INTRODUCTION TO GEOMETRIC ALGEBRA

DIEGO HERRERA

1. Basics of Newtonian Mechanics

Dot variables to denote time derivatives; *e.g.*, for $\mathbf{x}(t) \in \mathcal{E}^3$,

$$\dot{\mathbf{x}}(t) \doteq \frac{d}{dt}\mathbf{x}(t) \ .$$

Add another dot to denote second order time derivatives; *e.g.*,

$$\ddot{\mathbf{x}}(t) \doteq \frac{d}{dt}\dot{\mathbf{x}}(t) = \frac{d}{dt}\frac{d}{dt}\mathbf{x}(t) = \frac{d^2}{dt^2}\mathbf{x}(t) \ .$$

1.1. Basic Definitions.

Model *Newton's absolute time* by $t \in \mathcal{R}$, *Newton's absolute space* by $\mathcal{E}^3$, a *particle's position* in space, $\mathbf{x}$, by a map $t \mapsto \mathbf{x}(t)$, $\mathbf{x}(t) \in \mathcal{E}^3$, and the *particle's mass*, m, by a map $t \mapsto m(t)$, $m(t) \in \mathcal{E}^3$.

Shorten the value of functions by the functions name; *e.g.*, use $\mathbf{x}$ instead of $\mathbf{x}(t)$, m instead of $m(t)$ and so forth.

Consider a particle of mass m moving through space. Model the *particle's trajectory* via $\mathbf{x}$. Define the *particle's velocity* by $\mathbf{v} \doteq \dot{\mathbf{x}}$, the *particle's acceleration* by $\mathbf{a} \doteq \dot{\mathbf{v}}$ and the *particle's linear momentum* by $\mathbf{p} \doteq m\mathbf{v}$. *Newton's Second Law* governs particle's motion

$$\mathbf{f}(\mathbf{x}, \mathbf{v}) = \dot{\mathbf{p}} = m\mathbf{a} + \dot{m}\mathbf{v}\,,$$

where $\mathbf{f}$ denotes a force acting on a particle. Note that, in general, a force depends on the particle's position as well as on the particle's velocity. As a particular case, if the particle's mass remains constant during motion; *i.e.*, $\dot{m} = 0$, then Newton's law reduces to $\mathbf{f} = m\mathbf{a}$.

The *goal of classical mechanics* is to find $\mathbf{x}$ for all t from Newton's second law. Theorems on differential equations guarantee that it is indeed possible to find $\mathbf{x}$ for all t if $\mathbf{x}$ and $\mathbf{v}$ are given at an initial time t_0.

A *free particle* is a particle onto which no force acts; *i.e.*, a particle subject to no interaction. An *inertial frame of reference* is a frame in which a free particle with constant mass travels in a straight line: $\mathbf{x} = \mathbf{x}_0 + t\mathbf{v}$. Newton's first law states that such an inertial frame exists. Newton's second law is valid only in those frames.

1.1.1. *Angular Momentum.*

Consider a moving particle. Define the *particle's angular momentum* by $\mathbf{L} \doteq \mathbf{x} \times \mathbf{p}$ and a *particle's torque* by $\boldsymbol{\tau} \doteq \mathbf{x} \times \mathbf{f}$. Both $\mathbf{L}$ and $\boldsymbol{\tau}$ depend on where the origin is taken. In other words, angular momentum and torque are measured with respect to a particular point.

Find Newton's second law for angular momentum by cross multiplying Newton's second law by $\mathbf{x}$ and using the fact that $\dot{\mathbf{x}}$ and $\mathbf{p}$ are parallel (thus $\mathbf{p} \times \mathbf{v} = 0$): $\frac{d}{dt}(\mathbf{x} \times \mathbf{p}) = \mathbf{x} \times \dot{\mathbf{p}} + \mathbf{p} \times \mathbf{v}$; therefore,

$$\boldsymbol{\tau} = \dot{\mathbf{L}}\,.$$

Date: March 10, 2013.

Key words and phrases. geometric algebra, geometric product, scalars, vectors, multivectors, index notation.

1.1.2. *Conservation Laws.* Define a *central force* an interaction in which $\mathbf{f}$ is parallel to $\mathbf{x}$; *i.e.*, $\boldsymbol{\tau} = \mathbf{x} \times \mathbf{f} = 0$.

From Newton's second laws, two conservation laws follow:

(1) If $\mathbf{f} = 0$, then $\mathbf{p}$ is constant throughout the particle's motion; and
(2) If $\boldsymbol{\tau} = 0$, then $\mathbf{L}$ is constant throughout the particle's motion.

Note that $\boldsymbol{\tau} = 0$ does not require $\mathbf{f} = 0$, so a particle could still interact with a force and still not be subject to any torque. Gravitational interaction is an example of a central force.

1.1.3. *Energy.* Consider a particle of mass m describing a curve $\mathbf{x}$ in time t. Define the *particle's kinetic energy* by $2T \doteq m\mathbf{v} \cdot \mathbf{v} = m\dot{\mathbf{x}} \cdot \dot{\mathbf{x}}$.

Assume constant particle's mass. Calculate the rate of change of kinetic energy with time: $\dot{T} = \dot{\mathbf{p}} \cdot \mathbf{v} = \mathbf{f} \cdot \mathbf{v}$. If the particle travels from position $\mathbf{x}_1$ at time t_1 to position $\mathbf{x}_2$ at time t_2, then the change in kinetic energy becomes

$$T(t_2) - T(t_1) = \int_{t_1}^{t_2} \frac{dT}{dt}\, dt = \int_{t_1}^{t_2} \mathbf{f} \cdot \mathbf{v}\, dt = \int_{\mathbf{x}_1}^{\mathbf{x}_2} \mathbf{f} \cdot d\mathbf{x}\,,$$

where the last equality involving the integral of the force over the path is called the *work done by the force*.

A *conservative force* is a force that depends only on position, rather than on position and velocity; *i.e.*, $\mathbf{f}(t, \mathbf{x})$. The work done by a conservative force is path *independent.* In particular, for a closed path, the work done vanishes

$$\oint \mathbf{f} \cdot d\mathbf{x} = 0 \iff \nabla \times \mathbf{f} = 0\,.$$

This property implies that conservative forces may be written as a function of a *potential*, V:

$$\mathbf{f} = -\nabla V\,.$$

Since conservative forces depend only on position, then their potentials also depend on position only; *i.e.*, $V(\mathbf{x})$. Systems that admit a potential of this form include gravitational, electrostatic and interatomic forces.

Define the *total energy* by $E \doteq T + V$. This energy is conserved in systems interacting with conservatives forces.

Proof. In the equation for the work done by a force, replace the conservative force for its potential:

$$T(t_2) - T(t_1) = -\int_{\mathbf{x}_1}^{\mathbf{x}_2} \nabla V \cdot d\mathbf{x}\,,$$

or, rearranging terms,

$$T(t_1) + V(t_1) = T(t_2) + V(t_2)\ ,$$

where the total energy is a *constant of motion* or, in other words, energy is conserved. □

Call *Hamiltonian*, $\mathcal{H}$, the total energy that depends on position, $\mathbf{x}$, and momentum, $\mathbf{p}$; *i.e.*, $\mathcal{H}(\mathbf{x}, \mathbf{p})$.

1.1.4. *Examples.*

Example 1.1 (The Simple Harmonic Oscillator)**.** *Describe the dynamics of a simple harmonic oscillator. This is an object that moves in one-dimension subject to a force proportional to the distance to the origin.*

Solution. Denote by x the distance from the origin to the object. Thus, the force becomes $\mathbf{f}(x) = -kx$, where k denotes a proportionality constant. Because the particle moves in one-dimension, then angular momentum is undefined. Since the force depends only on position – a conservative force, then such a force arises from a potential $2V = kx^2$; but momentum is not conserved since $\mathbf{f} \neq \mathbf{0}$ – the object oscillates forwards and backwards. However, total energy is conserved and given by $2E = m\dot{x}^2 + kx^2$. □

Example 1.2 (The Damped Simple Harmonic Oscillator)**.** *Describe the dynamics of a damped simple harmonic oscillator. This is a simple harmonic oscillator subject to an additional force: friction. Assume a frictional force proportional to the object's velocity.*

Solution. Use the same notation as in the previous example. Find the frictional force by $\mathbf{f}(\dot{x}) = -\gamma\dot{x}$, where γ a proportionality constant. Thus, the total force on the object is given by $\mathbf{f}(x, \dot{x}) = -kx - \gamma\dot{x}$. Such a force is not conservative, thus energy is not conserved either. This system loses energy until it comes to rest. □

Example 1.3 (Particle Moving Under Gravity)**.** *Describe the dynamics of a particle pulled by gravity. Consider Newtonian gravity theory.*

Solution. Consider a particle of mass m moving in three dimensions under the gravitational pull of a much larger particle of mass M, $M \gg m$, and separated by a distance $\mathbf{x}$.

