic meter	$kg\ m^{-3}$
specific volume	υ	cubic meter per kilogram	$m^3\ kg^{-1}$
molar concentration (molarity) for substance X	$[X]$	mole per cubic meter	$mol\ m^{-3}$

Table 3. A selection of derived units [2].

Derived quantity		SI derived unit		
Name	Common symbol	Derived unit*	Name of derived unit	Symbol of derived unit
force	F	$1\ kg\ m\ s^{-2}$	newton	N
pressure	p	$1\ kg\ m^{-1}\ s^{-2}$, $1\ N\ m^{-2}$	pascal	Pa
energy	E	$1\ kg\ m^2\ s^{-2}$, $1\ N\ m$, $1\ Pa\ m^{-3}$	joule	J
power	W	$1\ kg\ m^2\ s^{-3}$, $1\ J\ s^{-1}$	watt	W

* Equivalent definitions in terms of derived units are given following the definition in terms of base units.

Table 4. SI prefixes [1,2].

Prefix	y	z	a	f	p	n	μ
Name	yocto	zepto	atto	femto	pico	nano	micro
Factor	10^{-24}	10^{-21}	10^{-18}	10^{-15}	10^{-12}	10^{-9}	10^{-6}
Prefix	m	c	d	da	h	k	M
Name	milli	centi	deci	deka	hecto	kilo	mega
Factor	10^{-3}	10^{-2}	10^{-1}	10^{1}	10^{2}	10^{3}	10^{6}
Prefix	G	T	P	E	Z	Y	
Name	giga	tera	peta	exa	zetta	yotta	
Factor	10^{9}	10^{12}	10^{15}	10^{18}	10^{21}	10^{24}	

There are a number of units that are in wide use but are not a part of the International System. Some are exactly equal to multiples of SI units. These include the liter (L), which is exactly 10^3 cm^3 (or 1 dm^3) and the atmosphere (atm), which is exactly 101.325 kPa. Others rely on the values of fundamental constants, and hence are liable to change when the values of the fundamental constants are modified by more accurate or more precise measurements. Thus, the size of the energy unit electronvolt (eV), the energy acquired by an electron that is accelerated through a potential difference of exactly 1 V, depends on the value of the charge of the electron, and the present (2005) conversion factor is 1 eV = $1.60217733 \times 10^{-19}$ J. Table 5 gives the conversion factors for a number of these convenient units [2].

Table 5. Some common units [2].

Physical quantity	Name of unit	Symbol for unit	Value
time	minute	min	60 s
	hour	h	3600 s
	day	d	85400s
length	ångstrom	Å	10^{-10} m
volume	liter	L, l	1 dm^3
pressure	bar	bar	10^5 Pa
	atmosphere	atm	101.325 Pa
energy	electronvolt	eV	96.48531 kJ mol^{-1}

2. Definitions of some physical quantities [8]

Mass, *m*. The *mass*, *m*, of a sample is a measure of the quantity of matter it contains. Thus, 2 kg of lead contains twice as much matter as 1 kg of lead and indeed twice as much matter as 1 kg of anything. The SI unit of mass is the kilogram (kg), with 1 kg currently defined as the mass of a certain block of platinum-iridium alloy preserved at Sèvres, outside Paris. For typical laboratory-sized samples it is usually more convenient to use a smaller unit and to express mass in grams (g), where 1 kg = 10^3 g.

Volume, *V*. The *volume*, *V*, of a sample is the amount of space it occupies. Thus, we write $V = 00$ cm^3 if the sample occupies 100 cm^3 of space. The units used to express volume (which include cubic meters, m^3; cubic decimeters, dm^3, or liters, L; milliliters, mL.

Displacement [4]**.** Motion involves the displacement of an object from one place in space and time to another. Describing motion requires some convenient coordinate system and a specified origin. A frame of reference is a choice of coordinate axes that defines the starting point for measuring any quantity, an essential first step in solving virtually any problem in mechanics. The *displacement* Δx of an object is defined as its change in position and is given by

$$\Delta x = x_f - x_i \tag{4}$$

where x_i is the coordinate of the initial position of the object and x_f is the coordinate of the objects' final position (he indices *i* and *f* stand for initial and final, respectively). The SI unit of displacement is a meter (m).

Because displacement has both a magnitude (size) and a direction, it's a vector quantity. In general, a vector quantity is characterized by having both a magnitude and a direction. By contrast, a scalar quantity has magnitude, but no direction. Scalar quantities such as mass and temperature are completely specified by a numeric value with appropriate units; no direction is involved.

Velocity and speed [4]. In everyday usage the terms *speed* and *velocity* are interchangeable. In physics, however, there's a clear distinction between them: speed is a scalar quantity, having only magnitude, whereas velocity is a vector, having both magnitude and direction. However, the two quantities share the same SI unit: meter per second (m/s).

The *average speed* of an object over a given time interval is the length of the path it travels divided by the total elapsed time:

$$\text{Average speed} = \frac{\text{path length}}{\text{elapsed time}} \tag{5}$$

Unlike average speed, *average velocity* is a vector quantity, having both a magnitude and a direction. The average velocity $\bar{v}$ during a time interval Δt is the displacement Δx divided by Δt:

$$\bar{v} = \frac{\Delta x}{\Delta t} = \frac{x_f - x_i}{t_f - t_i} \tag{6}$$

While the average speed is always positive, the average velocity of an object in one dimension can be either positive or negative, depending on the sign of the displacement. (The time interval Δt is always positive.)

The *instantaneous velocity v* is the limit of the average velocity as the time interval Δt becomes infinitesimally small:

$$v = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} \tag{7}$$

The notation $\lim_{\Delta t \to 0}$ means that the ratio $\Delta x / \Delta t$ is repeatedly evaluated for smaller and smaller time intervals Δt. As Δt gets extremely close to zero, the ratio $\Delta x / \Delta t$ gets closer and closer to a fixed number, which is defined as the instantaneous velocity.

The *instantaneous speed* of an object, which is a scalar quantity, is defined as the magnitude of the instantaneous velocity. Like average speed, instantaneous speed (which we will usually call, simply, "speed") has no direction associated with it and hence carries no algebraic sign.

Acceleration, *a* [4]. The changing of an object's velocity with time is called *acceleration*. The average acceleration $\bar{a}$ during the time interval Δt is the change in velocity Δv divided by Δt:

$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{v_f - v_i}{t_f - t_i} \tag{8}$$

Acceleration is a vector quantity having dimensions of length divided by the time squared. Acceleration has dimensions of meters per second per second ((m/s)/s, which is usually written as m s^{-2}). Acceleration is a vector quantity. For the case of motion in a straight line, the direction of the velocity of an object and the direction of its acceleration are related as follows: When the object's velocity and acceleration are in the same direction, the speed of the object increases with time. When the object's velocity and acceleration are in opposite directions, the speed of the object decreases with time.

The value of the average acceleration often differs in different time intervals, so it's useful to define the instantaneous acceleration, which is analogous to the instantaneous velocity discussed above. The instantaneous acceleration a is the limit of the average acceleration as the time interval Δt goes to zero:

$$a = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t} \tag{9}$$

Here again, the notation $\lim_{\Delta t \to 0}$ means that the ratio $\Delta v / \Delta t$ is evaluated for smaller and smaller values of Δt. The closer Δt gets to zero, the closer the ratio gets to a fixed number, which is the instantaneous acceleration.

Force, F. One of the most basic concepts of physical science is that of *force*. In classical mechanics, the mechanics originally formulated by Isaac Newton at the end of the seventeenth century, a body of mass m travels in a straight line at constant speed until a force acts on it. Then it undergoes an acceleration, a rate of change of velocity, given by Newton's second law of motion:

$$Force = mass \times acceleration \tag{10}$$
$$F = ma \tag{11}$$

The acceleration of a freely falling body at the surface of the Earth is 9.81 m s^{-2}, so the gravitational force acting on a mass of 1.0 kg is

$$F = (1.0 \text{ kg}) \times (9.81 \text{ m s}^{-2}) = 9.8 \text{ kg m s}^{-2} = 9.8 \text{ N}$$

The derived unit of force is the newton, N:

$$1 \text{ N} = 1 \text{ kg m s}^{-2} \tag{12}$$

Therefore, we can report the force we have just calculated as 9.8 N. It might be helpful to note that a force of 1 N is approximately the gravitational force exerted on a small apple (of mass 100 g). Force is a directed quantity, in the sense that it has direction as well as magnitude. For a body on the surface of the Earth, the force of gravitational attraction is directed toward the center of the Earth.

Work, W. When an object is moved through a distance against an opposing force, we say that *work* is done. The magnitude of the work is the product of the distance moved and the opposing force:

$$Work = force \times distance \tag{13}$$

Therefore, to raise a body of mass 1.0 kg on the surface of the Earth through a vertical distance of 1.0 m requires us to expend the following amount of work:

$$Work = (9.8 \text{ N}) \times (1.0 \text{ m}) = 9.8 \text{ N m} \tag{14}$$

As we shall see more formally in a moment, the unit 1 N m (or, in terms of base units, 1 kg m^2 s^{-2}) is called 1 joule (1 J). So, 9.8 J is needed to raise a mass of 1.0 kg through 1.0 m on the surface of the Earth.

Energy, E. A property that is central in physical science is, of course, the *energy*, E. Everyone uses the term "energy" in everyday language, but in science it has a precise meaning - energy is the capacity to do work. A fully wound spring can do more work than a half-wound spring (that is, it can raise a weight through a greater height or move a greater weight through a given height). A hot object has the potential for doing more work than the same object when it is cool and therefore has a higher energy. The SI unit of energy is the joule (J), named after the nineteenth-century scientist James Joule, who helped to establish the concept of energy. It is defined as

$$1 \text{ J} = 1 \text{ kg m}^2 \text{ s}^{-2} \tag{15}$$

A joule is quite a small unit, and in chemistry we often deal with energies of the order of kilojoules (1 kJ = 10^3 J).

Pressure, p. *Pressure*, p, is force, F, divided by the area, A, on which the force is exerted:

$$\text{Pressure} = \text{force} / \text{area} \qquad (16)$$
$$p = F / A \qquad (17)$$

Pressure can arise in many different ways. For example, from the gravitational pull of the Earth on an object. The impact of gas molecules on a surface gives rise to a force and hence to a pressure. If an object is immersed in the gas, it experiences a pressure over its entire surface because molecules collide with it from all directions. In this way, the atmosphere exerts a pressure on all the objects in it. We are incessantly battered by molecules of gas in the atmosphere and experience this battering as the "atmospheric pressure." The pressure is greatest at sea level because the density of air, and hence the number of colliding molecules, is greatest there.
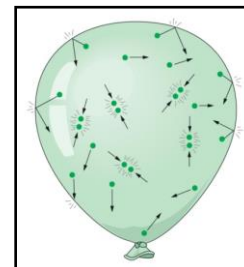
The atmospheric pressure is very considerable: it is the same as would be exerted by loading 1 kg of lead (or any other material) onto a surface of area 1 cm^2. We go through our lives under this heavy burden pressing on every square centimeter of our bodies. Some deep-sea creatures are built to withstand even greater pressures: at 1000 m below sea level the pressure is 100 times greater than at the surface. Creatures and submarines that operate at these depths must withstand the equivalent of 100 kg of lead loaded onto each square centimeter of their surfaces. The pressure of the air in our lungs helps us withstand the relatively low but still substantial pressures that we experience close to sea level.
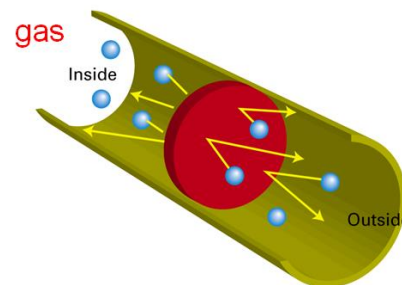
When a gas is confined to a cylinder fitted with a movable piston, the position of the piston adjusts until the pressure of the gas inside the cylinder is equal to that exerted by the atmosphere. When the pressures on either side of the piston are the same, we say that the two regions on either side are in *mechanical equilibrium*. The pressure of the confined gas arises from the impact of the particles: they batter the inside surface of the piston and counter the battering of the molecules in the atmosphere that is pressing on the outside surface of the piston. Provided the piston is weightless (that is, provided we can neglect any gravitational pull on it), the gas is in mechanical equilibrium with the atmosphere whatever the orientation of the piston and cylinder, because the external battering is the same in all directions.

The SI unit of pressure is the pascal, Pa:

$$1 \text{ Pa} = 1 \text{ kg m}^{-1} \text{ s}^{-2} \qquad (18)$$

The pressure of the atmosphere at sea level is about 10^5 Pa (100 kPa). This fact lets us imagine the magnitude of 1 Pa, for we have just seen that 1 kg of lead resting on 1 cm^2 on the surface of the Earth exerts about the same pressure as the atmosphere; so $1/10^5$ of that mass, or 0.01 g, will exert about 1 Pa, we see that the pascal is rather a small unit of pressure. Table 6 lists the other units commonly used to report pressure.[1] One of the most important in modern physical chemistry is the bar, where 1 bar = 10^5 Pa exactly. Normal atmospheric pressure is close to 1 bar.

