Chapter 1
Introduction

In this chapter, we describe the aims of perturbation theory in general terms, and give some simple illustrative examples of perturbation problems. Some texts and references on perturbation theory are [6], [7], and [10].

1.1 Perturbation theory

Consider a problem

$$P^\varepsilon(x) = 0 \tag{1.1}$$

depending on a small, real-valued parameter ε that simplifies in some way when $\varepsilon = 0$ (for example, it is linear or exactly solvable). The aim of perturbation theory is to determine the behavior of the solution $x = x^\varepsilon$ of (1.1) as $\varepsilon \to 0$. The use of a small parameter here is simply for definiteness; for example, a problem depending on a large parameter ω can be rewritten as one depending on a small parameter $\varepsilon = 1/\omega$.

The focus of these notes is on perturbation problems involving differential equations, but perturbation theory and asymptotic analysis apply to a broad class of problems. In some cases, we may have an explicit expression for x^ε, such as an integral representation, and want to obtain its behavior in the limit $\varepsilon \to 0$.

1.1.1 *Asymptotic solutions*

The first goal of perturbation theory is to construct a formal asymptotic solution of (1.1) that satisfies the equation up to a small error. For example, for each $N \in \mathbb{N}$, we may be able to find an asymptotic solution x_N^ε such that

$$P^\varepsilon\left(x_N^\varepsilon\right) = O(\varepsilon^{N+1}),$$

where $O(\varepsilon^n)$ denotes a term of the the order ε^n. This notation will be made precise in Chapter 2.

Once we have constructed such an asymptotic solution, we would like to know that there is an exact solution $x = x^\varepsilon$ of (1.1) that is close to the asymptotic solution when ε is small; for example, a solution such that

$$x^\varepsilon = x_N^\varepsilon + O(\varepsilon^{N+1}).$$

This is the case if a small error in the equation leads to a small error in the solution. For example, we can establish such a result if we have a stability estimate of the form

$$|x - y| \leq C\,|P^\varepsilon(x) - P^\varepsilon(y)|$$

where C is a constant independent of ε, and $|\cdot|$ denotes appropriate norms. Such an estimate depends on the properties of P^ε and may be difficult to obtain, especially for nonlinear problems. In these notes we will focus on methods for the construction of asymptotic solutions, and we will not discuss in detail the existence of solutions close to the asymptotic solution.

1.1.2 *Regular and singular perturbation problems*

It is useful to make an imprecise distinction between *regular perturbation problems* and *singular perturbation problems*. A regular perturbation problem is one for which the perturbed problem for small, nonzero values of ε is qualitatively the same as the unperturbed problem for $\varepsilon = 0$. One typically obtains a convergent expansion of the solution with respect to ε, consisting of the unperturbed solution and higher-order corrections. A singular perturbation problem is one for which the perturbed problem is qualitatively different from the unperturbed problem. One typically obtains an asymptotic, but possibly divergent, expansion of the solution, which depends singularly on the parameter ε.

Although singular perturbation problems may appear atypical, they are the most interesting problems to study because they allow one to understand qualitatively new phenomena.

The solutions of singular perturbation problems involving differential equations often depend on several widely different length or time scales. Such problems can be divided into two broad classes: layer problems, treated using the method of matched asymptotic expansions (MMAE); and multiple-scale problems, treated by the method of multiple scales (MMS). Prandtl's boundary layer theory for the high-Reynolds flow of a viscous fluid over a solid body is an example of a boundary layer problem, and the semi-classical limit of quantum mechanics is an example of a multiple-scale problem.

We will begin by illustrating some basic issues in perturbation theory with simple algebraic equations.

1.2 Algebraic equations

The first two examples illustrate the distinction between regular and singular perturbation problems.

Example 1.1 Consider the cubic equation

$$x^3 - x + \varepsilon = 0. \tag{1.2}$$

We look for a solution of the form

$$x = x_0 + \varepsilon x_1 + \varepsilon^2 x_2 + O(\varepsilon^3). \tag{1.3}$$

Using this expansion in the equation, expanding, and equating coefficients of ε^n to zero, we get

$$\begin{aligned} &x_0^3 - x_0 = 0, \\ &3x_0^2 x_1 - x_1 + 1 = 0, \\ &3x_0 x_2 - x_2 + 3x_0 x_1^2 = 0. \end{aligned}$$

Note that we obtain a nonlinear equation for the leading order solution x_0, and nonhomogeneous linearized equations for the higher order corrections x_1, x_2,.... This structure is typical of many perturbation problems.

