Supplement on Lagrangian, Hamiltonian Mechanics

Robert B. Griffiths
Version of 23 October 2009

Contents

1 Introduction

These notes are intended to supplement, and in one instance correct, the material on Lagrangian and Hamiltonian formulations of mechanics found in Thornton and Marion Ch. 7.

2 Partial derivatives

★ When there are several independent variables it is easy to make mistakes in taking partial derivatives. The fundamental rule is: always know *which set* of independent variables is in use, so that you are sure which are being held fixed during the process of taking a partial derivative.

★ A simple example will illustrate the problems that await the careless. Start with

$$f(x,y) = 2x + y; \quad \partial f/\partial x = 2. \tag{1}$$

But suppose we prefer to consider f as a function of x and $z = x + y$. Then

$$f(x,z) = x + z; \quad \partial f/\partial x = 1. \tag{2}$$

• How can we avoid drawing the silly conclusion that $1 = 2$? In particular, is there some helpful notation that will warn us? Indeed, there are (at least) two approaches to difficulties of this sort.

• Approach 1. Never use the same symbol f for two different functions. That is, $f(\cdot, \cdot)$ should always in a given discussion refer to a well-defined map from two real arguments to a real number. You give me the first and second arguments of f as two real numbers, and I will give you back another real number which I have looked up in a table. E.g., $f(0.2, 1.7)$ is always equal to 2.1. Period.

– In approach 1, we would never be so careless as to write down both (1) and (2), because the first tells us that $f(1,0) = 2$ and the second that $f(1,0) = 1$, and nonsense gives rise to nonsense. What we should have done is to write $\bar{f}$ in place of f in (2), and then there is no problem when $\partial f/\partial x$ is unequal to $\partial \bar{f}/\partial x$.

• Approach 2. Approach 1 is not popular with physicists, though perhaps it should be. The reason is that the symbols which occur in the equations of physics tend to refer to specific physical quantities. Thus in thermodynamics E (or U) refers to energy, T to temperature, V volume, P pressure, and so forth. Sometimes it is convenient to think of E as a function of T and V, and sometimes as a function of T and P, and sometimes as a function of other things. The habit of the physicist is to identify which function is under discussion by giving appropriate labels to the arguments. Thus $E(T, V)$ is not the same function as $E(T, P)$, and it is obvious (to the physicist) which is which, and if he sometimes writes $E(V, T)$ instead of $E(T, V)$, what is wrong with that?

○ Nothing is wrong with this approach, but when you write $\partial E/\partial T$, what do you mean by it? Are you thinking of $E(T, V)$ or $E(T, P)$? The two partials are *not* the same (except in special cases), and how are we to avoid confusion?

★ One standard way to avoid confusion is to use subscripts. Thus

$$\left(\frac{\partial E}{\partial T}\right)_V \tag{3}$$

means the partial derivative of energy with respect to temperature while the volume is held fixed. Or, to put it in other terms, the labels T and V, which told us which function we were referring to when writing $E(T, V)$ are again evident, one as the ∂T denominator and the other as a subscript.

• The subscript notation is effective but cumbersome. E.g., suppose the physical chemist is concerned with a system in which there are N_1 moles of species 1, N_2 moles of species 2, ... N_r moles of species r. Then a partial derivative of $E(T, V, N_1, \ldots N_r)$ might be written as

$$\left(\frac{\partial E}{\partial N_j}\right)_{T,V,N_1,\ldots N_{j-1},N_{j+1}\ldots N_r} \tag{4}$$

which is perfectly clear, but takes a lot of time to write down.

★ So in practice what happens is that the subscript list is often omitted, and which other variables are being held constant has to be determined from the context.

• Thus $\partial L/\partial \dot{q}_j$ in TM always means the variable list is $q_1, q_2, \ldots q_s, \dot{q}_1, \dot{q}_2, \ldots \dot{q}_s, t$, because by convention the Lagrangian L depends on these variables. We could make it depend on some other set of variables, but then it would be a good idea to warn people, perhaps by replacing L with $\bar{L}$ or some such.