Table 6. Pressure units and conversion factors [1].

Unit	Conversion factor*
pascal, Pa	$1 \text{ Pa} = 1 \text{ N m}^{-2}$
bar	$1 \text{ bar} = 10^5 \text{ Pa}$
atmosphere, atm	$1 \text{ atm} = \mathbf{101.325} \text{ kPa} = \mathbf{1.01325} \text{ bar}$
torr, Torr	$\mathbf{760} \text{ Torr} = 1 \text{ atm}$ $1 \text{ Torr} = 133.32 \text{ PA}$

* values in bold are exact

Mass density (density), ρ. The density, or more precisely, the volumetric mass density, of a substance is its mass per unit volume [5]. The symbol most often used for density is ρ (the lower case Greek letter rho). Mathematically, density is defined as mass divided by volume:[1]

$$\rho = m/V \tag{19}$$

where ρ is the density, m is the mass, and V is the volume. With the mass measured in kilograms and the volume in meters cubed, density is reported in kilograms per cubic meter (kg m^{-3}); however, it is equally acceptable and often more convenient to report mass density in grams per cubic centimeter (g cm^{-3}) or grams per milliliter (g mL^{-1}) [8]. The relation between these units is

$$1 \text{ g cm}^{-3} = 1 \text{ g mL}^{-1} = 10^3 \text{ kg m}^{-3} \tag{20}$$

Temperature, *T*. In everyday terms, the temperature is an indication of how "hot" or "cold" a body is. In science, temperature, *T*, is the property of an object that determines in which direction energy will flow when it is in contact with another object: energy flows from higher temperature to lower temperature. When the two bodies have the same temperature, there is no net flow of energy between them. In that case we say that the bodies are in thermal equilibrium. Temperature in science is measured on either the Celsius scale or the Kelvin scale. On the Celsius scale, in which the temperature is expressed in degrees Celsius (°C), the freezing point of water at 1 atm corresponds to 0°C and the boiling point at 1 atm corresponds to 100°C. This scale is in widespread everyday use. However, it turns out to be much more convenient in many scientific applications to adopt the Kelvin scale and to express the temperature in kelvin (K; note that the degree sign is not used for this unit). Whenever we use *T* to denote a temperature, we mean a temperature on the Kelvin scale. The Celsius and Kelvin scales are related by

$$T \text{ (in K)} = \text{temperature (in °C)} + 273.15 \tag{21}$$

Thus, water at 1 atm freezes at 273 K and boils at 373 K; a warm day (25°C) corresponds to 298.15 K.

The *absolute zero of temperature* is the temperature below which it is impossible to cool an object. The Kelvin scale ascribes the value $T = 0$ to this absolute zero of temperature. Note that we refer to absolute zero as $T = 0$, not $T = 0$ K. There are other "absolute" scales of temperature, all of which set their lowest value at zero. Insofar as it is possible, all expressions in science should be independent of the units being employed, and in this case the lowest attainable temperature is $T = 0$ regardless of the absolute scale we are using.
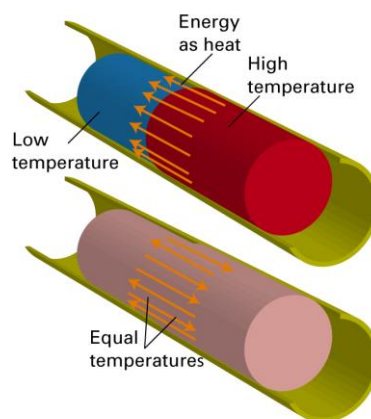
Amount of substance (***chemical amount, the number of moles***), ***n***. The mole (abbreviated mol) is the SI unit for *amount of substance* [6]. It is defined as the amount of a substance that contains the same number of entities as there are atoms in exactly 12 g (0.012 kg) of carbon-12. This number is called Avogadro's number (N_A), in honor of the 19th-century Italian physicist Amedeo Avogadro,

$$N_A = 6.022 \times 10^{23} \text{ mol}^{-1} \tag{22}$$

and as you can see, it is enormous. It follows that one mole (1 mol) of particles contains N_A (6.022×10^{23}) entities. Thus,

1 mol of carbon- 12 contains 6.022×10^{23} carbon-12 atoms

1 mol of H_2O contains 6.022×10^{23} H_2O molecules

1 mol of NaCI contains 6.022×10^{23} NaCl formula units

However, the mole is not just a counting unit like the dozen, which specifies only the number of objects. The definition of the mole specifies the number of objects in a fixed mass of substance. Therefore, 1 mole of a substance represents a fixed number of chemical entities and has a fixed mass.

Molar mass, M. The *molar mass* (*M*) of a substance is the mass per mole of its entities (atoms, molecules, or formula units). Thus, molar mass has units of grams per mole (g mol^{-1}). The periodic table is indispensable for calculating the molar mass of a substance. Here's how the calculations are done:

a). *Elements*. You find the molar mass of an element simply by looking up its atomic mass in the periodic table and then noting whether the element occurs naturally as individual atoms or as molecules.

- *Monatomic elements*. For elements that occur as individual atoms, the molar mass is the numerical value from the periodic table expressed in units of grams per mole. Thus, M(neon, Ne) = 20.18 g mol^{-1}, M(iron, Fe) is 55.85 g mol^{-1}, and M(gold, Au) is 197.0 g mol^{-1}.

- *Molecular elements*. For elements that occur as molecules, you must know the molecular formula to determine the molar mass. For example, oxygen exists normally in air as diatomic molecules, so the molar mass of O_2 molecules is twice that of O atoms:

$$M(O_2) = 2 \times M(O) = 2 \times 16.00 \text{ g/mol} = 32.00 \text{ g/mol} \tag{23}$$

The most common form of sulfur, S, exists as octatomic molecules, S_8:

$$M(S8 = 8 \times M(S) = 8 \times 32.07 \text{ g/mol} = 256.6 \text{ g/mol} \tag{24}$$

b). *Compounds*. The molar mass of a compound is the sum of the molar masses of the atoms of the elements in the formula. For example, the formula of sulfur dioxide (SO_2) tells us that 1 mol of SO_2 molecules contains 1 mol of S atoms and 2 mol of O atoms:

$$M(SO_2) = M(S) + 2 \times M(O) = 32.07 \text{ g/mol} + (2 \times 16.00 \text{ g/mol}) = 64.07 \text{ g/mol} \tag{25}$$

Similarly, for ionic compounds, such as potassium sulfide (K_2S), we have

$$M(K_2S) = 2 \times M(K) + M(S) = 2 \times 39. 10 \text{ g/mol} + 32.07 \text{ g/mol} = 110.27 \text{ g/mol} \tag{26}$$

A key point to note is that the subscripts in a formula refer to individual atoms (or ions), as well as to moles of atoms (or ions).

There is a very important relationship between the mass of a substance (*m*) and its molar mass (*M*). These two quantities are related by the amount of substance *n* (also called the chemical amount, see discussion above). The molar mass is equal to the mass of a sample divided by chemical amount of atoms, molecules, or formula units it contains:

$$M = m / n \tag{27}$$

Molar concentration (molarity), [J]. The *molar concentration* ('*molarity*', [J]) of a solute J in a solution is the amount of substance of the solute n_J divided by the volume of the solution (V).

$$[J] = n_J / V \tag{28}$$

Molar concentration is usually expressed in moles per decimetre cubed (mol dm^{-3} or mol L^{-1}; 1 dm^3 is identical to 1 L). A solution in which the molar concentration of the solute is 1 mol dm^{-3} is prepared by dissolving 1 mol of the solute in sufficient solvent to prepare 1 dm^3 of solution. Such a solution is widely called a '1 molar' solution and denoted 1 m.

Molality, b_J. Another way to express concentration of a solution is by using *molality*. The term molality refers to the amount of substance of solute J (n_J) divided by the mass of solvent used to prepare the solution, $m_{solvent}$. Its units are typically moles of solute per kilogram of solvent (mol kg^{-1}).

$$b_J = n_J / m_{solvent} \tag{29}$$

For dilute solutions of H_2O (*aqueous solutions*), [J] and b_J are very close because 1 L of water is very close to 1 kg ($\rho_{water} = 1$ kg L^{-1}). For concentrated aqueous solutions and non-aqueous solutions, $\rho \neq 1$ kg L^{-1}, so [J] $\neq b_J$.

3. Extensive and intensive physical properties

An *extensive property* is a property that depends on the amount of substance in the sample. Two examples are mass and volume. An *intensive property* is a property that is independent of the amount of substance in the sample. Examples are temperature, mass density (mass divided by volume), and pressure.

A molar property, X_m, is the value of an extensive property, X, of the sample divided by the amount of substance present in the sample:

$$X_m = X / n \tag{30}$$

A molar property is intensive. An example is the molar volume, V_m, the volume of a sample divided by the amount of substance in the sample (the volume per mole).

$$V_m = V / n \tag{31}$$

The one exception to the notation X_m is the molar mass, which is denoted M (see discussion above).

4. Fundamental physical constants [7]

A physical constant is a physical quantity that is generally believed to be both universal in nature and constant in time. It can be contrasted with a mathematical constant, which is a fixed numerical value, but does not directly involve any physical measurement.

There are many physical constants in science, some of the most widely recognized being the speed of light in vacuum c, the gravitational constant G, Planck's constant h, the electric constant ε_0, and the elementary charge e. A comprehensive list of fundamental physical constants is given in Table 7.

Table 7. Fundamental physical constants

Quantity	Symbol	Value	Power of ten	Units
Speed of light	c	2.997 925 58*	10^8	m s^{-1}
Elementary charge	e	1.602 176	10^{-19}	C
Boltzmann's constant	k	1.380 65	10^{-23}	J K^{-1}
Planck constant	h	6.626 08	10^{-34}	J s
	$\hbar = h/2\pi$	1.054 57	10^{-34}	J s
Avogadro's constant	N_A	6.022 14	10^{23}	mol^{-1}
Atomic mass constant	m_u	1.660 54	10^{-27}	kg
Mass				
electron	m_e	9.109 38	10^{-31}	kg
proton	m_p	1.672 62	10^{-27}	kg
neutron	m_n	1.674 93	10^{-27}	kg
Vacuum permittivity	$\varepsilon_0 = 1/c^2\mu_0$	8.854 19	10^{-12}	J^{-1} C^2 m^{-1}
	$4\pi\varepsilon_0$	1.112 65	10^{-10}	J^{-1} C^2 m^{-1}
Vacuum permeability	μ_0	4π	10^{-7}	J s^2 C^{-2} m^{-1} (= T^2 J^{-1} m^3)
Magneton				
Bohr	$\mu_B = e\hbar/2m_e$	9.274 01	10^{-24}	J T^{-1}
nuclear	$\mu_N = e\hbar/2m_p$	5.050 78	10^{-27}	J T^{-1}
g value of the electron	g_e	2.002 32		
Bohr radius	$a_0 = 4\pi\varepsilon_0\hbar^2/m_e e^2$	5.291 77	10^{-11}	m
Rydberg constant	$R = m_e e^4/8h^3 c\varepsilon_0^2$	1.097 37	10^5	cm^{-1}
Standard acceleration of free fall	g	9.806 65*		m s^{-2}

   * Exact value

5. The states of matter [8]

The broadest classification of matter is into one of three states of matter, or forms of bulk matter, namely gas, liquid, and solid. We distinguish the three states of matter by noting the behavior of a substance enclosed in a rigid container:

A *gas* is a fluid form of matter that fills the container it occupies.

A *liquid* is a fluid form of matter that possesses a well-defined surface and (in a gravitational field) fills the lower part of the container it occupies.

A *solid* retains its shape regardless of the shape of the container it occupies.

One of the roles of physical chemistry is to establish the link between the properties of bulk matter and the behavior of the particles - atoms, ions, or molecules - of which it is composed. We have the following models for the states of matter:

A gas is composed of widely separated particles in continuous rapid, disordered motion. A particle travels several (often many) diameters before colliding with another particle. For most of the time the particles are so far apart that they interact with each other only very weakly.

A liquid consists of particles that are in contact but are able to move past one another in a restricted manner. The particles are in a continuous state of motion but travel only a fraction of a diameter before bumping into a neighbor. The overriding image is one of movement but with molecules jostling one another.

A solid consists of particles that are in contact and unable to move past one another. Although the particles oscillate around an average location, they are essentially trapped in their initial positions and typically lie in ordered arrays.

The main difference between the three states of matter is the freedom of the particles to move past one another. If the average separation of the particles is large, there is hardly any restriction on their motion, and the substance is a gas. If the particles interact so strongly with one another that they are locked together rigidly, then the substance is a solid. If the particles have an intermediate mobility between these extremes, then the substance is a liquid. We can understand the melting of a solid and the vaporization of a liquid in terms of the progressive increase in the liberty of the particles as a sample is heated and the particles become able to move more freely.