Solving the leading-order perturbation equation, we obtain the three roots

$$x_0 = 0, \pm 1.$$

Solving the first-order perturbation equation, we find that

$$x_1 = \frac{1}{1 - 3x_0^2}.$$

The corresponding solutions are

$$x = \varepsilon + O(\varepsilon^2), \qquad x = \pm 1 - \frac{1}{2}\varepsilon + O(\varepsilon^2).$$

Continuing in this way, we can obtain a convergent power series expansion about $\varepsilon = 0$ for each of the three distinct roots of (1.2). This result is typical of regular perturbation problems.

An alternative — but equivalent — method to obtain the perturbation series is to use the Taylor expansion

$$x(\varepsilon) = x(0) + \dot{x}(0)\varepsilon + \frac{1}{2!}\ddot{x}(0)\varepsilon^2 + \ldots,$$

where the dot denotes a derivative with respect to ε. To compute the coefficients, we repeatedly differentiate the equation with respect to ε and set $\varepsilon = 0$ in the result.

For example, setting $\varepsilon = 0$ in (1.2), and solving the resulting equation for $x(0)$, we get $x(0) = 0, \pm 1$. Differentiating (1.2) with respect to ε, we get

$$3x^2\dot{x} - \dot{x} + 1 = 0.$$

Setting $\varepsilon = 0$ and solving for $\dot{x}(0)$, we get the same answer as before.

Example 1.2 Consider the cubic equation

$$\varepsilon x^3 - x + 1 = 0. \tag{1.4}$$

Using (1.3) in (1.4), expanding, and equating coefficents of ε^n to zero, we get

$$\begin{aligned} -x_0 + 1 &= 0, \\ -x_1 + x_0^3 &= 0, \\ -x_2 + 3x_0^2 x_1 &= 0. \end{aligned}$$

Solving these equations, we find that $x_0 = 1$, $x_1 = 1$, ..., and hence

$$x(\varepsilon) = 1 + \varepsilon + O(\varepsilon^2). \tag{1.5}$$

We only obtain one solution because the cubic equation (1.4) degenerates to a linear equation at $\varepsilon = 0$. We missed the other two solutions because they approach infinity as $\varepsilon \to 0$. A change in the qualitative nature of the problem at the unperturbed value $\varepsilon = 0$ is typical of singular perturbation problems.

To find the other solutions, we introduce a rescaled variable y, where

$$x(\varepsilon) = \frac{1}{\delta(\varepsilon)} y(\varepsilon),$$

and $y = O(1)$ as $\varepsilon \to 0$. The scaling factor δ is to be determined. Using this equation in (1.4), we find that

$$\frac{\varepsilon}{\delta^3} y^3 - \frac{1}{\delta} y + 1 = 0. \tag{1.6}$$

In order to obtain a nontrivial solution, we require that at least two leading-order terms in this equation have the same order of magnitude. This is called the principle of *dominant balance*.

Balancing the first two terms, we find that*

$$\frac{\varepsilon}{\delta^3} = \frac{1}{\delta},$$

which implies that $\delta = \varepsilon^{1/2}$. The first two terms in (1.4) are then $O(\varepsilon^{-1/2})$, and the third term is $O(1)$, which is smaller. With this choice of δ, equation (1.6) becomes

$$y^3 - y + \varepsilon^{1/2} = 0.$$

*Nonzero constant factors can be absorbed into y.

Solving this equation in the same way as (1.2), we get the nonzero solutions

$$y = \pm 1 - \frac{1}{2}\varepsilon^{1/2} + O(\varepsilon).$$

The corresponding solutions for x are

$$x = \pm\frac{1}{\varepsilon^{1/2}} - \frac{1}{2} + O\left(\varepsilon^{1/2}\right).$$

The dominant balance argument illustrated here is useful in many perturbation problems. The corresponding limit, $\varepsilon \to 0$ with $x(\varepsilon) = O(\varepsilon^{-1/2})$, is called a *distinguished limit*.

There are two other two-term balances in (1.6). Balancing the second and third terms, we find that

$$\frac{1}{\delta} = 1$$

or $\delta = 1$. The first term is then $O(\varepsilon)$, so it is smaller than the other two terms. This dominant balance gives the solution in (1.5). Balancing the first and third terms, we find that

$$\frac{\varepsilon}{\delta^3} = 1,$$

or $\delta = \varepsilon^{1/3}$. In this case, the first and third terms are $O(1)$, but the second term is $O(\varepsilon^{-1/3})$. Thus, it is larger than the terms that balance, so we do not obtain a dominant balance or any new solutions.

In this example, no three-term dominant balance is possible as $\varepsilon \to 0$, but this can occur in other problems.