• A similar convention applies to the Hamiltonian H. When one writes $\partial H/\partial q_j$ it is always understood that the full variable list is $q_1, q_2, \ldots q_x, p_1, p_2, \ldots p_s, t$, and thus one is thinking of $H(q_1, q_2, \ldots q_x, p_1, p_2, \ldots p_s, t)$, and thus

$$\frac{\partial H}{\partial q_j} \text{ means } \left(\frac{\partial H}{\partial q_j}\right)_{q_1,\ldots q_{j-1},q_{j+1},\ldots p_1,\ldots p_s,t} \tag{5}$$

And of course the same variable set is in mind whenever one writes $\partial H/\partial p_j$.

• This is what is behind the insistence of Thornton and Marion that one always write the Hamiltonian as a function of the q's and p's, and not of the q's and $\dot{q}$'s. That is, at this point they are employing Approach 1 to the problem of avoiding confusion when taking partial derivatives.

★ Let us see how this works in the case of a simple example: going from a Lagrangian to a Hamiltonian when there is only 1 degree of freedom, and we are considering a harmonic oscillator

$$L = L(x, \dot{x}) = (m/2)\dot{x}^2 - (k/2)x^2; \quad p = \partial L/\dot{x} = m\dot{x}. \tag{6}$$

○ In this case $\partial L/\partial t = 0$, and there is no harm in simply omitting the t argument from out list.

2

- Now define "the Hamiltonian" using the formula

$$\bar{H}(x, \dot{x}) = \dot{x}\partial L/\partial \dot{x} - L = (m/2)\dot{x}^2 + (k/2)x^2, \tag{7}$$

where we are being careful to distinguish the *function* $\bar{H}$ from the same quantity expressed as a *different* function

$$H(x, p) = p^2/2m + (k/2)x^2 = \bar{H}(x, \dot{x}(x, p)). \tag{8}$$

3   The difference between $\partial L/\partial t$ and dL/dt

★ Students of mechanics are often confused about the difference between $\partial L/\partial t$ and dL/dt. They are very different quantities, and it is important to understand just what is meant in each case. It will suffice to consider the case where there is a single (generalized) coordinate q with time derivative $\dot{q}$, and a Lagrangian

$$L(q, \dot{q}, t) = (m/2)\dot{q}^2 - (1 + a\cos bt)kq^2, \tag{9}$$

where k, a and b are constants.

- When considering *partial* derivatives of L one should take the perspective that all the other arguments of L are being held fixed when this one is varied. If one were to write $L(\alpha, \beta, \gamma)$, a function of three real variables, there would be no doubt at all as to what $\partial L/\partial \gamma$ means: it means the derivative of L with respect to γ with α and β held fixed.

○ So to find $\partial L/\partial t$ for L in (9) we should hold q and $\dot{q}$ absolutely constant while carrying out the derivative with respect to t. Consequently,

$$\partial L/\partial t = abk(\sin bt)q^2. \tag{10}$$

○ But what does this correspond to physically? Well, it might be that the spring of the harmonic oscillator is subject to some oscillating draft of air which makes its temperature go up and down, and therefore varies the spring constant. But surely one would not expect both q and $\dot{q}$, position and velocity, to remain fixed while the temperature varied? Indeed, one would not, and this is a case where one should carry out the math as prescribed by the symbols, and not try and interpret it as a physical process. Instead $\partial L/\partial t$ is a formal operation.

- Sometimes the situation $\partial L/\partial t = 0$ is described as one in which the Lagrangian has no *explicit* time dependence. This terminology is fine if one knows what it means. E.g., because of the $\cos bt$ term the Lagrangian in (9) has an explicit time dependence.

★ By contrast, when using the symbol dL/dt we *always* have in mind a situation where a definite *path* $q(t)$ is specified. Given this path, we can of course compute $\dot{q}(t)$, and insert $q(t)$ and $\dot{q}(t)$ into the Lagrangian as its first two arguments. This way L will (in general) change in time even in the case where $\partial L/\partial t = 0$, and of course if $\partial L/\partial t \neq 0$ this can be yet another source of time dependence.

- Consider (9) with $a = 0$: the oscillator will oscillate, and the value of L will oscillate in time, even though $\partial L/\partial t = 0$.

□ Exercise. Check this by working out dL/dt for L in (9), assuming $a = 0$ but that the oscillator is not stationary at the origin.