According to Newton's gravity law, the force between the particles is described by $\mathbf{f} = -(GMm/|\mathbf{x}|^2)\hat{\mathbf{x}}$, where G is a constant. This force arises from the potential $V = -GMm/|\mathbf{x}|$. Again, the linear momentum $\mathbf{p}$ of the smaller particle is not conserved ($\mathbf{f} \neq \mathbf{0}$), but the force is both central and conservative; *i.e.*, $\mathbf{f} \parallel \mathbf{x}$ and $\mathbf{f}(\mathbf{x})$. A central, conservative force ensures that the particle's total energy E and the angular momentum $\mathbf{L}$ are conserved. □

1.2. **The Principle of Least Action.** Consider a set of N particles moving in space due to conservative forces. Denote by $\mathbf{x}_i$ the position of the ith particle. Write the coordinates of the particles' positions as x_A, where $A = 1, \ldots, 3N$. Then, call *degrees of freedom* the coordinates of the particles' positions; *i.e.*, $3N$.

Since the particles interact with conservative forces, then use the potential instead the total force on the particles on Newton's equation of motion:

$$\dot{p}_A = -\frac{\partial V}{\partial x^A}\,,$$

where $p_A = m_A\dot{x}_A$. These parameterise a $3N$-dimensional space known as the configuration space C. Each point in C specifies a configuration of the system (*i.e.*, the positions of all N particles). Time evolution gives rise to a curve in C.

1.2.1. *Lagrangian.* The kinetic energy of all particles is given by

$$T = \frac{1}{2}\sum_A m_A(\dot{x}^A)^2$$

and the potential energy by $V\big(x^A\big)$.

Define the *Lagrangian* to be a function of the positions x^A and the velocities $\dot{x}^A$ of all particles by

$$\mathcal{L}\Big(x^A, \dot{x}^A\Big) \doteq T\Big(\dot{x}^A\Big) - V\Big(x^A\Big)\,.$$

Consider all smooth paths $x^A(t)$ in C with fixed end points so that

$$x^A(t_i) = x^A_{\text{initial}} \qquad \text{and} \qquad x^A(t_f) = x^A_{\text{final}}\,.$$

Only one of these paths is the actual one. To find this out, assign a number to each path, called the *action*, defined by

$$\mathcal{S}\Big(x^A(t)\Big) \doteq \int_{t_\text{i}}^{t_\text{f}} \mathcal{L}\Big(x^A(t)\,, \dot{x}^A(t)\Big)\,dt\,.$$

Note that the action is a functional (*i.e.*, a function of the path which is itself a function). The *principle of least action* is the following result:

Theorem 1.1. *The actual path taken by the system extremum of $\mathcal{S}$.*

Proof. The requirement that the action is an extremum says that $\delta\mathcal{S} = 0$ for all changes in the path $\delta x^A(t)$. This holds if and only if

$$\frac{\partial \mathcal{L}}{\partial x^A} - \frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{x}^A}\right) = 0\,,$$

for each $A = 1, 2, 3, \ldots, 3N$. Call the last equations *Lagrange's equations*. □

Lagrange's equations are equivalent to Newton's.

Some remarks on this theorem:

- ◇ This is an example of a variational principle.
- ◇ A *Principle of stationary action* would be a more accurate name. It is sometimes called *Hamilton's principle*.
- ◇ All the fundamental laws of physics can be written in terms of an action principle. This includes electromagnetism, general relativity, the standard model of particle physics, and attempts to go beyond the known laws of physics such as string theory.
- ◇ There are two very important reasons for working with Lagrange's equations rather than Newton's. The first is that Lagrange's equations hold in any coordinate system, while Newton's are restricted to an inertial frame. The second is the ease with which we can deal with constraints in the Lagrangian system.
- ◇ Part of the power of the Lagrangian formulation over the Newtonian approach is that it does away with vectors in favor of more general coordinates.

Example 1.4 (Hyperbolic Motion)**.** *Find the equations of motion of a particle that moves in the (x, y) plane with a force directed towards the origin $\mathcal{O}$ and magnitude proportional to the distance from $\mathcal{O}$.*

Solution. Motion happens in $\mathcal{E}^2$. Model then the particle's position by $\mathbf{x} = (x, y)$, where x and y are the distances (coordinates) to $\mathbf{x}$ from $\mathcal{O}$. (The particle's degree of freedom is two: x and y.)

Calculate the force acting on the particle: $\mathbf{f} = k\mathbf{x}$, where k is a proportionality constant. Since this force depends only on the position, then it's conservative and arises therefore from a potential; *viz.*, $\mathbf{f} = -\nabla V$. Find thus the potential energy:

$$\nabla V = -k\mathbf{x} \implies V = \frac{1}{2}k\mathbf{x}^2 = \frac{1}{2}k(x^2 + y^2)\,.$$

Additionally, find the kinetic energy:

$$T = \frac{1}{2}m\mathbf{v}^2 = \frac{1}{2}m\dot{\mathbf{x}}^2 = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2)$$

and the Lagrangian:

$$\mathcal{L} = \frac{1}{2}k(x^2 + y^2) - \frac{1}{2}m(\dot{x}^2 + \dot{y}^2)\,.$$

Construct next Lagrange's equations for all x^As; *i.e.*, construct

$$\frac{\partial \mathcal{L}}{\partial x} - \frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{x}}\right) = 0 \qquad \text{and} \qquad \frac{\partial \mathcal{L}}{\partial y} - \frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{y}}\right) = 0\,,$$

or, replacing variables and operating,

$$-kx - m\ddot{x} = 0 \qquad \text{and} \qquad -ky - m\ddot{y} = 0\,.$$

Finally, derive the equations of motion from Langrange's equations:

$$\ddot{x} = -\frac{k}{m}x \qquad \text{and} \qquad \ddot{y} = -\frac{k}{m}y\,;$$

that is,

$$\mathbf{a} = \dot{\mathbf{x}} = \ddot{\mathbf{x}} = -\kappa\mathbf{x}\,,$$

where $\kappa = k/m$. The last equation yields the desired result. □

1.3. **Constraints and Generalized Coordinates.** In writing $\dot{\mathbf{p}}_i = -\nabla_i V$, we implicitly assume that each particle can move freely in space $\mathcal{E}^3$. What if there are constraints? In Newtonian mechanics, we introduce *constraint forces*. These are things like the tension of ropes and normal forces applied by surfaces. In the Lagrangian formulation, we don't have to worry about such things. In this section, we'll show why.

Example 1.5 (The Simple Pendulum)**.** *Find the equations of motion of a linear pendulum of length l and mass m subject to gravitational interaction.*

Solution. The simple pendulum has a single dynamical degree of freedom θ, the angle the pendulum makes with the vertical. The position of the mass m in the plane is described by two Cartesian coordinates x and y subject to a constraint $x^2 + y^2 = l^2$. Parameterise this as $x = l\sin\theta$ and $y = l\cos\theta$. Employing Newton's method to solve this system, introduce the tension T and resolve the force vectors to find

$$m\ddot{x} = -Tx/l \qquad \text{and} \qquad m\ddot{y} = mg - Ty/l\,.$$

To determine the motion of the system, impose the constraints at the level of the equation of motion to find

$$\ddot{\theta} = -(g/l)\sin\theta \qquad \text{and} \qquad T = ml\dot{\theta}^2 + mg\cos\theta\,,$$

which was required. □

While this example was straightforward to solve using Newtonian methods, things get rapidly harder when we consider more complicated constraints. Moreover, you may have noticed that half of the work of the calculation went into computing the tension T. On occasion we'll be interested in this. (For example, we might want to know how fast we can spin the pendulum before it breaks). But often we won't care about these constraint forces, but will only want to know the motion of the pendulum itself. In this case it seems like a waste of effort to go through the motions of computing T. We'll now see how we can avoid this extra work in the Lagrangian formulation. Firstly, let's define what we mean by constraints more rigorously.

1.3.1. *Holonomic Constraints. Holonomic Constraints* are relationships between the coordinates of the form

$$f_\alpha(x_A, t) = 0 \qquad \alpha = 1, 2, 3, \ldots, 3N - n\,.$$

In general the constraints can be time dependent and our notation above allows for this. Holonomic constraints can be solved in terms of n *generalised coordinates* q_i, $i = 1, \ldots, n$. So

$$x_A = x_A(q_1, \ldots, q_n)\ .$$

The system is said to have n degrees of freedom. For the pendulum example above, the system has a single degree of freedom, $q = \theta$.

Often we don't care about the tension T or other constraint forces, but only want to know what the generalized coordinates q_i are doing. In this case we have the following useful theorem.