Example 1.3 A famous example of the effect of a perturbation on the solutions of a polynomial is Wilkinson's polynomial (1964),

$$(x-1)(x-2)\dots(x-20) = \varepsilon x^{19}.$$

The perturbation has a large effect on the roots even for small values of ε.

The next two examples illustrate some other features of perturbation theory.

Example 1.4 Consider the quadratic equation

$$(1-\varepsilon)x^2 - 2x + 1 = 0.$$

Suppose we look for a straightforward power series expansion of the form

$$x = x_0 + \varepsilon x_1 + O(\varepsilon^2).$$

We find that

$$x_0^2 - 2x_0 + 1 = 0,$$
$$2(x_0 - 1)x_1 = x_0^2.$$

Solving the first equation, we get $x_0 = 1$. The second equation then becomes $0 = 1$. It follows that there is no solution of the assumed form.

This difficulty arises because $x = 1$ is a repeated root of the unperturbed problem. As a result, the solution

$$x = \frac{1 \pm \varepsilon^{1/2}}{1 - \varepsilon}$$

does not have a power series expansion in ε, but depends on $\varepsilon^{1/2}$. An expansion

$$x = x_0 + \varepsilon^{1/2} x_1 + \varepsilon x_2 + O(\varepsilon^{3/2})$$

leads to the equations $x_0 = 1$, $x_1^2 = 1$, or

$$x = 1 \pm \varepsilon^{1/2} + O(\varepsilon)$$

in agreement with the exact solution.

Example 1.5 Consider the transcendental equation

$$xe^{-x} = \varepsilon. \tag{1.7}$$

As $\varepsilon \to 0^+$, there are two possibilities:

(a) $x \to 0$, which implies that $x = \varepsilon + \varepsilon^2 + O(\varepsilon^2)$;
(b) $e^{-x} \to 0$, when $x \to \infty$.

In the second case, x must be close to $\log 1/\varepsilon$.

To obtain an asymptotic expansion for the solution, we solve the equation iteratively using the idea that e^{-x} varies much more rapidly than x as $x \to 0$. Rewriting (1.7) as $e^{-x} = \varepsilon/x$ and taking logarithms, we get the equivalent equation

$$x = \log x + \log \frac{1}{\varepsilon}.$$

Thus solutions are fixed points of the function

$$f(x) = \log x + \log \frac{1}{\varepsilon}.$$

We then define iterates x_n, $n \in \mathbb{N}$, by

$$x_{n+1} = \log x_n + \log \frac{1}{\varepsilon},$$
$$x_1 = \log \frac{1}{\varepsilon}.$$

Defining

$$L_1 = \log \frac{1}{\varepsilon}, \qquad L_2 = \log \left(\log \frac{1}{\varepsilon} \right),$$

we find that

$$\begin{aligned} x_2 &= L_1 + L_2, \\ x_3 &= L_1 + \log(L_1 + L_2) \\ &= L_1 + L_2 + \frac{L_2}{L_1} + O\left(\left(\frac{L_2}{L_1}\right)^2\right). \end{aligned}$$

At higher orders, terms involving

$$L_3 = \log\left(\log\left(\log\frac{1}{\varepsilon}\right)\right),$$

and so on, appear.

The form of this expansion would be difficult to guess without using an iterative method. Note, however, that the successive terms in this asymptotic expansion converge very slowly as $\varepsilon \to 0$. For example, although $L_2/L_1 \to 0$ as $\varepsilon \to 0$, when $\varepsilon = 0.1$, $L_1 \approx 36$, $L_2 \approx 12$; and when $\varepsilon = 10^{-5}$, $L_1 \approx 19$, $L_2 \approx 1$.

1.3 Eigenvalue problems

Spectral perturbation theory studies how the spectrum of an operator is perturbed when the operator is perturbed. In general, this question is a difficult one, and subtle phenomena may occur, especially in connection with the behavior of the continuous spectrum of the operators. Here, we consider the simplest case of the perturbation in an eigenvalue.

Let $\mathcal{H}$ be a Hilbert space with inner product $\langle\cdot,\cdot\rangle$, and $A^\varepsilon : \mathcal{D}(A^\varepsilon) \subset \mathcal{H} \to \mathcal{H}$ a linear operator in $\mathcal{H}$, with domain $\mathcal{D}(A^\varepsilon)$, depending smoothly on a real parameter ε. We assume that:

(a) A^ε is self-adjoint, so that

$$\langle x, A^\varepsilon y\rangle = \langle A^\varepsilon x, y\rangle \qquad \text{for all } x, y \in \mathcal{D}(A^\varepsilon);$$

(b) A^ε has a smooth branch of simple eigenvalues $\lambda^\varepsilon \in \mathbb{R}$ with eigenvectors $x^\varepsilon \in \mathcal{H}$, meaning that

$$A^\varepsilon x^\varepsilon = \lambda^\varepsilon x^\varepsilon. \tag{1.8}$$

We will compute the perturbation in the eigenvalue from its value at $\varepsilon = 0$ when ε is small but nonzero.