- Analogy. The elevation at some fixed geographical point in Western Pennsylvania is, to a fairly good approximation, independent of time. Think of this as the Lagrangian. When you go on a hike, however, and feed your coordinates into this "Lagrangian," your elevation will vary with time because your geographical position is shifting. To be sure, if you have the misfortune of walking over a subsiding coal mine there might be a $\partial L/\partial t$ contribution as well.

★ All of this generalizes in an obvious way to the case where there is more than one generalized coordinate. When calculating $\partial L/\partial t$ one must hold *every* q_j and *every* $\dot{q}_j$ fixed. When calculating dL/dt one must be

thinking about a *collection* of functions $q_j(t)$, and insert them and their time derivatives as appropriate arguments in L to determine its time dependence.

★ The same comments apply in an obvious way to the difference between $\partial H/\partial t$ and dH/dt, except that when taking the partial derivative $\partial H/\partial t$ the variables held fixed are the q's and the p's; see the discussion in connection with (19) below.

4 Legendre Transformation and Differentials

★ The use of differentials speeds up certain formal derivations when one is dealing with functions of several independent variables, especially if one is changing from one independent set to another. The approach can be made mathematically rigorous; see Courant. The perspective here is a bit different from that of Courant, but for our purposes it will work equally well.

★ We choose illustrations from thermodynamics, where the use of differentials is common, and focus on a system with two independent variables, which could be the entropy S and the volume V, or temperature T and volume V, or S and pressure P, etc. For the following discussion the exact significance of the symbols is not important. We denote the energy by E, using that in preference to U because U is used by TM for potential energy. The thermodynamic energy E is the total kinetic plus potential energy of a system of particles in thermodynamic equilibrium, in which case E is some well-defined function of S and V, and in the standard notation one writes

$$dE = T\,dS - P\,dV, \tag{11}$$

where dE, dS, and dV are examples of *differentials*. How are we to understand this equation?

• A useful way of understanding it is to suppose that a path is defined in the space of independent variables S and V by giving each as a function of some parameter t, which can be thought of intuitively as the time, provided nothing is varied so rapidly as to take the system out of thermodynamic equilibrium. However, it really need not represent the time. All we want is that S and V are smooth functions of t, and we never allow the path to stop, i.e., for every t either dS/dt or dV/dt is nonzero.

○ And how does E vary along this path? Since it is a function of S and V, we write

$$\frac{dE}{dt} = \left(\frac{\partial E}{\partial S}\right)_V \frac{dS}{dt} + \left(\frac{\partial E}{\partial V}\right)_S \frac{dV}{dt}, \tag{12}$$

using the usual chain rule. This is the same as (11) if we make the identifications

$$T = \left(\frac{\partial E}{\partial S}\right)_V, \quad P = -\left(\frac{\partial E}{\partial V}\right)_S \tag{13}$$

and then divide both sides of (11) by dt.

○ Hence (11) may be regarded as an abbreviation for (12) and (13) taken together, with the advantage that it says things in a very compact way. Note in particular that if we choose a path on which V does not change, the dV term in (11) is equal to zero, and dividing both sides by dS we arrive at the first of the two relations in (13); likewise by setting dS equal to zero we arrive at the second term in (13). So we have at the very least a nice device for memorizing the physical significance of certain partial derivatives without having to write them out in longhand.

★ But the real power of differentials emerges when we change independent variables, as in a Legendre transformation. Let the Helmholtz free energy be written as

$$F = E - TS. \tag{14}$$

Now watch:

$$dF = dE - T\,dS - S\,dT = -S\,dT - P\,dV, \tag{15}$$

where $d(TS) = TdS + SdT$ is obvious if you divide through by dt, and the second equality in (15) employs the earlier differential relation (11). From (15) we see at once that if we express F as a function of T and V, then it is the case that

$$S = -\left(\frac{\partial F}{\partial T}\right)_V, \quad P = -\left(\frac{\partial F}{\partial V}\right)_T. \tag{16}$$

• Whereas one can derive the results in (16) from those in (13), given the definition of F in (14), in a moderately straightforward manner by means of chain rule arguments, the use of differentials is faster, and also good mathematics if one follows Courant, or if one divides through by dt in the manner indicated above.