Theorem 1.2. *For constrained systems, derive the equations of motion directly in generalized coordinates q_i; i.e., using the Langrangian*

$$\mathcal{L}(q_i, \dot{q}_i, t) = \mathcal{L}\Big(x^A(q_i, t)\,, \dot{x}^A(q_i, \dot{q}_i, t)\Big)\ .$$

Example 1.6 (The Simple Pendulum)**.** *Using generalized coordinates, find the equations of motion of a linear pendulum of constant length l and constant mass m subject to gravitational interaction.*

Solution. Parametrise the constraints in terms of the generalized coordinate θ by $x = l\sin\theta$ and $y = l\cos\theta$.

Calculate the kinetic energy. In Cartesian coordinates, the kinetic energy is given by $2T = m(\dot{x}^2 + \dot{y}^2)$. So the time derivatives become $\dot{x} = l\dot{\theta}\cos\theta$ and $\dot{y} = -l\dot{\theta}\sin\theta$, since the pendulum length is constant, $\dot{l} = 0$. Thus, $\dot{x}^2 + \dot{y}^2 = l^2\dot{\theta}^2$, because of the trigonometric identity $\sin^2\theta + \cos^2\theta = 1$. Finally, find the kinetic energy as a function of the parameter θ

$$T = \frac{1}{2}ml^2\dot{\theta}^2\,.$$

On the other hand, calculate the potential energy. Gravitational interaction depends only on position; so, it arises from a potential. Then, this potential is given by, using the parameter θ,

$$V = mgy = mgl\cos\theta\,.$$

Substitute the potential and kinetic energies into the Lagrangian, to get

$$\mathcal{L} = \frac{1}{2}ml^2\dot{\theta}^2 - mgl\cos\theta\,.$$

From the Lagrangian, find the equations of motion using Lagrange's equation

$$\frac{\partial\mathcal{L}}{\partial\theta} - \frac{d}{dt}\left(\frac{\partial\mathcal{L}}{\partial\dot{\theta}}\right) = ml^2\ddot{\theta} + mgl\dot{\theta}\sin\theta = 0\,,$$

or, equivalently, rearranging terms and simplifying,

$$\ddot{\theta} + \kappa\dot{\theta}\sin\theta = 0\,,$$

where $\kappa = g/l$. The last equation yields the desired result. □

Note that, as promised, we haven't calculated the tension T using this method. This has the advantage that we've needed to do less work. If we need to figure out the tension, we have to go back to the more laborious Lagrange multiplier method. Another advantage is that the degrees of freedom have reduced from two, x and y, to one, θ.

1.4. **Summary.** A system is described by n generalised coordinates q_i which define a point in a n-dimensional configuration space C. Time evolution is a curve in C governed by the Lagrangian $\mathcal{L}(t, q_i, \dot{q}_i)$ such that the q_i obey

$$\frac{\partial\mathcal{L}}{\partial q_i} - \frac{d}{dt}\left(\frac{\partial\mathcal{L}}{\partial\dot{q}_i}\right) = 0\,.$$

These are n coupled second order (usually) non-linear differential equations.

Before we move on, let's take this opportunity to give an important definition. The quantity

$$p_i = \frac{\partial\mathcal{L}}{\partial\dot{q}_i}$$

is called the *generalized momentum* conjugate to q_i. (It only coincides with the real momentum in Cartesian coordinates). We can now rewrite Lagrange's equations

$$\dot{p}_i = \frac{\partial\mathcal{L}}{\partial q_i}\,.$$

1.5. **Examples.**

Example 1.7. *Find the equations of motion for a particle of mass m moving under the interaction of a central potential ψ. Use Langrangian description of motion and spherical coordinates.*

Solution. Specify the position of the particle x by the radial distance of that point from a fixed origin r, the polar angle θ measured from a fixed zenith direction and the azimuth angle ϕ of its orthogonal projection on a reference plane that passes through the origin to the zenith, measured from a fixed reference direction on that plane. Call also the radial distance the radius or radial coordinate and the polar angle colatitude, zenith angle, normal angle or inclination angle. Additionally, this specification defines a frame $\{\gamma_k\} = \{\gamma_r, \gamma_\theta, \gamma_\phi\}$.

Determine then the trajectory of the particle at any time t by $x = x(t)$; *i.e.*, $r = r(t)$, $\theta = \theta(t)$ and $\phi = (t)$. Calculate the position of the particle by finding the components of x onto the frame

$$x = \gamma_r x^r + \gamma_\theta x^\theta + \gamma_\phi x^\phi = \gamma_r r + \gamma_\theta\theta + \gamma_\phi\phi\,.$$

Not only do the components change with t, so do the frame elements; thus the displacement vector $\mathrm{d}x$ becomes

$$\mathrm{d}x = \mathrm{d}r\,\gamma_r + r\,\mathrm{d}\theta\,\gamma_\theta + r\sin\theta\,\mathrm{d}\phi\,\gamma_\phi\,.$$

Agree on using the dot notation for time derivatives; *i.e.*, $\dot{x} \doteq \mathrm{d}x/\mathrm{d}t$.

To find the velocity of the particle v, divide the displacement vector by $\mathrm{d}t$

$$v = \dot{x} = \dot{r}\,\gamma_r + r\,\dot{\theta}\,\gamma_\theta + r\sin\theta\,\dot{\phi}\,\gamma_\phi\,.$$

The squared magnitude of the velocity is then $v^2 = vv = v \cdot v$, because v is colinear to itself, $v \wedge v = 0$. In terms of compontents, the squared magnitude is $v^2 = (\dot{r}^2 + r^2\,\dot{\theta}^2 + r^2\sin^2\theta\,\dot{\phi}^2)$.

Using the squared magnitude of the velocity, calculate the particle kinetic energy K defined by $2K \doteq mv^2$:

$$K = \frac{1}{2}\big(\dot{r}^2 + r^2\,\dot{\theta}^2 + r^2\sin^2\theta\,\dot{\phi}^2\big)\,,$$

since, by construction, the frame elements are orthogonal, then they satisfy $\gamma_k\gamma_l = 1$ for $k = l$.

Calculate next the Langragian $\mathcal{L}$ using the kinetic energy K and the potential energy $\psi = \psi(x)$:

$$\mathcal{L} = K - \psi = \frac{1}{2}\big(\dot{r}^2 + r^2\,\dot{\theta}^2 + r^2\sin^2\theta\,\dot{\phi}^2\big) - \psi\,,$$

Compute the equations of motion using the Langrange-Euler equations for $\{r, \theta, \phi\}$:

$$\begin{aligned}
\frac{\partial\mathcal{L}}{\partial r} - \frac{d}{dt}\left(\frac{\partial\mathcal{L}}{\partial\dot{r}}\right) = 0 &\implies m(\ddot{r} - r\dot{\theta}^2\cos^2\phi - r\dot{\phi}^2) + \psi_{,r} = 0\\
\frac{\partial\mathcal{L}}{\partial\theta} - \frac{d}{dt}\left(\frac{\partial\mathcal{L}}{\partial\dot{\theta}}\right) = 0 &\implies m(2\dot{r}\dot{\theta}\cos\phi + r\ddot{\theta}\cos\phi - 2r\dot{\theta}\dot{\phi}\sin\phi) + \psi_{,\theta} = 0\\
\frac{\partial\mathcal{L}}{\partial\phi} - \frac{d}{dt}\left(\frac{\partial\mathcal{L}}{\partial\dot{\phi}}\right) = 0 &\implies m(2\dot{r}\dot{\phi} + r\dot{\phi}^2\sin\phi\cos\phi + r\ddot{\phi}) + \psi_{,\phi} = 0\,,
\end{aligned}$$

which yields the desired result. □

1.6. **Halmiltonian.** Consider Newtonian mechanics. Consider a free particle of time-constant mass m moving with non-zero velocity v relative to an observer.

Since no forces, or, equivalently, no potentials, act on the particle – free particle – the particle kinetic energy K, given by $2K = mv^2$, equals the total energy E. Note that energy depends only on velocity: $E = E(v)$.

Next, find the derivative of K with respect to v: $\mathrm{d}K/dxv = mv$.; *i.e.*, the particle linear momentum p. Rewrite the total energy in terms of momentum: $E = p^2/m$. This is to say, $E = E(p) = p^2/m$.

The trick to learn is that velocity was replaced by momentum via derivation of the kinetic energy with respect to velocity.

2. Noether's Theorem and Symmetries

A function $F(q_i, \dot{q}_i, t)$ of the coordinates, their time derivatives and (possibly) time t is called a constant of motion (or a conserved quantity) if the total time derivative vanishes

$$\frac{dF}{dt} = \sum_{j=1}^{n}()\,.$$

3. Partial Derivatives

You probably first encountered *ordinary derivatives* when you learned how to find the slope of a line $m = dy/dx$ or how to determine the speed of an object given its position as a function of time $\mathbf{v}(t) = \dot{\mathbf{x}}(t)$. *Partial derivatives* are based on the same general concepts as ordinary derivatives, but extend those concepts to *functions of multiple variables*. You should never have any doubt as to which kind of derivative you're dealing with, because ordinary derivatives are written as d/dx or d/dt and partial derivatives are written as $\partial/\partial x$ or $\partial/\partial t$ (sometimes as ∂_x or ∂_t).