6. Equations of state [9]

We can specify the state of any sample of substance by giving the values of the following properties (all of which are defined above):
- V, the volume of the sample
- p, the pressure of the sample
- T, the temperature of the sample
- n, the amount of substance in the sample

However, an astonishing experimental fact is that these four quantities are not independent of one another. For instance, we cannot arbitrarily choose to have a sample of 0.555 mol H_2O in a volume of 100 cm^3 at 100 kPa and 500 K: it is found experimentally that that state simply does not exist. If we select the amount, the volume, and the temperature, then we find that we have to accept a particular pressure (in this case, close to 23 MPa). The same is true of all substances, but the pressure in general will be different for each one. This experimental generalization is summarized by saying the substance obeys an *equation of state*, an equation of the form

$$p = f(n,V,T) \tag{32}$$

This expression tells us that the pressure is some function of amount, volume, and temperature and that if we know those three variables, then the pressure can have only one value. The equations of state of most substances are not known, so in general we cannot write down an explicit expression for the pressure in terms of the other variables. However, certain equations of state are known. In particular, the equation of state of a low-pressure gas is known, and proves to be very simple and very useful. This equation is used to describe the behavior of gases taking part in reactions, the behavior of the atmosphere, as a starting point for problems in chemical engineering, and even in the description of the structures of stars.

7. The perfect gas equation of state [2,9]

The equation of state of a low-pressure gas was among the first results to be established in physical chemistry. The original experiments were carried out by Robert Boyle in the seventeenth century and there was a resurgence in interest later in the century when people began to fly in balloons. This technological progress demanded more knowledge about the response of gases to changes of pressure and temperature and, like technological advances in other fields today, that interest stimulated a lot of experiments.

The experiments of Boyle and his successors led to the formulation of the following perfect gas equation of state:

$$pV = nRT \tag{33}$$

This equation has the form of equation (32) when rearranged into

$$p = \frac{nRT}{V} \tag{34}$$

The *gas constant*, R, is an experimentally determined quantity that turns out to have the same value for all gases. It can be determined by evaluating

$$R = pV/nRT \tag{35}$$

for a gas in the limit of zero pressure (to guarantee that it is behaving perfectly). However, a more accurate value can be obtained by measuring the speed of sound in a low-pressure gas (argon is used in practice) and extrapolating its value to zero pressure. Values of R in different units are given in Table 8.

Table 8. The value of the gas constant R in various units [2]

8.314 47	$J\,K^{-1}\,mol^{-1}$
$8.205\ 74 \times 10^{-2}$	$dm^3\,atm\,K^{-1}\,mol^{-1}$
$8.314\ 47 \times 10^{-2}$	$dm^3\,bar\,K^{-1}\,mol^{-1}$
8.314 47	$Pa\,m^3\,K^{-1}\,mol^{-1}$
1 62.364	$dm^3\,Torr\,K^{-1}\,mol^{-1}$
1.987 21	$cal\,K^{-1}\,mol^{-1}$

The *perfect gas equation of state* - more briefly, the '*perfect gas law*' - is so-called because it is an idealization of the equations of state that gases actually obey. Specifically, it is found that all gases obey the equation ever more closely as $p \to 0$. That is, equation (33) is an example of a *limiting law*, a law that becomes increasingly valid as the pressure is reduced and is obeyed exactly in the limit of zero pressure.

A hypothetical substance that obeys equation (33) at all pressures is called a *perfect gas*. From what has just been said, an actual gas, which is termed a real gas, behaves more and more like a perfect gas at $p \to 0$. In practice, normal atmospheric pressure at sea level ($p \approx 100$ kPa) is already low enough for most real gases to behave almost perfectly, and unless stated otherwise we shall always assume in this text that the gases we encounter behave like a perfect gas. The reason why a real gas behaves differently from a perfect gas can be traced to the attractions and repulsions that exist between actual molecules and that are absent in a perfect gas.

Note that a perfect gas is widely called an 'ideal gas' and the perfect gas equation of state is commonly called 'the ideal gas equation'.

The equation of state of a gas at low pressure was established by combining a series of empirical laws:

$$\textbf{Boyle's law}: pV = \text{constant, at constant } n, T \tag{36}$$
$$\textbf{Charles's law}: V = \text{constant} \times T, \text{ at constant } n, p \tag{37}$$
$$p = \text{constant} \times T, \text{ at constant } n, V \tag{38}$$
$$\textbf{Avogadro's principle: } V = \text{constant} \times n \text{ at constant } p, T \tag{39}$$

Boyle's and Charles's laws are examples of a limiting law strictly true only when $p \to 0$. Avogadro's principle is commonly expressed in the form 'equal volumes of gases at the same temperature and pressure contain the same numbers of molecules'. In this form, it is increasingly true as $p \to 0$. Although these relations are strictly true only at $p = 0$, they are reasonably reliable at normal pressures ($p \approx 1$ bar) and are used widely throughout chemistry.

Figure 1a depicts the variation of the pressure of a sample of gas as the volume is changed. Each of the curves in the graph corresponds to a single temperature and hence is called an *isotherm*. According to Boyle's law, the isotherms of gases are hyperbolas. An alternative depiction, a plot of pressure against $1/V$, is shown in Fig. 1b. The linear variation of V with T summarized by Charles's law is illustrated in Fig. 1c. The lines in this illustration are examples of *isobars*, or lines showing the variation of properties at constant p. Figure 1d illustrates the linear variation of p with T. The lines in this diagram are *isochores*, or lines showing the variation of properties at constant V.

a)   **b)**   **c)**   **d)**
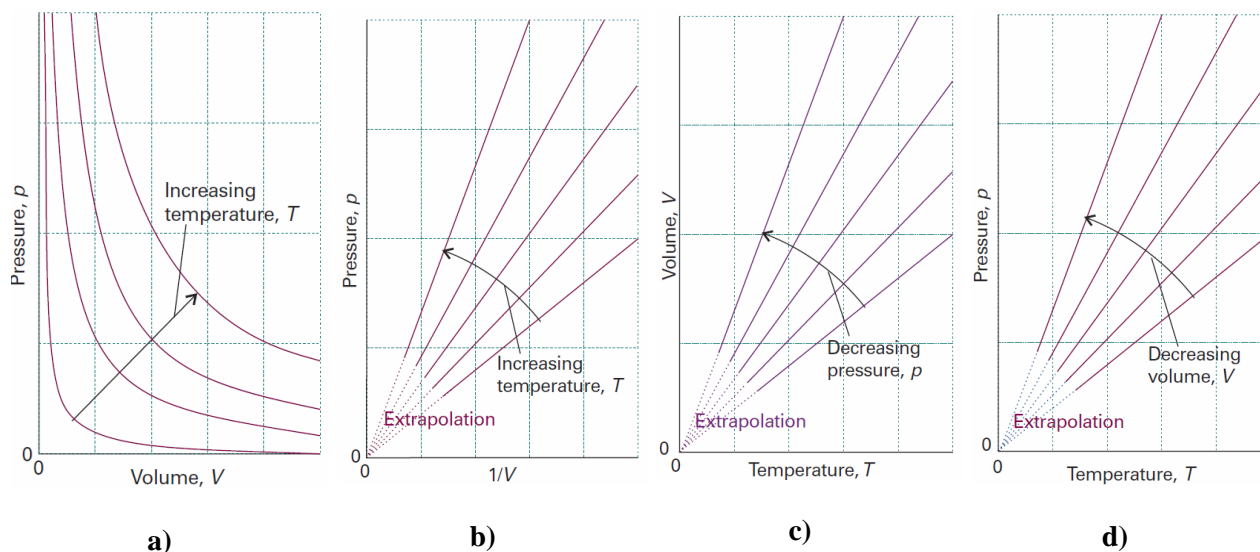
Figure 1. (a) The p–V dependence of a fixed amount of perfect gas at different T. Each curve is a hyperbola ($pV = $ const.) and is called an *isotherm*. **(b)** Straight lines are obtained when p pressure is plotted against $1/V$ at T=const. **(c)** The variation of V of a fixed amount of gas with T at p=const. Note that in each case the *isobars* extrapolate to zero volume at $T = 0$ (-273°C). **(d)** The pressure also varies linearly with the T at V=const. constant volume, and extrapolates to zero at $T = 0$ (-273°C). [2]

If we combine the perfect gas law (eq.33) and the expression for molar volume, $V_m = V/n$ (eq. 31), we get

$$V_m = \frac{V}{n} = \frac{nRT}{np} = \frac{RT}{p} \tag{40}$$

According to Avogadro's principle, V_m should be the same for all gases at the same p and T (provided they behave ideally). Table 9 lists V_m for several gases under ambient pressure and temperature (T=298.15 K, $p = 1$ bar).

Table 9. The molar volumes of gases at T=298.15 K, p = 1 bar [9]

Gas	V_m/(dm³ mol⁻¹)
Perfect gas	24.7896*
Ammonia	24.8
Argon	24.4
Carbon dioxide	24.6
Nitrogen	24.8
Oxygen	24.8
Hydrogen	24.8
Helium	24.8

This shows that many gases indeed behave ideally under ambient conditions.

8. Real gases [2,9]

Real gases do not obey the perfect gas law exactly. Deviations from the law are particularly important at high pressures and low temperatures, especially when a gas is on the point of condensing to liquid.

Real gases show deviations from the perfect gas law because molecules interact with one another. Repulsive forces between molecules assist expansion and attractive forces assist compression. Repulsive forces are significant only when molecules are almost in contact: they are short-range interactions, even on a scale measured in molecular diameters (Fig. 2). Because they are short-range interactions, repulsions can be expected to be important only when the average separation of the molecules is small. This is the case at high pressure, when many molecules occupy a small volume. On the other hand, attractive intermolecular forces have a relatively long range and are effective over several molecular diameters. They are important when the molecules are fairly close together but not necessarily touching (at the intermediate separations in Fig. 2). Attractive forces are ineffective when the molecules are far apart (well to the right in Fig. 2). Intermolecular forces are also important when the temperature is so low that the molecules travel with such low mean speeds that they can be captured by one another. At low pressures, when the sample occupies a large volume, the molecules are so far apart for most of the time that the intermolecular forces play no significant role, and the gas behaves virtually perfectly.

Figure 2. The variation of the potential energy of two molecules on their separation. High positive potential energy (at very small separations) indicates that the interactions between them are strongly repulsive at these distances. At intermediate separations, where the potential energy is negative, the attractive interactions dominate. At large separations (on the right) the potential energy is zero and there is no interaction between the molecules.

At moderate pressures, when the average separation of the molecules is only a few molecular diameters, the attractive forces dominate the repulsive forces. In this case, the gas can be expected to be more compressible than a perfect gas because the forces help to draw the molecules together. At high pressures, when the average separation of the molecules is small, the

repulsive forces dominate and the gas can be expected to be less compressible because now the forces help to drive the molecules apart.

The compression factor

The compression factor, Z, of a gas is the ratio of its measured molar volume, $V_m = V / n$, to the molar volume of a perfect gas, V_m^o, at the same pressure and temperature.

$$Z = \frac{V_m}{V_m^o} \tag{41}$$

Because the molar volume of a perfect gas is equal to RT/p, an equivalent expression is

$$Z = \frac{pV_m}{RT}, \tag{42}$$

which we can write as

$$pV_m = RTZ \tag{43}$$

Because for a perfect gas $Z = 1$ under all conditions, deviation of Z from 1 is a measure of departure from perfect behavior.

Some experimental values of Z are plotted in Fig. 3. At very low pressures, all the gases shown have $Z \approx 1$ and behave nearly perfectly. At high pressures, all the gases have $Z > 1$, signifying that they have a larger molar volume than a perfect gas. Repulsive forces are now dominant. At intermediate pressures, most gases have $Z < 1$, indicating that the attractive forces are reducing the molar volume relative to that of a perfect gas.
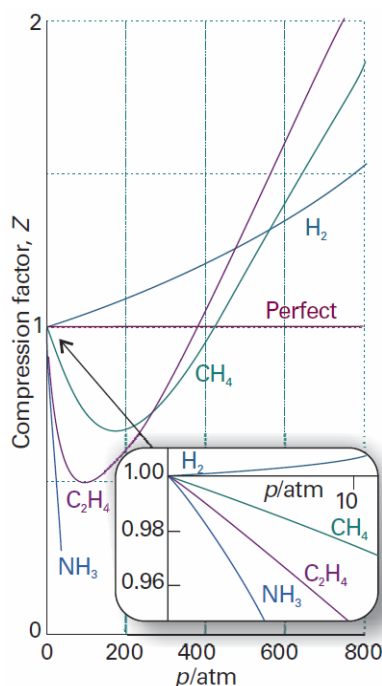
Figure 3. The variation of Z with p for several gases at 0°C. A perfect gas has $Z = 1$ at all pressures. Notice that, although the curves approach 1 as p → 0, they do so with different slopes.