○ There are, to be sure, questions about starting with (13) and solving for S as a single-valued, well-defined and differentiable function of T and V in order to obtain F as a function of these variables using (14). We shall not address these issues here. Courant again has an accessible treatment. Typical physicists simply go ahead without checking, until they get into trouble.

★ Back to mechanics, where with L the Lagrangian we write its differential as

$$dL = \sum_j p_j \, d\dot{q}_j + \sum_j (\partial L/\partial q_j) \, dq_j + (\partial L/\partial t) \, dt, \tag{17}$$

noting that p_j is defined to be $\partial L/\partial \dot{q}_j$, so this equation merely summarizes names of partial derivatives, and tells us that the independent variables we are concerned with are those whose differentials appear on the right hand side.

• Then we introduce the analog of the thermodynamic E to F transformation above, except for a sign change, writing the Hamiltonian as

$$H = \sum_j p_j \dot{q}_j - L. \tag{18}$$

Next take its differential:

$$dH = \sum_j p_j \, d\dot{q}_j + \sum_j \dot{q}_j \, dp_j - dL = \sum_j \dot{q}_j \, dp_j - \sum_j (\partial L/\partial q_j) \, dq_j - (\partial L/\partial t) \, dt$$

$$= \sum_j \dot{q}_j \, dp_j - \sum_j \dot{p}_j \, dq_j - (\partial L/\partial t) \, dt \tag{19}$$

where the second equality employs (17) and the final one invokes the Lagrange equations $\partial L/\partial q_j = dp_j/dt$. The right side of (19) at once identifies the significance of the partial derivatives of H when written as a function of the q's and p's. It is actually a compact way of writing the Hamiltonian equations of motion, and has the advantage that it leaves us in no doubt about which variables are being held fixed when we take partial derivatives. It also tells us that $\partial H/\partial t = -\partial L/\partial t$.

5 Quadratic Kinetic Energy

★ The transformation from a Lagrangian to a Hamiltonian is particularly simple in a situation in which one can write the Lagrangian in the form

$$L(\vec{q}, \dot{\vec{q}}, t) = T - U = \tfrac{1}{2} \sum_{jk} A_{jk}(\vec{q}, t)\dot{q}_j \dot{q}_k - U(\vec{q}, t), \tag{20}$$

which is to say the kinetic energy is quadratic in the $\dot{q}_j$ and the potential energy is independent of the $\dot{q}_j$. We assume that the matrix A is symmetrical

$$A_{jk} = A_{kj} \text{ or } A^{\mathrm{T}} = A. \tag{21}$$

5

• The notation A^{T} means the *transpose* of the matrix A. Students who have not seen matrices in linear algebra, or have forgotten the material, are referred to App. A for a fast review.

○ Note that $A_{jk}(\vec{q}, t)$ is allowed to depend on some of the q_j's and t, but does not have to. The key point is that it does *not* depend upon any of the $\dot{q}_j$; the dependence of L on the $\dot{q}_j$ can only be that shown *explicitly* in (20). Similarly, U is allowed to depend on the q_j and t, but cannot depend on any $\dot{q}_j$.

★ A consequence of (20) is

$$p_j = \partial L/\partial \dot{q}_j = \partial T/\partial \dot{q}_j = \sum_k A_{jk}\dot{q}_k, \quad \text{or} \quad p = A \cdot \dot{q}, \tag{22}$$

where the second form, $p = A \cdot \dot{q}$, is to be understood in terms of matrix multiplication, with $\dot{q}$ a column vector, A a square matrix, and p another column vector. (Again, see App. A if this is not familiar.)

○ The matrix product $A \cdot \dot{q}$ is often written as simply $A\dot{q}$, and we omit the connecting $\cdot$ in (23) and (24) below.

• Notice that when we differentiate the expression for T the factor $\frac{1}{2}$ no longer appears in (22). This is correct. To see that it is correct look at an example or two using a small matrix, and remember (21).