A concrete example is the perturbation in the eigenvalues of a symmetric matrix. In that case, we have $\mathcal{H} = \mathbb{R}^n$ with the Euclidean inner product

$$\langle x, y\rangle = x^T y,$$

and $A^\varepsilon : \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation with an $n \times n$ symmetric matrix (a_{ij}^ε). The perturbation in the eigenvalues of a Hermitian matrix corresponds to $\mathcal{H} = \mathbb{C}^n$ with inner product $\langle x, y\rangle = \overline{x}^T y$. As we illustrate below with the Schrödinger equation of quantum mechanics, spectral problems for differential equations can be formulated in terms of unbounded operators acting in infinite-dimensional Hilbert spaces.

We use the expansions

$$\begin{aligned} A^\varepsilon &= A_0 + \varepsilon A_1 + \ldots + \varepsilon^n A_n + \ldots, \\ x^\varepsilon &= x_0 + \varepsilon x_1 + \ldots + \varepsilon^n x_n + \ldots, \\ \lambda^\varepsilon &= \lambda_0 + \varepsilon \lambda_1 + \ldots + \varepsilon^n \lambda_n + \ldots \end{aligned}$$

in the eigenvalue problem (1.8), equate coefficients of ε^n, and rearrange the result. We find that

$$(A_0 - \lambda_0 I)\, x_0 = 0, \tag{1.9}$$

$$(A_0 - \lambda_0 I)\, x_1 = -A_1 x_0 + \lambda_1 x_0, \tag{1.10}$$

$$(A_0 - \lambda_0 I)\, x_n = \sum_{i=1}^{n} \left\{ -A_i x_{n-i} + \lambda_i x_{n-i} \right\}. \tag{1.11}$$

Assuming that $x_0 \neq 0$, equation (1.9) implies that λ_0 is an eigenvalue of A_0 and x_0 is an eigenvector. Equation (1.10) is then a singular equation for x_1. The following proposition gives a simple, but fundamental, solvability condition for this equation.

Proposition 1.6 Suppose that A is a self-adjoint operator acting in a Hilbert space $\mathcal{H}$ and $\lambda \in \mathbb{R}$. If $z \in \mathcal{H}$, a necessary condition for the existence of a solution $y \in \mathcal{H}$ of the equation

$$(A - \lambda I)\, y = z \tag{1.12}$$

is that

$$\langle x, z\rangle = 0,$$

for every eigenvector x of A with eigenvalue λ.

Proof. Suppose $z \in \mathcal{H}$ and y is a solution of (1.12). If x is an eigenvector of A, then using (1.12) and the self-adjointness of $A - \lambda I$, we find that

$$\begin{aligned} \langle x, z\rangle &= \langle x, (A - \lambda I)\, y\rangle \\ &= \langle (A - \lambda I)\, x, y\rangle \\ &= 0. \end{aligned}$$

□

In many cases, the necesary solvability condition in this proposition is also sufficient, and then we say that $A - \lambda I$ satisfies the Fredholm alternative; for example, this is true in the finite-dimensional case, or when A is an elliptic partial differential operator.

Since A_0 is self-adjoint and λ_0 is a simple eigenvalue with eigenvector x_0, equation (1.12) it is solvable for x_1 only if the right hand side is orthogonal to x_0, which implies that

$$\lambda_1 = \frac{\langle x_0, A_1 x_0 \rangle}{\langle x_0, x_0 \rangle}.$$

This equation gives the leading order perturbation in the eigenvalue, and is the most important result of the expansion.

Assuming that the necessary solvability condition in the proposition is sufficient, we can then solve (1.10) for x_1. A solution for x_1 is not unique, since we can add to it an arbitrary scalar multiple of x_0. This nonuniqueness is a consequence of the fact that if x^ε is an eigenvector of A^ε, then $c^\varepsilon x^\varepsilon$ is also a solution for any scalar c^ε. If

$$c^\varepsilon = 1 + \varepsilon c_1 + O(\varepsilon^2)$$

then

$$c^\varepsilon x^\varepsilon = x_0 + \varepsilon \left(x_1 + c_1 x_0 \right) + O(\varepsilon^2).$$

Thus, the addition of $c_1 x_0$ to x_1 corresponds to a rescaling of the eigenvector by a factor that is close to one.