The virial equation of state

At large molar volumes and high temperatures the real-gas isotherms do not differ greatly from perfect-gas isotherms. The small differences suggest that the perfect gas law is in fact the first term in an expression of the form

$$pV_{\mathrm{m}} = RT\left(1 + B'p + C'p^2 + D'p^3 + \ldots\right) \tag{44}$$

This expression is an example of a common procedure in physical chemistry, in which a simple law that is known to be a good first approximation (in this case $pV = nRT$) is treated as the first term in a series in powers of a variable (in this case p). A more convenient expansion for many applications is

$$pV_{\mathrm{m}} = RT\left(1 + \frac{B}{V_{\mathrm{m}}} + \frac{C}{V_{\mathrm{m}}^2} + \frac{D}{V_{\mathrm{m}}^3} + \ldots\right) \tag{45}$$

These two expressions are two versions of the *virial equation of state*. By comparing the expression with equation (43) we see that the term in parentheses can be identified with the compression factor, Z.

$$Z = 1 + \frac{B}{V_{\mathrm{m}}} + \frac{C}{V_{\mathrm{m}}^2} + \frac{D}{V_{\mathrm{m}}^3} \ldots \tag{46}$$

The coefficients B, C, D, ... , which depend on the temperature, are the *second*, *third*, *fourth* etc. *virial coefficients* (Table 1.4); the first virial coefficient is 1. The third virial coefficient, C, is usually less important than the second coefficient, B, in the sense that at typical molar volumes $C/V_{\mathrm{m}}^2 \ll B/V_{\mathrm{m}}$. When the molar volume is very large, the terms B/V_{m} and C/V_{m}^2 are both very small, and only the 1 inside the parentheses survives. In this limit, the equation of state approaches that of a perfect gas.

The van der Waals equation of state

Although it is the most reliable equation of state, the virial equation does not give us much immediate insight into the behavior of gases and their condensation to liquids. The van der Waals equation, which was proposed in 1873 by the Dutch physicist Johannes van der Waals, is only an approximate equation of state but it has the advantage of showing how the intermolecular interactions contribute to the deviations of a gas from the perfect gas law. We can view the van der Waals equation as another example of taking a soundly based qualitative idea and building up a mathematical expression that can be tested quantitatively.

The repulsive interactions between molecules are taken into account by supposing that they cause the molecules to behave as small but impenetrable spheres. The nonzero volume of the molecules implies that instead of moving in a volume V they are restricted to a smaller volume $V-nb$, where nb is approximately the total volume taken up by the molecules themselves. This argument suggests that the perfect gas law $p = nRT/V$ should be replaced by

$$p = \frac{nRT}{V - nb} \tag{47}$$

when repulsions are significant. The closest distance of two hard-sphere molecules of radius r, and volume $V_{\mathrm{molecule}} = \frac{4}{3}\pi r^3$, is $2r$, so the volume excluded is $\frac{4}{3}\pi(2r)^3$, or $8V_{\mathrm{molecule}}$. The volume excluded per molecule is one-half this volume, or $4V_{\mathrm{molecule}}$, so $b \approx 4V_{\mathrm{molecule}}N_{\mathrm{A}}$.

Excluded volume

The pressure depends on both the frequency of collisions with the walls and the force of each collision. Both the frequency of the collisions and their force are reduced by the attractive forces, which act with a strength proportional to the molar concentration, n/V, of molecules in the sample. Therefore, because both the frequency and the force of the collisions are reduced by the attractive forces, the pressure is reduced in proportion to the square of this concentration. If the

17
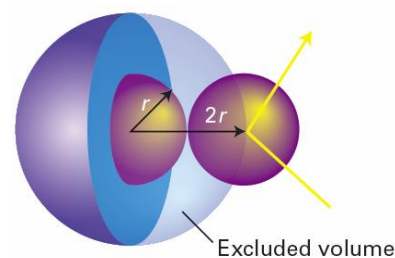
reduction of pressure is written as $-a(n/V)^2$, where a is a positive constant characteristic of each gas, the combined effect of the repulsive and attractive forces is the *van der Waals equation of state*:

$$p = \frac{nRT}{V - nb} - a\left(\frac{n}{V}\right)^2 \tag{48}$$

or in terms of molar volume

$$p = \frac{RT}{V_m - b} - \frac{a}{V_m^2} \tag{49}$$

where constants a and b are called the **van der Waals coefficients**.

This equation can be derived in other ways, but the present method has the advantage that it shows how to derive the form of an equation out of general ideas. The derivation also has the advantage of keeping imprecise the significance of the coefficients a and b: they are much better regarded as empirical parameters than as precisely defined molecular properties.

To show the resemblance of this equation to the perfect gas equation $pV = nRT$, equation (48) is sometimes rearranged as

$$\left(p + \frac{an^2}{V^2}\right)(V - nb) = nRT \tag{50}$$

Note that the at high temperature and low pressures, equation (50) reduces to the perfect gas law $pV = nRT$ because $V \gg nb$ and $V \gg an^2$.

Some additional equations of state

Note that the virial and van der Waals equations of state are not the only two that exist. Table 11 includes several other equations of state, but many more are available.

Table 11. Selected equations of state. Note that the letters a and b stand for constant parameters that have different values for different substances. These parameters do not necessarily have the values for the same substance in different equations of state. [2,10]

perfect gas	$p = \dfrac{RT}{V_m}$
van der Waals	$p = \dfrac{RT}{V_m - b} - \dfrac{a}{V_m^2}$
virial	$p = \dfrac{RT}{V_m}\left(1 + \dfrac{B}{V_m} + \dfrac{C}{V_m^2}\right)$
Berthelot	$p = \dfrac{RT}{V_m - b} - \dfrac{a}{T V_m^2}$
Redlich-Kwong	$p = \dfrac{RT}{V_m - b} - \dfrac{a}{T^{1/2} V_m (V_m + b)}$
Dieterici	$p = \dfrac{RT e^{-a/RTV_m}}{V_m - b}$

Note that the parameters for a given gas do not necessarily have the same values (and even units!) in different equations even if the same letters are used [10]. The accuracy of several of the equations of state has been evaluated [11,12]. The Redlich–Kwong equation of state seemed to perform better than the other two-parameter equations, with the van derWaals equation coming in second best.

Parameters for selected equations of state are listed in Table 12.

Table 12. Parameters for selected equations of state [10].

Parameters for the van der Waals equation of state			Parameters for the Berthelot equation of state		
Substance	a/Pa m^6 mol^{-2}	$b \times 10^5$/m^3 mol^{-1}	Substance	a/Pa m^6 mol^{-2} K	$b \times 10^5$/m^3 mol^{-1}
Ammonia	0.4225	3.707	Ammonia	171	3.70
Argon	0.1363	3.219	Argon	20.5	3.20
Carbon dioxide	0.3640	4.267	Carbon dioxide	111	4.28
Helium	0.003457	2.370	Helium	0.019	2.41
Hydrogen	0.02476	2.661	Methane	43.6	4.27
Methane	0.2283	4.278	Neon	0.98	1.77
Neon	0.8636	1.709	Nitrogen	17.3	3.87
Nitrogen	0.1408	3.913	Oxygen	21.3	3.18
Oxygen	0.1378	3.183	Xenon	121	5.13
Xenon	0.4250	5.105	Water	357	3.04
Water	0.5536	3.049			

Parameters for the Dieterici equation of state			Parameters for the Redlich-Kwong equation of state		
Substance	a/Pa m^6 mol^{-2}	$b \times 10^5$/m^3 mol^{-1}	Substance	a/Pa m^6 mol^{-2} K$^{1/2}$	$b \times 10^5$/m^3 mol^{-1}
Ammonia	0.540	4.00	Ammonia	8.59	2.56
Argon	0.174	3.47	Argon	1.69	2.22
Carbon dioxide	0.468	4.63	Carbon dioxide	6.44	2.96
Helium	0.0046	2.60	Helium	0.00835	1.67
Hydrogen	0.031	2.83	Hydrogen	0.14195	1.813
Methane	0.293	4.62	Methane	3.20	2.96
Neon	0.028	1.91	Neon	0.149	1.22
Nitrogen	0.176	4.19	Nitrogen	1.56	2.68
Oxygen	0.177	3.45	Oxygen	1.74	2.21
Xenon	0.536	5.56	Xenon	7.20	3.56
Water	0.709	3.29	Water	14.24	2.11

Table 12. (continued) Parameters for selected equations of state.

	Second virial coefficients, B ($cm^3\ mol^{-1}$)			
	100 K	273 K	373 K	600 K
Air	−167.3	−13.5	3.4	19.0
Ar	−187.0	−21.7	−4.2	11.9
CH_4		−53.6	−21.2	8.1
CO_2		−142	−72.2	−12.4
H_2	−2.0	13.7	15.6	
He	11.4	12.0	11.3	10.4
Kr		−62.9	−28.7	1.7
N_2	−160.0	−10.5	6.2	21.7
Ne	−6.0	10.4	12.3	13.8
O_2	−197.5	−22.0	−3.7	12.9
Xe		−153.7	−81.7	−19.6

Data: AIP, JL. The values relate to the expansion in eqn 1.22 of Section 1.3b; convert to eqn 1.21 using $B' = B/RT$.
For Ar at 273 K, $C = 1200\ cm^6\ mol^{-1}$.

9. Computational exercise. Objectives.

Suppose, an experimental chemist approaches you and says:

"*Hello Computational Scientist, I have collected a lot of pressure -volume data for many gases at different temperatures. Please write a computer program for me (in Mathematica, Maple, Matlab, Fortran, Pascal, C or C++ etc.) that can automatically read my data files (I will of course, provide the data file format) and fit the data using the method of non-linear least squares to an equation of state that is specified in the input file. I would also like your program to output statistics of a fit, including standard errors of optimized parameters, R and R^2. Oh, yes, one more thing. The program should have a user-friendly output, i.e. it should tell me what it does as it is running (remember, you are dealing with a chemist here), and at the very end I would like it to produce a plot of fitted equation of state with optimized parameters over the experimental data points. Finally, the runtime of the program should be under 1 hour for each dataset. It should be easy for you, right?*".

[Note: additional instructions are provided below]

Then, s(he) gives you the pressure-volume data file for, say, the argon (Ar) gas collected at 130K that you can use while writing and debugging your program. Here is the data file and explanations/instructions on how to read it.

```
Ar, vdW eos
vdW
10. 0.001
temp  130.0 K
m^3/mol Pa
     1.08088110000000E-01     9.99538762522963E+03
     5.40440550000000E-02     1.99653125543371E+04
     3.60293700000000E-02     2.99259020963550E+04
     2.70220275000000E-02     3.98535057502300E+04
     2.16176220000000E-02     4.97777690877801E+04
     1.80146850000000E-02     5.96668679380589E+04
     1.54411585714286E-02     6.95500237075953E+04
     1.35110137500000E-02     7.94659492082504E+04
     ........................................
```

Line 1: This is a comment line, one can put whatever s(he) wants there, but you need to read this comment and print it out before doing anything else.

Line 2: The requested equation of state for fitting. These are abbreviated as follows:

vdw	- van der Waals
rk	- Redlich-Kwong
dieterici	- Dieterici
berthelot	- Berthelot

Note that these keywords can be specified using either lower or upper case letters, or a mixture of those. The virial equation of state does not have to be included – it will be covered in Project 1a.

Line 3: initial guess for the fitted parameters

Line 4: The temperature of the experiment (always specified in kelvin)

Line 5: Keywords for units of measured V_m (first number) and p (second number). For V_m, these
can be as follows:

`dm^3/mol`	- dm^3 mol^{-1}
`m^3/mol`	- m^3 mol^{-1}
`cm^3/mol`	- cm^3 mol^{-1}
`L/mol`	- liter/mol (1 L = 10^{-3} m^3)

For p, there are more choices:

`pa`	- Pa
`megapa`	- MPa
`kilobar`	- Kbar
`bar`	- bar (1 bar = 10^5 Pa)
`atm`	- atm (1 atm = 101325 Pa)
`torr`	- torr (1 torr = $\frac{101325}{760}$ Pa)
`mmHg`	- millimeter of mercury (1 mmHg = 133.322387415 Pa)

Again, these keywords can be specified using either the lower or upper case letters, or a mixture of those.

Lines 6...: These are the actual data lines. The first number in each line is the value of V_m in units specified in line 5. The second number in each line, is obviously, p (in units specified in line 5). The number of data lines can vary from 10 to, say, 10^6, but you never know in advance how many data points a given file has.

You can assume that V_m and p data do not have associated standard deviations, so the only statistics you need to take into account are those related to your fit.

Additional instructions (read these carefully!)

The fit should be done using the non-linear least squares (NL LSQ) technique outlined below. You absolutely can not use built-in NL LSQ or any other optimization options that your software has (what fun is that anyway?!) in the final production code (which means you are allowed to use built-in minimization/optimization options for debugging only!). You are allowed to fit equations of states in one of the following forms (you choose which one):

$$pV_m = \ldots \tag{51}$$

$$p = \ldots \tag{52}$$

$$\frac{pV_m}{RT} = \ldots \tag{53}$$

whatever is more convenient for you. It is advised that before you perform the fit, you convert p-V data to standard SI units (that is, if input data are not in standard SI units): m^3 mol^{-1} for V_m, and Pa for p. Of course, you can do the fit in any units, but then you absolutely must convert your final optimized parameters to the standard SI units (see Table 12).