Ordinary derivatives come about when you're interested in the *change* of one variable with respect to another. For example, you may encounter a variable y which is a function of another variable x (which means that the value of y depends on the value of x). This can be written as $y = f(x)$, where y is called the » dependent variable « and x is called the » independent variable «. The ordinary derivative of y with respect to x (written as dy/dx) tells you how much the value of y changes for a small change in the variable x. If you make a graph with y on the vertical axis and x on the horizontal axis, then the slope of the line between any two points (x_1, y_1) and (x_2, y_2) on the graph is simply $(y_2 - y_1)/(x_2 - x_1) = \Delta y/\Delta x$. That's because the slope is defined as » the rise over the run «, and since the rise is y for a run x, the slope of the line between any two points must be $\Delta y/\Delta x$. But if you look closely at an expanded region, you'll notice that the graph of y versus x has a slight curve between points (x_1, y_1) and (x_2, y_2), so the slope is actually changing in that interval. Thus the ratio can't represent the slope everywhere between those points. Instead, it represents the average slope over this interval, as suggested by a line between points (x_1, y_1) and (x_2, y_2) (which by the mean value theorem does equal the slope somewhere in between these two points, but not necessarily in the middle). To represent the slope at a given point on the curve more precisely, all you have to do is to allow the run Δx to become very small. As Δx approaches zero, the difference between the dashed line and the curved line becomes negligible. If you write the incremental run as dx and the (also incremental) rise as dy, then the slope at any point on the line can be written as dy/dx. This is the reasoning that equates the derivative of a function to the slope of the graph of that function.

Now imagine that you have a variable z that depends on two other variables, say x and y, so $z = f(x, y)$. One way to picture such a case is to visualize a surface in three-dimensional space. The height of this surface above the xy plane is z, which gets higher and lower at different values of x and y. And since the height z may change at a different rate in different directions, a single derivative will not generally be sufficient to characterize the total change in height as you move from one point to another. You can see the height z changing at different rates; the slope of the surface is quite steep if you move in the direction of increasing y (while remaining at the same value of x), but the slope is almost zero if you move in the direction of increasing x (while holding your y-value constant).

This illustrates the usefulness of *partial* derivatives, which are derivatives formed by allowing one independent variable (such as x or y) to change while holding other independent variables constant. So the partial derivative $\partial z/\partial x$ represents the slope of the surface at a given location if you move only along the x-direction from that location, and the partial derivative $\partial z/\partial y$ represents the slope if you move only along the y-direction. You may find these partial derivatives written as $\partial z/\partial x|_y$ and $\partial z/\partial y|_x$, where the variables that appear in the subscript after the vertical line are held constant. As you've probably already guessed, the change in the value of z as either x or y changes is easily found using partial derivatives. If only x changes, $dz = \partial z/\partial x\, dx$, and if only y changes, then $dz = \partial z/\partial y\, dy$. And if both x and y change, then

$$dz = \frac{\partial z}{\partial x}\, dx + \frac{\partial z}{\partial y}\, dy\,.$$

The process of taking a partial derivative of a given function is quite straightforward; if you know how to take ordinary derivatives, you already have the tools you'll need to take partial derivatives. Simply treat all variables (with the exception of the one variable over which the derivative is being taken) as constants, and take the derivative as you normally would. This is best explained using an example.

Example 3.1. *Consider the function $z = f(x, y) = 6x^2y + 3x + 5xy + 10$. Find $\partial_x z$ and $\partial_y z$.*

Solution. To find $\partial_x z$, take the partial derivative of z with respect to x simply by treating the variable y as a constant:

$$\partial_x z = \frac{\partial z}{\partial x} = 12xy + 3 + 5y\,.$$

To find $\partial_y z$, take the partial derivative of z with respect to x simply by treating the variable x as a constant:

$$\partial_y z = \frac{\partial z}{\partial x} = 6x^2 + 5x\,.$$

Both equations yield the desired results. □

Before interpreting these derivative results, you may want to take a moment to make sure you understand why the process of taking the derivative of a function involves bringing down the exponent of the relevant variable and then subtracting one from that exponent (so $dx^2/dx = 2x$, for example). The answer is quite straightforward. Since the derivative represents the change in the function z as the independent variable x changes over a very small run, the formal definition for this derivative can be written as

$$\frac{dz}{dx} = \lim_{\Delta x \to 0} \frac{z(x + \Delta x) - z(x)}{\Delta x} .$$

So in the case of $z = x^2$, you have

$$\frac{dx^2}{dx} = \lim_{\Delta x \to 0} \frac{(x + \Delta x)^2 - x^2}{\Delta x} .$$

If you think about the term in the numerator, you'll see that this is $x^2 + 2x\Delta x + (\Delta x)^2 - x^2$, which is just $2x\Delta x + (\Delta x)^2$, and dividing this by Δx gives $2x + \Delta x$. But as Δx approaches zero, the Δx term becomes negligible, and this approaches $2x$. So where did the 2 come from? It's just the number of cross terms (that is, terms with the product of x and Δx) that result from raising $(x + \Delta x)$ to the second power. Had you been taking the derivative of x^3 with respect to x, you would have had three such cross terms. So you bring down the exponent because that's the number of cross terms that result from taking $x + \Delta x$ to that power. And why do you then subtract one from the exponent? Simply because when you take the change in the function z (that is, $(x + \Delta x)^2 - x^2$), the highest-power terms (x^2 in this case) cancel, leaving only terms of one lower power (x^1 in this case). It's a bit laborious, but the same analysis can be applied to show that $dx^3/dx = 3x$ and that $dx^n/dx = nx^{n-1}$.

Just as you can take *higher order ordinary derivatives* such as $d/dx(dz/dx) = d^2z/dx^2$ and $d/dy(dz/dy) = d^2z/dy^2$, you can also take higher-order partial derivatives. So for example $\partial_{xx}z = \partial^2 z/\partial x^2$ tells you the change in the x-direction slope of z as you move along the x-direction, and $\partial_{yy}z = \partial^2 z/\partial y^2$ tells you the change in the y-direction slope as you move along the y-direction.

It's important for you to realize that an expression such as $\partial^2 z/\partial x^2$ is the derivative of a derivative, which is *not* the same as $(\partial z/\partial x)^2$, which is the square of a first derivative. By convention the order of the derivative is always written between the d or ∂ and the function, as d^2z or $\partial^2 z$, so be sure to look carefully at the location of superscripts when you're dealing with derivatives.

You may also have occasion to use *mixed partial derivatives* such as $\partial_{xy}z = \partial/\partial x(\partial z/\partial y)$. If you've been tracking the discussion of partial derivatives as slopes of functions in various directions, you can probably guess that $\partial_{xy}z$ represents the change in the y-direction slope as you move along the x-direction and $\partial_{yx}z$ represents the change in the x-direction slope as you move along the y-direction. Thankfully, for well-behaved functions these expressions are interchangeable, so you can take the partial derivatives in either order.

There's another widely used aspect of partial derivatives you should make sure you understand, and that's the *chain rule*. Up to this point, we've been dealing with functions such as $z = f(x, y)$ without considering the fact that the variables x and y may themselves be functions of other variables. It's common to call these other variables u and v and to allow both x and y to depend on one or both of u and v; *i.e.*, $z = f(x(u, v), y(u, v))$. You may encounter situations in which you know the variation in u and v, and you want to know how much your function z will change due to those changes. In such cases, the chain rule for partial derivatives gives you the answer:

$$\partial_u z = \frac{\partial z}{\partial u} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial u}$$

and

$$\partial_v z = \frac{\partial z}{\partial v} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial v} .$$

The chain rule is a concise expression of the fact that z depends on both x and y, and since both x and y may change if u changes, the change in z with respect to u is the sum of two terms. The first term is the change in x due to the change in u ($\partial_u x$) times the change in z due to that change in x ($\partial_x z$), and the second term is the change in y due to the change in u ($\partial_u y$) times the change in z due to that change in y ($\partial_y z$). Adding those two terms together gives you $\partial_u z$. The same reasoning applied to changes in z caused by changes in v leads to $\partial_v z$.

4. Vectors as Derivatives

In many texts dealing with vectors and tensors, you'll find that vectors are equated to *directional derivatives* and that partial derivatives such as ∂_x and $\partial/\partial y$ are referred to as basis vectors along the coordinate axes.