This expansion can be continued to any order. The solvability condition for (1.11) determines λ_n, and the equation may then be solved for x_n, up to an arbitrary vector $c_n x_0$. The appearance of singular problems, and the need to impose solvabilty conditions at each order which determine parameters in the expansion and allow for the solution of higher order corrections, is a typical structure of many pertubation problems.

1.3.1 *Quantum mechanics*

One application of this expansion is in quantum mechanics, where it can be used to compute the change in the energy levels of a system caused by a perturbation in its Hamiltonian.

The Schrödinger equation of quantum mechanics is

$$i\hbar \psi_t = H\psi.$$

Here t denotes time and $\hbar$ is Planck's constant. The wavefunction $\psi(t)$ takes values in a Hilbert space $\mathcal{H}$, and H is a self-adjoint linear operator acting in $\mathcal{H}$ with the dimensions of energy, called the Hamiltonian.

Energy eigenstates are wavefunctions of the form

$$\psi(t) = e^{-iEt/\hbar}\varphi,$$

where $\varphi \in \mathcal{H}$ and $E \in \mathbb{R}$. It follows from the Schrödinger equation that

$$H\varphi = E\varphi.$$

Hence E is an eigenvalue of H and φ is an eigenvector. One of Schrödinger's motivations for introducing his equation was that eigenvalue problems led to the experimentally observed discrete energy levels of atoms.

Now suppose that the Hamiltonian

$$H^\varepsilon = H_0 + \varepsilon H_1 + O(\varepsilon^2)$$

depends smoothly on a parameter ε. Then, rewriting the previous result, we find that the corresponding simple energy eigenvalues (assuming they exist) have the expansion

$$E^\varepsilon = E_0 + \varepsilon \frac{\langle \varphi_0, H_1\varphi_0\rangle}{\langle \varphi_0, \varphi_0\rangle} + O(\varepsilon^2)$$

where φ_0 is an eigenvector of H_0.

For example, the Schrödinger equation that describes a particle of mass m moving in $\mathbb{R}^d$ under the influence of a conservative force field with potential $V : \mathbb{R}^d \to \mathbb{R}$ is

$$i\hbar\psi_t = -\frac{\hbar^2}{2m}\Delta\psi + V\psi.$$

Here, the wavefunction $\psi(x,t)$ is a function of a space variable $x \in \mathbb{R}^d$ and time $t \in \mathbb{R}$. At fixed time t, we have $\psi(\cdot,t) \in L^2(\mathbb{R}^d)$, where

$$L^2(\mathbb{R}^d) = \left\{u : \mathbb{R}^d \to \mathbb{C} \mid u \text{ is measurable and } \textstyle\int_{\mathbb{R}^d} |u|^2\, dx < \infty\right\}$$

is the Hilbert space of square-integrable functions with inner-product

$$\langle u, v\rangle = \int_{\mathbb{R}^d} \overline{u}(x)v(x)\, dx.$$

The Hamiltonian operator $H : \mathcal{D}(H) \subset \mathcal{H} \to \mathcal{H}$, with domain $\mathcal{D}(H)$, is given by

$$H = -\frac{\hbar^2}{2m}\Delta + V.$$

If u, v are smooth functions that decay sufficiently rapidly at infinity, then Green's theorem implies that

$$\begin{aligned}\langle u, Hv\rangle &= \int_{\mathbb{R}^d} \overline{u}\left(-\frac{\hbar^2}{2m}\Delta v + Vv\right) dx \\ &= \int_{\mathbb{R}^d} \left\{\frac{\hbar^2}{2m}\nabla\cdot(v\nabla\overline{u} - \overline{u}\nabla v) - \frac{\hbar^2}{2m}(\Delta\overline{u})v + V\overline{u}v\right\} dx\end{aligned}$$

$$\begin{aligned} &= \int_{\mathbb{R}^d} \overline{\left(-\frac{\hbar^2}{2m}\Delta u + Vu\right)} v \, dx \\ &= \langle Hu, v\rangle. \end{aligned}$$

Thus, this operator is formally self-adjoint. Under suitable conditions on the potential V, the operator can be shown to be self-adjoint with respect to an appropriately chosen domain.