10. Introduction to the Method of Least Squares

In science, it is often important to fit a curve to a given set of experimental measurements (observations):

$$x_1^{obs}, x_2^{obs}, x_3^{obs}, x_4^{obs}, \ldots \rightarrow x_i^{obs}, i = 1 \ldots N \tag{54}$$

$$y_1^{obs}, y_2^{obs}, y_3^{obs}, y_4^{obs}, \ldots \rightarrow y_i^{obs}, i = 1 \ldots N \tag{55}$$

where N is the number of observations and superscript '*obs*' signifies that these are indeed observations. We assume that the curve can be represented by a mathematical expression that evaluates y as a function of x:

$$y = f(x, a_1, a_2, a_3, \ldots, a_M) \tag{56}$$

At this stage the form of the function $f(\ldots)$ is not important, but we note that in addition to variable x, this function also depends on M parameters (coefficients) a_j ($j=1...M$) that are constants in this particular function $f(\ldots)$. Based on a set of experimental observations $\{x_i^{obs}, y_i^{obs}\}$, we need to find coefficients a_j so that function $y = f(x, a_1, a_2, a_3, \ldots, a_M)$ could be evaluated at any x that we may be interested in. Note that at the beginning, we only have a discrete set of N observations at which the values of function $y = f(x, a_1, a_2, a_3, \ldots, a_M)$ are known (y_i^{obs}).

When substituting experimental values of x_i^{obs} into equation (56) we get a set of function values y_i^{calc} calculated at x_i^{obs}:

$$y_1^{calc}, y_2^{calc}, y_3^{calc}, y_4^{calc}, \ldots = y_i^{calc}, i = 1 \ldots N \tag{57}$$

where

$$y_i^{calc} = f(x_i^{obs}, a_1, a_2, a_3, \ldots a_M) \tag{58}$$

i.e.,

$$y_1^{calc} = f(x_1^{obs}, a_1, a_2, a_3, \ldots a_M)$$
$$y_2^{calc} = f(x_2^{obs}, a_1, a_2, a_3, \ldots a_M)$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$y_N^{calc} = f(x_N^{obs}, a_1, a_2, a_3, \ldots a_M) \tag{59}$$

Obviously we want to choose coefficients a_j such as to minimize differences between observed values y_i^{obs} and calculated values y_i^{calc}:

$$\Delta_i = y_i^{obs} - y_i^{calc} \tag{60}$$

Thus, the *error function S* (also called "chi-squared", χ^2) to be minimized is

$$S = \sum_{i=1}^{N} \Delta_i^2 = \sum_{i=1}^{N} \left(y_i^{obs} - y_i^{calc} \right)^2 \tag{61}$$

Because we use *squares* of the differences between observed and calculated values (there are several reasons for doing that!) this approach is called the method of *least squares* (LSQ).

Note that in expression (61), y_i^{obs} are numbers (results of our measurements), while y_i^{calc} are values of the function (58). Also note that coefficients a_j are the same for all values of x. Thus, expression (61) can be rewritten as:

$$S = \sum_{i=1}^{N} \Delta_i^2 = \sum_{i=1}^{N} \left(y_i^{obs} - y_i^{calc} \right)^2 = \sum_{i=1}^{N} \left[y_i^{obs} - f\left(x_i^{obs}, a_1, a_2, a_3, \ldots, a_M \right) \right]^2 \tag{62}$$

Note how S is a function of coefficients a_j. As long as the number of observations N is greater than the number of coefficients in expression (62), we should be able to determine these coefficients.

We have already mentioned that function S should be minimized by coefficients a_j which (according to the language of calculus) means that the first partial derivative of S with respect to (w.r.t.) each of the coefficients should be zero:

$$\frac{\partial S}{\partial a_1} = 0 \ , \ \ \frac{\partial S}{\partial a_2} = 0, \ \ \frac{\partial S}{\partial a_3} = 0 \ , \ \dots \ , \ \ \frac{\partial S}{\partial a_M} = 0 \tag{63}$$

Once the form of the function (56) in equation (62) is chosen, differentiation of S is straightforward.

11. The *Linear* Regression Analysis

*Note: This section is here just to illustrate a concept -
you need to program non-linear least squares*

While the form of function (56) can in general be fairly complicated, in this section we shall consider the simplest function:

$$y = f(x, a_1, a_2, a_3, ..., a_M) = a_1 + a_2 x = a + bx \tag{64}$$

This is the equation of a straight line, i.e. function y is *linearly* dependent on variable x. Coefficient b is the slope of the function, while coefficient a is the intercept.

We shall always try to fit a straight line through a set of experimental observations. If a set of observations does not follow a straight line we will try to *linearize* the data, i.e. using mathematical manipulations we will convert data to a linear form (64). For example, function

$$z = pe^{qx} \tag{65}$$

is obviously not linear w.r.t. variable x, but taking the natural logarithm of both sides

$$\ln(z) = \ln(pe^{qx}) = \ln(p) + \ln(e^{qx}) = \ln(p) + qx \tag{66}$$

gives us a linear function in a form similar to that in expression (64), where

$$y = \ln(z)$$
$$a = \ln(p)$$
$$b = q \tag{67}$$

In this example, by taking the natural log we have *linearized* our original function (65).

While the linear function (64) is very simple, it is very useful because we can derive analytical expressions for calculation of coefficients a and b that minimize function S (62). First we put equation (64) into equation (62):

$$S = \sum_{i=1}^{N} \left[y_i^{obs} - f\left(x_i^{obs}, a_1, a_2, a_3, ... a_M\right) \right]^2 = \sum_{i=1}^{N} \left[y_i^{obs} - \left(bx_i^{obs} + a\right) \right]^2 \tag{68}$$

Then we take partial derivatives of S w.r.t. coefficients a and b, and set them to zero:

$$\frac{\partial S}{\partial a} = \frac{\partial \sum_{i=1}^{N} \left(y_i^{obs} - bx_i^{obs} - a\right)^2}{\partial a} = 0$$

$$\frac{\partial S}{\partial b} = \frac{\partial \sum_{i=1}^{N} \left(y_i^{obs} - bx_i^{obs} - a\right)^2}{\partial b} = 0 \tag{69}$$

Note that these equations contain only coefficients a and b, and observations $\{x_i^{obs}, y_i^{obs}\}$. After a lot of basic calculus and college algebra, we get the final expressions for coefficients a and b:

$$a = \frac{\sum\limits_{i=1}^{N} y_i \sum\limits_{i=1}^{N} x_i^2 - \sum\limits_{i=1}^{N} x_i \sum\limits_{i=1}^{N} x_i y_i}{N\sum\limits_{i=1}^{N} x_i^2 - \left(\sum\limits_{i=1}^{N} x_i\right)^2}$$

$$b = \frac{N\sum\limits_{i=1}^{N} x_i y_i - \sum\limits_{i=1}^{N} x_i \sum\limits_{i=1}^{N} y_i}{N\sum\limits_{i=1}^{N} x_i^2 - \left(\sum\limits_{i=1}^{N} x_i\right)^2} \tag{70}$$

For simplification, superscript '*obs*' has been dropped from x_i's and y_i's. By defining the mean values of *x* and *y* as

$$\bar{x} = \frac{1}{N}\sum\limits_{i=1}^{N} x_i \quad \text{and} \quad \bar{y} = \frac{1}{N}\sum\limits_{i=1}^{N} y_i \tag{71}$$

expressions (17) can be somewhat simplified: [13]

$$a = \frac{\bar{y}\left(\sum\limits_{i=1}^{N} x_i^2\right) - \bar{x}\sum\limits_{i=1}^{N} x_i y_i}{\sum\limits_{i=1}^{N} x_i^2 - N(\bar{x})^2}$$

$$b = \frac{\sum\limits_{i=1}^{N} x_i y_i - N\bar{x}\bar{y}}{\sum\limits_{i=1}^{N} x_i^2 - N(\bar{x})^2} \tag{72}$$

These can be rewritten in an even simpler form by defining the sums of squares:[13]

$$S_{xx} = \sum\limits_{i=1}^{N}(x_i - \bar{x})^2 = \sum\limits_{i=1}^{N} x_i^2 - N(\bar{x})^2 \tag{73}$$

$$S_{yy} = \sum\limits_{i=1}^{N}(y_i - \bar{y})^2 = \sum\limits_{i=1}^{N} y_i^2 - N(\bar{y})^2 \tag{74}$$

$$S_{xy} = \sum\limits_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}) = \sum\limits_{i=1}^{N} x_i y_i - N(\bar{x})(\bar{y}) \tag{75}$$

Then, coefficients *a* and *b* can be calculated simply as:[13]

$$b = \frac{S_{xy}}{S_{xx}} \tag{76}$$

$$a = \bar{y} - b\bar{x} \tag{77}$$

The overall quality of the fit is then parameterized in terms of a quantity known as the correlation coefficient *r* (sometimes also denoted as *R; not to be confused with the gas constant!*), defined by

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} \tag{78}$$

The correlation coefficient is a quantity that gives the quality of a least squares fitting to the original data, i.e. how closely one variable is related to another variable. *r* always falls between -1 and +1. *r* = -1 tells you that there is a perfect negative relationship between the two variables, i.e. as values on one variable increase there is a perfectly predictable decrease in values of the

other variable. $r = +1$ tells you that there is a perfect positive relationship between the two variables. This means that as values of one variable increase there is a perfectly predictable increase in values of the other variable. $r = 0$ tells you that there is a zero correlation, or no relationship, between the two variables. In other words, as one variable changes (goes up or down) you can't really say anything about what happens to the other variable [14].

To calculate the errors (standard deviation in the slope σ_b and intercept σ_a) the following additional formulas are needed [13]:

$$s = \sqrt{\frac{S_{yy} - bS_{xy}}{N-2}} \tag{79}$$

$$\sigma_b = \frac{s}{\sqrt{S_{xx}}} \tag{80}$$

$$\sigma_a = s\sqrt{\frac{1}{N} + \frac{(\bar{x})^2}{S_{xx}}} \tag{81}$$

12. Non-linear Least Squares

In the previous section, we went over a simple fit to a straight line. But what if the data you would like to fit are not linear, and can not be linearized? We now consider fitting when the model depends nonlinearly on the set of M unknown parameters (coefficients) a_j ($j=1...M$). Also, unlike in the previous section, we would like to have a general method for fitting the data with minimal changes when a new non-linear model needs to be fitted. In what follows, I rely heavily on several excellent sources of information for non-linear fitting - [15], [16], [17], and [18].

As discussed in section 10, we are seeking the minimum of the error function S

$$S = \sum_{i=1}^{N} \Delta_i^2 = \sum_{i=1}^{N}\left[y_i^{obs} - f\left(x_i^{obs}, a_1, a_2, a_3, ..., a_M\right)\right]^2 \tag{82}$$

w.r.t. parameters a_j ($j=1...M$). The minimum value of S occurs when the gradient of S w.r.t. each parameter a_j vanishes. So, we end up with M gradient equations:

$$\frac{\partial S}{\partial a_1} = 0 \ , \ \ \frac{\partial S}{\partial a_2} = 0, \ \ \frac{\partial S}{\partial a_3} = 0 \ , \ ... \ , \ \frac{\partial S}{\partial a_M} = 0 \tag{83}$$

which can be written in a compact form as

$$\frac{\partial S}{\partial a_j} = 0 \quad (j = 1, ..., M) \tag{84}$$

Differentiating (82) w.r.t. a_j gives

$$\frac{\partial S}{\partial a_j} = \frac{\partial \sum_{i=1}^{N}\left[y_i^{obs} - f\left(x_i^{obs}, a_1, a_2, a_3, ..., a_M\right)\right]^2}{\partial a_j} = 0 \quad (j = 1, ..., M) \tag{85}$$

$$\frac{\partial S}{\partial a_j} = -2\sum_{i=1}^{N}\left[\Delta_i \frac{\partial f\left(x_i^{obs}, a_1, a_2, a_3, ..., a_M\right)}{\partial a_j}\right] = 0 \quad (j = 1, ..., M) \tag{86}$$

In a non-linear system, the derivatives $\dfrac{\partial S}{\partial a_j}$ are functions of both the independent variable x_i and the parameters a_j, so these gradient equations do not have a closed solution. Instead, initial values must be chosen for the parameters a_j. Then, the parameters are refined iteratively, that is, the values are obtained by successive approximation,

$$a_j \approx a_j^{k+1} = a_j^k + \Delta a_j \quad (j = 1,\dots,M) \tag{87}$$

where k is the iteration (cycle) number, and Δa_j is the perturbation applied to parameter a_j that reduces S. Think of Δa_j as a step that takes S downhill. In fact this can resemble a good old steepest (gradient) descend method which updates parameters a_j in the direction opposite to the gradient of the function being minimized (in our case the function being minimized is S).

