THE WORLD ACCORDING TO QUANTUM MECHANICS

Why the Laws of Physics Make Perfect Sense After All

Ulrich Mohrhoff

World Scientific

THE WORLD ACCORDING TO QUANTUM MECHANICS

Why the Laws of Physics Make Perfect Sense After All

Ulrich Mohrhoff

World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.
5 Toh Tuck Link, Singapore 596224
USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601
UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

THE WORLD ACCORDING TO QUANTUM MECHANICS
Why the Laws of Physics Make Perfect Sense After All

ISBN-13 978-981-4293-37-2
ISBN-10 981-4293-37-7

Printed in Singapore.

Contents

A Closer Look 75

This page intentionally left blank

PART 1

Overview

This page intentionally left blank

Chapter 1

Probability: Basic concepts and theorems

The mathematical formalism of quantum mechanics is a probability calculus. The probability algorithms it places at our disposal—state vectors, wave functions, density matrices, statistical operators—all serve the same purpose, which is to calculate the probabilities of measurement outcomes. That's reason enough to begin by putting together what we already know and what we need to know about probabilities.

1.1 The principle of indifference

Probability is a measure of likelihood ranging from 0 to 1. If an event has a probability equal to 1, it is certain that it will happen; if it has a probability equal to 0, it is certain that it will not happen; and if it has a probability equal to 1/2, then it is as likely as not that it will happen.

Tossing a fair coin yields heads with probability 1/2. Casting a fair die yields any given natural number between 1 and 6 with probability 1/6. These are just two examples of the *principle of indifference*, which states:

If there are n mutually exclusive and jointly exhaustive possibilities (or possible events), and if we have no reason to consider any one of them more likely than any other, then each possibility should be assigned a probability equal to $1/n$.

Saying that events are *mutually exclusive* is the same as saying that at most one of them happens. Saying that events are *jointly exhaustive* is the same as saying that at least one of them happens.

1.2 Subjective probabilities versus objective probabilities

There are two kinds of situations in which we may have no reason to consider one possibility more likely than another. In situations of the first kind, there are objective matters of fact that would make it certain, *if we knew them*, that a particular event will happen, but we don't know any of the relevant matters of fact. The probabilities we assign in this case, or whenever we know some but not all relevant facts, are in an obvious sense *subjective*. They are *ignorance* probabilities. They have everything to do with our (lack of) *knowledge* of relevant facts, but nothing with the *existence* of relevant facts. Therefore they are also known as *epistemic* probabilities.

In situations of the second kind, there are no objective matters of fact that would make it certain that a particular event will happen. There may not even be objective matters of fact that would make it more likely that one event will occur rather than another. There isn't any relevant fact that we are ignorant of. The probabilities we assign in this case are neither subjective nor epistemic. They deserve to be considered *objective*. Quantum-mechanical probabilities are essentially of this kind.

Until the advent of quantum mechanics, all probabilities were thought to be subjective. This had two unfortunate consequences. The first is that probabilities came to be thought of as something *intrinsically* subjective. The second is that something that was not a probability at all—namely, a *relative frequency*—came to be called an "objective probability."

1.3 Relative frequencies

Relative frequencies are useful in that they allow us to measure the likelihood of possible events, at least approximately, provided that trials can be repeated under conditions that are identical in all relevant respects. We obviously cannot measure the likelihood of heads by tossing a single coin. But since we can toss a coin any number of times, we can count the number N_H of heads and the number N_T of tails obtained in N tosses and calculate the fraction $f_N^H = N_H/N$ of heads and the fraction $f_N^T = N_T/N$ of tails. And we can expect the difference $|N_H - N_T|$ to increase significantly slower than the sum $N = N_H + N_T$, so that

$$\lim_{N\to\infty} \frac{|N_H - N_T|}{N_H + N_T} = \lim_{N\to\infty} |f_N^H - f_N^T| = 0\,. \tag{1.1}$$

In other words, we can expect the relative frequencies f_N^H and f_N^T to tend to the probabilities p_H of heads and p_T of tails, respectively:

$$p_H = \lim_{N\to\infty} \frac{N_H}{N}\,, \qquad p_T = \lim_{N\to\infty} \frac{N_T}{N}\,. \tag{1.2}$$

1.4 Adding and multiplying probabilities

Suppose you roll a (six-sided) die. And suppose you win if you throw either a 1 or a 6 (no matter which). Since there are six equiprobable outcomes, two of which cause you to win, your chances of winning are 2/6. In this example it is appropriate to *add* probabilities:

$$p(1 \vee 6) = p(1) + p(6) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}\,. \tag{1.3}$$

The symbol $\vee$ means "or." The general rule is this:

Sum rule. Let $\mathcal{W}$ be a set of w mutually exclusive and jointly exhaustive events (for instance, the possible outcomes of a measurement), and let $\mathcal{U}$ be a subset of $\mathcal{W}$ containing a smaller number u of events: $\mathcal{U} \subset \mathcal{W}$, $u < w$. The probability $p(\mathcal{U})$ that one of the events $e_1, \dots, e_u$ in $\mathcal{U}$ takes place (no matter which) is the sum $p_1 + \cdots + p_u$ of the respective probabilities of these events.

One nice thing about relative frequencies is that they make a rule such as this virtually self-evident. To demonstrate this, let N be the total number of trials—think coin tosses or measurements. Let N_k be the total number of trials with outcome e_k, and let $N(\mathcal{U})$ be the total number of trials with an outcome in $\mathcal{U}$. As N tends to infinity, N_k/N tends to p_k and $N(\mathcal{U})/\mathcal{N}$ tends to $p(\mathcal{U})$. But

$$\frac{N(\mathcal{U})}{N} = \frac{N_1 + \cdots + N_u}{N} = \frac{N_1}{N} + \cdots + \frac{N_u}{N}\,, \tag{1.4}$$

and in the limit $N \to \infty$ this becomes

$$p(\mathcal{U}) = p_1 + \cdots + p_u\,. \tag{1.5}$$

Suppose now that you roll two dice. And suppose that you win if your total equals 12. Since there are now 6×6 equiprobable outcomes, only one of which causes you to win, your chances of winning are $1/(6 \times 6)$. In this example it is appropriate to *multiply* probabilities:

$$p(6 \wedge 6) = p(6) \times p(6) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}\,. \tag{1.6}$$

The symbol $\wedge$ means "and." Here is the general rule:

Product rule. The *joint probability* $p(e_1 \wedge \cdots \wedge e_v)$ of v *independent* events $e_1, \ldots, e_v$ (that is, the probability with which *all* of them happen) is the product of the probabilities $p(e_1), \ldots, p(e_v)$ of the individual events.