To understand this correspondence between vectors and derivatives, consider a path. You can think of this as a path along which you're traveling with velocity $\mathbf{v}$; for simplicity imagine that this path lies in the xy plane. Now imagine that you're keeping track of time as you move, so you assign a value to each point on the curve. By marking the curve with values, you have *parameterized* the curve (with t as your parameter). Note that there need not be equal distance along the curve between your parameter values (there definitely won't be if you choose time as your parameter and then change your speed as you move).

As a final bit of visualization, imagine that this curve lies in a region in which the air temperature is different at each location. So as you move along the curve, you will experience the spatial change in air temperature as a temporal change (in other words, you'll be able to make a graph of air temperature *vs.* time). Of course, how fast the air temperature changes for you will depend both on the distance between measurable changes in the temperature in the direction you're heading and on your speed (how fast you're covering that distance).

With this scenario in mind, the concept of a directional derivative is easy to understand. If the function $f(x, y)$ describes the temperature at each x, y location, the directional derivative df/dt tells you how much the value of the function f changes as you move a small distance along the curve (in time dt). But recall the chain rule:

$$\frac{df}{dt} = \frac{dx}{dt}\frac{\partial f}{\partial x} + \frac{dy}{dt}\frac{\partial f}{\partial y}\,,$$

or, in compact, alternate notation, $\dot{f} = \dot{x} f_x + \dot{y} f_y$.

Notation. An alternate notation involves index notation, dot derivative, comma derivative and Einstein summation convention. Thus, df/dt becomes

$$\dot{f} = \dot{x}^1 f_{,1} + \dot{x}^2 f^{\,,2} = \dot{x}^k f_{,k}\,.$$

This equation says simply that the directional derivative of the function f along the curve parameterized by t (that is, df/dt) equals the rate of change of the x-coordinate (dx/dt) as you move along the curve times the rate of change of the temperature function with x ($\partial_x f$) plus the rate of change of the y-coordinate (dy/dt) as you move along the curve times the rate of change of the temperature function with y $\partial_y f$. But dx/dt is just v_x, the x-component of your velocity, and dy/dt is v_y, the y-component of your velocity. And since you know that your velocity is a vector that is always tangent to the path on which you're moving, you can consider the directional derivative df/dt to be a vector with direction tangent to the curve and with length equal to the rate of change of f with t (that is, the time rate of change of the air temperature).

Now here's the important concept: since f can be any function, you can write the last expression as an *operator* equation (that is, an equation waiting to be fed a function on which it can operate):

$$\frac{d}{dt} = \frac{dx}{dt}\frac{\partial}{\partial x} + \frac{dy}{dt}\frac{\partial}{\partial y}\,.$$

The trick to seeing the connection between derivatives and vectors is to view this equation as a vector equation in which

Vector $=$ x-component times x basis vector $+$ y-component times y basis vector.

Comparing this to the operator equation, you should be able to see that the directional derivative operator d/dt represents the tangent vector to the curve, the dx/dt and dy/dt terms represent the x- and y-components of that vector, and the operators $\partial/\partial x$ and $\partial/\partial y$ represent the basis vectors in the direction of the x and y coordinate axes.

Of course, it's not just air temperature that can be represented by $f(x,y)$; this function can represent anything that is spatially distributed in the region around your curve. So $f(x,y)$ could represent the height of the road, the quality of the scenery, or any other quantity that varies in the vicinity of your curve. Likewise, you could have chosen to parameterize your path with markers other than time; had you assigned a value s or λ to each point on your path, the directional derivative d/ds or $d/d\lambda$ would still represent the tangent vector to the curve, dx/ds or $dx/d\lambda$ would still represent the x-component of that vector, dy/ds and $dy/d\lambda$ or would still represent the y-component of that vector.

If you plan to proceed on to the study of tensors, you will find that understanding this relationship between basis vectors along the coordinate axes and partial derivatives is of significant value.

4.1. **Nabla – the Del Operator.** The partial derivatives discussed in the previous section can be put to use in a wide range of problems, and when you come across such problems you may find that they involve equations that contain an inverted upper case delta wearing a vector hat $\bigtriangledown$. This symbol represents a vector differential operator called *nabla* or *del*, and its presence instructs you to take derivatives of the quantity on which the operator is acting. The exact form of those derivatives depends on the symbol following the del operator, with $\bigtriangledown()$ signifying *gradient*, $\bigtriangledown\cdot$ signifying *divergence*, $\bigtriangledown\times$ indicating *curl*, and $\triangle()$ signifying the *Laplacian*. Each of these operations is discussed in later sections; for now we'll just consider what an operator is and how the del operator can be written in Cartesian coordinates.

Like all good mathematical operators, del is an action waiting to happen. Just as $\sqrt{()}$ tells you to take the square root of anything that appears under its roof, $\bigtriangledown$ is an instruction to take derivatives in three directions. Specifically, in Cartesian coordinates

$$\bigtriangledown \doteq \hat{\imath}\frac{\partial}{\partial x} + \hat{\jmath}\frac{\partial}{\partial y} + \hat{k}\frac{\partial}{\partial z}\,,$$

where $\hat{\imath}$, $\hat{\jmath}$, and $\hat{z}$ are the unit vectors in the direction of the Cartesian coordinates x, y, and z. Alternatively, $\bigtriangledown$, in index notation,

$$\bigtriangledown \doteq \gamma^l\partial_l = \gamma^1\partial_1 + \gamma^2\partial_2 + \gamma^3\partial_3\,.$$

This expression may appear strange, since in this form it's lacking anything on which it can operate. However, if you follow the del with a scalar or vector field, you can extract information about how those fields change in space. In this context, *field* refers to an array or collection of values defined at various locations. A *scalar field* is specified entirely by its magnitude at these locations: examples of scalar fields include the air temperature in a room and the height of terrain above sea level. A *vector field* is specified by both magnitude and direction at various locations: examples include electric, magnetic, and gravitational fields. Specific examples of how the del operator works on scalar and vector fields are given in the following sections.

4.2. **Gradient.** When the del operator is followed by a scalar field, the result of the operation is called the *gradient of the field*. What does the gradient tell you about a scalar field? Two important things:

> the *magnitude* of the gradient indicates how quickly the field is changing over space, and
> the *direction* of the gradient indicates the direction in which the field is increasing most quickly with distance.

So although the gradient operates on a scalar field, the result of the gradient operation is a vector, with both magnitude and direction. Thus, if the scalar field represents terrain height, the magnitude of the gradient at any location tells you how steeply the ground is sloped at that location, and the direction of the gradient points *uphill* along the steepest slope.

The definition of the gradient of the scalar field ψ in Cartesian coordinates is

$$\operatorname{grad}\psi = \bigtriangledown\,\psi \doteq \hat{\imath}\frac{\partial\psi}{\partial x} + \hat{\jmath}\frac{\partial\psi}{\partial y} + \hat{k}\frac{\partial\psi}{\partial z}\,,$$

alternatively, in index notation,

$$\operatorname{grad}\psi = \bigtriangledown\,\psi \doteq \gamma^l\partial_l\psi = \gamma^1\partial_1\psi + \gamma^2\partial_2\psi + \gamma^3\partial_3\psi\,.$$

where the $\{\gamma^l = 1/\gamma_l\}$ are unit vectors in the directions of the Cartesian coordinates x, y and z and the $\partial_l\psi$s are the partial derivatives of ψ with respect to x, y and z.

Notation. Sometimes the *comma derivative notation* is used along with index notation:

$$\operatorname{grad}\psi = \nabla\,\psi \doteq \gamma^k\psi_{,k} = \gamma^1\psi_{,1} + \gamma^2\psi_{,2} + \gamma^3\psi_{,3}\,,$$

where the $\psi_{,l}$s are the partial derivatives of ψ with respect to x, y and z; *e.g.*, $\psi_{,1} = \psi_x = \partial\psi/\partial x^1 = \partial\psi/\partial x$.

Example 4.1. *Consider the scalar field $f(\mathbf{x}) = f(x,y,z) = x^2 + y^2 + z^2$. Find the gradient of f using the comma derivative notation.*

Solution. The gradient of a function $\psi(x^1, x^2, \ldots, x^n)$ in abstract form – using index notation, comma derivative notation and Einstein summation convention – is given by

$$\operatorname{grad}\psi = \nabla\,\psi = \gamma^k\psi_{,k}\,.$$

where the index k runs from 1 to n and $\{\gamma^k = 1/\gamma_k\}$ – in the case of rectangular coordinates, this case, $\{\gamma^k = \gamma_k\}$. Thus, to find the gradient of f, first write f in index notation:

$$f\left(x^k\right) = f(x^1, x^2, x^3) = (x^1)^2 + (x^2)^2 + (x^3)^2\,.$$

Then, calculate the partial derivatives of f with respect to x^1:

$$f_{,1} = 2x^1\,,$$

which is $f_{,1} = f_x = \partial f/\partial x^1 = \partial f/\partial x$. Do the same for the other partial derivatives

$$f_{,2} = 2x^2 \qquad \text{and} \qquad f_{,3} = 2x^3\,.$$

Finally, multiply each $f_{,k}$ by the correspondent γ^k and sum the terms to find the gradient of f:

$$\operatorname{grad} f = 2(\gamma^1x^1 + \gamma^2x^2 + \gamma^3x^3) = 2\gamma^kx^k\,,$$

which yields the desired result. □

4.3. **Divergence.** When dealing with vector fields, you may encounter the del operator followed by a dot: $\nabla\,\boldsymbol{\cdot}$, signifying the *divergence of a vector field* (in GA, divergence is called the *interior derivative of a vector field*). The concept of divergence often arises in the areas of physics and engineering that deal with the spatial variation of vector fields, because

> divergence describes the tendency of vectors to "flow" into or out of a point of interest.