★ If the matrix A is invertible with inverse A^{-1}, (22) implies that $\dot{q}$ and p, again understood as column vectors, are related by

$$\dot{q} = A^{-1}p \tag{23}$$

• Therefore we can write the kinetic energy in the form

$$T = \tfrac{1}{2}\dot{q}^{\mathrm{T}}A\dot{q} = \tfrac{1}{2}p^{\mathrm{T}}A^{-1}AA^{-1}p = \tfrac{1}{2}p^{\mathrm{T}}A^{-1}p = \tfrac{1}{2}\sum_{jk}(A^{-1})_{jk}p_j p_k. \tag{24}$$

Here we have used the fact that $(BC)^{\mathrm{T}} = C^{\mathrm{T}}B^{\mathrm{T}}$ for a matrix product, including the case where B or C is a row or column vector, see App. A; also $(A^{\mathrm{T}})^{-1} = (A^{-1})^{\mathrm{T}}$ for an invertible square matrix A; and $A^{\mathrm{T}} = A$, (21), for the particular matrix A we are interested in.

★ Thus the Hamiltonian expressed in terms of its standard set of independent variables is

$$H(\vec{q}, \vec{p}, t) = T + U = \tfrac{1}{2}p^{\mathrm{T}}A^{-1}p + U(\vec{q}, t), \tag{25}$$

where A^{-1} will depend on $\vec{q}$ and t if A depends on these quantities.

• Therefore when the Lagrangian has the form (20)—which is frequently but not always the case—we have a quick recipe for getting the Hamiltonian as a function of $\vec{q}$ and $\vec{p}$, the form that's useful when using it to obtain the equations of motion.

☐ Exercise. Fill in the spaces [...] in the following statement: Provided the Hamiltonian $H(\vec{q}, \vec{p}, t)$ is of the form [...], one can obtain from it the corresponding Lagrangian $L(\vec{q}, \dot{\vec{q}}, t)$ by the simple device of finding the matrix inverse of [...] and then writing L in the form [...].

6 Flows in phase space. Liouville's theorem

★ Phase space. It is customary to refer to the $2s$-dimensional space of Hamiltonian mechanics, spanned by $(q_1, q_2, \ldots q_s, p_1, p_2, \ldots p_s) = (\vec{q}, \vec{p})$, as *phase space* in contrast to the s-dimensional *configuration* space of the Lagrangian formulation. A point in this phase space moves in time with a "velocity" vector

$$\vec{v} = (\dot{q}_1, \dot{q}_2, \ldots \dot{q}_s, \dot{p}_1, \dot{p}_2, \ldots \dot{p}_s) = (\dot{\vec{q}}, \dot{\vec{p}}), \tag{26}$$

which has $2s$ components. The velocity at any point in phase space is determined by the Hamilton equations of motion:

$$\dot{q}_j = \partial H/\partial p_j, \quad \dot{p}_j = -\partial H/\partial q_j \tag{27}$$

As the velocity is well-defined, a point in phase space representing the mechanical state of a system at some time t_0 follows a well-defined "trajectory" at all later (and earlier) times.

○ As an example, consider a harmonic oscillator, where the trajectories are ellipses centered on the origin of the 2-dimensional ($s = 1$) phase space.

★ Let R be some collection of points in the phase space which constitute a region with well-defined boundaries. Then its "volume" can be written as a $2s$-dimensional integral

$$\mathcal{V}(R) = \int_R dq_1 \, dq_2 \cdots dq_s \, dp_1 \, dp_2 \cdots dp_s \tag{28}$$

★ Of course the collection of all the points that are in a region R_0 at time t_0 will at some later time t_1 be found in some other region R_1 determined by solving the equations of motion. Liouville's theorem states that

$$\mathcal{V}(R_0) = \mathcal{V}(R_1), \tag{29}$$

whatever the initial region R_0—assuming, of course, that the integral in (28) is well-defined—and whatever the times t_0 and t_1.

○ This is even true when the Hamiltonian is a function of time.

★ One way of stating Liouville's theorem is as follows. The time transformation from t_0 to t_1 produces an invertible map of the phase space onto itself. Associated with this map is a Jacobian. What Liouville says is that this Jacobian is equal to 1.

• For a proper mathematical proof of Liouville's theorem from this perspective see Gallavotti.