It must be stressed that the product rule only applies to independent events. Saying that two events a, b are *independent* is the same as saying that the probability of a is independent of whether or not b happens, and *vice versa.*

As an illustration of the product rule for two independent events, let $a_1, \ldots, a_J$ be mutually exclusive and jointly exhaustive events (think of the possible outcomes of a measurement of a variable A), and let $p^a_1, \ldots, p^a_J$ be the corresponding probabilities. Let $b_1, \ldots, b_K$ be a second such set of events with corresponding probabilities $p^b_1, \ldots, p^b_K$. Now draw a 1×1 square with coordinates x, y ranging from 0 to 1. Partition it horizontally into J strips of respective width p^a_j. Partition it vertically into K strips of respective width p^b_k. You now have a square partitioned into $J \times K$ rectangles with respective areas $p^a_j \times p^b_k$. Since a joint measurement of A and B is equivalent to throwing a dart in such a way that it hits a random position (x, y) within the square, the joint probability $p(a_j \wedge b_k)$ equals the corresponding area.

Problem 1.1. *We have seen that the probability of obtaining a total of 12 when rolling a pair of dice is 1/36. What is the probability of obtaining a total of (a) 11, (b) 10, (c) 9?*

Problem 1.2. $(*)$[1] *In 1999, Sally Clark was convicted of murdering her first two babies, which died in their sleep of sudden infant death syndrome. She was sent to prison to serve two life sentences for murder, essentially on the testimony of an "expert" who told the jury it was too improbable that two children in one family would die of this rare syndrome, which has a probability of 1/8,500. After over three years in prison, and five years of fighting in the legal system, Sally was cleared by a Court of Appeal, and another two and a half years later, the "expert" pediatrician Sir Roy Meadow was found guilty of serious professional misconduct. Amazingly, during the trial nobody raise the objection that an expert pediatrician was not likely to be an expert statistician. Meadow had argued that the probability of two sudden infant deaths in the same family was* $(1/8,500) \times (1/8,500) = 1/72,250,000$. *Explain why he was so terribly wrong.*

[1] A star indicates that a solution or a hint is provided in Appendix A.

1.5 Conditional probabilities and correlations

If the events a_j and b_k are *not* independent, we must distinguish between *marginal probabilities*, which are assigned to the possible outcomes of either measurement without taking account of the outcome of the other measurement, and *conditional probabilities*, which are assigned to the possible outcomes of either measurement depending on the outcome of the other measurement. If a_j and b_k are not independent, their joint probability is

$$p(a_j \wedge b_k) = p(b_k|a_j)\, p(a_j) = p(a_j|b_k)\, p(b_k)\,, \tag{1.7}$$

where $p(a_j)$ and $p(b_k)$ are marginal probabilities, while $p(b_k|a_j)$ is the probability of b_k conditional on the outcome a_j and $p(a_j|b_k)$ is the probability of a_j conditional on the outcome b_k. This gives us the useful relation

$$p(b|a) = \frac{p(a \wedge b)}{p(a)}\,. \tag{1.8}$$

Another useful rule is

$$p(a) = p(a|b)\, p(b) + p(a|\overline{b})\, p(\overline{b})\,, \tag{1.9}$$

where b and $\overline{b}$ are two mutually exclusive and jointly exhaustive events. (To obtain $\overline{b}$ is to obtain any outcome other than b.) The validity of this rule is again readily established with the help of relative frequencies. We obviously have that

$$\frac{N(a)}{N} = \frac{N(a \wedge b)}{N} + \frac{N(a \wedge \overline{b})}{N} = \frac{N(a \wedge b)}{N(b)}\frac{N(b)}{N} + \frac{N(a \wedge \overline{b})}{N(\overline{b})}\frac{N(\overline{b})}{N}\,, \tag{1.10}$$

where N is the number of joint measurements of two variables, one with the possible outcome a and one with the possible outcome b. In the limit $N \to \infty$, $N(a)/N$ (the left-hand side of Eq. 1.10) tends to the marginal probability $p(a)$, while the right-hand side of this equation tends to the right-hand side of Eq. (1.9), as will be obvious from a glance at Eq. (1.8).

An important concept is that of (probabilistic) correlation. Two events a, b are *correlated* just in case that $p(a|b) \neq p(a|\overline{b})$. Specifically, a and b are *positively* correlated if $p(a|b) > p(a|\overline{b})$, and they are *negatively* correlated if $p(a|b) < p(a|\overline{b})$. Saying that a and b are *independent* is thus the same as saying that they are *uncorrelated*, in which case $p(a|b) = p(a|\overline{b}) = p(a)$.

Problem 1.3. ($*$) Let's Make a Deal *was a famous game show hosted by Monty Hall. In it a player was to open one of three doors. Behind one door there was the Grand Prize (for example, a car). Behind the other doors* *there were booby prizes (say, goats). After the player had chosen a door, the host opened a different door, revealing a goat, and offered the player the opportunity of choosing the other closed door. Should the player accept the offer or should he stick with his first choice? Does it make a difference?*

Problem 1.4. (∗) *Which of the following statements do you think is true? (i) Event A happens more frequently because it is more likely. (ii) Event A is more likely because it happens more frequently.*

Problem 1.5. (∗) *Suppose we have a 99% accurate test for a certain disease. And suppose that a person picked at random from the population tests postive. What is the probability that this person actually has the disease?*

1.6 Expectation value and standard deviation

Another two important concepts associated with a probability distribution are the *expected/expectation value* (or *mean*) and the *standard deviation* (or *root mean square deviation from the mean*).

The expected value associated with the measurement of an observable with K possible outcomes v_k and corresponding probabilities $p(v_k)$ is

$$\langle v \rangle \overset{\text{Def}}{=} \sum_{k=1}^{K} p(v_k)\, v_k \,. \tag{1.11}$$

Note that the expected value doesn't have to be one of the possible outcomes. The expected value associated with the roll of a die, for instance, equals 3.5.