Electrostatic fields, for example, may be represented by vectors that point radially away from points at which positive electric charge exists, just as the flow vectors of a fluid point away from a source (such as an underwater spring). Likewise, electrostatic field vectors point toward locations at which negative charge is present, analogous to fluid flowing toward a sink or drain. It was the brilliant Scottish mathematical physicist James Clerk Maxwell who coined the term "convergence" for the mathematical operation which measures the rate of vector "flow" toward a given location. In modern usage we consider the opposite behavior (vectors flowing away from a point), and *outward flow is considered positive divergence.* In the case of fluid flow,

> the divergence at any point is a measure of the tendency of the flow vectors to diverge from that point (that is, to carry more material away from it than toward it). Thus points of positive divergence mark the location of sources, while points of negative divergence show you where the sinks are located.

To find the locations of positive divergence in each of fields, look for points at which the flow vectors either spread out or are larger pointing away from the location and shorter pointing toward it. Some authors suggest that you imagine sprinkling sawdust on flowing water to assess the divergence; if the sawdust is *dispersed*, you have selected a point of *positive divergence*, while if it becomes *more concentrated*, you've picked a location of *negative divergence.*

The divergence can be defined mathematically by: let $\mathbf{x} \in \mathcal{E}^n$ be the position vector and $\mathbf{u} = \mathbf{u}(\mathbf{x})$ be a vector field. Let ∇ be the geometric derivative. Then, define the *divergence of* $\mathbf{u}$, *aka* the *interior derivative of* $\mathbf{u}$, denoted $\operatorname{div}\mathbf{u}$, by

$$\operatorname{div}\mathbf{u} \doteq \langle \nabla\,\mathbf{u}\rangle_0 = \nabla \cdot \mathbf{u}\,.$$

In rectangular, *aka* Cartesian, coordinates, find the $\operatorname{div}\mathbf{u}$ by

$$\operatorname{div}\mathbf{u} = \nabla \cdot \mathbf{u} = \gamma^k \partial_k \cdot \gamma_l u^l = \partial_k \gamma^k \cdot \gamma_l u^l\,.$$

Use the symmetric property of the inner product and apply the definition of the reciprocal frame $\{\gamma^k\}$ to find that $\gamma^k \cdot \gamma_l = \gamma_k \cdot \gamma^l = \delta_l^k$. With this result, calculate $\operatorname{div}\mathbf{u}$:

$$\operatorname{div}\mathbf{u} = \gamma^k \cdot \gamma_l \partial_k u^l = \partial_k \delta_l^k u^l = \partial_k u^k\,.$$

Consider $\mathcal{E}^3$ and $\mathbf{u} = \mathbf{u}(\mathbf{x})$, then expand the index k to have $\operatorname{div}\mathbf{u}$

$$\operatorname{div}\mathbf{u} = \nabla \cdot \mathbf{u} = \partial_k u^k = \partial_1 u^1 + \partial_2 u^2 + \partial_3 u^3 = u_{\,,x} + u_{\,,y} + u_{\,,z} = \frac{\partial u^x}{\partial x} + \frac{\partial u^y}{\partial y} + \frac{\partial u^x}{\partial z}\,.$$

Thus the divergence of $\mathbf{u}$ is simply the change in its x-component along the x-axis plus the change in its y-component along the y-axis plus the change in its z-component along the z-axis. Notice that the divergence of a vector field is a scalar quantity; it has magnitude but no direction.

4.4. **Exterior Derivative.** Let $\mathbf{x} \in \mathcal{E}^n$ be the position vector and $\mathbf{u} = \mathbf{u}(\mathbf{x})$ be a vector field. Let ∇ be the geometric derivative. Then, define the *exterior derivative of* $\mathbf{u}$, denoted $\nabla \wedge \mathbf{u}$, by

$$\nabla \wedge \mathbf{u} \doteq \langle \nabla\,\mathbf{u}\rangle_2\,.$$

4.5. **Curl.** The del operator followed by a cross, $\nabla\times$, signifies the differential operation of *curl*.

> The curl of a vector field is a measure of the field's tendency to circulate about a point, much like the divergence is a measure of the tendency of the field to flow away from a point.

But unlike the divergence, which produces a scalar result, the curl produces a vector. The magnitude of the curl vector is proportional to the amount of circulation of the field around the point of interest, and the direction of the curl vector is perpendicular to the plane in which the field's circulation is a maximum.

The curl at a point in a vector field can be understood by considering the vector fields shown in Figure. To find the locations of large curl in each of these fields, look for points at which the flow vectors on one side of the point are significantly different (in magnitude, direction, or both) from the flow vectors on the opposite side of the point. Once again a thought experiment is helpful:

> imagine holding a tiny paddlewheel at each point in the flow. If the flow would cause the paddlewheel to rotate, the center of the wheel marks a point of non-zero curl. The direction of the curl is along the axis of the paddlewheel.

By convention, the positive-curl direction is determined by the right-hand rule: if you curl the fingers of your right hand along the circulation direction, your thumb points in the direction of positive curl.

Mathematically, let $\mathbf{x} \in \mathcal{E}^3$ be the position vector, i the unit pseudoscalar and $\mathbf{u} = \mathbf{u}(\mathbf{x})$ be a vector field. Let ∇ be the geometric derivative. Then, define the *curl of* $\mathbf{u}$, denoted $\operatorname{curl}\mathbf{u}$, by

$$\operatorname{curl}\mathbf{u} \doteq -i\langle \nabla\,\mathbf{u}\rangle_2 = -i(\nabla \wedge \mathbf{u}) = \nabla \times \mathbf{u}\,.$$

That is, in $\mathcal{E}^3$, the curl is the dual of the exterior derivative. However, unlike the exterior derivative, which exists in n-dimensions, the curl is *only* defined in 3-D.

In rectangular coordinates, the exterior derivative of $\mathbf{u}$ is

$$\nabla \wedge \mathbf{u} = \gamma^k \partial_k \wedge \gamma_l u^l = \gamma^k \wedge \gamma_l \partial_k u^l\,,$$

$\gamma^k = \gamma_k$ and $\gamma_k \wedge \gamma_l = \gamma_k \gamma_l$. So, expand both indices, to find

$$\nabla \wedge \mathbf{u} = \gamma_2\gamma_3(\partial_2 u^3 - \partial_3 u^2) + \gamma_3\gamma_1(\partial_3 u^1 - \partial_1 u^3) + \gamma_1\gamma_2(\partial_1 u^2 - \partial_2 u^1)\,.$$

Since $\nabla \times \mathbf{u} = -i(\nabla \wedge \mathbf{u})$ and $i\gamma_1 = \gamma_2\gamma_3$, $i\gamma_2 = \gamma_3\gamma_1$ and $i\gamma_3 = \gamma_1\gamma_2$, then

$$\nabla \times \mathbf{u} = \gamma_1(\partial_2 u^3 - \partial_3 u^2) + \gamma_2(\partial_3 u^1 - \partial_1 u^3) + \gamma_3(\partial_1 u^2 - \partial_2 u^1)\,.$$

In alternate notation, $\nabla \times \mathbf{u}$ becomes

$$\operatorname{curl}\mathbf{u} = \nabla \times \mathbf{u} = \hat{\imath}\left(\frac{\partial u_z}{\partial y} - \frac{\partial u_y}{\partial z}\right) + \hat{\jmath}\left(\frac{\partial u_x}{\partial z} - \frac{\partial u_z}{\partial x}\right) + \hat{k}\left(\frac{\partial u_y}{\partial x} - \frac{\partial u_x}{\partial y}\right)\,.$$

Notice that each component of the curl of $\mathbf{u}$ indicates the tendency of the field to rotate in one of the coordinate planes. If the curl of the field has a large x-component, it means that the field has significant circulation about that point in the yz plane. The overall direction of the curl represents the axis about which the rotation is greatest, with the sense of the rotation given by the right-hand rule.