Perhaps, now it is time to switch to a matrix notation which should significantly simply our discussion [or not, but we are going to do it anyway!]. Note that parameters a_j can be represented by a *row* vector **a** of length M:

$$\mathbf{a} = (a_1, a_2, a_3,\dots, a_M) \tag{88}$$

Then the notation for our function $f(x_i^{obs}, a_1, a_2, a_3,\dots,a_M)$ becomes

$$f(x_i^{obs}, a_1, a_2, a_3,\dots,a_M) = f(x_i^{obs}, \mathbf{a}) \tag{89}$$

and all functions $f(x_i^{obs}, \mathbf{a})$ are collected together in a *column* vector $\mathbf{f}(\mathbf{a})$ of length N with elements defined as:

$$\mathbf{f}(\mathbf{a}) = \begin{pmatrix} f(x_1^{obs}, \mathbf{a}) \\ f(x_2^{obs}, \mathbf{a}) \\ \dots \\ f(x_N^{obs}, \mathbf{a}) \end{pmatrix} \tag{90}$$

By the same token we can collect our observations y_i^{obs} into a column vector of length N:

$$\mathbf{y} = \begin{pmatrix} y_1^{obs} \\ y_2^{obs} \\ \dots \\ y_N^{obs} \end{pmatrix} \tag{91}$$

Obviously, the residuals Δ_i that enter the expression for S can be defined in the same way:

$$\Delta\mathbf{y} = \mathbf{y} - \mathbf{f}(\mathbf{a}) = \begin{pmatrix} y_1^{obs} - f(x_1^{obs}, \mathbf{a}) \\ y_2^{obs} - f(x_2^{obs}, \mathbf{a}) \\ \dots \\ y_N^{obs} - f(x_N^{obs}, \mathbf{a}) \end{pmatrix} \tag{92}$$

Then, in the matrix notation, the [scalar] error function S is given by

$$\begin{aligned} S(\mathbf{a}) = \sum_{i=1}^{N} \left[y_i^{obs} - f(x_i^{obs}, \mathbf{a}) \right]^2 &= \sum_{i=1}^{N} \Delta_i^2 \\ &= \Delta\mathbf{y}^{\mathrm{T}}\Delta\mathbf{y} \\ &= (\mathbf{y} - \mathbf{f}(\mathbf{a}))^{\mathrm{T}}(\mathbf{y} - \mathbf{f}(\mathbf{a})) \\ &= \mathbf{y}^{\mathrm{T}}\mathbf{y} - 2\mathbf{y}^{\mathrm{T}}\mathbf{f}(\mathbf{a}) + \mathbf{f}(\mathbf{a})^{\mathrm{T}}\mathbf{f}(\mathbf{a}) \end{aligned} \tag{93}$$

where superscript T indicates transposition. Recall that the *scalar product* of two column vectors **p** and **q** is defined as [19]

$$\mathbf{p}^T\mathbf{q} = \mathbf{q}^T\mathbf{p} = \sum_i p_i q_i \tag{94}$$

The gradient of the [scalar] error function S w.r.t. parameters given in the *row* vector **a** can also be represented by a *row* vector of length M:

$$\nabla S(\mathbf{a}) = \left(\frac{\partial S}{\partial a_1}, \frac{\partial S}{\partial a_2}, \frac{\partial S}{\partial a_3}, \ldots, \frac{\partial S}{\partial a_M} \right) \tag{95}$$

Now we apply the well-known rule for differentiating a product [19]

$$\nabla(\mathbf{p}^T\mathbf{q}) = (\nabla\mathbf{p}^T)\mathbf{q} + (\nabla\mathbf{q}^T)\mathbf{p} \tag{96}$$

to the expansion of $S(\mathbf{a})$ given in equation (93) to get the gradient of $S(\mathbf{a})$:

$$\begin{aligned}
\nabla S(\mathbf{a}) &= \nabla\left(\mathbf{y}^T\mathbf{y} - 2\mathbf{y}^T\mathbf{f}(\mathbf{a}) + \mathbf{f}(\mathbf{a})^T\mathbf{f}(\mathbf{a})\right) \\
&= \nabla\left(\mathbf{y}^T\mathbf{y}\right) - 2\nabla\left(\mathbf{y}^T\mathbf{f}(\mathbf{a})\right) + \nabla\left(\mathbf{f}(\mathbf{a})^T\mathbf{f}(\mathbf{a})\right) \\
&= 0 - 2\mathbf{y}^T\nabla\mathbf{f}(\mathbf{a}) + 2\mathbf{f}(\mathbf{a})^T\nabla\mathbf{f}(\mathbf{a}) \\
&= -2(\mathbf{y} - \mathbf{f}(\mathbf{a}))^T\nabla\mathbf{f}(\mathbf{a}) = -2\Delta\mathbf{y}^T\nabla\mathbf{f}(\mathbf{a})
\end{aligned} \tag{97}$$

where $\nabla\mathbf{f}(\mathbf{a}) = \dfrac{\partial\mathbf{f}(\mathbf{a})}{\partial\mathbf{a}}$ are the elements of the $N \times M$ *Jacobian matrix* which represents the local sensitivity of the function $\mathbf{f}(\mathbf{a})$ to variation in the parameters **a**:

$$\mathbf{J} = \begin{pmatrix}
\dfrac{\partial f(x_1^{obs}, \mathbf{a}^k)}{\partial a_1} & \dfrac{\partial f(x_1^{obs}, \mathbf{a}^k)}{\partial a_2} & \cdots & \dfrac{\partial f(x_1^{obs}, \mathbf{a}^k)}{\partial a_M} \\[2ex]
\dfrac{\partial f(x_2^{obs}, \mathbf{a}^k)}{\partial a_1} & \dfrac{\partial f(x_2^{obs}, \mathbf{a}^k)}{\partial a_2} & \cdots & \dfrac{\partial f(x_2^{obs}, \mathbf{a}^k)}{\partial a_N} \\[2ex]
\dfrac{\partial f(x_3^{obs}, \mathbf{a}^k)}{\partial a_1} & \dfrac{\partial f(x_3^{obs}, \mathbf{a}^k)}{\partial a_2} & \cdots & \dfrac{\partial f(x_3^{obs}, \mathbf{a}^k)}{\partial a_M} \\[2ex]
\dfrac{\partial f(x_4^{obs}, \mathbf{a}^k)}{\partial a_1} & \dfrac{\partial f(x_4^{obs}, \mathbf{a}^k)}{\partial a_2} & \cdots & \dfrac{\partial f(x_4^{obs}, \mathbf{a}^k)}{\partial a_M} \\[2ex]
\dfrac{\partial f(x_N^{...obs}, \mathbf{a}^k)}{\partial a_1} & \dfrac{\partial f(x_N^{...obs}, \mathbf{a}^k)}{\partial a_2} & \cdots & \dfrac{\partial f(x_N^{...obs}, \mathbf{a}^k)}{\partial a_M}
\end{pmatrix} \tag{98}$$

Note that in our case $N > M$. The Jacobian matrix is a function of constants, the independent variables and the parameters, so it changes from one iteration to the next. In terms of the Jacobian matrix **J**, the expression for $\nabla S(\mathbf{a})$ (equation 97) can be re-written simply as

$$\nabla S(\mathbf{a}) = -2\Delta\mathbf{y}^T\mathbf{J} \tag{99}$$

The Gradient (Steepest) Descent method

With equation (99) in place we are now in position to derive a formula for the *gradient (steepest) descent method* that minimizes S by following the direction opposite to the gradient at each step of the iterative procedure. Recall that equation (87) defined each iterative step as

$$a_j \approx a_j^{k+1} = a_j^k + \Delta a_j \quad (j = 1, \ldots, M) \tag{100}$$

which in matrix notation can be re-written as

$$\mathbf{a} \approx \mathbf{a}^{k+1} = \mathbf{a}^k + \Delta\mathbf{a} \tag{101}$$

According to equation (99), the perturbation $\Delta\mathbf{a}$, which is defined by a *row* vector in the form

$$\Delta \mathbf{a} = \left(\Delta a_1, \Delta a_2, \Delta a_3, ..., \Delta a_M \right), \tag{102}$$

applied at each step of the *gradient (steepest) descent method* can be calculated as

$$\Delta \mathbf{a} = \text{constant} \times \Delta \mathbf{y}^{\mathrm{T}} \mathbf{J} \tag{103}$$

where "constant" is a scalar parameter that determines the step in the steepest-descent (downhill) direction (and, incidentally includes the factor of 2 from equation 99). Note that should we wish to represent the perturbation by a *column* vector, equation (103) can be re-written as

$$\Delta \mathbf{a}^{\mathrm{T}} = \text{constant} \times \mathbf{J}^{\mathrm{T}} \Delta \mathbf{y} \tag{104}$$

The Gauss-Newton method

The Gauss-Newton method presumes that the objective function is approximately quadratic in the parameters near the optimal solution [16]. For moderately-sized problems the Gauss-Newton method typically converges much faster than gradient-descent methods [20].

The function evaluated with perturbed model parameters may be locally approximated using Taylor's theorem (expansion):

$$f(b) = f(a) + f'(a)(b-a) + \frac{f''(a)}{2!}(b-a)^2 + \frac{f'''(a)}{3!}(b-a)^3 + \cdots \tag{105}$$

Suppose, the minimum of the function f is given by $f(a)$, but we are currently located at b. We want to take as few steps as possible to get from b to a. In the Gauss-Newton method, Taylor's expansion is truncated at the first order:

$$f(a) \approx f(b) + f'(b)(a-b) \tag{106}$$

Applying eqn. (106) to our model function $\mathbf{f(a)}$, we get

$$\mathbf{f(a + \Delta a)} \approx \mathbf{f(a)} + \nabla \mathbf{f(a)} \Delta \mathbf{a}^{\mathrm{T}} = \mathbf{f(a)} + \mathbf{J} \Delta \mathbf{a}^{\mathrm{T}} \tag{107}$$

Substituting the approximation for the perturbed function, $\nabla \mathbf{f(a)} + \mathbf{J} \Delta \mathbf{a}^{\mathrm{T}}$, for $\mathbf{f(a + \Delta a)}$ in the expansion for $S(\mathbf{a + \Delta a})$ obtained from that for $S(\mathbf{a})$ (equation 93) results in

$$
\begin{aligned}
S(\mathbf{a + \Delta a}) &= \mathbf{y}^{\mathrm{T}}\mathbf{y} - 2\mathbf{y}^{\mathrm{T}}\mathbf{f(a + \Delta a)} + \mathbf{f(a + \Delta a)}^{\mathrm{T}}\mathbf{f(a + \Delta a)} \\
&\approx \mathbf{y}^{\mathrm{T}}\mathbf{y} - 2\mathbf{y}^{\mathrm{T}}\left(\mathbf{f(a)} + \mathbf{J}\Delta \mathbf{a}^{\mathrm{T}} \right) + \left(\mathbf{f(a)} + \mathbf{J}\Delta \mathbf{a}^{\mathrm{T}} \right)^{\mathrm{T}}\left(\mathbf{f(a)} + \mathbf{J}\Delta \mathbf{a}^{\mathrm{T}} \right) \\
&= \mathbf{y}^{\mathrm{T}}\mathbf{y} - 2\mathbf{y}^{\mathrm{T}}\mathbf{f(a)} - 2\mathbf{y}^{\mathrm{T}}\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}} + \mathbf{f(a)}^{\mathrm{T}}\mathbf{f(a)} + \mathbf{f(a)}^{\mathrm{T}}\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}} + \left(\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}}\right)^{\mathrm{T}}\mathbf{f(a)} + \left(\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}}\right)^{\mathrm{T}}\left(\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}}\right) \\
&= \mathbf{y}^{\mathrm{T}}\mathbf{y} - 2\mathbf{y}^{\mathrm{T}}\mathbf{f(a)} - 2\mathbf{y}^{\mathrm{T}}\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}} + \mathbf{f(a)}^{\mathrm{T}}\mathbf{f(a)} + \mathbf{f(a)}^{\mathrm{T}}\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}} + \mathbf{f(a)}^{\mathrm{T}}\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}} + \left(\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}}\right)^{\mathrm{T}}\left(\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}}\right) \\
&= \mathbf{y}^{\mathrm{T}}\mathbf{y} + \mathbf{f(a)}^{\mathrm{T}}\mathbf{f(a)} - 2\mathbf{y}^{\mathrm{T}}\mathbf{f(a)} - 2\mathbf{y}^{\mathrm{T}}\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}} + 2\mathbf{f(a)}^{\mathrm{T}}\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}} + \left(\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}}\right)^{\mathrm{T}}\left(\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}}\right) \\
&= \mathbf{y}^{\mathrm{T}}\mathbf{y} + \mathbf{f(a)}^{\mathrm{T}}\mathbf{f(a)} - 2\mathbf{y}^{\mathrm{T}}\mathbf{f(a)} - 2\mathbf{y}^{\mathrm{T}}\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}} + 2\mathbf{f(a)}^{\mathrm{T}}\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}} + \left(\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}}\right)^{\mathrm{T}}\left(\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}}\right) \\
&= \mathbf{y}^{\mathrm{T}}\mathbf{y} + \mathbf{f(a)}^{\mathrm{T}}\mathbf{f(a)} - 2\mathbf{y}^{\mathrm{T}}\mathbf{f(a)} - 2\left(\mathbf{y}^{\mathrm{T}} - \mathbf{f(a)}^{\mathrm{T}}\right)\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}} + \left(\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}}\right)^{\mathrm{T}}\left(\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}}\right) \\
&= \mathbf{y}^{\mathrm{T}}\mathbf{y} + \mathbf{f(a)}^{\mathrm{T}}\mathbf{f(a)} - 2\mathbf{y}^{\mathrm{T}}\mathbf{f(a)} - 2\left(\mathbf{y} - \mathbf{f(a)}\right)^{\mathrm{T}}\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}} + \left(\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}}\right)^{\mathrm{T}}\left(\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}}\right) \\
&= \mathbf{y}^{\mathrm{T}}\mathbf{y} + \mathbf{f(a)}^{\mathrm{T}}\mathbf{f(a)} - 2\mathbf{y}^{\mathrm{T}}\mathbf{f(a)} - 2\Delta \mathbf{y}^{\mathrm{T}}\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}} + \left(\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}}\right)^{\mathrm{T}}\left(\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}}\right) \\
&= \mathbf{y}^{\mathrm{T}}\mathbf{y} + \mathbf{f(a)}^{\mathrm{T}}\mathbf{f(a)} - 2\mathbf{y}^{\mathrm{T}}\mathbf{f(a)} - 2\Delta \mathbf{y}^{\mathrm{T}}\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}} + \Delta \mathbf{a}\mathbf{J}^{\mathrm{T}}\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}}
\end{aligned}
\tag{108}
$$