To calculate the *rms* deviation from the mean, Δv, we first calculate the squared deviations from the mean, $(v_k - \langle v \rangle)^2$, then we calculate their mean, and finally we take the root:

$$\Delta v = \sqrt{\sum_{k=1}^{K} p(v_k)(v_k - \langle v \rangle)^2} \,. \tag{1.12}$$

The standard deviation of a random variable V with possible values v_k is an important measure—albeit not the only one—of the variability or spread of V.

Problem 1.6. (∗) *Calculate the standard deviation for the sum obtained by rolling two dice.*

Chapter 2

A (very) brief history of the "old" theory

2.1 Planck

Quantum physics started out as a rather desperate measure to avoid some of the spectacular failures of what we now call "classical physics." The story begins with the discovery by Max Planck, in 1900, of the law that perfectly describes the radiation spectrum of a glowing hot object. (One of the things predicted by classical physics was that you would get blinded by ultraviolet light if you looked at the burner of your stove.) At first it was just a fit to the data—"a fortuitous guess at an interpolation formula," as Planck himself described his radiation law. It was only weeks later that this formula was found to imply the quantization of energy in the emission of electromagnetic radiation, and thus to be irreconcilable with classical physics. According to classical theory, a glowing hot object emits energy *continuously*. Planck's formula implies that it emits energy in *discrete* quantities proportional to the *frequency* ν of the radiation:

$$E = h\nu\,, \tag{2.1}$$

where $h = 6.626069 \times 10^{-34}\,\mathrm{Js}$ is the *Planck constant*. Often it is more convenient to use the *reduced Planck constant* $\hbar = h/2\pi$ ("h bar"), which allows us to write

$$E = \hbar\omega\,, \tag{2.2}$$

where the *angular frequency* $\omega = 2\pi\nu$ replaces ν.

2.2 Rutherford

In 1911, Ernest Rutherford proposed a model of the atom that was based on experiments conducted by Hans Geiger and Ernest Marsden. Geiger and Marsden had directed a beam of alpha particles (helium nuclei) at a thin gold foil. As expected, most of the alpha particles were deflected by at most a few degrees. Yet a tiny fraction of the particles were deflected through angles much larger than 90 degrees. In Rutherford's own words [Cassidy *et al.* (2002)],

> It was almost as incredible as if you fired a 15-inch shell at a piece of tissue paper and it came back and hit you. On consideration, I realized that this scattering backward must be the result of a single collision, and when I made calculations I saw that it was impossible to get anything of that order of magnitude unless you took a system in which the greater part of the mass of the atom was concentrated in a minute nucleus.

The resulting model, which described the atom as a miniature solar system, with electrons orbiting the nucleus the way planets orbit a star, was however short-lived. Classical electromagnetic theory predicts that an orbiting electron will radiate away its energy and spiral into the nucleus in less than a nanosecond. This was the worst quantitative failure in the history of physics, under-predicting the lifetime of hydrogen by at least forty orders of magnitude. (This figure is based on the experimentally established lower bound on the proton's lifetime.)

2.3 Bohr

In 1913, Niels Bohr postulated that the angular momentum L of an orbiting atomic electron was quantized: its possible values are integral multiples of the reduced Planck constant:

$$L = n\hbar, \quad n = 1, 2, 3 \ldots . \tag{2.3}$$

Observe that angular momentum and Planck's constant are measured in the same units.

Bohr's postulate not only explained the stability of atoms but also accounted for the by then well-established fact that atoms absorb and emit electromagnetic radiation only at specific frequencies. What is more, it enabled Bohr to calculate with remarkable accuracy the spectrum of atomic hydrogen—the particular frequencies at which it absorbs and emits light (visible as well as infrared and ultraviolet).

Apart from his quantization postulate, Bohr's reasoning at the time remained completely classical. Let us assume with Bohr that the electron's

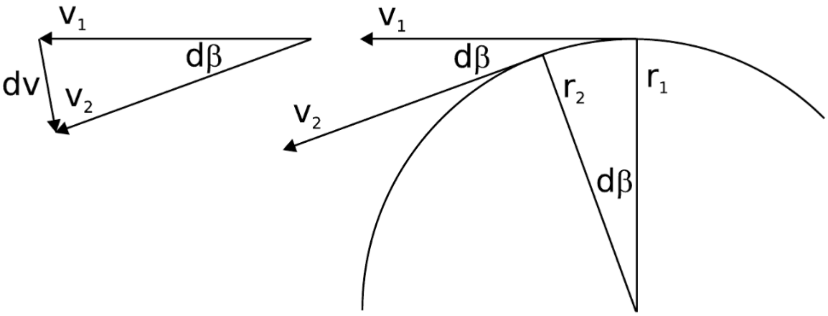

Fig. 2.1 Calculating the acceleration of an orbiting electron.

orbit is a circle of radius r. The electron's speed is then given by $v = r\,d\beta/dt$, where $d\beta$ is the small angle traversed during a short time dt, while the magnitude a of the electron's acceleration is the magnitude dv of the vector difference $\mathbf{v}_2 - \mathbf{v}_1$ divided by dt.[1] This equals $a = v\,d\beta/dt$, as we gather from Fig. 2.1. Eliminating $d\beta/dt$ by using $v = r\,d\beta/dt$, we arrive at $a = v^2/r$.

We want to calculate the electron's total energy as it orbits the nucleus (a proton). In Gaussian units, the magnitude of the Coulomb force exerted on the electron by the proton takes the particularly simple form $F = e^2/r^2$, where e is the absolute value of both the electron's and the proton's charge. Since $F = ma = mv^2/r$, we have that $mv^2 = e^2/r$. This gives us the electron's *kinetic* energy,

$$E_K = \frac{m_e v^2}{2} = \frac{e^2}{2r}\,, \tag{2.4}$$

where m_e is the electron's mass.