Now suppose that the potential V^ε depends on a parameter ε, and has the expansion

$$V^\varepsilon(x) = V_0(x) + \varepsilon V_1(x) + O(\varepsilon^2).$$

The perturbation in a simple energy eigenvalue

$$E^\varepsilon = E_0 + \varepsilon E_1 + O(\varepsilon^2),$$

assuming one exists, is given by

$$E_1 = \frac{\int_{\mathbb{R}^d} V_1(x)|\varphi_0(x)|^2 \, dx}{\int_{\mathbb{R}^d} |\varphi_0(x)|^2 \, dx},$$

where $\varphi_0 \in L^2(\mathbb{R}^d)$ is an unperturbed energy eigenfunction that satisfies

$$-\frac{\hbar^2}{2m}\Delta\varphi_0 + V_0\varphi_0 = E_0\varphi_0.$$

Example 1.7 The one-dimensional simple harmonic oscillator has potential

$$V_0(x) = \frac{1}{2}kx^2.$$

The eigenvalue problem

$$-\frac{\hbar^2}{2m}\varphi'' + \frac{1}{2}kx^2\varphi = E\varphi, \qquad \varphi \in L^2(\mathbb{R})$$

is exactly solvable. The energy eigenvalues are

$$E_n = \hbar\omega\left(n + \frac{1}{2}\right) \qquad n = 0, 1, 2, \ldots,$$

where

$$\omega = \sqrt{\frac{k}{m}}$$

is the frequency of the corresponding classical oscillator. The eigenfunctions are

$$\varphi_n(x) = H_n(\alpha x)e^{-\alpha^2 x^2/2},$$

where H_n is the nth Hermite polynomial,

$$H_n(\xi) = (-1)^n e^{\xi^2}\frac{d^n}{d\xi^n}e^{-\xi^2},$$

and the constant α, with dimensions of 1/length, is given by

$$\alpha^2 = \frac{\sqrt{mk}}{\hbar}.$$

The energy levels E_n^ε of a slightly anharmonic oscillator with potential

$$V^\varepsilon(x) = \frac{1}{2}kx^2 + \varepsilon\frac{k}{\alpha^2}W(\alpha x) + O(\varepsilon^2) \qquad \text{as } \varepsilon \to 0^+$$

where $\varepsilon > 0$ have the asymptotic behavior

$$E_n^\varepsilon = \hbar\omega\left\{n + \frac{1}{2} + \varepsilon\Delta_n + O(\varepsilon^2)\right\} \qquad \text{as } \varepsilon \to 0^+,$$

where

$$\Delta_n = \frac{\int W(\xi)H_n^2(\xi)e^{-\xi^2}\,d\xi}{\int H_n^2(\xi)e^{-\xi^2}\,d\xi}.$$

For an extensive and rigorous discussion of spectral perturbation theory for linear operators, see [9].

1.4 Nondimensionalization

The numerical value of any quantity in a mathematical model is measured with respect to a system of units (for example, meters in a mechanical model, or dollars in a financial model). The units used to measure a quantity are arbitrary, and a change in the system of units (for example, to feet or yen, at a fixed exchange rate) cannot change the predictions of the model. A change in units leads to a rescaling of the quantities. Thus, the independence of the model from the system of units corresponds to a scaling invariance of the model. In cases when the zero point of a unit is arbitrary, we also obtain a translational invariance, but we will not consider translational invariances here.

Suppose that a model involves quantities $(a_1, a_2, \ldots, a_n)$, which may include dependent and independent variables as well as parameters. We denote the dimension of a quantity a by $[a]$. A fundamental system of units is a minimal set of independent units, which we denote symbolically by $(d_1, d_2, \ldots, d_r)$. Different fundamental system of units can be used, but given a fundamental system of units any other derived unit may be constructed uniquely as a product of powers of the fundamental units, so that

$$[a] = d_1^{\alpha_1}d_2^{\alpha_2}\ldots d_r^{\alpha_r} \tag{1.13}$$

for suitable exponents $(\alpha_1, \alpha_2, \ldots, \alpha_r)$.

Example 1.8 In mechanical problems, a fundamental set of units is d_1 = mass, d_2 = length, d_3 = time, or $d_1 = M$, $d_2 = L$, $d_3 = T$, for short. Then velocity

$V = L/T$ and momentum $P = ML/T$ are derived units. We could use instead momentum P, velocity V, and time T as a fundamental system of units, when mass $M = P/V$ and length $L = VT$ are derived units. In problems involving heat flow, we may introduce temperature (measured, for example, in degrees Kelvin) as another fundamental unit, and in problems involving electromagnetism, we may introduce current (measured, for example, in Ampères) as another fundamental unit.