4.6. **Canonical Decomposition of the Geometric Derivative.** Let $\mathbf{x} \in \mathcal{E}^3$ be the position vector, i the unit pseudoscalar and $\mathbf{u} = \mathbf{u}(\mathbf{x})$ be a vector field. Then, write the *geometric derivative of* $\mathbf{u}$, denoted $\nabla\,\mathbf{u}$, by

$$\nabla\,\mathbf{u} = \nabla \cdot \mathbf{u} + \nabla \wedge \mathbf{u}\,.$$

That is, the geometric derivative is the addition of two terms: a symmetric scalar term – the *interior derivative* or *divergence*, and an anti-symmetric bivector term – the *exterior derivative*

In $\mathcal{E}^3$, the geometric derivative becomes

$$\nabla\,\mathbf{u} = \nabla \cdot \mathbf{u} + \nabla \wedge \mathbf{u} = \nabla \cdot \mathbf{u} + i(\nabla \times \mathbf{u})\,,$$

where the exterior derivative becomes the dual of the *curl*.

4.7. **Laplacian.** Once you know that the gradient operates on a scalar function and produces a vector and that the divergence operates on a vector and produces a scalar, it's natural to wonder whether these two operations can be combined in a meaningful way. As it turns out, the divergence of the gradient of a scalar function ϕ, written as $\nabla \cdot \nabla\,\phi$, is one of the most useful mathematical operations in physics and engineering. This operation, usually written as $\triangle\,\phi$, is called the *Laplacian of* ϕ in honor of Pierre-Simon Laplace.

Before trying to understand why the Laplacian operator is so valuable, you should begin by recalling the operations of gradient and divergence in Cartesian coordinates in $\mathcal{E}^3$:

- ⋄ gradient of a scalar function $\phi = \phi(\mathbf{x})$: $\nabla\,\phi = \gamma^k \partial_k \phi = \gamma^k \phi_{\,,k}$.
- ⋄ divergence of a vector function $\mathbf{A} = \mathbf{A}(\mathbf{x})$: $\operatorname{div}\mathbf{A} = \nabla \cdot \mathbf{A} = \partial_k A^k$.

Since the x-component of the gradient of $\mathbf{A}$ is $\phi_{,k}$, the y-component of the gradient of $\mathbf{A}$ is $\phi_{,y}$ and the z-component of the gradient of $\mathbf{A}$ is $\phi_{,z}$, the divergence of the vector produced by the gradient is

$$\triangle\,\phi = \operatorname{div}(\operatorname{grad}\phi) = \partial^k\partial_k\phi = \phi_{,xx} + \phi_{,yy} + \phi_{,zz} = \frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2}\,.$$

Motivation. Consider the position vector $\mathbf{x} \in \mathcal{E}^n$ and a scalar field $\phi = \phi(\mathbf{x})$. Find the gradient of ϕ: $\operatorname{grad}\phi = \nabla\,\phi = \gamma^k\partial_k\phi$. Since $\operatorname{grad}\phi$ results in a vector, apply then the divergence operator to it to find

$$\operatorname{div}(\operatorname{grad}\phi) = \nabla \cdot \nabla\,\phi = \gamma^l\partial_l \cdot \gamma^k\partial_k\phi = \partial_l\gamma^l \cdot \gamma^k\partial_k\phi\,.$$

Use the definition of the inner product to have: $\gamma^l \cdot \gamma^k = g^{lk}$. Thus, write the Laplacian as

$$\operatorname{div}(\operatorname{grad}\phi) = \partial_l g^{lk}\partial_k\phi = \partial^k\partial_k\phi = \partial^l\partial_l\phi\,.$$

This motivates the definition of the Laplacian. □

Just as the gradient, divergence, and curl represent differential operators, so too the Laplacian is an operator waiting to be fed a function. As you may recall, the gradient operator tells you the direction of greatest increase of the function (and how steep the increase is), the divergence tells you how strongly a vector function "flows" away from a point (or toward that point if the divergence is negative), and the curl tells you how strongly a vector function tends to circulate around a point. So what does the Laplacian, the divergence of the gradient, tell you?

If you write the Laplacian operator as $\triangle\,\phi = \partial^2\phi/\partial x^2 + \partial^2\phi/\partial y^2 + \partial^2\phi/\partial z^2$, it should help you see that this operator finds the *change in the change* of the function (if you make a graph, the change in the slope) in all directions from the point of interest. That may not seem very interesting, until you consider that acceleration is the change in the change of position with time, or that the maxima and minima of functions (peaks and valleys) are regions in which the slope changes significantly, or that one way to find blobs and edges in a digital image is to look for points at which the gradient of the brightness suddenly changes.

To understand why the Laplacian performs such a diverse set of useful tasks, it helps to understand that

> at each point in space, the Laplacian of a function represents the difference between the value of the function at that point and the average of the values at surrounding points.

And how does the difference between a function's value at a point and the average value at neighboring points relate to the divergence of the gradient of that function? To understand that, think about a point at which the function's value is greater than the surrounding average – such a point represents a local maximum of the function. Likewise, a point at which the function's value is less than the surrounding average represents a local minimum. This is the reason you may find the Laplacian described as a "concavity detector" or a "peak finder" – *the Laplacian finds points at which the value of the function sticks above or falls below the values at the surrounding points.*

To better understand how peaks and valleys relate to the divergence of the gradient of a function, recall that the gradient points in direction of steepest incline (or decline if the gradient is negative), and divergence measures the "flow" of a vector field out of a region (or into the region if the divergence is negative). Now consider the peak of the function shown in Figure 2.19(a) and the gradient of the function in the vicinity of that peak, shown in Figure 2.19(b). Near the peak, the gradient vectors "flow" toward the peak from all directions. Vector fields that converge upon a point have negative divergence, so this means that the divergence of the gradient in the vicinity of a peak will be a large negative number. This is consistent with the conclusion that the Laplacian is negative near a function's maximum point.

The alternative case is shown in Figures 2.20(a) and 2.20(b). Near the bottom of a valley, the gradient "flows" outward in all directions, so the divergence of the gradient is a large positive number in this case (again consistent with the conclusion that the Laplacian of a minimum point is positive). And what is the value of the Laplacian of a function away from a peak or valley? The answer to that question depends on the shape of the function in the vicinity of the point in question. As described in Section 2.9, the value of the divergence depends on how strongly the function "flows" away from a small volume surrounding the point of interest. Since the Laplacian involves the divergence of the gradient, the question is whether the gradient vectors "flow" toward or away from the point (in other words, whether the gradient vectors tend to concentrate toward or disperse away from that point). If the inward flow of gradient vectors equals the outward flow, then the Laplacian of the function is zero at that point. But if the length and direction of the gradient vectors conspire to make the outward flow greater than the inward flow at some point, then the Laplacian is positive at that point.

For example, if you're climbing out of a circularly symmetric valley with constant slope, the gradient vectors are spreading apart without changing in length, which means the divergence of the gradient (and hence the Laplacian) will have a positive value at that point. But if a different valley has walls for which the slope gets less steep (so the gradient vectors get shorter) as you move away from the bottom of the valley, it's possible for the reduced strength of the gradient vectors to exactly compensate for the spreading apart of those vectors, in which case the Laplacian will be zero.

To see how this works mathematically, consider a three-dimensional scalar function $\phi = \phi(\mathbf{r})$ whose value decreases in inverse proportion to the distance r from the origin. This function may be written as $\phi = k/r$, where k is just a constant of proportionality and r is the distance from the origin. Thus $r = (x^2 + y^2 + z^2)^{1/2}$ and $\phi = k/(x^2 + y^2 + z^2)^{1/2}$. You can find the value of the Laplacian for this case using its definition; the first step is to find the partial derivative of ϕ with respect to x

$$\phi_{,x} = \frac{\partial \phi}{\partial x} = \frac{-kx}{(x^2 + y^2 + z^2)^{3/2}}\,,$$

after which you take another partial with respect to x:

$$\phi_{,xx} = \frac{\partial^2 \phi}{\partial x^2} = \frac{-k}{(x^2 + y^2 + z^2)^{3/2}} + \frac{3kx^2}{(x^2 + y^2 + z^2)^{5/2}}\,.$$

The same approach for the second-order partials with respect to y and z gives

$$\phi_{,yy} = \frac{\partial^2 \phi}{\partial y^2} = \frac{-k}{(x^2 + y^2 + z^2)^{3/2}} + \frac{3ky^2}{(x^2 + y^2 + z^2)^{5/2}}$$

and

$$\phi_{,zz} = \frac{\partial^2 \phi}{\partial z^2} = \frac{-k}{(x^2 + y^2 + z^2)^{3/2}} + \frac{3kz^2}{(x^2 + y^2 + z^2)^{5/2}}\,.$$

Now it's just a matter of adding all three second-order partials:

$$\triangle\, \phi = \phi_{,xx} + \phi_{,yy} + \phi_{,zz} = 0\,.$$

So for a three-dimensional function with $1/r$-dependence, the Laplacian of the function is zero everywhere away from the origin. What about at the origin itself? That point requires special treatment, since the $1/r$-dependence of the function becomes problematic at $r = 0$. That special treatment involves the Dirac delta function and integral rather than differential techniques.