By convention, the electron's *potential* energy is 0 at $r = \infty$. Its potential energy at the distance r from the nucleus is therefore minus the work done by moving it from r to infinity,

$$E_P = -\int_r^\infty F\,dr = -\int_r^\infty \frac{e^2}{(r')^2}\,dr' = -\frac{e^2}{r}\,. \tag{2.5}$$

(You will do the integral in the next chapter.) So the electron's total energy is $E = E_K + E_P = -e^2/2r$.

Our next order of business is to express E as a function of L rather than r. Classically, $L = m_e vr$. Equation (2.4) allows us to massage E into

[1] To be precise, this holds in the limit in which dt, and hence $d\beta$ and dv, go to 0. See the next chapter for a brief introduction to vectors, differential quotients, and such.

the desired form:

$$E = -\frac{m_e e^4}{2\, m_e^2 v^2 r^2} = -\frac{m_e e^4}{2\, L^2}\,. \tag{2.6}$$

At this point Bohr simply substitutes $L = n\hbar$ for the classical expression $L = m_e v r$:

$$E_n = -\frac{1}{n^2}\left(\frac{m_e e^4}{2\,\hbar^2}\right), \quad n = 1, 2, 3, \dots \tag{2.7}$$

If $n\hbar$ ($n = 1, 2, 3, \dots$) are the only values that L can take, then these are the only values that the electron's energy can take. It follows at once that a hydrogen atom can emit or absorb energy only by amounts equal to the differences

$$\Delta E_{nm} = E_n - E_m = \left(\frac{1}{m^2} - \frac{1}{n^2}\right)\text{Ry}\,, \tag{2.8}$$

where the *Rydberg* (Ry) is an energy unit equal to $m_e e^4/2\hbar^2 = 13.605691\,\text{eV}$. It is also the ionization energy $\Delta E_{\infty 1}$ of atomic hydrogen in its ground state.

Considering the variety of wrong classical assumptions that went into the derivation of Eq. (2.8), it is remarkable that the frequencies predicted by Bohr via $\nu_{nm} = E_{nm}/h$ were in excellent agreement with the experimentally known frequencies at which atomic hydrogen emits and absorbs light.

2.4 de Broglie

In 1923, ten years after Bohr postulated that L comes in integral multiples of $\hbar$, someone finally hit on an explanation *why* angular momentum was quantized. In 1905, Albert Einstein had argued that electromagnetic radiation itself was quantized—not merely its emission and absorption, as Planck had held. Planck's radiation formula had implied a relation between a particle property and a wave property for the quanta of electromagnetic radiation we now call *photons*: $E = h\nu$. Einstein's explanation of the photoelectric effect established another such relation:

$$p = h/\lambda\,, \tag{2.9}$$

where p is the photon's momentum and λ is its wavelength. But if electromagnetic waves have particle properties, Louis de Broglie reasoned, why cannot electrons have wave properties?

Imagine that the electron in a hydrogen atom is a standing wave on a circle (Fig. 2.2) rather than a corpuscle moving in a circle. (The crests,

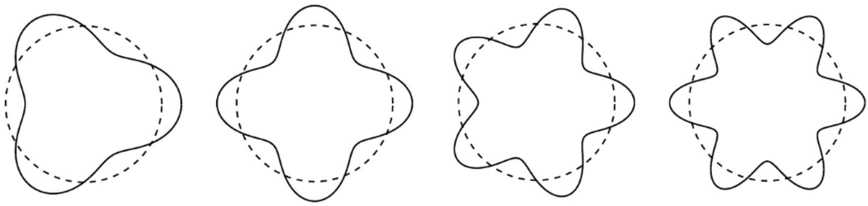

Fig. 2.2 Standing waves on a circle for $n = 3, 4, 5, 6$.

troughs, and nodes of a standing wave are stationary—they stay put.) Such a wave has to satisfy the condition

$$2\pi r = n\lambda\,, \quad n = 1, 2, 3, \ldots, \tag{2.10}$$

i.e., the circle's circumference $2\pi r$ must be an integral multiple of λ. Using $p = h/\lambda$ to eliminate λ from Eq. (2.10) yields $pr = n\hbar$. But $pr = mvr$ is just the angular momentum L of a classical electron moving in a circle of radius r. In this way de Broglie arrived at the quantization condition $L = n\hbar$, which Bohr had simply postulated.

This page intentionally left blank

Chapter 3

Mathematical interlude

3.1 Vectors

A *vector* is a quantity that has both a magnitude and a direction—for present purposes a direction in "ordinary" 3-dimensional space. Such a quantity can be represented by an arrow.

The sum of two vectors can be defined via the parallelogram rule: (i) move the arrows (without changing their magnitudes or directions) so that their tails coincide, (ii) duplicate the arrows, (iii) move the duplicates (again without changing magnitudes or directions) so that (a) their tips coincide and (b) the four arrows form a parallelogram. The resultant vector extends from the tails of the original arrows to the tips of their duplicates.

If we introduce a coordinate system with three mutually perpendicular axes, we can characterize a vector $\mathbf{a}$ by its components (a_x, a_y, a_z) (Fig. 3.1).

Problem 3.1. ($*$) *The sum* $\mathbf{c} = \mathbf{a} + \mathbf{b}$ *of two vectors has the components* $(c_x, c_y, c_z) = (a_x + b_x, a_y + b_y, a_z + b_z)$.

The *dot product* of two vectors $\mathbf{a}, \mathbf{b}$ is the number

$$\mathbf{a} \cdot \mathbf{b} \stackrel{\text{Def}}{=} a_x b_x + a_y b_y + a_z b_z \,. \tag{3.1}$$

We need to check that this definition is independent of the (rectangular) coordinate system to which the vector components on the right-hand side refer. To this end we calculate

$$\begin{aligned}(\mathbf{a}+\mathbf{b})\cdot(\mathbf{a}+\mathbf{b}) &= (a_x+b_x)^2+(a_y+b_y)^2+(a_z+b_z)^2\\ &= a_x^2+a_y^2+a_z^2+b_x^2+b_y^2+b_z^2+2\,(a_xb_x+a_yb_y+a_zb_z)\\ &= \mathbf{a}\cdot\mathbf{a}+\mathbf{b}\cdot\mathbf{b}+2\,\mathbf{a}\cdot\mathbf{b}\,. \end{aligned} \tag{3.2}$$