★ The essential geometrical idea behind Liouville's theorem can be stated in the following way: the velocity $\vec{v}$ defined in (26) as a function of position in the phase space, i.e., $\vec{v}(\vec{q}, \vec{p})$, has zero divergence

$$\nabla \vec{v} = \sum_j \left(\frac{\partial \dot{q}_j}{\partial q_j} + \frac{\partial \dot{p}_j}{\partial p_j} \right) = 0. \tag{30}$$

□ Exercise. Carry out the proof of (30). Consider a particular j in the sum, and use the fact that $\partial^2 H/\partial q_j \partial p_j = \partial^2 H/\partial p_j \partial q_j$.

• An incompressible fluid has a velocity field (in 3 dimensions) with zero divergence. And of course if the particles of such a fluid occupy a region R_0 at time t_0 they will occupy a region R_1 of equal volume at time t_1, or else the density would be different. Liouville is the $2s$-dimensional generalization of this, and one often says that the flow in phase space of a mechanical system with a Hamiltonian is "incompressible."

★ The perspective in TM is slightly different, and can be thought of in the following way. Suppose one has an enormous number of identical mechanical systems—same phase space, same Hamiltonian—but with different initial conditions. Then it is plausible that one can assign a "density" ρ, i.e., a number per unit "volume" in the phase space in the sense of (28) Given density and velocity in the phase space, we form the *current density* $\vec{J}$

$$\vec{J} = \rho \vec{v}, \tag{31}$$

This is the analog of a mass current density in fluid dynamics. Since the total number of systems represented by points in the phase space does not change as a function of time, we expect $\vec{J}$ to satisfy a conservation law

$$\partial \rho / \partial t = -\nabla \cdot \vec{J}, \tag{32}$$

which is the analog of mass conservation in a fluid, including the compressible kind. Charge conservation in electricity leads to the same type of equation. Thus we expect (but have not proved) that an equation of the form (32) will also hold in phase space.

∘ Indeed, Eqn. (7.195) on p. 277 of TM is precisely (32), and is correct, as are the conclusions they draw later down on that page. Their derivation of Eqn. (7.195), on the other hand, is badly flawed, even by the standards of physicists.

★ The final result in TM, their (7.198),

$$d\rho/dt = 0,$$ (33)

is to be interpreted in the following way. Think of ρ as a function of the q's and p's, $\rho(\vec{q}, \vec{p}, t)$. Then given $\vec{q}(t)$ and $\vec{p}(t)$ as functions of time on some phase space trajectory that satisfies the equations of motion, one can insert these in ρ, and $\rho(\vec{q}(t), \vec{p}(t), t)$ is, somewhat surprisingly, independent of time. Even though $\partial\rho/\partial t$ at some fixed point in phase space will in general depend on the time.

∘ One thinks of the group of 40 wealthy retired folk from Philadelphia who went together on a trip around around the world and were surprised to find American tourists everywhere they went.

• The result (33) is, at least intuitively, equivalent to Liouville's theorem in the form (29). The reason is that if phase space volume is preserved as time progresses, the number of points representing different systems in the region R_1 will be the same as the number of points in R_0, and since density ρ is number per unit volume, the density will remain fixed. Conversely, if one knows that ρ is fixed as a function of time in the sense of (33), i.e., in the vicinity of a point moving in phase space in accordance with Hamilton's equations, then the volume occupied by a set of points must remain fixed.

A   Appendix. Matrices and column/row vectors

★ A matrix is a square or rectangular array of numbers (or symbols, etc.) with m rows and n columns. The elements of a matrix M are labeled by subscripts, e.g., M_{jk}, where the first, j, labels the row and the second, k, the column. Thus

$$M = \begin{pmatrix} 1 & 0 & 2 \\ 3 & -3.5 & 16 \end{pmatrix}$$ (34)

is a matrix with $M_{11} = 1$, $M_{12} = 0$, $M_{13} = 2$, $M_{21} = 3$, $M_{22} = -3.5$, $M_{23} = 16$.

• Sometimes row and columns are labeled (indexed) starting with 0 instead of with 1. People who use such labels or indexing would say that $M_{00} = 1$, $M_{01} = 0$, ... $M_{12} = 16$.