The invariance of a model under the change in units $d_j \mapsto \lambda_j d_j$ implies that it is invariant under the scaling transformation

$$a_i \to \lambda_1^{\alpha_{1,i}} \lambda_2^{\alpha_{2,i}} \dots \lambda_r^{\alpha_{r,i}} a_i \qquad i = 1, \dots, n$$

for any $\lambda_1, \dots \lambda_r > 0$, where

$$[a_i] = d_1^{\alpha_{1,i}} d_2^{\alpha_{2,i}} \dots d_r^{\alpha_{r,i}}. \tag{1.14}$$

Thus, if

$$a = f(a_1, \dots, a_n)$$

is any relation between quantities in the model with the dimensions in (1.13) and (1.14), then f has the scaling property that

$$\lambda_1^{\alpha_1} \lambda_2^{\alpha_2} \dots \lambda_r^{\alpha_r} f(a_1, \dots, a_n) = f\left(\lambda_1^{\alpha_{1,1}} \lambda_2^{\alpha_{2,1}} \dots \lambda_r^{\alpha_{r,1}} a_1, \dots, \lambda_1^{\alpha_{1,n}} \lambda_2^{\alpha_{2,n}} \dots \lambda_r^{\alpha_{r,n}} a_n\right).$$

A particular consequence of the invariance of a model under a change of units is that any two quantities which are equal must have the same dimensions. This fact is often useful in finding the dimension of some quantity.

Example 1.9 According to Newton's second law,

$$\text{force} = \text{rate of change of momentum with respect to time.}$$

Thus, if F denotes the dimension of force and P the dimension of momentum, then $F = P/T$. Since $P = MV = ML/T$, we conclude that $F = ML/T^2$ (or mass $\times$ acceleration).

Example 1.10 In fluid mechanics, the shear viscosity μ of a Newtonian fluid is the constant of proportionality that relates the viscous stress tensor $\mathcal{T}$ to the velocity gradient $\nabla \mathbf{u}$:

$$\mathcal{T} = \frac{1}{2}\mu\left(\nabla \mathbf{u} + \nabla \mathbf{u}^T\right).$$

Stress has dimensions of force/area, so

$$[\mathcal{T}] = \frac{ML}{T^2}\frac{1}{L^2} = \frac{M}{LT^2}.$$

The velocity gradient $\nabla \mathbf{u}$ has dimensions of velocity/length, so

$$[\nabla \mathbf{u}] = \frac{L}{T}\frac{1}{L} = \frac{1}{T}.$$

Equating dimensions, we find that

$$[\mu] = \frac{M}{LT}.$$

We can also write $[\mu] = (M/L^3)(L^2/T)$. It follows that if ρ_0 is the density of the fluid, and $\mu = \rho_0 \nu$, then

$$[\nu] = \frac{L^2}{T}.$$

Thus ν, which is called the kinematical viscosity, has the dimensions of diffusivity. Physically it is the diffusivity of momentum. For example, in time T, viscous effects lead to the diffusion of momentum over a length scale of the order $\sqrt{\nu T}$.

Scaling invariance implies that we can reduce the number of quantities appearing in the problem by the introduction of dimensionless variables. Suppose that $(a_1, \ldots, a_r)$ are a set of (nonzero) quantities whose dimensions form a fundamental system of units. We denote the remaining quantities in the model by $(b_1, \ldots, b_m)$, where $r + m = n$. Then for suitable exponents $(\beta_{1,i}, \ldots, \beta_{r,i})$, the quantity

$$\Pi_i = \frac{b_i}{a_1^{\beta_{1,i}} \ldots a_r^{\beta_{r,i}}}$$

is dimensionless, meaning that it is invariant under the scaling transformations induced by changes in units. Such dimensionless quantities can often be interpreted as the ratio of two quantities of the same dimension appearing in the problem (such as a ratio of lengths, times, diffusivities, and so on). Perturbation methods are typically applicable when one or more of these dimensionless quantities is small or large.

Any relationship of the form

$$b = f(a_1, \ldots, a_r, b_1, \ldots, b_m)$$

is equivalent to a relation

$$\Pi = f(1, \ldots, 1, \Pi_1, \ldots, \Pi_m).$$

Thus, the introduction of dimensionless quantities reduces the number of variables in the problem by the number of fundamental units involved in the problem. In many cases, nondimensionalization leads to a reduction in the number of parameters in the problem to a minimal number of dimensionless parameters. In some cases, one may be able to use dimensional arguments to obtain the form of self-similar solutions.

Example 1.11 Consider the following IVP for the Green's function of the heat equation in $\mathbb{R}^d$:

$$\begin{aligned} u_t &= \nu \Delta u, \\ u(x,0) &= E\delta(x). \end{aligned}$$

Here δ is the delta-function. The dimensioned parameters in this problem are the diffusivity ν and the energy E of the point source. The only length and times scales are those that come from the independent variables (x,t), so the solution is self-similar.