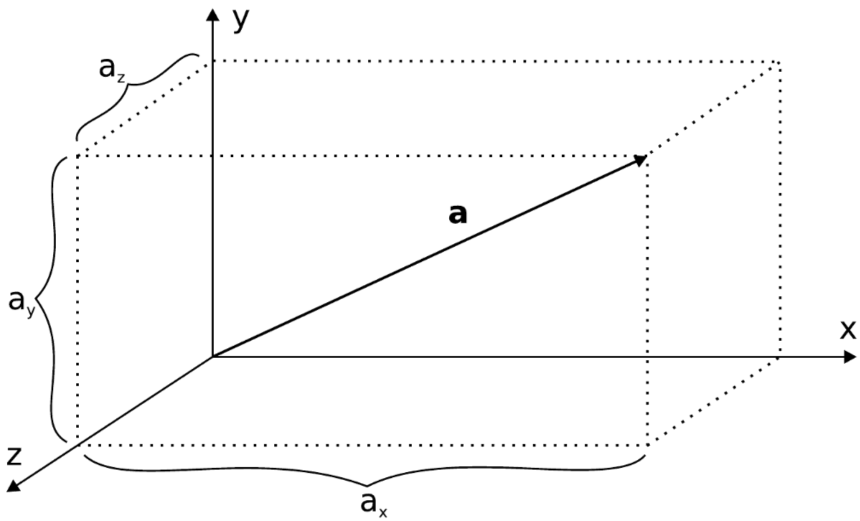

Fig. 3.1 The components of a vector.

According to Pythagoras, the magnitude a of a vector $\mathbf{a}$ equals $\sqrt{a_x^2 + a_y^2 + a_z^2}$. Because the left-hand side and the first two terms on the right-hand side of Eq. (3.2) are the squared magnitudes of vectors, they do not change under a coordinate transformation that preserves the magnitudes of all vectors. Hence the third term on the right-hand side does not change under such a transformation, and neither therefore does the product $\mathbf{a} \cdot \mathbf{b}$. But the coordinate transformations that preserve the magnitudes of vectors also preserve the angles between vectors. In particular, they turn a system of rectangular coordinates into another system of rectangular coordinates. Thus while the individual components on the right-hand side of Eq. (3.2) generally change under such a transformation, the dot product $\mathbf{a} \cdot \mathbf{b}$ does not.

By the term *scalar* we mean a number that is invariant under transformations of some kind or other. Since the dot product is invariant under translations and rotations of the coordinate axes—the transformations that preserve magnitudes and angles—it is also known as *scalar product.*

Problem 3.2. $(*)$ $\mathbf{a} \cdot \mathbf{b} = ab\cos\theta$, *where* θ *is the angle between* $\mathbf{a}$ *and* $\mathbf{b}$.

Another useful definition (albeit only in a 3-dimensional space) is the *cross product* of two vectors. If $\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}$ are unit vectors parallel to the coordinate

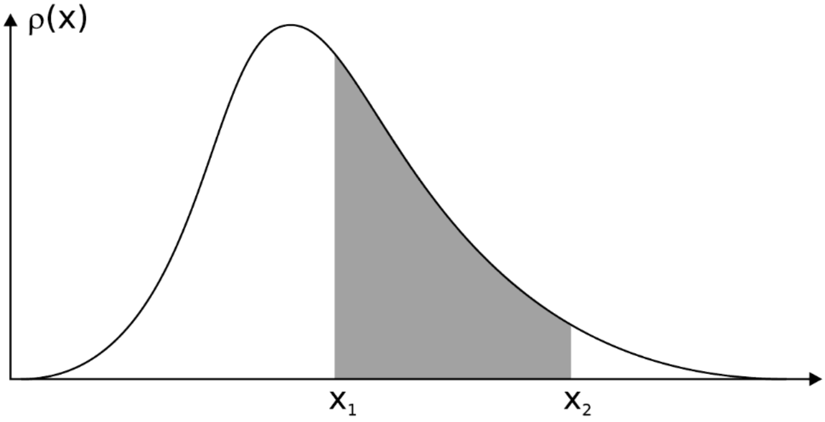

Fig. 3.2 The area corresponding to a definite integral.

axes, this is given by

$$\mathbf{a} \times \mathbf{b} \stackrel{\text{Def}}{=} (a_y b_z - a_z b_y)\,\hat{\mathbf{x}} + (a_z b_x - a_x b_z)\,\hat{\mathbf{y}} + (a_x b_y - a_y b_x)\,\hat{\mathbf{z}}\,. \tag{3.3}$$

Problem 3.3. *The cross product is antisymmetric:* $\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}$.

Problem 3.4. $(*)$ $\mathbf{a} \times \mathbf{b}$ *is perpendicular to both* $\mathbf{a}$ *and* $\mathbf{b}$.

Problem 3.5. $\hat{\mathbf{x}} \times \hat{\mathbf{y}} = \hat{\mathbf{z}}\,,\;\; \hat{\mathbf{y}} \times \hat{\mathbf{z}} = \hat{\mathbf{x}}\,,\;\; \hat{\mathbf{z}} \times \hat{\mathbf{x}} = \hat{\mathbf{y}}\,.$

By convention, the direction of $\mathbf{a} \times \mathbf{b}$ is given by the right-hand rule: if the first (index) and the second (middle) finger of your right hand point in the direction of $\mathbf{a}$ and $\mathbf{b}$, respectively, then your right thumb (pointing in a direction perpendicular to both $\mathbf{a}$ and $\mathbf{b}$) indicates the direction of $\mathbf{a} \times \mathbf{b}$.

Problem 3.6. $(*)$ *The magnitude of* $\mathbf{a} \times \mathbf{b}$ *equals* $ab\sin\theta$, *the area of the parallelogram spanned by* $\mathbf{a}$ *and* $\mathbf{b}$.