where we used the following relationships:

$$
\begin{aligned}
\left(\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}}\right)^{\mathrm{T}}\mathbf{f(a)} &= \mathbf{f(a)}^{\mathrm{T}}\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}} \\
\left(\mathbf{J}\Delta \mathbf{a}^{\mathrm{T}}\right)^{\mathrm{T}} &= \Delta \mathbf{a}\mathbf{J}^{\mathrm{T}}
\end{aligned}
\tag{109}
$$

This shows that the error function is approximately quadratic in the perturbation $\Delta \mathbf{a}$, and that the *Hessian matrix* of S is approximately $\mathbf{J}^{\mathrm{T}}\mathbf{J}$ (this matrix is also called a *curvature matrix* because

of its relationship to the curvature of the S function in the parameter space [17]). Next, we expand the gradient of S about $\mathbf{a}+\Delta\mathbf{a}$ by modifying equation (99) as follows:

$$\nabla S(\mathbf{a}+\Delta\mathbf{a})=-2\Delta\mathbf{y}_{\mathbf{a}+\Delta\mathbf{a}}^{\mathrm{T}}\mathbf{J}_{\mathbf{a}+\Delta\mathbf{a}} \tag{110}$$

Expansions for $\Delta\mathbf{y}_{\mathbf{a}+\Delta\mathbf{a}}$ and $\mathbf{J}_{\mathbf{a}+\Delta\mathbf{a}}$ are easy because we make use of equation (107):

$$\Delta\mathbf{y}_{\mathbf{a}+\Delta\mathbf{a}}^{\mathrm{T}}=\left(\mathbf{y}-\mathbf{f}(\mathbf{a}+\Delta\mathbf{a})\right)^{\mathrm{T}}\approx\left(\mathbf{y}-\mathbf{f}(\mathbf{a})-\mathbf{J}\Delta\mathbf{a}^{\mathrm{T}}\right)^{\mathrm{T}} \tag{111}$$

$$\mathbf{J}_{\mathbf{a}+\Delta\mathbf{a}}=\nabla\mathbf{f}(\mathbf{a}+\Delta\mathbf{a})\approx\nabla\left(\mathbf{f}(\mathbf{a})+\mathbf{J}\Delta\mathbf{a}^{\mathrm{T}}\right)=\nabla\mathbf{f}(\mathbf{a})+\nabla\mathbf{J}\Delta\mathbf{a}^{\mathrm{T}}=\mathbf{J}+\nabla\mathbf{J}\Delta\mathbf{a}^{\mathrm{T}} \tag{112}$$

Substituting equations (111) and (112) into equation (110) we get

$$\begin{aligned}\nabla S(\mathbf{a}+\Delta\mathbf{a})&=-2\Delta\mathbf{y}_{\mathbf{a}+\Delta\mathbf{a}}^{\mathrm{T}}\mathbf{J}_{\mathbf{a}+\Delta\mathbf{a}}\\ &\approx-2\left(\mathbf{y}-\mathbf{f}(\mathbf{a})-\mathbf{J}\Delta\mathbf{a}^{\mathrm{T}}\right)^{\mathrm{T}}\left(\mathbf{J}+\nabla\mathbf{J}\Delta\mathbf{a}^{\mathrm{T}}\right)\\ &=-2\left(\mathbf{y}^{\mathrm{T}}-\mathbf{f}(\mathbf{a})^{\mathrm{T}}-\Delta\mathbf{a}\mathbf{J}^{\mathrm{T}}\right)\left(\mathbf{J}+\nabla\mathbf{J}\Delta\mathbf{a}^{\mathrm{T}}\right)\\ &=-2\left(\mathbf{y}^{\mathrm{T}}\mathbf{J}-\mathbf{f}(\mathbf{a})^{\mathrm{T}}\mathbf{J}-\Delta\mathbf{a}\mathbf{J}^{\mathrm{T}}\mathbf{J}+\mathbf{y}^{\mathrm{T}}\nabla\mathbf{J}\Delta\mathbf{a}^{\mathrm{T}}-\mathbf{f}(\mathbf{a})^{\mathrm{T}}\nabla\mathbf{J}\Delta\mathbf{a}^{\mathrm{T}}-\Delta\mathbf{a}\mathbf{J}^{\mathrm{T}}\nabla\mathbf{J}\Delta\mathbf{a}^{\mathrm{T}}\right)\end{aligned} \tag{113}$$

Ignoring higher order terms involving the gradient of the Jacobian, $\nabla\mathbf{J}$, we obtain an approximate but simple expression for the gradient of the error function

$$\begin{aligned}\nabla S(\mathbf{a}+\Delta\mathbf{a})&\approx-2\left(\mathbf{y}^{\mathrm{T}}\mathbf{J}-\mathbf{f}(\mathbf{a})^{\mathrm{T}}\mathbf{J}-\Delta\mathbf{a}\mathbf{J}^{\mathrm{T}}\mathbf{J}\right)\\ &=-2\left(\left(\mathbf{y}^{\mathrm{T}}-\mathbf{f}(\mathbf{a})^{\mathrm{T}}\right)\mathbf{J}-\Delta\mathbf{a}\mathbf{J}^{\mathrm{T}}\mathbf{J}\right)\\ &=-2\left(\left(\mathbf{y}-\mathbf{f}(\mathbf{a})\right)^{\mathrm{T}}\mathbf{J}-\Delta\mathbf{a}\mathbf{J}^{\mathrm{T}}\mathbf{J}\right)\\ &=-2\Delta\mathbf{y}^{\mathrm{T}}\mathbf{J}+2\Delta\mathbf{a}\mathbf{J}^{\mathrm{T}}\mathbf{J}\end{aligned} \tag{114}$$

The perturbation $\Delta\mathbf{a}$ that minimizes S is found by setting

$$\nabla S(\mathbf{a}+\Delta\mathbf{a})=0 \tag{115}$$

which results in the normal equations for the Gauss-Newton perturbation

$$\Delta\mathbf{a}\mathbf{J}^{\mathrm{T}}\mathbf{J}=\Delta\mathbf{y}^{\mathrm{T}}\mathbf{J} \tag{116}$$

Recall that the perturbation $\Delta\mathbf{a}$ is a *row* vector. To get the perturbation in a form of a *column* vector we transpose both sides of equation (116):

$$\left(\Delta\mathbf{a}\left(\mathbf{J}^{\mathrm{T}}\mathbf{J}\right)\right)^{\mathrm{T}}=\left(\Delta\mathbf{y}^{\mathrm{T}}\mathbf{J}\right)^{\mathrm{T}} \tag{117}$$

Because the transpose of a product of two matrices $\mathbf{P}$ and $\mathbf{Q}$ is given by

$$\left(\mathbf{PQ}\right)^{\mathrm{T}}=\mathbf{Q}^{\mathrm{T}}\mathbf{P}^{\mathrm{T}} \tag{118}$$

equation (117) is simplified as

$$\left(\mathbf{J}^{\mathrm{T}}\mathbf{J}\right)\Delta\mathbf{a}^{\mathrm{T}}=\mathbf{J}^{\mathrm{T}}\Delta\mathbf{y} \tag{119}$$

The solution of equation (119) for $\Delta\mathbf{a}^{\mathrm{T}}$ is obtained as follows:

$$\left(\mathbf{J}^{\mathrm{T}}\mathbf{J}\right)^{-1}\left(\mathbf{J}^{\mathrm{T}}\mathbf{J}\right)\Delta\mathbf{a}^{\mathrm{T}}=\left(\mathbf{J}^{\mathrm{T}}\mathbf{J}\right)^{-1}\mathbf{J}^{\mathrm{T}}\Delta\mathbf{y} \tag{120}$$

Since the inverse of a product of matrices $\mathbf{P}$ and $\mathbf{Q}$ can be expressed as

$$\left(\mathbf{PQ}\right)^{-1}=\mathbf{Q}^{-1}\mathbf{P}^{-1}, \tag{121}$$

the solution for $\Delta\mathbf{a}^{\mathrm{T}}$ becomes simply

$$\Delta\mathbf{a}^{\mathrm{T}}=\mathbf{J}^{-1}\left(\mathbf{J}^{\mathrm{T}}\right)^{-1}\mathbf{J}^{\mathrm{T}}\Delta\mathbf{y} \tag{122}$$

$$\Delta\mathbf{a}^{\mathrm{T}}=\mathbf{J}^{-1}\Delta\mathbf{y} \tag{123}$$

The Levenberg-Marquardt method

The Levenberg-Marquardt algorithm adaptively varies the parameter updates between the gradient (steepest) descent update and the Gauss-Newton update,

$$\left(\mathbf{J}^{\mathrm{T}}\mathbf{J} + \lambda\mathbf{I}\right)\Delta\mathbf{a}^{\mathrm{T}} = \mathbf{J}^{\mathrm{T}}\Delta\mathbf{y} \tag{124}$$

where **I** is the identity matrix (sometimes also denoted **1**, or **E)** is defined such that [21]

$$\mathbf{I}(\mathbf{p}) \equiv \mathbf{p} \tag{125}$$

for all vectors **p**, and where *small* values of the algorithmic parameter λ result in a Gauss-Newton update and *large* values of λ result in a gradient descent update [16]. The parameter λ is initialized to be large. If an iteration happens to result in a worse approximation, λ is increased. As the solution approaches the minimum, λ is decreased, the Levenberg-Marquardt method approaches the Gauss-Newton method, and the solution typically converges rapidly to the local minimum [16,20].

Marquardt's suggested update relationship is given by [20]

$$\left[\mathbf{J}^{\mathrm{T}}\mathbf{J} + \lambda\,\mathrm{diag}\left(\mathbf{J}^{\mathrm{T}}\mathbf{J}\right)\right]\Delta\mathbf{a}^{\mathrm{T}} = \mathbf{J}^{\mathrm{T}}\Delta\mathbf{y} \tag{126}$$

where $\mathrm{diag}\left(\mathbf{J}^{\mathrm{T}}\mathbf{J}\right)$ is matrix of the same dimensions as $\mathbf{J}^{\mathrm{T}}\mathbf{J}$ that has the $\mathbf{J}^{\mathrm{T}}\mathbf{J}$ elements on the leading diagonal and zeros elsewhere (i.e. matrix $\mathbf{J}^{\mathrm{T}}\mathbf{J}$ with off-diagonal elements set to zero).

The Levenberg-Marquardt method is reasonably insensitive to the starting values of the parameters *unless they are outside of reasonable range*. Unlike the gradient-only based methods, it provides an estimate of the full error matrix and better calculation of the diagonal errors [17].

Suggested implementation of the Levenberg-Marquardt method

The implementation of the Levenberg-Marquardt method is that described in detail in references [15], [17], and [18].