• If $n = 1$, one column, the matrix is referred to as a *column vector* of *dimension* m, i.e., it has m elements. If $m = 1$, only one row, the matrix is called a *row vector* of *dimension* n.

★ An important operation on matrices is the *transpose*, denoted by M^{T}. This consists of interchanging rows and columns so that the first row of M is the first column of M^{T}, the second row is the second column, etc. Thus for the example in (34),

$$M^{\mathrm{T}} = \begin{pmatrix} 1 & 3 \\ 0 & -3.5 \\ 2 & 16 \end{pmatrix}$$ (35)

• The transpose of a column vector is a row vector and vice versa.
• Transposing the transpose results in the original: $(M^{\mathrm{T}})^{\mathrm{T}} = M$.

★ Matrices can be *multiplied*; this is perhaps their most important property. Thus if A is an $m \times n$ matrix and B is an $n \times p$ matrix, $C = A \cdot B$ or simply AB with no separating dot, is the $m \times p$ matrix

$$C_{jl} = \sum_{k=1}^{n} A_{jk} B_{kl}$$ (36)

• Note that multiplication is only defined when the number of columns of the first factor is equal to the number of rows of the second. Let us call this *compatibility*. In what follows we will assume multiplication involves compatible matrices.

• Matrix multiplication is associative, $A \cdot (B \cdot C) = (A \cdot B) \cdot C$, but is *noncommutative*: in general $A \cdot B$ and $B \cdot A$ are not the same. Order is important!

• Transpose of a product

$$(A \cdot B)^{\mathrm{T}} = B^{\mathrm{T}} \cdot A^{\mathrm{T}} \tag{37}$$

□ Exercise. Prove it. Find the corresponding expression for $(A \cdot B \cdot C)^{\mathrm{T}}$.

★ Let A be a matrix, c a column vector and r a row vector. Then $A \cdot c$ is a column vector, $r \cdot A$ is a row vector, $r \cdot c$ is a single number (1×1 matrix), and the matrix $c \cdot r$ is called a *dyad*.

□ Exercise. What are the compatibility conditions for each of the operations just mentioned?

□ Exercise. How big is the dyad in terms of the dimensions of c and r?

★ One is often interested in square matrices of a given size, say $n \times n$. A very important square matrix is the *identity* I defined by

$$I_{jk} = \delta_{jk} \quad \text{or} \quad I = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix} \tag{38}$$

It has the property that $I \cdot A = A \cdot I = A$ for any $n \times n$ square matrix A.

★ If for a square matrix M there exists another matrix N of the same size such that

$$N \cdot M = I; \quad M \cdot N = I, \tag{39}$$

N is called the *inverse* of M, and M is the inverse of N. Often one writes $N = M^{-1}$.

• An N which satisfies the first of these two equations is called a *left* inverse, and one that satisfies the second equation a *right* inverse. In the case of square matrices a left inverse is always a right inverse and vice versa. Furthermore the inverse, if it exists, is unique.

★ Not all square matrices have inverses. If M has an inverse it is said to be *invertible* or *nonsingular*; if there is no inverse it is said to be *singular*.

□ Exercise. A matrix all of whose elements are zero has no inverse. Prove this, and then give a less trivial example of a singular 2×2 matrix.

★ The inverse of a 1×1 matrix (r) is $(1/r)$ if $a \neq 0$. A 2×2 matrix has an inverse

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \tag{40}$$

provided the determinant $ad - bc$ is not zero.

□ Exercise. Check this by working out the product of this matrix with its inverse.

• An $n \times n$ matrix M has an inverse if and only if its determinant is nonzero. The topic of determinants lies outside the scope of these notes.

• Whereas formulas comparable to (40) exist for 3×3, 4×4, ... matrices, they are rather unwieldy, and it is often better to have your computer find the inverse.

★ The inverse of a product of several square matrices exists if and only if each individual matrix is nonsingular. The inverse of a product is given by the product of the inverses in reverse order:

$$(A \cdot B \cdot C \cdots Z)^{-1} = Z^{-1} \cdots B^{-1} \cdot A^{-1}. \tag{41}$$

□ Exercise. Prove it.

□ Exercise. Show that $(A^{-1})^{\mathrm{T}} = (A^{\mathrm{T}})^{-1}$ if A is invertible.