3.2 Definite integrals

We frequently have to deal with probabilities that are assigned to intervals of a continuous variable x (like the interval $[x_1, x_2]$ in Fig. 3.2). Such probabilities are calculated with the help of a *probability density function* $\rho(x)$, which is defined so that the probability with which x is found to

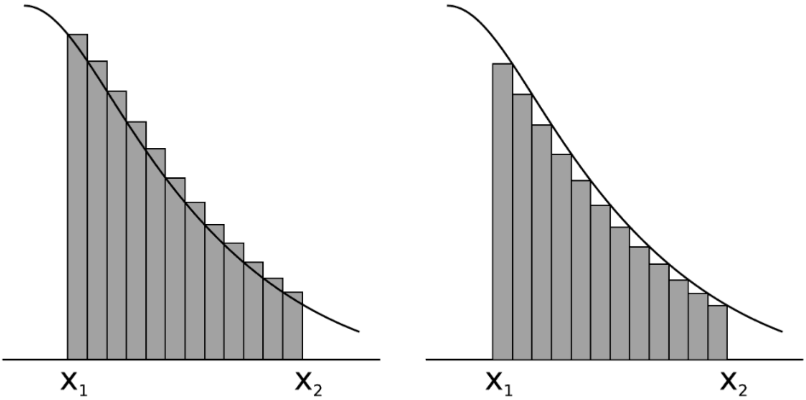

Fig. 3.3 Two approximations to the definite integral (3.4).

lie in the interval $[x_1, x_2]$ is given by the shaded area in Fig. 3.2. The mathematical tool for calculating this area is the *(definite) integral*

$$A = \int_{x_1}^{x_2} \rho(x)\, dx\,. \tag{3.4}$$

To define this integral, we overlay the shaded area of Fig. 3.2 with N rectangles of width $\Delta x = (x_2 - x_1)/N$ in either of the ways shown in Fig. 3.3. The sum of the rectangles in the left half of this figure,

$$A_+ = \sum_{k=0}^{N-1} \rho(x + k\,\Delta x)\,\Delta x\,, \tag{3.5}$$

is larger than the wanted area A, while the sum of the rectangles in the right half,

$$A_- = \sum_{k=1}^{N} \rho(x + k\,\Delta x)\,\Delta x\,, \tag{3.6}$$

is smaller. It is clear, though, that the differences $A_+ - A$ and $A - A_-$ decrease as the number of rectangles increases. The integral (3.4) is defined as the limit of either sum:

$$\lim_{N\to\infty} \sum_{k=1}^{N} \rho(x + k\,\Delta x)\,\Delta x = \int_{x_1}^{x_2} \rho(x)\, dx = \lim_{N\to\infty} \sum_{k=0}^{N-1} \rho(x + k\,\Delta x)\,\Delta x\,.$$

Another frequently used expression is the integral $\int_{-\infty}^{+\infty} \rho(x)\, dx$, which is defined as the limit

$$\lim_{a\to\infty} \int_{-a}^{+a} \rho(x)\, dx\,. \tag{3.7}$$

One often has to integrate functions of more than one variable. Take the integral

$$\int_R f(x,y,z)\, d^3r\,. \tag{3.8}$$

R is a region of 3-space, and $d^3r = dx\,dy\,dz$ is the volume of an infinitely small rectangular cuboid with sides dx, dy, dz. Instead of summing over infinitely many infinitely small intervals lying inside a finite interval, one now sums over infinitely many infinitely small rectangular cuboids lying inside a finite region R. (For more on infinitely many infinitely small things see the next section.)

3.3 Derivatives

A *function* $f(x)$ is a machine that has an input and an output. Insert a number x and out pops the number $f(x)$. [*Warning*: sometimes $f(x)$ denotes the machine itself rather than the number obtained after inserting a particular x.] We shall mostly be dealing with functions that are well-behaved. Saying that a function $f(x)$ is *well-behaved* is the same as saying that we can draw its graph without lifting up the pencil, and we can do the same with the graphs of its derivatives.

The (first) *derivative* of $f(x)$ is a machine $f'(x)$ that works like this: insert a number x, and out pops the slope of (the graph of) $f(x)$ at x. What we mean by the *slope* of $f(x)$ at a particular point $x = a$ is the slope of the *tangent* $t(x)$ on $f(x)$ at a.

Take a look at Fig. 3.4. The curve in each of the three diagrams is (the graph of) $f(x)$. The slope of the straight line $s(x)$ that intersects $f(x)$ at two points in the upper diagrams is given by the *difference quotient*

$$\frac{\Delta s}{\Delta x} = \frac{s(x+\Delta x) - s(x)}{\Delta x}\,. \tag{3.9}$$

This tells us how much $s(x)$ increases as x increases by Δx. The lower diagram shows the tangent $t(x)$ on the function $f(x)$ for a particular x.

Now consider the small black disk at the intersection of the functions $f(x)$ and $s(x)$ at $x+\Delta x$ in the upper left diagram. Think of it as a bead sliding along $f(x)$ towards the left. As it does so, the slope of $s(x)$ increases (compare the upper two diagrams). In the limit in which this bead occupies the same place as the bead sitting at x, $s(x)$ coincides with $t(x)$, as one gleans from the lower diagram. In other words, as Δx tends to 0, the

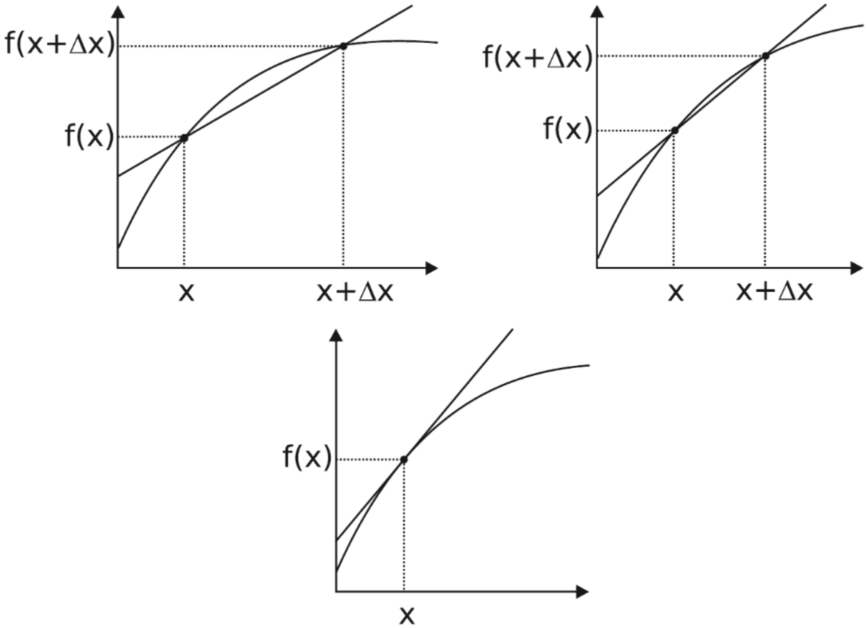

Fig. 3.4 Definition of the slope of a function $f(x)$ at x.