1. Derive and implement formulas for model functions $f(x,\mathbf{a})$ (recall, we need to perform fits to several functional forms) and their derivatives w.r.t. parameters **a**:

$$\frac{\partial f(x,\mathbf{a})}{\partial a_1}, \frac{\partial f(x,\mathbf{a})}{\partial a_2}, \frac{\partial f(x,\mathbf{a})}{\partial a_3}, \ldots, \frac{\partial f(x,\mathbf{a})}{\partial a_M} \tag{127}$$

2. Pick a value for λ. In [15,17] it is suggested to set λ=0.001. However, in [16] it says that "[t]*he parameter λ is initialized to be large*". Basically, it depends whether you would like the first step to be taken using the *steepest-descent* (*large λ*) or *Gauss-Newton* (*small λ*) method. It is your call!

3. Using the initial guess parameters **a** read from the input file, compute the error function $S(\mathbf{a})$:

$$S(\mathbf{a}) = \sum_{i=1}^{N}\left[y_i^{obs} - f\left(x_i^{obs},\mathbf{a}\right)\right]^2 = \sum_{i=1}^{N}\Delta_i^2 = \Delta\mathbf{y}^{\mathrm{T}}\Delta\mathbf{y} \tag{128}$$

4. Use parameters **a** to compute the gradient vector of the error function $\nabla S(\mathbf{a})$:

$$\nabla S(\mathbf{a}) = -2\Delta\mathbf{y}^{\mathrm{T}}\mathbf{J} \tag{129}$$

$$\frac{\partial S}{\partial a_m} = -2\sum_{i=1}^{N}\left[\Delta_i\,\frac{\partial f\left(x_i^{obs},\mathbf{a}\right)}{\partial a_m}\right] = 0 \quad (m=1,\ldots,M) \tag{130}$$

Alternatively, it may be convenient to remove the factor of 2 by defining the *column vector* $\boldsymbol{\beta}$ as

$$\boldsymbol{\beta} = \mathbf{J}^{\mathrm{T}}\Delta\mathbf{y} = \nabla\mathbf{f}(\mathbf{a})^{\mathrm{T}}\Delta\mathbf{y} \tag{131}$$

$$\beta_m = \sum_{i=1}^{N}\left[\Delta_i \frac{\partial f\left(x_i^{obs},\mathbf{a}\right)}{\partial a_m}\right] \quad (m = 1,\ldots,M) \tag{132}$$

Note that the gradient (steepest) descent formula (104) can be written in terms of $\boldsymbol{\beta}$ as

$$\Delta\mathbf{a}^{\mathrm{T}} = \mathrm{constant}\times\boldsymbol{\beta} \tag{133}$$

5. Compute the approximate Hessian matrix of $S(\mathbf{a})$ defined as $\mathbf{J}^{\mathrm{T}}\mathbf{J}$.

$$\boldsymbol{\alpha} = \mathbf{J}^{\mathrm{T}}\mathbf{J} \tag{134}$$

$$\alpha_{mn} = \sum_{i=1}^{N}\left[\frac{\partial f\left(x_i^{obs},\mathbf{a}\right)}{\partial a_m}\frac{\partial f\left(x_i^{obs},\mathbf{a}\right)}{\partial a_n}\right] \quad \left\{\begin{matrix} m = 1,\ldots,M \\ n = 1,\ldots,M \end{matrix}\right\} \tag{135}$$

Note that the Gauss-Newton normal equations (119) can be re-written in terms of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ as

$$\boldsymbol{\alpha}\Delta\mathbf{a}^{\mathrm{T}} = \boldsymbol{\beta} \tag{136}$$

$$\sum_{n=1}^{M}\alpha_{mn}\Delta a_n = \beta_m \quad (m = 1,\ldots,M) \tag{137}$$

6. Modify matrix $\boldsymbol{\alpha}$ as prescribed by Marquardt [20]. In references [15,17] it is suggested to define a new matrix $\boldsymbol{\alpha}'$ according to the following rule(s):

$$\alpha'_{mn} = \begin{cases} \alpha_{mn}(1+\lambda) & \text{for } m = n \\ \alpha_{mn} & \text{for } m \neq n \end{cases} \tag{138}$$

7. Solve the linear equations to get $\Delta\mathbf{a}$:

$$\boldsymbol{\alpha}'\Delta\mathbf{a}^{\mathrm{T}} = \boldsymbol{\beta} \tag{139}$$

$$\sum_{n=1}^{M}\alpha'_{mn}\Delta a_n = \beta_m \quad (m = 1,\ldots,M) \tag{140}$$

8. Evaluate $S(\mathbf{a}+\Delta\mathbf{a})$
   - If $S(\mathbf{a}+\Delta\mathbf{a}) \geq S(\mathbf{a})$, *increase* λ by a factor of 10 (or any other substantial factor) and go back to step 6.
   - If $S(\mathbf{a}+\Delta\mathbf{a}) < S(\mathbf{a})$, *decrease* λ by a factor of 10, update the trial solution $\mathbf{a}+\Delta\mathbf{a} \rightarrow \mathbf{a}$, replace $S(\mathbf{a}+\Delta\mathbf{a}) \rightarrow S(\mathbf{a})$, and go back to step 4.

Your code must print out the following information <u>for each iteration (cycle)</u>:
   a. cycle number,
   b. value of the lambda parameter (λ),
   c. error function value, and
   d. values of the fitted coefficients.

Also necessary is a condition for stopping [15]. Iterating to convergence (to machine accuracy or to the roundoff limit) is generally wasteful and unnecessary since the

minimum is at best only a statistical estimate of the parameters **a**. In practice, one might as well stop iterating after a few occurrences of S decreasing by a negligible amount, say either less than 10^{-5} absolutely or (in case roundoff prevents that being reached) fractionally. Don't stop after a step where S increases more than trivially: that only shows that λ has not yet adjusted itself optimally [15].

9. Once an acceptable minimum has been found, compute and print out the following statistics of the fit.

a) Print out "chi square", χ^2, which is (of course) the error function that was minimized

$$S(\mathbf{a}) = \sum_{i=1}^{N}\left[y_i^{obs} - f\left(x_i^{obs},\mathbf{a}\right)\right]^2 \qquad (141)$$

b) Compute and print out the *sample variance* σ^2 defined as [17]

$$\sigma^2 \cong s^2 = \frac{1}{N-M}\sum_{i=1}^{N}\left[y_i^{obs} - f\left(x_i^{obs},\mathbf{a}\right)\right]^2 \qquad (142)$$

where N-M is the number of degrees of freedom and is equal to the number of measurements minus the number of parameters determined from the fit [17]. Note that the true residual variance σ^2 is replaced by an estimate based on the minimized value of error function S [22].

c) Set $\lambda=0$ and compute the matrix **C**:

$$\mathbf{C} = \boldsymbol{\alpha}^{-1} \qquad (143)$$

which is the estimated the *variance-covariance matrix* (also called the *covariance matrix*) of the standard errors in the fitted parameters. For example, the *variance* on the parameter a_m can be estimated as [23,24] :

$$\mathrm{var}(a_m) = \sigma_{a_m}^2 = \sigma^2 C_{mm} \approx s^2 C_{mm} \qquad (144)$$

while the *covariance* between parameters a_m and a_n is given by

$$\mathrm{covar}(a_m,a_n) = \sigma_{mn}^2 \approx s^2 C_{mn} \qquad (145)$$

d) Calculate and print out *the standard deviation* (also called the *standard error*) for each fitted parameter a_m (σ_{a_m}) from the diagonal elements of the variance-covariance matrix (i.e. variances) and the sample variance σ^2 [23,24]

$$\sigma_{a_m} = \sqrt{s^2 C_{mm}} \qquad (146)$$

e) Calculate and print out the matrix of correlation coefficients ρ_{mn} between parameters a_m and a_n defined as [24]

$$\rho_{mn} = \frac{\sigma_{mn}^2}{\sigma_{a_m}\sigma_{a_n}} \qquad (147)$$

The correlation coefficient is a measure of the degree of correlation between the parameters [24]. The values of ρ_{mn} are in the range from -1 to 1. If the value is zero, then the parameters are uncorrelated (i.e., independent), if the value is 1, then they fall exactly on a line with a positive slope and if the value is -1 then the fall exactly on a line with a negative slope.

f) Calculate and print out the *coefficient of determination R²* (pronounced "R squared")
[25]

$$R^2 = 1 - \frac{S}{SS_{\text{tot}}} \tag{148}$$

where SS_{tot} is the total sum of squares (proportional to the sample variance)

$$SS_{\text{tot}} = \sum_{i=1}^{N}\left[y_i^{obs} - \bar{y}\right]^2 \tag{149}$$

and $\bar{y}$ is the mean of the observed data

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i^{obs} \tag{150}$$

Warning! Do not confuse the coefficient of determination (R²) with the gas constant (R).
[*Some students were known to do that!*]

g) Calculate and print out the *adjusted R²* (often written as $\bar{R}^2$ and pronounced "R bar
squared") [25]

$$\bar{R}^2 = 1 - \frac{S/(N-M-1)}{SS_{\text{tot}}/(N-1)} \tag{151}$$

The adjusted R^2 can be negative, and its value will always be less than or equal to that of
R^2. Adjusted R^2 does not have the same interpretation as R^2 - while R^2 is a measure of fit,
adjusted R^2 is instead a comparative measure of suitability of alternative nested sets of
explanators (*M*). As such, care must be taken in interpreting and reporting this statistic.
Adjusted R^2 is particularly useful in the feature selection stage of model building [25].

h) It may be useful to compute and print out the *R*-factor (sometimes called *residual
factor* or *reliability factor*) [26] which is used in X-ray crystallography and must be
reported for each determined crystal structure:

$$R-\text{factor} = \frac{\sum_{i=1}^{N}\left|y_i^{obs} - f\left(x_i^{obs}, \mathbf{a}\right)\right|}{\sum_{i=1}^{N}\left|y_i^{obs}\right|} \tag{152}$$

It is a measure of how well the model agrees with the measured (observed) data. The
value is also sometimes called the *discrepancy index*, as it mathematically describes the
difference between the experimental observations and the calculated values. The
absolute range of values is zero to one, with a large *R*-factor indicating a poor fit [26]. It
is customary to report *R*-factor in percent (%).

i) Finally, on the same graph plot the observed data and the fitted model.

**That's it! Do not forget to submit your code and all required files to D2L by the due
date/time.**

References

[1] http://physics.nist.gov

[2] Atkins, P; de Paula, J. *Physical Chemistry*, 8th edition, W. H. Freeman and Company, New York, 2006.

[3] Young, H. D., Freedman, R. A., Sandin, T. R., Ford, A. L. *Sears and Zemansky's University Physics With Modern Physics*, 13th edition, Addison-Wesley, 2010.

[4] Serway, R. A., Vuille, C. *College Physics*, 9th edition. Cengage Learning, 2011.

[5] http://en.wikipedia.org/wiki/Density

[6] Silberberg , M. S. *Principles of General Chemistry*, 1st edition, McGrawHill, 2006.

[7] http://en.wikipedia.org/wiki/Physical_constant

[8] Atkins, P; de Paula, J. *Physical Chemistry for the Life Sciences*, W. H. Freeman and Company, New York, 2006.

[9] Atkins, P; de Paula, J. *Elements of Physical Chemistry*, Fifth edition, W. H. Freeman, New York, 2009.

[10] Mortimer, R. G. *Physical Chemistry*, 3rd edition, Academic Press, 2008.

[11] B. Ott, J. R. Goates, and H. T. Hall, Jr., J. Chem. Educ., 48, 515 (1971);

[12] M.W. Kemp, R. E. Thompson, and D. J. Zigrang, J. Chem. Educ., 52, 802 (1975).

[13] http://mathworld.wolfram.com/LeastSquaresFitting.html

[14] Higgins. J. *The Radical Statistician*, 2010.

[15] Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, 3rd edition. Cambridge University Press, 2007.

[16] Gavin, J. P. *The Levenberg-Marquardt method for nonlinear least squares curve-fitting problems*. Department of Civil and Environmental Engineering, Duke University, 2013. http://people.duke.edu/~hpgavin/ce281/lm.pdf

[17] Bevington, P., Robinson, D. K. *Data Reduction and Error Analysis for the Physical Sciences*, 3rd edition. McGraw-Hill Science/Engineering/Math, 2002

[18] http://en.wikipedia.org/wiki/Non-linear_least_squares
http://en.wikipedia.org/wiki/Levenberg%E2%80%93Marquardt_algorithm

[19] Fletcher, R. *Practical Methods of Optimization*, 2nd edition, John Wiley & Sons, 1987.

[20] Marquardt, D.W. "*An algorithm for least-squares estimation of nonlinear parameters*" Journal of the Society for Industrial and Applied Mathematics, 11(2):431-441, 1963.

[21] http://mathworld.wolfram.com/IdentityMatrix.html

[22] http://en.wikipedia.org/wiki/Least_squares#Non-linear_least_squares

[23] http://en.wikipedia.org/wiki/Ordinary_least_squares

[24] Wolberg, J. *Data Analysis Using the Method of Least Squares: Extracting the Most Information from Experiments*, Springer, 2006.

[25] http://en.wikipedia.org/wiki/R_squared

[26] http://en.wikipedia.org/wiki/R-factor_crystallography
http://reference.iucr.org/dictionary/R_factor