We have $[u] = \theta$, where θ denotes temperature, and, since

$$\int_{\mathbb{R}^d} u(x,0)\,dx = E,$$

we have $[E] = \theta L^d$. Dimensional analysis and the rotational invariance of the Laplacian Δ imply that

$$u(x,t) = \frac{E}{(\nu t)^{d/2}} f\left(\frac{|x|}{\sqrt{\nu t}}\right).$$

Using this expression for $u(x,t)$ in the PDE, we get an ODE for $f(\xi)$,

$$f'' + \left(\frac{\xi}{2} + \frac{d-1}{\xi}\right) f' + \frac{d}{2} f = 0.$$

We can rewrite this equation as a first-order ODE for $f' + \frac{\xi}{2} f$,

$$\left(f' + \frac{\xi}{2} f\right)' + \frac{d-1}{\xi}\left(f' + \frac{\xi}{2} f\right) = 0.$$

Solving this equation, we get

$$f' + \frac{\xi}{2} f = \frac{b}{\xi^{d-1}},$$

where b is a constant of integration. Solving for f, we get

$$f(\xi) = a e^{-\xi^2/4} + b e^{-\xi^2/4} \int \frac{e^{-\xi^2}}{\xi^{d-1}}\,d\xi,$$

where a s another constant of integration. In order for f to be integrable, we must set $b = 0$. Then

$$u(x,t) = \frac{aE}{(\nu t)^{d/2}} \exp\left(-\frac{|x|^2}{4\nu t}\right).$$

Imposing the requirement that

$$\int_{\mathbb{R}^d} u(x,t)\,dx = E,$$

and using the standard integral

$$\int_{\mathbb{R}^d} \exp\left(-\frac{|x|^2}{2c}\right) dx = (2\pi c)^{d/2},$$

we find that $a = (4\pi)^{-d/2}$, and

$$u(x,t) = \frac{E}{(4\pi\nu t)^{d/2}} \exp\left(-\frac{|x|^2}{4\nu t}\right).$$

Example 1.12 Consider a sphere of radius L moving through a fluid with constant speed U. A primary quantity of interest is the total drag force D exerted by the fluid on the sphere. We assume that the fluid is incompressible, which is a good approximation if the flow speed U is much less than the speed of sound in the fluid. The fluid properties are then determined by the density ρ_0 and the kinematic viscosity ν. Hence,

$$D = f(U, L, \rho_0, \nu).$$

Since the drag D has the dimensions of force (ML/T^2), dimensional analysis implies that

$$D = \rho_0 U^2 L^2 F\left(\frac{UL}{\nu}\right).$$

Thus, the dimensionless drag

$$\frac{D}{\rho_0 U^2 L^2} = F(\mathrm{Re})$$

is a function of the Reynold's number

$$\mathrm{Re} = \frac{UL}{\nu}.$$

The function F has a complicated dependence on Re that is difficult to compute explicitly. For example, F changes rapidly near Reynolds numbers for which the flow past the sphere becomes turbulent. Nevertheless, experimental measurements agree very well with the result of this dimensionless analysis (see Figure 1.9 in [1], for example).

Dimensional analysis leads to continuous scaling symmetries. These scaling symmetries are not the only continuous symmetries possessed by differential equations. The theory of Lie groups and Lie algebras provides a systematic method for computing all continuous symmetries of a given differential equation [13]. Lie originally introduced the notions of Lie groups and Lie algebras precisely for this purpose.

Example 1.13 The full group of symmetries of the one-dimensional heat equation

$$u_t = u_{xx}$$

is generated by the following transformations [13]:

$$
\begin{aligned}
&u(x,t) \mapsto u(x-\alpha,t),\\
&u(x,t) \mapsto u(x,t-\beta),\\
&u(x,t) \mapsto \gamma u(x,t),\\
&u(x,t) \mapsto u(\delta x,\delta^2 t),\\
&u(x,t) \mapsto e^{-\varepsilon x+\varepsilon^2 t}u(x-2\varepsilon t,t),\\
&u(x,t) \mapsto \frac{1}{\sqrt{1+4\eta t}}\exp\left[\frac{-\eta x^2}{1+4\eta t}\right]u\left(\frac{x}{1+4\eta t},\frac{t}{1+4\eta t}\right),\\
&u(x,t) \mapsto u(x,t)+v(x,t),
\end{aligned}
$$

where $(\alpha,\ldots,\eta)$ are constants, and $v(x,t)$ is an arbitrary solution of the heat equation. The scaling symmetries involving γ and δ can be deduced by dimensional arguments, but the symmetries involving ε and η cannot.

For further discussion of dimensional analysis and self-similar solutions, see [1].