difference quotient (3.9) tends to the *differential quotient*

$$\frac{df}{dx} \stackrel{\text{Def}}{=} \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x}\,, \tag{3.10}$$

which is the same as $f'(x)$. The *differentials* dx and df are *infinitesimal* ("infinitely small") quantities. This sounds highly mysterious until one realizes that every expression containing such quantities is to be understood as the limit in which these tend to 0, one (here, dx) independently, the others (here, df) dependently.

To *differentiate* a function $f(x)$ is to obtain its first derivative $f'(x)$. By differentiating $f'(x)$, we obtain the second derivative $f''(x)$ of $f(x)$, for which we can also write d^2f/dx^2. To make sense of the last expression, think of d/dx as an operator. Like a function, an *operator* has an input and an output, but unlike a function, it accepts a function as input. Insert $f(x)$ into d/dx and get the function df/dx. Insert the output of d/dx into another operator d/dx and get the function $(d/dx)(d/dx)f(x) \stackrel{\text{Def}}{=} (d^2/dx^2)f(x) = d^2f/dx^2$.

By differentiating the second derivative we obtain the third, and so on.

Fig. 3.5 Illustration of the product rule.

Problem 3.7. *Find the slope of the straight line* $f(x) = ax + b$.

Problem 3.8. $(*)$ *Calculate* $f'(x)$ *for* $f(x) = 2x^2 - 3x + 4$.

Problem 3.9. $(*)$ *What does* $f''(x)$*—the slope of the slope of* $f(x)$*—tell us about the graph of* $f(x)$*?*

By definition, $(f + g)(x) = f(x) + g(x)$.

Problem 3.10. *If* a *is a number and* f *and* g *are functions of* x*, then*

$$\frac{d(af)}{dx} = a\frac{df}{dx} \qquad \text{and} \qquad \frac{d(f+g)}{dx} = \frac{df}{dx} + \frac{dg}{dx} .$$

A slightly more difficult task is to differentiate the product $h(x) = f(x)\,g(x)$. Think of f and g as the vertical and horizontal sides of a rectangle of area h. As x increases by Δx, the product fg increases by the sum of the areas of the three white rectangles in Fig. 3.5:

$$\Delta h = f(\Delta g) + (\Delta f)g + (\Delta f)(\Delta g) . \tag{3.11}$$

Hence

$$\frac{\Delta h}{\Delta x} = f\,\frac{\Delta g}{\Delta x} + \frac{\Delta f}{\Delta x}\,g + \frac{\Delta f\,\Delta g}{\Delta x} . \tag{3.12}$$

If we now let Δx go to 0, the first two terms on the right-hand side tend to $f\,g' + f'\,g$. What about the third term? Since it is the product of an expression (either $\Delta g/\Delta x$ or $\Delta f/\Delta x$) that tends to a finite number and an expression (either Δf or Δg) that tends to 0, it tends to 0. The bottom line:

$$h' = (f\,g)' = f\,g' + f'\,g . \tag{3.13}$$

Problem 3.11. $(*)$ $(f\,g\,h)' = f\,g\,h' + f\,g'\,h + f'\,g\,h$.

The generalization to products of n functions is straightforward. An important special case is the product of n identical functions:

$$(f^n)' = f^{n-1}\,f' + f^{n-2}\,f'\,f + \cdots + f'\,f^{n-1} = n\,f^{n-1}f'\,. \tag{3.14}$$

If $f(x) = x$, this boils down to

$$(x^n)' = n\,x^{n-1}. \tag{3.15}$$

Suppose now that g is a function of f, and that f is a function of x. An increase in x by Δx will cause an increase in f by $\Delta f \approx (df/dx)\Delta x$, and this will cause an increase in g by $\Delta g \approx (dg/df)\Delta f$ (the symbol $\approx$ means "is approximately equal to"). Thus

$$\frac{\Delta g}{\Delta x} \approx \frac{dg}{df}\frac{df}{dx}\,. \tag{3.16}$$

In the limit $\Delta x \to 0$, "approximately equal" becomes "equal," and Eq. (3.16) becomes the *chain rule*

$$\frac{dg}{dx} = \frac{dg}{df}\frac{df}{dx}\,. \tag{3.17}$$

Problem 3.12. *We have proved Eq. (3.15) for integers $n \geq 2$. Check that it also holds for $n = 0$ and $n = 1$.*

Problem 3.13. $(*)$ *Equation (3.15) also holds for negative integers n.*

Problem 3.14. $(*)$ *Equation (3.15) also holds for $n = 1/m$, where m is a natural number.*

Problem 3.15. *Use the chain rule (3.17) to show that if Eq. (3.15) holds for $n = a$ and $n = b$, then it also holds for $n = ab$.*

It follows from what you have just shown that Eq. (3.15) holds for all *rational numbers n*. Moreover, since every real number is the limit of a *sequence* of rational numbers, we can make sure that Eq. (3.15) holds for all *real numbers*, by defining it as the limit of some sequence in case n is an irrational number.

We often use functions with more than one input slot. The output of $f(t,x,y,z)$, for example, depends on the time coordinate t as well as the spatial coordinates x,y,z. If we choose a fixed set of values x,y,z, we obtain a function $f_{xyz}(t)$ of t alone. The *partial derivative* of $f(t,x,y,z)$ with respect to t is the derivative of $f_{xyz}(t)$, for which we write $\partial f/\partial t$ (usually without explicitly indicating that this function depends on the chosen set of values x,y,z). The partial derivatives of $f(t,x,y,z)$ with respect to the other variables are defined analogously.