4.8. **Directional Derivative.** In mathematics, the directional derivative of a multivariate differentiable function along a given vector $\mathbf{v}$ at a given point $\mathbf{x}$ intuitively represents the instantaneous rate of change of the function, moving through $\mathbf{x}$ with a velocity specified by $\mathbf{v}$. It therefore generalizes the notion of a partial derivative, in which the rate of change is taken along one of the coordinate curves, all other coordinates being constant.

Definition 4.1 (Directional Derivative)**.** *Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. Then, the directional derivative of a scalar function $f(\mathbf{x})$ along a vector $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ is the function defined by the limit*

$$\nabla_{\mathbf{v}}\, f(\mathbf{x}) = \lim_{h \to 0} \frac{f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x})}{h}\,.$$

If the function f is differentiable at $\mathbf{x}$, then the directional derivative exists along any vector $\mathbf{v}$, and one has

$$\nabla_{\mathbf{v}}\, f(\mathbf{x}) = \nabla\, f(\mathbf{x}) \cdot \mathbf{v}\,,$$

where the ∇ on the right denotes the gradient and $\cdot$ is the inner (dot) product. At any point $\mathbf{x}$, the directional derivative of f intuitively represents the rate of change in f moving at a rate and direction given by $\mathbf{v}$ at the point $\mathbf{x}$.

Directional derivatives can be also denoted by:

$$\nabla_{\mathbf{v}}\, f(\mathbf{x}) \sim \frac{\partial f(\mathbf{x})}{\partial v} \sim f'_{\mathbf{v}}(\mathbf{x}) \sim D_{\mathbf{v}} f(\mathbf{x}) \sim \mathbf{v} \cdot \nabla\, f(\mathbf{x})\ .$$

Example 4.2. *Let $f(x, y) = x^2 y$. a) Find $\nabla\, f(3, 2)$. b) Find the derivative of f in the direction of $(1, 2)$ at the point $(3, 2)$.*

Solution. a) The gradient is just the vector of partial derivatives. The partial derivatives of f at the point $(x, y) = (3, 2)$ are:

$$f_x(x, y) = 2xy \implies f_x(3, 2) = 12\,,$$
$$f_y(x, y) = x^2 \implies f_x(3, 2) = 9\,.$$

Therefore, the gradient is

$$\nabla f(x, y) = \left(2xy, x^2\right) \implies \nabla f(3, 2) = (12, 9)\ .$$

b) Let $\mathbf{u} = (u_1, u_2)$ be a unit vector. The directional derivative at $(3, 2)$ in the direction of $\mathbf{u}$ is

$$\begin{aligned}\nabla_{\mathbf{u}} f(\mathbf{x}) = \nabla_{\mathbf{u}} f(3, 2) &= \nabla f\,(3, 2) \cdot \mathbf{u}\,,\\ &= (12, 9) \cdot (u_1, u_2)\ ,\\ &= 12u_1 + 9u_2\,.\end{aligned}$$

To find the directional derivative in the direction of the vector $(1, 2)$, we need to find a unit vector in the direction of the vector $(1, 2)$. We simply divide by the magnitude of $(1, 2)$.

$$\mathbf{u} = \frac{\mathbf{u}}{|\mathbf{u}|} = \frac{(1, 2)}{\sqrt{5}} = \left(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right)\ . \tag{4.1}$$

Plugging this expression for $\mathbf{u} = (u_1, u_2)$ into the equation for the directional derivative, we find that the directional derivative at the point $(3, 2)$ in the direction of $(1, 2)$ is

$$\nabla_{\mathbf{u}} f(\mathbf{x}) = 12u_1 + 9u_2 = \frac{12}{\sqrt{5}} + \frac{9}{\sqrt{5}} = \frac{30}{\sqrt{5}}\,,$$

which was required. □

Example 4.3. *Let $f(x, y, z) = xye^{x^2+z^2-5}$. Calculate the gradient of f at the point $(1, 3, -2)$ and calculate the directional derivative $\nabla_{\mathbf{u}} f(\mathbf{x})$ at the point $(1, 3, -2)$ in the direction of the vector $v = (3, -1, 4)$.*

Solution. To calculate the gradient of f at the point $(1, 3, -2)$ we just need to calculate the three partial derivatives of f

$$\nabla f(x, y, z) = (f_x, f_y, f_z) = \left((y + 2x^2y)e^{x^2+z^2-5}, xe^{x^2+z^2-5}, 2xyze^{x^2+z^2-5}\right) = e^{x^2+z^2-5}\left((y + 2x^2y), x, 2xyz\right)\ .$$

Then, in ∇f we replace the coordinates of the point $(1, 3, -2)$

$$\nabla f(1, 3, -2) = e^{3^2+(-2)^2-5}\left((3 + 2(1)^2(3)), 1, 2(1)(3)(-2)\right) = (9, 1, -12)\ .$$

The directional derivative is $\nabla_{\mathbf{u}} f(x, y, z) = \nabla f(x, y, z) \cdot \mathbf{u}$ where $\mathbf{u}$ is a unit vector. To calculate $\mathbf{u}$ in the direction of $\mathbf{v}$, we just need to divide $\mathbf{v}$ by its magnitude, $|\mathbf{v}|$. Since $|\mathbf{v}| = \sqrt{3^2 + (-1)^2 + 4^2} = \sqrt{26}$,

$$\mathbf{u} = \frac{\mathbf{v}}{|\mathbf{v}|} = \left(\frac{3}{\sqrt{26}}, \frac{-1}{\sqrt{26}}, \frac{4}{\sqrt{26}}\right)$$

and

$$\nabla_{\mathbf{u}} f(1, 3, -2) = (9, 1, -12) \cdot \left(\frac{3}{\sqrt{26}}, \frac{-1}{\sqrt{26}}, \frac{4}{\sqrt{26}}\right) = \frac{-22}{\sqrt{26}}\,,$$

which yields the desired result. □

4.9. **Fourier's Law.** Position vector: $\mathbf{x} \in \mathcal{E}^3$. $q = q(\mathbf{x})$, $T = T(\mathbf{x})$, $\kappa \in \mathcal{R}$ and the index k running from 1 to 3.

- ◇ Vectorial form: $\mathbf{q} = -\kappa \nabla T$.
- ◇ Index notation with Einstein summation convention and basis vectors: $\gamma_k q^k = -\kappa\gamma_k\partial_k T$.
- ◇ Index notation without Einstein summation convention and basis vectors: $q^k = -\kappa\partial_k T$.
- ◇ Index notation, comma derivative notation without basis vectors: $q^k = -\kappa T^{,k}$.
- ◇ The one I like the most: Index notation, (upper) comma derivative notation, Einstein summation convention, basis vectors: $\gamma_k q^k = -\kappa\gamma_k T^{,k}$.

Discussion. Note the difference between index notation with and without basis vectors and Einstein summation convention. Index notation implies different equations. For instance, consider Fourier's law in vectorial notation: $\mathbf{q} = -\kappa \bigtriangledown T$. Using index notation and comma derivative notation, it becomes $q^k = -\kappa T^{,k}$ which, in turn, implies *three separate* equations:

$$q^1 = -\kappa T^{,1}\,, \qquad q^2 = -\kappa T^{,2} \qquad \text{and} \qquad q^3 = -\kappa T^{,3}\,,$$

One the other hand, Fourier's law written with index notation, comma derivative notation, Einstein summation convention and basis vectors becomes $\gamma_k q^k = -\kappa\gamma_k T^{,k}$ which, in turn, implies *one* equation:

$$\gamma_1 q^1 + \gamma_2 q^2 + \gamma_3 q^3 = -\kappa(\gamma_1 T^{,1} + \gamma_2 T^{,2} + \gamma_3 T^{,3})\,.$$

4.10. **Notation for Derivatives.** Other notation than the classical one is used in order to distinguish the derivative operator from the d variable.

(1) Classical notation: $df/dt = \partial f/\partial x\, dx/dt$.

(2) Other notation: $\mathrm{d}f/\mathrm{d}t = \partial f/\partial x\, \mathrm{d}x/\mathrm{d}t$.

I think the classical notation look better, but there's a lil risk of confusion, for example, what if d represents the diameter of a circle, then its derivative with respect to t would look like dd/dt... weird!!! Better to use another symbol for the diameter (D, thus dD/dt) of another or another symbol for the derivative (D, thus Dd/Dt)... The latter looks weird, better to use another symbol for the diameter ;).

4.11. **Polar Coordinates.** Typing some things: polar coordinates frame: Consider orthonormal frame $\{\gamma_r, \gamma_\theta\}$. Then, expand the position vector r in the frame:

$$r = \gamma_k r^k = \gamma_r r^r + \gamma_\theta r^\theta\,.$$