3.4 Taylor series

A well-behaved function can be expanded into a power series. This means that for all non-negative integers $k = 0, 1, 2, \ldots$ there are real numbers a_k such that

$$f(x) = \sum_{k=0}^{\infty} a_k\, x^k = a_0 + a_1\, x + a_2\, x^2 + a_3\, x^3 + a_4\, x^4 + \cdots . \qquad (3.18)$$

Let's calculate the first four derivatives using (3.15):

$$\begin{aligned} f'(x) &= a_1 + 2\, a_2 x + 3\, a_3\, x^2 + 4\, a_4\, x^3 + 5\, a_5\, x^4 + \cdots , \\ f''(x) &= 2\, a_2 + 2 \cdot 3\, a_3\, x + 3 \cdot 4\, a_4\, x^2 + 4 \cdot 5\, a_5\, x^3 + \cdots , \\ f'''(x) &= 2 \cdot 3\, a_3 + 2 \cdot 3 \cdot 4\, a_4\, x + 3 \cdot 4 \cdot 5\, a_5\, x^2 + \cdots , \\ f''''(x) &= 2 \cdot 3 \cdot 4\, a_4 + 2 \cdot 3 \cdot 4 \cdot 5\, a_5\, x + \cdots . \end{aligned}$$

Setting x equal to zero, we obtain the following values:

$$f(0) = a_0\,, \quad f'(0) = a_1\,, \quad f''(0) = 2\, a_2\,,$$
$$f'''(0) = 2 \times 3\, a_3\,, \quad f''''(0) = 2 \times 3 \times 4\, a_4\,.$$

Since we don't want to go on adding primes ($'$), we will write $f^{(n)}(x)$ for the n-th derivative of $f(x)$. If we also write $f^{(0)}(x)$ for $f(x)$, we have that $f^{(k)}(0)$ equals $k!\, a_k$, where the *factorial* $k!$ is defined as equal to 1 for $k = 0$ and $k = 1$, and as the product of all natural numbers $n \leq k$ for $k > 1$. Expressing the *coefficients* a_k in terms of the derivatives of $f(x)$ for $x = 0$, we arrive at the following power series—also known as the *Taylor series*—for $f(x)$:

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!}\, x^k. \qquad (3.19)$$

A remarkable result: if you know the value of a well-behaved function $f(x)$ and the values of *all* of its derivatives for a single value of x (in this case $x = 0$, but there is a similar series for any value of x), then you know $f(x)$ for *all* values of x.

3.5 Exponential function

We define the function $\exp(x)$ by requiring that $\exp'(x) = \exp(x)$ and $\exp(0) = 1$. In other words, the value of this function is everywhere equal to the slope of its graph, which intersects the vertical axis at the value 1.

Problem 3.16. *Sketch the graph of* $\exp(x)$ *using this information alone.*

Problem 3.17. *All derivatives of* $\exp(x)$ *are equal to* $\exp(x)$.

Thus $\exp^{(k)}(0) = 1$ for all k, whence a particularly simple Taylor series results:

$$\exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \cdots . \tag{3.20}$$

Problem 3.18. (∗) $\exp(x)$ *satisfies*

$$f(a)\, f(b) = f(a+b) . \tag{3.21}$$

It can be shown that every function satisfying Eq. (3.21) has the form $f(x) = a^x$. This means that there is a number e such that $\exp(x) = e^x$—hence the name "exponential function."

Problem 3.19. (∗) *Calculate* e.

Problem 3.20. $d(e^{ax})/dx = a\, e^{ax}$.

The *natural logarithm* $\ln x$ is the inverse of e^x, that is, $e^{\ln x} = \ln(e^x) = x$.

Problem 3.21. $\ln a + \ln b = \ln(ab)$.

Problem 3.22. (∗)

$$\frac{d \ln f(x)}{dx} = \frac{1}{f(x)} \frac{df}{dx} . \tag{3.22}$$

3.6 Sine and cosine

We define the function $\cos(x)$ by requiring that $\cos''(x) = -\cos(x)$, $\cos(0) = 1$, and $\cos'(0) = 0$.

Problem 3.23. (∗) *Sketch the graph of* $\cos(x)$, *making use of this information alone.*

Problem 3.24. *For* $n \geq 0$: $\cos^{(n+2)}(x) = -\cos^{(n)}(x)$.

Problem 3.25.

$$\cos^{(k)}(0) = \begin{cases} +1 \text{ for } k = 0, 4, 8, 12, \ldots \\ -1 \text{ for } k = 2, 6, 10, 14, \ldots \\ 0 \quad \text{for odd } k \end{cases}$$

We thus arrive at the following Taylor series:

$$\cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots . \tag{3.23}$$

The function $\sin(x)$ is defined by requiring that $\sin''(x) = -\sin(x)$, $\sin(0) = 0$, and $\sin'(0) = 1$. This leads to the Taylor series

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots . \tag{3.24}$$

3.7 Integrals

In Sec. 3.2 we defined the definite integral as a limit. How do we calculate this limit? The answer is elementary if we know a function $F(x)$ of which $f(x)$ is the first derivative, $f = dF/dx$, for we can then substitute dF for $f\,dx$:

$$\int_a^b f(x)\,dx = \int_a^b dF(x)\,. \tag{3.25}$$

On the face of it, we are still adding infinitely many infinitely small quantities, but look what this amounts to:

$$\begin{aligned}\int_a^b dF(x) = [F(a+dx) - F(a)]& \\ + [F(a+2\,dx) - F(a+dx)]& \\ + [F(a+3\,dx) - F(a+2\,dx)]& \\ + \cdots& \\ + [F(b-2\,dx) - F(b-3\,dx)]& \\ + [F(b-dx) - F(b-2\,dx)]& \\ + [F(b) - F(b-dx)]\,.&\end{aligned}$$

After all cancellations are done, we are left with $\int_a^b dF(x) = F(b) - F(a)$.

If $f(x)$ is the derivative of $F(x)$, $F(x)$ is known as an *integral* or *antiderivative* of $f(x)$—*an* integral rather than *the* integral because if $F(x)$ is an integral of $f(x)$ and c is a constant, then $F(x) + c$ is another integral of $f(x)$. To distinguish integrals from definite integrals, we also refer to them as *indefinite* integrals.

Problem 3.26. $(*)$ *Calculate* $\int_1^2 x^2\,dx$.